NMR studies of the structure of a conserved RNA motif of 23S ribosomal RNA and its interaction with peptidyl transferase antibiotics

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences

2010

John Paul King

School of Chemistry
# Table of Contents

I  
List of figures  

II  
List of tables  

III  
List of abbreviations  

IV  
List of symbols  

V  
Abstract  

VI  
Declaration  

VII  
Copyright statement  

VIII  
Acknowledgements  

**Chapter 1  Introduction**  

1.1  
Anti-biotic Resistance  
1.1.1  
The importance of resistance  
1.1.2  
The potential of *peptidyl transferase* antibiotics  

1.2  
Physical and chemical properties of RNA  

1.3  
The ribosome – mechanism of protein synthesis  
1.3.1  
Introduction to the ribosome (archaea, prokaryotes, eukaryotes)  
1.3.2  
Structure of the ribosome  
1.3.3  
Function of the ribosome  
1.3.4  
Conserved RNA motifs  

1.4  
*Peptidyl transferase* antibiotics  
1.4.1  
Amicetin  
1.4.2  
Blasticidin S  
1.4.3  
Gougerotin  
1.4.4  
Other aminohexose cytosine nucleoside antibiotics  

1.5  
Ribosome antibiotic complexes  
1.5.1  
Amicetin  
1.5.2  
Blasticidin S  
1.5.3  
Gougerotin  

1.6  
The predicted amicetin binding site motif  
1.6.1  
Comparison of the X-ray crystal structures for the motif  

1.7  
Principles of NMR  
1.7.1  
Basic principles  
1.7.2  
Longitudinal and transverse relaxation  
1.7.3  
Nuclear Overhauser effect (NOE)  
1.7.4  
One dimensional NMR  
1.7.5  
Multi dimensional NMR  

1.8  
Molecular modelling  
1.8.1  
*Ab initio* calculations  
1.8.2  
Molecular mechanics  
1.8.3  
Energy minimisation  
1.8.4  
Molecular dynamics  
1.8.5  
Simulated annealing
Chapter 2  Materials and methods

2.1 Preparation of antibiotic and RNA samples for NMR
2.1.1 Antibiotics
2.1.2 Antibiotic Sample preparation
2.1.3 RNA
2.1.4 RNA sample preparation
2.1.5 RNA quantification
2.1.6 Titration of the *E. coli* 29-mer RNA motif with amicetin

2.2 NMR instrumentation and experiments
2.2.1 NMR spectrometers
2.2.2 NMR acquisition parameters
2.2.3 NMR experiments
2.2.4 NMR processing and analysis

2.3 NMR techniques
2.3.1 1D homonuclear and heteronuclear experiments with decoupling
2.3.2 Longitudinal, $T_1$, relaxation measurement
2.3.3 Double quantum filtered correlation spectroscopy (DQF-COSY)
2.3.4 Total correlation spectroscopy (TOCSY)
2.3.5 Nuclear Overhauser effect spectroscopy (NOESY)
2.3.6 Rotation-frame Overhauser effect spectroscopy (ROESY)
2.3.7 Heteronuclear single quantum correlation (HSQC)
2.3.8 2D and 3D HSQC-TOCSY
2.3.9 2D and 3D HSQC-NOESY and 3D NOESY-HSQC
2.3.10 $^1$H-$^{31}$P CPMG-HSQC-NOESY
2.3.11 HCP
2.3.12 NMR titration
2.3.13 Variable temperature series

2.4 Assignment strategies
2.4.1 Antibiotics
2.4.2 Isotopically normal RNA
2.4.3 Uniformly $^{13}$C and $^{15}$N labelled RNA
2.4.4 RNA-antibiotic complexes

2.5 Constraint generation
2.5.1 Antibiotic ROE distance constraints
2.5.2 Antibiotic dihedral angle constraints
2.5.3 RNA NOE distance constraints
2.5.4 RNA dihedral angle constraints
2.5.5 RNA hydrogen bond constraints

2.6 Structure calculation protocol
2.6.1 Antibiotics
2.6.2 RNA
Chapter 3  NMR structure determination of *peptidyl transferase* antibiotics

3.1  Structure determination of blasticidin S  
3.1.1  NMR assignment of blasticidin S  
3.1.2  Assignment table of blasticidin S  
3.1.3  Effect of temperature on the NMR spectrum of blasticidin S  
3.1.4  Longitudinal relaxation time, $T_1$, measurement of blasticidin S  
3.1.5  NMR constraints of blasticidin S  
3.1.6  NMR solution structure of blasticidin S  

3.2  Structure determination of gougerotin  
3.2.1  NMR assignment of gougerotin  
3.2.2  Assignment table of gougerotin  
3.2.3  Effect of temperature on the NMR spectrum of gougerotin  
3.2.4  NMR constraints of gougerotin  
3.2.5  NMR solution structure of gougerotin  

3.3  Structure determination of amicetin  
3.3.1  NMR assignment of amicetin  
3.3.2  Assignment table of amicetin  
3.3.3  NMR constraints of amicetin  
3.3.4  NMR solution structure of amicetin  

3.4  Discussion and comparison of the structure and dynamics of blasticidin S, gougerotin, and amicetin

Chapter 4  NMR structure determination of isotopically normal and labelled 23S rRNA motifs

4.1  Assignment and structure determination of isotopically normal *E. coli* 29-mer RNA  
4.1.1  Assignment of the exchangeable proton resonances  
4.1.2  Assignment of the non-exchangeable proton and carbon resonances  
4.1.3  Assignment of the phosphorus resonances  
4.1.4  Assignment table of the *E. coli* 29-mer RNA motif  
4.1.5  NMR geometrical constraints of the *E. coli* 29-mer RNA motif  
4.1.6  NMR solution structure of the *E. coli* 29-mer RNA motif  

4.2  Assignment and structure determination of isotopically normal *H. h.* 29-mer RNA  
4.2.1  Assignment of the exchangeable proton resonances  
4.2.2  Assignment of the non-exchangeable proton and carbon resonances  
4.2.3  Assignment of the phosphorus resonances  
4.2.4  Assignment table of the *H. h.* 29-mer RNA motif  
4.2.5  NMR geometrical constraints of the *H. h.* 29-mer RNA motif  
4.2.6  NMR solution structure of the *H. h.* 29-mer RNA motif
Assignment and structure determination of uniformly $^{13}$C and $^{15}$N labelled H. h. 37-mer RNA

4.3.1 Use of the world's first 1 GHz NMR spectrometer
4.3.2 Assignment of the exchangeable proton and $^{15}$N resonances
4.3.3 Assignment of the non-exchangeable proton and carbon resonances
4.3.4 Assignment of the phosphorus resonances
4.3.5 Assignment table of the H. h. 37-mer RNA
4.3.6 NMR geometrical constraints of the H. h. 37-mer RNA
4.3.7 NMR solution structure of the H. h. 37-mer RNA motif

4.4 Discussion and comparison of the NMR structures of the E. coli and H. h. RNA motifs
4.4.1 Comparison of the NMR assignment
4.4.2 Comparison of the solution structures

Chapter 5 Interactions of isotopically normal 23S rRNA motifs with amicetin
5.1 Interactions of normal H. h. 29-mer rRNA-amicetin complex
5.2 Interactions of normal E. coli 29-mer rRNA-amicetin complex
5.3 Discussion of the RNA-antibiotic complexes of the E. coli and H. h. RNA motifs

Chapter 6 Conclusion and further work

Presentations and papers

References

Appendices
A Experimental scripts
A.1 Typical NMRPipe processing scripts
A.2 Antibiotic structure calculation scripts and files
A.3 RNA structure calculation scripts
B Experimental lists
B1 E. coli titration NMR experiments
C Additional experimental results
C.1 Additional helical parameters and dihedral angles for the E. coli 29-mer solution structure
C.2 Additional helical parameters and dihedral angles for the H. h. 29-mer solution structure
C.3 Additional helical parameters and dihedral angles for the H. h. 37-mer solution structure

Word count including appendices = 83,055

Word count excluding appendices = 65,334
List of figures

Figure 1.1.1: This chart illustrates the number of death certificates citing S. aureus infection per year in England and Wales. 24

Figure 1.2.1: An example of a RNA monophosphate nucleotide. 28

Figure 1.2.2: The two canonical base pairs of RNA (A-U and G-C). 28

Figure 1.2.3: A fully W-C base paired RNA duplex. 29

Figure 1.2.4: Images showing examples of RNA primary, secondary and tertiary structure. 30

Figure 1.2.5: Schematic representation of four major classes of RNA secondary structures. 30

Figure 1.2.6: Figure illustrating the location of the major and minor grooves on the tertiary structure of an A-form RNA helix. 31

Figure 1.2.7: Images showing the phosphate backbone dihedral angles, the glycosidic dihedral angles and the ribose ring dihedral angles. 32

Figure 1.2.8: A schematic representation of the common ribose ring pucker conformations. 32

Figure 1.2.9: The figure shows how the local helical axis is defined for the calculation of local helical parameters. 33

Figure 1.2.10: This figure shows the coordinate frame of reference used to describe helical parameters of the RNA. 33

Figure 1.2.11: The four base pair helical parameters. 34

Figure 1.2.12: The five base pair parameters. 34

Figure 1.2.13: The six base pair step parameters. 34

Figure 1.3.1: Two representations of the bacterial ribosome in two different orientations. 37

Figure 1.3.2: The 50S E. coli ribosomal subunit and 70S E. coli ribosome (PDB ID:2AW4). 37

Figure 1.3.3: The secondary structure of the PTC of E. coli, with the predicted amicetin binding motif labelled. 38

Figure 1.3.4: The PTC from the E. coli ribosome X-ray crystal structure (taken from PDB ID: 2AW4). 39

Figure 1.3.5: This figure shows the proposed process of translation of mRNA to protein. 40

Figure 1.3.6: Two schematic illustrations of tRNA, one interacting with the A-loop and one interacting with the P-loop. 42

Figure 1.3.7: The proposed mechanism for peptide bond formation in the ribosome. 42

Figure 1.3.8: The proposed mechanism for polypeptide release from the P-site tRNA. 43

Figure 1.3.9: The figure shows the extent of conservation of the PTC of the E.coli sequence compared to 592 sequences taken from a range of archaea, prokaryotes and eukaryotes. 44

Figure 1.4.1: The chemical structure of the peptidyl transferase antibiotic amicetin. 46

Figure 1.4.2: The chemical structure of the peptidyl transferase antibiotic blasticidin S. 47

Figure 1.4.3: The chemical structure of the peptidyl transferase antibiotic gougerotin. 49

Figure 1.4.4: The chemical structure of an analogue of the peptidyl transferase antibiotic gougerotin. 50

Figure 1.4.5: A partial gougerotin structure for which an X-ray crystal structure has been produced. 51

Figure 1.4.6: The chemical structure of the peptidyl transferase antibiotics, bamicetin and oxamicetin. 52

Figure 1.4.7: The chemical structure of the peptidyl transferase antibiotic plicacetin. 52

Figure 1.4.8: The chemical structure of cytimidine. 52

Figure 1.4.9: The chemical structure of cytosamine triacetate. 52

Figure 1.5.1: The H. marismortui peptidyl transferase centre, showing the two binding sites of blasticidin S in relation to the A- and P-loops (PDB ID: 1KC8). 55
Figure 1.6.1: The secondary structures of the amicetin binding site in the ribosomes of *E. coli* (2AW4), *D. radiodurans* (1NKW) and *H. marismortui* (3CC2).

Figure 1.6.2: The X-ray crystal structure of the amicetin binding site in the ribosomes of *E. coli* (2AW4), *D. radiodurans* (1NKW) and *H. marismortui* (3CC2).

Figure 1.6.3: A comparison of the amicetin binding site of the *H. marismortui* (3CC2) unbound crystal structure and *H. marismortui* (1KC8) blasticidin S bound crystal structure.

Figure 1.7.1: An image of the Karplus curve.

Figure 1.7.2: The figure illustrates longitudinal relaxation.

Figure 1.7.3: The figure illustrates transverse relaxation.

Figure 1.7.4: The figure illustrates the NOE effect.

Figure 1.7.5: A generic 1D pulse sequence.

Figure 1.7.6: A generic 2D pulse sequence.

Figure 1.8.1: The figure illustrates the process of simulated annealing.

Figure 1.9.1: The secondary structure and NMR structure of the uniformly $^{13}$C and $^{15}$N labelled 14-mer RNA motif (2KOC).

Figure 2.3.1: A standard homonuclear (proton) 1D pulse sequence.

Figure 2.3.2: A decoupled heteronuclear (X nuclei) 1D pulse sequence.

Figure 2.3.3: The $T_1$ measurement pulse sequence.

Figure 2.3.4: The DQF-COSY pulse sequence.

Figure 2.3.5: The figure illustrates the problem of linewidths which are larger than or equal to the coupling constant in the DQF COSY experiment.

Figure 2.3.6: The TOCSY/ROESY pulse sequence.

Figure 2.3.7: The NOESY pulse sequence.

Figure 2.3.8: The HSQC pulse sequence.

Figure 2.3.9: A basic 3D HSQC-TOCSY pulse sequence.

Figure 2.3.10: A basic 3D HSQC-NOESY pulse sequence.

Figure 2.4.1: The figure shows a brief overview of the tasks which were taken to carry out the NMR assignment of the antibiotics and the NMR spectra used to accomplish these tasks.

Figure 2.4.2: The figure shows a brief overview of the tasks which were taken to carry out the NMR assignment of the isotopically normal RNA motifs and the NMR spectra used to accomplish these tasks.

Figure 2.4.3: The figure shows the H1 to aromatic sequential and intra connectivity distances.

Figure 2.4.4: The figure shows a series of sequential and intra connectivity distances.

Figure 2.4.5: The chart shows a brief overview of the tasks which were taken to carry out the NMR assignment of the $^{13}$C and $^{15}$N labelled RNA motif and the NMR spectra used to accomplish these tasks.

Figure 3.1.1: The chemical structure of the blasticidin S antibiotic.

Figure 3.1.2: The image shows a 500 MHz 1D $^1$H spectrum of blasticidin S measured at pH 4.8, and a 500 MHz 1D $^1$H spectrum of blasticidin S measured at a pH of 7.3.

Figure 3.1.3: 500 MHz $^1$H-$^1$H correlated TOCSY spectrum of blasticidin S.

Figure 3.1.4: 500 MHz $^1$H-$^1$H correlated ROESY spectrum of blasticidin S.

Figure 3.1.5: 800 MHz $^1$H-NMR variable temperature series (2ºC to 50ºC) of blasticidin S.

Figure 3.1.6: Measurement of the longitudinal relaxation time constant $T_1$ of blasticidin S.

Figure 3.1.7: The figure shows the average NMR structure of blasticidin S and an overlay of the 10 best structures, with an all atom average RMSD of 0.56 Å.

Figure 3.2.1: The chemical structure of the gougerotin antibiotic.
Figure 3.2.2: 600 MHz $^1$H-NMR spectrum of gougerotin.

Figure 3.2.3: 600 MHz $^1$H-$^1$H correlated TOCSY spectrum of gougerotin.

Figure 3.2.4: 600 MHz $^1$H-$^1$H correlated ROESY spectrum of gougerotin.

Figure 3.2.5: Section of the 600 MHz $^1$H-$^1$H correlated ROESY spectrum of gougerotin.

Figure 3.2.6: 600 MHz $^1$H-NMR variable temperature series (2ºC to 40ºC) of gougerotin.

Figure 3.2.7: The figure shows the average NMR structure of gougerotin. and an overlay of the 10 best structures, with an all atom average RMSD of 0.66 Å.

Figure 3.3.1: The chemical structure of the antibiotic amicetin.

Figure 3.3.2: 600 MHz $^1$H-NMR spectrum of amicetin.

Figure 3.3.3: The figure shows the average NMR structure of amicetin and an overlay of the 10 best structures, with an all atom average RMSD of 1.00 Å.

Figure 3.4.1: The figure shows the average NMR structure of blasticidin S, the average NMR structure of gougerotin and the average NMR structure of amicetin.

Figure 3.4.2: The figure shows the crystal structure of the ribosome bound blasticidin S in its major binding site, and the average NMR structure of gougerotin.

Figure 3.4.3: The scheme illustrates the major parts that the antibiotics can be broken into and the lower structure shows a hypothetical hybrid antibiotic.

Figure 4.0.1: The figure shows the $E. coli$ 29-mer, the $H. h.~$29-mer and the uniformly $^{13}$C and $^{15}$N labelled $H. h.~$37-mer RNA secondary structure.

Figure 4.1.1: 600 MHz NOESY spectrum of $E. coli$ 29-mer RNA, dissolved in 90% $^1$H$_2$O + 10% $^2$H$_2$O showing the identification of the imino to imino through space sequential connectivities.

Figure 4.1.2: 600 MHz NOESY spectrum of $E. coli$ 29-mer RNA, dissolved in 90% $^1$H$_2$O + 10% $^2$H$_2$O showing the identification of the imino to amino through space connectivities.

Figure 4.1.3: The figure shows the 600 MHz NOESY spectrum of $E. coli$ 29-mer RNA, dissolved in 90% $^1$H$_2$O + 10% $^2$H$_2$O showing the identification of the imino to aromatic through space connectivities and the 700 MHz $^1$H-$^1$H TOCSY spectrum of the same sample, under the same conditions, showing the $^J_H^c$ H5-H6 correlation due to the cytosine and uracil bases of the RNA.

Figure 4.1.4: The figure shows the 800 MHz DQF-COSY spectrum of the $E. coli$ 29-mer RNA motif, dissolved in 100% $^2$H$_2$O, the spectrum shows the identification of the aromatic cytosine and uracil protons H5 and H6. The figure also shows the 700 MHz $^1$H-$^1$C HSQC spectrum of the same sample, under the same conditions, showing the $^J_{Hc}$ H5-C5 correlation due to the cytosine and uracil bases.

Figure 4.1.5: 700 MHz $^1$H-$^1$C HSQC spectrum of the $E. coli$ 29-mer RNA motif, dissolved in 100% $^2$H$_2$O. The spectrum shows the identification of the H2 adenine protons based on the observed scalar coupled ($^J_{Hhc}$) cross peaks.

Figure 4.1.6: The figure shows the 700 MHz NOESY spectrum of the $E. coli$ 29-mer RNA motif, dissolved in 100% $^2$H$_2$O. The spectrum shows the pattern of sequential connectivities which starts at the C1 intra peak and can be followed all the way to the U14 intra peak. The figure also shows the two sections of the 700 MHz $^1$H-$^1$C HSQC spectrum of the same sample, under the same conditions, the sections show the $^J_{Hhc}$ H6/8-C6/8 correlations and the $^J_{Hhc}$ H1'-C1' correlations.

Figure 4.1.7: The figure shows the 700 MHz NOESY spectrum of the $E. coli$ 29-mer RNA motif, dissolved in 100% $^2$H$_2$O containing. This spectrum shows how the sequence specific H1' assignment was used to gain sequence specific H2' assignment through observed scalar coupled ($^J_{H}$) cross peaks.

Figure 4.1.8: 800 MHz DQF-COSY spectrum of the $E. coli$ 29-mer RNA motif, dissolved in 100% $^2$H$_2$O containing. This spectrum shows how the sequence specific H1' assignment was used to gain sequence specific H2' assignment through observed scalar coupled ($^J_{H}$) cross peaks.
**Figure 4.1.9:** The figure shows the 700 MHz NOESY spectrum of the *E. coli* 29-mer RNA motif, dissolved in 100% $^{2}$H$_{2}$O. The spectrum shows the pattern of sequential connectivities which starts at the A21 intra peak and can be followed all the way to the G35 intra peak. The figure also shows two sections of the 600 MHz $^{1}$H-$^{13}$C PGMG-HSQC-NOESY spectrum of the same sample, under the same conditions, the sections show the H6/8 to phosphorus correlations and the H1’ to phosphorus correlations.

**Figure 4.1.10:** 81 MHz $^{31}$P (200 MHz, $^{1}$H) 1D spectrum of the *E. coli* 29-mer RNA motif, dissolved in 100% $^{2}$H$_{2}$O.

**Figure 4.1.11:** The inter residue “exchangeable” NOE distance constraints used in the structure determination calculation of the *E. coli* 29-mer RNA motif.

**Figure 4.1.12:** The inter residue “non-exchangeable” NOE distance constraints used in the structure determination calculation of the *E. coli* 29-mer RNA motif.

**Figure 4.1.13:** The figure shows an overlay of the ten best structures produced by the structure determination of the *E. coli* 29-mer RNA motif, the average structure of the motif and the observed secondary structure of the motif.

**Figure 4.2.1:** 600 MHz NOESY spectrum of the *H. h.* 29-mer RNA motif, dissolved in 90% $^{1}$H$_{2}$O and 10% $^{2}$H$_{2}$O showing the identification of the imino to imino through space sequential connectivities.

**Figure 4.2.2:** The figure shows the 600 MHz DQF-COSY spectrum of the *H. h.* 29-mer RNA motif, dissolved in 100% $^{1}$H$_{2}$O. The spectrum shows the identification of the aromatic cytosine and uracil protons H5 and H6. The figure also shows the 500 MHz $^{1}$H-$^{13}$C HSQC-NOESY spectrum of the same sample, under the same conditions, showing the $^{1}J_{H,C}$ H5-C5 correlation due to the cytosine and uracil bases.

**Figure 4.2.3:** The figure shows the 500 MHz NOESY spectrum of the *H. h.* 29-mer RNA motif, dissolved in 100% $^{2}$H$_{2}$O. The spectrum shows the pattern of sequential connectivities which starts at the A21 intra peak and can be followed all the way to the G35 intra peak. The figure also shows two sections of the 500 MHz $^{1}$H-$^{13}$C HSQC-NOESY spectrum of the same sample, under the same conditions, the sections show the $^{1}J_{H,C}$ H6/8-C6/8 correlations and the $^{1}J_{H,C}$ H1’-Cl’ correlations.

**Figure 4.2.4:** The inter residue“exchangeable” NOE distance constraints used in the structure determination calculation of the *H. h.* 29-mer RNA motif.

**Figure 4.2.5:** The inter residue “non-exchangeable” NOE distance constraints used in the structure determination calculation of the *H. h.* 29-mer RNA motif.

**Figure 4.2.6:** The figure shows an overlay of the ten best structures produced by the structure determination of the *H. h.* 29-mer RNA motif, the average structure of the motif and the observed secondary structure of the motif.

**Figure 4.3.1:** Bruker poster on the subject of the new 1 GHz NMR spectrometer, including data on the uniformly $^{13}$C and $^{15}$N labelled *H. h.* 37-mer RNA.

**Figure 4.3.2:** 600 MHz $^{1}$H-$^{14}$N HSQC spectrum of the *H. h.* 37-mer RNA motif, dissolved in 90% $^{1}$H$_{2}$O + 10% $^{2}$H$_{2}$O. The spectrum shows the identification of the imino resonances.

**Figure 4.3.3:** The panels show planes of the 1 GHz 3D NOESY-HSQC spectrum of the *H. h.* 37-mer RNA motif (0.2 mM), dissolved in 100% $^{2}$H$_{2}$O. The spectrum shows the pattern of sequential connectivities which starts at the A9 intra peak and can be followed to the C29 intra peak.

**Figure 4.3.4:** The figure shows a comparison of two sets of planes from the 3D $^{1}$H-$^{13}$C HSQC-NOESY (800 MHz) and the 3D NOESY-HSQC (1 GHz).

**Figure 4.3.5:** The figure shows a number of planes of the 700 MHz 3D $^{1}$H-$^{13}$C HSQC TOCSY spectrum ($\tau_{m}=5.44$ ms) of the *H. h.* 37-mer RNA motif, dissolved in 100% $^{2}$H$_{2}$O. The spectrum shows a pattern of $^{1}J_{H,H}$ coupling correlations which starts at the H1’ of C18 and can be followed through the ribose protons to H5’ and H5” of C18.

**Figure 4.3.6:** The figure shows the 800 MHz $^{1}$H-$^{13}$C HSQC spectrum of the *H. h.* 37-mer RNA motif dissolved in 100% $^{2}$H$_{2}$O at 25°C. The spectrum shows the H5’-S” to C5’ correlations ($^{1}J_{H,C}$). The top panels show planes of the 600 MHz 3D HCP of the same sample, under the same conditions. The planes show the intra residue $^{3}$P to C5’ to H5’ correlations.

**Figure 4.3.7:** The inter residue NOE distance constraints used in the structure determination calculation of the *H. h.* 37-mer RNA motif.
**Figure 4.3.8:** The figure shows an overlay of the ten best structures produced by the structure determination of the *H. h.* 37-mer RNA motif, the average structure of the motif and the observed secondary structure of the motif.

**Figure 4.3.9:** The figure shows the overlay of residues 9 to 29 of the ten best structures produced by the structure determination of the *H. h.* 37-mer RNA motif.

**Figure 4.4.1:** The figure shows the 700 MHz 1D $^1$H spectrum of the imino region of the *E. coli* 29-mer RNA and the 600 MHz 1D $^1$H spectrum of the imino region of the *H. h.* 29-mer RNA.

**Figure 4.4.2:** The top trace is the 600 MHz 1D $^1$H spectrum of the *E. coli* 29-mer RNA (1 mM, pH 6.2 in 100% $^2$H$_2$O) at 25ºC. The middle trace is the 500 MHz 1D $^1$H spectrum of the *H. h.* 29-mer RNA (2.4 mM, pH 6.2 in 100% $^2$H$_2$O) at 25ºC. The lower trace is the 800 MHz 1D $^1$H spectrum of the *H. h.* 37-mer RNA (0.2 mM, pH 6.4 in 100% $^2$H$_2$O) at 25ºC.

**Figure 4.4.3:** The figure shows the average NMR structure of the *E. coli* 29-mer RNA, the *H. h.* 29-mer RNA motif and the $^{13}$C and $^{15}$N labelled *H. h.* 37-mer RNA.

**Figure 4.4.4:** The figure shows the observed secondary structure of the *E. coli* 29-mer, the *H. h.* 29-mer RNA motif and the $^{13}$C and $^{15}$N labelled *H. h.* 37-mer RNA.

**Figure 5.1.1:** The figure shows the observed secondary structure of the *H. h.* 29-mer, indicating base pairs and non-base paired residues whose imino proton resonances vary the greatest upon titration with amicetin.

**Figure 5.1.2:** A series of 600 MHz 1D $^1$H spectra of the *H. h.* 29-mer RNA dissolved in 90% $^1$H$_2$O and 10% $^2$H$_2$O all with various molar equivalents (eq) of amicetin.

**Figure 5.2.1:** A series of 600 MHz 1D $^1$H spectra of the *E. coli* 29-mer RNA motif dissolved in 100% $^2$H$_2$O, all with various molar equivalents of amicetin.

**Figure 5.2.2:** An overlay of the 600 MHz 2D TOCSY spectra of the *E. coli* 29-mer RNA motif with various molar equivalents of amicetin ranging from 0 eq, to 0.390 eq.

**Figure 5.2.3:** The figure shows the 81 MHz 1D $^{31}$P spectrum of the *E. coli* 29-mer RNA motif and the 1D $^{31}$P spectrum of the *E. coli* 29-mer RNA with 1.24 equivalents of amicetin and the 81 MHz 1D $^{31}$P spectrum of the *E. coli* 29-mer RNA with 3.6 equivalents of amicetin.

**Figure 5.2.4:** The figure shows the 700 MHz 1D $^1$H spectrum of the *E. coli* 29-mer RNA motif dissolved in 90% $^1$H$_2$O and 10% $^2$H$_2$O at 2ºC, and the 700 MHz 1D $^1$H spectrum of the *E. coli* 29-mer RNA with 3.6 equivalents of amicetin, under the same conditions.
### List of tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.5.1</td>
<td>The table gives the distances used to constrain standard W-C base pairing.</td>
</tr>
<tr>
<td>Table 3.1.1</td>
<td>Table showing the chemical shifts ($\delta$, ppm) of assigned $^1$H resonances and the scalar coupling constants ($J_{\text{HH}}$, Hz) of Blasticidin S.</td>
</tr>
<tr>
<td>Table 3.1.2</td>
<td>This table shows the list of NMR experimental ROE constraints used in the XPLOR-NIH structure determination protocol for blasticidin S.</td>
</tr>
<tr>
<td>Table 3.1.3</td>
<td>Table showing the list of NMR experimental dihedral constraints used in the XPLOR-NIH structure determination protocol for blasticidin S.</td>
</tr>
<tr>
<td>Table 3.1.4</td>
<td>A list of possible hydrogen bonds, observed in the NMR structure, the free crystal structure and the major bound crystal structure.</td>
</tr>
<tr>
<td>Table 3.1.5</td>
<td>A list of important dihedral angles for the NMR structure, the free crystal structure and the major bound crystal structure.</td>
</tr>
<tr>
<td>Table 3.2.1</td>
<td>Table showing the proton chemical shifts ($\delta$, ppm) and the scalar coupling constants ($J_{\text{HH}}$, Hz) of gougerotin.</td>
</tr>
<tr>
<td>Table 3.2.2</td>
<td>This table shows the list of NMR experimental ROE constraints used in the XPLOR-NIH structure determination protocol for gougerotin.</td>
</tr>
<tr>
<td>Table 3.2.3</td>
<td>Table showing the list of NMR experimental dihedral constraints used in the XPLOR-NIH structure determination protocol for gougerotin.</td>
</tr>
<tr>
<td>Table 3.2.4</td>
<td>A list of possible hydrogen bonds observed in the NMR structure and in both confirmations observed in the crystal structure of the nucleoside fragment of gougerotin (structures A and B).</td>
</tr>
<tr>
<td>Table 3.2.5</td>
<td>A list of important dihedral angles for the NMR structure and in both confirmations observed in the crystal structure of the nucleoside fragment of gougerotin (structures A and B).</td>
</tr>
<tr>
<td>Table 3.3.1</td>
<td>Table showing the chemical shifts ($\delta$, ppm) of assigned $^1$H and $^{13}$C resonances and proton scalar coupling constants ($J$) of amicetin.</td>
</tr>
<tr>
<td>Table 3.3.2</td>
<td>This table shows the list of NMR experimental ROE/NOE constraints used in the XPLOR-NIH structure determination protocol for amicetin.</td>
</tr>
<tr>
<td>Table 3.3.3</td>
<td>Table showing the list of NMR experimental dihedral constraints, used in the XPLOR-NIH structure determination protocol for amicetin.</td>
</tr>
<tr>
<td>Table 3.3.4</td>
<td>A list of possible hydrogen bonds observed for the NMR structure and the free crystal structure.</td>
</tr>
<tr>
<td>Table 3.3.5</td>
<td>A list of important dihedral angles for the NMR structure and the free crystal structure.</td>
</tr>
<tr>
<td>Table 3.4.1</td>
<td>A comparison of important dihedral angles for the NMR structures of amicetin, blasticidin S and gougerotin.</td>
</tr>
<tr>
<td>Table 4.1.1</td>
<td>The $^1$H, $^{13}$C and $^{31}$P chemical shifts of the <em>E. coli</em> 29-mer RNA.</td>
</tr>
<tr>
<td>Table 4.1.2</td>
<td>The table details information on the six groups of NOEs and the parameters used in CcpNmr Analysis to produce the NOE based distance constraints.</td>
</tr>
<tr>
<td>Table 4.1.3</td>
<td>The table details how the ribose and glycosidic dihedral angles were constrained for the <em>E. coli</em> 29-mer RNA motif, and the information that these constraints were based on.</td>
</tr>
<tr>
<td>Table 4.1.4</td>
<td>The table gives the imino proton chemical shifts for the <em>E. coli</em> 29-mer RNA motif and whether they were constrained as W-C base pairs in the structure determination.</td>
</tr>
<tr>
<td>Table 4.1.5</td>
<td>The table shows the base pairs observed in the solution structure of the <em>E. coli</em> 29-mer RNA motif.</td>
</tr>
<tr>
<td>Table 4.1.6</td>
<td>The table shows the local base pair step parameters and the form of the RNA at a given step for the <em>E. coli</em> 29-mer RNA motif.</td>
</tr>
<tr>
<td>Table 4.2.1</td>
<td>The $^1$H, $^{13}$C and $^{31}$P chemical shifts of the <em>H. h.</em> 29-mer RNA.</td>
</tr>
<tr>
<td>Table 4.2.2</td>
<td>Details information on the six groups of NOEs and the parameters used in CcpNmr Analysis to produce the NOE based distance constraints for the <em>H. h.</em> 29-mer RNA motif.</td>
</tr>
</tbody>
</table>
Table 4.2.3: The table details how the ribose and glycosidic dihedral angles were constrained for the H. h. 29-mer RNA motif, and the information that these constraints were based on.

Table 4.2.4: The table gives the imino proton chemical shifts of the H. h. 29-mer RNA and whether they were constrained as W-C base pairs in the structure determination.

Table 4.2.5: The table shows the base pairs observed in the NMR structure of H. h. 29-mer RNA.

Table 4.2.6: The table shows the local base pair step parameters, and the form of the RNA at a given step.

Table 4.3.1: The table shows the $^1$H, $^{13}$C, $^{15}$N and $^{31}$P chemical shifts of the H. h. 37-mer RNA.

Table 4.3.2: The table details information on the six groups of NOEs and the parameters used in CcpNmr Analysis to produce the NOE based distance constraints for the H. h. 37-mer RNA.

Table 4.3.3: The table details how the ribose and glycosidic dihedral angles were constrained for the H. h. 37-mer RNA motif, and the information that these constraints were based on.

Table 4.3.4: The table gives the imino proton chemical shift of the H. h. 37-mer RNA and whether they were constrained as W-C base pairs in the structure determination.

Table 4.3.5: The table shows the base pairs observed in the NMR structure of H. h. 37-mer RNA.

Table 4.3.6: The table shows the local base pair step parameters, and the form of the RNA at a given step.

Table 4.4.1: The table shows the difference in exchangeable proton chemical shifts between the H. h. 29-mer, the H. h. 37-mer and the E. coli 29-mer.

Table 4.4.2: The table shows the difference in non-exchangeable proton chemical shifts between the H. h. 29-mer, the H. h. 37-mer and the E. coli 29-mer.

Table 4.4.3: The table shows the difference carbon and phosphorus chemical shifts between the H. h. 29-mer, the H. h. 37-mer and the E. coli 29-mer.

Table 5.1.1: The table showing some of the chemical shifts of assigned $^1$H resonances of amicetin, and the complexed amicetin (with H. h. 29-mer RNA). The nuclei that are shown are potentially important in the binding process.

Table B.1.1: The table shows the NMR spectra carried out during the E. coli 29-mer – amicetin titration.

Table C.1.1: The table shows the local base pair step helical parameters of the E. coli 29-mer NMR structure.

Table C.1.2: The table shows the local base pair parameters of the E. coli 29-mer NMR structure.

Table C.1.3: The table shows the ribose dihedral angles, $\nu_0$ to $\nu_4$, the amplitude of pseudorotation of the sugar ring, tm, the phase angle of pseudorotation of the sugar ring, P, and the resultant type of puckering for the first strand of the E. coli 29-mer NMR structure.

Table C.1.4: The table shows the ribose dihedral angles, $\nu_0$ to $\nu_4$, the amplitude of pseudorotation of the sugar ring, tm, the phase angle of pseudorotation of the sugar ring, P, and the resultant type of puckering of the second strand of the E. coli 29-mer NMR structure.

Table C.1.5: The table shows the backbone dihedral angles, for the base paired residues of the first strand of the E. coli 29-mer NMR structure.

Table C.1.6: The table shows the backbone dihedral angles, for the base paired residues of the second strand of the E. coli 29-mer solution structure.

Table C.2.1: The table shows the local base pair step parameters of the H. h. 29-mer NMR structure.

Table C.2.2: The table shows the local base pair parameters of the H. h. 29-mer NMR structure.

Table C.2.3: The table shows the ribose dihedral angles, $\nu_0$ to $\nu_4$, the amplitude of pseudorotation of the sugar ring, tm, the phase angle of pseudorotation of the sugar ring, P, and the resultant type of puckering of the first strand of the H. h. 29-mer NMR structure.

Table C.2.4: The table shows the ribose dihedral angles, $\nu_0$ to $\nu_4$, the amplitude of pseudorotation of the sugar ring, tm, the phase angle of pseudorotation of the sugar ring, P, and the resultant type of puckering of the second strand of the H. h. 29-mer NMR structure.
Table C.2.5: The table shows the backbone dihedral angles for the base paired residues of the first strand of the *H. h.* 29-mer NMR structure.

Table C.2.6: The table shows the backbone dihedral angles for the base paired residues of the second strand of the *H. h.* 29-mer NMR structure.

Table C.3.1: The table shows the local base pair step parameters and the form of the RNA at a given step of the *H. h.* 37-mer NMR structure.

Table C.3.2: The table shows the local base pair parameters of the *H. h.* 37-mer NMR structure.

Table C.3.3: The table shows the ribose dihedral angles, $\nu_0$ to $\nu_4$, the amplitude of pseudorotation of the sugar ring, $\tau_m$, the phase angle of pseudorotation of the sugar ring, $\psi$, and the resultant type of puckering of the *H. h.* 37-mer NMR structure.

Table C.3.4: The table shows the backbone dihedral angles, for the base paired residues of the second strand of the *H. h.* 37-mer NMR structure.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>One dimensional</td>
</tr>
<tr>
<td>2D</td>
<td>Two dimensional</td>
</tr>
<tr>
<td>A-loop</td>
<td>Aminoacyl loop</td>
</tr>
<tr>
<td>A-site</td>
<td>Aminoacyl site</td>
</tr>
<tr>
<td>bsd</td>
<td>Blasticidin S deaminase</td>
</tr>
<tr>
<td>COSY</td>
<td>Correlation spectroscopy</td>
</tr>
<tr>
<td>CPD</td>
<td>Composite pulse decoupling</td>
</tr>
<tr>
<td>CPMG</td>
<td>Carr, Purcell, Meiboom and Gill</td>
</tr>
<tr>
<td>DFT</td>
<td>Density functional theory</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DQF-COSY</td>
<td>Double quantum filtered correlation spectroscopy</td>
</tr>
<tr>
<td>E-site</td>
<td>Exit site</td>
</tr>
<tr>
<td>EF</td>
<td>Elongation factor</td>
</tr>
<tr>
<td>eq</td>
<td>Equivalents</td>
</tr>
<tr>
<td>FID</td>
<td>Free induction decay</td>
</tr>
<tr>
<td>HCP</td>
<td>Hydrogen-Carbon-Proton</td>
</tr>
<tr>
<td>HPLC</td>
<td>High-performance liquid chromatography</td>
</tr>
<tr>
<td>HF</td>
<td>Hartree-Fock</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear single quantum correlation</td>
</tr>
<tr>
<td>IF</td>
<td>Initiation factor</td>
</tr>
<tr>
<td>INEPT</td>
<td>Insensitive Nuclei Enhanced by Polarization Transfer</td>
</tr>
<tr>
<td>MLEV</td>
<td>Malcolm Levitt's CPD sequence</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>NDM-1</td>
<td>New Delhi metallo-ß-lactamase-1</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser effect</td>
</tr>
<tr>
<td>NOESY</td>
<td>Nuclear Overhauser effect spectroscopy</td>
</tr>
<tr>
<td>P-loop</td>
<td>Peptidyl transferase loop</td>
</tr>
<tr>
<td>P-site</td>
<td>Peptidyl transferase site</td>
</tr>
<tr>
<td>PTC</td>
<td>Peptidyl transferase centre</td>
</tr>
<tr>
<td>RDC</td>
<td>Residual dipolar coupling</td>
</tr>
<tr>
<td>RF</td>
<td>Release factor</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root mean squared deviation</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>ROE</td>
<td>Rotating frame Overhauser effect</td>
</tr>
<tr>
<td>ROESY</td>
<td>Rotating frame Overhauser effect spectroscopy</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal ribonucleic acid</td>
</tr>
<tr>
<td>TOCSY</td>
<td>Total correlation spectroscopy</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer ribonucleic acid</td>
</tr>
<tr>
<td>WALTZ</td>
<td>A broad-band composite pulse decoupling scheme</td>
</tr>
</tbody>
</table>
### List of symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>Ångström</td>
</tr>
<tr>
<td>Ångs</td>
<td>Ångström</td>
</tr>
<tr>
<td>mÅ</td>
<td>Ångström</td>
</tr>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>B₀</td>
<td>Applied magnetic field</td>
</tr>
<tr>
<td>B₁</td>
<td>Local magnetic field</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>Da</td>
<td>Daltons</td>
</tr>
<tr>
<td>Δ</td>
<td>Shift</td>
</tr>
<tr>
<td>Δₓ</td>
<td>x displacement</td>
</tr>
<tr>
<td>Δᵧ</td>
<td>y displacement</td>
</tr>
<tr>
<td>Δᵧ</td>
<td>y displacement</td>
</tr>
<tr>
<td>Dz</td>
<td>Rise</td>
</tr>
<tr>
<td>E</td>
<td>Energy</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>h</td>
<td>Planck’s constant</td>
</tr>
<tr>
<td>H</td>
<td>Hamiltonian</td>
</tr>
<tr>
<td>H₅</td>
<td>Hamiltonian due to chemical shift</td>
</tr>
<tr>
<td>H₆</td>
<td>Hamiltonian due to coupling</td>
</tr>
<tr>
<td>I</td>
<td>NOE intensity</td>
</tr>
<tr>
<td>Iₛ</td>
<td>Dipolar coupled (coupled to S)</td>
</tr>
<tr>
<td>J</td>
<td>Scalar Coupling constant</td>
</tr>
<tr>
<td>k</td>
<td>Boltzmann's constant</td>
</tr>
<tr>
<td>kₐ</td>
<td>Association constant</td>
</tr>
<tr>
<td>M₀</td>
<td>Bulk magnetisations</td>
</tr>
<tr>
<td>mₛ</td>
<td>Magnetic spin state</td>
</tr>
<tr>
<td>Jₙ</td>
<td>n bond scalar coupling constant</td>
</tr>
<tr>
<td>ppm</td>
<td>Parts per million</td>
</tr>
<tr>
<td>r</td>
<td>Distance between I and S</td>
</tr>
<tr>
<td>Sₛ</td>
<td>Svedberg sedimentation constant</td>
</tr>
<tr>
<td>Sₛ</td>
<td>Saturated dipolar coupled spin (coupled to I)</td>
</tr>
<tr>
<td>SLₘ₅ₐₓ</td>
<td>Spin lock mixing time</td>
</tr>
<tr>
<td>Sₓ</td>
<td>Shear</td>
</tr>
<tr>
<td>Sᵧ</td>
<td>Stretch</td>
</tr>
<tr>
<td>S₂</td>
<td>Stagger</td>
</tr>
<tr>
<td>t₁</td>
<td>Evolution or detection period in NMR pulse sequence</td>
</tr>
<tr>
<td>t₂</td>
<td>Evolution or detection period in NMR pulse sequence</td>
</tr>
<tr>
<td>t₃</td>
<td>Evolution or detection period in NMR pulse sequence</td>
</tr>
<tr>
<td>T</td>
<td>Temperature</td>
</tr>
<tr>
<td>T₁</td>
<td>Longitudinal relaxation</td>
</tr>
<tr>
<td>T₂</td>
<td>Transverse relaxation due to intra and intermolecular interactions</td>
</tr>
<tr>
<td>T₂*</td>
<td>Transverse relaxation</td>
</tr>
<tr>
<td>T₂(ΔB₀)</td>
<td>Transverse relaxation due to inhomogeneity in the magnetic field</td>
</tr>
<tr>
<td>U</td>
<td>Uracil</td>
</tr>
<tr>
<td>W₀</td>
<td>Zero quantum transition</td>
</tr>
<tr>
<td>W₁</td>
<td>Single quantum transition</td>
</tr>
<tr>
<td>W₂</td>
<td>Double quantum transition</td>
</tr>
<tr>
<td>α</td>
<td>RNA phosphate backbone dihedral angle (defined in Figure 1.2.7)</td>
</tr>
<tr>
<td>β</td>
<td>RNA phosphate backbone dihedral angle (defined in Figure 1.2.7)</td>
</tr>
<tr>
<td>γ</td>
<td>RNA phosphate backbone dihedral angle (defined in Figure 1.2.7)</td>
</tr>
<tr>
<td>γ</td>
<td>Gyromagnetic ratio</td>
</tr>
<tr>
<td>ρ</td>
<td>RNA phosphate backbone dihedral angle (defined in Figure 1.2.7)</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>δ</td>
<td>Chemical shift</td>
</tr>
<tr>
<td>ΔE</td>
<td>Difference in energy</td>
</tr>
<tr>
<td>ε</td>
<td>RNA phosphate backbone dihedral angle (defined in Figure 1.2.7)</td>
</tr>
<tr>
<td>ζ</td>
<td>RNA phosphate backbone dihedral angle (defined in Figure 1.2.7)</td>
</tr>
<tr>
<td>η</td>
<td>Inclination</td>
</tr>
<tr>
<td>θ</td>
<td>Tip</td>
</tr>
<tr>
<td>ϑ</td>
<td>Dihedral angle</td>
</tr>
<tr>
<td>κ</td>
<td>Buckle</td>
</tr>
<tr>
<td>ν₀</td>
<td>Larmour frequency</td>
</tr>
<tr>
<td>ν₁</td>
<td>RNA ribose ring dihedral angle (defined in Figure 1.2.7)</td>
</tr>
<tr>
<td>ν₂</td>
<td>RNA ribose ring dihedral angle (defined in Figure 1.2.7)</td>
</tr>
<tr>
<td>ν₃</td>
<td>RNA ribose ring dihedral angle (defined in Figure 1.2.7)</td>
</tr>
<tr>
<td>ν₄</td>
<td>RNA ribose ring dihedral angle (defined in Figure 1.2.7)</td>
</tr>
<tr>
<td>ν₅</td>
<td>Chemical shift frequency</td>
</tr>
<tr>
<td>ρ</td>
<td>Roll</td>
</tr>
<tr>
<td>τ</td>
<td>Tilt</td>
</tr>
<tr>
<td>τ</td>
<td>Delay time in a NMR pulse sequence</td>
</tr>
<tr>
<td>τₑ</td>
<td>Correlation time</td>
</tr>
<tr>
<td>τ_m</td>
<td>Mixing time</td>
</tr>
<tr>
<td>χ</td>
<td>RNA glycosidic dihedral angle (defined in Figure 1.2.7)</td>
</tr>
<tr>
<td>Ω</td>
<td>Twist</td>
</tr>
<tr>
<td>ω</td>
<td>Propeller twist</td>
</tr>
</tbody>
</table>
In this project a number of peptidyl transferase antibiotics were studied, specifically a group of aminohexose cytosine nucleoside antibiotics and their interaction with a selected number of highly conserved ribonucleic acid (RNA) motifs, designed to represent their possible binding site within the ribosome. This group of antibiotics shows a wide range of interesting properties, including antiviral and anti-tumour activity, and as they bind to a particularly conserved region in the ribosome, they are likely to be difficult for microorganisms to develop resistance to. It is hoped that once the mechanism of action of these antibiotics is better understood, that modifications to the antibiotics can be effectively made to create new or hybrid antibiotics with more selective antibacterial, or indeed antiviral or anti-tumour properties.

The nuclear magnetic resonance (NMR) structure of the RNA binding, peptidyl transferase inhibitor antibiotics amicetin, blasticidin S and gougerotin, in their native solution states, have been successfully determined. The structures all exhibit a stable conformation, stabilised by intramolecular hydrogen bonds. Amicetin was observed to be folded, distinctly different from the linear, extended conformation of amicetin previously determined by X-ray crystallography. The structure of blasticidin S was found to be very similar to its X-ray crystal structure. Gougerotin was shown to form a similar conformation to blasticidin S, save that the end chain of gougerotin was bent at right angles to the rest of the molecule, forming a structure similar to that of the major bound X-ray crystal structure of blasticidin S. All the solution structures showed a similar conformation in the analogous regions of their chemical structure, suggesting that hybrid antibiotics could be produced.

Two highly conserved RNA motifs of Halobacterium halobium (H. h.) and Escherichia coli (E. coli) 23S ribosomal RNAs were chosen to investigate their interaction with amicetin. The NMR structure of the H. h. and E. coli 29-mer RNA motifs have been determined; the motifs both form well folded A-form RNA conformations. The E. coli NMR structure differs from the X-ray crystal structure of the motif contained within the ribosome, as a highly conserved adenine residue, which resides in a bulge strongly implicated with amicetin binding, folds into the helix as opposed to being flipped out. Instead, an adjacent cytosine residue partially flips out; whereas in the crystal structure, it is folded within the helix. The NMRstructures of the H. h. motif differs from the X-ray crystal structure of the motif, contained within the ribosome, as none of the bases are flipped out and a number of non-canonical base pairs are formed in the solution structure. To continue this study, a fully $^{13}$C and $^{15}$N isotopically labelled version of the H. h. RNA sample has been partially assigned, and an initial structure determination has been performed, using ultra high field 1 GHz spectroscopy.

Addition of amicetin to both the H. h. and E. coli 29-mer RNA samples were accompanied by discrete changes to the spectra, suggesting weak interaction between the two components. These can be qualitatively interpreted to changes induced in the local conformation of the RNA motifs and the amicetin arising from the formation of a complex, between the amicetin and the bulge region of the particular motif.
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on presentation of Theses.
**Acknowledgements**

I would like to acknowledge and thank:

- Dr Vasudevan Ramesh for his generous support and advice.
- Tony Cheung, Misbah Nareen, Bo Yang, Usman Rasul, Sadia Mohammed, Misbah Ghafoor, Nick Chan, and Dr. Christos Shammas for their helpful discussion and comments.
- The Sir Richard Stapley Educational Trust and The Grundy Educational Trust for the awards.
- A family friend who provided monetary and moral support throughout my PhD.
- Barbara Knowles, and Tanya Aspinall, for valuable help proof reading.
- The MRC Biomedical NMR Centre, N.I.M.R., Mill Hill, London for Highfield NMR support (in particular Tom Frankiel, Geoff Kelly and Alain Oregioni).
- The Euro-NMR, Lyon, France for 1 GHz NMR support (in particular Moreno Lelli).
- The Manchester Chemistry department NMR centre (in particular Roger Speak).
- My wife and family for supporting me throughout my PhD.

“It is the glory of God to conceal a matter; to search out a matter is the glory of kings.”

(Proverbs 25:2 )
Chapter 1

Introduction

This chapter will discuss the significance and background information of this project. This will include the importance of antibiotic resistance, an introduction to the antibiotics this project will examine and, the antibiotics target, the peptidyl transferase centre (PTC) of the ribosome. A short introduction to the techniques used will also be given, including nuclear magnetic resonance (NMR), molecular modelling and NMR structure determination. Additionally the background and earlier work carried out for this project will be discussed followed by the aims of the current project.

1.1 Antibiotic resistance

1.1.1 The importance of resistance

Microorganisms are rapidly becoming resistant to our current range of antibiotics in various fields, the most disturbing is in the case of clinical antibiotics.¹ An example of this in medicine, is the case of Staphylococcus aureus (S. aureus) infections; this Gram-positive bacteria used to be treated with penicillin, but penicillin resistant S. aureus is now common.¹ Methicillin type antibiotics therefore, are now used, but resistances to even methicillin type antibiotics have begun to occur. The antibiotic vancomycin is often used in cases of methicillin resistant S. aureus infection, but there are now even reports of vancomycin resistant S. aureus infections.¹ According to recent government statistics, the number of fatal cases of S. aureus infections have reduced, but the percentage which are methicillin resistant have continued to increase. Figure 1.1.1 indicates the number of death certificates citing S. aureus infections in England and Wales from 1993 to 2008, and the proportion that are due to methicillin resistant Staphylococcus aureus.
Recently a new antibiotic resistance mechanism has emerged in Gram-negative bacteria, which confers resistance to the antibiotic carbapenem, by New Delhi metallo-β-lactamase 1 (NDM-1). It has so far been found in India, Pakistan and the UK. This is particularly worrying, as carbapenem is currently a reserved antibiotic; an antibiotic of last resort. There are no new antibiotics which are active against NDM-1 producers in the pharmaceutical pipeline. This again emphasises the need for research into antibiotic resistance.

Resistance can be divided into two main categories, intrinsic and acquired. Intrinsic resistance is where the microbe is inherently resistant to the antibiotic, due to a feature of the microbe of which its primary function is not aimed at conferring resistance. Acquired resistance occurs when a population of microbes is exposed to an antibiotic, and in response to this selection pressure microbes which have or develop a resistance to the antibiotic are selected for. The acquired resistance can arise via two mechanisms; the first mechanism is when a random mutation in a microbe happens to result in resistance; this
mutation can then be passed down to daughter cells. The second, more common, method is the transfer of resistance genes between microbial cells of different strains or species.

There are three main methods of antibiotic resistance:

- Antibiotic inactivation.
- Decreasing the uptake or increasing the efflux into or out of the cell respectively.
- Drug target alteration.

Antibiotic inactivation occurs when the antibiotic is modified by an enzyme, expressed by the microbe, which results in the antibiotic's activity being abolished or reduced. Decreasing the uptake or increasing the efflux works by simply keeping the concentration of the antibiotic in the cell low, either by reducing the amount entering the microbe or increasing the rate of its removal. Drug target alteration works by either altering the target of the antibiotic, so that the antibiotic no longer interacts as effectively, or by overproducing the target, which ensures that there is enough of the target to perform its function despite interference from the antibiotic.

A continued stream of new antibiotics is therefore required to treat infections, as microbes become resistant to older classes of antibiotics. Unfortunately new classes of antibiotics appear to be getting harder and harder to find. Controlling the use of antibiotics can help slow the rate of development of antibiotic resistance, but these measures are only likely to buy time. Crucially, this valuable time can be used to develop new antibiotics, which potentially will be more of a challenge for the microbes to develop resistance to. It is this goal that this project aims to further.
1.1.2 The potential of peptidyl transferase antibiotics

The peptidyl transferase antibiotics are the largest group among the naturally occurring antibiotics. They show great potential as they target the PTC of the ribosome, which is largely structurally conserved throughout a wide range of organisms. This may mean that developing resistance may be difficult. Unfortunately, this high degree of structural conservation also creates a great difficulty, causing this class of antibiotics to tend to be universal in their mode of action. This means they affect all types of organisms; probably due to the structural conservation of the ribosome between different species. Although some of the antibiotics have been shown to be more selective between different types of organism. For example, the widely used clinical antibiotic azithromycin, which binds to the ribosome and inhibits protein synthesis, is much more effective in inhibiting the bacterial Deinococcus radioduran ribosome than in the archaea ribosome Haloarcula marismortui. This selectivity is also shown by the fact that it can be used clinically.

Alterations to the antibiotics which increase their affinity for a particular type of ribosome maybe possible, but also may suggest that a non-harmful mutation could provide a means of resistance, as the PTC of most organisms is very similar. Therefore, it could be more effective to take advantage of other differences between cells, for instance the rate of uptake into the cell. If this class of antibiotics is better understood, it may be possible to see how to modify individual antibiotics to retain activity and increase selectivity.

Another useful property of the peptidyl transferase antibiotics is that the gene for the ribosome exists in several places. So if a resistance mutation occurs via modification of the target ribosome's DNA, the result of the mutation will only occur in a small proportion of the ribosomes as the mutation will be localised to one copy of the ribosomal DNA,
leaving the rest unchanged. This will leave the organism still partially susceptible to the antibiotic, buying more time for the antibiotic to kill the microbe before it passes on the resistance or generates further resistance. This could significantly reduce the rate of resistance occurring, if the main source of resistance occurs via the drug target alteration pathway. Also as there are only four natural bases to chose from in RNA, as opposed to the twenty amino acids in the case of proteins, there are less options for mutation in order to get the right balance of resistance and functionality.

Another interesting feature of certain peptidyl transferase antibiotics is that they not only have antimicrobial activity but also anti-tumour and antiviral properties. These antibiotics selectively inhibit protein synthesis in tumour and virally infected cells. Also some peptidyl transferase antibiotics are used as crop protection agents; in particular blasticidin S. An increased interest in environmental matters, and subsequent new environmental protection legislation has resulted in renewed interest in the use of natural products in farming. Antibiotics such as blasticidin S could potentially play an important role in farming over the coming years.
1.2 Physical and chemical properties of RNA \(^{14}\)

The PTC is entirely composed of RNA, so it is important to have a good understanding of its structure. RNA is a polymer composed of four different monomer subunits. A generic monomer is shown in Figure 1.2.1. The monomer is made up of three sections, the ribose ring, the phosphate group and the heterocyclic base.

![RNA monophosphate nucleotide](image)

**Figure 1.2.1:** An example of a RNA monophosphate nucleotide (with a uracil base), showing the standard ribose numbering.

The base in the nucleotide monomer of RNA can be adenine (A), uracil (U), cytosine (C) or guanine (G). These bases can come together to form base pairs, the standard Watson-Crick (W-C) base pairs are shown in Figure 1.2.2. The RNA monomers when combined together form polynucleotides and can form a duplex by base pairing with a polynucleotide with a complementary set of bases, as illustrated in Figure 1.2.3.

![RNA base pairs](image)

**Figure 1.2.2:** The two canonical base pairs of RNA (A-U and G-C); the bases show the standard base numbering.
Figure 1.2.3: A fully W-C base paired RNA duplex indicating the 5' and 3' ends of each strand.

Three levels of RNA structure are discussed: primary, secondary, and tertiary, these are illustrated in Figure 1.2.4.

Defining primary and secondary structure

The primary structure of these RNA polynucleotides is the sequence that the different types of nucleotides appear in. The sequence is typically started from the 5' end of the RNA. The 5' end of the RNA is the end where the 5' carbon of the ribose is closest to the end of the RNA chain. The other end of the same RNA chain is known as the 3' end following the same reasoning. An example of a primary structure would be “A U G C U C A A”, where each of the letters represents a different base, Figure 1.2.4. shows another example of a primary structure.
Figure 1.2.4: Image A shows an example of a RNA primary structure corresponding to a 17-mer RNA sequence, image B shows the secondary structure formed by hydrogen bonded base pairing with a complementary strand of the same RNA. Image C shows the corresponding tertiary structure due to the folding of the RNA duplex.

The secondary structure is concerned primarily with a two dimensional (2D) shape of the molecule based on which bases form base pairs with each other. Figure 1.2.5. shows four basic classes of RNA secondary structure, the simplest being the duplex where one polynucleotide is fully base paired to a complementary strand of polynucleotide RNA. Secondary structures can be complicated, and therefore interesting, when the structure does not follow standard base pairing.

Figure 1.2.5: Schematic representation of four major classes of RNA secondary structures. The solid lines represent the nucleic acid back bone, the dotted lines represent the hydrogen bonded base pairs.
Defining the tertiary structure

The secondary structure is not fully informative and hence it is important to study the tertiary structure (the three dimensional structure) of the RNA. A RNA duplex will tend to form a double helix structure, in particular an A-form helix, as shown in Figure 1.2.4. Two important features of the A-form RNA helix are the major and minor grooves of the helix, Figure 1.2.6 shows their location on an A-form helix.

![Figure 1.2.6: Figure illustrates the location of the major and minor grooves on the tertiary structure of an A-form RNA helix.](image)

Deviations in the secondary structure from a fully complementary W-C base pair, such as the insertion of an extra base into one of the strands, will distort the A-form helix. Therefore, it is important to be able to quantify these distortions, or changes in the RNA tertiary structure to enable adequate description and the observation of possibly significant patterns.

One important way to begin to define the tertiary structure is via a set of dihedral angles $\alpha$, $\beta$, $\gamma$, $\delta$, $\epsilon$ and $\zeta$ which define the backbone conformation of RNA, $\chi$ which defines the
glycosidic dihedral angle, and $\nu_0$, $\nu_1$, $\nu_2$, $\nu_3$ and $\nu_4$ which define the conformation of the ribose ring. These dihedral angles are defined in Figure 1.2.7. There are certain standard forms that nucleic acid tertiary structure can take. Deoxyribonucleic acid (DNA) tends to take on a B-form type conformation with the deoxyribose taking a C$_{2'}$-endo conformation and RNA an A-form type conformation with the ribose sugar taking a C$_{3'}$-endo conformation, Figure 1.2.8 shows a representation of these two common ribose ring pucker conformations. These types of tertiary structure conformations can be defined by the aforementioned dihedral angles.

Figure 1.2.7: In image A a section of a polyribonucleotide chain is shown with the phosphate backbone dihedral angles labelled $\alpha$ to $\zeta$; the glycosidic dihedral angle $\chi$ is also shown. In image B the dihedral angles of the ribose ring are labelled $\nu_0$ to $\nu_4$.

Figure 1.2.8: A schematic representation of the common ribose ring pucker conformation C$_{3'}$-endo, as seen in RNA, is shown in the left image and C$_{2'}$-endo, as seen in DNA, is shown in the right image.
Another way of describing aspects of the tertiary structure of a RNA helix is by using helical parameters.\textsuperscript{15} There are many helical parameters that describe different aspects of RNA conformation. The helical parameters can be measured based on either a global or local helical axis. A global helical axis is plotted along the centre of the RNA helix, as calculated by software, and is used as the z dimension. How the local helical axis is defined is illustrated in Figure 1.2.9, the local helical parameters are calculated using this. Figure 1.2.10 shows the coordinate frame used by Figures 1.2.11 to 1.2.12 to describe the base pair helical, base pair and base pair step parameters. Local base pair helical parameters make little sense as the axes are defined by the orientation of the base pair, therefore the value of all four would always equal zero. However, local base pair step helical parameters can be measured. This is where the previous base pair's local axis is taken as the reference axis and so it is the deviation from this new reference axis that the base pair in question's local base pair step helical parameters are measured.

**Figure 1.2.9:** The figure shows how the local helical axis is defined for the calculation of local helical parameters due to the two strands I and II of the RNA.

**Figure 1.2.10:** This figure shows the coordinate frame of reference used in figures 1.2.11 to 1.2.13 to describe helical parameters of the RNA.
Figure 1.2.11: The four base pair helical parameters ($dx$, $dy$, $\eta$ and $\theta$), drawn with reference to the coordinate frame shown in figure 1.2.10.

Figure 1.2.12: The five base pair parameters ($S_x$, $S_y$, $S_z$, $\kappa$ and $\omega$), drawn with reference to the coordinate frame shown in figure 1.2.10.

Figure 1.2.13: The six base pair step parameters ($D_x$, $D_y$, $D_z$, $\rho$, $\tau$ and $\Omega$), drawn with reference to the coordinate frame shown in figure 1.2.10.
1.3 The ribosome – mechanism of protein synthesis

1.3.1 Introduction to the ribosome (archaea, prokaryotes, eukaryotes)

The ribosome is composed of both proteins and RNA and is vital to protein synthesis and therefore vital to life itself. The Nobel prize for chemistry was in fact awarded in 2009 "for studies of the structure and function of the ribosome" to Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath. The ribosome as a whole acts like a large enzyme, holding the various components required for protein synthesis in place. However, the main catalytic centre, the PTC, consists entirely of RNA with the closest protein over 18Å away. The ribosome works at an astonishing pace and can produce peptide bonds at a rate of $\geq 300 s^{-1}$ when bound to acceptable substrates.

The three evolutionary kingdoms of life, archaea, prokaryotes and eukaryotes, all contain ribosomes, although there are some distinct differences between the ribosomes of the three types. Common to all three types is the fact that each ribosome is composed of two subunits, known as the large and small subunits. The ribosome's size is commonly measured by the Svedberg sedimentation constant (S). Archaea and prokaryote ribosomes have a sedimentation constant of 70S, with a 50S large subunit and a 30S small subunit. Eukaryote ribosomes have a sedimentation constant of 80S, with a 60S large subunit and a 40S small subunit.

The RNA that makes up the ribosome is known as ribosomal RNA (rRNA). Within the cell there are two other types of RNA, known as messenger RNA (mRNA) and transfer RNA (tRNA). The mRNA is a RNA copy of a section of DNA, its role is to take the genetic information from the DNA to the ribosome, so that the information it contains can
be read and a protein produced. The tRNA is a section of RNA that brings the required amino acid to the ribosome as the mRNA is being read in order to produce the protein coded for.

1.3.2 Structure of the ribosome

The three important sites of protein synthesis are located in the large subunit of the ribosome, they are:

- The aminoacyl or attachment site (A-site).
- The peptidyl transferase site (P-site).
- The exit site (E-site).

The orientation of the two subunits and the approximate locations of the E, P and A-sites are shown in diagrammatic form in Figure 1.3.1.

In the case of the bacterial ribosome, the main component of the 30S small ribosomal subunit is the 16S rRNA; the rest is made up of 20-21 ribosomal proteins. The prokaryotic ribosome 50S subunit is made up of two rRNA, the 23S and 5S rRNAs, and 31-35 more ribosomal proteins. Figure 1.3.2 shows the X-ray crystal structure of the *E. coli* (prokaryotic) ribosome, with each of the major components individually colour coded to help depict the structure.
Figure 1.3.1: Shown above are two representations of the bacterial ribosome, in two different orientations. The 50S and 30S ribosomal subunits, in each, are shown along with the locations of the E, P and A-sites, and their orientation with respect to a strand of messenger RNA.\textsuperscript{21}

The 50S subunit contains the PTC, with the PTC located within domain V of the 23S rRNA. Figure 1.3.3 shows the secondary structure of the PTC of the \textit{E. coli} ribosome.\textsuperscript{22}

The A and P-loops are shown, as they are important components of the A and P-sites, respectively. Figure 1.3.4 shows the PTC of the \textit{E. coli} ribosome X-ray crystal structure.\textsuperscript{20}

The structure is colour coded to match the secondary structure shown in Figure 1.3.3. The structure shows how close the P and A-loops are to each other in the ribosome structure.
Figure 1.3.3: The secondary structure of the PTC, found within domain V of the 23S ribosomal RNA of *E. coli*, with the predicted amicetin binding motif labelled (see section 1.6). The secondary structure is colour coded to match Figure 1.3.4. The 5' and 3' ends are labelled and every tenth nucleotide is marked with a line, and every 50th is numbered according to the *E. coli* ribosomal numbering scheme. A line is drawn between nucleotides to indicate a canonical base pair and a dot is drawn between nucleotides to indicate a G-U base pair.
1.3.3 Function of the ribosome

The function of the ribosome is to enable protein synthesis. A condensed outline of the protein synthesis pathway is shown in Figure 1.3.5. The PTC is the catalytic site of protein synthesis.\textsuperscript{4,16} It is thought that the PTC catalyses protein synthesis by holding together the specific components in the correct orientation, while allowing the tRNA to provide the chemical component of the catalysis reaction.\textsuperscript{23}
Figure 1.3.5: This figure shows the proposed process of translation of mRNA to protein, indicating the various processes that occur. The key shows the identity of the various components.
The process of protein synthesis is thought to begin when the mRNA, which contains the genetic information to produce a protein, binds to the small ribosomal subunit. This is followed by a tRNA, which binds to the mRNA and small ribosomal subunit, and carries the correct amino acid according to the codon. At the start of the translation, the first tRNA would bind to the start codon on the mRNA.

A codon is a set of three nucleotides on the mRNA which code for a particular amino acid. At the start of translation the first tRNA binds to the start codon, (AUG) which codes for the start of the protein as well as a particular amino acid (methionine). A codon is paired to an anti-codon which is the complementary W-C base pairing sequence. The anti-codon is located on a tRNA, there is at least one type of tRNA for each amino acid, the tRNA transports the type of amino acid coded for by the codon to the ribosome.

After the initial tRNA has bound to the start codon, the large ribosomal subunit binds to the small subunit, the tRNA is now located at the P-site. Another tRNA with an anti-codon fitting the next codon in the mRNA sequence then carries the appropriate amino acid to the ribosome and binds to the A-site.

A major role of the ribosome is to hold the A-site tRNA and the P-site tRNA in the correct positions to help catalyse the formation of peptide bonds to form the protein. All known tRNAs have a CCA sequence at their 3’ end,24 a P-site bound tRNA’s main contact with the P-site is at the P-loop with the residues G2251 and G2252 (E. coli numbering; all following numbering will be for E. coli unless otherwise stated) forming base pairs with the conserved CCA 3’ end specifically C75 and C74, of the tRNA.24,16 An A-site bound tRNA's main contact with the A-site is between G2553 of the A-loop and C75 of the tRNA.25,16 Both sets of interactions are shown in Figure 1.3.6. The mechanism by which
the peptide bond is thought to form is catalysed by the 2'-OH of the P-site tRNA by a proton shuttle mechanism as seen in Figure 1.3.7. The 2'-OH of A2451 of the 23S rRNA also plays an important role by hydrogen bonding to the 2'-OH of the P-site tRNA, therefore helping to hold the P-site tRNA in place.\textsuperscript{26,19}

Figure 1.3.6: The left image shows a schematic illustration of tRNA interacting with the A-loop. The right image shows tRNA interacting with the P-loop. The long continuous line indicates the RNA backbone, the short straight line represents base pairs, the 5' and 3' ends of the tRNA are also marked.\textsuperscript{27}

Figure 1.3.7: The proposed mechanism for peptide bond formation in the ribosome, via the proton shuttle mechanism, indicating the role of A2451.\textsuperscript{26,19}
Once the peptide bond is formed between the two amino acids, the first amino acid is no longer attached to the P-site tRNA. This first spent tRNA moves to the E-site and then leaves the ribosome. The tRNA currently held in the A-site is then transferred to the P-site, and a new tRNA enters the A-site and binds to the next codon. Another peptide bond is then formed. This process continues until a stop codon is reached. At this point the release factor protein (RF) binds to the stop codon at the A-site and causes the release of the new protein from the P-site tRNA and subsequently from the ribosome. The protein exits the ribosome via the nascent peptide tunnel.\(^4\) The chemical mechanism by which this separation from the P-site tRNA is thought to take place is shown in Figure 1.3.8.\(^4\) The ribosome then dissociates ready to start the process all over again.

![Figure 1.3.8: The proposed mechanism for polypeptide release from the P-site tRNA.](image)

### 1.3.4 Conserved RNA motifs

Despite the differences in the ribosome between the three different kingdoms (archaea, prokaryote and eukaryote), a considerable amount of the ribosome is highly conserved. This is illustrated in Figure 1.3.9, where the PTC of the 23S rRNA of the *E. coli* ribosome has been colour coded to indicate the extent of conservation of the *E. coli* sequence compared to 592 sequences taken from a range of archaea, prokaryotes and eukaryotes.\(^22\) Large sections of the ribosome are over 98% conserved. This is surprising given the broad range of sequences covered, and suggests that these regions are likely to be important, if not essential, to the functionality of the ribosome.
Figure 1.3.9: The figure shows the extent of conservation of the PTC of the *E.coli* domain V sequence compared to 592 sequences taken from a range of archaea, prokaryotes and, eukaryotes. The 5' and 3' ends are labelled and every tenth nucleotide is marked with a line, and every 50th is numbered, a line is drawn between nucleotides to indicate a canonical base pair and a dot is drawn between nucleotides to indicate a G-U base pair. Residue U2438 is highlighted which is thought to be of particular importance to the binding of the group of antibiotics which were studied in this project. Additionally its mutation to C strongly impairs ribosome function.
1.4 Peptidyl transferase antibiotics

The subject of this study is a sub-group of peptidyl transferase antibiotics, called the aminohexose cytosine nucleosides, which all share a number of common structural features. Specifically, the three antibiotics that are the subject of this study are amicetin, blasticidin S and gougerotin. These all share a cytosine moiety and a “saccharide” type ring, and all have at least one amide bond. These commonalities between the structures appear to be important to their functionality, particularly the cytosine ring. For example if the cytosine ring is replaced in blasticidin S by a uracil the activity drops by up to a 1000 fold. Each of these antibiotics target the PTC and then acts to inhibit the vital activity of protein synthesis. This results in inhibition of growth, replication and repair.

1.4.1 Amicetin

Amicetin was first isolated from a Streptomyces sp., which was found in a soil sample from Kalamazoo, Michigan. The amicetin structure is shown in Figure 1.4.1. It can be isolated from cultures of Streptomyces fascilatus, Streptomyces plecatus and Streptomyces vinaceus-drappus. Amicetin is a universal antibiotic, inhibiting the ribosomes of archaea, prokaryotes and eukaryotes. The antibiotic was initially found to have activity against a number of bacteria, including Mycobacterium tuberculosis and E. coli. In general, it was found to be most active against Gram-positive and acid fast bacteria (particularly Mycobacteria). Amicetin also showed antiviral activity; in a mono cell layer assay designed to find new antiviral drugs, amicetin showed activity against the herpes virus.
Figure 1.4.1: The chemical structure of the peptidyl transferase antibiotic amicetin. The atom specific labelling is shown in Arabic numerals (1 to 18, 1’ to 6’ and 1* to 8*). Stereochemical non-equivalent saccharide protons are distinguished as either axial or equatorial and given the sub label a or e respectively, other stereochemical non-equivalent protons are distinguished by either an a or b sub-label.32

There have been a number of structural studies into the structure of amicetin; including X-ray crystallography, NMR, and computational studies. The earliest study was a simple Dreiding stereo-model.29 This was followed by an X-ray crystallography study of amicetin in its free-state, which predicted a structure with a highly extended conformation with a number of intramolecular hydrogen bonds.32 It showed a dihedral angle of 25.7º along the glycosyl linkage between the saccharide ring and the cytosine moiety (O5’-C1’-N1-C6), which is low compared to other pyranosyl nucleosides.32,35

A more recent structural study carried out on amicetin was a full NMR solution structure determination.32 It predicted a folded conformation, held together by a network of hydrogen bonds,32 but the structure has been revised with additional constraints and better stereochemistry as part of this project (described in section 3.3).

1.4.2 Blasticidin S

Blasticidin S was first isolated from Streptomyces griseochromogenes by Takeuchi et al,36 its structure is shown in Figure 1.4.2.35 The chemical structure of blasticidin S was then determined.37,38 Blasticidin S is also a universal antibiotic, inhibiting the ribosomes of all three so called evolutionary domains.33,34
Figure 1.4.2: The chemical structure of the *peptidyl transferase* antibiotic blasticidin S. The atom specific labelling is shown in Arabic numerals (1 to 15, 1' to 7' and 6''). The labelling is loosely based on the system set out and used for the X-ray crystal structure, changes needed to be made to allow individual labelling of all proton groups.35 Stereochemical non-equivalent protons are distinguished by either an a or b sub-label.

Blasticidin S has been used as an agricultural fungicide.12,13 Its importance in agriculture is less significant today as lower toxicity pathogen-specific synthetic rice blast products have been introduced.13 But, as concern is being raised over synthetic agriculture products and their impact on the environment, it may yet have a resurgence in use for this purpose. Blasticidin S is said to have no deleterious effects on the environment and is relatively non-hazardous to non-target organisms when used correctly.13

More recently blasticidin S has been used as a selection agent in molecular biology. This use came about when several blasticidin S resistant organisms were discovered to produce a blasticidin S deaminase (bsd). This enzyme catalyses the deamination of the cytosine moiety, which results in a non toxic deaminohydroxy derivative.39 Once the gene was identified and isolated it could be added to a plasmid and given to a selected microbe.39

Blasticidin S has been shown to have antiviral activity10 in addition to its antibacterial and anti-fungal activity. The antiviral activity was originally thought to be due to an increase in the permeability of the cell membrane when exposed to viral infection. However, a later study into the antiviral activity of the related antibiotic gougerotin, indicated that this is probably not the case (this study will be discussed further in section 1.4.3).11 Blasticidin S has also been shown to have some anti-tumour activity.6
It is of interest that a fluorinated blasticidin S has been produced (5-fluoro-blasticidin S). This could be of use in further studies into the antibiotic, particularly given the favourable NMR properties of fluorine.\textsuperscript{40} A simple study of the fluorinated blasticidin S showed it to have similar antibiotic properties to the non-fluorinated version.\textsuperscript{40} Another analogue of blasticidin S has been produced where the 4NH\textsubscript{2} was replaced with SH. This analogue has been shown to have anti-tumour and antibacterial properties.\textsuperscript{41} The fluorinated blasticidin S is resistant to enzymatic 4-deamination, like that initiated by the bsd enzyme.\textsuperscript{42} It is likely that the blasticidin S 4SH analogue may also show resistance to this process.

There have been a number of studies into the structure of blasticidin S. Again, the earliest was a Dreiding stereo-model.\textsuperscript{29} This study indicated a fairly linear structure, with the cytosine ring at approximately right angles to the pseudo saccharide ring (along the glycosyl linkage) and with a hydrogen bond between O7 and NH\textsuperscript{9}. The second study was an X-ray crystal structure of blasticidin S hydrochloride pentahydrate.\textsuperscript{35} This predicted another linear structure with a different intramolecular hydrogen bond between O6\textsuperscript{9} and NH\textsuperscript{9}; there were also a number of intermolecular hydrogen bonds within the crystal structure. The X-ray crystal structure also showed a dihedral angle of 86\textdegree{} along the glycosyl linkage between the saccharide ring and the cytosine moiety (O5'-C1'-N1-C6), which is a similar value to other pyranosyl nucleosides (and to that indicated in the Dreiding stereo model).\textsuperscript{35}
1.4.3 Gougerotin

Gougerotin is an antibiotic very similar to blasticidin S, and is thought to function in the same way. Its structure is shown in Figure 1.4.3. Gougerotin can be isolated from Streptomyces gougerotii, and it is also a universal antibiotic. Although gougerotin has not been used commercially, it does however show a wide range of biological activity. It has been shown to have antiviral, anti-tumour, and of course antibacterial properties.

![Figure 1.4.3: The chemical structure of the peptidyl transferase antibiotic gougerotin. The atom specific labelling is shown in Arabic numerals (1 to 14, 1' to 7' and 6''). The labelling is based on the system set out and used for blasticidin S. Stereochemical non-equivalent protons are distinguished by either an a or b sub-label.]

A number of analogues of gougerotin have been produced. One analogue simply replaces the 6" CONH$_2$ with a CH$_2$OH group; this analogue has been shown to inhibit peptide synthesis. Another analogue of gougerotin that does not inhibit peptide synthesis is shown in Figure 1.4.4, this is thought to be due to a change in stereochemistry from the native gougerotin structure. A hybrid of blasticidin S and gougerotin has also been produced.

49
The antiviral activity of gougerotin is thought to be due to an increased efficiency in protein inhibition within viral infected cells. This was initially thought to be due to an increase in the permeability of the cell membrane, but a later study showed that this was unlikely to be the dominant factor as, in most cases, viral infection does not increase the rate of uptake of gougerotin into the cell. Gougerotin is likely to be absorbed into the cell via endosomes; it is the fate of these endosomes when they enter the cell which is likely to change and cause an increase in the efficiency of inhibition. Another possibility is that although the measured rate of absorption did not increase, there may be a small increase in permeability, caused by an alternate route into the cell being opened up. This route could bring the gougerotin more effectively to its site of action.

There have been relatively few investigations into the structure of gougerotin. Again the first investigation into its structure was a Dreiding stereo-model. The Dreiding stereo-model produced is very similar to the structure of blasticidin S generated by the same technique. However, there is a kink in its tail, causing the end section of the chain to run perpendicular to the two rings. It also indicates a hydrogen bond between O7 and OH9. An X-ray crystal structure has been generated for a derivative of gougerotin which is

Figure 1.4.4: The chemical structure of an analogue of the peptidyl transferase antibiotic gougerotin.
composed simply of the two structural rings, its structure is shown in Figure 1.4.5. There were two distinct conformations found within the crystal, in which the bases from the two types were found to stack with each other. Both structures were found to form a hydrogen bond between NH6' and O6''. The structures also both showed a dihedral angle of approximately 73º along the glycosyl linkage between the saccharide ring and the cytosine moiety (O5'-C1'-N1-C6), which is a similar value to other pyranosyl nucleosides including the value observed for blasticidin S in its X-ray crystal structure.33,44

![Figure 1.4.5: A partial gougerotin structure, for which an X-ray crystal structure has been produced.](image)

1.4.4 Other aminohexose cytosine nucleoside antibiotics

There are several antibiotics that are very closely related to amicetin, these include bamicetin, oxamicetin and plicacetin, their structures are shown in Figure 1.4.6 and Figure 1.4.7. These amicetin related antibiotics are all peptidyl transferase antibiotics which are active against Gram-positive and Gram-negative bacteria. In assays using the 70S ribosome, oxamicetin and amicetin were found to have the greatest inhibitory effect, with plicacetin being 10 times less active as an inhibitor compared to amicetin (bamicetin was not included in the assay). It was also found that cytimidine, an analogue of amicetin missing the disaccharide portion of the molecule, was 50 times less active, and cytosamine triacetate was 100 times less active compared to amicetin, their structure is shown in Figure 1.4.9 and Figure 1.4.10.
Figure 1.4.6: The chemical structure of the peptidyl transferase antibiotics, bamicetin where R1 and R2 equal H and oxamicetin where R1 equals CH$_3$ and R2 equals OH.$^{29}$

Figure 1.4.7: The chemical structure of the peptidyl transferase antibiotic plicacetin.$^{29}$

Figure 1.4.8: The chemical structure of cytimidine, an analogue of the peptidyl transferase antibiotic amicetin.$^{29}$

Figure 1.4.9: The chemical structure of cytosamine triacetate, an analogue of the peptidyl transferase antibiotic amicetin.$^{47}$

There are also several other aminohexose cytosine nucleoside antibiotics which include mildiomycin.$^{48,49}$ Mildiomycin acts by interfering with the peptidyl transferase centre$^{12,50}$ and has been used primarily to control powdery mildews in Japan.$^{12,50,51}$ Mildiomycin appears to have a low toxicity in both mammals and fish.$^{48}$ It has also been shown to have
difficulty crossing the cell membrane, when the cell membrane is removed or when the cell membrane is made more permeable by a virus, its inhibitory activity increases. Although, doubt has been cast on whether the viral infection generally permeablises the cells with respect to this class of antibiotics.\textsuperscript{52} Previous studies have shown that viral infection does not increase the rate of uptake of gougerotin into the cell, so the increased inhibitory effect must be due to another factor.\textsuperscript{11}

Other antibiotics belonging to the aminohexose cytosine nucleoside group include bagougeramines A and B,\textsuperscript{53,54} which inhibit both Gram-positive and Gram-negative bacteria plus some fungi, and the two spotted spider mite.\textsuperscript{53} Additionally there is arginomycin,\textsuperscript{55} which inhibits both Gram-positive bacteria and fungi.\textsuperscript{55} Finally there is anthelmycin (also called hikizimycin),\textsuperscript{56,57} which inhibits fungi (particularly phytopathogenic fungi) and has also been shown to have antiviral activity in picorna-viruses,\textsuperscript{10,58} and has been shown to inhibit protein synthesis.\textsuperscript{59}

There are clearly a large number of antibiotics in this group, each with a similar pattern of inhibition, and possibly a similar mode of action, but all with slightly different characteristics. Making a thorough study of these antibiotics may give clues on how they can be modified to tune their activity and thus create antibiotics to serve useful purposes, possibly clinically and agriculturally.
1.5 **Ribosome antibiotic complexes**

1.5.1 **Amicetin**

The most important study into the binding of amicetin was performed by Leviev *et al.* This study revealed that when *H. halobium* cells are exposed to a sub inhibiting concentration of amicetin for one and a half months, a specific mutation occurs in the 23S rRNA, which confers a resistance to amicetin. The site of this mutation is the U2438 residue (*E. coli* numbering), this residue mutates to C. The amicetin binding site is likely to be close to the site of mutation. When *E. coli* is exposed to a sub inhibiting concentration of amicetin the same mutation does not occur, but the amicetin protects the base A2070. This acts to reduce the probability that amicetin interferes with the binding of the aminoacyl tRNA 3’ end to the A-site. It also acts to move the binding site towards the P-loop.

This information was used in an unconstrained molecular modelling study of amicetin to a potential binding site on a 35-mer RNA motif. The study suggested that it may bind to the bulged U2068 residue *via* a number of hydrogen bonds between the cytosine moiety and one of the glycosidic linkage of amicetin. However, the binding site of U2068 seems unlikely, due to its position within the ribosome crystal structure. The base points away from the PTC and into the surrounding ribosome and therefore, in this region of the ribosome, there does not appear to be much space for amicetin to bind. U2068 does hydrogen bond with another residue, A2432, which ties together two sections of rRNA. Therefore, U2068 could be structurally important, and if its hydrogen bonding were to be disrupted it could cause an inhibition to protein synthesis. It is noteworthy however that U2068 is occasionally absent in some rRNA sequences. So far there is no crystal structure of the ribosome bound to the amicetin antibiotic.
1.5.2 Blasticidin S

There have been several studies into the binding of blasticidin S. One of the most recent studies shows an X-ray crystal structure of two molecules of blasticidin S, bound to the P-loop of the PTC within the 50S subunit of the ribosome.\textsuperscript{61} Up to this point there had been some uncertainty to where in the PTC blasticidin S binds, whether to the P or the A-site.\textsuperscript{29} It has now been suggested, that blasticidin S transiently interferes with the A-site before moving on to bind at the P-site.\textsuperscript{62} There are two binding sites at the P-loop as shown by Figure 1.5.1. It is thought that blasticidin S initially binds to site A and then at high concentrations it begins to bind at site B.\textsuperscript{61} At site A it is interesting to note that the end of the blasticidin S chain appears to interact with A2439 (\textit{E. coli} numbering). This residue is part of the bulge region which resides next to U2438, which when mutated to C in \textit{H. halobium} gives resistance to amicetin.\textsuperscript{28}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{The \textit{H. marismortui} peptidyl transferase centre, showing the two binding sites of blasticidin S in relation to the A- and P-loops (PDB ID: 1KC8). Site A is the major binding site and B the minor.\textsuperscript{61}}
\end{figure}
The bound X-ray crystal structures are useful, but they show the structure in a rather unnatural state. An equivalent NMR solution structure would show the structure in a more natural solution state. However, current technology does not allow an NMR solution structure of the 50S subunit to be produced, as it is too large. However, NMR solution structures of small sections of the PTC could be produced, followed by binding studies of the antibiotic to these sections. However, these sections of RNA may not hold their original conformation when removed from the structure of the ribosome, therefore comparison of the two types of structures is required.

1.5.3 Gougerotin

Gougerotin is assumed to bind in a very similar way to blasticidin S, although there have not been many studies conducted to verify this; however based on its similar mode of function and structure, it is likely that binding will occur via a similar mechanism.\(^{63}\) One relatively recent study into the binding of gougerotin tried to find important sites of interaction with the PTC of the 23S rRNA.\(^{64}\) It did this via a cross-linking study, measuring whether the presence of the antibiotic effected the rate of formation of cross-links being formed between tRNA and several specific sites. However, this study did not include the P-loop (the presumed major binding site of gougerotin), or the region around U2438 (the predicted binding site of amicetin and possibly a site of interaction of gougerotin). The results indicated that for one of the sites, U2506, gougerotin actually increased the rate of cross-linking; for the other sites there was no significant change.\(^{64}\) The fact that the addition of gougerotin had an effect is evidence that it could be binding close to this site. In fact, this residue is located very near to the site of binding of blasticidin S and to both the A and P-loop as well as the nascent peptide tunnel.
1.6 The predicted amicetin binding site motif

The predicted site of amicetin binding is shown in Figure 1.3.3. The binding site comprises of a section of duplex rRNA with two bulges and is located just off the central domain V circle. This motif was chosen due to its high degree of secondary structure and sequence conservation, as seen in Figure 1.3.9, but mainly due to the presence of U2438 which in *H. h*. organisms spontaneously mutates to a C to give resistance to amicetin. However, it has been noted in another study into a group of different peptidyl transferase antibiotics that 50% of mutations that give rise to resistance occur at >6Å from the affected bound drug. But this was still determined to be a good place to start studying the binding as the other 50% of mutations occur <6Å away from the affected bound drug and at the very least the motif is likely to be of importance to protein synthesis due to its high degree of conservation between species. Additionally, once the NMR structure of the motif is produced the mutant can be studied and compared to the wild-type, to see how the mutation could affect amicetin binding. The motif also contains a number of sites that when mutated lead to inhibition of ribosome function, which suggests that it is an important motif in ribosome function. Work has previously been done on the equivalent *H. h.* sequence, this work will also be extended in this project.

1.6.1 Comparison of the X-ray crystal structures for the motif

As we have already seen, there is a great deal of sequence and structural conservation throughout the PTCs of different species. Figure 1.6.1 shows a comparison of the secondary structures of the amicetin binding motif taken from three X-ray crystal structures. While some differences exist between the sequences and secondary structures, the vast majority of the structure is conserved. This similarity is also conserved...
in the tertiary structures as can be seen in Figure 1.6.2. In all of the structures the A2439 and U2068 bulge residues flip out of the helix and the C2440 bulge residue is folded into the helix. The only major difference is between how A2430 and U2068 base pair, *D. radiodurans* and *H. marismortui* form a standard W-C base pair, whereas *E. coli* forms an unusual base pair.

**Figure 1.6.1:** The secondary structure of the amicetin binding site in the ribosomes of *E. coli* (2AW4, *left*), *D. radiodurans* (1NKW, *centre*) and *H. marismortui* (3CC2, *right*) as indicated by their respective crystal structures. The single and double solid lines represent standard A-U and G-C W-C base pairs respectively. A dashed line between residues indicates that a non-canonical base pair is formed and a dot indicates a G-U wobble base pair. The dotted lines connect consecutive bases which cannot be placed directly next to each other due to the secondary structure. 

**Figure 1.6.2:** The X-ray crystal structure of the amicetin binding site in the ribosomes of *E. coli* (2AW4), *D. radiodurans* (1NKW) and *H. marismortui* (3CC2) as seen in their respective ribosome crystal structures. Some important residues are numbered. In the *E. coli* structure a non-canonical base pair can be observed between the A 2430 and U 2068, in contrast to the equivalent canonical base pair observed in the other two structures.
Figure 1.6.3 shows the relationship between the unbound *H. marismortui* amicetin binding motif and the blasticidin S bound *H. marismortui* amicetin binding motif, taken from bound and unbound X-ray crystal structures of the large ribosomal subunit. As can be seen, there is very little difference in the RNA conformation, but the guanidine tail of blasticidin S can be seen to be interacting with A2439. The adjacent base is U2438, which in the equivalent *H. h.* sequence, undergoes spontaneous mutation to C when the organism is exposed to sub-inhibiting quantities of amicetin, resulting in amicetin resistance. This suggests that this motif could be important to the binding of all three antibiotics.

![Figure 1.6.3](image)

**Figure 1.6.3:** A comparison of the amicetin binding site of the *H. marismortui* (3CC2) unbound crystal structure (*left image*), and *H. marismortui* (1KC8) blasticidin S bound crystal structure (*right image*). The main point of interaction appears to be with the guanidine tail of blasticidin S and the highly conserved A2474 (*A2439* *E. coli* numbering). It can be noted that there is very little conformational change observed upon binding to blasticidin S in this region.
1.7 Principles of NMR

1.7.1 Basic principles

Nuclear magnetic resonance spectroscopy is a very powerful spectroscopic technique with a wide range of uses, for example determining a compound's chemical structure, measuring dynamic effects, three dimensional structure determination, measurement of binding constants and more. A number of the important aspects of NMR will now be discussed.

Chemical shift

Chemical shift (δ) is one of the main reasons why NMR is such a useful technique. The frequency at which a certain nuclei resonates is down to the local magnetic field at that nuclei. Nuclei can be shielded, or deshielded from the overall magnetic field applied, and this occurs due to its chemical environment.

The frequency of precession, and thus of the frequency of the resulting peak in a spectrum, is dependent on the strength of the magnetic field, which is dependent on the spectrometer, Equation 1.7.1 is used to convert the value to parts per million (ppm), which in turn is independent of field strength. Where δ_{ppm} is the chemical shift value in ppm, ν₁ is the chemical shift value in Hz and ν₀ is a standard reference frequency. This allows spectra from spectrometers of varying field strength to be more easily compared.

\[
\delta_{ppm} = \frac{(ν₁ - ν₀) × 10^6}{ν₀} \quad \text{Equation: 1.7.1}
\]
Line intensity

Typically in proton NMR, the line intensities of the peaks are proportionate to the ratio of protons in the molecule. Hence if there were three proton resonances with an intensity ratio of 1:3:1, there would be three groups of protons in distinct chemical environments with a ratio of protons 1:3:1. However, we cannot say how many protons are in the group unless we know the total number of protons in the molecule from another source.

Scalar coupling

Scalar coupling ($J$) is the interaction of two spins through chemical bonds. If we take the example of two spin 1/2 nuclei, A and X, which are connected together by a number of bonds, X can exist in either the $\alpha$ or $\beta$ state. If it is in the $\alpha$ state its magnetic moment is aligned with the field, and this effect is communicated by the electrons of the bond which causes a difference in the local magnetic field at A, therefore shifting the resonance frequency of A. X however, can also exist in the $\beta$ state. In this case the same but opposite effect is felt by A causing its resonance frequency to shift by the same amount in the opposite direction.

The extent of splitting is known as the coupling constant and is measured in Hz. The coupling between A and X would be represented by $nJ_{AX}$, where $n$ is the number of bonds separating A and X and $J$ is the coupling constant. The coupling constant is independent of field strength as it is caused by the magnetic moment of the two coupled nuclei, which is independent of field. Three bond proton-proton coupling constants ($^3J_{HH}$) can be used to predict the dihedral angle between the protons using the Karplus equation, which is stated in Equation 1.7.2 where $^3J$ is the three bond coupling constant, $\theta$ is the dihedral angle.
between the coupled protons and X, Y and Z are constants dependent on the bond types present. An example of a resulting Karplus curve is illustrated in Figure 1.7.1.

\[ 3J = (X \cos^2 \theta - Y \cos \theta + Z) \]  

**Equation : 1.7.2**

![Figure 1.7.1: An image of the Karplus curve indicating graphically the relationship between coupling constant \( J \) and dihedral angle (\( \theta \)).](image)

### 1.7.2 Longitudinal and transverse relaxation

Relaxation is the process of the bulk magnetisation returning to equilibrium after it has been disturbed. There are two main types of relaxation process, longitudinal relaxation and transverse relaxation.

**Longitudinal relaxation**

Longitudinal relaxation is the recovery of the magnetisation along the +z-axis after it has been perturbed, this process is illustrated in Figure 1.7.2. Once the bulk magnetisation has been moved away from the +z-axis, more nuclei will be in a higher energy state. This energy needs to be lost to allow the bulk magnetisation to return to its equilibrium position.
at the +z-axis. Stimulated emission is the main mechanism of relaxation; to cause
relaxation by this mechanism, magnetic fields oscillating at the Larmor frequency are
required. These can occur via molecular motions and these therefore induce relaxation.
However, these magnetic fields oscillating at the Larmor frequency are quite rare leading
to long longitudinal relaxation times ($T_1$) of 0.5 to 5 seconds for medium sized organic
molecules. Longitudinal relaxation times, are in fact time constants, and can be measured
(see section 2.3.2).

Figure 1.7.2: The figure illustrates longitudinal relaxation, when the bulk magnetisation (red arrow) starts off
in the x-y plane with no component in z-magnetisation. As it relaxes back to the z-axis the component in the
y- dimension (horizontal blue arrow) reduces and the component along the z-axis increases (vertical blue
arrow) until the y component is reduced to zero and all the magnetisation lies along the z-axis and that the
magnetisation has fully returned to equilibrium.

Transverse relaxation

Transverse relaxation is the process by which magnetisation in the x-y plane is lost. The
way this occurs is by groups of like spins experiencing slightly different local magnetic
fields. Some spins will experience a larger magnetic field and so will precess slightly
faster whilst others will experience slightly weaker fields and precess at a slightly slower
rate. The signal will slowly disperse until there is no net magnetisation present in the x-y
plane, this is illustrated in Figure 1.7.3. There are two causes of this; the first is
inhomogeneity in applied magnetic field, which is usually due to imperfections in the
magnetic field, and the second is intra and intermolecular interactions. The transverse
relaxation resulting from both of these mechanisms can be described by the time constant
$T_2^*$. The time constant due to the inhomogeneity in the magnetic field can be described by
the time constant $T_{2(\Delta B_0)}$. The time constant due to the intra and intermolecular interactions can be described by the time constant $T_2$. Equation 1.7.3 shows the relationship between the different transverse relaxation time constants.

\[
\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_{2(\Delta B_0)}} \quad \text{Equation : 1.7.3}
\]

The linewdths of the NMR resonances are inversely proportional to $T_2^*$, with long $T_2^*$ resulting in sharp peaks and short $T_2^*$'s resulting in broad peaks. This is because the shorter the $T_2^*$ the faster the dephasing of the resonance will be, which means the greater the range of frequencies which leads to a broader peak.

Figure 1.7.3: The figure illustrates transverse relaxation. The top series of axes show how the magnetisation can start to fan out as the separate nuclei precess at different rates as they experience slightly different magnetic fields (green arrows). The lower series of axes show what happens to the bulk magnetisation in the x-y plane as the magnetisation fans out, eventually leading to no bulk magnetisation in the x-y plane (red arrows).

The time constant $T_1$ can never be shorter than $T_2$ as once the magnetisation has returned to the z-axis there can, by definition, be no component along the x-y plane. However, there can be no magnetisation in the x-y plane and yet the z magnetisation may not yet have completely recovered back to its equilibrium magnitude.
1.7.3 Nuclear Overhauser Effect (NOE)

The nuclear Overhauser effect (NOE) is defined as the change in intensity to a spin's resonance, which occurs when the equilibrium population of a second spin, which is dipolar coupled to the first, is perturbed. Dipolar coupling is coupling between spins which occurs directly through space, as opposed to scalar coupling which is mediated through bonds.

If the example is taken of a two spin homonuclear system I and S, where I and S are dipolar coupled to each other and are both spin 1/2 nuclei, we can construct an energy level diagram such as the one drawn in the top left of Figure 1.7.4. We can approximate that the populations (N) of the αβ and βα states are identical (and equal to N) as the difference in chemical shift relative to the Larmor frequency will be negligible. There will be an excess of spins in the αα state (N+Δ) and a deficit in the ββ state (N-Δ). When the resonance frequency of S is saturated causing the population difference between the S α and β states to be zero, this results in the situation shown in top right of Figure 1.7.4. After the excitation there are a number of transitions available which result in the reinstatement of equilibrium at S. The outcome depends on the relative probabilities of theses transitions, which are illustrated in the lower section of Figure 1.7.4. There is the probability of a single quantum transitions (W₁), the probability of a zero quantum transition (W₀) and the probability of a double quantum transition (W₂).
Figure 1.7.4: The figure illustrates the NOE effect. The dark blue boxes represent the magnitude of populations (N) of each of the four quantum states, the difference between the populations has been exaggerated for clarity. The top left scheme illustrates the equilibrium position of two dipolar coupled spins I and S, both with a spin quantum number of a 1/2. The top right scheme illustrates excitation of S to saturation so that the \( \alpha \) and \( \beta \) states of S have equal population. The light blue indicates increased population of a state, the white box indicating a reduction in population of that state. The bottom scheme illustrates the possible relaxation processes that occur to reinstate equilibrium at S but which may also affect the population difference between the \( \alpha \) and \( \beta \) states of I, where \( W_1 \) is the probability of a single quantum transition, \( W_2 \) is the probability of a double quantum transition, \( W_0 \) is the probability of a zero quantum transition and \( \Delta \) is a set difference in the population of a state.

All these transitions are dependent on stimulated emission, and so depend on the frequency of molecular motions. The \( W_1 \) quantum transition probability has no effect on I but it increases the rate of relaxation of S, so that if its probability is too high an NOE may not be observed. This is simply because the system may fully relax before the zero and double quantum transitions could have any significant effect. The single quantum transition is stimulated by oscillating magnetic fields near the Larmor frequency.
When the $W_0$ quantum transition probability is high, this leads to a negative NOE. This occurs as $\beta\alpha$ goes to $\alpha\beta$ which reinstates the population difference between the S $\alpha$ and $\beta$. However, this transition also causes a decrease in the population difference between the $\alpha$ and $\beta$ states of I and therefore a reduction in its resonance intensity. The zero quantum transition is stimulated by oscillating magnetic fields at low frequencies, equivalent to the difference in chemical shift between I and S. The zero quantum transition is therefore more likely for larger slow moving molecules.

When the $W_2$ quantum transition probability is high, this leads to a positive NOE. This occurs as $\beta\beta$ goes to $\alpha\alpha$, which again helps reinstat the population difference between the S $\alpha$ and $\beta$. This transition however, causes an increase in the population difference between the $\alpha$ and $\beta$ states of I, and therefore an increase in its resonance intensity. The double quantum transition is stimulated by oscillating magnetic fields at high frequencies, equivalent to approximately double the Larmor frequency. The double quantum transition is therefore more likely for smaller, faster moving molecules. There is a distance dependence on the NOE intensity as shown in Equation 1.7.4, where $I$ is the NOE intensity, and the $r$ is the distance between the I and S nuclei. This useful feature enables distance constraints to be estimated between nuclei thus enabling structure determination.

$$I \propto \frac{1}{r^6} \quad \text{Equation : 1.7.4}$$
1.7.4 One dimensional NMR

The 1D experiment is composed of two sections; the first section being the preparation and the second being detection, as illustrated in Figure 1.7.5. During preparation the bulk magnetisation of the sample is perturbed by a pulse or group of pulses dependent on the purpose of the experiment in preparation for detection. During detection the FID is detected and recorded. The FID is subsequently Fourier transformed from time domain to frequency domain data to form the spectrum.

Figure 1.7.5: A generic 1D pulse sequence illustrating the two sections of a normal 1D pulse sequence, the preparation and the detection period ($t_2$). The FID is also illustrated.

The standard 1D experiment consists of a delay period to allow the magnetisation to return to equilibrium, followed by a 90º pulse to rotate the magnetisation by 90º fully into the x-y plane. The pulse typically last around 10 μs for proton NMR. After the pulse comes the detection period, where the magnetisation is allowed to freely precess and the FID is measured.

1.7.5 Multi dimensional NMR

In multi dimensional NMR there are at least two dimensions, as in 1D NMR the first comes from the detection period. If the example of 2D NMR is taken, there are four periods in the pulse sequence, first a preparation time analogous to that of 1D NMR, an
evolution period, a mixing period and finally a detection time, again analogous to that of 1D NMR, see Figure 1.7.6. The mixing period contains a number of pulses and delays dependent on the purpose of the experiment. The evolution period is an incremental delay time where the second dimension of the spectrum originates from. The pulse sequence is repeated many times with the evolution period being increased incrementally. The data is Fourier transformed with respect to the detection period, and then with respect to the evolution time to produce the 2D spectrum. Any higher dimensioned NMR spectrum is analogous to the 2D spectrum, i.e. for every extra dimension there is an additional evolution and mixing period inserted just before detection.

**Figure 1.7.6:** A generic 2D pulse sequence illustrating the four sections of a 2D pulse sequence, the preparation, the evolution ($t_1$), the mixing period and the detection ($t_2$) time intervals. The FID is also illustrated.
1.8 Molecular modelling

Molecular modelling can involve a number of methods, from making physical models to computational calculations. It is used for a wide range of purposes including to find the energy of conformations of molecules/systems and in turn to find the structures with the lowest energies. It can also be used to simulate molecular motions, and in more complicated examples is used to calculate physical information about that structure, such as its NMR, ultra-violet and infra-red spectra. There are two major classes of molecular modelling calculations: *Ab initio* type calculations and molecular mechanics. Molecular mechanics is based mainly on classical mechanics. *Ab initio* techniques are based on the more accurate quantum mechanical theory, and require much more CPU time.

1.8.1 *Ab initio* calculations

These calculations try to predict the properties of the molecules from first principles, as opposed to the empirical techniques of molecular mechanics. There are a number of classes of these types of calculation, such as Hartree-Fock, Møller-Plesset, and density functional.

The Hartree-Fock method is an approximate method for solving the Schrödinger equation, using the variation theorem. This theorem states that energy calculated from an approximation to the true wavefunction will always be higher than the true value. Therefore, the lower the energy of the calculated wavefunction, the better the wavefunction will be.
The Møller-Plesset method is based on the Hartree-Fock method, with an additional correction. This correction is known as the Møller-Plesset correlation energy correction and acts to take electron correlation into account, using perturbation theory. If the correction is curtailed at the second order, the calculation is known as MP2, at the third MP3 and so on.\textsuperscript{72}

Density functional theory (DFT) calculations are now preferred to the Hartree-Fock and MP2 methods, and employ functionals of the electron density to evaluate exchange and electron correlation contributions to the electronic energy.\textsuperscript{73,74}

\subsection{1.8.2 Molecular mechanics}

Molecular mechanics, as opposed to \textit{ab initio} calculations, is based on equations which are related to classical mechanics and modified to obtain results which get as close as possible to experimental data. There are two main components to a molecular mechanics calculation: the force field, and the parameter set. The force field consists of a number of equations designed to define a set of energy terms; together these energy terms form the empirical energy. The parameter set is a list of parameters to be used in those terms for a given atom type. Together the force field and the parameter set are designed to reproduce certain properties of the molecule, such as molecular geometry, usually for a specific set of molecules, such as nucleic acids. Equation 1.8.1. shows a standard equation to calculate the empirical energy of a molecule, where \( E_{\text{empirical}} \) is the empirical energy of the system and \( E_N \) is the empirical energy of the system due to \( N \). For each energy term listed, there will be a relatively simple equation within the force field to calculate that particular energy.

\begin{equation}
E_{\text{empirical}} = E_{\text{bond}} + E_{\text{bond angle}} + E_{\text{improper}} + E_{\text{Dihedral}} + E_{\text{vdW}} + E_{\text{electrostastics}}
\end{equation}
1.8.3 Energy minimisation

This is a method for finding a minimum energy conformation. Energy minimisation calculations can be based on either quantum mechanical principles or on molecular mechanics. A molecule has $3N-6$ degrees of freedom, where $N$ is the number of atoms, during an energy minimisation calculation. These degrees of freedom are all varied in order to find the energy minima. This variation need not bear any resemblance to how the starting conformation may naturally change \textit{(in vitro)} to reach the energy minima. There are a number of energy minimisation algorithms; these are optimised for different purposes.

1.8.4 Molecular dynamics

Molecular dynamics is a technique which simulates the movement of a molecule given a certain amount of energy. Molecular dynamic simulations can also be based on either quantum mechanical principles or on molecular mechanics. The general premise is that every atom of the system is initially given a velocity. These velocities are usually based on the temperature that the calculation is given. Using the initial atom coordinates, the various forces which are acting on each atom are calculated. The computer programme then goes on to calculate new positions and velocities of the atoms after a short time period. These positions and velocities are then used to calculate the next set of velocities and coordinates after the next time period. A calculation usually has many thousands of these steps.
A slight variation on molecular dynamics is restrained molecular dynamics, where constraints are added to the calculation. For example if a distance constraint is added, the atoms in question would receive an energy penalty if they moved outside the set bounds, and so the movement would be restricted.

1.8.5 Simulated annealing

This is a molecular dynamics technique, which simulates the heating up of a molecular system and then the slow cooling to encourage the molecules to find a low energy stable conformation see Figure 1.8.1. It usually takes the form of a series of high temperature dynamic simulations followed by a cooling molecular dynamic simulation, where the simulated temperature is dropped slowly over the calculation. This is usually followed by an energy minimisation.

![Figure 1.8.1: The process of simulated annealing involves the application of a high temperature at the start of a molecular dynamic calculation to reduce the possibility of the system becoming trapped in a local minimum, and to allow it to explore a wide conformational range. The system is then slowly cooled to encourage it to a low energy conformation.](image-url)
1.9 NMR structure determination

NMR structure determination is a technique that uses NMR to find geometrical constraints for a molecular system and then applies these constraints during a series of molecular mechanics calculations to produce a three dimensional model of the system. These calculations typically include restrained simulated annealing calculations.

The example of the NMR structure determination of a particular UUCG RNA tetra loop is described here. Its secondary structure is shown in Figure 1.9.1. It is a uniformly $^{13}$C and $^{15}$N labelled 14-mer RNA motif. It has been assigned by NMR using a range of 2D and 3D homonuclear and heteronuclear techniques. Using this assignment, distance constraints were then generated using the NOE effect and they were also set to constrain base pairs. Dihedral angle constraints were generated by measuring certain coupling constants.

![Figure 1.9.1: The left image shows the secondary structure of the uniformly $^{13}$C and $^{15}$N labelled 14-mer RNA motif. The single and double solid lines represent standard W-C A-U and C-G base pairs, respectively, dotted lines represent non-canonical base pairs. The right image shows the NMR structure of the 14-mer RNA motif (2KOC).](image)

The final structure obtained is shown in Figure 1.9.2. An overall root mean squared deviation (RMSD) of 0.37 Å was obtained. This exceptionally low RMSD is due partially to wide range of techniques applied to gain constraints and the fact that the RNA is also a small hair-pin motif, both of these factors help to gain a good NMR structure. A similar procedure can be carried out for other RNA motifs, and other molecules to obtain their solution structure via NMR.
1.10 Background and earlier work

A large amount of work has previously been carried out on this project. Initially an unconstrained molecular modelling study into the binding of amicetin to the *H. h.* 29-mer amicetin binding motif was carried out.\textsuperscript{60}

The assignment and NMR structure determination of amicetin has also been determined. A recalculation of this structure is shown in section 3.3 using a greater number of constraints.\textsuperscript{32} An initial assignment and NMR structure determination of the *H. h.* 29-mer amicetin binding motif was carried out. This was corrected, expanded and completed as part of this project and it is shown in section 4.2.\textsuperscript{66}

The binding of amicetin to the *H. h.* 29-mer has also been characterised and the bound NMR solution structure of the amicetin determined. The binding constant was calculated to be 0.43 mM\textsuperscript{-1} and the dissociation constant to be 5.17 mM\textsuperscript{-1}, which suggested weak binding.\textsuperscript{66} The results of the NMR titration will be reinterpreted in section 5.1 based on the new corrected and expanded assignment.\textsuperscript{60,66}

The most recent piece of work carried out previously, as part of this project was the initial assignment of the uniformly $^{13}$C $^{15}$N doubly labelled *H. h.* 37-mer RNA motif. The RNA sequence was based on the unlabelled *H. h.* 29-mer motif. Further work has been carried out on this motif as part of the current project and will be discussed in section 4.3.\textsuperscript{66}
1.11 **Aim of the Project**

The overall aim of the project is to better understand the binding and mechanism of action of the *peptidyl transferase* antibiotics, in particular the aminohexose cytosine nucleoside antibiotics, in order that this understanding may be used to combat the growing problem of antibiotic resistance. It is also hoped that through this project knowledge may also be gained about the structure and function of the ribosome.

It is proposed to apply state of the art NMR techniques to carry out the above structural investigation, including measurement at ultra high field (1 GHz).

These aims will be accomplished by determining the NMR structure of a number of aminohexose cytosine nucleoside antibiotics and by beginning to study their dynamics. Specifically these antibiotics will include blasticidin S, gougerotin and amicetin.

This will be followed by determining the NMR structure for a number of amicetin binding RNA motif, in particular the isotopically normal *E. coli* 29-mer and *H. h.* 29-mer amicetin binding RNA motifs and the uniformly doubly labelled $^{13}$C and $^{15}$N *H. h.* 37-mer amicetin binding RNA. Finally the binding of amicetin to the 29-mer amicetin binding RNA motifs will be characterised through NMR.
Chapter 2

Materials and methods

This chapter describes the materials and methods used to obtain the NMR solution structures for the antibiotics (amicetin, blasticidin S and gougerotin) and the predicted amicetin binding RNA motifs (the *E. coli* and *H. h. 29-mer* RNA motifs and the *H. h. 37-mer* RNA motif). The chapter also describes the methods used to gain information concerning the *E. coli* 29-mer - amicetin complex and the *H. h. 29-mer* - amicetin complex. Additionally methods used to acquire information on dynamics are also described.

2.1 Preparation of antibiotic and RNA samples for NMR

2.1.1 Antibiotics

**Blasticidin S**

The blasticidin S was supplied by Novabiochem (UK). Three separate NMR samples were prepared following the methodology in section 2.1.2: (i) 2.5 mM in 100% $^2$H$_2$O pH 6.2, (ii) 7.4 mM in 90% $^1$H$_2$O and 10% $^2$H$_2$O pH 4.8 and (iii) 7.3 mM in 90% $^1$H$_2$O and 10% $^2$H$_2$O pH 7.3.

**Gougerotin**

The gougerotin was supplied by Sigma-Aldrich (UK). One NMR sample was prepared following the methodology in section 2.1.2: (i) 1.5 mM in 90% $^1$H$_2$O and 10% $^2$H$_2$O pH 6.2.
Amicetin

The amicetin was kindly supplied by Phamacia, UK. Four NMR samples were prepared following the methodology in section 2.1.2: (i) 2.5 mM in $^2$H$_2$O pH 5.7, (ii) 1.6 mM in 90% $^1$H$_2$O and 10% $^2$H$_2$O pH 7.6 and (iii) 2.5 mM in 90% $^1$H$_2$O and 10% $^2$H$_2$O pH 4.8. The fourth sample was prepared as part of the *E. coli* 29-mer RNA motif – amicetin titration, and is described in section 2.1.6.

2.1.2 Antibiotic sample preparation

The antibiotic NMR samples were all prepared in a similar way. The antibiotics were dissolved in either 90% $^1$H$_2$O (filtered Q water) and 10% $^2$H$_2$O or 100% $^2$H$_2$O both containing phosphate buffer (20 mM PO$_4^{3-}$ and 20 mM NaCl pH 6.2). The pH was then adjusted, if necessary, with NaOH or H$_2$SO$_4$.

2.1.3 RNA

*E. coli* 29-mer RNA motif

The *E. coli* 29-mer RNA sample was chemically synthesised and HPLC purified by Metabion as two separate strands; a 14-mer (900 nmol) and a 15-mer (720 nmol). Following the methodology in section 2.1.4 two NMR samples were prepared: (i) 1 mM in 90% $^1$H$_2$O and 10% $^2$H$_2$O pH 6.2 and (ii) 1 mM 100% $^2$H$_2$O pH 6.2.
The *H. h.* 29-mer RNA motif sample was synthesised and purified via high-performance liquid chromatography (HPLC) by Oswel (UK) as two separate strands; a 14-mer and a 15-mer. Following the methodology in section 2.1.4 two NMR samples were prepared: (i) 2.4 mM 90% $^1$H$_2$O and 10% $^2$H$_2$O pH of 6.0 and (ii) 2.4 mM 100% $^2$H$_2$O pH 6.0.

$^{13}$C and $^{15}$N labelled *H. h.* 37-mer RNA motif

The *H. h.* $^{13}$C and $^{15}$N (95% enriched) labelled 37-mer RNA motif was enzymatically synthesised by *in-vitro* transcription and then purified by HPLC by BioQuantis (Germany). Following the methodology in section 2.1.4 two NMR samples were prepared in a Shigemi tube: (i) 0.2 mM 90% $^1$H$_2$O and 10% $^2$H$_2$O pH 6.4 and (ii) 0.2 mM 100% $^2$H$_2$O pH 6.4.

2.1.4 RNA Sample preparation

For RNA duplex preparation the two strands were both dissolved separately in 1 ml each of filtered Q water. The absorbance was then measured to confirm the sample concentration as described in section 2.1.5. The samples were then lyophilised, before being redissolved in 0.5 ml of phosphate buffer (20 mM PO$_4^{3-}$ and 20 mM NaCl, pH 6.2). The two samples were then mixed, trying to ensure equimolar amounts of each solution to obtain the final sample. For single stranded samples the RNA was initially dissolved in approximately 1 ml phosphate buffer (20 mM PO$_4^{3-}$ and 20 mM NaCl, pH 6.2), the absorbance was then measured to confirm the sample concentration, see section 2.1.5. Both types of samples were then lyophilised and redissolved in 60 μl of $^2$H$_2$O and 600 μl of filtered Q water (standard NMR tube) or 25 μl and 225 μl respectively (Shigemi tubes). The samples were
then annealed by heating in a heating block to ~80°C and then slowly cooling to ~2°C, this was then repeated. This produced the NMR samples in 90% $^1$H$_2$O and 10% $^2$H$_2$O and a pH of ~6.2. The samples were later lyophilised several times in 99.9% $^2$H$_2$O and placed in ~600 μl (standard tube) or 250 μl (Shigemi tube) of 100% $^2$H$_2$O to produce the NMR samples.

### 2.1.5 RNA quantification

The concentration of the RNA samples were calculated by measuring their UV absorbance. When samples were initially dissolved in filtered Q water, typically 1 ml, a small quantity, typically 5 μl, was then removed and diluted with filtered Q water to 1 ml. The absorbance of the sample was then measured at 260 nm to determine the RNA concentration.

### 2.1.6 Titration of the *E. coli* 29-mer RNA motif with amicetin

An amicetin sample was prepared ready for titration with the *E. coli* 29-mer RNA motif. A reference amicetin NMR sample was also prepared in the same way. The amicetin came from the same source as the amicetin NMR samples. The amicetin sample was prepared in 99.9% $^2$H$_2$O, with a concentration of 0.85 mM and no buffer, part was used as a reference NMR sample and the other part for titration. This sample was slowly titrated with the *E. coli* 29-mer RNA motif. When the sample volume became too large the sample was lypholyesd and redissolved in 600 μl of 100% $^2$H$_2$O. Once the titration and NMR experiments were complete the resulting sample was lypholyesd and placed in a solvent of 90% $^1$H$_2$O and 10% $^2$H$_2$O.
2.2 NMR instrumentation and experimentation

2.2.1 NMR spectrometers

The NMR data was acquired on Varian Inova 800 MHz, 600 MHz and Unity plus 500 MHz spectrometers and on Bruker Avance 1000 MHz (world’s first 1 GHz, NMR spectrometer recently installed in Lyon, France), 700 MHz, 600 MHz, 500 MHz, 400 MHz and 200 MHz spectrometers. The Varian 800 MHz and Bruker 1 GHz, 700 MHz and 600 MHz spectrometers were equipped with cryogenically cooled probes and all the spectrometers except the Bruker 500 MHz and 200 MHz were equipped with variable temperature units. The Varian 600 MHz and the Bruker 400 MHz and 200 MHz spectrometers were equipped with probes capable of detecting $^{31}\text{P}$ and $^{19}\text{F}$ nuclei. All the spectrometers except the Bruker 200 MHz spectrometer were equipped with triple resonance pulsed field gradient (PFG) probes.

2.2.2 NMR acquisition parameters

The acquisition parameters used to collect spectra were as follows. The number of scans for a 1D $^1\text{H}$ was typically 128. Commonly for 2D $^1\text{H}$-$^1\text{H}$ experiments the number of $t_1$ increments was 800, the number of $t_2$ points measured was 4000 and the number of scans was 32. Generally for 2D $^1\text{H}$-$^{13}\text{C}$/-$^{15}\text{N}$ experiments the number of $t_1$ increments was 128 ($^{13}\text{C}$/-$^{15}\text{N}$), the number of $t_2$ points measured was 2000 ($^1\text{H}$) and the number of scans was around 32, for the $^{13}\text{C}$ and $^{15}\text{N}$ labelled sample and around 256 for the isotopically normal sample. Typically for 2D $^1\text{H}$-$^{31}\text{P}$ experiments the number of $t_1$ increments was 256 ($^{31}\text{P}$), the number of $t_2$ points measured was 128 ($^1\text{H}$) and the number of scans was 128. Generally for 3D experiments involving $^1\text{H}$ and $^{13}\text{C}$ nuclei the number of $t_1$ increments was
around 256 ($^1\text{H}$) and the number of $t_2$ increments was around 128 ($^{13}\text{C}$), the number of $t_3$ points measured was around 1280 ($^1\text{H}$) and the typical number of scans was between 4 and 8. Typical for 3D experiments involving $^1\text{H}$, $^{13}\text{C}$ and $^{31}\text{P}$ nuclei the number of $t_1$ increments was 96 ($^{13}\text{C}$) and the number of $t_2$ increments was 64 ($^{31}\text{P}$), the number of $t_3$ points measured for was 1000 ($^1\text{H}$) and the number of scans was between 4 and 8.

Both WATERGATE and presaturation water suppression techniques were each tested and then the most one effective was used.\textsuperscript{77,78} For samples in $^2\text{H}_2\text{O}$, the presaturation technique was favoured and for samples in $^1\text{H}_2\text{O}$, WATERGATE was favoured. The typical proton spectral width for the $^2\text{H}_2\text{O}$ samples was 10 ppm; for $^1\text{H}_2\text{O}$ samples this was increased 24 ppm, to allow low field exchangeable protons to be observed. The typical $^{15}\text{N}$ spectral width for $^1\text{H}_2\text{O}$ samples was 100 ppm. The typical $^{13}\text{C}$ spectral width for the $^2\text{H}_2\text{O}$ and $^1\text{H}_2\text{O}$ samples was 120 ppm. The typical $^{31}\text{P}$ spectral width for the $^2\text{H}_2\text{O}$ samples was around 8 ppm in 2D and 3D spectra, in 1D spectra the $^{31}\text{P}$ spectral width was generally 400 ppm.

### 2.2.3 NMR experiments

#### Antibiotics

Several NMR experiments were used for the assignment and structural calculation of blasticidin S and gougerotin. These included a series of ROESY experiments measured at 2°C with a range of spin lock mixing times and a TOCSY at 2°C with a 70 ms spin lock mixing time. A 1D $^1\text{H}$ variable temperature series for both antibiotics were measured. For blasticidin S, the 1D $^1\text{H}$ spectrum was measured at pH 4.8 and 7.3, and a qualitative $T_1$ relaxation experiment was also carried out.
For amicetin all the experiments for the assignment and structure determination had been performed previously by C. Shammas. These were reinterpreted to check the assignment and correct the original NMR solution structure.

**E. coli 29-mer RNA motif**

A large number of NMR experiments have been carried out on the *E. coli* 29-mer RNA motif. For the sample in $^1\text{H}_2\text{O}$, a 1D $^1\text{H}$ variable temperature series from 25°C to 2°C, two TOCSY spectra at 2°C with 75 ms and 50 ms mixing times, and four NOESY spectra at 2°C with a range of mixing time were carried out. For the sample in 100% $^2\text{H}_2\text{O}$ a 1D $^1\text{H}$, four NOESYs with a range of mixing times, a TOCSY ($\text{SL}_{\text{mix}}=75$ ms), and a DQF-COSY were carried out. Also for this sample a number of heteronuclear experiments were performed, this includes two $^1\text{H}$-$^13\text{C}$ HSQC experiments, a $^1\text{H}$-$^13\text{C}$ HSQC-TOCSY ($\text{SL}_{\text{mix}}=75$ ms) and a $^1\text{H}$-$^13\text{C}$ HSQC-NOESY ($\tau_{\text{m}}=250$ ms).

A 1D phosphorus experiment was carried out on the 200 MHz (81 MHz $^{31}\text{P}$ frequency) spectrometer. Also a $^1\text{H}$-$^{31}\text{P}$ CPMG-HSQC-NOESY ($\tau_{\text{m}}=500$ ms) at 600 MHz ($^1\text{H}$ frequency).

**E. coli 29-mer RNA motif - amicetin complex**

The *E. coli* 29-mer RNA sample was titrated with a designated amicetin titration sample in $^2\text{H}_2\text{O}$. 1D $^1\text{H}$ spectra were collected at 25°C, after every addition of amicetin, and short TOCSY spectra ($\text{SL}_{\text{mix}}=75$ ms) and NOESY spectra were also collected at various points. This was done to allow observation of any RNA-antibiotic intermolecular NOEs, and to monitor any changes in NOE intensity patterns and any changes in chemical shifts.
certain points of the titration the NMR experiments were also carried out at 5°C, to see if the lower temperature would induce stronger binding. Also at selected points along the titration 1D $^{31}$P spectra were collected to observe amicetin's effect on the phosphate back bone of the RNA, via differences in the $^{31}$P resonances. After the titration was completed the sample was lyophilised and placed in 90% $^1$H$_2$O and 10% $^2$H$_2$O, to observe any difference in the exchangeable proton resonances with a 1D $^1$H spectrum and 2D NOESY at 2°C.

**H. h 29-mer RNA motif**

For the *H. h* 29-mer RNA motif all the experiments for the assignment and structure determination had been performed previously by C. Shammas. These were reinterpreted to correct the assignment and subsequently the NMR solution structure was recalculated. A titration of amicetin with the *H. h* 29-mer RNA motif in $^1$H$_2$O was also previously performed with amicetin from 0.1 to 5.0 molar equivalents, this was reinterpreted based on the reassignment.

**$^{12}$C and $^{15}$N labelled *H. h* 37-mer RNA motif**

For the *H. h* 37-mer RNA motif some of the experiments for the assignment had been previously carried out by C. Shammas. Further experiments were carried out in order to check, correct and expand the previous assignment, and achieve NMR constraints for structure determination. For the sample in $^1$H$_2$O, a 1D $^1$H, a $^1$H-$^{13}$C HSQC and a $^1$H-$^{15}$N HSQC-NOESY ($\tau_m$=50 ms) were carried out at 2°C. For the sample in 100% $^2$H$_2$O all the experiments were run at 25°C, these included a 1D $^1$H spectrum a $^1$H-$^{13}$C HSQC, a number of 2D $^1$H-$^{13}$C HSQC-NOESY spectra at various mixing times, a number of 2D $^1$H-$^{13}$C
HSQC-TOCSY spectra at various spin lock mixing times, two 3D $^1$H-$^{13}$C HSQC-NOESY spectra at 100 ms and 300 ms mixing time, a 3D NOESY-HSQC ($\tau_m$=300 ms), three 3D $^1$H-$^{13}$C HSQC-TOCSY spectra with various spin lock mixing times, and a 3D HCP.

2.2.4 NMR processing and analysis

The 1D NMR spectra were processed using SpinWorks. The programme provided quick and easy NMR data processing. Typically a Lorentzian apodisation function was used, the line broadening was adjusted depending on how noisy the spectrum was.

The 2D NMR experiments were processed using NMRPipe, which is a UNIX based command line program, and comes with a graphical user interface called NMRDraw, to allow easier processing and conversion. Initial processing was based on a typical conversion script and then modified, to correct phasing, remove water suppression if required and also removing the baseline correction, if required. A standard data processing script is included in Appendix A.1. Typically, 2D data processing included zero filling, a Gauss apodisation function in the first dimension and a zero filling and an adjustable sine apodisation function in the second. For 3D processing typically the first two dimensions are processed similarly to the two dimensions of 2D spectra and the third is typically processed with zero filling and an adjustable sine apodisation function.

The processed data was then analysed using Sparky (after conversion of the data to the correct format) for the 2D antibiotic data and CcpNMR Analysis for the RNA 2D and 3D data. Sparky and CcpNMR Analysis are graphical programs for the analysis of NMR spectra, they are particularly aimed at biomolecular polymers such as nucleic acids, but are also suitable for small molecules.
2.3 NMR techniques

2.3.1 1D homonuclear and heteronuclear experiments with decoupling

These are relatively simple experiments used to record 1D spectra of NMR active nuclei. A typical pulse sequence used for a 1D experiment is shown in Figure 2.3.1. A decoupling pulse can be used to remove the complicated multiplicity patterns that can occur due to heteronuclear coupling with other nuclei. A typical pulse sequence used for such an experiment is shown in Figure 2.3.2.

**Figure 2.3.1**: A standard homonuclear (proton) 1D pulse sequence, where the A nuclei is detected. The preparation and detection time periods are marked below the pulse sequence.

**Figure 2.3.2**: A decoupled heteronuclear (X nuclei) 1D pulse sequence, where the X nuclei is detected using the upper part of the pulse sequence and A is the nuclei which is being decoupled using the lower part of the pulse sequence. The preparation and detection time periods are marked below the pulse sequences.
The decoupling pulse works, for example in spin 1/2 nuclei, by flipping the nuclei rapidly between the $\alpha$ and $\beta$ states. If this flipping is fast, relative to the magnitude of coupling constant, $J$, the multiplicity pattern collapses into a singlet. Composite pulse decoupling (CPD), is a group of decoupling pulse techniques which are now used as standard, they include such sequences such as MLEV-16 and WALTZ-16.\textsuperscript{83,84}

2.3.2 Longitudinal, $T_1$, relaxation measurement

A quick measurement of $T_1$ can be made using the inversion recovery pulse sequence, (Figure 2.3.3) a series of 1D spectra are collected with different incremental time delays.

![Figure 2.3.3: The $T_1$ measurement pulse sequence. The preparation, incremental delay ($\tau$) and detection ($t_2$) time are marked on the pulse sequence.](image)

The incremental delay, $\tau$, is initially set to zero and the resulting spectrum is phased to be negative absorption. The delay period, $\tau$, is increased slowly on the scale of ms up to a few seconds (for protons) depending on the size of the molecule. From inspection of the spectra a value of the delay time which gives a null intensity for a particular resonance can be estimated, $\tau_{\text{null}}$, which can then be used in Equation 2.3.1 in order to estimate a relative $T_1$ value for the given nuclei.

$$T_1 = \frac{\tau_{\text{null}}}{\ln 2} \quad \text{Equation 2.3.1}$$
2.3.3 Double quantum filtered correlation spectroscopy (DQF-COSY)\textsuperscript{85,86}

The DQF-COSY experiment is a 2D NMR technique where cross peaks arise between protons which are directly scalar coupled, usually \textit{via} three bond coupling. The basic DQF-COSY pulse sequence is shown in Figure 2.3.4.

![Figure 2.3.4: The DQF-COSY pulse sequence. The preparation, evolution, mixing period (including a fixed $\tau_0$ delay (usually a few micro seconds) and detection time are marked on the pulse sequence.](image)

The DQF-COSY experiment is based on the simpler correlation spectroscopy (COSY) experiment. The DQF-COSY experiment enables cross peaks much closer to the diagonal to be observed more easily. However, the DQF-COSY is much less sensitive compared to the COSY experiment and thus takes much longer to achieve a similar level of signal to noise. The DQF-COSY experiment can also be used to determine coupling constants, the splitting between the peaks theoretically corresponds to the $^3J$ value. However, when the line-width of the peak approaches or is greater than the coupling constant, the positive and negative components of the peak can cancel each other out, leading to reduced intensity and apparent increase in the coupling constant. This effect is illustrated in Figure 2.3.5. Thus in most cases the coupling constant can only be approximated.\textsuperscript{1,2}
Figure 2.3.5: The figure illustrates the problems of linewidths which are larger than or equal to the coupling constant in the DQF-COSY experiment. As the linewidth increases (black trace-low linewidth to green trace-high line-width), the intensity falls as the negative and positive contributions of the peak cancel each other out. Along with this a more problematic effect occurs, the most intense points of the peaks move further apart leading to an apparent increase in the coupling constant.

2.3.4 Total correlation spectroscopy (TOCSY)\textsuperscript{87,88}

The total correlation spectroscopy experiment (TOCSY) is a 2D NMR technique where cross peaks arise between protons from the same spin system. An example of a basic TOCSY pulse sequence is shown in Figure 2.3.6.

Figure 2.3.6: The TOCSY/ROESY pulse sequence. The preparation, evolution and mixing period and detection time are marked on the pulse sequence. The difference between the TOCSY and ROESY pulse sequence is contained in the nature of the spin lock pulse (\textit{SL}_{mix}).

The TOCSY experiment is performed using a spin lock pulse, of which there are several variations, but in its simplest and original form it is a long series of closely separated 180°
pulses spanning the mixing time. The spin lock pulse keeps the magnetisation within the x-y plane. The energy of the system (the Hamiltonian, $H$) (shown in Equation 2.3.2) has two major components. One of these components is due to the chemical shift ($H_\delta$) and another to the scalar coupling ($H_J$).

\[
H = H_\delta + H_J + \ldots \quad \text{Equation 2.3.2}
\]

Under a large magnetic field, $B_0$, the component resulting from the chemical shift far outweighs that of the component due to scalar coupling and so $H \approx H_\delta$. But under the influence of the spin lock pulse, the spins are effectively not under the influence of the large magnetic field, but only under the influence of the magnetic field resulting from the radio-frequency pulse. This magnetic field is much smaller and the situation is now reversed, therefore $H \approx H_J$. The reason for this is that $\Delta E$, which leads to the chemical shift value, is strongly dependent on the magnetic field strength, whereas the scalar coupling constants are independent of field strength. This now essentially means that the Larmor frequency of precession around the applied magnetic field produced by the spin lock pulse is dependent on the coupling constant and not the chemical shift. This therefore allows magnetisation transfer between coupled protons. This effect then begins to spill over and allows magnetisation transfer between all spins within a spin system.

As the magnetisation transfer between spins takes time, a longer mixing time enables magnetisation transfer throughout the spin system. For example for short mixing times the magnetisation transfer may only occur between directly coupled spins.
The nuclear Overhauser spectroscopy experiment (NOESY) is a 2D NMR technique where cross peaks arise between protons which are close (<5Å) in space. The NOESY pulse sequence is shown in Figure 2.3.7.

Figure 2.3.7: The NOESY pulse sequence. The preparation, evolution (t₁), mixing period and detection time (t₂) are marked on the pulse sequence.

This 2D technique takes advantage of the previously discussed NOE effect (section 1.7.3). During the mixing time, τₘ, magnetisation is transferred between dipolar coupled nuclear spins up to ~5Å away. The intensity of the cross peak is related to the distance via equation 1.7.6. This allows the distance between the nuclei to be approximately measured, and these distances can then be used as constraints in structure calculations. However, if the mixing time is too long, spin diffusion can become a significant problem. Spin diffusion is where magnetisation is transferred indirectly via another nuclei. This can lead to an increase in intensity of the cross peaks which in turn can cause an underestimation of the distance between the two nuclei. Spin diffusion can also cause cross peaks to be observed between spins which are >5Å apart.

Another complicating factor is that the intensity of the NOE cross peak depends on the correlation time (τₖ) of the molecule. The correlation time is approximately the time taken...
for the molecule to rotate by one radian about any axis, this can be approximated by $\tau_c = 10^{-12}$ Da, where Da is the molecular mass in Daltons. For small molecules the NOE cross peak appear negative and for large they appear positive, for medium sized molecules the NOE intensity can approach zero. An NOE intensity of zero will obviously be a problem, the NMR technique ROESY offers a solution and is discussed in the next section.

### 2.3.6 Rotating-frame Overhauser effect spectroscopy (ROESY)

The rotating-frame Overhauser spectroscopy experiment (ROESY) is a 2D NMR technique where cross peaks arise between resonances which are close together (<5Å) through space. An example of a basic ROESY pulse sequence is shown in Figure 2.3.6. The pulse sequence is very similar to that of the TOCSY experiment, but due to a difference in the spin lock pulse, magnetisation is transferred via dipolar coupling rather than scalar.

The ROESY spectrum is very similar to the NOESY spectrum, but there are a number of important differences. The rotating-frame Overhauser effect (ROE) cross peaks are always in the opposite phase to the diagonal peaks, independent of the correlation time. Peaks due to chemical exchange are in the same phase as the diagonal. These are both obvious advantages over the NOESY spectrum, particularly if the molecular mass in the medium range, which in the NOESY experiment can lead to zero intensity cross peaks. However, ROESY also has the disadvantage that it is prone to certain spurious cross peaks, which include COSY and TOCSY type cross peaks. The TOCSY peaks do appear positive and the COSY peaks appear as anti phase peaks, and so they are both easy to spot, except when a TOCSY style peak overlays with a ROESY cross peak. This will obviously lead to the sum of the two and so this can cause confusion. magnetisation is transferred via dipolar coupling rather than scalar coupling.
2.3.7 Heteronuclear single quantum correlation (HSQC)\textsuperscript{96,97,98}

The heteronuclear single quantum correlation experiment (HSQC) is a 2D NMR technique where cross peaks arise between the nuclei with a large Boltzmann population difference (usually $^1$H) and another scalar coupled (usually directly bonded) nuclei, such as $^{13}$C or $^{15}$N, with low magnetogyric ratios. An example of a basic HSQC pulse sequence is shown in Figure 2.3.8.

![HSQC Pulse Sequence](image)

**Figure 2.3.8:** The HSQC pulse sequence. The preparation, evolution ($t_1$) and mixing period and detection time ($t_2$) are marked on the pulse sequence. The fixed delay $\tau$ is equal to the reciprocal of the scalar coupling constant between the two types of nuclei, which is to be observed ($\tau = 1/J_{HX}$).

The experiment uses the insensitive nuclei enhanced by polarization transfer (INEPT) pulse sequence which is designed to enhance the signals of nuclei with low magnetogyric ratio. It does this by transferring the magnetisation from a nuclei with a large Boltzmann population difference (for example a $^1$H) to the second nuclei (for example a $^{13}$C), where the fixed delay $\tau$ of the INEPT sequence is equal to $1/J_{HC}$. The magnetisation is then transferred back from the second nuclei ($^{13}$C) to the first ($^1$H), where it is then observed.

This experiment is particularly useful for RNA as it helps divide proton resonances which occur in the same region of the proton spectrum, by their directly bonded $^{13}$C nuclei's chemical shifts, which are more disperse than the corresponding protons. This is
particularly useful for the ribose component where the bulk of the protons all occur in the same region of the $^1$H spectrum, but their attached $^{13}$C nuclei have distinct chemical shift ranges.

### 2.3.8 2D and 3D HSQC-TOCSY

The 2D and 3D HSQC-TOCSY experiments are combinations of the HSQC and TOCSY experiments as the names indicates, a basic 3D HSQC-TOCSY pulse sequence is shown in Figure 2.3.9. The first half of the pulse sequences work in the same way as a standard HSQC pulse sequence. Once the HSQC section of the pulse sequence has transferred the magnetisation back to the first nuclei (the $^1$H) the TOCSY half of the pulse sequence transfers the magnetisation on the first nuclei, to all the nuclei within its spin system. For example the 2D $^1$H-$^{13}$C HSQC-TOCSY spectrum has a $^1$H and $^{13}$C dimension, $F_2$ and $F_1$, cross peaks occur between directly bonded, scalar coupled $^1$H and $^{13}$C nuclei ($J_{CH}$), as in a straight 2D $^1$H-$^{13}$C HSQC. Since there is a TOCSY section cross peaks are also observed between $^{13}$C resonances and the $^1$H resonances in the $^{13}$C nucleus' directly scalar coupled $^1$H's spin system, subject to a long enough spin lock mixing time.

![Figure 2.3.9: A basic 3D HSQC-TOCSY pulse sequence. The preparation, evolution, mixing periods and detection times are marked on the pulse sequences, where $\tau$ is equal to the reciprocal of the scalar coupling constant between the two types of nuclei, where X is the second NMR active nuclei ($\tau = 1/J_{XX}$). Simply removing the second evolution period ($t_2$) leaves you with a 2D HSQC-TOCSY pulse sequence.](image)
The 3D $^1$H-$^{13}$C HSQC-TOCSY spectrum obviously has an additional third dimension, which arises due to a second incremented evolution time (Figure 2.3.9), resulting in a second proton dimension. In a 3D $^1$H-$^{13}$C HSQC-TOCSY spectrum, cross peaks occur at the chemical shift of a proton ($F_3$), its directly bonded $^{13}$C ($F_1$), and any resonant protons in the initial protons spin system ($F_2$), subject to a long enough spin lock mixing time. The advantage of the 3D version over the 2D version is the greater dispersion of the peaks, allowing clearer assignment. The disadvantage is the longer time required for the experiment, if a 2D type spectrum quality is required. However, the resolution of the first two dimensions can be reduced, in comparison to the 2D; due to the increased peak separation gained by the addition of the third.

2.3.9 2D and 3D HSQC-NOESY and 3D NOESY-HSQC

The 2D and 3D HSQC-NOESY and 3D NOESY-HSQC experiments are combinations of the individual HSQC and NOESY experiments, (Figure 2.3.10). The experiments function similarly to the HSQC-TOCSY experiments, except that in the case of 2D and 3D HSQC-NOESY, in the second half of the pulse sequence, magnetisation is transferred via dipolar (through space) coupling as opposed to scalar (through bond) coupling. The 3D NOESY-HSQC is similar to the 3D HSQC-NOESY except that the NOESY section of the pulse sequence precedes the HSQC, but the resulting spectrum is nearly identical. For example the 2D $^1$H-$^{13}$C HSQC-NOESY spectrum has a $^1$H and $^{13}$C dimension, $F_2$ and $F_1$, cross peaks occur between directly bonded, scalar coupled $^1$H and $^{13}$C nuclei ($J_{CH}$), similar to a straight 2D $^1$H-$^{13}$C HSQC. Additional cross peaks are also seen, due to the NOESY section of the pulse sequence, between a $^{13}$C resonance and the proton resonances which are close in space ($\approx$5Å) to the $^{13}$C nucleus' directly scalar coupled proton, subject to a long enough mixing time.
Figure 2.3.10: A basic 3D HSQC-NOESY pulse sequence. The preparation, evolution, mixing periods and detection times are marked on the pulse sequences, where $\tau$ is equal to the reciprocal of the scalar coupling constant between the two types of nuclei, where X is the second NMR active nuclei ($\tau = 1/J_{\text{HX}}$).

The 3D spectrum has an additional dimension, which arises from a second incremented evolution time (Figure 2.3.10). Taking the example of a 3D $^1$H-$^{13}$C HSQC-NOESY spectrum, peaks occur at the chemical shift of a proton ($F_3$), its directly bonded $^{13}$C ($F_1$), and any proton which is close through space ($<\sim5\text{Å}$) to the initial proton ($F_2$), given a long enough mixing time.

2.3.10 $^1$H-$^{31}$P Carr-Purcell-Meiboom-Gill-HSQC-NOESY (CPMG-HSQC-NOESY)

The $^1$H-$^{31}$P Carr-Purcell-Meiboom-Gill-HSQC-NOESY (CPMG-HSQC-NOESY) experiment is designed specifically for use with RNA. The spectrum is used to correlate the easily identifiable H1' and aromatic H6/H8 proton resonances to their intra and sequential $^{31}$P resonances. The most important peaks the spectrum produces, as they allow clear assignment, are H1'$_i$ – P$_i$, H6/8$_i$ – P$_i$, H1'$_{i+1}$ – P$_i$ and H6/8$_{i+1}$ – P$_i$, where “$i$” is the residue number.

These peaks not only allow assignment of the $^{31}$P resonances but also enables two further sets of sequential connectivities to be followed making it a very useful technique to independently confirm the proton assignment.
The experiment works by using a CPMG pulse train, a series of closely spaced 180° pulses, during the periods of magnetisation transfer, instead of the single 180° pulse used in a standard HSQC pulse sequence. The magnetisation is first transferred from the H3', H5' and H5'' protons to their closest $^{31}$P nuclei, then back to the starting H3', H5' and H5'' protons. The NOESY section that follows this transfers the magnetisation to protons less than ~5Å away though space, which includes the easily distinguishable H1' and aromatic H6/H8 protons.

The CPMG pulse train is used as broad $^{31}$P resonances, due to significant conformational change along the backbone, can lead to non-ideal refocusing of the magnetisation. The CPMG pulse is used to eliminate this effect and therefore help improve the signal to noise. Another problem overcome by using this particular form of the CPMG pulse train (XY-16 expansion scheme), is that the magnitude of some $^1$H-$^1$H couplings (~1-10.5 Hz) are in the same range of the $^1$H-$^{31}$P couplings (~2-23 Hz) of interest.\textsuperscript{102} If just a standard HSQC pulse sequence is used, this would cause distortions in the line shape.

2.3.11 HCP\textsuperscript{103,104}

The 3D HCP experiment is designed particularly for sequence specific assignment of the $^{31}$P atoms of uniformly $^{13}$C labelled RNAs. The technique transfers magnetisation via INEPT transfer steps, from the H3', H4', H5' and H5'' protons to their directly bonded carbon, and then to their respective $^{31}$P nuclei ($J_{C4',P} \sim 10$ Hz, $J_{C3'/5'/5'',P} \sim 3-6$ Hz)\textsuperscript{118}. The magnetisation is then transferred back to the protons via the carbon for detection. This produces a 3D spectrum with $^{13}$C, $^{31}$P and $^1$H dimensions. Cross peaks can be observed between H3'\textsubscript{i-1} – C3'\textsubscript{i-1} – P\textsubscript{i}, H4'\textsubscript{i-1} – C4'\textsubscript{i-1} – P\textsubscript{i}, H4'\textsubscript{i} – C4'\textsubscript{i} – P\textsubscript{i} and H5'/H5''\textsubscript{i} – C5'\textsubscript{i} – P\textsubscript{i}, where “i” is the residue number.
2.3.12 NMR titration

The NMR titration is a relatively straightforward technique. In the case of two molecules which are expected to interact with each other, one is slowly titrated with the other and the changes in their NMR signals are monitored. These changes can include:

- changes in chemical shift
- changes in the shape and linewidth of peaks
- changes in the intensities of intramolecular NOEs
- the appearance of any intermolecular NOEs

The changes in chemical shift and linewidth can be monitored simply by using a series of 1D spectra, or various 2D techniques can be utilised to track these changes more clearly. 2D experiments which are commonly used to do this are $^1$H-$^1$H TOCSY and $^1$H-$^1$H NOESY, additionally if labelled samples are available HSQC type experiments are also often used. Changes in chemical shift and line shape can used to predict the parts of the molecules that are interacting with each other, i.e. the binding site. This is because as the two molecules interact, they will alter the chemical environment of the nuclei at and around the site of binding and thus alter their chemical shifts. The relaxation properties of the two molecules, particularly around the site of binding, will also be altered and so the linewidth will be affected.

Two particular advantages of the $^1$H-$^1$H NOESY are that changes in intramolecular NOEs can be observed and that intermolecular NOEs can sometimes be seen between the two interacting molecules. Both of these can be used to give information on how the structure of the molecules has changed between their free and bound state. It may even be possible, using a NOESY spectrum, to calculate a solution structure of the two bound molecules.
The NMR titration can be interpreted to give information on whether the binding is strong, medium or weak and can even be used to calculate an equilibrium association constant, $k_a$.

For instance, in a strong binding system where the ligand is in excess of the target, two sets of sharp peaks for the ligand will usually be observed, one for the bound ligand and the other for the free ligand. For the same system, where the strength of binding is in the medium range, the two sets of ligand peaks become broader and the equivalent peaks from the bound and unbound states begin to merge together. For the same system, where the strength of binding is in the weak range, the two sets of ligand peaks completely merge into one set.

2.3.13 Variable temperature series

A variable temperature series experiment involves measuring a series of NMR spectra (usually 1D) at a number of increasing temperature points, before then returning to the original temperature to ensure the sample has not degraded. The change in line shapes and chemical shifts can be monitored to see if anything interesting is revealed about the structure. It is typically used to monitor the extent of retardation of exchangeable proton resonances in a 90% $^1$H$_2$O 10% $^2$H$_2$O sample, this can give useful information on the extent and strength of hydrogen bonding.
2.4 Assignment strategies

2.4.1 Antibiotics

The assignment strategy for the antibiotics is illustrated in Figure 2.4.1. The strategy started by identifying the exchangeable protons either by identifying the peaks which reduce markedly in intensity upon increase of temperature, or unambiguously by noting their absence in the 1D spectrum measured in $^{2}$H$_{2}$O.

<table>
<thead>
<tr>
<th>Task</th>
<th>Spectra required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification of exchangeable proton resonances</td>
<td>1D $^{1}$H $^{2}$H$_{2}$O, 25°C</td>
</tr>
<tr>
<td></td>
<td>1D $^{1}$H $^{2}$H$_{2}$O, 2°C</td>
</tr>
<tr>
<td></td>
<td>Variable temperature series</td>
</tr>
<tr>
<td>Classification of resonances into their spin systems</td>
<td>2D $^{1}$H $^{1}$H TOCSY</td>
</tr>
<tr>
<td>Assignment of resonances</td>
<td>2D $^{1}$H $^{1}$H ROESY</td>
</tr>
<tr>
<td>Confirmation of assignment</td>
<td>2D $^{1}$H $^{1}$H ROESY</td>
</tr>
</tbody>
</table>

**Figure 2.4.1:** The chart shows a brief overview of the tasks which were taken to carry out the NMR assignment of the antibiotics and the NMR spectra used to accomplish these tasks.

Subsequently, a 2D TOCSY spectrum was used to classify the resonances into their individual spin systems based on the antibiotic's chemical structure. This was done by first matching the number of resonances observed in the spin system with the number of protons in the spin systems, deduced from the chemical structure. Doing this allowed the identification of most of the larger spin systems. The next step in completing the identification of the spin systems was to identify the characteristic resonances within the spin system, such as methyl protons, which give intense high field resonances, or aromatic protons which give low field resonances. The identification of such resonances aids and confirms the assignment of the spin systems by matching the resonances to their potential nuclei on the chemical structure.
Once the majority of the spin systems had been identified, the individual resonances that they contained can be assigned much more easily as the assignment task has already been broken up into much smaller and more manageable sections. To unambiguously assign these protons, the chemical shift, line integral (intensity), coupling constants and splitting patterns (obtained from high resolution 1D spectra) were used. If assignment based on these factors was not possible the ROESY spectra were used to provide additional information on the proximity through space of the various protons. This data from the ROESY spectrum was given less weight as it can be misleading (the molecule may be in an unexpected conformation producing unexpected ROE peaks). The data from the ROESY spectra was finally used to help confirm the rest of the assignment.

2.4.2 Isotopically normal RNA

A brief overview of how the assignment of both the isotopically normal E. coli and H. h. RNAs was undertaken is shown in Figure 2.4.2. The process began with the assignment of the exchangeable protons. First, the imino NH protons were assigned, this was done by separating the observed imino proton resonances into imino protons involved in canonical W-C base pairing and those involved in unusual base pairs. The former have imino proton resonances which tend to fall into the chemical shift range of 12-15 ppm, the latter tend to fall into the chemical shift range of 10-12 ppm. The imino protons involved in W-C base pairing were separated further into guanine and uracil imino protons, the former generally fall into the chemical shift range of 12-14 ppm and the later tend to fall into the chemical shift range of 13-15 ppm.
Figure 2.4.2: The chart shows a brief overview of the tasks which were taken to carry out the NMR assignment of the isotopically normal RNA motifs and the NMR spectra used to accomplish these tasks.

Using this categorisation, the sequential NOE connectivity patterns between the imino proton resonances, in a long mixing time $^1$H-$^1$H NOESY spectrum recorded in 90% $^1$H$_2$O and 10% $^2$H$_2$O, were traced. NOE connectivities can usually be observed in standard A-form RNA between imino protons which are on adjacent residues on the same strand. Inter-strand NOE connectivities can also be observed between the imino proton of one residue and any imino proton of a residue directly connected to the residue which base pairs to the initial residue. In sections of RNA with a non-A-form conformation these connectivities may still be seen and other such connectivities as well.
The imino proton assignment was then confirmed using the imino-to-aromatic region of the spectrum, once the aromatic non-exchangeable protons had been assigned, as described later. The aromatic assignment was transferred from the non-exchangeable proton assignment and a number of characteristic aromatic-to-imino NOE connectivities were observed which enabled the assignment to be confirmed. Because the exchangeable and non-exchangeable proton assignments were done at two different temperatures the aromatic protons' chemical shifts varied slightly. The most characteristic imino-to-aromatic connectivity observed is between a uracil imino proton and its base paired adenine H2. Which gives rise to a very intense sharp peak.

The amino protons were assigned based on their NOE connectivities to imino protons, their chemical shifts and their characteristic broad NOE peaks. The adenine H61 and H62 amino proton resonances can be specifically identified by their respective chemical shift ranges of ~7.5-8.0 ppm and ~6.25-7.0 ppm and their NOE connectivity to their base paired uracil's imino proton. The cytosine H41 and H42 amino protons can be identified by their respective chemical shift ranges of ~8.0-8.75 ppm and ~6.75-7.25 ppm and their NOE connectivity to their base paired guanine's imino proton. The guanine H21 and H22 amino protons can be identified by their respective chemical shift ranges of ~6.75-7.25 ppm and ~5.5-6.25 ppm and their NOE connectivity to their own guanine's imino proton. The cytosine H42 amino proton and guanine H21 amino proton chemical shift ranges coincide and as they tend to show an NOE connectivity to potentially the same guanine imino proton, it was difficult to distinguish between the two types of amino protons.

Next came the non-exchangeable proton assignment, this began by identifying the different types of proton resonances. This was done so that a series of sequential NOE connectivity patterns could be followed, the most important of which being the pattern observed
between the H1' ribose protons and the H6/8 aromatic protons. The H1' resonances tended
to fall in the chemical shift range of ~5.0-6.5 ppm, but the H5 protons also occurred within
that range. The $^1$H-$^{13}$C HSQC spectrum can be used to distinguish the two: the C1' and C5
resonances fall into different chemical shift ranges ~89-96 ppm and ~96-106 ppm
respectively. C5 resonances, and thus the H5 resonances, can be further split into cytosine
C5 and uracil C5 resonances; which tended to fall in the chemical shift ranges of ~96-100
ppm and ~103-106 ppm respectively. The H2, H6 and H8 aromatic protons all fall in the
same proton chemical shift range (~6.9-8.5 ppm), again the $^1$H-$^{13}$C HSQC spectrum was
used to distinguish the C2 resonances, and thus the H2 resonances, as the C2 resonances
fall into a distinct chemical shift range of ~152-156 ppm. However, the C8 and C6
chemical shift ranges overlapped (~135-144 ppm and ~140-145 ppm). These were
however separated by observing the H5 to H6 scalar coupling correlations in either the
TOCSY, DQF-COSY or $^1$H-$^{13}$C HSQC-TOCSY spectra. The H2', H3', H4', H5' and H5''
resonances were similarly identified. All of these proton resonances fall into the same
region (~3.9-4.9 ppm). But the associated C2', C3', C4' and C5' resonances all fall into
distinct chemical shift ranges, ~75-78 ppm, ~69-75 ppm, ~81-85 ppm and ~61-68 ppm
respectively, although the C2' and C3' ranges overlapped.

Once the different types of resonances were identified, the next step was for them to be
sequentially assigned. The basis of the sequential assignment is H1' to H6/8 intra and
sequential NOE connectivity patterns. This is as the NOE connectivity peaks between the
H1' and H6/8 occurred in a relatively uncrowded region of the spectrum, all other sequential
connectivity patterns occurred in a very overlapped region of the spectrum. The distances
concerned in this series of sequential connectivities are illustrated in Figure 2.4.3
The H1’-H6/8 NOE sequential connectivity pattern is simply the series of cross peaks formed between H6/8 resonance and the H1' resonance of the same residue and then the NOE peak formed between H1' of that residue and the H6/8 of the next and so on throughout the RNA. So long as the distances between protons did not greatly exceed 5Å, which is the case for standard A-form RNA. Therefore the pattern of peaks in tandem with the previous identification of the different types of resonances, was used to sequentially assign the H6/8 and H1' nuclei.

The sequence specific assignment was then easily extended to the H5 by observing the H5 to H6 cross peaks in the TOCSY and DQF-COSY spectra. Characteristic aromatic-to-aromatic NOE connectivities were observed, these were vital in aiding the observation of the H1'-H6/8 NOE sequential connectivity pattern, and confirming the assignment.

Next the H2 protons were sequence specifically assigned by observing characteristic NOE connectivities between the H2 resonances and the sequentially assigned H1's. The H2' resonances were then sequentially assigned by observing their H1' to H2' scalar coupled correlation in the DQF-COSY. In A-form RNA, the coupling constant between H1' to H2' is small therefore not all correlation peaks were observed.
To confirm and expand the sequential assignment a number of other sequential connectivity patterns were followed, as shown in Figure 2.4.4. However, these patterns occurred in a heavily overlapped region of the spectrum and so were not always easy to follow. Following these patterns helped expand the assignment to H3'. Characteristic NOE connectivities were also observed between H6/8 and the H4', H5' and H5'' resonances of the same residue and also between H1' to H3', H4', H5' and H5'' resonances on the same residue, which allowed further assignment. However, these NOE cross peaks occurred in a very crowded region of the spectrum. Additionally it is difficult to stereo specifically assign H5' and H5'', so they are quite often assigned non-stereo specifically as H5'a and H5'b.

Figure 2.4.4: A series of sequential and intra connectivity distances, sequential (blue) and intra (red), are indicated with arrows on the image. The distances in standard A-form RNA are all below 5Å, for the left image the distances are: sequential ~2.1-2.4 Å and intra ~3.5-3.8 Å, for the centre image the distances are: sequential ~3.2-3.6 Å and intra ~2.6-3.3 Å and for the right image the distances are: sequential ~3.7-5.0 Å and intra ~2.7-2.9 Å. Therefore NOE connectivities are all easily observed (save for the problem of overlap) and can be followed in standard A-form RNA, as indicated by the image, along the chain of residues.

NOE connectivities can be observed between the H2', H3', H4', H5' and H5'', but as they all occur in the same proton chemical shift region, the region is extremely overlapped preventing it being extensively used for assignment. This same region was observed both in the TOCSY and the DQF-COSY, revealing scalar coupling correlations between the ribose protons on the same residue. It is possible to trace these connectivities through the entire ribose sugar to gain assignment, particularly in the less overlapped DQF-COSY. But
the region is still very overlapped and in practice was difficult to analyse. The $^{1}H$-$^{13}C$ HSQC-NOESY and $^{1}H$-$^{13}C$ HSQC-TOCSY spectra were used to try to observe some more of these connectivities, as they were spread out along the $^{13}C$ dimension, to gain additional assignment. Unfortunately, in practice, overlap still caused a problem.

The sequence specific assignment was then easily extended to any carbon nuclei directly bonded to an assigned proton using the $^{1}H$-$^{13}C$ HSQC spectrum to acquire the $^{13}C$ assignment.

To assign the phosphorus resonances the $^{1}H$-$^{31}P$ CPMG-HSQC-NOESY was used. The $^{31}P$ resonances were assigned using the aromatic to $^{31}P$ and $H1'$ to $^{31}P$ sequential connectivity patterns. Using the $H1'_i - P_{i}$, $H6/8_i - P_{i}$, $H1'_{i-1} - P_{i}$ and $H6/8_{i-1} - P_{i}$ peaks which were observed and the already sequentially assigned $H6/8$ and $H1'$ resonances the spectrum was assigned. Unfortunately, on sensitivity grounds, the $^{1}H$-$^{31}P$ CPMG-HSQC-NOESY spectrum did not work well for the *H. h.* 29-mer RNA sample

2.4.3 Uniformly $^{13}C$ and $^{15}N$ labelled RNA

A brief overview of the actions taken to assign the *H. h.* 37-mer doubly labelled RNA motif is shown in Figure 2.4.5. The process began with the assignment of the exchangeable protons and associated $^{15}N$ resonances, these were attained using $^{15}N$ HSQC and $^{15}N$ HSQC-NOESY spectra recorded in 90% H$_2$O and 10% D$_2$O. The $^{15}N$ imino chemical shifts fall into two distinct ranges, the guanine imino nitrogens which appear around ~140-150 ppm, with the standard W-C base pairing imino nitrogens occuring around ~145-150 ppm. The uracil imino nitrogens appear around ~155-165 ppm, with the standard W-C base pairing imino nitrogens occuring around ~160-165 ppm. This information was used
to provisionally identify the different imino resonances. The $^1$H-$^{15}$N HSQC-NOESY was then used to follow the sequential connectivity patterns, analogous to the isotopically normal RNA $^1$H-$^1$H NOESY, to gain sequence specific assignment. The process was aided by comparison to the previously assigned H. h. 29-mer sample and literature data on the UUCG tetra nucleotide loop.\textsuperscript{75,76}

**Figure 2.4.5:** The chart shows a brief overview of the tasks which were taken to carry out the NMR assignment of the $^{13}$C and $^{15}$N labelled RNA motif and the NMR spectra used to accomplish these tasks.

The amino resonances were then assigned based on the observed NOE connectivities in the $^1$H-$^{15}$N HSQC-NOESY and on the characteristic amino proton (as described in section 2.4.2.) and nitrogen chemical shifts. The guanine N2 amino nitrogens appear around $\sim$70-78 ppm, the cytosine N4 amino nitrogens occur around $\sim$94-100 ppm and the adenine N6 nitrogens are usually found at $\sim$74-84 ppm.
The non-exchangeable proton assignment began as with the isotopically normal RNA motifs by identifying the different types of resonances, so sequential connectivity patterns could be followed. This was done in exactly the same manner as the isotopically normal RNA motifs, except that the 3D $^1$H-$^{13}$C HSQC-NOESY, and $^1$H-$^{13}$C HSQC-TOCSY spectra were used to observe the intra NOE and scalar coupling connectivities between H5 and H6 to allow separation of the H6 and H8 proton resonances.

Once this was done, the H1' to H6/8 sequential NOE connectivity pattern could then be followed in the 3D $^1$H-$^{13}$C HSQC-NOESY and the 3D $^1$H-$^{13}$C NOESY-HSQC. The additional $^{13}$C dimension allowed for reduction in overlap of the signals to allow the sequential connectivities to be followed more easily (otherwise the connectivities are followed in the same way as in the isotopically normal samples). The sequence specific assignment can then be easily extended to the H5 by observing the H5 to H6 cross peaks in the $^1$H-$^{13}$C HSQC-TOCSY and 3D $^1$H-$^{13}$C HSQC-NOESY spectra.

Next the H2 protons were sequence specifically assigned in the same manner as in the isotopically normal RNA. The H2' resonances were then sequentially assigned by observing their H1' to H2' scalar coupled correlation in the very low mixing time 3D $^1$H-$^{13}$C HSQC-TOCSY spectrum. This spectrum should only show three bond coupling correlations, due to its very low mixing time, and thus connectivities were traced from H1' to H2', and then from H2' to H3', from H3' to H4' and then finally from H4' to H5' and H5''. Due to the 3D nature of the spectrum, overlap is reduced as compared to a standard 2D low mixing time TOCSY or DQF-COSY, and so many more assignments were possible, although overlap was still an issue. A long mixing time 3D $^1$H-$^{13}$C HSQC-TOCSY which shows scalar coupling correlations through out the entire spin system, was used to aid the assignment. Despite the 3D nature of this experiment, due to the increased number of
peaks compared to the equivalent low mixing time spectrum overlap was still a large problem.

To confirm and expand the sequential assignment, a number of other sequential connectivity patterns were followed, as they were for the isotropically normal RNA, as shown in Figure 2.4.2. The same characteristic NOE connectivities can be observed in the doubly labelled RNA samples as in the isotopically normal samples. However, in the case of the doubly labelled RNA sample, following these connectivities was much easier due to the decreased overlap in the 3D \( ^1\text{H}-^1\text{C} \) HSQC-NOESY spectra. As stated earlier the 3D \( ^1\text{H}-^1\text{C} \) HSQC-NOESY separate the various types of ribose protons into different regions by virtue of the differing \( ^1\text{C} \) chemical shifts of their directly bonded \( ^1\text{C} \) atoms. Following and observing these connectivities allowed confirmation of and further assignment to be produced, particularly for the ribose protons.

The sequential assignment can be transferred to the carbons directly bonded to assigned protons by using the \( ^1\text{H}-^1\text{C} \) HSQC spectrum to acquire and confirm the \( ^1\text{C} \) assignment, potentially already gathered from the 3D \( ^1\text{H}-^1\text{C} \) HSQC-NOESY and 3D \( ^1\text{H}-^1\text{C} \) HSQC-TOCSY spectra.

The assignment of the \( ^3\text{P} \) resonances was performed using the 3D HCP spectrum. The most powerful method of assignment using the HCP is to follow the H4' to \( ^3\text{P} \) sequential connectivity pattern as cross peaks occur between H4'\(_{i-1}\) – C4'\(_{i-1}\) – P\(_i\), H4'\(_i\) – C4'\(_i\) – P\(_i\). Assignment can also be gleaned via H3', H5' and H5'' connectivities to the \( ^3\text{P} \) (H3'\(_{i-1}\) – C3'\(_{i-1}\) – P\(_i\), H5'\(_i\) – C5'\(_i\) – P\(_i\)). Due to the two ribose protons connected to one carbon the H5' and H5'' to \( ^3\text{P} \) connectivities can be particularly clear and easy to spot.
2.4.4 RNA-antibiotic complexes

The RNA-antibiotic complexes were assigned mainly by comparison with the assignment of their constituent parts and by following the chemical shifts of the resonances through the stages of the titration. To confirm this assignment the cross peaks patterns and sequential connectivity patterns observed in the 2D spectra, were compared to those of their constituent parts. Where necessary some of the same techniques used to originally assign the antibiotics and RNA were used to gain and confirm assignments.
2.5 Constraint generation

2.5.1 Antibiotic ROE distance constraints

The NMR input for the structure determination of antibiotics consisted primarily of $^1\text{H}-^1\text{H}$ distance constraints, derived from ROESY experiments carried out at different spin lock mixing times. A number of molecular distances were used as distance rulers to help obtain distance constraints from the measured ROE peak volumes; such as distances between aromatic protons, along with a number of other proton-proton distances found within the non-aromatic rings of the antibiotics. The distances were measured from the calculated starting structures (see section 2.6.1). These empirical distances were plotted against the corresponding ROE cross peak volume (measured using Sparky)\textsuperscript{81}. This was done separately for each set of experimental conditions used to obtain a ROESY spectrum. These plots were used with the data from their respective spectrum to place ROE cross peaks within distance ranges of 1.8 - 3.0 Å (strong), 2.8 - 4.0 Å (medium) and 3.8 - 5.0 Å (weak).

Based on the above, the distances suggested from each spectrum were then compared and an overall set of constraints was generated. At long mixing times, ROE intensities become a complicated function of many inter nuclear distances due to the onset of spin diffusion. Another difficulty with the ROESY experiment is that at high mixing times some cross peaks can arise from spin-spin scalar correlation. Therefore, the distance constraints obtained only from the 400 ms ROESY spectrum were given large error bounds (3.8 - 6.0 Å). For ROEs between exchangeable protons the error bounds for the distance constraints were set to give a distance range starting from 1.8 Å (i.e. A and B are between 1.8 Å and X Å apart), to allow for possible lower intensities due to protons exchanging with water.
Other ROE constraints were given larger error bounds if they had particularly low intensities, and so were difficult to distinguish from noise, or if they were consistently violated during structure calculation.

2.5.2 Antibiotic dihedral angle constraints

Dihedral angle constraints were derived from vicinal $^1$H-$^1$H scalar coupling constants. The scalar coupling constants were measured with high digital resolution 1D $^1$H spectra, using SpinWorks. MestReJ software was used to interpret the coupling constant data to give the possible dihedral angles for each specific case. If multiple possibilities for the dihedral angle were obtained, an attempt was made to reduce this number by studying the structure of the antibiotic to see if certain combinations of dihedral angles were possible and which were the most likely. If multiple options still seemed possible, then those dihedral angles were not constrained. As the coupling constants are also dependent on other factors, such as the electronegativity of the attached substituents and bond lengths, the information was used conservatively. Because of this large error limits were used, from ±20° to ±40°, depending on the quality of the spectra.

Other dihedral angle constraints were set to help maintain the correct sterochemistry in the rings of the molecule. This had to be done carefully, so as not to artificially restrict the conformation of the molecule. They were used as the structure calculation software did not account for the sterochemistry and so on some occasions without additional constraints within the ring the calculation would generate large numbers of structures with incorrect sterochemistry. For these dihedral angle constraints error bounds were set to ±30°.
2.5.3 RNA NOE distance constraints

The NOE distance constraints were generated using CcpNmr Analysis using low mixing time NOESY spectra (~100 ms). The cross peaks were classified into separate groups, based mainly on how overlapped the peaks were, so that lower error bounds can be applied to distances from cross peaks where the peak volume is known more accurately. The intensities were measured by peak volume and the reference distance, reference intensity and intensity function were calibrated using known distances and distance ranges, in the canonical regions of the RNA. The various constraint sets can then be exported as an ARIA constraint list which can be used in Xplor-NIH directly with minor modification (the Xplor-NIH software was used for structure calculation).

2.5.4 RNA dihedral angle constraints

Three main sets of dihedral angles were constrained; the ribose sugar pucker torsion angles, the glycosidic torsion angle and the backbone dihedral angles.

First the sugar pucker was determined by using approximate H1'-H2' coupling constants, gained from the DQF-COSYs and $^{13}$C chemical shift information. H1'-H2' coupling constants were estimated to be ~1.3 Hz if the cross peak was not visible or very low intensity in the DQF-COSY, thus the sugar pucker was constrained to be C$_{3'}$-endo. If the peak was of high intensity, the coupling constant was estimated to be ~7.6 Hz, thus the sugar pucker was constrained to be C$_{2'}$-endo. To confirm this, or where a DQF-COSY spectrum was not available, the canonical coordinate (can1) method was used. This method is based on equation 2.5.1, values of can1 greater than -6.25 ppm tend to indicate a C$_{3'}$-endo sugar pucker and values less than -6.25 ppm tend to indicate a C$_{2'}$-endo sugar
pucker, where $\delta_{C1'}$ is the chemical shift of the C1' resonance of the given residue, $\delta_{C4'}$ is the chemical shift of the C4' resonance of the given residue and $\delta_{C5'}$ is the chemical shift of the C5' resonance of the given residue.\textsuperscript{75,110} The sugar puckers were constrained using three dihedral angles $\nu_1$, $\nu_2$ and $\nu_3$. For C3-endo conformations these were set respectively to -25.0º +/- 30º, 37.3º +/- 30º and 80.0º +/- 20º and for C2-endo to 25.0º +/- 30º, -35.0º +/- 30º and 140.0º +/- 20º.

\begin{equation}
\text{can1} = 0.179 \delta_{C1'} - 0.225 \delta_{C4'} - 0.0585 \delta_{C5'}.
\end{equation}

The backbone dihedral angles were constrained to a standard A-form conformation, in canonical base pairing sections of the RNA where sugar pucker was found to be C3-endo and the phosphorus chemical shift fell within the standard A-form RNA range (approximately -3.75 to -4.80 ppm). The backbone was constrained using five dihedral angles $\alpha$, $\beta$, $\gamma$, $\varepsilon$ and $\zeta$, when constraining the backbone to an A-form conformation these were set respectively to -68.0º +/- 30º, 178.0º +/- 30º, 54.0º +/- 30º, -153.0º +/- 30º and -71.0º +/- 30º.

The glycosidic angle $\chi$, was constrained as anti when the intra H1' to H6/8 NOE was less intense than both the corresponding intra H2' to H6/8 and H3' to H6/8 NOEs. If only one of the intra, NOEs of the latter was observed, the angle was still constrained as anti as long as the observed NOE was of greater intensity than the corresponding H1' to H6/8 NOE. Where the above intra NOEs were not available, possibly due to overlap of peaks or the peaks not being assigned, residues in canonical W-C base pairing regions which were also constrained, as C3-endo were also constrained in the anti conformation. The glycosidic angle when constrained as anti was set to -150.0º +/- 90º.
2.5.5 RNA hydrogen bond constraints

Several factors were taken into account when constraining a base pair as a canonical W-C base pair. The first being whether the imino proton resonance for the base pair is observed in its canonical region of the spectrum. The second was the intensity of the imino peak, and the third was if the corresponding amino peaks were observed. All of these factors were taken into account, but the main emphasis was placed on the chemical shift of the imino proton resonance. Thus G-C base pairs were constrained with six distance constraints and A-U base pairs were also constrained with six distance constraints. Details of the distance constraints are given in Table 2.5.1, all the distance constraints were given error bonds of +/- 0.2 Å. These distances were based on those used in the Xplor NIH example scripts.\textsuperscript{108,109}

<table>
<thead>
<tr>
<th>For base pair:</th>
<th>Atom 1</th>
<th>Atom 2</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-C</td>
<td>Guanine N1</td>
<td>Cytosine N3</td>
<td>2.81</td>
</tr>
<tr>
<td>G-C</td>
<td>Guanine H1</td>
<td>Cytosine N3</td>
<td>2.87</td>
</tr>
<tr>
<td>G-C</td>
<td>Guanine O6</td>
<td>Cytosine N4</td>
<td>1.86</td>
</tr>
<tr>
<td>G-C</td>
<td>Guanine N2</td>
<td>Cytosine O2</td>
<td>2.81</td>
</tr>
<tr>
<td>G-C</td>
<td>Guanine N2</td>
<td>Cytosine N3</td>
<td>3.58</td>
</tr>
<tr>
<td>G-C</td>
<td>Guanine O6</td>
<td>Cytosine N3</td>
<td>3.63</td>
</tr>
<tr>
<td>A-U</td>
<td>Adenine N1</td>
<td>Uracil N3</td>
<td>2.92</td>
</tr>
<tr>
<td>A-U</td>
<td>Adenine N1</td>
<td>Uracil H3</td>
<td>1.91</td>
</tr>
<tr>
<td>A-U</td>
<td>Adenine N6</td>
<td>Uracil O4</td>
<td>2.89</td>
</tr>
<tr>
<td>A-U</td>
<td>Adenine N2</td>
<td>Uracil O2</td>
<td>2.94</td>
</tr>
<tr>
<td>A-U</td>
<td>Adenine N1</td>
<td>Uracil O4</td>
<td>3.69</td>
</tr>
<tr>
<td>A-U</td>
<td>Adenine N1</td>
<td>Uracil O2</td>
<td>3.67</td>
</tr>
</tbody>
</table>

Table 2.5.1: The table gives the distances used to constrain standard W-C base pairing, the first column state, which type of base pair the distance is for the second and third state the pair of atoms for which the distance between is constrained and the fourth states that distance.
2.6 Structure calculation protocol

The structure determination of both the nucleic acids and antibiotics were calculated using the Xplor-NIH structure determination software.\textsuperscript{123,124} Before structure determination could begin, several files were required; these included a topology file, a parameter file, a starting structure (only for antibiotics), sequence files (only for nucleic acids) and the various constraint files. Topology files contain information on a specific molecule or on a class such as nucleic acids. They include information such as atom types, charges, mass of atoms, the location of bonds and improper angles. Parameter files include such information as ideal bond lengths, bond angles, dihedral angles and improper angles and their respective force constants. There are standard files for nucleic acids included in Xplor-NIH, but for antibiotics a new topology file must be written and a more general parameter file used, as described subsequently, which still may need modification to work well for the antibiotic in question.

2.6.1 Antibiotics

A topology file was first generated (the topology files used are shown in Appendix A.2), where each atom was named and listed, and given an atom type as defined by the parameter file (described later). Next the bonds were defined along with the improper angles and hydrogen bond donors and acceptors (the bond angles and dihedral angles were defined automatically). The mass of each atom type was also defined and given standard masses. Each atom was also given a charge, this was calculated by first generating an “ideal” structure for the antibiotic by using Gaussian 03,\textsuperscript{111} with Gabedit,\textsuperscript{112} to run a B3LYP\textsuperscript{73,74} 6-31G energy minimisation on the antibiotic. This was followed by an MP2\textsuperscript{72} 6-31G single point energy calculation along with the Merz-Singh-Kollman scheme (MK or
ESP) of population analysis to finally generate charges. The structure resulting from
the energy minimisation was used as the starting structure. The GAFF (General Amber
Force Field) parameter file was used and converted into Xplor-NIH format (Appendix
A.2). Small modifications were required as it did not adequately describe a small number
of bond angles, and improper angles. Therefore extra parameters were added, by
comparison to similar described bond angles and improper angles already described in the
parameter file (see Appendix A.2). Once all the files had been collected the NMR
Structure determination of the antibiotic could then begin. This was done using the Xplor-
NIH software package.

Randomisation

The first step in the NMR structure determination of the antibiotics was randomisation.
This section of the script was based on the Xplor-NIH example script, random.inp. A
typical randomisation script is shown in Appendix A.2. For this initial step the structures
were given randomised coordinates. The process of randomisation produced 200 to 400
random coordinate structures for the antibiotics, depending on the number of stereo centres
in that particular antibiotic. The structure calculation protocol does not take the
stereochemistry into account so a racemic mixture of structures are produced, the incorrect
enantiomers therefore have to be removed at the end of the calculations. Therefore, for
every instance this problem occurs, in order to get ~100 structures with the correct stereo
chemistry the number of structures calculated has to be doubled. The stereo chemistry of
the rings can be controlled with additional dihedral angle constraints, as described
previously, so where possible these dihedral angle constraints were applied, but only when
it could be confidently assumed that they would not lead to an incorrect structure.
The resulting structures were then put through a period (96 ps) of high temperature (1000 K) restrained molecular dynamics. During this period the weight of the bonding and angle constraints term were slowly increased (0.00005 to 0.01), the weight of the van der Waals term was set to zero. This was followed by a short period (4.5 ps) of high temperature (1500 K) restrained molecular dynamics. The van der Waals, bond and bond angle terms were weighted very low and were slowly increased to a higher, yet a still relatively low weight. This was done to allow atoms to move relatively freely, to facilitate the sampling of conformational space and thus improve convergence. The weights were slowly increased to start to move the structures towards convergence and a more realistic model.

Simulated annealing

The randomisation step was followed by a simulated annealing step, this step was based on the Xplor-NIH example script dgsa.inp. A typical simulated annealing script is shown in Appendix A.2. The simulated annealing step started with two energy minimisations (stoped after 100 steps and 1000 steps or when the norm of the energy gradient is smaller than 0.0001). This was followed by a period (800 ps) of restrained high temperature (1500 K) molecular dynamics. This was followed by a long simulated annealing step (1500 ps) where the temperature is reduced from 1500 K to 300 K This step was then followed by a long energy minimisation (10,000 steps, minimum norm of the energy gradient 0.0001).

Refinement

The structures from the simulated annealing step were then put through a series of refinement steps, the refinement script was based on the refinement script used for the previous NMR structure determination of amicetin. An example of a typical refinement
script is shown in Appendix A.2. The distance and the dihedral experimental constraints were used with an increased weight throughout the refinement calculations. The first refinement step was started by an energy minimisation (500 steps, minimum norm of the energy gradient 0.0001), followed by a short period (0.01 ps) of high temperature (1000 K) molecular dynamics. This was followed by a longer period (67.5 ps) of high temperature molecular dynamics (1000 K) where the dihedral angle force constant was slowly increased (5 to 45, in steps of 5).

The second period of refinement started with energy minimisation (500 steps, minimum norm of the energy gradient 0.0001), followed by a short period (0.01 ps) of high temperature (1000 K) molecular dynamics. This was followed by another long period (67.5 ps) of high temperature molecular dynamics (1000 K) where the dihedral angle force constant again was slowly increased (5 to 45, in steps of 5). This was followed by a period (72.5 ps) of simulated annealing, where the temperature was slowly dropped (1000.1K to 275.1K, in steps of 25K). This was then followed by another energy minimisation (10,000 steps, minimum norm of the energy gradient 0.0001).

The final refinement period simply consisted of a long period on energy minimisation (100,000 steps, minimum norm of the energy gradient 0.0001). This final energy minimisation step was run with non-bonded energy terms such as electrostatics and attractive van der Waals.
Structure acceptance and validation

The resulting structures for each antibiotic were all examined as part of structure acceptance procedures and validation. Structures with large violations to the ROE and dihedral constraints were then disregarded. Structures with bond lengths, bond angles and dihedral angles that were considered too strained were also disregarded. This was done by using an acceptance script within Xplor-NIH,\(^{108,109}\) to select structures with low violations to constraints and low deviations from ideal geometry. The acceptance script was based on a Xplor-NIH example script accept.inp.\(^{108,109}\) A typical acceptance script is shown in Appendix A.2. The resulting structures were then checked manually for any with incorrect stereochemistry, any structures with incorrect stereochemistry were disregarded. If a large number of the structures were unacceptable the constraints were checked and modified if required, the calculation was then re-run.

From the finally accepted structures the top ten structures were chosen (based on low energy, low violations and good overlay with other good structures). This was done by dividing the structures into to groups of structures which over laid well and then selecting the best group and from that group selecting ten with the lowest overall all atom average RMSD. An average structure was then produced from these structures and then put through a period of energy minimisation, using the average structure script shown in Appendix A.2, which is based on the Xplor-NIH example script average.inp.\(^{108,109}\) The average structures were then checked to ensure they had low violations and low energy. The average structure and the overlay were then used to illustrate the NMR solution structure.
Further validating small molecule NMR structures can be difficult. One of the best methods used in general for NMR structure validation is to set aside a small percentage of constraints and not use them in the structure calculation. The distance or dihedral angle values of these set aside constraints can then be compared to those observed in the NMR structure. If the constraint values match well with the values observed in the NMR structure, this indicates the structure is reliable. However, in small molecules there is only a low number of constraints and so including all available constraints is important to gain an accurate NMR structure. Therefore this method was not used.

Another method of structure validation is to compare the results to X-ray crystal structures of the small molecule. Although we may expect some differences between an NMR solution model and an X-ray crystal model it is likely that there should be similarities. Therefore an NMR structure which is completely different to its corresponding X-ray structure indicates that one of the two is likely to be wrong. This method was utilised.

2.6.2 RNA

Structure determination was carried out using the python interface of Xplor-NIH, using scripts based on the python example scripts in Xplor-NIH. Typical simulated annealing and refinement scripts are shown in Appendix A.3. The standard Xplor-NIH nucleic acid parameter and topology files were used and NMR constraints, as described in sections 2.5.3 to 2.5.5, were employed.
Simulated annealing

A single extended, random structure, was generated. From this 100 structures were generated through restrained simulated annealing calculations. The initial temperature was set to 3500 K, and a high temperature section of torsion angle dynamics were run for 800 ps. During this process only phosphorus – phosphorus non bonded terms were used. The non bonding repulsion energy constant was ramped during this stage from 0.004 to 4 (default 100). This was done to allow better sampling of conformational space, by allowing all atoms save phosphorus to move through each other and allowing the phosphorus atoms to get very close to each other.

Subsequently a torsion angle simulated annealing calculation was performed, the initial temperature was set to 3500 K and was dropped by 12.5 K every 0.2 ps during the calculation until 25 K was reached. All of the non-bonded terms were turned back on during this stage. This was then followed by a torsion angle energy minimisation. After this, a Cartesian-space simulated annealing step was performed, using the same set-up as the torsion angle simulated annealing. This was then followed by a Cartesian-space energy minimisation (500 steps, minimum norm of the energy gradient 0.0001). The structure was then outputted and the process repeated 100 times to gain 100 structures.

The best structure (low energy, low violations, good overlay with other good structures) was then chosen. This was done by selecting the 30 lowest energy structures, including energy contributions from the constraints, and then selecting the ten with the lowest overall all atom average RMSD, and then selecting the structure from this group with the lowest all atom average RMSD. The first ten were then analysed to see if they were acceptable structures, this was based mainly on the extent and consistency of any violations and great
deviation from the ideal RNA parameters particularly in their standard W-C base pairing regions. If there were any problems with the structures the constraints were re-checked, particularly the violated constraints, to see if they were set correctly. If any corrections were made the structure calculation was rerun from the beginning. If the structures proved acceptable, the one structure with the lowest average all atom RMSD, out of the 10, was then taken on to be used in the refinement stage of the calculation.

**Refinement**

From the resulting structure of the simulated annealing stage, another 100 structures were generated through the refinement calculation. All the same constraints were used with the addition of database constraints, to aid convergence. First, there was an initial energy minimisation (20 steps, minimum norm of the energy gradient 0.0001), followed by a second longer torsion angle minimisation (1000 steps, minimum norm of the energy gradient 0.0001). Next an initial high temperature step was started, set up similarly to the previous high temperature step in the simulated annealing calculation, except that it was set to be 10 ps long and the temperature was set to 2000 K. This was followed by a torsion angle simulated annealing calculation. The temperature was initially set to 2000 K and reduced by 25 K every 0.2 ps till the final temperature of 25K was reached. A torsion angle energy minimisation and a Cartesian-space energy minimisation was then run to produce the final structures (500 steps each, minimum norm of the energy gradient 0.0001).
Structure acceptance and validation

From the 100 structures which resulted from the refinement step, the top ten structures were chosen (based on low energy, low violations and good overlay with other good structures). This was done by selecting the 30 lowest energy structures and then selecting the ten with the lowest overall all atom average RMSD. These structures were then analysed to see whether they were acceptable, as in the annealing step, making corrections and restarting from the annealing step using corrected constraints as necessary. Once ten acceptable structures were produced an average structure was calculated from them and then run through a period of energy minimisation to produce a single structure to be used for analysis. A typical average structure generation script is shown in Appendix A.3.

The average structure was then analysed using the w3DNA web-server, which gives information about the hydrogen bonding and helical parameters of the nucleic acid's structure. This helped validate the structures by showing any irregular and unusual values in the parameters that were measured. Any problems found were then investigated, and corrections made to the structure calculation if necessary.

The structure validation method of setting aside a small percentage of constraints as described for the antibiotic NMR structures (see section 2.6.1) can also be used for RNA structures, but again this method was not used as compared to proteins, for which this technique is commonly used, RNA has a low proton density and so a relatively low number of NOE distance constraints are generated. Therefore all available constraints are required to generate an accurate structure.
Chapter 3

NMR structure determination of peptidyl transferase inhibitor antibiotics

As described in the introduction, the peptidyl transferase antibiotics show promise with regard to being useful in combating the growing problem of antibiotic resistance. Therefore, this chapter will look at a number of antibiotics from the aminohexose cytosine nucleoside group of peptidyl transferase antibiotics. Specifically, this chapter describes the NMR structure determination of blasticidin S, gougerotin and amicetin. The structure determination and assignment of these antibiotics allows their binding to the ribosome to be studied by NMR. The binding of amicetin to the 29-mer motif, predicted to be important for the binding of these antibiotics in the ribosome, is discussed in chapter 5. Additionally, aspects of the dynamics of the antibiotics are discussed. The individual structures were analysed further in the current chapter and with all three structures compared and conclusions drawn and discussed. One particularly interesting prospect which emerged from the results is the potential for hybrid antibiotics.

3.1 Structure determination of blasticidin S

3.1.1 NMR assignment of blasticidin S

The chemical structure of blasticidin S is shown in Figure 3.1.1 and the assigned 1D $^1$H NMR spectrum measured at two different pH values is shown in Figure 3.1.2.
Figure 3.1.1: The chemical structure of the blasticidin S antibiotic. The labelling of the atoms, in Arabic numerals (1-15, 1’-8’ and 6”) are shown for NMR assignment purposes as discussed in the text. The labelling is based empirically on the labelling system set out and used in the free X-ray crystal structure. Changes however needed to be made to allow individual labelling of all proton groups. The three major spin systems are shown as α, β and γ. Stereo chemistry is indicated at the carbon atoms.

Figure 3.1.2: The top trace shows the 500 MHz 1D ¹H spectrum of blasticidin S in 90% ¹H₂O + 10% ²H₂O, at 2°C, measured at pH 4.8, (7.3 mM), and the bottom trace shows the 500 MHz 1D ¹H spectrum of blasticidin S in 90% ¹H₂O + 10% ²H₂O, at 2°C measured at a pH of 7.3, (7.4 mM). The 4’ peak was obscured by the water peak. The spectrum shows the assignment of all available proton resonances, including the exchangeable proton resonances 6’, 9’, 14/15, 4a and 4b.
To assign the protons of blasticidin S, the resonances were first divided into resonances belonging to exchangeable or non-exchangeable protons, as described in section 2.4.1. Subsequently, the spin systems were identified and then assigned, then the individual protons of those spin systems were assigned. The full assignment is shown in Figure 3.1.2 and later in Table 3.1.1.

Identification of spin systems

There are three distinct spin systems within blasticidin S; α, β, γ (as labelled on Figure 3.1.1). The α, β, γ spin systems were easily observable in the TOCSY spectrum, every resonance in the groups showed scalar correlations to every other resonance of its own group (Figure 3.1.3). The spin systems (α, β and γ) were subsequently assigned to their various groups of corresponding resonances, from the number of peaks and chemical shifts of those peaks within each group.
Assignment of non-exchangeable proton resonances

There are an additional five resonances that were not observed to belong to one of the three spin systems, of these there is one non-exchangeable resonance. This was easily assigned as it consisted of the one intense signal at 3.05 ppm, with a chemical shift and intensity that fitted well with the methyl protons in H13. The low field shift was induced by the attached electronegative nitrogen atom N12. The assignment of the remaining exchangeable proton resonances will be discussed later.
Assignment of spin system $\alpha$

Spin system $\alpha$ was first assigned (Figure 3.1.1), which contains the protons H5 and H6. The protons were specifically assigned to their individual resonances through the ROESY spectrum (Figure 3.1.4). H6 is closer on the molecule to the $\beta$ spin system than H5 and the peak at 7.63 ppm gave ROEs to a number of protons within the $\beta$ spin system, whereas the peak at 6.07 ppm does not show any ROEs. Therefore, the peak at 7.63 ppm represents H6 and H5 is represented by the peak at 6.07 ppm. The H5 resonance appears broad compared to the H6 resonance. This is also observed in the NMR spectrum of an isolated cytosine ring, additionally the chemical shifts in the isolated cytosine compare well with the cytosine ring within blasticidin S, helping to confirm the assignment of the spin system $\alpha$. 
Figure 3.1.4: 500 MHz $^1$H-$^1$H correlated ROESY spectrum of blasticidin S (7.4 mM) in 90% $^1$H$_2$O + 10% $^2$H$_2$O, pH 4.8 at 2°C, measured with a spin lock time of 400 ms. A number of important ROEs are highlighted, including a series of ROEs from the exchange retarded H6' proton which resonates at 8.62 ppm.

Assignment of spin system β

Spin system β consists of the protons H1' to H6'. H6' is a retarded amide proton and so was easily assigned to the peak at 8.62 ppm due to the high chemical shift and its absence in the $^2$H$_2$O spectrum. There are four doublets (including H6'), one singlet and one peak obscured by the water signal, which can however, be seen as a doublet in the 1D spectrum measured in $^2$H$_2$O. This seemed unexpected at first glance, as given the structure protons H2' and H3' both have two protons three bonds away ($^3J_{H,H}$), therefore suggesting that a triplet or doublet of doublets for each should be observed. The double bond between H2' and H3' means that the dihedral angle between them is fixed at 0° and thus will give a large coupling constant of ~10 Hz. Therefore the dihedral angle between H2' and H1' should be
approximately 90° as well as the dihedral angle between H3’ and H4’, thus producing a coupling constant close to zero in both and so resulting in doublet signals for both H2’ and H3’. H1’ is therefore a good candidate for the singlet signal at 6.50 ppm, due to the low coupling constant with H2’. The relatively high chemical shift also fits well as C1’ is directly bonded to O5’ and N1. Only doublet signals are left, this fits as H2’ is a doublet from its coupling with H3’ and vice versa H3’ a doublet from its coupling with H2’. H4’ is a doublet in the 2H2O spectra due to coupling with H5’ and finally H5’ is a doublet due to its coupling with H4’

H2’ and H3’ are alkene protons and as a result have relatively high chemical shift values, therefore the resonances at 5.88 and 6.11 ppm correspond to H2’ and H3’. The resonances both have a coupling constant of ~10 Hz, indicating that they are scalar coupled to each other (\(^3J_{H,H}\)). To specifically assign the protons the ROESY spectrum was used. The signal at 6.11 ppm shows ROE cross peaks to H4’ (assignment discussed later) and H6’ whereas the signal at 5.88 ppm does not. Therefore, the resonance at 6.11 ppm corresponds to H3’ and the resonance at 5.88 ppm to H2’.

This leaves H4’ and H5’ and the resonances at 4.65 ppm and 4.11 ppm. H6’ appears to be strongly retarded and so it is expected at low pH and at low temperature for coupling to be observed to H4’. In the 1D spectrum measured at pH 4.8 and at 2°C the peak at around 4.65 ppm is obscured, but a triplet resonance can be observed upon close inspection; the same resonance at pH 6.2 appears as a doublet. No difference in the splitting is observed between these spectrum for the signal at 4.11 ppm. To further confirm the assumption that H6’ is coupling with H4’ at low pH, the H6’ resonance in the spectrum measured at pH 4.8 is observed as a doublet. Therefore, the resonance at 4.65 ppm was assigned as H4’ and the resonance at 4.11 ppm was assigned as H5’. The ROE patterns observed for these resonances also confirms this assignment.
Assignment of spin system $\gamma$

The $\gamma$ spin system is composed of protons H8 to H11. There are five resonances; one triplet, two doublet of doublets and two broad multiplet peaks. H11 was observed to be represented by the triplet resonance at 3.48 ppm, this was as it was seen to have two protons three bonds away (the two H10 protons), which through scalar coupling ($^3J_{H,H}$) caused the H11 resonance to form a triplet. The resonance also showed an ROE cross peak to H14/15 (assignment described later), no other resonances in the spin system (except H10, assignment discussed subsequently) show such a cross peak. Additionally the resonance at 3.48 ppm was seen to have an intensity which fitted that of a CH$_2$ group, thus confirming the assignment of the resonance as H11.

H10 represents another two protons in a CH$_2$ group, it has three protons three bonds away (two H11 protons and one H9 proton) and C10 is connected to two carbons. Therefore, the broad multiplet at 2.06 ppm was assigned as H10, due to its appropriate intensity, multiplet characteristics and low chemical shift.

H9 represents one proton, which has four protons three bonds away (two H10 and two H8 protons), not including the exchangeable H9' protons. Also C9, is directly bonded to an electronegative nitrogen. Therefore, the broad multiplet at 3.67 ppm was assigned as H9, due to its appropriate intensity, multiplet characteristics and relatively high chemical shift. This left the two doublet of doublets, they both appear to be split by a characteristic two bond, geminal coupling constant of 16 Hz ($^2J_{H,H}$). This indicated that the resonances represent two protons held in separate chemical environments yet attached to the same carbon. The resonances are then further split into doublets of doublets indicating a proton three bonds away. Therefore, the two doublets of doublets at 2.76 and 2.64 ppm were assigned as H8a and H8b. H8a and H8b are observed to be held in separate environments.
This could be due to the possibility of a ring being formed via hydrogen bonding, containing H8a and H8b, and so helping to hold the two protons in separate environments. However, the separate chemical environments could be due to the fact that C8 is bonded to the chiral C9, but this same effect in the case of the H10 group does not seem to cause the two H10 protons to exist in separate environments.

**Assignment of the exchange retarded amide protons**

Assignment of the H6' exchange retarded proton has been described above. Consequently, this has left four additional exchangeable resonances and three groups of exchangeable proton groups (H9', H14/15 and H4) to assign. From the 1D spectrum taken at pH 4.8, the peak at 6.86 ppm has a large intensity, which indicates protons H14/15, and an ROE correlation is also seen from it to both H10 and H11.

There are no ROE signals or TOCSY signals for the remaining three exchangeable proton resonances, therefore the assignment of these peaks is quite tentative. Due to there being three peaks, it is likely that the protons in H9' or H4 are held in different environments, either by a partial double bond (H4) or by hydrogen bonding. Looking ahead to the less tentative assignment of gougerotin (see section 3.2.1), we see that the equivalent gougerotin proton group H4 gives two resonances at 7.41 ppm and 6.95 ppm. Therefore, the resonances at the similar chemical shifts of 7.58 ppm and 7.06 ppm were assigned as H4a and H4b, and H9' to 8.04 ppm. H4a and H4b are likely to be held in two different chemical environments due to the double bond character in the bond between C4 and N4.
3.1.2 Assignment table of blasticidin S

Table 3.1.1 shows the assignment of the proton chemical shifts and the associated coupling constants.

<table>
<thead>
<tr>
<th>Proton</th>
<th>Chemical Shifts (δ, ppm)</th>
<th>Coupling constant(s) (J_H,H, Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4a</td>
<td>7.58</td>
<td>s</td>
</tr>
<tr>
<td>4b</td>
<td>7.06</td>
<td>s</td>
</tr>
<tr>
<td>5</td>
<td>6.07</td>
<td>d 3J_{5,6} = 8</td>
</tr>
<tr>
<td>6</td>
<td>7.63</td>
<td>d 3J_{6,5} = 8</td>
</tr>
<tr>
<td>1'</td>
<td>6.50</td>
<td>s</td>
</tr>
<tr>
<td>2'</td>
<td>5.88</td>
<td>d 3J_{2,3} = 10</td>
</tr>
<tr>
<td>3'</td>
<td>6.11</td>
<td>d 3J_{3,2} = 10</td>
</tr>
<tr>
<td>4'</td>
<td>4.65*</td>
<td>d 3J_{4,5} = 8*</td>
</tr>
<tr>
<td>5'</td>
<td>4.11</td>
<td>d 3J_{5,6} = 9</td>
</tr>
<tr>
<td>6'</td>
<td>8.62</td>
<td>d 3J_{6,5} = 9</td>
</tr>
<tr>
<td>9'</td>
<td>8.04</td>
<td>s</td>
</tr>
<tr>
<td>8a</td>
<td>2.76</td>
<td>d of d 3J_{8a,9} = 4 and 2J_{8a,8b} = 16</td>
</tr>
<tr>
<td>8b</td>
<td>2.64</td>
<td>d of d 3J_{8b,9} = 8 and 2J_{8b,8a} = 16</td>
</tr>
<tr>
<td>9</td>
<td>3.67</td>
<td>m</td>
</tr>
<tr>
<td>10</td>
<td>2.06</td>
<td>q 3J_{10,9} = 7 and 3J_{10,11} = 8</td>
</tr>
<tr>
<td>11</td>
<td>3.48</td>
<td>t 3J_{11,10} = 8</td>
</tr>
<tr>
<td>13</td>
<td>3.05</td>
<td>s</td>
</tr>
<tr>
<td>14/15</td>
<td>6.86</td>
<td>s</td>
</tr>
</tbody>
</table>

Table 3.1.1: Table showing the chemical shifts (δ, ppm) of assigned ¹H resonances, and the scalar coupling constants (J_H,H, Hz) of Blasticidin S (7.4 mM in 90% ¹H₂O + 10% ²H₂O, pH 4.8 at 2°C). Values marked with an * were obtained in ²H₂O at pH 6.2 at 5.5°C. “s” indicates that the peak was observed as a singlet, “d” as a doublet, “t” as a triplet and “q” as a quartet and “m” as a multiplet. If the peak was observed as doublet of doublets this is indicated by “d of d”.

3.1.3 Effect of temperature on the NMR spectrum of blasticidin S

To begin to probe the dynamics of blasticidin S in solution, a series of 1D ¹H-NMR spectra were measured at 800 MHz over a temperature range of 2-50°C (see Figure 3.1.5). The temperature-induced change, in both chemical shift and linewidth for all of the exchangeable proton resonances is quite striking. Despite the increase in temperature, the amide proton H6’ and the guanidine protons H14/H15 demonstrate retardation to exchange against solvent water protons and the residual resonances are both clearly visible even up to 50°C. The proton resonances for H4a and H4b are also both clearly visible up to 20°C.
This retardation to exchange suggests that these protons may be involved in hydrogen bonding within the molecule, particularly in the case of H6'; these dynamic events have implications for the NMR solution structure.

![Figure 3.1.5: 800 MHz 1H-NMR variable temperature series (2°C to 50°C) of blasticidin S in 90% 1H2O + 10% 2H2O, pH 4.8 (7.4 mM). The temperature induced change in chemical shift and linewidth can clearly be seen for all the exchangeable proton resonances.](image)

None of the aliphatic or aromatic protons of blasticidin S produced significant changes with respect to temperature. This indicated that blasticidin S is endowed with a thermodynamically stable structure.
3.1.4 Longitudinal relaxation time, $T_1$, measurement of blasticidin S

To further probe the dynamics, the longitudinal relaxation time, $T_1$, of the aliphatic and aromatic protons of blasticidin S were qualitatively measured using the Bruker Avance 600 MHz spectrometer (see Figure 3.1.6). This produced the interesting result that all the protons on the chain relaxed faster than that of the protons on the two rings. In addition, the protons on the cytosine moiety seemed to relax slightly slower than those on the pseudo saccharide ring. From this experiment an approximate overall $T_1$ relaxation time for blasticidin S (under the standard sample conditions) was calculated at 1.2 seconds.

![Figure 3.1.6: Measurement of the longitudinal relaxation time constant $T_1$ of blasticidin S (2.5 mM) at 600 MHz at 25°C in 100% $^2$H$_2$O. The arrow marks the main point of transition.](image)

3.1.5 NMR constraints of blasticidin S

The structure was determined using 29 inter-atomic distance constraints, shown in Table 3.1.2, and 4 dihedral angle constraints, which are shown in Table 3.1.3. All of these dihedral angle constraint were derived from coupling constants. These constraints were generated following the methodology outlined in sections 2.5.1 and 2.5.2.
<table>
<thead>
<tr>
<th>Proton Group A</th>
<th>Proton Group B</th>
<th>ROE constraints (Å)</th>
<th>Positive error value</th>
<th>Negative error value</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1’</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6</td>
<td>2’</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6</td>
<td>3’</td>
<td>3.40</td>
<td>3.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6</td>
<td>4’</td>
<td>3.40</td>
<td>3.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6</td>
<td>5’</td>
<td>4.40</td>
<td>0.80</td>
<td>0.60</td>
</tr>
<tr>
<td>3’</td>
<td>6’</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>3’</td>
<td>8a</td>
<td>4.40</td>
<td>1.00</td>
<td>0.60</td>
</tr>
<tr>
<td>3’</td>
<td>8b</td>
<td>4.40</td>
<td>1.20</td>
<td>0.60</td>
</tr>
<tr>
<td>3’</td>
<td>9</td>
<td>4.40</td>
<td>1.60</td>
<td>0.60</td>
</tr>
<tr>
<td>4’</td>
<td>6’</td>
<td>3.40</td>
<td>0.60</td>
<td>1.60</td>
</tr>
<tr>
<td>5’</td>
<td>6’</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>5’</td>
<td>10</td>
<td>4.40</td>
<td>1.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6’</td>
<td>8a</td>
<td>3.40</td>
<td>0.60</td>
<td>1.60</td>
</tr>
<tr>
<td>6’</td>
<td>8b</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6’</td>
<td>9</td>
<td>4.40</td>
<td>0.60</td>
<td>2.60</td>
</tr>
<tr>
<td>6’</td>
<td>10</td>
<td>4.40</td>
<td>1.60</td>
<td>2.60</td>
</tr>
<tr>
<td>6’</td>
<td>11</td>
<td>4.40</td>
<td>1.20</td>
<td>2.60</td>
</tr>
<tr>
<td>8a</td>
<td>10</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>8b</td>
<td>10</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>8a</td>
<td>11</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>8b</td>
<td>11</td>
<td>3.40</td>
<td>1.60</td>
<td>0.60</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>9</td>
<td>13</td>
<td>3.40</td>
<td>3.60</td>
<td>0.60</td>
</tr>
<tr>
<td>9</td>
<td>14/15</td>
<td>4.40</td>
<td>1.00</td>
<td>2.60</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>10</td>
<td>14/15</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>11</td>
<td>14/15</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>13</td>
<td>14/15</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table 3.1.2: This table shows the list of NMR experimental ROE constraints used in the XPLOR-NIH structure determination protocol for blasticidin S.

<table>
<thead>
<tr>
<th>Dihedral Angles</th>
<th>Dihedral Angle Constraints</th>
<th>Error value</th>
</tr>
</thead>
<tbody>
<tr>
<td>H5’,C5’,C4’,H4’</td>
<td>164</td>
<td>30</td>
</tr>
<tr>
<td>H1’,C1’,C2’,H2’</td>
<td>90</td>
<td>40</td>
</tr>
<tr>
<td>H4’,C4’,C3’,H3’</td>
<td>90</td>
<td>40</td>
</tr>
<tr>
<td>H4’,C4’,N6’,H6’</td>
<td>180</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 3.1.3: Table showing the list of NMR experimental dihedral constraints, used in the XPLOR-NIH structure determination protocol for blasticidin S.
3.1.6 NMR solution structure of blasticidin S

The final NMR solution structure was calculated, as described in section 2.6.1, the resulting structure is shown in Figure 3.1.7.

![Figure 3.1.7](image)

**Figure 3.1.7:** The left image shows the average NMR structure of blasticidin S. The blasticidin S structure reveals three structurally significant intramolecular hydrogen bonds. The hydrogen bonds are shown by the dotted lines and the distances between donor and acceptor atoms (in Å) are indicated. The right image shows an overlay of the 10 best structures, with an all atom average RMSD of 0.56 Å.

The structure fits the constraints applied and forms two hydrogen bonds acting to hold the structure in its conformation. The ten best structures shown overlayed in Figure 3.1.7 have a low all atom RMSD of 0.56 Å. The structure also fits well with features observed in the NMR spectra that were not used to generate constraints. The most important feature being the hydrogen bond between H6' and N9', causing the retardation to exchange up to 50°C and beyond. This hydrogen bond appears to be stabilised by a weaker hydrogen bond between H9' and O8', indicated in the NMR spectra by a much lesser degree of retardation (only up to 20°C), as compared to H6' and by the structure showing a greater hydrogen bond length. Another feature in the NMR spectrum potentially explained by the structure is the separate resonances observed for H8a and H8b as mentioned earlier and why only
one resonance is observed for proton group 10 whose carbon is similarly bonded to the chiral C9. The solution structure explains this, as C8 is within a H-bonded ring system held together by the H9' to O8 and H6' to N9' hydrogen bonds and so give an additional factor holding the two H8 protons in separate environments.

H4a and H4b are also slightly retarded, but no hydrogen bonds are observed to hold them in position; given the chemical structure it is unlikely that any hydrogen bonds would be possible, whatever the conformation. Therefore the retardation observed must be due to other means

A number of hydrogen bonds are observed in the blasticidin S solution structure which hold together the conformation of blasticidin S. Table 3.1.4 details these hydrogen bonds and compares them to those observed in the free and major bound crystal structures. All the hydrogen bonds are potentially shared between the free structures, but two are lost in the bound crystal structure leaving only one potential intramolecular hydrogen bond.

<table>
<thead>
<tr>
<th>D-H···A</th>
<th>D···A Solution structure (Å)</th>
<th>D···A free crystal structure (Å)</th>
<th>D···A bound crystal structure (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N9'-H9'···O8'</td>
<td>3.50</td>
<td>2.67</td>
<td>Not observed</td>
</tr>
<tr>
<td>O8'-H8'···O5'</td>
<td>2.92</td>
<td>-</td>
<td>2.77</td>
</tr>
<tr>
<td>N6'-H6'···N9'</td>
<td>2.83</td>
<td>-</td>
<td>Not observed</td>
</tr>
</tbody>
</table>

Table 3.1.4: A list of possible hydrogen bonds, where D is the donor atom, H is the hydrogen and A is the acceptor. The distances given are measured between the donor and acceptor atoms, for the NMR structure, the free crystal structure and the major bound crystal structure, a “-” indicates that this value was not available but that the hydrogen bond is a possibility.35,61
Table 3.1.5 details important dihedral angles of the solution structure and compares these to those observed in the bound and free crystal structure. The differences between both the free structures are quite small around the pseudo saccharide ring and the beginning of the chain section of the molecule. The dihedral angles generally only vary moderately. The difference between the free solution structure and bound crystal structure is much larger as would be expected. The equivalent of the glycosidic dihedral angle (O5'-C1'-N1-C6) also varies significantly between all three structures.

<table>
<thead>
<tr>
<th>Dihedral angle</th>
<th>NMR structure (°)</th>
<th>Free crystal structure (°)</th>
<th>Bound crystal structure (°)</th>
<th>Δfree (°)</th>
<th>Δbound (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O5'-C1'-N1-C6</td>
<td>41.6</td>
<td>86</td>
<td>20.3</td>
<td>-44.4</td>
<td>21.3</td>
</tr>
<tr>
<td>C1'-C2'-C3'-C4'</td>
<td>-1.6</td>
<td>3</td>
<td>0.5</td>
<td>-4.6</td>
<td>-2.1</td>
</tr>
<tr>
<td>C2'-C3'-C4'-C5'</td>
<td>22.3</td>
<td>15</td>
<td>6.8</td>
<td>7.3</td>
<td>15.5</td>
</tr>
<tr>
<td>C3'-C4'-C5'-O5'</td>
<td>-55.0</td>
<td>-50</td>
<td>-39.0</td>
<td>-5.0</td>
<td>-16.0</td>
</tr>
<tr>
<td>C4'-C5'-O5'-C1'</td>
<td>71.6</td>
<td>68</td>
<td>66.4</td>
<td>3.6</td>
<td>5.2</td>
</tr>
<tr>
<td>C5'-O5'-C1'-C2'</td>
<td>-48.7</td>
<td>-50</td>
<td>-60.5</td>
<td>1.3</td>
<td>11.8</td>
</tr>
<tr>
<td>O5'-C1'-C2'-C3'</td>
<td>12.8</td>
<td>15</td>
<td>25.70</td>
<td>-2.2</td>
<td>-12.9</td>
</tr>
<tr>
<td>C5'-C4'-N6'-C7</td>
<td>-104.0</td>
<td>-91</td>
<td>-131.0</td>
<td>-13.0</td>
<td>27.0</td>
</tr>
<tr>
<td>C3'-C4'-N6'-C7</td>
<td>134.6</td>
<td>140</td>
<td>94.4</td>
<td>-5.4</td>
<td>40.2</td>
</tr>
<tr>
<td>C4'-N6'-C7-C8</td>
<td>176.2</td>
<td>170</td>
<td>178.3</td>
<td>6.2</td>
<td>-2.1</td>
</tr>
<tr>
<td>N6'-C7-C8-C9</td>
<td>-70.9</td>
<td>-94</td>
<td>-79.3</td>
<td>23.1</td>
<td>8.4</td>
</tr>
<tr>
<td>C7-C8-C9-C10</td>
<td>-178.0</td>
<td>-177</td>
<td>-60.6</td>
<td>-1.0</td>
<td>-117.4</td>
</tr>
<tr>
<td>C8-C9-C10-C11</td>
<td>59.7</td>
<td>74</td>
<td>122.9</td>
<td>14.3</td>
<td>-63.2</td>
</tr>
<tr>
<td>C9-C10-C11-N12</td>
<td>179.4</td>
<td>155</td>
<td>173.3</td>
<td>24.4</td>
<td>6.1</td>
</tr>
<tr>
<td>C10-C11-N12-C14</td>
<td>-129.6</td>
<td>-106</td>
<td>-96.8</td>
<td>-23.6</td>
<td>-32.8</td>
</tr>
<tr>
<td>C11-N12-C14-N14</td>
<td>-90.0</td>
<td>-156</td>
<td>164.6</td>
<td>66</td>
<td>105.4</td>
</tr>
<tr>
<td>C11-N12-C14-N15</td>
<td>91.9</td>
<td>20</td>
<td>1.1</td>
<td>71.9</td>
<td>90.8</td>
</tr>
</tbody>
</table>

Table 3.1.5: A list of important dihedral angles. The dihedral angles are given for the NMR structure, the free crystal structure, the major bound crystal structure, and the difference between them is indicated in the two delta columns. The difference between the NMR structure and the free crystal structure is shown in the “Δfree” column and the difference between the NMR structure and the bound crystal structure is shown in the “Δbound” column. Where the difference is greater than 10° the cell is coloured yellow and where the difference is greater than 25° the cell is coloured red.35,61
3.2 Structure determination of gougerotin

3.2.1 NMR assignment of gougerotin

The chemical structure of gougerotin is shown in Figure 3.2.1 and the assigned 1D $^1$H NMR spectrum is shown in Figure 3.2.2. Gougerotin was assigned using the same method as blasticidin S as described in sections 2.4.1 and 3.1.1. The full assignment is tabulated later in Table 3.2.1.

Figure 3.2.1: The chemical structure of the gougerotin antibiotic. The labelling of the atoms, in Arabic numerals (1-14, 1'-7' and 6") are shown for NMR assignment purposes as discussed in the text. The labelling is based on the scheme used in the NMR solution structure of blasticidin S. The three major spin systems are shown as α, β and γ. Stereo chemistry is indicated at the carbon atoms.

Figure 3.2.2: 600 MHz $^1$H-NMR spectrum of gougerotin (1.5 mM) in 90% $^1$H$_2$O + 10% $^2$H$_2$O, at 2°C pH 6.2, including an expansion of an overlapped region. The spectrum shows the assignment of all observed proton resonances, including the exchangeable proton resonances 6’, 7, 13, 4a and 4b.
Identification of spin systems

There are three distinct spin systems within gougerotin; α, β, γ (as labelled on Figure 3.2.1). The α, β, γ spin systems were easily observable in the TOCSY spectrum, with every resonance in the groups showing scalar correlations to every other resonance of its own group (Figure 3.2.3) (except for the exchangeable proton HN10, which was not observed). The spin systems (α, β and γ) were subsequently assigned to their various groups of corresponding resonances, from the number of peaks and chemical shifts of those peaks within each group.

Figure 3.2.3: 600 MHz $^1$H-$^1$H correlated TOCSY spectrum of gougerotin (1.5 mM) in 90% $^1$H$_2$O + 10% $^2$H$_2$O, at 2°C, pH 6.2, measured with a spin lock mixing time of 70 ms. The three major spin systems are shown, labelled α, β and γ, key cross peaks are also labelled.
Assignment of the non-exchangeable protons

There are an additional seven resonances that were not observed to belong to one of the three spin systems, out of these there are three non-exchangeable resonances (Figure 3.2.2). The resonance at 2.75 ppm was easily assigned, as it consisted of one intense signal with a chemical shift and intensity that fitted well with the methyl protons in the H14 proton groups. The low field shift of H14 being induced by the attached electronegative nitrogen atom N13. The two remaining non exchangeable proton resonances are at 3.97 and 3.93 ppm, and each has a 16 Hz coupling constant which indicates that these two protons are geminally coupled ($^2J_{H,H}$). This led to the assignment of the two resonances to the two protons of H12. This indicated that the two protons of H12 were held in slightly different chemical environments, therefore the resonances at 3.97 and 3.93 ppm were assigned to H12a and H12b respectively. The resonances also have a strong ROE cross peak between themselves and H14 (assigned above), and so further confirms this assignment (see Figure 3.2.4). The assignment of the remaining exchangeable resonances is discussed later.
First, spin system α (Figure 3.2.1) was assigned, this spin system contains the protons H5 and H6. These two aromatic protons from the cytosine moiety were easily assigned as these appear at characteristic chemical shifts, thus the resonances at 6.10 and 7.80 ppm were assigned to H5 and H6 respectively. To confirm the assignment the H6 proton shows a number of ROEs to adjacent protons of spin system β and H5 shows an ROE to the exchangeable H4a resonance (assignment described later) (both highlighted in Figure 3.2.4).
Assignment of spin system $\beta$

Spin system $\beta$ consists of the protons H1' to H6'. The exchangeable proton H6' was first assigned to the resonance at 8.65 ppm; the resonance shows a clear set of cross peaks to the rest of the protons in spin system $\beta$ in the TOCSY (Figure 3.2.3), and is observed to diminish in intensity upon increase in temperature. H1' was assigned to the resonance at 5.73 ppm as it is a clear doublet and is the furthest down field of the non-exchangeable protons in the $\beta$-spin system. This is due to the C1' being directly bonded to a nitrogen and an oxygen atom, acting to deshield H1'. H5' was assigned to 4.14 ppm; it is a clear doublet and is next furthest low field, due to the C5' being directly bonded to an oxygen atom and the CONH$_2$ group. The resonance also shows an ROE to H7' (assignment discussed later), to further distinguish it from H1' (Figure 3.2.4). H4' was assigned to the resonance at 4.07 ppm, as the resonance is a clear triplet at high temperatures and a distorted triplet at low temperatures. This may be due to the effect of the adjacent exchangeable H6'. If so, H6' will be exchanging more rapidly at higher temperatures and so will have no effect on coupling.

The two resonances remaining in spin system $\beta$ were observed to occupy the overlapped region around 3.8 ppm. This was expected as the two proton groups H2' and H3' are both in very similar chemical environments. This region was characterised using the ROESY spectrum as shown in Figure 3.2.5. Two cross peaks can be observed in this region. The first cross peak occurs between the resonance at 3.86 ppm and the H4' resonance (4.08 ppm), the other cross peak occurs between the resonance at 3.82 ppm and H5' (4.14 ppm). Based on these ROEs and the stereochemistry of the saccharide ring, the resonance at 3.86 ppm was assigned to H2' and the resonance at 3.82 ppm to H3'.
Figure 3.2.5: Section of the 600 MHz 'H-'H correlated ROESY spectrum of gougerotin (1.5 mM) in 90% \(^1\text{H}_2\text{O} + 10\% \ ^2\text{H}_2\text{O}, \text{at } 2^\circ\text{C}, \text{pH } 6.2, \text{measured with a spin lock mixing time of 400 ms. Cross peaks arising from three different resonances, within the overlapped region, are clearly visible (labelled), each showing a distinct splitting pattern.}

*Assignment of spin system \(\gamma\)*

The \(\gamma\) spin system is composed of protons H8 to H10, H10 is an exchangeable proton and was not observed. The H9 proton group was initially assigned to the resonance at 3.84 ppm, as H9 is in a similar chemical environment to H2' and H3'. Therefore, by process of elimination H8 is assigned to the resonance at 4.53 ppm. This is confirmed by the low signal intensity which suggests that the group is composed of one proton. The splitting pattern also indicates that H8 is coupled to two non-equivalent protons. This fits the fact that when the shape of the H9 resonance is examined through the ROE cross peaks shown in Figure 3.2.5, the splitting pattern indicates that both H9 protons are held in slightly different environments and are coupling with each other *via* two bond geminal coupling. This lead to the assignment of the resonances at 3.87 ppm and 3.81 ppm to H9a and H9b respectively.
Assignment of the exchangeable protons

The exchangeable proton resonances were easily identified from the variable temperature 1D experiments (Figure 3.2.6) and from the positive exchange cross peaks they gave to the water signal in the ROESY spectrum (Figure 3.2.4). Assignment of the exchange retarded H6' proton has been described above. The resonance at 7.73 ppm shows ROE cross peaks to H4', H5', and H8. The resonance at 7.31 ppm also shows this cross peak pattern, as well as a possible ROE to H12. The resonance at 7.31 ppm also shows a possible TOCSY cross peak to H14. This has lead to the tentative assignment of the resonances at 7.73 ppm and 7.31 ppm to H7' and H13 respectively. There is also a strong peak in the TOCSY spectrum between H7' and H13, and an ROE correlation between the two. This may suggest that these two protons are close though space and possibly even connected via a hydrogen bond, or that the protons exchange between the two proton groups.

The resonance at 7.41 ppm was initially assigned to H4 as it gave ROE cross peaks to H5. The resonance at 6.95 ppm was difficult to assign confidently, as it shows a strong cross peak in the TOCSY to the H4 resonance. Additionally in the ROESY an ROE correlation is observed between the two, also no other ROE correlations are seen. There are two major possibilities at this point, either the resonance is for H10 and there is considerable folding, which would appear to put a large amount of strain on the molecule, or that the H4 group is held in two separate environments due to partial double bond character, and so leads to two resonances. The second option was thought to be the more likely, and so the resonances at 7.41 ppm and 6.95 ppm were assigned as H4a and H4b respectively.
3.2.2 Assignment table of gougerotin

Table 3.2.1 shows the assignment of the proton chemical shifts of gougerotin and the associated coupling constants.

<table>
<thead>
<tr>
<th>Proton</th>
<th>Chemical shift (δ, ppm)</th>
<th>Coupling constant(s) (J_{HH}, Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4a</td>
<td>7.41</td>
<td>s</td>
</tr>
<tr>
<td>4b</td>
<td>6.95</td>
<td>s</td>
</tr>
<tr>
<td>5</td>
<td>6.10</td>
<td>d $^3J_{5,6} = 8$</td>
</tr>
<tr>
<td>6</td>
<td>7.80</td>
<td>d $^3J_{6,5} = 8$</td>
</tr>
<tr>
<td>1'</td>
<td>5.73</td>
<td>d $^3J_{1',2'} = 9$</td>
</tr>
<tr>
<td>2'</td>
<td>3.86</td>
<td>m</td>
</tr>
<tr>
<td>3'</td>
<td>3.82</td>
<td>m</td>
</tr>
<tr>
<td>4'</td>
<td>4.07*</td>
<td>m $^3J_{4',5'} = 10^<em>$ and $^3J_{4',6} = 10^</em>$</td>
</tr>
<tr>
<td>5'</td>
<td>4.14</td>
<td>d $^3J_{5',4'} = 10$</td>
</tr>
<tr>
<td>6'</td>
<td>8.65</td>
<td>d $^3J_{6',5'} = 6$</td>
</tr>
<tr>
<td>7'</td>
<td>7.73</td>
<td>s</td>
</tr>
<tr>
<td>8</td>
<td>4.53</td>
<td>t $^3J_{8,9a} = 6$ and $^3J_{8,9b} = 6$</td>
</tr>
<tr>
<td>9a</td>
<td>3.87</td>
<td>m</td>
</tr>
<tr>
<td>9b</td>
<td>3.81</td>
<td>m</td>
</tr>
<tr>
<td>12a</td>
<td>3.97</td>
<td>d $^2J_{12a,12b} = 16$</td>
</tr>
<tr>
<td>12b</td>
<td>3.93</td>
<td>d $^2J_{12b,12a} = 16$</td>
</tr>
<tr>
<td>13</td>
<td>7.31</td>
<td>s</td>
</tr>
<tr>
<td>14</td>
<td>2.75</td>
<td>s</td>
</tr>
</tbody>
</table>

Table 3.2.1: Table showing the proton chemical shifts (δ, ppm), and the scalar coupling constants (J_{HH}, Hz) of gougerotin (1.5 mM in 90% $^1$H$_2$O + 10% $^2$H$_2$O at 2°C) Values marked with a * were obtained in $^1$H$_2$O at 40°C. “s” indicates that the peak was observed as a singlet, “d” as a doublet, “t” as a triplet and “m” as a multiplet.

3.2.3 Effect of temperature on the NMR spectrum of gougerotin

To probe the dynamics of gougerotin in solution, a series of 1D $^1$H-NMR spectra were measured at 600 MHz field strength in a temperature range of 2°C to 40°C, as shown in Figure 3.2.6. The temperature-induced changes, in both chemical shift and linewidth for all of the exchangeable proton resonances, are again quite striking. Despite the increase in temperature, the amide proton H6' and the proton groups H7', H13, H4a and H4b all demonstrate retardation to exchange against solvent water protons and the H7' and H13 resonances are clearly visible up to 40°C. It is interesting to note that the H6' proton
appears to be much more strongly retarded in blasticidin S. This retardation to exchange suggests that these protons may be involved in hydrogen bonding within the molecule. These dynamic events have implications for the NMR solution structure.

![Figure 3.2.6: 600 MHz $^1$H-NMR variable temperature series (2°C to 40°C) of gougerotin (1.5 mM) in 90% $^1$H$_2$O + 10% $^2$H$_2$O, at pH 6.2. The temperature induced change in chemical shift and linewidth can clearly be seen for all the exchangeable proton resonances.]

**3.2.4 NMR constraints of gougerotin**

The structure was determined using 15 interatomic distance constraints, as shown in Table 3.2.2, and 6 dihedral angle constraints, as shown in Table 3.2.3. The constraints were produced following the methodology set out in sections 2.5.1 and 2.5.2. All of the dihedral angles were derived from scalar coupling constants except one (number 6, Table 3.2.3). This additional dihedral angle was within the saccharide ring and was consistent with the conformation of the saccharide rings as indicated by the observed coupling constants. This additional dihedral angle constraint was used as the structure calculation software did not account for the sterochemistry and so without this additional constraint the calculation would generate a larger numbers of structures with incorrect sterochemistry within the saccharide rings.
### Table 3.2.2: This table shows the list of NMR experimental ROE constraints used in the XPLOR-NIH structure determination protocol for gougerotin.

<table>
<thead>
<tr>
<th>Proton Group A</th>
<th>Proton Group B</th>
<th>ROE constraints (Å)</th>
<th>Positive error value</th>
<th>Negative error value</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>12</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6</td>
<td>2'</td>
<td>3.40</td>
<td>0.60</td>
<td>1.60</td>
</tr>
<tr>
<td>6</td>
<td>1'</td>
<td>4.40</td>
<td>0.60</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>5'</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>7'</td>
<td>4'</td>
<td>4.40</td>
<td>0.60</td>
<td>2.60</td>
</tr>
<tr>
<td>6'</td>
<td>5'</td>
<td>3.40</td>
<td>0.60</td>
<td>1.60</td>
</tr>
<tr>
<td>7'</td>
<td>5'</td>
<td>4.40</td>
<td>0.60</td>
<td>2.60</td>
</tr>
<tr>
<td>6'</td>
<td>8</td>
<td>4.40</td>
<td>0.60</td>
<td>2.60</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>4.40</td>
<td>2.60</td>
<td>0.60</td>
</tr>
<tr>
<td>7'</td>
<td>8</td>
<td>4.40</td>
<td>2.60</td>
<td>2.60</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>4.40</td>
<td>0.60</td>
<td>2.60</td>
</tr>
<tr>
<td>5'</td>
<td>13</td>
<td>4.40</td>
<td>0.60</td>
<td>2.60</td>
</tr>
<tr>
<td>4'</td>
<td>13</td>
<td>4.40</td>
<td>1.00</td>
<td>2.60</td>
</tr>
<tr>
<td>7'</td>
<td>13</td>
<td>3.40</td>
<td>0.60</td>
<td>1.60</td>
</tr>
</tbody>
</table>

### Table 3.2.3: Table showing the list of NMR experimental dihedral constraints, used in the XPLOR-NIH structure determination protocol for gougerotin. An “X” indicates that this was not the constraints origin and a “√” indicates that this was its origin, for example a “√” in the $^3J$ column indicates the constraint was derived directly from coupling constants.

<table>
<thead>
<tr>
<th>Serial Number</th>
<th>Dihedral Angles Constraints</th>
<th>Dihedral Angle Constraints</th>
<th>Error value</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H8,C8,C9,H9a</td>
<td>320.3</td>
<td>40</td>
<td>$^3J$√ X</td>
</tr>
<tr>
<td>2</td>
<td>H8,C8,C9,H9b</td>
<td>231.0</td>
<td>40</td>
<td>$^3J$√ X</td>
</tr>
<tr>
<td>3</td>
<td>H5',C5',C4',H4'</td>
<td>180.0</td>
<td>30</td>
<td>$^3J$√ X</td>
</tr>
<tr>
<td>4</td>
<td>H3',C3',C4',H4'</td>
<td>180.0</td>
<td>30</td>
<td>$^3J$√ X</td>
</tr>
<tr>
<td>5</td>
<td>H1',C1',C2',H2'</td>
<td>180.0</td>
<td>30</td>
<td>$^3J$√ X</td>
</tr>
<tr>
<td>6</td>
<td>H3',C3',C2',H2'</td>
<td>180.0</td>
<td>30</td>
<td>X√</td>
</tr>
</tbody>
</table>
3.2.5 NMR solution structure of gougerotin

The final NMR solution structure was calculated, as described in section 2.6.1 and is shown in Figure 3.2.7.

![Figure 3.2.7](image)

**Figure 3.2.7:** The *left image* shows the average NMR structure of gougerotin. The gougerotin structure in solution reveals a number of structurally important intramolecular hydrogen bonds. The hydrogen bonds are shown by the dotted lines and the distances between donor and acceptor atoms (in Å) are indicated. The *right image* shows an overlay of the 10 best structures, with an all atom average RMSD of 0.66 Å.

The structure fits the constraints applied and forms six intramolecular hydrogen bonds acting to hold the structure in its conformation. The ten best structures shown overlayed in Figure 3.2.7 have a low all atom RMSD of 0.66 Å. The structure also fits well with features observed in the NMR spectra that were not used to generate constraints. The most significant feature being the hydrogen bond between H13 and O6", causing H13 to be retarded to exchange up to 40°C and beyond. This hydrogen bond is also key in holding the antibiotic in this conformation. The retardation of H7' is also explained by the structure, in particular by the hydrogen bond between H7' and O5'. Although there does appear to be some flexibility in this part of the structure allowing CONH₂ group to rotate to
bring the NH$_2$ 7' group close to H13, as can be seen in the overlay of the best structures shown in Figure 3.2.7, which may help to explain the TOCSY peak observed between them. H6' is also retarded up to 40ºC, there are no specific hydrogen bonds observed in the average structure, but it is held in a loop of the molecule where a number of hydrogen bonds are possible given slight rearrangement. A water molecule may also be able to be held within this “crevice”. H4a and H4b are also slightly retarded, with no hydrogen bonds to hold them in position.

Another feature is that the two protons in each of the two groups H9 and H12 are held in very slightly different environments. This is explained in the structure for H12, by the hydrogen bond between HN13 and O6", which places H12 within a ring holding H12a and H12b into slightly different chemical environments. A similar explanation is seen for H9, by the presence of a hydrogen bond between HO9 and O7, producing a ring containing H9. Although for H9 this could be due to the fact that C9 is bonded to the chiral C8, which could lead to the two H9 protons being observed in different chemical environments.

Table 3.2.4 lists the hydrogen bonds observed in the gougerotin solution structure. These help hold together the conformation of the gougerotin solution structure. Table 3.2.4 also compares them to those observed in the two confirmations (A and B) observed in the crystal structure of the nucleoside fragment of gougerotin.$^{44}$ Most of the hydrogen bonds in the nucleoside section of the solution structure are potentially shared with that of the the crystal structure, except that between H2' and O2, due to a difference in the glycosidic dihedral angle (O5'-C1'-N1-C6), as later noted.
In the solution structure it is interesting to note that there are a string of hydrogen bonds following a long a series of O and OH groups along the right hand side of the molecule (as shown in Figure 3.2.7). This string of hydrogen bonds must act greatly to hold and stabilise the solution structure.

<table>
<thead>
<tr>
<th>D-H···A</th>
<th>D···A Solution structure (Å)</th>
<th>D···A Crystal structure A (Å)</th>
<th>D···A Crystal structure B (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N6'-H6'···O6''</td>
<td>2.95</td>
<td>2.14</td>
<td>2.71</td>
</tr>
<tr>
<td>O2'-H2'···O2</td>
<td>3.36</td>
<td>Not observed</td>
<td>Not observed</td>
</tr>
<tr>
<td>O2'-H2'···O3'</td>
<td>2.79</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>O3'-H3'···O2'</td>
<td>2.79</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>O3'-H3'···O7</td>
<td>2.58</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>O9-OH9···O7</td>
<td>2.72</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>N7'-H7'···O5'</td>
<td>2.59</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>N13-H13···O6''</td>
<td>2.77</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 3.2.4: A list of possible hydrogen bonds, where D is the donor atom, H is the hydrogen and A is the acceptor. The distances given are measured between the donor and acceptor atoms, for the NMR structure, and both confirmations observed in the crystal structure of the nucleoside fragment of gougerotin (structures A and B). A “-” indicates that this value was not available but that the hydrogen bond is a possibility and NA indicates that the hydrogen bond is not possible as this part of the molecule is not present in the nucleoside fragment. 

Table 3.2.5 details important dihedral angles of the solution structure and compares them to those observed in the free crystal structure of the nucleoside fragment of gougerotin. The differences in dihedral angles between the structures are quite small in general, except again for the equivalent of the glycosidic dihedral angle (O5'-C1'-N1-C6). This again varies significantly between the crystal and solution structures. This difference prevents the H2' and O2 hydrogen bond in the crystal structure.
<table>
<thead>
<tr>
<th>Dihedral angle</th>
<th>NMR structure (°)</th>
<th>Crystal structure A (°)</th>
<th>Crystal structure B (°)</th>
<th>ΔA (°)</th>
<th>ΔB (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O5'-C1'-N1-C6</td>
<td>47.4</td>
<td>75.0</td>
<td>71.7</td>
<td>-27.6</td>
<td>-24.3</td>
</tr>
<tr>
<td>C1'-C2'-C3'-C4'</td>
<td>-59.3</td>
<td>-54.9</td>
<td>-51.4</td>
<td>-4.4</td>
<td>-7.9</td>
</tr>
<tr>
<td>C2'-C3'-C4'-C5'</td>
<td>58.3</td>
<td>56.3</td>
<td>51.0</td>
<td>2.0</td>
<td>7.3</td>
</tr>
<tr>
<td>C3'-C4'-C5'-O5'</td>
<td>-58.3</td>
<td>-57.9</td>
<td>-57.7</td>
<td>-0.4</td>
<td>-0.6</td>
</tr>
<tr>
<td>C4'-C5'-O5'-C1'</td>
<td>62.2</td>
<td>61.7</td>
<td>66.3</td>
<td>0.5</td>
<td>-4.1</td>
</tr>
<tr>
<td>C5'-O5'-C1'-C2'</td>
<td>-62.0</td>
<td>-62.4</td>
<td>-66.6</td>
<td>0.4</td>
<td>4.6</td>
</tr>
<tr>
<td>O5'-C1'-C2'-C3'</td>
<td>59.0</td>
<td>56.7</td>
<td>58.3</td>
<td>2.3</td>
<td>0.7</td>
</tr>
<tr>
<td>C5'-C4'-N6'-C7</td>
<td>-162.84</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>C3'-C4'-N6'-C7</td>
<td>76.17</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>C4'-N6'-C7-C8</td>
<td>175.62</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>N6'-C7-C8-C9</td>
<td>-174.54</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>N6'-C7-C8-N10</td>
<td>-51.25</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>C7-C8-C9-O9</td>
<td>-53.79</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>C7-C8-N10-C11</td>
<td>-56.1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>C8-N10-C11-C12</td>
<td>-179.8</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>N10-C11-C12-N13</td>
<td>-55.1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>C11-C12-N13-C14</td>
<td>-151.8</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 3.2.5: A list of important dihedral angles. The dihedral angles are given for the NMR structure, and both confirmations observed in the crystal structure of the nucleoside fragment of gougerotin (structures A and B). The difference between them is indicated in the two delta columns, the difference between the solution structure and the crystal structure A is shown in the “ΔA” column and the difference between the solution structure and the crystal structure B is shown in the “ΔB” column. Where the difference is greater than 10° the cell is coloured yellow and where the difference is greater than 25° the cell is coloured red, NA indicates that the torsion angle is not present in the nucleoside fragment of gougerotin.\(^4\)
3.3 Structure determination of amicetin

3.3.1 NMR assignment of amicetin

The chemical structure of amicetin is shown in Figure 3.3.1, its structure was determined previously by C. Shammas.\(^{32,66}\) However, as part of this project the assignment was independently checked and this confirmed the original assignment. The assigned 1D \(^1\)H spectrum is shown in Figure 3.3.2.

![Chemical structure of amicetin](image1)

**Figure 3.3.1:** The chemical structure of the antibiotic, amicetin. The labelling of the atoms, in Arabic numerals (1*-8*, 1'-6' and 1-18), are shown for NMR assignment purposes as discussed in the text. The labelling is based on the labelling system set out and used in the X-ray crystal structure of amicetin.\(^{31}\)

![NMR spectrum of amicetin](image2)

**Figure 3.3.2:** 600 MHz \(^1\)H-NMR spectrum of amicetin (2.5 mM) in 90\% \(^1\)H\(_2\)O 10\% \(^2\)H\(_2\)O, pH 4.8 at 2°C. The spectrum shows the assignment of all available proton resonances, including the low-field shifted exchangeable proton resonance at 10.61 ppm, according to the proton labeling scheme given in Figure 3.3.1.
### 3.3.2 Assignment table of amicetin

Table 3.3.1 shows the assignment of the proton chemical shifts, the associated coupling constants and the $^{13}$C chemical shifts.

<table>
<thead>
<tr>
<th>Assignment</th>
<th>$^1$H Chemical shift ($\delta$, ppm)</th>
<th>$^1$H Coupling constant(s) (J, Hz)</th>
<th>$^{13}$C Chemical shift ($\delta$, ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7*/8*</td>
<td>2.95</td>
<td>s</td>
<td>44.1</td>
</tr>
<tr>
<td>4*</td>
<td>3.13</td>
<td>t $^3$J$<em>{4s,5s}$ = 10, $^3$J$</em>{4s,3s}$ = 10 ($^1$H$_2$O) s ($^1$H$_2$O)</td>
<td>72.4</td>
</tr>
<tr>
<td>6*</td>
<td>1.48</td>
<td>d $^3$J$_{6e,5e}$ = 6</td>
<td>20.8</td>
</tr>
<tr>
<td>5*</td>
<td>4.13</td>
<td>m</td>
<td>66.3</td>
</tr>
<tr>
<td>3*</td>
<td>4.04</td>
<td>t $^3$J$<em>{3s,4s}$ = 10, $^3$J$</em>{3s,2s}$ = 10</td>
<td>69.3</td>
</tr>
<tr>
<td>2*</td>
<td>3.69</td>
<td>d of d $^3$J$<em>{2s,3s}$ = 10, $^3$J$</em>{2s,1s}$ = 4</td>
<td>74.5</td>
</tr>
<tr>
<td>1*</td>
<td>5.14</td>
<td>d $^3$J$_{1s,2s}$ = 4</td>
<td>96.9</td>
</tr>
<tr>
<td>4'</td>
<td>3.51</td>
<td>t of d $^3$J$<em>{4's,3's}$ = 10, $^3$J$</em>{4's,5's}$ = 10, $^3$J$_{4's,6's}$ = 4</td>
<td>77.1</td>
</tr>
<tr>
<td>6'</td>
<td>1.36</td>
<td>d $^3$J$_{6'e,5'e}$ = 6</td>
<td>20.6</td>
</tr>
<tr>
<td>5'</td>
<td>3.85</td>
<td>m</td>
<td>79.5</td>
</tr>
<tr>
<td>3'a</td>
<td>1.67</td>
<td>q of d $^3$J$<em>{3'a,2'a}$ = 3, $^3$J$</em>{3'a,4'a}$ = 10, $^3$J$<em>{3'a,1'a}$ = 11, $^3$J$</em>{3'a,3'e}$ = 12</td>
<td>29.0</td>
</tr>
<tr>
<td>3'e</td>
<td>2.42</td>
<td>m $^3$J$<em>{3'e,2'e}$ = 12, $^3$J$</em>{3'e,4'e}$ = 4, $^3$J$_{3'e,5'e}$ = 4</td>
<td>29.0</td>
</tr>
<tr>
<td>2'a</td>
<td>1.89</td>
<td>q of d $^3$J$<em>{2'a,3'a}$ = 13, $^3$J$</em>{2'a,2'e}$ = 4, $^3$J$<em>{2'a,3'e}$ = 3, $^3$J$</em>{2'a,1'a}$ = 2</td>
<td>31.4</td>
</tr>
<tr>
<td>2'e</td>
<td>2.15</td>
<td>m $^3$J$<em>{2'e,3'e}$ = 13, $^3$J$</em>{2'e,3'a}$ = 4, $^3$J$<em>{2'e,5'e}$ = 3, $^3$J$</em>{2'e,1'a}$ = 2</td>
<td>31.4</td>
</tr>
<tr>
<td>1'</td>
<td>5.82</td>
<td>d of d $^3$J$<em>{1',2'}$ = 11, $^3$J$</em>{1',3'}$ = 2</td>
<td>86.2</td>
</tr>
<tr>
<td>6</td>
<td>8.24</td>
<td>d $^3$J$_{6,5}$ = 7</td>
<td>148.6</td>
</tr>
<tr>
<td>5</td>
<td>7.50</td>
<td>d $^3$J$_{5,6}$ = 7</td>
<td>101.5</td>
</tr>
<tr>
<td>13/9</td>
<td>7.68</td>
<td>d $^3$J$_{13/9,12/10}$ = 9</td>
<td>132.1</td>
</tr>
<tr>
<td>12/10</td>
<td>7.99</td>
<td>d $^3$J$_{12/10,11/9}$ = 9</td>
<td>124.4</td>
</tr>
<tr>
<td>18</td>
<td>1.62</td>
<td>s</td>
<td>21.6</td>
</tr>
<tr>
<td>17a</td>
<td>3.84</td>
<td>d $^3$J$_{17a,16b}$ = 12</td>
<td>67.9</td>
</tr>
<tr>
<td>17b</td>
<td>4.13</td>
<td>d $^3$J$_{17b,16a}$ = 12</td>
<td>67.9</td>
</tr>
<tr>
<td>14</td>
<td>10.61</td>
<td>s</td>
<td>NA</td>
</tr>
</tbody>
</table>

**Table 3.3.1**: Table showing the chemical shifts ($\delta$, ppm) of assigned $^1$H and $^{13}$C resonances and proton scalar coupling constants (J) of amicetin (3.0 mM in 100% $^2$H$_2$O, pH 5.7, 25°C, except for the exchangeable proton 14 which was measured at 2.5 mM in 90% $^1$H$_2$O 10% $^2$H$_2$O, pH 4.8 at 2°C). “S” indicates that the peak was observed as a singlet, “d” as a doublet, “t” as a triplet, “q” as a quartet and “m” as a multiplet. If the peak was observed as, for example, a quartet of doublets this is indicated by “q of d” and similarly for other combinations.
3.3.3 NMR constraints of amicetin

The structure was determined using 60 inter-atomic distance constraints (Table 3.3.2) and 17 dihedral angle constraints (Table 3.3.3). The constraints were produced following the methodology set out in sections 2.5.1 and 2.5.2. Ten of the dihedral angles were derived directly from coupling constants (1-10, Table 3.3.3) and seven additional dihedrals were set to help maintain the correct stereochemistry in the saccharide rings (11-17 Table 3.3.3). These additional dihedral angles were consistent with the conformation of the saccharide rings as indicated by the observed coupling constants. They were used as the structure calculation software did not account for the stereochemistry and so without the additional constraints, the calculation would generate large numbers of structures with incorrect stereochemistry within the saccharide rings.
<table>
<thead>
<tr>
<th>Proton Group A</th>
<th>Proton Group B</th>
<th>ROE constraints (Å)</th>
<th>Positive error value</th>
<th>Negative error value</th>
</tr>
</thead>
<tbody>
<tr>
<td>7*/8*</td>
<td>6*</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>7*/8*</td>
<td>5*</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>7*/8*</td>
<td>3*</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>7*/8*</td>
<td>2*</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>7*/8*</td>
<td>4'</td>
<td>4.40</td>
<td>2.60</td>
<td>0.60</td>
</tr>
<tr>
<td>7*/8*</td>
<td>6'</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>7*/8*</td>
<td>5'</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>4*</td>
<td>6*</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>4*</td>
<td>2*</td>
<td>3.40</td>
<td>1.60</td>
<td>1.60</td>
</tr>
<tr>
<td>6*</td>
<td>5*</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6*</td>
<td>3*</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6*</td>
<td>2*</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6*</td>
<td>1*</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6*</td>
<td>4'</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>5*</td>
<td>1*</td>
<td>4.40</td>
<td>2.60</td>
<td>1.60</td>
</tr>
<tr>
<td>5*</td>
<td>4'</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>5*</td>
<td>6'</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>3*</td>
<td>1*</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>3*</td>
<td>6'</td>
<td>3.40</td>
<td>0.60</td>
<td>1.60</td>
</tr>
<tr>
<td>3*</td>
<td>3'a</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>2*</td>
<td>1'</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>1*</td>
<td>4'</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>1*</td>
<td>6'</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>1*</td>
<td>5'</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>1*</td>
<td>3'a</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>1*</td>
<td>3'e</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>4'</td>
<td>6'</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>4'</td>
<td>5</td>
<td>4.40</td>
<td>2.60</td>
<td>0.60</td>
</tr>
<tr>
<td>4'</td>
<td>3'e</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>4'</td>
<td>2'a</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>4'</td>
<td>1'</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>4'</td>
<td>6</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6'</td>
<td>5'</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6'</td>
<td>1'</td>
<td>3.40</td>
<td>1.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6'</td>
<td>6</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>5'</td>
<td>3'a</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>5'</td>
<td>3'e</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>5'</td>
<td>2'a</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>5'</td>
<td>1'</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>5'</td>
<td>6</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>3'a</td>
<td>2'e</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>3'a</td>
<td>1'</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>3'a</td>
<td>6</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>3'e</td>
<td>2'a</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>3'e</td>
<td>2'e</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>3'e</td>
<td>6</td>
<td>2.40</td>
<td>2.60</td>
<td>0.60</td>
</tr>
<tr>
<td>2'a</td>
<td>5</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>2'a</td>
<td>1'</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>2'e</td>
<td>6</td>
<td>4.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>1'</td>
<td>6</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>2.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>4.40</td>
<td>1.60</td>
<td>2.60</td>
</tr>
<tr>
<td>13/9</td>
<td>14</td>
<td>4.40</td>
<td>0.60</td>
<td>2.60</td>
</tr>
<tr>
<td>13/9</td>
<td>18</td>
<td>4.40</td>
<td>0.60</td>
<td>2.40</td>
</tr>
<tr>
<td>13/9</td>
<td>17b</td>
<td>4.40</td>
<td>0.60</td>
<td>2.60</td>
</tr>
<tr>
<td>12/10</td>
<td>17a</td>
<td>3.40</td>
<td>0.60</td>
<td>2.60</td>
</tr>
<tr>
<td>12/10</td>
<td>17b</td>
<td>4.40</td>
<td>0.60</td>
<td>2.60</td>
</tr>
<tr>
<td>18</td>
<td>17a</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>18</td>
<td>17b</td>
<td>3.40</td>
<td>0.60</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table 3.3.2: This table shows the list of NMR experimental ROE/NOE constraints used in the XPLOR-NIH structure determination protocol for amicetin.
<table>
<thead>
<tr>
<th>Serial, Number</th>
<th>Dihedral Angles</th>
<th>Dihedral Angle Constraints</th>
<th>Error value</th>
<th>( ^{1}J )</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H2'a,C2',C3',H3'a</td>
<td>180.0</td>
<td>20</td>
<td>( \checkmark )</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>H2'e,C2',C3',H3'e</td>
<td>60.0</td>
<td>20</td>
<td>( \checkmark )</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>H6,C6,C5,H5</td>
<td>35.0</td>
<td>40</td>
<td>( \checkmark )</td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>H4',C4',C5',H5'</td>
<td>170.0</td>
<td>20</td>
<td>( \checkmark )</td>
<td>X</td>
</tr>
<tr>
<td>5</td>
<td>H5*,C5*,C4*,H4*</td>
<td>180.0</td>
<td>20</td>
<td>( \checkmark )</td>
<td>X</td>
</tr>
<tr>
<td>6</td>
<td>H4*,C4*,C3*,H3*</td>
<td>180.0</td>
<td>20</td>
<td>( \checkmark )</td>
<td>X</td>
</tr>
<tr>
<td>7</td>
<td>H3*,C3*,C2*,H2*</td>
<td>172.0</td>
<td>20</td>
<td>( \checkmark )</td>
<td>X</td>
</tr>
<tr>
<td>8</td>
<td>H2'a,C2',C1',H1'</td>
<td>180.0</td>
<td>30</td>
<td>( \checkmark )</td>
<td>X</td>
</tr>
<tr>
<td>9</td>
<td>H4',C4',C3',H3'a</td>
<td>180.0</td>
<td>30</td>
<td>( \checkmark )</td>
<td>X</td>
</tr>
<tr>
<td>10</td>
<td>H7,C7,C6,H6</td>
<td>55.0</td>
<td>30</td>
<td>( \checkmark )</td>
<td>X</td>
</tr>
<tr>
<td>11</td>
<td>O1*,C4',C3',H3'e</td>
<td>-55.0</td>
<td>30</td>
<td>X</td>
<td>( \checkmark )</td>
</tr>
<tr>
<td>12</td>
<td>O1*,C4',C5',H5'</td>
<td>55.0</td>
<td>30</td>
<td>X</td>
<td>( \checkmark )</td>
</tr>
<tr>
<td>13</td>
<td>O1*,C1*,C2*,H2*</td>
<td>180.0</td>
<td>30</td>
<td>X</td>
<td>( \checkmark )</td>
</tr>
<tr>
<td>14</td>
<td>O2*,C2*,C3*,O3*</td>
<td>55.0</td>
<td>30</td>
<td>X</td>
<td>( \checkmark )</td>
</tr>
<tr>
<td>15</td>
<td>N4*,C4*,C3*,O3*</td>
<td>-55.0</td>
<td>30</td>
<td>X</td>
<td>( \checkmark )</td>
</tr>
<tr>
<td>16</td>
<td>N4*,C4*,C5*,C6*</td>
<td>55.0</td>
<td>30</td>
<td>X</td>
<td>( \checkmark )</td>
</tr>
<tr>
<td>17</td>
<td>H2'e,C2',C1',N1</td>
<td>-55.0</td>
<td>30</td>
<td>X</td>
<td>( \checkmark )</td>
</tr>
</tbody>
</table>

Table 3.3.3: Table showing the list of NMR experimental dihedral constraints, used in the XPLOR-NIH structure determination protocol for amicetin. An “X” indicates that this was not the constraint's origin and a “\( \checkmark \)” indicates that this was it's origin, for example a “\( \checkmark \)” in the \( ^{1}J \) column indicates the constraint was derived directly from coupling constants.
3.3.4 NMR solution structure of amicetin

The final NMR solution structure was calculated, as described in the methodology section 2.6.1 and the resultant structure is shown in Figure 3.3.3.

![Figure 3.3.3: The left image shows the average NMR structure of amicetin. The amicetin structure reveals a number of structurally significant intramolecular hydrogen bonds, the hydrogen bonds are shown by the dotted lines and the distances between donor and acceptor atoms (in Å) are indicated. The right image shows an overlay of the 10 best structures, with an all atom average RMSD of 1.00 Å.]

The structure fits the constraints applied and forms four hydrogen bonds helping to hold the structure in its conformation. The ten best structures shown overlayed in Figure 3.3.3 have a low all atom RMSD of 1.00 Å. The structure also fits well with features observed in the NMR spectra that were not used to generate constraints. For instance, a hydrogen bond is observed between HN14 and N16, offering an explanation as to why the HN14 resonance is retarded against exchange.

It is also of note that the cytosine H5 and H6 chemical shifts (7.50 and 8.24 ppm, respectively) are shifted to low field compared to the chemical shifts for the equivalent protons of both blasticidin S (6.07 and 7.63 ppm respectively) and gougerotin (6.10 and 7.80 ppm respectively). This may be due to the effect of the stacking of the two aromatic rings as seen in the structure.
A number of hydrogen bonds are observed in the amicetin solution structure which influence the conformation of amicetin. Table 3.3.4 details these hydrogen bonds and compares them to those observed in the crystal structure. A significant number of hydrogen bonds are potentially shared between both structures.

<table>
<thead>
<tr>
<th>D-H···A</th>
<th>D···A Solution structure (Å)</th>
<th>D···A Crystal structure (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O3*-H3*···N4*</td>
<td>2.82</td>
<td>-</td>
</tr>
<tr>
<td>O3*-H3*···O2*</td>
<td>2.85</td>
<td>-</td>
</tr>
<tr>
<td>O2*-H2*···O3*</td>
<td>2.85</td>
<td>-</td>
</tr>
<tr>
<td>O2*-H2*···O1*</td>
<td>2.80</td>
<td>-</td>
</tr>
<tr>
<td>N14-HN14···N16</td>
<td>2.65</td>
<td>2.66</td>
</tr>
<tr>
<td>C5-H5···O7</td>
<td>Not observed</td>
<td>2.82</td>
</tr>
</tbody>
</table>

Table 3.3.4: A list of possible hydrogen bonds, where D is the donor atom, H is the hydrogen and A is the acceptor. The distances given are measured between the donor and acceptor atoms for the NMR structure and the free crystal structure, a “-” indicates that this value was not available but that the hydrogen bond is a possibility.31

Table 3.3.5 details important dihedral angles of the solution structure and compares these to those observed in the crystal structure. The dihedral angles were found to be generally very similar throughout the saccharide rings, but vary greatly around the peptide moieties. There was also found to be a very large difference in the conformation of the peptide bond between the two aromatic rings. The crystal structure shows an elongated conformation of amicetin, whereas the solution structure is folded around the N4 to C7 peptide bond, allowing the stacking of the two aromatic rings. The equivalent of the glycosidic dihedral angle (O5'-C1'-N1-C6) also varies significantly by 28.3º. The dihedral angles were observed to differ greatly in the alpha methylserine moiety between the solution and crystal structure. There is a great deal of variation between the 10 best structures obtained through the solution structure determination. This may have been due to the low number of constraints in this section or the moiety could genuinely be fairly mobile, as illustrated by the fact this section is observed as disordered in the crystal structure of the antibiotic (see table 3.3.5).31
<table>
<thead>
<tr>
<th>Dihedral angle</th>
<th>NMR structure (º)</th>
<th>Crystal structure (º)</th>
<th>Δ (º)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O5'-C1'-N1-C6</td>
<td>54.0</td>
<td>25.7</td>
<td>28.3</td>
</tr>
<tr>
<td>C1'-C2'-C3'-C4'</td>
<td>-57.5</td>
<td>-53.6</td>
<td>-3.9</td>
</tr>
<tr>
<td>C2'-C3'-C4'-C5'</td>
<td>58.3</td>
<td>52.7</td>
<td>5.6</td>
</tr>
<tr>
<td>C3'-C4'-C5'-O5'</td>
<td>-59.3</td>
<td>-54.1</td>
<td>-5.2</td>
</tr>
<tr>
<td>C4'-C5'-O5'-C1'</td>
<td>62.8</td>
<td>58.9</td>
<td>3.9</td>
</tr>
<tr>
<td>C5'-O5'-C1'-C2'</td>
<td>-62.0</td>
<td>-62.7</td>
<td>0.7</td>
</tr>
<tr>
<td>O5'-C1'-C2'-C3'</td>
<td>57.9</td>
<td>58.2</td>
<td>-0.3</td>
</tr>
<tr>
<td>C5'-C4'-O1*-C1*</td>
<td>-169.6</td>
<td>-162.8</td>
<td>-6.8</td>
</tr>
<tr>
<td>C3'-C4'-O1*-C1*</td>
<td>70.7</td>
<td>76.4</td>
<td>-5.7</td>
</tr>
<tr>
<td>C4'-O1*-C1*-C2*</td>
<td>-179.6</td>
<td>-163.8</td>
<td>-15.8</td>
</tr>
<tr>
<td>C4'-O1*-C1*-C5*</td>
<td>61.3</td>
<td>72.6</td>
<td>-11.3</td>
</tr>
<tr>
<td>C1*-C2*-C3*-C4*</td>
<td>-59.4</td>
<td>-50.8</td>
<td>-8.6</td>
</tr>
<tr>
<td>C2*-C3*-C4*-C5*</td>
<td>59.0</td>
<td>53.0</td>
<td>6.0</td>
</tr>
<tr>
<td>C3*-C4*-C5*-O5*</td>
<td>-57.5</td>
<td>-55.7</td>
<td>-1.8</td>
</tr>
<tr>
<td>C4*-C5*-O5*-C1*</td>
<td>59.6</td>
<td>58.2</td>
<td>1.4</td>
</tr>
<tr>
<td>C5*-O5*-C1*-C2*</td>
<td>-60.0</td>
<td>-55.2</td>
<td>-4.8</td>
</tr>
<tr>
<td>O5*-C1*-C2*-C3*</td>
<td>58.2</td>
<td>51.6</td>
<td>6.6</td>
</tr>
<tr>
<td>N3-C4-N4-C7</td>
<td>-114.7</td>
<td>175.7</td>
<td>69.6</td>
</tr>
<tr>
<td>C5-C4-N4-C7</td>
<td>65.5</td>
<td>-5.2</td>
<td>70.7</td>
</tr>
<tr>
<td>C4-N4-C7-O7</td>
<td>178.5</td>
<td>18.6</td>
<td>159.9</td>
</tr>
<tr>
<td>C4-N4-C7-C8</td>
<td>-0.2</td>
<td>-159.2</td>
<td>159.0</td>
</tr>
<tr>
<td>N4-C7-C8-C9</td>
<td>-91.2</td>
<td>26.6</td>
<td>-117.8</td>
</tr>
<tr>
<td>N4-C7-C8-C13</td>
<td>87.2</td>
<td>-160.6</td>
<td>-112.2</td>
</tr>
<tr>
<td>C12-C11-N14-C15</td>
<td>90.24</td>
<td>167.4</td>
<td>-77.16</td>
</tr>
<tr>
<td>C10-C11-N14-C15</td>
<td>-90.2</td>
<td>-15.2</td>
<td>-75.0</td>
</tr>
<tr>
<td>C11-N14-C15-O15</td>
<td>0.3</td>
<td>-0.8</td>
<td>1.1</td>
</tr>
<tr>
<td>C11-N14-C15-C16</td>
<td>179.8</td>
<td>-178.9</td>
<td>-1.3</td>
</tr>
<tr>
<td>O15-C15-C16-C18</td>
<td>68.9</td>
<td>53.2</td>
<td>15.7</td>
</tr>
<tr>
<td>O15-C15-C16-C17</td>
<td>-50.22</td>
<td>-64.3</td>
<td>-125.2</td>
</tr>
<tr>
<td>O15-C15-C16-N16</td>
<td>-170.9</td>
<td>174.4</td>
<td>14.7</td>
</tr>
<tr>
<td>N14-C15-C16-N16</td>
<td>9.6</td>
<td>-7.4</td>
<td>17.0</td>
</tr>
<tr>
<td>N14-C15-C16-C18</td>
<td>-110.5</td>
<td>-128.6</td>
<td>18.1</td>
</tr>
<tr>
<td>N14-C15-C16-C17A</td>
<td>160.29</td>
<td>113.8</td>
<td>46.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>103.2</td>
<td>37.1</td>
</tr>
<tr>
<td>O17-C17-C16-C15</td>
<td>-141.6</td>
<td>174.0</td>
<td>44.4</td>
</tr>
<tr>
<td>O17-C17-C16-N16</td>
<td>-19.2</td>
<td>-65.7</td>
<td>46.5</td>
</tr>
<tr>
<td>O17-C17-C16-C18</td>
<td>123.8</td>
<td>56.8</td>
<td>-54.8</td>
</tr>
</tbody>
</table>

Table 3.3.5: A list of important dihedral angles. The dihedral angles are given for the NMR structure and the free crystal structure and the difference between them is indicated in the delta (Δ) column, where the difference is greater than 10 degrees the cell is coloured yellow and where the difference is greater than 25 degrees the cell is coloured red. Where the columns are split this indicates that this region was observed as disordered in the crystal structures and so two sets of dihedral angles are given.31
The amicetin NMR structure determined previously by C. Shammas was distinctly different from the structure determined in this project. This is as a greater number of constraints were used in the later structure determination. Originally 28 constraints were used in total, for the later structure determination 77 constraints were used. These additional constraints helped to pin down the structure and also to remove the sterochemical problems. The later structure bears more similarity to the solution structures observed for the analogous antibiotics blasticidin S and gougerotin and also to its own X-ray crystal structure.
3.4 Discussion and comparison of the structure and dynamics of blasticidin S, gougerotin, and amicetin

In the light of the three high quality NMR structures which have been determined, it is worthwhile to spend some time comparing some of their important features. As can be seen in Figure 3.4.1, all three antibiotics show conformational similarities, particularly around the cytosine and the attached saccharide type rings. There also seems to be some similarity in the location of certain electronegative atoms, particularly between gougerotin and amicetin.

![Figure 3.4.1: The left image shows the average NMR structure of blasticidin S, the centre image shows the average NMR structure of gougerotin and the right image shows the average NMR structure of amicetin. The structures reveal structurally significant intramolecular hydrogen bonds, the hydrogen bonds are shown by the dotted lines and the distances between donor and acceptor atoms (in Å) are indicated.](image)

The common dihedral angles between the three antibiotics are compared in Table 3.4.1. These angles are centred around the saccharide type ring which is attached to the cytosine moiety. The equivalent of the glycosidic dihedral angle (O5'-C1'-N1-C6) is similar throughout the antibiotics, the largest difference being between amicetin and blasticidin S. This could be as amicetin has a second aromatic ring and peptide moiety which may cause the dihedral angle to rotate. Also gougerotin has the potential of a hydrogen bond between O2 and HO2', which may also promote rotation. Neither of these factors exist for blasticidin S.
The dihedral angles (2-7 Table 3.4.1) of the nucleoside saccharide ring are very similar between amicetin and gougerotin, however due to the presence of a double bond in the blasticidin S pseudo saccharide ring, the dihedral angles are quite dissimilar compared to the two other antibiotics.

<table>
<thead>
<tr>
<th>Serial Number</th>
<th>Dihedral angle</th>
<th>Blasticidin S (º)</th>
<th>Gougerotin (º)</th>
<th>Amicetin (º)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O5'-C1'-N1-C6</td>
<td>41.6</td>
<td>47.4</td>
<td>54.0</td>
</tr>
<tr>
<td>2</td>
<td>C1'-C2'-C3'-C4'</td>
<td>-1.6</td>
<td>-59.3</td>
<td>-57.5</td>
</tr>
<tr>
<td>3</td>
<td>C2'-C3'-C4'-C5'</td>
<td>22.3</td>
<td>58.3</td>
<td>58.3</td>
</tr>
<tr>
<td>4</td>
<td>C3'-C4'-C5'-O5'</td>
<td>-55.0</td>
<td>-58.3</td>
<td>-59.3</td>
</tr>
<tr>
<td>5</td>
<td>C4'-C5'-O5'-C1'</td>
<td>71.6</td>
<td>62.2</td>
<td>62.8</td>
</tr>
<tr>
<td>6</td>
<td>C5'-O5'-C1'-C2'</td>
<td>-48.7</td>
<td>-62.0</td>
<td>-62.0</td>
</tr>
<tr>
<td>7</td>
<td>O5'-C1'-C2'-C3'</td>
<td>12.8</td>
<td>59.0</td>
<td>57.9</td>
</tr>
<tr>
<td>8</td>
<td>C5'-C4'-N6'-C7</td>
<td>-104.0</td>
<td>-162.8</td>
<td>NA</td>
</tr>
<tr>
<td>9</td>
<td>C3'-C4'-N6'-C7</td>
<td>134.6</td>
<td>76.2</td>
<td>NA</td>
</tr>
<tr>
<td>10</td>
<td>C4'-N6'-C7-C8</td>
<td>176.2</td>
<td>175.6</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 3.4.1: A comparison of important dihedral angles for the NMR structures of amicetin, blasticidin S and gougerotin.

Further, it is interesting to note that the free gougerotin solution structure shows distinct similarity to the crystal structure of the ribosome bound blasticidin S in its major binding site, as can be seen in Figure 3.4.2. The locations of the electronegative atoms are interestingly more comparable between bound blasticidin S and free gougerotin, then it is between free blasticidin S and free gougerotin. The free solution structure of amicetin is also more similar to the bound blasticidin S than to the free NMR solution structure.
Figure 3.4.2: The left image shows the crystal structure of the ribosome bound blasticidin S in its major binding site, the right image shows the average NMR structure of gougerotin. The structures reveal structural significant intramolecular hydrogen bonds, the hydrogen bonds are shown by the dotted lines and the distances between donor and acceptor atoms (in Å) are indicated.61

From the NMR structures determined, it is clear to see that the chemical structure of each antibiotic can be broken into five components, as shown in the top half of Figure 3.4.3. Different variants of these sections can be taken from the functional antibiotics of the aminohexose cytosine nucleoside group discussed in section 1.4 and the conserved structures discussed above. If these are recombined, over 1700 potential new antibiotics can be produced. Using the knowledge gained from the structure determination of these antibiotics and binding studies, structures more likely to share peptidyl transferase antibiotic activity can be selected. The lower half of Figure 3.4.3 illustrates this methodology, suggesting a potential new hybrid antibiotic which could potentially be synthesised. Some of these antibiotics could have more favourable qualities and could be clinically, agriculturally or scientifically useful.
Figure 3.4.3: The upper scheme illustrates the major parts that the antibiotics can be broken into, the number of variants of each category from the functional antibiotics which are discussed in section 1.4 are indicated in the brackets. When the different parts are reassembled in all possible combinations, it leads to a predicted total of 1728 potential antibiotics! The lower structure shows a hypothetical hybrid antibiotic produced by following the above method. The different parts of the new hybrid antibiotic are labelled.

Comparison of the dynamics

To compare the dynamics of the three antibiotics, the effects on the 1D $^1$H NMR spectra of temperature can be compared. For amicetin, the analysis of the 1D $^1$H temperature series was carried out by C. Shammas.$^{32,66}$
The most striking exchangeable resonance in blasticidin S is the exchange retarded H6' resonance, which is retarded up to at least 50°C. The equivalent H6' proton of gougerotin behaves similarly except that it is only retarded up to 40°C and the peak appears generally much broader throughout the temperature series when compared to blasticidin S. This indicates that the H6' protons exchange rate in gougerotin is considerably faster than in blasticidin S. The likely reason for this is illustrated in the solution structures, as the H6' of blasticidin has a definite hydrogen bond whereas the H6' of gougerotin has no specific hydrogen bonds, but it is located pointing into a loop with plenty of hydrogen bond acceptors nearby which could potentially form a hydrogen bond with a small amount of rearrangement. It may also be possible that a water molecule maybe involved in hydrogen bonding within the ring. It should however be noted that the temperature series of blasticidin S was carried out at a slightly lower pH 4.8 as compared to 6.2 for gougerotin. This retardation to exchange of each of these protons indicates a great stability with respect to temperature, at least in the aspect of the structure that is responsible for the retardation of the H6' protons.

A further two comparable resonances between blasticidin S and gougerotin that display interesting behaviour upon temperature change are the ones due to the exchangeable H4a and H4b of the cytosine moiety. These were observed to behave almost identically in both blasticidin S and gougerotin, both sets of resonances are retarded to exchange up to 20-30°C.
Chapter 4

NMR structure determination of isotopically normal and labelled 23S rRNA motifs

The aim of this chapter is to describe the NMR assignment and structure determination of the conserved E. coli and H. h. 29-mer RNA motifs, and to describe the assignment of the H. h. 37-mer uniformly $^{13}$C and $^{15}$N labelled RNA. The structure determination of these motifs allows the study of their binding to the peptidyl transferase antibiotics, in particular amicetin. The structure determination of the motifs also allows observation of this important motif of the ribosome in a more natural solution state, as opposed to the current crystal structures of the intact ribosome. Amicetin's binding to the 29-mer motifs is described in chapter 5.

The predicted secondary structures for the three RNA motifs used for NMR investigation are shown in Figure 4.0.1. They were determined by use of the Dinamelt Server$^{117,118}$ and by comparison to the secondary structure observed for the motif in various ribosome crystal structures.

The residues of the various motifs have been numbered, based on the original single stranded 35-mer RNA motif previously studied in our lab. This leads to the labels of “a” and “b” for the first two residues of the 37-mer RNA motif. This numbering scheme is illustrated in Figure 4.0.1

The single stranded H. h. 37-mer uniformly $^{13}$C and $^{15}$N labelled RNA motif was produced as a 37-mer rather than a labelled 29-mer duplex, for reasons of cost and ease of synthesis. It was also thought that a single stranded structure would be more stable. To further stabilise the structure an additional end base pair aG-bC was added for similar reasons.
Figure 4.0.1: The left image shows the *E. coli* 29-mer RNA (duplex) secondary structure, the centre image shows the *H. h.* 29-mer RNA (duplex) secondary structure and the right image shows the uniformly $^{13}$C and $^{15}$N labelled *H. h.* 37-mer RNA (single stranded) secondary structure. The residue numbering scheme used for assignment is indicated on the secondary structures. The single and double solid lines represent standard A-U and G-C W-C base pairs respectively.
4.1 Assignment and structure determination of isotopically normal \textit{E. coli} 29-mer RNA

4.1.1 Assignment of the exchangeable proton resonances

The 1D spectrum at 2\textdegree{}C shows seven clear imino (NH) peaks in the standard W-C base pairing region of the spectrum (see section 2.4.2). However, when the 1D temperature series was analysed it was clear that there was overlap and that there are in fact nine resonances in this region. There are no resonances in the region of the spectrum associated with non-canonical base pairs. From the predicted secondary structure, 13 resonances would be expected in the region of the spectrum associated with W-C base pairs. It would be quite likely that imino protons located in the bulge or fraying ends would not be observed. It could be expected that the U6 imino proton may be observed in the region of the spectrum associated with non-canonical base pairs, but it may simply have been exchanging too rapidly with solvent to be observed.

The imino protons of the \textit{E. coli} 29-mer RNA were then assigned by following the imino to imino sequential NOE connectivities, as illustrated in Figure 4.1.1. Through this method, the majority of the imino protons were assigned. The imino protons that were not assigned are located adjacent to the bulges and at the fraying ends of the RNA and so are likely to have been exchanging too rapidly with the solvent to be observed. The fact that these imino resonances are missing helped confirm the predicted secondary structure.
Figure 4.1.1: 600 MHz NOESY (τ_m = 250 ms) spectrum of E. coli 29-mer RNA (1 mM), dissolved in 90% 
\( ^1\text{H}_2\text{O} + 10\% ^2\text{H}_2\text{O} \) containing 20 mM PO_4^{3-} (pH 6.2), 20 mM NaCl at 2°C. The E. coli 29-mer RNA secondary structure is also shown on the right. The spectrum shows the identification of the imino to imino through space sequential connectivities and the sequence specific assignment is shown at the top of the spectrum. The sequential connectivities are indicated both on the spectrum and on the sequence with two sets of corresponding arrows, single headed arrows for the segment of sequential walk from U13 to G25, and double headed arrows for the segment of sequential walk from G35 to G32.

It was discovered, by using the amino to imino region (Figure 4.1.2), that the U12 and U13 imino protons overlapped as shown in Figure 4.1.1. The three main indications of this are firstly; the presence of two characteristic NOE peaks between the uracil imino protons and the base paired adenine H2s, secondly; that there appeared to be two separate lines of peaks at two very similar, but distinct, imino proton chemical shifts and finally; that in the 1D temperature series the peak splits in to two at 10°C.

The imino to imino sequential NOE connectivities fall into two separate sections, both sets of connectivities are clear to follow, with distinct cross peaks, the uracil and guanine imino protons clearly fall into their characteristic chemical shift range. The first series of
sequential connectivities (single arrow head, Figure 4.1.1) starts at the overlapping U13 and U12 imino resonances and a strong NOE connectivity links it to G24, which is in turn linked to G25 via another strong NOE. This first series then ends as the bulge region is reached. The second series of sequential connectivities (double arrow head, Figure 4.1.1) starts at the G35 resonance and is linked to G34 via a very weak NOE peak, as can be expected as there is likely to be fast exchange of the imino protons with solvent water at the fraying ends of the RNA. The G34 imino proton was then linked to G33 and G33 to G32 by two strong NOE connectivities. Again the bulge region was reached and the pattern of sequential connectivities disappeared.

The amino protons were then assigned following the method described in section 2.4.2, using the imino proton assignment as a basis. The assignment is shown in Figure 4.1.2.
Figure 4.1.2: 600 MHz NOESY (τm=250 ms) spectrum of *E. coli* 29-mer RNA (1 mM), dissolved in 90% 
1H2O + 10% 2H2O containing 20 mM PO4³− (pH 6.2), 20 mM NaCl at 2°C. The spectrum shows the 
identification of the imino to amino through space connectivities. Their sequence specific assignment is 
shown at the top (imino) and side (amino) of the spectrum. Dotted lines are drawn across the spectrum at the 
amino chemical shifts with ‘X’s marking the the through space correlations to the imino protons. The *E. coli*
29-mer RNA secondary structure is also shown.
The imino proton assignment has further been confirmed using the imino to aromatic proton correlations, as shown in Figure 4.1.3. The aromatic proton assignment, described later in section 4.1.2, was used here to confirm both the imino proton and aromatic assignment. A significant number of inter strand imino to aromatic NOE connectivities were observed, as well as a number of sequential NOE connectivities. These observed NOE connectivities fit the pattern expected and so help confirm not only the assignment of the protons involved, but also the predicted secondary structure.
Figure 4.1.3: The large left panel shows the 600 MHz NOESY ($\tau_m=250$ ms) spectrum of E. coli 29-mer RNA (1 mM), dissolved in 90% $^1$H$_2$O + 10% $^2$H$_2$O containing 20 mM PO$_4$$^{3-}$ (pH 6.2), 20 mM NaCl at 2°C. The spectrum shows the identification of the imino to aromatic through space connectivities. Their sequence specific assignment is shown at the top (imino) and side (aromatic) of the spectrum. Dotted lines are drawn across the spectra at the aromatic chemical shifts with 'X's marking the through space correlations to the imino protons. The two right panels show the 700 MHz $^1$H-$^1$H TOCSY ($\tau_m=75$ ms) spectrum of the same sample, under the same conditions, showing the $^3$$J_{HH}$ H5-H6 correlation due to the cytosine and uracil bases of the RNA. The E. coli 29-mer RNA secondary structure is also shown.
The exchangeable proton assignment confirms thus far the sequence of the RNA motif. The imino proton resonances fall into their respective characteristic chemical shift ranges, and the NOE cross peaks observed connect protons together in the expected manner. The predicted secondary structure is also confirmed by the imino to imino connectivities observed in the stem regions of the RNA and their absence adjacent to and in the bulge regions.

4.1.2 Assignment of the non-exchangeable proton and carbon resonances

The basis of the method for the sequence specific assignment of the RNA motif was to follow the characteristic sequential NOE connectivities observed between the H6/8 and H1' resonances in the NOESY spectrum, as described in section 2.4.2.

Identification of resonance types

The resonances due to the H5/H6 protons of the pyrimidine bases C and U were first identified using the DQF-COSY spectrum, by locating the H5-H6 scalar coupled cross peaks. The H5/H6 resonances were then classified as to whether they were from uracil or cytosine bases, using the characteristic C5 chemical shift as illustrated in Figure 4.1.4, the sequence specific assignment is described later. For most resonances it is easy to classify them as either from cytosine or uracil bases. However, due to features such as the overlap in the H5 chemical shifts for the residues C4 and U6, this is not as straight forward for a small number of residues. Their classification as either cytosine or uracil resonances has to be confirmed via sequential NOE connectivities observed in the NOESY spectra. It is also interesting to note that the bulge residue C28's H5-H6 cross peak in the DQF-COSY is of particularly low intensity. This could indicate a degree of flexibility in this region.
Figure 4.1.4: The Right panel shows the 800 MHz DQF-COSY spectrum of the *E. coli* 29-mer RNA motif (1 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4^{3-}$ (pH 6.2), 20 mM NaCl at 25°C. The spectrum shows the identification of the aromatic cytosine and uracil protons H5 and H6 based on the observed scalar coupled ($^3$J) cross peaks. Their sequence specific assignment is shown at the top (H6) and side (H5) of the spectrum. The left and middle panels show the 700 MHz $^1$H-$^13$C HSQC spectrum of the same sample, under the same conditions, showing the $^1$J$_{H-C}$ H5-C5 correlation due to the cytosine (left) and uracil (middle) bases of the RNA respectively. Dotted lines mark the H5 chemical shifts, the 29-mer cross peaks are marked with ‘X’s and the impurity peaks are marked with black dots.

The H2 proton resonances appear in the same region of the spectrum as the H6 and H8 resonances. Therefore, in order to prevent confusion, the $^1$H-$^13$C HSQC was then used to identify the H2 resonances. The $^1$H-$^13$C HSQC was used as the H2 to C2 correlations fall in a separate characteristic region of the spectrum, as shown in Figure 4.1.5, making their identification clear.
Figure 4.1.5: 700 MHz $^1$H-$^{13}$C HSQC spectrum of the E. coli 29-mer RNA motif (1 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4$$^3−$ (pH 6.2), 20 mM NaCl at 25ºC. The spectrum shows the identification of the H2 adenine protons based on the observed scalar coupled ($^1$J$_{H-C}$) cross peaks. Their sequence specific assignment is shown at the top (H2). Dotted lines mark the H2 chemical shifts, the 29-mer correlations are marked with ‘X’s and the impurity peaks are marked with black dots.

The $^1$H-$^{13}$C HSQC was used to identify the H8 protons through the H8 to C8 correlations, but these peaks occur in the same region as the H6-C6 correlations. Fortunately as the H6 resonances were already identified, the two groups were easily separated. The H1'-C1' correlations also appear in a characteristic region and so are easily identified (see Figures 4.1.6 and 4.1.7).

Sequential assignment

The H1'-H6/8 sequential connectivities, as described in section 2.4.2, could then be followed fully along both strands despite the bulges and this allows sequence specific assignment, as shown in Figure 4.1.6 for the first strand (C1-U14) and Figure 4.1.7 for the second (A21-G35). For example, if the sequential connectivities from the G32 intra NOE peak are taken they can be clearly traced, first to the G32-G33 inter NOE peak, then to the G33 intra peak, subsequently to the G33-G34 inter peak, then to the G34 intra peak, then to the G34-G35 inter peak and finally to the G35 intra peak.
Figure 4.1.6: The central panel shows the 700 MHz NOESY spectrum ($\tau_m=250$ ms) of the *E. coli* 29-mer RNA motif (1 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4$$^3-$ (pH 6.2), 20 mM NaCl at 25ºC. The spectrum shows the pattern of sequential connectivities which starts at the C1 intra peak (marked start on centre panel) and can be followed all the way to the U14 intra peak (marked end). The solid black line with arrows marks the path of the sequential connectivities. Their sequence specific assignment is shown at the top (H6/8) and side (H1') of the spectrum. The top and right hand panel show the 700 MHz $^1$H-$^13$C HSQC spectrum of the same sample, under the same conditions, showing the the $^1J_{H-C}$ H6/8-C6/8 correlations and the $^1J_{H-C}$ H1'-C1' correlations respectively. The *E. coli* 29-mer RNA secondary structure is also shown.
Figure 4.1.7: The central panel shows the 700 MHz NOESY spectrum ($\tau_m=250$ ms) of the *E. coli* 29-mer RNA motif (1 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4$ (pH 6.2), 20 mM NaCl at 25ºC. The spectrum shows the pattern of sequential connectivities which starts at the A21 intra peak (marked start on centre panel) and can be followed all the way to the G35 intra peak (marked end). The solid black line with arrows marks the path of the sequential connectivities. Their sequence specific assignment is shown at the top (H6/8) and side (H1') of the spectrum. The top and right hand panel show the 700 MHz $^1$H-$^13$C HSQC spectrum of the same sample, under the same conditions, showing the the $^1$J$_{H-C}$ H6/8-C6/8 correlations and the $^1$J$_{H-C}$ H1'-C1' correlations respectively. The *E. coli* 29-mer RNA secondary structure is also shown.

The identified H2 protons were also assigned sequence specifically by identifying their through space connectivities. The process of assigning the H2 resonances also aided the identification of the H1', H6 and H8 resonances, through characteristic NOE connectivities.

Overall, the peaks in this H1' to aromatic sequential region are clear and well separated, which aided the sequence specific assignment. However, there are regions of overlap, particularly around 7.8 ppm (in the H1' - H6/H8 region of the spectrum), which made following the sequential NOE connectivities more difficult. The overlap of the G24 and
G25 H8 resonances also caused a particular puzzle, but once realised, it was clear to see and allowed progress in the assignment. Also it was observed that the chemical shifts observed for residues at the fraying ends and in the bulge region had a higher tendency to lie at the extremes of the chemical shift range. For example C1 H1' and H6, A8 H1', A27 H1' and H8 and A21 H8.

Further assignment

Once the sequence specific assignment of the H1's, H6s, and H8s were complete, the sequence specific assignment could be expanded to other proton groups. The sequence specific H5 assignment was simply done by utilising H5 to H6 correlations observed in the DQF-COSY, shown previously in Figure 4.1.4. The same was done for H1' to H2' correlations observed in the DQF-COSY, shown in Figure 4.1.8. This was complicated as ribose rings with C$_{3'}$-endo conformation, which is standard for A-form RNA, have a very small $^3J_{H1'-H2'}$ coupling constant (~1.3 Hz). Not only is the magnetisation less efficiently transferred than for a large coupling constant, but as the linewidth tended to be greater than the coupling constant, the positive and negative aspects of the peak began to cancel each other out; as a result not all H1' to H2' correlations were observed. A number of high intensity peaks were observed, suggesting a large $^3J_{H1'-H2'}$ coupling constant, which in turn suggests a possible C$_{2'}$-endo conformation ($^3J_{H1'-H2'} = 7.6$ Hz). This information not only helps confirm the assignment, as the large peaks are all at or next to end residues or bulges, but also the overall conformation of the RNA.
Figure 4.1.8: 800 MHz DQF-COSY spectrum of the *E. coli* 29-mer RNA motif (1 mM), dissolved in 100% \(^2\text{H}_2\text{O}\) containing 20 mM PO\(_4^{3-}\) (pH 6.2), 20 mM NaCl at 25°C. This spectrum shows how the sequence specific H1' assignment was used to gain sequence specific H2' assignment through observed scalar coupled (\(^3J\)) cross peaks. Their sequence specific assignment is shown at the top (H1') and side (H2') of the spectrum. Dotted lines mark the H2' chemical shifts, the 29-mer cross peaks are marked with ‘X’s and the impurity peaks are marked with black dots.

The remaining ribose sugar protons and carbons were difficult to assign due to overlap and thus were not completely assigned. A number of methods were used to gain as many assignments as possible, these methods are outlined in section 2.4.2. The most successful method was to observe the aromatic proton to ribose protons through space connectivities and the H1' to ribose proton through space connectivities, and the relative intensities of the peaks observed in the NOESY spectra. These were then correlated to the relevant regions of the \(^1\text{H}-\text{\textsuperscript{13}C}\) HSQC spectra where the different type of ribose carbons fall into different chemical shift ranges.
The assignment of the non exchangeable proton and carbon resonances has further confirmed the secondary structure through the sequence specific assignment due to a number of inter strand NOEs being observed. The overall conformation has also been indicated as the NOE connectivity patterns for the non-bulge regions matched that expected of A-form RNA, this is further confirmed by the peak intensities of the H1'-H2' correlations in the DQF-COSY.

### 4.1.3 Assignment of the phosphorus resonances

Assignment of the phosphorus resonances was completed by using a $^1$H-$^{31}$P CPMG-HSQC-NOESY spectrum as described in section 2.3.12 and 2.4.2. Only a partial pattern of sequential connectivities were observed, but with the already completed sequence specific proton assignment, it was enough to obtain the majority of the assignments. Figures 4.1.9 illustrates this. The $^{31}$P assignment is shown in Figure 4.1.10 on the 1D $^{31}$P spectra. A pattern is easily observed from this as the purine and pyrimidine residues fall clearly into different regions, with residues which occur in or near bulges residing at the extremities of those regions.
Figure 4.1.9: The central panel shows the 700 MHz NOESY spectrum ($\tau_m=250$ ms) of *E. coli* 29-mer RNA (1 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4^{3-}$ (pH 6.2), 20 mM NaCl at 25ºC. The spectrum shows the pattern of sequential connectivities which starts at the A21 intra peak (marked start on centre panel) and can be followed all the way to the G35 intra peak (marked end). The solid black line with arrows marks the path of the sequential connectivities. Their sequence specific assignment is shown at the top (H6/8) and side (H1') of the spectrum. The top and right hand panel 600 MHz $^1$H-$^31$P CPMG-HSQC-NOESY ($\tau_m=500$ ms) spectrum of the same sample, under the same conditions, showing the H6/8 to phosphorus correlations and the H1' to phosphorus correlations respectively. The *E. coli* 29-mer RNA secondary structure is also shown.
**Figure 4.1.10**: 81 MHz 1D $^{31}$P (200 MHz, $^1$H) spectrum of the *E. coli* 29-mer RNA motif (1 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4^{3-}$ (pH 6.2), 20 mM NaCl at 25°C. The sequence specific $^{31}$P assignment is indicated along the top of the spectrum.
4.1.4 Assignment table of the *E. coli* 29-mer RNA motif

The $^1$H, $^{13}$C and $^{31}$P chemical shifts of the *E. coli* 29-mer motif are shown in Table 4.1.1.

Table 4.1.1 The $^1$H, $^{13}$C and $^{31}$P chemical shifts of the *E. coli* 29-mer RNA, based on the assignment of the exchangeable (measured at 2°C) and non-exchangeable proton, carbon and phosphorus (measured at 25°C) resonances described in this chapter. A “~” indicates that the chemical shift was not obtained during assignment.
4.1.5 NMR geometrical constraints of the \textit{E. coli} 29-mer RNA motif

\textbf{NOE constraints}

Following the methods outlined in 2.5.3, distance constraints were generated from the various NOESY spectra. The observed NOE cross peaks are classified into a number of groups: exchangeable, non-overlapped, overlapped, very overlapped, H2 and A21. The exchangeable group consisted of constraints generated from the NOEs of the exchangeable protons, measured from the $^1\text{H}^1\text{H}$ NOESY spectrum measured in 90\% $^1\text{H}_2\text{O}$ and 10\% $^2\text{H}_2\text{O}$ with a mixing time of 50 ms at 2ºC.

The non-overlapped, overlapped and very overlapped groups consist of constraints generated from NOEs measured from the $^1\text{H}^1\text{H}$ NOESY spectrum measured in 100\% $^2\text{H}_2\text{O}$ with a mixing time of 100 ms at 25ºC. The NOEs are then divided into each group dependent on how overlapped the NOE in question was with other peaks in the spectrum. For example an NOE which does not overlap with any other peak is placed in the non-overlapped group, an NOE which is in a heavily overlapped region is placed in the very overlapped group and those NOEs with only some overlap with other peaks are placed in the overlapped group. This was done so that tighter error bounds could be used on the non-overlapped NOEs and conversely so that the upper limits of the error bounds could be increased for the groups containing overlapped NOEs. The error bounds were increased due to the apparent increase in intensity that these NOEs had due to overlap with other peaks.

The H2 group was used for constraints generated from the same spectrum as the overlapped groups, but only contained NOEs from H2 protons, as it was found that these
NOEs appeared to have higher intensities relative to their inter proton distances and so were placed in a separate group. The distance constraints from this NOE set were calculated separately with higher upper error bounds. The final group was the A21 group which consists of NOEs again taken from the same spectrum as the overlapped groups, but involving NOEs from the terminal A21 residue. NOEs involving this residue also appear to have a high intensity and so again were treated separately, and given a different reference intensity and a slightly wider error bound range.

Table 4.1.2 lists the parameters used to turn the NOE peaks into constraints and the number of constraints in that group, for each of the NOE groups. The non-ideal distance function of $i^{-1/4}$ (where $i$ is intensity) was used for the non-exchangeable NOEs as this appeared to give constraints which better fitted a range of reference distances in the canonical regions of the RNA. This non-ideal behaviour may be due to a small amount of spin diffusion.

<table>
<thead>
<tr>
<th>NOE group</th>
<th>Reference intensity</th>
<th>Reference distance (Å)</th>
<th>Upper/lower fractional error</th>
<th>Distance function</th>
<th>Number of constraints generated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchangeable</td>
<td>$6.04 \times 10^7$</td>
<td>3.0</td>
<td>0.40/0.40</td>
<td>$i^{-1/6}$</td>
<td>43</td>
</tr>
<tr>
<td>Non-overlapped</td>
<td>$2.42 \times 10^8$</td>
<td>3.7</td>
<td>0.38/0.38</td>
<td>$i^{-1/4}$</td>
<td>194</td>
</tr>
<tr>
<td>overlapped</td>
<td>$2.42 \times 10^8$</td>
<td>3.7</td>
<td>0.60/0.38</td>
<td>$i^{-1/4}$</td>
<td>153</td>
</tr>
<tr>
<td>Very overlapped</td>
<td>$2.42 \times 10^8$</td>
<td>3.7</td>
<td>1.20/0.38</td>
<td>$i^{-1/4}$</td>
<td>67</td>
</tr>
<tr>
<td>“H2”</td>
<td>$2.42 \times 10^8$</td>
<td>3.7</td>
<td>1.00/0.38</td>
<td>$i^{-1/4}$</td>
<td>26</td>
</tr>
<tr>
<td>“A21”</td>
<td>$7.92 \times 10^8$</td>
<td>3.7</td>
<td>0.40/0.4</td>
<td>$i^{-1/4}$</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 4.1.2: The table details information on the six groups of NOEs and the parameters used in CcpNmr Analysis to produce the NOE based distance constraints.

The exchangeable and non-exchangeable inter residue NOE constraints used are illustrated in Figure 4.1.11 and 4.1.12. A number of key inter strand connectivities are seen; some of the most important constraints are found in the non-exchangeable group and are between G5 and G7, indicating that U6 is flipped out of the bulge. There are constraints between A27 and U29 indicating that C28 may be flipped out, however there is also a constraint
from C28 H1' to A7 H2, suggesting the opposite. There are additionally a number of constraints between A9 and A27, also suggesting that A27 is likely to be held within the helix.

**Figure 4.1.11:** The inter residue “exchangeable” (i.e. extracted from the NOESY spectrum measured in 90% $^{1}$H$_2$O + 10% $^{2}$H$_2$O) NOE distance constraints used in the structure determination calculation, are shown by lines drawn between the protons of the various residues.
The inter-residue “non-exchangeable” (i.e. extracted from the NOESY spectrum measured in $^2$H$_2$O) NOE distance constraints used in the structure determination calculation, are shown by lines drawn between the protons of the various residues.

**Torsion angle constraints**

Following the methodology set out in 2.5.4, torsion angle constraints were produced. Whether the different residues were constrained as C$_2$-endo, C$_3$-endo or left unconstrained was dependent on the approximate $^3J$ H1' to H2' coupling constant measured from the DQF-COSY (C$_2$-endo $\sim$7.6 HZ, C$_3$-endo $\sim$1.3 Hz) and on “can1”, which is a function of the C1', C4' and C5' chemical shifts (see 2.5.4). These values are displayed in Table 4.1.3, along with the result of how the residue was constrained.
<table>
<thead>
<tr>
<th>Residue</th>
<th>Approximate $^1$J$_{H1'H2'}$ (Hz)</th>
<th>$can1$ (ppm)</th>
<th>Ribose constrained as</th>
<th>H2'-H6/8 NOE intensity &gt; H1'-H6/8 NOE intensity?</th>
<th>H3'-H6/8 NOE intensity &gt; H1'-H6/8 NOE intensity?</th>
<th>Glycosidic angle ($\chi$) constrained as</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Small</td>
<td>-5.7</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>C2</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>C3</td>
<td>Small</td>
<td>-5.63</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>C4</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>G5</td>
<td>Large</td>
<td>-6.04</td>
<td>Unconstrained</td>
<td>Yes</td>
<td>No</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>U6</td>
<td>Small</td>
<td>~</td>
<td>Unconstrained</td>
<td>Yes</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>G7</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>A8</td>
<td>Small</td>
<td>-5.89</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>A9</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>C10</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>No</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>C11</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>U12</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>U13</td>
<td>Large</td>
<td>-5.79</td>
<td>Unconstrained</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>U14</td>
<td>Large</td>
<td>-6.49</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>A21</td>
<td>Large</td>
<td>-6.32</td>
<td>C$_3$-endo</td>
<td>No</td>
<td>Yes</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>A22</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>A23</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>G24</td>
<td>Small</td>
<td>-5.69</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>G25</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>U26</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>A27</td>
<td>Large</td>
<td>~</td>
<td>Unconstrained</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>C28</td>
<td>Small</td>
<td>~</td>
<td>Unconstrained</td>
<td>Yes</td>
<td>No</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>U29</td>
<td>Small</td>
<td>-6.44</td>
<td>Unconstrained</td>
<td>Yes</td>
<td>Yes</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>C30</td>
<td>Large</td>
<td>~</td>
<td>Unconstrained</td>
<td>No</td>
<td>Yes</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>C31</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>G32</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>G33</td>
<td>Small</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>G34</td>
<td>Small</td>
<td>-5.58</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>G35</td>
<td>Large</td>
<td>-6.08</td>
<td>Unconstrained</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
</tbody>
</table>

Table 4.1.3: The second column of the table indicates the approximate size of the $^1$J$_{H1'H2'}$ coupling constants (small ~1.3 Hz, large ~7.6 Hz), the third column indicates the value of “$can1$” for each residue and how the ribose of the residue was constrained based on this information is displayed in the fourth column. The fifth and sixth columns indicate whether the H2' to H6/8 NOEs are greater in intensity than the corresponding H1' to H6/8 and whether H3' to H6/8 NOEs are greater that the corresponding H1' to H6/8 respectively. The final column indicates how the glycosic ($\chi$) torsion angle was constrained based on the information in columns 5 and 6. A “$can1$” value of greater than -6.25ppm indicates a C$_3$-endo conformation, less than -6.25 indicates a C$_2$-endo conformation, a small coupling constant indicates C$_3$-endo and as a large coupling constant a C$_2$-endo conformation. When the H2' to H6/8 and or H3' to H6/8 NOEs have a greater intensity than the corresponding H1' to H6/8 NOE, this indicates a glycosidic torsion angle with an anti conformation. A “~” indicates that the information was not available or was unclear.

The glycosidic torsion angle was determined by observing intra residue H6/8 to H1', H2' and H3' NOEs as described in section 2.5.4. Table 4.1.3 details which residues were constrained as anti and which were left unconstrained.
For residues of the RNA which were constrained as C$_3$-endo, which were also in sections of the RNA that have standard W-C base pairing and where the $^{31}$P were within the standard range (approximately -3.75 to -4.80 ppm), the backbone was constrained with standard A-form RNA constraints.

Hydrogen bond constraints were set following the method detailed in 2.5.5, Table 4.1.4 details which base pairs were constrained.

<table>
<thead>
<tr>
<th>Base pair</th>
<th>Imino proton chemical shift ($\delta$, ppm)</th>
<th>Base pair constrained?</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-G35</td>
<td>13.24</td>
<td>Yes</td>
</tr>
<tr>
<td>C2-G34</td>
<td>12.93</td>
<td>Yes</td>
</tr>
<tr>
<td>C3-G33</td>
<td>12.62</td>
<td>Yes</td>
</tr>
<tr>
<td>C4-G32</td>
<td>12.44</td>
<td>Yes</td>
</tr>
<tr>
<td>G5-C31</td>
<td>~</td>
<td>No</td>
</tr>
<tr>
<td>G7-C30</td>
<td>~</td>
<td>No</td>
</tr>
<tr>
<td>A8-U29</td>
<td>~</td>
<td>No</td>
</tr>
<tr>
<td>A9-U26</td>
<td>~</td>
<td>No</td>
</tr>
<tr>
<td>C10-G25</td>
<td>13.31</td>
<td>Yes</td>
</tr>
<tr>
<td>C11-G24</td>
<td>13.00</td>
<td>Yes</td>
</tr>
<tr>
<td>U12-A23</td>
<td>14.01</td>
<td>Yes</td>
</tr>
<tr>
<td>U13-A22</td>
<td>14.01</td>
<td>Yes</td>
</tr>
<tr>
<td>U14-A21</td>
<td>~</td>
<td>No</td>
</tr>
</tbody>
</table>

**Table 4.1.4:** The table gives the imino proton chemical shift and whether they were constrained as W-C base pairs in the structure determination.
4.1.6 NMR solution structure of the *E. coli* 29-mer RNA

The NMR solution structure of the *E. coli* 29-mer RNA has been successfully completed and is shown in Figure 4.1.13. The structure was calculated following the method described in section 2.6.2. The thirty lowest energy structures from the one hundred refined structures were selected, and from them the ten structures with the lowest all atom RMSD were selected. From these structures an average structure was calculated and then this structure was energy minimised. The energy minimised, final structure was then analysed using w3DNA.116

![Figure 4.1.13: The left image is the overlay of the ten best structures produced by the structure determination, they have an all atom average RMSD of 3.11 Å. The centre image is the average structure calculated from the ten best structures and has been used for purpose of analysis, some important residues are labelled. The right image is the observed secondary structure of the *E. coli* 29-mer based on the NMR solution structure. The single and double solid lines represent standard A-U and G-C W-C base pairs respectively, doted lines represent non-standard base pairs.](image)

Based on the solution structure and the w3DNA analysis the observed secondary structure is shown in Figure 4.1.13. The observed secondary structure matches the predicted secondary structure well, the major difference is that a G7 to C30 base pair was not seen in the observed structure.
The U6 is not only unpaired as predicted in the predicted secondary structure, but is clearly flipped out of the helix. The A27 residue is folded into the helix, although it is observed to be unpaired. However, there appears to be a possibility of an interaction between A9 and A27. C28 is also seen to be unpaired and appears to be slightly flipped out of the helix, accommodating the constraints previously mentioned that suggested it is likely to be flipped out and also that it is likely to be folded in. In contrast, the X-ray crystal structure of the ribosome shows A27 to be flipped out and C28 is folded into the helix. In Table 4.1.5 detailed information is given on the base pairs found in the structure.

<table>
<thead>
<tr>
<th>Base pair</th>
<th>W-C type?</th>
<th>First hydrogen bond</th>
<th>Distance (Å)</th>
<th>Second hydrogen bond</th>
<th>Distance (Å)</th>
<th>Third hydrogen bond</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-G35</td>
<td>Yes</td>
<td>O2 - N2</td>
<td>3.37</td>
<td>N3 - N1</td>
<td>3.02</td>
<td>N4 - O6</td>
<td>2.67</td>
</tr>
<tr>
<td>C2-G34</td>
<td>Yes</td>
<td>O2 - N2</td>
<td>3.09</td>
<td>N3 - N1</td>
<td>2.83</td>
<td>N4 - O6</td>
<td>2.53</td>
</tr>
<tr>
<td>C3-G33</td>
<td>Yes</td>
<td>O2 - N2</td>
<td>2.93</td>
<td>N3 - N1</td>
<td>2.78</td>
<td>N4 - O6</td>
<td>2.65</td>
</tr>
<tr>
<td>C4-G32</td>
<td>Yes</td>
<td>O2 - N2</td>
<td>3.05</td>
<td>N3 - N1</td>
<td>2.83</td>
<td>N4 - O6</td>
<td>2.60</td>
</tr>
<tr>
<td>G5-C31</td>
<td>No</td>
<td>N2 - O2</td>
<td>3.10</td>
<td>N1 - N3</td>
<td>2.62</td>
<td>O6 - N4</td>
<td>2.50</td>
</tr>
<tr>
<td>A8-U29</td>
<td>No</td>
<td>N1 - O2'</td>
<td>3.42</td>
<td>N6 - O2</td>
<td>3.77</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A9-U26</td>
<td>Yes</td>
<td>N6 - O4</td>
<td>3.60</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>C10-G25</td>
<td>Yes</td>
<td>O2 - N2</td>
<td>3.29</td>
<td>N3 - N1</td>
<td>2.98</td>
<td>N4 - O6</td>
<td>2.60</td>
</tr>
<tr>
<td>C11-G24</td>
<td>Yes</td>
<td>O2 - N2</td>
<td>3.23</td>
<td>N3 - N1</td>
<td>3.00</td>
<td>N4 - O6</td>
<td>2.67</td>
</tr>
<tr>
<td>U12-A23</td>
<td>Yes</td>
<td>N3 - N1</td>
<td>2.75</td>
<td>O4 - N6</td>
<td>2.60</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U13-A22</td>
<td>Yes</td>
<td>N3 - N1</td>
<td>2.77</td>
<td>O4 - N6</td>
<td>2.53</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U14-A21</td>
<td>Yes</td>
<td>N3 - N1</td>
<td>3.10</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

Table 4.1.5: The table shows the base pairs observed in the solution structure of the E. coli 29-mer RNA motif, whether they are standard W-C base pairs, their hydrogen bonds and their distances. The mark “---” indicates that the particular field is not relevant to the particular base pair.

Table 4.1.6 shows the local base pair step parameters and the form the RNA takes at that step. All the steps were judged to be A-form except those occurring near the bulge regions and the fraying ends, it was also seen that the steps near the bulges have the greatest deviation from the values displayed by the other steps. In particular, this was obvious with respect to the tilt and roll. Further helix parameters are detailed in Appendix C.1.
<table>
<thead>
<tr>
<th>Step</th>
<th>Shift  ($D_x$)</th>
<th>Slide  ($D_y$)</th>
<th>Rise  ($D_z$)</th>
<th>Tilt ($\tau$)</th>
<th>Roll ($\rho$)</th>
<th>Twist ($\Omega$)</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-C2/G34-G35</td>
<td>-0.30</td>
<td>-1.03</td>
<td>4.76</td>
<td>-0.17</td>
<td>5.59</td>
<td>31.18</td>
<td>~</td>
</tr>
<tr>
<td>C2-C3/G33-G34</td>
<td>0.37</td>
<td>-1.48</td>
<td>4.46</td>
<td>-1.52</td>
<td>8.11</td>
<td>25.61</td>
<td>A</td>
</tr>
<tr>
<td>C3-C4/G32-G33</td>
<td>0.22</td>
<td>-1.83</td>
<td>3.97</td>
<td>-4.96</td>
<td>12.01</td>
<td>24.60</td>
<td>A</td>
</tr>
<tr>
<td>C4-G5/C31-G32</td>
<td>-0.72</td>
<td>-1.63</td>
<td>3.95</td>
<td>8.39</td>
<td>13.21</td>
<td>23.18</td>
<td>A</td>
</tr>
<tr>
<td>A8-A9/U26-U29</td>
<td>-1.87</td>
<td>-0.19</td>
<td>4.69</td>
<td>-24.45</td>
<td>16.57</td>
<td>77.80</td>
<td>~</td>
</tr>
<tr>
<td>A9-C10/G25-U26</td>
<td>0.74</td>
<td>-1.60</td>
<td>4.11</td>
<td>4.53</td>
<td>9.76</td>
<td>24.30</td>
<td>A</td>
</tr>
<tr>
<td>C10-C11/G24-G25</td>
<td>-0.92</td>
<td>-1.00</td>
<td>4.38</td>
<td>-4.91</td>
<td>8.52</td>
<td>29.60</td>
<td>A</td>
</tr>
<tr>
<td>C11-U12/A23-G24</td>
<td>0.65</td>
<td>-1.31</td>
<td>4.22</td>
<td>-3.27</td>
<td>5.97</td>
<td>29.63</td>
<td>A</td>
</tr>
<tr>
<td>U12-U13/A22-A23</td>
<td>-0.34</td>
<td>-1.59</td>
<td>4.20</td>
<td>-0.06</td>
<td>8.88</td>
<td>28.68</td>
<td>A</td>
</tr>
<tr>
<td>U13-U14/A21-A22</td>
<td>-0.88</td>
<td>-1.07</td>
<td>4.94</td>
<td>20.59</td>
<td>7.77</td>
<td>30.48</td>
<td>~</td>
</tr>
</tbody>
</table>

Table 4.1.6: The table shows the local base pair step parameters and the form of the RNA at a given step. "~" indicates that the information could not be given.
4.2 **Assignment and structure determination of isotopically normal *H. h* 29-mer RNA**

4.2.1 **Assignment of the exchangeable proton resonances**

The 1D spectrum at 2°C showed nine clear imino peaks in the standard W-C base pairing region of the spectrum and two in the non-canonical region. From the predicted secondary structure 13 resonances were expected in the standard W-C base pairing region, and possibly one resonance in the non-canonical region associated with the non-base paired U6. It was expected that some of the imino protons associated with bases at the fraying ends and near the bulges of the RNA might not have been observed. It was also possible that a base pair in the bulge region might have given rise to at least one of the peaks in the non-canonical region of the spectrum.

The imino proton resonances were assigned by following the imino to imino sequential connectivities as illustrated in Figure 4.2.1. This was done by following three sets of sequential connectivity patterns. The first series (single arrow heads, Figure 4.2.1) started at the U13 imino proton resonance and then went to the overlapping U12 imino resonance. This connectivity was very weak but was confirmed by imino to aromatic NOE connectivities. A strong NOE connectivity then connected U12 to the G24 imino proton, and another NOE then connected G24 to G10, followed by a final NOE, which occurred very close to the diagonal, connecting G10 to U26. The second connectivity pattern (double headed arrows, Figure 4.2.1) started at G35, a weak NOE peak connected this to the G34 imino proton. There is then a series of strong NOE connectivity peaks first from G34 to G33, and then to the slightly high field U4 imino resonance. The third set of connectivities (triple headed arrows, Figure 4.2.1) began at U31, then to the high field U6 imino proton *via* a strong NOE and then to the slightly high field U30 imino proton *via* a weaker NOE.
Almost full assignment was obtained, except for the imino proton of the end residue U14. This was presumably due to the imino proton exchanging too rapidly with the solvent water to have been observed.

**Figure 4.2.1:** 600 MHz NOESY (\(\tau_m=250 \text{ ms}\)) spectrum of the *H. h.* 29-mer RNA motif (2.4 mM), dissolved in 90% \(^1\text{H}_2\text{O}\) and 10% \(^2\text{H}_2\text{O}\) containing 20 mM PO₄³⁻ (pH 6.0), 20 mM NaCl at 25°C. The *H. h.* 29-mer RNA secondary structure is also shown on the right. The spectrum shows the identification of the imino to imino through space sequential connectivities and their sequence specific assignment is shown at the top of the spectrum. The sequential connectivities are indicated both on the spectrum and on the sequence with three sets of corresponding arrows, single headed arrows for the segment of sequential walk from U13 to U26, double headed arrows for the segment of sequential walk from G35 to U4 and triple headed arrows for the segment of sequential walk from U31 to U30. The peak due to the U6a imino proton appears to be from a second conformation.

Three additional imino resonances were observed, none of which appeared to have been U14. These peaks were presumed to have been from at least one additional minor conformation. The peak at 10.63 ppm has been assigned as U6a (see Figure 4.2.1), as this peaks connectivities appeared to match those of U6, except that the peaks were of lower
intensity. This was presumably due to a lower population in this state, though possibly due to a faster exchange rate of the U6a imino proton with solvent water. The other two additional resonances were connected via a low intensity NOE cross peak to each other and their imino to amino and imino to aromatic were also low intensity, so it was difficult to identify these resonances further.

The imino resonances of G10 and G24 appeared to overlap; this was confirmed by observing the imino to aromatic connectivities, particularly the weak H2 connectivities from both A9 H2 and A23 H2.

The imino proton resonance of U31 was well within the standard chemical shift range for a W-C base pair and showed good connectivity to U6 and U30. However, the U30 imino resonance (12.99 ppm) fell outside the standard A-U W-C base pair range (~14.5 to 13.5 ppm), indicating that A7-U30 might not have formed a canonical base pair. The U6 imino resonance fell far outside the standard range adding weight to the proposition that U6 was not base paired.

The amino protons were then assigned using the imino proton assignment as a basis. This was done as described in section 2.4.2, based on the cross peak intensities and chemical shifts, (the process is illustrated for the E. coli 29-mer in Figure 4.1.2 the H. h. 29-mer assignments are tabulated in Table 4.2.1).

The imino proton assignment has been confirmed using the imino to aromatic proton NOEs, (the process is illustrated for the E. coli 29-mer in Figure 4.1.3, the H. h. 29-mer assignments are tabulated in Table 4.2.1). The aromatic assignment is described in section 4.2.2 and was used here to separately confirm both the imino proton and aromatic
assignment. A good number of inter strand, sequential and intra residue connectivities were observed, which helped confirm the secondary structure. The inter strand NOE connectivities observed between A32 H2 to U4 imino, A5 H2 to U31 imino and A7 H2 to U30 imino acted to confirm the predicted secondary structure.

The exchangeable proton assignment confirmed thus far the sequence of the RNA motif. The assignment also suggested that the predicted secondary structure was correct, although it also suggested that there was at least some instability or multiple conformations, as was indicated particularly by the multiple U6 imino resonances. There was also the low chemical shift value observed for the imino proton of the base pair A7-U30, which indicated that this base pair might not be involved in standard W-C base pairing.

4.2.2 Assignment of the non-exchangeable proton and carbon resonances

Identification of resonance types

The methodology for the sequence specific assignment of the RNA motif was to follow the characteristic sequential NOE connectivities observed between the H6/8 and H1' proton resonances in the NOESY spectrum, as detailed in section 2.4.2.

The resonances due to the H5/H6 protons of the pyrimidine U and C bases were first identified using the DQF-COSY spectrum, via their scalar correlations (Figure 4.2.2, right panel). These resonances were then classified as to whether they were from uracil or cytosine bases, using the characteristic C5 chemical shift as illustrated in Figure 4.2.2. For most resonances, it was easy to classify the resonances as either from cytosine or uracil bases. However, due to the presence of an impurity peak this was not the case for C29 H5. Its classification as either cytosine or uracil resonances had to be confirmed via sequential
NOE connectivities observed in the NOESY spectra, as well as by a process of elimination. Unfortunately, a \(^1\text{H}-^{13}\text{C}\) HSQC-NOESY spectrum was used to classify the H5 resonances, instead of the more suitable \(^1\text{H}-^{13}\text{C}\) HSQC experiment. This was due to the \(^1\text{H}-^{13}\text{C}\) HSQC data being unusable. The \(^1\text{H}-^{13}\text{C}\) HSQC-NOESY has been used in a number of other circumstances, as stated later.

**Figure 4.2.2:** The right panel shows the 600 MHz DQF-COSY spectrum of the *H. h.* 29-mer RNA motif (2.4 mM), dissolved in 100% \(^2\text{H}_2\text{O}\) containing 20 mM PO\(_4^3-\) (pH 6.0), 20 mM NaCl at 25°C. The spectrum shows the identification of the aromatic cytosine and uracil protons H5 and H6 based on the observed scalar coupled (\(^1\text{J}\)) cross peaks. Their sequence specific assignment is shown at the top (H6) and side (H5) of the spectrum. The left and middle panels show the 500 MHz \(^1\text{H}-^{13}\text{C}\) HSQC-NOESY spectrum \((\tau_m=250\text{ ms})\) of the same sample, under the same conditions, shows the \(^1\text{J}_{\text{H-C}}\) H5-C5 correlation due to the cytosine (left) and uracil (middle) bases of the RNA are shown respectively. Dotted lines mark the H5 chemical shifts, the 29-mer cross peaks are marked with ‘X’s and the impurity peaks are marked with black dots.
The H2 protons resonances appeared in the same region of the spectrum as the H6 and H8 resonances. Therefore, in order to avoid ambiguity, the $^{1}H-^{13}C$ HSQC NOESY was then used to identify the H2 resonances. This was done by identifying the H2 to C2 correlations, which appeared in a separate characteristic region of the spectrum making their identification clear (the process is illustrated for the *E. coli* 29-mer in Figure 4.1.5, the *H. h.* 29-mer assignments are tabulated in Table 4.2.1).

The $^{1}H-^{13}C$ HSQC-NOESY was then used to identify H8 protons. This was done as the corresponding C8 resonances appeared in a characteristic region, together with the C6 resonance. Fortunately, as the H6 resonances have already been identified, the two groups were easily separated. The H1'-C1' correlations also appeared in a characteristic region and so were easily identified (see Figure 4.2.3).

**Sequential assignment**

The H1'-H6/H8 sequential connectivities could then be followed, as described in section 2.4.2, to allow sequence specific assignment as shown in Figure 4.2.3 for the second strand (A21-G35). For example, if we take the sequential connectivities from the A32 intra NOE peak, it can be clearly connected to the A32-G33 inter NOE peak, then to the G33 intra peak, subsequently to the G33-G34 inter peak, then to the G34 intra peak, then to the G34-G35 inter peak and finally to the G35 intra peak. The first strand's H1'-H6/H8 sequential connectivities were followed in a similar manner (not shown, the assignments generated are tabulated in Table 4.2.1)
Figure 4.2.3: The central panel shows the 500 MHz NOESY spectrum ($\tau_m=400$ ms) of the H. h. 29-mer RNA motif (2.4 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4$$^3-$ (pH 6.0), 20 mM NaCl at 25ºC. The spectrum shows the pattern of sequential connectivities which starts at the A21 intra peak (marked start on centre panel) and can be followed all the way to the G35 intra peak (marked end). The solid black line with arrows mark the path of the sequential connectivities. Their sequence specific assignment is shown at the top (H6/8) and side (H1') of the spectrum. The top and right hand panel and shows the 500 MHz $^1$H-$^1$C HSQC-NOESY spectrum ($\tau_m=250$ ms) of the same sample, under the same conditions, shows the $^1$J$_{H\alpha C}$ H6/8-C6/8 correlations and the $^1$J$_{H\beta C}$ H1'-C1' correlations respectively. The H. h. 29-mer RNA predicted secondary structure is also shown.

The previously identified H2 protons were also assigned sequence specifically by identifying their through space connectivities, which eliminated the impurity peaks in the H2 region mentioned earlier. The process of assigning the H2 resonances also aided the identification of the H1', H6 and H8 resonances.
Overall, the peaks in this region were very clear and well separated, which made the sequence specific assignment very clear. However, similar to the *E. coli* 29-mer RNA, there were regions of overlap. There was a large degree of overlap again around 7.8 ppm, (in the H1'-H6/H8 region) which made following the sequential NOE connectivities a little more difficult in this region. Also it was seen that the chemical shifts observed for residues at the fraying ends or in the bulge region have a higher tendency to lie at the extremes of the chemical shift range. For example C1 H1' and H6, U6, H1' and H6, A27 H1' and H8 and A21 H8.

Another issue was the presence of additional NOE connectivity peaks which linked into the series of assigned NOE connectivities. This was presumed to be more evidence of additional and localised alternative conformations. These additional connectivities appeared to be associated with residues located near the U6 bulge region.

**Further assignment**

Once the sequence specific assignment of the H1's, H6s and H8s were complete, the sequence specific assignment were then expanded to other proton groups. The sequence specific H5 assignment was simply done by utilising H5 to H6 correlations observed in the DQF-COSY, shown previously in Figure 4.2.2. The same was done for H1' to H2' correlations observed in the DQF-COSY (not shown, the process is illustrated for the *E. coli* 29-mer in Figure 4.1.8 the *H. h.* 29-mer assignments are tabulated in Table 4.2.1)

Ribose rings with C₃*-endo* conformation, which is standard for A-form RNA, have very small $^3J_{H1'-H2'}$ coupling constants (~1.3 Hz). The magnetisation is therefore less efficiently transferred in these cases than for cases where there are large coupling constants. As the $^3J_{H1'-H2'}$ coupling constants were often smaller than the linewidth, the positive and negative
aspects of the peak therefore began to cancel each other out. Therefore not all H1' to H2' correlations were observed, although most were. There were a number of high intensity peaks suggesting a large $^3J_{H1'-H2'}$ coupling constant ($\sim 7.6$ Hz), which in turn suggested a possible C$_2$-endo conformation. This information not only helped confirm the assignment, as the large peaks all occurred at or next to end residues or bulge regions, but also the overall conformation of the RNA.

The remaining ribose sugar protons and carbons were difficult to assign due to overlap, and thus were not completely assigned. A number of methods were used to gain as many assignments as possible following the methods outlined in section 2.4.2. The most successful method was to observe the aromatic proton to ribose protons and the H1' to ribose protons through space connectivities and their relative intensities as observed in the NOESY spectra, and then to correlate them to the relevant regions of the $^1$H-$^13$C HSQC-NOESY.

The assignment of the non-exchangeable proton and carbon resonances had further confirmed the sequence and secondary structure, through the sequence specific assignment and the observation of a number of inter strand NOEs. The overall conformation had also been indicated as the NOE connectivity patterns for the non-bulge regions matched that expected of A-form RNA: this was further confirmed by the peak intensities of the H1'-H2' correlations in the DQF-COSY. There has also been more evidence of a possibly localised second conformation.

**4.2.3 Assignment of the phosphorus resonances**

Assignment of the phosphorus resonances of *H. h.* 29-mer RNA was not possible as $^1$H- $^{31}$P CPMG-HSQC-NOESY did not have sufficient sensitivity.
4.2.4 Assignment table of the *H. h.* 29-mer RNA motif

The $^1$H, $^{13}$C chemical shifts for the *H. h.* 29-mer RNA motif is shown in table 4.2.1.

**Table 4.2.1:** The $^1$H, $^{13}$C and $^3$P chemical shifts of the *H. h.* 29-mer RNA based on the assignment of the exchangeable (measured at 2°C) and non-exchangeable proton, carbon and phosphorus (measured at 25°C) resonances described in this chapter. A “−” indicates that the chemical shift was not obtained during assignment.
4.2.5 NMR geometrical constraints of the *H. h.* 29-mer RNA motif

**NOE constraints**

Following the methods outlined in 2.5.3, distance constraints were extracted from the assigned NOESY spectra. The NOE cross peaks were split into a number of groups: exchangeable, non-overlapped, overlapped, very overlapped and very, very overlapped. The exchangeable group consisted of constraints generated from the NOEs of the exchangeable protons, measured from the NOESY spectrum measured in 90% $\text{H}_2\text{O}$ and 10% $\text{D}_2\text{O}$ with a mixing time of 250 ms at 2ºC. The non-overlapped, overlapped, very overlapped and very, very overlapped groups consisted of constraints generated from NOEs measured from the NOESY spectrum measured in 100% $\text{D}_2\text{O}$ with a mixing time of 75 ms at 25ºC.

The NOEs were divided into each group dependent on how overlapped the NOE in question was with other peaks in the spectrum. For example an NOE which did not overlap with any other peak was placed in the non-overlapped group and an NOE which was in a heavily overlapped region was placed in the very, very overlapped group and those NOEs with only some overlap with other peaks were placed in the overlapped or very overlapped groups. This was done so that tighter error bounds could be used on the non-overlapped NOEs and conversely so that the upper error bounds could be increased for the groups containing overlapped NOEs. The error bounds were increased due to the apparent increase in intensity that these NOEs had due to overlap with other peaks.
For each of these categories, a different set of parameters were used in calculating the distance constraints. Table 4.2.2 lists the parameters used to turn the NOE peaks into constraints and the number of constraints in that group. The non-ideal distance function of $i^{-1/4}$ (where $i$ is intensity) was used for the exchangeable NOEs as this appeared to give constraints which better fitted a range of reference distances in the canonical regions of the RNA. This non-ideal behaviour may have been due to spin diffusion.

<table>
<thead>
<tr>
<th>NOE group</th>
<th>Reference intensity</th>
<th>Reference distance</th>
<th>Upper/ lower fractional error</th>
<th>Distance function</th>
<th>Number of constraints generated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchangeable</td>
<td>$7.99 \times 10^6$</td>
<td>3.2</td>
<td>1.00/1.00</td>
<td>$i^{-1/4}$</td>
<td>35</td>
</tr>
<tr>
<td>Non-overlapped</td>
<td>$7.65 \times 10^5$</td>
<td>3.7</td>
<td>0.30/0.30</td>
<td>$i^{-1/6}$</td>
<td>53</td>
</tr>
<tr>
<td>Overlapped</td>
<td>$7.65 \times 10^5$</td>
<td>3.7</td>
<td>0.60/0.30</td>
<td>$i^{-1/6}$</td>
<td>55</td>
</tr>
<tr>
<td>Very overlapped</td>
<td>$7.65 \times 10^5$</td>
<td>3.7</td>
<td>0.90/0.30</td>
<td>$i^{-1/6}$</td>
<td>77</td>
</tr>
<tr>
<td>Very, very overlapped</td>
<td>$7.65 \times 10^5$</td>
<td>3.7</td>
<td>1.20/0.30</td>
<td>$i^{-1/6}$</td>
<td>36</td>
</tr>
</tbody>
</table>

Table 4.2.2: Details information on the six groups of NOEs and the parameters used in CcpNmr Analysis to produce the NOE based distance constraints.

The exchangeable and non-exchangeable inter residue NOE constraints that were used are illustrated in Figure 4.2.4 and 4.2.5. A number of key inter strand connectivities were seen. One of the most important constraints was between A9 H2 and C28 H1' with a distance range of 2.6 to 5.9 Å. This distance suggested that the C28 bulge residue was likely to be folded in to the helix.
Figure 4.2.4: The inter residue “exchangeable” (ie. extracted from the NOESY spectrum measured in 90% $^1$H$_2$O +10% $^2$H$_2$O) NOE distance constraints used in the structure determination calculation, are shown by lines drawn between the protons of the various residues.
Figure 4.2.5: The inter residue “non-exchangeable” (i.e., extracted from the NOESY spectrum measured in $^2$H$_2$O) NOE distance constraints used in the structure determination calculation, are shown by lines drawn between the protons of the various residues.

**Torsion angle constraints**

Following the methodology set out in 2.5.4, torsion angle constraints were produced. Whether the different residues were constrained as C$_2$-endo, C$_3$-endo or left unconstrained was dependent on the approximate $^3J_{H1' \to H2'}$ coupling constant estimated from the DQF-COSY, where a large coupling constant is ~7.6 Hz and a small coupling constant is ~1.3 Hz. These approximations are listed in Table 4.2.3, along with how the residue was constrained.
<table>
<thead>
<tr>
<th>Residue</th>
<th>Approximate $J_{H1'H2'}$ (Hz)</th>
<th>Ribose constrained as</th>
<th>Glycosidic angle constrained as:</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Small</td>
<td>C$_3$-endo</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>C2</td>
<td>Small</td>
<td>C$_3$-endo</td>
<td>anti</td>
</tr>
<tr>
<td>C3</td>
<td>Small</td>
<td>C$_3$-endo</td>
<td>anti</td>
</tr>
<tr>
<td>U4</td>
<td>Small</td>
<td>C$_3$-endo</td>
<td>anti</td>
</tr>
<tr>
<td>A5</td>
<td>Small</td>
<td>C$_3$-endo</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>U6</td>
<td>Small</td>
<td>Unconstrained</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>A7</td>
<td>Small</td>
<td>Unconstrained</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>G8</td>
<td>Large</td>
<td>Unconstrained</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>A9</td>
<td>Small</td>
<td>Unconstrained</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>G10</td>
<td>Small</td>
<td>C$_3$-endo</td>
<td>anti</td>
</tr>
<tr>
<td>C11</td>
<td>Small</td>
<td>C$_3$-endo</td>
<td>anti</td>
</tr>
<tr>
<td>U12</td>
<td>Small</td>
<td>Unconstrained</td>
<td>anti</td>
</tr>
<tr>
<td>U13</td>
<td>Large</td>
<td>C$_2$-endo</td>
<td>anti</td>
</tr>
<tr>
<td>U14</td>
<td>Large</td>
<td>C$_2$-endo</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>A21</td>
<td>Large</td>
<td>C$_2$-endo</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>A22</td>
<td>Small</td>
<td>C$_3$-endo</td>
<td>anti</td>
</tr>
<tr>
<td>A23</td>
<td>Small</td>
<td>Unconstrained</td>
<td>anti</td>
</tr>
<tr>
<td>G24</td>
<td>Small</td>
<td>C$_3$-endo</td>
<td>anti</td>
</tr>
<tr>
<td>C25</td>
<td>Small</td>
<td>C$_3$-endo</td>
<td>anti</td>
</tr>
<tr>
<td>U26</td>
<td>Large</td>
<td>Unconstrained</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>A27</td>
<td>Large</td>
<td>Unconstrained</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>C28</td>
<td>Small</td>
<td>Unconstrained</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>C29</td>
<td>Small</td>
<td>Unconstrained</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>U30</td>
<td>Small</td>
<td>Unconstrained</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>U31</td>
<td>Small</td>
<td>Unconstrained</td>
<td>anti</td>
</tr>
<tr>
<td>A32</td>
<td>Large</td>
<td>Unconstrained</td>
<td>anti</td>
</tr>
<tr>
<td>G33</td>
<td>Small</td>
<td>C$_3$-endo</td>
<td>anti</td>
</tr>
<tr>
<td>G34</td>
<td>Small</td>
<td>C$_3$-endo</td>
<td>anti</td>
</tr>
<tr>
<td>G35</td>
<td>Large</td>
<td>C$_2$-endo</td>
<td>Unconstrained</td>
</tr>
</tbody>
</table>

**Table 4.2.3:** The table indicates the approximate size of the $J_{H1'H2'}$ for each residue and how the ribose of the residue was constrained based on this information. The final column indicates how the glycosic ($\chi$) torsion angle was constrained.

The glycosidic torsion angles were constrained in an *anti* conformation in the residues which were in standard W-C base pairing regions of the RNA, these constraints were given wide error bounds. Table 4.2.3 details which glycosidic torsion angles were constrained.

For sections of the RNA which were constrained as C$_3$-endo, and were in regions of the RNA which have standard W-C base pairing, the backbone in these regions was constrained with standard A-form RNA constraints.
Hydrogen bond constraints were set following the method detailed in 2.5.5, Table 4.2.4 details which base pairs were constrained.

<table>
<thead>
<tr>
<th>Base pair</th>
<th>Imino chemical shift ($\delta$, ppm)</th>
<th>Base pair constrained?</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-G35</td>
<td>12.56</td>
<td>Yes</td>
</tr>
<tr>
<td>C2-G34</td>
<td>12.76</td>
<td>Yes</td>
</tr>
<tr>
<td>C3-G33</td>
<td>12.99</td>
<td>Yes</td>
</tr>
<tr>
<td>U4-A32</td>
<td>13.22</td>
<td>Yes</td>
</tr>
<tr>
<td>A5-U31</td>
<td>13.95</td>
<td>Yes</td>
</tr>
<tr>
<td>A7-U30</td>
<td>12.99</td>
<td>No</td>
</tr>
<tr>
<td>G8-C29</td>
<td>13.22</td>
<td>No</td>
</tr>
<tr>
<td>A9-U26</td>
<td>13.62</td>
<td>No</td>
</tr>
<tr>
<td>G10-C25</td>
<td>13.50</td>
<td>Yes</td>
</tr>
<tr>
<td>C11-G24</td>
<td>13.50</td>
<td>Yes</td>
</tr>
<tr>
<td>U12-A23</td>
<td>14.01</td>
<td>Yes</td>
</tr>
<tr>
<td>U13-A22</td>
<td>13.92</td>
<td>Yes</td>
</tr>
<tr>
<td>U14-A21</td>
<td>~</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Table 4.2.4*: The table gives the imino proton chemical shift, and whether they were constrained as W-C base pairs in the structure determination.
4.2.6 NMR solution structure of the *H. h.* 29-mer RNA

The NMR solution structure of the *H. h.* 29-mer RNA was successfully completed and is shown in Figure 4.2.6. The structure was calculated following the method described in section 2.6.2. The thirty lowest energy structures from the one hundred refined structures were selected, and from them the ten structures with the lowest average all atom RMSD were selected. From these structures an average structure was calculated and then then this structure was energy minimised. The energy minimised, final structure was then analysed using w3DNA\textsuperscript{116}.

![Figure 4.2.6](image)

*Figure 4.2.6:* The left image is the overlay of the ten best structures produced by the structure determination, they have an all atom average RMSD of 1.28. The centre image is the average structure calculated from the ten best structures and has been used for purpose of analysis, some important residues are labelled. The right image shows the observed secondary structure of the *H. h.* 29-mer based on the NMR solution structure. The single and double solid lines represent standard A-U and G-C W-C base pairs respectively and the doted lines represent non-standard base pairs.

The average all atom RMSD of the best 10 structures is 1.28, which is very good. Based on the solution structure and the w3DNA\textsuperscript{116} analysis, the observed secondary structure is...
shown in Figure 4.2.6. As was seen in the observed structure, the predicted A7 to U30 base pair is not seen and a non-canonical U6 to U30 base pair was formed in its place. Similarly the predicted A9 to U26 base pair was also not observed and another non-canonical A9 to A27 base pair was formed. It is also interesting to note that none of the residues were flipped out of the helix. In Table 4.2.5 detailed information is given on the base pairs found in the structure.

<table>
<thead>
<tr>
<th>Base pair</th>
<th>W-C type?</th>
<th>First hydrogen bond</th>
<th>Distance (Å)</th>
<th>Second hydrogen bond</th>
<th>Distance (Å)</th>
<th>Third hydrogen bond</th>
<th>Distance (Å)</th>
<th>Fourth hydrogen bond</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1C-G35</td>
<td>Yes</td>
<td>O2 - N2</td>
<td>2.99</td>
<td>N3 - N1</td>
<td>2.78</td>
<td>N4 - O6</td>
<td>2.92</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>2C-G34</td>
<td>Yes</td>
<td>O2 - N2</td>
<td>2.40</td>
<td>N3 - N1</td>
<td>2.37</td>
<td>N4 - O6</td>
<td>2.35</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>3C-G33</td>
<td>Yes</td>
<td>O2 - N2</td>
<td>2.72</td>
<td>N3 - N1</td>
<td>2.62</td>
<td>N4 - O6</td>
<td>2.55</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>4U-A32</td>
<td>Yes</td>
<td>N3 - N1</td>
<td>2.82</td>
<td>O4 - N6</td>
<td>2.58</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>5A-U31</td>
<td>Yes</td>
<td>N1 - O2</td>
<td>2.40</td>
<td>N1 - N3</td>
<td>2.60</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>6U-U30</td>
<td>No</td>
<td>N3 - O2</td>
<td>4.24</td>
<td>O4-N3</td>
<td>4.46</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>7A-C29</td>
<td>No</td>
<td>N1 - N3</td>
<td>3.33</td>
<td>N1-N4</td>
<td>3.31</td>
<td>N6-N4</td>
<td>3.16</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>8G-C28</td>
<td>No</td>
<td>N1 - N3</td>
<td>3.50</td>
<td>O6 - N3</td>
<td>2.76</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>9A-A27</td>
<td>No</td>
<td>N3 - N1</td>
<td>3.39</td>
<td>N1 - N3</td>
<td>3.93</td>
<td>N6 - N7</td>
<td>4.15</td>
<td>N7-N6</td>
<td>4.43</td>
</tr>
<tr>
<td>10G-C25</td>
<td>Yes</td>
<td>N2 - O2</td>
<td>3.01</td>
<td>N1 - N3</td>
<td>2.77</td>
<td>O6 - N4</td>
<td>2.50</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>11C-G24</td>
<td>Yes</td>
<td>O2 - N2</td>
<td>2.25</td>
<td>N3 - N1</td>
<td>2.33</td>
<td>N4 - O6</td>
<td>2.33</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>12U-A23</td>
<td>Yes</td>
<td>O2 - N1</td>
<td>2.72</td>
<td>N3 - N1</td>
<td>2.67</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>13U-A22</td>
<td>Yes</td>
<td>N3 - N1</td>
<td>2.64</td>
<td>O4 - N6</td>
<td>2.75</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>14U-A21</td>
<td>Yes</td>
<td>N3 - N1</td>
<td>2.49</td>
<td>O4 - N6</td>
<td>2.94</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

**Table 4.2.5:** The table shows the base pairs observed in the NMR structure of *H. h.* 29-mer RNA; whether they are standard Watson-Crick base pairs, their hydrogen bonds and their distances. The mark “---” indicates that the particular field is not relevant to the particular base pair.

Table 4.2.6 shows the local base pair step parameters and the form the RNA takes at each step. Most of the steps were judged to be A-form, those that were judged not to be A-form tended to occur near the bulge regions and the fraying ends of the RNA. It was also seen that the steps near the bulges had the greatest deviation from the base pair step parameters displayed by the steps with e W-C base pairing regions. In particular this was clear with respect to the tilt and roll. Other helix parameters are shown in Appendix C.2.
<table>
<thead>
<tr>
<th>Step</th>
<th>Shift ($D_x$)</th>
<th>Slide ($D_y$)</th>
<th>Rise ($D_z$)</th>
<th>Tilt ($\tau$)</th>
<th>Roll ($\rho$)</th>
<th>Twist ($\Omega$)</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-C2/G34-G35</td>
<td>0.73</td>
<td>-1.35</td>
<td>3.83</td>
<td>-13.84</td>
<td>12.24</td>
<td>27.74</td>
<td>A</td>
</tr>
<tr>
<td>C2-C3/G33-G34</td>
<td>0.15</td>
<td>-1.41</td>
<td>3.60</td>
<td>15.13</td>
<td>7.21</td>
<td>27.42</td>
<td>A</td>
</tr>
<tr>
<td>C3-U4/A32-G33</td>
<td>-1.45</td>
<td>-1.13</td>
<td>3.80</td>
<td>-5.84</td>
<td>4.24</td>
<td>29.32</td>
<td>A</td>
</tr>
<tr>
<td>U4-A5/U31-A32</td>
<td>0.22</td>
<td>-1.78</td>
<td>3.23</td>
<td>-3.91</td>
<td>4.45</td>
<td>29.12</td>
<td>A</td>
</tr>
<tr>
<td>A5-U6/U30-U31</td>
<td>0.58</td>
<td>-1.35</td>
<td>3.16</td>
<td>6.48</td>
<td>0.77</td>
<td>34.27</td>
<td>A</td>
</tr>
<tr>
<td>U6-G8/C29-U30</td>
<td>-0.96</td>
<td>-1.41</td>
<td>5.91</td>
<td>16.73</td>
<td>3.26</td>
<td>46.73</td>
<td>~</td>
</tr>
<tr>
<td>G8-A9/A27-C29</td>
<td>-0.46</td>
<td>0.18</td>
<td>5.55</td>
<td>-12.35</td>
<td>13.56</td>
<td>67.04</td>
<td>~</td>
</tr>
<tr>
<td>A9-G10/C25-A27</td>
<td>3.37</td>
<td>-1.98</td>
<td>5.07</td>
<td>-22.34</td>
<td>5.28</td>
<td>30.30</td>
<td>A</td>
</tr>
<tr>
<td>G10-C11/G24-C25</td>
<td>0.01</td>
<td>-2.12</td>
<td>4.55</td>
<td>-1.59</td>
<td>9.24</td>
<td>26.97</td>
<td>A</td>
</tr>
<tr>
<td>C11-U12/A23-G24</td>
<td>-0.40</td>
<td>-1.64</td>
<td>3.63</td>
<td>6.55</td>
<td>13.55</td>
<td>24.49</td>
<td>~</td>
</tr>
<tr>
<td>U12-U13/A22-A23</td>
<td>1.39</td>
<td>-0.83</td>
<td>3.92</td>
<td>-3.19</td>
<td>-7.32</td>
<td>31.30</td>
<td>~</td>
</tr>
<tr>
<td>U13-U14/A21-A22</td>
<td>0.32</td>
<td>-1.36</td>
<td>4.31</td>
<td>-10.39</td>
<td>-11.27</td>
<td>27.92</td>
<td>~</td>
</tr>
</tbody>
</table>

Table 4.2.6: The table shows the local base pair step parameters, and the form of the RNA at a given step, a ~ indicates that the information could not be given.
4.3 Assignment and structure determination of uniformly $^{13}$C and $^{15}$N labelled $H. h.$ 37-mer RNA

4.3.1 Use of the world's first 1 GHz NMR spectrometer

The world's first 1 GHz NMR spectrometer has recently been installed in Lyon (December 2009) and has been used to perform a 3D NOESY-HSQC on the uniformly $^{13}$C and $^{15}$N labelled $H. h.$ 37-mer RNA. The results have been used to demonstrate the benefits of a 1 GHz spectrometer frequency over 800 MHz in a Bruker poster, shown in Figure 4.3.1.

One benefit of the higher field strength is an increased signal to noise ratio, due to an increased energy gap between the spin states and therefore a greater excess of spins in the lower energy state. This is particularly useful for RNA as sample concentration is generally limited to 1 mM or below, due to high sample costs and a tendency towards aggregation above 1 mM concentrations. A second benefit particularly useful for RNA is that a greater dispersion of the signals is obtained, due to the larger field strength. This is due to the fact that there is a larger difference in energy between the resonances of the nuclei with different chemical shifts. Both these advantages are highly useful for the uniformly $^{13}$C and $^{15}$N labelled $H. h.$ 37-mer RNA sample as the concentration is low (0.2 mM), and the signal dispersion in RNA is generally low. The signal dispersion is particularly low in the ribose proton region where the H2', H3', H4', H5' and H5'' protons chemical shifts all occur in the same narrow region, causing a great deal of overlap.
4.3.2 Assignment of the exchangeable proton and $^{15}$N resonances

The imino region of the 1D $^1$H spectrum, taken at 2°C in 90% $^1$H$_2$O and 10% $^2$H$_2$O, showed 14 clear imino peaks, with the possibility that some peaks were hidden due to overlap. There are 19 imino protons, 15 of which (based on the predicted secondary structure) should occur in the canonical W-C base pairing region (12-15 ppm); in fact nine peaks occurred in this region. The final four imino peaks should be observed in the non W-C base pairing region (~10-12 ppm); in fact five peaks were observed in this region. This may be due to a similar situation as the $H. h.$ 29-mer RNA motif where the U6 imino
proton is observed at two chemical shift values suggesting multiple conformations, alternatively the predicted secondary structure may be incorrect.

A $^1$H-$^{15}$N HSQC of the imino region was collected. The uracil imino nitrogens appeared in a distinct $^{15}$N chemical shift range from the guanine imino nitrogens, as can be seen in Figure 4.3.2. A $^1$H-$^{15}$N HSQC-NOESY spectrum was also performed, and the NOE connectivities were traced to assign the imino resonances. Unfortunately, only a few NOE peaks were observed, and so full assignment was not possible. Fortunately, the imino proton chemical shifts were found to match well with those of the $H. h.$ 29-mer RNA (see section 4.2.1) and that of the previously assigned UUCG tetra loop and the closing C-G base pair.$^{75,76}$ Thus, the imino (NH) protons of the $H. h.$ 37-mer RNA were assigned by comparison with the imino assignments of the $H. h.$ 29-mer RNA, and that of the tetra loop and closing base pair. The fact that the chemical shifts matched well acts to confirm the assumption that the structures of the $H. h.$ 37-mer RNA would be similar to that of the $H. h.$ 29-mer RNA. The chemical shifts of the tetra loop and closing C-G base pair also matched well, indicating that the loop is taking on a standard UUCG tetra loop conformation. The uracil and guanine imino nitrogens all fell into the expected chemical shift regions, acting to not only confirm the assignment of the $H. h.$ 37-mer RNA, but also the $H. h.$ 29-mer RNA, as well as the previously assigned tetra loop and closing C-G base pair.$^{75,76}$
Figure 4.3.2: 600 MHz $^{1}$$H$$^{15}$N HSQC spectrum of *H. h.* 37-mer RNA motif (0.2 mM), dissolved in 90% $^{1}$$H_{2}O + 10%^{2}$$H_{2}O$ containing 20 mM $PO_4^{3-}$ (pH 6.4) and 20 mM NaCl at 2°C. The spectrum shows the identification of the imino resonances. Their sequence specific assignment is shown at the top and right hand side of the spectrum stars mark the unassigned imino resonances and “X”s mark the assigned imino resonances. The *H. h.* 37-mer RNA secondary structure is also shown.

The imino resonance assignment was then checked and modified using the few NOE peaks observed in the imino region of the $^{1}$$H$$^{15}$N HSQC-NOESY spectrum. One NOE between “G33” and “U30”, using the assignment based just on comparison with the *H. h.* 29-mer, was observed. This cannot be the case as the imino protons of these residues are highly likely to be too far apart for an NOE to be observed, particularly given the predicted secondary structure. Therefore, the initial assignment of the imino resonances of U30 and U4, whose chemical shifts differ by 0.23 ppm, were switched. The peak intensities fit better for this altered assignment, particularly for U30, which has a very weak intensity. The U30 residue occurs in the middle of the bulge region of the motif. The intensity of U4 imino resonance is higher, which fits with the predicted secondary structure where the
residue occurs at the edge of the bulge region. It is possible that these imino protons were originally incorrectly assigned in the *H. h.* 29-mer RNA motif, but the presence of an NOE connecting the U30 imino proton and the U6 H5 proton in the *H. h.* 29-mer RNA spectra renders it unlikely, as U4 imino proton to U6 H5 would be an unlikely NOE connectivity to observe. The majority of the imino protons were thus assigned. The imino protons that were not assigned were Ga, G35, U13 and U14. These are either from the additional end base pair or from regions next to the additional bases of the *H. h.* 37-mer, as compared to the *H. h.* 29-mer, or they are next to or are in base pairs which were not assigned in the *H. h.* 29-mer RNA motif.

It is gratifying to note that four unassigned peaks are observed, two in the uracil imino region and two in the guanine imino region, these likely correspond to U13, U14, Ga and G35. In the $^1$H-$^{15}$N HSQC-NOESY, it should be noted that the only other imino to imino NOE peak observed was a high intensity peak between U6 and U6a.

Some amino nitrogen and proton chemical shifts were assigned using the $^1$H-$^{15}$N HSQC-NOESY. This was done as described in section 2.4.3, however due to overlap, particularly in the $^{15}$N dimension, it only allowed a few sequence specific assignments. These assignments are shown in the assignment table in section 4.3.5.
4.3.3 Assignment of the non-exchangeable proton and carbon resonances

The basis of the non-exchangeable assignment for the doubly $^{15}$N and $^{13}$C labelled 37-mer is the sequential connectivities observed between the H6/8 and H1' protons. However, as opposed to isotopically normal RNA, this can now be followed in the 3D $^1$H-$^{13}$C HSQC-NOESY spectrum. This procedure is described in section 2.4.3.

Identification of resonance types

To begin following the sequential connectivity patterns, the H6 proton resonances were assigned as being uracil or cytosine H6 resonances. This was done using the 3D $^1$H-$^{13}$C HSQC-NOESY spectrum. First, the U H5 to U H6 NOE connectivity peaks were isolated from the 3D $^1$H-$^{13}$C HSQC-NOESY by observing the H5, H6, uracil C6 region of the 3D spectrum. The same method was then followed for the cytosine H5 to H6 NOE connectivities.

Although not all the uracil H5s were sequentially assigned (shown later), all ten C5 to H5 correlations were observed in the $^1$H-$^{13}$C HSQC, although notably only eight H5 to H6 NOEs are clearly seen in the 3D $^1$H-$^{13}$C HSQC-NOSEY. A similar story was observed for cytosine, all ten expected C5 to H5 correlations, plus two impurities, are seen in the $^1$H-$^{13}$C HSQC spectrum but only eight H5 to H6 NOEs are clearly observed in the 3D $^1$H-$^{13}$C HSQC-NOSEY. Since the H5 to H6 NOEs are one of the most distinct in the spectrum, it is unclear why some were not observed.

Once the C and U H6 resonances were identified, the H8 resonances were identified by elimination. This was necessary as the H8 and C8 chemical shift ranges overlap with that
of H6 and C6 chemical shift ranges. Again it was not possible to sequence specifically assign all the H8 and C8 resonances, as there were more than the expected 37 peaks in this region due to impurities or additional conformations. An interesting feature of the $^1$H-$^{13}$C HSQC spectrum in this region was that the H6 to C6 correlation peaks were approximately twice as long in the $^{13}$C dimension, compared to that of the H8 to C8 correlation peaks. This is likely due to a coupling effect between the N1 and the C6, this feature was a useful confirmation of the assignment, although it did lead to broader peaks.

The adenine H2 resonances were then identified. This was because the H1' to H2 NOEs can be observed while the pattern of H6/8 to H1' sequential connectivities are traced. These NOEs can be helpful in confirming the H6/8 to H1' and for assigning the H2 protons themselves. Four of the expected eight H2 to C2 correlations were clearly observed, and sequentially assigned alongside the H6/8 and H1's (described later). At least four more H2 to C2 correlations were observed in the region, which were not subsequently sequential assigned.

The H1' nuclei were then identified, through the $^1$H-$^{13}$C HSQC, as their directly scalar coupled C1' resonances appeared in a distinct chemical shift range.

**Sequential assignment**

The H1' to H6/8 sequential connectivities were then followed using the 3D $^1$H-$^{13}$C HSQC-NOESY, see Figure 4.3.3. The advantage of using a 3D $^1$H-$^{13}$C HSQC-NOESY over a straight $^1$H-$^1$H NOESY is that there is a third $^{13}$C dimension, allowing overlap to be alleviated. Sequential assignment was then achieved from residue A9 through to C29, with a single gap at U14. This was done following the subsequent process, for example, if the
sequential connectivities starting from the A9 H1' to H8 intra NOE peak are taken, they can be clearly traced, first to the A9 H1'-G10 H8 inter NOE peak, both in the A9 C1' plane, then to the G10 H1' to H8 intra peak in the G10 C1' plane, and subsequently to the G10 H1'-C11 H6 inter peak in the same plane. This pattern of connectivities can then be followed all the way thorough to the C29 intra peak (with a gap at U14), as shown in Figure 4.3.3.

The remaining residues were not assigned sequentially due to a lack of key NOEs in this section of the RNA. This lack of NOEs towards the fraying ends of the sequence could be due to dynamic effects or multiple conformations forming towards the fraying ends, causing a reduction in the NOE intensity.

Observing the characteristic H2 connectivities to the H1' protons allowed the sequence specific assignment of most of the H2 protons in the assigned region.

The sequential connectivities that were observed act to confirm the sequence of the RNA for that region. It also indicates an A-form conformation, in the standard base pairing sections due to the characteristic NOE patterns observed.

A 3D NOESY-HSQC was subsequently recorded on the 1 GHz spectrometer in Lyon; this provided some additional NOEs, as illustrated in Figure 4.3.4, which confirmed the assignment. However, key NOEs were still missing and so unfortunately no additional assignments could be made.
Figure 4.3.3: The panels show planes of the 1 GHz 3D NOESY-HSQC spectrum ($\tau_m=300$ ms) of the $H$. $h$. 37-mer RNA motif (0.2 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4^{3-}$ (pH 6.4) and 20 mM NaCl at 25°C. The spectrum shows the pattern of sequential connectivities which starts at the A9 intra peak and can be followed all the way to the C29 intra peak. With a single break in the pattern at U14 and a missing sequential peak between C18 and G19 in the loop region. The solid black line with arrows mark the path of the sequential connectivities. Their sequence specific assignment is shown at the top (H1') and side (H6/8) of the spectrum and the $^{13}$C chemical shifts of the various planes (the associated C1' chemical shift) is indicated at the top of each panel.
**Figure 4.3.4:** The left image shows the $^{13}$C plane at 138.96 ppm of the 1 GHz 3D NOESY-HSQC ($\tau_m=300$ ms) (left) and the 800 MHz 3D $^1$H-$^{13}$C HSQC-NOESY spectrum ($\tau_m=300$ ms) (right) of the $H. h.$ 37-mer RNA motif (0.2 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4$ (pH 6.4) and 20 mM NaCl at 25ºC. The right image shows the $^{13}$C plane at 139.29 ppm of the same spectra taken at 1 GHz (left) and 800 MHz (right). The H8 assignment is shown along the top of the spectra and further assignments are shown along the right of the spectra. Additional peaks can be clearly seen in these regions of the 1 GHz spectrum as compared to the 800 MHz equivalent.

**Further assignment**

The sequential assignment of the aromatic and H1' nuclei was then expanded to the rest of the ribose protons and carbons by using a number of methods as detailed in section 2.4.3. The most straightforward and least ambiguous method was to use the 3D $^1$H-$^{13}$C HSQC-TOCSY spectrum with a short mixing time (5.44 ms). Assignment following this method is shown for residue C18 in Figure 4.3.5. Unfortunately, for most residues this was not possible as overlap of peaks prevented the sequences of peaks being followed accurately. This method was backed up by using the 3D $^1$H-$^{13}$C HSQC-TOCSY, with a mixing time of 80 ms. However, the larger number of peaks present led to more extensive overlap which impeded the assignment.
Figure 4.3.5: The panels show planes of the 700 MHz 3D $^1$H-$^1$H $^1$C HSQC TOCSY spectrum ($\tau_m=5.44$ ms) of the $H. h.$ 37-mer RNA motif (0.2 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4^{3-}$ (pH 6.4) and 20 mM NaCl at 25°C. The spectrum shows a pattern of $^3J_{\text{H-H}}$ coupling correlations which starts at the H1' of C18 and can be followed through the ribose protons to H5' and H5'' of C18. The solid black line with arrows mark the path of the $^3J_{\text{H-H}}$ coupling correlations. Their sequence specific assignment is shown at the top (ribose protons) and side (ribose protons) of the spectrum and the $^1$C chemical shifts of the various planes (the associated ribose carbon chemical shift) is indicated at the top of each panel.

Other sequential connectivity patterns were followed in the 3D $^1$H-$^1$C HSQC-NOESY to help expand and confirm the ribose assignment, for instance the sequential connectivity pattern between H6/8 and H3' (see Figure 2.4.2). This series of sequential connectivities were seen particularly clearly from A9 to U12.
By following such patterns in the 3D $^1$H-$^{13}$C HSQC-NOESY other nuclei were assigned, for instance in canonical regions of the RNA, the H5' protons are only close enough to the intra H6/8 and not the inter H6/8 to form NOE connectivities. This allowed the assignment of many of the H5' protons.

**Conclusion**

The vast majority of non-exchangeable proton and associated carbon nuclei were assigned for the sequentially assigned residues (A9-C29), due to the $^{13}$C and $^{15}$N labelling of the RNA, and the reduced overlap yielded by the 3D spectra. However, a number of residues were not sequentially assigned due to the lack of key NOEs in the spectra as discussed earlier.

The assignment for the tetra loop and closing C-G base pair has confirmed the previous assignment$^{75,76}$ of the loop and closing base pair. This suggests again that a standard UUCG tetra loop confirmation has been assumed by the RNA. Within the canonical base pairing regions of the RNA that have been assigned, standard A-form NOE patterns have been observed. Inter strand H2 NOE connectivities which are observed confirm the predicted secondary structure in the region between base pair U12-A23 and base pair C15-G20.
4.3.4 Assignment of the phosphorus resonances

The $^{31}$P resonances were assigned using the 3D HCP spectrum, using a number of methods detailed in section 2.4.3. One of the most powerful methods is to follow a series of sequential scalar correlations from the H4' resonances to the residue's own phosphorus and also to the subsequent residue's phosphorus. Unfortunately, despite the three dimensions of the spectrum, the overlap in the H4'-C4'-P region was still too large to be able to extensively follow the pattern of sequential correlations. However, a number of assignments were made through this method. Another method is to observe the intra residue H5'/5'' to C5' to $^{31}$P correlations. The results of this are shown in Figure 4.3.6, for a number of tetra loop and closing base pair residues, other phosphorus resonances were also assigned using this method. As can be observed clearly for C18, two correlations occur in the HCP spectrum, a C18 H5' to C5' to P correlation and a C18 H5'' to C5' to P correlation. This enabled the assignment of the C18 P resonance, as the C18 carbon and proton H5' shifts were already known.
Figure 4.3.6: The lower panel shows the 800 MHz $^1$H-$^{13}$C HSQC spectrum of the *H. h.* 37-mer RNA motif (0.2 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4^{3-}$ (pH 6.4) and 20 mM NaCl at 25°C. The spectrum shows the H5'/5'' to C5' correlations ($^1$J$_{H-C}$). The top panels show planes of the 600 MHz 3D HCP of the same sample, under the same conditions. The planes show the intra residue $^{31}$P to C5' to H5' correlations. Their sequence specific assignment is shown at the top (H5') upper right hand side (P) and lower right hand side (C5') of the spectra. Dotted lines mark the H5' chemical shifts, the assigned peaks that are shown are marked with 'X's.

The assigned $^{31}$P resonances for the loop and closing base pair fall outside the standard chemical shift range (approximately -3.75 to -4.80 ppm) except U16, indicating the non-A-form conformation of the loop. Also, the assigned $^{31}$P resonances in the canonical base pairing region of the RNA came within the standard A-form $^{31}$P chemical shift range, confirming the A-form conformation of those sections of the RNA.
4.3.5 Assignment table of the *H. h.* 37-mer RNA

The $^1$H, $^{13}$C, $^{15}$N and $^{31}$P chemical shifts of the *H. h.* 37-mer are shown in Table 4.3.1.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.3.1** The $^1$H, $^{13}$C, $^{15}$N and $^{31}$P chemical shifts of the *H. h.* 37-mer RNA based on the assignment of the exchangeable (measured at 2°C) and non-exchangeable proton, carbon and phosphorus (measured at 25°C) resonances described in this chapter. A “~” indicates that the chemical shift was not obtained during assignment.
4.3.6 NMR geometrical constraints of the H. h. 37-mer RNA

NOE constraints

Following the methods outlined in section 2.5.3, distance constraints were calculated. The NOE cross peaks were split into two groups, non-overlapped and overlapped, based on the extent of overlap of the peak in question, no overlap in the non-overlapped group and overlapped peaks in the overlapped group. For both categories, the NOEs were measured from the same spectrum (800 MHz 3D $^1$H-$^{13}$C HSQC-NOESY spectrum ($\tau_m=100$ ms) of H. h. 37-mer RNA (0.2 mM, pH 6.4 in 100% $^2$H$_2$O) at 25°C ) and the same set of parameters were used in calculating the distance constraints, except that a higher upper error bound was applied to overlapped group.

Table 4.3.2 lists the values used to turn the NOE peaks into constraints and the number of constraints in that group. The inter nucleotide and inter strand NOE constraints that were used are illustrated in Figure 4.3.7. A number of key inter strand connectivities can be seen, which helped confirm the assignment

<table>
<thead>
<tr>
<th></th>
<th>Reference intensity</th>
<th>Reference distance (Å)</th>
<th>Upper/lower fractional error</th>
<th>Distance function</th>
<th>Number of constraints generated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-overlapped</td>
<td>$5.72 \times 10^7$</td>
<td>3.7</td>
<td>0.38/0.38</td>
<td>$i^{16n}$</td>
<td>236</td>
</tr>
<tr>
<td>Overlapped</td>
<td>$5.72 \times 10^7$</td>
<td>3.7</td>
<td>0.90/0.38</td>
<td>$i^{11n}$</td>
<td>145</td>
</tr>
</tbody>
</table>

Table 4.3.2: The table details information on the six groups of NOEs and the parameters used in CcpNmr Analysis to produce the NOE based distance constraints. Constraints calculated based on the 3D $^1$H-$^{13}$C HSQC-NOESY spectrum, measured in 100% $^2$H$_2$O with a 100 ms mixing time.
Figure 4.3.7: The inter residue NOE distance constraints used in the structure determination calculation, are shown by lines drawn between the protons of the various residues.

**Torsion angle constraints**

Following the methodology set out in section 2.5.4, torsion angle constraints were produced. Whether the different residues were constrained as C\textsubscript{2}-endo, C\textsubscript{3}-endo or left unconstrained was dependent on the “can1”, which is a function of the C1', C4' and C5' chemical shifts (see section 2.5.4). As the “can1” value was not available for large sections of the RNA the sugar pucker in canonical base pairing regions of the RNA were constrained as C\textsubscript{3}-endo. The “can1” values are displayed in Table 4.3.3, including how the residues were constrained.
<table>
<thead>
<tr>
<th>Residue</th>
<th>$can1$ (ppm)</th>
<th>Ribose constrained as:</th>
<th>H2'-H6/8 NOE intensity &gt; H1'-H6/8 NOE intensity?</th>
<th>H3'-H6/8 NOE intensity &gt; H1'-H6/8 NOE intensity?</th>
<th>Glycosidic angle ($\chi$) constrained as:</th>
</tr>
</thead>
<tbody>
<tr>
<td>G0</td>
<td>~</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>C1</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>C2</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>C3</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>C4</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>G5</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>U6</td>
<td>~</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>G7</td>
<td>~</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>A8</td>
<td>~</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>A9</td>
<td>-6.06</td>
<td>C$_3$-endo</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>C10</td>
<td>-5.7</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>C11</td>
<td>-5.44</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>U12</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>U13</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>U14</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>C15</td>
<td>-5.39</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>U16</td>
<td>~</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>U17</td>
<td>-7.56</td>
<td>Unconstrained</td>
<td>Yes</td>
<td>No</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>C18</td>
<td>-6.98</td>
<td>Unconstrained</td>
<td>Yes</td>
<td>No</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>G19</td>
<td>-5.8</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>G20</td>
<td>-6.01</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>A21</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>A22</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>A23</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>G24</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>G25</td>
<td>-5.4</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>U26</td>
<td>~</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>A27</td>
<td>-5.69</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>C28</td>
<td>-5.63</td>
<td>Unconstrained</td>
<td>Yes</td>
<td>Yes</td>
<td>anti</td>
</tr>
<tr>
<td>U29</td>
<td>~</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>C30</td>
<td>~</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>C31</td>
<td>~</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>G32</td>
<td>~</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>G33</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>G34</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>G35</td>
<td>~</td>
<td>C$_3$-endo</td>
<td>~</td>
<td>~</td>
<td>anti</td>
</tr>
<tr>
<td>C36</td>
<td>~</td>
<td>Unconstrained</td>
<td>~</td>
<td>~</td>
<td>Unconstrained</td>
</tr>
</tbody>
</table>

Table 4.3.3: The second column of the table indicates the value of “$can1$” for each residue and the third column indicates how the ribose of the residue was constrained based on this information. The fourth and fifth columns indicate whether the H2' to H6/8 NOEs are greater in intensity than the corresponding H1' to H6/8 and whether H3' to H6/8 NOEs are greater that the corresponding H1' to H6/8 respectively and the final column indicates how the glycosic ($\chi$) torsion angle was constrained based on this information. A “$can1$” value of greater than -6.25 ppm indicates a C$_3$-endo conformation, less than -6.25 ppm indicates a C$_2$-endo conformation. When the H2' to H6/8 and or H3' to H6/8 NOEs have a greater intensity than the corresponding H1' to H6/8 NOE, this indicates a glycosidic torsion angle with an anti conformation. A “~” indicates that the information was not available or was unclear.
The glycosidic torsion angle was determined by observing intra residue H6/8 to H1', H2' and H3' NOEs. If a large H3' or H2' to H6/8 NOE was observed and a small H6/8 to H1', the residue was constrained to an *anti* conformation. If a small H3' or H2' to H6/8 NOE was observed and a large H6/8 to H1', the residue was left unconstrained. However, this information was sparse and so the glycosidic angle was also constrained as *anti* in canonical base pairing regions of the RNA. Residues where these conditions were not met were left unconstrained. Table 4.3.3 details which residues were constrained in such a way.

For residues of the RNA which were constrained as C3'-*endo*, which were also in sections of the RNA which have canonical W-C base pairing and where the $^{31}$P chemical shifts were within the standard range (approximately -3.75 to -4.80 ppm), the backbone was constrained with standard A-form RNA constraints. However, as there was a lack of $^{31}$P assignments towards the fraying end of the RNA, the requirement for the $^{31}$P resonances to be known to be in the required range was dropped.

Hydrogen bond constraints were set following the method detailed in section 2.5.5, Table 4.3.4 details which base pairs were constrained. Where there was a lack of experimental evidence, base pairs were constrained based on the predicted secondary structure, though base pairs next to the bulges or loops were not constrained.
<table>
<thead>
<tr>
<th>Base pair</th>
<th>Imino chemical shift (δ, ppm)</th>
<th>Base pair constrained?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ga-Cb</td>
<td>~</td>
<td>No</td>
</tr>
<tr>
<td>C1-G35</td>
<td>~</td>
<td>Yes</td>
</tr>
<tr>
<td>C2-G34</td>
<td>12.80</td>
<td>Yes</td>
</tr>
<tr>
<td>C3-G33</td>
<td>12.74</td>
<td>Yes</td>
</tr>
<tr>
<td>U4-A32</td>
<td>12.97</td>
<td>Yes</td>
</tr>
<tr>
<td>A5-U31</td>
<td>13.96</td>
<td>Yes</td>
</tr>
<tr>
<td>A7-U30</td>
<td>13.20</td>
<td>No</td>
</tr>
<tr>
<td>G8-C29</td>
<td>13.29</td>
<td>No</td>
</tr>
<tr>
<td>A9-U26</td>
<td>13.60</td>
<td>No</td>
</tr>
<tr>
<td>G10-C25</td>
<td>13.47</td>
<td>Yes</td>
</tr>
<tr>
<td>C11-G24</td>
<td>13.41</td>
<td>Yes</td>
</tr>
<tr>
<td>U12-A23</td>
<td>13.97</td>
<td>Yes</td>
</tr>
<tr>
<td>U13-A22</td>
<td>~</td>
<td>Yes</td>
</tr>
<tr>
<td>U14-A21</td>
<td>~</td>
<td>Yes</td>
</tr>
<tr>
<td>C15-G20</td>
<td>13.54</td>
<td>No</td>
</tr>
</tbody>
</table>

**Table 4.3.4:** The table gives the imino proton chemical shift, and whether they were constrained as W-C base pairs in the structure determination.
4.3.7 NMR solution structure of the *H. h.* 37-mer RNA

The NMR solution structure of the *H. h.* 37-mer RNA has been completed and is shown in Figure 4.3.8. The structure was calculated following the method described in section 2.6.2. The thirty lowest energy structures from the one hundred refinement structures were selected. From these, the ten structures with the lowest average all atom RMSD were selected, giving increased weight to a low average all atom RMSD for the upper section of the motif, which was constrained by NOE distance constraints, as opposed to the overall average all atom RMSD. From these structures an average structure was calculated and then this structure was energy minimised. The energy minimised, final structure was then analysed using w3DNA.\textsuperscript{116}

![Figure 4.3.8](image_url)

*Figure 4.3.8:* The left image is the overlay of the ten best structures produced by the structure determination, they have an all atom average RMSD of 3.61 Å. The centre image is the average structure calculated from the ten best structures and has been used for purpose of analysis, some important residues are labelled. The right image is the observed secondary structure of the *H. h.* 37-mer based on the NMR solution structure. The single and double solid lines represent standard A-U and G-C W-C base pairs respectively and the dotted lines represent non-standard base pairs.
The average all atom RMSD of the best ten structures is 3.61 Å, this number indicates that the degree of variation is very large. This is due to a lack of constraints in the lower half of the RNA. The section of RNA which contains NOE distance constraints (residues 9 to 29) has an all atom average RMSD of 1.72 Å, for all 10 structures (Figure 4.3.9).

**Figure 4.3.9:** The image is the overlay of residues 9 to 29 of the ten best structures produced by the structure determination, they have an all atom average RMSD of 1.72 for residues 9 to 29.

Based on the solution structure and the w3DNA analysis, the observed secondary structure is shown in Figure 4.3.8. In Table 4.3.5 detailed information is given on the base pairs found in the structure. Due to the lack of distance constraints the lower half of the RNA is relatively unstructured; particularly around the bulge regions, as there are also few dihedral constraints in this region. The loop itself forms a conformation similar to that observed in other instances of the UUCG tetra loop (Figure 1.9.1).
Table 4.3.5: The table shows the base pairs observed in the solution structure of *H. h*. 37-mer RNA; whether they are standard W-C base pairs, their hydrogen bonds and their distances. The mark “---” indicates that the particular field is not relevant to the particular base pair.

Table 4.3.6 show the base pair step parameters and the form the RNA takes at that step. All but one of the steps are judged to be A-form, excepting those occurring near the bulge regions and the loop. It can also be seen that the steps near the bulges and the loop have the greatest deviation from the values displayed by the other steps. Other helix parameters are shown in Appendix C.3.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>aG-Cb</td>
<td>No</td>
<td>O6 - O2</td>
<td>4.19</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1C-G35</td>
<td>Yes</td>
<td>O2 - N1</td>
<td>3.14</td>
<td>N3 - N2</td>
<td>2.52</td>
<td>N4 - N1</td>
<td>2.66</td>
<td>N4 - O6</td>
<td>2.10</td>
<td>---</td>
</tr>
<tr>
<td>2C-G34</td>
<td>Yes</td>
<td>O2 - N1</td>
<td>2.85</td>
<td>N3 - N2</td>
<td>2.23</td>
<td>N4 - O6</td>
<td>2.37</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>3C-G33</td>
<td>Yes</td>
<td>O2 - N1</td>
<td>2.62</td>
<td>N3 - N2</td>
<td>3.00</td>
<td>N3 - O6</td>
<td>2.78</td>
<td>N4 - O6</td>
<td>2.14</td>
<td>---</td>
</tr>
<tr>
<td>4U-A32</td>
<td>No</td>
<td>O2 - N1</td>
<td>2.75</td>
<td>N3 - N1</td>
<td>2.05</td>
<td>O4 - N6</td>
<td>2.36</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>9A-U26</td>
<td>No</td>
<td>N6 - O2</td>
<td>2.69</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>10G-C25</td>
<td>Yes</td>
<td>N2 - O2</td>
<td>2.54</td>
<td>N1 - N3</td>
<td>2.64</td>
<td>O6 - N4</td>
<td>2.85</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>11C-G24</td>
<td>Yes</td>
<td>O2 - N2</td>
<td>2.84</td>
<td>N3 - N1</td>
<td>2.41</td>
<td>N4 - O6</td>
<td>2.37</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>12U-A23</td>
<td>Yes</td>
<td>N3 - N1</td>
<td>2.45</td>
<td>O4 - N6</td>
<td>2.61</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>13U-A22</td>
<td>Yes</td>
<td>N3 - N1</td>
<td>2.47</td>
<td>O4 - N6</td>
<td>2.90</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>14U-A21</td>
<td>No</td>
<td>O2 - N1</td>
<td>3.25</td>
<td>N3 - N1</td>
<td>2.40</td>
<td>O4 - N6</td>
<td>2.48</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>15C-18C</td>
<td>No</td>
<td>O2 - N4</td>
<td>3.96</td>
<td>N3 - N3</td>
<td>5.25</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>16U-20G</td>
<td>No</td>
<td>N3 - N2</td>
<td>2.94</td>
<td>O4 - N2</td>
<td>3.07</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

Table 4.3.6: The table shows the local base pair step parameters, and the form of the RNA at a given step, a ~ indicates that the information could not be given.
4.4 Discussion and comparison of the NMR structures of the *E. coli* and *H. h* RNA motifs

4.4.1 Comparison of the NMR assignment

The chemical shifts and the pattern of NOEs of the *E. coli* and *H. h*. RNA motifs should be comparable particularly that of the *H. h*. 37-mer and the *H. h*. 29-mer, due to their conserved sequence and secondary structure. Table 4.4.1 compares the chemical shifts of the *E. coli* 29-mer, *H. h*. 29-mer and *H. h*. 37-mer RNA exchangeable protons. Approximately 90% of the exchangeable chemical shifts of the two *H. h*. RNA motifs vary by 0.25 ppm or less and ~60% vary by only 0.1 ppm or less. This obviously indicates that the structures of the two motifs are likely to be very similar.

Table 4.4.1 The table shows the difference in exchangeable proton chemical shifts between the *H. h*. 29-mer and the *H. h*. 37-mer (A), the *H. h*. 29-mer and the *E. coli* 29-mer (B), and the *H. h*. 37-mer and the *E. coli* 29-mer (C). Where the delta value of the chemical shift is less than or equal to 0.10 the cell is coloured green, if the value is greater than 0.10 and less than or equal to 0.25 are coloured yellow, if it is greater than 0.25 and less than or equal to 0.50 the cells were coloured orange and if the value was greater than 0.50 the cells were coloured red. A colour coded *H. h*. and *E. coli* sequence are shown at either side of the table sequence. A “~” marks cells where the data was not available, a greyed out cell indicates that the value is not applicable.
The exchangeable proton chemical shift differences between the \textit{E. coli} 29-mer and the two \textit{H. h.} RNA motifs, are larger than the differences observed between the two \textit{H. h.} RNA motifs, as would be expected. These differences are illustrated in Figure 4.4.1 which shows the imino region of the 1D $^1$H of the \textit{E. coli} 29-mer and \textit{H. h.} 29-mer RNA. In the regions where the sequences matched the chemical shifts tended to differ only a little, except for the G35 imino proton which differs in chemical shift between the two 29-mers by 0.68 ppm.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure441.png}
\caption{The \textit{top trace} is the 700 MHz 1D $^1$H spectrum of the imino region of the \textit{E. coli} 29-mer RNA motif (1 mM), dissolved in 90\% $^1$H$_2$O + 10\% $^2$H$_2$O containing 20 mM PO$_4$$^{3-}$ (pH 6.2), 20 mM NaCl at 2\textdegree C. The \textit{lower trace} is the 600 MHz 1D $^1$H spectrum of the imino region of the \textit{H. h.} 29-mer RNA motif (2.4 mM), dissolved in 90\% $^1$H$_2$O and 10\% $^2$H$_2$O containing 20 mM PO$_4$$^{3-}$ (pH 6.0), 20 mM NaCl at 25\textdegree C. The sequence specific assignment is indicated on both traces. The observed secondary structure of each of the RNA motifs are also shown, \textit{E.coli upper, H. h. lower}.}
\end{figure}
Table 4.4.2 shows the differences in the non-exchangeable proton chemical shifts between the three RNA motifs. As can be seen from the table ~80% of the chemical shifts only differ by 0.25 ppm or less and ~55% of these differ by less than 0.10 ppm. The large differences occur at or near variations in the sequence, and at the bulge regions. The differences near the bulge regions particularly at A27 may indicate variability in the structure of the different motifs at these points. This could also be due to the inherent increased instability in these regions. Figure 4.4.2 helps illustrate the similarity and differences in the non-exchangeable chemical shifts. The three 29-mer motifs produce similar 1D $^1$H spectra as shown in Figure 4.4.2. The *H. h.* 37-mer RNA motif shows its ribose proton region (minus H1') taking up a wider chemical shift range as opposed to the 29-mers, as might be expected due to the addition of the eight extra residues, with some in non-canonical regions of the RNA. Additionally the peaks in the aromatic proton region of the 37-mer spectrum appear to be much broader than the 29-mers, which may be an indication of a more dynamic structure in the case of the 37-mer.
Table 4.4.2: The table shows the difference in non-exchangeable proton chemical shifts between the *H. h.* 29-mer and the *H. h.* 37-mer (A), the *H. h.* 29-mer and the *E. coli* 29-mer (B), and the *H. h.* 37-mer and the *E. coli* 29-mer (C). Where the delta value of the chemical shift is less than or equal to 0.10 the cell is coloured green, if the value is greater than 0.10 and less than or equal to 0.25 are coloured yellow, if it is greater than 0.25 and less than or equal to 0.50 the cells were coloured orange and if the value was greater than 0.50 the cells were coloured red. A colour coded *H. h.* and *E. coli* sequence are shown at either side of the table sequence. A “~” marks cells where the data was not available, a greyed out cell indicates that the value is not applicable.
Figure 4.4.2: The top trace is the 600 MHz 1D $^1$H spectrum of the *E. coli* 29-mer RNA motif (1 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4$$^{3-}$ (pH 6.2), 20 mM NaCl at 25°C. The middle trace is the 500 MHz 1D $^1$H spectrum of the *H. h.* 29-mer RNA motif (2.4 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4$$^{3-}$ (pH 6.0), 20 mM NaCl at 25°C. The lower trace is the 800 MHz 1D $^1$H spectrum of the *H. h.* 37-mer RNA motif (0.2 mM), dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4$$^{3-}$ (pH 6.4) and 20 mM NaCl at 25°C.

Table 4.4.3 illustrates the differences in carbon chemical shift between the *E. coli* 29-mer and the two *H. h.* RNA motifs. Some of the $^{13}$C chemical shifts appeared to differ significantly between the motifs, however, the largest chemical shifts differences occurred where there are differences in the sequence or adjacent to these differences, eg A21, in the 29-mers. The A21 is residue a terminal residue in the two 29-mers, but in the 37-mer it is adjacent to another C-G base pair which is followed by the tetra loop. Additionally there appears to be significant differences in the A27 $^{13}$C chemical shifts, indicating that there may be differences in the conformation of the structure around this point. A22 C1’ chemical shift difference between the two 29-mers is -2.66 ppm and it is unclear here why there is such a large difference.
Table 4.4.3: The table shows the difference carbon and phosphorus chemical shifts between the *H. h.* 29-mer and the *H. h.* 37-mer (A), the *H. h.* 29-mer and the *E. coli* 29-mer (B), and the *H. h.* 37-mer and the *E. coli* 29-mer (C). Where the delta value of the $^{13}$C chemical shift is less than or equal to 0.25 the cell is coloured green, if the value is greater than 0.25 and less than or equal to 0.50 are coloured yellow, if it is greater than 0.50 and less than or equal to 1.00 the cells were coloured orange and if the value was greater than 1.00 the cells were coloured red. Where the delta value of the $^{31}$P chemical shift is less than or equal to 0.10 the cell is coloured green, if the value is greater than 0.10 and less than or equal to 0.25 are coloured yellow, if it is greater than 0.25 and less than or equal to 0.50 the cells were coloured orange and if the value was greater than 0.50 the cells were coloured red. A colour coded *H. h.* and *E. coli* sequence are shown at either side of the table sequence. A “~” marks cells where the data was not available, a greyed out cell indicates that the value is not applicable.
Phosphorus chemical shifts were only assigned for the *E. coli* 29-mer and the *H. h.* 37-mer, the difference between the assigned resonances differed quite considerably as shown in Table 4.4.3.

### 4.4.2 Comparison of the solution structures

The *H. h.* 29-mer RNA and *E. coli* 29-mer solution structures have both been fully determined, but the *H. h.* 37-mer has unfortunately only been partly determined. Therefore only the two 29-mers will be discussed in detail. The three average structures are shown in Figure 4.4.3, the figure shows the two 29-mers forming well folded A-form RNAs. The 37-mer structure as stated, has been largely undetermined and so a large amount is unstructured, particularly near the regions of interest (the bulges). The secondary structures revealed in the average solution structures are shown in Figure 4.4.4. Here differences can be clearly seen between the 29-mers which are discussed subsequently, and the unstructured nature of the *H. h.* 37-mer solution structure is clearly visible.

![Figure 4.4.3: The left image is the average NMR structure of the *E. coli* 29-mer RNA. The centre image is the average NMR structure of the *H. h.* 29-mer RNA. The right image is the NMR solution structure of the *13C* and *15N* labelled *H. h.* 37-mer RNA. Some important residues are labelled.](image-url)
The most obvious difference between the two 29-mer structures is the position of U6. It is clearly flipped out in the *E. coli* 29-mer, but is situated within the helix in the *H. h.* 29-mer. Various X-ray crystal structures of the ribosome indicate that U6 in the ribosome is base paired to an adenine base from another section of the RNA (see Figure 1.6.3). This is the case for the *E. coli* ribosome crystal structure, but there is not an X-ray crystal structure for the *H. h.* ribosome, however it would be unlikely to deviate from the observed pattern. The strong G-C base pair below U6 in *E. coli* is likely to help hold the base outside the helix without it base pairing to an additional adenine. Replacing the two flanking potential G-C base pairs found in the *E. coli* sequence with two weaker A-U base pairs, may allow the U6 into the helix. But it may also be the case that if there was an additional adenine outside the helix to base pair with the U6, then the U6 may favour being flipped out. If this is true it may be that the *H. h.* 29-mer RNA motif will not make a good model of this.
section of the RNA as it is within the ribosome. The fact that in solution the U6 of the *E. coli* 29-mer is flipped out and exposed to solvent, explains why the U6 imino proton is available for exchange with the solvent and so is not visible in the 1D $^1$H spectrum recorded in 90% $^1$H$_2$O + 10% $^2$H$_2$O (see Figure 4.5.1). The U6 imino proton of the *H. h.* 29-mer is visible in such a spectrum and so fits with the U6 being within the helix, thus the imino proton being protected from exchange. In fact two peaks due to the U6 imino proton are visible in the *H. h.* 29-mer suggesting that the U6 is held within two distinct conformations, both held within the helix. The U6 being situated within the helix leads to a non-canonical U6-U30 base pair leaving A7 unpaired.

Another major difference between the two 29-mer structures is that in the *H.h.* 29-mer A9 and A27 appear to form a base pair leaving U26 unpaired. In the *E. coli* 29-mer solution structure A9 forms a base pair with U26. However, A27 is still within the helix and with very little rearrangement of the conformation a base pair could be formed between A9 and A27. This apparent flexibility in the *E. coli* 29-mer appears less likely to be found in the *H. h.* 29-mer solution structure as there would have to be a more extensive change in conformation to form an A9 to U26 base pair.

Overall, despite the distinct differences, the two 29-mer structures are very similar as would be expected. Both structures have good all atom average RMSD values, and a stable A-form conformation. The 29-mer structures provide good models for this section of the RNA within the ribosome in a more natural solution state, than that observed in the intact ribosome crystal structures, particularly around the A27 residue.
Chapter 5

Interactions of isotopically normal 23S rRNA motifs with amicetin

The aim of this chapter is to discuss the interaction of amicetin with the *H. h.* 29-mer and *E. coli* 29-mer RNA motifs. The binding of amicetin to the ribosome has so far not been characterised by X-ray crystallography. Through NMR, this project aims to gather more information on the binding. The interaction of amicetin with the two motifs is expected to be weak in magnitude, as previously demonstrated for the *H. h.* 29-mer motif by C. Shammas. In the subsequently described two titrations, four signs of interaction were looked for:

- changes in chemical shift
- changes in the shape and linewidth of peaks
- changes in the intensities of intramolecular NOEs
- the appearance of any inter RNA-antibiotic NOEs

5.1 **Interactions** of normal *H. h.* 29-mer RNA-amicetin complex

The titration of the *H. h.* 29-mer RNA with amicetin was carried out previously. The data will now be reassessed in light of the corrected *H. h.* 29-mer RNA assignment (section 4.2).
Changes to the *H. h.* RNA spectrum

The most distinct chemical shift and intensity changes observed occurred for the imino resonances. The largest changes in the imino resonances chemical shifts and intensities occurred for U13, G10, U26, U30, U6, U6a and U31, as indicated in Figure 5.1.1 and as can be seen in Figure 5.1.2. These residues all occur close to the bulge regions of the RNA motif except for U13, which occurs near the end of the RNA molecule. This indicates that the binding of amicetin is likely to occur in the bulge region. However, in contrast to the previous assignment, this evidence does not indicate a particular region of the bulge where amicetin is likely to interact.

![RNA secondary structure diagram](image)

**Figure 5.1.1:** The figure shows the observed secondary structure of the *H. h.* 29-mer. The boxed base pairs and boxed non-base paired residues are those with imino protons whose resonances vary the greatest upon titration. The double and single solid lines represent standard A-U and G-C W-C base pairs respectively, dotted lines represent non-standard base pairs.
Figure 5.1.2: A series of 600 MHz 1D $^1$H spectra of the H. h. 29-mer RNA motif (2.4 mM), dissolved in 90% $^1$H$_2$O and 10% $^2$H$_2$O containing 20 mM PO$_4$$^{3-}$ (pH 6.0) and 20 mM NaCl at 2°C all with various molar equivalents (eq) of amicetin as indicated to the right of the spectrum.

A number of other chemical shift changes were observed in the non-exchangeable proton resonances, the largest being the U6 H6 chemical shift with a change of 0.1 ppm followed by the U6 H5 with a shift of 0.08 ppm. These changes in chemical shift indicate binding, potentially around the U6 residue but as stated smaller changes are observed across the RNA.
The previously reported disappearance of a number of H5 H6 NOE peaks upon addition of amicetin were mistaken. They can be clearly seen in the spectrum; this error may have been due to over use of water suppression during data processing as the “lost” peaks appear close to the water resonance.

Changes to the amicetin spectrum

Chemical shift differences were also observed for the amicetin resonances. Some of these differences are shown in Table 5.1.1. It is interesting to note that the largest chemical shift differences occur around the two aromatic rings and around the proton groups 7*/8* and 4*, which occur at the end of the molecule. Resonances 13/9 and 12/10 are due to protons situated on the benzene like aromatic ring and were not observed, despite the fact that the observed aromatic resonances 6 and 5 should have had a weaker intensity. It is also of note that the H2'a resonance had a significant change in chemical shift, yet H2'e does not, despite H2'e being attached to the same carbon as H2'a. This could indicate that part of the RNA is specifically interacting with H2'a, or that there is a conformational change in the antibiotic and this is causing part of the amicetin molecule to specifically interact with H2'a. For instance, a change in the stacking observed between the two aromatic rings of amicetin could go some way in explaining the chemical shift differences observed in this region.
### Table 5.1.1:
The table showing some of the chemical shifts (δ, ppm) of assigned ¹H resonances of amicetin (3 mM in 90% ¹H₂O and 10% ²H₂O, pH 6.15 2°C)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>¹H Chemical shift (δ, ppm)</th>
<th>¹H Chemical shift (δ, ppm) Complex</th>
<th>Δδ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7*/8*</td>
<td>3.02</td>
<td>2.88</td>
<td>0.14</td>
</tr>
<tr>
<td>4*</td>
<td>3.25</td>
<td>2.99</td>
<td>0.26</td>
</tr>
<tr>
<td>2'a</td>
<td>1.9</td>
<td>1.72</td>
<td>0.18</td>
</tr>
<tr>
<td>2'e</td>
<td>2.16</td>
<td>2.13</td>
<td>0.03</td>
</tr>
<tr>
<td>1'</td>
<td>5.83</td>
<td>5.64</td>
<td>0.19</td>
</tr>
<tr>
<td>6</td>
<td>8.27</td>
<td>8.05</td>
<td>0.22</td>
</tr>
<tr>
<td>5</td>
<td>7.53</td>
<td>7.17</td>
<td>0.36</td>
</tr>
<tr>
<td>13/9</td>
<td>7.68</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12/10</td>
<td>7.96</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>1.68</td>
<td>1.57</td>
<td>0.11</td>
</tr>
</tbody>
</table>

The nuclei shown are potentially important in the binding process. The difference in chemical shift between the complex amicetin and free amicetin is shown in the Δδ column. Where the difference is greater than 0.10 the cell is coloured yellow and where the difference is greater than 0.25 the cell is coloured red.

It is also of note that chemical shift differences in resonances 7*/8* and 4* are observed, as this is the equivalent region to the region of blasticidin S shown to interact with the RNA motif through X-ray crystal structures of the blasticidin S bound ribosome. It should also be noted that these two proton groups appear very sensitive to changes in sample conditions.

**H. h. RNA-amicetin NOEs**

Inter RNA antibiotic NOEs were searched for in the complex NOESY spectra, but as reported previously none were found.

**Summary**

In summary the data indicates a weak interaction between the aromatic rings and surrounding area of amicetin, and also the area near the 7*/8* and 4* proton groups of amicetin with the bulge region of H. h. 29-mer RNA motif. There are some indications that the amicetin may interact with the region of the bulge surrounding U6.
5.2 Interactions of normal *E. coli* 29-mer RNA-amicetin complex

The titration of *E.coli* 29-mer RNA was carried out by collecting a series of NMR spectra after the addition of increasing amounts of amicetin. This included a series of 1D $^1$H, 2D $^1$H-$^1$H TOCSY and 2D $^1$H-$^1$H NOESY spectra, see Appendix B.1 for a full list of experiments carried out during the titration.

**Changes to the RNA spectrum**

Figure 5.2.1 shows a stack plot of the proton 1D spectra measured in $^2$H$_2$O at 25ºC. The figure shows the chemical shift range of 9.0 to 0.5 ppm. Small but clear differences can be seen in the RNA resonances displayed in this spectrum, indicating a weak interaction between the amicetin and the RNA. However, these changes are difficult to characterise just from the 1D spectrum.
Figure 5.2.1: A series of 600 MHz 1D $^1$H spectra of the *E. coli* 29-mer RNA motif (1 mM) dissolved in 90% $^1$H$_2$O and 10% $^2$H$_2$O containing 20 mM PO$_4$$^{3-}$ (pH 6.2) and 20 mM NaCl at 25ºC, all with various molar equivalents of amicetin as indicated to the left of the spectrum.

To look at the changes in the RNA spectra in detail, a series of $^1$H-$^1$H TOCSY spectra were measured. Using these, a number of chemical shift movements were observed, particularly around the bulges and fraying ends of the RNA. For example, Figure 5.2.2 shows an overlay of TOCSY spectra taken at various equivalents of amicetin (0 to 0.390 eq) of the H5-H6 correlations. Changes in the H5-H6 chemical shifts of up to 0.1 ppm can be observed; the largest chemical shift differences interestingly occur around the C28 residue. This residue's correlation also appears to broaden as amicetin is added, indicating binding. The C28 residue is close to the U26 residue which, in the *H. h.* ribosome, spontaneously mutates to C when exposed to sub inhibiting concentrations of amicetin to C. This region of the motif is where it has been shown, by X-ray crystallography, that blasticidin S interacts upon binding to the intact ribosome.
Figure 5.2.2: An overlay of the 600 MHz 2D TOCSY spectra of the *E. coli* 29-mer RNA motif (1 mM) dissolved in 100% $^2$H$_2$O containing 20 mM PO$_4^{3-}$ (pH 6.2) and 20 mM NaCl at 25°C, with various molar equivalents of amicetin ranging from 0 eq (light orange), to 0.390 eq (dark blue). The arrows indicate the changes in chemical shift of the correlation peaks. The C28 correlation is particularly weak and so is not visible on the figure, though is observable in the original data.

Additionally through the $^1$H-$^1$H TOCSY spectra, it can be observed that the H1' resonance of A27 shifts by 0.04 ppm, that the H1' of U6 shifts by 0.05 ppm and that the H1' of U29 shifts by 0.04 ppm. These shifts were all observed from 0 to 0.39 equivalents (eq).

$^1$H-$^1$H NOESY spectra were also measured at 0 and 0.39 eq. It was observed in these spectra that there was a small change in a number of chemical shifts, one of the largest that was easily observable was G5 H8 by 0.04 ppm. There were also some clear intensity changes in some of the NOE cross peaks, such as those involving A27 H8, A8 H8, and A9 H8, which appeared weaker and those involving C28 H8, U29 H6, 7G H8 and U6 H6, which were not clearly visible. These changes in intensity appear to be due to broadening of the resonances.
When a comparison was made between the NOESY spectrum measured at 0 eq and 3.6 eq, it was revealed that the two spectra were very similar with respect to the RNA peaks. One of the largest differences was a change in chemical shift of 0.02 ppm for the C28 H6. The NOESY spectrum measured at 1.17 eq also displayed a very similar spectrum to the 0 eq NOESY spectrum. This phenomenon can also be observed in the 1D spectra as displayed in Figure 5.2.1, where the spectrum measured at 1.17 eq is more similar to the spectrum measured at 0 eq than to the one measured at 0.39 eq. Closer inspection reveals that from the addition of 0.78 eq of amicetin (also coinciding with a lyophilisation to reduce sample volume), a pH change appears to have begun to occur (inferred from the change in amicetin chemical shifts and comparison to a previous pH titration) which appears to inhibit the interaction of amicetin to the RNA. Therefore, titration points past 0.39 equivalents must be examined using extra caution.

Figure 5.2.3 shows the 1D $^{31}$P spectrum of the E. coli 29-mer and the 1D $^{31}$P spectra of the E. coli 29-mer amicetin complex. Small but definite differences are observed between the spectra, but most are difficult to identify due to overlap of resonances. The A27 $^{31}$P resonance shifts from -3.78 ppm in the uncomplexed sample, then to -3.88 ppm with 1.24 eq of amicetin and then back to -3.77 ppm with 3.6 eq of amicetin. This is an unusual pattern and the reason for it is unclear. Overall, it is difficult to make any further assessment from this series other than to say that there are some indications of changes in the conformation of the RNA backbone, which may suggest binding.
Figure 5.2.3: The top spectrum is the 81 MHz 1D $^{31}$P (200 MHz, $^1$H) spectrum of the *E. coli* 29-mer RNA motif (1 mM)dissolved in 100% $^{2}$H$_2$O containing 20 mM PO$_4^{3-}$ (pH 6.2) and 20 mM NaCl at 27ºC, the middle spectrum is the 81 MHz 1D $^{31}$P spectrum of the *E. coli* 29-mer RNA with 1.24 equivalents of amicetin, under the same conditions, the lower spectrum is the 81 MHz 1D $^{31}$P spectrum of the *E. coli* 29-mer RNA (1 mM) dissolved in 90% $^{1}$H$_2$O and 10% $^{2}$H$_2$O containing 20 mM PO$_4^{3-}$ (pH 6.2) and 20 mM NaCl with 3.6 equivalents of amicetin, at 27ºC. The region of greatest change is indicated with a series of lines on the spectra.

Figure 5.2.4 shows the imino region of the *E. coli* 29-mer and the imino region with 3.6 eq of amicetin. There was however, negligible change observed in this region. There does appear to be differences in intensity for residues U13, U12 and G35 at the fraying ends of the RNA.
Figure 5.2.4: The *top spectrum* is the 700 MHz 1D $^1$H spectrum of the *E. coli* 29-mer RNA motif (1 mM) dissolved in 90% $^1$H$_2$O and 10% $^2$H$_2$O containing 20 mM PO$_4$$^{3-}$ (pH 6.2) and 20 mM NaCl at 2°C, the *lower spectrum* is the 700 MHz 1D $^1$H spectrum of the *E. coli* 29-mer RNA with 3.6 equivalents of amicetin, under the same conditions. The sequence specific assignment is shown at the top (imino) of the spectrum.

Changes to the amicetin spectrum

As the amicetin peaks become visible in the titration data, the peaks were observed to be much broader than would be expected for amicetin in isolation (Figure 5.2.1). The broader peaks indicate a faster $T_2$ relaxation period, which in turn indicates at least some interaction between the amicetin and the RNA, which would cause a faster transverse relaxation. Unfortunately, it was not possible to analyse the amicetin resonances in detail as most were only clearly visible after 0.39 eq and as has been observed previously, the amicetin resonance can be quite pH dependent. However, one resonance that has a clear chemical shift change from 0.07 eq to 0.39 eq of 0.21 ppm is H7*/8*. Notably, its chemical shift at 0.07 eq is 2.77 ppm, at 0.39 eq it is 2.98 and in free amicetin at a pH of 6.15 it is 3.02 ppm. This looks like a weak binding system moving from a point where the majority of
the amicetin is bound, to a point where most is unbound. However, this resonance is very sensitive to sample conditions and so this evidence must be looked upon with caution.

Inter RNA-antibiotic NOEs

$^1$H-$^1$H NOESY spectra were conducted at various points throughout the titration in order to locate any inter RNA-antibiotic NOEs, and to allow the possibility of NMR solution structure of the complex. No intermolecular NOEs were observed, even in the data collected on the 1 GHz spectrometer at 3.6 eq at 2ºC, or in the spectra collected at 0.39 eq.

To summarise the information gained strongly indicated weak binding at low equivalents of amicetin. Unfortunately, the possible change in pH at high equivalences of amicetin appears to have inhibited binding later on in the titration. The binding appears to occur in the bulge region and potentially centres around the C28 residue, the U6 or perhaps other sites are involved. It is difficult to say which parts of the amicetin may be strongly involved in the binding but as has been observed, there are indications that the two methyls, $7^*/8^*$, are involved.
5.3 Discussion of the RNA-antibiotic complexes of the *E. coli* and *H. h.* rRNA motifs

Evidence of amicetin binding to the *E. coli* 29-mer RNA motif and the *H. h.* 29-mer RNA motif were observed.

The changes in the RNA spectra appear to indicate that for both the *E. coli* and the *H. h.* 29-mer RNA motifs, the amicetin binds within the bulge region. However, the data for the *H. h.* 29-mer seems to slightly favour a site near the U6 residue and the *E. coli* 29-mer seems to favour a site near the flipped out C28 residue. Both systems show indications that the region around the 7*/8* methyl groups of amicetin may be important for binding. The *H. h.* 29-mer system also shows indications that the regions around the two aromatic rings may be important for binding.

Although through the binding data obtain in this project the exact binding location cannot be determined, valuable information has been gained, which starts to build up a picture of the bound state of the amicetin to the RNA motifs. This data further indicates that the binding is weak and as previously stated gives an indication about the possible site of binding.
Chapter 6

Conclusion and further work

Conclusion

The project offered an opportunity to apply state of the art NMR techniques to elucidate a biologically significant area of research in RNA structural biology. Further the NMR structures of several RNA binding antibiotics have been determined leading to the proposal of a new scheme to generate novel hybrid antibiotics (see Figure 3.4.3).

From this project, the 3D structural similarity between the aminohexose cytosine nucleoside antibiotics has been clearly demonstrated, and each has been shown to have a number of structurally important intramolecular hydrogen bonds. The NMR structures of the proposed amicetin binding RNA motif from *E. coli* and *H. h.* have also been determined. The binding of amicetin to the *H. h.* and *E. coli* 29-mer RNA motifs have also been characterised via NMR.

Antibiotics

Blasticidin S and gougerotin were successfully assigned and their NMR structures determined via NMR to a high quality. In addition, the previously determined amicetin NMR structure\textsuperscript{32} was corrected. All three structures were shown to be very similar, sharing certain structural aspects despite the differences observed in chemical structure. They were all found to have a number of intramolecular hydrogen bonds that act to explain the retardation of a number of exchangeable protons as observed by NMR. Some of these protons were found to be retarded up to relatively high temperatures, up to 50°C for blasticidin S, 40°C for gougerotin and 25°C for amicetin. This indicated that these hydrogen bonds and consequently the surrounding structure are thermodynamically very stable.
The determined NMR structures of blasticidin S gougerotin and amicetin shared similarities with their equivalent crystal structures, although there were important observable differences, such as a differing set of intramolecular hydrogen bonds. The amicetin solution structure is the most distinctive from its crystal structure, due to the fact that the structure folds to allow the two aromatic rings to stack in solution, whereas the crystal structure is very linear.

Through the determination of the solution structures of a number of aminohexose cytosine nucleoside antibiotics, particularly after the correction of the amicetin solution structure, it became clear how the similar, yet differing chemical structures lead to similar three dimensional structures. From this observation it became clear that the molecule could be clearly divided into different sections, which could be interchanged with the equivalent sections from the other antibiotics of the group. As a consequence of this mixing of different parts, it became clear that hybrid antibiotics may be able to be formed, leading to the opportunity to discover new hybrid antibiotics.

**RNA**

The isotopically normal *E. coli* and *H. h.* 29-mer RNA motifs were both successfully assigned and subsequently their NMR solution structures were successfully determined. Both motifs formed good A-form RNA helices. The *E. coli* 29-mer motif formed a similar structure to that observed for the motif in the X-ray crystal structure of the ribosome, except that the C28 was flipped out instead of folded into the helix and the A27 was folded into the helix instead of flipped out. There were also some rearrangements in the surrounding residues observed. The *H. h.* 29-mer RNA motifs structure differed significantly from the standard structure of the motif, as observed in the crystal structures.
of the ribosome. For example, U6 was folded into the helix and this appears to be unlikely to be the case in the ribosome. Additionally, all the other residues were also folded into the helix, in contrast to what was observed both in the crystal structures of the ribosomes and to the E.coli 29-mer solution structure. These differences raise the question of whether the H. h. 29-mer RNA motif is a good model for how the motif would behave in the intact ribosome.

The doubly $^{13}$C and $^{15}$N labelled H. h. 37-mer RNA motif has been partially assigned and an initial solution structure determined. The sample is potentially too weakly concentrated to allow for the assignment and subsequent structure determination of the region of the RNA past the first bulge towards the 5' and 3' ends of the RNA. Alternatively, it may be that there are multiple conformations in that region of the RNA, or that the structure is simply not stable.

**Complex**

There was strong evidence of weak binding of the amicetin to the E. coli 29-mer motif. The data indicated that the binding occurred in the bulge region, probably around the C28 motif. There were also indications that the 7*/8* methyl groups of the amicetin played an important role in binding.

There is strong evidence for weak binding of the amicetin with the H. h. 29-mer motif. The changes to the RNA resonances and peaks suggested that the amicetin binds in the bulge region of the RNA. The changes in the amicetin spectrum suggested the antibiotic made contact with the RNA around the two aromatic rings and the 7*/8* methyls.
Further Work

There are several potential routes to take to try to further the work of this project. One potentially useful next step is to attempt to synthesise a selection of hybrid antibiotics and subsequently test their antibiotic activity and then determine their solution structures, to enable correlation between activity and structure. The synthesis could be aided by previous studies into the synthesis of the aminohexose cytosine nucleoside antibiotics, for example, a study into the chemical synthesis of gougerotin.¹¹⁹

The assignment of the doubly labelled $^{13}$C, $^{15}$N H. h. 37-mer RNA motif has only been partially completed. There appears to be a problem of, some missing key NOEs, particularly towards the 5' and 3' ends of the motif, required for assignment. Therefore, one possible way to complete the assignment may be to run a 2D $^1$H-$^1$H NOESY with a large number of scans to help locate these key missing NOEs. Once the assignment is complete the structure determination could then be completed.

Once the structure determination of the 37-mer is complete it may be appropriate for it to be titrated with amicetin to further characterise the binding. Performing the titration with a doubly labelled RNA sample would allow changes in the $^{13}$C and $^{15}$N chemical shifts to be observed. It also would allow the use of isotope filtered NOE experiments where only NOEs between the labelled RNA and unlabelled antibiotics are observed, therefore removing the problem of intermolecular NOEs being missed due to overlap with intramolecular NOEs.¹²⁰ It would also be possible to simply observe the antibiotic intra NOE peaks or just the RNA intra peaks to get a clearer image of the different components of the complex.¹²⁰
The *E. coli* 29-mer, amicetin titration should also be repeated, to enable more accurate measurements of the titration after 0.39 eq. This is needed as the pH of the sample appeared to alter past this point in the titration apparently inhibiting the binding of amicetin.

Another piece of work which would be useful to carry out is the structure determination of the analogous *H. h.* 29-mer amicetin resistant RNA motif (U26 to C).\(^{33,28}\) Knowing the structure of the resistant motif and the differences between the two may give insight into how the resistance arises from the mutation, it may also give additional information on how the amicetin binds.

The X-ray crystal structure of blasticidin S bound to the ribosome shows the cytosine moiety of blasticidin S base pairing with a flipped out guanine of the P-loop, and with its “tail” interacting with the proposed amicetin binding motif. The corrected structure determination of amicetin has shown the similarity between the antibiotics more clearly, therefore it may be worthwhile to investigate the binding of amicetin to the P-loop via NMR. Once the this study has been completed, titration of hybrid antibiotics to the main binding motif (as determined by the this later study) may also yield interesting information, which may enable the better design of further hybrid antibiotics.
Presentations and papers

Presentations

The results of this project have been presented on a number of occasions, the most significant are listed below.

Poster presented at the RSC NMR Discussion Group meeting, Department of Chemistry, University of Cambridge, April 2008 (Awarded competitive bursary). Additionally an image of the antibiotic solution structures was used on the front cover of the meetings programme.

Oral presentation at the RSC NMR Discussion Group, Post-Graduate meeting, School of Chemistry, University of Manchester, June 2009.

Papers to be published

The results of this project (chapters 3-5) will be published in peer reviewed journals and they are listed below.


270
References


Kirillov, S. V.; Porse, B. T. & Garrett, R. A. (1999), 'Peptidyl transferase antibiotics perturb the relative positioning of the 3'-terminal adenosine of P/P'-site-bound tRNA and 23S rRNA in the ribosome', *RNA* **5**, 1003-1013.


http://www.umanitoba.ca/chemistry/nmr/spinworks/


T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco


http://www.gaussian.com/

http://gabedit.sourceforge.net/


Appendices

Appendix A: Experimental scripts

A.1. Typical NMRPipe processing script

2D processing script

```
bruk2pipe -in /home/john/JKdata_18062009/23/ser \
    -bad 0.0 -noaswap -DMX -decim 24 -dspfs 12 -grpdly -1 \
    -xN 6144 -yN 600 \
    -xT 2999 -yT 300 \
    -xMODE DQD -yMODE States-TPPI \
    -xSW 7183.908 -ySW 7183.908 \
    -xOBS 599.927 -yOBS 599.924 \
    -xCAR 4.7735 -yCAR 4.7735 \
    -xLAB 1Hx -yLAB 1Hy \
    -ndim 2 -aq2D States \ 
    -out /home/john/JKdata_18062009/23/test.fid -verb -ov 
sleep 5
nmrPipe -verb -in /home/john/JKdata_18062009/23/test.fid \
    | nmrPipe -fn POLY -time \ 
    | nmrPipe -fn SOL -fl 32 \ 
    | nmrPipe -fn GMB -lb -2.0 -gb 0.08 -c 0.5 \ 
    | nmrPipe -fn ZF -auto \ 
    | nmrPipe -fn FT -auto \ 
    | nmrPipe -fn PS -p0 -39 -p1 -17 -di \ 
    | nmrPipe -fn POLY -auto -ord 2 -verb \ 
    | nmrPipe -fn TP \ 
    | nmrPipe -fn SP -off 0.5 -end 0.98 -c 0.5 \ 
    | nmrPipe -fn ZF -auto \ 
    | nmrPipe -fn FT -auto \ 
    | nmrPipe -fn PS -p0 -92.4 -p1 177 -di \ 
    | nmrPipe -fn POLY -auto -ord 2 -verb \ 
    | nmrPipe -fn TP \ 
    | nmrPipe -ov -verb -out /home/john/JKdata_18062009/23/Ec29mer+200microlAmicetinTOCSY75ms.ft2
```

3D processing script

```
var2pipe -in /windows/D1/JKdata_130709/37mer_HCP.fid/fid -noaswap \
    -xN 1000 -yN 96 -zN 64 \ 
    -xT 500 -yT 48 -zT 32 \ 
    -xMODE Complex -yMODE Complex -zMODE Complex \ 
    -xSW 5000.000 -ySW 5000.000 -zSW 1199.994 \ 
    -xOBS 599.893 -yOBS 150.854 -zOBS 242.840 \ 
    -xCAR 4.773 -yCAR 78.996 -zCAR -3.582 \ 
    -xLAB H1 -yLAB C13 -zLAB P31 \ 
    -ndim 3 -aq2D States \ 
    -out /windows/D1/JKdata_130709/37mer_HCP.fid/data/test%03d.fid -verb -ov 
sleep 5
xyz2pipe -in /windows/D1/JKdata_130709/37mer_HCP.fid/data/test%03d.fid -x -verb \
    | nmrPipe -fn POLY -time \ 
    | nmrPipe -fn GMB -lb -2.0 -gb 0.08 -c 0.5 \ 
    | nmrPipe -fn ZF -auto \ 
    | nmrPipe -fn FT -auto \ 
    | nmrPipe -fn PS -p0 -122 -p1 0 -di \ 
    | nmrPipe -fn TP \ 
    | nmrPipe -fn SP -off 0.5 -end 0.98 -c 0.5 \ 
    | nmrPipe -fn ZF -auto \ 
    | nmrPipe -fn FT -auto \ 
    | nmrPipe -fn PS -p0 0 -p1 0 -di \ 
    | pipe2xyz -out /windows/D1/JKdata_130709/37mer_HCP.fid/data/test%03d.ft2 -y
```

xyz2pipe -in /windows/D1/JKdata_130709/37mer_HCP.fid/data/test%03d.ft2 -z -verb \
A.2. Antibiotic structure calculation scripts and files

Blasticidin S topology file

remark file topallhdg.dna
remark geometric energy function topology for distance geometry and
remark simulated annealing,
remark
remark history:
remark
remark The first version of this file is the fault of MP and LN remark (1984).
remark CHARGES DERIVED FROM toprna10/
remark H charges set to 0.035 and slack taken on C
remark SOME HINTS TAKEN FROM GELENI AND ROSSKY KARPLUS RAHMAN PAPER.
remark GROUPING REMOVED. (ONE GROUP/RESIDUE PUT BACK).
remark H-BOND CHARGES FOR R_DIEL & EPS=1.0
remark REDUCED PHOSPHATE CHARGES. CH-stretches and bends from
remark uracil normal mode fit.
remark Sugar dihedrals adapted from W.Olson JACS (1982)
remark planarity of hydrogens on bases is now maintained by impropers
remark instead of dihedrals, and all impropers governing the planes
remark have their force constants increased to 80 (G.M. CLORE
remark 16/5/85)
set echo=false end

AUTOGENERATE ANGLES=true END
AUTOGENERATE dihe=true END

{*==================================*}

{* DNA/RNA default atomic masses *}
MASS  h1 1.00800
MASS  h2 1.00800
MASS  h4 1.00800
MASS  ha 1.00800
MASS  hc 1.00800
MASS  ho 1.00800
MASS  hn 1.00800
MASS  h* 1.00800
MASS  c 12.01100
MASS  ch 12.01100
MASS  c2 12.01100
MASS  c3 12.01100
MASS  ca 12.01100
MASS  cd 12.01100
MASS  ce 12.01100
MASS  cc 12.01100
MASS  c1 12.01100
MASS  n 14.00670
MASS  ns 14.00670
MASS  n2 14.00670
MASS  na 14.00670
MASS  nb 14.00670
MASS  nc 14.00670
MASS  nd 14.00670
MASS  ns 14.00670
MASS  nh 14.00670
MASS  nq 14.00670
MASS  n3 14.00670

282
MASS nf 14.00670
MASS ne 14.00670
MASS nh2e 16.02270
MASS nh3 14.02270
MASS n* 14.02270
MASS o 15.99940
MASS o2 15.99940
MASS oS 15.99940
MASS oh 15.99940
MASS o* 15.99940
MASS sd 22.98980
MASS p 30.97400
MASS br 79.90400
MASS br 79.90400
MASS mg 24.30500

========================================================================
RESidue BLA
GROUP
ATOM N14 TYPE=n2 CHARGE=-0.998355 END
ATOM C14 TYPE=c2 CHARGE=0.936999 END
ATOM N15 TYPE=na CHARGE=-1.052021 END
ATOM N12 TYPE=na CHARGE=-0.421425 END
ATOM C13 TYPE=c3 CHARGE=-0.366859 END
ATOM C11 TYPE=c3 CHARGE=-0.050288 END
ATOM C10 TYPE=c3 CHARGE=-0.039812 END
ATOM C9 TYPE=c3 CHARGE=-0.584821 END
ATOM N9 TYPE=n3 CHARGE=-1.055180 END
ATOM C8 TYPE=c3 CHARGE=-0.584022 END
ATOM C7 TYPE=c CHARGE=0.853684 END
ATOM O7 TYPE=o CHARGE=-0.603791 END
ATOM N6A TYPE=n CHARGE=-0.871954 END
ATOM C4A TYPE=c3 CHARGE=-0.579328 END
ATOM C5A TYPE=c3 CHARGE=-0.050764 END
ATOM C3A TYPE=c2 CHARGE=-0.195031 END
ATOM O5A TYPE=os CHARGE=-0.548061 END
ATOM C2A TYPE=c2 CHARGE=-0.344805 END
ATOM C1A TYPE=c3 CHARGE=-0.444997 END
ATOM C6B TYPE=c CHARGE=-0.866985 END
ATOM O7A TYPE=o CHARGE=-0.583355 END
ATOM O8A TYPE=oh CHARGE=-0.724882 END
ATOM N1 TYPE=n CHARGE=-0.537141 END
ATOM C6 TYPE=cc CHARGE=-0.248867 END
ATOM C5 TYPE=cd CHARGE=-0.686481 END
ATOM C4 TYPE=cd CHARGE=-1.27495 END
ATOM N3 TYPE=nc CHARGE=-0.959995 END
ATOM C2 TYPE=cc CHARGE=-0.853684 END
ATOM O2 TYPE=o CHARGE=-0.644828 END
ATOM N4 TYPE=nh CHARGE=-1.1381850 END
ATOM H14 TYPE=hn CHARGE=0.456129 END
ATOM H15A TYPE=hn CHARGE=0.415985 END
ATOM H15B TYPE=hn CHARGE=0.425145 END
ATOM H13A TYPE=h1 CHARGE=0.151820 END
ATOM H13B TYPE=h1 CHARGE=0.147190 END
ATOM H13C TYPE=h1 CHARGE=0.153463 END
ATOM H11A TYPE=h1 CHARGE=0.059599 END
ATOM H11B TYPE=h1 CHARGE=0.099155 END
ATOM H10A TYPE=hc CHARGE=0.000188 END
ATOM H10B TYPE=hc CHARGE=0.003915 END
ATOM H9 TYPE=h1 CHARGE=0.042933 END
ATOM H9A TYPE=hn CHARGE=0.386578 END
ATOM H9B TYPE=hn CHARGE=0.376750 END
ATOM H8A TYPE=hc CHARGE=0.127239 END
ATOM H8B TYPE=hc CHARGE=0.153227 END
ATOM H6A TYPE=hn CHARGE=0.355031 END
ATOM H4A TYPE=h1 CHARGE=-0.010138 END
ATOM H5A TYPE=h1 CHARGE=-0.102231 END
ATOM H3A TYPE=h4 CHARGE=0.142218 END
ATOM H2A TYPE=h4 CHARGE=-0.178147 END
ATOM H1A TYPE=h2 CHARGE=-0.121128 END
ATOM H8C TYPE=ho CHARGE=0.509879 END
ATOM H6 TYPE=h4 CHARGE=0.129005 END
ATOM H5 TYPE=ha CHARGE=0.229709 END
ATOM H4B TYPE=hn CHARGE=0.479532 END
ATOM H4C TYPE=hn CHARGE=0.461735 END

BOND N14 C14
BOND N14 H14
ACCE O2  C2
ACCE N4  C4
ACCE N1  C1A
ACCE N1  C2
ACCE N1  C6
ACCE N3  C4
ACCE N3  C2

!X -X -c -o
IMPR N3 N1 C2 O2
IMPR O8A C5A C6B O7A
IMPR N6A C8 C7 O7

!X -X -n -hn
IMPR C4A C7 N6A H6A

!X -X -nh-hn
!!IMPR H4B C4 N4 H4C
!!IMPR H4C C4 N4 H4B
!c -c3-n -hn
IMPR C7 C4A N6A H6A
!c3-o -c -oh
IMPR C5A O7A C6B O8A

!X -X -n2-hn
!!IMPR ? C14 N14 H14

!X -X -na-hn
IMPR H15A C14 N15 H15B
IMPR H15B C14 N15 H15A

!X -X -cd-nh
IMPR C5 N3 C4 N4

END {* BLA *}
set echo=true end
end

Gougerotin topology file

remark  file topallhdg.dna
remark  geometric energy function topology for distance geometry and
remark  simulated annealing.
remark  history:
remark  The first version of this file is the fault of MP and LN
remark  CHARGES DERIVED FROM toprna10/
remark  H charges set to 0.035 and slack taken on C
remark  SOME HINTS TAKEN FROM GELIN AND ROSSKY KARPLUS RAHMAN PAPER.
remark  GROUPING REMOVED. (ONE GROUP/RESIDUE PUT BACK).
remark  H-BOND CHARGES FOR R, DIE & EPS=1.0
remark  REDUCED PHOSPHATE CHARGES. CH-stretches and bends from
remark  uracil normal mode fit.
remark  Sugar dihedrals adapted from W.Olson JACS (1982)
remark  planarity of hydrogens on bases is now maintained by impropers
remark  instead of dihedrals, and all impropers governing the planes
remark  have their force constants increased to 80 (G.M. CLORE
remark  16/5/85)
set echo=false end

AUTOGENERATE ANGLES=true END
AUTOGENERATE dihe=true END
MASS  h1     1.00800
MASS  h2     1.00800
MASS  h4     1.00800
MASS  ha     1.00800
MASS  hc     1.00800
MASS  ho     1.00800
MASS  hn     1.00800
MASS  h*     1.00800
MASS  c     12.01100
MASS  ch    12.01100
MASS  c2    12.01100
MASS  c3    12.01100
MASS  ca    12.01100
MASS  cb    12.01100
MASS  cd    12.01100
MASS  ce    12.01100
MASS  cf    12.01100
MASS  cm    12.01100
MASS  cs    12.01100
MASS  c1    12.01100
MASS  n     14.00670
MASS  ns    14.00670
MASS  n2    14.00670
MASS  na    14.00670
MASS  nb    14.00670
MASS  nc    14.00670
MASS  nd    14.00670
MASS  nh    14.00670
MASS  nq    14.00670
MASS  n*    14.00670
MASS  o     15.99940
MASS  o2    15.99940
MASS  os    15.99940
MASS  oh    15.99940
MASS  o*    15.99940
MASS  sd    22.98980
MASS  p     30.97400
MASS  br    79.90400
MASS  mg    24.30500

RESIdue Gou
GROUP
ATOM C15 TYPE=c  CHARGE=0.6113 END
ATOM C  TYPE=cd CHARGE=1.0985 END
ATOM C1 TYPE=c CHARGE=1.0184 END
ATOM C2 TYPE=cd CHARGE=0.6800 END
ATOM C3 TYPE=cc CHARGE=0.1895 END
ATOM C4 TYPE=c3 CHARGE=0.3013 END
ATOM C5 TYPE=c3 CHARGE=0.2084 END
ATOM C6 TYPE=c3 CHARGE=0.1539 END
ATOM C7 TYPE=c3 CHARGE=0.3510 END
ATOM C8 TYPE=c CHARGE=0.7384 END
ATOM C9 TYPE=c3 CHARGE=0.1442 END
ATOM C10 TYPE=c3 CHARGE=0.1228 END
ATOM C11 TYPE=c3 CHARGE=0.0521 END
ATOM C12 TYPE=c3 CHARGE=0.6754 END
ATOM C13 TYPE=c3 CHARGE=0.0928 END
ATOM C14 TYPE=c CHARGE=0.8355 END
ATOM N  TYPE=nh CHARGE=1.1229 END
ATOM N1 TYPE=nc CHARGE=0.9396 END
ATOM N2 TYPE=n CHARGE=0.3446 END
ATOM N3 TYPE=n3 CHARGE=0.8726 END
ATOM N4 TYPE=n CHARGE=0.9962 END
ATOM N5 TYPE=n CHARGE=1.0624 END
ATOM N6 TYPE=n CHARGE=0.7222 END
ATOM O  TYPE=o CHARGE=0.6136 END
ATOM O1 TYPE=oh CHARGE=0.7635 END
ATOM O2 TYPE=o CHARGE=0.6087 END
ATOM O3 TYPE=os CHARGE=0.6200 END

286
ATOM O4  TYPE=oh  CHARGE=-0.7981  END
ATOM O5  TYPE=oh  CHARGE=-0.7806  END
ATOM O6  TYPE=o  CHARGE=-0.6066  END
ATOM O7  TYPE=o  CHARGE=-0.5894  END
ATOM H24  TYPE=hn  CHARGE=0.3982  END
ATOM E23  TYPE=hn  CHARGE=0.4534  END
ATOM H1  TYPE=hn  CHARGE=0.4529  END
ATOM H2  TYPE=hn  CHARGE=0.4725  END
ATOM H3  TYPE=ha  CHARGE=0.2292  END
ATOM H4  TYPE=ho  CHARGE=0.4984  END
ATOM H5  TYPE=hn  CHARGE=0.4273  END
ATOM A26  TYPE=hl  CHARGE=0.0427  END
ATOM H7  TYPE=hl  CHARGE=-0.0178  END
ATOM H8  TYPE=hl  CHARGE=-0.1214  END
ATOM H9  TYPE=hl  CHARGE=-0.0957  END
ATOM A30  TYPE=hl  CHARGE=-0.0699  END
ATOM H11  TYPE=hl  CHARGE=-0.0552  END
ATOM H12  TYPE=hl  CHARGE=-0.0051  END
ATOM B26  TYPE=hl  CHARGE=-0.0262  END
ATOM D26  TYPE=hl  CHARGE=-0.0010  END
ATOM H15  TYPE=ho  CHARGE=-0.4977  END
ATOM H16  TYPE=ho  CHARGE=-0.4993  END
ATOM E27  TYPE=hn  CHARGE=-0.4627  END
ATOM A18  TYPE=hl  CHARGE=-0.0730  END
ATOM B18  TYPE=hl  CHARGE=-0.0561  END
ATOM B30  TYPE=hl  CHARGE=-0.0568  END
ATOM H21  TYPE=hl  CHARGE=-0.0708  END
ATOM H22  TYPE=hn  CHARGE=-0.3797  END

BOND H N
BOND N H1
BOND N C
BOND N1 C
BOND N1 C1
BOND O C1
BOND C C2
BOND C1 N2
BOND C2 H3
BOND C2 C3
BOND H2 C4
BOND N2 C3
BOND N2 C4
BOND O1 H4
BOND O1 C5
BOND C3 H8
BOND C4 C5
BOND C4 O3
BOND C5 H11
BOND C5 C7
BOND O2 C8
BOND H7 C7
BOND N3 C6
BOND N3 C10
BOND O3 C9
BOND C6 B26
BOND C6 D26
BOND H9 C9
BOND A30 C11
BOND C7 O4
BOND C7 C13
BOND H12 C12
BOND C8 C10
BOND C8 N4
BOND C9 C13
BOND C9 C14
BOND O4 H16
BOND H15 O5
BOND O5 C11
BOND C10 A18
BOND C10 B18
BOND C11 C12
BOND C11 B30
BOND E27 N5
BOND C12 N4
BOND C12 C15
BOND N4 H22
Amicetin topology file

remark file topallhgd.dna
remark geometric energy function topology for distance geometry and
remark simulated annealing.
remark history:
remark The first version of this file is the fault of MP and LN (1984).
remark CHARGES DERIVED FROM toprna10/
remark H charges set to 0.035 and slack taken on C
remark SOME HINTS TAKEN FROM GELIN AND ROSSKY KARPLUS RAHMAN PAPER.
remark GROUPING REMOVED. (ONE GROUP/RESIDUE PUT BACK).
remark REDUCED PHOSPHATE CHARGES. CH-stretches and bends from
remark uracil normal mode fit.
remark Sugar dihedrals adapted from W. Olson JACS (1982)
remark planarity of hydrogens on bases is now maintained by impropers
remark instead of dihedrals, and all impropers governing the planes
remark have their force constants increased to 80 (G.M. CLORE
remark 16/5/85)
set echo=false end

AUTOGENERATE ANGLES=true dihedrals=true END

{* DNA/RNA default atomic masses *}

MASS h1  1.00800
MASS h2  1.00800
MASS h4  1.00800
MASS ha  1.00800
MASS hc  1.00800
MASS ho  1.00800
MASS hu  1.00800
MASS h*  1.00800
MASS c  12.01100
MASS ch  12.01100
MASS c2  12.01100
MASS c3  12.01100
MASS ca  12.01100
MASS ca  12.01100
MASS eb  12.01100
MASS ed  12.01100
MASS ce  12.01100
MASS cf  12.01100
MASS cm  12.01100
MASS cs  12.01100
MASS ct  12.01100
MASS cc  12.01100
MASS c1  12.01100
MASS n  14.00670
MASS ns  14.00670
MASS n2  14.00670
MASS na  14.00670
MASS nb  14.00670
MASS nc  14.00670
MASS nd  14.00670
MASS ns  14.00670
MASS nh  14.00670
MASS nq  14.00670
MASS n3  14.00670
MASS nh2e 16.02270
MASS nh3 14.02270
MASS n* 14.02270
MASS o  15.99940
MASS o2  15.99940
MASS oS  15.99940
MASS oh  15.99940
MASS o*  15.99940
MASS sd  22.98980
MASS p  30.97400
MASS br  79.90400
MASS mg  24.30500

!========================================================================
<table>
<thead>
<tr>
<th>Atom</th>
<th>Type</th>
<th>Charge</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>C18A</td>
<td>c3</td>
<td>-0.38</td>
<td></td>
</tr>
<tr>
<td>H18A</td>
<td>hc</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>H18B</td>
<td>hc</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>C18C</td>
<td>hc</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>C18</td>
<td>c3</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>C17A</td>
<td>ca</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>H17A</td>
<td>ha</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>C16A</td>
<td>ca</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>H16A</td>
<td>ha</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>CE</td>
<td>ca</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>C16B</td>
<td>ca</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>H16B</td>
<td>ha</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>C17B</td>
<td>ca</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>H17B</td>
<td>ha</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>CN5</td>
<td>ca</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>COE</td>
<td>c</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>O5</td>
<td>o</td>
<td>-0.46</td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>c3</td>
<td>-0.38</td>
<td></td>
</tr>
<tr>
<td>H1B</td>
<td>h1</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>H1C</td>
<td>h1</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>H1A</td>
<td>h1</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>C1A</td>
<td>c3</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>H18A</td>
<td>h1</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>H1C</td>
<td>h1</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>H1A</td>
<td>h1</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>N1</td>
<td>n3</td>
<td>-0.78</td>
<td></td>
</tr>
<tr>
<td>O5</td>
<td>oh</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>H5</td>
<td>h2</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>H7</td>
<td>h2</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>c3</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td>H6</td>
<td>h1</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>H5</td>
<td>h1</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>H4</td>
<td>h1</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>C9A</td>
<td>c3</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>H9C</td>
<td>hc</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>H9A</td>
<td>hc</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>H9B</td>
<td>hc</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>H12E</td>
<td>hc</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>C12</td>
<td>c3</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>H12A</td>
<td>hc</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>C11</td>
<td>c3</td>
<td>-0.31</td>
<td></td>
</tr>
<tr>
<td>H11E</td>
<td>hc</td>
<td>-0.08</td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td>c3</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>OC</td>
<td>os</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>C13</td>
<td>c3</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>H13</td>
<td>h2</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td>c3</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>H8</td>
<td>h1</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>OB</td>
<td>os</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>H11A</td>
<td>hc</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>H10</td>
<td>h1</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>N2</td>
<td>n</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>C14</td>
<td>cc</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>H14</td>
<td>h4</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>C15</td>
<td>cd</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>H15</td>
<td>ha</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>CN3</td>
<td>cd</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>N4</td>
<td>n</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>HN4</td>
<td>h4</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>N3</td>
<td>nc</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>COD</td>
<td>c</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>OD</td>
<td>o</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>COF</td>
<td>c</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>OF</td>
<td>o</td>
<td>0.51</td>
<td></td>
</tr>
</tbody>
</table>
ATOM N5 TYPE=n CHARGE=-0.83 END
ATOM HN5 TYPE=hn CHARGE=0.33 END
ATOM N6 TYPE=n3 CHARGE=-0.83 END
ATOM HN6 TYPE=hn CHARGE=0.33 END
ATOM HN7 TYPE=hn CHARGE=0.35 END
ATOM O19 TYPE=oh CHARGE=-0.55 END
ATOM HO19 TYPE=ho CHARGE=0.32 END
ATOM H19A TYPE=h1 CHARGE=0.01 END
ATOM C19 TYPE=c3 CHARGE=-0.18 END
ATOM H19B TYPE=h1 CHARGE=0.01 END
BOND C18A H18A
BOND C18A H18B
BOND C18A C18
BOND C18 N6
BOND N6 HN6
BOND N6 HN7
BOND C18 C19
BOND C19 H19B
BOND C19 H19A
BOND C19 O19
BOND O19 HO19
BOND C18 COF
BOND COF OF
BOND COF N5
BOND N5 HN5
BOND N5 CN5
BOND CN5 C17B
BOND CN5 C17A
BOND C17A C16A
BOND C16A CE
BOND CE C16B
BOND C16B C17B
BOND C16B H16B
BOND C17B H17B
BOND C17A H17A
BOND C16A H16A
BOND CE COE
BOND COE OE
BOND COE N4
BOND N4 CN3
BOND N4 HN4
BOND CN3 N3
BOND CN3 C15
BOND N3 COD
BOND COD OD
BOND COD N2
BOND N2 C14
BOND C14 C15
BOND C15 H15
BOND C14 H14
BOND N2 C13
BOND C13 H13
BOND C13 OC
BOND C13 C12
BOND C12 H12A
BOND C12 H12E
BOND C12 C11
BOND C11 H11A
BOND C11 H11E
BOND C11 C8
BOND C9 H10
BOND C9 C9A
BOND C9A H9B
BOND C9A H9A
BOND C9A H9C
BOND C9 OC
BOND C8 OB
BOND C8 H8
BOND C8 C9
BOND OB C7
BOND C7 OA
BOND C7 H7
BOND C7 C6
BOND C6 H6
BOND C6 O6
BOND O6 HO6
BOND OA C3
Blasticidin S and gougerotin parameter file

set echo=false end

BOND  ow  hw  553.0  0.9572  ! TIP3P water
BOND  c   c   290.0  1.550  ! 3
BOND  c   c1  379.8  1.460  ! 0
BOND  c   c2  449.9  1.406  ! 3
BOND  c   ca  349.7  1.487  ! 1
BOND  c   cc  377.4  1.462  ! 3
BOND  c   cd  377.4  1.462  ! 3
BOND  c   ce  363.8  1.474  ! 1
BOND  c   ha  325.1  1.101  ! 3
BOND  c   nc  438.8  1.371  ! 3
BOND  c   nd  438.8  1.371  ! 3
BOND  c   o   648.0  1.214  ! 1
BOND  c   oh  466.4  1.306  ! 1
BOND  c   n   478.2  1.345  ! 1
BOND  c2  c2  589.7  1.324  ! 1
BOND  c2  c3  328.3  1.508  ! 1
BOND  c2  h4  348.6  1.084  ! 1
BOND  c2  n2  581.1  1.288  ! 1
BOND  c2  na  411.1  1.391  ! 1
BOND  c3  c3  303.1  1.535  ! 1
BOND  c3  cc  337.3  1.499  ! 3
BOND  c3  cd  337.3  1.499  ! 3
BOND  c3  h1  335.9  1.093  ! 3
BOND  c3  h2  326.4  1.100  ! 3
BOND  c3  h3  335.9  1.093  ! 2
BOND  c3  hc  337.3  1.092  ! 3
BOND  c3  hx  338.7  1.091  ! 3
BOND  c3  n  330.6  1.460  ! 1
BOND  c3  n3  320.6  1.470  ! 1
BOND  c3  na  334.7  1.456  ! 1
BOND  c3  nh  332.7  1.458  ! 3
BOND  c3  nd  334.7  1.456  ! 3
BOND  c3  nc  334.7  1.456  ! 3
BOND  c3  nh  334.7  1.456  ! 3
BOND  c3  os  301.5  1.439  ! 1
BOND  ca  ca  478.4  1.387  ! 1
BOND  ca  ha  344.3  1.087  ! 3
BOND  ca  n  372.3  1.422  ! 3
BOND  cc  cc  418.3  1.429  ! 1
BOND  cc  cd  504.0  1.371  ! 3
BOND  cc  h4  350.1  1.083  ! 3
BOND  cc  ha  347.2  1.085  ! 3
BOND  cc  n  426.0  1.380  ! 3
BOND  cc  na  438.8  1.371  ! 3
BOND  cc  nh  431.6  1.376  ! 1
BOND  cc  nd  494.6  1.335  ! 3
BOND  cc  nh  449.0  1.364  ! 3
BOND  cc  oh  405.9  1.347  ! 3
BOND  cc  os  376.1  1.370  ! 3
BOND  cd  cd  418.3  1.429  ! 1
BOND  cd  cd  504.0  1.371  ! 3
BOND  cd  h4  350.1  1.083  ! 3
BOND  cd  ha  347.2  1.085  ! 3
BOND  cd  n  426.0  1.380  ! 3
BOND  cd  nh  449.0  1.364  ! 3
BOND  cd  oh  405.9  1.347  ! 3
BOND  cd  os  376.1  1.370  ! 3
BOND  cd  nc  494.6  1.335  ! 3
BOND  cc  nc  381.8  1.414  ! 3
BOND  cc  nh  484.7  1.341  ! 3
BOND  hn  n  410.2  1.009  ! 3
BOND  hn  n2  375.5  1.029  ! 3
BOND  hn  n3  394.1  1.018  ! 3
BOND  hn  na  406.6  1.011  ! 3
BOND  hn  nh  401.2  1.014  ! 3
BOND   ho   o  357.9  0.981   !  3
BOND   ho   oh  369.6  0.974   !  3
BOND   n    n  469.7  1.390   !  3
BOND   n    n3  443.3  1.408   !  3
BOND   n    nh  437.7  1.412   !  3
BOND   n    o  646.6  1.264   !  3
BOND   n    oh  395.4  1.410   !  3
BOND   n    os  372.3  1.429   !  3
BOND   n3   nh  426.7  1.420   !  1
BOND   n3   o  564.0  1.303   !  3
BOND   n3   os  359.6  1.440   !  1
BOND   nh   nh  453.3  1.401   !  1
BOND   nh   o  596.2  1.287   !  3
BOND   nh   oh  359.6  1.440   !  3
BOND   nh   os  358.5  1.441   !  3
BOND   o    o  384.3  1.430   !  3
BOND   o    oh  294.6  1.517   !  3
BOND   o    os  306.3  1.504   !  3
BOND   oh   oh  340.5  1.469   !  3
BOND   oh   os  342.6  1.467   !  3
BOND   os   os  343.6  1.466   !  1

ANGLE   hw   ow   hw   100.0  104.52   ! TIP3P water
ANGLE   hw   hw   ow   0.0   127.74   ! (found in crystallographic water with 3 bonds)
ANGLE   n    c    o   75.8  122.03   ! 3
ANGLE   c    c    n   69.9  104.81   ! 3
ANGLE   c3   c    n   67.9  115.15   ! 3
ANGLE   c3   c    nh  68.2  113.58   ! 0
ANGLE   c3   c    o   68.0  123.11   ! 3
ANGLE   c3   c    oh  69.8  112.20   ! 3
ANGLE   c3   c    os  77.4  122.88   ! 3
ANGLE   ca   c    n   69.4  112.03   ! 3
ANGLE   ca   c    o   68.7  123.44   ! 3
ANGLE   n   c    nc  73.0  116.84   ! 3
ANGLE   nc   c    o   73.9  125.29   ! 3
ANGLE   c2   c2   c3  64.3  123.42   ! 3
ANGLE   c2   c2   h4  49.3  124.68   ! 3
ANGLE   c3   c2   h4  43.8  127.53   ! 3
ANGLE   n2   c2   na  71.7  123.62   ! 3
ANGLE   na   c2   na  73.7  109.33   ! 3
ANGLE   c    c3   c3  63.8  110.53   ! 3
ANGLE   c    c3   h1  47.6  107.66   ! 3
ANGLE   c    c3   hc  47.2  109.68   ! 3
ANGLE   c    c3   n   66.7  111.56   ! 3
ANGLE   c    c3   n3  65.8  113.91   ! 3
ANGLE   c    c3   os  68.0  108.88   ! 3
ANGLE   c2   c3   c3  63.7  110.96   ! 3
ANGLE   c2   c3   h1  47.0  110.46   ! 3
ANGLE   c2   c3   h2  47.0  110.46   ! 3 added by JK based on c2 c3 h1
ANGLE   c2   c3   os  68.3  108.88   ! 3
ANGLE   c2   c3   n   65.9  112.13   ! 3 added by JK based on c3 c3 n
ANGLE   c3   c3   c3  63.2  110.63   ! 3
ANGLE   c3   c3   cc  64.7  108.10   ! 3
ANGLE   c3   c3   cd  64.7  108.10   ! 3
ANGLE   c3   c3   h1  46.4  110.07   ! 3
ANGLE   c3   c3   h2  46.0  111.59   ! 3
ANGLE   c3   c3   hc  46.4  110.05   ! 3
ANGLE   c3   c3   n   65.9  112.13   ! 3
ANGLE   c3   c3   nq  65.9  112.13   ! 3
ANGLE   c3   c3   n3  66.2  110.38   ! 3
ANGLE   c3   c3   nh  66.5  109.78   ! 3
ANGLE   c3   c3   na  65.8  112.59   ! 3
ANGLE   c3   c3   oh  67.7  109.43   ! 3
ANGLE   c3   c3   os  67.8  108.42   ! 3
ANGLE   h1   c3   h1  39.2  109.55   ! 3
ANGLE   h1   c3   n   49.8  109.32   ! 3
ANGLE   h1   c3   nq  49.8  109.32   ! 3

294
| ANGLE | h1 c3 n3 | 49.4 | 109.92 | 13 |
| ANGLE | h1 c3 nc | 50.1 | 108.57 | 13 |
| ANGLE | h1 c3 nh | 49.7 | 109.96 | 13 |
| ANGLE | h1 c3 o  | 52.0 | 117.19 | 13 |
| ANGLE | h1 c3 oh | 51.0 | 109.88 | 13 |
| ANGLE | h1 c3 os | 50.8 | 108.82 | 13 |
| ANGLE | h1 c3 na | 49.9 | 109.45 | 13 |
| ANGLE | h2 c3 os | 50.8 | 108.58 | 13 |
| ANGLE | h2 c3 n  | 50.0 | 108.31 | 13 |
| ANGLE | he c3 he | 39.4 | 108.35 | 13 |
| ANGLE | n c3 n  | 68.8 | 113.81 | 13 |
| ANGLE | n c3 os | 71.2 | 109.19 | 13 |
| ANGLE | os c3 os | 71.7 | 110.24 | 13 |
| ANGLE | c ca ca | 64.6 | 120.14 | 13 |
| ANGLE | ca ca ca | 67.2 | 119.97 | 13 |
| ANGLE | ca ca ha | 48.5 | 120.01 | 13 |
| ANGLE | ca ca n  | 68.0 | 119.57 | 13 |
| ANGLE | cd cc b4 | 47.2 | 129.11 | 13 |
| ANGLE | cc cd ha | 48.4 | 122.89 | 13 |
| ANGLE | cc cd n  | 70.7 | 115.52 | 13 |
| ANGLE | ed cc n  | 70.7 | 115.52 | 13 |
| ANGLE | ed cd ed | 68.2 | 114.19 | 13 |
| ANGLE | ed cd ha | 46.6 | 123.74 | 13 |
| ANGLE | h4 cc n  | 50.4 | 117.62 | 13 |
| ANGLE | nc ed nh | 72.4 | 120.11 | 13 |
| ANGLE | ed ed nh | 68.6 | 118.98 | 13 |
| ANGLE | ed ed nc | 71.1 | 112.56 | 13 |
| ANGLE | cc cc nd | 71.1 | 112.56 | 13 |
| ANGLE | ed ed nd | 70.0 | 113.42 | 13 |
| ANGLE | cc cc nc | 70.0 | 113.42 | 13 |
| ANGLE | cd ed n  | 67.9 | 119.89 | 13 |
| ANGLE | n cd nc  | 70.9 | 123.86 | 13 |
| ANGLE | c n c3   | 63.9 | 121.35 | 13 |
| ANGLE | c n ca   | 64.3 | 123.71 | 13 |
| ANGLE | c n cc   | 65.2 | 124.19 | 13 |
| ANGLE | c n cd   | 65.2 | 124.19 | 13 |
| ANGLE | c n hn   | 49.2 | 118.46 | 13 |
| ANGLE | c n o    | 71.6 | 118.90 | 13 |
| ANGLE | c n oh   | 69.5 | 113.39 | 13 |
| ANGLE | c n os   | 69.1 | 113.14 | 13 |
| ANGLE | c n ca   | 64.3 | 123.71 | 13 |
| ANGLE | c3 n c3  | 64.0 | 112.62 | 13 |
| ANGLE | c3 n cc  | 63.3 | 121.17 | 13 |
| ANGLE | c3 n cd  | 63.3 | 121.17 | 13 |
| ANGLE | c3 n hn  | 46.0 | 116.78 | 13 |
| ANGLE | ca n hn  | 47.6 | 114.59 | 13 |
| ANGLE | hn n hn  | 39.7 | 117.85 | 13 |
| ANGLE | ed n hn  | 47.5 | 121.52 | 13 |
| ANGLE | c2 n2 hn | 52.6 | 110.33 | 13 |
| ANGLE | c3 n3 c3 | 64.0 | 110.90 | 13 |
| ANGLE | c3 n3 hn | 47.1 | 109.92 | 13 |
| ANGLE | hn n3 hn | 41.3 | 107.13 | 13 |
| ANGLE | c2 na c3 | 64.2 | 117.20 | 13 |
| ANGLE | c2 na hn | 47.6 | 119.28 | 13 |
| ANGLE | c3 na c3 | 60.7 | 125.59 | 13 |
| ANGLE | hn na hn | 39.8 | 116.80 | 10 |
DIHEDRAL  c  n  c3  c  0.85000000  1  180.0 !
DIHEDRAL  c  n  c3  c  0.80000000  1  0.0 !
DIHEDRAL  c3  n  c  c3  0.50000000  1  180.0 !
DIHEDRAL  c3  n  c  c3  0.15000000  1  0.0 !
DIHEDRAL  c3  c3  n  c  0.53000000  1  180.0 !
DIHEDRAL  c3  c3  n  c  0.07000000  1  0.0 !
DIHEDRAL  c2  ne  p5  o  2.30000000  1  0.0 !
DIHEDRAL  c2  nf  p5  o  2.30000000  1  0.0 !
DIHEDRAL  ce  ne  p5  o  2.30000000  1  0.0 !
DIHEDRAL  ce  nf  p5  o  2.30000000  1  0.0 !
DIHEDRAL  cf  ne  p5  o  2.30000000  1  0.0 !
DIHEDRAL  cf  nf  p5  o  2.30000000  1  0.0 !
DIHEDRAL  hn  n  c  o  2.50000000  1  180.0 !
DIHEDRAL  hn  n  c  o  2.00000000  1  0.0 !
DIHEDRAL  c3  ss  ss  c3  3.50000000  1  0.0 !
DIHEDRAL  c3  ss  ss  c3  0.60000000  1  0.0 !
DIHEDRAL  c3  n3  nh  ca  1.90000000  1  0.0 !
DIHEDRAL  c3  n3  nh  ca  1.90000000  1  0.0 !
DIHEDRAL  c3  n3  p5  o  3.00000000  1  180.0 !
DIHEDRAL  c3  n3  p5  o  2.30000000  1  0.0 !
DIHEDRAL  ca  nh  oh  ho  1.20000000  1  0.0 !
DIHEDRAL  ca  nh  oh  ho  3.00000000  2  0.0 !
DIHEDRAL  oh  p5  os  c3  0.25000000  1  0.0 !
DIHEDRAL  oh  p5  os  c3  1.20000000  1  0.0 !
DIHEDRAL  h1  c3  c  o  0.80000000  1  0.0 !
DIHEDRAL  h1  c3  c  o  0.08000000  1  180.0 !
DIHEDRAL  hc  c3  c  o  0.80000000  1  0.0 !
DIHEDRAL  hc  c3  c  o  0.08000000  1  180.0 !
DIHEDRAL  hc  c3  c3  hc  0.15000000  1  0.0 !
DIHEDRAL  hc  c3  c3  hc  0.16000000  1  0.0 !
DIHEDRAL  hc  c3  c2  c2  0.38000000  1  180.0 !
DIHEDRAL  hc  c3  c2  c2  1.15000000  1  0.0 !
DIHEDRAL  ho  oh  c3  0.16000000  1  0.0 !
DIHEDRAL  ho  oh  c3  0.25000000  1  0.0 !
DIHEDRAL  ho  oh  c  o  2.30000000  1  180.0 !
DIHEDRAL  ho  oh  c  o  1.90000000  1  0.0 !

DIHEDRAL  c2  c2  c  o  2.17500000  1  180.0 !
DIHEDRAL  c2  c2  c  o  0.30000000  1  0.0 !
DIHEDRAL  c3  c2  c2  c  6.65000000  1  180.0 !
DIHEDRAL  c3  c2  c2  c  1.90000000  1  180.0 !
DIHEDRAL  c3  c3  c  c3  1.90000000  1  180.0 !
DIHEDRAL  c3  c3  c  c3  0.18000000  1  0.0 !
DIHEDRAL  c3  c3  c3  c  0.25000000  1  180.0 !
DIHEDRAL  c3  c3  c3  c  0.20000000  1  180.0 !
DIHEDRAL  c3  c3  n3  c  0.30000000  1  0.0 !
DIHEDRAL  c3  c3  n3  c  0.48000000  1  180.0 !
DIHEDRAL  c3  c3  os  c  0.38300000  1  0.0 !
DIHEDRAL  c3  c3  os  c  0.10000000  1  180.0 !
DIHEDRAL  c3  c3  os  c  0.38300000  1  0.0 !
DIHEDRAL  c3  c3  os  c  0.80000000  1  180.0 !
DIHEDRAL  c3  c3  os  c  0.10000000  1  0.0 !
DIHEDRAL  c3  c3  os  c  0.85000000  1  180.0 !
DIHEDRAL  c3  c3  os  c  1.35000000  1  180.0 !
DIHEDRAL  c3  c3  na  0.38300000  1  0.0 !
DIHEDRAL  c3  c3  na  0.65000000  1  0.0 !

DIHEDRAL  o  c  os  c  2.70000000  1  180.0 !
DIHEDRAL  o  c  os  c  1.40000000  1  180.0 !
DIHEDRAL  os  c3  na  c2  0.00000000  1  0.0 !
DIHEDRAL  os  c3  na  c2  2.50000000  1  0.0 !
DIHEDRAL  os  c3  os  c3  0.14400000  1  0.0 !
DIHEDRAL  os  c3  os  c3  1.17500000  1  0.0 !
DIHEDRAL  os  c3  oh  0.14400000  1  0.0 !
DIHEDRAL  os  c3  oh  1.17500000  1  0.0 !
DIHEDRAL  oh  c3  oh  0.14400000  1  0.0 !
DIHEDRAL  oh  c3  oh  1.17500000  1  0.0 !

DIHEDRAL  f  c3  c3  1.20000000  1  180.0 !
DIHEDRAL  cl  c3  c3  0.45000000  1  180.0 !
DIHEDRAL  br  c3  c3  0.00000000  1  180.0 !

DIHEDRAL  h1  c3  c3  0.25000000  1  0.0 !
DIHEDRAL  h1  c3  c3  0.25000000  1  0.0 !
DIHEDRAL  h1  c3  c3  0.19000000  1  0.0 !
DIHEDRAL  h1  c3  c3  0.25000000  1  0.0 !
DIHEDRAL  h1  c3  c3  br   0.55000000     1     0.0  !
DIHEDRAL  hc  c3  c3  br   0.55000000     1     0.0  !
DIHEDRAL  hc  c3  c3  oh   0.25000000     1     0.0  !
DIHEDRAL  hc  c3  c3  cl   0.25000000     1     0.0  !
DIHEDRAL  hc  c3  c3  br   0.55000000     1     0.0  !

IMPROPER   X   c   o   10.50000000     2   180.0  !
IMPROPER   X   n   hn   1.10000000     2   180.0  !
IMPROPER   c3  n   hn   1.10000000     2   180.0  !
IMPROPER   c3  n   c3   1.10000000     2   180.0  !
IMPROPER   c3  hc   1.10000000     2   180.0  !
IMPROPER   hc  o   c3   1.10000000     2   180.0  !
IMPROPER   X   X   c3  o2   1.10000000     2   180.0  !
IMPROPER   c3  o2   1.10000000     2   180.0  !
IMPROPER   X   X   c3  cl   1.10000000     2   180.0  !
IMPROPER   hc  c3  cl   1.10000000     2   180.0  !

IMPROPER   X   X   cd  nh   1.10000000     2   180.0  ! added by JK to flatten Cytosine ring
IMPROPER   X   X   nh  hn   1.10000000     2   180.0  ! added by JK to flatten Cytosine ring

{"nonbonding parameter section *}
{" ============================ *

nbonds
atom cdie shift eps=1.0  e14fac=0.4   tolerance=0.5
cutnb=9.0 ctonnb=7.5 ctofnb=8.0
nbxmod=5 vswitch wmin 1.0
end

!                   Emin       sigma    Emin/2     sigma
!                (kcal/mol)     (A)       (for 1-4's)
NONBonded  h       0.015700     1.069078    0.007850    1.069078   7.2382  0.9435
NONBonded  ho      0.000000     1.069078    0.000000    1.069078   2.1447  0.6035
NONBonded  hs      0.015700     1.069078    0.007850    1.069078   2.1447  0.6035
NONBonded  hc      0.015700     2.649532    0.007850    2.649532   9.2028  1.0285
NONBonded  h1      0.015700     2.471352    0.007850    2.471352   9.2028  1.0285
NONBonded  h2      0.015700     2.293173    0.007850    2.293173   9.2028  1.0285
NONBonded  h3      0.015700     2.114993    0.007850    2.114993   9.2028  1.0285
NONBonded  hP      0.015700     1.959977    0.007850    1.959977   9.2028  1.0285
NONBonded  ha      0.015000     2.599642    0.007500    2.599642   9.2028  1.0285
NONBonded  h4      0.015000     2.510552    0.007500    2.510552   9.2028  1.0285
NONBonded  h5      0.015000     2.421462    0.007500    2.421462   9.2028  1.0285
NONBonded  hw      0.000000     1.069078    0.000000    1.069078   2.1447  0.6035
NONBonded  hn      0.015700     1.069078    0.007850    1.069078   7.2382  0.9435
NONBonded  o       0.210000     2.959921    0.105000    2.959921   14.1372  1.1985
NONBonded  o2      0.210000     2.959921    0.105000    2.959921   14.1372  1.1985
NONBonded  ow      0.152000     3.150752    0.076000    3.150752   14.1372  1.1985
NONBonded  oh      0.210400     3.066473    0.105200    3.066473   14.1372  1.1985
NONBonded  os      0.170000     3.000012    0.085000    3.000012   14.1372  1.1985
NONBonded  ct      0.109400     3.399669    0.054700    3.399669   20.5796  1.1592
NONBonded  ca      0.086000     3.399669    0.043000    3.399669   20.5796  1.1592
NONBonded  c       0.086000     3.399669    0.043000    3.399669   20.5796  1.1592
NONBonded  cm      0.086000     3.399669    0.043000    3.399669   20.5796  1.1592
NONBonded  cs      0.086000     3.399669    0.043000    3.399669   20.5796  1.1592
NONBonded  cb      0.086000     3.399669    0.043000    3.399669   20.5796  1.1592
NONBonded  cc      0.086000     3.399669    0.043000    3.399669   20.5796  1.1592
NONBonded  c1      0.086000     3.399669    0.043000    3.399669   20.5796  1.1592
NONBonded  c2      0.086000     3.399669    0.043000    3.399669   20.5796  1.1592
NONBonded  c3      0.109400     3.399669    0.054700    3.399669   20.5796  1.1592
NONBonded  c4      0.170000     3.249998   0.085000    3.249998   15.5986  1.1592
NONBonded  c5      0.170000     3.249998   0.085000    3.249998   15.5986  1.1592
NONBonded  c6      0.170000     3.249998   0.085000    3.249998   15.5986  1.1592

298
AEXP=4  REXP=6  HAEX=4  AAEX=2
! "all" possible combinations of HB-pairs in nucleic acids:
! WELL DEPTHS DEEPENED BY 0.5 KCAL TO IMPROVE BASEPAIR ENERGIES /LN
! AND DISTANCES INCREASED BY 0.05

hbond n* n* -3.5 3.05
hbond n* o* -4.0 2.95
hbond o* o* -4.75 2.80
hbond o* n* -4.50 2.90

set echo=true end

Amicetin Parameter file

REMARKS  *  AMBER Cornell et al. (parm94) force field conversion
REMARKS  *  nucleic acid part converted by tec3, march 1997
REMARKS  *  protein part converted by tec3, feb 1999
REMARKS  *
set echo=false end

! This is a hand-conversion of the Cornell et. al. force field
! parm94.dat file (except for the torsion which were converted via
! a perl script). All the atom type names are the same except where
! otherwise noted.
!
! Any problems? Let me know...
! cheatham@helix.nih.gov, cheatham@cgl.ucsf.edu
!
! note: types N* remaned to NS
!              C* renamed to CS
!
! note: the HO and HW nonbonded parameters, although zero in
! Cornell et al. should not have a 0.0 vdw radius in CHARMM
! to avoid difficulties with the 0.0/0.0 in the FAST OFF
! van der Waal code...
!
! note: the default nonbonded options chosen here are to match AMBER.
! It is not recommended that users actually run with GROUP based
! truncation and a switch unless really trying to match AMBER.
! Better would be ATOM based FSHIFT VSHIFT, however note that this
! does modify the energies (and effectively the parameters).
!
BOND  ow  hw  553.0  0.9572  TIP3P water
BOND  c  c  290.0  1.550  3
BOND  c  c1 379.8  1.460  0
BOND  c  c2 449.9  1.406  3
BOND  c  c3 328.3  1.508  1
BOND  c  ca 349.7  1.487  1
BOND  c  cc 377.4  1.462  3
BOND  c  cd 377.4  1.462  3
BOND  c  ce 363.8  1.474  1
BOND  c  ha 325.1  1.101  3
BOND  c  nc 438.8  1.371  3
BOND  c  nd 438.8  1.371  3
BOND  c  o 648.0  1.214  1
BOND  c  n 478.2  1.345  1
BOND  c3 c3 303.1  1.355  1
ANGLE  e  c3  n  66.7  111.56  ! 3
ANGLE  e  c3  n3  65.8  113.91  ! 3
ANGLE  e  c3  os  68.0  109.82  ! 3
ANGLE  c3  c3  e  63.2  110.63  ! 3
ANGLE  c3  c3  cc  64.7  108.10  ! 3
ANGLE  c3  c3  h1  46.4  110.07  ! 3
ANGLE  c3  c3  h2  46.0  111.59  ! 3
ANGLE  c3  c3  hc  46.4  110.05  ! 3
ANGLE  c3  c3  n  65.9  112.13  ! 3
ANGLE  c3  c3  nq  65.9  112.13  ! 3
ANGLE  c3  c3  n3  66.2  110.38  ! 3
ANGLE  c3  c3  nh  66.5  109.78  ! 3
ANGLE  c3  c3  oh  67.7  109.43  ! 3
ANGLE  c3  c3  os  67.8  108.42  ! 3
ANGLE  h1  c3  h1  39.2  109.55  ! 3
ANGLE  h1  c3  n  49.8  109.32  ! 3
ANGLE  h1  c3  nq  49.8  109.32  ! 3
ANGLE  h1  c3  n3  49.4  109.92  ! 3
ANGLE  h1  c3  nc  50.1  108.57  ! 3
ANGLE  h1  c3  nh  49.7  109.96  ! 3
ANGLE  h1  c3  o  52.0  117.19  ! 3
ANGLE  h1  c3  oh  51.0  109.88  ! 3
ANGLE  h1  c3  os  50.8  108.82  ! 3
ANGLE  h2  c3  os  50.8  108.58  ! 3
ANGLE  hc  c3  hc  39.4  108.35  ! 3
ANGLE  n  c3  n  68.8  113.81  ! 3
ANGLE  n  c3  os  71.2  109.19  ! 3
ANGLE  h2  c3  n  50.0  108.31  ! 3
ANGLE  os  c3  os  71.7  110.24  ! 3
ANGLE  e  ca  ca  64.6  120.14  ! 3
ANGLE  ca  ca  ca  67.2  119.97  ! 3
ANGLE  ca  ca  ha  48.5  120.01  ! 3
ANGLE  ca  ca  n  68.0  119.57  ! 3
ANGLE  ed  cc  h4  47.2  129.11  ! 3
ANGLE  cc  ed  ha  48.4  122.89  ! 3
ANGLE  cc  ed  n  70.7  115.52  ! 3
ANGLE  ed  cc  n  70.7  115.52  ! 3
ANGLE  cc  ed  ed  68.2  114.19  ! 3
ANGLE  ed  ca  ha  46.6  123.74  ! 3
ANGLE  h4  cc  n  50.4  117.62  ! 3
ANGLE  ec  ed  nh  72.4  120.11  ! 3
ANGLE  ed  cd  nh  68.6  118.98  ! 3
ANGLE  ed  ed  nc  71.1  112.56  ! 3
ANGLE  ec  cc  nd  71.1  112.56  ! 3
ANGLE  ed  ed  nd  70.0  113.42  ! 3
ANGLE  cc  cc  nc  70.0  113.42  ! 3
ANGLE  ed  ed  n  67.9  119.89  ! 3
ANGLE  ed  ed  nc  70.9  123.86  ! 3
ANGLE  c  n  c3  63.9  121.35  ! 3
ANGLE  c  n  ca  64.3  123.71  ! 3
ANGLE  c  n  cc  65.2  124.19  ! 3
ANGLE  c  n  ed  65.2  124.19  ! 3
ANGLE  c  n  hn  49.2  118.46  ! 3
ANGLE  c  n  o  71.6  118.90  ! 3
ANGLE  c  n  oh  69.5  113.39  ! 3
ANGLE  c  n  os  69.1  113.14  ! 3
ANGLE  c  n  ca  64.3  123.71  ! 3
ANGLE  c  n  c3  64.0  112.62  ! 3
ANGLE  c  n  cc  63.3  121.17  ! 3
ANGLE  c  n  ed  63.3  121.17  ! 3
ANGLE  c  n  hn  46.0  116.78  ! 3
ANGLE  ca  n  hn  47.6  114.59  ! 3
ANGLE  hn  n  hn  39.7  117.85  ! 3
ANGLE  cd  n  hn  47.5  121.52  ! 3
ANGLE  c3  n3  c3  64.0  110.90  ! 3
ANGLE  c3  n3  hn  47.1  109.92  ! 3
ANGLE  hn  n3  hn  41.3  107.13  ! 3
ANGLE  c    nc  cd  67.0  119.25  ! 3
ANGLE  cc  nh  hn  48.9  117.16  ! 3
ANGLE  cd  nh  hn  48.9  117.16  ! 3
ANGLE  hn  nh  hn  40.1  114.43  ! 3
ANGLE  c3  oh  ho  47.1  108.16  ! 3
ANGLE  c3  os  c3  62.1  113.41  ! 3

DIHEDRAL  X  c  c  X  1.20000000  4  180.0  !
DIHEDRAL  X  c3  X  0.00000000  6  180.0  !

DIHEDRAL  X  c  ca  X  14.50000000  4  180.0  !

DIHEDRAL  X  c  cc  X  11.50000000  4  180.0  !
DIHEDRAL  X  c  cd  X  11.50000000  4  180.0  !
DIHEDRAL  X  c  n  X  10.00000000  4  180.0  !
DIHEDRAL  X  c  nq  X  10.00000000  4  180.0  !
DIHEDRAL  X  c  nc  X  8.50000000  2  180.0  !
DIHEDRAL  X  c3  c3  X  1.40000000  9  180.0  !
DIHEDRAL  X  c3  n  X  0.00000000  6  180.0  !
DIHEDRAL  X  c3  nq  X  0.00000000  6  180.0  !
DIHEDRAL  X  c3  c3  X  1.70000000  3  180.0  !
DIHEDRAL  X  c3  c3  X  2.00000000  3  180.0  !

302
### Nonbonded Interactions

<table>
<thead>
<tr>
<th>Type</th>
<th>Atom 1</th>
<th>Atom 2</th>
<th>Atom 3</th>
<th>Atom 4</th>
<th>Emin</th>
<th>sigma</th>
<th>Emin/2</th>
<th>sigma</th>
</tr>
</thead>
<tbody>
<tr>
<td>NONBonded</td>
<td>c3</td>
<td>c3</td>
<td>n3</td>
<td>c3</td>
<td>0.30000000</td>
<td>1</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>NONBonded</td>
<td>c3</td>
<td>c3</td>
<td>n3</td>
<td>c3</td>
<td>0.40000000</td>
<td>1</td>
<td>180.0</td>
<td>1.5</td>
</tr>
<tr>
<td>NONBonded</td>
<td>c3</td>
<td>os</td>
<td>c3</td>
<td>os</td>
<td>0.38300000</td>
<td>1</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>NONBonded</td>
<td>c3</td>
<td>os</td>
<td>c3</td>
<td>os</td>
<td>0.10000000</td>
<td>1</td>
<td>180.0</td>
<td>1.5</td>
</tr>
</tbody>
</table>

### Improper Interactions

<table>
<thead>
<tr>
<th>Type</th>
<th>Atom 1</th>
<th>Atom 2</th>
<th>Atom 3</th>
<th>Atom 4</th>
<th>Elin</th>
<th>sigma</th>
<th>Elin/2</th>
<th>sigma</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPROPER</td>
<td>X</td>
<td>c</td>
<td>o</td>
<td>10.50000000</td>
<td>2</td>
<td>180.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMPROPER</td>
<td>X</td>
<td>n</td>
<td>hn</td>
<td>1.10000000</td>
<td>2</td>
<td>180.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMPROPER</td>
<td>c</td>
<td>n</td>
<td>hn</td>
<td>1.10000000</td>
<td>2</td>
<td>180.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMPROPER</td>
<td>X</td>
<td>ca</td>
<td>ha</td>
<td>1.10000000</td>
<td>2</td>
<td>180.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Other Important Parameters

- nbonds = 303
- cutnb = 9.0
ctonnb = 7.5
ctofnb = 8.0
- nbond = 5
- vswitch = 1.0

The program parameters include a tolerance of 0.5 and a cut-off distance of 9.0. The nonbonding parameter section is followed by a list of improper and dihedral angles with associated energy values.
Antibiotic randomisation script

parameter @gaff_amec_20090218.dna end
topology @TOPPAR:topamilatestbcongaffD.dna
COOR @AME6.pdb
segment
name=amec
mole
name=AME end end
COOR @AME6.pdb
write structure output=ame6.psf end
stop
xplor
remarks nmr/random.inp
remarks The ultimate simulated annealing protocol for NMR structure
remarks determination!
remarks The starting structure for this protocol can be completely
remarks arbitrary, such as random numbers. Note: the resulting
remarks structures need to be further processed by the dgsa.inp protocol.
remarks Author: Michael Nilges

evaluate ($init_t = 1000 )  (* Initial simulated annealing temperature.*)
parameter  (*Read the parameter file.*)
@gaff_amec_20090218.dna end
structure @ame6.psf end  (*Read the structure file.*)
noe
nres=3000  (*Estimate greater than the actual number of NOEs.*)
class all
@ame112jd.tbl end  (*Read NOE distance ranges.*)
restraints dihe
nassign=50
@AMETORj7.tbl end  (*Read dihedral angle restraints.*)
noe
ceiling=1000  (*Parameters for NOE effective energy term.*)
averaging * cent
potential * soft
scale * 1
sqoffset * 0.0
sqconstant * 1.0
sqexponent * 2
soexponent * 1
asymptote * 2.0
rswitch * 1.0

set message=off end

evaluate ($end_count=400)
{* Number of structures. *}
evaluate ($count = 0)
while ($count < $end_count ) loop main
evaluate ($count=$count+1)

{* Generate a starting structure. *}
vector do (x = (random()-0.5)*20) (all)
vector do (y = (random()-0.5)*20) (all)
vector do (z = (random()-0.5)*20) (all)
vector do (fbeta=10) (all)      {*Friction coefficient for MD heatbath.*}
vector do (mass=100) (all)              {*Uniform heavy masses to speed*}
{molecular dynamics.          *}

parameter nbonds
atom cutnb 100 tolerance 45 repel=1.2
rexp=2 rexp=2 rcon=1.0 nbxmod 4
end end

flags exclude * include plan bonds angle impr vdw noe cdih harm end

evaluate ($knoe = 0.5)
evaluate ($kbond = 0.00005 )  {* Bonds.   *}
evaluate ($kang = 0.00005 )   {* Angles.  *}
evaluate ($kimp = 0.0 )          {* Impropers.   *}
evaluate ($kvdw = 0.1)           {* Vdw.     *}

constraints
interaction (not name ca) (all)
weights bond $kbond angl $kang impr $kimp vdw 0 elec 0 end
interaction (name ca) (name ca)
weights bond $kbond angl $kang impr $kimp vdw $kvdw end

{* ======================================== High temperature dynamics. *}
vector do (vx = maxwell($init_t)) (all)
vector do (vy = maxwell($init_t)) (all)
vector do (vz = maxwell($init_t)) (all)
evaluate ($timestep = 0.04)
evaluate ($nstep = 100)
while ($kbond < 0.01) loop stage1

    evaluate (Skbon = min(0.25, Skbon * 1.25))
evaluate (Skang = Skbon)
evaluate (Skimp = 0)

    noe scale * $knoe end
    restraints dihed scale 0. end

    constraints
    interaction (not name ca) (all)
    weights bond $kbond angl $kang impr $kimp vdw 0 elec 0 end
    interaction (name ca) (name ca)
    weights bond $kbond angl $kang impr $kimp vdw $kvdw end

    dynamics verlet
    nstep=$nstep timestep=$timestep iasvel=current
tcoupling=true tbath=$init_t nprint=50 iprfiq=0
end loop stage 1

restraints dihed scale 0. end

noe scale * 5 end

parameter

\{ Parameter for the repulsive energy term. \}

nbonds

repel=0.9 \{ Initial value for repel - modified later. \}

nbxmod=3 \{ Initial value for nbxmod - modified later. \}

wmin=0.01
cutnb=4.5 ctonnb=2.99 ctofnb=3.
tolerance=0.5
end

end constraints

interaction (all) (all)

weights bond 0.02 angl 0.02 impr 0 vdw 0.002 elec 0 end

end

end

dynamics verlet

nstep=500 timestep=0.003 iasvel=maxwell

firstt=1500
tcoupling=true

tbath=1500 nprint=50 iprfrq=0
end

end

end

end

end

dynamics verlet

nstep=500 timestep=0.003 iasvel=current tcoupling=true

tbath=1500 nprint=50 iprfrq=0
end

end

end

dynamics verlet

nstep=500 timestep=0.003 iasvel=current tcoupling=true

tbath=1500 nprint=50 iprfrq=0
end

end

end

end

end

end

end

end loop main

stop

---

{\* Write out the final structure(s). \*

remarks

overall,bonds,angles,improper,vdw,noe,cdih

remarks energies: $ener, $bond, $angl, $impr, $vdw, $noe, $cdih

\{ Name(s) of the family of final structures. \}

evaluate ($filename=random_=encode($count)+.pdb)

write coordinates output =$filename end

end loop main

stop
Antibiotic simulated annealing script

xplor

evaluate ($init_t = 1500)
evaluate ($high_steps = 800000)
evaluate ($cool_steps = 3000000)

parameter @gaff_amec_20090218.dna end
structure @ame6.psf end

set echo=false end
set message=off end
set echo=off end

flags include plan end
set message=on end
set echo=on end

noe
nres=600
class=all
@ame112jd.tbl end

restraints dihedral
nassign=50
@AMETORj7.tbl end

vector do (fbeta=10) (all)
vector do (mass=100) (all)

noe
celling=100
averaging * cent
potential * square
sqconstant * }
sglexponent * 2
scale * 60.10
end

parameter
nbonds
repel=0.5
rexp=2 irexp=2 rcon=1.
nbxmod=-2
wmin=0.01
cutnb=4.5 ctonnb=2.99 ctofnb=3.
tolerance=0.5
end

restraints dihedral
scale=510
end

evaluate ($end_count=400)
evaluate ($count=0)
while ($count < $end_count ) loop main

evaluate ($count=$count+1)
evaluate ($filename="random_"+encode($count)+".pdb")
for $image in ( 1 -1 ) loop imag
coor initialize end
coor @i@$filename
vector do (x=x * $image) ( known )
vector identity (store1) (not known) {*Set store1 to unknowns.*}

{* ------------------------------------------------------ Minimization of bonds, VDWs, and NOEs.*}
restraints dihedral scale=5. end
parameter nbonds nbxmod=-2 repel=0.5 end end
flags exclude * include bond vdw noe cdih end
constraints interaction (all) (all) weights * 1. vdw 20. end end
minimize powell nstep=100 nprint=10 end

{% "================================== Include angles. " %}
flags include angle end
minimize powell nstep=1000 nprint=10 end
flags include impr plan end
evaluate ($nstep1 = int($high_steps/8))
evaluate ($nstep2 = int($high_steps/2))

constraints interaction (all) (all) weights * 0.1 impr 0.05 vdw 20. end end
dynamics verlet
 nstep=$nstep1 time=0.001 iasvel=maxwell firstt=$init_t
tcoup=true tbath=$init_t nprint=100 iprfrq=0
end
constraints interaction (all) (all) weights * 0.2 impr 0.1 vdw 20. end end
dynamics verlet
 nstep=$nstep1 time=0.001 iasvel=maxwell firstt=$init_t
tcoup=true tbath=$init_t nprint=100 iprfrq=0
end

parameter nbonds repel=0.9 end end
constraints interaction (all) (all) weights * 0.2 impr 0.2 vdw 0.01 end end
dynamics verlet
 nstep=$nstep1 time=0.001 iasvel=maxwell firstt=$init_t
tcoup=true tbath=$init_t nprint=100 iprfrq=0
end

parameter nbonds nbxmod=-3 end end
constraints interaction (all) (all) weights * 0.4 impr 0.4 vdw 0.003 end end
dynamics verlet
 nstep=$nstep2 time=0.001 iasvel=maxwell firstt=$init_t
tcoup=true tbath=$init_t nprint=100 iprfrq=0
end

constraints interaction (all) (all) weights * 1.0 impr 1.0 vdw 0.003 end end
dynamics verlet
 nstep=$nstep1 time=0.001 iasvel=maxwell firstt=$init_t
tcoup=true tbath=$init_t nprint=100 iprfrq=0
end

if ($image = 1) then
  vector do (store7=x) ( all ) {*Store first image in stores.*}
  vector do (store8=y) ( all )
  vector do (store9=z) ( all )
  vector do (store4=vx) ( all )
  vector do (store5=vy) ( all )
  vector do (store6=vz) ( all )
end if
end loop imag

energy end
evaluate ($e_minus=$ener)
coor copy end
vector do (x=store7) ( all )
vector do (y=store8) ( all )
vector do (z=store9) ( all )
energy end

{% "Uncomment the following lines if a test for the correct*%
 {% " enantiomer is desired based on an rms difference from a*%
 {% " reference structure. " %}
 !coor disp=comp @AME4.pdb {*Read reference structure.*}
 !coor fit sele=( known ) end
 !coor rms sele=( known ) end
 !evaluate (sold_rms=$result)
 !vector do (x=-x) ( known )
restraints dihedral scale=200. end
evaluate ($final_t = 300) { K }  
evaluate ($tempstep = 50) { K }  
evaluate ($nstep = int($cool_steps/$ncycle))  
evaluate ($ini_rad = 0.9)  
evaluate ($fin_rad = 0.75)  
evaluate ($ini_con= 0.003)  
evaluate ($fin_con= 4.0)  
evaluate ($bath = $init_t)  
evaluate ($k_vdw = $ini_con)  
evaluate ($k_vdwfact = ($fin_con/$ini_con)^(1/$ncycle))  
evaluate ($radius= $ini_rad)  
evaluate ($radfact = ($fin_rad/$ini_rad)^(1/$ncycle))

evaluate ($i_cool = 0)
while ($i_cool < $ncycle) loop cool
  evaluate ($i_cool=$i_cool+1)
evaluate ($bath = $bath - $tempstep) 
evaluate ($k_vdw=min($fin_con,$k_vdw*$k_vdwfact))  
evaluate ($radius=max($fin_rad,$radius*$radfact))  
parameter nbonds repel=$radius end end
constraints interaction (all) (all) weights * 1. vdw $k_vdw end end
dynamics verlet
  nstep=$nstep timestep=0.0005 iasvel=current firstt=$bath 
tcoup=true tbath=$bath nprint=$nstep iprfrq=0
end

evaluate ($critical=$temp/$bath)  
if ($critical > 10. ) then  
display ****& & rerun job with smaller timestep (i.e., 0.003)  
stop  
end if
end loop cool
minimize powell nstep= 100000 nprint=25 end

print threshold=0.1 noe
evaluate ($rms_noe=$result)  
evaluate ($violations_noe=$violations)  
print threshold=5. cdih
evaluate ($rms_cdih=$result)  
evaluate ($violations_cdih=$violations)  
print thres=0.05 bonds
evaluate ($rms_bonds=$result)  
print thres=5. angles
evaluate ($rms_angles=$result)  
print thres=5. impropers
evaluate ($rms_impropers=$result)  
remarks =------------------------------------------------------------------------------------------------------------------
remarks overall,bonds,angles,improper,vdw,noe,cdih  
remarks energies: Sener, $bond, $angl, $impr, $vdw, $noe, $cdih  
remarks =------------------------------------------------------------------------------------------------------------------
remarks bonds,angles,improper,noe,cdih  
remarks rms-d: $rms_bonds,$rms_angles,$rms_impropers,$rms_noe,$rms_cdih  
remarks =------------------------------------------------------------------------------------------------------------------
remarks noe, cdih  
remarks violations.: $violations_noe, $violations_cdih  
remarks =------------------------------------------------------------------------------------------------------------------
evaluate ($filename="amecdgsa_"+encode($count)+".pdb")
write coordinates output =$filename end
end loop main
stop
Antibiotic refinement script

xplor

structure @ame6.psf end
parameter @gaff_amec_20090218.dna end

set echo=false end
set message=off end
set echo=off end
flags include plan end
set message=on end
set echo=on end
evaluate ($kcdi = 5.0)
noe
nrestraints=1000
class=all
@ame112jd.tbl
scale * 150
ceiling 1000
potential * squarewell
sqconstant * 1.0
average * center
end
flags
include cdih noe
end
restraints dihe
nassign=50
@AMETORj7.tbl
scale $kcdi
end
evaluate ($end_count=400)
evaluate ($count=0)
while ($count < $end_count ) loop main
evaluate ($count=$count+1)
evaluate ($filename="amecdgsa_"+encode($count)+".pdb")
coordinates @@$filename
flags include bonds vdw noe cdih end
vector do (fbeta=10) (all)
vector do (mass=100) (all)
noe
celling=1000
averaging * cent
potential * square
scale * 50
sqoffset * 0.0
sqconstant * 2
soexponent * 1
end
parameters
nbonds
wmin=0.01
cutnb=4.5 ctonnb=2.5 etofnb=3.5
toler 0.5
repel=0.8
rexp=2
irex=2
tcon=1
end
end

310
restraints dihedral
scale=5
end

constraints interaction (all) (all) weights
bonds 1.0
impropers 1.0
angles 1.0
end

minimize powell nstep=500 nprint=100 end

constraints interaction (all) (all) weights
bonds 1.0
impropers 1.0
angles 1.0
vdw 4
end

vector do (fbeta = 10.0) (all)
end

vector do (mass=100) (all)
end

dynamics verlet
nstep=15000 timestep=0.0005 iasvel=current firstt=1000
tcoup=true tbath=1000 nprint=250 iprfrq=100
end

evaluate ($kcdi = min(50.1, $kcdi + 5.0))
restraints dihedral
scale=$kcdi
end

end loop stg1
evaluate ($filename="amecrefb_"+encode($count)+".pdb")
write coordinates output =$filename end

end loop main
stop

xplor

structure @ame6.psf end
parameter @gaff_amec_20090218.dna end

evaluate ($kcdi = 5.0)
set echo=false end
set message=off end
set echo=off end
restraints plane
@plane.tbl
end
flags include plan end
set message=on end
set echo=on end

noe
nrestraints=1000
class=all @ame112jd.tbl
scale * 75
ceiling 1000
potential * squarewell
sqconstant * 1.0
average * center
end
flags
include cdih noe
end

restraints dihe
nassign=50
@AMETORJ7.tbl
scale $kcdi
end

evaluate ($end_count=400)
evaluate ($count=0)
while ($count < $end_count ) loop main
   evaluate ($count=$count+1)
evaluate ($filename="amecrefb_"+encode($count)+".pdb")
   coordinates @@$filename
   flags include bonds vdw noe cdih end
   vector do (fbeta=10) (all)
   vector do (mass=100) (all)

   noe
   ceiling=1000
   averaging * cent
   potential * square
   scale b 50
   scale h 200
   sqoffset * 0.0
   sqconstant * 2
   soexponent * 1
end

parameters
nbonds
wmin=0.01
cutnb=4.5 ctomnb=2.5 ctofnb=3.5
toler 0.5
repsf=0.8
irex=2
rcon=1
end
end

restraints dihedral
scale=5
end

constraints interaction (all) (all) weights
bonds 1.0
improper 1.0
angles 1.0
end
end

minimize powell nstep=500 nprint=100 end

constraints interaction (all) (all) weights
bonds 1.0
improper 1.0
angles 1.0
vdw 4
end
end

vector do (fbeta = 10.0) (all)
vector do (mass=100) (all)

dynamics verlet
nstep=10 timestep=0.001 iasvel=maxwell firstt=1000
tcoup=true tbath=1000 nprint=10 iprfrq=100
ntrf = 999999999
end
evaluate ($kcdi=5)
while ($kcdi<50.0) loop stg1
  dynamics verlet
  nstep=15000 timestep=0.0005 iasvel=current firstt=1000
tcoup=true tbath=1000 nprint=250 iprfrq=100
end

evaluate ($kcdi = \min(50.1, kcdi + 5.0))
restraints dihedral
scale=$kcdi
end

end loop stg1
evaluate ($sbath = 1000.1)
while ($sbath gt 300) loop stg3
  dynamics verlet
  nstep=5000 time=0.0005 iasvel=current
tcoup=true tbath=$sbath nprint=500 iprfrq=100
  ntrfr = 0
end
evaluate ($sbath = $sbath - 25)
end loop stg3
constraints interaction (all) (all) weights
bonds 2.0
impropers 2.0
angles 2.0
vdw 4
end
end

minimize powell nstep=10000 drop=10 nprint=500 end

evaluate ($filename="amecrefa_"+encode($count)+".pdb")
write coordinates output =$filename end

end loop main
stop
xplor
structure @ame6.psf end
parameter @gaff_amec_20090218.dna end

set echo=false end
set message=off end
set echo=off end
restraints plane
@plane.tbl
?
end
flags include plan end
set message=on end
set echo=on end
evaluate ($knoe = 75.0)
evaluate ($kcdi = 50.0)
noe
nrestraints=1000
class=all
@ame112jd.tbl
scale * $knoe
ceiling 1000
potential * squarewell
sqconstant * 1.0
sqexponent * 2
average * center
end
flags
include cdih noe end
restraints dihe
nassign=50
@@AMETORj7.tbl
scale $kcdi
end

evaluate ($rcon=100)
parameters
nbonds
wmin=1.4
cutnb=8.5
toler 0.3
repel=0.0
rexp=2
irex=2
rcon=$rcon
end
end

constraints interaction (all) (all)
weights
angles 2.0
impropers 2.0
bonds 2.0
end
end

parameter nbonds nbxmod=5 end end
evaluate ($end_count=400)
evaluate ($count=0)
while ($count < $end_count ) loop main
evaluate ($count=$count+1)
evaluate ($filename="amecrefa_"+encode($count)+".pdb")
coordinates @@$filename

minimize powell nstep=100000 drop=10 nprint=500 end

print threshold=0.1 noe
evaluate ($rms_noe=$result)
print threshold=5. cdih
evaluate ($rms_cdih=$result)
print threshold=0.05 bonds
evaluate ($rms_bonds=$result)
print threshold=5. angles
evaluate ($rms_angles=$result)
print threshold=5. impropers
evaluate ($rms_impropers=$result)
remarks
remarks energies: $ener, $bond, $angl, $impr, $vdw, $noe, $cdih
remarks violations.: $violations_noe, $violations_cdih

evaluate ($filename="amecref_"+encode($count)+".pdb")
write coordinates output =$filename end
end loop main
stop
Antibiotic acceptance script

rem file nmr/accept.inp
rem Analysis of a family of NMR structures--
rem generation of a subfamily of "acceptable" structures

parameter {*Read the parameter file.*}

@end
@gaff_amec_20090218.dna
end

structure @ame6.psf end {*Read the structure file.*}

noe

nres=3000 {*Estimate greater than the actual number of NOEs.*}
class all
}@AME112jd.tbl {*Read NOE distance ranges.*}
end

restraints dihedral

nassign=110
}@AMETORj7.tbl {*Read dihedral angle restraints.*}
end

noe {*Parameters for NOE effective energy term.*}

ceiling=1000
 averaging * cent
 potential * square
 sqconstant * 1.0
 sqexponent * 2
 scale * 50.
end

parameter {*Parameters for the repulsive energy term.*}

nbonds
 repel=0.75
 rexp=2 irexp=2 rcon=4.
 nbxmod=3
 wmin=0.01
 cutnb=4.5 ctominb=2.99 ctofnb=3.
tolerance=0.5
end

restraints dihedral
 scale=200.
end

flags exclude * include bonds angle impr vdw noe cdih end

set precision=4 end

evaluate ($end_count=400) {*Loop through a family of 100 structures.*}
evaluate ($accept_count =0)
evaluate ($count =0)
while ($count < $end_count ) loop main
evaluate ($count=$count+1)

{=====>} {*Filename(s) for embedded coordinates.*}
evaluate ($filename="amecref_"+encode($count)+".pdb")
coor @@$filename
evaluate ($accept=0)

{*Print all NOE violations larger than 0.3 A*}
{*and compute RMS difference between observed*}
{*and model distances.*}
print threshold=0 noe
evaluate ($rms_noe=$result)
evaluate ($violations_noe=$violations)
if ($violations_noe > 0.2) then evaluate ( $accept=$accept + 1) end if

{*Print all dihedral angle restraint*}
{*violations.                          *}
print threshold=5. cdih
evaluate ($rms_cdih=$result)
evaluate ($violations_cdih=$violations)
if ($violations_cdih > 0) then evaluate ( $accept=$accept + 1) end if

print thres=0.01 bonds        {*Print deviations from ideal geometry.*}
evaluate ($rms_bonds=$result)
if ($result > 0.1) then evaluate ( $accept=$accept + 1) end if

print thres=3 angles
evaluate ($rms_angles=$result)
if ($result > 10) then evaluate ( $accept=$accept + 1) end if

print thres=5. impropers
evaluate ($rms_impropers=$result)

distance from=( not hydrogen ) to=( not hydrogen ) cutoff=1.5 end

{*Acceptance criteria: no NOE violations greater than 0.5 A,*}
{*no dihedral angle restraint violations > 5 deg.   *}
{*rms difference for bond deviations from ideality < 0.01 A,*}
{*rms difference for angle deviations from ideality < 2 deg.*}
energy end

if ($accept = 0 ) then
evaluate ( $accept_count=accept_count+1)
evaluate ($filename2="accept10_"+encode($accept_count)+".pdb")

remars ==================================================================
remars overall,bonds,angles,vdw,noe,cdih
remars energies: $ener, $bond, $angl, $vdw, $noe, $cdih, $impr
remars ==================================================================
remars bonds, angles, impropers, noe, cdih
remars rms-d: $rms_bonds,$rms_angles,$rms_impropers,$rms_noe,$rms_cdih
remars
remars noe, cdih
remars violations.: $violations_noe, $violations_cdih
remars $filename
write coordinates output=$filename2 end
end if

end loop main
stop

**Antibiotic average script**

remars file  nmr/average.inp
remars Computes the average structure, atomic rms differences from the
remars mean for a family of structures, and average overall rms
remars difference between the family and the mean structure.

{====>}
structure @ame6.psf end              {*Read the structure file.*}
{====>} "Backbone" selection--this example is typical for proteins.*}
vector idend ( store9 ) ( name ** or name * )

{------------------ The first stage consists of computing the mean structure.*}
{====>}
{Loop through the family of 8 accepted structures.*}
evaluate ($end_count=10)
eval ($nfile=0)
vector do (store1=0) (all)
vector do (store2=0) (all)
vector do (store3=0) (all)
vector do (store4=0) (all)
evaluate ($count = 0)
while ($count < $end_count ) loop main
  evaluate ($count=$count+1)
  {====> } {*This is the name of the family of structures.*)
evaluate ($filename="accept10_"+encode($count)+".pdb")
  coor @@$filename
  if ($count=1) then
    coor copy end {*Store first structure in comparison set.*)
  end if
  coor sele=( recall9 ) fit end
  {* swap equivalent groups to minimize rms difference *}
  @../geomanal/rotares.inp
  vector do (store1=store1+x) (all)
  vector do (store2=store2+y) (all)
  vector do (store3=store3+z) (all)
  vector do (store4=store4+x*x+y*y+z*z) (all)
eval ($nfile=$nfile+1)
end loop main

  vector do (x = store1 / $nfile) (all)
  vector do (y = store2 / $nfile) (all)
  vector do (z = store3 / $nfile) (all)
evaluate ($ave_rmsd_all=0.)
evaluate ($ave_rmsd_back=0.)
coor copy end

evaluate ($count = 0)
while ($count < $end_count ) loop main
  evaluate ($count=$count+1)
  {====> } {*This is the name of the family of structures.*)
evaluate ($filename="accept10_"+encode($count)+".pdb")
  coor @@$filename
  coor fit sele=( recall9 ) end
  coor rms sele=( recall9 ) end
  evaluate ($ave_rmsd_back=$ave_rmsd_back + $result)
  coor rms sele=( not hydrogen ) end
  evaluate ($ave_rmsd_all =$ave_rmsd_all + $result)
end loop main

evaluate ($ave_rmsd_back=$ave_rmsd_back / $nfile)
evaluate ($ave_rmsd_all =$ave_rmsd_all  / $nfile)
display ave. rms diff. to the mean struct. for non-h atoms= $ave_rmsd_all
display ave. rms diff. to the mean struct. for the backbone= $ave_rmsd_back

  {====== Finally, the average structure and RMSDs are written to a file.*)
  coor swap end
  vector do (b=bcomp) ( all )
  remarks unminimized average over $nfile files
  remarks ave. rms diff. to the mean struct. for non-h atoms= $ave_rmsd_all
  remarks ave. rms diff. to the mean struct. for the backbone= $ave_rmsd_back
  remarks b array (last column) is the rms difference from the mean
{*Write average coordinates and RMSDs to specified file.*}

```
write coordinates output=average_ame.pdb end

stop

xplor

structure @ame6.psf end
parameter @gaff_amec_20090218.dna end

set echo=false end
set message=off end
set echo=off end

restraints plane
@@plane.tbl
?
end

flags include plan end
set message=on end
set echo=on end
evaluate ($knoe = 50.0)
evaluate ($kcdi = 50.0)

noe
nrestraints=1000
class=all (@ame112jd.tbl
scale * $knoe
celling 1000
potential * squarewell
sqconstant * 1.0
sqexponent * 2
average * center
end

flags
include cdih noe
end

restraints dihe
nassign=50
@AMETORj7.tbl
scale $kcdi
end

evaluate ($rcon=100)
parameters
nbonds
wmin=1.4
cutnb=8.5
toler 0.3
repel=0.0
rexp=2
irex=2
rcon=$rcon
end
end

constraints interaction (all) (all)
weights
angles 10.0
impropers 2.0
bonds 2.0
noe 50
end
end

parameter nbonds nbxmod=5 end end
coordinates @average_ame.pdb

minimize powell nstep=100000 drop=10 nprint=500 end

print threshold=0.1 noe
evaluate ($rms_{\text{noe}}=$result)
evaluate ($violations_{\text{noe}}=$violations)
print threshold=5. cdih
evaluate ($rms_{\text{cdih}}=$result)
evaluate ($violations_{\text{cdih}}=$violations)
print thres=0.05 bonds
evaluate ($rms_{\text{bonds}}=$result)
print thres=5. angles
evaluate ($rms_{\text{angles}}=$result)
print thres=5. impropers
evaluate ($rms_{\text{impropers}}=$result)
remarks
remarks            overall,bonds,angles,improper,vdw,noe,cdih
remarks energies: $ener, $bond, $angl, $impr, $vdw, $noe, $cdih
remarks
remarks            bonds,angles,improper,vdw,noe,cdih
remarks rms-d: $rms_{\text{bonds}},$rms_{\text{angles}},$rms_{\text{impropers}},$rms_{\text{noe}},$rms_{\text{cdih}}
remarks
remarks            noe, cdih
remarks violations.: $violations_{\text{noe}},$violations_{\text{cdih}}
remarks
write coordinates output =average_em_ame_.pdb
del loop main
stop
A.3. RNA structure calculation scripts

RNA simulated annealing script

from pdbTool import PDBTool
from xplorPot import XplorPot
from rdcPotTools import create_RDCPot
from varTensorTools import create_VarTensor
import varTensorTools
from ivm import IVM
from potList import PotList
import protocol
from protocol import initMinimize
from ivm import IVM
from xplor import command
import random
from atomAction import SetProperty
from simulationTools import StructureLoop
from vec3 import Vec3
from psfGen import seqToPSF
from xplorPot import XplorPot
from varTensorTools import create_VarTensor
import varTensorTools
from ivm import IVM
from potList import PotList
import protocol
from avePot import AvePot
from simulationTools import StructureLoop, AnnealIVM, FinalParams
from simulationTools import AnnealIVM
from monteCarlo import randomizeTorsions

xplor.parseArguments()
# this checks for typos on the command-line. User-customized arguments can
# also be specified

command = xplor.command
from noePotTools import create_NOEPot
protocol.initParams("nucleic")
protocol.initTopology("nucleic")

# parameters to ramp up during the simulated annealing protocol
#
rampedParams=[]
highTempParams=[]

init_t = 3500.  # Need high temp and slow annealing to converge
final_t=25
bathTemp=2000

seqToPSF(open('Hh14mer.seq').read(), seqType='rna')
seqToPSF(open('Hh15mer.seq').read(), seqType='rna', startResid=21)

#command("write psf output=29mer.psf end")
for atom in AtomSel("all"):
    atom.setPos( Vec3(float(atom.index())/10,
    random.uniform(-0.5,0.5),
    random.uniform(-0.5,0.5)) )
pass

protocol.fixupCovalentGeom(useVDW=1,maxIters=100)
pots = PotList()

noex = create_NOEPot("noex",
            "Hh29merex250b.tbl")

320
noex.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultiRamp(0.2,30,"noex.setScale(VALUE)") )

noeno = create_NOEPot("noeno",
    "Hh29mernonovl2ac.tbl")
noeno.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultiRamp(2,30,"noeno.setScale(VALUE)") )

noeo = create_NOEPot("noeo",
    "Hh29merovl2c.tbl")
noeo.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultiRamp(2,30,"noeo.setScale(VALUE)") )

noevo = create_NOEPot("noevo",
    "Hh29mervovl2c.tbl")
noevo.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultiRamp(2,30,"noevo.setScale(VALUE)") )

hbon = create_NOEPot("hbon",
    "Hhhbon2.tbl")
hbon.setPotType("hard")
rampedParams.append( MultiRamp(0.2,30,"hbon.setScale(VALUE)") )

#hbs = create_NOEPot("hbs",
#    "hbonsoft2.tbl")
#hbs.setPotType("hard")
rampedParams.append( MultiRamp(0.2,30,"hbs.setScale(VALUE)") )

protocol.initDihedrals("Hhtor2.tbl",
    scale=5,          #initial force constant
    useDefaults=0)

xplor.command("@planeqe.tbl")  
## radius of gyration term
##
#protocol.initCollapse(Rtarget=10.16)
#rampedParams.append( XplorPot('COLL') )

pots.add( XplorPot('CDIH') )
pots['CDIH'].setThreshold( 5 )

xplor.command("@planeqe.tbl")

rampedParams.append( MultiRamp(0.4,1.0,"pots['ANGL'].setScale(VALUE)") )
rampedParams.append( MultiRamp(0.1,1.0,"pots['IMPR'].setScale(VALUE)") )

pots.add( XplorPot("ANGL") )
pots.add( XplorPot("IMPR") )

rampedParams.append( MultiRamp(0.9,0.78,
    "xplor.command('param nbonds repel VALUE end end')") )
rampedParams.append( MultiRamp(0.004,4,
    "xplor.command('param nbonds rcon VALUE end end')") )

pots.add(noex)
pots.add(noeno)
pots.add(noeo)
pots.add(noevo)
pots.add(hbon)

#pots.append(AvePot(XplorPot("plan",xplor.simulation)) )
# IVM setup
# the IVM is used for performing dynamics and minimization in torsion-angle space, and in Cartesian space.

from selectTools import IVM_groupRigidSidechain
from selectTools import IVM_breakRiboses

dyn = IVM()
protocol.initDynamics(dyn, potList=pots)
#IVM_groupRigidSidechain(dyn)
#IVM_breakRiboses(dyn, sel=0, breakSelStr="name O4' or name C1'")
protocol.torsionTopology(dyn)

minc = IVM()
protocol.initMinimize(minc, potList=pots)
IVM_groupRigidSidechain(minc)
#IVM_breakRiboses(minc, sel=0, breakSelStr="name O4' or name C1'")
protocol.cartesianTopology(minc, "not resname ANI")

# object which performs simulated annealing
#
from simulationTools import AnnealIVM
cool = AnnealIVM(initTemp =init_t,
    finalTemp=final_t,
    tempStep =12.5,
    ivm=dyn,
    rampedParams = rampedParams)

cart_cool = AnnealIVM(initTemp =init_t,
    finalTemp=25,
    tempStep =12.5,
    ivm=minc,
    rampedParams = rampedParams)

def calcOneStructure( structData):
    randomizeTorsions(dyn)
    # initialize parameters for high temp dynamics.
    InitialParams( rampedParams )
    # high-temp dynamics setup - only need to specify parameters which # differ from initial values in rampedParams
    InitialParams( highTempParams )
    # high temperature bit - using only P-P nonbonded terms
    protocol.initNBond(repel=1.2,
        cutnb=100,
        tolerance=45,
        selStr="name P")

    protocol.initDynamics(dyn,
        potList=pots, # potential terms to use
        bathTemp=init_t,
        initVelocities=1,
        numSteps=1000,
        printInterval=100)

    dyn.setETolerance( init_t/100 ) #used to det. stepsize. default: t/1000
dyn.run()

    protocol.initNBond() #reset to include all atoms
    # initialize parameters for cooling loop
    InitialParams( rampedParams )

    # initialize integrator for simulated annealing
    #
    protocol.initDynamics(dyn,
        potList=pots,
        numSteps=100, #at each temp: 100 steps or
        finalTime=2 , # .2ps, whichever is less
        printInterval=100)

    # perform simulated annealing
    #
cool.run()

    # final torsion angle minimization
    #
    protocol.initMinimize(dyn,
        printInterval=50)

dyn.run()
protocol.initDynamics(minc,
    potList=pots,
    numSteps=100,       #at each temp: 100 steps or
    finalTime=.4 ,       # .2ps, whichever is less
    printInterval=100)
cart_cool.run()
    # final all- atom minimization
    #
    protocol.initMinimize(minc,
        potList=pots,
        dEPred=10)

minc.run()
structData.writeStructure(pots)
simWorld.setRandomSeed( 785 )
outPDBFilename = 'SCRIPT_STRUCTURE.pdb'
StructureLoop(numStructures=100,
    pdbTemplate=outPDBFilename,
    structLoopAction=calcOneStructure,
    genViolationStats=1,
    averageTopFraction=0.3, #report stats on best 30% of structs
    averageContext=FinalParams(rampedParams),
    averageSortPots=[pots['BOND'],pots['ANGL'],pots['IMPR'],
        noeno,noeo, noevo, noevo,  noex.pots['CDIH'],
        averageFilename="SCRIPT_ave.pdb",    #generate regularized ave structure
        averageFitSel="name P",
        averagePotList=pots).run()

RNA refinement script

from pdbTool import PDBTool
from xplorPot import XplorPot
from rdcPotTools import create_RDCPot
from varTensorTools import create_VarTensor
import varTensorTools
from ivm import IVM
from potList import PotList
import protocol
from ivm import IVM
from xplor import command
import random
from atomAction import SetProperty
from simulationTools import StructureLoop
from vec3 import Vec3
from psfGen import seqToPSF
from xplorPot import XplorPot
from ivm import IVM
from potList import PotList
import protocol
from avePot import AvePot
from simulationTools import MultRamp, StaticRamp, InitialParams, StructureLoop, AnnealIVM
from simulationTools import AnnealIVM
xplor.parseArguments()
    # this checks for typos on the command-line. User-customized arguments can
    # also be specified.
    #
command = xplor.command
from noePotTools import create_NOEPot
protocol.initParams("nucleic")
protocol.initTopology("nucleic")

seed=56789
numberOfStructures=100
startStructure=0
outFilename = "SCRIPT_STRUCTURE.pdb"

rampedParams=[]

init_t=2000
final_t=25
bathTemp=2000

startFile="annealHh29mer_34.pdb"

simWorld.setRandomSeed(seed)

seqToPSF(open('Hh14mer.seq').read(), seqType='rna')
seqToPSF(open('Hh15mer.seq').read(), seqType='rna', startResid=21)

#command("write psf output=29mer.psf end")

# starting coords
#
protocol.initCoords(startFile)

protocol.covalentMinimize()

# list of potential terms used in refinement
pots = PotList()
crossTerms=PotList('cross terms') # can add some pot terms which are not
# refined against- but included in analysis

noex = create_NOEPot("noex", "Hh29merex250b.tbl")
noex.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultRamp(0.2,30.,"noex.setScale( VALUE )") )

noeno = create_NOEPot("noeno", "Hh29merovl2c.tbl")
noeno.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultRamp(2,30.,"noeno.setScale( VALUE )") )

noeo = create_NOEPot("noeo", "Hh29merovl2c.tbl")
noeo.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultRamp(2,30.,"noeo.setScale( VALUE )") )

noevo = create_NOEPot("noevo", "Hh29mervovl2c.tbl")
noevo.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultRamp(2,30.,"noevo.setScale( VALUE )") )

noevvo = create_NOEPot("noevvo", "Hh29mervvovl2c.tbl")
noevvo.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultRamp(2,30.,"noevvo.setScale( VALUE )") )

hbon = create_NOEPot("hbon", "Hhhbon2.tbl")
hbon.setPotType("hard")
rampedParams.append( MultRamp(0.2,30.,"hbon.setScale( VALUE )") )

hbs = create_NOEPot("hbs", "hbonsoft2.tbl")
hbs.setPotType("hard")
rampedParams.append( MultRamp(0.2,30.,"hbs.setScale( VALUE )") )

protocol.initDihedrals("Hhtor2.tbl", scale=5) #initial force constant
pots.append(AvePot(XplorPot,"cdih")
rampedParams.append( StaticRamp("pots['CDIH'].setScale(200)") )

324
protocol.initRamaDatabase("nucleic")
pots.append(AvePot(XplorPot("rama")))
rampedParams.append(MultRamp(1.1,"xplor.command('rama scale VALUE end')"))

#xplor.command('@planeqe.tbl')

pots.add(XplorPot("BOND"))
pots.add(XplorPot("DIHE"))
pots.add(XplorPot("ANGL"))
pots.add(XplorPot("IMPR"))
rampedParams.append(MultRamp(0.4,1.0,"pots['ANGL'].setScale(VALUE)"))
rampedParams.append(MultRamp(0.1,1.0,"pots['IMPR'].setScale(VALUE)"))

protocol.initNBond(cutnb=4.5)
pots.add(XplorPot("VDW"))
rampedParams.append(MultRamp(0.9,0.78,"xplor.command('param nbonds repel VALUE end end')"))
rampedParams.append(MultRamp(0.004,4,"xplor.command('param nbonds rcon VALUE end end')"))
pots.add(noex)
pots.add(noeno)
pots.add(neco)
pots.add(noevo)
pots.add(novo)
pots.add(hbon)

#pots.add(hbs)
#pots.append(AvePot(XplorPot("plan",xplor.simulation)))

mini = IVM()  # initial alignment of orientation tensor axes
from selectTools import IVM_groupRigidSidechain
from selectTools import IVM_breakRiboses

IVM_groupRigidSidechain(mini)
#IVM_breakRiboses(mini, sel=0, breakSelStr="name O4' or name C1'")
protocol.cartesianTopology(mini,"not resname ANI")

protocol.initMinimize(mini,
    numSteps=20)
mini.run("not resname ANI")

# this initial minimization is not strictly necessary

# uncomment to allow Da, Rh to vary
# for medium in ('bic1','phg1'): media[medium].setFreedom("varyDa, varyRh")
# for medium in ('bic2'),:
#    media[medium].setFreedom("varyDa, varyRh, fixAxisTo bic1")
# for medium in ('phg2','phg3',):
#    media[medium].setFreedom("varyDa, fixAxisTo phg1, fixRhTo phg1")

dyn = IVM()
protocol.initDynamics(dyn,potList=pots)

#protocol.cartesianTopology(dyn)

# uncomment to allow Da, Rh to vary
# for medium in ('bic1','phg1'): media[medium].setFreedom("varyDa, varyRh")
# for medium in ('bic2'),:
#    media[medium].setFreedom("varyDa, varyRh, fixAxisTo bic1")
# for medium in ('phg2','phg3',):
#    media[medium].setFreedom("varyDa, fixAxisTo phg1, fixRhTo phg1")

anneal= AnnealIVM(initTemp =init_t,
    finalTemp=25,
    tempStep =25,
    ivm=dyn,
    rampedParams = rampedParams)

# initialize parameters for initial minimization.
InitialParams( rampedParams )

# initial minimization
protocol.initMinimize(dyn,
    numSteps=1000)
dyn.run()

from simulationTools import testGradient
#testGradient(potList,eachTerm=1)

def calcOneStructure( structData ):
    # initialize parameters for high temp dynamics.
    InitialParams( rampedParams )

    # high temperature bit - using only P-P nonbonded terms
    protocol.initNBond(rep=1.2,
        cutnb=100,
        tolerance=45,
        selStr="name P")

    protocol.initDynamics(dyn,
        initVelocities=1,
        bathTemp=init_t,
        potList=pots,
        finalTime=10)
dyn.run()
    protocol.initNBond() #reset to include all atoms

    # perform simulated annealing
    #
    protocol.initDynamics(dyn,
        finalTime=0.2,  #time to integrate at a given temp.
        numSteps=0,     # take as many steps as necessary
        eTol_minimum=0.001 # cutoff for auto-TS det.
    )
    anneal.run()
    #
    # torsion angle minimization
    #
    protocol.initMinimize(dyn)
dyn.run()
    ##
    ##all atom minimization
    ##
    minc.run()

    structData.writeStructure(pots,crossTerms)

def accept(potList):
    return True if current structure meets acceptance criteria
    
    #if pots['noex'].violations()>2:
    #    return False
    #if pots['noeno'].violations()>2:
    #    return False
    #if pots['noenvo'].violations()>2:
    #    return False
    #if pots['noenno'].violations()>2:
    #    return False
    #if pots['noenA21'].violations()>2:
    #    return False
    #if pots['CDIH'].violations()>2:
    #    return False
    #if pots['BOND'].violations()>2:
    #    return False
    #if pots['ANGL'].violations()>0:
    #    return False
    #if pots['IMPR'].violations()>2:
    #    return False
from simulationTools import StructureLoop, FinalParams
StructureLoop(numStructures=numberOfStructures,
startStructure=startStructure,
pdbTemplate=outFilename,
structLoopAction=calcOneStructure,
genViolationStats=1,
averagePotList=pots,
averageSortPots=[pots['BOND'],pots['ANGL'],pots['IMPR'],noeno,noevo,noeo,noevvo,noex,pots['CDIH'],hbon],
averageTopFraction=0.25, #report only on best 25% of structs
#averageAccept=accept, #only use structures which pass accept()
averageContext=FinalParams(rampedParams),
averageFilename="SCRIPT_ave.pdb", #generate regularized ave structure
averageFitSel="name P", averageCompSel="not resname ANI and not name H"
).run()
{"The second stage consists of computing an overall rms difference.*/}
evaluate ($ave_rmsd_all=0.)
evaluate ($ave_rmsd_back=0.)
coor copy end
evaluate ($count = 0)
while ($count < $end_count ) loop main
  evaluate ($count=$count+1)
  {====>} {*This is the name of the family of structures.*}
evaluate ($filename="refineHh29mer_"+encode($count)+".pdb")
  coor @@$filename
  coor fit sele=( recall9 ) end
  coor rms selection=( recall9 )end
  evaluate ($ave_rmsd_back=$ave_rmsd_back + $result)
  coor rms selection=( not hydrogen )end
  evaluate ($ave_rmsd_all =$ave_rmsd_all + $result)
end loop main

evaluate ($ave_rmsd_back=$ave_rmsd_back / $nfile)
evaluate ($ave_rmsd_all =$ave_rmsd_all / $nfile)
display ave. rms diff. to the mean struct. for non-h atoms= $ave_rmsd_all
display ave. rms diff. to the mean struct. for the backbone= $ave_rmsd_back

{=====> Finally, the average structure and RMSDs are written to a file.*}
coor swap end
vector do (b=bcomp) ( all )
remarks unminimized average over $nfile files
remarks ave. rms diff. to the mean struct. for non-h atoms= $ave_rmsd_all
remarks ave. rms diff. to the mean struct. for the backbone= $ave_rmsd_back
remarks b array (last column) is the rms difference from the mean

{=====>} {*Write average coordinates and RMSDs to specified file.*}
write coordinates output=averageb.pdb end
stop

from pdbTool import PDBTool
from xplorPot import XplorPot
from rdcPotTools import create_RDCPot
from varTensorTools import create_VarTensor
import varTensorTools
from ivm import IVM
from potList import PotList
import protocol
from protocol import initMinimize
from xplor import command
import random
from atomAction import SetProperty
from simulationTools import StructureLoop
from vec3 import Vec3
from psfGen import seqToPSF
from xplorPot import XplorPot
from xplor import IVM
from potList import PotList
import protocol
from avePot import AvePot
from simulationTools import MultRamp, StaticRamp, InitialParams, StructureLoop, AnnealIVM
from simulationTools import AnnealIVM
xplor.parseArguments()
# this checks for typos on the command-line. User-customized arguments can
# also be specified.
#
command = xplor.command
from noePotTools import create_NOEPot

protocol.initParams("nucleic")
protocol.initTopology("nucleic")

seed=56789
numberOfStructures=1
startStructure=0
outFilename = "SCRIPT_STRUCTURE.pdb"
rampedParams=[]

init_t=2000
final_t=25
bathTemp=2000

startFile="averageb.pdb"

simWorld.setRandomSeed(seed)
seqToPSF(open('Hh14mer.seq').read(), seqType="rna")
seqToPSF(open('Hh15mer.seq').read(), seqType="rna", startResid=21)

#command("write psf output=29mer.psf end")
#
# starting coords
#
protocol.initCoords(startFile)
protocol.covalentMinimize()

# list of potential terms used in refinement
pots = PotList()
crossTerms=PotList('cross terms') # can add some pot terms which are not
# refined against- but included in analysis

noex = create_NOEPot("noex",
                     "Hh29merex250.tbl")
noex.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultRamp(0.2,30,"noex.setScale( VALUE y)"))

noeno = create_NOEPot("noeno",
                      "Hh29mernovl2ac.tbl")
noeno.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultRamp(2,30,"noeno.setScale( VALUE y)"))

noeo = create_NOEPot("noeo",
                     "Hh29merovl2c.tbl")
noeo.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultRamp(2,30,"noeo.setScale( VALUE y)"))

noevo = create_NOEPot("noevo",
                      "Hh29mervovl2c.tbl")
noevo.setPotType("hard") #if incorrect noes suspected set soft if not set hard
rampedParams.append( MultRamp(2,30,"noevo.setScale( VALUE y)"))

hbon = create_NOEPot("hbon",
                     "Hhhbon2.tbl")
hbon.setPotType("hard")
rampedParams.append( MultRamp(0.2,30,"hbon.setScale( VALUE y)"))

protocol.initDihedrals("Hhtor2.tbl",
                        scale=5) #initial force constant
pots.append(AvePot(XplorPot,"cdih")
rampedParams.append( StaticRamp("pots['CDIH'].setScale(200)"))
protocol.initRamaDatabase("nucleic")
pots.append(AvePot(XplorPot,"rama")
)  
rampedParams.append(  
    MultiRamp(1,1,"xplor.command('rama scale VALUE end')")
)  
pots.add( XplorPot("BOND") )  
pots.add( XplorPot("DIHE") )  
pots.add( XplorPot("ANGL") )  
pots.add( XplorPot("IMPR") )

rampedParams.append(  
    MultiRamp(0.4,1.0,"pots['ANGL'].setScale(VALUE)")
)  
rampedParams.append(  
    MultiRamp(0.1,1.0,"pots['IMPR'].setScale(VALUE)")
)  

protocol.initNBond(cutnb=4.5)

pots.add( XplorPot("VDW") )

rampedParams.append(  
    MultiRamp(0.9,0.78,
        "xplor.command('param nbonds repel VALUE end end')")
)  
rampedParams.append(  
    MultiRamp(.004,4,
        "xplor.command('param nbonds rcon VALUE end end')")
)  

pots.add(noex)
pots.add(noeno)
pots.add(noeo)
pots.add(noevo)
pots.add(noevvo)
pots.add(hbon)

mini = IVM()  #initial alignment of orientation tensor axes
from selectTools import IVM_groupRigidSidechain
from selectTools import IVM_breakRiboses

dyn = IVM()
protocol.initDynamics(dyn,potList=pots)
IVM_groupRigidSidechain(dyn)
#IVM_breakRiboses(dyn, sel=0, breakSelStr="name O4' or name C1'"
#protocol.cartesianTopology(dyn,"not remame ANI")
protocol.torsionTopology(dyn)

from selectTools import IVM_groupRigidSidechain
minc = IVM()
protocol.initMinimize(minc,potList=pots,numSteps=2000)
#IVM_groupRigidSidechain(minc)
#IVM_breakRiboses(minc, sel=0, breakSelStr="name O4' or name C1'"
protocol.cartesianTopology(minc,"not remame ANI")

InitialParams( rampedParams )
minc.run()

command("write coordinates output=averagelim.pdb end")
from simulationTools import StructureLoop, FinalParams
StructureLoop(numStructures=numberOfStructures,
    startStructure=startStructure,
    pdbTemplate=outFilename,
    #structLoopAction=calcOneStructure,
    genViolationStats=1,
    averagePotList=pots,
    averageSortPots=[pots['BOND'],pots['ANGL'],pots['IMPR'],
        ],
    averageTopFraction=0.25, #report only on best 25% of structs
    #averageAccept=accept,  #only use structures which pass accept()
    #averageContext=FinalParams(rampedParams),
    #averageFilename="SCRIPT_ave.pdb",    #generate regularized ave structure
    averageFitSel="name P",
    averageCompSel="not remame ANI and not name H**
).run()
Appendix B: Experimental lists

B.1. *E. coli* titration NMR experiments

<table>
<thead>
<tr>
<th>Equivalents of amicetin</th>
<th>Experiment</th>
<th>Spectrometer Frequency (MHz)</th>
<th>Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>1D $^1$H</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.000</td>
<td>1D $^3$P</td>
<td>81</td>
<td>27</td>
</tr>
<tr>
<td>0.000</td>
<td>TOCSY (75ms)</td>
<td>700</td>
<td>25</td>
</tr>
<tr>
<td>0.000</td>
<td>NOESY (250ms)</td>
<td>700</td>
<td>25</td>
</tr>
<tr>
<td>0.000</td>
<td>NOESY (250ms)*</td>
<td>600</td>
<td>2</td>
</tr>
<tr>
<td>0.000</td>
<td>1D $^1$H*</td>
<td>600</td>
<td>2</td>
</tr>
<tr>
<td>0.033</td>
<td>1D $^1$H</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.065</td>
<td>1D $^1$H</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.130</td>
<td>1D $^1$H</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.130</td>
<td>TOCSY (75ms)</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.130</td>
<td>NOESY (250ms)</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.195</td>
<td>1D $^1$H</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.260</td>
<td>1D $^1$H</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.260</td>
<td>TOCSY (75ms)</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.390</td>
<td>1D $^1$H</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.390</td>
<td>TOCSY (75ms)</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.390</td>
<td>NOESY (150ms)</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.390</td>
<td>NOESY (250ms)</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.780</td>
<td>1D $^1$H</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>0.780</td>
<td>TOCSY (75ms)</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>1.170</td>
<td>1D $^1$H</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>1.170</td>
<td>TOCSY (75ms)</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>1.170</td>
<td>NOESY (250ms)</td>
<td>600</td>
<td>25</td>
</tr>
<tr>
<td>1.240</td>
<td>1D $^1$H</td>
<td>400</td>
<td>25</td>
</tr>
<tr>
<td>1.240</td>
<td>1D $^1$H</td>
<td>400</td>
<td>5</td>
</tr>
<tr>
<td>1.240</td>
<td>1D $^3$P</td>
<td>81</td>
<td>27</td>
</tr>
<tr>
<td>1.630</td>
<td>1D $^1$H</td>
<td>400</td>
<td>25</td>
</tr>
<tr>
<td>1.63</td>
<td>1D $^1$H</td>
<td>800</td>
<td>5</td>
</tr>
<tr>
<td>3.6</td>
<td>1D $^1$H</td>
<td>800</td>
<td>25</td>
</tr>
<tr>
<td>3.6</td>
<td>1D $^1$H</td>
<td>800</td>
<td>5</td>
</tr>
<tr>
<td>3.6</td>
<td>1D $^3$P*</td>
<td>81</td>
<td>27</td>
</tr>
<tr>
<td>3.6</td>
<td>NOESY (100ms)</td>
<td>800</td>
<td>25</td>
</tr>
<tr>
<td>3.6</td>
<td>NOESY (250ms)</td>
<td>800</td>
<td>25</td>
</tr>
<tr>
<td>3.6</td>
<td>NOESY (150ms)</td>
<td>800</td>
<td>5</td>
</tr>
<tr>
<td>3.6</td>
<td>TOCSY (75ms)</td>
<td>800</td>
<td>25</td>
</tr>
<tr>
<td>3.6</td>
<td>1D $^1$H*</td>
<td>1000</td>
<td>2</td>
</tr>
<tr>
<td>3.6</td>
<td>NOESY (250ms)*</td>
<td>1000</td>
<td>2</td>
</tr>
</tbody>
</table>

Table B.1.1: Indicates the NMR spectra carried out and at what temperature and at what equivalents of amicetin to the *E. coli* 29-mer RNA in 100% $^2$H$_2$O. An “*” indicates that this experiment was carried out on the same sample but in 90% $^1$H$_2$O and 10% $^2$H$_2$O.
Appendix C: Additional experimental results

C.1. Additional helical parameters and dihedral angles for the *E. coli* 29-mer NMR solution structure

The local step parameters are shown in Table C.1.1 and the local base pair parameters are shown in Table C.1.2. The ribose dihedral angles of the base paired residues, their sugar pucker, the amplitude of pseudorotation of the sugar ring and the phase angle of pseudorotation of the sugar ring are shown in Table C.1.3 and Table C.1.4. Table C.1.5 and Table C.1.6 show the backbone and glycosidic torsion angles of all base paired residues.

<table>
<thead>
<tr>
<th>Step</th>
<th>X-displacement</th>
<th>Y-displacement</th>
<th>Inclination</th>
<th>Tip</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-C2/G34-G35</td>
<td>-3.36</td>
<td>0.50</td>
<td>10.29</td>
<td>0.32</td>
</tr>
<tr>
<td>C2-C3/G33-G34</td>
<td>-5.87</td>
<td>-1.30</td>
<td>17.71</td>
<td>3.33</td>
</tr>
<tr>
<td>C3-C4/G32-G33</td>
<td>-6.94</td>
<td>-1.76</td>
<td>25.99</td>
<td>10.73</td>
</tr>
<tr>
<td>C4-G5/C31-G32</td>
<td>-6.70</td>
<td>3.59</td>
<td>29.02</td>
<td>-18.43</td>
</tr>
<tr>
<td>A8-A9/U26-U29</td>
<td>-0.77</td>
<td>0.49</td>
<td>12.77</td>
<td>18.85</td>
</tr>
<tr>
<td>A9-C10/G25-U26</td>
<td>-6.53</td>
<td>-0.18</td>
<td>21.87</td>
<td>-10.15</td>
</tr>
<tr>
<td>C10-C11/G24-G25</td>
<td>-4.02</td>
<td>0.48</td>
<td>16.12</td>
<td>9.30</td>
</tr>
<tr>
<td>C11-U12/A23-G24</td>
<td>-3.98</td>
<td>-2.06</td>
<td>11.49</td>
<td>6.30</td>
</tr>
<tr>
<td>U12-U13/A22-A23</td>
<td>-5.30</td>
<td>0.65</td>
<td>17.41</td>
<td>0.11</td>
</tr>
<tr>
<td>U13-U14/A21-A22</td>
<td>-3.10</td>
<td>5.18</td>
<td>12.89</td>
<td>-34.15</td>
</tr>
</tbody>
</table>

Table C.1.1: Shows the local base pair step helical parameters of the *E. coli* 29-mer NMR solution structure, “~” indicates that the information could not be given.

<table>
<thead>
<tr>
<th>Base pair</th>
<th>Shear</th>
<th>Stretch</th>
<th>Stagger</th>
<th>Buckle</th>
<th>Propeller twist</th>
<th>Opening</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-G35</td>
<td>0.02</td>
<td>-0.18</td>
<td>-0.72</td>
<td>3.29</td>
<td>12.28</td>
<td>-13.70</td>
</tr>
<tr>
<td>C2-G34</td>
<td>0.17</td>
<td>-0.43</td>
<td>-0.81</td>
<td>-0.95</td>
<td>7.66</td>
<td>-11.66</td>
</tr>
<tr>
<td>C3-G33</td>
<td>0.02</td>
<td>-0.35</td>
<td>-0.53</td>
<td>-2.72</td>
<td>15.21</td>
<td>-8.14</td>
</tr>
<tr>
<td>C4-G32</td>
<td>0.30</td>
<td>-0.39</td>
<td>0.68</td>
<td>0.47</td>
<td>13.71</td>
<td>-4.02</td>
</tr>
<tr>
<td>5G-C31</td>
<td>0.24</td>
<td>-0.71</td>
<td>-0.97</td>
<td>8.46</td>
<td>24.85</td>
<td>-20.53</td>
</tr>
<tr>
<td>A8-U29</td>
<td>-6.02</td>
<td>0.98</td>
<td>-0.42</td>
<td>-23.19</td>
<td>16.28</td>
<td>63.06</td>
</tr>
<tr>
<td>A9-U26</td>
<td>1.46</td>
<td>0.77</td>
<td>1.44</td>
<td>12.47</td>
<td>-7.36</td>
<td>-20.14</td>
</tr>
<tr>
<td>C10-G25</td>
<td>0.29</td>
<td>-0.19</td>
<td>0.18</td>
<td>9.82</td>
<td>6.00</td>
<td>-9.82</td>
</tr>
<tr>
<td>C11-G24</td>
<td>0.00</td>
<td>-0.31</td>
<td>1.06</td>
<td>2.22</td>
<td>-1.31</td>
<td>-9.97</td>
</tr>
<tr>
<td>U12-A19</td>
<td>-0.32</td>
<td>-0.68</td>
<td>1.48</td>
<td>1.76</td>
<td>3.33</td>
<td>-4.43</td>
</tr>
<tr>
<td>U13-A20</td>
<td>-0.54</td>
<td>-0.70</td>
<td>1.58</td>
<td>-1.71</td>
<td>1.34</td>
<td>-9.10</td>
</tr>
<tr>
<td>U14-A21</td>
<td>0.52</td>
<td>-0.77</td>
<td>-1.42</td>
<td>-6.64</td>
<td>57.56</td>
<td>-42.60</td>
</tr>
</tbody>
</table>

Table C.1.2: Shows the local base base parameters of the *E. coli* 29-mer NMR solution structure.
Table C.1.3: Shows the ribose dihedral angles, $\nu_0$ to $\nu_4$, the amplitude of pseudorotation of the sugar ring, $tm$, the phase angle of pseudorotation of the sugar ring, $P$, and the resultant type of puckering for the first strand of the *E. coli* 29-mer NMR solution structure.

<table>
<thead>
<tr>
<th>Residue</th>
<th>$\nu_0$</th>
<th>$\nu_1$</th>
<th>$\nu_2$</th>
<th>$\nu_3$</th>
<th>$\nu_4$</th>
<th>$tm$</th>
<th>$P$</th>
<th>Puckering</th>
</tr>
</thead>
<tbody>
<tr>
<td>G35</td>
<td>-5.7</td>
<td>-16.9</td>
<td>31.2</td>
<td>-36.0</td>
<td>26.3</td>
<td>35.4</td>
<td>28.0</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>G34</td>
<td>-15.0</td>
<td>-7.5</td>
<td>25.3</td>
<td>-35.3</td>
<td>31.7</td>
<td>35.0</td>
<td>43.7</td>
<td>C$_4$-exo</td>
</tr>
<tr>
<td>G33</td>
<td>-18.3</td>
<td>-3.8</td>
<td>22.5</td>
<td>-34.5</td>
<td>33.2</td>
<td>34.9</td>
<td>49.8</td>
<td>C$_4$-exo</td>
</tr>
<tr>
<td>G32</td>
<td>-1.7</td>
<td>-19.8</td>
<td>32.2</td>
<td>-34.6</td>
<td>22.9</td>
<td>34.7</td>
<td>21.6</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>C31</td>
<td>5.1</td>
<td>-24.2</td>
<td>32.7</td>
<td>-30.9</td>
<td>16.3</td>
<td>33.2</td>
<td>10.1</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>C30</td>
<td>2.8</td>
<td>-19.6</td>
<td>27.7</td>
<td>-27.1</td>
<td>15.4</td>
<td>28.5</td>
<td>13.3</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>U29</td>
<td>-11.5</td>
<td>-11.4</td>
<td>28.1</td>
<td>-36.2</td>
<td>30.1</td>
<td>35.4</td>
<td>37.5</td>
<td>C$_4$-exo</td>
</tr>
<tr>
<td>G25</td>
<td>-8.8</td>
<td>-13.8</td>
<td>29.4</td>
<td>-35.9</td>
<td>28.1</td>
<td>35.1</td>
<td>33.1</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>G24</td>
<td>0.1</td>
<td>-22.0</td>
<td>33.9</td>
<td>-35.3</td>
<td>22.1</td>
<td>35.8</td>
<td>18.7</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>A23</td>
<td>-4.6</td>
<td>-17.9</td>
<td>31.9</td>
<td>-35.9</td>
<td>25.5</td>
<td>35.5</td>
<td>26.1</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>A22</td>
<td>3.8</td>
<td>-24.7</td>
<td>34.7</td>
<td>-33.9</td>
<td>19.0</td>
<td>35.6</td>
<td>12.9</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>A21</td>
<td>-15.6</td>
<td>27.4</td>
<td>-28.1</td>
<td>19.9</td>
<td>-2.9</td>
<td>28.8</td>
<td>166.9</td>
<td>C$_2$-endo</td>
</tr>
</tbody>
</table>

Table C.1.4: Shows the ribose dihedral angles, $\nu_0$ to $\nu_4$, the amplitude of pseudorotation of the sugar ring, $tm$, the phase angle of pseudorotation of the sugar ring, $P$, and the resultant type of puckering of the second strand of the *E. coli* 29-mer NMR solution structure.
Table C.1.5: Shows the backbone dihedral angles, for the base paired residues of the first strand of the *E. coli* 29-mer NMR solution structure. The mark “---” indicates that the particular dihedral angle is not applicable to the particular residue.

<table>
<thead>
<tr>
<th>Residue</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>δ</th>
<th>ε</th>
<th>ζ</th>
<th>χ</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>---</td>
<td>---</td>
<td>56.7</td>
<td>90.4</td>
<td>-157.0</td>
<td>-76.9</td>
<td>-158.7</td>
</tr>
<tr>
<td>C2</td>
<td>-74.0</td>
<td>-174.2</td>
<td>57.9</td>
<td>90.5</td>
<td>-155.3</td>
<td>-68.0</td>
<td>-151.1</td>
</tr>
<tr>
<td>C3</td>
<td>-71.4</td>
<td>-177.7</td>
<td>58.5</td>
<td>86.6</td>
<td>-159.6</td>
<td>-66.2</td>
<td>-152.4</td>
</tr>
<tr>
<td>C4</td>
<td>-70.7</td>
<td>-175.5</td>
<td>56.2</td>
<td>85.3</td>
<td>-157.9</td>
<td>-69.0</td>
<td>-152.3</td>
</tr>
<tr>
<td>G5</td>
<td>-73.9</td>
<td>177.3</td>
<td>63.5</td>
<td>105.9</td>
<td>-145.6</td>
<td>14.6</td>
<td>-139.3</td>
</tr>
<tr>
<td>G7</td>
<td>-72.9</td>
<td>-168.3</td>
<td>55.8</td>
<td>86.6</td>
<td>-153.8</td>
<td>-80.0</td>
<td>-160.5</td>
</tr>
<tr>
<td>A9</td>
<td>-79.5</td>
<td>-172.9</td>
<td>57.1</td>
<td>87.4</td>
<td>-153.8</td>
<td>-63.3</td>
<td>-151.4</td>
</tr>
<tr>
<td>C10</td>
<td>-73.8</td>
<td>179.3</td>
<td>58.6</td>
<td>87.5</td>
<td>-163.8</td>
<td>-86.6</td>
<td>-148.8</td>
</tr>
<tr>
<td>C11</td>
<td>-78.7</td>
<td>-179.7</td>
<td>58.9</td>
<td>87.2</td>
<td>-156.8</td>
<td>-65.1</td>
<td>-153.2</td>
</tr>
<tr>
<td>U12</td>
<td>-73.2</td>
<td>-177.6</td>
<td>59.9</td>
<td>86.0</td>
<td>-162.3</td>
<td>-66.7</td>
<td>-153.0</td>
</tr>
<tr>
<td>U13</td>
<td>-75.4</td>
<td>-174.6</td>
<td>59.0</td>
<td>84.6</td>
<td>-160.5</td>
<td>-61.5</td>
<td>-151.1</td>
</tr>
<tr>
<td>U14</td>
<td>-75.1</td>
<td>-166.7</td>
<td>54.4</td>
<td>133.2</td>
<td>---</td>
<td>---</td>
<td>-120.4</td>
</tr>
</tbody>
</table>

Table C.1.6: Shows the backbone dihedral angles, for the base paired residues of the second strand of the *E. coli* 29-mer NMR solution structure. The mark “---” indicates that the particular dihedral angle is not applicable to the particular residue.

<table>
<thead>
<tr>
<th>Residue</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>δ</th>
<th>ε</th>
<th>ζ</th>
<th>χ</th>
</tr>
</thead>
<tbody>
<tr>
<td>G35</td>
<td>-73.0</td>
<td>-174.6</td>
<td>55.1</td>
<td>83.3</td>
<td>---</td>
<td>---</td>
<td>-143.1</td>
</tr>
<tr>
<td>G34</td>
<td>-73.0</td>
<td>-177.8</td>
<td>56.8</td>
<td>86.9</td>
<td>-165.8</td>
<td>-72.9</td>
<td>-145.4</td>
</tr>
<tr>
<td>G33</td>
<td>-71.4</td>
<td>-179.9</td>
<td>57.4</td>
<td>87.7</td>
<td>-165.6</td>
<td>-75.3</td>
<td>-149.2</td>
</tr>
<tr>
<td>G32</td>
<td>-71.2</td>
<td>-175.7</td>
<td>57.8</td>
<td>87.5</td>
<td>-157.7</td>
<td>-65.6</td>
<td>-154.9</td>
</tr>
<tr>
<td>C31</td>
<td>-76.7</td>
<td>-178.1</td>
<td>55.8</td>
<td>91.2</td>
<td>-155.0</td>
<td>-68.6</td>
<td>-157.0</td>
</tr>
<tr>
<td>U29</td>
<td>-58.5</td>
<td>164.2</td>
<td>74.2</td>
<td>95.7</td>
<td>-119.6</td>
<td>149.2</td>
<td>-170.7</td>
</tr>
<tr>
<td>U26</td>
<td>-70.5</td>
<td>-177.9</td>
<td>63.0</td>
<td>85.9</td>
<td>-166.9</td>
<td>-66.4</td>
<td>-152.8</td>
</tr>
<tr>
<td>G25</td>
<td>-69.3</td>
<td>177.0</td>
<td>59.4</td>
<td>86.2</td>
<td>-169.0</td>
<td>-73.6</td>
<td>-149.5</td>
</tr>
<tr>
<td>G24</td>
<td>-76.1</td>
<td>-175.9</td>
<td>58.8</td>
<td>86.8</td>
<td>-161.5</td>
<td>-62.5</td>
<td>-145.4</td>
</tr>
<tr>
<td>A23</td>
<td>-76.2</td>
<td>-177.6</td>
<td>61.6</td>
<td>86.2</td>
<td>-164.5</td>
<td>-74.4</td>
<td>-152.7</td>
</tr>
<tr>
<td>A22</td>
<td>-76.7</td>
<td>-179.4</td>
<td>63.7</td>
<td>88.2</td>
<td>-159.6</td>
<td>-63.5</td>
<td>-151.5</td>
</tr>
<tr>
<td>A21</td>
<td>---</td>
<td>---</td>
<td>56.3</td>
<td>141.1</td>
<td>-172.6</td>
<td>-111.1</td>
<td>-118.2</td>
</tr>
</tbody>
</table>
C.2. Additional helical parameters and dihedral angles for the *H. h.* 29-mer NMR solution structure

The local step parameters are shown in Table C.2.1 and the local base pair parameters are shown in Table C.2.2. The ribose dihedral angles of the base paired residues, their sugar pucker, the amplitude of pseudorotation of the sugar ring and the phase angle of pseudorotation of the sugar ring are shown in Table C.2.3 and Table C.2.4. Table C.2.5 and Table C.2.6 shows the backbone and glycosidic torsion angles of all the base paired residues.

<table>
<thead>
<tr>
<th>Step</th>
<th>X-displacement</th>
<th>Y-displacement</th>
<th>Inclination</th>
<th>Tip</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-C2/G34-G35</td>
<td>-4.53</td>
<td>-1.92</td>
<td>23.90</td>
<td>15.16</td>
</tr>
<tr>
<td>C2-C3/G33-G34</td>
<td>-6.02</td>
<td>3.94</td>
<td>20.48</td>
<td>-16.39</td>
</tr>
<tr>
<td>C3-U4/A32-G33</td>
<td>-3.29</td>
<td>-0.74</td>
<td>2.34</td>
<td>6.95</td>
</tr>
<tr>
<td>U4-A5/U31-A32</td>
<td>-5.40</td>
<td>-1.25</td>
<td>13.52</td>
<td>10.93</td>
</tr>
<tr>
<td>A5-U6/U30-U31</td>
<td>-1.81</td>
<td>-2.19</td>
<td>-2.22</td>
<td>-12.72</td>
</tr>
<tr>
<td>U6-A7/C29-U30</td>
<td>-1.06</td>
<td>2.94</td>
<td>2.62</td>
<td>-22.29</td>
</tr>
<tr>
<td>A7-G8/C28-A27</td>
<td>-4.06</td>
<td>5.18</td>
<td>31.01</td>
<td>11.55</td>
</tr>
<tr>
<td>G8-A9/A27-C28</td>
<td>-1.91</td>
<td>-3.26</td>
<td>3.35</td>
<td>2.11</td>
</tr>
<tr>
<td>G10-C11/G24-C25</td>
<td>-7.77</td>
<td>-0.67</td>
<td>25.44</td>
<td>-0.23</td>
</tr>
<tr>
<td>C11-U12/A23-G24</td>
<td>-5.60</td>
<td>3.29</td>
<td>26.47</td>
<td>-19.61</td>
</tr>
<tr>
<td>U12-U13/A22-A23</td>
<td>-0.84</td>
<td>-0.96</td>
<td>-5.20</td>
<td>4.36</td>
</tr>
<tr>
<td>U13-U14/A21-A22</td>
<td>1.04</td>
<td>-5.37</td>
<td>-19.27</td>
<td>16.97</td>
</tr>
</tbody>
</table>

Table C.2.1: Shows the local base step parameters of the *H. h.* 29-mer NMR solution structure.
### Table C.2.2: Shows the local base pair parameters of the *H. h.* 29-mer NMR solution structure.

<table>
<thead>
<tr>
<th>Base pair</th>
<th>Shear</th>
<th>Stretch</th>
<th>Stagger</th>
<th>Buckle</th>
<th>Propeller</th>
<th>Opening</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-G35</td>
<td>-0.37</td>
<td>-0.38</td>
<td>-0.86</td>
<td>-3.81</td>
<td>29.95</td>
<td>-10.85</td>
</tr>
<tr>
<td>C2-G34</td>
<td>-0.10</td>
<td>-0.67</td>
<td>0.19</td>
<td>13.09</td>
<td>11.35</td>
<td>-0.03</td>
</tr>
<tr>
<td>C3-G33</td>
<td>-0.32</td>
<td>-0.98</td>
<td>-1.67</td>
<td>3.99</td>
<td>7.42</td>
<td>-11.22</td>
</tr>
<tr>
<td>U4-A32</td>
<td>0.14</td>
<td>-0.31</td>
<td>-0.62</td>
<td>1.51</td>
<td>-2.76</td>
<td>-8.36</td>
</tr>
<tr>
<td>A5-U31</td>
<td>0.22</td>
<td>-1.02</td>
<td>1.43</td>
<td>0.72</td>
<td>-17.19</td>
<td>-10.66</td>
</tr>
<tr>
<td>U6-U30</td>
<td>-1.87</td>
<td>-0.48</td>
<td>1.30</td>
<td>-7.40</td>
<td>-26.19</td>
<td>12.37</td>
</tr>
<tr>
<td>A7-C29</td>
<td>0.96</td>
<td>-0.34</td>
<td>-2.07</td>
<td>-7.35</td>
<td>-6.18</td>
<td>-1.92</td>
</tr>
<tr>
<td>G8-C28</td>
<td>-0.07</td>
<td>-0.62</td>
<td>-2.33</td>
<td>-7.86</td>
<td>-21.46</td>
<td>-52.50</td>
</tr>
<tr>
<td>A9-A27</td>
<td>1.06</td>
<td>-0.14</td>
<td>-0.89</td>
<td>-21.26</td>
<td>10.74</td>
<td>-25.28</td>
</tr>
<tr>
<td>G10-C25</td>
<td>0.19</td>
<td>-0.28</td>
<td>0.02</td>
<td>12.03</td>
<td>14.60</td>
<td>-8.98</td>
</tr>
<tr>
<td>C11-G24</td>
<td>0.02</td>
<td>-0.71</td>
<td>-0.23</td>
<td>0.23</td>
<td>5.81</td>
<td>-1.55</td>
</tr>
<tr>
<td>U12-A23</td>
<td>0.14</td>
<td>-0.89</td>
<td>-1.71</td>
<td>6.95</td>
<td>7.04</td>
<td>-10.68</td>
</tr>
<tr>
<td>U13-A22</td>
<td>0.02</td>
<td>-0.80</td>
<td>-1.68</td>
<td>14.84</td>
<td>12.68</td>
<td>-10.37</td>
</tr>
<tr>
<td>U14-A21</td>
<td>0.35</td>
<td>-0.56</td>
<td>0.69</td>
<td>-0.98</td>
<td>30.60</td>
<td>14.97</td>
</tr>
</tbody>
</table>

### Table C.2.3: Shows the ribose dihedral angles, $\nu_0$ to $\nu_4$, the amplitude of pseudorotation of the sugar ring, $tm$, the phase angle of pseudorotation of the sugar ring, $P$, and the resultant type of puckering of the first strand of the *H. h.* 29-mer NMR solution structure.

<table>
<thead>
<tr>
<th>Residue</th>
<th>$\nu_0$</th>
<th>$\nu_1$</th>
<th>$\nu_2$</th>
<th>$\nu_3$</th>
<th>$\nu_4$</th>
<th>tm</th>
<th>P</th>
<th>Puckering</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>-14.2</td>
<td>-9.1</td>
<td>27.0</td>
<td>-36.6</td>
<td>31.8</td>
<td>36.0</td>
<td>41.5</td>
<td>C$_4$-exo</td>
</tr>
<tr>
<td>C2</td>
<td>-6.2</td>
<td>-17.3</td>
<td>32.3</td>
<td>-37.4</td>
<td>27.5</td>
<td>36.7</td>
<td>28.4</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>C3</td>
<td>-15.5</td>
<td>-8.4</td>
<td>27.1</td>
<td>-37.3</td>
<td>33.2</td>
<td>37.0</td>
<td>43.0</td>
<td>C$_4$-exo</td>
</tr>
<tr>
<td>U4</td>
<td>-6.6</td>
<td>-15.0</td>
<td>29.7</td>
<td>-34.8</td>
<td>25.7</td>
<td>34.2</td>
<td>29.7</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>A5</td>
<td>-1.7</td>
<td>-20.1</td>
<td>32.5</td>
<td>-35.1</td>
<td>23.2</td>
<td>35.0</td>
<td>21.7</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>U6</td>
<td>-0.4</td>
<td>-23.8</td>
<td>37.1</td>
<td>-38.7</td>
<td>24.7</td>
<td>39.4</td>
<td>19.3</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>A7</td>
<td>-17.0</td>
<td>-1.9</td>
<td>18.4</td>
<td>-29</td>
<td>28.8</td>
<td>30.0</td>
<td>52.2</td>
<td>C$_4$-exo</td>
</tr>
<tr>
<td>G8</td>
<td>-24.4</td>
<td>6.9</td>
<td>11.3</td>
<td>-26</td>
<td>31.7</td>
<td>31.1</td>
<td>68.6</td>
<td>C$_4$-exo</td>
</tr>
<tr>
<td>A9</td>
<td>-18.7</td>
<td>-2.6</td>
<td>21.1</td>
<td>-32.8</td>
<td>31.9</td>
<td>33.7</td>
<td>51.3</td>
<td>C$_4$-exo</td>
</tr>
<tr>
<td>G10</td>
<td>-7.1</td>
<td>-16.1</td>
<td>31.4</td>
<td>-37</td>
<td>27.7</td>
<td>36.3</td>
<td>29.9</td>
<td>C$_3$-endo</td>
</tr>
<tr>
<td>C11</td>
<td>-18.6</td>
<td>-5.5</td>
<td>25.7</td>
<td>-37.8</td>
<td>35.3</td>
<td>38.0</td>
<td>47.4</td>
<td>C$_4$-exo</td>
</tr>
<tr>
<td>U12</td>
<td>-32.8</td>
<td>23.6</td>
<td>-6.5</td>
<td>-12.7</td>
<td>28.6</td>
<td>32.4</td>
<td>101.5</td>
<td>O$_2$-endo</td>
</tr>
<tr>
<td>U13</td>
<td>-28.4</td>
<td>33.0</td>
<td>-25.4</td>
<td>9.5</td>
<td>11.9</td>
<td>32.8</td>
<td>140.8</td>
<td>C$_1$-exo</td>
</tr>
<tr>
<td>U14</td>
<td>-26.9</td>
<td>36.2</td>
<td>-31.2</td>
<td>16.5</td>
<td>6.3</td>
<td>35.6</td>
<td>151.1</td>
<td>C$_2$-endo</td>
</tr>
</tbody>
</table>

---

336
Table C.2.4: Shows the ribose dihedral angles, $\nu_0$ to $\nu_4$, the amplitude of pseudorotation of the sugar ring, $t_m$, the phase angle of pseudorotation of the sugar ring, $P$, and the resultant type of puckering of the second strand of the H. h. 29-mer NMR solution structure.

Table C.2.5: Shows the backbone dihedral angles, for the base paired residues of the first strand of the H. h. 29-mer NMR solution structure. The mark “---” indicates that the particular dihedral angle is not applicable to the particular residue.
<table>
<thead>
<tr>
<th>Residue</th>
<th>alpha</th>
<th>beta</th>
<th>gamma</th>
<th>delta</th>
<th>epsilon</th>
<th>zeta</th>
<th>chi</th>
</tr>
</thead>
<tbody>
<tr>
<td>G35</td>
<td>-64.3</td>
<td>-176.4</td>
<td>50.9</td>
<td>142.6</td>
<td>---</td>
<td>---</td>
<td>-108.2</td>
</tr>
<tr>
<td>G34</td>
<td>-64.0</td>
<td>179.4</td>
<td>52.9</td>
<td>82.4</td>
<td>-157.3</td>
<td>-70.7</td>
<td>-158.3</td>
</tr>
<tr>
<td>G33</td>
<td>-61.8</td>
<td>172.3</td>
<td>53.9</td>
<td>82.0</td>
<td>-156.6</td>
<td>-69.5</td>
<td>-157.7</td>
</tr>
<tr>
<td>A32</td>
<td>-64.0</td>
<td>170.2</td>
<td>53.9</td>
<td>82.1</td>
<td>-157.1</td>
<td>-66.8</td>
<td>-158.6</td>
</tr>
<tr>
<td>U31</td>
<td>-67.7</td>
<td>172.1</td>
<td>53.3</td>
<td>82.3</td>
<td>-154.6</td>
<td>-61.3</td>
<td>-157.9</td>
</tr>
<tr>
<td>U30</td>
<td>-64.8</td>
<td>172.1</td>
<td>55.8</td>
<td>82.3</td>
<td>-161.1</td>
<td>-91.0</td>
<td>-157.1</td>
</tr>
<tr>
<td>C29</td>
<td>-65.1</td>
<td>170.7</td>
<td>51.4</td>
<td>110.0</td>
<td>-175.7</td>
<td>-66.1</td>
<td>-132.7</td>
</tr>
<tr>
<td>A27</td>
<td>-58.6</td>
<td>171.6</td>
<td>54.7</td>
<td>83.3</td>
<td>-155.0</td>
<td>-62.0</td>
<td>-149.1</td>
</tr>
<tr>
<td>U26</td>
<td>-66.6</td>
<td>173.8</td>
<td>51.8</td>
<td>112.1</td>
<td>177.5</td>
<td>-96.2</td>
<td>-123.8</td>
</tr>
<tr>
<td>C25</td>
<td>-63.6</td>
<td>-179.3</td>
<td>53.5</td>
<td>82.0</td>
<td>-157.9</td>
<td>-74.4</td>
<td>-157.9</td>
</tr>
<tr>
<td>G24</td>
<td>-64.1</td>
<td>178.0</td>
<td>53.9</td>
<td>82.4</td>
<td>-156.5</td>
<td>-68.1</td>
<td>-157.5</td>
</tr>
<tr>
<td>A23</td>
<td>-63.3</td>
<td>171.9</td>
<td>53.4</td>
<td>82.4</td>
<td>-156.0</td>
<td>-67.6</td>
<td>-155.8</td>
</tr>
<tr>
<td>A22</td>
<td>-106.9</td>
<td>159.0</td>
<td>91.0</td>
<td>83.1</td>
<td>-155.9</td>
<td>-74.0</td>
<td>-164.1</td>
</tr>
<tr>
<td>A21</td>
<td>---</td>
<td>---</td>
<td>51.9</td>
<td>143.5</td>
<td>176.8</td>
<td>-95.3</td>
<td>-120.0</td>
</tr>
</tbody>
</table>

Table C.2.6: Shows the backbone dihedral angles, for the base paired residues of the second strand of the H. h. 29-mer NMR solution structure. The mark “---” indicates that the particular dihedral angle is not applicable to the particular residue.

C.3. Additional helical parameters and dihedral angles for the H. h. 37-mer NMR solution structure

The local base pair step parameters are shown in Table C.3.1 and the local base pair parameters are shown in Table C.3.2. The ribose dihedral angles of the base paired residues, their sugar pucker, the amplitude of pseudorotation of the sugar ring and the phase angle of pseudorotation of the sugar ring are shown in Table C.3.3. Table C.3.4 shows the backbone and glycosidic torsion angles of all base paired residues.
<table>
<thead>
<tr>
<th>Step</th>
<th>X-displacement</th>
<th>Y-displacement</th>
<th>Inclination</th>
<th>Tip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ga-C1/G35-Cb</td>
<td>-1.93</td>
<td>2.01</td>
<td>2.64</td>
<td>-12.98</td>
</tr>
<tr>
<td>C1-C2/G34-G35</td>
<td>-5.88</td>
<td>0.75</td>
<td>16.99</td>
<td>-2.16</td>
</tr>
<tr>
<td>C2-C3/G33-G34</td>
<td>-6.00</td>
<td>-2.11</td>
<td>18.91</td>
<td>7.08</td>
</tr>
<tr>
<td>C3-U4/A32-G33</td>
<td>-6.38</td>
<td>0.04</td>
<td>18.18</td>
<td>1.89</td>
</tr>
<tr>
<td>A9-G10/C25-U26</td>
<td>-3.01</td>
<td>-0.94</td>
<td>13.28</td>
<td>7.95</td>
</tr>
<tr>
<td>G10-C11/G24-C25</td>
<td>-5.63</td>
<td>4.45</td>
<td>19.97</td>
<td>-4.15</td>
</tr>
<tr>
<td>C11-U12/A23-G24</td>
<td>-5.80</td>
<td>1.72</td>
<td>16.11</td>
<td>-3.10</td>
</tr>
<tr>
<td>U12-U13/A22-A23</td>
<td>-4.54</td>
<td>-6.32</td>
<td>11.13</td>
<td>6.70</td>
</tr>
<tr>
<td>U13-U14/A21-A22</td>
<td>-7.90</td>
<td>0.38</td>
<td>27.12</td>
<td>0.25</td>
</tr>
<tr>
<td>U14-C15/G20-A21</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>U14-C15/C18-A21</td>
<td>-0.94</td>
<td>1.07</td>
<td>35.43</td>
<td>-15.65</td>
</tr>
<tr>
<td>C15-U16/G20-C18</td>
<td>2.18</td>
<td>-0.88</td>
<td>25.02</td>
<td>-15.66</td>
</tr>
</tbody>
</table>

Table C.3.1: Shows local base pair step parameters, and the form of the RNA at a given step of the H. h. 37-mer NMR solution structure, a ~ indicates that the information could not be given.

<table>
<thead>
<tr>
<th>Base pair</th>
<th>Shear</th>
<th>Stretch</th>
<th>Stagger</th>
<th>Buckle</th>
<th>Propeller</th>
<th>Opening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ga-Cb</td>
<td>-6.49</td>
<td>-0.82</td>
<td>2.09</td>
<td>42.04</td>
<td>45.80</td>
<td>2.69</td>
</tr>
<tr>
<td>C1-G35</td>
<td>-0.63</td>
<td>-1.32</td>
<td>-1.06</td>
<td>24.56</td>
<td>23.78</td>
<td>-8.47</td>
</tr>
<tr>
<td>C2-G34</td>
<td>-0.31</td>
<td>-2.31</td>
<td>-1.90</td>
<td>3.14</td>
<td>14.59</td>
<td>-8.50</td>
</tr>
<tr>
<td>C3-G33</td>
<td>0.20</td>
<td>-1.65</td>
<td>-1.07</td>
<td>-13.58</td>
<td>12.61</td>
<td>0.05</td>
</tr>
<tr>
<td>U4-A32</td>
<td>0.38</td>
<td>-1.01</td>
<td>-0.02</td>
<td>-30.77</td>
<td>35.99</td>
<td>-8.90</td>
</tr>
<tr>
<td>G8-A27</td>
<td>3.84</td>
<td>-0.13</td>
<td>-1.21</td>
<td>-47.68</td>
<td>-20.60</td>
<td>-51.38</td>
</tr>
<tr>
<td>A9-U26</td>
<td>-4.13</td>
<td>-1.01</td>
<td>-0.83</td>
<td>-0.69</td>
<td>20.68</td>
<td>-7.39</td>
</tr>
<tr>
<td>G10-C25</td>
<td>-0.12</td>
<td>-0.50</td>
<td>-0.72</td>
<td>10.84</td>
<td>21.98</td>
<td>-3.53</td>
</tr>
<tr>
<td>C11-G24</td>
<td>-0.47</td>
<td>-0.95</td>
<td>-1.24</td>
<td>8.48</td>
<td>21.57</td>
<td>-20.27</td>
</tr>
<tr>
<td>U12-A23</td>
<td>-0.26</td>
<td>-1.21</td>
<td>-1.50</td>
<td>-1.51</td>
<td>25.92</td>
<td>-30.32</td>
</tr>
<tr>
<td>U13-A22</td>
<td>-0.16</td>
<td>-0.69</td>
<td>-1.18</td>
<td>-0.98</td>
<td>15.36</td>
<td>1.49</td>
</tr>
<tr>
<td>U14-A21</td>
<td>-0.07</td>
<td>-0.70</td>
<td>-0.70</td>
<td>-8.44</td>
<td>12.21</td>
<td>-4.52</td>
</tr>
<tr>
<td>C15-C18</td>
<td>0.98</td>
<td>-1.38</td>
<td>-1.46</td>
<td>-31.84</td>
<td>-31.29</td>
<td>-108.05</td>
</tr>
<tr>
<td>U16-G20</td>
<td>-3.10</td>
<td>1.56</td>
<td>-1.95</td>
<td>18.32</td>
<td>52.68</td>
<td>64.78</td>
</tr>
</tbody>
</table>

Table C.3.2: Shows the local base pair parameters of the H. h. 37-mer NMR solution structure.
<table>
<thead>
<tr>
<th>Residue</th>
<th>ν₀</th>
<th>ν₁</th>
<th>ν₂</th>
<th>ν₃</th>
<th>ν₄</th>
<th>tm</th>
<th>P</th>
<th>Puckering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ga</td>
<td>-14.9</td>
<td>-11.3</td>
<td>31.5</td>
<td>-41.5</td>
<td>35.1</td>
<td>40.9</td>
<td>39.6</td>
<td>C₄'-exo</td>
</tr>
<tr>
<td>C1</td>
<td>-10.3</td>
<td>-14.1</td>
<td>31.4</td>
<td>-39.2</td>
<td>31.2</td>
<td>38.1</td>
<td>34.6</td>
<td>C₃'-endo</td>
</tr>
<tr>
<td>C2</td>
<td>-14.9</td>
<td>-11.0</td>
<td>30.7</td>
<td>-41.2</td>
<td>35.5</td>
<td>40.4</td>
<td>40.4</td>
<td>C₄'-exo</td>
</tr>
<tr>
<td>C3</td>
<td>-14.0</td>
<td>-11.7</td>
<td>31.2</td>
<td>-41.0</td>
<td>34.7</td>
<td>40.2</td>
<td>39.1</td>
<td>C₄'-exo</td>
</tr>
<tr>
<td>U4</td>
<td>-17.0</td>
<td>-9.7</td>
<td>30.6</td>
<td>-42.4</td>
<td>37.9</td>
<td>41.8</td>
<td>43.0</td>
<td>C₄'-exo</td>
</tr>
<tr>
<td>A5</td>
<td>12.1</td>
<td>-40.0</td>
<td>51.0</td>
<td>-45.6</td>
<td>21.2</td>
<td>51.3</td>
<td>5.3</td>
<td>C₃'-endo</td>
</tr>
<tr>
<td>A9</td>
<td>3.9</td>
<td>-22.2</td>
<td>30.5</td>
<td>-29.7</td>
<td>16.4</td>
<td>31.2</td>
<td>12.1</td>
<td>C₃'-endo</td>
</tr>
<tr>
<td>G10</td>
<td>-17.0</td>
<td>-8.5</td>
<td>28.8</td>
<td>-40.0</td>
<td>35.8</td>
<td>39.8</td>
<td>43.6</td>
<td>C₄'-exo</td>
</tr>
<tr>
<td>C11</td>
<td>-4.4</td>
<td>-21.0</td>
<td>36.3</td>
<td>-40.6</td>
<td>28.6</td>
<td>40.1</td>
<td>25.2</td>
<td>C₃'-endo</td>
</tr>
<tr>
<td>U12</td>
<td>-10.5</td>
<td>-16.0</td>
<td>34.0</td>
<td>-41.8</td>
<td>33.1</td>
<td>40.8</td>
<td>33.5</td>
<td>C₃'-endo</td>
</tr>
<tr>
<td>U13</td>
<td>-7.9</td>
<td>-17.7</td>
<td>34.6</td>
<td>-40.7</td>
<td>30.6</td>
<td>39.9</td>
<td>30.0</td>
<td>C₃'-endo</td>
</tr>
<tr>
<td>U14</td>
<td>-22.2</td>
<td>-2.7</td>
<td>24.5</td>
<td>-38.7</td>
<td>38.3</td>
<td>39.8</td>
<td>52.0</td>
<td>C₄'-exo</td>
</tr>
<tr>
<td>C15</td>
<td>5.2</td>
<td>-24.5</td>
<td>33.0</td>
<td>-31.5</td>
<td>16.7</td>
<td>33.5</td>
<td>10.4</td>
<td>C₃'-endo</td>
</tr>
<tr>
<td>U16</td>
<td>2.1</td>
<td>-24.4</td>
<td>35.7</td>
<td>-36.4</td>
<td>21.5</td>
<td>37.1</td>
<td>15.9</td>
<td>C₃'-endo</td>
</tr>
<tr>
<td>C18</td>
<td>-33.1</td>
<td>22.7</td>
<td>-4.6</td>
<td>-14.8</td>
<td>30.0</td>
<td>33.0</td>
<td>97.9</td>
<td>O₄'-endo</td>
</tr>
<tr>
<td>G20</td>
<td>22.0</td>
<td>-39.0</td>
<td>40.1</td>
<td>-29.1</td>
<td>4.6</td>
<td>41.1</td>
<td>347.5</td>
<td>C₂'-exo</td>
</tr>
<tr>
<td>A21</td>
<td>-9.4</td>
<td>-16.0</td>
<td>33.1</td>
<td>-40.4</td>
<td>31.5</td>
<td>39.3</td>
<td>32.7</td>
<td>C₃'-endo</td>
</tr>
<tr>
<td>A22</td>
<td>-17.5</td>
<td>-8.3</td>
<td>28.8</td>
<td>-40.6</td>
<td>36.6</td>
<td>40.2</td>
<td>44.2</td>
<td>C₄'-exo</td>
</tr>
<tr>
<td>A23</td>
<td>-14.2</td>
<td>-11.9</td>
<td>31.4</td>
<td>-41.1</td>
<td>35.1</td>
<td>40.4</td>
<td>39.1</td>
<td>C₄'-exo</td>
</tr>
<tr>
<td>G24</td>
<td>-9.1</td>
<td>-17.0</td>
<td>34.4</td>
<td>-41.3</td>
<td>31.9</td>
<td>40.5</td>
<td>31.7</td>
<td>C₃'-endo</td>
</tr>
<tr>
<td>C25</td>
<td>-13.2</td>
<td>-12.8</td>
<td>31.8</td>
<td>-41.0</td>
<td>34.2</td>
<td>40.2</td>
<td>37.7</td>
<td>C₄'-exo</td>
</tr>
<tr>
<td>U26</td>
<td>-11.6</td>
<td>-13.5</td>
<td>31.5</td>
<td>-39.6</td>
<td>32.0</td>
<td>38.8</td>
<td>35.7</td>
<td>C₃'-endo</td>
</tr>
<tr>
<td>A27</td>
<td>58.2</td>
<td>-31.9</td>
<td>-5.5</td>
<td>40.8</td>
<td>-64.0</td>
<td>63.6</td>
<td>265.1</td>
<td>O₄'-exo</td>
</tr>
<tr>
<td>A32</td>
<td>-21.4</td>
<td>-1.8</td>
<td>22.5</td>
<td>-35.8</td>
<td>35.4</td>
<td>37.1</td>
<td>52.7</td>
<td>C₄'-exo</td>
</tr>
<tr>
<td>G33</td>
<td>-8.6</td>
<td>-15.6</td>
<td>32.2</td>
<td>-38.8</td>
<td>30.0</td>
<td>38.0</td>
<td>32.0</td>
<td>C₃'-endo</td>
</tr>
<tr>
<td>G34</td>
<td>-16.3</td>
<td>-8.9</td>
<td>28.9</td>
<td>-39.9</td>
<td>35.5</td>
<td>39.5</td>
<td>43.0</td>
<td>C₄'-exo</td>
</tr>
<tr>
<td>G35</td>
<td>-12.5</td>
<td>-13.3</td>
<td>32.0</td>
<td>-41.1</td>
<td>34.0</td>
<td>40.1</td>
<td>37.0</td>
<td>C₄'-exo</td>
</tr>
<tr>
<td>Cb</td>
<td>3.9</td>
<td>-29.0</td>
<td>41.1</td>
<td>-40.5</td>
<td>23.2</td>
<td>42.3</td>
<td>13.7</td>
<td>C₃'-endo</td>
</tr>
</tbody>
</table>

Table C.3.3: Shows the ribose dihedral angles, ν₀ to ν₄, the amplitude of pseudorotation of the sugar ring, tm, the phase angle of pseudorotation of the sugar ring, P, and the resultant type of puckering of the *H. h. 37-mer* NMR solution structure.
<table>
<thead>
<tr>
<th>Residue</th>
<th>alpha</th>
<th>beta</th>
<th>gamma</th>
<th>delta</th>
<th>epsilon</th>
<th>zeta</th>
<th>chi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ga</td>
<td>---</td>
<td>---</td>
<td>53.8</td>
<td>80.7</td>
<td>-157.7</td>
<td>-63.5</td>
<td>-159.2</td>
</tr>
<tr>
<td>C1</td>
<td>-56.8</td>
<td>-176.6</td>
<td>52.1</td>
<td>80.8</td>
<td>-154.1</td>
<td>-66.9</td>
<td>-158.0</td>
</tr>
<tr>
<td>C2</td>
<td>-64.9</td>
<td>-179.0</td>
<td>51.6</td>
<td>80.4</td>
<td>-154.4</td>
<td>-65.8</td>
<td>-157.2</td>
</tr>
<tr>
<td>C3</td>
<td>-62.6</td>
<td>-178.6</td>
<td>52.6</td>
<td>80.6</td>
<td>-155.5</td>
<td>-67.1</td>
<td>-157.5</td>
</tr>
<tr>
<td>U4</td>
<td>-62.6</td>
<td>-174.5</td>
<td>52.0</td>
<td>79.0</td>
<td>-153.2</td>
<td>-72.2</td>
<td>-157.0</td>
</tr>
<tr>
<td>A5</td>
<td>-64.1</td>
<td>171.3</td>
<td>52.6</td>
<td>78.9</td>
<td>-154.8</td>
<td>-68.5</td>
<td>-177.7</td>
</tr>
<tr>
<td>A9</td>
<td>-48.0</td>
<td>-144.1</td>
<td>139.4</td>
<td>90.1</td>
<td>-149.4</td>
<td>-56.9</td>
<td>-157.4</td>
</tr>
<tr>
<td>G10</td>
<td>-63.8</td>
<td>172.0</td>
<td>53.6</td>
<td>81.0</td>
<td>-156.8</td>
<td>-88.4</td>
<td>-160.9</td>
</tr>
<tr>
<td>C11</td>
<td>-68.8</td>
<td>177.5</td>
<td>52.4</td>
<td>81.0</td>
<td>-154.7</td>
<td>-69.3</td>
<td>-157.6</td>
</tr>
<tr>
<td>U12</td>
<td>-65.1</td>
<td>-179.9</td>
<td>52.9</td>
<td>81.4</td>
<td>-153.9</td>
<td>-65.3</td>
<td>-157.9</td>
</tr>
<tr>
<td>U13</td>
<td>-56.8</td>
<td>168.9</td>
<td>54.6</td>
<td>82.4</td>
<td>-156.9</td>
<td>-67.4</td>
<td>-158.0</td>
</tr>
<tr>
<td>U14</td>
<td>-62.2</td>
<td>179.2</td>
<td>55.2</td>
<td>81.8</td>
<td>-155.8</td>
<td>-84.6</td>
<td>-157.9</td>
</tr>
<tr>
<td>C15</td>
<td>119.4</td>
<td>-164.8</td>
<td>-179.0</td>
<td>85.5</td>
<td>-151.7</td>
<td>-73.9</td>
<td>-163.7</td>
</tr>
<tr>
<td>U16</td>
<td>-73.5</td>
<td>164.7</td>
<td>52.3</td>
<td>82.3</td>
<td>-155.6</td>
<td>-61.2</td>
<td>-158.5</td>
</tr>
<tr>
<td>U17</td>
<td>-151.0</td>
<td>109.2</td>
<td>74.6</td>
<td>146.9</td>
<td>-87.2</td>
<td>-58.6</td>
<td>-126.3</td>
</tr>
<tr>
<td>C18</td>
<td>151.3</td>
<td>-135.1</td>
<td>-174.4</td>
<td>85.9</td>
<td>-158.5</td>
<td>-76.1</td>
<td>-163.5</td>
</tr>
<tr>
<td>G19</td>
<td>169.9</td>
<td>-171.1</td>
<td>52.1</td>
<td>82.5</td>
<td>-158.6</td>
<td>-59.8</td>
<td>-158.2</td>
</tr>
<tr>
<td>G20</td>
<td>-80.0</td>
<td>171.5</td>
<td>36.8</td>
<td>110.9</td>
<td>-72.4</td>
<td>70.2</td>
<td>-135.6</td>
</tr>
<tr>
<td>A21</td>
<td>-65.1</td>
<td>-179.1</td>
<td>52.5</td>
<td>80.9</td>
<td>-154.2</td>
<td>-66.5</td>
<td>-157.5</td>
</tr>
<tr>
<td>A22</td>
<td>-66.1</td>
<td>178.7</td>
<td>52.6</td>
<td>80.9</td>
<td>-155.3</td>
<td>-88.9</td>
<td>-157.4</td>
</tr>
<tr>
<td>A23</td>
<td>-64.0</td>
<td>170.5</td>
<td>53.2</td>
<td>81.2</td>
<td>-155.8</td>
<td>-69.3</td>
<td>-157.5</td>
</tr>
<tr>
<td>G24</td>
<td>-64.6</td>
<td>178.6</td>
<td>53.0</td>
<td>81.8</td>
<td>-154.6</td>
<td>-63.4</td>
<td>-157.5</td>
</tr>
<tr>
<td>C25</td>
<td>-65.1</td>
<td>178.7</td>
<td>53.3</td>
<td>81.9</td>
<td>-154.8</td>
<td>-67.1</td>
<td>-157.7</td>
</tr>
<tr>
<td>U26</td>
<td>-63.2</td>
<td>162.2</td>
<td>55.4</td>
<td>81.2</td>
<td>-160.7</td>
<td>-78.1</td>
<td>-157.6</td>
</tr>
<tr>
<td>A27</td>
<td>-68.3</td>
<td>-136.2</td>
<td>75.3</td>
<td>162.9</td>
<td>61.8</td>
<td>94.2</td>
<td>-18.4</td>
</tr>
<tr>
<td>A32</td>
<td>-62.1</td>
<td>-179.3</td>
<td>55.3</td>
<td>81.7</td>
<td>-161.9</td>
<td>-72.3</td>
<td>-146.0</td>
</tr>
<tr>
<td>G33</td>
<td>-62.6</td>
<td>177.2</td>
<td>51.8</td>
<td>81.2</td>
<td>-155.1</td>
<td>-66.4</td>
<td>-156.8</td>
</tr>
<tr>
<td>G34</td>
<td>-63.4</td>
<td>-180.0</td>
<td>52.6</td>
<td>81.0</td>
<td>-155.2</td>
<td>-68.5</td>
<td>-157.3</td>
</tr>
<tr>
<td>G35</td>
<td>-64.3</td>
<td>-179.3</td>
<td>52.2</td>
<td>81.2</td>
<td>-153.4</td>
<td>-66.0</td>
<td>-157.9</td>
</tr>
<tr>
<td>Cb</td>
<td>-65.3</td>
<td>-179.0</td>
<td>53.2</td>
<td>80.1</td>
<td>---</td>
<td>---</td>
<td>-172.0</td>
</tr>
</tbody>
</table>

Table C.3.4: Shows the backbone dihedral angles, for the base paired residues of the second strand of the H. h. 37-mer NMR solution structure. The mark “---” indicates that the particular dihedral angle is not applicable to the particular residue.