Complete 3D description of dynamic behaviour of enzyme mimics: role of various structural elements in catalysis and interactions with bio-target

A thesis submitted to the University of Manchester for the degree of Doctor in Philosophy in the Faculty of Biology, Medicine and Health

2018

Linda T. Trivoluzzi

The University of Manchester

School of Health Sciences

Division of Pharmacy and Optometry
# Table of Contents

List of Figures ................................................................. 7  
List of Tables ................................................................. 20  
List of Schemes ............................................................... 21  
List of abbreviations ....................................................... 22  
Abstract ........................................................................... 25  
Declaration ......................................................................... 26  
Copyright Statement .......................................................... 27  
Acknowledgements .............................................................. 28  
Publications ....................................................................... 29  

## Chapter 1: Introduction

1.1 Low molecular weight drugs: decline in drug discovery process ............. 30  
1.2 RNA targeting .................................................................. 33  
  1.2.1 Rationale, possible impact and advantages of RNA targeting ........... 33  
  1.2.2 MicroRNAs as attractive therapeutic targets ............................... 37  
1.3 Spontaneous, enzymatic and chemical catalysis of RNA cleavage ........... 38  
  1.3.1 RNA chemical stability .................................................... 38  
  1.3.2 Spontaneous RNA transesterification ...................................... 39  
  1.3.3 Base catalysis .................................................................. 40  
  1.3.4 Metal ion catalysis ............................................................. 41  
  1.3.5 RNA cleaving enzymes ...................................................... 42  
  1.3.5.1 Ribonucleases ............................................................. 42  
  1.3.5.2 Ribozymes ................................................................. 44  
1.4 Oligonucleotide-based chemical ribonucleases ..................................... 44  
  1.4.1 Metal-dependent artificial ribonucleases ................................... 47  
  1.4.2 Metal-independent artificial ribonucleases ................................ 51  
  1.4.3 Drawbacks of existing ribonucleases ....................................... 54  
  1.4.4 Peptidyl oligonucleotide conjugates ..................................... 54  
  1.4.5 Mechanism of RNA cleavage by peptidyl-oligonucleotide conjugates 60  
   1.4.5.1 Gaps in knowledge .................................................... 63  
1.5 Aims and objectives of the project ..................................................... 63  
1.6 Structure and dynamics of macromolecules by NMR ............................. 66  
  1.6.1 Introduction to NMR methods ............................................. 66  
  1.6.2 Common NMR Isotopes .................................................... 69  
  1.6.3 Chemical shifts ................................................................ 70  
  1.6.4 Spin-Spin or J-coupling .................................................... 71  
  1.6.5 Relaxation in NMR Spectroscopy ......................................... 74  
  1.6.6 Fourier Transform in NMR ................................................ 75  
  1.6.7 Multidimensional NMR .................................................... 76
1.6.8 NMR techniques ........................................................................................................... 79
  1.6.8.1 Water suppression ....................................................................................................... 79
1.6.9 COSY: Correlation Spectroscopy .................................................................................. 81
1.6.10 DQF-COSY: Double-Quantum Filtered Correlation Spectroscopy ......................... 82
1.6.11 TOCSY: Total Correlation Spectroscopy ................................................................... 83
1.6.12 NOESY: Nuclear Overhauser Effect Spectroscopy ..................................................... 84
1.6.13 HSQC: Heteronuclear Single-Quantum Correlation .................................................. 87
  1.6.13.1 CT-HSQC: Constant-time Heteronuclear Single-Quantum Correlation .......... 88
1.6.14 2D HSQC-NOESY ...................................................................................................... 88
1.6.15 2D ω1-ω2-filtered NOESY ............................................................................................ 89
1.6.16 3D HNHA ................................................................................................................... 91
1.7 Combination of NMR and Molecular Modelling for structural analysis ..................... 92
  1.7.1 Molecular Modelling ................................................................................................... 92
    1.7.1.1 Molecular Mechanics .............................................................................................. 93
    1.7.1.2 Energy Minimisation and Geometry Optimization ................................................ 93
    1.7.1.3 Molecular Dynamics .............................................................................................. 94
    1.7.1.4 Restrained Molecular Dynamics ........................................................................... 96
    1.7.1.5 Simulated Annealing and NMR assisted folding .................................................... 96
  1.7.2 Analysis ....................................................................................................................... 98
    1.7.2.1 Geometrical Analysis .............................................................................................. 98
    1.7.2.2 RMSD ..................................................................................................................... 99
    1.7.2.3 Cluster Analysis ...................................................................................................... 100
1.8 References ...................................................................................................................... 101

Chapter 2: Materials and methods

2.1 Source of materials ........................................................................................................... 115
2.2 Mass Spectrometry .......................................................................................................... 115
2.3 Synthesis of peptidyl-oligonucleotide conjugates ........................................................ 116
  2.3.1 DNA oligonucleotide ..................................................................................................... 117
  2.3.2 Solid phase peptide synthesis with Fmoc Chemistry .................................................... 117
  2.3.3 Peptidyl-oligonucleotide conjugate synthesis .............................................................. 120
  2.3.4 Synthesis of the selectively (\(^{13}\)C, \(^{15}\)N) labelled conjugates ................................ 122
2.4 Conjugate quantification ................................................................................................. 123
2.5 Ribonuclease activity assay ............................................................................................. 123
2.6 Nuclear Magnetic Resonance Spectroscopy ................................................................. 124
  2.6.1 Spectrometers ............................................................................................................. 124
  2.6.2 NMR Sample preparation ........................................................................................... 124
  2.6.3 Data acquisition .......................................................................................................... 125
    2.6.3.1 NMR spectra acquired for assignment of the unlabelled conjugate ..................... 125
    2.6.3.2 NMR spectra acquired for assignment of the labelled conjugates ....................... 127
    2.6.3.3 NMR spectra acquired at different temperature and pH ...................................... 127
  2.6.4 Data processing and analysis ....................................................................................... 128
Chapter 3: Solution structure and dynamics of the model conjugate by NMR spectroscopy

3.1 Introduction .................................................................................................................. 139
3.2 Design of the Peptidyl-oligonucleotide conjugate .................................................... 141
3.3 Design of the Peptidyl-oligonucleotide labelled conjugate .......................................... 141
3.4 Synthesis of the unlabelled and site-specifically $^{13}$C- and $^{15}$N-labelled peptidyl-oligonucleotide conjugates ........................................................................................................ 142
   3.4.1 Solid phase synthesis of the peptide NH$_2$-[Leu-Arg]-$^4$Gly-CONH$_2$ and site-specifically $^{13}$C- and $^{15}$N-labelled peptides ........................................................................................................ 143
   3.4.2 Peptide Characterization (Mass spectrometry) .......................................................... 144
   3.4.3 Coupling reaction between the oligonucleotide 5'-d(pTCAATC)-3' and peptide NH$_2$-[Leu-Arg]-$^4$Gly-CONH$_2$ ........................................................................................................ 145
   3.4.4 Conjugate Characterization (Mass spectrometry) ....................................................... 146
3.5 Cleavage of 5$'$-[$^{32}$P]-microRNA-24 by the peptidyl-oligonucleotide conjugate ................. 147
3.6 NMR spectroscopy and structure determination of the conjugate ................................. 151
   3.6.1 Preliminary NMR studies on the conjugate ............................................................... 151
   3.6.2 Optimisation of solution conditions for NMR experiments .................................... 153
   3.6.3 NMR strategies for nucleic acid resonance assignments ......................................... 158
      3.6.3.1 Identification of base protons and sugar proton spin systems ............................... 158
      3.6.3.2 Assignment of the exchangeable proton resonances .......................................... 160
      3.6.3.3 Sequential backbone assignment ....................................................................... 161
      3.6.3.4 Resonance assignments from $^{31}$P correlated experiments ................................. 167
      3.6.3.5 Nucleic acid chemical shifts ............................................................................. 168
   3.6.4 NMR strategies for conjugate peptide resonance assignments .................................. 170
      3.6.4.1 Backbone resonance assignment ..................................................................... 170
      3.6.4.2 Sequential backbone assignment .................................................................... 173
      3.6.4.3 Side chain assignments .................................................................................. 174
      3.6.4.4 Peptide medium and long-range assignments .................................................. 179
   3.6.5 Identification of intramolecular NOEs .................................................................... 180
   3.6.6 Chemical shift temperature dependence ................................................................. 183
   3.6.7 Assignment of the peptide protons assisted with isotopic labelling ......................... 185
   3.6.8 Solution structure of the conjugate ...................................................................... 195
      3.6.8.1 Experimental restraints ............................................................................... 195
Chapter 4: miRNases: Novel peptide-oligonucleotide bioconjugates that silence miR-21 in lymphosarcoma cells

4.0. DECLARATION

4.1 Abstract

4.2 Introduction

4.3 Material and Methods

4.3.1 Instrumentation

4.3.2 Oligonucleotides

4.3.3 Peptide synthesis and purification

4.3.3.1 Acetyl-(LeuArg)-Gly-(LeuArg)-Gly-COOH

4.3.4 Synthesis of the conjugates

4.3.5 Conjugate purification

4.3.6 Characterisation of conjugates

4.3.6.1 Conjugate 5’-16

4.3.6.2 Conjugate 5’-h-6/14

4.3.6.3 Conjugate 5’-h-9/14

4.3.6.4 Conjugate 5’-h-6/16

4.3.6.5 Conjugate 5’-h-9/16

4.3.6.6 Conjugate 5’-luc-h-9/14

4.3.7 5’-RNA labelling

4.3.8 Gel-retardation assay

4.3.9 Ribonuclease activity assay

4.3.10 Molecular Modelling

4.3.11 Transfection of tumour cells with conjugates

4.3.12 qPCR

4.3.13 Western Blot

4.3.14 MTT-test and doubling time determination

4.3.15 Statistics

4.4 Results

4.4.1 Design of miR-21 targeted oligonucleotides and peptide-oligonucleotide conjugates

4.4.2 Selection of oligonucleotides for POCs synthesis

4.4.3 POCs synthesis and characterization

4.4.4 Hybridization properties of POCs

4.4.5 Site-selective ribonuclease activity of POCs

4.4.6 Ribonuclease activity of control non-complementary POCs against miR-21

4.4.7 Molecular modelling

4.4.8 Effect of conjugate 5’-h-9/14 on the level of miR-21 in lymphosarcoma cells RLS40
4.4.9 Effect of conjugate 5'-h-9/14 on the level of miR-21 protein target PDCD4 in RLS40 lymphosarcoma cells ........................................................ ............................................................ 243
4.4.10 Effect of conjugate 5'-h-9/14 on proliferative potential of RLS40 lymphosarcoma cells...245
4.5 Discussion ........................................................................................................................................................................... 246
4.6 Conclusion ......................................................................................................................................................................... 252
4.7 Acknowledgements .......................................................................................................................................................... 253
4.8 References ....................................................................................................................................................................... 254

Chapter 5: Conclusion and future work

Conclusions and future work ................................................................................................................................. 262
Appendix A. Atoms nomenclature for the peptide and nucleotides .............................................................. 265
Appendix B. 13C- and 15N-labelled peptides and conjugate characterization (Mass spectrometry) ........................................................................................................................................... 267
Appendix C. Amber 14 Files - Simulated Annealing - chirality restraints .............................................. 272
Appendix D. Supplementary material for Chapter 4 ......................................................................................... 276

Total word count: 62546
List of Figures

Chapter 1

Figure 1.1 Illustration of the decline in Drug Discovery and Development of small-molecule drugs. Adapted from Scannell et al.1 ............................................................... 31

Figure 1.2 Shift of Big Pharma to include large molecules. Adapted from Otto et al.6 ............... 32

Figure 1.3 Mechanism of protein expression arrest by antisense nucleotides: inside the cell the antisense oligonucleotide aligns with the target mRNA via base complementarity. The heteroduplex formed in this way halts protein synthesis by interfering with a variety of processing pathways involved in either mRNA maturation or translation into proteins. Adapted from researchers6,55 ............................................................. 35

Figure 1.4 Breakdown of the RNA internucleotide linkage: (a) Internal phosphoester transfer by 2'-OH nucleophilic attack on 3'-phosphorus atom. (b) Formation of the penta-coordinated phosphorane intermediate. (c) Cleavage products. (d) Subsequent hydrolysis of the 2', 3'-cyclic monophosphate into a mixture of 2'- and 3'-nucleotide monophosphates53 .......................................................... 40

Figure 1.5 Scission of RNA phosphodiester bond promoted by metal ions (Mn+) and their hydrated forms58 .......................................................... 41

Figure 1.6 Reaction pathway of RNA cleavage catalysed by ribonuclease A: a) The His12 in the active site abstracts a proton from the RNA ribose 2'-OH activating the oxygen nucleophilic attack on the phosphorus. b) The negative charge builds up on the transition state phosphate stabilised by positively charged residues Lys34 and Phe120 and departure of the adjacent ribose 5'-OH promoted by donation of a proton from His119 allow formation of the cyclic nucleotide intermediate. c) The general acid-base activity of the His12 and His319 side chain imidazole rings promote hydroxyl attack on the 2', 3'-cyclic nucleotide monophosphate to generate either 2'- or 3'- nucleoside monophosphates ................................................................. 43

Figure 1.7 Schematic representation of Lanthanide ion based chemical ribonuclease structures: (1) Lutetium (Lu3+) ion complexed with chelate conjugates containing negatively charged carboxylate groups86; (2) Thorium (IV) ion complexed to DOTA derivative87; (3) Europium (Eu3+) macrocyclic conjugate containing a pyridine cyclophane89; (4) Dysprosium (Dy3+) thexaphyrin attached to the oligonucleotides via an internally positioned glyceryl linker88 .......................................................... 49

Figure 1.8 Structures of artificial ribonucleases based on Zn2+ ion: (5) dinuclear Zn (II) complex of triazacyclododecane derivative92; (6) mononuclear Zn (II) complex of 2,6-bis[2-(guanidiniumethyl)pyrimidin-4-yl]pyridine94 .......................................................... 50

Figure 1.9 Structures of the oligonucleotide conjugates of Cu2+ and Zn2+ chelates: (7) Cu (II)-terpyridine conjugate; (8) and (9) Zn (II)-1,4,7-triazacyclododecane conjugates95 .......................................................... 51
Figure 1.10 Schematic representation of the chemical ribonucleases based on metal independent catalytic structures: (10) diethylenetriamine-peptide nucleic acid conjugate\(^9\), (11) bis (guanidinium) derivative\(^8\), (12) imidazole based construct-oligonucleotide conjugate\(^\)………………… 53

Figure 1.11 Tentative mechanism suggested by Barbier and Brack\(^\) of an oligoribonucleotide hydrolysis by poly (Leu-Lys)………………………………………………………………. 56

Figure 1.12 Structure of the peptidyl-oligodeoxyribonucleotide conjugates\(^10\),\(^12\)………………… 56

Figure 1.13 Schematic representation of the cleavage site of the oligonucleotide-peptide conjugate in tRNA\(^3\) upon perfect base-pair matching\(^8\)…………………………………………………………………………… 58

Figure 1.14 Chemical structure of the peptide [(LeuArg)\(^2\)Gly\(^2\)] with C-terminal carboxamide modification attached to the 5'-end phosphate of the 17-mer antisense oligonucleotide\(^\)………………… 59

Figure 1.15 Structure of the dual conjugate showing the peptide [(LeuArg)\(^2\)Gly\(^2\)] (in purple) modified at the C-terminus conjugated to two flanking oligonucleotide recognition motif (black) along with added linker at the 3’end\(^\)………………… 60

Figure 1.16 Catalytic mechanism of two guanidinium groups\(^\) intramolecular cooperation in the cleavage of the phosphodiester bond in the RNA model\(^11\): the guanidinium unit acts as an electrophilic activator by stabilising the four-membered cyclic intermediate via hydrogen bonding with the phosphate group, and a neutral guanidinium unit acts as a general base to aid deprotonation of water to use as a nucleophile to attack the phosphate unit………………………………………………………………………………………………………………… 62

Figure 1.17 Spin states: In the presence of a static magnetic field \(B_0\) nuclear spin with a magnetic moment may adopt \(2I+1\) possible orientations relative to \(B_0\). For spin-\(\frac{1}{2}\) nuclei, there are two possible orientations, either parallel (\(\alpha\)) or antiparallel (\(\beta\))…………………………………… 67

Figure 1.18 Larmor precession. The nuclear spin precesses about the magnetic field \(B_0\) axis with an angular frequency, known as the Larmor frequency, which is dependent on the magnetic field strength and the spin magnetogyric ratio………………………………………………………………………………………………………………… 68

Figure 1.19 Excitation with a rf pulse: The application of a rf field \(B_1\) (for example along the \(x\) axis), rotates the spin vectors from the \(z\) axis towards the \(x\)-\(y\) plane. The vectors precess about the \(z\) axis at their Larmor frequency. The rotating magnetization vector induces voltage in the coil surrounding the sample; these are the electrical signals detected to produce the observed NMR signals………………………………………………………………………………………………………………… 69

Figure 1.20 (a) the phi (\(\phi\)) dihedral angle describes the rotation of the polypeptide backbone about the N-C\(_\alpha\) bond and involves the C(O)-N-C\(_\alpha\)-C(O) bonds. (b) Karplus curve describing the variation of \(3J_{\text{HNI\text{H}a}}\) with backbone dihedral angle \(\phi\)………………………………………………………………………………………………………………… 73

Figure 1.21 Longitudinal relaxation (\(T_1\)) in the rotating frame representation. The net magnetisation relaxation diminishes the \(y\)-component and re-establishes the \(z\)-component, until magnetisation comes to its equilibrium position finally along the \(z\) axis………………………………………………………………………………………………………………… 74

Figure 1.22 Transverse relaxation (\(T_2\)). Local molecular environment differences cause spins to precess with slightly different Larmor frequencies, diminishing the \(x\) and \(y\) components and eventually leading to magnetisation loss in the \(x\)-\(y\) plane………………………………………………………………………………………………………………… 75
Figure 1.23  Schematic diagram of a time-domain NMR signal (FID signal) and its equivalent frequency NMR spectrum on Fourier Transform………………………………………… 76

Figure 1.24  Diagram of 2D pulse sequence of a FT-experiment………………………………………… 77

Figure 1.25  Left: 2D NMR spectrum as a function of time (t1 and t2). Middle: the data upon t2 Fourier transform into frequency ω2. Right: 2D NMR spectrum upon t1 transformation to ω1 as a contour plot (concentric circles represent peak intensity)…………………………………… 77

Figure 1.26  3D pulse sequence. 2D experiments are combined to create a 3D experiment (i.e. 3D HNHA). In a 3D experiment the first two evolution periods, which are incremented, correspond to the two indirectly-detected time domains (t1 and t2), and the third, which is the acquisition time, corresponds to the directly-detected time domain (t3). After a Fourier transform (FT) of the three time domains the 3D spectrum has three frequencies, ω1 (1H), ω2 (X) and ω3 (1H)………………………………………………… 78

Figure 1.27  Solvent suppression methods. (a) Presaturation, involves the application of a long, low-power pulse on the solvent frequency prior to the pulse sequence. (b) WATERGATE, uses a gradient spin-echo sequence which combines a single selective 180° pulse and two pulsed field gradients to dephase the net solvent magnetisation. (c) Excitation sculpting, applies a double gradient spin-echo sequence……………………………………………………………………… 80

Figure 1.28  2D COSY. Comprises of a two-pulse sequence with an evolution period lying in between. This experiment correlates the chemical shift of nuclei of the same kind through magnetisation transfer between scalar J coupled nuclei following the second 90° pulse…………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………
Figure 1.33 2D HSQC-NOESY. The pulse sequence combines HSQC and NOESY experiments. The first two steps are an HSQC preparation followed by evolution periods, during which magnetisation is transferred to the directly-bonded nuclei via one-bond $J_{HX}$ coupling. The last step consists of a NOESY mixing period, where magnetisation transfer between closely spaced protons occurs via dipolar coupling.

Figure 1.34 2D $\omega_1$-X-filtered NOESY. The pulse sequence consists of a NOESY sequence in which the first 90° pulse is replaced by an X half-filter. This experiment removes signals of the protons directly-bonded to NMR active heteronuclei X.

Figure 1.35 3D HNHA. In the initial part of the pulse sequence the flow of magnetisation starts at $^1\text{H}^\text{n}$ after the first 90° pulse on the proton and is transferred to $^{15}\text{N}$. The second 90° $^1\text{H}$ pulse transfers the $^1\text{H}^\text{n}$ antiphase magnetisation to $^1\text{H}^\text{a}$ coherence, during which $^{15}\text{N}$ evolution $t_1$ takes place. The magnetisation transfer from the $^1\text{H}^\text{n}$ to the $^1\text{H}^\text{a}$ coherences is proportional to the length of the delay period $\delta_2$.

Figure 1.36 Graphical illustration of the Leapfrog-like Verlet algorithm in AMBER and its characteristic scheme of integration. The coordinates and velocities are updated using the same time step, but the values are shifted by half a time step.

Chapter 2

Figure 2.1 Peptidyl-oligonucleotide conjugates sequence and nomenclature. Left: Sequence of the 6-mer oligonucleotide and the 9-mer peptide linked at the 5' and at the N terminus respectively within the conjugate via a phosphoramidate bond. Right: Nomenclature of the conjugates used in the structure determination. The amino acids leucine (Leu) and arginine (Arg) labelled with $^{13}\text{C}$- and $^{15}\text{N}$ in the four site-specifically labelled conjugates are underlined and shown in green.

Chapter 3

Figure 3.1 Chemical structure of the peptidyl-oligonucleotide conjugate NH$_2$CO-Gly-[Arg-Leu]-NH-5'-d(pTCAA TC)-3'.

Figure 3.2 MS-analysis (MALDI-TOF) of the unlabelled peptide 1. Formula C$_{50}$H$_{96}$N$_{22}$O$_9$, molecular species observed [M-H]$^+$, 1150.81, found 1150.80.

Figure 3.3 MS-analysis (MALDI-TOF) of the unlabelled conjugate S1. Formula C$_{108}$H$_{170}$N$_{42}$O$_{45}$P$_6^-$ [M +H]$^+$, 2962.10, found 2962.10.

Figure 3.4 Analysis of miR-24 cleavage by the conjugate. (a) Autoradiograph of 12% denaturing polyacrylamide/8M urea gel after catalytic hydrolysis. Lanes Im and T1: RNA cleavage by imidazole buffer and partial RNA digestion with RNase T1, respectively. Lanes C1 and C2 (incubation control): RNA incubated in the absence of conjugate for 0 and 48 hours. Lanes 0-48, incubation of miR-24 in the presence of the conjugate (40-fold excess) at 37°C for 0, 3, 5, 8, 24 and 48 hours, respectively. (b) Sites of cleavage of miR-24 induced by the conjugate (indicated by arrows). (c) Kinetics of miR-24 cleavage by the conjugate as a function of time.
Figure 3.5  (a) 1D $^1$H-NMR spectrum of the peptidyl-oligonucleotide conjugate (2 mM in 90% water and 10% D$_2$O recorded at 800 MHz and 5°C) showing the different regions of the oligonucleotide (in black) and peptide resonances (in blue). The region around the downfield-shifted resonances of the amino acid amide protons are enlarged in the inset (b).  

Figure 3.6  2D NOESY spectrum of the conjugate showing H1’ and aromatic regions of oligonucleotide component (5.9-8.3 ppm) against the aliphatic region of the peptide (0.8-1.8 ppm) in D$_2$O. The cross peaks in this region represent intramolecular NOEs exclusively between the nucleic acid and the peptide side chain non-exchangeable protons.

Figure 3.7  600 MHz $^1$H variable temperature series of the unlabelled conjugate (1 mM), dissolved in 90% water and 10% D$_2$O (pH 6.5). The A-3 amine protons exchange slowly with water over the temperature range; the cytosine –NH$_2$ groups and the 3’-OH peaks are very broad and become visible at lower temperatures.

Figure 3.8  NMR-based titration curves recorded for the oligonucleotide aromatic protons of the unlabelled conjugate. A model with two pK$_a$’s was selected to obtain pK$_a$ values of the bases. All the bases exist in their physiological charge state in the range of pH 6-9.

Figure 3.9  1D $^1$H NMR spectra of the conjugate in the amide and aromatic region acquired at different pH values (4 to 10). The aromatic proton resonances upfield shifts are observable at acidic pH. The peptide amide protons chemical shifts are not affected at the titration pH range.

Figure 3.10 Assignment of the base H5 (or -CH$_3$) and H6 resonances of cytidine or thymine residues, respectively, of the conjugate nucleic acid, in an 800 MHz $^1$H-$^1$H TOCSY spectrum (2 mM, D$_2$O at pH 6.5, 25 °C). Note that the H5 and H6 of the cytosine C$_2$ and C$_6$ are the only aromatic scalar-coupled protons, hence identified by the two strong aromatic cross peaks, split in the $^1$H o$_2$ dimension. The weak cross peaks connecting the aromatic region to the up-field aliphatic region are the result of the four-bond coupling between thymine residues methyl (H7*) and H6 protons. The right hand plots show the 800 MHz $^1$H-13C HSQC H7*-C7 and H5-C5 regions for cytosine and thymine residues.

Figure 3.11 Assignment of the exchangeable amino protons of two adenine (A$_3$ and A$_4$ H61/H62) and cytosine (C$_2$ and C$_6$ H41/H42) nucleobases in a 800 MHz $^1$H-$^1$H NOESY spectrum (2 mM, H$_2$O 90%, D$_2$O 10%, at pH 6.5, 5 °C). The top and right plots represent sections of $^1$H-$^{15}$N HSQC spectra showing the $^{15}$N and base amino proton correlations of adenine and cytosine respectively.

Figure 3.12 Sections of 2D TOCSY spectrum (2 mM, H$_2$O 90%, D$_2$O 10%, at pH 6.5, 5 °C) showing the H41 and H42 correlations of the cytosine bases C$_2$ and C$_6$ (6.5-7.3 ppm), H61/H62 and H2 correlations of adenine bases A3 and A4 (6.7-7.9 ppm), and 3’-terminal C$_6$ hydroxyl proton correlations to its own sugar H3’ and H4’ (4.1-6.5 ppm).

Figure 3.13 Expanded NOESY spectrum (300 ms mixing time, D$_2$O at pH 6.5, 25 °C) of the conjugate bases to H1’ region. A representative sequential NOE walk, base protons H6/H8 and the sugar H1’ protons from the 5’end (T$_1$) to the 3’end (C$_6$) are shown as a green trace. The cytidine C$_2$ and C$_6$ base H5-H6 cross-peaks, which are the most intense ones, are connected to their corresponding H1’-H6 cross peaks by dashed lines for clarity.
Figure 3.14  Regions (a) and (b) of the 2D NOESY spectrum. (a) drawn lines connect cross peaks involving intranucleotide contacts base (i)-H2', H2'' and internucleotide contacts H2'' (i)- base (i+1), included those of 3'-terminal nucleotide C6. The H2' and H2'' of the adenines resonate at lower field as a result of the deshielding effect of the stronger aromatic ring current of purines. Region (b) shows the base-H3' cross peaks.

Figure 3.15  Section of DQF-COSY spectrum showing the cross peaks linking H1' with H2' and H2'' (2 mM, D2O at pH 6.5, 25 °C; 800 MHz). The differences between the coupling constants between the 2’CH2 protons and H1’ is apparent in the cross peak multiplet splitting pattern. For A4 the multiplet pattern of the single peak shows that H2' and H2'' resonate at very similar frequency. The large JH1'-H2' coupling and a small JH1'-H2'' coupling is indicative of a C2'-endo conformation of the sugar pucker.

Figure 3.16  Pictorial representation of intranucleotide and sequential NOE connectivities of the oligonucleotide fragment. The intraresidue and sequential short 1H-1H distances from the 5’ to 3’ direction (indicated by black arrows and red arrows, respectively), are used for sequential assignments of the non-exchangeable protons to confirm the oligonucleotide sequence.

Figure 3.17  Expanded plot of the region containing H3', H4', H5' and H5'' of a 1H-1H NOESY spectrum obtained under the same conditions as Figure 3.14.

Figure 3.18  A representative 31P-1H HSQC spectrum of the conjugate (2 mM, D2O at pH 6.5, 25 °C, 400 MHz) showing the backbone 31P resonances versus the H3'/H4'/H5', H5'' region.

Figure 3.19  Section of the 15N HQSC spectrum of the amide protons recorded for the unlabelled conjugate (800 MHz, 5°C). The cross peak of Leu7 in the inset is located outside the main plotted region of the spectrum and appears at $\omega_1 =$133.09 ppm and $\omega_2 =$3.63 ppm. HN-15N cross-peaks have been labelled with their correlating residue number.

Figure 3.20  The top two panels a) show the correlations 1H-1H “fingerprint region” in the 800 MHz TOCSY spectrum of the conjugate (2 mM, H2O 90%, D2O 10%, at pH 6.5, 5 °C). The Gly15 Hα and Hα2 are connected by a horizontal line. The residue number next to cross-peaks indicates 1H and 1H resonance assignments. b) 1H 13C-HSQC spectrum showing the 13C-1H region.

Figure 3.21  Expanded region of the NOESY spectrum of the conjugate acquired at 5°C and pH 6.5, with a mixing time of 300 ms. a) NOE connectivities between the amide protons HN (8.2-8.7 ppm region) from Leu to Gly15. The amide proton assignments are listed along the diagonal and their sequential connectivities are traced by dotted lines. The long-range amide –amide NOEs between Leu6 and Leu11 are labelled with a **, and the NOE between Leu7, HN and Leu6 HN with *. b) NOE connectivities between the HN and Hα (3.6-4.4 ppm) protons. Sequential NOEs to neighbouring amino acids (i +1) are indicated by dotted lines. Intra-residue cross peaks are labelled with the corresponding residue numbers according to the assignments based on sequential NOEs.
Figure 3.22  Expanded region of the TOCSY spectrum of the conjugate (2 mM, D<sub>2</sub>O at pH 6.5, 25 °C, 800 MHz) showing the correlation between α-protons and leucine and arginine side chain protons. The cross peaks of the arginine and leucine H<sup>α</sup>, H<sup>β</sup> and H<sup>γ</sup> appear heavily overlapped and the leucine H<sup>α1</sup> and H<sup>β2</sup> resonances were not distinguishable but instead appear as a single unresolved peak.  

Figure 3.23  a) ct-H<sup>13</sup>C-HSQC spectrum (H<sub>2</sub>O 90%, D<sub>2</sub>O 10%, at pH 6.5, 5 °C, 26 ms, 800 MHz) of the leucine and arginine side chain protons showing similarity of the chemical shifts. Individual labelling of proton resonances is not possible due to severe signal overlap. Distinct regions of each type of proton are identified by inverted signal sign of peaks arising from <sup>13</sup>C<sup>3</sup>-J-coupled to an even number of carbon atoms. b) TOCSY spectrum (H<sub>2</sub>O 90%, D<sub>2</sub>O 10%, at pH 6.5, 5 °C) which correlates the amide protons to the arginine and leucine side chains. 

Figure 3.24  The expanded region of the TOCSY spectrum of the conjugate (H<sub>2</sub>O 90%, D<sub>2</sub>O 10%, at pH 6.5, 5 °C, 800 MHz) which correlates H<sup>α</sup> protons of the four arginine residues with the other protons from the same spin system. 

Figure 3.25  The expanded region of the TOCSY spectrum of the conjugate (H<sub>2</sub>O 90%, D<sub>2</sub>O 10%, at pH 6.5, 5 °C, 800 MHz) which correlates the arginine H<sup>α1</sup> to H<sup>β2</sup>. Selective assignment of these protons to the individual arginine residues is prevented due to considerable overlap of their broad signals. The label Arg<sup>a</sup> indicates either Arg<sub>8</sub>, Arg<sub>10</sub>, Arg<sub>12</sub> and/orArg<sub>14</sub>. 

Figure 3.26  Sections of the NOESY spectrum (H<sub>2</sub>O 90%, D<sub>2</sub>O 10%, at pH 6.5, 5 °C, 300 ms mixing time, 800 MHz) showing NOE cross-peaks between the peptide protons and nucleotide protons, located far apart in the conjugate sequence. The summary of these interactions is given in Table 3.3. 

Figure 3.27  Plots of the correlation between the backbone amide proton chemical shift and temperature for the peptide residues from Leu<sub>7</sub> to Gly<sub>15</sub>. The residue amide proton chemical shift temperature coefficient derived from <sup>1</sup>H chemical shift measured at 5, 10, 20, 25, 30 and 35 °C. The best-fit lines are indicated. 

Figure 3.28-a<sup>H</sup>  <sup>13</sup>C-HSQC spectra (2 mM, D<sub>2</sub>O at pH 6.5, 25 °C, 800 MHz) of the samples contained <sup>13</sup>C, <sup>15</sup>N labelled arginine and leucine residues introduced at positions: Leu<sub>7</sub>Arg<sub>8</sub> (a) and Leu<sub>8</sub>Arg<sub>10</sub> (b). The <sup>1</sup>H-<sup>13</sup>C correlations seen in each spectrum could be assigned to the labelled residues. 

Figure 3.28-b<sup>H</sup>  <sup>13</sup>C-HSQC spectra (2 mM, D<sub>2</sub>O at pH 6.5, 25 °C, 800 MHz) of the samples contained <sup>13</sup>C, <sup>15</sup>N labelled arginine and leucine residues introduced at positions: Leu<sub>11</sub>Arg<sub>12</sub> (c) and Leu<sub>12</sub>Arg<sub>14</sub> (d). 

Figure 3.29  Expanded regions of the 2D <sup>1</sup>H <sup>13</sup>C HSQC-NOESY spectra (2 mM, D<sub>2</sub>O at pH 6.5, 25 °C, 800 MHz) recorded for the four <sup>13</sup>C, <sup>15</sup>N labelled samples Leu<sub>7</sub>Arg<sub>8</sub> (S2) (a), Leu<sub>8</sub>Arg<sub>10</sub> (S3) (b), Leu<sub>11</sub>Arg<sub>12</sub> (S4) (c), Leu<sub>13</sub>Arg<sub>14</sub> (S5) (d). Leu<sub>7</sub>Arg<sub>8</sub> conjugate (a) and Leu<sub>8</sub>Arg<sub>10</sub> (b) display the larger number of NOE contacts as compared with Leu<sub>11</sub>Arg<sub>12</sub> (c) and Leu<sub>13</sub>Arg<sub>14</sub> conjugates (d), respectively, showing the gradual decrease in number of NOE interactions between distal oligonucleotide and peptide residues within the conjugate. No intra-molecular NOEs were observed between Leu<sub>11</sub> and the oligonucleotide part of the conjugate.
Figure 3.30  Expanded region of the 2D $^1$H $^{13}$C HSQC-NOESY spectrum (2 mM, D$_2$O at pH 6.5, 25 °C, 800 MHz) recorded for the $^{13}$C, $^{15}$N labelled [Leu$_9$Arg$_{10}$] conjugate, showing the two NOE cross peaks between Leu$^7$ and Arg$_{10}$ in the spectral region of strong interference from direct $^1$H-$^{13}$C correlation peaks.

Figure 3.31  a) Expanded region of the 2D $^1$H $^{15}$N HSQC-NOESY spectra (H$_2$O 90%, D$_2$O 10%, at pH 6.5, 5 °C, 300 ms mixing time, 800 MHz) recorded for $^{13}$C, $^{15}$N labelled [Leu$_9$Arg$_{10}$] conjugate, which displays the assigned sequential NOE connectivities between amide protons and their own and sequential H$^\alpha$, H$^\delta$ protons. b) The residues from a) are shown in red for the arginine and green for the leucine residues, respectively, and yellow for T$_1$ with the H$^N$-H$^N$ and H$^N$-H$^\alpha$ highlighted.

Figure 3.32  Part of the 2D $^1$H $^{15}$N HSQC-NOESY spectrum showing the guanidine NH$^\eta$ signals of Arg$_{10}$ in [Leu$_9$RArg$_{10}$] conjugate (S3) recorded at 800 MHz at 5°C. The four resolved signals detected for the NH$^\eta$ nuclei in Arg$_{10}$ are labelled. The cross peaks show the correlations between protons in the two NH$^\eta$ groups, the NOE correlations between guanidino protons and their own H$^\delta$, as well as NOE correlations between the H$^\epsilon$ and their own H$^\delta$ and H$^\beta$.

Figure 3.33  Summary of short, medium and long range NOEs for the peptide. The thickness of the vertical black bars indicates the relative intensities of the NOE connectivities. Below the NOE patterns, the temperature coefficients of the amide protons are given in units 10$^{-9}$ ppm/K and the experimental $^3$J$_{HNH^\alpha}$ are shown.

Figure 3.34  Solution NMR structure of the peptidyl-oligonucleotide conjugate (a) Views of the backbone traces for the refined twenty structures overlaid on the backbone heavy atoms in the secondary structure elements. The peptide backbone in ribbon representation is blue in the tight turn of the segment [Arg$_8$-Leu$_{11}$] and gray for the extended backbone following Arg$_{12}$. The oligonucleotide backbone is shown in orange, the bases in yellow and the phosphates in red. The average root-mean square deviations (RMSD) are as follows: all 1.33 Å for all non-hydrogen atoms, 0.963 Å for the backbone, peptide backbone excluding C-terminal tail 0.665 Å. (b) Organisation of the lowest energy conformation of the conjugate NMR structure ensemble represented in stick and ribbon. (c) Surface (heavy atoms of the oligonucleotide) and stick representation of the lowest energy structure. The oligonucleotide surface is painted according to location of the hydrophobic and hydrophilic groups with red indicating the phosphates and the base polarizable sites and green indicating the hydrophobic regions. Surface transparency is used to convey the interactions between the positively charged arginine residues and the oligonucleotide phosphates. (d) View of the same structure as (c) at a 180° rotation round the y-axis.

Figure 3.35  Ramachandran plots showing the backbone conformation of the ensemble of twenty refined structures. 98.3% of the peptide residues lie in the most favoured regions and 1.7 % in the allowed regions. The $\phi$ and $\psi$ values for the lowest energy conformation of the conjugate NMR structure ensemble are shown in yellow.

Figure 3.36  Interactions between the peptide and oligonucleotide components of the conjugate. Views of the residue contacts in the overlay of the 20 refined NMR structures (left), single structures showing the most frequent observed salt bridges/hydrogen bonds (right; see Table 3.9).
Figure 3.37 The most frequent hydrogen bonding occurring within the conjugate peptide in the 20 refined NMR structures. \( H^N-H^N \) interactions are shown in green and the \( H^N-H^\eta \) in orange.  

Figure 3.38 Representative conformations of the four clusters identified in the 1\( \mu \)s simulations: a) Cl1, b) Cl2, c) Cl3 and d) Cl4.  

Figure 3.39 Alternative orientations observed for Arg12, Arg10, Arg12 and Arg14 in the four cluster representative structures a) Cl1, b) Cl2, c) Cl3 and d) Cl4. Dashed lines indicate salt bridges and hydrogen bonds between these arginines and the conjugate oligonucleotide.  

Figure 3.40 Time series of heavy atom distances of the most frequent intra-molecular interactions in the conjugate simulation. (P)-(CZ/N) distance (d) \( \leq 4.5 \text{ Å} \): red, \( 4.5 \leq d \leq 5.5 \text{ Å} \): green \( 5.5 \leq d \leq 6.5 \text{ Å} \): blue, \( d \geq 6.5 \text{ Å} \): yellow; (O)-(N) distance (d) \( \leq 3.5 \text{ Å} \): red, \( 3.5 \leq d \leq 4.5 \text{ Å} \): green \( 4.5 \leq d \leq 5.5 \text{ Å} \): blue, \( d \geq 5.5 \text{ Å} \): yellow. The topmost three plots show the Arg12 and Arg10 side chain interactions with the oligonucleotide phosphates. The three central plots show the peptide backbone/T1 phosphate salt bridges of Arg8, Arg9 and Arg10 and the three bottom plots show the most frequently occurring \((i,i+4)\) and \((i,i+3)\) hydrogen bonds within the peptide backbone.  

Figure 3.41 Ramachandran plots of the distribution of the \((\phi, \psi)\) backbone dihedrals for the conjugate peptide nine residues in the four clusters. Residues Leu7 to Leu11 are in the \( \alpha \)-helix region in all four clusters; residues Arg12 to Arg14 lie in the \( \beta \)-sheet region in Cl1/Cl2 and in \( \alpha \)-helix region in Cl3/Cl4 (axes range from \(-180^\circ \) to \( 180^\circ \)). The N-terminal Gly15 is found in six different regions in each cluster.  

Chapter 4  

Figure 4.1 The structure of anti-miR-21 peptide-oligonucleotide conjugates (POCs). A. Schematic representation of a general structure of anti-miR-21 POCs. Labels ‘10-16 n’, ‘16 n’ or ‘14-16 n’ indicate the length ranges of the antisense ‘recognition’ sequence used in different conjugates. Label ‘6-9 n’ indicates the length range of the stem. B-C. The chemical structure of the peptide (\textit{pep}) acetyl-(LeuArg)\textsubscript{2}-Gly-(LeuArg)\textsubscript{2}-Gly-COOH. The peptide was conjugated via the carboxylic group of the C-terminus to the aminohexyl linker, which was attached to either the 5’- or 3’-terminal phosphate of the antisense oligonucleotide (asON) (B and C, respectively). Amino acids are designated by numbers which are used for the interpretation of the data from molecular dynamics simulation.  

Figure 4.2 Hybridization of oligonucleotides with 5’-[\textsuperscript{32}P]-miR-21 (\textit{*miR-21}). Autoradiograph of 12% native PAGE. \textit{*miR-21} (1\( \mu \)M) and one of the oligonucleotides (1 \( \mu \)M) were incubated at 37°C for 1 h. The samples were loaded onto the running gel immediately after quenching of the reaction with an interval of 2 min.
Figure 4.3 Hybridization of POCs and corresponding addressing oligonucleotides with *miR-21. 
A. Autoradiograph of 12% native PAGE. *miR-21 (1 µM) and conjugates or oligonucleotides (5 µM) were incubated at 37°C for 20 min. The samples were loaded onto the running gel immediately after quenching of the reaction with an interval of 2 min. B. The concentration dependency of binding of oligonucleotide 5'-h-9/14 or conjugate 5'-h-9/14 with miR-21. *miR-21 (1 µM) and oligonucleotide or conjugate in different concentrations were incubated at 37°C for 20 min……………………………..

Figure 4.4 Kinetic and concentration analysis of *miR-21 cleavage by 5'-conjugates. A. Autoradiograph of 18% polyacrylamide/8 M urea gel. Lanes Im and T1, imidazole ladder and partial RNA digestion with RNase T1, respectively. Control – RNA incubated in the absence of conjugate for 0 and 72 h. *miR-21 (1 µM) and 5'-conjugates (20 µM) were incubated at 37°C for 0 – 72 h. The conjugate type and incubation time are shown at the top. B. Positions of miR-21 cleavage induced by 5'-conjugates 5'-h-6/16 and 5'-h-6/14 (indicated by arrows). C. Kinetics of *miR-21 cleavage by conjugates. *miR-21 (1 µM) and one of the conjugates (20 µM) were incubated at 37°C for 0 – 72 h. D. Concentration dependence of *miR-21 cleavage at the time point 24 h. *miR-21 (1 µM) and conjugates (1 – 50 µM) were incubated at 37°C for 24 h…………………………………………………………………………………..

Figure 4.5 Concentration analysis of miR-21 cleavage by complementary conjugate 5'-h-9/14 and non-complementary conjugate 5'-luc-9/14. A. Autoradiograph of 18% polyacrylamide/8 M urea gel. Lanes Im and T1, imidazole ladder and partial RNA digestion with RNase T1, respectively. C – RNA incubated in the absence of conjugate for 24 h. *miR-21 (1 µM) and conjugates (1 – 20 µM) were incubated at 37°C for 24 h. B. Concentration dependency of *miR-21 cleavage at the time point 24 h.………………………………………………………………………………………………………..

Figure 4.6 Results of the molecular dynamics simulation of the 5'-h-6/14/miR-21 complex. A, B. Structure of 5'-h-6/14 complex with miR-21 after 1 µs of molecular dynamics simulation in implicit solvent, viewed from two alternative orientations. For 5'-h-6/14, oligonucleotide in white, Leu residues in yellow, Gly in white and Arg by atom type. For miR-21 residues 1-14 in pink; for residues 15-22, Ade in purple, Ura in green and Gua in mauve. The two parts of the system folded together in a compact spherical-like shape at the end of the double helix due to the network of electrostatic interactions involving polar groups in the two parts of the system: salt bridges between the arginine guanidino and RNA phosphate groups, and hydrogen bonds involving arginine side chains, base polar atoms and sugar hydroxyls. C. Distribution of angle τ over the 1 µs simulation. D. O2' [G15]···P[A16] distance d over the 1 µs simulation…………………………………………………………………………………………………………………..
Figure 4.7 Biological activity of the conjugate 5'-h-9/14 in RLSa0 lymphosarcoma cells. Expression level of miR-21 (A) and let7-g (B) in RLSa0 cells after transfection with 1 μM of antisense oligonucleotide 5'-h-9/14, control conjugate luc-h-9/14 and anti-miR-21 conjugate 5'-h-9/14. Transfection was performed using Lipofectamine 2000. LF – RLSa0 cells incubated with Lipofectamine 2000 only. The expression of miRNAs was normalized to U6 and Rpl30. C. Western blot analysis of PDCD4: 48 h after transfection. GAPDH served as an internal control. 1 – intact RLSa0 cell; 2 – RLSa0 cells incubated with Lipofectamine 2000; 3, 4, 5 – RLSa0 cells incubated with 1 μM control conjugate luc-h-9/14, antisense oligonucleotide 5'-h-9/14, and anti-miR-21 conjugate 5'-h-9/14, respectively. D. The bar graph shows the semi-quantitative analysis of the western blot results for PDCD4. E. Proliferative potential of RLSa0 cells 72 h after transfection with 1 μM of antisense oligonucleotide 5'-h-9/14, control conjugate luc-h-9/14 and anti-miR-21 conjugate 5'-h-9/14. Data are given as medians calculated from three independent experiments.

Appendices

Figure A.1 Atom nomenclatures in the peptide amino acid backbone and side chains

Figure A.2 Nomenclature and atom numbering for the nucleotides bases and sugars

Figure B.1 MS-analysis (ESI-TOF): mass calculated for labelled peptide 2 at L7/R8 C3813C12H98N15 15N3O9 [M-H]+, 1167.81, found 1167.85

Figure B.2 MS-analysis (ESI-TOF): mass calculated for labelled peptide 3 at L9/R10, C38 15C12H98N15 15N3O9 [M-H]+, 1167.81, found 1167.84

Figure B.3 MS-analysis (ESI-TOF): mass calculated for labelled peptide 4 at L11/R12 C3813C12H98N15 15N3O9 [M+H]+, 1168.81, found 1168.90

Figure B.4 MS-analysis (ESI-TOF): mass calculated for labelled peptide 5 at L13/R14 C3813C12H98N15 15N3O9 [M+H]+, 1168.81, found 1168.90

Figure B.5 MS-analysis (ESI-TOF): mass calculated for labelled conjugate L7/R8 2 C9613C12H70N37 15N3O42P6 [M +2H+ +2Na +2H]3+, 2979.10, found 2979.10

Figure B.6 MS-analysis (ESI-TOF): mass calculated for labelled conjugate L9/R10 3 C9613C12H70N37 15N3O42P6 [M +2Na +2H]3+, 2979.10, found 2979.39

Figure B.7 MS-analysis (ESI-TOF): mass calculated for labelled conjugate L11/R12 4 C9613C12H70N37 15N3O42P6 [M+H]+, 2979.10, found 2979.10

Figure B.8 MS-analysis (ESI-TOF): mass calculated for labelled conjugate L13/R14 5 C9613C12H70N37 15N3O42P6 [M+H]+, 2979.10, found 2979.10

Figure S1 Representative 1H-NMR spectra of free peptide acetyl-(LeuArg)2-Gly-(LeuArg)2-Gly-COOH (A) and conjugate 5'-16 (B), showing the addition of characteristic resonance signals from TCAGTCTGATAAGCTA oligonucleotide. Spectra were recorded in D2O at 25°C using Bruker Avance II+ 400 Ultrashield spectrometer.
Figure S2  Cleavage of *miR-21 by 3'-conjugates 3'-16, 3'-h-6/16, 3'-h-9/16, 3'-h-9/10. Autoradiograph of 18% polyacrylamide/8 M urea gel. Lanes Im and T1, imidazole ladder and partial RNA digestion with RNase T1, respectively. Control – RNA incubated in the absence of conjugate for 0 and 144 h. *miR-21 (1 µM), 3'-conjugates (50 µM) were incubated at 37 °C for 0 – 144 h. The conjugate type and incubation time are shown at the top.

Figure S3  Accumulation of miR-21 products after hydrolysis by conjugates 5’-h-6/14 (A), 5’-h-9/14 (B), 5’-h-6/16 (C), 5’-h-9/16 (D) and 5’-16 (E) at individual sites. *miR-21 (1 µM) and conjugates (20 µM) were incubated at 37 °C for 0 – 72 h.

Figure S4  Backbone RMSD (in Å) of 5’-h-6/14 complex with miR-21 over course of one microsecond MD trajectory in implicit solvent.

Figure S5  Snapshots from trajectory of 5’-h-6/14 complex with miR-21 over 1µs of molecular dynamics simulation in implicit solvent. Peptide in yellow, oligonucleotide in green, miR-21 in blue.

Figure S6  Time series of (A) Arg10 and (B) Arg5 interactions with phosphorus atom of G15-A16 linkage.

Figure S7  Time series of Arg10 interactions with base atoms of G15.

Figure S8  Time series of Arg8 interactions with base atoms of G15.

Figure S9  A putative in-line orientation of 2’-OH group at G18-U19 linkage formed by 5’-h-6/14 complex with miR-21 during 1 µs molecular dynamics simulation.

Figure S10  Flow cytometry analysis of the accumulation of FITC-labeled peptide-oligonucleotide conjugate 5’-h-9/14 in RLS40 and KB-8-5 cells. (A) and (C) – RLS40 and KB-8-5, transfected with FITC-labeled peptide-oligonucleotide conjugate 5’-h-9/14 (1 µM) precomplexed with Lipofectamine 2000; (B) and (D) RLS40 and KB-8-5 cells incubated with Lipofectamine 2000 only. Experimental procedure was as follows: 4 h post-transfection cells were washed with saline solution and fixed in 2% formaldehyde in PBS. Cells were analyzed using Cytomics FC 500 (Beckman Coulter, USA) flow cytometer (excitation wave length 488 nm, emission 530±30 nm). Two parameters reflecting the efficiency of the process were used for the comparison: the transfection efficiency estimated as the percentage of cells with green fluorescence exceeded the maximum level of the auto-fluorescence of untreated cells and the mean fluorescence intensity of cells measured in relative fluorescent units (‘RFU’). The MFI for RLS40 and KB-8-5 cells was 12.6 and 40 RFU, respectively.
Figure S11  Confocal microscopy study of cellular accumulation of FITC-labeled peptide-oligonucleotide conjugate 5'-h-6/14 precomplexed with Lipofectamine 2000. A. Control KB-8-5 cells incubated with Lipofectamine 2000 only. B. KB-8-5 cells, transfected by FITC-5'-h-6/14 precomplexed with Lipofectamine 2000. [N.B. Human KB-8-5 cells were deliberately used for confocal microscopy: these are adherent cells forming stable monolayers, while RLSa0 cells are suspension cell lines, which are known to produce poor samples for confocal microscopy]. Experimental procedure was as follows: KB-8-5 cells in the exponential phase of growth were plated on the glass coverslips in 24-well plates at a density of 1.3 × 10⁵ cells/well one day before the experiment. FITC-labeled conjugate 5'-h-6/14 precomplexed with Lipofectamine 2000 was then added to the cells to final concentration 1 µM. In 4 h after transfection the cells were washed with PBS, fixed in 2% formaldehyde/PBS, permeabilized in 0.1% Triton-X100/PBS. The cells on the coverslips were then mounted on the object-plate in DAPI/Antifade solution (Millipore, USA). The analysis of FITC-labeled 5'-h-9/14 localization was performed using a confocal fluorescent microscope LSM 710 Meta (Carl Zeiss, Germany) at 60 × Magnification. Images (right column) merge two channels: blue (cell nuclei stained with DAPI, left column) and green (peptide-oligonucleotide conjugate 5'-h-6/14 labeled with Fluorescein, middle column). …….  

Figure S12  Expression level of miR-21 in RLSa0 cells after transfection of anti-miR-21 conjugate 5'-h-6/14 or with control conjugate luc-h-9/14. The conjugate concentration varied from 0.1 µM to 5 µM. Transfection was performed using Lipofectamine 2000. LF – RLSa0 cells incubated with Lipofectamine 2000 only. The expression of miR-21 was normalized to U6. Data are given as mean calculated from three independent experiments ± SEM…………………………………………………………  

Figure S13  Expression level of miR-17 and miR18a in RLSa0 cells after transfection with 1 µM of control conjugate luc-h-9/14, anti-miR-21 antisense oligonucleotide 5'-h-6/14 and anti-miR-21 conjugate 5'-h-6/14. Transfection was performed using Lipofectamine 2000. LF – RLSa0 cells incubated with Lipofectamine 2000 only. The expression of miRNAs was normalized to U6 and Rpl30. Data are given as mean calculated from three independent experiments ± SEM…………………………………………………………  

Figure S14  Stability of asON 5'-16 and conjugate 5'-16 in 10% FBS. A. Image of 15% PAAG/8M urea, stained with Stains-All. asON or conjugate was incubated in 10% FBS at 37 °C for 0, 2, 4, 8, 24 and 48 h. B. Kinetics of degradation of asON or conjugate in 10% FBS………………………………………………………………………………  

Figure S15  Cleavage of miR-21/oligonucleotide(asON) and miR-21/conjugate duplexes by RNase H. *miR-21 (1 µM) and oligonucleotide 5'-h-9/14 (5 or 30 µM) or conjugate 5'-h-9/14 (5 or 30 µM) were incubated at 37 °C for 0 and 30 min. Lanes Im and T1, imidazole ladder and partial RNA digestion with RNase T1, respectively. C – RNA incubated in the absence of oligonucleotide, conjugate or RNase H for 30 min…………  

19
List of Tables

Chapter 1

Table 1.1 List of microRNA candidates in clinical trials. Adapted from48................................. 38
Table 1.2 Parameters for peptide and protein conformations.................................................... 73

Chapter 3

Table 3.1 Chemical shifts (ppm.) of the nucleic acid component of the unlabelled conjugate ........169
Table 3.2 Medium and long range sequential peptide assignments............................................. 180
Table 3.3 Intramolecular NOEs between the peptide and the nucleotide protons within the conjugate (Summary of the data presented in Figure 3.26)...................................................... 182
Table 3.4 NMR experiments used for the assignments and structure determination of the labelled conjugates (S2) - (S5) containing residue-specific $^{13}$C, $^{15}$N labelled peptide. The purpose of each experiment, along with the extracted information, is indicated.......................... 186
Table 3.5 Chemical shifts (ppm) of the peptide a within the conjugate........................................194
Table 3.6 NOEs contacts in the peptidyl-oligonucleotide conjugate........................................... 196
Table 3.7 Experimental $^{3}J_{\text{HNH}}$ coupling constants of the conjugate peptide............................ 197
Table 3.8 NMR constraints and structure statistics.......................................................... 202
Table 3.9 Salt bridges and hydrogen bonds for the peptidyl-oligonucleotide conjugate from the 20 refined NMR structures a .................................................................................................. 204

Chapter 4

Table 4.1 Oligodeoxyribonucleotides used in the study: sequences and nomenclature.............. 232

Appendices

Table S1 RT and PCR primers used in the study................................................................. 276
Table S2 Hybridization of the conjugates with miR-21 and efficiency and specificity of miR-21 cleavage.................................................................................................................. 277
List of Schemes

Chapter 2

Scheme 2.1  Solid phase synthesis of the peptide NH$_2$-[Leu-Arg]$_4$-Gly-CONH$_2$ with Fmoc chemistry…………………………………………………………………………………………………………………………….. 119

Scheme 2.2  Synthesis of the peptidyl-oligonucleotide conjugates. A solution of aldrithiol-2 (Py$_2$S$_2$)$_1$ and triphenylphosphine (Ph$_3$P)$_2$ in anhydrous DMSO was stirred with the oligonucleotide 3 for 5 minutes. DMAP 4 was added to give the oligonucleotide with activated 5¢-phosphate 5. The peptide 6 dissolved in anhydrous DMSO was added to the solution and the reaction was allowed to proceed for two hours at 40 °C. The two fragments in the resulting conjugates 7 are linked via a phosphoramidate bond………….. 121
List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>Angstrom</td>
</tr>
<tr>
<td>ACN</td>
<td>acetonitrile</td>
</tr>
<tr>
<td>$^{13}$C</td>
<td>carbon-13</td>
</tr>
<tr>
<td>CTAB</td>
<td>cetyltrimethylammonium bromide</td>
</tr>
<tr>
<td>COSY</td>
<td>correlation spectroscopy</td>
</tr>
<tr>
<td>2D NMR</td>
<td>two-dimensional nuclear magnetic resonance</td>
</tr>
<tr>
<td>3D</td>
<td>three-dimensional</td>
</tr>
<tr>
<td>DIPEA</td>
<td>N,N-diisopropylethylamine</td>
</tr>
<tr>
<td>DMAP</td>
<td>4- (dimethylamino) pyridine</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DOTA</td>
<td>1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid</td>
</tr>
<tr>
<td>DSS</td>
<td>4,4-dimethyl-4-silapentane-1-sulfonic acid</td>
</tr>
<tr>
<td>Dy$^{3+}$</td>
<td>Dysprosium</td>
</tr>
<tr>
<td>Eu$^{3+}$</td>
<td>europium ion</td>
</tr>
<tr>
<td>FID</td>
<td>free induction decay</td>
</tr>
<tr>
<td>Fmoc</td>
<td>Fluorenylmethyloxycarbonyl</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>HBTU</td>
<td>$N,N,N',N'$-tetramethyl-$O$-(1H-benzotriazol-1-yl)uranium hexafluorophosphate</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum correlation</td>
</tr>
<tr>
<td>Hz</td>
<td>hertz</td>
</tr>
<tr>
<td>His</td>
<td>Histidine</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>J</td>
<td>scalar spin-spin coupling constant</td>
</tr>
<tr>
<td>kHz</td>
<td>kilohertz</td>
</tr>
<tr>
<td>Lu$^{3+}$</td>
<td>Lutetium</td>
</tr>
<tr>
<td>LiClO$_4$</td>
<td>Lithium perchlorate</td>
</tr>
<tr>
<td>MALDI</td>
<td>Matrix-assisted laser desorption spectroscopy</td>
</tr>
<tr>
<td>MHz</td>
<td>megahertz</td>
</tr>
<tr>
<td>$^{15}$N</td>
<td>nitrogen-15</td>
</tr>
<tr>
<td>nm</td>
<td>nanometer</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>ns</td>
<td>nanosecond</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------</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>-O⁻</td>
<td>oxyanion</td>
</tr>
<tr>
<td>O.D.</td>
<td>optical density</td>
</tr>
<tr>
<td>ODN</td>
<td>oligodeoxynucleotide</td>
</tr>
<tr>
<td>OH</td>
<td>hydroxyl</td>
</tr>
<tr>
<td>p</td>
<td>phosphate</td>
</tr>
<tr>
<td>pKₐ</td>
<td>acid dissociation constant</td>
</tr>
<tr>
<td>ppm</td>
<td>parts per million</td>
</tr>
<tr>
<td>ps</td>
<td>picosecond</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase H</td>
<td>ribonuclease H</td>
</tr>
<tr>
<td>RNase P</td>
<td>ribonuclease P</td>
</tr>
<tr>
<td>RP-HPLC</td>
<td>reverse-phase high performance liquid chromatography</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>THAP</td>
<td>2',4',6'-Trihydroxyacetophenone monohydrate</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer ribonucleic acid</td>
</tr>
<tr>
<td>Sₙ2</td>
<td>nucleophilic substitution bimolecular</td>
</tr>
<tr>
<td>TFA</td>
<td>Trifluoroacetic acid</td>
</tr>
<tr>
<td>τc</td>
<td>rotational correlation time</td>
</tr>
<tr>
<td>τm</td>
<td>mixing time</td>
</tr>
<tr>
<td>φ</td>
<td>dihedral angle</td>
</tr>
<tr>
<td>TD</td>
<td>number of data points</td>
</tr>
<tr>
<td>TOCSY</td>
<td>total correlation spectroscopy</td>
</tr>
<tr>
<td>UV-Vis</td>
<td>ultraviolet-visible</td>
</tr>
<tr>
<td>µM</td>
<td>micromolar</td>
</tr>
<tr>
<td>µs</td>
<td>microsecond</td>
</tr>
</tbody>
</table>
### Abbreviations – Amino Acids/Nucleotides

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>3-letter Abbreviation</th>
<th>1-Letter Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
</tr>
<tr>
<td>Aspartic Acid</td>
<td>Asp</td>
<td>D</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
</tr>
<tr>
<td>Glutamic Acid</td>
<td>Glu</td>
<td>E</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>3-letter Abbreviation</th>
<th>1-Letter Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenine</td>
<td>Ade</td>
<td>A</td>
</tr>
<tr>
<td>Cytosine</td>
<td>Cyt</td>
<td>C</td>
</tr>
<tr>
<td>Guanine</td>
<td>Gua</td>
<td>G</td>
</tr>
<tr>
<td>Thymine</td>
<td>Thy</td>
<td>T</td>
</tr>
<tr>
<td>Uracil</td>
<td>Ura</td>
<td>U</td>
</tr>
</tbody>
</table>
Abstract

Chemical ribonucleases mimics represent an important class of artificial enzymes, which can be generated by conjugation of short, catalytically inactive oligopeptides and oligonucleotides to produce biologically active conjugates. The most remarkable attribute of these hybrid structures is that the covalently attached oligonucleotide moiety seems to trigger catalytic activity of a previously inactive peptide, which becomes capable of cleaving RNA sequences. These biocatalytic structures mimicking the active centres of natural ribonucleases may provide a basis for the development of novel sequence-specific gene silencing approaches. This discovery may open new exciting avenues for advanced therapeutic interventions, where a disease caused by overexpression of a pathogenic protein could be treated through controlled down-regulation of disease-relevant mRNA, viral genomic RNA or micro-RNA. However, very little is known about structural mechanisms behind the biocatalytic activity of these artificial enzymes and their interactions with RNA sequences. The research presented in this thesis attempted to investigate structural and dynamics properties of these enzyme mimetics, both alone and in the complex with the target RNA. The first part of this research was focused on a high-resolution structural analysis of these enzyme mimics using high-field NMR spectroscopy and molecular dynamics of the model conjugates with the aim to provide a complete 3D description of their structure and dynamic behaviour. In order to overcome the challenges associated with the highly repetitive nature of the peptide component, we used site-specific $^{13}$C/$^{15}$N isotopic labelling of the model conjugates, which allowed us to considerably increase the number of the NMR derived distance constraints and thus provide a more accurate description of the most populated conformational ensembles. The second part of this research was focused on the investigation of the structural aspects of RNA targeting by this class of peptidyl-oligonucleotide conjugates, which were carefully designed to hybridise and cleave the highly oncogenic microRNA-21 in a sequence-specific manner. To gain structural insights into specific interactions between these catalytic conjugates and their bio-targets, we carried out 1 μs molecular dynamics simulations of the hybridised complex between miR-21 and one of the most efficient chemically engineered miRNase. The results of this thesis are vital for the future design of such conjugates, which may guide future development of novel, more efficient therapeutic interventions based on the RNA targeting.
Declaration

No portion of the work referred in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or 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, trademarks 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://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University’s policy on Presentation of Theses.
Acknowledgements

First, I would like to thank my supervisor, Dr. Elena Bichenkova who took me on this PhD project although I came from a totally different field. I really learnt a lot and I am grateful for this chance.

I am very much obliged to Dr. Richard Bryce, who let me work in the computational chemistry lab, which was the key to the successful finish of my PhD study. I also thank Dr Kepa Burusco very much for his patient unceasing effort to transfer at least some of his great computational chemistry knowledge to me, for help with the program scripts used in my projects and thus saved me a vast amount of time. I thank Dr Irfan Alibay and Dr Rupesh Chikhale for answering even the most trivial questions and for many valuable hints.

It has been a great pleasure collaborating with the colleagues from the Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, I am very thankful for the cosy atmosphere. Namely I would like to thank Prof. Marina Zenkova and Dr Olga Patutina. I want to acknowledge the company C4X Discovery for providing me with the training and the research facilities in the initial stage of my project. The financial support from C4X Discovery and BBSRC Industrial Case study is greatly appreciated.

Also, I thank Dr Wojtek Augustyniak, who did a wonderful job with the NMR training and who was always there when I needed advice. I also acknowledge those that trained and helped me during my time in C4X Discovery, Mr Michael Denison, Dr Sadia Mohammed, Dr Zoltan Takacs, and special thanks go to Dr Matt Cliff (MIB) for friendly environment and useful consultations.

I would also like to mention people from the chemistry lab, Mr Neil O’Hara and Dr Aled Williams for their valuable assistance and Dr Nikolai Ulyanov (UCSF) for his assistance in using the MARDIGRAS software.
Publications


1.1 Low molecular weight drugs: decline in drug discovery process

Despite extensive efforts undertaken by both academia and pharmaceutical companies in the drug discovery process, many conventional lead generation approaches to develop successful small-molecule drugs often fail due to a number of reasons. Analysis of the trends in drug development launched in the UK over the recent decades (Figure 1.1) indicates that, despite the increase in investment into drug research and development, the rate of new drug introductions has increased little and declined considerably relative to the investment made\textsuperscript{1,2}. Even though the significant scientific and technological progress has been made by both academia and the pharmaceutical industry over the past 60 years and in spite of the considerable money injection into the drug development process, the number of new approved drugs per billion invested US dollars has decreased by a factor of 2 in every 9-year time stretch. Instead of the expected increase in the cost-efficiency of commercialisation of new drugs, the effectiveness of the research and development of low molecular weight drugs is rapidly falling.

In addition, usual strategies for development of small-molecule therapeutics often suffer from poor selectivity towards target proteins or receptors, which potentially cause undesirable side-effects and toxicity in humans\textsuperscript{3,4}. One possible reason for this could be insufficient numbers of functional groups within low molecular weight drugs, leading to limited opportunities to form stable, selective interactions with bio-targets. Another reason could be a lack of ‘clean’ targets associated with a particular disease, which may lead to undesirable consequence that the developed drugs often hit many bio-targets, thus triggering side effects and cytotoxicity. In addition, an existing drug paradigm relies on finding a target, which is unique to a disease state only. However, in reality the majority of known bio-targets are present in both health and disease, and thus hitting any target with a small-molecule drug may inevitably lead to toxicity in humans.

Oversimplified concepts of membrane penetration (\textit{e.g.} Lipinski’s “Rule of 5”) may also contribute to drug failure. Indeed, unspecific penetration of cellular membranes by lipophilic compounds may inevitably lead to increased off-target side effects in healthy tissues. Also, orally-active drugs falling outside Lipinski’s ‘Rule of 5’, whose structures
may potentially provide motifs for absorption via biological transport, may be simply excluded from consideration.

Figure 1.1: Illustration of the decline in Drug Discovery and Development of small-molecule drugs. Adapted from Scannell et al.¹

Finally, every new protein target requires a separate screening process for selection of potential lead candidates. Consequently, the drug discovery & development process of small-molecule drugs remains very lengthy (10-12 years) and extremely expensive: average cost to develop and gain marketing approval for a new drug was pegged at $2.558 billion.⁵

Biopharmaceuticals may represent an attractive alternative in selective disease targeting⁶, since they have had a higher success rate than small molecules once they leave research and enter clinical trials. There was an approximately 32% approval rate for biologics versus an approximately 13% approval rate for small-molecule drugs first tested in humans between 1993 and 2004. The opportunity in biopharmaceuticals is big and growing too.
rapidly to ignore. Consequently, major pharmaceutical companies worldwide are increasingly shifting their R&D and sourcing focus to large-molecule products (Figure 1.2). Today biopharmaceuticals generate global revenues making up to 37% of the Pharma market (2012).

![Graph showing change in % of revenue from biopharma, 2000-12 and total revenue % of biopharma, 2012.]

**Figure 1.2:** Shift of Big Pharma to include large molecules. Adapted from Otto et al. ⁶

Being the most sophisticated and elegant achievements of 3D structure-based molecular design in combination with the most advanced experimental techniques (e.g. X-ray, NMR), biopharmaceuticals may become the core of the pharmaceutical industry in the near future, as they can offer high efficacy, unique selectivity and fewer (or negligible) side effects due to their unique shape, precise molecular recognition and extensive network of atomic interactions with bio-target. However, this precision in molecular design comes at great cost, and the development of biopharmaceuticals will require considerable investment for transformation of core laboratory facilities and for development of advanced technologies. This thesis is focused on the development of a particular class of such biopharmaceuticals for targeting pathogenic RNAs over-expressed in disease states.
1.2 RNA targeting

1.2.1 Rationale, possible impact and advantages of RNA targeting

The depletion of biological targets amenable to conventional small-molecule drugs, and the recent advances in understanding of the molecular mechanisms that dictate disease onset and development, underscores the importance of expanding the range of targets available for the development of novel therapeutic agents.

The rapidly growing knowledge in genetics has given rise to many follow-up studies that have supported a correlation between genetic mutation and the susceptibility of individuals to disease\(^7\)\(^,\)\(^8\)\(^,\)\(^9\). A DNA molecule can be mutated by replacement, addition or loss of specific bases in the genome, with consequent production of faulty proteins, which could be potentially harmful to the organism\(^10\) thus leading to a disease state or physiological disorder. After the Human Genome Project had been finalised, which led to the identification and mapping of all genes in human genome, DNA-based therapeutic interventions began to be exploited to provide new, exciting avenues in drug discovery. Missing or faulty genes can be supplied or replaced in a form of stable oligonucleotide analogues. In the case of oligonucleotide-based therapy, genes abnormally expressed in the disease states can be inhibited using oligonucleotides that hybridise the complementary regions of the nuclear dsDNA to form a triplex DNA, leading to selective inhibition of DNA transcription\(^6\). However, there has been a strong concern that intervening with the genetic information at the level of DNA may lead to irreversible consequences, as this approach involves manipulation with the essential genetic information\(^11\).

An alternative approach involves targeting of RNA as a potential bio-target to allow “upstream” targeting of the disease state by blocking protein biosynthesis in a ‘natural way’ that is often pursued by cells.

A multitude of therapeutic approaches, which are based on gene silencing, is mediated by antisense oligonucleotides, which are synthetic, single-stranded nucleic acids that hybridize to accessible regions within the target messenger RNA in a sequence-specific fashion\(^12\)\(^,\)\(^13\). Antisense oligonucleotides may also be potentially directed against disease-
causing genes or proteins\textsuperscript{14,15,16}, as well as employed as probes to detect specific sequences within target nucleic acids for diagnostic purposes\textsuperscript{17}.

RNA molecules take part in a multitude of essential cellular pathways, acting both as information carriers and as catalytic agents. For example, RNA molecules assembled with protein complexes are fundamental components of the spliceosome and play a key role in the splicing process of nascent messenger RNAs by removal of introns and linkage of exons in mature mRNA transcripts\textsuperscript{18}. The mature mRNA transcripts convey the genetic instructions from DNA to ribosomes, themselves RNA-protein complexes, which catalyse the build-up of proteins by sequentially adding amino acids from transfer RNA molecules to the growing protein chain\textsuperscript{19}. Furthermore, some RNA molecules are known to be involved in crucial signalling pathways that orchestrate regulation of a wide range of gene expressions through translational repression of specific target mRNAs. These short non-coding single and double-stranded RNA molecules (\textit{i.e.} micro RNAs and small interfering RNAs respectively) play an important role in gene silencing through post-transcriptional restraint and RNA degradation\textsuperscript{20,21}. Not surprisingly therefore, RNA has garnered a wide interest due to its potential applications in diagnosis and medication both as a therapeutic agent and a pharmaceutical target.

The RNA-mediated approach offers a number of crucial advantages over conventional gene-silencing, in which DNA transcription is restrained. Indeed, suppression of protein expression by antisense RNA-targeting oligonucleotides can occur via a variety of cellular pathways (Figure 1.3), which comprise (i) RNA digestion by RNase H following heteroduplex formation\textsuperscript{22,23}, (ii) mRNA splicing disruption\textsuperscript{24} or (iii) interference with the translation process by hindering mRNA binding to the ribosome complex\textsuperscript{25}. Therefore, gene silencing can be accomplished by disrupting a vast repertoire of RNA processing mechanisms, which involve either RNA maturation or translation.

Another reason why RNA is a more suitable target for gene silencing therapy than DNA is related to its inherent chemical instability and higher reactivity compared to DNA. In fact, the human genome is more stable against modification, degradation or other forms of attack from either endogenous (\textit{e.g.} reactive oxygen species, free radicals) or exogenous (\textit{e.g.} UV radiation, pathogens, drugs) agents compared to RNA. The explanation for the reduced stability of RNA compared to DNA is due to some minor differences in their
chemical structures. While RNA and DNA are both macromolecules that consist of nucleotides joined together via 3', 5'-phosphodiester bonds, the RNA backbone contains an additional hydroxyl group at the 2'-position of its ribose sugar, which can be involved in the degradation of the sugar-phosphate backbone.

Figure 1.3: Mechanism of protein expression arrest by antisense nucleotides: inside the cell the antisense oligonucleotide aligns with the target mRNA via base complementarity. The heteroduplex formed in this way halts protein synthesis by interfering with a variety of processing pathways involved in either mRNA maturation or translation into proteins. Adapted from 26.

As a result, in a basic environment, RNA is susceptible to hydrolysis through the 2'-hydroxyl group which can attack the neighbouring phosphodiester linkage and lead to the molecule being cleaved 27,28. This degradation route is not available in DNA due to the absence of the 2'-OH group. Additionally, the presence of the 2'-hydroxyl group causes RNA duplexes to adopt the A-form helix rather than the B-form found in DNA. As such, the minor groove is wider in double-stranded RNAs, resulting in a greater propensity to enzymatic attacks 29. Furthermore, RNA exists mostly as single-stranded molecules and is hence more exposed to attack from various chemical and biological agents than DNA, which is protected in double and superhelical conformations.

There are many other advantages of RNA-based gene silencing over DNA-based strategies. RNA structural elements, known as stem-loops and bulges, which usually contain single-stranded oligonucleotide segments, may be involved in RNA-protein interactions. These sites could potentially be exploited as anchorage sites for antisense nucleotides.
oligonucleotides by complementary base alignment\textsuperscript{30}. Moreover, RNA is distributed in most subcellular compartments and exhibits a ubiquitous presence in both nucleus and cytoplasm, which makes RNA an easily accessible target. In contrast to DNA, which is promptly checked and can potentially be repaired by highly effective genome repair mechanisms, RNA repair pathways are not known. Instead, damaged RNA appears to be quickly and purposefully degraded by ribonucleases\textsuperscript{31,32}.

RNA-gene silencing is also free from the risks associated with DNA-based gene therapy that involves DNA manipulations (e.g. random insertion of exogenous genes) that may cause genome alterations deleterious for function. The rapid turnover of RNA also potentially affords a high level of treatment control in terms of durability and dose adjustment according to individual response.

RNA-viruses also provide a basis for possible application of RNA-based gene therapy as an appealing alternative to conventional treatments of RNA virus infections\textsuperscript{33}. In fact, viral reproduction can be halted by preventing the expression of critical viral proteins. Moreover, specificity towards viral genes prevents hosts genomes from being affected thus minimising the occurrence of side effects\textsuperscript{34}.

Another aspect of RNAs that makes them appealing to consider as targets for the treatment of genetic disorders is the ability of some RNAs to fold into tertiary structures and act as enzymes, such as RNase P (a ribonucleoprotein)\textsuperscript{35} and ribozymes\textsuperscript{36}, which can bind specifically to RNA molecules, break them down and trigger a new cleavage process with other target RNAs\textsuperscript{37}.

This particular property of some RNA molecules has inspired the development of catalytic agents capable of cleaving segments of single stranded RNA. The challenging aspect here is the design of such constructs that are capable of accelerating RNA degradation with cleavage rates comparable to those of natural enzymes. The lower rate enhancements observed so far for such artificial catalysts are likely to be due to non-optimal binding interactions with RNA substrates and steric clashes. Therefore, a better grasp of the factors responsible for RNA phosphodiester bond scission and the cleavage pathways for biological catalysts such as natural nucleases may lead to more efficient catalytic agents.
1.2.2 MicroRNAs as attractive therapeutic targets

MicroRNAs are a class of natural, small, non-coding, single-stranded RNAs (ssRNAs) containing sequences of approximately 20-25 bases in length. They are involved in the regulation of approximately 30% of human genes and in the post-transcriptional gene-silencing in diverse biological processes, such as cell proliferation and differentiation via base-pairing with 3’ untranslated segments of mRNAs. This results in degradation of the targeted mRNAs and consequently reduced or modified translation in proteins.

Aberrant expression of microRNAs has been reported to be closely associated with the development of a spectrum of human diseases ranging from autoimmune disorders, cardiovascular diseases, inflammatory diseases and all types of tumour. In some cancers, microRNA expression is increased and they gain the function of inhibiting tumour suppressor genes that control proliferation and apoptosis of cells, causing the progressive development of cancer.

Their unique characteristic of being small and with highly-conserved sequences has made the microRNA an attractive target in antisense therapy. Potent, synthetic oligonucleotides have been developed against microRNA, which possess high affinity and selectivity towards the target and have reached clinical trials. Binding of microRNA to antisense oligonucleotides with base pair complementarity triggers various mechanisms depending on the therapeutics, chemical structures and design, which results in reduced or modified protein expression. The different antisense therapeutics developed for inhibiting microRNA and currently in clinical trials are depicted in Table 1.1.

Downstream modulation of gene expression by targeting microRNA offers an opportunity to downregulate protein expression in the cytosol. Considerable work is still needed in the development of efficient vehicles to deliver the antisense oligonucleotides to the target tissues or organs. Currently the oligonucleotides are administered through local or parental injection routes and technological advancements in oral delivery formulation are thus an important step to further use miRNAs as therapeutic targets in patients.
Despite the outstanding obstacles, the increasing research and development of microRNA therapeutic programs provides grounds for further exploitation as a viable option for modulation of microRNA expression in human disease and promises the translation of their therapeutic potential into available drugs in the coming years.

<table>
<thead>
<tr>
<th>Name (company)</th>
<th>miRNA therapeutic</th>
<th>Delivery system</th>
<th>Indication</th>
<th>Clinical trial details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mirvirasen (Santaris Pharma)</td>
<td>AntimiR-122</td>
<td>LNA-modified antisense inhibitor</td>
<td>Hepatitis C</td>
<td>Multicentre phase II</td>
</tr>
<tr>
<td>RG-101 (Regulus Therapeutics)</td>
<td>AntimiR-122</td>
<td>GaINAc-conjugated antimiR</td>
<td>Chronic hepatitis C</td>
<td>Multiple phase II</td>
</tr>
<tr>
<td>RG-125/AZD4076 (Regulus Therapeutics)</td>
<td>AntimiR-103/107</td>
<td>GaINAc-conjugated antimiR</td>
<td>Type 2 diabetes and liver diseases</td>
<td>Single-centre phase I/IIa</td>
</tr>
<tr>
<td>MRG-106 (miRagen Therapeutics)</td>
<td>AntimiR-155</td>
<td>LNA-modified antisense inhibitor</td>
<td>Cutaneous T cell lymphoma and mycosis fungoides</td>
<td>Multicentre phase I</td>
</tr>
<tr>
<td>MesomiR-1 (EnGeneIC)</td>
<td>miR-16 mimic</td>
<td>EnGeneIC delivery vehicle</td>
<td>Mesothelioma, lung cancer</td>
<td>Multicentre Phase I</td>
</tr>
<tr>
<td>MRX34 (Mirna Therapeutics)</td>
<td>miR-34 mimic</td>
<td>LNPs</td>
<td>Solid tumours</td>
<td>Multicentre phase I</td>
</tr>
<tr>
<td>MRG-201 (miRagen Therapeutics)</td>
<td>miR-29 mimic</td>
<td>Cholesterol-conjugated miRNA duplex</td>
<td>Scleroderma</td>
<td>Single-centre phase I</td>
</tr>
</tbody>
</table>

**Table 1.1:** List of microRNA candidates in clinical trials. Adapted from\(^{48}\).

1.3 Spontaneous, enzymatic and chemical catalysis of RNA cleavage

1.3.1 RNA chemical stability

Despite the common view that RNA is easily susceptible to spontaneous cleavage, non-enzymatic degradation of the internucleotide sugar-phosphate backbone in the RNA molecules (especially in the absence of other catalysts) is extremely slow. The estimated half-life of the RNA spontaneous hydrolysis under biologically-relevant conditions (\(i.e. 23^\circ C, \text{pH 7}\)) is approximately 1 year\(^{49,27}\).
The rates of phosphodiester bond cleavage are greatly affected by the molecular environment. For example, double-stranded helical structures and nucleotide base compositions that favour base stacking and hydrogen bond formation may further retard cleavage rates.\textsuperscript{50,51}

In order for RNA cleavage to be biologically relevant RNA instability has to be enhanced by at least a million-fold\textsuperscript{52}, which is currently achievable only through the participation of catalysts such as enzymes or metal ion complexes.

### 1.3.2 Spontaneous RNA transesterification

Natural (\textit{i.e.} non-enzymatic) cleavage of the RNA phosphodiester bonds occurs via internal phosphoester transfer (\textit{i.e.} transesterification) involving nucleophilic in-line attack of the 3’ phosphorus in the target linkage by the adjoining 2’-OH functional group. The fragments yielded by this reaction carry 2’, 3’-cyclic monophosphate (subsequently hydrolysed to a phosphate monoester) and 5’-hydroxyl termini. This one-step reaction proceeds through a mechanism similar to S\textsubscript{N}2 and in the absence of an enzyme occurs with first order kinetics, the rate constants being strongly dependent on pH.\textsuperscript{49} In fact, extreme pH conditions (\textit{i.e.} either acidic or basic pH) can accelerate spontaneous RNA cleavage by factors up to 10\textsuperscript{6}. Figure 1.4 schematically represents the spontaneous transesterification reaction at physiological pH.

At neutral pH, the oxygen at the 2’position exists in a protonated form (hydroxyl). The two phosphate oxygens, which are not directly involved in formation of the nucleotidic bridge, have an excess of electrons and a resonant double bond to the phosphorus centre. The nucleophilic attack at phosphorous atom by 2’-OH group gives rise to a penta-coordinated intermediate with a double negative charge. The fragment containing 5’-OH leaves as an oxyanion and becomes protonated upon scission.

In the absence of a catalyst, the weak nucleophilicity of the 2’-OH group, the high electron density at the non-bridging oxygens developed upon attack at phosphorus atom and the negatively charged 5’-O- departing group all contribute to the low rate of spontaneous phosphodiester scission.
Figure 1.4: Breakdown of the RNA internucleotide linkage: (a) Internal phosphoester transfer by 2’-OH nucleophilic attack on 3’-phosphorus atom. (b) Formation of the penta-coordinated phosphorane intermediate. (c) Cleavage products. (d) Subsequent hydrolysis of the 2’, 3’-cyclic monophosphate into a mixture of 2’- and 3’-nucleotide monophosphates\textsuperscript{53}.

Indeed, according to Westheimer’s rule\textsuperscript{54}, spontaneous cleavage occurs only if both the nucleophilic and the leaving oxygens are in-line with the phosphorus atom. In the in-line configuration the penta-coordinate intermediates assume trigonal bipyramidal geometry, with the 2’ and 5’ oxygens occupying apical positions approximately 180° from one to another. However, an in-line arrangement can be properly accomplished only via interaction with an enzyme active site that constrains the atoms in a position favourable for the formation of the penta-coordinate intermediates\textsuperscript{55}.

1.3.3 Base catalysis

Enhancement of the RNA phosphodiester cleavage rate to a considerable extent can be achieved by base catalysis (hydroxide ion), wherein the 2’ oxygen exists as an oxyanion that exhibits a nucleophilic power up to $10^8$-fold higher than the protonated species (hydroxyl). The reaction rate constants increase in a log-linear fashion when the reaction mixture pH values rise above 5. This is due to the steep augmentation of the population of the nucleophilic deprotonated form over that of the hydroxyl species. Conversely, at pH values near 5 the 2’-hydroxyl group is the dominant nucleophile form present and is a much poorer nucleophile\textsuperscript{49}.
1.3.4 Metal ion catalysis

The scission of the RNA phosphodiester linkage can be promoted by a wide range of metal ions such as Mg$^{2+}$, Ca$^{2+}$, Zn$^{2+}$ and Co$^{3+}$. With the exception of lanthanide ions, cleavage catalysis usually requires the participation of single metal ion catalysts.

Trivalent Lanthanide ions (e.g. Eu$^{3+}$, Tb$^{3+}$) are the most efficient catalysts with reaction rate enhancement of factors up to $10^4$, whereas alkaline earth metal ions (e.g. Mg$^{2+}$, Ca$^{2+}$) catalyse phosphodiester cleavage only to some extent over the spontaneous reaction rate$^{56}$. Typically, metal ions with a low $pK_a$ cleave RNA more efficiently than less acidic cations$^{57}$. Apparently, catalysis of the RNA cleavage occurs as the result of three concurrent processes (see Figure 1.5): (i) The metal ion assists deprotonation of the 2'-OH group by lowering the nucleophile $pK_a$; (ii) Interaction of the hydrated metal ion with the non-bridging oxygens decreases the phosphorus electron cloud and consequently increase its electrophilicity and thus susceptibility to nucleophilic attack; (iii) Ultimate departure of the leaving 5'-OH fragment is thermodynamically favoured by oxyanion protonation upon coordination with a water molecule donated by the hydrated metal ion$^{58}$.

**Figure 1.5:** Scission of RNA phosphodiester bond promoted by metal ions ($M^{n+}$) and their hydrated forms$^{58}$. 
1.3.5 RNA cleaving enzymes

1.3.5.1 Ribonucleases

Ribonucleases (RNases) are proteins exerting RNA cleavage activity in important metabolic processes such as the degradation of redundant or faulty ribonucleotides. Ribonucleases are also involved in the maturation of both messenger RNAs\(^59\) and non-coding RNAs\(^60,61\) as well as in RNA end turnover\(^62\). Structural features shared by all RNases include structural motifs to ensure specific alignment with substrate bases, binding to phosphate groups and residues catalysing scission of the phosphodiester bond. Typically, the amino acid residues present in the ribonuclease’s active centre leading to phosphodiester bond scission include histidines and glutamines. Target specificity of RNases is dictated by the organization of their nucleotide binding domains. RNases can be grouped in three major families (A, T1 and T2) depending on their active site residue composition, size and base specificity.

Enzymes belonging to the RNase A family cleave single-stranded RNAs and exhibit specificity for cytosine and uracil\(^63\) while T1 family RNases cleave the phosphate group of unmatched guanine residues\(^64\). T2 family RNases are enzymes that cleave the phosphodiester bond of all four ribonucleotides. Bovine pancreatic ribonuclease A represents an A-family RNase and is an endonuclease that has been particularly thoroughly studied. The nucleotide binding domain of this enzyme consists of the side chains of three residues; the hydroxyl group and the amidic nitrogen atom of threonine and serine residues, which bind either the cytosine or uracil heterocycle of the substrate\(^65\) and the aromatic ring of a phenylalanine, which engages in partial \(\pi\)-\(\pi\) stacking interactions with the pyrimidine heterocycles. The catalytic groups of the enzyme active site include the imidazole ring of histidine 12 and histidine 119, the backbone amide of phenylalanine 120 and the \(\varepsilon\)-amino group of lysine 41. A generally accepted mechanism for RNase catalysis is described in Figure 1.6. In this mechanism, the basic nitrogen of the histidine 12 side chain abstracts a proton from the sugar ring 2’-OH of the nucleotide target, which then becomes a more powerful nucleophile. Consequently the attack on the adjacent 3’-phosphate in the phosphodiester bond of the RNA backbone is energetically favoured\(^66\) and proceeds in-line via an associative mechanism similar to \(S_N2\)^\(^67\), generating a 2’, 3’-cyclic pentavalent phosphodiester intermediate.
Figure 1.6: Reaction pathway of RNA cleavage catalysed by ribonuclease A: a) The His12 in the active site abstracts a proton from the RNA ribose 2'-OH activating the oxygen nucleophilic attack on the phosphor. b) The negative charge builds up on the transition state phosphate stabilised by positively charged residues Lys41 and Phe120 and departure of the adjacent ribose 5'-OH promoted by donation of a proton from His119 allow formation of the cyclic nucleotide intermediate. c) The general acid-base activity of the His12 and His119 side chain imidazole rings promote hydroxyl attack on the 2', 3'-cyclic nucleotide monophosphate to generate either 2' or 3' nucleoside monophosphates.

In the RNA hydrolysis catalysed by RNase A, the 2', 3'-cyclic phosphate formation occurs through a transition state having a dianionic phosphorane group where the phosphorus is bonded to five oxygens. A partial negative charge that builds up on the scissile phosphate is neutralised by interactions with the positively charged lysine 41 and phenylalanine 120 during the transition state.
The aromatic heterocycle of histidine 119 acts as the general acid and protonates the 5’-oxyanion of the susceptible phosphodiester bond in the transition state. This favours departure of the 5’-leaving group and consequently phosphorus detachment from the adjacent sugar, resulting in the RNA hydrolysis.

For the second step of the cleavage reaction, the imidazole ring acid-base catalysis of the two histidine residues is reversed. Histidine 12 protonates the 2’, 3’-cyclic nucleotide monophosphate intermediate, while histidine 119 draws away a proton from water and provide an OH- ion which attacks the same 2’, 3’-cyclic intermediate to yield either a 2’ or a 3’-nucleotide monophosphates when hydrolysed.

1.3.5.2 Ribozymes

Ribozymes are nucleic acid enzymes that breakdown RNA using a transesterification reaction mechanism.

Reaction rate enhancement by riboyme catalysis can be substantial, with rate constants comparable to those of protein enzymes. Similarly to ribonucleases, ribozymes (i) stabilize the transition state by adding or abstracting protons, (ii) align the RNA target to a position susceptible to reaction and (iii) optimise non-covalent binding interactions with non-reacting RNA sites to lower the transition state energy barrier. Ribozymes catalyse sequence-specific internal transesterification reactions by using base-pairing to orient the cleavage site within the enzyme catalytic domain.

Eight different types of ribozymes have been identified in nature, four of which are self-cleaving molecules containing 40-160 nucleotide residues. They include hairpin, hammerhead, hepatitis delta virus and Neurospora VS RNAs. These ribozymes cleave RNA in a site-specific manner and do not bind ions to direct catalysis, in contrast to other ribozymes such as self-splicing group I and group II introns and RNase P.

1.4 Oligonucleotide-based chemical ribonucleases

In the last decade, chemical catalysts, which are capable of mimicking ribonucleases and ribozymes and cleave specific sequences within RNA rather than specific bases, have attracted tremendous attention from the scientific community and inspired many research groups.
The advantage of artificial ribonucleases as potential gene silencing therapeutics over catalytically inactive antisense oligonucleotides is their inherent or induced RNA catalytic activity. Indeed, antisense oligonucleotides themselves do not directly induce RNA degradation, which is instead triggered by RNase H upon formation of the RNA-DNA hybrid duplex via hybridization. Consequently, their interaction with the RNA target occurs with a 1:1 stoichiometry. Conversely, catalytic therapeutics may have a potential to inactivate multiple targets at a time and their employment in therapy would allow lower doses of drug to be administered to achieve comparable effectiveness. This may considerably increase their potency, decrease the cost of therapeutic treatment, reduce undesired side-effects and minimize drug toxicity.

Chemical ribonucleases usually consist of two main moieties connected via a linker (i) oligonucleotides or other nucleic acid analogues for specific hybridization with a target RNA sequence and (ii) a catalytically active construct at the termini, which is delivered by the oligonucleotide recognition motif to the targeted phosphodiester bond to achieve site-specific cleavage outside the RNA sequence that is base paired to the carrier oligonucleotide.

In such constructs, the enzyme catalytic turnover could be affected by the length of the base-paired regions and the binding strength to the target chain. To achieve the desired level of catalytic turnover, it is important to ensure that the hybridization affinity is higher in the intact RNA-artificial ribonuclease duplex than in the cleaved counterpart.

Several natural ribonuclease-oligonucleotide hybrids have been reported to catalyse RNA cleavage in a sequence-specific manner. However, use of natural ribonucleases for antisense therapy presents some drawbacks compared to synthetic counterparts. Although expression of natural nucleases is feasible in cells and could be a possible advantage, residue replacement with more stable analogues for extended half-lives dramatically reduces their catalytic activity. In fact, since natural enzyme functionality relies on a precise and complicated tertiary conformation, which is achievable only when natural residues occupy key positions, any changes in the enzyme structure may lead to a reduction or complete loss of catalytic activity. Furthermore, delivery of a whole protein molecule into the cell is complicated by the considerable size of the natural enzyme and any attempts to favour internalization into cells by chemical modifications may result in
decrease in catalytic activity. However, nuclease mimics are smaller constructs possessing the minimal set of residues required for catalysis and thus have the advantage of being delivered more easily into the cell. Furthermore, in such enzyme mimics the binding and cleavage sites are physically separated, so that the sequence-specific recognition motif and the catalytic site can be designed and synthesised independently.

Based on the mechanism of RNA cleavage in the biological environment, a potential candidate for an enzyme mimic should contain functional groups capable of enhancing nucleophilicity of the 2’-OH group and reduce the electron density on the phosphorus centre to facilitate the attack from 2’-oxygen. These functional groups should promote reduction of the negative charge at the leaving 5’-oxyanion group upon cleavage of the P-O5’ linkage. According to Westheimer’s rule, the nucleophilic attack at the phosphorus centre and the departure from the penta-coordinate phosphorane transition state is permitted exclusively through an apical position. Within double-stranded helices, departure of the leaving 5’-O’ nucleotide fragment in the apical position is hindered by base-stacking interactions between the bases of the hybridised double-stranded region. The rational design of artificial ribonucleases should therefore be based on oligonucleotide targeting of RNA sequences that are not engaged in base-pairing within RNA secondary structures (also as hairpin loops or bulges) nor in double-helical structures formed upon binding of RNA with the recognition elements of chemical ribonucleases. Furthermore, sequences with high frequencies of purines should not be selected as target sites since they tightly bind and interact with neighbouring bases, hampering RNA cleavage.

Synthetic ribonuclease constructs developed to date and capable of RNA cleavage in physiologically relevant conditions can be grouped into two main chemical types. This classification depends mainly on the features of their catalytic moieties and their mechanism of cleavage: (i) metal ion-dependent complexes and (ii) metal ion-independent conjugate groups. In the case of metal-dependent catalysts, metal complexes (lanthanide ion chelates, Cu^{2+} and Zn^{2+} chelates) are attached to molecules capable of RNA recognition. In metal-independent catalysts the active site of natural ribonucleases are mimicked by moderately basic groups engaged directly in catalysis such as polyamines or guanidinium or imidazole derivatives.
Barbier and Brack\textsuperscript{83} found that marginal ribonuclease activity was also exhibited by polycationic polypeptides made up of alternating basic hydrophilic (Lys, Arg) and hydrophobic (Ala, Leu) residues. In a later study by Mironova and co-workers\textsuperscript{84} a significant improvement of the catalytic activity of these peptides towards RNA was observed when these peptides were conjugated with short oligodeoxynucleotides.

1.4.1 Metal-dependent artificial ribonucleases

For the purpose of RNA strand cleavage at specific sites of phosphodiester bond, the metal ion must be firmly bound to cage-like structures that allow its conjugation to sequence recognizing elements. Tri and tetravalent lanthanide ions display exceptional catalytic activity\textsuperscript{85} under physiological conditions and represent the most efficient ribonucleases developed thus far in this category. This is consistent with the hard Lewis acid nature of the lanthanide ions and complexes with nitrogen-containing macrocyclic ligands (\textit{i.e.} DOTA) are favoured for their higher kinetic inertness to lanthanide dissociation.

The first developed metal ion chelates\textsuperscript{86} consisted of lanthanide ions (\textit{i.e.} Th\textsuperscript{3+} and Lu\textsuperscript{3+}) complexed with iminodiacetic acid group conjugated to the terminus of oligodeoxynucleotides (1), which provided site selectivity. These ribozyme mimics cleaved 39-mer synthetic ribonucleotide targets outside the double helix region with an efficiency of almost 20\% after 8 hours. Despite their good stability, complexation with negatively charged groups leads to a substantial loss of catalytic activity of the lanthanide ion and lack of substrate turnover. Promising results have been obtained with neutral macrocycles like N-substituted azacrowns that are more suitable for retaining the reactivity of the metal ion and, indeed, such lanthanide complexes conjugated with oligonucleotides have shown considerably increased cleaving efficiency.

In work by Morrow\textsuperscript{87}, tetraamide derivatives of 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA) have been reported to form highly stable complexes with trivalent lanthanides and with Th\textsuperscript{4+} (2), which showed the highest degree of affinity towards the target sequence. 50\% cleavage of the RNA analogues by the Th\textsuperscript{4+} complex under physiological conditions occurred within 4.5 min, with an efficiency forty times greater than that of trivalent lanthanides. Its higher reactivity is attributed to the ninth coordination site that confers to Th\textsuperscript{4+} a greater Lewis acidity and thereby allows a strong interaction
with the dianionic phosphorane transition state and promotes the phosphate ester cleavage. Magda and co-workers developed artificial ribozymes by conjugating Eu\(^{3+}\) (3) and Dy\(^{3+}\) chelated by texaphyrin macrocycles to the 5'-end of short oligodeoxynucleotide probes. These conjugates specifically cleaved at the phosphodiester immediately 3’ to the last base pair and were found to be the most efficient mimics displaying 50% cleavage of total RNA in 10 and 2 hours respectively at pH 7.5 and 37˚C. Intra-chain derivatives of Dy\(^{3+}\) (4) texaphyrin and Eu\(^{3+}\) pyridine cyclophane conjugates were additionally prepared by tethering the chelate to an internally positioned glyceryl linker to ensure turnover of the substrate RNA. Double helical RNA is less susceptible to metal ion-promoted cleavage than single-stranded RNA, since strong base-stacking interactions prevent the 5’-nucleoside leaving group from attaining an apical orientation with respect to the 3’-nucleoside nucleophile group within the phosphorane intermediate. For this reason the Sessler group developed these lanthanide chelate conjugates bearing a non-nucleosidic linker internally to the oligonucleotide in order to create bulges in the target chain following hybridization and thereby promote cleavage. These lanthanide chelate based cleavers also exhibited efficient RNA substrate turnover. A schematic representation of lanthanide ion based chemical ribonuclease structures is shown in Figure 1.7.

Mononuclear Zn\(^{2+}\) complexes have also been broadly studied as catalysts for RNA cleavage, although their cleavage activity was less efficient than that of lanthanide chelates. Zn\(^{2+}\) macrocyclic complexes that formed a strong interaction with RNA analogue bases were based on triazacrown derivatives, which exhibit high binding constants for uridine and thymidine and are linearly related to the Zn\(^{2+}\) complex pK\(_a\). Based on this finding, dinuclear triaazamacrocycle Zn\(^{2+}\) complexes that bind two uridine residues have been developed (5). Uridine-containing RNA analogues formed complexes with mono- and dinuclear Zn\(^{2+}\) complexes. However, the target RNA was more readily cleaved by the dinuclear Zn\(^{2+}\) catalyst compared to the mononuclear counterpart.
Stabilization of the transition state from proton transfer via acid-base catalysis was not observed in the rate limiting step of the reaction, which posits that the large catalytic rate acceleration of the dinuclear Zn$^{2+}$ complex is exclusively due to stabilization of the transition state for RNA cleavage by electrostatic binding between the zinc di-cations and the anion charges of the bridging alkoxy groups$^{93}$. 

**Figure 1.7:** Schematic representation of Lanthanide ion based chemical ribonuclease structures: (1) Lutetium (Lu$^{3+}$) ion complexed with chelate conjugates containing negatively charged carboxylate groups$^{86}$; (2) Thorium (IV) ion complexed to DOTA derivative$^{87}$; (3) Europium (Eu$^{3+}$) macrocyclic conjugate containing a pyridine cyclophane$^{89}$; (4) Dysprosium (Dy$^{3+}$) thaxypyrin attached to the oligonucleotides via an internally positioned glyceryl linker$^{88}$. 
Enhancement of the catalytic activity of mononuclear Zn\(^{2+}\) complexes was achieved by including hydrogen bond donors such as guanidinium or ammonium groups\(^{94}\) (6). A putative mechanism of RNA analogue hydrolysis includes stabilization of the phosphorane by coordination to the Zn\(^{2+}\) ion and to the guanidinium/ammonium group, followed by proton shuttling from the nucleophilic 2'-OH to the leaving 5'-O- mediated by the Zn-OH anchored to the phosphate. The structure of the mononuclear and dinuclear Zn\(^{2+}\)-catalysts are shown in Figure 1.8.

Artificial ribonucleases based on Zn\(^{2+}\) or Cu\(^{2+}\) complexes conjugated to oligonucleotides have also been developed to deliver the catalytic component to specific regions within the cleavage site through complementary base pairing, as shown in Figure 1.9. Bashkin and co-workers\(^{95}\) synthesized artificial ribonucleases by conjugating Cu\(^{2+}\) terpyridine to a 17-mer oligonucleotide for molecular recognition. The catalytic activity was optimized by tethering an abasic serinol backbone to the catalyst component (a Cu\(^{2+}\) terpyridine complex) (7) resulting in increased flexibility of the proximal catalytic unit. This chemical modification relieved the duplex oligonucleotide-target rigidity, which was thought to be responsible for hampering the formation of the RNA transesterification transition state and consequently produced enhancement of the cleavage efficiency.

Catalytic turnover and cleavage of complementary ribonucleotides was also achieved by Zn\(^{2+}\) azacrown complexes such as 1,5,9-triazacyclododecane chelates conjugated to a 13-mer oligonucleotide\(^{29}\). Attachment of the macrocycle to the oligonucleotide 3'- or 5'-terminus was achieved via disulphide (8) or amide bond (9) linkers. Both conjugates
catalytically cleaved the RNA target with first order kinetics (the target concentration was in excess). The cleavage rates seemed not to be affected significantly by the linker length, whereas it was unexpectedly sensitive to the presence of the disulphide bond in the linker. Indeed, the disulphide tethered complexes afforded an eight times faster RNA cleavage than the amide bond equivalent.

![Diagram of oligonucleotide conjugates](image)

**Figure 1.9:** Structures of the oligonucleotide conjugates of Cu$^{2+}$ and Zn$^{2+}$ chelates: (7) Cu (II)-terpydine conjugate; (8) and (9) Zn (II)-1,4,7-triazacyclododecane conjugates.

### 1.4.2 Metal-independent artificial ribonucleases

In metal independent artificial ribonucleases the catalytic moiety consists of small organic functional groups such as polyamines, imidazole and guanidine, which mimic the active site of natural ribonucleases (Figure 1.10). These fragments can be attached to antisense deoxyribonucleotides to ensure sequence-selective binding to a specific sequence within the RNA target.
The simplest artificial nucleases that promote RNA hydrolysis were based on diethylenetriamine conjugated to the 5’-end of oligodeoxyribonucleotides\textsuperscript{96}. The phosphodiester bond scission of a tRNA\textsuperscript{Phe} occurs when the tRNA forms a duplex with the 19-mer sequence of the oligonucleotide. The selective hydrolysis of the phosphodiester linkage at the 3’-side of the hybrid was attributed to the acid-base catalysis mediated by cooperation of the ammonium cation and neutral amino group, although they were poorly efficient. In a later study by Verheijen and co-workers\textsuperscript{97} peptide nucleic acids (PNAs) that specifically bind to a synthetic 22-mer RNA were tethered to the diethylenetriamine fragment via a urea linker (10). A remarkably high catalytic activity was observed with 50% of RNA cleavage reached in 5 hours under physiological conditions.

Various compounds have also been studied using guanidinium groups due to their ability to interact with the target via hydrogen bonds and accelerate phosphate ester hydrolysis in non-aqueous solvents. In particular bis(guanidinium) derivatives (11) were shown to promote the cleavage of a ribonucleotide model, a phosphate diester containing an hydroxylalkyl chain that mimics the sugar 2’-OH, with the substantial rate enhancements of 290-fold over the rate of an uncatalysed reaction. When oriented into the optimal position, the guanidinium groups seem to increase the phosphorus electrophilicity by coordinating two phosphate groups via four hydrogen bonds favouring the nucleophilic attack by the hydroxyl group.

Apart from polyamine and guanidine conjugates, constructs comprising imidazole derivatives with the aim of mimicking the histidine residues of the ribonuclease active centre have been studied. Vlassov’s group\textsuperscript{98} developed artificial nucleases by conjugating bis-imidazole compounds to the 5’ end of oligodeoxyribonucleotides through a flexible linker (12). The ribonuclease activity of the conjugates was assessed using yeast tRNA\textsuperscript{Phe} as the target, with a site in the tRNA T-arm complementary to the oligonucleotide part. The conjugates exhibited high catalytic activity by efficiently binding to the complementary bases on the target T-arm and by directing the cleaving part, the imidazole residues, towards the sequence adjacent to the binding site, C61-ACA-G65, known to display inherent lability. Ribonuclease activity was observed under physiological condition resulting in a rapid hydrolysis of the C63-A64 bond within 1-2 hours.
It was also found that the catalytic imidazole groups were conformationally highly flexible and no stable contacts to the RNA cleavage sites were detected, implying the absence of stable cleaving conformations. This suggests that improvement of such conjugates can be attained by more exact placement of the imidazole residues of the cleaving constructs.

A later study\[^8\] designed artificial ribonucleases based on nucleotides binding the target tRNA sequences conjugated to multiple imidazole groups as the catalytic moiety. The cleavage activity was demonstrated to be greatly affected by (i) the number of imidazole groups, (ii) the linker flexibility and length that connects the oligonucleotide to the catalytic fragment and (iii) the nature of the anchor group between the linker and the nucleotide. The cleavage construct containing four imidazole groups readily cleaved the phosphodiester bond when closely located towards the RNA backbone, whereas a higher number of imidazole groups hinder the RNA hydrolysis, presumably, due to steric hindrance.

\[\text{Figure 1.10: Schematic representation of the chemical ribonucleases based on metal independent catalytic structures: (10) diethylenetriamine-peptide nucleic acid conjugate}\[^9\]; (11) bis (guanidium) derivative\[^9\]; (12) imidazole based construct-oligonucleotide conjugate\[^9\].]
1.4.3 Drawbacks of existing ribonucleases

Artificial ribonucleases have been proposed as an attractive option to the traditional gene silencing strategy. Indeed, while in the antisense oligonucleotide approach drugs bind their targets in a stoichiometric 1:1 molar ratio, the attachment of a catalytic part to the artificial nuclease construct has a potential to (i) irreversibly destroy the selected RNA target and (ii) inactivate multiple copies of RNA targets, if RNA cleavage can trigger rapid dissociation of the hybridised complex to liberate the artificial nuclease for a new act of cleavage. This would increase efficiency of the treatment, reduce the cost of therapy and allow the use of a lower drug dosage, thereby minimising adverse reactions. However, artificial nucleases developed to date still lack optimal balance between catalytic turnover and specificity of RNA cleavage. Metal ion complexes have shown a high ribonuclease activity, but tend to suffer from metal leakage and metal ion exchange reactions under intracellular conditions. Additionally, metal ion complexes capable of hydrolysing selective sequences within a structured RNA molecule have not yet been developed. These drawbacks of metal ion complexes make metal independent artificial nucleases more attractive.

Although the metal-independent nucleases exhibit a limited catalytic turnover, a greater RNA hydrolysis rate might be achieved by optimising the nuclease construct in terms of flexibility and proper orientation of the catalytic groups.

1.4.4 Peptidyl oligonucleotide conjugates

Recently, a group of researchers led by Prof. M. Zenkova (Russian Academy of Science, Novosibirsk) developed novel chemical ribonucleases with the potential to degrade a specific region of RNA in a catalytic manner99. These RNAses biomimetics have been evaluated in vitro against transcripts of human tRNA3Lys and a 96mer fragment of HIV-1 RNA (123-218 nt) comprising the primer binding sites, as well as a short synthetic RNA sequences. Design of these enzyme mimics was based on the fusion of a recognition oligonucleotide portion with a second catalytic portion comprising an oligopeptide (RL)4G with alternating basic (Arg) and hydrophobic (Leu) residues to mimic fragments of natural enzyme active site.
The design approach of these peptidyl oligonucleotides was inspired by previous work\textsuperscript{100,83} that suggested a nuclease activity of long polypeptides containing a regular alternation of hydrophobic (Ala, Leu) and basic (Lys, Arg) amino acid residues. At physiological pH, these basic amino acids exist mainly in their protonated form.

Within the ribonuclease active site the positively charged amino groups or guanidinium groups of the basic amino acid residues do not directly induce cleavage. Instead, by forming hydrogen bond interactions with the non-bridging oxygens they seem to stabilize the penta-coordinate phosphorane transition state favouring the subsequent phosphodiester transfer\textsuperscript{101}.

Barbier and Brack\textsuperscript{100,83} found that dipeptides composed of Leu-Lys sequence exhibited no cleavage activity and that introduction of proline residues or D,L-iso mer of leucine and lysine caused changes in the polypeptide structure that resulted in a substantial decline of the catalytic activity. These findings suggest that the basic long polypeptides exert nuclease activity when the residue functional groups are arranged in a given spatial conformation.

Indeed, for peptides consisting of regular alternation of hydrophobic and basic residues increase in hydrophobicity is correlated with higher catalytic activity, with poly (Leu-Lys) sequence displaying the highest cleavage rate. It is known that hydrophobicity confers upon the polypeptide chain the capability to fold into alpha helix or beta-sheet structures. In this backbone conformation, two factors seem to favour chemical activity, \textit{i.e.} (i) conformational rigidity and (ii) structural properties. Namely, the distance between two positively charged amino acids (6.9 Å) and two contiguous phosphodiester bonds of ribonucleotides (6.2 Å) should match. It is assumed that the target ribonucleotide forms a complex with the polypeptide through ionic interactions between the RNA phosphate groups and the two parallel rows of protonated basic side chains of the artificial nucleases, present in both the alpha-helix and beta-sheet structure. The phosphodiester bond cleavage occurs upon attack at phosphate groups by the basic amino acid’s free base in its side chain (Figure 1.11). The subsequent hydrolysis of the cyclic phosphorane is facilitated by the proton transfer from the nitrogen atom of a lysine side chain.
Figure 1.11: Tentative mechanism suggested by Barbier and Brack\textsuperscript{100} of an oligoribonucleotide hydrolysis by poly (Leu-Lys).

Pyshnyi and co-workers\textsuperscript{102} studied the catalytic activity of nucleotide-based artificial nuclease conjugated with short peptide sequences consisting of alternating arginine and leucine residues. When unconjugated, neither oligopeptides nor oligonucleotides possessed RNA cleavage activity but, surprisingly, were able to hydrolyse phosphodiester bonds when chemically conjugated together\textsuperscript{102,103}.

In these studies the peptide used had the sequence (ArgLeu)$_n$Gly-amide and was conjugated to deoxyribonucleotide pTCAATC complementary to the tRNA$^{\text{Phe}}$ anticodon arm (Figure 1.12).

Figure 1.12: Structure of the peptidyl-oligodeoxyribonucleotide conjugates\textsuperscript{103,102}. 
The conjugate’s catalytic activity was shown to be dependent on the oligopeptide length. Conjugates consisting of three repeats of Arg/Leu sequence (i.e. (Arg(Leu))_3 Gly-amide) conjugated to the oligonucleotide pTCAATC did not cleave RNA but showed a moderate cleavage activity (20%) when the conjugate structure was modified with a phenazinium anchor, probably favouring formation of a stable RNA/ODN duplex. Unmodified conjugates of pTCAATC with (ArgLeu)_4 Gly-amide displayed the highest cleavage rate (80%). The corresponding conjugates with phenazinium anchor cleaved RNA to a lesser extent (40%). The conjugates bearing shorter peptide (i.e. oligonucleotide-(ArgLeu)_2 Gly-amide) induced RNA hydrolysis two-fold less efficiently.

Oligodeoxynucleotides pCCAAACA conjugated to H-Gly-(Arg-Leu)_4 at the 5’end exhibited a significant catalytic activity and catalytic turnover (up to 175 in 24 hours) at the 5’GpX-3’ sites of the target RNA and as well as at the more recurrent 5’CpA-3’ and 5’UpA-3’ sites. The RNA strand scission is determined by the intrinsic instability of the scissile internucleotide linkage and does not occur within specific sequences. Videlicet, the oligodeoxynucleotide moiety is not always involved in the target recognition by base pairing; nevertheless, it plays a primary role in augmenting the cleavage activity, presumably, by affecting the peptide conformation and chemical/physical features in such a way that an optimal hydrogen bond network takes place between the arginine guanidinium group and the nucleobase nitrogen and oxygen atoms. The efficiency of catalysis was considerably enhanced by conjugation of nonadeoxyribonucleotide of unique sequence and peptide (LR)_4 G-NH₂ connected by the flexible linker comprising of three abasic deoxyribonucleotides.

Observation of sequence selectivity was also reported in a study of tRNA<sub>Lys³</sub> cleavage<sup>103</sup>. A 16-mer oligodeoxyribonucleotide complementary to the tRNA<sub>Lys³</sub> within the regions of 38-53 was conjugated to the peptide sequence H-Leu-Arg-(Leu-Arg)_4 Gly-NH₂. The peptidyl-oligonucleotide cleaved the scissile bond preferentially between the C56 and A57 at the region of the tRNA<sub>Lys³</sub> bulge<sup>103</sup>, which was three nucleotides away from the ODN/RNA double helix (Figure 1.13). In addition to the perfect alignment to the sequence 38-53 in tRNA<sub>Lys³</sub>, the oligonucleotide forms weak complexes by non-perfect base matching in the D loop as well as in the anticodon and TψC- loop. It was also observed
that the unmodified oligodeoxynucleotide moiety competitively inhibited the cleavage at this site.

The most efficient peptidyl-oligonucleotide artificial nuclease to date was developed by Mironova and co-workers\(^9\). This catalyst consisted of the nonadeoxyribonucleotide pGGATCTCTTT sequence coupled to the peptide (Leu-Arg)\(_4\)-Gly-NH\(_2\) via a spacer containing three abasic deoxyribonucleotides. The conjugate exhibited specificity towards guanine residues within the single stranded RNA regions and multiple reaction turnover similarly to natural RNase T. The spontaneous RNA linkage hydrolysis was accelerated by a factor of \(10^8\), which corresponded to only a \(10^5\) lower efficiency than RNase T1. Since the oligonucleotide moiety was not complementary to the RNA molecule and not directly involved in the RNA alignment, affinity to the target is determined by an optimal peptide conformation for the interaction and binding with RNA that is induced by the attached oligonucleotide fragments. Any changes in the peptide (Leu-Arg)\(_4\) length and the deoxyribonucleotide sequence and removal of the spacer nullified the catalytic activity and switched cleavage selectivity towards Pyr-A linkages. Therefore, the nonadeoxyribonucleotide sequence is essential for the G-X specificity. Furthermore, the linker appears to favour an intramolecular interaction network of the peptide and nucleotide moiety, which determines the correct peptide orientation towards the target phosphodiester linkage.
Studies undertaken in the research group led by Dr E. Bichenkova (University of Manchester, UK) and Prof M. Zenkova (Russian Academy of Science, Novosibirsk)\textsuperscript{104} have shown that when an antisense 17-mer oligonucleotide, complementary to the T arm of the target tRNA\textsuperscript{Phe}, is attached to a peptide containing the repeating polycationic motif [LR]\textsubscript{4} with an additional central glycine residue and/or C-terminal carboxamide modification (Figure 1.14) the cleaving activity of the conjugates is increased 6-fold. Up to 100\% site-specific tRNA\textsuperscript{Phe} cleavage was observed for the above peptidyl-oligonucleotide conjugates over a 24 h period reaching a plateau after around 8 h of reaction, with the major cuts at Pyr-A sites proximal to the target sequence.

The increase in cleavage efficacy observed for the C-terminal amide versus COOH-terminus could be attributed to the additional hydrogen bonding between the amide nitrogen and the RNA bases/phosphates which may help to shape the system into conformations with a more favourable O2’-P-O5’ angle for bond breaking. The insertion of the flexible glycine residue in the middle of the peptide chain also improved the conjugate cleavage activity, presumably by enhancing the conformational freedom of the catalytic elements, thus enhancing sampling of competent conformations for RNA phosphate binding.

![Chemical structure of the peptide (LeuArg)\textsubscript{2}Gly\textsubscript{2} with C-terminal carboxamide modification attached to the 5’-end phosphate of the 17-mer antisense oligonucleotide.](image)

**Figure 1.14:** Chemical structure of the peptide [(LeuArg)\textsubscript{2}Gly]\textsubscript{2} with C-terminal carboxamide modification attached to the 5’-end phosphate of the 17-mer antisense oligonucleotide\textsuperscript{104}.

In a later study of ‘dual’ conjugates\textsuperscript{105} a higher degree of conformational freedom to the cleaving construct was provided by the incorporation of a flexible aminohexyl linker between the peptide C-terminus and the 5’-end of the oligonucleotide, other than by the extra glycine in the middle of the peptide sequence (Figure 1.15). Gain in the peptide conformational freedom was sought to optimise the interactions of its polar groups with the
target RNA sequence to promote phosphodiester bond cleavage, which can only occur \textit{via} the ‘in-line’ geometry\textsuperscript{53} necessary for the negatively charged 2’-oxygen attack on the bridging phosphorus atom.

![Structure of the dual conjugate showing the peptide [(LeuArg)\textsubscript{2}Gly]\textsubscript{2} (in purple) modified at the C-terminus conjugated to two flanking oligonucleotide recognition motif (black) along with added linker at the 3’end\textsuperscript{105}.](image)

This change in the type of conjugation led to a specificity switch from C-A/ U-A to G-X phosphodiester bonds within the target region and showed complete site-selective cleavage of the target tRNA\textsubscript{Phe} within 4 h of reaction.

Recently, studies based on the analysis of artificial ribonuclease spatial conformations by 2D NMR and computational chemistry indicated that both oligonucleotide and peptide moieties reciprocally modulated each other’s spatial conformation positively affecting the functional property of the conjugate. Specifically, the oligonucleotide mediator appears to induce changes in the peptide conformational ensemble such that catalytic performance is greatly enhanced. This work also demonstrated that the artificial ribonucleases developed in this research group are characterized by a highly flexible structure that can assume more than one spatial conformation, presumably with varying affinity and activity against the target.

\textbf{1.4.5 Mechanism of RNA cleavage by peptidyl-oligonucleotide conjugates}

Since the cleavage products produced by the ‘dual’ conjugates\textsuperscript{105} showed electrophoretic mobilities comparable to those of oligonucleotides generated by RNase T1, we can postulate that the cleavage reaction takes place through an S\textsubscript{N}2 mechanism to form a cyclic di-anionic phosphorane intermediate, followed by release of the 5’-linked nucleotide with
concomitant formation of the 2’, 3’ cyclic phosphate\textsuperscript{106}. The RNA P-O5’ bond cleavage rate can be enhanced by introducing various functional groups into the conjugate molecule to mimic the catalytic centre of the natural enzymes, which may promote (i) deprotonation of the attacking 2’-OH, (ii) protonation of the departing 5’-O, (iii) stabilization of the anionic phosphorane intermediate by electrostatic interactions and/or hydrogen-bond anchoring of the phosphate to the protonated cleaving agent.

The peptidyl-oligonucleotide conjugates reported above contain amphipathic peptides bearing four guanidinium groups, which participate in the catalytic steps of RNA cleavage promoted by these artificial nucleases.

Although the nature of interactions between the nucleotide phosphates and the strongly basic arginine guanidinium groups present in the natural nuclease active sites\textsuperscript{107} are well known, the experimental evidence for the role of the peptide’s four guanidinium groups in the RNA cleavage is still meagre.

The hypothesis that arginine residues would form similar contacts to achieve RNA cleavage can be formulated in the peptidyl-oligonucleotide conjugates.

A tentative interpretation for the cleavage reaction is the participation of such constructs directly or indirectly in more than one catalytic step: acting as a general acid-base the guanidinium groups may shuttle protons from the attacking nucleophile\textsuperscript{108} to the non-bridging oxygen atoms in the phosphorane transition state\textsuperscript{101,109} and then to the departing 5’-oxygen\textsuperscript{110}; additionally, binding of the four guanidinium units to the RNA, presumably provide a H-bonding pattern favourable for the in-line conformation that accelerate the intramolecular transesterification at the cleavage site.

In a recent paper by Salvio et al.\textsuperscript{111} further studies were carried out on guanidinium promoted cleavage of phosphodiester bonds with the aim of better understanding of the mechanistic details of the cleavage reaction of artificial phosphodiesterases. The work shows that a requisite for an efficient catalysis of phosphate transfer is the incorporation of two guanidinium functionalities in the same molecular framework. It was shown that bifunctional artificial phosphodiesterases based on diphenylmethane derivatives provided with two guanidine units were efficient catalysts of the transesterification of an RNA model 2-hydroxypropyl-\textit{p}-nitrophenyl phosphate (HPNP) substrate \textit{via} a general acid/base
mechanism. Namely, a neutral guanidine unit acts as a general base by assisting nucleophilic attack on the 2'-OH in RNA and the protonated guanidinium unit acts as Lewis acid and electrophilic activator by forming two strong hydrogen bonds, each from different NHs to the phosphate oxygens (Figure 1.16).

In these bifunctional catalysts, a crucial feature for enhanced hydrolysis of the RNA analogues was the molecular framework connecting the guanidine units. The performance of the catalysts seemed to depend on an efficient molecular scaffold that should keep the guanidinium groups at an optimal distance in order to achieve a correct balance of pre-organisation and flexibility. This feature ensures adaptability of these synthetic enzymes to the altered substrate in the transition state. However, the resulting reduced conformational entropy may be detrimental for the catalytic efficiency.

Figure 1.16: Catalytic mechanism of two guanidinium groups intramolecular cooperation in the cleavage of the phosphodiester bond in the RNA model\(^1\); the guanidinium unit acts as an electrophilic activator by stabilising the four-membered cyclic intermediate via hydrogen bonding with the phosphate group, and a neutral guanidinium unit acts as a general base to aid deprotonation of water to use as a nucleophile to attack the phosphate unit.

The improved understanding of the mechanistic details of the phosphodiester hydrolysis in these studies provides invaluable information about the construct requirements for efficient enzyme mimics and undoubtedly helps scientists generate artificial catalysts capable of cleaving RNA with a catalytic activity comparable to that achieved by metal cation based synthetic systems.

Nevertheless, further mechanistic speculations will be needed to elucidate the role of multiple guanidinium functionalities of the oligonucleotide-based catalysts in the cleavage
of natural RNA substrate under physiological conditions, as well as the mechanism underlying the conjugate base specificity in particular regions of the RNA structure.

1.4.5.1 Gaps in knowledge

Development of RNase mimetics based on antisense oligonucleotides and peptidyl-oligonucleotides could provide useful therapeutics for treatment of various genetic disorders and for potential applications in anticancer and antiviral therapies\textsuperscript{112,113}. However, very often these oligonucleotide-based artificial nucleases lack specificity towards RNA and/or significant catalytic turnover. The study of the structural principles that govern the interactions with RNA and provide catalytic activity for this class of artificial ribonucleases may lead to the clarification of the biochemical mechanisms of the RNA cleavage and target recognition, which could then be exploited to design highly efficient therapeutic agents. Following the promising results of the recent findings further studies are therefore needed to better understand how these peptidyl-oligonucleotide conjugates interact with the target. This can be achieved by first identifying the conformations adopted by the artificial ribonucleases, both free and within the complex with RNA. This in turn will allow the role played by the two domains to be identified in the specific recognition and degradation of RNA molecules.

In this research, NMR will be used to determine the three dimensional shapes that can be adopted by these highly flexible molecules in solution. The combination of NMR with computational approaches will allow us to determine the most populated conformations, which play a pivotal role in bio-catalysis and interactions with RNA targets.

1.5 Aims and objectives of the project

The fast growing family of the peptidyl-oligonucleotide artificial ribonucleases\textsuperscript{114,104}, and the accumulating knowledge about their mechanism of recognition and cleavage of specific ribonucleotides sites\textsuperscript{105} have provided a real perspective for the design of highly selective, more potent ribonucleic acid-targeting agents of potential therapeutic value. Their efficiency as antisense agents that interfere with RNA functions is not completely satisfactory yet. Therefore, considerable attention is devoted to the improvement of the characteristic of these constructs.
A common structural element of these artificial ribonucleases is an amphiphilic, positively charged peptide whose conjugation to targeting oligonucleotides has proved to be fundamental in delivering the catalytic peptide to the appropriate RNA reactive sites.

It was postulated that the covalently attached oligonucleotides induce an active conformation in the peptide moiety through a network of ionic interactions and hydrogen bond formation e.g. between backbone oligonucleotide charged phosphate groups and peptide guanidinium groups and therefore modulate the peptide cleavage properties. Understanding the conformational behaviour of the peptide within the conjugate and how specific structural features contribute to their performance could provide valuable information to help in the development of more effective compounds and thus shorten the research cycle and synthetic chemistry workflows.

More importantly, in order to improve the biological, therapeutic and clinical potential of such agents, it is crucial to understand the structural aspects and dynamic behaviour of this family of catalytic conjugates with respect to their biological target.

We have therefore identified two key aims of this research work:

(i) The first aim is to determine the solution structure of the model conjugate \( \text{NH}_2\text{CO-} \) Gly-[Arg-Leu]_4-NH-pTCAATC and its isotopically labelled analogues using \(^1\text{H}\) NMR spectroscopy in combination with molecular dynamics simulations in order to provide the information on the orientation of the peptide within the free conjugate.

(ii) The second aim is to investigate the structural and dynamical properties of the catalytically active peptidyl-oligonucleotide conjugate 5'-h-6/14 in its complex with the highly oncogenic microRNA-21 (miR-21) in order to correlate its structure with the cleavage activity (to be evaluated from \textit{in vitro} experiments by our colleagues from the Institute of Chemical Biology & Fundamental Medicine (ICBFM), Russia).

The aim is to determine the solution structure of the model conjugate \( \text{NH}_2\text{CO-Gly-[Arg-Leu]_4-NH-pTCAATC} \) and its isotopically labelled analogues using \(^1\text{H}\) NMR spectroscopy in combination with molecular dynamics simulations in order to provide the information on the orientation of the peptide within the free conjugate.
The goal (i) will be achieved by addressing the following objectives:

1. Chemical synthesis of the non-labelled and \((^{13}\text{C}, ^{15}\text{N})\) isotopically labelled peptidyl-oligonucleotide conjugates, including:
   (a) Solid phase peptide synthesis with Fmoc Chemistry;
   (b) Selective \((^{13}\text{C}, ^{15}\text{N})\) labelling of peptide at various positions \((L_7R_8), (L_9R_{10}), (L_{11}R_{12})\) and \((L_{13}R_{14})\);
   (c) Chemical conjugation of the oligonucleotide and peptide components.
2. Isolation and purification of the conjugates by reverse phase HPLC.
3. Full characterization of the peptidyl-oligonucleotide conjugate by mass spectrometry as well as \(^1\text{H}\) and \(^{31}\text{P}\) NMR spectroscopy.
4. Assessment of the conjugate cleavage activity against \([^32\text{P}\text{-miR-24}}\) (to be carried out in the Institute of Chemical Biology and Fundamental Medicine Novosibirsk, Russia).
5. Multi-dimensional, high-field NMR (600 MHz and 800 MHz) analysis of the:
   (a) unlabelled conjugate;
   (b) site-specifically labelled \((^{13}\text{C}, ^{15}\text{N})\) conjugates
6. Assignment of the peptide and oligonucleotide \(^1\text{H}, ^{13}\text{C}, ^{31}\text{P}\) and/or \(^{15}\text{N}\) NMR signals using a range of multidimensional NMR experiments.
7. Exploration of the conjugate favoured conformations using molecular dynamics simulations to provide a converged estimate of the conjugate secondary structure.

The goal (ii) will be achieved by addressing the following objectives:

8. Synthesis of the sequence-specific peptidyl-oligonucleotide conjugates carrying the acetyl-\([\text{LeuArg}_2\text{Gly}_2]\) peptide and miR-21-recognition motifs (to be synthesised in the group of Dr. Elena Bichenkova).
9. Assessment of \textit{in-vitro} cleavage activity and site-specificity of these conjugates against oncogenic miR-21 (to be carried out by the colleagues from ICBFM (Russia)).
10. Molecular dynamics simulation of the hybridized complex between the most active conjugate and miR-21 to obtain structural insights into specific interactions between all components of the system.
The outcome of the experimental and computational results in this research work will provide a substantial contribution to the knowledge of the molecular aspects related to the RNA targeting using artificial ribonucleases, which will offer a promising tool to rational development of more potent and specific variants of the current peptidyl-oligonucleotide conjugates.

1.6 Structure and dynamics of macromolecules by NMR

Studies of bioactive molecule function (drugs or natural products) mainly rely upon knowledge of their structural properties. In most cases their biological activities are performed in solution, where both average ensemble and dynamic processes responsible for activity play an important role. Therefore, knowledge about the mode of interaction of these molecules with the target allows acceleration of conversion of lead compounds into drug candidates with improved quality of binding interactions. To achieve these goals, investigation of the structural aspect of the molecule of interest has to be undertaken both in the free-from and bound-form to the biological target.

In solution, the molecules may inter-convert among diverse possible conformations. The advances in high field NMR spectrometers and in the technology allow the structural properties of free compounds and the target-bound form to be studied at high resolution.

1.6.1 Introduction to NMR methods

The majority of the chemical elements exhibit more than one isotopic forms (mass number), each characterised by a different magnetic property. Nuclei of all atoms are made of protons and neutrons, which are characterised by a spin quantum number, \( I \), equal to \( \frac{1}{2} \).

The isotopes such as \( ^{12}\text{C} \), \( ^{16}\text{O} \) have even numbers of protons and even numbers of neutrons: the particles can pair up with antiparallel spins resulting in no overall dipole and thus a net zero spin. These isotopes have no magnetic moment and consequently do not exhibit any NMR absorption spectra. If the numbers of protons and neutrons in a given isotope are both odd, the overall spin is an integer greater than zero, because the total number of unpaired particles is even (\( i.e. ^{2}\text{H} \) and \( ^{14}\text{N} \) have \( I = 1 \)).
The isotopes that are even-proton-odd-neutron, or vice versa, possess half-integer spin, i.e. \( I = \frac{1}{2} \) such as \(^1\text{H}, \ ^{13}\text{C}, \ ^{15}\text{N}, \ ^{19}\text{F} \) and \(^{31}\text{P} \). These nuclei have a magnetic dipole and thus are observable by NMR.

The spinning nuclei are associated with a magnetic moment, \( \mu \), originated by the motion of the magnetic dipole. The magnetic moment of a spin is directly connected to its spin angular momentum \( P \) by the magnetogyric ratio, \( \gamma \), which is constant for any given nuclei, via the relation:

\[
\mu = \gamma P
\]

Equation 1.6-1

The spin angular momentum \( P \) is a vector quantity with a quantized magnitude and orientation. In the absence of a magnetic field the nuclear spins adopt random orientations since all the spin states have the same energy.

In a static magnetic field \( B_0 \) this energy degeneracy is removed and the magnetic moment orientation is no longer arbitrary, but has \( 2I+1 \) projections along the static field direction (by convention the \( z \) axis).

For the spin-\( \frac{1}{2} \) nucleus, the two projections, which correspond to the two spin states at higher (the \( \alpha \) state) and lower energy (the \( \beta \) state), correspond to the parallel or antiparallel orientations adopted by the magnetic moment of the nucleus with respect to the static magnetic field (Figure 1.17).

**Figure 1.17:** Spin states: In the presence of a static magnetic field \( B_0 \) nuclear spin with a magnetic moment may adopt \( 2I+1 \) possible orientations relative to \( B_0 \). For spin-\( \frac{1}{2} \) nuclei, there are two possible orientations, either parallel (\( \alpha \)) or antiparallel (\( \beta \)).
When placed in a strong external magnetic field, the spins’ magnetic moment experiences a torque oriented along the field direction due to its magnetic moment being out of alignment with the field (Figure 1.18).

**Figure 1.18**: Larmor precession. The nuclear spin precesses about the magnetic field $B_0$ axis with an angular frequency, known as the Larmor frequency, which is dependent on the magnetic field strength and the spin magnetogyratic ratio.

The nucleus’s axis of rotation will precess about the applied field axis at a rate, defined by the angular velocity $\omega$ (rad/s) or frequency $\nu$ (hertz), that is proportional to both the field strength ($B_0$) and the spin magnetogyratic ratio ($\gamma$) (Equations 1.6-2 and 1.6-3). This motion is known as the resonance or *Larmor precession* of the spin.

$$\omega = -\gamma B^0$$  \hspace{1cm} \text{Equation 1.6-2}

or

$$\nu = (\gamma B_0)/2\pi$$  \hspace{1cm} \text{Equation 1.6-3}

Due to interference from the strong, main external magnetic field $B_0$, the frequency of the spins can only be observed when the net magnetisation lies at an angle to $B_0$.
The perturbation of the nuclear spin alignment occurs through the transfer of energy from an electromagnetic radiation pulse to the spin, which takes place when the oscillating frequency of the applied field matches that of the spin Larmor precession frequency, a physical phenomenon called resonance. At the resonance frequency there is exactly the correct energy to flip the spins (nuclei) from the lower energy state to the excited state. The energy involved in the two states transition being given, according to the Planck relationship, by:

\[ \Delta E = h\nu = (h\gamma B_0)/2\pi \quad \text{Equation 1.6-4} \]

In NMR experiments, this is accomplished by another magnetic field \( B_1 \), whose axis is perpendicular to \( B_0 \). \( B_1 \) rotates the net magnetisation into the transverse plane, conventionally labelled \( x-y \). This arrangement allows the net magnetisation to move 90° away from the \( z \)-axis and thereby generate a significant NMR signal (Figure 1.19).

**Figure 1.19:** Excitation with a rf pulse: The application of a rf field \( B_1 \) (for example along the \( x \) axis), rotates the spin vectors from the \( z \) axis towards the \( x-y \) plane. The vectors precess about the \( z \) axis at their Larmor frequency. The rotating magnetization vector induces voltage in the coil surrounding the sample; these are the electrical signals detected to produce the observed NMR signals.

### 1.6.2 Common NMR Isotopes

Protons (\(^1\text{H}\)) are the most studied nuclei since they are particularly abundant in nature (99.985%), occupy different positions in a molecule, and show high sensitivity to NMR.
The $^{31}\text{P}$ nucleus is the only phosphorus isotope occurring in biomolecules (100% abundant) and is mainly present in the phosphate group of nucleic acids; however its sensitivity to NMR is much lower than $^1\text{H}$. The $^{13}\text{C}$, $^{15}\text{N}$ nuclei are used less extensively in NMR analysis due to their low abundance, 1.1% and 0.37% respectively, and poor sensitivity (higher number of scans is required than for $^1\text{H}$). Nevertheless, their use is increasing in molecule isotope labelling, particularly nucleotides to resolve resonance overlap$^{118,119}$.

1.6.3 Chemical shifts

The radiofrequency pulse created by the NMR spectrometer magnet induces transitions between the two allowed spin states $+1/2$ (aligned with the applied magnetic field) and $-1/2$ (against it). Spins of the same nucleus type resonate at similar frequencies; however resonances of different nuclei of a given type do not occur at the same frequency, which can significantly vary depending on the chemical and electronic environment. These differences, described as chemical shifts ($\delta$), arise from the shielding effect of the induced magnetic field generated by sigma electrons in the surrounding bonds which opposes the external magnetic field. The chemical shifts scale linearly with the strength of the applied field; therefore, the resonance frequency of the nuclei increases with increasing strength of the magnetic field. For this reason, the chemical shifts are described as resonance frequencies relative to that of a reference compound defined to be zero (due to the high level of shielding of the nuclei in these compounds), which, for $^1\text{H}$ and $^{13}\text{C}$ NMR, is the methyl groups in trimethylsilyl propionate (TSP), 85% phosphoric acid ($\text{H}_3\text{PO}_4$) in D$_2$O for $^{31}\text{P}$, and nitric acid (HNO$_3$) for $^{15}\text{N}$.

Chemical shift values of nuclei are usually reported in parts per million (ppm) in literature and are defined in terms of difference in frequencies between the sample signal ($v_0$) and the reference compound $v_{\text{ref}}$. Dividing by the frequency of the reference compound, eliminates this dependence on the magnetic field strength used to measure $\delta$ (Equation 1.6-5).

$$
\delta(\text{ppm}) = 10^6 \times \frac{v_0 - v_{\text{ref}}}{v_{\text{ref}}} \quad \text{Equation 1.6-5}
$$
Thus, a frequency difference of 600 Hz for a proton at 400 MHz corresponds to a chemical shift difference of \((600/400) = 1.5 \times 10^{-6}\), which, expressed as parts per million, is 1.5 ppm. The separation of the chemical shifts dictated by the changes in local chemical environment is crucial for assignment and interpretation of resonance signals.

1.6.4 Spin-Spin or \(J\)-coupling

The nuclear spins are also sensitive to the small magnetic field generated by neighbouring non-equivalent nuclei. The magnetic interactions with other spins in their surroundings are described as spin-spin coupling, also referred to as scalar coupling or \(J\)-coupling. Spin-spin coupling is transmitted primarily through covalent bonds and causes splitting of the NMR signals into two or more peaks, whose patterns provide insight into molecule structure and conformation. Such couplings fall off rapidly upon increasing the number of bonds and are not observable between spins separated by more than 4-5 covalent bonds. Surprisingly, \(J\)-couplings are also detected across hydrogen bonds\(^{120}\) and van der Waals interactions\(^{121}\). The splitting pattern of a nucleus \(j\) depends on the energy state of the non-equivalent neighbouring nucleus \(i\). A high energy state (aligned with the external field) shifts the resonance frequency slightly downfield, whereas a low energy state (against the field) shifts the resonance frequency up-field. The presence of several neighbouring non-equivalent nuclei, gives rise to further splitting of the nucleus \(j\) resonance and increases complexity of the splitting pattern.

The spin-spin coupling effect is independent of the applied magnetic field strength. This effect can be quantified by using the coupling constant \((J)\), which is the spacing between two adjacent lines in a split signal and is measured in units of Hertz (Hz). The coupling constant can be calculated by multiplying the distance between two adjacent split lines by the spectrometer resonance frequency on which the spectrum is obtained. Coupling constants for vicinal protons are 7 to 8 Hz, and around zero for more than four intervening bonds. Unsaturated bonds enhance coupling as well as planar zig-zag configuration over 4 to 5 bonds.

Spin-spin coupling is mutual: both coupled nuclei show \(J\)-coupling splitting and the split lines of their resonance signals have the same coupling constant. In most cases this allows
one to distinguish if two closely spaced lines are coupled \( (J) \) or shifted \( (\delta) \) by observing the spectrum appearance.

When the chemical shift differences \( (\Delta \nu) \) of two or more non-equivalent nuclei are similar in magnitude to the coupling constants, the analysis of the splitting pattern becomes more complicated. In particular when \( \Delta \nu/J \) is \(< 5\) the resonance intensity and position are perturbed: intensity of the outer lines becomes weaker and the inner peaks become larger and the multiplets are skewed towards each other.

The most useful \( J \)-couplings for conformational analysis are those involving protons separated by three covalent bonds \( (3J_{HH}) \), via H-C-C-H bonds. The value of a \( 3J \) -coupling constant varies with the dihedral angle \( (\theta) \) formed by the three covalent bonds according to the ‘Karplus relation’ \(^{122}\) (equation 1.6-6):

\[
3J = A \cos^2 \theta + B \cos \theta + C \quad \text{Equation 1.6-6}
\]

where the constants \( A \), \( B \) and \( C \) values are empirically derived for each type of nuclei involved in the coupling.

Typical values reported in literature for peptides and proteins backbone amide-alpha protons\(^{123}\) \( J \)-couplings \( (3J_{HNHa}) \) are \( A = 6.4 \) Hz, \( B = -1.4 \) Hz, \( C = 1.9 \) Hz.

This relationship within coupling constants and dihedral angles finds valuable applications in the determination of peptides and proteins structures by NMR. For example, the three-bond couplings between the amide proton H\( ^{N} \) and H\( ^{\alpha} \) in the backbone of a protein chain provide information on the protein secondary structure as it is directly related to the backbone dihedral angle phi \( (\phi) \) (Figure 1.20-a) by the Equation 1.6-7.

\[
3J_{HNHa} = A \cos^2(\phi - 60) + B \cos(\phi - 60) + C \quad \text{Equation 1.6-7}
\]

In particular, the two most common types of secondary structure in proteins \( (i.e. \) alpha helices and beta sheets) have characteristic values for \( \phi \), approximatively \(-60^\circ\) and \(-120^\circ\) respectively, which result in significant differences in the corresponding \( 3J_{HNHa} \) value (Table 1.2 and Figure 1.20-b).
Table 1.2: Parameters for peptide and protein conformations

<table>
<thead>
<tr>
<th>Secondary structure</th>
<th>$\phi$</th>
<th>residues per turn</th>
<th>$^3J_{HNHA}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>right-handed alpha helix</td>
<td>-57$^\circ$</td>
<td>3.6</td>
<td>3.9 Hz</td>
</tr>
<tr>
<td>right-handed $3_{10}$ helix</td>
<td>-60$^\circ$</td>
<td>3.0</td>
<td>4.3 Hz</td>
</tr>
<tr>
<td>antiparallel beta sheet</td>
<td>-139$^\circ$</td>
<td>2.0</td>
<td>8.9 Hz</td>
</tr>
<tr>
<td>parallel beta sheet</td>
<td>-119$^\circ$</td>
<td>2.0</td>
<td>9.7 Hz</td>
</tr>
<tr>
<td>left-handed alpha helix</td>
<td>57$^\circ$</td>
<td>4.4</td>
<td>6.9 Hz</td>
</tr>
</tbody>
</table>

Thus, values of $^3J_{HNHa}$ far away from the extrema expected from the Karplus relation are not used for determination of secondary structure because they may reflect motional averaging of folded and unfolded conformations.$^{124}$

Figure 1.20: (a) the phi ($\phi$) dihedral angle describes the rotation of the polypeptide backbone about the N-C$\alpha$ bond and involves the C(O)-N-C$\alpha$-C(O) bonds. (b) Karplus curve describing the variation of $^3J_{HNHa}$ with backbone dihedral angle $\phi$. 

73
$^3J_{\text{HNH}_\alpha}$ values smaller than 6 Hz are diagnostic of alpha helix structure, while couplings larger than about 8 Hz generally indicate an extended beta sheet conformation. $^3J_{\text{HNH}_\alpha}$ values in the range 6-8 Hz arise from extensive backbone motion and occur generally in short, linear peptides and in random coil segments of proteins.

1.6.5 Relaxation in NMR Spectroscopy

An understanding of spin relaxation is important for correct interpretation of NMR spectra. Relaxation is a process by which a spin excited by the radio-frequency pulse returns to the equilibrium state along the direction of the static magnetic field. This process is described by two time constants, the ‘spin-lattice’ relaxation time ($T_1$) and spin-spin relaxation time $T_2$. $T_1$ corresponds to the process of magnetization vector return to equilibrium along the applied magnetic field, conventionally the $z$-axis (Figure 1.21). The time required for spins to relax fully after excitation by a 90° pulse is at least $5 \times T_1$.

![Figure 1.21: Longitudinal relaxation ($T_1$) in the rotating frame representation. The net magnetisation relaxation diminishes the $y$-component and re-establishes the $z$-component, until magnetisation comes to its equilibrium position finally along the $z$ axis.](image)

$T_2$ describes the decay of the excited magnetization $x$ and $y$ components, perpendicular to the applied magnetic field (Figure 1.22). $T_2$ is directly related to the observed width of NMR lines. The width of the signal at half maximum height ($\nu_-(1/2)$) is given by $\nu_-(1/2) = 1 / \pi T_2$, where short $T_2$ (fast decay) leads to broad lines, long $T_2$ (slow decay) to sharper signals.

Spin relaxation is not a spontaneous process, but it is driven by local magnetic fields fluctuating at the Larmor frequency of the spins. The magnetic fields which induce the nuclear spin transitions are due to the spin interactions with the surroundings (the ‘lattice’).
The most important mechanism that causes relaxation is dipolar coupling. Relaxation arises from intramolecular or intermolecular magnetic interactions between nearby spins. The local magnetic field experienced at one spin as a result of the nearby spin is modulated by random molecular tumbling in solution and therefore are time-dependent.

The random molecular motion rate can be defined by the rotational correlation time ($\tau_c$), which is the average time it takes the molecule to rotate one radian ($\approx 60^\circ$) and is in the order of picoseconds.

The correlation time ($\tau_c$) is highly dependent on the molecule size: small molecules tumble rapidly and possess shorter correlation time, while large molecules move more slowly and have long $\tau_c$. Longer correlation times result in more effective dipole-dipole relaxation.

**Figure 1.22:** Transverse relaxation ($T_2$). Local molecular environment differences cause spins to precess with slightly different Larmor frequencies, diminishing the $x$ and $y$ components and eventually leading to magnetisation loss in the $x$-$y$ plane.

### 1.6.6 Fourier Transform in NMR

In an NMR experiment, the spins are excited by a 90° short pulse of radio-frequency radiation, which tips the magnetization from the $z$ axis to $x$-$y$ plane. The frequencies re-emitted by the spins in the excited state induce voltage in the detecting coil surrounding the sample, and give rise to a signal which oscillates at the Larmor frequency, referred to as free induction decay (FID signal). The FID signal is acquired as a function of time domain. Conversion of the time domain signal to frequency domain is done by Fourier Transform
(FT), which results in a spectrum of the intensity plotted against the frequency (Figure 1.23).

![Schematic diagram of a time-domain NMR signal (FID signal) and its equivalent frequency NMR spectrum on Fourier Transform.](image)

**Figure 1.23:** Schematic diagram of a time-domain NMR signal (FID signal) and its equivalent frequency NMR spectrum on Fourier Transform.

### 1.6.7 Multidimensional NMR

In 1D NMR multiplets of complex molecules, signals corresponding to the various functional groups can be heavily overlapped, making interpretation of the spectrum difficult. In two dimensional NMR this challenge is circumvented by introducing a second frequency variable in the plot, which yields spectra easier to resolve and provides more informative data. In 1D NMR experiments spins are hit with a pulse of radio-frequency under an external magnetic field, and the re-emitted signal, through Fourier transform, results as 1D spectrum as a function of resonance frequency.

Spectra obtained by a 2D experiment can be divided into a sequence of four periods (or times) as shown in Figure 1.24. The preparation period consists in one or more pulses that excite the sample, in most cases to generate transverse magnetization. The consecutive period is the evolution period $t_1$, during which the spins evolve at their own frequency in the $x$-$y$ plane; FID is recorded at each experiment with an incremented delay time to generate an indirectly-detected frequency dimension. The mixing time is a combination of a further pulse (or pulses) and delay periods which induce magnetization transfer from one spin to another as a result of either scalar-coupling or NOE interaction in the case of proximal nuclei. The detection period consists in recording a series of FIDs as a function of an acquisition time $t_2$ for each different $t_1$ value.
This sequence of rf pulses, delays and acquisition times is called a pulse sequence and the exact nature of the excitation pulse and timing parameters determines the information contained in the spectrum.

![Diagram of 2D pulse sequence of a FT-experiment.](image)

**Figure 1.24:** Diagram of 2D pulse sequence of a FT-experiment.

A two-dimensional spectrum with two frequency axes can be produced by Fourier transformation of the time variables which are converted to frequency signals. This is indicated schematically in Figure 1.25.

The additional information is provided by the spins exchanging magnetization during the mixing time, which can occur through two different mechanisms: scalar coupling or dipolar interaction.

![Diagram of 2D NMR spectrum and transformations.](image)

**Figure 1.25:** Left: 2D NMR spectrum as a function of time ($t_1$ and $t_2$). Middle: the data upon $t_2$ Fourier transform into frequency $\omega_2$. Right: 2D NMR spectrum upon $t_1$ transformation to $\omega_1$ as a contour plot (concentric circles represent peak intensity).

2D spectra show two types of peaks. A diagonal of signals, arising as a result of magnetization coherences not exchanged during the mixing time, correspond to peaks in a 1D experiment and have the same resonance frequencies in both dimensions. The cross
signals (peaks that are not on the diagonal) are originated by exchanged magnetization between spins during the mixing time, whose evolution is at different frequencies during $t_1$ and $t_2$.

Three-dimensional experiments overcome problems associated with overcrowded NMR spectra of large molecules, due to the high content of $^1$Hs in the sample, by spreading the coherences over three different frequency dimensions.

A three-dimensional NMR spectrum can be obtained by inserting an extra indirect evolution and mixing time between the first mixing time and detection in a 2D pulse sequence (Figure 1.26) to obtain data $t_1$, $t_2$ and $t_3$. Fourier transform of the data in $\omega_1$, $\omega_2$ and $\omega_3$ gives the three-dimensional NMR spectrum, which can be analysed as 2D planes.

There are no additional resonances (or cross peaks) in a 3D spectrum, on the other hand the information provided is presented in the most resolved form.

Figure 1.26: 3D pulse sequence. 2D experiments are combined to create a 3D experiment (i.e. 3D HNHA). In a 3D experiment the first two evolution periods, which are incremented, correspond to the two indirectly-detected time domains ($t_1$ and $t_2$), and the third, which is the acquisition time, corresponds to the directly-detected time domain ($t_3$). After a Fourier transform (FT) of the three time domains the 3D spectrum has three frequencies, $\omega_1$ ($^1$H), $\omega_2$ (X) and $\omega_3$ ($^1$H).

In fact, the cross peaks ‘density’ of the 2D experiments is diminished, as a result of the resonances being filtered and spread out over the additional $\omega_1$, $\omega_2$ and $\omega_3$ dimensions. The high sensitivity of a 3D NMR experiment results from the high efficiency of the
heteronuclear magnetisation transfer which is accomplished through relatively large scalar, one-bond $J$ couplings (i.e. $\text{^1J}_{\text{CH}} = 119\text{ Hz}$, $\text{^1J}_{\text{NH}} = 91\text{ Hz}$).

### 1.6.8 NMR techniques

#### 1.6.8.1 Water suppression

Most compounds of biological or medicinal interest, such as proteins and nucleic acids, are studied by NMR in deuterated solvents to avoid the signal of the solvent which is present in large excess. In order to observe vitally important labile protons, however, the NMR $^1\text{H}$ spectra must be recorded in water ($90\% \text{^1H}_2\text{O}/10\% \text{^2H}_2\text{O}$) solutions.

Water signals are very intense and broad due its high proton concentration (at approximately 105 M), which may mask resonances of the molecule of interest.

A number of methods for solvent suppression are available to successfully reduce the magnitude of the intense water singlet.

Presaturation is the simplest and classical procedure to suppress the water signal$^{126}$ (Figure 1.27-a).

A low-power continuous radiofrequency irradiation is applied at the water frequency immediately prior to the pulse sequence, rendering the water proton spins saturated such that no signal can fully accumulate and therefore be observable. The disadvantage of this method is that the intensity of the exchangeable protons is affected and the resonances close to the water frequency also are obliterated. This problem can be overcome effectively by altering the sample temperature and pH. The water signal undergoes large variations in chemical shift due to fast proton exchange, whereas the other proton resonances are less affected.

The most effective method to suppress unwanted solvent resonance in the spectrum is based on the use of PFGs, which produce zero net solvent magnetisation.

Most commonly employed approach is the water suppression through gradient tailored excitation (WATERGATE). The solvent suppression is achieved using a single PFG spin-echo experiment of sequence $G_1$-$S$-$G_1$, in which a selective $180^\circ$ pulse is introduced.
between two symmetrical pulsed field gradients\textsuperscript{127} Figure 1.27-b). This provides a $180^\circ$ inversion to proton resonances except for the water resonance which is left unchanged.

![Figure 1.27](image)

\textbf{Figure 1.27}: Solvent suppression methods. (a) Presaturation, involves the application of a long, low-power pulse on the solvent frequency prior to the pulse sequence. (b) WATERGATE, uses a gradient spin-echo sequence which combines a single selective $180^\circ$ pulse and two pulsed field gradients to dephase the net solvent magnetisation. (c) Excitation sculpting, applies a double gradient spin-echo sequence.

After excitation of all resonances by the first $90^\circ$ pulse the first PFG dephases all coherences. Then, a selective $180^\circ$ pulse provides a $180^\circ$ inversion to proton resonances except for the water resonance which is left unchanged.

The second PFG, which acts in opposition to the first PFG, rephases all the resonances that have experienced the selective inversion pulse and are therefore retained, whereas the water coherence is not refocused, instead, it remains fully dephased in the transverse plane.

This selective technique yields better suppression compared with other methods (\textit{i.e.} presaturation) and allows the observation of exchangeable protons without loss in signal intensity, although signals very close to the water resonance are also suppressed.
An improved approach, termed ‘excitation sculpting’, makes use of a double PFG spin-echo sequence for selective removal of the solvent resonance (Figure 1.27-c). The gradient-echo is repeated with a different gradient strength, $G_2$, leading to removal of any remaining phase errors and improved phase and amplitude characteristics.

### 1.6.9 COSY: Correlation Spectroscopy

The COSY experiment correlates the chemical shift of scalar-coupled protons (mediated by electrons in a covalent bond) and its main purpose is to obtain structural information from the determination of nuclear connectivity within a molecule.

The 2D COSY experiment requires two 90° pulses separated by a $t_1$ delay, which is changed incrementally\(^\text{128}\) (Figure 1.28). The first pulse produces transverse magnetization in the $x$-$y$ plane on each nucleus, which then evolves at its characteristic precession frequency during the evolution time. Once the spins have come to thermal equilibrium a second 90° pulse is applied, which induces magnetization transfer to each coupled nucleus. The nuclei then evolve under the influence of their chemical shift frequency and coupling constants during detection. The correlation peaks appear at the intersection of frequencies of two nuclei that are connected by scalar coupling.

\[ ^3\text{H} \]

**Figure 1.28:** 2D COSY. Comprises of a two-pulse sequence with an evolution period lying in between. This experiment correlates the chemical shift of nuclei of the same kind through magnetisation transfer between scalar $J$ coupled nuclei following the second 90° pulse.

The diagonal peaks correspond to the peaks of nuclei that couple to themselves, namely magnetization is not transferred to another nucleus but remains on the same spin upon application of a second 90° pulse. This passive coupling during both the applied first and
second 90° pulses generate multiplet peaks that are in-phase. The cross peak, that occurs symmetrically across the diagonal, correlates the chemical shift of the two diagonal peaks and indicates coupling between the two nuclei. This allows assignment of the peak to adjacent nuclei, providing additional information about the molecular structure.

Under conditions of high resolution, fine structure associated with correlation peaks can be observed and this can be used to calculate the coupling constants that are obscured in overcrowded 1D NMR spectra.

The multiplet peaks arising from active coupling are naturally antiphase, namely one-half signal is up in the multiplet and the other half down\textsuperscript{129}. When the couplings are quite small compared to the signal width the spectral negative and positive component can cancel one another out, causing cross peaks to become very weak and hence difficult to identify.

1.6.10 DQF-COSY: Double-Quantum Filtered Correlation Spectroscopy

The DQF (double-quantum filtered)-COSY experiment is a variation of the conventional COSY experiment\textsuperscript{130}, in which a double quantum filter is incorporated to allow observation of signals from all coupled spin systems but removes the intense peaks of singlets, which do not share a $J$-coupling.

The pulse sequence of a DQF-COSY differs from the standard COSY experiment by the addition of a DQF after the variable evolution period (Figure 1.29). Thus, the DQF-COSY pulse sequence consists of three 90° pulses. The first 90° pulse creates single-quantum coherences which evolve during $t_1$ period of time. After the second pulse is applied to generate double-quantum coherences a final pulse converts these resonances back into single quantum coherences, which are visible to the detector.

The advantage of DQF-COSY is that diagonal multiplets have anti-phase absorption-mode line shapes and they are removed from uncoupled nuclei. Consequently, the resolution of the cross peaks adjacent to the diagonal peaks is greatly enhanced. This aids the analysis of crowded spectra and allows multiplet fine structure to be studied to extract coupling constant information.
1.6.11 TOCSY: Total Correlation Spectroscopy

The 2D TOCSY (Total Correlation Spectroscopy) experiment permits to analyse scalar $J$ couplings networks and to establish correlations between all protons within a given spin system, provided every intervening proton is directly coupled to adjacent protons.

TOCSY is able to relay magnetisation from one proton to its neighbour along a chain of protons, i.e. A-B-C-D, as long as A-B, B-C and C-D are directly coupled. The propagation of magnetisation along the chain, therefore, provides correlations between all protons within the same spin system (i.e. A-C and A-D).

A TOCSY spectrum is a proton-proton 2D chemical shift correlation map in which cross peaks arising from both direct and relayed connectivities are present.

TOCSY data are acquired in phase-sensitive mode and the cross-peaks have positive in-phase absorption-mode line shapes.

The TOCSY pulse sequence follows the same principles of a COSY experiment. After exciting the spins with a 90° pulse, transverse magnetisation evolves during evolution period $t_1$. In place of a single mixing pulse, the TOCSY experiment employs an isotropic mixing sequence (also known as a spin-lock) to force the relayed magnetisation transfer into a coupling spin network (Figure 1.30).

The length of the mixing time allows determination of the extension of the magnetisation propagated to determine the extension of the magnetisation along the coupled chain. Mixing times in the range 30-60 ms enable correlations to be observed between protons separated by four or fewer covalent bonds. Longer mixing times (> 60 ms) are required to promote the long-range magnetization transfer up to 5 or 6 bonds (i.e. that of a peptide.
amide proton to the protons at the end the side chain), but usually is less than 120 ms to avoid sensitivity losses caused by relaxation of spin-locked magnetisation.

This proves useful in the analysis of oligonucleotide and peptide spectra, molecules that are comprised of discrete subunits (spin systems) *i.e.* nucleotide sugar rings and amino acid side chains.

*Figure 1.30:* 2D TOCSY. The pulse sequence follows the same principle as the COSY sequence, but in place of the second 90°-pulse a *spin-lock* sequence is applied, a continuous sequence of 180° pulses separated by short periods. This experiment shows chemical shift correlations between nuclei that are within the same spin system.

**1.6.12 NOESY: Nuclear Overhauser Effect Spectroscopy**

Nuclear Overhauser Effect Spectroscopy (NOESY)\(^1\) is a 2D homonuclear \(^1\)H NMR technique which allows the analysis of biomolecular conformations by probing through-space interactions among proximal spins. The NOE cross-peaks can be distinguished from those observed in a COSY spectrum, for instance, where the spin-spin correlation arises from 2- or 3- bond \(^J\)-coupled spins.

NOESY exhibits cross-relaxation between spins close to one another in space that possess magnetic dipoles, and is identified by detecting mutual spin-lattice relaxation caused by dipolar coupling. To achieve this, the nuclei are excited by three 90° pulses (Figure 1.31). The first 90° pulse generates transverse magnetization, which precesses during the evolution period \(t_1\). The time delay is incremented by a certain amount to provide resonance frequency information in the \(t_1\) domain. The second pulse gives rise to magnetization in \(z\) axis again. In the following mixing period, \(\tau_m\), the spin relaxation occurs between spatially close nuclei, and stimulates magnetization changes in adjacent spins.
through a dipole-dipole mechanism. The $\tau_m$ value is kept constant during a NOESY experiment.

**Figure 1.31:** 2D NOESY. The basic pulse sequence consists of three $90^\circ$ pulses. The first and second pulses are separated by a variable evolution $t_1$ period, followed by a constant mixing period $\tau_m$, during which cross relaxation occurs. The third $90^\circ$ pulse creates transverse magnetisation from remaining magnetisation, which is detected. The NOE is created by the exchange of the magnetisation between nuclei that share a mutual dipolar coupling during relaxation, which are up to 5 Å apart.

Finally, the last $90^\circ$ pulse creates transverse magnetization which is detected as a function of the evolution time $t_2$. Cross peaks appear at the frequencies of the spins involved if the inter-nuclear distance is $\leq 5\text{Å}$ allowing mutual relaxation to occur. The intensity of the cross-peaks depends on the distance between the spins and the molecular internal motion, which is described by the correlation time, $\tau_c$.

The intensity of the NOESY cross-peak formed by the interacting protons $i$ and $j$ is proportional to the cross-relaxation rate $\sigma_{ij}$ and is related to the interproton distance by the following equation:

$$\sigma_{ij} = \frac{\hbar^2 \gamma^4}{10} \frac{1}{\tau_c^6} \frac{6\pi}{1 + 4(\omega \tau_c)^2}$$  \hspace{1cm} \text{Equation 1.6-8}$$

where $\sigma_{ij}$ is the cross-relaxation rate between the interacting protons $i$ and $j$, $r_{ij}$ the distance between them, $\omega$ is the spectrometer frequency in radians and $\tau_c$ is the correlation time of the molecular species.
The correlation time, $\tau_c$, also referred to as the inverse of the rate of molecule tumbling in solution, depends in large part on the molecule weight and shape, solvent viscosity and temperature. For large molecules, the magnitude of the NOE effect at short mixing times grows as correlation time increases as well as being proportional to the inverse of the distance between two spins $i$ and $j$ to the 6th power, $(r_{ij})^6$.

Thus, it is possible to obtain inter-nuclear distances by comparing the NOE signal for a fixed distance of a proton pair, such as the cytosine H5 and H6 contour lines$^{133}$ (2.45 Å) with a NOE signal of an unknown inter-proton distance by the following equation:

$$ r_{ij}^6 = r_{\text{ref}}^6 \times \left( \frac{\text{NOE}_{\text{ref}}}{\text{NOE}_{ij}} \right) $$  

Equation 1.6-9

This approach is often called the ‘isolated spin-pair approximation’, which is based on the assumption that NOE is caused by a cross-relaxation between two isolated nuclei $i$ and $j$. However, in actual chemical and biological systems, each nucleus is normally incorporated into a complex net of interacting nuclei, leading to additional magnetisation transfer mechanisms through indirect cross-relaxation pathways via closely located spins. These indirect cross-relaxation mechanisms are often called ‘spin diffusion,’$^{134}$ which can be strongly affected by long mixing time $\tau_m$. Spin-diffusion may cause imprecise estimation of distances between interacting spins $i$ and $j$, and ultimately lead to inaccurate description of 3D structures.

In order to minimise the impact of spin-diffusion mechanisms one can use relatively short $\tau_m$ in the NOESY experiments, but this may lead to a loss of valuable long-distance NOE interactions and thus compromise the structural analysis. Instead, it is possible to use the full relaxation matrix approach (i.e. MARDIGRAS), which allows one to measure distances between nuclei more precisely taking into account all possible multi-spin diffusion mechanisms$^{135}$. 

86
1.6.13 HSQC: Heteronuclear Single-Quantum Correlation

The HSQC (heteronuclear single-quantum correlation)\textsuperscript{136} experiment makes it possible to establish correlations between the proton coherences and the directly bonded ($J$-coupled) heteronucleus, X, coherences. This experiment is an inverse technique, which involves the transfer of magnetisation from the proton to its attached NMR visible heteronucleus isotope (which is usually $^{13}$C or $^{15}$N) and finally back to proton for detection.

The resulting 2D HSQC spectrum has frequency of protons on the x-axis (F2 dimension) and frequency of heteronuclei on the y-axis (F1 dimension) and each cross-peak represents a proton at a specific frequency that is bound to X via a direct heteronuclear coupling $^{1}J_{HX}$.

The HSQC pulse sequence consists of four distinct steps (Figure 1.32): proton magnetisation is transferred to the heteronucleus through $J$ coupling; the magnetisation then evolves under the effect of the heteronuclear chemical shift during the following evolution period, with a 180° pulse applied on the proton at its midpoint refocusing $^{1}J_{HX}$ couplings and hence removes the coupling in $f_{1}$ dimension (X); after a time delay ($t_{1}$) the heteronuclear magnetisation is converted to in-phase proton magnetisation; proton observation with X decoupling.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{hsqc_pulse_sequence}
\caption{2D HSQC. The pulse sequence comprises a preparation period, $t_{1}$ evolution which allows observation of the indirect X (i.e. $^{13}$C or $^{15}$N) frequency to be observed, mixing period and $t_{2}$ evolution period for direct $^{1}$H frequency detection with X decoupling. The HSQC experiment correlates X and proton chemical shifts via the direct heteronuclear coupling $^{1}J_{HX}$.}
\end{figure}
1.6.13.1 CT-HSQC: Constant-time Heteronuclear Single-Quantum Correlation

The 2D Constant-Time HSQC (CT-HSQC)\(^{137}\) experiment, a version of the standard 2D HSQC, addresses the issue of multiplet patterns of signals in the \(^{13}\)C dimension arising from aliphatic homonuclear one-bond \(^1\)J\(_{CC}\) couplings which reduce spectral resolution. In this experiment a constant-time period is introduced to replace the variable \(^{13}\)C evolution period allowing refocussing of the \(^1\)J\(_{CC}\) evolution. The resulting CT-HSQC spectrum shows cross-peaks with collapsed multiplet patterns and singlet responses for the \(^{13}\)C resonances in the \(f_1\) dimension.

This experiment also provides information on the number of directly-bonded carbon nuclei to the observed \(^{13}\)C nucleus. If the constant-time period (\(T\)) length is \(T = 1 / ^1\)J\(_{CC}\) (i.e. 27 ms) the sign of the signal of \(^{13}\)C nuclei attached to an even number of aliphatic carbons will appear opposite (negative) to those attached to an odd number (positive), whereas \(T = 2 / ^1\)J\(_{CC}\) (i.e. 54 ms) all signals will have the same sign.

1.6.14 2D HSQC-NOESY

The 2D HSQC-NOESY is an inverse experiment which combines the enhanced identification of \(^{13}\)C labelled units from the HSQC experiment with the internuclei interactions detection from the NOESY building block\(^{138}\).

The 2D HSQC-NOESY pulse sequence is based on the HSQC experiment in which a NOESY block consisting of the element 90° pulse, a constant mixing time and a 90° pulse is incorporated prior to acquisition. The first step is an HSQC preparation time in which proton magnetisation is transferred to the directly-bonded X nucleus (\(^{13}\)C or \(^{15}\)N). After the evolution of X nucleus chemical shifts during \(t_1\), the magnetisation is transferred back to the directly-bonded proton nuclei via one-bond \(^1\)J\(_{HX}\) coupling. The following step is a NOESY mixing period, where the magnetisation transfer from the proton (attached to \(^{13}\)C) to other spatially-close protons is achieved \textit{via} dipolar coupling during nuclei relaxation.

The resulting spectrum is a 2D map that correlates frequencies of heteronucleus X (\(f_1\) dimension) with the proton frequencies (\(f_2\) dimension).

Two different cross peaks appear at the chemical shift of each heteronucleus which arise from two cumulative magnetisation transfers: at the chemical shift of directly-bonded
proton (s) through the strong one-bond heteronuclear $^1J_{HX}$ correlation (HSQC sequence) and at the proton chemical shift to which X-labelled proton has NOE(s) (NOESY block).

This is an experiment that can be used to resolve overlapped $^1$H-$^1$H NOE connectivities at the resonances of the directly-bonded heteronuclei ($^{13}$C and/or $^{15}$N) in crowded spectra, resulting in a significant increase in spectral resolution.

Figure 1.33: 2D HSQC-NOESY. The pulse sequence combines HSQC and NOESY experiments. The first two steps are an HSQC preparation followed by evolution periods, during which magnetisation is transferred to the directly-bonded nuclei via one-bond $^1J_{HX}$ coupling. The last step consists of a NOESY mixing period, where magnetisation transfer between closely spaced protons occurs via dipolar coupling.

1.6.15 2D $\omega_1$-X-filtered NOESY

The 2D $\omega_1$-X-filtered NOESY experiment is a variant of the 2D NOESY, which removes the signals of the protons directly-bonded to NMR active nuclei X (i.e. $^{13}$C and/or$^{15}$N) and retains the signals of the protons attached to $^{12}$C/$^{14}$N in the $f_1$ dimension.

This experiment is useful for identifying NOEs unambiguously in large molecules or complexes (i.e. nucleic acid-protein/peptide or molecular chimeras) where the two interacting components of a complex or macromolecule present opposite labelling (i.e. when either the nucleic acid at natural abundance or the protein/peptide is $^{13}$C/$^{15}$N labelled). In the 2D $\omega_1$-X-filtered NOESY an isotope filter (a ‘purge element’) is applied just prior the $t_1$ evolution period to reject the magnetisation caused from the labelled component$^{39}$. 
The pulse sequence of the 2D $\omega_1$-X-filtered NOESY is based on a 2D NOESY experiment in which the initial 90° pulse is replaced with an X half-filter (purge) element, a heteronuclear filter on the $t_1$ dimension of the 2D experiment (Figure 1.34).

The purge element consists of a spin-echo period during which two 180° pulses are applied on both the $^1$H and the X nuclei. At the end of the first isotope filter the unlabelled proton magnetisation will be on the y axis, whereas the heteronucleus-coupled magnetisation will have evolved to $-2$HxCz. The interpulse delay is optimised to $1/2^1J_{HX}$ resulting in antiphase magnetisation for a $^1$H-X spin system. After the second 180° pulse is applied on the X nuclei the sign of the antiphase magnetisation changes, leading to cancellation upon summation. The use of the half-filter block can be improved by replacing the 180° pulse on the X nuclei by two consecutive 90° pulses and applying a refocusing delay.

Since there is no X decoupling during the evolution period $t_1$, all peaks of ‘X spin-labelled’ protons appear as doublets in $f_1$ domain due to $^1J_{XH}$.

The resulting spectrum is a typical 2D NOESY map with a reduced number of cross peaks in the $f_1$ dimension, which connect unlabelled protons solely and, therefore, is amenable to further detailed spectra analysis.

**Figure 1.34:** 2D $\omega_1$-X-filtered NOESY. The pulse sequence consists of a NOESY sequence in which the first 90° pulse is replaced by an X half-filter. This experiment removes signals of the protons directly-bonded to NMR active heteronuclei X.
The 3D HNHA experiment is a method for the accurate measurement of homonuclear $^3J_{HNHa}$ coupling constants in $^{15}$N enriched protein/peptide using the quantitative $J$ correlation measurement\textsuperscript{140}. This technique relies on a quantitative analysis of the diagonal-cross-peak intensity ratio which provides a direct measure for the magnitude of scalar coupling $^3J_{HNHa}$.

This method is referred to as the HNHA experiment as it correlates the intraresidue $^1$HN, $^1$Ha and $^{15}$N resonances.

In the initial part of the pulse sequence a 90° $^1$H pulse generates $^1$HN magnetisation which dephases due to $^3J_{HNHa}$ (Figure 1.35). A 90° pulse applied on $^{15}$N generates multiple quantum coherences in a HMQC-type fashion. Chemical shift labelling of the $^{15}$N spins occurs in a constant-time period during the subsequent delays. The multiple quantum coherence evolution is not affected by $^1J_{NH}$ during the constant-time period and it is inserted into the periods during which the antiphase amide magnetisation is transferred to the $^1$Ha coherence by the subsequent 90° pulse on the proton. At the end of a short evolution time, the antiphase $^1$Hn magnetisation is converted back to antiphase amide magnetisation.

The subsequent 90° pulse on the $^{15}$N converts the $^{15}$N-$^1$HN multiple quantum coherences back to observable $^1$HN magnetisation. Following the 90° purge pulse, only the in-phase amide proton magnetisation is observed during acquisition.

![Figure 1.35: 3D HNHA](image)

In the initial part of the pulse sequence the flow of magnetisation starts at $^1$HN after the first 90° pulse on the proton and is transferred to $^{15}$N. The second 90° $^1$H pulse transfers the $^1$HN antiphase magnetisation to $^1$Ha coherence, during which $^{15}$N evolution $t_1$ takes place. The magnetisation transfer from the $^1$HN to the $^1$Ha coherences is proportional to the length of the delay period $\delta_2$. 

91
The 3D HNHA affords two types of signals: autocorrelation or diagonal signals at $f_1^{(15\text{N})}$, $f_2^{(1\text{H})}$, $f_3^{(1\text{H})}$ and cross-peak signals at $f_1^{(15\text{N})}$, $f_2^{(1\text{H}^\alpha)}$, $f_3^{(1\text{H})}$ with opposite phase. The intensity ratio of the diagonal and cross peaks in $f_2$ dimension provides a direct measure for the magnitude of $^3J_{\text{HNHa}}$ (Equation 1.6-10).

$$\frac{\text{(Intensity)cross}}{\text{(Intensity)diagonal}} = -\tan^2(2\pi J_{\text{HNHa}}\delta_2)$$

Equation 1.6-10

1.7 Combination of NMR and Molecular Modelling for structural analysis

The inter-nuclear distance constraints and torsion angle values obtained from NMR can potentially be used for the structure calculation using restrained molecular modelling RMM, which applies mathematical equations of classical physics and experimental parameters (e.g. from NMR). RMM calculates chemical structures via adjustment of bond length, angles and torsion angles to satisfy the equilibrium values\textsuperscript{141}.

1.7.1 Molecular Modelling

Molecular modelling\textsuperscript{142} is concerned with ways to mimic the behaviour of molecules and molecular systems. In theory there are different approaches to create molecular models. However, since computational techniques have revolutionised this area of knowledge in the last 20 years, molecular modelling today is invariably associated with computational models. Molecular modelling uses different techniques, such as quantum mechanics, molecular mechanics, minimisation, molecular dynamics and simulations in general, conformational analysis and a variety of computer based methods for understanding and predicting the behaviour of molecular systems.

In the area of biomolecular simulations\textsuperscript{143}, especially protein and nucleic acid modelling\textsuperscript{144}, molecular mechanics methodologies have become standard tools thanks to their capability to effectively build large models with a considerable number of atoms providing predictions in good agreement with experimental results at a highly efficient computational cost.
1.7.1.1 Molecular Mechanics

In contrast with quantum chemistry\textsuperscript{145}, that uses quantum mechanics and Schrödinger’s equation to describe the nature of chemical bonds, molecular orbitals and spectroscopic information (IR, NMR), molecular mechanics uses classical mechanics (Newtonian mechanics) to describe the 3D structure and kinetic properties of molecules and chemical systems.

In the molecular mechanics paradigm, a molecule is just a group of point masses (atoms) connected with springs (bonds). In this model the interactions between atoms are expressed by simple equations for the bonded interactions (bonds, angles and torsions) and non-bonded interactions (Coulomb and van der Waals). The components of the potential equation together with the parameters required to model the individual interactions is known as a “Force Field”:

\[
V(r) = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2} (1 + \cos[n\phi - \delta]) + \sum_{\text{vdW},ij} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \sum_{\text{Coulomb},ij} \frac{q_i q_j}{r_{ij}}
\]

Developing a force field is a demanding and challenging process that requires, in the first place, collecting extended quantities of information from different sources: mainly spectroscopic, quantum mechanics and thermochemistry; and then, secondly, fitting all of the data to the model using advanced statistical techniques, minimising the global error of the fitting.

There are different force fields designed for different types of molecular and biomolecular systems, organic systems, and material systems. The selection of the most suitable one to run the simulation will strongly depend on the type of molecular system under study and the type of information needed.

1.7.1.2 Energy Minimisation and Geometry Optimization

Energy minimisation, sometimes called geometry optimization, or relaxation, is a type of calculation that provides the closest 3D geometry to the starting conformation that is at a minimum in the potential energy surface. The potential energy surface in molecular
mechanics is described by the force field; however, this calculation can be performed using a quantum chemistry approach.

There are different algorithms with different speed and convergence properties. Nevertheless, with the computational power available today, even a big macromolecular system can be fully minimised in a very reasonable amount of time with good convergence and robustness.

This type of calculation, as mentioned, will provide the closest minimum to the starting conformation only, meaning that the absolute minimum or any other minimum cannot be guaranteed. This situation prompted the study and development of other methodologies aiming to explore the conformational space of molecules such as Monte Carlo calculations, molecular dynamics and other types of conformational search techniques.

1.7.1.3 Molecular Dynamics

Molecular Dynamics (MD) is a type of calculation that describes the evolution of a molecular system with time. Even though the foundations of this technique remain still the same today, the methodology has evolved considerably in the last 30 years incorporating ideas from Lagrangian and Hamiltonian mechanics and statistical mechanics to effectively correlate the computational results with experimental data through thermodynamics.

Any MD calculation starts with a particular conformation of a system, i.e. input coordinates of the atoms, and propagates the Cartesian coordinates all over the phase space integrating Newton’s laws of motion.

\[
\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i}
\]

Newton’s equation, however, is not practical in that form and it requires a transformation into a finite difference form to be used in computer programs. In principle there are different algorithms available but, for stability and theoretical reasons, most of the software packages use variants of the Verlet form. In AMBER, for example, they use a variant of
the Verlet form that is close to the Leapfrog and velocity Verlet algorithms\textsuperscript{150} (Figure 1.36).

\[ r_i(t + \delta t) = r_i(t) + \delta t \cdot v_i \left( t + \frac{\delta t}{2} \right) \]

\[ v_i \left( t + \frac{\delta t}{2} \right) = v_i \left( t - \frac{\delta t}{2} \right) + \delta t \cdot \frac{f_i[r_i(t)]}{m_i} \]

Figure 1.36: Graphical illustration of the Leapfrog-like Verlet algorithm in AMBER and its characteristic scheme of integration. The coordinates and velocities are updated using the same time step, but the values are shifted by half a time step.

The kinetic energy of the system, required for the initial velocities of the atoms, is in agreement with the temperature of the simulation. The set of initial velocities is drawn from a Maxwell-Boltzmann distribution of velocities compatible with the macroscopic temperature of the system. Temperature and pressure are kept under strict control during the simulation using thermostats and barostats.

The output of a MD calculation is not a single structure, but a trajectory: a collection of conformations and their energies –sometimes thousands or hundreds of thousands of data-sampling a particular thermodynamic ensemble. In molecular simulations the most common ensembles are: the NVT or Canonical ensemble (related to Helmholtz free energy) and the NPT or Isothermal-Isobaric ensemble (related to the Gibbs free energy) because their results can be correlated with experimental results.

Statistical analysis of the energy and geometric output results can be correlated with experimental data and/or used as a predictive tool to improve understanding of the behaviour of macromolecular systems at atomistic level.
1.7.1.4 Restrained Molecular Dynamics

Restrained Molecular Dynamics (rMD) simulations are MD simulations where geometrical restraints, normally harmonic restraints, have been added to the total potential function.

The most common types of restraints are: i) positional (keeps atoms in a certain position in space), ii) distance (keeps the distance between two atoms), iii) angle (keeps the value of the angle between three atoms constant), and iv) torsional (keeps stable the torsion between four atoms). In some cases the restraints can be more sophisticated involving, for example, distance restraints between centres of several groups of atoms.

As mentioned before, the restraint is not completely rigid and usually follows a harmonic equation allowing for small fluctuations around the equilibrium value:

\[
V(r) = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2} (1 + \cos[n\phi - \delta]) + \sum_{vdW,ij} \frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^6} + \sum_{Coulomb,ij} \frac{q_i q_j}{r_{ij}} + \sum_{position} \text{al restraints} \ K_p \Delta_p^2 + \sum_{distance} \text{restraints} \ K_d \Delta_d^2 + \sum_{angle} \text{restraints} \ K_\alpha \Delta_\alpha^2 + \sum_{torsion} \text{restraints} \ K_\tau \Delta_\tau^2
\]

The implementation of restraints in MD calculations opened the door to advanced methodologies in biased ensembles, also called non-Boltzmann ensembles that require re-weighting techniques to post-process the results.

There are other applications where restraints can be very useful in biological simulations, for example, in protein folding.

1.7.1.5 Simulated Annealing and NMR assisted folding

Simulated Annealing (SA) was first introduced in computer simulations by Kirkpatrick, Gelatt and Vecchi using a Monte Carlo style implementation. The idea of simulated annealing was well known before in material science and used to obtain high quality and well-structured crystals.

The simulated annealing process in material sciences consists of, first, heating a material – for example a metal- at high temperature to allow the system to escape from a local
minimum with many crystal defects and then, secondly, leave it to cool very slowly to allow for the system to "crystallize" with a very low number of structural defects.

The same principle has been applied to macromolecular systems with a large conformational space like proteins. If the system is trapped in a local minimum (not the absolute minimum), it is first required to move the system out of that potential energy minimum by raising its temperature to allow exploration of other higher energy regions. Then the system is allowed to cool very slowly into a lower energy minimum condensed state. This process can be done repeatedly using iterations and collecting a full ensemble of “crystallized” conformations to analyse.

The practical implementation of the SA methodology is very straightforward. It is a molecular dynamics simulation where the system temperature is raised to a very high value –usually several hundreds of degrees or even over a thousand degrees- and kept at that value for a certain time to allow exploration of the conformational space and escape from local minima, and then the temperature is lowered very slowly until 0 Kelvin, which is the minimum (condensed state).

This calculation is usually done using implicit solvent models and is very efficient in perturbing full systems promoting unfolding-folding events. Unfortunately, in big proteins and macromolecular systems the number of steps required to reach a lower energy folding state is still unaffordable and there is no guarantee that the natural folded state will be obtained. In this sense, NMR-assisted folding has become a very useful tool to predict reliable 3D folding.

NMR-assisted folding uses the SA technique in combination with restrained molecular dynamics to fold proteins using experimental information. NMR information from NOE experiment can be used to determine distances from atoms close in space and other geometrical information as well, like angles or dihedrals. This information can be introduced in simulated annealing calculations by using different types of geometrical restraints to drive each iteration effectively towards a conformation in very good agreement with the NMR data. If the calculation is successful, several iterations of the SA algorithm can produce an ensemble of conformations of the protein, all of them in reasonably good agreement with the NMR data. Further geometrical analysis can be
applied to the ensemble to classify, group and filter the results and produce a 3D model of a folded protein.

1.7.2 Analysis

A common question when analysing results from molecular dynamics (MD) calculations, Monte Carlo (MC) calculations, or any other simulation that produces an ensemble of conformations and data as the output is what structure can be labelled as the solution of the MD – or MC, or SA- calculation. The answer is simple: all.

Molecular dynamics calculations produce a full trajectory containing several thousands of conformations and energy data in a particular thermodynamic ensemble, normally NVT or NPT. In order to obtain the results, the trajectory and energy files need to be post-processed. This step is as important as the actual MD simulation and sometimes is very demanding. The analysis may involve from very basic analysis of distributions of data to more advance statistical analysis including clustering, principal component analysis (PCA), Markov state models or normal modes analysis or ensemble re-weighting.

The types of statistical analyses employed in this research are briefly explained below and include geometrical analysis, RMSD and cluster analysis.

1.7.2.1 Geometrical Analysis

This analysis involves monitoring and studying the evolution with time of certain geometrical parameters like distances, angles, torsions, distances between groups of atoms, hydrogen bonding interactions, radial distribution functions and any other properties that can be evaluated from direct analysis of the Cartesian coordinates in the trajectory file.

This analysis is frequently done in two ways: on the one hand, the plot of the parameter vs. time. This approach, though very basic, can provide useful information about convergence of the calculations, NOE distances, conformational changes and lifetimes of loop motions or long range motions.
On the other hand time is removed and data are treated and analysed as distributions (once the system is fully equilibrated) where ensemble averages can be correlated with experimental data or can be used to predict molecular behaviour.

1.7.2.2 RMSD

The root-mean-square-distance (RMSD) of two structures is defined as the square root of the normalized sum of the distances of equivalent atoms in space:

\[ \text{RMSD} = \sqrt{\frac{\sum_{i=1}^{N_{\text{atoms}}} d_i^2}{N_{\text{atoms}}} } \]

If this parameter is going to be used as an index to measure how different two conformations of the same molecule are, then a fitting procedure is first required. Fitting is a mathematical operation where two –or more- conformations of the same molecule are translated and oriented in space in a way that particular atoms are optimally superimposed upon each other. If fitting is not applied, two identical conformations shifted and/or rotated in space will give a RMSD value different than 0.0, which is incorrect.

In MD calculations for example, the plot of RMSD vs. time is a very common descriptor used to measure how far the conformations in a trajectory fall from a certain conformation used as reference (sometimes the initial, the last one, or the crystal structure). Conformational changes, especially big ones, usually involve changes in the RMSD value.

This parameter, however, has to be used with care because it involves a great reduction of dimensionality, meaning that the entire position of 3N values in space will be reduced into a single scalar, involving therefore a great loss of information. There may be a situation where the movement of certain groups of atoms in space change while at the same time being compensated by other groups of atoms moving in other directions. That will lead to a net conformational change but the RMSD value might very well still remain almost the same. This means that when the RMSD value changes, then a conformational change occurs; however, the fact that the RMSD value remains stable does not guarantee that the molecular conformations are the same.
1.7.2.3 Cluster Analysis

Cluster analysis\textsuperscript{152,153}, refers to a very broad set of techniques for finding subgroups, or clusters, in a data set. It has also been defined as a classification tool aiming to group together similar objects. Molecular dynamics simulations generate trajectories containing thousands of conformations that is unpractical to analyse by inspection. In these cases it is highly desirable to extract a smaller data set that is representative of the entire trajectory: a set of structures that can cover all the statistical information of the original data set but with a minimal number of conformations.

Cluster analysis processes a complete MD trajectory and bins all the conformations in groups or “clusters” according to a certain similarity criterion. This criterion is a geometrical one, frequently the RMSD between pairs of conformations. This procedure can easily reduce a trajectory of tens of thousands of conformations to a group of ten or twenty, representing more than 80\% of the system. Sometimes, the first most populated clusters –the first 4 or 5- are usually enough to explain most of the behaviour of the system.

In AMBER, the cluster analysis implementation relies on RMSD of either distance between pairs of atoms or dihedrals. Also, the clustering algorithm used in this study was the DBSCAN\textsuperscript{154}.

Clustering is today an entire area of knowledge in statistics under expansion, with applications to data mining and Big Data.
1.8 References


32. M. S. Lalonde, Y. Zuo, J. Zhang, X. Gong, S. Wu, A. Malhotra, Z. Li, Exoribonuclease R in Mycoplasma genitalium can carry out both RNA processing
and degradative functions and is sensitive to RNA ribose methylation, *Rna*, 2007, 13, 1957-68.


61. Z. Li, S. Pandit, M. P. Deutscher, RNase G (CafA protein) and RNase E are both required for the 5’ maturation of 16S ribosomal RNA, *Embo J*, 1999, **18**, 2878-85.


70. A. T. Torelli, J. Krucinska, J. E. Wedekind, A Comparison of Vanadate to a 2'-5' Linkage at the Active Site of a Small Ribozyme Suggests a Role for Water in Transition-State Stabilization, *RNA*, 2009, **13**:1052.


83. A. Brack, B. Barbier, Chemical activity of simple basic peptides, *Orig Life Evol Biosph*, 1990, 20, 139-44.


98. N. L. Mironova, D. V. Pyshnyi, E. M. Ivanova, M. A. Zenkova, H. J. Gross, V. V. Vlassov, Covalently attached oligodeoxyribonucleotides induce RNase activity of a


108. R. Salvio, L. Mandolini, C. Savelli, Guanidine–Guanidinium Cooperation in Bifunctional Artificial Phosphodiesterases Based on Diphenylmethane Spacers;


Chapter 2: Materials and methods

2.1 Source of materials

Chemicals and solvents were used without further purification. N-α-Fmoc-N-ω-(2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulphonyl)-L-arginine, N-α-Fmoc-L-leucine, 2,2’-dipyridyl disulphide, hexadecyltrimethylammonium bromide (CTAB), triphenylphosphine, 4-(dimethylamino)pyridine (DMAP), N,N',N'-tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU) and the solvents – Acetonitrile (ACN) HPLC Grade 99.9% min, Acetone HPLC Grade 99.5% min, and Dichloromethane for peptide synthesis 99.8% min, dimethyl sulfoxide (DMSO), trifluoroacetic acid (TFA), piperidine (20% in DMF), N,N-diisopropylethylamine (DIPEA), toluene and N,N-dimethylformamide (DMF) were purchased from Sigma-Aldrich (Dorset, UK).

Fmoc-Gly-Rink-Amide MBHA resin, isotopically labelled amino acids Fmoc-(Pbf)-arginine (99% $^{13}$C$_6$ and 99% $^{15}$N$_4$) and Fmoc-leucine (99% $^{13}$C$_6$ and 99% $^{15}$N) were supplied by Anaspec (Cambridge, UK).

The 5’-phosphate modified oligodeoxyribonucleotide 5’-d(pTCAATC)-3’ was purchased from ATDBio Ltd. (Southampton, UK) and purified by HPLC.

Inorganic water was purified and de-ionized using a Milli-Q purification system Purelab Ultra (ELGA, 18.2 MΩ·cm resistivity).

2.2 Mass Spectrometry

Electrospray ionisation (ESI) mass spectra were collected on a Thermofisher LTQ Orbitrap XL mass spectrometer at the EPSRC National Mass Spectrometry Centre (NMSF, Swansea, UK). Samples were solvated in MeOH and analysed using THAP matrix with added diammonium citrate. Doubly charged ion clusters were observed by positive ion nanospray.
Matrix-Assisted Laser Desorption Ionization (MALDI) mass spectrometry analysis was used to characterize the peptidyl-oligonucleotide conjugates. The analysis was performed using a MALDI-TOF/TOF spectrometer (Ultraflex II) in the Mass Spectrometry Facility in the Manchester Institute of Biotechnology (MIB).

Analysis of the conjugates by MALDI at a concentration of 2 μg/μL was obtained with a gel matrix solution of α-cyano-4-hydroxycinnamic acid or sinapinic acid in aqueous acetonitrile (1:1, v/v) at 10 mg/mL concentration. A multi-layer sandwich method was used by depositing 1 μL of the matrix mixture on the MALDI plate, air-drying, then the sample layer added (1 μL, 2 μM) and finally coating with a second matrix layer. The ions were observed in positive-ion mode.

The molecular weight of the peptide-oligonucleotide conjugates were reconstructed from their mass-to-charge (m/z) ratio values of the multiply charged fragments.

### 2.3 Synthesis of peptidyl-oligonucleotide conjugates

Five different types of the conjugate were synthesised (Figure 2.1): one unlabelled conjugate (S1) incorporating L7R8 L9R10 L11R12L13R14G15 peptide and the other four being site-specifically 13C- and 15N-labelled samples at L and R residues in different positions of the same peptide, specifically at L7R8 (S2), L9R10 (S3), L11R12 (S4) and L13R14 (S5) sites (see Figure 2.1).

Sequences used for NMR and MD are as follows:

\[
\begin{array}{c}
1 & 3 & 6 \\
5'-pTCAATC-3' \\
\text{NH-} & \text{LeuArgLeuArgLeuArgLeuArgGly-CO-NH}2
\end{array}
\]

(S1): oligo-LeuArgLeuArgLeuArgLeuArgGly
(S2): oligo-LeuArgLeuArgLeuArgLeuArgGly
(S3): oligo-LeuArgLeuArgLeuArgLeuArgGly
(S4): oligo-LeuArgLeuArgLeuArgLeuArgGly
(S5): oligo-LeuArgLeuArgLeuArgLeuArgGly

**Figure 2.1:** Peptidyl-oligonucleotide conjugates sequence and nomenclature. Left: Sequence of the 6-mer oligonucleotide and the 9-mer peptide linked at the 5’ and at the N terminus respectively within the conjugate via a phosphoramidate bond. Right: Nomenclature of the conjugates used in the structure determination. The amino acids leucine (Leu) and arginine (Arg) labelled with 13C- and 15N in the four site-specifically labelled conjugates are underlined and shown in green.
2.3.1 DNA oligonucleotide

The lyophilised unlabelled 6-mer oligonucleotide 5’-d(pTCAATC)-3’ was purchased from ATDbio in the lyophilised form. Accurate quantifications of the oligonucleotide stock solutions prepared in H$_2$O were obtained by measuring the ultraviolet (UV) absorbance ($A_{260}$) of the test samples, measured as an optical density OD$_{260}$, at 260 nm, the wavelength at which the aromatic heterocyclic bases absorb the ultraviolet light.

The quantity of the oligonucleotide present in solution was calculated by applying the Beer-Lambert law,

$$A_{260} = \log_{10}\left(\frac{I_0}{I}\right) = \varepsilon_{260} \times l \times c$$

where $I_0$ is incident light intensity, $I$ intensity of the light passing through the sample, $A_{260}$ is UV absorbance measured at 260 nm; $\varepsilon_{260}$ is the molar extinction coefficient of the corresponding oligonucleotide at 260 nm ($M^{-1} \times cm^{-1}$), $l$ is the path length of the cuvette ($cm$), which was 1 cm in all experiments, $c$ is the concentration of the sample ($mM$).

The extinction coefficient for the oligonucleotide 5’-d(pTCAATC)-3’ was predicted by simply summing the extinction coefficients of the mononucleotides (A, C, T)$^\dagger$ and was found to be 57000 $M^{-1} \times cm^{-1}$. The purity of the oligonucleotide was assessed by $^1H$ and $^{31}P$ NMR spectroscopy prior to use: $^1H$ NMR (1 mM in D$_2$O, 400 MHz): 1.07-3.11 (m, 18H, 6 × H2’ and 6 × H2” ribose protons, 2 × CH$_3$ of 2 × dT), 4.45-5.20 (m, 24H, H3’/H4’/H5’/H5” ribose protons, 5.02-6.19 (m, 8H, 6 × H1’ ribose protons, 2 × H5 dC), 6.93-8.54 (s, d, s, 6H, Ar-H from dA (×2), dC (×2) and dT (×2). $^{31}P$ NMR (1 mM in D$_2$O, 160 MHz): δ -1.36 to -0.87 (6P, PO$_4^-$).

2.3.2 Solid phase peptide synthesis with Fmoc Chemistry

The unlabelled peptide NH$_2$-[Leu-Arg]$_4$-Gly-CNH$_2$ and the four (Leu/Arg) site-specifically $^{13}C$- and $^{15}N$-labelled peptides were prepared using the same procedure. Solid phase synthesis was used with 9-fluorenylmethyloxycarbonyl (Fmoc)-chemistry and
\(N,N,N',N'-\text{tetramethyl-O-(1H-benzotriazol-1-yl)}\) uranium hexafluorophosphate (HBTU) as the condensation agent.

The peptides were synthesized on a Rink Amide MBHA resin (430 mg) preloaded with Fmoc-L-Glycine (0.17 mmoles). Dichloromethane (DCM) was added (10 mL of DCM per 1g resin) and the resin was left to swell in DCM overnight at room temperature. The resin was then washed with DCM (15 mL \(\times\) 2) and DMF (15 mL \(\times\) 2) and the Fmoc group was removed by treatment with piperidine-DMF solution (2:8) (2 \(\times\) 15 minutes).

Scheme 2.1 represents a detailed procedure of the peptide solid phase syntheses.

Each amino acid coupling was achieved by adding to the resin 3 equiv. Fmoc (9-fluorenylmethoxy-carbonyl)-protected amino acids, pre-activated with 2.9 equiv. HBTU (O-benzotriazole-\(N,N,N',N'\)-tetramethyluronium hexafluorophosphate) in the presence of 5 equiv. DIPEA (\(N,N'\)-diisopropylethylamine) in DMF (\(N,N'\)-dimethylformamide). The use of the \textit{in situ} coupling agent HBTU, under basic conditions by DIPEA, forms an active ester during the coupling reaction. The coupling reactions were allowed to proceed for 30 minutes on a shaker. The resin was then drained and washed with 15 mL of DMF (\(\times\) 2) and DCM (\(\times\) 2) to remove excess reagents.

Every de-protection and subsequent amino acid coupling completion was verified by performing the Kaiser Test, a qualitative colorimetric test for the presence or absence of free terminal amino groups. A few beads of resin were removed using a Pasteur pipette and added to the test solution mixture (ninhydrin 2.5%-pyridine 1:1) heated at 70ºC for 2-3 minutes using a heat gun. Blue-purple resin beads indicated a complete Fmoc de-protection (positive test for free primary amine) and clear resin beads indicated complete Fmoc amino acid coupling (test negative for free primary amino groups). Protected amino acid derivatives were \(N\alpha\)-Fmoc-\(N\alpha\)-Pbf-Arginine and \(N\alpha\)-Fmoc-Leucine. The removal of the protecting group 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl group (Pbf) from the arginine side chain and cleavage from the resin was carried out by treatment of resin-bound peptide with 10 mL of the cleavage mixture TFA/TIS/H\(_2\)O (95:2.5:2.5; v/v/v) for 3 hours. The cleaved peptide in solution was then drained and the resin was washed with TFA 95% 5mL four times.
Scheme 2.1: Solid phase synthesis of the peptide \( \text{NH}_2-[^{\text{Leu-Arg}}]_4-\text{Gly}-\text{CONH}_2 \) with Fmoc chemistry.
After removal of excess TFA (trifluoroacetic acid) by rotary evaporation, the peptide was dropped in cold t-butyl methyl ether and kept at -20°C overnight to ensure complete precipitation. The precipitated peptide was then centrifuged (4000 rpm, 10 minutes at 4°C) and washed twice with t-butyl methyl ether to remove any residual scavengers. The peptide was air-dried, dissolved in H₂O-ACN-TFA (60:40:0.1) and purified by reverse phase on a Phenomenex Luna C18 (5 μm 10 × 250mm) column at room temperature using a flow rate of 1.5 mL min⁻¹. The mobile phase A was 0.1 % TFA in H₂O and the mobile phase B was 0.1% TFA in acetonitrile. The elution gradient was 10-100% B in 60 min, isocratic 100% B for 10 min and 100-10% B for 1 min.

Peak absorbance was monitored at absorbance wavelength 210 nm to detect the amide bond and at 280 nm to verify complete removal of arginine protecting group, Pbf. HPLC retention times for both unlabelled and labelled peptides were 36 min. Yield of TFA salts of the peptides varied in the range from 67 to 73%.

MS-analysis (ESI-TOF): mass calculated for unlabelled peptide 1 C₅₀H₉₈N₂₂O₉ [M-H]²⁺, 1150.81, found 1150.80; labelled peptide in L₇R₈ 2 C₃₈¹³C₁₂H₉₈N₁₅¹⁵N₅O₉ [M-H]⁴⁺, 1167.81, found 1167.85; labelled peptide in L₉R₁₀ 3 C₃₈¹³C₁₂H₉₈N₁₅¹⁵N₅O₉ [M-H]⁴⁺, 1167.81, found 1167.84; labelled peptide in L₁₁R₁₂ 4 C₃₈¹³C₁₂H₉₈N₁₅¹⁵N₅O₉ [M+H]⁺, 1167.81, found 1168.90; labelled peptide in L₁₃R₁₄ 5 C₃₈¹³C₁₂H₉₈N₁₅¹⁵N₅O₉ [M+H]⁺, 1167.81, found 1168.90.

2.3.3 Peptidyl-oligonucleotide conjugate synthesis

The 5’-phosphate modified oligonucleotide 5’-d(pTCAATC)-3’ was converted into the corresponding cetyltrimethylammonium salt, a complex soluble in the conjugation reaction medium DMSO. To achieve this, the oligonucleotide dissolved in water (0.35 μmol in 100 μL H₂O) was precipitated by stepwise addition of cetyltrimethylammonium bromide (CTAB) stock solution (4% CTAB in H₂O w/v). The insoluble oligonucleotide-CTAB salts were isolated from solution at each step by centrifugation (13400 rpm, 4 min). This procedure was repeated until no precipitate was observed from CTAB solution addition. The clear supernatant was then removed and the pellet was washed with water (500 μL × 2) and lyophilised. Prior to conjugation, the 5’-phosphate group of the oligonucleotide was converted to a more reactive intermediate capable of chemically attaching to the peptide.
terminal α-amino group. To the lyophilised oligonucleotide (0.35 μmol in 50 μL DMSO) were added aldrithiol-2 (10 mg, 45.4 μmol) and triphenylphosphine (10 mg, 38 μmol). The reaction mixture was vortexed and left to stand for 5 minutes, whereupon 4-dimethylaminopyridine (DMAP, 5 mg, 41 μmol) was added and the phosphate activation reaction was complete 10 minutes later.

The conjugation reaction, which relies on the formation of a phosphoramidate bond between the oligonucleotide with the activated 5′-phosphate and the amino-terminus of the 9-mer peptide is depicted in Scheme 2.2.

Scheme 2.2: Synthesis of the peptidyl-oligonucleotide conjugates. A solution of aldrithiol-2 (Py₂S₂) 1 and triphenylphosphine (Ph₃P) 2 in anhydrous DMSO was stirred with the oligonucleotide 3 for 5 minutes. DMAP 4 was added to give the oligonucleotide with activated 5′-phosphate 5. The peptide 6 dissolved in anhydrous DMSO was added to the solution and the reaction was allowed to proceed for two hours at 40 °C. The two fragments in the resulting conjugates 7 are linked via a phosphoramidate bond.

1.8 mg (1.04 μmol, 3 equiv.) of peptide dissolved in DMSO (20 μL) and pre-treated with DMAP (3 mg, 25 μmol) were added to a solution of the activated oligonucleotide (0.35 μmol) in DMSO (50 μL) in an Eppendorf tube, followed by vortexing. The conjugation
reaction was allowed to proceed for 2 hours at 40 °C. At reaction completion, the excess peptide was removed by adding 4% LiClO$_4$ in acetone (1.8 mL) and the peptidyl-oligonucleotide conjugates were isolated via precipitation in acetone overnight at -80 °C in the form of lithium salts. Following centrifugation for 4 minutes at 13400 rpm the supernatant was discarded and the pellet containing the conjugates and unreacted starting reagents was washed with acetone (1 mL) and dried in air before re-suspension in 120 μL 3M LiClO$_4$ in H$_2$O. The suspension was centrifuged (4 min, 13400 rpm) to separate the conjugate and unreacted oligonucleotides in solution from the undissolved reagents. The supernatant was then collected and transferred into a solution of 4% LiClO$_4$ in acetone to re-precipitate the conjugates. To ensure complete precipitation the solution was kept at -80°C overnight, whereupon it was centrifuged and the isolated pellet was left to air dry. 0.35 μmol (× 8) of the crude conjugates dissolved in aqueous solution of 1M LiClO$_4$ was purified by reverse phase HPLC using a Phenomenex Luna 5μm C18 (2) (10 × 250 mm) column and a 0-100% gradient of Buffer B (0.05M LiClO$_4$ in HPLC grade acetonitrile) in Buffer A (0.05M LiClO$_4$ in water) run over 42 minutes at a flow rate of 2 mL/min. All runs were performed at the temperature of 40°C. The conjugate runs were monitored at 260 nm and 210 nm. The final product of the conjugation reaction eluted fairly later than the unreacted reagents and side-products; the conjugates could thus be obtained at approximatively 95% purity. The final yield of the isolated conjugate after removal of buffer salts by size exclusion on NAP-10 Sephadex columns was calculated to be 55% by UV absorbance at 260 nm. MALDI-TOF: mass calculated for unlabelled conjugate S1 C$_{108}$H$_{170}$N$_{42}$O$_{45}$P$_6$²⁻ [M +H]$^+$, 2962.10, found 2962.10.

2.3.4 Synthesis of the selectively ($^{13}$C, $^{15}$N) labelled conjugates

Conjugates isotopically ($^{13}$C, $^{15}$N) labelled at selective amino acids residues, were generated and isolated using the same synthesis and purification procedure as for the unlabelled conjugate by incorporating isotopically labelled amino acids Fmoc-(Pbf)-arginine ($^{13}$C$_6$ and $^{15}$N$_4$) and Fmoc-leucine ($^{13}$C$_6$ and $^{15}$N) during the solid phase peptide synthesis. Each sample contained (L+R) labelled blocks at the four different positions, i.e. (L$_7$R$_8$), (L$_9$R$_{10}$), (L$_{11}$R$_{12}$) and (L$_{13}$R$_{14}$).
2.4 Conjugate quantification

The concentrations of the peptidyl-oligonucleotide conjugates in the stock samples were determined using ultraviolet (UV) absorption measurements at room temperature on a Varian Cary 4000 UV-visible spectrometer in 1 cm quartz cuvettes. Precise concentrations were determined by measuring ultraviolet absorbance at 260 nm of the test sample containing a small aliquot (5 µL) dissolved in 1 mL of water. The concentration of the stock corresponding solution was then calculated using the dilution factor (i.e. 200-fold) and the average extinction coefficient $\varepsilon_{260}$ of the oligonucleotide 5'-d(pTCAATC)-3' (57.3 mM$^{-1}$ cm$^{-1}$) calculated according to Gray et al$^1$.

2.5 Ribonuclease activity assay

Ribonuclease activity assay has been measured in the Research Institute of Chemical Biology and Fundamental Medicine (Novosibirsk, Russia) under the supervision of Dr. Olga Patutina.

The reaction mixture (18 µL) contained 600 counts per minute (cpm, Cherenkov’s counting) of 5'-[32P] labelled miR-21 5'-UCCACAGAAUCAGAUCUGAUGUUGA-3' and corresponding unlabeled RNA at a concentration of 1 µM to facilitate precipitation steps, buffer (50mM Tris-HCl pH 7, 200 mM KCl, 1 mM EDTA) and conjugate at concentrations of 10 and 50 µM. The RNA cleavage reactions were carried out at 37˚C for up to 48 hours and quenched by precipitation of RNA and RNA fragments with 30 µL of 2% lithium perchlorate in acetone. After collection by centrifugation the generated fragments were mixed with loading solution (8M urea, 0.025% bromophenol blue, 0.025% xylene cyanol) and separated according to their size and charge on denaturing 18% polyacrylamide-8M urea sequencing gel in TBE buffer (100 mM Tris-borate, pH 8.3, 2 mM EDTA). A ladder for the assignment of the site cleavage in each fragment, here an imidazole ladder and G-ladder produced by partial tRNA$^{\text{Phe}}$ digestion, was run in parallel. Gels were dried at 80 °C in vacuo.
Intensity of the radioactive 5’-[\(^{32}\)P] RNA bands was quantified using phosphoimaging (Molecular Imager FX, BioRad) and the extent of cleavage at specific sites was determined using Quantity One software.

### 2.6 Nuclear Magnetic Resonance Spectroscopy

#### 2.6.1 Spectrometers

The majority of NMR experiments were acquired on high field NMR spectrometers at Manchester Institute of Biotechnology (MIB) and included a Bruker Avance III 800 MHz spectrometer equipped with a TCI probe, \(^1\)H/\(^{13}\)C/\(^{15}\)N with cold preamplifiers for \(^1\)H and \(^{13}\)C and a 600 MHz Bruker four-channel spectrometer equipped with Triple Resonance Inverse (TXI) cryoprobe. The array of \(^{31}\)P experiments were conducted on a 500 MHz Avance spectrometer equipped with a 5 mm Broadband Observe (BBO) cryoprobe at Manchester School of Chemistry and a 400 MHz Bruker Avance-II\(^+\) spectrometer equipped with a BBI \(^1\)H/D-BB Z-GRD Z8202/0347 probe.

#### 2.6.2 NMR Sample preparation

NMR samples of the conjugates were produced by dissolving lyophilised conjugates in a final volume of 600 \(\mu\)L of either 100% D\(_2\)O or mixture of 90% H\(_2\)O and 10% D\(_2\)O for lock purposes. The pH was adjusted to 6.5 with 0.1M HCl and 0.1M NaOH (giving approx. 0.5 mM Na\(^+\)).

NMR samples for observation of proton-1 and phosphorus-31 spectra in 1D and in 2D NMR experiments were prepared by dissolving 1.4 \(\mu\)mol of conjugate in 0.6 mL of D\(_2\)O to give 2.3 mM final concentration, 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) was added as a chemical shift reference and the spectra were recorded in 5mm NMR tubes. The pH for D\(_2\)O samples was 6.17. In order to detect non-exchangeable protons, the samples were repeatedly freeze-dried in Eppendorf tubes and re-dissolved in D\(_2\)O to remove any traces of water (\(^1\)H\(_2\)O).
The signals from –(CH₃)₃ group of DSS (9H; 0.000 ppm) was used as an internal reference to calibrate ¹H NMR spectra. The concentration of DSS in the NMR tube was 0.1 mM to generate a reference signal of integral intensity comparable to those from the protons of the sample.

Samples were reconstituted in 90% H₂O and 10% D₂O for detection of exchangeable protons.

To observe influence of pH on NMR signals, the sample was recorded at different pH. To achieve this, the sample was removed from the NMR tube and pH was adjusted by addition of small amounts of concentrated acid and/or base (DCl or NaOD), followed by replacing the sample into the tube.

2.6.3 Data acquisition

NMR data were acquired at temperature of 25°C and/or 5°C, unless stated otherwise. Multidimensional experiments were performed in a phase sensitive mode by applying the States-TPPI method in all indirect frequency (f₁) dimensions to achieve quadrature detection. In order to maximise the water signal suppression the ¹H carrier was set at the frequency of water resonance (around 4.7- 4.8 ppm). Experiments were performed on the unlabelled and labelled conjugates at the concentration of either 1 mM or 2 mM.

2.6.3.1. NMR spectra acquired for assignment of the unlabelled conjugate

Conventional strategies were employed for the collection of 1D ¹H and ³¹P NMR spectra and the array of 2D experiments COSY, DQF-COSY, TOCSY, HSQC, HMBC and NOESY.

One-dimensional ¹H NMR spectra of the conjugates were performed using the following acquisition parameters. Data collection was carried out with an acquisition time of 1.4 s, for a relaxation delay between scans of 2.0 s over a spectral width of 12 kHz (15 ppm). The number of scans in 1D NMR spectra was 32 collecting 32 K points of data.
The homonuclear 2D $^1$H-$^1$H nuclear Overhauser enhancement spectroscopy (NOESY) spectra of samples in 90% H$_2$O and 10% D$_2$O mixture for the assignments of exchangeable resonances were acquired at 5°C in phase-sensitive mode with WATERGATE pulse sequence using States-TPPI method for quadrature detection. The spectral width was set to 12 kHz (15 ppm) for both proton dimensions. A mixing time $\tau_m$ of 300 ms and a relaxation delay of 2.0 s were applied. A total of 384 FIDs of 4096 complex data points were collected. Homonuclear 2D, correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY) and NOESY spectra of samples in D$_2$O with natural isotope abundance were acquired at 25°C, using gradient pulses for selection with presaturation in the COSY experiment and excitation sculpting in the NOESY and TOCSY experiments for the removal of the residual HDO signal. The spectra were run with the carrier frequency set at the residual HDO peak.

The spectral width was set to 9.6 kHz (12 ppm) for both dimensions. The spectra were acquired with 512 FIDs of 2048 complex data and a relaxation delay of 2 s between the scans. For processing, the acquired data were zero filled in $\omega_1$ prior to Fourier transformation giving a processed data matrix 2K $\times$ 2K; a sine-bell function was applied to enhance spectral resolution. The NOESY spectra in D$_2$O were also collected at 25°C with mixing times $\tau_m$ = 80, 100, 200, 300, 500 ms and 2.0 s interscan delay to monitor the linearity of the cross-relaxation build-up. These spectra were acquired with 16 transients for each $t_1$ time domain (FID) with a delay of 2.0 s between transients. The 2D $^1$H-$^1$H TOCSY spectra were collected using a mixing time of 60 ms, acquisition times of 170 ms ($t_1$, $^1$H), 22 ms ($t_2$, $^1$H) and eight transients for FID were accumulated.

2D double-quantum filtered COSY (DQF-COSY) spectra were collected in D$_2$O at 25°C using a spectral width of 8 kHz (10 ppm) in dimensions, 16 transients for each $t_1$ and 4096 and 256 data points along $\omega_2$ and $\omega_1$ dimensions, respectively.

2D $^1$H-$^{13}$C heteronuclear single quantum coherence (HSQC) ($^{13}$C at natural abundance, data matrix of 4096 $\times$ 640, $^{13}$C sweep width of 170 ppm) was performed in D$_2$O with the carbon carrier frequency set to 85 ppm and acquisition times of 170 ms ($t_1$, $^1$H), 10 ms ($t_2$, $^1$H). 2D $^1$H-$^{15}$N HSQC ($^{15}$N at natural abundance, data matrix of 4096 $\times$ 640, $^{15}$N sweep width of 80 ppm) was recorded at 5°C with nitrogen carrier frequency set at 105 ppm, and acquisition times of 85 ms ($t_1$, $^1$H), 32 ms ($t_2$, $^1$H).
For the assignments of the carbon atoms in the oligonucleotide ribose, the 2D constant time (HSQC) experiments for one-bond $^1\text{H}$$-^{13}\text{C}$ correlation were recorded using the unlabelled conjugate samples with $^{13}\text{C}$ decoupling during acquisition using GARP.

One-dimensional $^{31}\text{P}$$-[^1\text{H}]$ NMR spectra were recorded with composite pulse proton decoupling (CPD) at 25°C in D$_2$O (85% phosphoric acid as the external $^{31}\text{P}$ reference standard). The sweep width and frequency offset were set at 8 kHz (20 ppm) and 1.2 kHz (3 ppm) respectively. 3000 scans were accumulated for the conjugate at the concentration of 2mM and 4K complex data points were collected.

2D $^1\text{H}$$-^{31}\text{P}$ HSQC NMR data were acquired over spectral widths equivalent to 4 kHz (10 ppm) in $\omega_1$ and 6 kHz (15 ppm) in $\omega_2$ centred at $\delta^{1}\text{H} = 4.8$ ppm and $\delta^{31}\text{P} = 2$ ppm respectively with 152 transients and a relaxation delay of 2.0 s.

### 2.6.3.2 NMR spectra acquired for assignment of the labelled conjugates

For the assignments of the carbon atoms of the $^{13}\text{C}$, $^{15}\text{N}$ labelled amino acids residues within the peptide moiety of the conjugates, 2D $^1\text{H}$$-^{13}\text{C}$ HSQC experiments were recorded with constant-time period.

Distance constraints for the non-exchangeable resonances for subsequent structure calculation were derived from 2D $^{13}\text{C}$ HSQC-NOESY, recorded in D$_2$O with a mixing time of 200 and 300 ms and acquisition times of 230 ms ($t_1$, $^1\text{H}$) and 23 ms ($t_2$, $^{13}\text{C}$). For the assignments of the exchangeable resonances 2D $^1\text{H}$$-^{15}\text{N}$ HSQC-NOESY) were collected with acquisition time of 160 ms ($t_1$, $^1\text{H}$) and 23 ms ($t_2$, $^{15}\text{N}$). To extend these resonances 2D $\omega_1$,$^{13}\text{C}$, $^{15}\text{N}$ – filtered/edited NOESY$^{3,4}$ were recorded in H$_2$O with a mixing time of 200 and 300 ms with acquisition times of 80 ms ($t_1$, $^1\text{H}$) and 46 ms ($t_2$, $^1\text{H}$).

### 2.6.3.3 NMR spectra acquired at different temperature and pH

Seven additional 1D $^1\text{H}$ spectra were acquired at 5, 10, 15, 20, 25, 30 and 35 °C in order to determine the temperature dependence of the amide proton and aromatic proton resonances.
1D, HSQC and TOCSY were also acquired at temperatures ranging from 5 to 35 °C, in 5 °C increments to follow the movement of the peaks.

Spectra were acquired at different pH values ranging from pH 4.0 to 10.0 for measurements of pK_a of the titrating groups which give an estimate of the through-space intramolecular contacts. The chemical shift changes of peaks moving were monitored as a function of pH. The apparent pK_a (s) values were obtained from non-linear least-squares fitting to a one- or two-site model based on the Henderson-Hasselbach equation.

2.6.4 Data processing and analysis

One-dimensional (1D) spectra were processed and analysed with Topspin (version 3.2, Bruker); 2D spectra were processed using the program NMRPipe^5 and peaks were assigned and integrated with the analysis program Sparky^6 in preparation for molecular structure calculations. The 3-D NMR structures were displayed in Visual Molecular Dynamics^7 (VMD) and the distances between the atoms of interest were measured in Molecular Operating Environment^8 (MOE).

2.7 NMR assignment methods and constraint generation

Five samples, namely, unlabelled sample (S1) and four type-specifically ^13C, ^15N labelled samples at L/R residues of the peptide moiety (L_7R_8 (S2), L_9R_10 (S3), L_11R_12 (S4) and L_13R_14 (S5)) were used to obtain nearly complete resonance assignments of the conjugate.

Intra-residual ^1H signals were identified by applying a standard approach based on 2D TOCSY and COSY connectivities. The assignments of the peptide side chains and oligonucleotide ribose protons were corroborated by analysing ^1H-^13C HSQC and ^1H-^15N HSQC.

Non-exchangeable protons of the oligonucleotide residues were easily assigned from 2D NOESY spectra recorded in D_2O by identifying base-base, base (H_6/H_8) – sugar ring protons (H_1’, H_2’/H_2” and H_3’) via sequential walks^9,^10.
Assignments for the ribose H5’/H5” were completed with the help of constant time $^{13}$C HSQC spectrum. Assignments of the exchangeable adenine and cytosine amino protons were straightforward based on $^1$H-$^1$H NOESY spectra in water.

The assignment of the backbone $^{31}$P resonances was achieved by carrying out assignments of the correlations between $^{31}$P and H3’, H4’ and H5’/H5” resonances in the 2D $^1$H-$^{31}$P HSQC experiment.

The program TALOS+ was utilised to detect possible tendencies of the peptide moiety to adopt secondary structure (i.e. $\beta$-sheet or $\alpha$-helix) and the backbone order parameters based on the observed chemical shifts of H$^\alpha$, $^{15}$N, $^{13}$C$^\alpha$ and $^{13}$C$^\beta$. The experimental chemical shift values are compared to a database of resonance assignments of tripeptide segments from solved protein structure to find the 10 best matches which cluster in a ‘consistent’ region of the Ramachandran (phi/psi angle) map.

Experimental H-H coupling constants were extracted from resonance multiplet patterns in the 1D $^1$H NMR spectra at different pH values and from homonuclear 2D DQF-COSY spectra.

The NOESY spectra of the unlabelled conjugated, with mixing time of 200 and 300 ms, in either D$_2$O and mixture of 90% H$_2$O and 10% D$_2$O were examined to identify correlation peaks arising from short, medium and long range interactions to provide distance constraints for structure determination calculations.

The cross peaks, required for the backbone resonance assignment, were well resolved except for the overlap of the peptide resonances. The homogeneous nature of the peptide sequence and insufficient spectral resolution led to strategic isotope labelling of the peptide at alternate positions. 2D HSQC-NOESY and 2D $\omega_1$ filtered-NOESY spectra of the type-specifically labelled samples were recorded at mixing times of 200 and 300 ms in either D$_2$O and mixture of 90% H$_2$O and 10% D$_2$O to resolve this issue.
2.7.1 Distance restraints

Two-dimensional nuclear Overhauser effect experiments serve as the primary method of deriving inter-proton distances which are used as structural constraints in determining biomolecular structure\textsuperscript{12}. NOEs result from through-space dipolar coupling interactions rather than through-bond interactions. Thus, NOEs contain information on the separation of pairs of protons that are closer than 6 Å even though the protons are distant in the primary sequence.

The pick lists of assigned NOE distance constraints from two-dimensional $^1$H-$^1$H NOESY, 2D HSQC-NOESY and $\omega_{1-13}$C, $^{15}$N – filtered/edited NOESY were used to generate the peptidyl-oligonucleotide conjugate 3-D structure in the simulated annealing calculation.

2.7.1.1 Generation of restraint file

In 2D NOESY spectra, only fairly well resolved cross-peaks were used to calculate distance restraints. Accurate intensities of the D$_2$O homonuclear NOE cross-peaks in both $f_1$ and $f_2$ were integrated with Sparky by line fitting of the peaks to bell-shaped Gaussian function.

The measured NOE cross-peaks volumes were employed in the subsequent calculation of distances between non-exchangeable protons by using a complete relaxation matrix approach, MARDIGRAS\textsuperscript{13} with 800 MHz $^1$H-$^1$H NOESY data sets recorded with mixing times of 200 and 300 ms.

The starting models for the MARDIGRAS calculations were three structures generated by simulated annealing and energy minimised using the program AMBER 14.

In order to determine quantitative upper and lower bounds for subsequent structure calculation and refinement the RANMARDI\textsuperscript{14} modification of MARDIGRAS was used, which accounts for experimental errors in the intensities due to indirect relaxation, so-called spin diffusion\textsuperscript{15}.

The procedure was repeated using a series of correlation times $\tau_c$; however, ultimately three different values (3, 4 and 5 ns) were used to represent the isotropic motion of the molecule. These correlation time values were selected since they best reproduced fixed inter-proton distances (such as pyrimidine H5-H6 distance, adenine H2-H8 distance) and...
certain intra-sugar distances whose changes are independent of the glycosidic conformation.

Overall, eighteen MARDIGRAS calculations were carried out on the different NOE intensity sets obtained for two mixing times, three correlation times and for the three starting models.

NOE intensities were corrected by adding experimental noise in the relaxation matrix calculations, whose level was set to the size of the weakest integrated peak intensity. The NOE peak integration errors of 20% were also incorporated to the input intensities for RANMARDI to obtain standard deviations (SD) of the distances, therefore defining distance bounds more accurately. Each RANMARDI was run using 30 randomly varied intensities. The resulting distances computed by MARDIGRAS were combined to give the distance values and standard deviations for each data set. The upper and lower bounds for each proton pair were determined from the average distances ± SD.

Upper bound distance constraint of 6.0 Å were given to peaks observed in the 2D NOESY spectra, in cases when the restraints could not be quantified from the NOEs intensity due to spectral overlap.

Precise distance restraints involving the exchangeable protons cannot be always determined from the NOE intensity due to broadening of resonances by exchange with solvent. However, restraints for structure calculation can be derived from visual inspection of the cross-peaks intensities in the water NOESY spectra. The NOE cross-peaks were thus, translated to upper bound separations of either 4.0 or 6.0 Å, for strong and weak NOEs observed respectively, rather than fixed distance restraints.

The intra and inter-proton distances to methyl protons were calculated using a three-jump model determined with an accounting for the rapid spin of methyl groups compared to overall correlation time.$^{16}$

Torsional angle restraints for the ribose sugar conformation were based on the analysis of the 2D DQF-COSY spectrum. Sugars showing a strong H1'-H2’ cross peak were restrained to C2’-endo. Backbone torsion angle restraints were consistent with a B conformation.
2.8 Restrained molecular dynamics (rMD)

Restrained molecular dynamics is a computational procedure which provides a means to explore the conformational space in order to find a set of molecular structures that best explain the experimentally determined restraints.

The simulated annealing (SA) method, in conjunction with a MD simulation algorithm, is a convenient way to find the global minimum in energy conformation. The annealing protocol entails a heating period which enables the simulated molecule to get over high energy barriers due to its elevated kinetic energy and therefore to sample a wide range of conformational space. High temperatures are used in the initial stages to permit large structural rearrangements. A gradual cooling enables structure changes to proceed at descending energy levels towards the lowest energy state in accordance to Boltzmann distribution. At absolute zero, the system should occupy the lowest-energy state. Several starting structures and different initial trajectories may enhance probability that a series of molecules achieve the same low-energy conformations.

All MD calculations were performed in implicit solvent. The effect of counter-ions was modelled by neutralising the effective charges on the phosphodiester groups.

Upper and lower bounds computed by MARDIGRAS were employed to determine the structure of the conjugate by restrained molecular dynamics simulations (rMD). The spatial structure of the peptidyl-oligonucleotide conjugate was calculated by molecular mechanics using AMBER 14\textsuperscript{17} suite of programs with the Cornell \textit{et al.}\textsuperscript{18} based FF14SB force field along with the generalised Born (GB) implicit solvation model\textsuperscript{19} to mimic solvent.

The phosphoramidate bond is not parameterised in the FF14SB force field and the missing parameter was retrieved from GAFF and parm10 force fields and dihedral angle terms for phosphoramidate linkage atoms developed by Cieplak \textit{et al.}\textsuperscript{20}. xLEaP module of Amber was used to build the initial structure of the peptidyl-oligonucleotide conjugate and to generate parameter/topology and coordinate files required for minimisation and molecular dynamics SA.
The initial models used at the start of AMBER calculations for the conjugate were random chain conformations generated by simulated annealing, energy-minimised and without distance restraints. Local geometry violations are unavoidable while accommodating experimental restraints in high temperature simulated annealing. The planarity of aromatic bases and sugar bond angles are especially vulnerable to distortion\(^2\). To prevent this, geometry constraints for each base and ribose bond lengths and bond angles, as well as the peptide omega bond in trans configuration constraints were generated based on geometric parameters derived from high-resolution crystal structures of nucleic acids and proteins\(^2\).

The simulated annealing was run for a period representing 80 ps and consisted of 80000 steps. The molecular dynamics started at a low temperature (0 K) and low force constants of 0 kcal/mol·Å\(^2\). The temperature was increased in small increments to 600 K to allow structural relaxation with a simultaneous increase in the force constant for the distance restraints (\(k_{\text{NOE}}\)) and dihedral angle restraints (\(k_{\text{DIH}}\)) typically to 35 kcal/mol·Å\(^2\) and 35 kcal/mol·rad\(^2\) respectively. The restraints were gradually built up over 80 ps of molecular dynamics. Square well potentials were used for all NMR restraint energy function and the SHAKE algorithm\(^2\) was employed to keep bond lengths involving hydrogens fixed. The runs with these extreme values of temperature and force constant were typically 20000 steps long over 20 ps. During the cooling period, the temperature was gradually decreased to 0 K in 60000 steps over 60 ps. The resulting structures were subjected to 2000 iterations of constrained energy minimization (an amalgam of steepest descent algorithm followed by a conjugate gradient algorithm).

A starting list of unambiguous NOEs assignments was used to calculate the first set of 3D structures of the peptidyl-oligonucleotide conjugate in solution. These initial conformations were analysed and used to tentatively assign further NOEs and to continue the next rounds of structure calculations. Rounds of NOEs assignment and simulated annealing calculations were performed until a set of conjugate conformations with inconsistent distance violations, low RMSD values and an energy gap of \(\approx 20\) kcal mol\(^{-1}\) were achieved.

The pairwise RMSD values were generated by superimposing the resulting structures either on all atoms and the heavy atoms. A set of twenty conjugate conformations were energy minimised with NMR restraints and explicit solvation using the Sander program of the Amber suite to optimise the hydrogen-bond network within the conjugate molecule. The quality of the conjugate structures was monitored using VMD and MOE\(^8\) and analysed by MolProbity\(^2\).
Additional analysis to detect hydrogen bond interactions and arginine-phosphate salt bridges were performed using MOE and the cpptraj module of AmberTools 15. The program Curves+ was used to analyse the oligonucleotide helical and backbone conformations of the structures obtained from the ensemble by applying the helical parameters used to describe a nucleic acid double helix.

2.9 MD simulations

The MD simulations were carried out by using the PMEMD (Particle Mesh Ewald Molecular Dynamics) module in AMBER 14. The initial structure used at the start of AMBER calculations for the conjugate was the conformation having the lowest overall energy generated by SA with NMR restraints. The structure was placed in a truncated octahedron periodic box containing 10 Å TIP3P water molecules. Subsequently, two atoms of sodium were added as counter-ions to neutralise the excess phosphate negative charge of the system. The solvated system was minimised with 1000 steps of steepest descent and conjugate gradient while the conjugate atoms were restrained with a force constant of 500 kcal/mol·Å². Continuing with the same positional restraints with a force constant of 10 kcal/mol·Å², the system was then gradually heated from 0 K to 298 K over 500 ps at constant volume before equilibration at a constant pressure for further 1 ns. Subsequently, the harmonic positional restrained were gradually released over the course of 500 ps. A final equilibration step consisted of relaxation period of 200 ps at constant volume without constraints on atom motions, followed by a production calculation, run for a period of 1 µs, at constant pressure (1 atm) and temperature (298 K). The temperature was controlled using a Langevin dynamics with a collision frequency of 2 ps⁻¹ while the pressure was kept constant by a Berendsen coupling algorithm with a pressure relaxation time of 2 ps. During the MD simulation, the hydrogen atoms motion was restrained by employing the SHAKE algorithm, which permitted a 2 fs integration step to be used.

Structures PDB were visualised and analysed in MOE and PyMOL. Hydrogen bond and salt bridge distance measurements were calculated and analysed using cpptraj module. Clustering of the resulting trajectory (previously freed of the water molecules and ions) was performed in cpptraj using the following cluster algorithm and parameter settings:
DBSCAN (density-based), RMSD on structures heavy atom coordinates as similarity metric and sieve set to 10.

Representative structures from the most populated clusters in the simulation were chosen by their proximity to the centroid of the dominant cluster.
2.10 References


29. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

Chapter 3: Solution structure and dynamics of the model conjugate by NMR spectroscopy

3.1 Introduction

The aim of this work was to determine the solution structure of a model peptidyl-oligonucleotide conjugate \( \text{NH}_2\text{CO-Gly-[Arg-Leu]}_4\text{-NH-5'}-\text{d(pTCAATC)-3'} \) (Figure 3.1) capable of cleaving RNA sequences. A detailed characterization of the conjugate conformation in solution can shed a light on the induced conformational alteration in the peptide upon conjugation and provide suggestions of how the network of interactions between the two distinct parts of this macromolecule contributes to the shaped structural features, which seems to induce cleavage of RNA targets. Understanding of the effects of nucleic acid conjugation on a peptide conformation and the resulting cleavage activity would serve as a basis for rational design of optimised conjugate structures in the future.

An accurate description of the conjugate structure and the interaction between peptide and oligonucleotide components required collection of NMR data of both peptide and oligonucleotide backbone and side-chains. These included NOEs between nucleic acid aromatic bases/sugar rings and peptide amide/aliphatic nuclei to evaluate distances between interacting protons. The 2D NOESY spectra recorded in deuterium oxide (\( \text{D}_2\text{O} \)) and water (\( \text{H}_2\text{O} \)) displayed all these NOEs and experimental data from natural abundance of \( ^{13}\text{C} \) and \( ^{15}\text{N} \) NMR spectra alone provided enough information for accurate and complete assignments of the individual oligonucleotide moieties and inter-residue interactions between them. Nevertheless, the highly repetitive nature of the peptide sequence due to the lack of amino acid variation led to severe signal overlaps. Consequently, the number of unambiguous assignments of inter-nuclear interactions involving the peptide component of the conjugate was relatively low which proved to be detrimental for precise spatial structure determination.

Therefore, it was necessary to prepare conjugates optimally labelled at the peptide part with NMR-active isotopes (\( i.e. ^{13}\text{C} \) and \( ^{15}\text{N} \)) in order to expand the repertoire of NMR experiments available with enhanced sensitivity and ability to resolve strongly overlapping signals. Higher resolution structural information has been obtained by the NMR analysis of four site-specifically labelled conjugates, each contained two \( ^{13}\text{C} \)- and \( ^{15}\text{N} \) labelled residues.
(leucine and arginine) at different positions of the peptide sequence (see Figure 2.1 in Section 2.3 for nomenclature). Subsequently, complete assignments of the $^1\text{H}$, $^{13}\text{C}$ and $^{15}\text{N}$ resonance frequencies were achieved which enabled NOE distance restraints to be derived from NMR experiments in order to calculate the structures of the conjugate conformations.

Figure 3.1: Chemical structure of the peptidyl-oligonucleotide conjugate $\text{NH}_2\text{CO-Gly-[Arg-Leu-]}_4\text{-NH-}5'-\text{d(pTCAA TC)-3'}$.
3.2 Design of the Peptidyl-oligonucleotide conjugate

The design of the RNA-cleaving construct was based upon a previously studied peptidyl-oligonucleotide conjugate with proven ability to enhance the rate of spontaneous cleavage of phosphodiester backbones of RNA targets. A short amphipathic 9-mer peptide containing leucine-arginine (Leu-Arg) repeating units was equipped with a targeting fragment, a 6-mer oligonucleotide TCAATC, with the ability to specifically recognise RNA target sequences. The approach adopted for the synthetic route of the conjugate was to locate the catalytic peptide at one end of the oligonucleotide structure by activating the terminal phosphate and by further reaction of the latter with the N-terminal amino acid to form a phosphoramidate linkage.

3.3 Design of the Peptidyl-oligonucleotide labelled conjugate

Isotope labelling strategies increase sensitivity and resolution in the NMR experiments, which results in a marked simplification of otherwise crowded spectra. Uniform isotope labelling enables structure determination through resonance assignments with multidimensional heteronuclear NMR experiments but in drug design an alternative approach is residue-selective labelling or even segmental labelling. In fact, labelling of specific residues or small regions of the molecule of interest enables ligand-binding studies by chemical shift changes or intermolecular NOE effects by isotope-filtering and/or isotope-detecting heteronuclear NMR experiments. In this way, $^1$H resonance assignments are facilitated, and thus more information regarding the structure and dynamics of the molecule of interest is provided.

In this study discrete fragments of the peptide were labelled. Two adjacent amino acids, a leucine and arginine, at specific positions in the peptide sequence, a leucine and arginine, were synthetically replaced with uniformly $^{13}$C and $^{15}$N enriched correspondent amino acid analogues. Therefore, we produced four distinct conjugates with the peptide labelled at different positions (see figure 2.1 in Section 2.3 for nomenclature) in order to identify the interaction sites between the peptide and oligonucleotide parts of the conjugate.
3.4 Synthesis of the unlabelled and site-specifically $^{13}$C- and $^{15}$N-labelled peptidyl-oligonucleotide conjugates

Several protocols have been described for the conjugation reactions of these chimeric molecules$^{2,3}$; protection of the functional groups, which are not involved in the coupling reaction, and covalent attachment of the peptide and oligonucleotide components through additional activation of the selected functional groups to promote formation of the desired linkage between the two moieties.

Two principal strategies can be adopted for the covalent attachment of peptides to oligonucleotides: in-line solid phase synthesis and fragment conjugation. In the in-line solid phase synthesis the peptide-oligonucleotide conjugates are synthesized by stepwise coupling of amino acids and subsequent nucleotide assembly on the same solid support or vice versa$^4$. In the fragment conjugation the two moieties are synthesized, de-protected and purified independently, and reactive groups (i.e. amine, thiols or oxime) are introduced to favour the covalent linkage of the two fragments in the post-synthetic coupling$^5,6$.

For the synthesis of conjugates containing arginine residues, the stepwise solid phase method can be a difficult task. In fact, protective groups of arginine side chains require strong acidic conditions in order to be removed, which can lead to loss of the oligonucleotide purine of the oligonucleotide during the deprotection stage. Therefore, to circumvent this incompatibility of protecting groups during sequential assembly, in this work the peptidyl-oligonucleotide conjugates were prepared using the solution-phase fragment coupling strategy.

This post-assembly conjugation involved the formation of a phosphoramidate bond between the 5’-terminus of a phosphate modified oligonucleotide with the α-amino-terminus of the oligopeptide. The overall synthesis of the peptidyl-oligonucleotide conjugate NH$_2$-[Leu-Arg]$_4$-Gly-COH-5’-d(pTCAATC)-3’ entails multiple steps summarised below (see also ‘Materials and Methods’ chapter for specific details):

i. Removal of the N-α protecting group from the Fmoc-Gly-Rink Amide-MBHA resin compound using 20% piperidine in DMF.

ii. N-α-Fmoc and N-ω-Pbf protected Arginine residue activation in situ with HBTU and subsequent coupling to the terminal primary group of the Glycine residue attached to the Rink Amide- MBHA resin to form the first amide bond.
iii. Sequential addition of N-α-Fmoc protected Leucine residues and N-α-Fmoc and N-ω-Pbf protected Arginine residues in an iterative fashion by repeated steps involving N-α de-protection/coupling reactions to generate NH₂-[Leu-Arg(Pbf)]₄-Gly-CNH₂ peptide.

iv. The removal of the Pbf protecting group from the Arginine residue followed by cleavage of the produced peptide from the resin and subsequent purification.

v. Activation of the 5’-phosphate of the oligonucleotide - 5’-d(pTCAATC)-3’ \textit{in situ} with PPh₃, (Py₂S)₂ and DMAP prior to coupling to the peptide.

vi. Coupling of the activated phosphate in 5’- of the oligonucleotide with the N-terminus of the peptide to give the conjugate NH₂-[Leu-Arg]₄-Gly-CNH-5’-d(pTCAATC)-3’.

Synthesis of the four double labelled conjugates (see Figure 2.1) was achieved by incorporating the uniformly $^{13}$C and $^{15}$N labelled leucine and arginine residues into the elongating peptide at specific positions in the peptide chain, respectively at N-terminus (in L₇R₈), internally (L₉R₁₀, L₁₁R₁₂), and at the C-terminus (L₁₃R₁₄).

3.4.1 Solid phase synthesis of the peptide NH₂-[Leu-Arg]₄-Gly-CNH₂ and site-specifically $^{13}$C- and $^{15}$N-labelled peptides

The peptide was synthesized according to the procedure detailed in the ‘Material and Methods’ chapter (see Section 2.3.2). Briefly, the peptide was prepared as a C-terminal amide by sequential conjugation of amino acid residues (\textit{i.e.} Fmoc-Arg(Pbf)-OH or Fmoc-Leu-OH) to Rink Amide MBHA resin preloaded with Fmoc-Glycine and applying Fmoc chemistry (see Scheme 2.1). The use of \textit{in situ} coupling agent HBTU, under basic (DIPEA) conditions, generated an active ester during the coupling reaction.

Protection of the guanidinium group in the arginine side chain was required during peptide synthesis due to the propensity of this group in its non-protonated form to be involved in side reactions such as acylation as a result of its nucleophilicity.

For the Fmoc strategy the most recommended arginine protecting group is 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl group (Pbf). However, its removal from synthesised peptides with multiple arginine residues remains problematic as a result of its
high stability to TFA and requires high TFA concentrations (95%) and long de-protection times (at least three hours) in the peptide-resin cleavage step.

All the coupling and de-protection steps proceeded smoothly at each stage of the peptide synthesis and were completed in 30 minutes, as indicated by the Kaiser test used for the detection of free (i.e. uncoupled) primary amines.

After completion of the synthesis the peptide was released from the resin by treatment with the cleavage mixture and isolated by precipitation in ether, as described in ‘Material and methods’ section.

The peptide was purified by reverse phase HPLC using a stepwise gradient of acetonitrile in water (with 0.1% TFA in both phases) and detected at retention time of approximately 34-36 minutes with UV detection at 210 nm, which corresponded to the wavelength of absorbance for the amide bond. No absorbance was observed at 280 nm, which is typical of the Pbf-protected peptide, indicating complete removal of the Arginine protecting group. The peptide was isolated as trifluoroacetate salts with a 60% overall yield.

3.4.2 Peptide Characterization (Mass spectrometry)

Prior to coupling to the oligonucleotide the unlabelled and the four $^{13}$C- and $^{15}$N-labelled peptides were characterized by MALDI-TOF mass spectrometry and NMR spectroscopy to confirm the identity and purity of the compound. The MALDI-TOF spectrum recorded for the unlabelled peptide in negative mode shows a base peak at $m/z$ 576.40, consistent with the mass of the molecular species $[M-H]^{2+}$ (Figure 3.2). The MALDI-TOF spectra recorded in negative mode for the labelled peptides in L$_7$R$_8$ and in L$_9$R$_{10}$ show a base peak at $m/z$ 292.96 consistent with the mass of the molecular species $[M-H]^{4+}$, and MS spectra recorded in positive mode for the labelled peptides in L$_{11}$R$_{12}$ and in L$_{13}$R$_{14}$ show a base peak at $m/z$ 1168.90 consistent with the mass of the molecular species $[M+H]^{+}$. 
The measured mass of the $^{13}$C- and $^{15}$N-labelled peptides is consistent with the theoretical mass (1167.81) which corresponds to 100% incorporation.

The mass spectra of the four $^{13}$C- and $^{15}$N-labelled peptides can be found in Appendix B.

### 3.4.3 Coupling reaction between the oligonucleotide 5'-d(pTCAATC)-3' and peptide NH$_2$-[Leu-Arg]$_4$-Gly-COH$_2$

The synthesis of the conjugates was achieved following a fragment coupling strategy using a post-synthetic modification of oligonucleotide 5'-d(pTCAATC)-3’ and peptide NH$_2$-[Leu-Arg]$_4$-Gly-COH$_2$. The step-wise synthesis of the peptidyl-oligonucleotide conjugate using the in-line strategy was hindered by the fact that Pbf, the arginine protecting group, requires strong acidic conditions for its removal (95% TFA), which may induce purine base hydrolysis.

On the other hand, the protecting groups used in the oligonucleotide synthesis are usually removed in basic conditions which can hydrolyse amide bonds or cause side reactions such as racemization of the amino acids. The incompatibility of the protection strategies for the peptide and oligonucleotide components was circumvented by a separation of these two
synthetic routes followed by the post-synthetic conjugation of the two components as described in the ‘Materials and Methods’ chapter.

The 5’-phosphate of the oligonucleotide was activated in situ with the activating agents \( \text{i.e.} \ Ph_3, (\text{Py}_2\text{S})_2\) and DMAP) prior to the coupling with the peptide. After coupling with a three-fold excess of the peptide, the reaction product was purified by reverse phase chromatography using a C18 RP-HPLC column to separate peptidyl-oligonucleotide conjugates from excess of the activating agents and unreacted oligonucleotides using a stepwise gradient of acetonitrile in water (with 0.05M LiClO\(_4\) present in both phases). The starting reagents were for the most part present in the flow-through, while the conjugate eluted at a higher acetonitrile/water ratio after 26 minutes and was identified during chromatography by UV detector set at the wavelength of 210, 260 and 280 nm. Only one peak showed an absorbance at both wavelengths of 210 nm (maximum absorbance of the amide bonds in the peptide) and 260 nm (maximum absorbance of the aromatic rings of oligonucleotide) and was considered to contain the conjugate.

Estimation of the conjugation reaction efficiency was determined by comparing the integral area of the conjugate peak with that of the unreacted parent oligonucleotide in the HPLC chromatograph. The yield of the crude product was found to be approximatively 95%. The conjugate fractions were collected, freeze-dried and freed from salt excess. The yield of the isolated conjugate in form of lithium salt was calculated to 55% by UV absorbance at 260 nm.

### 3.4.4 Conjugate Characterization (Mass spectrometry)

Correct identity of the conjugates was confirmed by MALDI-TOF MS, revealing a single charged peak at the expected average \( m/z \) of 2962.10 (Figure 3.3) and a purity of > 95% for all conjugates was detected by RP-HPLC.

The mass spectra of the four \(^{13}\text{C}\)- and \(^{15}\text{N}\)-labelled conjugates can be found in Appendix B.
Figure 3.3: MS-analysis (MALDI-TOF) of the unlabelled conjugate S1. Formula C\(_{108}\)H\(_{170}\)N\(_{42}\)O\(_{45}\)P\(_6\)^{2-} [M +H]^+, 2962.10, found 2962.10.

3.5 Cleavage of 5’-\([^{32}\text{P}]\)-microRNA-24 by the peptidyl-oligonucleotide conjugate

In earlier publications it was shown that peptidyl-oligonucleotide conjugates containing alternating leucine and arginine residues possess per se a nuclease activity degrading the RNA\(^{8,9}\) without the aid of a catalyst such as metal ions.

In order to assess whether the peptidyl-oligonucleotide conjugate was able to hydrolyse phosphodiester bonds within an RNA molecule, ribonuclease activity was assayed with respect to oncogenic microRNA-24 (miR-24) under physiological conditions and at 10-50fold excess of conjugate over RNA substrate (as described in the ‘Materials and methods’ section).

The kinetics of RNA cleavage as a function of time is shown in Figure 3.4. In the above experiment the miR-24 cleavage by the conjugate over time was monitored through the change of radioactive emission intensity of the breakdown products.

The reaction mixture containing the 5’-\([^{32}\text{P}]\) labeled miR-24 was left to incubate with the conjugate over a 48 hour period and quenched at various times by precipitation of RNA
with 2% lithium perchlorate in acetone and subjected to separation by gel electrophoresis. The 5'-[^32P] labelled breakdown products were separated from each other as discrete bands according to their chain length under denaturing conditions. Identification of cleavage sites was allowed by simultaneous electrophoresis in adjacent lanes of partial digests carried out at the base phosphodiester bonds in the RNA by imidazole (Im) and RNase T1 (T1), which cleaves RNA exclusively at G-X bonds.

While miR-24 proved to be quite stable in the absence of the conjugate (control) (see lane C1 and C2, in Figure 3.4-a), a significant cleavage was observed within the first three hours. The material at the origin, which corresponds to the undegraded miR-24, progressively disappeared to give rise to parallel and independent cleavage of two main fragments predominantly (lanes 0-48, Figure 3.4-a). Almost complete degradation was achieved after 48 hours.

Although the complementary conjugate oligonucleotide does not form a stable complex with the miR-24 (data not shown) the peptidyl-oligonucleotide conjugate cleaves miR-24, mainly at a G-A sequence. The most effective cleavage occurred at the G_{7}A_{8} and G_{13}A_{14} phosphodiester bonds (Figure 3.4-b) with cleavage extent of 63% and 14% respectively in 48 hours (Figure 3.4-c).

The percentage of total radioactivity recovered in band G_{7} and G_{13} as a function of time is shown in Figure 3.4-c. The breakdown fragments accounted for 95% of cleavage extent of miR-24 by the conjugate (Figure 3.4-c).

One may presume that the conjugate binds to the partly complementary U_{16}G_{17}A_{18} and/or A_{21}G_{22}G_{23}U_{24}-RNA segment inducing the cleavage within the neighbouring hairpin loop at the 5’-side of G_{13} and/or at the 5’-side of G_{7} within a bulge spatially close to A_{21} as a result of the flexibility of the RNA-cleaving peptide of the conjugate (the secondary structure pattern (Figure 3.4-b) was predicted by using the Fold algorithm of the RNAstructure webserver^10 suggesting a preference for single stranded regions.

Additionally, the cleavage products yielded migrated similarly to the oligoribonucleotides by RNase T1 treatment^8 allowing the postulate that the conjugate acts as an artificial enzyme mimicking RNase T1.
Figure 3.4: Analysis of miR-24 cleavage by the conjugate. (a) Autoradiograph of 12% denaturing polyacrylamide/8M urea gel after catalytic hydrolysis. Lanes Im and T1: RNA cleavage by imidazole buffer and partial RNA digestion with RNase T1, respectively. Lanes C1 and C2 (incubation control): RNA incubated in the absence of conjugate for 0 and 48 hours. Lanes 0-48, incubation of miR-24 in the presence of the conjugate (40-fold excess) at 37°C for 0, 3, 5, 8, 24 and 48 hours, respectively. (b) Sites of cleavage of miR-24 induced by the conjugate (indicated by arrows). (c) Kinetics of miR-24 cleavage by the conjugate as a function of time.
Interestingly, the intensity of the G$_{13}$A$_{14}$ band increases up to 8 hour incubation and, after reaching a plateau, remains constant before slightly diminishing after 24 hour incubation, whereas the G$_7$A$_8$ band increases in intensity over 48 hours. This can be explained by assuming that the RNA degradation at the two predominant sites occur simultaneously with the generation of two different RNA fragments. After 8 hours the 13-mer fragment undergoes additional degradation within the complex resulting in the shorter secondary RNA fragments being formed (C$_5$A$_6$ and C$_3$A$_4$).

Considerably less efficient cleavage was observed at linkages at the 5'-side of A$_{17}$, A$_{20}$ and A$_{23}$ phosphodiester bonds within the substrate sequence and the neighbouring bulge. No miR-24 cleavage is detected when incubated with either the parent oligodeoxyribonucleotide or free peptide [Leu-Arg]$_4$-Gly-amide (Figure 3.4-a). These results are consistent with previous studies showing that peptide [Leu-Arg]$_4$-Gly exhibits ribonuclease activity only when conjugated with the oligonucleotide$^{11}$. The obtained data suggest existence of an ‘active’ conformation of the conjugate promoted by a network of intramolecular contacts between oligonucleotide and peptide residues. It is known that the arginine has a large hydrogen-bonding capability in the side chain and can make two H-bond contacts to the nucleic acid bases. Arginine almost always binds to guanosine, probably due to a partial negative charge on O6 and N7 of the guanosine base$^{12}$. In particular, the arginine forms a ‘fork’ motif in which the side chain forms a pseudo-Hoogsteen pairing with guanosine. The arginine guanidinium can also bind and orient phosphoester groups. It is postulated that a single guanidinium group can bridge two phosphates; this commonly occurs in single stranded regions or near RNA loops and bulges.

Thus, arginine residues of the peptide, other than providing a catalytic activity, also seem to play an important role in the G-A affinity of the conjugate, presumably, by stabilizing the RNA/conjugate complex at the specific cleavage site via a network of hydrogen bonds with the guanine residues. It is proposed that the multiple contacts with the guanine base acceptor groups may disrupt stacking with adjacent bases (in the single-stranded regions or at the junction of the double strand) facilitating conformational changes to the guanosine ribose phosphate followed by nucleophilic attack of the ribose 2’-OH on the 5’-phosphate in line with the 5’-C-O-P bond and subsequent cleavage of the phosphodiester bond$^{13}$. 

150
3.6 NMR spectroscopy and structure determination of the conjugate

3.6.1 Preliminary NMR studies on the conjugate

The sensitivity of NMR spectroscopy is rather low when compared to other methods such as mass spectrometry, and thus, requires high sample concentrations, ideally at the millimolar range, to record NMR spectra of sufficient quality. On the other hand, high concentration solutions can lead to increased solution viscosity or even compound aggregation, which should be avoided, as these factors represent a challenge for structural studies with this technique, i.e. can lead to significant line broadening. Therefore, it was important to achieve the highest possible concentration of a sample (in order to obtain NMR spectra of sufficient quality), but without any associated problems. Samples of unlabelled conjugate were, therefore, prepared at concentrations between 200 μM and 2 mM (as higher concentrations led to high viscosity) and 1D ¹H NMR spectra were recorded to test the spectra quality. Similar ¹H signals linewidths for resolved resonances were observed for all the spectra which indicated independence of spectra quality from concentration within the investigated range. Therefore, 2 mM concentration was selected for all subsequent experiments.

¹H-¹³C HSQC and ¹H-¹⁵N HSQC experiments were also recorded and inspected for the presence of thermodynamically stable multiple conformers by counting cross peaks and examining the backbone signals. Neither signal broadening nor duplication were observed which would otherwise be indicative of intermediate or slow conformational exchange. Instead, a single set of resonances were observed in these experiments which was consistent with a unimolecular structure. Nevertheless, these preliminary results do not totally exclude coexistence of multiple conformations in rapid exchange in solution.

The ¹H NMR spectrum in Figure 3.5-a shows the up-field shifted methyl groups of the leucine residues at 0.8-0.9 ppm displaying a narrow signal line-width indicative of rapid methyl rotation and the downfield-shifted amide protons of the eight amino acid residues except for the amide proton of the leucine directly involved in the formation of the phosphoramidate bond with the oligonucleotide, which resonates in an abnormal region (~3.6 ppm). The amide proton signals are characterised by broad line-widths and lower intensity in contrast to the non-labile proton signals (Figure 3.5-b), arising from the proton solvent exchange and presence of fraction of D₂O (10%) in solution. The region between
5.5 and 6.5 ppm displays signals originating from nucleic acid cytosine H5 and sugar H1’ only whereas the remaining sugar protons are grouped together in the region from 2 to 5 ppm.

Figure 3.5: (a) 1D $^1$H-NMR spectrum of the peptidyl-oligonucleotide conjugate (2 mM in 90% water and 10% D$_2$O recorded at 800 MHz and 5°C) showing the different regions of the oligonucleotide (in black) and peptide resonances (in blue). The region around the downfield-shifted resonances of the amino acid amide protons are enlarged in the inset (b).

At lower field, the aromatic protons of the nucleotide bases were found in partial overlap with the peptide amide protons. The peptide backbone H$\alpha$ signals resonate between 3.6 and 4.3 ppm and the aliphatic protons in side chains resonate at upper field between 1.4 and 1.9 ppm. These regions of the spectrum strongly overlap with the sugar proton resonances of the nucleic acid.

Since no imino proton signals were observed in the 1D $^1$H NMR spectrum in the downfield region (at 9-11 ppm for those imino protons that are not involved in hydrogen bonding of Watson-Crick base pairs) which could potentially display interactions between the nucleic acid and peptide within the conjugate, 2D NOESY experiments in D$_2$O were used to estimate the presence and extent of through space interactions.

The presence of intramolecular NOEs between nucleic acid components and the peptide was ascertained by the analysis of cross peaks between resonances in the aromatic region.
(Figure 3.6), where only base resonances are observed in D$_2$O and the up-field region between 0.8 and 2 ppm where mainly peptide aliphatic proton resonances are found.

Figure 3.6: 2D NOESY spectrum of the conjugate showing H1’ and aromatic regions of oligonucleotide component (5.9-8.3 ppm) against the aliphatic region of the peptide (0.8-1.8 ppm) in D$_2$O. The cross peaks in this region represent intramolecular NOEs exclusively between the nucleic acid and the peptide side chain non-exchangeable protons.

3.6.2 Optimisation of solution conditions for NMR experiments

In order to obtain the optimal signal line-shape and intensity, the influence of temperature and pH values on spectral quality were investigated to achieve high sensitivity in the NMR experiments without affecting other factors. The factors which can potentially be affected by pH media and/or temperature include for example, chemical shifts of sugar H3’, which are located in close proximity to water signals, the broadening of the labile protons due to chemical exchange (i.e. the amide protons), or compound solubility. For the spectra recorded for the unlabelled conjugate sample between 5 °C to 35 °C in water, the signal intensities of the base protons were seen to increase monotonically upon temperature increase, which can be explained by the effect of temperature on the solution viscosity. At higher temperatures the molecule tumbling rate increases with a resulting in a longer $T_1$ relaxation time.
The reduction in temperature decreased the exchange rate of \(^1\)H with water and therefore, the resonance intensity of some exchangeable groups increased. In particular, the arginine guanidinium protons and cytosine amine protons, which were not observable at room temperature, became visible at 10 °C, with chemical shifts at around 7.0 and 6.5 ppm, 7.3 and 6.7 ppm, respectively (Figure 3.7). These labile protons probably form salt bridges or hydrogen bonds.

![Figure 3.7: 600 MHz \(^1\)H variable temperature series of the unlabelled conjugate (1 mM), dissolved in 90% water and 10% D\(_2\)O (pH 6.5). The A-3 amine protons exchange slowly with water over the temperature range; the cytosine –NH\(_2\) groups and the 3’-OH peaks are very broad and become visible at lower temperatures. The peaks of slowly-exchanging protons H61 and H62 of adenine residues were always visible over this temperature range. The water resonance frequency moved by ~ 0.2 ppm to higher field over this temperature range. The resonances of sugar H3’ which appeared downfield and up-field of the water signal were equally affected at low and high temperatures. The amide proton intensities were greatest at 5 °C and appeared weak at room temperature as expected since \(^1\)H exchange rate with solvent is strongly retarded at low temperatures. Therefore, in order to]
increase resonance intensity of the exchangeable protons to allow the identification of the majority of the NOEs generated by their interaction with nearby groups, the experiments in water were performed at the lowest temperature of 5 °C.

Another important parameter to optimise was the pH value of the solution. Molecule solubility is lowest at the pH equivalent to its isoelectric point (pI), at which positive and negative charges are equally present and the molecule is at neutral charge. In order to avoid conditions that favour precipitation, the pH of the solution should be at least 1 pH unit away from the molecule pI.

The theoretical pI of a molecule can be calculated from its residue composition and by knowing corresponding pK\textsubscript{a} values. However, the presence of either hydrogen bonding or salt-bridges, with nearby polar groups, can potentially change the residue pK\textsubscript{a} dramatically. Since local ionisation events can produce variation in the NMR chemical shifts of an ionisation site itself as well as of the nuclei in close proximity, NMR pH titrations serve as an excellent reporter of these nuclei chemical shift changes undergoing ionisation and are used in this work to determine their pK\textsubscript{a} values.

The conjugate nucleobases have ionizable groups within a specific pH range. For example, adenine N1 on the aromatic ring is expected to exist predominantly in a protonated form at a pH below 3.7 and cytosine N3 will be mainly protonated at a pH lower than 4.5, according to their theoretical pK\textsubscript{a} values. Thymine base N3 is instead expected to be mainly deprotonated at a pH higher than 9.7. However, an aromatic base pK\textsubscript{a} is not only dependent on the effect of the type of neighbouring nucleotides but can also change substantially due to formation of conformations involving the protonated/deprotonated base.

The bases’ pK\textsubscript{a}’s were determined through an indirect-detection approach by monitoring the non-exchangeable protons resonances over a pH range of 4-10. The pH titration was not extended below 4.0 in order to prevent depurination. This approach was used neither to determine the pK\textsubscript{a} values of the side-chain ionizable groups of the arginine residues (expected to be ~12 in the absence of stable interactions with other groups) due to severe signal overlap, nor to measure pK\textsubscript{a} values of the peptide C- and N- terminus due to absence of charge.
The chemical shifts for the base titrations were fitted using non-linear regression of the Henderson-Hasselbalch equation, expressed in terms of NMR chemical shifts for two pK\(_a\) values:

\[
\delta_{\text{peak}} = \delta_{\text{HA}} + \Delta\delta_1 / \left[ 1 + 10^{(pK_{a1} - \text{pH})} \right] + \Delta\delta_2 / \left[ 1 + 10^{(pK_{a2} - \text{pH})} \right]
\]

where \(\delta_{\text{peak}}\) is the observed chemical shift of a particular proton, \(\delta_{\text{HA}}\) is the chemical shift of the fully protonated base and \(\Delta\delta\) is the difference between \(\delta_{\text{HA}}\) and the chemical shift of the fully deprotonated base governed by pK\(_{ai}\).

By measuring \(\delta_{\text{peak}}\) at the different pH in the range 4-10 it was not possible to extract accurate base proton pK\(_a\) values because both of the plateau chemical shifts of the protonated/de-protonated form on the titration curve are poorly defined. Nevertheless, by fitting the titration curve with a model involving two independently ionizable groups (pK\(_a\) of the given base and pK\(_a\) of a neighbouring base) the calculated pK\(_a\)s did not significantly deviate from those expected, which suggests an absence of strong interaction involving the bases. It was also found that the protonation state of the aromatic bases does not change in the pH range 6-9 (Figure 3.8).
Figure 3.8: NMR-based titration curves recorded for the oligonucleotide aromatic protons of the unlabelled conjugate. A model with two $pK_a$'s was selected to obtain $pK_a$ values of the bases. All the bases exist in their physiological charge state in the range of pH 6-9.

The amide protons did not experience sizeable pH titration shifts, estimated to be smaller than -0.030 ppm. Since at pH values above 7.5 the amide proton exchange with water is too fast for the resonance to be readily observed, titration shifts that might possibly arise from deprotonation of the side chain of the arginine residues could not be estimated (Figure 3.9).

Slightly acidic pH values gave better resolved amide signals due to slow solvent exchange. Nevertheless, a pH of 6.5 was chosen as this gives the lowest solvent exchange rates of the peptide amides while maintaining the molecule in its physiological charge state.
Figure 3.9: 1D $^1$H NMR spectra of the conjugate in the amide and aromatic region acquired at different pH values (4 to 10). The aromatic proton resonances upfield shifts are observable at acidic pH. The peptide amide protons chemical shifts are not affected at the titration pH range.

3.6.3 NMR strategies for nucleic acid resonance assignments

3.6.3.1 Identification of base protons and sugar proton spin systems

The NMR experiments for the identification of the oligonucleotide base and sugar protons were recorded in D$_2$O to remove the NOE peak ambiguities arising from overlapping signals from peptide amide protons and aromatic base protons.

The first step for complete assignments of the conjugate oligonucleotide was to identify the nucleobase spin systems ($H_8$, $H_6$, $H_5$ and $H_2$) and sugar moiety protons ($H_1'$, $H_2'$, $H_3'$, $H_4'$, $H_5'$, $H_5''$), which was achieved by using a combination of $^1$H-$^1$H TOCSY, DQF-COSY and $^1$H-$^1$H NOESY. These experiments link the nucleobase spin-systems to the sugar spin system via through-bond correlations and by observation of NOE contacts between protons that are within 5-6 Å of each other. $^1$H-$^{13}$C HSQC and HMQC spectra were used to verify the assignments of the corresponding $^{13}$C chemical shifts. The experiments were also performed at temperatures in the range 5 to 25 °C, which resulted in
slight shifts of sugar proton signals and consequently aided with the assignments of partially overlapping peaks.

From the TOCSY spectrum the homonuclear $J$-connectivities between the cytosines C$_2$ and C$_6$ aromatic protons H5 and H6, which have a fixed distance of 2.4 Å, were easily established by the presence of a very strong cross peak in correspondence with their resonances and by the 7 Hz signal split; the thymine residues T$_1$ and T$_5$ were identified by the two weak cross peaks connecting the two $J$-coupled (via their four bond spin-spin coupling) thymidine-methyls in the aliphatic region and H6 in the downfield aromatic region (Figure 3.10). Both cytosines H5 were further distinguished from the other bases by their characteristic $^{13}$C chemical shifts at ~ 99 ppm, determined in a natural abundance $^1$H-$^{13}$C HSQC spectrum. The remaining signals in the aromatic region are from the purines A$_3$ and A$_4$ nucleobase protons.

Adenine H2 resonances were distinguished by their characteristic sharp singlet lines (long T1 relaxation times) and by the chemical shift of their attached carbon atom in the $^1$H-$^{13}$C HSQC spectrum well separated from other C8-H8 and C6-H6 signals, further confirmed by a $^1$H-$^{13}$C HMQC experiment at natural abundance. Adenine H8 signals were subsequently assigned via a process of elimination.

Figure 3.10: Assignment of the base H5 (or -CH$_3$) and H6 resonances of cytidine or thymine residues, respectively, of the conjugate nucleic acid, in an 800 MHz $^1$H-$^1$H TOCSY spectrum (2 mM, D$_2$O at pH 6.5, 25 °C). Note that the H5 and H6 of the cytosine C$_2$ and C$_6$ are the only aromatic scalar-coupled protons, hence identified by the two strong aromatic cross peaks, split in the $^1$H $\omega_2$ dimension. The weak cross peaks connecting the aromatic region to the up-field aliphatic region are the result of the four-bond coupling between thymine residues methyl (H7*) and H6 protons. The right hand plots show the 800 MHz $^1$H-$^{13}$C HSQC H7*-C7 and H5-C5 regions for cytosine and thymine residues.
3.6.3.2 Assignment of the exchangeable proton resonances

The exchangeable $^1$H resonances were assigned from NMR spectra acquired in H$_2$O solution. The assignment of the conjugate nucleic acid was performed using the procedure outlined in Section 2.7. The absence of imino protons peaks of the oligonucleotide in H$_2$O solution in the spectral region ($\omega_1 = 9$-14 ppm, $\omega_2 = 9$-14 ppm) implies that there is no involvement of the bases in stable hydrogen bonding network (i.e. normally occurring in double-strand base pairing or with the peptide amide/guanidine group), which can be detected on the NMR time scale. Instead, these protons have complete access to solvent and are in fast exchange with H$_2$O.

Assignments of the adenines A$_3$ and A$_4$ N6 and cytosines C$_2$ and C$_6$ N4 amino groups were determined by analysis of $^1$H-$^15$N HSQC (80 ppm and 94 ppm respectively) (Figure 3.11) and TOCSY spectra in the amino proton spectral region (6.5-7.3 ppm), where the cytosine H41 and H42 correlations and adenosine correlations between H61/H62 and H2 were observable (Figure 3.12). The NOEs between the protons of the N6 amino group of the two adenosines were observed in the spectral region 6.7-6.9 ppm (Figure 3.11).

Figure 3.11: Assignment of the exchangeable amino protons of two adenine (A$_3$ and A$_4$H61/H62) and cytosine (C$_2$ and C$_6$ H41/H42) nucleobases in a 800 MHz $^1$H-$^1$H NOESY spectrum (2 mM, H$_2$O 90%, D$_2$O 10%, at pH 6.5, 5 °C). The top and right plots represent sections of $^1$H-$^15$N HSQC spectra showing the $^{15}$N and base amino proton correlations of adenine and cytosine respectively.
The H41 and H42 protons of both C2 and C6 amino group resonate around 6.7 ppm and 7.3 ppm respectively, and display NOESY cross-peaks with protons of the same group as well as with the cytosine aromatic protons H5 and H6. These cross peaks can be distinguished from each other on the basis of stronger NOE intensity to the H5 (Figure 3.11).

The C6 3'-OH 1H resonance was observed at 6.5 ppm. The assignment was established based upon TOCSY cross-peaks from the 3'-OH of C6 to H3' and to H4' (Figure 3.12).

![Figure 3.12: Sections of 2D TOCSY spectrum (2 mM, H2O 90%, D2O 10%, at pH 6.5, 5 °C) showing the H41 and H42 correlations of the cytosine bases C2 and C6 (6.5-7.3 ppm), H61/H62 and H2 correlations of adenine bases A3 and A4 (6.7-7.9 ppm), and 3'-terminal C6 hydroxyl proton correlations to its own sugar H3' and H4' (4.1-6.5 ppm).](image)

3.6.3.3 Sequential backbone assignment

By using the known first-stage identification of the base protons (see section 3.6.3.1), the assignments of the individual protons to the specific nucleotide residues were achieved through identification of sequential base to ribose nuclear Overhauser effect (NOE) patterns, similar to the procedure originally utilised for DNA studies15 (see the ‘Material and methods’ section).

All aromatic H8/H6 protons show NOE cross peaks to their own sugar H1’ and to the neighbour nucleotide in the 5’ direction. The sequential NOE connectivities were identified
in a 200 ms mixing-time NOESY experiment (to minimise possible spin-diffusion), confirmed by a $^1$H-$^{13}$C HSQC spectrum where the H1’ signals appear in a well resolved spectral region.

As outlined in a NOESY contour plot of the conjugate in D$_2$O in Figure 3.13, the NOE cross peaks of thymine at the 5’-terminal position served as a starting point. A thymine H6 at 7.32 ppm gives an NOE to its own deoxyribose H1’ at 6.01 ppm and to the H1’ of an adenine at 6.18 ppm, whereas another thymine H6 at 7.52 ppm displays a single cross peak with its own H1’ at 6.2 ppm. Therefore the 7.32 ppm H6 was identified as the aromatic proton of the thymine residue in position 5 (T$_5$H6) adjacent to an adenine residue and the 7.52 ppm H6 as the aromatic proton of the 5’-terminal thymine residue (T$_1$H6). H1’ of A$_4$ was thus assigned at 6.18 ppm on the basis of NOESY cross-peaks to the assigned T$_5$ H6. The resonances of the other oligonucleotide aromatic and sugar H1’ protons were then traced along to C$_2$ H1’ and continued all the way to the residue T$_6$ (Table 3.1).

Aromatic H2 NOEs to their own and the H1’ of the next residue (i.e. A$_3$H2 to T$_3$H1’), as well as sequential aromatic base H6 NOEs to base H5 of the next residue (i.e. C$_6$H5 to T$_5$ H6) assignments were also identified, which further confirmed the sequential connectivities.

**Figure 3.13:** Expanded NOESY spectrum (300 ms mixing time, D$_2$O at pH 6.5, 25 °C) of the conjugate bases to H1’ region. A representative sequential NOE walk, base protons H6/H8 and the sugar H1’ protons from the 5’end (T$_1$) to the 3’end (C$_6$) are shown as a green trace. The cytidine C$_2$ and C$_6$ base H5-H6 cross-peaks, which are the most intense ones, are connected to their corresponding H1’-H6 cross peaks by dashed lines for clarity.
Once these sequence specific assignments H6/H8-H1’ were completed, the relevant intra-nucleotide NOEs were used to further assign the other sugar proton resonances with respect to the H2’, H2” and consecutively to H3’ resonances by means of the scalar coupling and NOE between H3’ and H2”, H2”.

The sequential connectivities from the H2’ H2” to H8/H6 are shown in Figure 3.14-a. The H2” of T1 connect to the H6 of C2 at 7.55 ppm and the other couplings $^3J_{H1'-H2'}$ and $^3J_{H1'-H2''}$ were identified in a similar manner. The weak intensity of the NOE cross peaks in adenine compared to thymine and cytosine is attributed to the longer distance between the aromatic and H2’ protons in purine than in pyrimidine. The sugar H2’ and H2” protons were stereo-specifically assigned based on the larger intra-residue NOE between the H2’ and the aromatic protons, whereas the intranucleotide NOE between H1’ and H2” protons is stronger since the distance between these two protons is usually shorter in all sugar pucker conformations.
Figure 3.14: Regions (a) and (b) of the 2D NOESY spectrum. (a) drawn lines connect cross peaks involving intranucleotide contacts base \(i\)-H2’, H2” and internucleotide contacts H2” \(i\)-base \((i+1)\), included those of 3’-terminal nucleotide C6. The H2’ and H2” of the adenines resonate at lower field as a result of the deshielding effect of the stronger aromatic ring current of purines. Region (b) shows the base-H3’ cross peaks.
The correlations of the H1’ and H2’ and H2” were checked also by DQF-COSY (Figure 3.15).

**Figure 3.15:** Section of DQF-COSY spectrum showing the cross peaks linking H1’ with H2’ and H2” (2 mM, D2O at pH 6.5, 25 °C; 800 MHz). The differences between the coupling constants between the 2’CH2 protons and H1’ is apparent in the cross peak multiplet splitting pattern. For A4, the multiplet pattern of the single peak shows that H2’ and H2” resonate at very similar frequency. The large $^3J_{H1'-H2'}$ coupling and a small $^3J_{H1'-H2''}$ coupling is indicative of a C2’-endo conformation of the sugar pucker.

The H3’ protons were assigned by NOESY by using the cross-peaks of aromatic H8/H6 to H3’ and continuous sequential base proton connectivities to their 3’- neighbouring bases through the nucleic acid led to their full assignments (Figure 3.14-b).

A schematic representation of the intranucleotide and sequential NOE connectivities observed is given in Figure 3.16, which outlines three continuous networks of short $^1$H-$^1$H distances, specifically, H1’/H2”/H3’ (n) with the H6/H8 (n) and H6/H8 (n+1).
Figure 3.16: Pictorial representation of intranucleotide and sequential NOE connectivities of the oligonucleotide fragment. The intraresidue and sequential short $^1$H-$^1$H distances from the 5’ to 3’ direction (indicated by black arrows and red arrows, respectively), are used for sequential assignments of the non-exchangeable protons to confirm the oligonucleotide sequence.

The H3’ protons are scalar coupled to their own H4’, giving rise to fairly strong cross peaks, which were assigned in TOCSY experiments. The H4’ protons in turn are scalar coupled to H5’ and 5’”. The H5’ and H5’” strong coupling complicated the analysis of the NOESY spectra and their resonances are all crowded in a small region near 4 ppm, overlapped with H4’ cross peaks generated by connectivities to H4’ and with the diagonal. The H5’ and H5’” protons were, therefore, assigned by using the corresponding intra-residue NOEs to H3’ (Figure 3.17). Stereospecific assignment of H5’/H5’” protons from NOESY was carried out on the assumption that the H5’ protons normally resonate upfield from H5” because of the negative charge on the phosphate.
3.6.3.4 Resonance assignments from $^{31}$P correlated experiments

The phosphorus resonances of the conjugate nucleotides were assigned using a 1D $^{31}$P NMR spectrum and 2D $^{31}$P-$^1$H HSQC experiment (Figure 3.18). The phosphorus resonances were partially overlapped in the region -4.4 to -3.8 ppm with T$_1$ showing unique chemical shift at 2.6 ppm.

In this experiment, the phosphorus nuclei are coupled to the preceding H3' (i-1) and H4' (i-1) resonances on the 5' side and the following H5'i/H5''i and H4'i on the 3' side of the phosphodiester backbone.

Strong peaks corresponded to the sequential H3'(i-1)-P(i) connectivities, consistent with the $\epsilon$ (C4'-C3'-O3'-P) dihedral angle trans conformation (observed in DNA double helices$^{17}$); the H3'-$^{31}$P correlations were all visible and assigned, with an exception for the T$_1$ base which is not bound to a nucleotide at the 5' end.

Strong-medium $^1$H-$^{31}$P correlations were observed for the H5', H5'' resonances. Absence of H4'-$^{31}$P correlations for A$_4$ and A$_3$ were observed.
Figure 3.18: A representative $^{31}$P-$^1$H HSQC spectrum of the conjugate (2 mM, D$_2$O at pH 6.5, 25 °C; 400 MHz) showing the backbone $^{31}$P resonances versus the H3'/H4'/H5', H5'" region.

3.6.3.5 Nucleic acid chemical shifts

The chemical shifts of assigned non-exchangeable and exchangeable protons, as well as those of the other nuclei of the conjugate (i.e. $^{13}$C, $^{15}$N and $^{31}$P) are listed in Table 3.1.
### Table 3.1 Chemical shifts (ppm.) of the nucleic acid component of the unlabelled conjugate

The protons were assigned using a combination of NOESY, TOCSY and DQF-COSY as described in the text. All chemical shifts are reported relative to internal 4,4-dimethylsilapentane-1-sulphonate.

<table>
<thead>
<tr>
<th></th>
<th>H8/H6</th>
<th>H5/H2</th>
<th>H1'</th>
<th>H2'</th>
<th>H2''</th>
<th>H3'</th>
<th>H4'</th>
<th>H5'</th>
<th>H5''</th>
<th>H7*</th>
<th>amino&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>7.64</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.19</td>
<td>2.25</td>
<td>2.42</td>
<td>4.78</td>
<td>4.22</td>
<td>3.95</td>
<td>3.91</td>
<td>1.88</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>C2</td>
<td>7.54</td>
<td>5.94</td>
<td>6.03</td>
<td>1.77</td>
<td>2.24</td>
<td>4.72</td>
<td>4.11</td>
<td>4.25</td>
<td>4.12</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.68/7.35</td>
</tr>
<tr>
<td>A3</td>
<td>8.21</td>
<td>7.89</td>
<td>5.99</td>
<td>2.61</td>
<td>2.74</td>
<td>4.96</td>
<td>4.32</td>
<td>4.04</td>
<td>3.94</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.69</td>
</tr>
<tr>
<td>A4</td>
<td>8.35</td>
<td>7.84</td>
<td>6.17</td>
<td>2.70</td>
<td>2.75</td>
<td>5.08</td>
<td>4.44</td>
<td>4.20</td>
<td>4.20</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.84</td>
</tr>
<tr>
<td>T5</td>
<td>7.27</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.01</td>
<td>2.12</td>
<td>2.34</td>
<td>4.80</td>
<td>4.20</td>
<td>3.95</td>
<td>3.95</td>
<td>1.64</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>C6</td>
<td>7.79</td>
<td>5.99</td>
<td>6.22</td>
<td>2.36</td>
<td>2.24</td>
<td>4.54</td>
<td>4.09</td>
<td>4.09</td>
<td>4.03</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.71/7.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>C8/C6</th>
<th>C5/C2</th>
<th>C7</th>
<th>C1'</th>
<th>C2'</th>
<th>C3'</th>
<th>C4'</th>
<th>C5'</th>
<th>&lt;sup&gt;31&lt;/sup&gt;P</th>
<th>N4/N6</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>140.11</td>
<td>114.38</td>
<td>14.65</td>
<td>87.80</td>
<td>40.25</td>
<td>79.04</td>
<td>87.53</td>
<td>67.16</td>
<td>2.61</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>C2</td>
<td>143.64</td>
<td>99.28</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>88.02</td>
<td>40.33</td>
<td>78.51</td>
<td>86.70</td>
<td>66.87</td>
<td>-4.21</td>
<td>94.79</td>
</tr>
<tr>
<td>A3</td>
<td>141.84</td>
<td>155.17</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>85.48</td>
<td>39.53</td>
<td>80.13</td>
<td>87.75</td>
<td>68.27</td>
<td>-4.04</td>
<td>80.00</td>
</tr>
<tr>
<td>A4</td>
<td>141.78</td>
<td>155.13</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>84.98</td>
<td>40.42</td>
<td>78.95</td>
<td>87.28</td>
<td>67.83</td>
<td>-4.29</td>
<td>80.26</td>
</tr>
<tr>
<td>T5</td>
<td>138.96</td>
<td>113.98</td>
<td>14.36</td>
<td>86.30</td>
<td>40.38</td>
<td>76.64</td>
<td>85.94</td>
<td>67.98</td>
<td>-4.24</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>C6</td>
<td>144.13</td>
<td>99.02</td>
<td>NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>88.33</td>
<td>42.23</td>
<td>72.96</td>
<td>87.71</td>
<td>67.35</td>
<td>-3.86</td>
<td>94.02</td>
</tr>
</tbody>
</table>

<sup>a</sup> non-exchangeable proton chemical shifts are at 25°C.<br/>
<sup>b</sup> amino proton chemical shifts are at 5°C.<br/>
<sup>c</sup> not applicable
3.6.4 NMR strategies for conjugate peptide resonance assignments

As for the nucleic acid, the assignments of the peptide protons were attempted by adopting the sequential assignments strategy developed by Wüthrich and co-workers. The strategy is based on identifying protons within the same spin system i.e. by making use of scalar couplings between amide protons and the corresponding α-protons and between the latter with the protons in the side chain. The arginine guanidinium groups are an exception to this rule as they are not scalar coupled to the rest of the side chain and therefore represent an isolated spin system.

A key prerequisite to successful and complete peptide sequential assignments is the correctly chosen experimental conditions that allow observation of the backbone amide proton signals, which can be normally attained at low temperature and pH below neutral values where the exchange with the solvent is slow.

3.6.4.1 Backbone resonance assignment

The backbone H\text{N}, N\text{H}, C\text{α} and H\text{α} assignments were made by using a combination of $^1$H $^{13}$C-HSQC, $^1$H $^{15}$N-HSQC, TOCSY (both in D\text{2}O and H\text{2}O, pH 6.5, 5°C) for the assignments of the spin systems, followed by NOESY to achieve a sequence-specific assignment of the backbone protons.

All nine amide proton signals were identified using $^1$H $^{15}$N-HSQC in water (Figure 3.19), including C-terminal Gly\text{15}. Eight of these resonances were observed in the range 8.2 to 8.7 ppm, which corresponded to the amide protons of the amino acid residue Arg\text{8} to Gly\text{15}. The H\text{α} chemical shift (3.6 ppm) and the H\text{N} (3.7 ppm) of Leu\text{7} have values outside the minimum and maximum deposited at the BioMagResBank. These unusual upfield chemical shifts can be explained by the proximity of that amino acid to the attachment site with the oligonucleotide via a phosphoramidate bond. A key factor of the blocked de-shielding effect of the phosphoramidite N-atom lone pair on the bonded Leu\text{7} H\text{α} is that the N-atom non-bonding electron pair is delocalised on the asymmetrical three-center antibonding $\sigma^{*}_{\text{OPO}}$ orbital of the (O)\text{2}P-N bond. As a result, the electron density in the 1s orbital of the Leu\text{7} amide protons is not reduced by an inductive effect as it would be normally in the presence of an electron withdrawing amide N, and therefore the Leu\text{7} H\text{α} resonance is shifted upfield. The amide resonances did not exhibit substantial broadening,
which indicates that the proton exchange with the solvent is slow enough to make signals observable, probably due to involvement in hydrogen bonding within the molecule.

Figure 3.19: Section of the $^{15}$N HQSC spectrum of the amide protons recorded for the unlabelled conjugate (800 MHz, 5°C). The cross peak of Leu$_7$ (in the inset) is located outside the main plotted region of the spectrum and appears at $\omega_1 = 133.09$ ppm and $\omega_2 = 3.63$ ppm. H$^N$-$^{15}$N cross-peaks have been labelled with their correlating residue number.

The chemical shift dispersion of the peptide was judged by analysis of the backbone amide $^1$H$^N$ resonances in $^1$H-$^{15}$N HSQC spectrum which are normally better resolved than the $\alpha$-protons (Figure 3.19). The amino acid chemical shifts are not affected by the type of the neighbouring residues. For this reason, any variation from the random coil reference value (see BioMagResBank database) is due to either direct interaction with close in space functional groups or the consequent conformational changes themselves. Some conjugate residue amide protons (i.e. Arg$_8$ to Arg$_{10}$ and Arg$_{14}$) were observed resonating slightly more downfield, presumably due to polarization induced by hydrogen bonds of a weak and transient nature. The other amide proton signals are poorly dispersed, predicting flexibility.
in this peptide region. The NMR signals of the peptide backbone C$\alpha$ and H$\alpha$ residues were assigned using a combination of TOCSY (Figure 3.20-a) and a constant time $^1$H $^{13}$C-HSQC experiment (Figure 3.20-b), which enhanced spectral resolution and aided the resolving of the cross peaks overlapping in this overcrowded region.

Figure 3.20: The top two panels a) show the correlations $^1$H$\alpha$-$^1$H$N$ “fingerprint region” in the 800 MHz TOCSY spectrum of the conjugate (2 mM, H$_2$O 90%, D$_2$O 10%, at pH 6.5, 5 °C). The Gly$_{15}$ H$\alpha^1$ and H$\alpha^2$ are connected by a horizontal line. The residue number next to cross-peaks indicates $^1$H$\alpha$ and $^1$H$N$ resonance assignments. b) $^1$H $^{13}$C-HSQC spectrum showing the $^{13}$C-$^1$H region.
The Gly15 residue was unambiguously identified in the TOCSY spectrum recorded (Figure 3.20-a) in H2O based on its high intensity peak and characteristic fine structure. The single cross peak, corresponding to the correlation HN-Hα1/Hα2, shows a prominent large doublet splitting as a result of the geminal coupling constant of 15.0 Hz. No peak was detected for the two α-protons as they resonate at nearly identical chemical shifts.

3.6.4.2 Sequential backbone assignment

The strategy adopted for the assignment of the resonances sequence specifically was the Wüthrick walks18, which involves identification of neighbouring residues along the peptide chain by observing sequential NOE connectivities \( d_{NN} \) and \( d_{aN} \) between residue \( i \) and residue \( i+1 \).

Tracing of sequential amide-amide connectivities was attempted by starting from a Leu7 HN cross peak in a NOESY spectrum, which was identified in an upfield region, distinct from that of amide proton resonances. The sequential NOE walk to the HN of the residues \( i+1 \) (in the direction of the C-terminus) was then traced out in the downfield region (Figure 3.21-a). The Gly15 lacked a connection to an HN \( (i+1) \) in the fingerprint region. However, it was readily identified by its characteristic \(^{15}\text{N} \) shift in the \(^1\text{H}-^{15}\text{N} \) HSQC spectrum, well separated from the other backbone amide proton resonances.

An attempt to trace out sequential NOE connectivities between amide HN protons and the same residue Hα and between the former and the preceding Hα was made from Leu7 and Gly15.

The expanded NOESY spectrum exhibiting NOEs between the amide protons (8.2-8.7 ppm) and Hα (3.6-4.4 ppm) is outlined in Figure 3.21-b.

However, the repetitiveness of the amino acids arginine and leucine in the sequence and very similar chemical shifts for HN and Hα for several residues resulted in partial overlap of some cross peaks in the NOESY spectra. Consequently, the sequential \( d_{NN} (i, i+1) \) and \( d_{aN} (i+1) \) assignments were only possible by combining analysis of the spin system proton resonances in HSQC and TOCSY experiments. Furthermore, the poor dispersion of the Hα and HN chemical shift of residues L11 to R14 hindered the intensity of the \( d_{aN} \) NOEs, and consequently prevented the derivation of NOE distance restraints, an essential first step towards secondary structural determination.
Figure 3.21: Expanded region of the NOESY spectrum of the conjugate acquired at 5°C and pH 6.5, with a mixing time of 300 ms. a) NOE connectivities between the amide protons H^N (8.2-8.7 ppm region) from Leu_7 to Gly_{15}. The amide proton assignments are listed along the diagonal and their sequential connectivities are traced by dotted lines. The long-range amide–amide NOEs between Leu_{9} and Leu_{11} are labelled with a **, and the NOE between Leu_{7} H^N and Leu_{9} H^N with *. b) NOE connectivities between the H^N and H^α (3.6-4.4 ppm) protons. Sequential NOEs to neighbouring amino acids (i +1) are indicated by dotted lines. Intra-residue cross peaks are labelled with the corresponding residue numbers according to the assignments based on sequential NOEs.

3.6.4.3 Side chain assignments

Assignments of the proton resonances within each side chain spin system were attempted using a combination of TOCSY and NOESY experiments (recorded both in D_2O and H_2O), which allowed connecting individual spin-systems to specific residues through the strong NOEs between covalently linked protons, including H^β, H^γ, and H^δ side chain protons. The constant time $^1$H$^{13}$C-HSQC (ct-HSQC) experiment was used in an attempt to correlate the proton resonances to the corresponding $^{13}$C carbon chemical shifts.
The ct-HQSC experiment gave considerable overlap of $^{13}$C signals in the peptide resonance region despite the removal of the $^{13}$C-$^{13}$C couplings. Therefore, it was not possible to assign the side chain $^{13}$C resonances to their individual amino acid residues. However, the use of each group set of $^{13}$C chemical shifts extracted from the BioMagResBank database and the discernment of the two very similar resonances of Arg $C^\gamma$ from Leu $C^\gamma$ through their cross-peak sign-inversion (been $J$ coupled to an even and odd number of carbons respectively) allowed identification of narrower chemical shift regions for the side chain protons of the two types of residues.

Since the conjugate peptide contains a repetitive sequence of alternating arginine and leucine residues, it was not possible to assign individual cross peaks to specific amino acid residues in the TOCSY spectral regions of the side chain resonance regions (from 0.8 to 1.9 ppm) due to severe overlap (Figure 3.22).

**Figure 3.22:** Expanded region of the TOCSY spectrum of the conjugate (2 mM, D$_2$O at pH 6.5, 25 °C, 800 MHz) showing the correlation between $\alpha$-protons and leucine and arginine side chain protons. The cross peaks of the arginine and leucine $H^\alpha$, $H^\beta$ and $H^\gamma$ appear heavily overlapped and the leucine $H^\delta$ and $H^{\delta 2}$ resonances were not distinguishable but instead appear as a single unresolved peak.
The identification of the arginine and leucine $H^{\alpha}$ and $H^{\beta2}/H^{\beta3}/H^{\gamma}$ spin systems was hampered by chemical shift degeneracies since almost all residue $H^{\alpha}-H^{\beta2}/H^{\beta3}$ correlations gave rise to extensive overlapping of cross peaks. Furthermore, the cross peaks from $\beta$-protons with nearly identical chemical shifts were located near the diagonal and thus were masked by the closely located diagonal cross peaks. This caused further ambiguity in the assignments of the peptide protons in the individual residues.

Although the peptide contains four leucine residues cross peaks are reasonably separated in the leucine methyl and $\alpha$-protons resonance regions (Figure 3.22). However, stereospecific assignments of the leucine $H^{\delta1/\delta2}$ were prevented by their signals being merged in a single unresolved peak.

In contrast, successful assignments were attained for the well separated cross peaks of Leu$_7$ $H^{\alpha}$-$H^{\beta}$ and Leu$_7$ $H^{\alpha}$-$H^{\gamma}$. Less crowded and partially resolved cross peaks were seen in the amide-aliphatic region and assignments of the L$_9$ and L$_{11}$ $H^{\beta}$ and $H^{\gamma}$, and R$_{12}$ $H^{\beta}$ were accomplished (Figure 3.23).

The spin system of the four arginine residues was identified until $H^{\epsilon}$ protons. The dispersion observed for the $H^{\epsilon}$ proton chemical shifts is not large enough to allow separation of these.
Figure 3.23: a) ct-1H^{13}C-HSQC spectrum (H$_2$O 90%, D$_2$O 10%, at pH 6.5, 5 °C, 26 ms, 800 MHz) of the leucine and arginine side chain protons showing similarity of the chemical shifts. Individual labelling of proton resonances is not possible due to severe signal overlap. Distinct regions of each type of proton are identified by inverted signal sign of peaks arising from $^{13}$C-$J$-coupled to an even number of carbon atoms. b) TOCSY spectrum (H$_2$O 90%, D$_2$O 10%, at pH 6.5, 5 °C) which correlates the amide protons to the arginine and leucine side chains.

Peripheral protons in side chains, instead, gave rise to two coalesced signals at 7.28 and 7.33 ppm corresponding to Arg$_8$/Arg$_{10}$ and Arg$_{12}$/Arg$_{14}$ respectively (Figure 3.24).
Figure 3.24: The expanded region of the TOCSY spectrum of the conjugate (H₂O 90%, D₂O 10%, at pH 6.5, 5 °C, 800 MHz) which correlates H\(^{\epsilon}\) protons of the four arginine residues with the other protons from the same spin system.

The four arginine side chain H\(^{\eta}\) proton correlations were observable in TOCSY recorded in H₂O at 5°C (Figure 3.25). However, it was not possible to distinguish H\(^{\eta11}\) (H\(^{\eta12}\)) and H\(^{\eta21}\) (H\(^{\eta22}\)) resonances since the arginine residues exhibited two broad unresolved coalesced signals at 6.54 and 6.99 ppm, corresponding to the averaged chemical shifts of their four \(\eta\) proton resonances; most likely to be due to fast exchange with the solvent.
Figure 3.25: The expanded region of the TOCSY spectrum of the conjugate (H$_2$O 90%, D$_2$O 10%, at pH 6.5, 5 °C, 800 MHz) which correlates the arginine H$^\eta_1$ to H$^\eta_2$. Selective assignment of these protons to the individual arginine residues is prevented due to considerable overlap of their broad signals. The label Arg* indicates either Arg$_8$, Arg$_{10}$, Arg$_{12}$ and/or Arg$_{14}$.

The same pattern was observed in the NOESY spectrum recorded in H$_2$O at 5°C and this can indicate that the signals’ averaging arises from exchange between the NH$^\eta$ nuclei caused by free rotations about the N$^\varepsilon$-C$^\zeta$ but hindered rotation about C$^\zeta$-N$^\eta_b$ bonds. This normally occurs when the guanidinium group is involved in formation of a network of transient intramolecular salt bridges and therefore is highly dynamic.

3.6.4.4 Peptide medium and long-range assignments

Based on the knowledge of previously assigned chemical shifts of the peptide, an attempt to assign medium- and long-range NOEs within the peptide was made for the prediction of the peptide secondary structure by recording NOESY experiments in H$_2$O.

The well-resolved cross peaks observable in the NOESY spectrum was the medium-range H$^N$-H$^N$ NOEs between the pairs Leu$_7$-Leu$_9$, and Leu$_9$-Leu$_{11}$ and H$^\alpha$ /H$^\beta_2$-$^\beta_3$-H$^N$ NOEs between Leu$_7$ and Leu$_9$.

Long-range backbone H$^N$ - H$^N$ NOEs were also observed in the amide region between Leu$_7$ and Leu$_{11}$, as well as H$^\alpha$ -H$^N$ NOE between Leu$_7$ H$^\alpha$ and Leu$_9$ H$^N$. The presence of short distances between these two pairs of residues ($i$+2) and between Leu$_7$ and Leu$_{11}$ ($i$+4), are summarised in Table 3.2, is indicative of a $\beta_{10}$ helix in this peptide segment.
Table 3.2 Medium and long range sequential peptide assignments

<table>
<thead>
<tr>
<th>amino acid (i)</th>
<th>proton type</th>
<th>amino acid (i+2), (i+4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu7</td>
<td>H_N</td>
<td>Leu9 H_N</td>
</tr>
<tr>
<td></td>
<td>H_a</td>
<td>Leu9 H_N, Leu11 H_N</td>
</tr>
<tr>
<td></td>
<td>H_b</td>
<td>Leu9 H_N, Leu11 H_N</td>
</tr>
<tr>
<td>Leu9</td>
<td>H_N</td>
<td>Leu11 H_N</td>
</tr>
</tbody>
</table>

3.6.5 Identification of intramolecular NOEs

Identification of potential intramolecular interactions of the peptide with the linked oligonucleotide is needed to determine the structure of the complex. The information contained in the NOESY spectrum recorded in H_2O is high. However, this is only partially accessible due to the presence of a large number of NOE interactions leading to extensive cross peak overlap, particularly in the crowded aliphatic and aromatic regions, which contain both peptide and nucleic acid resonances.

The NOE cross peaks (Figure 3.26) distinctly visible were from Leu_7 and Leu_9 amide protons to T_1 and A_3 aromatic and sugar protons. From Leu_7, a larger number of NOE interactions with the nucleotide were observed, most likely because this amino acid residue is directly linked to the nucleic acid part of the conjugate. In addition, Leu_7 resonates in a spectral region which is free from both peptide and nucleotide resonances and consequently it is less affected by ambiguous assignments.

The multiple NOE interactions from A_3 amine groups to arginine backbone and side chain protons as well as to leucine methyl groups are of particular interest. This shows that the A_3 base is positioned very closely to the peptide in the conjugate structure. However, the distances could not be extracted from NOE-connectivities observed from this standard NOESY experiment, since the failure to resolve individual side chain protons of the arginine and leucine residues did not allow identification of those specific residues which were responsible for NOE contacts with A_3. The weak NOEs observed to arginine guanidine H_η1/ H_η2 could indicate the existence of transient hydrogen bonds with both A_3 amine hydrogens.
Figure 3.26: Sections of the NOESY spectrum (H$_2$O 90%, D$_2$O 10%, at pH 6.5, 5 °C, 300 ms mixing time, 800 MHz) showing NOE cross-peaks between the peptide protons and nucleotide protons, located far apart in the conjugate sequence. The summary of these interactions is given in Table 3.3.
The NOE restraints derived from the NOESY spectrum were not sufficient to determine the conjugate structure with high accuracy. Therefore, a different experimental approach is needed to identify each NOE without ambiguity by employing for example, site-selective $^{13}$C and $^{15}$N labelling of the peptide residues to distinguish the NOEs arising from protons of the same residue type.

Nevertheless, the NOE data were sufficient to extract some initial information for the prediction of the conjugate structure. In fact, the observation of NOEs between the oligonucleotide residues and peptide component, including those which are located far apart within the conjugate sequence, allowed delineation of a compact conformation of the conjugate in which the peptide seems to bend towards the nucleic acid. This U-like shape is presumably driven by electrostatic interactions between polarised groups, which were identified from the NOEs between the A3 amine protons and the arginine guanidinium group, as well as from possible hydrogen bonding suggested by the presence of NOEs between the base protons of oligonucleotides and arginine or leucine amide protons (see Table 3.3).

**Table 3.3** Intramolecular NOEs between the peptide and the nucleotide protons within the conjugate (Summary of the data presented in Figure 3.26).

<table>
<thead>
<tr>
<th>amino acid</th>
<th>proton type</th>
<th>nucleotide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu7</td>
<td>H$^N$</td>
<td>T1 (H6, H1’, H2’, H2”, H3’, H5’, H5”)</td>
</tr>
<tr>
<td></td>
<td>H$^\alpha$</td>
<td>T1 (H6, H1’, H2’, H2”, H3’, H5’, H5”)</td>
</tr>
<tr>
<td></td>
<td>H$^{\beta2/\beta3}$</td>
<td>T1 (H6, H2’, H2”, H3’, H5’, H5”)</td>
</tr>
<tr>
<td>Arg8</td>
<td>H$^N$</td>
<td>A3 H61</td>
</tr>
<tr>
<td>Leu9</td>
<td>H$^N$</td>
<td>T1 (H6, H1’, H2’, H2”, H3’, H5’, H5”)</td>
</tr>
<tr>
<td>Arg*</td>
<td>H$^{\gamma2/\gamma3}$</td>
<td>A3 H61, A3 H62</td>
</tr>
<tr>
<td>Arg*</td>
<td>H$^\eta$</td>
<td>A3 H61, A3 H62</td>
</tr>
<tr>
<td>Leu*</td>
<td>H$^{\delta1/\delta2}$</td>
<td>A3 H61, A3 H62</td>
</tr>
</tbody>
</table>

Arg* either Arg8, Arg10, Arg12 and/or Arg14
Leu* either Leu7, Leu9, Leu11 and/or Leu13
3.6.6 Chemical shift temperature dependence

A series of $^1$H variable temperature experiments was carried out to investigate the effect of temperature on the $^{1}H$ peptide resonances, thereby providing important information on the peptide structure within the conjugate. The temperature was raised from 5 °C to 35 °C in increments of 5 °C and changes in chemical shifts recorded. At higher temperatures, the peptide amide signals are reduced in intensity, presumably due to increased exchange with the solvent.

All amino acid amide protons manifested an up-field shift as the temperature increased to 35 °C. The residues L7 to R10 $H^N$ showed a weakened downfield shift, presumably due to their involvement in hydrogen bonding with carbonyl groups or nucleic acid bases. Significantly more negative shifts were observed for the other amino acids $\geq 0.25$ ppm from L11 towards the C-terminal residue, as expected for a flexible region. Temperature coefficients were determined for the amide bond of each residue by monitoring changes in chemical shift of the amide protons. Hydrogen bond restraints can be generated from these.

The dependence of the amide proton chemical shifts with temperature is related to the presence of hydrogen bonds and their up-field or downfield shift give a measure of the distance between the amide proton and the acceptor atom. The lengthening of the hydrogen bond caused by an increase in temperature leads to a reduced deshielding effect from the acceptor resulting in a shift toward higher field of the amide proton resonances as was observed experimentally.

An empirical rule was proposed by Cierpicki et al.\textsuperscript{21} that a $^1H^N$ is hydrogen bonded with predictivity higher than 85% when its temperature coefficient is more negative than $-4.6 \times 10^{-3}$ ppm/K. More positive values are observed for amide protons involved in hydrogen bonds in different secondary structures. For temperature coefficients in the range -11 to -5 \times 10^{-3} ppm/K the amide protons are not involved in hydrogen bonds and are exposed on the molecule surface and interact with water molecules. These values are normally found in random coil peptides\textsuperscript{22}. The equation used to calculate the amide bond temperature coefficient was $\delta / dT$.

The amide proton chemical shifts displayed linear temperature dependence with representative correlations shown in Figure. 3.27. Slopes of the linear fits in each plot corresponded to the chemical shift temperature coefficient. It was observed that Leu$^7$ to Arg$^{10}$ and Gly$^{15}$ have a less negative proton chemical shift temperature coefficient,
indicative of presence of secondary structure in this peptide region. The values of the temperature coefficient more negative than $-6 \times 10^{-3}$ ppm/K for peptide residues L_{11} to Arg_{14} amide protons predictive of either a random-coil structure of the peptide in this region or that these amide protons are involved in transient hydrogen bonding and thus the rise of temperature causes loss of secondary structure.

**Figure 3.27:** Plots of the correlation between the backbone amide proton chemical shift and temperature for the peptide residues from Leu$_7$ to Gly$_{15}$. The residue amide proton chemical shift temperature coefficient derived from $^1$H chemical shift measured at 5, 10, 20, 25, 30 and 35 $^\circ$C. The best-fit lines are indicated.
3.6.7 Assignment of the peptide protons assisted with isotopic labelling

Although the oligonucleotide base and sugar protons have been unambiguously assigned in the unlabelled conjugate, assignments of the peptide side chain remained ambiguous especially for the arginine residues, due to severe signal overlap in the spectra obtained by the standard 2D NMR experiments. This gave rise to indistinguishable NOE peak patterns in the peptide resonance regions. Intramolecular NOEs between the peptide Leu7/Leu9 and the nucleic acid T1 and A3 aromatic and sugar protons have already been observed in 2D NOESY spectra of the unlabelled conjugate recorded both in D2O and H2O. In order to fully determine a high-resolution 3D structure, it is essential to obtain a sufficient number of NOE-derived distance restraints within the conjugate, which requires complete assignment of all protons involved in the interaction network.

Site-specific uniform 13C and 15N arginine and leucine residue labelling was beneficial for this peptide with highly repetitive nature of [Leu-Arg] building blocks to achieve assignments of the backbone and the side chain resonances. Four samples were chemically synthesised to determine the structure of the peptide attached to the conjugate, in which two amino acids in four pairs were chosen for selective labelling: 13C, 15N labelled arginine and leucine residues were introduced alternatively at specific positions. This was achieved by incorporating isotopically labelled amino acids arginine (13C6 and 15N4) and leucine (13C6 and 15N) during the solid phase peptide synthesis. Each of these four samples contained a [Leu-Arg] labelled block at a specific position, i.e. at Leu7Arg8 (S2), Leu9Arg10 (S3), Leu11Arg12 (S4) and Leu13Arg14 (S5).

The drawback of this approach is that four samples needed to be synthesised and a set of experiments needed to be recorded for each of these samples. The sample conditions and the types of experiments that were applied for the site-specific uniformly 13C, 15N labelled conjugates are illustrated in Table 3.4.
Table 3.4 NMR experiments used for the assignments and structure determination of the labelled conjugates (S2) - (S5) containing residue-specific $^{13}$C, $^{15}$N labelled peptide. The purpose of each experiment, along with the extracted information, is indicated.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>experiments</th>
<th>purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>D$_2$O</td>
<td>ct-$^{1}H$ $^{13}$C-HSQC</td>
<td>peptide C$\alpha$, H$\alpha$ and side chain assignment</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>$^{1}H-^{15}$N HSQC (5°C)</td>
<td>peptide H$\gamma$, N$\epsilon$, H$\delta$ and H$\eta$ assignment</td>
</tr>
<tr>
<td>D$_2$O</td>
<td>2D $^{1}H$ $^{13}$C-HSQC-NOESY</td>
<td>extracting $^{13}$C labelled peptide a.a.-nucleic acid NOE</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>2D $^{1}H-^{15}$N HSQC-NOESY (5°C)</td>
<td>extracting $^{15}$N labelled peptide a.a.-unlabelled peptide part and nucleic acid NOE</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>2D $^{13}$C, $^{15}$N - filtered/edited NOESY (5°C)</td>
<td>extracting NOEs within unlabelled peptide part and between unlabelled peptide a.a. and nucleic acid</td>
</tr>
</tbody>
</table>

The combination of the two types of labelled amino acid (i.e. arginine and leucine) gave an optimal simplification of the constant time $^{1}H$ $^{13}$C-HSQC spectra, in which the overlap was reduced in the crowded regions of arginine and leucine side chain protons H$\beta$, H$\gamma$, H$\delta$. Indeed, arginine backbone/side chain carbon correlations to H$\beta$/H$\gamma$/H$\delta$ were well-separated from the leucine $^{13}$C- H$\beta$/H$\gamma$/H$\delta$ correlations. The severe signal overlap induced by the repetitiveness in the peptide sequence was circumvented, and thus the amino acid sidechain chemical shifts that could not be assigned earlier due to overlapped resonances were identified with the exceptions of the degenerated protons such as leucine methyl protons and all four arginine side chain H$\delta$ protons (Figure 3.28).
Figure 3.28-a. $^1$H $^{13}$C-HSQC spectra (2 mM, D$_2$O at pH 6.5, 25 °C, 800 MHz) of the samples contained $^{13}$C, $^{15}$N labelled arginine and leucine residues introduced at positions: Leu$_7$Arg$_8$ (a) and Leu$_9$Arg$_{10}$ (b). The $^1$H-$^{13}$C correlations seen in each spectrum could be assigned to the labelled residues.
The 2D $^1$H $^{13}$C HSQC-NOESY in D$_2$O allowed identification of NOEs between the protons attached to the $^{13}$C and the other protons while suppressing NOEs arising from protons attached to $^{12}$C. This experiment was particularly useful in detecting the intramolecular interactions between the nucleic acid aromatic/sugar H1' protons and the labelled peptide side chain protons (Figure 3.29). Interestingly, all six nucleotides showed NOEs with the peptide residues, including those which were distant from the peptide components in the conjugate sequence, indicative of a folded conformation where the two component residues face each other. However, no NOEs were observed for Leu$_{11}$ and Leu$_{13}$ presumably
because of lack of interactions with the oligonucleotide component. On the other hand, weak NOEs were observed for peptide residues Arg_{12} to Arg_{14} with the oligonucleotide 3' end (i.e. T_5 and C_6). The drawback was that only a section of the NOESY spectrum (i.e. 5.9 to 8.4 ppm in $\omega_1$ dimension, and 23 to 56 ppm in $\omega_2$ dimension) can be analysed in this labelling method. The higher field spectral region (i.e. 0.8 to 5 ppm) suffers from contributions of strong direct $^1$H-$^{13}$C correlations of the labelled arginine/leucine side chain protons which masked the contribution from $^1$H-$^1$H NOE.

**Figure 3.29:** Expanded regions of the 2D $^1$H-$^{13}$C HSQC-NOESY spectra (2 mM, D$_2$O at pH 6.5, 25 °C, 800 MHz) recorded for the four $^{13}$C, $^{15}$N labelled samples Leu$_7$Arg$_8$ (S2) (a), Leu$_9$Arg$_{10}$ (S3) (b), Leu$_{11}$Arg$_{12}$ (S4) (c), Leu$_{13}$Arg$_{14}$ (S5) (d). Leu$_7$Arg$_8$ conjugate (a) and Leu$_9$Arg$_{10}$ (b) display the larger number of NOE contacts as compared with Leu$_{11}$Arg$_{12}$ (c) and Leu$_{13}$Arg$_{14}$ conjugates (d), respectively, showing the gradual decrease in number of NOE interactions between distal oligonucleotide and peptide residues within the conjugate. No intra-molecular NOEs were observed between Leu$_{11}$ and the oligonucleotide part of the conjugate.
The only detectable NOEs in the upfield region were those from the well-separated signals of Leu\textsubscript{7}. Long range sequential NOEs (i+3) were observed with Arg\textsubscript{10} H\textsubscript{δ}, supporting the presence of helical conformation for the peptide in this segment (Figure 3.30).

**Figure 3.30:** Expanded region of the 2D $^1$H$^{13}$C HSQC-NOESY spectrum (2 mM, D\textsubscript{2}O at pH 6.5, 25 °C, 800 MHz) recorded for the $^{13}$C, $^{15}$N labelled [Leu\textsubscript{9}Arg\textsubscript{10}] conjugate, showing the two NOE cross peaks between Leu\textsubscript{7} and Arg\textsubscript{10} in the spectral region of strong interference from direct $^1$H-$^{13}$C correlation peaks.

Assignments of the H\textsuperscript{N} and nitrogen chemical shifts as well as sequential and through space H\textsuperscript{N} and H\textsuperscript{α}/H\textsuperscript{β2/β3} connectivities were achieved in 2D $^1$H-$^{15}$N HSQC-NOESY experiments at 5°C in combination with 2D $^1$H-$^{15}$N HSQC in H\textsubscript{2}O. For the sample of $^{13}$C, $^{15}$N labelled [Leu\textsubscript{13}Arg\textsubscript{14}] conjugate only direct $^1$H-$^{15}$N correlations were observed, presumably indicating a high flexibility of the peptide backbone in proximity to the C-terminus.

The spectra of the $^{13}$C, $^{15}$N labelled [Leu\textsubscript{9}Arg\textsubscript{10}] conjugate (S3) in particular (Figure 3.31-a), provided crucial information for the determination of the peptide residue spatial arrangement by revealing short and medium sequential $^1$H-$^1$H-NOEs among the residues Leu\textsubscript{7}, Leu\textsubscript{9}, Leu\textsubscript{10} and Leu\textsubscript{11} (as outlined in Figure 3.31-b). This crucial information allowed derivation of additional distance restraints for the conjugate structure calculation.

The residue Arg\textsubscript{8} signals were not visible as they were masked by the strong $^1$H-$^{15}$N correlations.

The presence of the medium range NOEs between the H\textsuperscript{α} proton of Leu\textsubscript{7} and the N\textsuperscript{H} of the residue i+3 Arg\textsubscript{10} and between Leu\textsubscript{7} H\textsuperscript{N} and Leu\textsubscript{9} N\textsuperscript{H}, and of NOEs between labelled Arg\textsubscript{10} with T\textsubscript{1} sugar protons H5'/H5'' (Figure 3.31-b) provided further evidence for the helicity of this region and compact conformation of the conjugate.
Figure 3.31: a) Expanded region of the 2D $^1$H$^{15}$N HSQC-NOESY spectra ($H_2O$ 90%, $D_2O$ 10%, at pH 6.5, 5 °C, 300 ms mixing time, 800 MHz) recorded for $^{13}$C, $^{15}$N labelled [Leu$^9$Arg$^{10}$] conjugate, which displays the assigned sequential NOE connectivities between amide protons and their own and sequential H$\alpha$, H$\delta$ protons. b) The residues from a) are shown in red for the arginine and green for the leucine residues, respectively, and yellow for T1 with the H$N$-H$N$ and H$N$-H$\alpha$ highlighted.
H\(^n\) protons resonances were observed in \(^1\)H-\(^{15}\)N HSQC spectrum for the residue Arg\(_{10}\) for the labelled [Leu\(_9\)Arg\(_{10}\)] conjugate (S3) only, which showed four resolved signals for the four NH\(^n\) protons and two N\(^n\) nitrogens (Figure 3.32). The high field pair of signals produces resonances at 6.99 ppm resulting from averaging of H\(^{\eta11}\) and H\(^{\eta12}\) and 70.61 and 72.87 ppm from and N\(^{\eta1}\) and N\(^{\eta2}\) signals. These chemical shifts values, which were very similar to those of a random coil, presumably indicate that the guanidine group is not involved in stable hydrogen bonding interaction.

**Figure 3.32:** Part of the 2D \(^1\)H-\(^{15}\)N HSQC-NOESY spectrum showing the guanidine NH\(^n\) signals of Arg\(_{10}\) in [Leu\(_9\)RArg\(_{10}\)] conjugate (S3) recorded at 800 MHz at 5°C. The four resolved signals detected for the NH\(^n\) nuclei in Arg\(_{10}\) are labelled. The cross peaks show the correlations between protons in the two NH\(^n\) groups, the NOE correlations between guanidino protons and their own H\(^\delta\), as well as NOE correlations between the H\(^\epsilon\) and their own H\(^\delta\) and H\(^\beta\).

No assignment has been achieved for the guanidine H\(^n\) in the other arginine residues, which is most likely to be due to their transient association with the phosphates of the oligonucleotide sugar-phosphate backbone. This fast exchange between different states and conformations makes their resonances undetectable on the NMR time-scale.

The two-dimensional \(^{13}\)C, \(^{15}\)N–filtered/edited NOESY experiments were crucial in further identification of the conjugate intramolecular NOEs. These NOESY spectra recorded in D\(_2\)O for the four labelled samples detected only NOE cross peaks involving
$^{12}$C-attached protons of both oligonucleotide and peptide components as the source and the destination proton while suppressing all other NOE cross-peaks by isotope filtering.

The $\omega_1^{13}$C, $^{15}$N–filtered/edited NOESY experiments recorded in H$_2$O were used to identify NOEs from H$^N$ and other exchangeable protons attached to $^{14}$N of the unlabelled residue protons in the presence of the two $^{13}$C, $^{15}$N labelled arginine and leucine residues.

The unambiguous assignments from the HSQC-NOESY and $\omega_1$-filtered NOESY experiments have provided a complementary dataset that further disambiguated the highly overlapping NOEs of the repetitive peptide sequence, which helped expand the repertoire of the NOE assignments.

The previously missing side chain proton assignments of the peptide have been completed. The fully assigned conjugate peptide is given in Table 3.5.
### Table 3.5 Chemical shifts (ppm) of the peptide within the conjugate

The protons were assigned by using a combination of NOESY, HSQC, 2D HSQC-NOESY and $\omega_1$-filtered NOESY as described in the text. All chemical shifts are reported relative to internal 4,4-dimethylsilapentane-1-sulphonate.

<table>
<thead>
<tr>
<th></th>
<th>$H^N$</th>
<th>$H^\alpha$</th>
<th>$H^\beta$</th>
<th>$H^\gamma$</th>
<th>$H^\delta$</th>
<th>$H^\varepsilon$</th>
<th>$H^\eta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu7</td>
<td>3.72</td>
<td>3.63</td>
<td>1.43, 143</td>
<td>1.63</td>
<td>0.86, 0.83</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Arg8</td>
<td>8.48</td>
<td>4.25</td>
<td>1.87, 177</td>
<td>1.63, 1.57</td>
<td>3.16</td>
<td>7.33</td>
<td>-</td>
</tr>
<tr>
<td>Leu9</td>
<td>8.69</td>
<td>4.27</td>
<td>1.70, 1.54</td>
<td>1.59</td>
<td>0.91, 0.83</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Arg10</td>
<td>8.48</td>
<td>4.22</td>
<td>1.79</td>
<td>1.62, 1.57</td>
<td>3.12</td>
<td>7.32</td>
<td>6.99, 6.55</td>
</tr>
<tr>
<td>Leu11</td>
<td>8.25</td>
<td>4.26</td>
<td>1.64, 153</td>
<td>1.58</td>
<td>0.88, 0.82</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Arg12</td>
<td>8.34</td>
<td>4.29</td>
<td>1.74, 1.79</td>
<td>1.65, 1.58</td>
<td>3.18</td>
<td>7.29</td>
<td>-</td>
</tr>
<tr>
<td>Leu13</td>
<td>8.36</td>
<td>4.33</td>
<td>1.65, 1.54</td>
<td>1.56</td>
<td>0.90, 0.85</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Arg14</td>
<td>8.50</td>
<td>4.31</td>
<td>1.81, 1.89</td>
<td>1.63, 1.67</td>
<td>3.20</td>
<td>7.29</td>
<td>-</td>
</tr>
<tr>
<td>Gly15</td>
<td>8.54</td>
<td>3.94, 3.89</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$C^\alpha$</th>
<th>$C^\beta$</th>
<th>$C^\gamma$</th>
<th>$C^\delta$</th>
<th>$N^H$</th>
<th>$N^\varepsilon$</th>
<th>$N^\eta^1$</th>
<th>$N^\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu7</td>
<td>57.40</td>
<td>45.59</td>
<td>27.02</td>
<td>25.08, 23.95</td>
<td>NA</td>
<td>133.09</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Arg8</td>
<td>56.54</td>
<td>30.70</td>
<td>27.50</td>
<td>43.35</td>
<td>-</td>
<td>120.57</td>
<td>84.31</td>
<td>-</td>
</tr>
<tr>
<td>Leu9</td>
<td>55.98</td>
<td>42.20</td>
<td>27.21</td>
<td>24.97, 23.62</td>
<td>NA</td>
<td>123.40</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Arg10</td>
<td>56.65</td>
<td>30.48</td>
<td>27.44</td>
<td>43.37</td>
<td>-</td>
<td>121.47</td>
<td>84.36</td>
<td>72.88, 70.57</td>
</tr>
<tr>
<td>Leu11</td>
<td>55.56</td>
<td>42.20</td>
<td>27.18</td>
<td>25.05, 23.39</td>
<td>NA</td>
<td>122.33</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Arg12</td>
<td>56.34</td>
<td>30.50</td>
<td>27.35</td>
<td>43.43</td>
<td>-</td>
<td>121.93</td>
<td>84.32</td>
<td>-</td>
</tr>
<tr>
<td>Leu13</td>
<td>55.33</td>
<td>42.21</td>
<td>27.21</td>
<td>24.87, 23.25</td>
<td>NA</td>
<td>123.28</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Arg14</td>
<td>56.45</td>
<td>30.60</td>
<td>27.40</td>
<td>43.42</td>
<td>-</td>
<td>122.00</td>
<td>84.31</td>
<td>-</td>
</tr>
<tr>
<td>Gly15</td>
<td>45.03</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>111.15</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

\(a\) non-exchangeable proton chemical shifts are at 25 °C.

\(b\) labile protons and N chemical shifts are at 5 °C.

\(c\) not applicable
3.6.8 Solution structure of the conjugate

3.6.8.1 Experimental restraints

The primary NMR data used in the calculation of the secondary structure of the conjugate was the distance constraints based on the analysis of the following experiments used in combination: 2D-NOESY experiments in D$_2$O and in H$_2$O recorded for the unlabelled conjugate, 2D $^{13}$C-HSQC-NOESY in D$_2$O, 2D $^{15}$N-HSQC-NOESY in H$_2$O, 2D $\omega_1$-$^{13}$C, $^{15}$N–filtered/edited NOESY in H$_2$O recorded for the four labelled conjugates. Distance-restraint information was extracted by measuring the volumes of the NOE cross-peaks and converting into distances using MARDIGRAS.

88 inter-residue restraints were identified between the 9-mer peptide and the 6-mer oligonucleotide in the conjugate molecule in solution. These inter-residue NOEs are listed in Table 3.6 and establish that the residues Leu$_7$ to Arg$_{10}$ backbone and side chain protons are involved in a significant number of NOEs with the conjugate oligonucleotide part, particularly with T$_1$ H6 and its sugar protons. It should be noted that Leu$_9$ exhibits NOEs with the majority of all six nucleotide sugars. In addition, Leu$_9$ is in extensive contact with the A$_3$ base protons (H2, and/or H61/H62) through its H$^N$, H$^{b2/b3}$ and H$^{b1/b2}$ protons. Arg$_{12}$ and Arg$_{14}$ form very few contacts with the oligonucleotide, suggesting a dynamic conformation of the conjugate structure in this region. In contrast, only intra-peptide contacts are observed for Leu$_{11}$, Leu$_{13}$ and Gly$_{15}$.

An initial attempt to identify the peptide structure was made from the three-bond $^3J_{HNH_\alpha}$ coupling constants and NOE patterns involving backbone amide protons.

The NMR experiment 3D HNHA$^{23}$, which correlates the $\alpha$-protons with their own amide, was used to yield the $^3J_{HNH_\alpha}$ coupling constants by measuring the intensity ratio of the diagonal-peak and H$^N$-H$^\alpha$ cross-peak in $\omega_2$ dimension.

Seven $^3J_{HNH_\alpha}$ couplings were measured and subsequently used to estimate the values of the $\phi$ backbone dihedral angle (see equation 1.6-10 and 1.6-7). The Arg$_{12}$ H$^N$-H$^\alpha$ cross-peak was not visible in the HNHA spectrum and thus, no measurement was made for this residue. The three-bond $^3J_{HNH_\alpha}$ values measured for the conjugate peptide are summarised in Table 3.7.
Table 3.6 NOEs contacts in the peptidyl-oligonucleotide conjugate

<table>
<thead>
<tr>
<th>amino acid</th>
<th>proton type</th>
<th>nucleotide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu7</td>
<td>H_N</td>
<td>T1 (H6, H2″, H3′, H4′, H5′, H5″), C2 H5</td>
</tr>
<tr>
<td>Leu7</td>
<td>H_α</td>
<td>T1 (H6, H1′, H2″, H3′, H5′, H5″)</td>
</tr>
<tr>
<td>Leu7</td>
<td>H^β2/β3</td>
<td>T1 (H6, H1′, H2″, H3′, H5′, H5″)</td>
</tr>
<tr>
<td>Leu7</td>
<td>H_δ1/δ2</td>
<td>T1 (H6, H3′)</td>
</tr>
<tr>
<td>Leu7</td>
<td>H^δ1/δ2</td>
<td>T1 (H6, H1′, H2″, H3′, H5′, H5″)</td>
</tr>
<tr>
<td>Arg8</td>
<td>H_N</td>
<td>T1 (H6, H1′, H2″, H3′, H5′, H5″), C2 (H5, H6), A3 H61</td>
</tr>
<tr>
<td>Arg8</td>
<td>H_α</td>
<td>T1 (H6, H3′)</td>
</tr>
<tr>
<td>Arg8</td>
<td>H^β2/β3</td>
<td>T1 (H1′, H2″)</td>
</tr>
<tr>
<td>Arg8</td>
<td>H^δ2/γ3</td>
<td>T1 (H6, H1′, H3′)</td>
</tr>
<tr>
<td>Arg8</td>
<td>H^δ2/δ3</td>
<td>T1 (H1′, H5′)</td>
</tr>
<tr>
<td>Leu9</td>
<td>H_N</td>
<td>T1 (H6, H1′, H3′, H4′, H5′, H5″), A3 H2</td>
</tr>
<tr>
<td>Leu9</td>
<td>H^β2/β3</td>
<td>T1 (H2″, H3′, H5″), A3 H2, T5 (H1′, H2″, H5″)</td>
</tr>
<tr>
<td>Leu9</td>
<td>H^δ1/δ2</td>
<td>T1 H5′, A3 (H2, H1′, H61, H62), A4 H1′, T5 H1′, C6 H5′</td>
</tr>
<tr>
<td>Arg10</td>
<td>H_N</td>
<td>T1 (H6, H1′, H3′, H5′, H5″), A3 H2</td>
</tr>
<tr>
<td>Arg10</td>
<td>H^β2/β3</td>
<td>T1 H5′</td>
</tr>
<tr>
<td>Arg10</td>
<td>H^δ2/δ3</td>
<td>T1 (H1′, H5′), C2 (H5, H1′)</td>
</tr>
<tr>
<td>Arg12</td>
<td>H^δ2/δ3</td>
<td>T5 H1′, C6 H1′</td>
</tr>
<tr>
<td>Arg14</td>
<td>H^δ2/δ3</td>
<td>T5 H1′, C6 H1’</td>
</tr>
</tbody>
</table>

The ϕ values found for the peptide residues are indicative of 3_{10}-helix conformations from residues Leu7 to Leu11. For Arg12 no J coupling was observed in the spectrum and Leu13 gave a ϕ value typical of a random coil conformation. Arg14 clearly does not contribute to a helix conformation, therefore is very likely to be involved in different network of hydrogen bonding through its amide.
Table 3.7 Experimental $^3J_{HNHa}$ coupling constants of the conjugate peptide

<table>
<thead>
<tr>
<th>Residue</th>
<th>$^3J_{HNHa}$ (Hz)</th>
<th>$\phi$ values (-30° to -180°)</th>
<th>predicted structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu7</td>
<td>5.6</td>
<td>-88</td>
<td>3$_{10}$-helix</td>
</tr>
<tr>
<td>Arg8</td>
<td>5.5</td>
<td>-86</td>
<td>3$_{10}$-helix</td>
</tr>
<tr>
<td>Leu9</td>
<td>5.3</td>
<td>-82</td>
<td>3$_{10}$-helix</td>
</tr>
<tr>
<td>Arg10</td>
<td>5.8</td>
<td>-95</td>
<td>3$_{10}$-helix</td>
</tr>
<tr>
<td>Leu11</td>
<td>5.8</td>
<td>-95</td>
<td>3$_{10}$-helix</td>
</tr>
<tr>
<td>Arg12</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Leu13</td>
<td>6.22</td>
<td>-106</td>
<td>random coil</td>
</tr>
<tr>
<td>Arg14</td>
<td>5.5</td>
<td>-86</td>
<td>3$_{10}$-helix</td>
</tr>
<tr>
<td>Gly15</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

In order to estimate $^3J_{HNHa}$ for Arg$_{12}$ and Gly$_{15}$, as well as restraints for the backbone, $\phi$ and $\psi$ dihedral angles were also predicted by using the program TALOS+, which compares the peptide H$_N$, N, C', H$_{\alpha}$, C$_{\alpha}$, and C$_{\beta}$ chemical shift assignments against fragments with similar chemical shifts and backbone geometry selected from the BMRB and PDB databases of assigned proteins of known structure. TALOS+ calculated backbone $\phi$ and $\psi$ dihedral angles for residues from Leu$_9$ to Arg$_{14}$, which were predictive of 3$_{10}$-helix conformations for Leu$_9$ and Leu$_{11}$ and of $\beta$-sheet conformations for Arg$_{12}$ to Arg$_{14}$. The experimental $^3J_{HNHa}$ value higher than 6 Hz was found for Leu$_{13}$ while the value calculated by TALOS+ is typical of a $\beta$-sheet strand indicates the possibility of a rapidly converting conformation of the peptide in proximity of the C-terminus non-detectable on the NMR time scale.

The existence of the helical structure is supported by the H$_N$-H$_N$ and H$_N$-H$_{\alpha}$ NOEs between Leu$_7$ and Arg$_{10}$ and by the H$_N$-H$_{\alpha}$ NOEs between Leu$_7$ and Leu$_{11}$ in the amide region of the 2D NOESY and 2D $^{15}$N-HSQC-NOESY spectra recorded in H$_2$O.

The temperature coefficient of their amide protons (Figure 3.27) which shows $d\delta/dT$ of approximately $-4.6 \times 10^{-3}$ ppm/K is indicative of the protons temperature dependence and therefore of their involvement in stabilising the $\alpha$-helix by hydrogen bonding (Figure 3.33).
Figure 3.33: Summary of short, medium and long range NOEs for the peptide. The thickness of the vertical black bars indicates the relative intensities of the NOE connectivities. Below the NOE patterns, the temperature coefficients of the amide protons are given in units $10^{-9}$ ppm/K and the experimental $^3J_{HNH\alpha}$ are shown.

The second type of geometry restraints that were used for structure calculation were the dihedral angle constraints from DQF-COSY cross peaks for the pucker ring. The conjugate oligonucleotide exhibited NMR spectral characteristics similar to base-stacking geometry in a B-form DNA duplex.

In particular, from the inspection of the pattern of the intra-residue and inter-residue NOE cross peaks generated by correlations between the aromatic and the ribose protons, it is clear that the conformation of the sugar puckers was predominantly C2’-endo, with strong intensity observed for intra-residue H6/H8-H2’, whereas the intra-residue H6/H8- H3’ were weaker (Figure 3.13). This is consistent with shorter distances between intra-residue H6/H8-H2’ (in the range 3.5-4.0 Å) and ~ 1 Å longer intra-residue H6/H8-H3’ distances in B-form DNA.¹⁵

Further evidence of this assumption was the higher field resonance of the sugar H2’ compared to H2” for all six nucleotides and by different multiplet splitting patterns of the
DQF-COSY cross peaks which showed large $^3J_{H1'H2'}$ with exception of terminal C$_6$. This instead exhibited an averaged $J$ coupling for H1’-H2’ and H1-H2’’ (presumably due to fast conformational exchange), and a small value $^3J_{H2'H3'}$ (< 2 Hz) which could not be detected in the multiplet lineshape, and this is consistent with C2’-endo ribose conformation (Figure 3.14).

The glycosidic angles, which reflect the orientation of the base relative to the sugar pucker also known as torsion angles $\chi$, were also determined to evaluate the preferred orientation adopted by the oligonucleotide residues. This information was derived by analysis of the intranucleotide NOE distances between the base protons and the sugar H1’, which were found in the range 3.5-3.7 Å. These distances correspond to the anti conformation, where the adenine rings and the oxygens of the thymine and cytosine point away from the ribose and the H6/H8 protons lie above the sugar pucker.

3.6.8.2 Simulated Annealing

A list of NOE distances derived from assignments of exclusively well-resolved cross peaks was used for the first round of distance geometry calculations. The procedure used to obtain an NMR structure for the conjugate involved the employment of 100 different starting structures, randomly modified by simulated annealing (SA).

The iterative rounds of simulated annealing calculations were continued, and adjustments were made to the NOESY peak lists between each calculation upon analysis of the initial 3D structures. These preliminary lists were expanded with additional unambiguous NOE assignments that were previously ambiguous and, therefore not included, by referring to the analysed structures. Additionally, consistently violated restraints, the NOEs misassigned from artefacts and more frequently generated by spin diffusion in proximity to adenine H2’ were re-evaluated or removed. At each round new calculations were then performed with a new set of restraints until no further improvement was observed in terms of NOE violations and an acceptable final ensemble of structures was obtained.
The lowest twenty energy conformations that satisfied together all the distance restraints, geometry and torsion angle geometry (non-violated by more than 0.5 Å) were then selected.

A final refinement calculation consisted in energy minimization of the twenty conformation ensemble (Figure 3.34-a) with NMR restraints performed in explicit water using the Sander module of the Amber software\textsuperscript{25} to optimise the hydrogen-bond network.

3.6.8.3 Structural analysis

The conformation of the oligonucleotide component of the conjugate NMR structure is precisely defined, and the peptide backbone is ordered from Leu\textsubscript{7} to Arg\textsubscript{14} (Figure 3.34-a). The precision of the NMR ensemble was evaluated by calculating the root-mean square deviation (RMSD) from one structure to another. Superposition of the twenty structures yielded an average root mean square deviation of 1.33 ± 0.2 Å for the heavy atoms, 0.963 ± 0.19 Å for the backbone and 0.665 ± 0.08 Å for the peptide backbone excluding C-terminal tail. The NMR structure from the ensemble that has the lowest energy and RMSD value was chosen to represent the whole ensemble (Figure 3.34-b). The refined NMR structures were then validated against the distance and dihedral angle constraints using MolProbity (Molprobity score was 1.65)\textsuperscript{26}.

The summary of the NMR constraints and structural statistics on the twenty refined structures are given in Table 3.8. Analysis of the ensemble of the twenty structures using RAMPAGE\textsuperscript{27} revealed that 98.3% of the peptide residues lie in the most favoured regions and 1.7% in the allowed regions of the Ramachandran $\phi$, $\psi$ dihedral-angle plot (Figure 3.35).
Figure 3.34: Solution NMR structure of the peptidyl-oligonucleotide conjugate (a) Views of the backbone traces for the refined twenty structures overlaid on the backbone heavy atoms in the secondary structure elements. The peptide backbone in ribbon representation is blue in the tight turn of the segment [Arg_{8}-Leu_{11}] and gray for the extended backbone following Arg_{12}. The oligonucleotide backbone is shown in orange, the bases in yellow and the phosphates in red. The average root-mean square deviations (RMSD) are as follows: all non-hydrogen atoms, 0.963 Å for the backbone, peptide backbone excluding C-terminal tail 0.665 Å. (b) Organisation of the lowest energy conformation of the conjugate NMR structure ensemble represented in stick and ribbon. (c) Surface (heavy atoms of the oligonucleotide) and stick representation of the lowest energy structure. The oligonucleotide surface is painted according to location of the hydrophobic and hydrophilic groups with red indicating the phosphates and the base polarizable sites and green indicating the hydrophobic regions. Surface transparency is used to convey the interactions between the positively charged arginine residues and the oligonucleotide phosphates. (d) View of the same structure as (c) at a 180° rotation round the y-axis.
**Table 3.8: NMR constraints and structure statistics**

**NMR distances and dihedral constraints**

<table>
<thead>
<tr>
<th>Distance restraints</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>275</td>
</tr>
<tr>
<td>2D NOESY</td>
<td>130</td>
</tr>
<tr>
<td>2D $^{13}$C/$^{15}$N HSQC-NOESY</td>
<td>103</td>
</tr>
<tr>
<td>2D $\omega$, $^{13}$C, $^{15}$N – filtered/edited NOESY</td>
<td>42</td>
</tr>
<tr>
<td>Oligonucleotide intramolecular distance constraints</td>
<td>111</td>
</tr>
<tr>
<td>Peptide intramolecular distance constraints</td>
<td>76</td>
</tr>
<tr>
<td>Intraresidue</td>
<td>27</td>
</tr>
<tr>
<td>Sequential ($</td>
<td>i-j</td>
</tr>
<tr>
<td>Medium range ($2 \leq</td>
<td>i-j</td>
</tr>
<tr>
<td>Long range ($</td>
<td>i-j</td>
</tr>
<tr>
<td>Peptide-oligonucleotide inter-residue NOEs</td>
<td>88</td>
</tr>
</tbody>
</table>

**Total dihedral angle restraints**

| Sugar pucker                                            | 10  |
| Backbone*                                               | 64  |

**Structural statistics on the final set of 20 simulated annealing structures of the conjugate**

<table>
<thead>
<tr>
<th>NOE violations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number $&gt; 0.2$ Å</td>
</tr>
<tr>
<td>Maximum violations (Å)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average pairwise r.m.s. deviation (Å)$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptide (residue L$<em>7$-R$</em>{14}$)</td>
</tr>
<tr>
<td>Heavy</td>
</tr>
<tr>
<td>Backbone</td>
</tr>
<tr>
<td>Oligonucleotide</td>
</tr>
</tbody>
</table>

**Ramachandran analysis$^c$**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in favoured regions</td>
<td>98.3%</td>
</tr>
<tr>
<td>Residues in allowed regions</td>
<td>1.7%</td>
</tr>
</tbody>
</table>

*a* geometric parameters for DNA bases derived by X-ray crystallography; peptide chirality restraints generated using Amber package.

*Pairwise r.m.s. deviation was calculated among twenty refined structures

*Based on ensemble of twenty refined structures
Figure 3.35: Ramachandran plots showing the backbone conformation of the ensemble of twenty refined structures. 98.3% of the peptide residues lie in the most favoured regions and 1.7% in the allowed regions. The $\phi$ and $\psi$ values for the lowest energy conformation of the conjugate NMR structure ensemble are shown in yellow.

The peptide within the conjugate adopts an S-shaped conformation, characterised by a $3_{10}$-helix conformation from residues Leu$_7$-Leu$_{11}$ and an extended conformation from the following residues (Figure 3.34-a,b). This is evidenced by the measured backbone $\phi$ and $\psi$ dihedral angles values, which fall within the $\alpha$-helical regions of the Ramachandran plot for the Leu$_7$-Leu$_{11}$ segment and beta sheet region for the Arg$_{12}$-Arg$_{14}$ segment (Figure 3.35). Gly$_{15}$, which is located on the peptide N-terminal tail and does not exhibit NOEs, except for those with neighbouring Arg$_{14}$, is not well defined in the conjugate structure. All the four arginine residues, together with Leu$_7$ and Leu$_9$, are oriented towards the oligonucleotide component whereas Leu$_{11}$ and Leu$_{13}$ are projected away from it. The oligonucleotide part adopts a bent conformation, stabilised by multiple contacts with peptide, mostly through the sugar-phosphate backbone rather than with the bases (Figure 3.34). All the sugars and the bases adopt a C2'-endo and anti conformation, respectively.
The six nucleotides are wrapped around Leu$_9$, forming a hydrophobic pocket around it (Figure 3.34-c,d). As a result, extensive hydrophobic contacts are seen for the side chain of Leu$_9$ that points its hydrophobic methyl groups towards all nucleotide sugars simultaneously as well as the A$_3$ base (Table 3.6).

The observed folded conformation of the conjugate is driven by numerous electrostatic and hydrophobic interactions between all six nucleotide and the peptide side chain guanidinium groups and/or backbone amides. The most frequently observed interactions for the peptidyl-oligonucleotide conjugate are described in Table 3.9.

**Table 3.9:** Salt bridges and hydrogen bonds for the peptidyl-oligonucleotide conjugate from the 20 refined NMR structures$^a$

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Percentage (%)</th>
<th>Distance (Å)</th>
<th>Angle (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intramolecular interactions between the conjugate peptide and oligonucleotide</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 5’-phosphate Arg10 H$^{\eta 22}$/H$^{\eta 11}$/H$^{\eta 21}$/H$^{\epsilon}$/H</td>
<td>85</td>
<td>2.75 ± 0.06</td>
<td>162 ± 9</td>
</tr>
<tr>
<td>T1 5’-phosphate Leu9 H</td>
<td>55</td>
<td>2.81 ± 0.09</td>
<td>158 ± 7</td>
</tr>
<tr>
<td>C2 5’-phosphate Arg10 H$^{\eta 12}$/H$^{\eta 21}$/H$^{\eta 11}$</td>
<td>50</td>
<td>2.77 ± 0.03</td>
<td>170 ± 9</td>
</tr>
<tr>
<td>T5 O2 Arg8 H$/H^{\eta 21}$</td>
<td>90</td>
<td>2.94 ± 0.25</td>
<td>140 ± 7</td>
</tr>
<tr>
<td>C6 5’-phosphate Arg14 H$^{\eta 11}$</td>
<td>95</td>
<td>2.76 ± 0.17</td>
<td>164 ± 3</td>
</tr>
<tr>
<td>C6 5’-phosphate Arg12 H$^{\eta 11}$/H$^{\epsilon}$</td>
<td>65</td>
<td>2.78 ± 0.06</td>
<td>158 ± 6</td>
</tr>
<tr>
<td>C6 5’-phosphate Gly15 H/H$^{N2}$</td>
<td>90</td>
<td>2.81 ± 0.03</td>
<td>156 ± 8</td>
</tr>
<tr>
<td><strong>Intramolecular hydrogen bonds within the conjugate peptide</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu7 O Leu11 H</td>
<td>60</td>
<td>2.98 ± 0.15</td>
<td>166 ± 3</td>
</tr>
<tr>
<td>Leu9 O Arg12 H</td>
<td>80</td>
<td>2.96 ± 0.23</td>
<td>142± 14</td>
</tr>
<tr>
<td>Leu9 O Arg14 H$^{\eta 12}$/H$^{\eta 22}$</td>
<td>90</td>
<td>2.84 ± 0.04</td>
<td>152± 9</td>
</tr>
<tr>
<td>Arg12 O Arg14 H$^{\eta 12}$</td>
<td>90</td>
<td>2.85 ± 0.09</td>
<td>158± 8</td>
</tr>
</tbody>
</table>

Analysis of the interactions between the conjugate peptide and oligonucleotide parts in all 20 NMR structures was performed using the cpptraj Amber module. The salt bridges and hydrogen-bonds were assigned based on distance between the peptide guanidine/amide protons and the nucleotide phosphate oxygens/bases acceptor groups and their orientation. The interactions that occur in ten structures or more and related occurrence frequency are listed.
Figure 3.36: Interactions between the peptide and oligonucleotide components of the conjugate. Views of the residue contacts in the overlay of the 20 refined NMR structures (left), single structures showing the most frequent observed salt bridges/hydrogen bonds (right; see Table 3.9).
Three positively charged side chains, Arg$_{10}$, Arg$_{12}$ and Arg$_{14}$, repeatedly form salt bridges with the oligonucleotide backbone phosphates (Table 3.9). Of these residues, Arg$_{10}$ forms a bidentate interaction with T$_1$ and C$_2$ phosphate groups simultaneously in 50% of the structures (Table 3.9, Figure 3.36-b), whereas Arg$_{12}$ and Arg$_{14}$ are in contact with C$_6$ phosphate group in the 65% and 95% of the structures, respectively (Table 3.9, Figure 3.36-c-d). Furthermore, in several conformers, the Arg$_8$ contacts the carbonyl of T$_5$ with a bifurcated hydrogen bond by its H$_{\eta 21}$ and H$_e$ atoms (Table 3.9, Figure 3.33-a). Additionally, two further salt bridges stabilise the conjugate in the folded conformation: the backbone amide of Leu$_9$ and Gly$_{15}$ are in contact with the phosphate groups of T$_1$ and C$_6$, respectively (Table 3.9, Figure 3.36-d).

In addition to the extensive hydrophobic and electrostatic interactions that provide a folded conformation of the conjugate, there is a dense network of hydrogen bonds between side chain and amide protons in the peptide residues (Table 3.9, Figure 3.37). Hydrogen bonds between Leu$_7$ CO and Leu$_{11}$ H$_N$ are repeatedly observed, which further indicates the peptide α-helical conformation in the first five residue backbones (Figure 3.37).

**Figure 3.37:** The most frequent hydrogen bonding occurring within the conjugate peptide in the 20 refined NMR structures. H$_N$-H$_N$ interactions are shown in green and the H$_N$-H$_e$ in orange.
3.7 Molecular modelling

Despite the fact that the conjugate structure in solution reported here is well defined (Table 3.8), there are variations in the frequency and patterns of salt bridges and hydrogen bonds observed in each NMR structure of the ensemble (Table 3.9). This reflects possible competition in the dynamics of local sub-states within the conjugate rather than persistent interactions. Consequently, it would be inappropriate to describe the interaction patterns by NMR alone as selecting only one of all possible conformations would represent a rather ‘static’ view of the molecular structure and limit the opportunity to evaluate the dynamic behaviour of such a flexible macromolecule.

Therefore, molecular dynamics (MD) simulations were performed, starting from the folded NMR structure in explicit solvent, in order to interpret and possibly expand the available structural information.

Four clusters were identified throughout the simulation of the conjugate, all having a folded structure with a consistent α-helical structure for the Leu7-Leu11 segment, while residues Arg12-Arg14 undergoing structural transition from β-sheet to α-helix. Representative conformers were extracted for each of the four clusters and their structural characteristics were analysed. The first observation is that that there is similarity between the first two clusters Cl1 and Cl2 (Figure 3.38-a and b, respectively), as well as between the other two clusters Cl3 and Cl4 (Figure 3.38-c and d, respectively).

The main feature distinguishing the clusters Cl1/Cl2 from Cl3/Cl4 is the position of the side chains of two arginines, Arg12 and Arg14. By analysing the frequency of salt bridges and hydrogen bonds occurrence (using cpptraj Amber module), it was observed that in Cl1 and Cl2, the Arg12 guanidinium group intermittently forms hydrogen bonds with the carbonyl oxygen of Arg8 and Leu9 or, more rarely, forms a salt bridge with the phosphate of C6 (Figure 3.39-a,b). In contrast, in Cl3 and Cl4 Arg12 binds the oligonucleotide backbone at the 3’-end forming a bidentate interaction with the phosphate of C6 and T5 (Figure 3.39-c,d). Similarly, in Cl1 and Cl2, Arg14 forms a bidentate interaction with the phosphate of C6 and the carbonyl oxygen of Leu9 and, intermittently, of Arg12, whereas in Cl3 and Cl4 Arg14 is not in contact with the oligonucleotide, instead, its side chain is outwards away from the molecule’s functional sites.
Figure 3.38: Representative conformations of the four clusters identified in the 1μs simulations: a) Cl1, b) Cl2, c) Cl3 and d) Cl4.

In addition, in Cl1 and Cl2 Arg₁₀ forms a bidentate interaction with T₁ and C₂ phosphate groups, whereas in Cl3 and Cl4 Arg₁₀ forms bifurcated bond with C₂ phosphate groups. One common feature that differs in the four clusters is the orientation of the Arg₈ side chain. In Cl1, Arg₈ is very close to T₁ phosphate backbone and rarely forms hydrogen bonds with O5’ and O4’; in Cl2, instead, Arg₈ forms a bidentate interaction with T₁ O2 and T₅ O2; in Cl3 its guanidinium group lies between the C₅ and T₆ bases and forms a salt bridge with C₆ phosphate; in Cl4, an interaction with T₅ O2 is observed instead.
Another feature in common was the location of Leu$_0$ in the hydrophobic core of the conjugate structure, also observed for the NMR structures.

The different orientations observed for the four arginines and their interactions with the conjugate oligonucleotide are illustrated in Figure 3.39.
Figure 3.39: Alternative orientations observed for Arg\textsubscript{8}, Arg\textsubscript{10}, Arg\textsubscript{12} and Arg\textsubscript{14} in the four cluster representative structures a) Cl1, b) Cl2, c) Cl3 and d) Cl4. Dashed lines indicate salt bridges and hydrogen bonds between these arginines and the conjugate oligonucleotide.
The most frequent interactions that appear in the conjugate 1 μs simulation are illustrated by the time series of arginine side chain/backbone-oligonucleotide phosphate distances in Figure 3.40. The distance fluctuations of the hydrogen bonds between amino acid residues within the conjugate peptide sequence are also presented.

**Figure 3.40**: Time series of heavy atom distances of the most frequent intra-molecular interactions in the conjugate simulation. (P)-(CZ/N) distance (d) ≤4.5 Å: red, 4.5 ≤ d ≤ 5.5 Å: green 5.5 ≤ d ≤ 6.5 Å: blue, d ≥ 6.5 Å: yellow; (O)-(N) distance (d) ≤3.5 Å: red, 3.5 ≤ d ≤ 4.5 Å: green 4.5 ≤ d ≤ 5.5 Å: blue, d ≥ 5.5 Å: yellow. The topmost three plots show the Arg_{12} and Arg_{10} side chain interactions with the oligonucleotide phosphates. The three central plots show the peptide backbone/T₁ phosphate salt bridges of Arg₈, Arg₉ and Arg₁₀ and the three bottom plots show the most frequently occurring (i,i+4) and (i,i+3) hydrogen bonds within the peptide backbone.

In the simulation Arg₁₂ often formed an additional salt bridge with T₅ phosphate, which is absent in the NMR ensemble and cannot be proved based on the NMR data (Figure 3.39, Figure 3.40). Similarly, MD predicted a fully stable interaction between Arg₈ amide and T₁ phosphate which is missing from the NMR structures. This salt bridge is presumably stabilised by a network of interactions that involve the T₁ phosphates and the neighbouring Leu₉ and Arg₁₀ backbone, as well as the continuous contacts between Arg₁₀ side chain and
$T_1$ and $C_2$ phosphates (Figure 3.40). Those interactions were also observed in the NMR structures (see Table 3.9).

The molecular dynamics simulation revealed that the Arg$_{12}$ and Arg$_{14}$ residues are flexible, which is in agreement with the fewer observed NOE signals. In particular, in the simulation, the Arg$_{12}$ and Arg$_{14}$ side chain residence time close to the 3’-phosphate backbone differs in the Cl1/Cl2 and Cl3/Cl4 clusters. Furthermore, when not involved in interactions with the oligonucleotide component, Arg$_{12}$ side chain interacts with the Arg$_8$ and Leu$_9$ carbonyl oxygens, while Arg$_{14}$ side chain moves away from the phosphate backbone and fluctuates into the solvent.

MD also revealed that the peptide backbone conformation, which shows an S-shape in the NMR ensemble structures as a result of the concomitant Arg$_{12}$/Arg$_{14}$/C$_6$ interactions, exists as shifting conformational sub-states, as observed in the four clusters. Therefore, instead of a structural conformation defined by the Arg$_{12}$/Arg$_{14}$/C$_6$ sub-state as captured in the NMR, there may also be sub-states Arg$_{12}$/C$_6$ and Arg$_{14}$/C$_6$ where either the Arg$_{12}$ or Arg$_{14}$ residues are not interacting with the oligonucleotide. This may explain the change in the conjugate peptide geometry predicted in the simulation.

Because of the conformation shift of the peptide segment Arg$_{12}$-Gly$_{15}$ during the simulation, the observed NOE distances may be an average of the actual distance between these arginine side chains and the nucleotides at the 3’-end.

To determine whether the peptide geometry extracted by the experimental NMR data is consistent with that predicted by the molecular dynamics simulation, the dihedral angles were measured for all nine residues of the four cluster conformations. From the analysis of the Ramachandran plots for each residue for the four clusters (Figure 3.41), it is readily apparent that the simulation produces two distinct secondary structures for the segment Arg$_{12}$-Arg$_{14}$.

In Cl1 and Cl2, Figure 3.41, the two segments Leu$_7$-Leu$_{11}$ and Arg$_{12}$-Arg$_{14}$ are characterised by $\alpha$-helix and $\beta$-sheet secondary structures, respectively. Instead, the segment Arg$_{12}$-Arg$_{14}$ flips from $\beta$-sheet to $\alpha$-helix character in Cl3 and Cl4.

Being part of the flexible tail at the peptide C-terminus Gly$_{15}$ is largely exposed to the solvent and its dihedral angles change over the simulation. This reflects the quite dynamic character of the peptide at this residue position.
Figure 3.41: Ramachandran plots of the distribution of the $(\phi, \psi)$ backbone dihedrals for the conjugate peptide nine residues in the four clusters. Residues Leu7 to Leu11 are in the $\alpha$-helix region in all four clusters; residues Arg12 to Arg14 lie in the $\beta$-sheet region in Cl1/Cl2 and in $\alpha$-helix region in Cl3/Cl4 (axes range from -180° to 180°). The N-terminal Gly15 is found in six different regions in each cluster.
Previous circular dichroism (CD) studies\textsuperscript{28} suggested that the free peptide adopts a random coil structure, and that the apparent ordered conformation of the peptide strongly depends on the conjugation with the oligonucleotide.

The results from the NMR and MD studies of the conjugate demonstrated that the peptide conformation is represented by $\alpha$-helix secondary structure for the segment Leu\textsubscript{7} to Leu\textsubscript{11}, and a dynamic conformation for the segment Arg\textsubscript{12}-Arg\textsubscript{14} which flips between $\alpha$-helix and $\beta$-sheet. The found ordered structure of the peptide is more likely a consequence of the positively charged arginine residues interacting with the oligonucleotide backbone phosphates.

The NMR and MD investigation provide evidence that the oligonucleotide backbone reduces the positive charge of the four arginine residues through a network of interactions and enable formation of the $\alpha$-helical conformation. The $\beta$-sheet conformation of the peptide observed in Cl1 and Cl2 is very likely induced by the transient interaction of the Arg\textsubscript{14} guanidinium group with the oligonucleotide phosphate backbone at its 3’-end.

The flexibility indicated in this study of the peptide segment Arg\textsubscript{12}-Gly\textsubscript{15} may play a role in the RNA substrate hydrolysis, presumably by enabling the peptide to sample various subtle conformations of which the appropriate ones are capable of cleaving RNA with a catalytic activity.
3.8 References


11. N. L. Mironova, D. V. Pyshnyi, E. M. Ivanova, M. A. Zenkova, H. J. Gross, V. V. Vlassov, Covalently attached oligodeoxyribonucleotides induce RNAse activity of a


Chapter 4: miRNases: Novel peptide-oligonucleotide bioconjugates that silence miR-21 in lymphosarcoma cells

4.0. DECLARATION

This chapter consists of one published paper:


This paper has been reproduced in an unchanged format with the Supplementary Material, except for minor adjustments to incorporate them into this thesis.

I carried out the molecular modelling study and analysed the data.
miRNases: novel peptide-oligonucleotide bioconjugates that silence miR-21 in lymphosarcoma cells

Olga A. Patutina1, Elena V. Bichenkova2, Svetlana K. Miroshnichenko1, Nadezhda L. Mironova1, Linda T. Trivoluzzi2, Kepa K. Burusco2, Richard A. Bryce2, Valentin V. Vlassov1, Marina A. Zenkova1*

1Laboratory of Nucleic Acids Biochemistry, Institute of Chemical Biology and Fundamental Medicine SB RAS, Lavrentiev ave., 8, Novosibirsk, 630090, Russia
2School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Oxford Road, Manchester, UK, M13 9PT.

4.1 Abstract

MicroRNAs (miRNAs) are active regulators in malignant growth and constitute potential targets for anticancer therapy. Consequently, considerable effort has focused on identifying effective ways to modulate aberrant miRNA expression. Here we introduce and assess a novel type of chemically engineered biomaterial capable of cleaving specific miRNA sequences, i.e. miRNA-specific artificial ribonucleases (hereafter ‘miRNase’). The miRNase template presented here consists of the catalytic peptide acetyl-[(LeuArg)2Gly]2 covalently attached to a miRNA-targeting oligonucleotide, which can be linear or hairpin. The peptide C-terminus is conjugated to an aminohexyl linker located at either the 3’- or 5’-end of the oligonucleotide. The cleavage efficacy, structural aspects of cleavage and biological relevance of a set of these designed miRNases was assayed with respect to highly oncogenic miR-21. Several miRNases demonstrated effective site-selective cleavage of miR-21 exclusively at G-X bonds. One of the most efficient miRNase was shown to specifically inhibit miR-21 in lymphosarcoma cells and lead to a reduction in their proliferative activity. This report provides the first experimental evidence that metallo-independent peptide-oligonucleotide chemical ribonucleases are able to effectively and selectively down-regulate oncogenic miRNA in tumour cells, thus suggesting their
potential in development of novel therapeutics aimed at overcoming overexpression of disease-related miRNAs.

4.2 Introduction

In the last decade, non-coding RNAs and in particular miRNAs have been the focus of considerable research due to their association with a range of pathological conditions, from cancer to neurodegenerative, cardiovascular and autoimmune diseases\textsuperscript{1-3}. MiRNAs represent a new class of regulatory molecule; they are 18-25 nucleotides in length and, through binding to specific mRNAs, are capable of post-transcriptional repression of gene expression, either by the RNA interference pathway or by translational arrest\textsuperscript{4-6}. A large amount of evidence suggests that miRNAs exert control over fundamental physiological processes, both within the cell and at the level of the whole body\textsuperscript{7, 8}. Disturbance in normal miRNA expression results in changes in the activity of target genes and is often associated with the initiation and progression of a wide spectrum of diseases, including oncology\textsuperscript{9-12}. Tumor development is accompanied by an excess or deficiency of certain miRNAs compared to normal tissues. Increased expression of oncogenic miRNAs contributes to the development of neoplasia by suppression of tumor-suppressor genes, whereas a significant lack of tumor-suppressor miRNAs results in the overexpression of oncogenes\textsuperscript{13, 14}. Modulation of activity of tumor-associated miRNAs is therefore of great scientific, biomedical and clinical interest. To date, a number of approaches have been developed for microRNA-based and miRNA-targeted therapies\textsuperscript{6, 15}. Restoring the level of deficient miRNA can be achieved by using miRNA mimics, representing synthetically prepared miRNAs\textsuperscript{16, 1}; or viral constructs, encoding for miRNAs\textsuperscript{18, 19}. Suppression of miRNA activity can be achieved using small-molecule inhibitors, acting at the transcriptional level\textsuperscript{20}; via miRNA sponges, representing transcripts that contain multiple tandem-binding sites adsorbing deleterious miRNAs\textsuperscript{21, 22}; by miR-mask oligonucleotides or target protectors, that are fully complementary to predicted miRNA binding sites in the 3′-UTR of the target mRNA\textsuperscript{23-25}; and using antisense oligonucleotides, complementary to the target miRNA and inducing either its degradation or steric blockage\textsuperscript{26, 27}. Numerous positive results have been achieved using strategies based on inhibition of oncogenic miRNAs. Indeed, the approaches based on miRNA suppression using miRNA-sponges, representing molecular traps, were shown to restore the activity of tumour-suppressor genes. For example, miRNA sponges designed for modulation of miR-10b,
miR-21, miR-155 and miR-221/222 mediate an impact on the activity of many protein targets, such as HOXD-10, PDCD4, Smad4, SRC3, Bcl-2 Bim, FOXO3a, PTEN and RhoA. Thus they provide inhibition of proliferation, activation of apoptosis and increase in the sensitivity of tumour cells to chemotherapy\textsuperscript{28-30}. Suppression of oncogenic activity of miR-522 in a non-small cell lung cancer model was successfully achieved by the use of miRNA masking\textsuperscript{31}. Antisense technology also has proven successful in blocking miRNA activity: synthetic oligonucleotides of different chemistries targeted to known oncogenic miRNAs, such as miR-17, miR-21, miR-155, and miR-221/222 promoted not only induction of apoptosis and inhibition of proliferation, but also tumour regression and metastasis suppression \textit{in vivo}\textsuperscript{32-38}. The drugs Miravirsen and Regulus RG-101, aimed at suppressing miRNA-122 for the treatment of hepatitis C, have successfully progressed into clinical trials, and thus provide grounds to believe that, in the near future, effective antisense-based anti-miRNA therapies will be developed to combat oncopathology\textsuperscript{39, 40}. Effective downregulation of miRNA levels in cells can be achieved by its selective, irreversible cleavage using agents that are capable of recognizing particular miRNA sequences. A direct approach to create such an artificial site-selective ribonuclease can be based on design of conjugates comprising of (i) antisense oligonucleotides (asON), which can form a complementary complex with a specific miRNA, and (ii) chemical moieties able to cleave phosphodiester bonds. These include metal complexes, imidazoles or cleaving peptides\textsuperscript{41}. Over the last couple of decades, some progress has been achieved in the field of designing site-selective artificial ribonucleases\textsuperscript{42-54}. It was shown that short peptides, containing either alternating leucine and arginine residues or imidazole-based catalytic groups, conjugated to antisense oligonucleotides targeting tRNA, were able to hydrolyze linkages adjacent to an oligonucleotide-binding site without involvement of exogenous species such as metal ions, enzymes or cofactors (\textit{e.g.}, RISC, RNase H)\textsuperscript{42, 44, 45, 47, 50, 51}. Effective cleavage of complementary substrates was also demonstrated for tris(2-aminobenzimidazole) ribonuclease conjugated to PNA oligomers\textsuperscript{49}. It is important to emphasize though that none of the above developments have been demonstrated against clinically relevant RNA sequences, and most of the studies in this area have been carried out so far using either short, linear RNA sequences or model RNAs (\textit{e.g.} tRNAs).

Despite the fact that considerable success has been achieved in the area of development of site-selective, metal-free artificial ribonucleases, to date no studies have been reported demonstrating a successful downregulation of clinically significant miRNAs by such metallo-independent sequence-specific catalytic bioconjugates in eukaryotic cells. The key
challenge of our research therefore was to assess whether such chemically engineered peptide-oligonucleotide conjugates (POCs) could potentially induce detectable downregulation of specific oncogenic miRNA in tumour cells, which is essential factor for future application in anticancer therapy. Here we report the development of miRNA-specific peptide-oligonucleotide conjugate (‘miRNases’) against highly oncogenic miR-21, and present the first experimental evidence of efficient, site-selective cleavage of this miRNA exclusively at G-X linkages by the most successful structural variants. Furthermore, we demonstrate here that one of the most efficient conjugates is capable of inducing specific inhibition of miR-21 in lymphosarcoma cells. This leads to reactivation of tumour-suppressor protein PDCD4, the direct target of mir-21, and subsequent reduction in proliferative activity of lymphosarcoma cells. To obtain structural insights into specific interactions between miRNases and miR-21 and to guide the future design of such conjugates, we carried out 1 μs molecular dynamics simulations of the hybridized complex between 5'-h-6/14 conjugate and miR-21.

4.3 Material and Methods

4.3.1 Instrumentation

The details of chemicals, reagents and facilities used in this research have already been provided in our previous publications. The software Topspin 3.2 was used to analyze the NMR data obtained. After electrophoresis the gels were analyzed using Molecular Imager FX (Bio-Rad, USA). PCR amplification was carried out using Bio-Rad iQ5 (Bio-Rad, USA).

4.3.2 Oligonucleotides

Oligodeoxyribonucleotides with an aminohexyl linker attached to either the 5′- or 3′-terminal phosphate of the oligonucleotide sequence were synthesized in the Laboratory of Medicinal Chemistry, ICBFM, Russia, by the standard phosphoramidite protocol on an ASM-800 synthesizer (Biosset, Russia) using solid support, nucleoside phosphoramidites and chemical phosphorylation reagent from Glenn Research (USA). Oligonucleotides were
isolated by consecutive ion-exchange (Polysil SA-500 columns, Russia) and reverse-phase HPLC (LiChrosorb RP-18 columns, Merck, Germany) according to standard protocols.

4.3.3 Peptide synthesis and purification

Catalytic peptide Acetyl-(LeuArg)_2-Gly-(LeuArg)_2-Gly-COOH was synthesized by manual solid-phase methodology on Fmoc-Gly-Wang resin using the Fmoc/tBu strategy as described in 50. Following completion of the peptide sequence, the N-terminus was acetylated by shaking with acetic anhydride (10 mmol) and DIPEA (10 mmol) in DMF (10 mL) for 30 min, followed by resin washing with DMF (2 × 10 mL), DCM (2 × 10 mL and DMF (2 × 10 mL). After acetylation of the sequence, the peptide was cleaved from the resin using the same protocol reported in 50. Crude lyophilized peptide was solubilized in 30% acetic acid and purified using RP-HPLC as described earlier 50.

4.3.3.1 Acetyl-(LeuArg)_2-Gly-(LeuArg)_2-Gly-COOH: Fractions collected at 33 min were combined and lyophilized to yield the TFA salt of the peptide as a fluffy white material. ESI-MS: m/z = 626.4 for [M+H+H]^2+ (MW = 1250.80 calcd. for [C_{54}H_{102}N_{22}O_{12}]). ^1H NMR (D_{2}O with TSP (0.1 μM), 400 MHz): δ 0.83-0.97 (m, 24H, Leu-Hβ), 1.49-1.94 (m, 28H, 8 × Arg-Hβ, 8 × Arg-Hγ, 8 × Leu-Hβ, 4 × Leu-Hγ), 2.04 (s, 3H, CH₃), 3.21 (m, 8H, Arg-Hδ), 3.82 - 4.41 (m, 12H, 2 x Gly-H₂, 8 x Leu/Arg-Hα) (see also Figure S1).

4.3.4 Synthesis of the conjugates

Oligonucleotides containing an aminohexyl linker attached to either the 5′- or 3′-end of the oligonucleotide, (50 nmol) in H₂O (100 μL) were converted into cetyltrimethylammonium salt soluble in DMSO 50, 51. Acetyl-(LeuArg)_2-Gly-(LeuArg)_2-Gly-COOH (2.3 μmol,) and 4-dimethylaminopyridine (3.5 μmol) were dissolved in a minimal volume of anhydrous DMSO (≈20 μL) before N,N'-dicyclohexylcarbodiimide (3.45 μmol) was added and vortexed. The resulting peptide solution was added directly to oligonucleotide solution and left at 60°C for 12 h. The reaction product (along with the unreached starting oligonucleotide) was precipitated with 4% LiClO₄ in acetone using similar protocol described in 50, 51 followed RP-HPLC purification of the corresponding peptide-oligonucleotide conjugate 50.
4.3.5 Conjugate purification

Crude conjugates were dissolved in water and purified using RP-HPLC (2.0 mL/min flow rate, eluent A – 0.05 M LiClO₄, eluent B – 0.05 M LiClO₄ in AcCN; 100% A for 5 min, 0% B to 50% B in 35 min). Fractions containing conjugate were collected, combined and lyophilized. The excess of LiClO₄ was removed by precipitation in 4% LiClO₄ in acetone (w/v) as reported earlier in 50, 51.

4.3.6 Characterisation of conjugates

The identity and purity of all peptide-oligonucleotide conjugates were confirmed by RP-HPLC, ¹H-NMR spectroscopy and mass spectrometry using matrix-assisted laser desorption ionisation (MALDI) spectroscopy (see below and Figure S1). Conjugates were identified as mono- or di-sodium adducts with experimental masses in agreement with theoretical calculations.

4.3.6.1 Conjugate 5’-16

MALDI-MS: m/z = 6192 [M+Na]⁺ (MW = 6170 calcd for 5’-16). ¹H NMR (D₂O, 400 MHz): δ 0.62-0.93 (m, 24H, Leu-δ), 1.11-1.47 (m, 12H, 6×CH₂ (aminohexyl linker)), 1.49-2.18 (m, 43H, 5×CH₃ of 5×dT, 8×Arg-β, 8×Arg-γ, 8×Leu-β, 4×Leu-γ), 2.03 (s, 3H, CH₃), 2.31-2.92 (m, 32H, 16×H₂’ and 16×H₂” sugar ring protons), 3.05-3.14 (m, 8H, 8×Arg-δ), 3.72-4.55 (m, 60H, 16×H₄’, 16×H₅’ and 16×H₅” sugar ring protons, 4×Gly-H, 8×Leu/Arg-H), 5.52-6.53 (m, 19H, 16×H₁’ sugar ring protons, 3×H₅ of dC), 7.25-8.28 (s or d, 21H, 21×Ar-H from 3×dG(H₈), 5×dA(H₈), 5×dA(H₂), 3×dC(H₆) and 5×dT(H₆)). The H₃’ sugar ring proton regions (4.56-5.2 ppm) were not analysed due to suppression of residual water signal at 4.78ppm.

4.3.6.2 Conjugate 5’-h-6/14

MALDI-MS: m/z = 10554 [M+H]⁺ (MW = 10553 calcd. for 5’-h-6/14). ¹H NMR (D₂O, 400 MHz): δ 0.62-0.93 (m, 24H, Leu-δ), 1.11-1.47 (m, 12H, 6×CH₂ (aminohexyl linker)), 1.49-2.18 (m, 46H, 6×CH₃ of 6×dT, 8×Arg-β, 8×Arg-γ, 8×Leu-β, 4×Leu-γ), 2.03 (s, 3H, CH₃), 2.31-2.92 (m, 60H, 30×H₂’ and 30×H₂” sugar ring protons), 3.05-3.14 (m, 8H, 8×Arg-δ), 3.72-4.55 (m, 102H, 30×H₄’, 30×H₅’ and 30×H₅” sugar ring protons, 4×Gly-
H, 8 × Leu/Arg-Hα), 5.52-6.53 (m, 36H, 30×H1’ sugar ring protons, 6×H5 of dC), 7.22-8.31 (s or d, 40H, 40×Ar-H from 8×dG(H8), 10×dA(H8), 10×dA(H2), 6×dC(H6) and 6×dT(H6)). The H3’ sugar ring proton regions (4.56-5.2 ppm) were not analysed due to suppression of residual water signal at 4.78ppm.

4.3.6.3 Conjugate 5’-h-9/14
MALDI-MS: m/z = 12424 [M+Na]+ (MW = 12404 calcd. for 5’-h-9/14). 1H NMR (D2O, 400 MHz): δ 0.62-0.93 (m, 24H, Leu-Hδ), 1.11-1.47 (m, 12H, 6×CH2 (aminohexyl linker)), 1.49-2.18 (m, 52H, 8×CH3 of 8×dT, 8×Arg-Hβ, 8×Arg-Hγ, 8×Leu-Hβ, 4×Leu-Hγ), 2.03 (s, 3H, CH3), 2.31-2.92 (m, 72H, 36×H2’ and 36×H2” sugar ring protons), 3.05-3.14 (m, 8H, 8×Arg-Hδ), 3.72-4.55 (m, 120H, 36×H4’, 36×H5’ and 36×H5” sugar ring protons, 4 × Gly-H, 8 × Leu/Arg-Hδ), 5.52-6.53 (m, 43H, 36×H1’ sugar ring protons, 7×H5 of dC), 7.22-8.31 (s or d, 48H, 48×Ar-H from 9×dG(H8), 12×dA(H8), 12×dA(H2), 7×dC(H6) and 8×dT(H6)). The H3’ sugar ring proton regions (4.56-5.2 ppm) were not analysed due to suppression of residual water signal at 4.78ppm.

4.3.6.4 Conjugate 5’-h-6/16
MALDI-MS: m/z = 11165 [M+Na]+ (MW = 11145 calcd. for 5’-h-6/16). 1H NMR (D2O, 400 MHz): δ 0.62-0.93 (m, 24H, Leu-Hδ), 1.11-1.47 (m, 12H, 6×CH2 (aminohexyl linker)), 1.49-2.18 (m, 49H, 7×CH3 of 7×dT, 8×Arg-Hβ, 8×Arg-Hγ, 8×Leu-Hβ, 4×Leu-Hγ), 2.03 (s, 3H, CH3), 2.31-2.92 (m, 64H, 32×H2’ and 32×H2” sugar ring protons), 3.05-3.14 (m, 8H, 8×Arg-Hδ), 3.72-4.55 (m, 108H, 32×H4’, 32×H5’ and 32×H5” sugar ring protons, 4 × Gly-H, 8 × Leu/Arg-Hδ), 5.52-6.53 (m, 39H, 32×H1’ sugar ring protons, 7×H5 of dC), 7.22-8.31 (s or d, 42H, 42×Ar-H from 8×dG(H8), 10×dA(H8), 10×dA(H2), 7×dC(H6) and 7×dT(H6)). The H3’ sugar ring proton regions (4.56-5.2 ppm) were not analysed due to suppression of residual water signal at 4.78ppm.

4.3.6.5 Conjugate 5’-h-9/16
MALDI-MS: m/z = 12999 [M+H]+ (MW = 12997 calcd. for 5’-h-9/16). 1H NMR (D2O, 400 MHz): δ 0.62-0.93 (m, 24H, Leu-Hδ), 1.11-1.47 (m, 12H, 6×CH2 (aminohexyl linker)), 1.49-2.18 (m, 55H, 9×CH3 of 9×dT, 8×Arg-Hβ, 8×Arg-Hγ, 8×Leu-Hβ, 4×Leu-Hγ), 2.03 (s,
3H, CH$_3$), 2.31-2.92 (m, 76H, 38×H2’ and 38×H2’’ sugar ring protons), 3.05-3.14 (m, 8H, 8×Arg-H$^\beta$), 3.72-4.55 (m, 120H, 36×H4’, 36×H5’ and 36×H5’’ sugar ring protons, 4×Gly-H, 8×Leu/Arg-H$^\beta$), 5.52-6.53 (m, 46H, 38×H1’ sugar ring protons, 8×H5 of dC), 7.22-8.31 (s or d, 50H, 50×Ar-H from 9×dG(H8), 12×dA(H8), 12×dA(H2), 8×dC(H6) and 9×dT(H6)). The H3’ sugar ring proton regions (4.56-5.2 ppm) were not analysed due to suppression of residual water signal at 4.78 ppm.

4.3.6.6 Conjugate 5’-luc-h-9/14

MALDI-MS: $m/z$ = 12429 [M+Na]$^+$ (MW = 12399 calcd. for 5’-luc-h-9/14). $^1$H NMR (D$_2$O, 400 MHz): \(\delta\) 0.62-0.93 (m, 24H, Leu-H$^\delta$), 1.11-1.47 (m, 12H, 6×CH$_2$ (aminohexyl linker)), 1.49-2.18 (m, 46H, 6×CH$_3$ of 6×dT, 8×Arg-H$^\beta$, 8×Arg-H$^\gamma$, 8×Leu-H$^\beta$, 4×Leu-H$^\gamma$), 2.03 (s, 3H, CH$_3$), 2.31-2.92 (m, 72H, 36×H2’ and 36×H2’’ sugar ring protons), 3.05-3.14 (m, 8H, 8×Arg-H$^\beta$), 3.72-4.55 (m, 120H, 36×H4’, 36×H5’ and 36×H5’’ sugar ring protons, 4×Gly-H, 8×Leu/Arg-H$^\beta$), 5.52-6.53 (m, 44H, 36×H1’ sugar ring protons, 8×H5 of dC), 7.22-8.31 (s or d, 49H, 48×Ar-H from 9×dG(H8), 13×dA(H8), 13×dA(H2), 8×dC(H6) and 6×dT(H6)). The H3’ sugar ring proton regions (4.56-5.2 ppm) were not analysed due to suppression of residual water signal at 4.78 ppm.

4.3.7 5’-RNA labelling

MiR-21 5’-UAGCUUAUCAGACUGAUUGA-3’ was synthesized by Dr. Maria I. Meschaninova (ICBFM SB RAS). 5’-end labelling using $[^{32}\text{P}]$ATP and T4 polynucleotide kinase (Thermo Scientific, USA) and isolation of $[^{32}\text{P}]-\text{miR-21}$ was carried out according to procedure previously described in $^{55,77}$.

4.3.8 Gel-retardation assay

The reaction mixture (10 µl) containing 50 mM Tris-HCl, pH 7.0, 0.2 M KCl and 1 mM EDTA, 6000 cpm (Cherenkov’s counting) of $[^{32}\text{P}]-\text{miR-21}$, 1 µM unlabeled miR-21, antisense oligonucleotide (asON) at a concentration ranging from 0.5 to 100 µM or conjugate at a concentration ranging from 1 to 50 µM, was incubated at 37°C for 1 h and quenched by adding of a loading buffer (20% ficoll, 0.025% bromophenol blue, 0.025% xylene cyanol). The samples were loaded onto the running gel immediately after
The quenching of the reaction with an interval of 2 min. Formation of the complex miR-21/asON or miR-21/conjugate was analyzed by electrophoresis in 12% native PAAG at 4°C. The gels were analyzed using Molecular Imager FX. The extent of binding of oligonucleotide to miR-21 was determined as a ratio of radioactivity measured in the complex to the total radioactivity applied onto the gel lane.

4.3.9 Ribonuclease activity assay

The reaction mixture (5 µl) contained 8000 cpm (Cherenkov’s counting) of [³²P]-miR-21, 1 µM unlabeled miR-21 and one of the conjugates at a concentration ranging from 1 to 50 µM, 50 mM Tris-HCl, pH 7.0, 0.2 M KCl, 1 mM EDTA. The mixture was incubated at 37°C (for various times) and quenched by precipitation of RNA with 2% LiClO₄ in acetone (50 µl). RNA was collected by centrifugation and dissolved in loading buffer (8 M urea, 0.025% bromophenol blue, 0.025% xylene cyanol). RNA cleavage products were analyzed in 18% PAAG/8M urea using TBE (100 mM Tris-borate, pH 8.3, 2 mM EDTA) as running buffer. To identify cleavage sites, an imidazole and T1-ladders produced by partial RNA cleavage with 2 M imidazole buffer (pH 7.0) and with RNase T1, respectively, were run in parallel. To obtain quantitative data, gels were dried and analyzed using Molecular Imager FX. The total extent of RNA cleavage and the extent of RNA cleavage at each individual site were determined Quantity One software.

4.3.10 Molecular Modelling

Computational studies were performed with AMBER 15 (AMBER 14 and AMBERTOOLS 15) and MOE 2014.09 molecular modelling packages. The initial 3D conformation of the complex between 5'-h-6/14 conjugate and miR-21 was generated via assembly of distinct components: the RNA-DNA hybrid double helix was modelled in the type-A form using the make-na online server (http://structure.usc.edu/make-na/server.htm). The hairpin structure was modelled from a NMR structure (PDB code 2K71 [60]). An initial conformation of the peptide component was generated from homology using the SWISS-MODEL online server. Finally, the linker moiety joining the peptide to the oligonucleotide was modelled in and the components of the complex assembled via xLEaP. For molecular dynamics (MD) simulations, parameters were assigned to the complex from the ff14SBonlysc and GAFF force fields. The complex was solvated
using the generalised Born implicit solvation model (igb=5) with a Debye-Huckel salt concentration of 0.2 M. MD simulations employed SHAKE on all atoms involving hydrogen a timestep of 2 fs and a cut-off of 999 Å for non-bonded interactions. Temperature was controlled using Langevin dynamics with a collision frequency of 2 ps. Subsequent to stepwise relaxation of solute constraints, the solvated complex was equilibrated for 10 ns. A production run of 1 μs was then acquired. Coordinates and energies were saved for analysis every ps. Simulations were performed using the pmemd.CUDA module of the AMBER 14 software package.

4.3.11 Transfection of tumour cells with conjugates

Mouse lymphosarcoma cells RLS were placed in a serum-free Iscove’s Modified Dulbecco’s Medium (IMDM) into 24-well plates immediately before transfection. The cells were incubated at 37°C in a humidified atmosphere with 5% CO₂ for 4 h with 1 µM asON 5’-h-9/14, or conjugate 5’-h-9/14, or control conjugate luc-h-9/14 precomplexed with Lipofectamine 2000 (Invitrogen, USA) in an Opti-MEM medium. Then the medium was replaced with a culture medium containing 10% fetal bovine serum (FBS) and 1% antibiotic antimycotic solution (10 000 µg/ml streptomycin, 10 000 IU/ml penicillin, and 25 µg/ml amphotericin), and the cells were cultivated for 24 – 72 h under the same conditions. At time points 24, 48 and 72 h, total RNA was isolated from the cells using TRIzol Reagent (Invitrogen, USA) according to the manufacturer’s protocol. At the time point 48 h, cell lysates were prepared using RIPA buffer (Thermo Scientific, USA).

4.3.12 qPCR

Expression of miRNA in RLS cells was analyzed using stem-loop PCR technology. cDNA synthesis was carried out using SuperScript III reverse transcriptase (SSIII RT, Invitrogen, USA) as previously described. The RT and PCR primers used in the study are presented in Table S1, “Supplementary data”. PCR amplification was carried out in a total volume of 20 µl using Maxima Hot Start Taq DNA polymerase (Thermo Scientific, USA), 1×PCR Buffer, 1.5 mM MgCl₂, 0.2 mM dNTPs, 1×EvaGreen (Biotium, Hayward, USA), and 0.2 mM of PCR sense and antisense primers. The reaction was performed with initial preheating at 94°C for 4 min and 40 cycles of 94°C for 40 s, 60°C for 30 s, 72°C for 30 s, followed by a melting point determination. The obtained PCR data were analyzed by
standard Bio-Rad iQ5 v.2.0 software. For each sample, the threshold cycle (Ct) was determined. Quantitative assessment of the level of transcripts representation and relative miRNA expression was performed by comparing the Ct values for miRNA and references U6 and Rpl30.

4.3.13 Western Blot

Cell lysates were separated in 12.5 % SDS-PAGE and transferred to a PVDF membrane using a semi-dry transfer. The membranes were blocked for 18 h in 3% nonfat dried milk in TBST (20 mM Tris-HCl, pH 7.6, 137 mM NaCl, 0.1 % Tween), and incubated with primary antibodies against PDCD4 (ab79405, Abcam, UK, 1:3000) and reference protein GAPDH (ab9485, Abcam, UK, 1:1000) at room temperature for 1 h. After three washes with TBST, membranes were incubated with secondary HRP-conjugated goat anti-rabbit antibodies (ab6721, Abcam, UK) at room temperature for 1 h. After three washes with TBST, proteins were detected using Chemiluminescent Reagent (ab79907, Abcam, UK).

4.3.14 MTT-test and doubling time determination

The influence of oligonucleotide 5'-h-9/14 and conjugate 5'-h-9/14 on the proliferation activity of RLS₄₀ cells was tested by MTT assay and evaluation of the cell doubling time. After transfection with the conjugates, RLS₄₀ cells were seeded into 96-well plates and incubated for 72 h under standard conditions at 37°C in a humidified atmosphere with 5% CO₂. MTT assay was performed as described in. To evaluate the cell doubling time, at 4 h and 72 h after transfection the number of cells was counted with a cytometer. The doubling time was determined by the calculation of the number of doubling according to the equation: n = 3.32 × log(x1/x0), where n is the number of doubling, x0 is the number of cells at the starting point, x1 is the number of cells at 72 h after transfection.

4.3.15 Statistics

The data obtained were statistically processed using one-way ANOVA and post-hoc Tukey test (p ≤0.05). The statistics package STATISTICA version 10.0 was used for this analysis.

4.4 Results
4.4.1 Design of miR-21 targeted oligonucleotides and peptide-oligonucleotide conjugates

Among the many miRNAs already identified as regulators of neoplastic transformation and tumorigenesis, miR-21 has emerged as a key miRNA which is dysregulated and overexpressed in major types of tumour and immortalized eukaryotic cell lines. Indeed, miR-21 has been found to be implicated in almost every cancer network\(^{71-74}\). Thus miR-21 represents an attractive target for anticancer therapy, where its inhibition could greatly reduce the malignant potential of tumour cells. For this reason, we target miR-21 using POCs to perform selective silencing of this miRNA. Moreover, our target *hsa*-miR-21 of *Homo sapiens* (denoted from this point as miR-21) has 100% homology with the sequence of *mmu*-miR-21 of *Mus musculus*. This provides the option to carry out experiments in either human or mouse tumour models.

Taking into account that miR-21 is 22 nt in length, the limiting factor for anti-miRNA POC design is the length of antisense oligonucleotide to form (i) a stable duplex with the target and (ii) to have potential cleavage sites in miR-21. In this respect, the design of the oligonucleotide part, to ensure selective and effective binding with the miRNA-target molecule, is an important step in creating anti-miRNA POCs.

As a guiding component in POCs, we explored antisense oligonucleotides of different length and structure, *i.e.* linear antisense oligonucleotides and hairpin oligonucleotides designed on the principle of primer generation used in stem-loop PCR\(^{67}\). The hairpin oligonucleotides consist of antisense sequences ranging from 10 to 16 nucleotides; hairpins containing a five-nucleotide loop flanked by two GC pairs\(^{75}\); and an additional stem of 4 or 7 nucleotides, so that the total length of the stem was 6 or 9 bp (Figure 4.1). Upon binding with miRNA, additional stabilization of the complex is achieved due to efficient stacking interactions. The complete list of oligonucleotides, used for miR-21 targeting, is given in Table 1.

The miRNA-specific conjugates were constructed based on our previous design strategies for the ‘single’ and ‘dual’ peptide-oligonucleotide conjugates\(^{50, 51}\), which demonstrated efficient cleavage of the model RNA (*i.e.* yeast tRNA\(^{\text{phe}}\)). Such conjugates incorporated unmodified oligodeoxyribonucleotide(s) complementary to this RNA and a catalytic peptide containing [Leu-Arg]\(_n\) building blocks, which act as a cleaving domain.
Figure 4.1: The structure of anti-miR-21 peptide-oligonucleotide conjugates (POCs). A. Schematic representation of a general structure of anti-miR-21 POCs. Labels ‘10-16 n’, ‘16 n’ or ‘14-16 n’ indicate the length ranges of the antisense ‘recognition’ sequence used in different conjugates. Label ‘6-9 n’ indicates the length range of the stem. B-C. The chemical structure of the peptide (pep) acetyl-(LeuArg)_{2}-Gly-(LeuArg)_{2}-Gly-COOH. The peptide was conjugated via the carboxylic group of the C-terminus to the aminohexyl linker, which was attached to either the 5′- or 3′-terminal phosphate of the antisense oligonucleotide (asON) (B and C, respectively). Amino acids are designated by numbers which are used for the interpretation of the data from molecular dynamics simulation.
Table 4.1: Oligodeoxyribonucleotides used in the study: sequences and nomenclature.

<table>
<thead>
<tr>
<th>Name</th>
<th>Oligodeoxyribonucleotide sequence (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3'-16</td>
<td>TCAACATCAGTCTGAT</td>
</tr>
<tr>
<td>5'-14</td>
<td>TCAGTCTGATAAGC</td>
</tr>
<tr>
<td>5'-16</td>
<td>TCAGTCTGATAAGCTA</td>
</tr>
<tr>
<td>5'-h-6/10</td>
<td>TGATAAGCTAGTCAGCGAAAGCTGAC</td>
</tr>
<tr>
<td>5'-h-9/10</td>
<td>TGATAAGCTACAAGTCAGCGAAAGCTGACTTG</td>
</tr>
<tr>
<td>5'-h-6/12</td>
<td>TCTGATAAGCTAGTCAGCGAAAGCTGAC</td>
</tr>
<tr>
<td>5'-h-9/12</td>
<td>TCTGATAAGCTACAAGTCAGCGAAAGCTGACTTG</td>
</tr>
<tr>
<td>5'-h-6/14</td>
<td>AGTCTGATAAGCTAGTCAGCGAAAGCTGAC</td>
</tr>
<tr>
<td>5'-h-9/14</td>
<td>AGTCTGATAAGCTACAAGTCAGCGAAAGCTGACTTG</td>
</tr>
<tr>
<td>5'-h-6/16</td>
<td>TCAGTCTGATAAGCTAGTCAGCGAAAGCTGAC</td>
</tr>
<tr>
<td>5'-h-9/16</td>
<td>TCAGTCTGATAAGCTACAAGTCAGCGAAAGCTGACTTG</td>
</tr>
<tr>
<td>3'-h-6/16</td>
<td>GTCAGCGAAAGCTGACTCAACATCAGTCTGAT</td>
</tr>
<tr>
<td>3'-h-9/16</td>
<td>CAAGTCAGCGAAAGCTGACTTGTCACACATCAGTCTGAT</td>
</tr>
<tr>
<td>3'-h-9/10</td>
<td>CAAGTCAGCGAAAGCTGACTTGTCACACATCAG</td>
</tr>
<tr>
<td>5'-luc-h-9/14</td>
<td>CGATAAAATAACGCAGCAGTCAACATCAGCGAAAGCTGACTTG</td>
</tr>
</tbody>
</table>

Regular font – sequence complementary to miR-21; underlined font – a hairpin. Oligonucleotides were designated as follows: 3’- or 5’ indicates the end of oligonucleotide used further for conjugation with a peptide; h indicates the presence of a hairpin in the structure; 6 or 9 indicates the length of the stem; and the labels from 10 to 16 indicate the length of the antisense sequence complementary to miR-21. In the text the names of oligonucleotides are shown in italic and synthesized conjugates have the same nomenclature as corresponding oligonucleotides, but shown in bold.

Since the incorporation of an extra glycine residue in the middle of the peptide chain consistently showed the highest level of ribonuclease activity, we selected the corresponding peptide -[(LeuArg)2]2- to generate these novel miRNA-specific conjugates. Based on the experimental evidence that any considerable changes in the conjugate structure (i.e. chemical modifications of the oligonucleotide part) may lead to a significant reduction of the cleavage activity, in this study we decided to preserve the ‘natural’ sugar-phosphate backbone of the oligonucleotide recognition moieties. Our previous studies of ‘dual’ conjugates have illuminated an importance of the conformational flexibility on their cleavage efficiency. Therefore, in this research we incorporated a flexible aminohexyl linker between the oligonucleotide and peptide to
enhance overall conformational freedom for the catalytic moiety. To achieve this, the catalytic peptide was attached via its C-terminus to the amine group of the aminohexyl linker located either at the 3’- or 5’- end of the oligonucleotide (Figure 4.1).

In order to avoid undesirable peptide self-conjugation and/or cyclization during the amide coupling reaction, the peptide was protected at the N-termini by an acetyl group to generate Acetyl-(LeuArg)2-Gly-(LeuArg)2-Gly-COOH sequence. The general structure of anti-miR-21 POCs is presented in Fig. 1 showing this new type of conjugation, which has never been used before for synthesis of ‘single’ peptide-oligonucleotide conjugates.

Two types of oligonucleotides were considered for conjugate design: (1) 3’-aminohexyl oligonucleotides, targeting the 3’-end of miR-21; these are intended to be conjugated with the peptide at the 3’-end, in order to cleave linkages at the 5’-end of miRNA; and (2) 5’-aminohexyl oligonucleotides, targeting the 5’-end of miR-21; these are intended to be conjugated with the peptide at the 5’-end, in order to perform cleavage of linkages at the 3’-end of miRNA.

### 4.4.2 Selection of oligonucleotides for POCs synthesis

To find the optimal addressing component in the POC structure, screening of oligodeoxyribonucleotides was performed to evaluate the efficiency of binding to the target. The ability of the oligonucleotides to bind with 5’-[32P]-miR-21 (denoted after as *miR-21) was analyzed at 37°C by gel-retardation assay (Figure 4.2). Binding efficacy of linear oligonucleotides with miR-21 at equimolar concentrations strongly depends on the length of antisense part: we observed an increase in hybridization extent from 60% for 5’-14 to 98% for 3’-16 and 5’-16. Hairpin oligonucleotides with the length of complementary sequence of 10 – 12 nucleotides exhibit poor binding with RNA. With the increase in length of complementary sequence to 14 – 16 nucleotides, a significant increase in binding extent is observed, up to 97 – 100%. We note that duplex stabilization effect of the hairpin is achieved for oligonucleotides with complementary sequence of 14 nucleotides: an increase in binding efficacy from 57% for 5’-14 to about 97% for 5’-h-6/14 and 5’-h-9/14 was observed (Figure 4.2). Nevertheless, elongation of the stem from 6 to 9 bp does not lead to significant increase in binding efficacy. In the case of oligonucleotides with complementary sequence of 16 nucleotides, the duplex stabilization effect of the hairpin is not so obvious because the binding efficacy of both 5’-16 and 3’-16 reach 100% at equimolar concentration.
Based on their hybridization properties, oligonucleotides 3’-16, 3’-h-6/16, 3’-h-9/16, 5’-h-6/14, 5’-h-9/14, 5’-16, 5’-h-6/16 and 5’-h-9/16 were chosen as addressing components for synthesis of POCs. Hairpin oligonucleotide 3’-h-9/10, containing a 10-mer antisense sequence with poor binding with miR-21, was also included for synthesis and used in the study as a control. The synthesized conjugates have the same nomenclature as corresponding oligonucleotides, but shown in bold.

![Figure 4.2: Hybridization of oligonucleotides with 5’-[32P]-miR-21 (*miR-21). Autoradiograph of 12% native PAGE. *miR-21 (1μM) and one of the oligonucleotides (1 μM) were incubated at 37°C for 1 h. The samples were loaded onto the running gel immediately after quenching of the reaction with an interval of 2 min.](image)

### 4.4.3 POCs synthesis and characterization

POCs, with the peptide acetyl-[(LeuArg)₂Gly]₂ attached to the aminohexyl linker located at either the 5’- or 3’-end of the addressing oligonucleotide via its C-terminal carboxylic group, were synthesized, purified and characterized as described in ‘Materials and methods’ section. The general structure of anti-miR-21 POCs and amino acid sequence of the peptide are shown in Figure 4.1.

Peptide was synthesized using an Fmoc-Gly-Wang resin (C-terminal carboxylic acid) and manual solid-phase methodology utilizing the common Fmoc/tBu protocol. The detailed information on synthesis and purification protocols is given in 50.

An array of various POCs incorporating the same catalytic peptide, but different miR-21 specific oligonucleotides (see Figure 4.1 and Table 4.1) were synthesized by activation of C-terminal carboxylic group of the peptide with N,N'-dicyclohexylcarbodiimide in the presence of 4-dimethylaminopyridine under anhydrous conditions. The acetyl-protected N-terminus of the peptide prevented the self-condensation of the peptide. The activated
peptide was added to the DMSO-soluble cetyltrimethylammonium salt of oligonucleotide and incubated as described in ‘Materials and methods’ section. Various reaction conditions (e.g. reagent excess, volume, temperature, time) were sequentially varied to find the optimum for conjugate yield, with the biggest contributory factors being minimal reaction volume and reagent excess (data not shown). Conjugates were isolated by precipitation with 4% LiClO4 (w/v) in acetone and purified by RP-HPLC. Reproducible yields between 70-90% were achieved. Identity and purity of the synthesized POCs have been confirmed by RP-HPLC, 1H-NMR and mass spectrometry (see ‘Materials and methods’ section and Supplementary Information).

4.4.4 Hybridization properties of POCs

The ability of conjugates to hybridize with *miR-21 was studied at 37°C by gel-retardation assay. From the data shown in Figure 4.3-A, it is apparent that at 5-fold excess, all studied conjugates, except 3'-h-9/10, bind to the target as efficiently as the corresponding oligonucleotides, demonstrating quantitative binding. Conjugates 3'-h-9/16 and 5'-h-9/16 bind to miR-21 with the formation of complexes of different stoichiometry (Figure 4.3-A).

![Figure 4.3: Hybridization of POCs and corresponding addressing oligonucleotides with *miR-21. A. Autoradiograph of 12% native PAGE. *miR-21 (1 µM) and conjugates or oligonucleotides (5 µM) were incubated at 37°C for 20 min. The samples were loaded onto the running gel immediately after quenching of the reaction with an interval of 2 min. B. The concentration dependency of binding of oligonucleotide 5'-h-9/14 or conjugate 5'-h-9/14 with miR-21. *miR-21 (1 µM) and oligonucleotide or conjugate in different concentrations were incubated at 37°C for 20 min.](image)

Detailed analysis of the concentration dependencies shows that the attachment of the peptide to an oligonucleotide decreases binding with miR-21 (Figure 4.3-B): at equimolar
concentration, the binding efficiency of 5'-h-9/14 is only one third of the binding efficiency of corresponding oligonucleotide, but at 5-fold excess, the binding of both oligonucleotide and conjugate reaches 100%. The reduced hybridization properties of conjugates in comparison with parent oligonucleotides could be explained by (at least) two reasons. First, non-hybridized conjugates may form intermolecular complexes via self-assembly induced by electrostatic interaction between the positively charged peptide moiety of one conjugate molecule and the negatively charged oligonucleotide motif from different molecular species. This molecular assembly may decrease the effective concentration of monomeric form of conjugates in the solution, which is essential for hybridization with miRNA target. In addition, electrostatic interactions between the arginine residues of the peptide and the phosphate groups of the oligonucleotide within the same conjugate molecule may stabilise intramolecular interactions and thus induce some folded conformations, which could be detrimental for the hybridization of the conjugate with the approaching miRNA target due to an induced steric hindrance.

4.4.5 Site-selective ribonuclease activity of POCs

Site-selective miR-21 cleavage by POCs bearing peptide either at the 3’-end of the corresponding oligonucleotide (denoted as 3’-conjugates) or the 5’-end (denoted as 5’-conjugates) was studied. This was performed at 20 – 50-fold excess of POC over RNA to ensure detection even for low cleavage extent. *miR-21 (1 μM) was incubated with one of the conjugates for 0 – 144 h. As a control of cleavage selectivity, conjugate 3’-h-9/10 (which was shown to be unable to form a stable duplex with miR-21) was used.

Analysis of miR-21 cleavage by 3’-conjugates shows that cleavage activity of these molecules is very low, even if a 50-fold excess of the conjugates is used. All 3’-conjugates cleave miR-21 at a single linkage G₃C₄ and cleavage efficacy does not exceed 5% in 144 h (Figure S2). In spite of the fact that we do not detect formation of stable complementary complex between miR-21 and conjugate 3’-h-9/10, this compound also cleaves miR-21 at the site G₃C₄ with similar efficacy, thus raising the question of whether the cleavage occurs within a specific POC/miR-21 complex (Figure S2).

By contrast, 5’-conjugates exhibit high cleavage activity (Figure 4.4). The cleavage extent was shown to vary from 57% to 99% at 20-fold excess in 72 h (Figure 4.4, Table S2).
Conjugates 5’-16, 5’-h-6/14 and 5’-h-9/14 seem to be the most efficient and were found to cleave about 65 – 85% of miR-21 in 24 h, whereas conjugates 5’-h-6/16 and 5’-h-9/16 cleave only 18-20% of miR-21 for the same time interval (Figure 4.4-C, Table S2). Kinetic curves for conjugates 5’-16, 5’-h-6/14 and 5’-h-9/14 plateau after 48 h at an RNA cleavage extent 93 – 99%. For the same time interval, the kinetics for conjugates 5’-h-6/16 and 5’-h-9/16 are linear or close to linear and reach 57 and 45%, respectively (Figure 4.4-C).

Figure 4.4: Kinetic and concentration analysis of *miR-21 cleavage by 5’-conjugates. A. Autoradiograph of 18% polyacrylamide/8 M urea gel. Lanes Im and T1, imidazole ladder and partial RNA digestion with RNase T1, respectively. Control – RNA incubated in the absence of conjugate for 0 and 72 h. *miR-21 (1 µM) and 5’-conjugates (20 µM) were incubated at 37°C for 0 – 72 h. The conjugate type and incubation time are shown at the top. B. Positions of miR-21 cleavage induced by 5’-conjugates 5’-h-6/16 and 5’-h-9/14 (indicated by arrows). C. Kinetics of *miR-21 cleavage by conjugates. *miR-21 (1 µM) and one of the conjugates (20 µM) were incubated at 37°C for 0 – 72 h. D. Concentration dependence of *miR-21 cleavage at the time point 24 h. *miR-21 (1 µM) and conjugates (1 – 50 µM) were incubated at 37°C for 24 h.
The half-lives calculated from the kinetics for conjugates 5’-h-9/16 and 5’-h-6/16 are 43.7±1.2 and 59.9±2.0 h respectively (Figure 4.4-D, Table S2), whereas the rates of miR-21 cleavage by conjugates 5’-16, 5’-h-6/14 and 5’-h-9/14 are significantly higher – the half-lives are 15.1±0.2, 10.0±0.17 and 17±0.4 h, respectively (Table S2).

Figure 4D shows the concentration dependencies of miR-21 cleavage by POCs at time point 24 h. The most active POCs 5’-16, 5’-h-6/14 and 5’-h-9/14 exhibit similar dependencies, with a plateau reached at a POC concentration of about 10 – 20 µM. Interestingly, at this time point in concentration range 20 – 50 µM, the maximum extent of miR-21 cleavage does not exceed 60-65% (Figure 4.4-D). These data show that complex formation is not the rate limiting step of site-selective RNA cleavage. Significantly less active POCs display a linear concentration dependency (5’-h-9/16) or a curve which plateaus at 20 µM of POC concentration (5’-h-6/16) when RNA cleavage is 20%. Thus, comparison of kinetics and concentration dependencies for these POCs shows that the behavior of 5’-h-6/16 is similar to the most active conjugates and 5’-h-9/16 exhibits a strong dependence of cleavage extent versus concentration (linear curve), while the kinetics has a tendency to plateau.

Depending on the length of antisense component of POCs, two or three cleavage sites are detected (Figure 4.4-A, 4.4-B). Within the first two hours, parallel and independent cleavage of three linkages, G15A16, G18U19 and G21A22, for 14-mer conjugates 5’-h-6/14, 5’-h-9/14 appear; and two linkages emerge, G18U19 and G21A22, for conjugates with a longer antisense part 5’-16, 5’-h-6/16, 5’-h-9/16 (Figure. 4.4-A, 4.4-B, Figure S3). Up to 72 h, the linkages located closer to the peptide are the main cleavage sites (63 – 68% of cleavage); and as the distance from the peptide to the linkage increases, the cleavage intensity drops: for 5’-h-6/14 G15A16 >> G18U19 >> G21A22; for 5’-h-6/16 G18U19 >> G21A22 (Figure 4.4-A, 4.4-B).

4.4.6 Ribonuclease activity of control non-complementary POCs against miR-21

In order to evaluate the biological activity of developed miRNases, the control conjugate 5’-luc-h-9/14 was synthesized based on the same design concept as that of the conjugate 5’-h-9/14. However, the oligonucleotide part of the 5’-luc-h-9/14 control conjugate was replaced with the fragment of firefly luciferase gene, sequence of which is not found in the mammalian genome. The cleavage of miR-21 by control non-complementary conjugate 5’-luc-h-9/14 was studied.
It was revealed that in the absence of complementary sequence within the target, the conjugate 5′-luc-h-9/14 was also able to cleave the substrate, but with much lower efficiency. The cleavage of miR-21 by this conjugate was observed at all G-X linkages along the RNA molecule with similar kinetics (Figure 4.5), whereas the cleavage of miR-21 with the sequence-specific conjugate 5′-h-9/14 occurs in a different manner, with the most effective scission at G-X bonds which is closest to the peptide location. The analysis of concentration dependencies shows that the total cleavage extent of miR-21 by non-complementary conjugate 5′-luc-h-9/14 is approximately 2-fold lower than by the miR-21 specific conjugate 5′-h-9/14 (Figure 4.5).

**Figure 4.5:** Concentration analysis of miR-21 cleavage by complementary conjugate 5′-h-9/14 and non-complementary conjugate 5′-luc-h-9/14. A. Autoradiograph of 18% polyacrylamide/8 M urea gel. Lanes Im and T1, imidazole ladder and partial RNA digestion with RNase T1, respectively. C – RNA incubated in the absence of conjugate for 24 h. *miR-21 (1 µM) and conjugates (1 – 20 µM) were incubated at 37 °C for 24 h. B. Concentration dependency of *miR-21 cleavage at the time point 24 h.

This suggests that, in the absence of complementarity, conjugate 5′-luc-h-9/14 behaves as a usual non-specific ribonuclease, which we observed in earlier works. This could be attributed to some type of opportunistic cleavage during transient electrostatic interactions or imperfect hydrogen bonding between the ribonuclease and RNA substrate. As soon as the complementary stretch appears within the target, this non-specific behavior is controlled, and the conjugate has no other choice than to cleave site-selectively.
4.4.7 Molecular modelling

In order to obtain insight into putative conjugate/miR-21 structures, a model of the 5'-h-6/14 /miR-21 complex was constructed using homology and NMR structural information (see ‘Materials and Methods’). A subsequent microsecond molecular dynamics simulation of the complex was performed, with the effect of aqueous solvation incorporated via generalized Born implicit water. The backbone RMSD of the complex stabilizes after 0.2 μs, with a minor subsequent shift at 0.65 μs (Figure S4). The duplex and hairpin structures remain intact throughout the simulation (Figure S5). The peptide moiety, AcO–(LeuArg)2-Gly-(LeuArg)2-Gly-COOH, forms persistent interactions with miR-21 (Figure 4.6-A, 4.6-B and Figure S5).

Specifically, the arginine residues form a network of ionic interactions with backbone phosphate groups in the 15-22 region of miR-21 (Figure 4.6-A, 4.6-B). Of particular interest, the guanidino sidechains of Arg5 and Arg10 (see Figure 4.1-B for amino acids numbering) interact strongly with the phosphate oxygens of the G15A16 site, the site which provides the major ‘end-point’ cleavage product after 48 h of incubation. Indeed, at 1 μs, there are Arg10 Hη1 and Hη2 distances to a G15A16 P=O oxygen atom of 1.93 and 2.34 Å respectively; for Arg5, these distances, to the other P=O oxygen, are 1.68 and 2.09 Å (Figure 4.6-A, 4.6-B; Figure S6). Interestingly, ancillary hydrogen bonding interactions of the Arg10 sidechain with the N7 and carbonyl O atoms of G15 are also evident from the simulation (Figure S7). The shift in RMSD at 0.65 μs corresponds to the involvement of Arg5 in interacting with the G15A16 phosphate backbone (Figure S6B) with reduced engagement of Arg8 with the base of G15 (Figure S8).

Following Soukup and Breaker\(^{78}\), it is possible to estimate a propensity for phosphodiester cleavage based on angle \(\tau\), defined by the oxygen atom of the 2'-hydroxyl group, the phosphorus atom and the oxygen atom of the departing 5'-group; and on distance \(d\) between the 2'-hydroxyl group and the phosphorus atom. We find here that relatively favorable values of \(\tau\) and \(d\) can be achieved at the G15A16 site in the simulation, although not predominant (Figure 4.6-C, 4.6-D). Based on simultaneously sampled \(\tau\) and \(d\) values from the trajectory of 133° of 3.3 Å respectively, a computed in-line fitness \(F\) of 0.51 is obtained using the equation of Soukup and Breaker (where a \(F\) value of 1 indicates the highest possible predisposition towards cleavage, and 0 least disposed)\(^{78}\).
Figure 4.6: Results of the molecular dynamics simulation of the 5'-h-6/14/miR-21 complex. A, B. Structure of 5'-h-6/14 complex with miR-21 after 1 μs of molecular dynamics simulation in implicit solvent, viewed from two alternative orientations. For 5'-h-6/14, oligonucleotide in white, Leu residues in yellow, Gly in white and Arg by atom type. For miR-21 residues 1-14 in pink; for residues 15-22, Ade in purple, Ura in green and Gua in mauve. The two parts of the system folded together in a compact spherical-like shape at the end of the double helix due to the network of electrostatic interactions involving polar groups in the two parts of the system: salt bridges between the arginine guanidino and RNA phosphate groups, and hydrogen bonds involving arginine side chains, base polar atoms and sugar hydroxyls. C. Distribution of angle \( \tau \) over the 1 μs simulation. D. O2' [G15]⋯P[A16] distance \( d \) over the 1 μs simulation.

We do note however other favorable in-line conformations are achievable at other phosphate linkages in the 15-22 region of miR-21. For example, during the MD simulation, a (\( \tau, d \)) orientation of (153°, 2.9 Å) is achieved at the G18U19 site, the second preferred site of cleavage found experimentally (Figure 4.4 and S9), yielding an in-line fitness \( F \) of 0.86. Indeed, complete sampling of the complex network(s) of ionic contacts may be to some extent limited here by the use of an implicit solvent model, which could overestimate the stability of these interactions. While future work could consider more computationally
demanding explicit solvent calculations, this molecular dynamics simulation nevertheless is indicative of potentially reactive conformations of the conjugate in its hybrid with miR-21.

4.4.8 Effect of conjugate 5'-h-9/14 on the level of miR-21 in lymphosarcoma cells RLS₄₀

From the point of view of therapeutic importance, the principal question is to ascertain whether the developed miRNases are able to inactivate miR-21 in the biological system. The intracellular accumulation of therapeutic compounds is a very important point. For this reason prior to cellular experiments the accumulation of FITC-labeled peptide-oligonucleotide conjugate 5'-h-9/14 (FITC-5’-h-9/14) in murine lymphosarcoma cells RLS₄₀ and human cervical cancer cells KB-8-5 mediated by Lipofectamine 2000 was studied using flow cytometry and confocal microscopy. Recently, we have shown that the treatment of murine lymphosarcoma cells with Lipofectamine decreased their proliferation rate by no more than 10%, which was within the experimental error of MTT assay. We found that 4 h after transfection of cells with 1 µM FITC-5’-h-9/14/Lipofectamine lipoplexes, 87% of RLS₄₀ cells and 86% of KB-8-5 cells became FITC-positive with the average cell fluorescence of 12.6 and 40 RFU for RLS₄₀ and KB-8-5 cells, respectively (Figure S10). The difference in RFU values between the studied cells can be attributed to the dissimilarity in the cell size. Indeed, human KB-8-5 cells are considerably larger than RLS₄₀ cells, and thus accumulate greater amount of FITC-POC leading to a stronger fluorescence signal in comparison with murine cells. It was also shown in these experiments that in the absence of the transfection agent FITC-5’-h-9/14 was able to penetrate approx. 2% of cells only.

Confocal microscopy showed that FITC-5’-h-9/14-POC/Lipofectamine lipoplexes are effectively accumulated in the cytoplasm of KB-8-5 cells 4 h after transfection, but did not enter the nucleus (Figure S11). Thus, both flow cytometry and confocal microscopy showed that the conjugation of oligonucleotides with short peptides neither noticeably affect the lipoplexes formation nor alter their cellular localization.

The ability of the conjugates to affect miR-21 level in tumour cells was tested using conjugates 5’-h-6/14 and 5’-h-9/14 and mouse lymphosarcoma cells RLS₄₀, which are characterized by elevated levels of miR-21. The concentration dependence of miR-21 suppression in tumor cells by the conjugate 5’-h-6/14 (0.1 – 5 µM) 24 h after transfection
of cells with the conjugate/Lipofectamine complexes measured by stem-loop qPCR is shown in Figure S12. It can be seen that control conjugate 5'-luc-h-9/14 did not decrease miR-21 level, whereas even a 0.5 µM concentration of 5'-h-6/14 resulted in a noticeable decrease in miR-21. The most effective dose was found to be 1 µM, and any further increase in 5'-h-6/14 concentration from 1 to 5 µM did not affect the efficiency of miR-21 silencing. Based on this data, 1 µM concentration of the conjugate was chosen for testing its biological activity in the following experiments.

In experiments on the kinetics of miR-21 suppression RLS40 cells were transfected with 1 µM oligonucleotide 5'-h-9/14, control conjugate 5'-luc-h-9/14 or anti-miR-21 conjugate 5'-h-9/14 precomplexed with Lipofectamine 2000 and the level of miR-21 was evaluated at 24, 48 and 72 h after transfection using stem-loop qPCR. Analysis of qPCR data (Figure 4.7-A) shows that 24 h post transfection, conjugate 5'-h-9/14 provides more than a two-fold reduction in miR-21 level in tumour cells; this is statistically significant in comparison with all groups (p=0.0002 for 5'-h-9/14 vs control, LF and 5'-h-9/14 asON; p=0.0003 for 5'-h-9/14 vs 5'-luc-h-9/14) and to the time point 72 h the level of miRNA is restored to 80% from the initial level (Figure 4.7-A). Conjugate 5'-luc-h-9/14 causes only a slight reduction in the level of miR-21 seen at 24 h after transfection, and this difference is statistically insignificant (Figure 4.7-A). The asON 5'-h-9/14 does not decrease miR-21 levels or it cannot be detected by qPCR analysis (Figure 4.7-A).

To confirm the specificity of miR-21 silencing by the conjugate, the level of the other miRNAs, let7-g, miR-17 and miR-18a was also monitored. The study of expression of let7-g, miR-17 and miR-18a in RLS40 cells shows that there is no statistically significant decrease in the level of miRNAs after transfection, confirming the selectivity of miR-21 silencing with the conjugate 5'-h-9/14 (Figure 4.7-B, Figure S13).

4.4.9. Effect of conjugate 5'-h-9/14 on the level of miR-21 protein target PDCD4 in RLS40 lymphosarcoma cells

The decrease in the level of oncogenic miR-21 in tumour cells under the effect of the conjugate should promote the restoration of normal activity of tumour-suppressive target genes of miR-21 and as a consequence their protein products. Due to the decline in miRNA concentration in the cells, changes in the level of target proteins are usually observed in approximately 48–72 h. In order to determine if there was any alteration in the level of protein PDCD4, the direct target of miR-21, Western blot analysis was performed 48 h
after transfection of RLS_{40} cells with 1 μM oligonucleotide 5'-h-9/14, control conjugate 5'-luc-h-9/14 or anti-miR-21 conjugate 5'-h-9/14 (Figure 4.7-C, 4.7-D).

**Figure 4.7:** Biological activity of the conjugate 5'-h-9/14 in RLS_{40} lymphosarcoma cells. Expression level of miR-21 (A) and let7-g (B) in RLS_{40} cells after transfection with 1 μM of antisense oligonucleotide 5'-h-9/14, control conjugate luc-h-9/14 and anti-miR-21 conjugate 5'-h-9/14. Transfection was performed using Lipofectamine 2000. LF – RLS_{40} cells incubated with Lipofectamine 2000 only. The expression of miRNAs was normalized to U6 and Rpl30. C. Western blot analysis of PDCD4 48 h after transfection. GAPDH served as an internal control. 1 – intact RLS_{40} cell; 2 – RLS_{40} cells incubated with Lipofectamine 2000; 3, 4, 5 – RLS_{40} cells incubated with 1 μM control conjugate luc-h-9/14, antisense oligonucleotide 5'-h-9/14, and anti-miR-21 conjugate 5'-h-9/14, respectively. D. The bar graph shows the semi-quantitative analysis of the western blot results for PDCD4. E. Proliferative potential of RLS_{40} cells 72 h after transfection with 1 μM of antisense oligonucleotide 5'-h-9/14, control conjugate luc-h-9/14 and anti-miR-21 conjugate 5'-h-9/14. Data are given as medians calculated from three independent experiments.
The data obtained show that there is no statistically significant change in the level of PDCD4 in tumour cells incubated with control conjugate 5'-luc-h-9/14 or antisense oligonucleotide 5'-h-9/14. By contrast, in the cells incubated with anti-miR-21 conjugate 5'-h-9/14 (when the reduction in miR-21 level is observed), the level of PDCD4 increases 1.9-fold in comparison with control (Figure 4.7-C, 4.7-D).

4.4.10 Effect of conjugate 5'-h-9/14 on proliferative potential of RLS40 lymphosarcoma cells

The decrease in the level of oncogenic miR-21 and reactivation of its target proteins in tumour cells under the effect of the conjugate should reverse the malignant behavior of tumour cells, in particular inhibiting cell growth, inducing apoptosis and reducing the invasive and migratory properties of the cells. To assess the presence of conjugate-mediated effects on the behavior of tumour cells, we examined the ability of tumour cells to proliferate after exposure to anti-miR-21 conjugate. For this purpose RLS40 cells were transfected with 1 μM oligonucleotide 5'-h-9/14, control conjugate 5'-luc-h-9/14 or anti-miR-21 conjugate 5'-h-9/14 using Lipofectamine 2000. To determine the growth rate of lymphosarcoma cells, the doubling time of cell populations was calculated by counting the number of cells 72 h after transfection. Data analysis shows that cell doubling time for intact RLS40 cells is 22.3±0.7 h. For cells treated with Lipofectamine 2000 or control conjugate 5'-luc-h-9/14, this parameter is 22.7±0.9 and 22.9±1.0 h, respectively. Transfection of RLS40 cells with anti-miR-21 antisense oligonucleotide 5'-h-9/14 results in an increase in the cell doubling time to 25.7±1.0 h. The most noticeable decrease in cell growth rate is observed for cells treated with anti-miR-21 conjugate 5'-h-9/14, where the cell doubling time increases to 1.4-fold and is 33.3±2.2 h. The data obtained are supported by an analysis of the proliferative properties of lymphosarcoma cells by MTT assay (Figure 4.7-E). Data analysis reveals that transfection of cells with Lipofectamine 2000 or control conjugate 5'-luc-h-9/14 causes to no statistically significant change in their proliferative properties. Antisense oligonucleotide exerts a 30% suppression of proliferation of lymphosarcoma cells ($p=0.00012$ for 5'-h-9/14 asON vs control, Figure 4.7-E). In the case of conjugate 5'-h-9/14, the viability of tumour cells falls to 50% as compared to control conjugate ($p=0.0003$ for 5'-h-9/14 vs control, Figure 4.7-E) and the
effect is 1.4-fold stronger as compared with antisense oligonucleotide (p=0.05 for 5’-h-9/14 vs 5’-h-9/14 asON, Figure 4.7-E).

4.5 Discussion

The present investigation is focused on the development of metal-free miRNA-specific peptide-oligonucleotide conjugates, intended for miRNA degradation and silencing in tumour cells without the need for exogenous species (e.g., metal ions). To date, the main method for miRNA inhibition is antisense technology that uses oligonucleotides complementary to miRNA for performing either RNA degradation by endogenous RNase H or steric blockage. The latter mechanism of action is more prevalent because chemical modifications required for improvement of nuclease stability and affinity of oligonucleotides to the target molecule lead to a loss of the ability to recruit endogenous RNase H, but provide steric blockage of target miRNA. Nevertheless the approaches aimed at destruction of miRNAs seem to be preferable due to irreversibility and therefore higher efficiency.

MiRNA-specific POCs, termed ‘miRNases’, represent a hybrid technology that combines (i) specific recognition and binding with target miRNA via antisense oligonucleotide and (ii) cleavage by the conjugated artificial ribonuclease. Preceding investigations of sequence-specific artificial ribonucleases have been largely limited to the use of model RNA targets (tRNAs, synthetic RNA-substrates, etc.)42-51, (see reviews53, 54) with limited attempts to apply these promising strategy to biologically or clinically relevant systems. Only a few examples of application of metallo-dependent sequence-specific artificial nucleases for degrading of biologically significant RNA molecules have been described81, 82. It has been reported that in vitro transcripts of 571- and 2977-nucleotide long human c-raf-1 RNA were efficiently cleaved by 2’-methoxyethoxy-modified oligonucleotides bearing a europium complex in a sequence specific manner81. The scission technique based on acridine-modified DNA and lanthanide (III) ion as a catalyst was shown to promote sequence-specific cleavage of a 40-mer fragment of human apolipoprotein E gene; this allowed identification of a single nucleotide polymorphism at the site of interest82. Application of sequence-specific conjugates in cell culture has been explored only in a few studies, demonstrating the advantage of applying of conjugates against corresponding antisense oligonucleotides in modulating RNA83, 84. The lanthanide macrocycle complex,
Eu(THED)$_{3}^{3+}$, attached to antisense oligonucleotide potentiated the inhibition of ICAM-1 protein expression in endothelial cells more efficiently than unconjugated oligonucleotide. Antisense phosphorothioate oligonucleotide bearing the 5’-imidazole group provided sequence-specific cleavage of a complementary target sequence in the gag mRNA from HIV-1 and inhibited the replication of HIV-1 in MT-4 cells more effectively than the parent antisense oligonucleotide.

Since miRNA molecules are short, the design of site-selective chemical nucleases for miRNA targeting represents a challenging task: an appropriate balance needs to be achieved between the length of the miRNA-binding component, sufficient for efficient complex formation, and a fragment of RNA accessible for cleavage by the conjugate.

To date, there is only one published work describing a miRNA-targeted metallo-dependent ribonuclease. Conjugates of peptide nucleic acid (PNA) – PEG–PNA–PEG – with HGG·Cu or DETA as cleaving domains, targeted to the hsa-miR-1323, demonstrated effective cleavage of RNA. These conjugates engaged with the miRNA molecule and performed cleavage, with the release of one or two nucleotides from the 3’-end of the miRNA-target (elongated by 5 nucleotides). In spite of clear success demonstrated by these metal-dependent artificial ribonucleases, they may potentially suffer from the risk of metal-loss from the coordinating ligand and uncontrolled diffusion of metal ions, which may lead to degradation of non-target biopolymers and cause undesirable toxicity.

Here, for the first time we apply a metallo-independent peptide-oligonucleotide chemical ribonuclease towards a clinically relevant biological target – miRNA-21. The oligonucleotides of two different structures were used for the design of the addressing domain of miRNA-targeting POCs – linear and hairpin oligonucleotides with the length of complementary part ranging from 10 to 16 nucleotides. The analysis of hybridization properties of oligonucleotides has shown that oligonucleotides with a length less than 14 nucleotides are not suitable for conjugate engineering due to very weak binding affinity for the target miR-21. We demonstrated that hairpin oligonucleotides with a 14-mer complementary part exhibit significantly enhanced hybridization properties in comparison with linear ones; however, elongation of the stem in the hairpin oligonucleotide from 6 to 9 nucleotides does not provide any advantage for binding. The addition of the hairpin to the 16-mer oligonucleotides 5’-h-6/16 and 5’-h-9/16 does not increase its binding affinity as the binding efficiency of linear 16-mer oligonucleotide is almost 100%. Therefore, linear and hairpin oligonucleotides with the length of the complementary part of 14 and 16 nucleotides were selected for POC synthesis.
Screening of the miR-21 cleaving activity of all designed 3’- and 5’-POCs was performed under large excess of POCs over the target to ensure that even faint cleavage would not be missed. Analysis of ribonuclease activity of 3’-POCs shows that this type of conjugate is not able to perform effective miR-21 cleavage. We can assume that the conjugate, which is obtained through attachment of the peptide at the 3’-end of oligonucleotide, does not form the effective intramolecular contacts between the oligonucleotide and peptide required for generation of a catalytically active conformation. Analysis of ribonuclease activity of 5’-POCs shows that this type of conjugate provides effective miR-21 cleavage at G-X bonds exceptionally and that the main cleavage sites are located near the point of peptide attachment. POCs 5’-h-9/14 and 5’-h-9/14 are the most effective and the total cleavage extent of miR-21 by these POCs reaches 100% (Figure 4.4, Table S2). Interestingly, POCs 5’-h-6/16 and 5’-h-9/16, which bear the elongated 2 nucleotides RNA-binding part, cleave miR-21 less efficiently: cleavage reached only 57-83% (Figure 4.4, Table S2). Several factors may be responsible. RNA cleavage by the peptide-oligonucleotide conjugate is a complex, multi-step process that involves (i) the hybridization via Watson-Crick hydrogen bonding and (ii) the formation of the catalytically-active conformation followed by (iii) the catalytic act itself. Our recent data50, 51, 76, 77 clearly demonstrate that even slight structural modifications of conjugates (i.e. variations in oligonucleotide length and composition as well as change of accessibility and base composition of the RNA cleavage sites) may dramatically affect the catalytic efficiency of such molecules. In the case of 5’-h-6/16 and 5’-h-9/16 POCs, the antisense part of the conjugate was elongated from 14 to 16 nucleotide residues thus leading to shortening of the target single-stranded region of miR-21 within the hybridized complex. It was shown earlier51 that narrowing of the RNA single-stranded target region upon binding with the conjugate may result in substantial decrease or even complete loss of catalytic activity, presumably due to reduced conformational freedom for the catalytic peptide. Additionally, the most accessible bond to the cleavage by POCs 5’-h-6/14 and 5’-h-9/14 seems to be G-A, whereas POCs 5’-h-6/16 and 5’-h-9/16 tend to attack G-U bond due to elongation of the oligonucleotide part, which is known to be less sensitive to the cleavage by nucleases. Finally, due to reduced conformational flexibility, the elongated duplexes miR-21/5’-h-9/16 and miR-21/5’-h-6/16 may obstruct the formation of the in-line conformation that is essential for cleavage to occur50, 51. Although the ribonuclease activity seen for 5’-16, 5’-h-6/14 and 5’-h-9/14 conjugates is very similar, hairpin conjugates seem to be more advantageous for application in cells. First, the only target of hairpin conjugates
is the mature miRNA molecule. Also, whereas conjugate 5'-16 provides cleavage of the tail region of miRNA, conjugates 5'-h-6/14 and 5'-h-9/14, possessing about the same efficiency of binding to the miRNA-target, catalyze the cleavage close to the middle part of the miRNA molecule, which can result in the loss of miRNA function with higher probability and reliability.

The specificity of conjugates deserves special attention. The developed nucleases mimic T1 RNase specificity and perform cleavage exclusively at G-X linkages within the RNA target. Previously we have shown that when the peptide [(LeuArg)2Gly]2 was attached to an oligonucleotide through N-terminus with a linker of zero length, the conjugates demonstrated Pyr-A specificity50. However, as soon as the type of conjugation is changed by introducing a new attachment point (i.e. via peptide C-terminus) by incorporation of the flexible linker into the conjugate structure, the specificity of chemical nucleases is switched to G-X. Previously, G-X specificity of RNA cleavage was observed for specially designed non-site selective artificial ribonuclease pep-976,77 and ‘dual’ conjugates targeted to tRNA, both contained long flexible linkers, and the attachment via C- and N-terminus of the peptide was used51.

It is worth considering in more detail the precise site-selectivity of the developed conjugates: the most efficient cleavage is accomplished at the phosphodiester bonds adjacent to the peptide attachment point and cleavage efficiency decreases with the remoteness from the reactive group. Nevertheless, we cannot rule out completely the possibility of formation of imperfect complexes of conjugates with RNA molecules: it was revealed that non-complementary conjugate 5'-luc-h-9/14 performs non-site-selective cleavage of miR-21 however with much reduced efficiency (Figure 4.5).

The preferential specificity of RNA cleavage by POCs at phosphodiester bonds after guanine residues can be explained by the high affinity of arginine to guanine. Arginine can form hydrogen bonds with acceptor groups in the RNA molecule such as phosphate oxygens, the 2’-OH oxygen of ribose, and acceptor groups in nucleobases (Gua O6 and N7, Ura O4 in the major groove, Gua N3 and Ura O2 in the minor groove). Specific Arg-Gua recognition was widely shown. With the example of onconase, it was shown that replacement of Glu91 by an arginine residue in an onconase binding centre that specifically binds with guanine nucleobase in d(AUGA) increased the guanine preference and afforded an onconase variant with the highest known kcat/KM value86. A direct interaction between an arginine residue Arg144 in the active site of E. coli prolyl-tRNA synthetase and the G72 residue in the acceptor stem of tRNAPro was found to be a key
contact providing the allostERIC interaction between the anticodon domain and the aminoacylation active site. It was shown that the structural basis for RNA recognition by ZRANB2 protein was the formation of a special network of hydrogen bonds between the GGU-containing ssRNA-motif and Arg81 and Arg82 of ZRANB2 via bidentate interactions of both Gua with side chains of arginines.

Indeed, here from 1 μs molecular dynamics simulations of the hybridized complex between 5'-h-6/14 and miR-21, we found that strong interactions are formed between the Arg5 and Arg10 sidechains of the peptide with the phosphate oxygens of the main cleavage site on miR-21, G15-A16. Additional hydrogen bonding by Arg10 with the N7 and O6 atoms of G15 is also observed over the course of the simulation (Figure 4.6-A, 4.6-B). The overall conformation of the complex from the microsecond trajectory displays tight binding of the oligonucleotide part of the POC with the target RNA, forming a network of ionic and hydrogen bonding contacts between the peptide and nucleotide residues (Figure 4.6-A, 4.6-B). A similar pattern of interactions was observed for G18-U19 (Figure S9).

Experimental data showed that it is formation of the complex that leads to effective site-selective cleavage, rather than cleavage from solution (Figure 4.5). Non-selective cleavage could take place due to the binding of the arginine residues of the peptide with so-called arginine binding sites, known to be GX or GGX motifs, especially GGU. Conformational stress of miRNA in the sites of these non-sequence-selective POCs binding lead to proximity of the 2'-OH and phosphate groups, with formation of an “in-line”-like conformation that facilitates internal transesterification. This process is not effective, since only faint cleavage is observed at G3C4 and G21A22 sites.

The most significant result of this study lies in the fact that the developed conjugates are able to reduce the level of miR-21 in tumour cells. This results in reactivation of target protein PDCD4 and inhibition of cell proliferation (Figure 4.7). At the same time, we do not observe a decrease in the levels of miR-21 after tumor cell exposure to asON 5'-h-9/14. The study of stability of asON and corresponding conjugate in 10% serum shows that during the first 4 h conjugate is approximately 2-fold more stable than asON, and after this time both asON and conjugate are completely degraded (Figure S14). This increase in the serum stability of the conjugate is not a reason to believe that the effect of the conjugate is based on the protective effect of the peptide.

As a recognition domain in the structure of conjugates, we used non-modified DNA oligonucleotide, and the degradation of miRNA in the cell can occur as a result of engaging of intracellular RNase H. Our findings show that the level of specific
downregulation of miR-21 by the action of the conjugate is higher than the effect of antisense oligonucleotide. In this regard, one question is whether the decrease in miR-21 level is realized either (i) at the expense of the conjugate’s ability to function or (ii) we observe an additive effect, composed of nuclease activity of the conjugate and RNase H-mediated cleavage of miRNA in heteroduplex. Experiments on the cleavage of duplex miR-21/5'-h-9/16 by RNase H show that the presence of peptide in the miRNase scaffold affects RNase H-activity and causes 2-fold decrease in miR-21 cleavage by RNase H within this complex. Nevertheless the total cleavage of miR-21 by RNase H in complex miR-21/5'-h-9/16 is rather notable: 43% vs 97% in complex miR-21/5'-h-9/16 (Figure S15). The contribution of RNase H into the degradation of miRNA in tumour cells can be considered as a positive event, as we may assume that it can facilitate dissociation of miR-21 fragments from the complex with the conjugate and promote the catalytic mode of action, thereby enhancing the inhibitory effect of the conjugate. Regardless of the possible RNase H contribution to the intracellular activity of conjugate, the obtained effect is comparable with the effect of anti-miRNA antisense oligonucleotides widely studied in recent years. It has been shown that chemically modified phosphorothioate and 2'-O-methyl oligonucleotides promote 2-fold decrease in miR-21 levels in various cell lines that leads to an increased expression of miR-21 protein targets such as PDCD4, PTEN and TPM1, up to 2-3-fold\textsuperscript{89, 90}, 2.2-fold\textsuperscript{91} and 1.4-fold\textsuperscript{92}, respectively. Comparing the results obtained earlier with the data obtained in this work, it can be stated that the efficiency of engineered miRNases is not inferior to the activity of the oligonucleotides previously proposed.

Due to the fact that we observed some cleavage of miR-21 by non-complementary (control) POC 5'-luc-9/14 \textit{in vitro}, it raises the question of off-target effects of the developed miRNAses in cells. It is worth mentioning that cleavage of miR-21 by non-complementary (control) POC was detected at relatively high concentration of the control conjugate (i.e. at 20-30 µM), which corresponded to 20-30 molar excess of the conjugate over miR-21 substrate (1 µM). In the \textit{in vitro} and \textit{ex vivo} experiments, the concentration of the conjugates was reduced to 1 µM. At this concentration, the level of non-specific cleavage of miR-21 with 5'-luc-9/14 did not exceed 1-2%, which was comparable with the spontaneous RNA cleavage during the incubation. Taking into account a very strong hybridization affinity of the anti-miR-21 conjugates towards the miR-21 driven by the fully complementary oligonucleotide recognition motif incorporated in their structure, we expect
this non-specific cleavage to be minimized or completely avoided *in vivo*, so that the targeted cleavage of miR-21 will dominate catalysis.

In order to further confirm the absence of noticeable off-target effect for the developed POCs, we examined the expression levels of several other miRNAs let7-g, miR-17 and miR-18a, which do not contain sequences complementary to those of the designed conjugates. From the presented data (Figure 4.7-B and Figure S13) it is seen that there is no statistically significant difference between the test samples and the controls. Moreover we compared the expression profile of several housekeeping genes, such as Gapdh, Ubc, Rpl30, Hprt1, Ywhaz and U6 in RLS40 cells after the treatment with POCs. To confirm the stability of the selected reference genes for all studied samples (LF, *luc-h*-9/14, *5*-h-9/14 asON and *5*-h-9/14 POC) the M-values were calculated (primary data not shown). We found that the differences between the genes were not more than 10-15 %, confirming the absence of non-specific effect of POC on gene expression.

Thus, designed miRNases performed *in vitro* cleavage of non-complementary RNA-substrates exhibited no off-target effects when transfected into the cells: we show that *5*-h-9/16 has no effect on the levels of several miRNAs as well as house-keeping genes while downregulating target miR-21 (Figure 4.7). This selectivity is provided by high affinity of the specifically designed oligonucleotides of the miRNase to the target RNA.

**4.6 Conclusion**

In this study, we design novel metallo-independent miRNA-specific artificial ribonucleases (‘miRNases’), incorporating oligonucleotide recognition motifs and catalytic peptide. These miRNases were demonstrated to cleave site-selectively their complementary miRNA-target with high efficiency leading to a detectable, statistically significant reduction of miRNA level in tumour cells. This seems to trigger a subsequent restoration of the expression of miR-targeted proteins and negatively affect tumour cell proliferation.

We also provided here a structural insight into the interactions between miRNases and miR-21 using molecular modelling simulations to offer a design platform for future development of novel, more efficient structural variants. Among the currently developed strategies oriented to suppress miRNA functioning, miRNases may take its well-deserved place in the field of development of anti-miRNA agents as an independent approach. Regarding future potential, miRNases may be a prospective candidate for drug
development in cases where overexpression of particular miRNAs is associated with a disease state.

4.7 Acknowledgements

This research has been supported by the Russian Science Foundation (Grant No. 14-44-00068), BBCRC (Grant No. BB/K012622/1) and EPSRC (Grant No. EP/E003400/1). The authors would like to thank Dr. Aled Williams for synthesis of 3’-conjugates, and Dr. Maria I. Meschaninova for synthesis of miR-21. The authors would like to thank Prof. David J. Clarke for fruitful discussions and manuscript editing. The authors are grateful to the EPSRC National Mass Spectrometry Centre in Swansea for MS analysis of the peptide and peptide-oligonucleotide conjugate samples.
4.8 References


34. T. Li, D. Li, J. Sha, P. Sun, Y. Huang, MicroRNA-21 directly targets MARCKS and promotes apoptosis resistance and invasion in prostate cancer cells, *Biochem Biophys Res Commun.*, 2009, **383**, 280–285.


38. E. Brognara, E. Fabbri, G. Montagner, J. Gasparello, A. Manicardi, R. Corradini, High levels of apoptosis are induced in human glioma cell lines by co-administration of peptide nucleic acids targeting miR-221 and miR-222, *Int J Oncol.*, 2016, **48**, 1029–1038.

39. S. Ottosen, T. B. Parsley, L. Yang, K. Zeh, L. J. van Doorn, E. van der Deer, In vitro antiviral activity and preclinical and clinical resistance profile of Miravirsen, a


42. N. G. Beloglazova, V. N. Sil'nikov, M. A. Zenkova, V. V. Vlassov, Cleavage of yeast trNAPhe with complementary oligonucleotide conjugated to a small ribonuclease mimic, \textit{FEBS}, 2000, \textbf{481}, 277–280.


70. O. A. Patutina, N. L. Mironova, N. A. Popova, V. I. Kaledin, V. P. Nikolin, V. V. Vlassov, The siRNA targeted to mdr1b and mdr1a mRNAs in vivo sensitizes murine lymphosarcoma to chemotherapy, *BMC Cancer*, 2010, **10**, 204-215.


Chapter 5: Conclusion and future work

The catalytic performance of the peptidyl-oligonucleotide conjugates and their potential biomedical and pharmaceutical impact rely on the conjugation of four consecutive arginine-leucine [RL]₄ peptide blocks to the oligonucleotide fragment, well before they can start demonstrating any catalytic activity against pathogenic RNA sequences. However the structural changes of the peptide upon conjugation and the way it may trigger the peptide catalytic activity has not been well understood so far.

In this project NMR experiments have been performed in order to characterise the overall conformation of the model conjugate and to reveal the peptide secondary structure for the first time. The nature of interactions between the two conjugate components has been elucidated, thus highlighting the essential role of the repetitiveness of arginine and leucine building blocks in the peptide segment for catalysis.

The employment of site-selective $^{13}\text{C}$ and $^{15}\text{N}$ labelled conjugates allowed us to use NMR techniques with enhanced sensitivity. A complete assignment of the peptide NMR signals was achieved which enabled us to determine the structure of the conjugate. In the NMR resolved structure the peptide folds around the oligonucleotide and assumes a helical structure in the region proximal to the oligonucleotide. The guanidinium groups in the arginine residues of the catalytic peptide form a network of interactions with the oligonucleotide component through (i) salt bridges with the phosphates of the oligonucleotide component and through (ii) their hydrogen bonds with the aromatic bases, which often involve the peptide amide protons in this hydrogen-bonding network.

This results in a folded conformation of the conjugate stabilized by the multiple interactions with the peptide retaining flexibility to some extent in the region proximal to the C-terminus.

The molecular dynamics simulations of the conjugate, introduced at the later stage of the research, complemented and expanded the NMR structural study and allowed identification of some important features that were not obvious from the NMR experiments due to the fast conformational exchange, not detectable at the NMR (i.e. millisecond) time.
scale. The MD simulation predicted the existence of two main conformational substrates for the conjugate, rather than a single static structure, in which the peptide is expected to exist in ordered secondary structure conformations. It was also possible to identify dynamic interactions on a microsecond timescale between the peptide and oligonucleotide components of the conjugate.

Another structural feature revealed from the NMR studies and confirmed by the work with molecular dynamics simulations is that the peptide side chains, in particular the arginine guanidinium groups, are externally exposed with the exception of the Leu9 side chain which is instead located in the hydrophobic core of the conjugate structure. The Arg8 side chain forms short-lived contacts with the oligonucleotide phosphate groups and aromatic bases but remains very mobile, as seen in the simulation trajectory of the conjugate. The Arg10 side chain and backbone form multiple contacts within the conjugate and appears to contribute to the stability of the α-helix conformation of the peptide. The Arg12 side chain appears to be involved in dynamic interactions with either the oligonucleotide or the peptide backbone. The role of guanidinium groups in Arg8, Arg10 and Arg12 appears to stabilise the conjugate in a compact conformation. In contrast, the Arg14 side chain forms only intermittent contacts with the other functional sites of the molecule and protrudes into the solvent, while moving away from the oligonucleotide backbone. In fact fewer NOEs were detectable from this residue to the rest of the structure, which provided a good correspondence between molecular modelling simulations and NMR experiments. The role of this Arg14 residue, which seems not to be essential for structural maintenance, might be to anchor the negative charge within the RNA target and direct the other positively charged groups of the conjugate into the proximity of the RNA phosphate groups, followed by catalytic cleavage of the phosphodiester bonds. This hypothesis is also supported by the dynamical nature of the interactions within the conjugate revealed in the molecular dynamics simulations. This would allow the conjugate the essential flexibility, which is needed to adapt to the structural shape of the target, to cleave and leave. The latter is the prerequisite for efficient catalytic activity of the future successful therapeutics.

As revealed from the dynamics studies of the arginine side chain interactions within the conjugate on the microsecond timescale in the simulation, more work is needed to study dynamics in the peptide side chains of the conjugate. Future consideration should involve the quantitative analysis of molecular motions of the peptide and oligonucleotide within
the conjugate by relaxation rate measurements using NMR. Nuclear Overhauser effect relaxation rates for $^{15}$N nuclei, and measurements of longitudinal ($R_1$) and transverse ($R_2$) relaxation rates could provide information on dynamics at the amide $^{15}$N sites along the conjugate peptide backbone on a subnanosecond timescale and therefore allowing identification of local intrinsically disordered regions. Along with relaxation rates, measurements of residual dipolar coupling ($H^N, N$) RDCs could extract dynamics information on the conjugate conformation local motions. Furthermore, $^{15}$N-$^{31}$P scalar coupling could allow examination of the arginine side chain/oligonucleotide phosphates salt bridges dynamics.

Further research with molecular dynamics simulations and NMR restraints also needs to be carried out to refine the conjugate NMR structure. This method will allow deriving a better representation of the structure in solution since the conjugate conformation is not rigid and the methods applied in this thesis using simulated annealing, hide the dynamics observed in the unrestrained molecular dynamics simulation.

Another future direction is to investigate the catalytic activity of the conjugate with a natural microRNA target by combining NMR techniques with restrained MD simulations to yield important insights into the cleavage activity of the peptidyl-oligonucleotide conjugate and therefore, to aid the design and development of more efficient therapeutic agents.
Appendix A

Atoms nomenclature for the peptide and nucleotides

Figure A.1: Atom nomenclatures in the peptide amino acid backbone and side chains
Figure A.2: Nomenclature and atom numbering for the nucleotides bases and sugars
Appendix B

$^{13}$C- and $^{15}$N-labelled peptides and conjugate characterization (Mass spectrometry)

Figure B.1: MS-analysis (ESI-TOF): mass calculated for labelled peptide 2 at L7R8 $^{13}$C$_{38}$$^{15}$C$_{12}$H$_{98}$N$_{15}$ $^{15}$N$_{5}$O$_{9}$ [M-H]$^{4+}$, 1167.81, found 1167.85
Figure B.2: MS-analysis (ESI-TOF): mass calculated for labelled peptide 3 at L\textsubscript{9}R\textsubscript{10}, C\textsubscript{38}\textsuperscript{13}C\textsubscript{12}H\textsubscript{98}N\textsubscript{15}\textsuperscript{15}N\textsubscript{5}O\textsubscript{9} [M-H]\textsuperscript{4+}, 1167.81, found 1167.84

Figure B.3: MS-analysis (ESI-TOF): mass calculated for labelled peptide 4 at L\textsubscript{11}R\textsubscript{12}, C\textsubscript{38}\textsuperscript{13}C\textsubscript{12}H\textsubscript{98}N\textsubscript{15}\textsuperscript{15}N\textsubscript{5}O\textsubscript{9} [M+H]\textsuperscript{+}, 1167.81, found 1168.90
**Figure B.4:** MS-analysis (ESI-TOF): mass calculated for labelled peptide 5 at L_{13}R_{14} C_{38}\textsuperscript{13}C_{12}H_{96}N_{15}\textsuperscript{15}N_{3}O_{9} [M+H]^\ast, 1167.81, found 1168.90

**Figure B.5:** MS-analysis (ESI-TOF): mass calculated for labelled conjugate L_{7}R_{8} 2 C_{96}\textsuperscript{13}C_{12}H_{170}N_{37}\textsuperscript{15}N_{5}O_{4}P_{6}^{2-} [M+2H+H]^\ast, 2979.10, found 2979.10
Figure B.6: MS-analysis (ESI-TOF): mass calculated for labelled conjugate L₉R₁₀ 3
C₉₀¹³C₁₂H₁₇₀N₃⁷¹⁵N₃O₅P₆²⁺ [M +2H+ 2Na +2H]⁺. 2979.10, found 2979.39

Figure B.7: MS-analysis (ESI-TOF): mass calculated for labelled conjugate L₁₁R₁₂ 4
C₉₀¹³C₁₂H₁₇₀N₃⁷¹⁵N₃O₅P₆²⁺ [M+H]⁺. 2979.10, found 2979.10
Figure B.8: MS-analysis (ESI-TOF): mass calculated for labelled conjugate L_{13}R_{14} 5 
C_{96}^{13}C_{12}H_{176}N_{37}^{15}N_{3}O_{45}P_{6}^{2} [M+H]^+, 2979.10, found 2979.10
Appendix C

Amber 14 Files - Simulated Annealing - chirality restraints

&rst iat= 14, 13, 25, 23, 46, 45, 55, 54, 80, 84, 85, 87, 112, 116, 117, 119, 140, 139, 151, 149, 172, 171, 181, 180, r1= -10.0, r2 = 0.0, r3= 0.0, r4= 10.0, rk2 = 1000.0, rk3 1000.0, &end
&rst iat= 73, 75, 76, 77, 120, 107, 108, 109, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 73, 75, 76, 78, 105, 107, 108, 110, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 73, 75, 88, 79, 105, 107, 120, 111, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 73, 75, 88, 87, 105, 107, 120, 119, r1= -10.0, r2 = 0.0, r3= 0.0, r4= 10.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 73, 75, 76, 77, 120, 107, 108, 110, 111, r1= -10.0, r2 = 0.0, r3= 0.0, r4= 10.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 77, 76, 78, 109, 108, 110, 111, 112, 113, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 80, 79, 76, 112, 111, 110, 108, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 88, 79, 76, 120, 111, 110, 108, r1= -10.0, r2 = 0.0, r3= 0.0, r4= 10.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 79, 80, 78, 76, 110, 111, 120, 107, 108, r1= -10.0, r2 = 0.0, r3= 0.0, r4= 10.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 75, 88, 87, 85, 107, 120, 119, 117, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 88, 87, 85, 120, 119, 117, 118, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 80, 84, 85, 116, 112, 117, 118, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 78, 79, 80, 81, 110, 111, 112, 113, r1= -10.0, r2 = 0.0, r3= 0.0, r4= 10.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 79, 80, 81, 120, 111, 112, 113, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 85, 84, 80, 81, 117, 116, 112, 113, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 79, 80, 81, 111, 112, 113, 114, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 84, 80, 81, 83, 116, 112, 113, 115, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 25, 13, 14, 15, 55, 45, 46, 47, 151, 139, 140, 141, 181, 171, 172, 173, r1= 170.0, r2= 180.0, r3= 180.0, r4= 190.0, rk2 = 1000.0, rk3 = 1000.0, &end
&rst iat= 12, 11, 10, 44, 43, 42, 74, 73, 72, 106, 105, 104, 138, 137, 136, 170, 169, 168  r1= 104.0,  r2= 109.0,  r3= 110.0,  r4= 115.0,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 12, 11, 13, 44, 43, 45, 138, 137, 139, 170, 169, 171,  r1= 103.0,  r2= 108.0,  r3= 109.0,  r4= 114.0,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 74, 73, 106, 105, 107,  r1= 102.9,  r2= 107.9,  r3= 108.9,  r4= 113.9,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 30, 29, 60, 59, 43, 92, 91, 73, 124, 123, 105, 156, 155, 137, 186, 185, 169,  r1= 106.5,  r2= 111.5,  r3= 112.5,  r4= 117.5,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 30, 29, 27, 60, 59, 57, 92, 91, 89, 124, 123, 121, 156, 155, 153, 186, 185, 183,  r1= 105.0,  r2= 110.0,  r3= 111.0,  r4= 116.0,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 28, 27, 29, 58, 57, 90, 89, 122, 121, 123, 154, 153, 155, 184, 183, 185,  r1= 104.0,  r2= 109.0,  r3= 110.0,  r4= 115.0,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 28, 27, 8, 58, 57, 40, 90, 89, 70, 122, 121, 102, 154, 153, 134, 184, 183, 166,  r1= 103.3,  r2= 108.3,  r3= 109.3,  r4= 114.3,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 28, 27, 32, 58, 57, 62, 90, 89, 94, 122, 121, 126, 154, 153, 158, 184, 183, 188,  r1= 105.3,  r2= 110.3,  r3= 111.3,  r4= 116.3,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 9, 8, 27, 41, 40, 57, 71, 70, 89, 103, 102, 121, 135, 134, 153, 167, 166, 183,  r1= 102.0,  r2= 107.0,  r3= 108.0,  r4= 113.0,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 9, 8, 5, 41, 40, 37, 71, 70, 67, 103, 99, 135, 134, 131, 167, 166, 163,  r1= 102.5,  r2= 107.5,  r3= 108.5,  r4= 113.5,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 9, 8, 10, 41, 40, 42, 71, 70, 72, 103, 102, 104, 135, 134, 136, 167, 166, 168,  r1= 102.5,  r2= 107.5,  r3= 108.5,  r4= 113.5,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 11, 29, 43, 59, 73, 91, 105, 123, 137, 155, 169, 185,  r1= 1.03,  r2= 1.53,  r3= 1.53,  r4= 2.03,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 9, 8, 27, 41, 40, 57, 71, 70, 89, 103, 102, 121, 135, 134, 153, 184, 183, 185,  r1= 1.04,  r2= 1.54,  r3= 1.54,  r4= 2.04,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 27, 8, 57, 40, 89, 70, 121, 102, 153, 134, 183, 166,  r1= 1.03,  r2= 1.53,  r3= 1.53,  r4= 2.03,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 27, 32, 57, 62, 89, 94, 121, 126, 153, 158, 183, 188,  r1= 0.92,  r2= 1.42,  r3= 1.42,  r4= 1.92,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 8, 10, 40, 42, 70, 72, 102, 104, 134, 136, 166, 168,  r1= 0.93,  r2= 1.43,  r3= 1.43,  r4= 1.93,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 10, 11, 42, 43, 72, 73, 104, 105, 136, 137, 168, 169,  r1= 0.93,  r2= 1.43,  r3= 1.43,  r4= 1.93,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 8, 5, 40, 37, 70, 67, 102, 99, 134, 131, 166, 163,  r1= 1.04,  r2= 1.54,  r3= 1.54,  r4= 2.04,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 5, 4, 37, 36, 67, 66, 99, 98, 131, 130, 163, 162,  r1= 0.92,  r2= 1.42,  r3= 1.42,  r4= 1.92,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 4, 1, 36, 33, 66, 63, 98, 95, 130, 127, 162, 159,  r1= 1.11,  r2= 1.61,  r3= 1.61,  r4= 2.11,  rk2= 1000.0,  rk3= 1000.0,  &end
&rst iat= 33, 32, 63, 62, 95, 94, 127, 126, 159, 158,  r1= 1.10,  r2= 1.60,  r3= 1.60,  r4= 2.10,  rk2= 1000.0,  rk3= 1000.0,  &end
Appendix D

Supplementary material for Chapter 4

Table S1: RT and PCR primers used in the study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Oligodeoxyribonucleotide sequence (5’ – 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-21 RT primer</td>
<td>GTCGTATCCAGTGCGAGGGTCCGAGGTATTCCGACTGGGATACGACTCAACATCAG</td>
</tr>
<tr>
<td>let-7g RT primer</td>
<td>GTCGTATCCAGTGCGAGGGTCCGAGGTATTCCGACTGGGATACGACAACTGTACAA</td>
</tr>
<tr>
<td>miR-17 RT primer</td>
<td>GTCGTATCCAGTGCGAGGGTCCGAGGTATTCCGACTGGATACGACCTACCTGCAC</td>
</tr>
<tr>
<td>miR-18a RT primer</td>
<td>GTCGTATCCAGTGCGAGGGTCCGAGGTATTCCGACTGGATACGACTATCTGCACT</td>
</tr>
<tr>
<td>U6 RT primer</td>
<td>GTCGTATCCAGTGCGAGGGTCCGAGGTATTCCGACTGGATACGACAAAAATATGGAACG</td>
</tr>
<tr>
<td>Rpl30 RT primer</td>
<td>AAGGATAACCAACTTCCGCCTTTG</td>
</tr>
<tr>
<td>miR-21 sense</td>
<td>AGACTAGCTTATCAGACTGA</td>
</tr>
<tr>
<td>let-7g sense</td>
<td>AACGCTGAGGTAGTGTGTGT</td>
</tr>
<tr>
<td>miR-17 sense</td>
<td>AGACAAAAGTGCTTACAGTGCA</td>
</tr>
<tr>
<td>miR-18a sense</td>
<td>GCGTAAGGTGCTACTAGTG</td>
</tr>
<tr>
<td>U6 sense</td>
<td>CTGGCTTCGGCAGCACA</td>
</tr>
<tr>
<td>Universal antisense for miR-21, let-7g, miR-17, miR-18a and U6</td>
<td>GTGCAGGGTCCGAGGT</td>
</tr>
<tr>
<td>Rpl30 sense</td>
<td>CTCTCTCTGTCTCCTGTGTAT</td>
</tr>
<tr>
<td>Rpl30 antisense</td>
<td>AAGGATAACCAACTTCCGCCTTTG</td>
</tr>
</tbody>
</table>
Table S2: Hybridization of the conjugates with miR-21 and efficiency and specificity of miR-21 cleavage

<table>
<thead>
<tr>
<th>Conjugate</th>
<th>Binding extent, % 1)</th>
<th>Cleavage sites</th>
<th>Total cleavage extent, % 2)</th>
<th>Total cleavage extent, % 3)</th>
<th>Cleavage at specific sites, % 3)</th>
<th>Half-life, h</th>
</tr>
</thead>
<tbody>
<tr>
<td>3'-16</td>
<td>100</td>
<td>G\textsubscript{3}C\textsubscript{4}</td>
<td>0</td>
<td>0</td>
<td>G\textsubscript{3}C\textsubscript{4} (2%)</td>
<td>–</td>
</tr>
<tr>
<td>3'-h-6/16</td>
<td>99</td>
<td>G\textsubscript{3}C\textsubscript{4}</td>
<td>0</td>
<td>0</td>
<td>G\textsubscript{3}C\textsubscript{4} (2%)</td>
<td>–</td>
</tr>
<tr>
<td>3'-h-9/16</td>
<td>96</td>
<td>G\textsubscript{3}C\textsubscript{4}</td>
<td>0</td>
<td>0</td>
<td>G\textsubscript{3}C\textsubscript{4} (2%)</td>
<td>–</td>
</tr>
<tr>
<td>3'-h-9/10</td>
<td>0.1</td>
<td>G\textsubscript{3}C\textsubscript{4}</td>
<td>0</td>
<td>0</td>
<td>G\textsubscript{3}C\textsubscript{4} (2%)</td>
<td>–</td>
</tr>
<tr>
<td>5'-h-6/14</td>
<td>100</td>
<td>G\textsubscript{15}A\textsubscript{16}, G\textsubscript{18}U\textsubscript{19}, G\textsubscript{21}A\textsubscript{22}</td>
<td>87 %</td>
<td>99 %</td>
<td>G\textsubscript{15}A\textsubscript{16} (68%), G\textsubscript{18}U\textsubscript{19} (30%), G\textsubscript{21}A\textsubscript{22} (1%)</td>
<td>10.0±0.1</td>
</tr>
<tr>
<td>5'-h-9/14</td>
<td>100</td>
<td>G\textsubscript{15}A\textsubscript{16}, G\textsubscript{18}U\textsubscript{19}, G\textsubscript{21}A\textsubscript{22}</td>
<td>70 %</td>
<td>98 %</td>
<td>G\textsubscript{15}A\textsubscript{16} (48%), G\textsubscript{18}U\textsubscript{19} (46%), G\textsubscript{21}A\textsubscript{22} (4%)</td>
<td>17±0.4</td>
</tr>
<tr>
<td>5'-16</td>
<td>100</td>
<td>G\textsubscript{15}A\textsubscript{16}, G\textsubscript{18}U\textsubscript{19}, G\textsubscript{21}A\textsubscript{22}, G\textsubscript{3}C\textsubscript{4}</td>
<td>71 %</td>
<td>93 %</td>
<td>G\textsubscript{3}C\textsubscript{4} (5%), G\textsubscript{15}A\textsubscript{16} (4%), G\textsubscript{18}U\textsubscript{19} (78%), G\textsubscript{21}A\textsubscript{22} (6%)</td>
<td>15.1±0.2</td>
</tr>
<tr>
<td>5'-h-6/16</td>
<td>99</td>
<td>G\textsubscript{18}U\textsubscript{19}, G\textsubscript{21}A\textsubscript{22}</td>
<td>22 %</td>
<td>83 %</td>
<td>G\textsubscript{18}U\textsubscript{19} (63%), G\textsubscript{21}A\textsubscript{22} (20%)</td>
<td>59.9±2.0</td>
</tr>
<tr>
<td>5'-h-9/16</td>
<td>97</td>
<td>G\textsubscript{18}U\textsubscript{19}, G\textsubscript{21}A\textsubscript{22}</td>
<td>20%</td>
<td>57 %</td>
<td>G\textsubscript{18}U\textsubscript{19} (33%), G\textsubscript{21}A\textsubscript{22} (24%)</td>
<td>43.7±1.2</td>
</tr>
</tbody>
</table>

1) [miR-21]=1 µM, [conjugate]=5 µM, 37 °C, 20 min

2) [miR-21]=1 µM, [conjugate]=20 µM, 37 °C, 24 h

3) [miR-21]=1 µM, [conjugate]=20 µM, 37 °C, 72 h
**Figure S1**: Representative $^1$H-NMR spectra of free peptide acetyl-(LeuArg)$_2$-Gly-(LeuArg)$_2$-Gly-COOH (A) and conjugate 5'-16 (B), showing the addition of characteristic resonance signals from TCAGTCTGATAAGCTA oligonucleotide. Spectra were recorded in D$_2$O at 25°C using Bruker Avance II+ 400 Ultrashield spectrometer.
**Figure S2:** Cleavage of *miR-21* by 3’-conjugates 3’-16, 3’-h-6/16, 3’-h-9/16, 3’-h-9/10. Autoradiograph of 18% polyacrylamide/8 M urea gel. Lanes Im and T1, imidazole ladder and partial RNA digestion with RNase T1, respectively. Control – RNA incubated in the absence of conjugate for 0 and 144 h. *miR-21* (1 µM), 3’-conjugates (50 µM) were incubated at 37 °C for 0 – 144 h. The conjugate type and incubation time are shown at the top.
Figure S3: Accumulation of miR-21 products after hydrolysis by conjugates 5'-h-6/14 (A), 5'-h-9/14 (B), 5'-h-6/16 (C), 5'-h-9/16 (D) and 5'-16 (E) at individual sites. *miR-21 (1 µM) and conjugates (20 µM) were incubated at 37 °C for 0 – 72 h.
Figure S4: Backbone RMSD (in Å) of 5'-h-6/14 complex with miR-21 over course of one microsecond MD trajectory in implicit solvent.

Figure S5: Snapshots from trajectory of 5'-h-6/14 complex with miR-21 over 1 µs of molecular dynamics simulation in implicit solvent. Peptide in yellow, oligonucleotide in green, miR-21 in blue.
Figure S6: Time series of (A) Arg10 and (B) Arg5 interactions with phosphorus atom of G\textsubscript{15}\textendashA\textsubscript{16} linkage.

Figure S7: Time series of Arg10 interactions with base atoms of G\textsubscript{15}

Figure S8: Time series of Arg8 interactions with base atoms of G\textsubscript{15}
Figure S9: A putative in-line orientation of 2'-OH group at G_{18}-U_{19} linkage formed by 5'-h-6/14 complex with miR-21 during 1 μs molecular dynamics simulation.
Figure S10: Flow cytometry analysis of the accumulation of FITC-labeled peptide-oligonucleotide conjugate 5'-h-9/14 in RLS40 and KB-8-5 cells. (A) and (C) – RLS40 and KB-8-5, transfected with FITC-labeled peptide-oligonucleotide conjugate 5'-h-9/14 (1 µM) precomplexed with Lipofectamine 2000; (B) and (D) RLS40 and KB-8-5 cells incubated with Lipofectamine 2000 only. Experimental procedure was as follows: 4 h post-transfection cells were washed with saline solution and fixed in 2% formaldehyde in PBS. Cells were analyzed using Cytomics FC 500 (Beckman Coulter, USA) flow cytometer (excitation wave length 488 nm, emission 530±30 nm). Two parameters reflecting the efficiency of the process were used for the comparison: the transfection efficiency estimated as the percentage of cells with green fluorescence exceeded the maximum level of the auto-fluorescence of untreated cells and the mean fluorescence intensity of cells measured in relative fluorescent units (‘RFU’). The MFI for RLS40 and KB-8-5 cells was 12.6 and 40 RFU, respectively.
**Figure S11:** Confocal microscopy study of cellular accumulation of FITC-labeled peptide-oligonucleotide conjugate 5'-h-6/14 precomplexed with Lipofectamine 2000. **A.** Control KB-8-5 cells incubated with Lipofectamine 2000 only. **B.** KB-8-5 cells, transfected by FITC-5'-h-6/14 precomplexed with Lipofectamine 2000. [N.B. Human KB-8-5 cells were deliberately used for confocal microscopy: these are adherent cells forming stable monolayers, while RLS<sub>40</sub> cells are suspension cell lines, which are known to produce poor samples for confocal microscopy].

Experimental procedure was as follows: KB-8-5 cells in the exponential phase of growth were plated on the glass coverslips in 24-well plates at a density of 1.3 × 10<sup>5</sup> cells/well one day before the experiment. FITC-labeled conjugate 5'-h-6/14 precomplexed with Lipofectamine 2000 was then added to the cells to final concentration 1 µM. In 4 h after transfection the cells were washed with PBS, fixed in 2% formaldehyde/PBS, permeabilized in 0.1% Triton-X100/PBS. The cells on the coverslips were then mounted on the object-plate in DAPI/Antifade solution (Millipore, USA). The analysis of FITC-labeled 5'-h-9/14 localization was performed using a confocal fluorescent microscope LSM 710 Meta (Carl Zeiss, Germany) at 60 × Magnification. Images (right column) merge two channels: blue (cell nuclei stained with DAPI, left column) and green (peptide-oligonucleotide conjugate 5'-h-6/14 labeled with Fluorescein, middle column).
Figure S12: Expression level of miR-21 in RLS40 cells after transfection of anti-miR-21 conjugate 5'-h-6/14 or with control conjugate luc-h-9/14. The conjugate concentration varied from 0.1 µM to 5 µM. Transfection was performed using Lipofectamine 2000. LF – RLS40 cells incubated with Lipofectamine 2000 only. The expression of miR-21 was normalized to U6. Data are given as mean calculated from three independent experiments ± SEM.

Figure S13: Expression level of miR-17 and miR18a in RLS40 cells after transfection with 1 µM of control conjugate luc-h-9/14, anti-miR-21 antisense oligonucleotide 5'-h-6/14 and anti-miR-21 conjugate 5'-h-6/14. Transfection was performed using Lipofectamine 2000. LF – RLS40 cells incubated with Lipofectamine 2000 only. The expression of miRNAs was normalized to U6 and Rpl30. Data are given as mean calculated from three independent experiments ± SEM.
Figure S14: Stability of asON 5'-16 and conjugate 5'-16 in 10% FBS. A. Image of 15% PAAG/8M urea, stained with Stains-All. asON or conjugate was incubated in 10% FBS at 37 °C for 0, 2, 4, 8, 24 and 48 h. B. Kinetics of degradation of asON or conjugate in 10% FBS.

Figure S15: Cleavage of miR-21/oligonucleotide(asON) and miR-21/conjugate duplexes by RNase H. *miR-21 (1 μM) and oligonucleotide 5'-h-9/14 (5 or 30 μM) or conjugate 5'-h-9/14 (5 or 30 μM) were incubated at 37 °C for 0 and 30 min. Lanes Im and T1, imidazole ladder and partial RNA digestion with RNase T1, respectively. C – RNA incubated in the absence of oligonucleotide, conjugate or RNase H for 30 min.