

**Potential prebiotic roles of (amino-)acylation in the synthesis  
and function of RNA**

2013

Christopher Ken Wai Chan

Supervised by Professor John D Sutherland

A dissertation submitted for the degree of Doctor of Philosophy at the MRC Laboratory  
of Molecular Biology

Gonville and Caius College

University of Cambridge

# Contents

<b>Summary</b>	<b>5</b>
<b>Declaration</b>	<b>6</b>
<b>Acknowledgements</b>	<b>7</b>
<b>Abbreviations</b>	<b>8</b>
<b>Numbering and Nomenclature</b>	<b>11</b>
<b>1. Introduction</b>	<b>12</b>
1.1. What prebiotic chemistry hopes to achieve	12
1.2. The early Earth and the prebiotic environment	12
1.3. Modern biology and a common ancestry	14
1.4. Theories for the origin of life	17
1.4.1. Autotrophic origin of life	18
1.4.2. Heterotrophic origin of life	19
1.4.3. The RNA world hypothesis	19
1.5. Chemistry towards an abiogenesis of RNA and proteins	21
1.5.1. The Miller-Urey experiment	22
1.5.2. The traditional disconnection of RNA	23
1.5.3. Recent successes: synthesis of the activated pyrimidine nucleotides bypassing preformed ribose	26
1.6. Abiotic synthesis of polymeric RNA	32
1.6.1. Oligomerisation of activated 5'-nucleotides	33
1.6.2. Oligomerisation of nucleoside-2',3'-cyclic phosphates	36
1.6.3. Oligonucleotide ligation facilitated by chemoselective acetylation	41
1.7. From RNA towards peptides	46
1.7.1. The search for a primitive aminoacylation	48
1.7.2. A linked prebiotic origin of RNA and coded peptides	53
1.8. Project aims	59
<b>2. Solid-phase synthesis of 2'/3'-O-acetylated RNA oligonucleotides</b>	<b>60</b>
2.1. Background	60
2.1.1. Incompatibilities of conventional RNA oligonucleotide synthesis	60
2.1.2. An acetyl compatible protecting group strategy	66
2.2. Synthesis of the phosphoramidites	69

2.2.1.	Proposed synthetic route to the phosphoramidites	69
2.2.2.	Nucleobase protection	71
2.2.3.	Synthesis of the 2'/3'- <i>O</i> -acetyl RNA phosphoramidites	77
2.2.4.	Synthesis of the 2'/3'- <i>O</i> -TBDMS phosphoramidites	87
2.3.	Synthesis of the photolabile linker and preparation of the solid-support	89
2.4.	Solid-phase synthesis of acetyl-RNA	96
2.4.1.	Optimisation of the automated synthesis of acetyl-RNA oligonucleotides	96
2.4.2.	Method optimisation for the deprotection, cleavage and purification of the acetylated-RNA oligonucleotides	102
2.5.	Synthesis of acetyl-RNA oligonucleotides	107
<b>3.</b>	<b>Properties of 2'/3'-<i>O</i>-acetyl RNA oligonucleotides and consequences for the non-enzymatic replication of RNA</b>	<b>110</b>
3.1.	Background	110
3.2.	Duplex stability of acetyl-RNA assessed by UV melting analysis	112
3.2.1.	Consequences of acetyl-RNA for the non-enzymatic replication of RNA	120
3.3.	Stability of an acetylated hairpin structure	123
3.4.	Future work and conclusions	129
3.4.1.	Disruption of A-minor and tertiary structure	130
3.4.2.	Summary	132
<b>4.</b>	<b>Potentially prebiotic aminoacylation of RNA</b>	<b>134</b>
4.1.	Background	134
4.1.1.	The Iron-Sulphur World and the prebiotic plausibility of amino thioacids	137
4.2.	Organic synthesis of amino thioacids	139
4.3.	Prebiotically plausible aminoacylation of nucleoside phosphates with amino thioacids	140
4.3.1.	Aminoacylation with thiovaline 134 and cyanoacetylene 7	140
4.3.2.	Aminoacylation with thiovaline 134 and alternative electrophilic activators	146

4.3.3. Aminoacylation of oligonucleotides with thiovaline and cyanoacetylene	152
<b>5. Conclusions</b>	<b>156</b>
<b>6. Experimental</b>	<b>158</b>
6.1. General procedures	158
6.2. Experimental for Chapter 2	161
6.2.1. Synthetic procedures for the synthesis of the phosphoramidites	161
6.2.2. Preparation of the Solid-Phase Support	201
6.2.3. Synthesis of oligonucleotides	203
6.3. Procedures for Chapter 3	208
6.4. Procedures for Chapter 4	214
6.4.1. Synthetic procedures for materials used in the aminoacylation reactions	214
6.4.2. Procedures for aminoacylation using amino thioacids and various electrophilic activators	218
<b>7. References</b>	<b>237</b>

## Summary

The Sutherland group recently demonstrated that from a mixture of oligoribonucleotide-2'- or 3'-phosphates the latter is chemoselectively acetylated. This is shown to mediate a template-directed ligation to give predominantly 3',5'-linked RNA that is acetylated at the ligation junction (acetyl-RNA). It was suggested that RNA emerged prebiotically *via* acetyl-RNA and also is proposed to have favourable genotypic properties due to greater propensity to form duplex structure. To study the properties of acetyl-RNA, their synthesis by solid-phase chemistry was required and described is the design of a 2'/3'-*O*-acetyl orthogonal protecting group strategy. Key to the orthogonal protecting group strategy is the use of (2-cyanoethoxy)carbonyl for the protection of the nucleobase exocyclic amines and a photolabile solid-phase linker group that allowed partial on-column deprotection. The synthesis of the 2'/3'-*O*-acetyl and 2'/3'-*O*-TBDMS phosphoramidites, in addition to preparation of a photolabile solid-phase support, are described. With the materials to hand the procedures for an automated synthesis of acetyl-RNA were optimised and several acetyl-RNA oligonucleotides were synthesised.

The duplex stability of acetyl-RNA with up to four sites of 2'-*O*-acetylation were assessed by UV melting curve analysis. Remarkably, the acetyl groups caused a consistent decrease in  $T_m$  of between 3.0-3.2 °C. Thermodynamic parameters indicated a decrease in duplex stability that was consistent with a decrease in hydration of the minor groove resulting in a reduction of the stabilising hydrogen bonding network. The stability of a tetraloop was also found to decrease on acetylation. The acetylated-tetraloop it is able to form duplex at lower concentrations than the natural tetraloop. Additionally, it is more stable at high concentrations, indicating that acetyl-RNA favours duplex over other secondary structure. These properties are considered to give acetyl-RNA competitive advantage for their non-enzymatic replication.

Aminoacylation of RNA is an important process in modern biology but the intermediacy of aminoacyl-adenylates is considered to be prebiotically implausible. A potentially prebiotic aminoacylation of nucleoside-3'-phosphates, selective for the 2'-hydroxyl, is presented. However, it was thought the aminoacylation yields could be improved and so a search for an alternative activator was conducted. Oligoribonucleotide-3'-phosphates were exposed to the aminoacylation conditions and selective aminoacylation at only the 2'-hydroxyl of the 3'-end was observed. In particular, the aminoacylation of a trimer lends support to Sutherland's theory of a linked origin of RNA and coded peptide synthesis.

## **Declaration**

This dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated in the text.

This dissertation does not exceed the word limit set by the biology degree committee.

Word count: 50,960

## **Acknowledgements**

First and foremost I would like to thank my supervisor, John Sutherland, for his guidance, knowledge and enthusiasm, which have been invaluable throughout my studies. It has been a roller coaster in many ways but I'm grateful for the opportunity that he gave me.

I am indebted to the tireless efforts of Frank; his proofreading and suggestions have made this thesis far better than it otherwise could have been. I extend many thanks to Colm and Jianfeng 'Jimmy' for their collaboration in the work described in Chapter 2 – their efforts have helped bring this work to fruition. I have also had the pleasure of working with Béatrice, Lello, Lee, Claire, Dougal, Bhavesh, Claudia and Sam. I am particularly grateful to Béatrice for her counsel during my first year and Lello for providing ample amusement and entertainment – I thank you all.

My biggest thanks must go to my family, especially my parents, for their continual support and encouragement in whichever path I take. And finally, to the long suffering Christina for your unshakable patience, encouragement, support, inspiration and love that have carried me through the good and bad times!

## Abbreviations

A	adenine
Ac	acetyl
Ar	aryl
ARS	aminoacyl-tRNA synthetase
ADP	adenosine diphosphate
ATP	adenosine triphosphate
aq.	aqueous
B	nucleic acid base
Boc	<i>tert</i> -butyloxycarbonyl
<sup>t</sup> Bu	<i>tert</i> -butyl
BTT	5-benzylthio-1H-tetrazole
°C	degrees Celsius
C	cytosine
<i>ca.</i>	<i>circa</i>
calc.	calculated
ce	cianoethyl
ceoc	(2-cyanoethoxy)carbonyl
CIP	calf intestinal phosphatase
cm <sup>-1</sup>	wavenumber
CoA	Coenzyme A
conc.	concentrated
COSY	correlated spectroscopy (NMR)
CPG	controlled pore glass
δ	chemical shift
DAMN	diaminomaleonitrile
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DCI	4,5-dicyanoimidazole
DIAD	<i>N,N'</i> -diisopropylazodicarboxylate
DIPEA	diisopropylethylamine
DMAP	4-(dimethylamino)-pyridine
DMF	<i>N,N</i> -dimethylformamide
dmf	dimethylformamidine
DMSO	dimethylsulfoxide
DMTr	4,4'-dimethoxytrityl
DNA	deoxyribonucleic acid
Eds.	Editors
<i>ee</i>	enantiomeric excess
ESI	electrospray ionization
Est.	estimated
<i>et al.</i>	<i>et alia</i>
ETT	5-ethylthio-1H-tetrazole
eq.	equivalent(s)
FADH	flavin adenine dinucleotide
G	guanine
h	hour(s)
HMDS	hexamethyldisilazane
hν	electromagnetic irradiation (UV)

HPLC	high performance liquid chromatography
Hz	Hertz
<i>i</i>	<i>iso</i>
IBCF	isobutyl chloroformate
<i>i.e.</i>	<i>id est</i>
IR	infrared
<i>J</i>	NMR coupling constant measured in Hertz
LCAA	long chain alkylamine
lit.	literature (reference)
m	milli
M	molar
MALDI-TOF	Matrix-assisted laser desorption/ionization-time of flight
Me	methyl
MeCN	acetonitrile
MHz	megahertz
min	minute
mL	millilitre
mmol	millimole
M.P.	melting point
mRNA	messenger ribonucleic acid
MS	mass spectrometry
μL	microliter
μM	micromolar
<i>m/z</i>	mass/charge ratio
NADH	nicotinamide adenine dinucleotide
NAI	<i>N</i> -acetylimidazole
NCI	<i>N</i> -cyanoimidazole
NCA	<i>N</i> -carboxyanhydride
NMR	nuclear magnetic resonance
NP	normal phase
npe	<i>p</i> -nitrophenylethyl
<i>p</i>	<i>para</i>
PBS	phosphate buffered saline
Ph	phenyl
P <sub>i</sub>	inorganic phosphate
PP <sub>i</sub>	inorganic pyrophosphate
ppm	parts per million
py.	pyridine
quant.	quantitative yield
R	unspecified group
<i>rac-</i>	racemic mixture
RP	reverse phase
rRNA	ribosomal ribonucleic acid
RNA	ribonucleic acid
RT	room temperature
sat.	saturated
<i>sca-</i>	scalemic
soln.	solution
<i>t</i>	tertiary
<i>tert</i>	tertiary

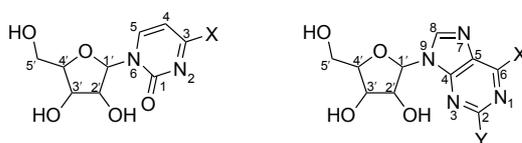
T	thymine
TBDMS	<i>tert</i> -butyldimethylsilyl
TCA	trichloroacetic acid
TFA	trifluoroacetic acid
THF	tetrahydrofuran
T <sub>m</sub>	melting temperature
TMS	trimethylsilyl
TLC	thin layer chromatography
TREAT.HF	triethylamine trihydrofluoride
tRNA	transfer ribonucleic acid
t <sub>1/2</sub>	half life
U	uracil
UMP	uridine-5'-monophosphate
UV	ultraviolet

# Numbering and Nomenclature

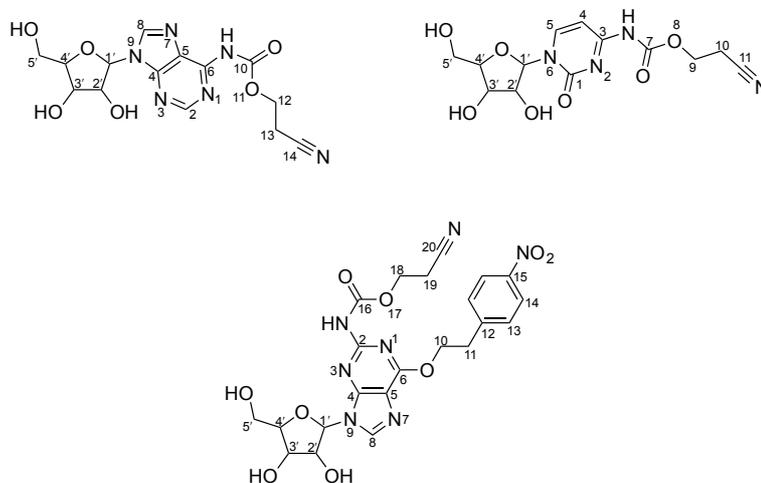
## Nucleobases



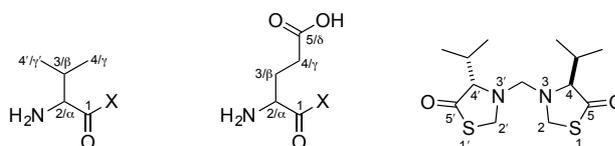
## Nucleosides



## Protected nucleosides



## Amino acid numbering



# 1. Introduction

## 1.1. What prebiotic chemistry hopes to achieve

The question of “How did life begin?” has been discussed and debated over the millennia by religion, philosophers and scientists, but this question is inherently complex as life is not easy to define. Luisi and Abel point out that no two definitions are exactly the same, although the most popular seems to be NASA’s official definition:<sup>[1-3]</sup>

“Life is a self-sustaining chemical system capable of Darwinian evolution”.

If physicists investigate the beginnings of the universe and biologists reduce the complexity of known life to its minimal requirements, then it is chemists who must discover how a “self-sustaining chemical system” could have emerged from inanimate chemicals. In research towards the origin of life there are many variables that have to be taken into account, such as where life started, the identity of starting materials and planetary conditions. We can define, to a degree of certainty, the molecules that are required for life such as RNA, peptides, lipids and essential metabolites. But with no direct clues to the conditions on the early Earth at the origin of life there is an immensely wide field of possible starting points and paths to follow. And so the chemist has to demonstrate how life *may* have occurred. This view is shared by Albert Eschenmoser who summed up this sentiment perfectly:<sup>[4]</sup>

“The origin of life cannot be discovered, it has to be re-invented”.

## 1.2. The early Earth and the prebiotic environment

The Earth was formed around 4.5 billion years ago from the gravitational aggregation of cosmic gas clouds and dust orbiting the Sun. During this process the gravitational forces would have resulted in immense heat and a molten surface. Also, up until about 3.9 billion years ago, large asteroids frequently impacted the earth and these would have sterilised the surface.<sup>[2, 5, 6]</sup> Although there is some debate over validity,<sup>[7]</sup> it is generally accepted that fossils found in western Australia are of organisms that resemble cyanobacteria, which have been dated to approximately 3.5 billion years ago.<sup>[8]</sup> Other fossil evidence puts the existence of photosynthetic cyanobacteria to a time point of

around 2.7 billion years ago.<sup>[9]</sup> These fossils suggest some sort of cellular life was present between 3.5 and 2.7 billion years ago, and the transition from non-living chemicals to living entities must have occurred within a relatively short time period of 400 million years after meteoric bombardment had ceased (Figure 1).

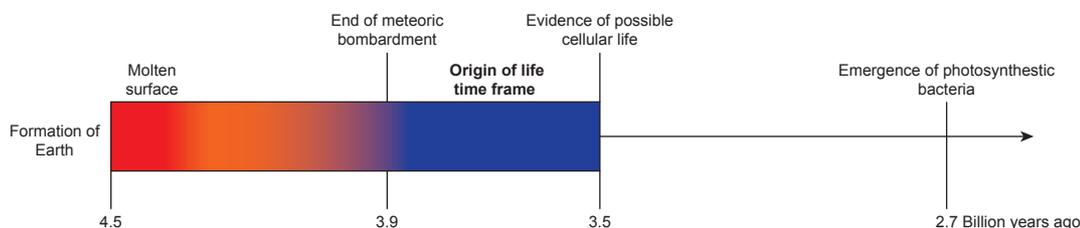


Figure 1. Timescale for the emergence of life.

The atmosphere at the beginning of life was thought to have been weakly reducing or close to neutral and contained mostly carbon dioxide and nitrogen with traces of carbon monoxide, hydrogen and reduced sulphur gases.<sup>[5, 10]</sup> Oxygen would not have been present until the dawn of biological activity that resulted in a slow but not immediate rise of oxygen, and so a protective ozone layer would have formed after the emergence of life.<sup>[10]</sup> The identity of reactive molecules on the early earth is not known but observations from the atmosphere of the gas giants,<sup>[11, 12]</sup> spark discharge/UV experiments,<sup>[13, 14]</sup> and analysis of carbonaceous meteorites after arrival on Earth<sup>[15, 16]</sup> suggest that the feedstock molecules in Figure 2 were important for prebiotic chemistry. On cooling of the Earth, water vapour would have condensed to form the oceans. After cooling sufficiently they were likely to have been at a neutral pH due to the buffering action of basalt and other minerals.<sup>[17]</sup>

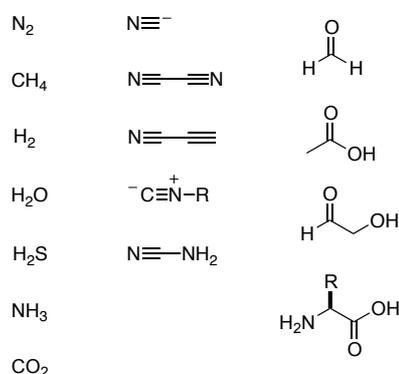


Figure 2. A selection of potentially prebiotic feedstock molecules.

### 1.3. Modern biology and a common ancestry

Life is overwhelmingly diverse, complex and interdependent. Despite this complexity, all life is related by the ‘Central Dogma of Molecular Biology’, which was put forward by Crick and underpins modern biology (Figure 3).<sup>[18]</sup>

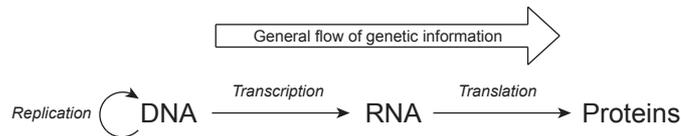


Figure 3. The 'Central Dogma of Molecular Biology'.

The Central Dogma describes a flow of genetic information that, once translated into proteins, cannot flow back towards nucleic acids. The processes described in Figure 3 are those that are common to most cells and animals, yet there are some exceptions such as retroviruses whose genetic data is held as RNA. This class of virus, which includes human immunodeficiency virus 1 (HIV-1), possess an enzyme called a reverse transcriptase that copies the viral RNA into DNA within a host.<sup>[19]</sup>

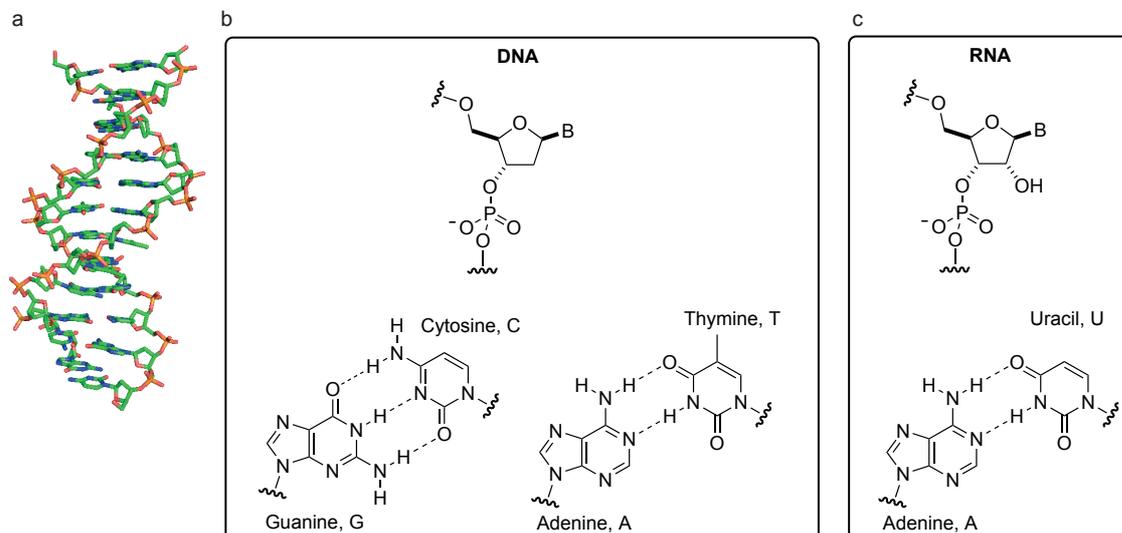


Figure 4. a) B-form DNA double helix, redrawn from PDB file 1bna using MacPyMOL.<sup>[20]</sup> b) The sugar-phosphate backbone structure of DNA and the Watson-Crick base pairing of the nucleobases. c) The sugar-phosphate backbone structure of RNA and the Watson-Crick base pairing of A:U that replaces A:T. B = nucleobases

In all living beings, genetic data is stored within the double helix of DNA, but this information must be copied before it can be passed on to progeny.<sup>[21]</sup> The double helix

of DNA comprises two anti-parallel strands, each with a deoxyribose-phosphate backbone located on the exterior and the nucleobases are found within the interior (Figure 4a). Genetic information is coded along a strand by four nucleobases, two purines (adenine (A) and guanine (G)) and two pyrimidines (thymine (T) and cytosine (C)). It is the specific hydrogen bonds formed between purine and pyrimidine bases (A=T and G=C) that are the key to the hereditary function of DNA, and these are named (Watson-Crick) base pairs (Figure 4b). The replication of DNA requires enzymes called DNA polymerases that utilise the specific base pairing to direct the accurate transfer of genetic information.

Like DNA, RNA is composed of a sugar-phosphate backbone and nucleobases. However, it differs in two aspects; firstly the base thymine (T) is replaced by another pyrimidine, uracil (U) (Figure 4c); secondly the sugar present is ribose that contains a 2'-hydroxyl not present in DNA. During transcription, RNA polymerase enzymes use the DNA strands as templates with the specific base pairing to accurately 'transcribe' the genetic code into RNA. The RNA products of transcription are processed into messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA) that will be used in the next step of information transfer. The mRNA is carried to the cytoplasm where the small and large subunits of the ribosome enclose the 'start' end of the mRNA. The aminoacyl-tRNA that has been previously charged with the correct amino acid according to the anti-codon by an aminoacyl-tRNA synthetase enzyme is then recruited by the ribosome. If the aminoacyl-tRNA anti-codon is not complementary to the codon on the mRNA strand, then the aminoacyl-tRNA falls away. If the aminoacyl-tRNA anti-codon is complementary, the ribosome catalyses the formation of a peptide bond and the amino acid is incorporated into the growing polypeptide (Figure 5). The specific folding of the 1D polypeptide chain into a 3D tertiary structure is the factor that determines the catalytic properties of proteins.

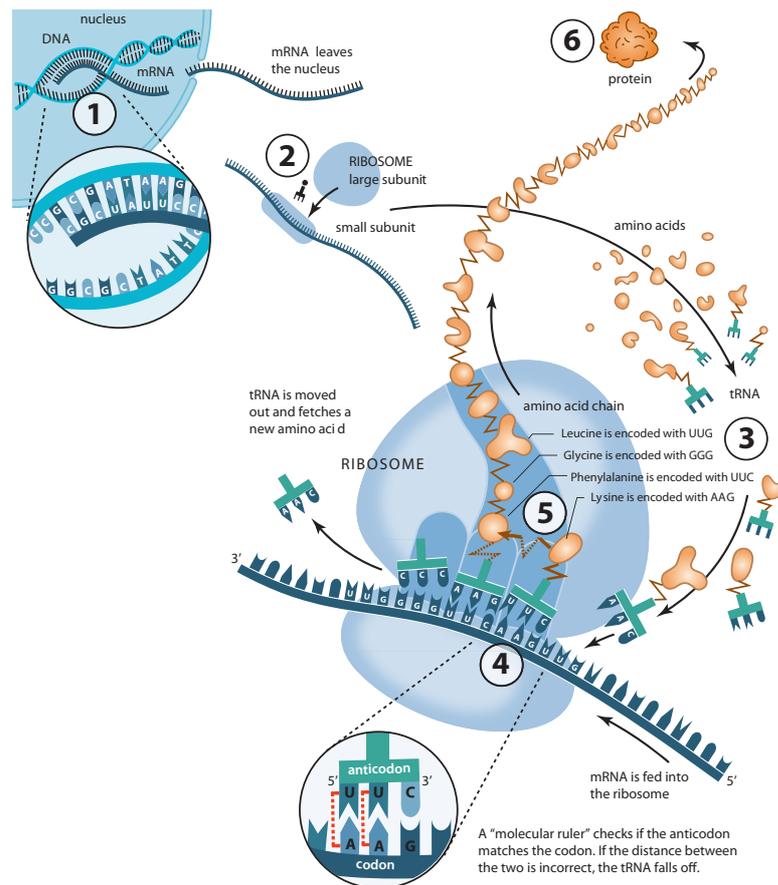


Figure 5. 1) DNA is transcribed to give mRNA. 2) The small and large subunit of the ribosome bind to the mRNA. 3) tRNAs are charged with the correct amino acid that corresponds to the anticodon. 4) At the ribosome the aminoacyl-tRNA binds with the mRNA and if the codon-anticodon does not match the aminoacyl-tRNA is released. 5) If the codon-anticodon match the amino acid is incorporated into the growing polypeptide and the ribosome moves along one codon step to allow another aminoacyl-tRNA to bind. 6) The polypeptide chain is folded into a protein. Adapted from reference<sup>[22]</sup>.

The sequence of three nucleobases in a triplet codon represents a particular amino acid. Gamov initially suggested the triplet codon theory after reading Watson and Crick's discovery of the structure of DNA.<sup>[23]</sup> In 1961, Crick presented preliminary evidence in support of Gamov's three-letter code,<sup>[24]</sup> but it was the efforts of Nirenberg and Khorana that led to the deciphering of the 64 possible codons of the genetic code (Figure 6).<sup>[25-35]</sup> From inspection of the standard genetic code it is clearly highly degenerate, and the amino acids that are coded by two or more codons are called synonyms. Although some slight variations exist between species the genetic code is largely universal.<sup>[19]</sup>

		Second Base				
		G	A	U	C	
G	G	Gly	Asp	Val	Ala	U
		Gly	Asp	Val	Ala	C
		Gly	Glu	Val	Ala	A
		Gly	Glu	Val	Ala	G
A	A	Ser	Asn	Ile	Thr	U
		Ser	Asn	Ile	Thr	C
		Arg	Lys	Ile	Thr	A
		Arg	Lys	Met	Thr	G
U	U	Cys	Tyr	Phe	Ser	U
		Cys	Tyr	Phe	Ser	C
		Stop	Stop	Leu	Ser	A
		Trp	Stop	Leu	Ser	G
C	C	Arg	His	Leu	Pro	U
		Arg	His	Leu	Pro	C
		Arg	Gln	Leu	Pro	A
		Arg	Gln	Leu	Pro	G

Figure 6. The standard genetic code.

Due to its ubiquitous presence in all organisms, ribosomal RNA from many life forms has been sequenced and compared to give clues to the evolutionary relationships between species.<sup>[36-38]</sup> These relationships can be plotted in a so called phylogenetic tree, which shows that the ancestry of all life forms can be traced back to a common beginning (Figure 7). The life form of this common beginning has become known as the ‘Last Universal Common Ancestor’ or *LUCA*.<sup>[39, 40]</sup>

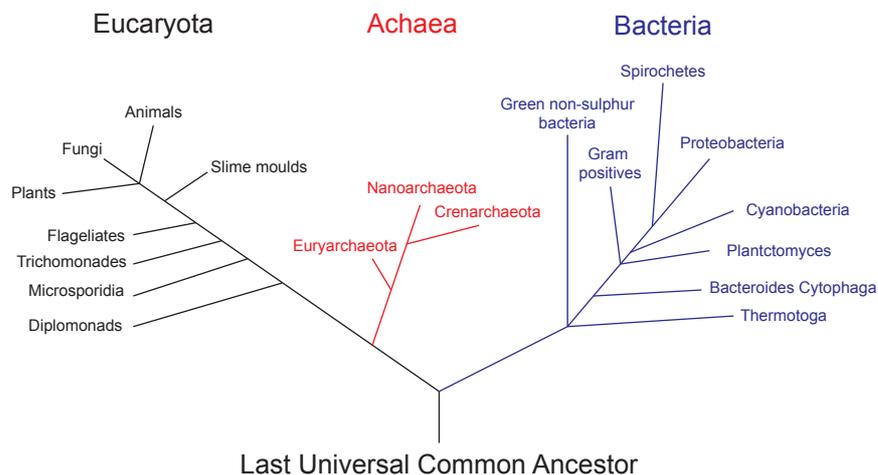


Figure 7. The phylogenetic tree and the three kingdoms of life.<sup>[38]</sup>

## 1.4. Theories for the origin of life

Although life can be traced back to a common ancestor, the question of how such an organism may have emerged from the abiotic mix of feedstock molecules has led

researchers to suggest a more ancient life form. Moreover, DNA and RNA both act as propagators of genetic information, but their replication relies on enzymes, whose structure and function arise from the information carried within the sequence of DNA. This close interdependence has led to suggestions that the beginnings of life were of a much simpler form. The theories put forward are generally divided into two schools of thought, which are the autotrophic and heterotrophic theories of the origin of life. The former favours the emergence of autocatalytic cycles catalysed on inorganic mineral surfaces, because they are considered more likely to have emerged from the early geological environment than the complex organic molecules that are ubiquitous in biology. However, it is the spontaneous formation of these complex organic molecules that the latter heterotrophic group favours, in particular a self-replicating informational molecule that is also endowed with catalytic capacity.<sup>[41]</sup> This thesis assumes a heterotrophic origin of life but does not dismiss the autotrophic theories from which important ideas can be used to piece together a deeper understanding of the origin of life.

#### **1.4.1. Autotrophic origin of life**

One theory of an autotrophic origin of life was proposed by Cairns-Smith who suggested that the beginnings of genotypic evolution could have occurred on the edges of ordered layers of clay minerals such as silicates.<sup>[42]</sup> It was suggested that the interaction of organic materials by adsorption onto mineral surfaces could produce an organic polymer that eventually would take over phenotypic functions (*i.e.* genetic takeover). However, a more popular autotrophic origin of life is the ‘Iron-Sulphur World’ advocated by Wächterhäuser, who suggested that the first organism was an autotroph that derived energy from the conversion of FeS (pyrrhotite) to FeS<sub>2</sub> (pyrite) by H<sub>2</sub>S exhaled from hydrothermal vents or volcanic sites. Using the reductive power of pyrite formation this early organism was said to be able to fix carbon by reducing atmospheric CO and CO<sub>2</sub>.<sup>[43, 44]</sup> This theory is backed by the observation that the metal-sulphur minerals bear resemblance to the FeS and (Fe,Ni)S clusters of corrinoid iron-sulphur protein (CFeSP) and carbon monoxide dehydrogenase-acetyl-CoA-synthase (CODH-ACS).<sup>[45]</sup> These enzymes take part in the Wood-Ljungdahl or reductive acetyl-coenzyme A pathway, which is considered to be a primitive metabolic cycle, and is present in many early branching thermophilic archaea and bacteria.<sup>[46]</sup>

### 1.4.2. Heterotrophic origin of life

Rather than favouring metabolic-cycles, this theory is based upon the gradual build-up of organic materials from the ‘aggregation’ of smaller molecules in the sea or on the surface of the earth. These organic materials are proposed to eventually form nucleic acids, proteins and other substances needed for life.<sup>[47]</sup> The organic material would then self-organise, possibly assisted by energetic species also available in the environment, to eventually result in a replicating life form.<sup>[4]</sup> The advantages of this theory are that it does not require ambiguous ‘genetic-takeover’ steps and is supported by a greater body of experimental work.

However, the nature of how these organic molecules assembled to form a replicating system is a major problem. In particular, the interdependence of replication, transcription and translation raises the classic ‘chicken and egg’ question of “Which came first; nucleic acids or proteins?”. This question was considered in three complementary works by Woese<sup>[48]</sup>, Crick<sup>[49]</sup> and Orgel<sup>[50]</sup> in the late 1960’s, where they suggested that RNA, in addition to propagating genetic information (genotypic), was also able to act catalytically (phenotypic) in a primitive fashion. This was based on the observation that DNA was known to form duplex structure and only acted as a template for RNA replication. Also it was known that ribosomes were composed mainly of RNA and that the adapter molecules tRNA were comprised of only RNA. This idea, however, was not developed until the discovery of catalytic RNA by Cech<sup>[51]</sup> and Altman.<sup>[52]</sup>

### 1.4.3. The RNA world hypothesis

In 1982 Cech *et al.* studied the rRNA genes of *Tetrahymena thermophila*. Within the coding region of the 26S rRNA subunit is a 413 basepair (bp) intervening sequence (IVS). The gene coding for the 26S subunit was transcribed into pre-rRNA that under enzyme-free conditions underwent splicing to remove the IVS. It was concluded that the IVS once transcribed was able to act like an enzyme to break and reform phosphodiester bonds, thus catalysing its own splicing.<sup>[51]</sup> Soon after in 1983, Altman *et al.* were studying the post-translational processing of tRNA by a ribonucleoprotein ribonuclease P. This ribonuclease P was deproteinised and, when incubated with the

correct co-factors and divalent metals, was still able to catalytically cleave the pre-tRNA into tRNA.<sup>[51]</sup> The discovery of these catalytic functions of RNA prompted Gilbert to suggest a time when RNA carried genetic information and was also able to carry out catalytic functions; he coined this ‘The RNA world’.<sup>[53]</sup> He proposed that the RNA was able to develop catalytic activities to enable self-replication. Continued evolution would have enabled RNA to synthesise proteins by first developing adapter RNA molecules like tRNA. The first proteinogenic enzymes were thought to have catalysed the same reactions as ribozymes but with greater efficiency and the final major step would have been transferring the phenotypic responsibility to DNA.

There is strong evidence for the involvement of RNA in the first primitive life form. Ribonucleotides are ubiquitous in modern biological processes and are constituents of many co-enzymes; this prompted White to consider co-enzymes as fossils of an earlier metabolism.<sup>[54]</sup> Ribonucleotides are used as the currency of energy storage (ATP), co-enzymes (NADH, FADH, CoA) and signalling in cells (cyclic ADP ribose).<sup>[55]</sup> The *de novo* biosynthesis of nucleotides invariably begins with ribose-5-phosphate, which is activated by ribose phosphate pyrophosphokinase to 5-phosphoribosyl 1- $\alpha$ -pyrophosphate (PRPP) from which the nucleoside-5'-phosphates are synthesised (Figure 8). Moreover, DNA nucleotides are formed from their corresponding ribonucleotides by reduction catalysed by ribonucleotide reductase.<sup>[55]</sup>

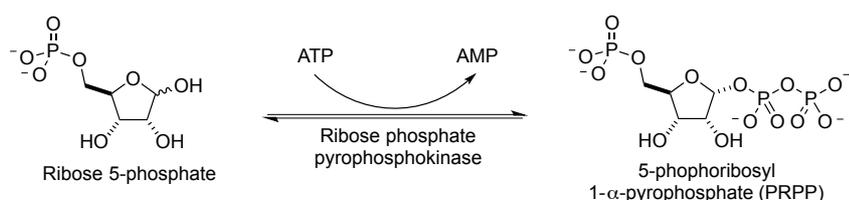
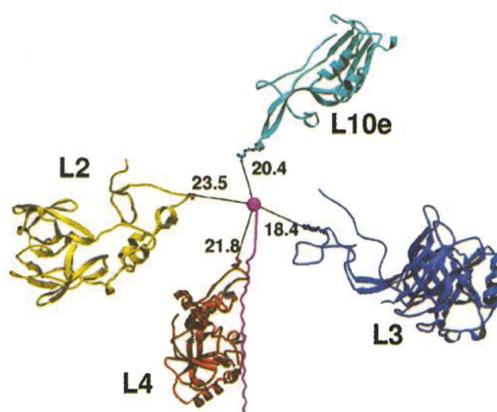


Figure 8. Biosynthesis of the key ribosyl precursor, 5-phosphoribosyl 1- $\alpha$ -pyrophosphate (PRPP), for the *de novo* synthesis of the nucleoside-5'-phosphates.

Persuasive evidence for an RNA world can be seen from the X-ray crystallographic data of contemporary ribosomes, where the site of peptidyl-bond formation (peptidyl transferase centre, PTC) is comprised of only ribosomal RNA and the closest proteins are 18.4 Å away (Figure 9).<sup>[56-59]</sup> Additionally, Noller *et al.* showed that two different bacterial ribosomes retained peptidyl transferase activity after extensive treatment with proteinase K and SDS. Thus, presenting the first evidence that catalysis by the ribosome

is RNA based.<sup>[60]</sup> These discoveries led Steitz to suggest that the “ribosome is a ribozyme” and that the first primitive ribosome comprised entirely RNA.<sup>[61]</sup> More over, Yonath has observed that the catalytic core of the ribosome is semi-symmetrical, and has suggested that a primitive or proto-ribosome could have comprised an all RNA self-assembled dimer.<sup>[62]</sup> The origin of ribosomal proteins has also been considered, many have long non-globular extensions that protrude in towards the centre of ribosome and clearly stabilise the structure of the ribosome. They fill the gaps between the RNA subunits and also neutralise the negative charges of the phosphates through the positively charged basic amino acid residues. It is therefore a possibility that the first peptide-synthesising RNA produced peptides that were useful in stabilising its structure, thereby improving its catalytic activity.<sup>[57, 61, 62]</sup>



*Figure 9. The peptidyl transferase centre (PTC) is represented with the RNA removed and is located at the magenta sphere. Also in magenta is a modelled polypeptide product. The closest proteins, L2, L3, L4 and L10e are shown and only approach within 18.4 Å (all distances are quoted in Ångström). Reprinted from reference <sup>[56]</sup> with permission from the copyright holder, American association for the advancement of science.*

## 1.5. Chemistry towards an abiogenesis of RNA and proteins

This thesis does not adhere to a strict RNA world, but considers that the abiogenesis of the first life form may have involved complementary or interrelated types of molecules. Nonetheless, RNA is recognised to have played an important role, and Chapter 1 will describe prebiotic chemistry related to the abiotic formation, oligomerisation and aminoacylation of RNA. Finally, an alternative theory of a linked origin of RNA and peptides will also be introduced.

### 1.5.1. The Miller-Urey experiment

Stanley Miller and Harold Urey's famous spark-discharge experiments stimulated origin of life research in 1953. Miller had read Oparin's book on the origins of life in which he suggested that the atmosphere of the early earth was highly reduced and contained CH<sub>4</sub>, NH<sub>3</sub>, H<sub>2</sub>O and H<sub>2</sub>.<sup>[47]</sup> From this atmosphere it was suggested that organic compounds could be formed and so Miller devised an experiment where these simple gases and water vapour were subjected to spark discharges to simulate the action of lightning.<sup>[14]</sup> The products were collected over a period of a week; many organic compounds were isolated, including HCN **1** and formaldehyde **2**.<sup>[63]</sup> Following strong acidic work-up of the reaction mixtures they were found to contain  $\alpha$ -amino acids that included aspartic acid, glycine, valine and alanine.<sup>[14, 63-65]</sup> The formation of the amino acids is thought proceed *via* Stecker-type chemistry (Figure 10).<sup>[66]</sup> However, the plausibility of this chemistry is questionable in that only low yields of amino acids were formed, and it is now generally thought that the atmosphere of the early earth was neutral and dominated by N<sub>2</sub> and CO<sub>2</sub>.

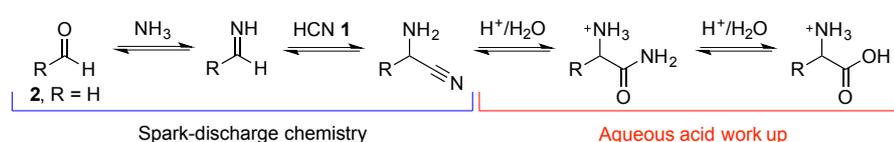


Figure 10. Amino acid formation in the Miller-Urey spark discharge experiments that occur via Strecker-type chemistry.

Spark discharge experiments were revisited under neutral atmospheric conditions.<sup>[67]</sup> An atmosphere of N<sub>2</sub>, CO<sub>2</sub> and water vapour were subjected to spark discharge for 48 hours and the products subjected to acidic work-up. The results showed that amino acids serine, glutamic acid, glycine and alanine could be produced but yields were lower than those obtained under reducing conditions. It has been proposed that the low yields of organic products from neutral atmospheres is due to limited formation of HCN **1** which is a key reagent in the Strecker synthesis of amino acids.<sup>[68]</sup> Both spark discharge experiments suffer from low yields suggesting limited prebiotic significance. Despite this, similar amino acids have been detected on the Murchison meteorite in comparable abundances to spark discharge experiments, suggesting that some other plausible synthesis maybe found.<sup>[69]</sup>

### 1.5.2. The traditional disconnection of RNA

Much of the prebiotic chemistry of RNA has focused on a structurally obvious disconnection (Figure 11). Firstly, polymeric RNA was proposed to have formed through the polymerisation of activated monomers. These monomers could be in the form of a 5'-phosphate **3** (X = leaving group), or of a 2'/3'-phosphate. Attack by a 3'-hydroxyl of a monomer on an activated 5'-phosphate **3** would lead to oligomerisation. Activation of 2'- or 3'-phosphates, however, would lead to monomer cyclisation through intramolecular attack by the adjacent hydroxyl group to give a nucleoside-2',3'-cyclic phosphate **4**. These species are 'stably' activated due to the slight ring strain; attack at cyclic phosphate by a 5'-hydroxyl group of another monomer would again give oligomerisation. For a detail discussion of the oligomerisation of activated nucleotide monomers see Chapter 1.6. The final disconnection of the activated monomers had long been unquestioned and seemed to be the most obvious. The assumption was that the monomers were derived from D-ribose **5**, a preformed heterocyclic base and phosphate **6** (Figure 11).

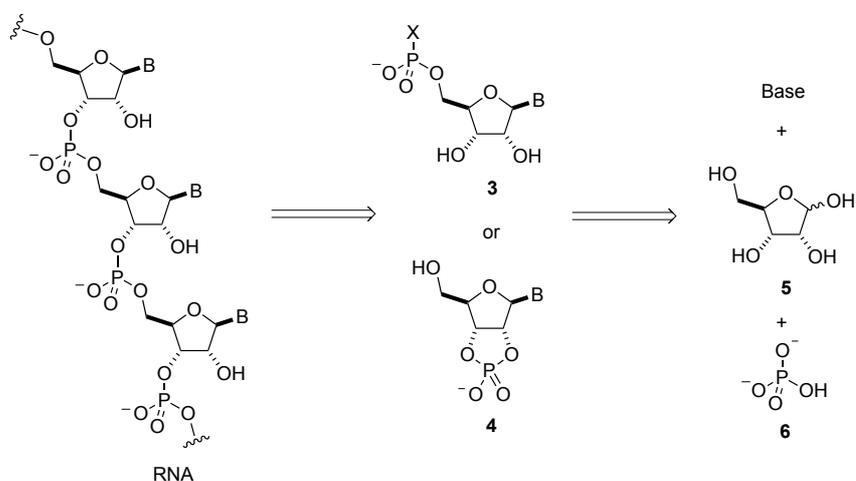


Figure 11. The traditional and obvious disconnection of RNA.

Although there is a wide field of work with this disconnection in mind, there are many problems associated with it and these are briefly summarised here. The purine nucleobases can be synthesised from HCN **1** and the pyrimidine nucleobases from cyanoacetylene **7**, both of which are formed in the spark discharge experiments (Figure 12). However, the yield of purine bases is low, and attempts to improve the yield have been met with little success.<sup>[70-76]</sup> The formation of the pyrimidine nucleobases is a little

more efficient and the highest yields (up to 19%) are obtained with cytosine **8**.<sup>[77, 78]</sup> Historically, Butlerow's formose reaction has been considered the most plausible way to form sugars and involves the polymerisation of formaldehyde **2** *via* repeated aldol condensations catalysed by an alkaline Earth metal, such as calcium hydroxide.<sup>[79]</sup> However, chromatographic analysis of the products from the formose reaction revealed a complex mixture of sugars with lack of both regio- and stereo-selectivity. Indeed, the yield of *rac*-ribose **5** is less than 1%, and of additional concern is the instability of ribose **5** under the formose conditions ( $t_{1/2} < 3$ hrs, pH = 10, 55°C).<sup>[80-82]</sup> Despite the lack of experimental evidence for a selective and efficient prebiotic method of forming D-ribose **5** or efficient synthesis of the nucleobases, direct attachment of a preformed base to **5** has been attempted. The best results have been reported by Orgel *et al.* who demonstrated that by heating D-ribose **5** with adenine **10** in the presence of MgCl<sub>2</sub> or seawater salts, β-D-adenosine **11** can be formed in only 4% yield.<sup>[83, 84]</sup> Even worse, the pyrimidines **8-9** are completely unreactive due to delocalisation of the N1 lone pairs into the carbonyl groups (Figure 12).<sup>[85, 86]</sup>

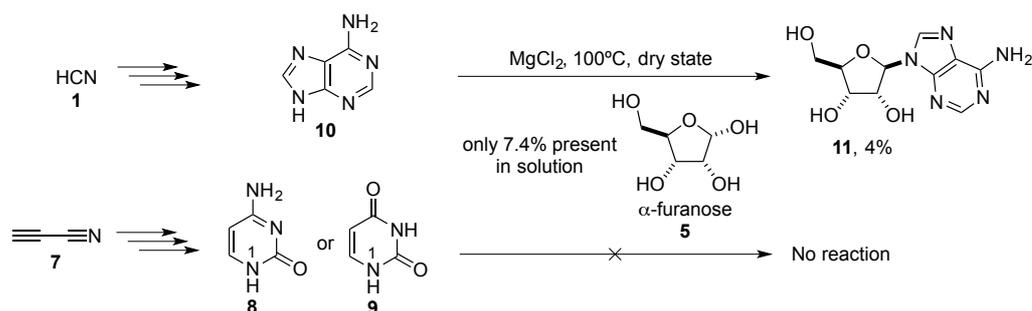


Figure 12. Nucleobase formation and the difficult glycosidation reaction.

The difficulty of direct attachment of nucleobase to sugar prompted Orgel and Sanchez to pursue a stepwise assembly of the pyrimidine nucleosides.<sup>[87]</sup> D-Ribose **5** or D-ribose-5-phosphate **5-5P** was reacted with cyanamide **12** to give the D-ribofuranosyl aminooxazolines *ribo*-**13**/*-5'P*. These were then subsequently treated with **7** to furnish α-D-ribofuranosyl cytidines α-**14**/*-5'P* in good yield, which are unfortunately the incorrect anomers (Figure 13a). Others have had success utilising D-arabinose to form the natural β-ribonucleotides<sup>[87, 88]</sup>; in particular Sutherland *et al.* have shown that **15** could be formed from D-arabinose-3-phosphate D-**16**-3P by sequential addition of **12** and **7** (Figure 13b).<sup>[89]</sup> The transformation of **15** into **17** and **18** was brought about under

prebiotically plausible conditions (pH = 7, sodium counterions) to give a conversion ratio of 1:4 (**17**:**18**) with an overall yield (from D-**16-3P**) of 3.5%.

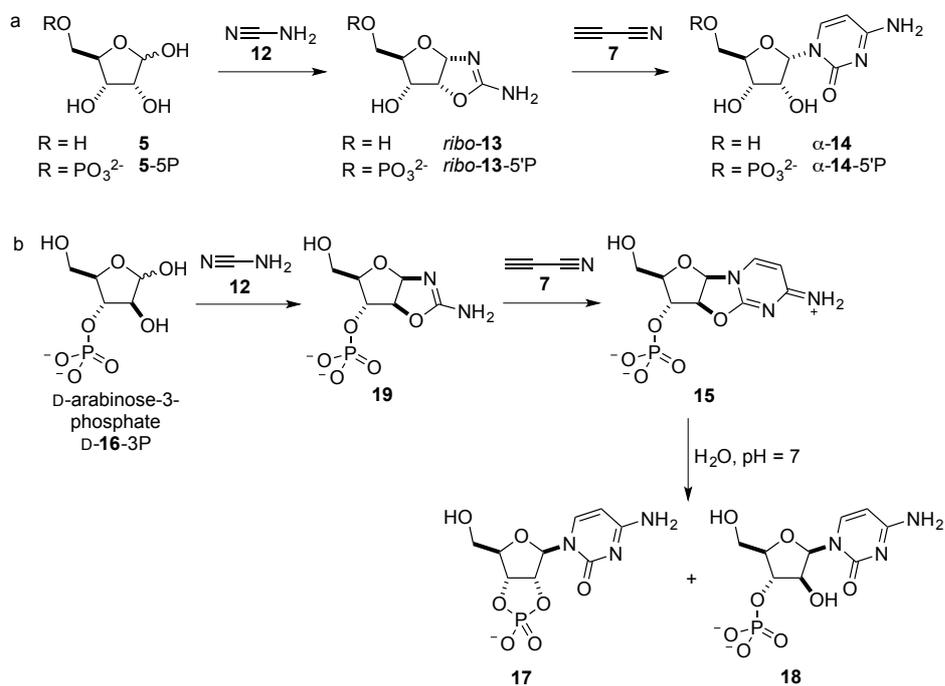


Figure 13. Stepwise assembly of pyrimidine nucleosides and nucleotides upon a) a ribose sugar and b) an arabinose sugar template.

Stepwise assembly of pyrimidine nucleobase upon a sugar (phosphate) template avoids the need for direct assembly of preformed sugar and nucleobase. However, the synthesis still requires pure D-**5**, D-**5-5P** or D-**16-3P** of which there is, as yet, no prebiotically plausible synthesis. Moreover, the hydrolysis to arabino-configured **18** is preferred over cyclisation to the desired  $\beta$ -D-ribofuranosyl cytidine-2',3'-cyclic phosphate **17**, which contributes to the low overall yields.

Despite some successes towards the direct synthesis of nucleoside and nucleotides, accumulation of prebiotically plausible feedstock molecules in pure form seemed impossible. In particular, the need for a preformed sugar and lack of an efficient synthesis of nucleosides and nucleotides were major hurdles towards considering the involvement of RNA at the origin of life. These problems led Joyce, Schwartz, Miller and Orgel to a dejected conclusion<sup>[90, 91]</sup>:

“It is possible that some efficient prebiotic synthesis of the  $\beta$ -ribosides, or some method of separating the  $\beta$ -ribosides from closely related isomers, will be discovered, but there is no basis in organic chemistry for optimism.”

### 1.5.3. Recent successes: synthesis of the activated pyrimidine nucleotides bypassing preformed ribose

The stepwise assembly of the pyrimidine nucleobases on a sugar template, whilst still requiring a preformed sugar, brought to light possible etiologically relevant intermediates: aminooxazolines (such as **13** and **19**). The aminooxazolines confer selectivity towards the furanose sugars due to the obligate 1',2'-*cis*-relationship and additionally their stability was found to be greater than the free sugars.<sup>[82]</sup> Powner *et al.* questioned whether the aminooxazoles (*ribo-/arabino-13*) could be formed from two simpler molecules; glyceraldehyde **20** and 2-aminooxazole **21**.<sup>[92]</sup> It is known that 2-aminooxazole **21** can be formed from the reaction of glycolaldehyde **22** with cyanamide **12**,<sup>[93]</sup> both of which are thought to be prebiotically available. This alternative disconnection of **13** led Powner *et al.* to investigate a new route to pyrimidine ribonucleotides that bypassed free sugars and nucleobase (spatially separated oxygenous and nitrogenous chemistries) where the first step involves mixed oxygenous-nitrogenous chemistry to give **21** (Figure 14).<sup>[94]</sup>

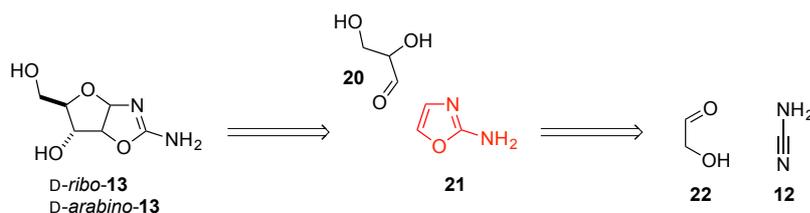


Figure 14. Disconnection of the aminooxazolines **13**.

The condensation between glycolaldehyde **22** and cyanamide **12** had been previously conducted in highly basic aqueous THF solution, and more prebiotically plausible conditions (*i.e.* neutral conditions) were thus sought.<sup>[93]</sup> However, in aqueous conditions and at neutral pH the condensation of **22** and **12** was found to be low yielding. It was suspected that formation of **21** was slowed by lack of specific base catalysis. Phosphate was chosen as an ideal general acid-base catalyst as its second  $pK_a$  is close to neutrality

and because it would ultimately be incorporated into the nucleotides. The addition of 1 M phosphate, in a repeat reaction of **22** and **12** conducted at pH = 7, proved to be an excellent choice as the 2-aminooxazole **21** was formed in >80% yield with excellent suppression of by-products (Figure 15).

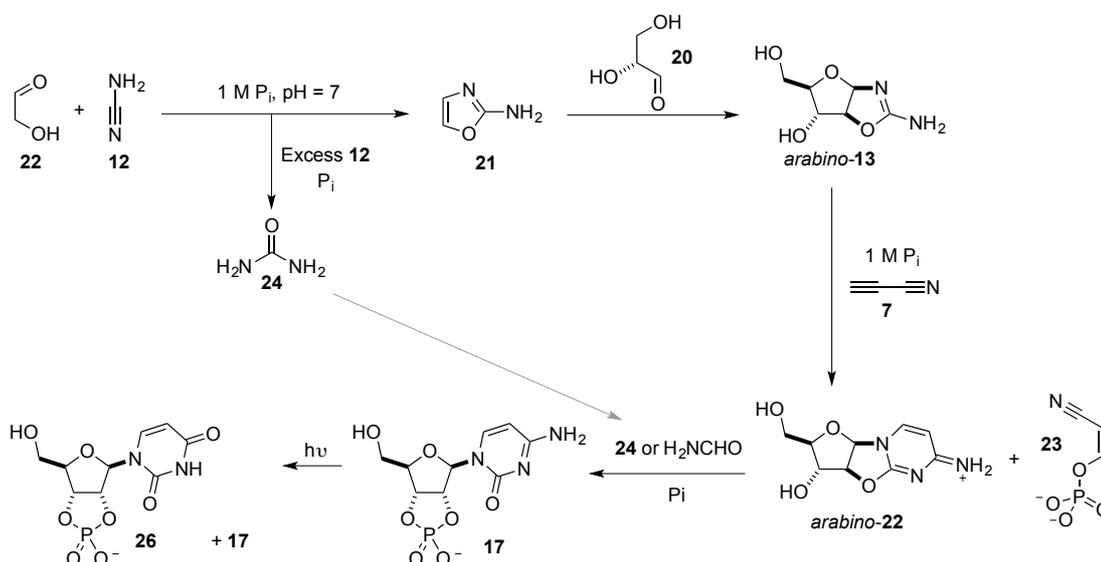


Figure 15. Summary of the synthesis of the activated pyrimidine nucleosides by Powner et al..

The condensation of glyceraldehyde **20** and 2-aminooxazole **21** was a crucial step as it would bypass preformed sugars. To simulate conditions on the early Earth, glyceraldehyde **20** was directly added to 2-aminooxazole **21** that had been freshly made from cyanamide **12** and glycolaldehyde **22** in the presence of phosphate. Results show that the reaction was tolerant to phosphate and that all four pentose aminooxazolines were formed in 50% yield over two steps (*ribo:arabino:lyxo:xylo* 25:15:6:4).<sup>[94]</sup> Of the major products, *ribo-13* has been found to be less soluble than *arabino-13*,<sup>[87]</sup> and additionally *ribo-13* is the least soluble of the all the pentose aminooxazolines.<sup>[82]</sup> By cooling the product mixture from the reaction of glyceraldehyde **20** and 2-aminooxazole **21**, *ribo-13* was selectively crystallised to give *arabino-13* as the major product in solution.

Sanchez and Orgel have previously shown that reaction of arabinose aminooxazoline *arabino-13* with excess cyanoacetylene **7**, in an unbuffered aqueous solution, gives  $\beta$ -arabinocytidine **18** (without 3'P) in low yield.<sup>[87]</sup> Powner *et al.* subsequently found that, during the reaction, a rise in pH causes hydrolysis of the anhydronucleoside

intermediates (e.g. *arabino-22*) and excess cyanoacetylene **7** reacts with the hydroxyl groups. To control the pH rise, a buffer was needed, and as in the first reaction phosphate was utilised. At pH = 6.5, the reaction was very clean with little evidence of anhydronucleoside hydrolysis. The excess cyanoacetylene **7** was also revealed to have reacted with phosphate to give cyanovinyl phosphate **23** instead of reacting with the anhydronucleoside hydroxyl groups. The use of phosphate, which acts as both a pH and a chemical buffer, allowed the formation of *arabino-22* in an extremely high yield of 92%.

With the arabinose-anhydronucleoside *arabino-22* available through an efficient, prebiotically plausible route, the next step involved a combined phosphorylation-rearrangement to convert *arabino-22* to the activated ribonucleotide **17**.<sup>[88, 95]</sup> The formation of cyanovinyl-phosphate **23** presented an alternative phosphorylating agent as it is known that **23** reacts with inorganic phosphate to form pyrophosphate.<sup>[96]</sup> Dry-state phosphorylation<sup>[97]</sup> in urea **24** was of particular interest as it is produced in the first step (formation of **21**) if cyanamide **12** is initially present in excess. Thus, heating arabinose-anhydronucleoside *arabino-22* with 0.5 equivalents of pyrophosphate in urea **24** gave  $\beta$ -ribocytidine-2',3'-cyclic phosphate **17** (32%) as the major product. Alternatively, **17** could be formed in even greater yield (46%) by heating *arabino-22* with inorganic phosphate and urea **24** in formamide solution.<sup>[98]</sup> Formation of **17** is thought to proceed by initial phosphorylation of the 3'-hydroxyl group of *arabino-22* to give the intermediate **25** that then undergoes an intramolecular nucleophilic substitution (Figure 16). This was remarkable, as it seemed that selectivity was for 3'-phosphorylation over 5'-phosphorylation, which is contrary to conventional knowledge as primary hydroxyls are normally less sterically hindered.

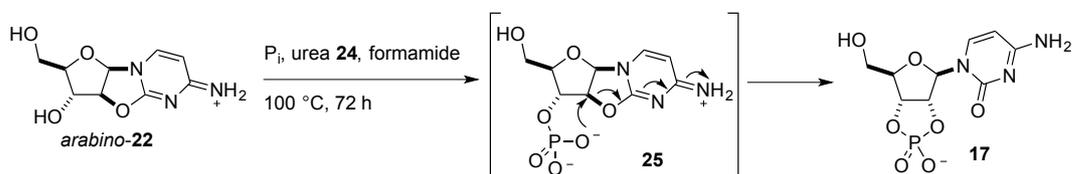


Figure 16. Mechanism of the phosphorylation-rearrangement of *arabino-22* to give  $\beta$ -ribocytidine-2',3'-cyclic phosphate **17**.

The X-ray crystal structure of *arabino-22* revealed a C(4')-*endo* sugar pucker with an added consequence that the 5'-OH is in a short contact ( $r_{O...C} = 2.70 \text{ \AA}$ ) with C2 (Figure

17a).<sup>[94]</sup> Sutherland *et al.*, assuming that this solid-state conformation was also the predominant conformation in solution, suggested that phosphorylation at 5'-OH was more sterically hindered. Choudhary *et al.* carried out computational studies on *arabino-22* and found that the short contact was preserved in the calculated structure ( $r_{O\cdots C} = 2.88 \text{ \AA}$ ).<sup>[99]</sup> Further calculations showed that electron density of the lone pair ( $n$ ) of the O5' is delocalised over the antibonding orbital ( $\pi^*$ ) of the C2=N3 bond (*i.e.* an  $n \rightarrow \pi^*$  electronic delocalisation, Figure 17b). Furthermore, the O5'...C2=N3 angle of  $99.2^\circ$  is close to the Bürgi-Dunitz trajectory ( $\sim 107^\circ$ ) such that the  $n \rightarrow \pi^*$  electronic delocalisation is reminiscent of nucleophilic attack on carbonyl groups.<sup>[100]</sup> This study suggested that 5'-OH reactivity was diminished towards phosphorylation as the proximity of O5' and C2' increased steric demand near O5', in support of Sutherland's suggestion. Additionally, the 5'-OH is engaged in an  $n \rightarrow \pi^*$  electronic delocalisation that decreased the intrinsic nucleophilicity of O5'. Both of these factors did not affect the O3', which underwent efficient phosphorylation.

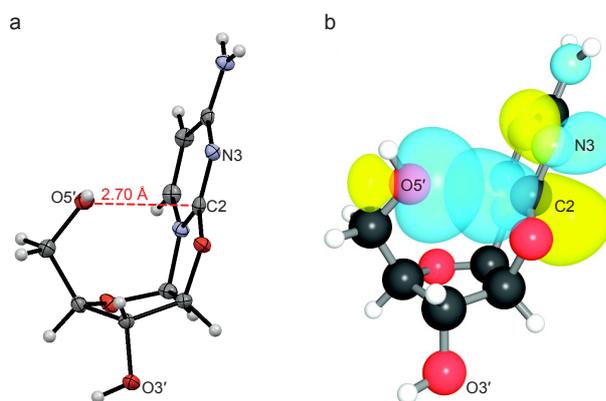


Figure 17. a) X-ray crystal structure of arabino-22. The dashed line shows the short contact distance between O5' and C2=N3 (2.70 Å) (adapted from reference <sup>[94]</sup>) and is in agreement with the gas-phase optimised geometry distance of 2.88Å. b) The gas-phase optimised geometry of arabino-22 that shows the overlap between the n of O5' and  $\pi^*$  of C2=N3. Reprinted from reference <sup>[99]</sup> with permission from the copyright holder American Chemical Society.

The presence of the major nucleoside/nucleotide by-products that accompany **17** would likely interfere with incorporation of activated **17** into polymeric RNA. A means to selectively destroy these and a way to convert some of **17** into  $\beta$ -ribouridine-2',3'-cyclic phosphate **26** was sought. Remarkably, it turned out that UV irradiation ( $\lambda = 254 \text{ nm}$  for 3 days at pH = 6.5) accomplished both of these goals.  $\beta$ -Ribocytidine-2',3'-cyclic

phosphate **17** underwent very little destructive photochemistry but underwent significant hydrolysis to  $\beta$ -ribofuranose-2',3'-cyclic phosphate **26** (42%, with 43% of **17** recovered).

This route to the activated pyrimidine ribonucleotides as their 2',3'-cyclic phosphates **17** and **26** was the first to overcome the problem of the availability of preformed sugar and the need for direct attachment of nucleobase to sugar. It therefore demonstrates that, under the correct geochemical conditions, the prebiotic formation of activated pyrimidine nucleotides can be viewed as predisposed. However, one of the issues with this chemistry is that the glyceraldehyde **20** used was racemic and both enantiomers of the pyrimidine nucleotides are produced. Moreover, it is known that enantiopure monomers are required for the template-directed ligation of activated RNA nucleotides, as enantiomeric cross-inhibition is a severe limitation for continual replication.<sup>[101]</sup>

As previously described, racemic *ribo*-**13** selectively crystallises from a solution of all four aminooxazolines **13**.<sup>[82]</sup> Additionally, Anastasi *et al.* have studied the formation of the aminooxazolines **13** from 2-aminooxazole **21** and scalemic glyceraldehyde **20** and found that if the *ee* of **20** is  $\geq 60\%$ , crystallisation of *ribo*-**13** gives enantioenriched crystals that are optically pure.<sup>[92]</sup> Powner and Sutherland have now shown that inorganic phosphate can catalyse the interconversion of enantiopure *ribo*-**13** to *arabino*-**13** (Figure 18).<sup>[102]</sup> The interconversion mechanism proceeds *via* the ring opened *ribo*-**13** to give the iminium species **27** which, can undergo phosphate-mediated deprotonation at C2' to give the C5 substituted 2-aminooxazole **28**. From this intermediate reprotonation of C2' can regenerate **27** or give **29** that after ring-closure generates *arabino*-**13**. As enantiopure *ribo*-**13** can be obtained by crystallisation, this provides a way to transfer enantiopurity from the *ribo* series to the *arabino* and finally to the *ribo* activated pyrimidine nucleotides.

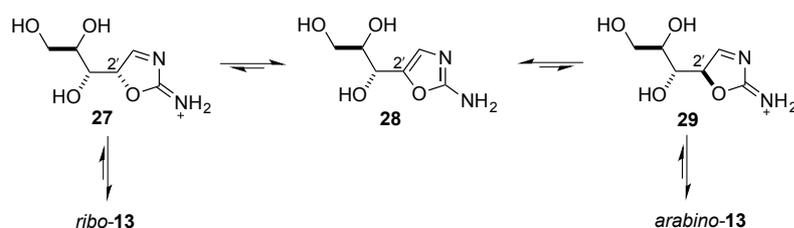


Figure 18. The phosphate-mediated interconversion of *ribo*-**13** and *arabino*-**13**.

Hein *et al.* have also addressed the enantiopurity problem at the stage of aminooxazoline **13** synthesis.<sup>[103]</sup> Utilising slight enantiomeric imbalances of proteogenic amino acids results in a kinetic resolution of the natural D-aminooxazolines D-**16**. They showed that when proline **30** with an initial 1% *ee* of the natural L-enantiomer was added to a reaction of *rac*-glyceraldehyde *rac*-**20** and 2-aminooxazole **21**, aminooxazolines of D-*ribo*-**13** and D-*arabino*-**13** could be produced in 20-80% *ee*. On cooling the solution, enantiopure crystals of D-*ribo*-aminooxazoline D-*ribo*-**13** could be produced, which can be interconverted to D-*arabino*-**13** by inorganic phosphate as previously discussed. The enantioenrichment is attributed to 2.5-fold faster reaction rates between glyceraldehyde **20** and proline **30** for the D-D (or L-L) when compared to the D-L (or L-D) sugar-amino acid interaction (Figure 19). Therefore, the enantioenriched L-proline L-**30** effectively sequesters the L-glyceraldehyde **20** to form the three-component product **31** and equally the D-proline D-**30** effectively sequesters the D-glyceraldehyde **20** in the same way. However, the enantiomeric deficiency of D-**30** leaves some of the natural D-glyceraldehyde **20** unreacted which goes on to react with 2-aminooxazole **21** to give the natural D-configured aminooxazolines D-**13**. A combination of kinetic resolution by amino acids and physical enantioenrichment by crystallisation provide strong evidence that a prebiotically plausible synthesis of enantiomerically pure pyrimidine nucleotides is possible. Encouragingly, only a small asymmetry in the amino acid enantiomers is required that could have occurred by chance. Moreover, small *ee* values of the L-amino acids have been observed in chondritic meteorites.<sup>[104]</sup>

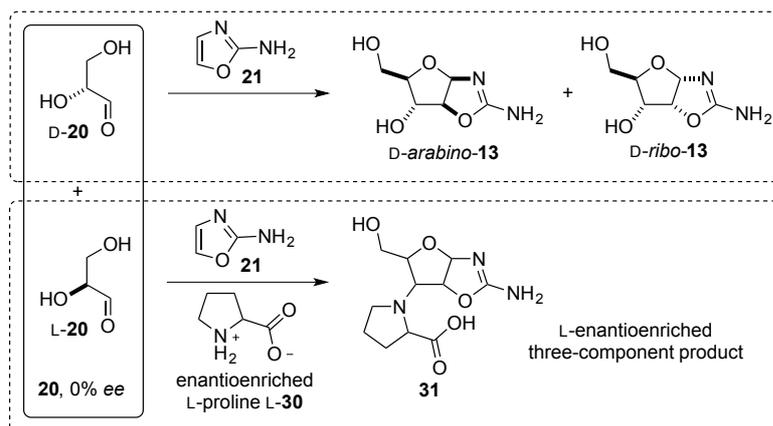


Figure 19. In the presence of enantioenriched L-proline **30**, the reaction of rac-**20** and **21** results in the diastereoselective formation the three-component product **31**. Due to reduced rate of reaction between D-L/L-D sugar-amino acid and the enantio-deficiency of D-proline **30** there is incomplete conversion of D-glyceraldehyde D-**20** that is then involved in formation of D-amino-oxazolines **13**.

## 1.6. Abiotic synthesis of polymeric RNA

The disconnection of RNA reduces the polymer into monomers of either nucleoside-5'-phosphates or nucleoside-2',3'-cyclic phosphates (see Chapter 1.5.2). Oligomerisation of both species has been extensively studied in the absence (non-templated) and presence (templated) of existing RNA oligomers.<sup>[105]</sup> There are several requirements for a possible monomer oligomerisation to be considered plausible. Biology almost exclusively contains 3',5'-internucleotide linkages in both RNA and DNA and any prebiotic process overall should selectively give the natural connectivity. Recent work from Szostak's lab however, shows that 10-25% of the wrong 2',5'-linkage isomers are tolerated in functional RNA, suggesting some linkage heterogeneity is allowed.<sup>[106]</sup> For RNA to successfully take part in replication, it should be able to form complementary duplex structure and use Watson-Crick base pairing to accurately transfer genetic information.<sup>a</sup>

<sup>a</sup> Oligomers of RNA are herein referred to as oligonucleotides rather than oligoribonucleotides for simplicity.

### 1.6.1. Oligomerisation of activated 5'-nucleotides

In contemporary biology, RNA is enzymatically polymerised using nucleoside-5'-triphosphates **32** (NTPs), which are high-energy phosphate esters (adenosine-5'-triphosphate **ATP** (**32**, B = A, Figure 20); standard free-energy of hydrolysis  $\Delta G'^{\circ} = -45.6 \text{ kJmol}^{-1}$ ).<sup>[107]</sup> However, in aqueous solution they are relatively stable and react slowly without enzymatic catalysis, so it is difficult to envisage how the prebiotic world would have utilised **32**.<sup>[91]</sup> Additionally, regiocontrol is a severe problem since polymerisation of activated NTPs can occur from attack of either the 2'-OH or 3'-OH. The 2'-OH is known to be 6-9 times more reactive than the 3'-OH and so leads to the unnatural linkage isomerism.<sup>[108]</sup> However, different salts, the identity of the activating agent and stereochemical orientation of monomers can vary the ratio of 2',5'- or 3',5'-linked products and these shall be discussed hereafter.

Nucleoside-5'-phosphorimidazolides **33** (Figure 20, where R = H, alternative nomenclature - ImpN where N = A, C, G or U) have been commonly used to study the oligomerisation of 5'-activated nucleotides. They were chosen due to ease of preparation and showed convenient reaction rates in aqueous solution, so should be considered as model systems.<sup>[109]</sup> In non-templated experiments and in the absence of catalysts, ImpN polymerise to give complex mixtures of short linear and cyclic oligomers.<sup>[105]</sup> Various metal ions have been found to catalyse the polymerisation reaction.<sup>[110, 111]</sup> In particular,  $\text{Pb}^{2+}$ <sup>[112, 113]</sup> and  $[\text{UO}_2]^{2+}$  have been found to produce longer oligomers. The uranyl ion has been found to efficiently catalyse the self-condensation of ImpA, ImpC or ImpU up to 16 nucleotides (nt).<sup>[114, 115]</sup> However, one major problem with the chemistry described so far is that the newly formed internucleotide linkages constitute greater than 80% of the unnatural 2',5'-linkage isomer. Impressive work by Ferris and co-workers has shown that the clay-mineral montmorillonite is also a very effective catalyst.<sup>[116-118]</sup> Using activated nucleoside-5'-phosphates based on 1-methyladenine **34** (Figure 20), it has been found that under mildly alkaline aqueous conditions (pH = 8), **34** can be oligomerised up to lengths of 50nt. Remarkably, the internucleotide linkages were found to be approximately 80% of the natural 3',5'-linkage. Although the exact mechanism is not known, Ferris suggests that the selectivity is due to the intercalation of monomers **34** into the layers of

montmorillonite. This adsorption on the surface of the mineral brings the monomers into close proximity and an orientation that favours attack of the 3'-OH.<sup>[119]</sup>

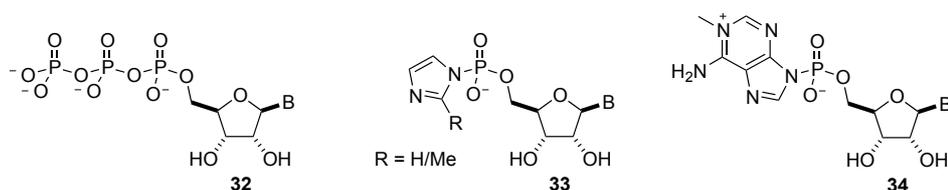


Figure 20. Activated nucleoside-5-phosphates.

Orgel and co-workers conducted much of the early work on the templated oligomerisation of activated nucleoside-5'-phosphates such as **33**. In the presence of a polyuridine template (polyU), the condensation of ImpA yields mainly dimers and trimers with principally the unnatural 2',5'-linkages (95%).<sup>[108]</sup> Further work by Ninio *et al.* found that oligomerisation of ImpG on a polyC template gave oligomers with predominantly the natural 3',5'-linkages (65%).<sup>[120]</sup> The oligomerisation can be enhanced using Pb<sup>2+</sup> ions, and in the presence of polyU, ImpA can oligomerise to give products longer than 5nt that contain more than 75% of the natural 3',5'-linkages.<sup>[113]</sup> Conversely, ImpG can be polymerised on a polyC template in the presence of Pb<sup>2+</sup> to very effectively give polymers of at least 40nt; however 2',5'-linkages now total 90%.<sup>[121]</sup> Using the more prebiotically plausible metal ions Zn<sup>2+</sup> and Mg<sup>2+</sup>, Bridson and Orgel found that the templated oligomerisation of ImpG on polyC produced oligomers of 9-10nt and surprisingly with purely 3',5'-linkages.<sup>[122]</sup> Despite this success, the catalysis by Zn<sup>2+</sup> and Mg<sup>2+</sup> could not be transferred to the corresponding reaction of ImpA on polyU. Overall, the templated condensations using nucleoside-5'-phosphorimidazoles are inconsistent and the internucleotide connectivity of the products are highly dependent upon the identity of the metal ions and the monomer/template.

In related work with nucleoside-5'-phospho-2-methylimidazole **33** (Figure 20, where R = Me, alternative nomenclature - 2-MeImpN where N = A, C, G or U), Inoue *et al.* found that 2-MeImpG can be oligomerised very efficiently upon a polyC oligonucleotide<sup>[123]</sup>: maximally 89% of the monomer could be converted to oligomeric material up to 50nt and which constituted exclusively 3',5'-linkages. However, the corresponding condensation of 2-MeImpA cannot be achieved as the polyU template

forms triple helices rather than double helices. Conversely, the oligomerisation of 2-MeImpC is prevented as polyG forms very stable G-quadruplexes.<sup>[124]</sup> With regards to templates of random sequence, the sequence can be faithfully copied using mixtures of 2-MeImpN (N = A, C, G and U) with very low rates of misincorporation but only if the sequence contains 60% C residues.<sup>[125]</sup> The requirement for at least 60% C residues is a major roadblock for replication in this system since the progeny contains 40% C residues.<sup>[126]</sup>

The ligation of short 5'-phosphate oligonucleotides activated as the 5'-phosphorimidazolides has seen less interest as the prebiotic accumulation of this type of chemical substrate has been doubted.<sup>[105, 127]</sup> Using more biotically relevant 5'-triphosphate oligonucleotides, Rohatgi *et al.* have described a template-directed non-enzymatic oligonucleotide ligation reaction that shows strong preference for the formation of native 3',5'-internucleotide bonds.<sup>[128, 129]</sup> Using a complementary 13nt template, a 10nt 'primer' oligonucleotide was ligated to a 5'-triphosphate-activated 7nt 'ligator' oligonucleotide at pH = 7.4, 37 °C and in the presence of divalent metal ions (Figure 21a). Ligation occurred through attack by the 3'-OH (or 2'-OH) of the primer on the 5'-triphosphate of the ligator. In this system attack by the 3'-OH was 60-80 times faster than attack of the 2'-OH, leading to favoured formation of 3',5'-linked-over-2',5'-linked phosphodiester bonds.

The selectivity of ligation was not affected by the identity of the base pair at the 3'-end of the attacking primer. Divalent metal ions were found to be essential for ligation, with Mn<sup>2+</sup> and Mg<sup>2+</sup> the most efficient catalysts. The requirement for divalent ions was rationalised by the association of a metal ion to the β- and γ-phosphates, which stabilised the developing negative charge of the leaving pyrophosphate. A metal ion is also thought to bind to the α-phosphate and the 3'-OH through a bridging hydroxide. This is proposed to assist in deprotonation and stabilising the transition state of the attacking nucleophile (Figure 21b). However, this ligation is extremely slow ( $t_{1/2} \approx 15$ -30 years at pH 7.4 and 100 mM Mg<sup>2+</sup>) and at higher pH = 8.9, where the ligation rate is higher, only 0.2% yield is observed after 100 hours. An important observation is that, in the context of a double helix, an isolated 2',5'-linkage suffers hydrolysis 50-100 times faster than a 3',5'-linkage in the same location. This indicates that the natural 3',5'-linked oligonucleotides would have accumulated in preference to the unnaturally-linked

oligonucleotides. However, the rate of ligation is so slow it would be in competition with hydrolysis of the products, which would be a major hindrance to significant accumulation of longer oligonucleotides.

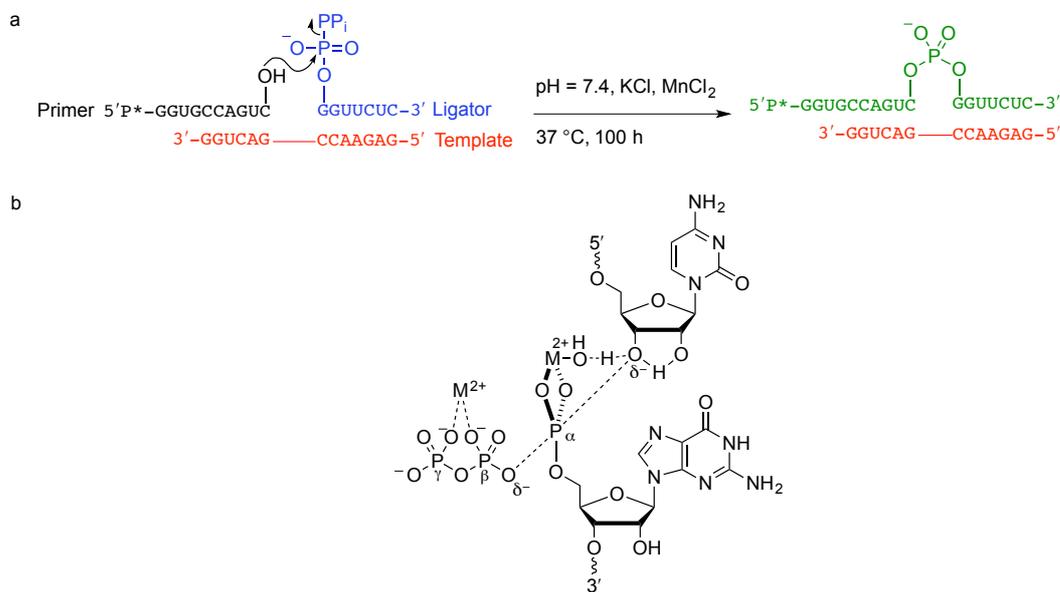


Figure 21. a) Ligation of 5'-triphosphate activated oligonucleotides. The primer and ligators are aligned by Watson-Crick base-pairing on a complementary template. Asterisk denotes a  $^{32}\text{P}$ -labelled phosphate. b) Suggested transition state for the ligation reaction. A divalent metal ( $\text{M}^{2+}$ ) ion binds to the  $\beta$ - and  $\gamma$ -phosphates to stabilise the negative charge. A second metal ion is suggested to bind to the  $\alpha$ -phosphate and the 3'-OH (through a bridging hydroxide) of the primer. This aids in deprotonation of the 3'-OH. Adapted from references.<sup>[128, 129]</sup>

In summary, although formation of purely 3',5'-linked oligonucleotides can be formed from 5'-activated (oligo)ribonucleotides, these experiments have tended to use 5'-activation chemistry that is considered to be prebiotically implausible. Ligation reactions using 5'-triphosphate activated oligonucleotides have proven to be very slow, suggesting that accumulation of oligomeric RNA would have been difficult. Moreover, prebiotic syntheses of  $\beta$ -D-ribonucleotide-5'-phosphates and their activated derivatives have not been reported, so it is difficult to envisage RNA assembly by this pathway.

### 1.6.2. Oligomerisation of nucleoside-2',3'-cyclic phosphates

The prebiotically plausible synthesis of the activated pyrimidine ribonucleotides **17** and **26** has provided experimental evidence that these RNA building blocks could have

existed on the early earth and gives support to RNA's significance at the origin of life.<sup>[130, 131]</sup> A related prebiotic pathway to the activated purine ribonucleotides would lead to further support and studies are on going.<sup>[132]</sup> Nucleoside-2',3'-cyclic phosphates are possible monomers for oligomerisation as identified from the disconnection of RNA (Chapter 1.5.2). These nucleotides retain some activation because of ring strain and have for many years been considered monomers for RNA synthesis by polymerisation.<sup>[133-135]</sup> An issue with these activated nucleotides is that, if oligomerisation reactions are to take place in aqueous solution, it is inevitable that they will undergo competing hydrolysis to a mixture of the nucleoside-2'- and 3'-phosphates **35-2'P/3'P** and thus become deactivated (Figure 22). Oligomerisations could be carried out in the dry-state, but hydrolysis of **4** is likely to occur during evaporation to form the dry-state mixtures, possibly limiting oligomerisation yields. Hydrolysis is a pervasive deactivation pathway, but this problem can be alleviated if the hydrolysis products can be converted back to the cyclic products **4** by phosphate activation. It is also for this reason that **35-2'P/3'P** are poor candidates for the oligomerisation as cyclisation back to **4** is favoured.

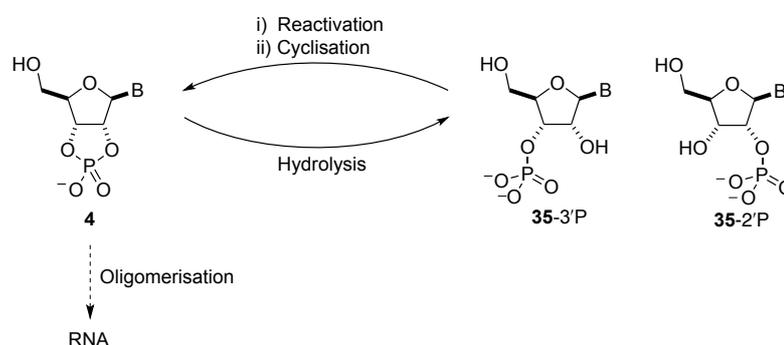


Figure 22. Nucleoside-2',3'-cyclic phosphates are susceptible to hydrolysis, which depletes stock available for oligomerisation. A continual activation is desirable to allow regeneration of **4** to provide monomers for the oligomerisation to RNA.

Prebiotically plausible activating agents for the cyclisation of **35** back to the nucleoside-2',3'-cyclic phosphates **4** have been investigated.<sup>[136]</sup> The activating agents studied include cyanofornamide, cyanamide **12** and cyanate. The most successful of these was **12** (0.8 M, pH = 5.0 and 65 °C), which was able to bring about conversion to **4** in 73% yield over 6 days. Alternatively, Sutherland and co-workers have described cyanoacetylene **7** as a possible activating agent, which is considered to be prebiotically

plausible and is a building block used in the synthesis of the nucleotides (Figure 15).<sup>[137, 138]</sup>  $\beta$ -D-Cytidine-monophosphates **36-2'P/3'P** can be converted to **17** in 55-60% yield at 60 °C with 6 equivalents of cyanoacetylene **7** (Figure 23). Nucleobase modification by **7** is a competing reaction, but encouragingly it was found to be reversible and could be minimised by addition of L-alanine (6 equivalents) to buffer the reaction. Nucleobase modification by **7** is not restricted to cytidine nucleotides; Furukawa *et al.* have shown that the base of adenosine nucleotides undergoes irreversible addition to **7**.<sup>[139]</sup>

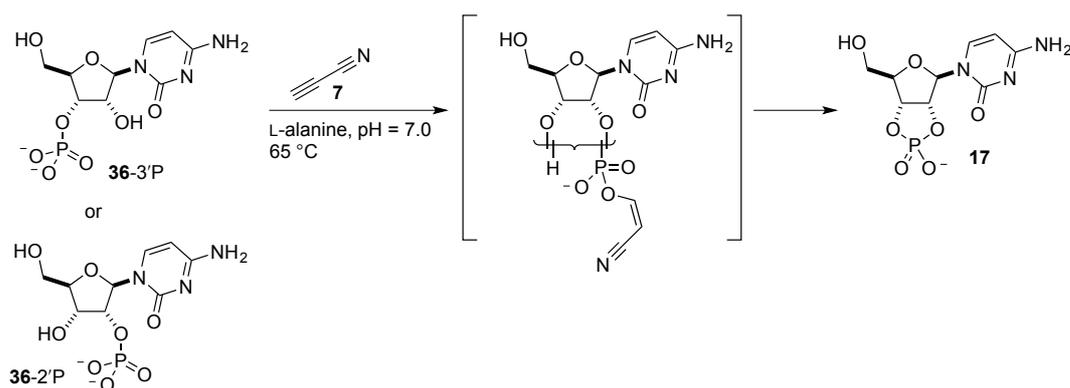


Figure 23. Formation of **17** by activation of **36-3'P/-2'P** with cyanoacetylene **7**.

Another efficient but selective phosphate activation was described by Mullen *et al.*, who's work was inspired by the multicomponent Ugi reaction<sup>[140]</sup> (and the related Passerini reaction<sup>[141]</sup>). They postulated that the initial reaction of isocyanide, an aldehyde, and  $\text{NH}_4\text{Cl}$  would form an intermediate that, instead of activating a carboxylic acid as in the classic Ugi reaction, could be used to activate a 2'-/3'-phosphate to give intermediate **37**. The byproducts of the reaction also produce derivatives of  $\alpha$ -amino acids (Figure 24a).<sup>[142]</sup> A second pathway was also envisioned, though later found not to occur, whereby an adjacent 2'- or 3'-hydroxyl could undergo aminoacylation *via* a 7-membered transition state instead of a 5-membered transition state required for phosphate cyclisation. This type of aminoacyl transfer has experimental precedence, where nucleoside-3'-phosphates **35-3'P** were reacted with *N*-carboxyanhydrides (see Chapter 4.1). Nonetheless, the nucleoside-2',3'-cyclic phosphates **4** were produced in near quantitative yields without any apparent modification of the nucleobases. Side-products of these reactions were also found to include the amino acid derivative **39** (Figure 24b).

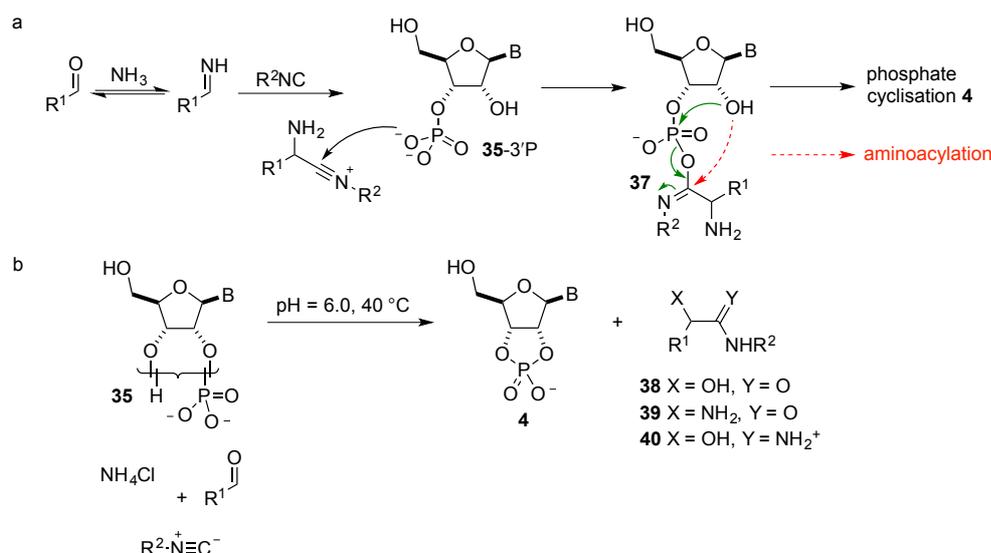


Figure 24. a) Possible mechanism for the phosphate cyclisation or C2' aminoacylation by Ugi/Passerini type reactions. b) Multicomponent reaction of nucleoside-2'/3'-phosphates **35** to give **4**. The side-products **38-40** are also formed, with **39** as the major byproduct.

The non-templated oligomerisation of adenosine-2',3'-cyclic phosphate **41** catalysed by various amines under dry-state conditions has been extensively studied by Verlander *et al.*<sup>[133, 134]</sup> Mixtures of **41** and catalysts such as aliphatic amines, amino acids and imidazole salts were evaporated to dryness under vacuum over P<sub>2</sub>O<sub>5</sub>. Once dry the residues were maintained at temperatures obtainable at the surface of the earth (25-85 °C). The most efficient catalyst was ethylenediamine **42** which gave 69% oligomeric material after 3 days of heating at 85 °C (Figure 25). Under more prebiotically plausible conditions whereby mixtures of **41** and **42** were allowed to evaporate under ambient conditions, 25% oligomeric material was obtained. Upon analysis of the reaction products, oligomeric material of hexamer and longer totalled 7.5%. Detailed analysis of this portion was found to contain significant amounts of oligomer in the range of 13-14nt long.<sup>[134]</sup> The most interesting finding of these studies was that the natural 3',5'-linkage dominated over the unnatural 2',5'-linkage (ratio of 3',5':2',5': dimer 1.85:1, trimer 1.65:1), as it suggests that the formation of 3',5'-linked RNA is chemically predisposed.

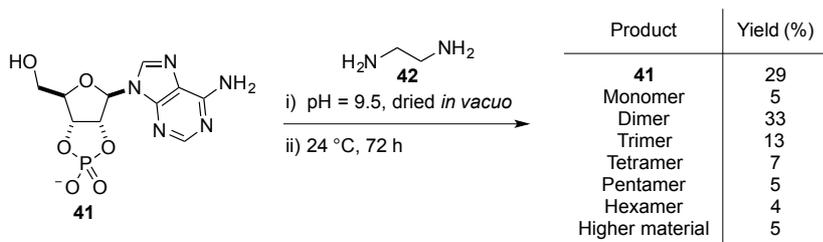


Figure 25. Yield of oligomers from the non-templated oligomerisation of **41** catalysed by 1,2-diaminoethane **42**.

Other catalysts that were used included imidazole, spermidine, NaCN and glycine, and under similar conditions (pH = 9.0-10.5, 85 °C, 5-8 hours) these prebiotically available compounds give total polymeric material in the range of 16-44% yield. Interestingly, these catalysts also cause significant amounts of hydrolysis of **41** to give the adenosine-2'/3'-monophosphates, with yields ranging from 25% to as high as 76%. A large proportion of the oligomeric products are also terminated as 2'/3'-monophosphates, and it seems apparent that deactivation of the 2',3'-cyclic phosphate by hydrolysis is unavoidable. This behaviour is important for an alternative oligonucleotide ligation method described in chapter 1.6.3.

In contrast to these results, the template-directed condensation of **41** in the presence of catalysts is very inefficient.<sup>[135]</sup> Hydrolysis of **41** is the predominant process with only a small percentage of dimers and trimers formed. To compound this problem further, the oligomeric products were 97% 2',5'-linked. Usher and McHale have shown that the ligation yields can be improved by using Watson-Crick base pairing in oligomers. Utilising a polyU template to direct the ligation of 2',3'-cyclic phosphate-terminated adenosine-hexamers in the presence of ethylenediamine **42** at pH = 8 was found to give 12mer (24%) and 18mer (5%) in moderate yields.<sup>[143]</sup> However, again it was found that the newly formed internucleotide phosphodiester bonds constituted 95% the 2',5'-linkage. Lutay *et al.* have attempted to improve the yields of template-directed ligation of 2',3'-cyclic phosphate oligomers using divalent metal ions but without success.<sup>[144]</sup> Again it is predominantly 2',5'-linkages that are formed. In agreement with findings from Rohatgi *et al.* (Chapter 1.6.1), the newly formed 2',5'-phosphodiester linkages underwent a higher rate of hydrolysis.

It is clear that the condensation of nucleoside-2',3'-cyclic phosphates **4** to form the natural 3',5'-linkage is difficult, but it is encouraging that the natural linkage isomer is nonetheless favoured in the dry-state reactions. Unfortunately, it is almost impossible to form 3',5'-linked RNA in templated reactions. Given that non-enzymatic template-directed ligation must surely have occurred at some point, it seems that 2',3'-cyclic phosphates could not have led to the synthesis of RNA (Figure 26). It is also apparent that hydrolysis is a pervasive problem but it has been found that nucleoside-2',3'-cyclic phosphates **4** can be hydrolysed cleanly with L-serine to **35-3'P** and **35-2'P** in a ratio of 2:1 and this is important for a recent prebiotically plausible ligation of oligonucleotides.<sup>[145]</sup>

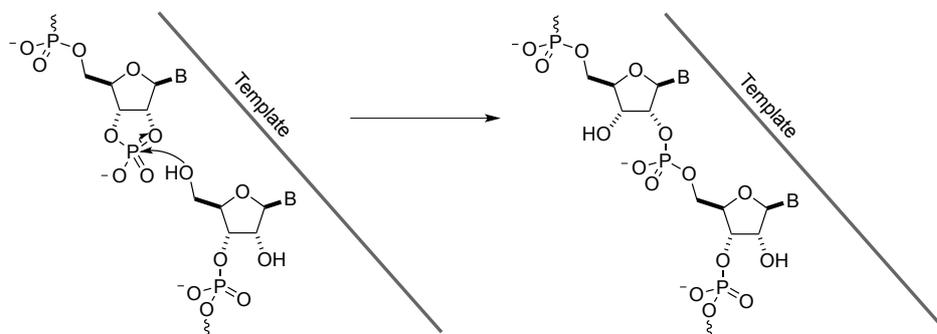


Figure 26. The template-directed oligomerisation/ligation of monomers/oligomers exclusively leads to the formation of the unnatural 2',5'-linkage.

### 1.6.3. Oligonucleotide ligation facilitated by chemoselective acetylation

Further to the discussion in chapter 1.6.2, Bowler *et al.* postulated that longer RNA could be synthesised from the short strands generated by the dry-state condensation of nucleoside-2',3'-cyclic phosphates **4**.<sup>[146]</sup> These short oligomers terminate in mixed 2'/3'-phosphates, and activation of the phosphate would only reform the 2',3'-cyclic phosphate before ligation could take place. It is also suggested that the evolutionary transition to fully 3',5'-linked nucleic acids would have been easier if prebiotically formed RNA was significantly enriched in 3',5'-linkages. As demonstrated by Rohatgi *et al.* and Lutay *et al.* the preferential hydrolysis of 2',5'-linkages can enrich for 3',5'-linkages but at the expense of chain cleavage. The key postulate of Bowler *et al.* is that, if a prebiotically plausible selective protection of the 2'-OH of a terminal 3'-phosphate could be found, subsequent phosphate activation would not lead to cyclisation. If this

acetylated and phosphate activated species was brought into proximity to the 5'-OH of another oligonucleotide by annealing to a complementary template this would lead to ligation with enrichment of 3',5'-ligation junctions.

The protecting group used in this work was the acetyl group; thioacetate **43** was chosen as it is considered prebiotically available from Wächterhäuser iron-sulphur world chemistry.<sup>[147]</sup> Thioacetate **43** can be converted to an acetylating agent by either electrophilic or oxidative activation.<sup>[148, 149]</sup> To activate **43**, the electrophile cyanoacetylene **7** was first chosen as it is involved in the synthesis of the activated pyrimidine nucleotides.<sup>[94]</sup> Thus, treating a mixture of adenosine-3'-phosphate **A3'P** and sodium thioacetate **43** with cyanoacetylene **7** in D<sub>2</sub>O at pD = 6.5 resulted in selective acetylation of the 2'-OH to give adenosine-2'OAc-3'-phosphate **A3'P-2'OAc** in 52% yield (Figure 27a).<sup>b</sup> A precipitate also formed rapidly and proved to be tetradeuterio-β,β-dicyanovinyl thioether **44**. Extending this chemistry to other nucleoside-3'-phosphates **N3'P**, they also showed selective 2'-OH acetylation with the following rough reactivity trends **A3'P** ~ **I3'P** > **G3'P** > **C3'P** > **U3'P**. When the chemistry was applied to nucleoside-2'-phosphates **N2'P**, 3'-OH acetylation was still observed but at significantly reduced efficiency. Additionally, relative to the acetylation of **N3'P**, a greater amount of phosphate cyclisation was observed. In mixtures of **N3'P** and **N2'P**, the reduced acetylation efficiency of the 3'-OH of the **N2'P** resulted in selective 2'-OH acetylation of **N3'P** (Figure 27b).

Alternative electrophiles to **7** were efficient and acetylation was equally selective (Figure 27c). These included cyanogen **45**, methyl isonitrile **46** and *N*-cyanoimidazole **47**; the latter was thought to bring about acetylation by intermediacy of *N*-acetylimidazole **48**. Direct acetylation with **48** was later used as a generic prebiotic acetylating reagent with **47** serving as the subsequent phosphate-activating agent. Oxidative activation was also found to be effective with thioacetate **43** and ferricyanide **49** affording high yields of **A3'P-2'OAc** and low yields of **A2'P-3'OAc**.

---

<sup>b</sup> The following numbering system will be used in this chapter for simplicity: nucleotides shall be referred to by either **A**, **C**, **G** etc. with the position and nature of the phosphate and/or modification indicated. For example, adenosine-3'-phosphate is numbered **A3'P**, adenosine-2',3'-cyclic phosphate is numbered **A>P** and adenosine-2'-OAc-3'-phosphate **A3'P-2'OAc**.

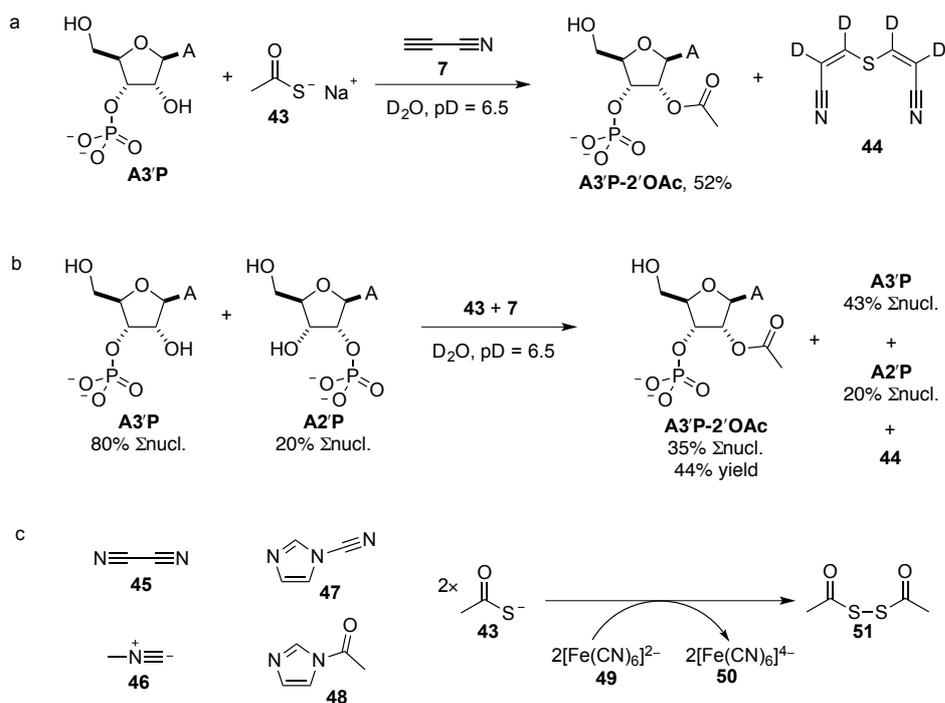


Figure 27. a) Adenosine-3'-phosphate **A3P** (100 mM) was treated with sodium thioacetate **43** (100 mM) and cyanoacetylene **7** (200 mM) in  $D_2O$  at  $pD = 6.5$  for 24 hours. b) **A3P** (80 mM) and **A2P** (20 mM) treated in the same way as b) results in exclusive 2'-acetylation of **A3P**. c) Additional electrophiles **45-47** have been shown to drive the acetylation of ribonucleotides with **43**. Direct acetylation with **48** is also possible, as it the oxidative activation of **43** with ferricyanide **49** to afford ferrocyanide **50** and the dimeric acetylating agent **51**.

The mechanism of the acetylation chemistry is proposed to proceed by reaction between the phosphate dianion of **N3'P** and the acetylating agent to give **N3'P(OAc)-2'OH** (Figure 28). This mixed carboxy-phosphate anhydride then undergoes rearrangement, resulting in an acyl-transfer of the acetate to give nucleoside-2'OAc-3'-phosphate **N3'P-2'OAc**. The selectivity of acetylation is rationalised in two ways: firstly, the reaction  $pD$  (or  $pH$ ) is close to the  $pK_a$  of the 2'- and 3'-phosphates and the slightly lower  $pK_a$  of the 3'-phosphate (0.2-0.5 units) may influence the selectivity,<sup>[150]</sup> secondly, the intermediate carboxy-phosphate anhydrides (e.g. **N3'P(OAc)-2'OH**) behave differently such that, **N3'P(OAc)-2'OH** favours attack at carbon rather than phosphate and the **N2'P** intermediate attacks either phosphate or carbon.

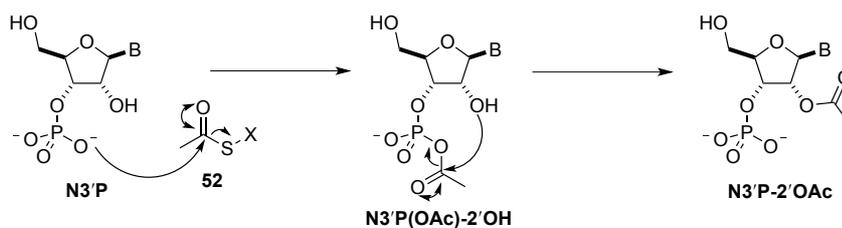


Figure 28. Proposed mechanism of acetylation of  $N3'P$  by activated thioacetate **52** ( $X =$  leaving group).

The acetylation chemistry was also shown to be selective for 3'-phosphate terminated dimer and trimer oligonucleotides. The chemistry was then extended to oligonucleotides in order to determine whether selective acetylation of 3'-phosphate terminated RNA would assist templated-ligation to form the natural 3',5'-linkage. Oligonucleotide sequences used by Rohatgi *et al.* (see Chapter 1.6.1, Figure 21) were chosen to allow comparison of the chemistry; the primer in this work terminated in an adenosine-3'-phosphate. Thus, an upstream 10nt 'primer' was treated with *N*-acetylimidazole (NAI) **48** that was found to be >70% acetylated (estimated from mass spectrometry peak integrals). This acetylated primer was mixed with a downstream 7nt 'ligator' strand and a 13nt 'template' RNA, and the resultant gapped duplex was activated for ligation with *N*-cyanoimidazole (NCI) **47** (Figure 29). Denaturing gel electrophoresis revealed successful ligation of the primer and ligator to give the 17nt product; control reactions showed that acetylation followed by phosphate activation was required for efficient ligation. Further experiments with a 5'-fluorescently labelled primer demonstrated that a yield of 49% of the 17nt ligation product was obtained after 19 hours. Ligation was also achieved when the primer was acetylated as part of a gapped duplex, albeit with a lower yield of 23% of the 17nt product (Figure 29). In comparison to Rohatgi's ligation of 5'-triphosphates, the yields were an order of magnitude higher and the rate of ligation was far greater (ligation of the gapped duplex was essentially complete after 4 hours).

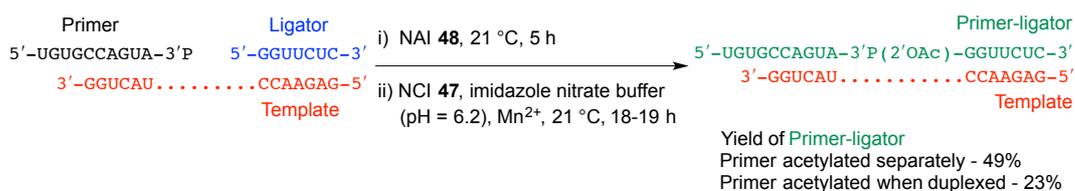
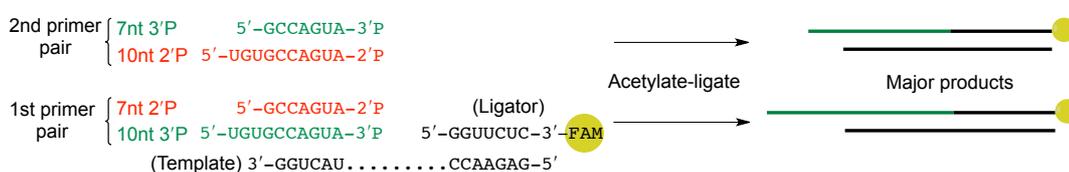


Figure 29. Efficient ligation of the chemoselectively acetylated 3'-phosphate oligonucleotide primer is high yielding both when the primer is acetylated separately and as a gapped duplex.

The relative selectivity for ligation of 2'-P- and 3'-P-terminated RNA oligonucleotides was also assessed with the original 10nt 3'-P primer and a 7nt 2'-P primer to permit resolution by gel electrophoresis. To control for the differing overhang length, ligations of 7nt 3'-P primers and 10nt 2'-P primers were also conducted. In all cases a 3'-fluorescently labelled ligator was used. Using mass spectrometry and estimates from integration of the peak areas, the 3'-P primers were found to consistently give 2- to 3-fold higher acetylation yields. Ligation of these 3'-P primers was found to be highly selective, affording yields up to 700-fold greater than for 2'-P primers, including when the primers were in competition (Figure 30).



*Figure 30. Sequences of oligonucleotides that were used to assess the selectivity of ligation. Illustrated is the acetylation-ligation reactions, showing that the ligation of the 3'-P terminated primers were the major products.*

The partially acetylated-RNA molecules produced by this ligation chemistry required deacetylation to give the native RNA. To demonstrate this, a 13nt RNA oligonucleotide with an internal 2'-O-acetyl (5'-GCAGUA-3', 5'-(2'OAc)-GGUUCUC-3'), prepared using newly developed phosphoramidites and a solid-phase synthesis protocol described in this thesis, see Chapter 2) was subjected to ammonolysis (~5 M aq. ammonia) for 1 hour at pH = 9.2 and 40 °C. The products were examined by mass spectrometry and showed almost complete removal of the acetyl group without noticeable hydrolysis of the backbone. More prebiotically plausible conditions (~1 M aq. ammonia, 21 °C, 48 hours) also efficiently (but more slowly) removed the acetyl group. Importantly, this showed that removal of the acetyl group is much faster than hydrolysis of extant 3',5'-phosphodiester bonds.

The stability of the 3',5'- or 2',5'-phosphodiester bonds in the context of duplex was further assessed. Synthetic standards of the 17nt ligation products (synthesis described in Chapter 2) containing an internal 2'/3'-O-acetyl group at the ligation junction (*i.e.* both the acetylated 2',5'- and 3',5'-linkage isomers at the ligation junction) were synthesised. In the presence of the complementary 13nt template, the two 17nt ligation

products were analysed by HPLC pre- and post-ammonolysis. The results showed that the 3',5'-phosphodiester bonds were stable to the deacetylation conditions but that the single 2',5'-phosphodiester bond was susceptible to hydrolysis after deacetylation had occurred. It is suggested that the deacetylation step, required to reveal the native RNA, would provide a way to enrich for 3',5'-bonds that are more resistant to hydrolysis, thus allowing catalytic properties of RNA to emerge. The partial 2'-*O*-acetylated RNA is also hypothesised to favour duplex structure due to reduced A-minor interactions and increased North-type sugar puckering. Reduced secondary structure is thought to facilitate replication of partially 2'-*O*-acetylated RNA relative to RNA. The properties of partially 2'-*O*-acetylated RNA and consequences for replication are returned to in Chapter 3.

In summary, this work showed that the natural 3',5'-linkages in RNA are selected for by an initial chemoselective acetylation that favours ligation of 3'-phosphate terminated oligonucleotides. The required deacetylation is effectively accomplished by aqueous ammonia; the same deacetylation conditions lead to increased hydrolysis of 2',5'- *versus* 3',5'-linkages, thus providing a further pathway towards enrichment of the latter.

## 1.7. From RNA towards peptides

Modern biomolecular machinery required to synthesise peptides is complex and highly evolved. Key to the process are the ribosome, messenger RNA (mRNA), the full set of transfer-RNA (tRNA) molecules and 20 aminoacyl-tRNA synthetases (ARSs).<sup>[19]</sup> The ARSs are enzymes (proteins) crucial for the correct charging of amino acids (each amino acid has a unique ARS) onto the correct tRNAs, which is considered to be the most important step in accurate translation of the genetic code.<sup>[151, 152]</sup> The elucidation of how this biological process arose and evolved is a major goal for origins of life research.

The ARSs are responsible for essentially two chemical reactions.<sup>[19, 152]</sup> Firstly, an amino acid is activated with adenosine-5'-triphosphate (ATP) to form an aminoacyl-adenylate (aa-AMP) that is tightly bound to the active site (Figure 31a). Secondly, the ARS with the bound aa-AMP recruits its cognate tRNA and the activated amino acid is

transferred to either the 2'- or 3'-OH of the tRNA molecule to form aminoacyl-tRNA (aa-tRNA, Figure 31b).

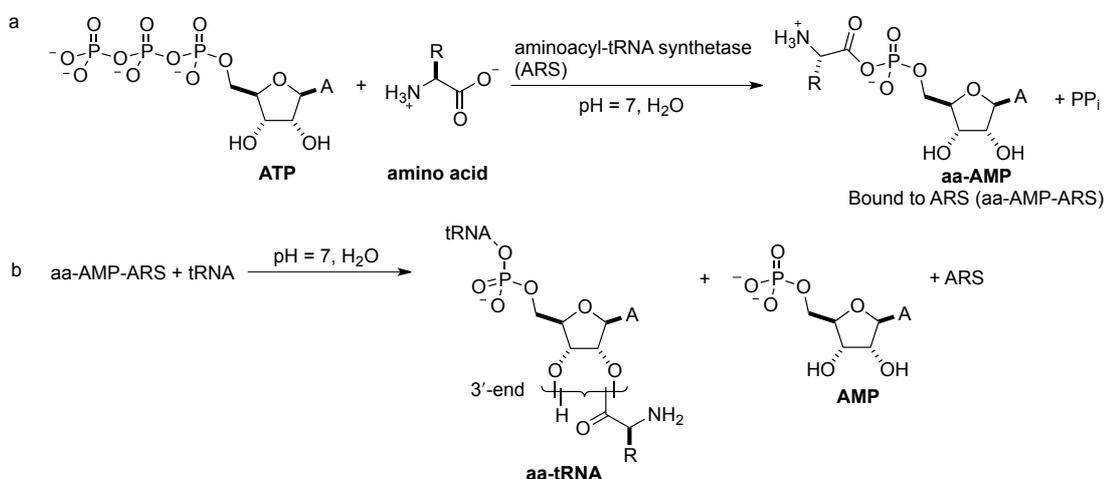


Figure 31. a) Formation of the activated aminoacyl-adenylate (aa-AMP). b) Transfer of the aminoacyl group to the 3'-terminus of tRNA. Both reactions are catalysed by the aminoacyl-tRNA synthetase.

ARSs achieve accurate aminoacylation of the corresponding tRNA by utilising particular molecular interactions in the amino acid binding site, before the tRNA is charged, such as hydrogen bonding to the hydroxyl group in threonine **53** to discriminate from valine **54** (Figure 32).<sup>[19]</sup> The ARS then in most cases carries out further 'proof-reading' of the charged tRNA, and if it is incorrect, the amino acid is hydrolysed from the tRNA.<sup>[153]</sup> In addition to these processes, the ARS also selects the correct tRNA by interacting with the anticodon loop and the acceptor stem (green portion of tRNA molecule in Figure 33). At the origin of life, a primitive translation system must have been in operation. The strongly supported idea of an RNA world has inspired several workers to investigate possible RNA-only aminoacylation systems and to also look for simplified systems.

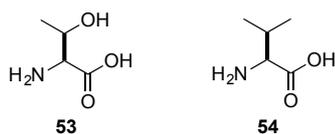


Figure 32. Threonine **53** and valine **54** are discriminated by hydrogen bonding to the  $\beta$ -hydroxyl.

### 1.7.1. The search for a primitive aminoacylation

Schimmel and co-workers have searched for minimal tRNA structures that are still aminoacylated by their cognate ARS. It was found that a single G:U basepair (3:70) in the acceptor stem of the tRNA<sup>Ala</sup> was the main determinant for the alanyl-ARS to recognise the correct tRNA.<sup>[152]</sup> Efficient aminoacylation could be achieved when the anti-codon and D-loops (and later also the T $\psi$ C-loop) were dispensed with to give a minihelix based solely upon the acceptor stem (Figure 33).<sup>[154]</sup> Further work has found other minihelix or minihelix-like structures that are substrates for nine ARSs, which include histidine, glycine, valine **54** and methionine.<sup>[152, 155, 156]</sup> Even smaller substrates that resemble the acceptor stem have been shown to be aminoacylated by their cognate ARS.<sup>[157]</sup> These smaller substrates were composed of a 4-basepair stem stabilised by a tetraloop, and the results indicated that for glycine, alanine and histidine only the first three basepairs are required to overcome other deleterious effects of minimising the tRNA. These studies suggested that tRNA may have evolved from smaller species, and that early synthetases discriminated between different primitive tRNAs by the base pairs closest to the amino acid attachment site: such that the anticodon was a later addition.<sup>[155]</sup>

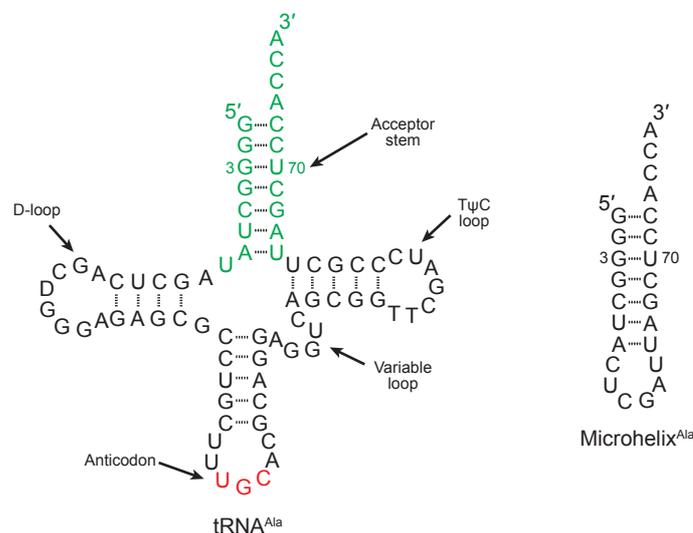


Figure 33. The main determinant for recognition of alanyl-tRNA by alanyl-tRNA-synthetase is the G:U (3:70) basepair. Reducing the tRNA structure to a microhelix based upon only the acceptor stem confirmed that the anticodon was not necessary for accurate recognition.

Schimmel and Tamura then looked for a way to aminoacylate these minihelices without the use of enzymes. They took a minihelix derived from *E. coli* tRNA<sup>Ala</sup> (minihelix<sup>Ala</sup>) and showed that its 3'-hydroxyl could be non-enzymatically aminoacylated by an aminoacyl donor oligonucleotide.<sup>[158, 159]</sup> This reaction was actually an aminoacyl-transfer where a chemically synthesised 5'-aminoacyl phosphate oligonucleotide was brought into proximity with the minihelix by hybridisation with a complementary bridging oligonucleotide (Figure 34). The approximate 15% yield of the aminoacylated minihelix<sup>Ala</sup> was limited by the hydrolysis of the aminoacyl-5'-phosphate oligonucleotide.

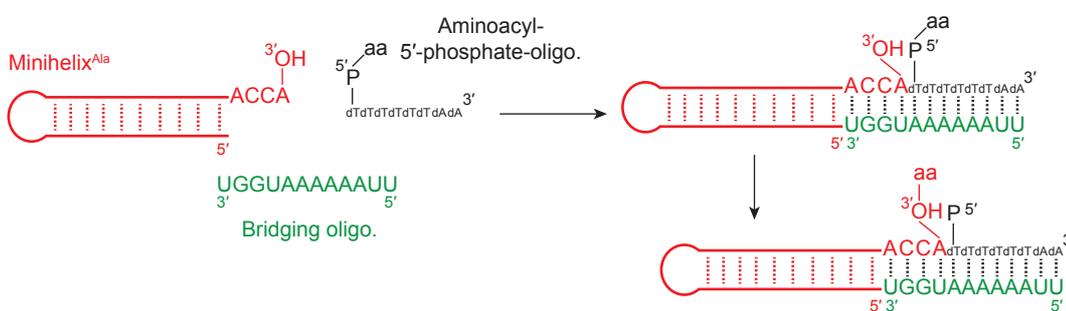


Figure 34. The aminoacylation of a minihelix by an aminoacyl-oligonucleotide-5'-phosphate brought together by hybridisation to a bridging oligonucleotide.

Although the work by Schimmel described an non-enzymatic aminoacylation of RNA, it was not catalytic and also used a DNA aminoacyl-donor oligonucleotide. In the search of a catalytic RNA aminoacylation, Yarus *et al.* isolated RNAs that were able to catalyse their own aminoacylation.<sup>[160, 161]</sup> Randomised RNAs were incubated with phenylalanyl-5'-adenylate (Phe-AMP) and then the amino groups derivatised with a hydrophobic groups. This altered the chromatographic behaviour of any aminoacylated RNAs to allow separation from non-aminoacylated RNA. In this way a pool of self-aminoacylating RNAs was formed; in particular a 95nt RNA was isolated and named 'isolate 29'. This isolate 29 was subsequently found to aminoacylate its own 3'-terminus and was able to use activated amino acids such seryl-AMP and alanyl-AMP.<sup>[161]</sup> In a later publication, isolate 29 was reduced in size to give a 29nt oligonucleotide that was also found to aminoacylate itself.<sup>[162]</sup> Additionally, a second product characterised as the diphenyl-RNA suggested that small RNAs could have catalysed the formation of peptide bonds.

By using a different selection protocol (SELEX<sup>[163-165]</sup>) Yarus *et al.* were able to create a pool of self-aminoacylating RNAs of varying size that contained three conserved nucleotides at the aminoacyl-transfer site. Sequencing of these oligonucleotides suggested that the aminoacyl-transfer centre consisted of a helix-loop-helix junction with a 5'-NGU or longer loop and a 3'-U aminoacyl-acceptor (Figure 35a).<sup>[166]</sup> The most active self-aminoacylating RNA was found to be the C3 RNA (Figure 35a). Divalent metal ions were nonessential for activity, and the C3 RNA accepted several different activated aminoacyl species (Phe-AMP, Phe-UMP and Met-AMP) with yields of the aminoacylated-RNA ranging from 65-95%. Interestingly, when the C3 RNA was incubated with unnatural D-Phe-AMP, aminoacylation was much slower, suggesting the active site is stereospecific. This observation may be important for the emergence of biological homochirality. In the pool of RNA, the 3'-U was substituted with A, C and G and in all cases the rate of reaction was greatly reduced. Aminoacylation is most efficient when in the 5'-NGU loop sequence N = U (*i.e.* 5'-UGU in C3 RNA) but this position is tolerant of all four nucleobase residues. Through computational studies, the aminoacyl-transfer centre was found to interact only with the amino, carbonyl and phosphate groups of the phenyl-AMP. This suggests that any amino acid or phosphate-leaving group could be utilised and indicates a universal aminoacylating RNA.

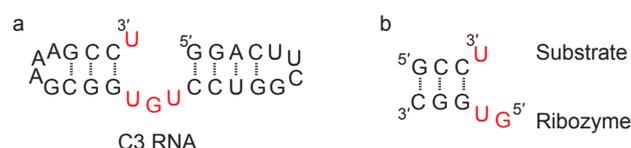


Figure 35. a) The C3 RNA that was one of the selected self-aminoacylating RNAs. In red are the conserved nucleotides. b) The small trans-aminoacylating RNA complex.

So far the aminoacylating RNAs discussed have not been true enzymes, as they either do not show multiple turn-over or the RNA is self-aminoacylated. Yarus *et al.* eluded to the possibility that the 5'-helix of C3 RNA was not crucial for activity.<sup>[166]</sup> In fact, this turned out to be the case and the C3 RNA was reduced to a 5nt oligonucleotide ribozyme that now functioned in *trans* to aminoacylate a tetramer substrate (Figure 35b).<sup>[167]</sup> The 5nt ribozyme shows similar behaviour to the C3 RNA in that it is selective for the 2'-OH of the 3'-diol of the substrate and the 5'-G is indispensable for activity. It is also capable of accepting different activated aminoacyl species (Phe-UMP and Met-AMP). When GCCU (20  $\mu$ M), the ribozyme GUGGC (10  $\mu$ M), KCl (100 mM)

and MgCl<sub>2</sub> (5 mM) were incubated with Phe-AMP (18.2 mM) at pH = 7 and 4 °C, multiple peptidyl products are observed.<sup>[168]</sup> These products include the full range of GCCU-polypeptides up to pentapeptides. Additionally, as the aminoacyl group is susceptible to 2'/3'-OH migration, GCCU-bis(2'/3'-O-Phe) and higher peptides are observed (Figure 36). Although, peptide bond formation is accelerated in this system, the ribozyme is not a peptide-bond forming catalyst. Rather the amino group of GCCU-Phe is thought to attack free Phe-AMP. The ribozyme is essentially catalytic as when the substrate:ribozyme ratio is 10:1, 50% of the substrates are aminoacylated. Thus, the ribozyme is acting on multiple substrates. The small size of this ribozyme suggests that catalytically active RNAs would have been in existence very soon after the oligomerisation of RNA monomers commenced. The work by Yarus shows the capabilities of small RNAs and also suggests that, once small RNAs were formed, aminoacylation and peptides would have shortly followed.

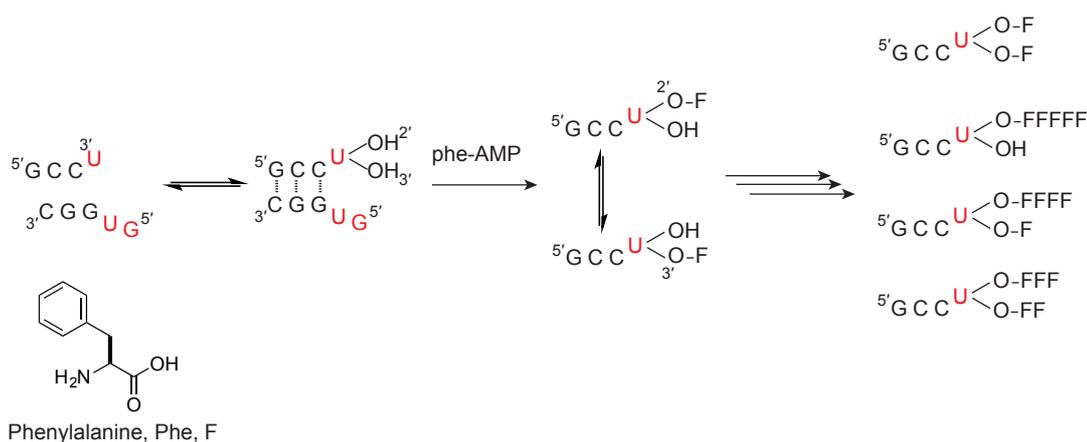
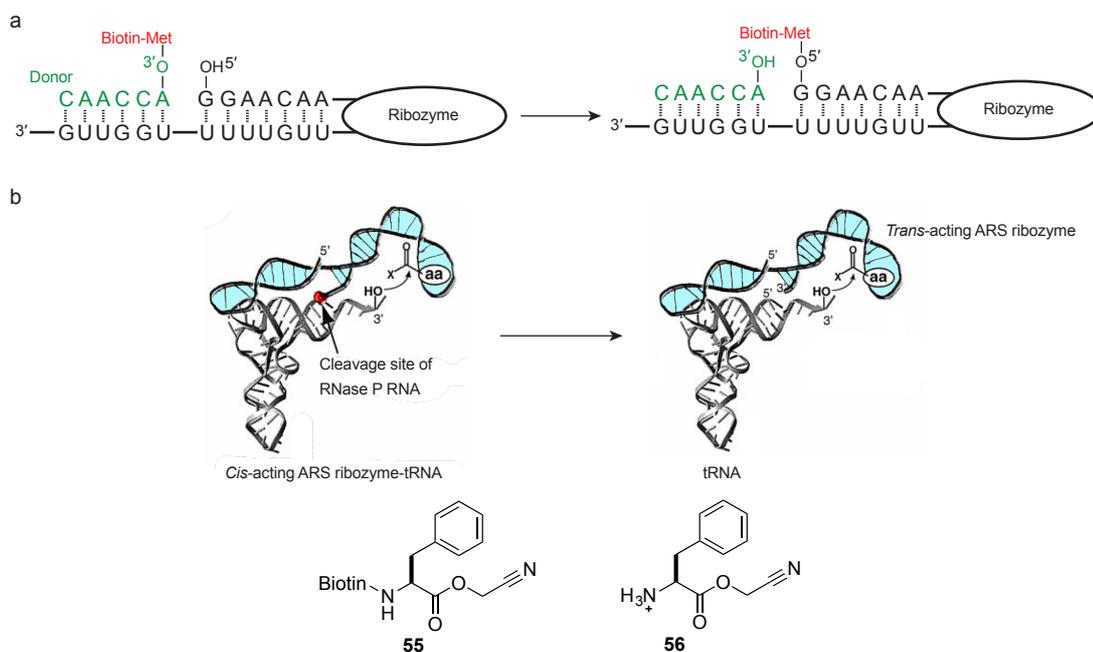


Figure 36. Aminoacylation of GCCU by the 5nt ribozyme, and a selection of the products formed.

An alternative aminoacylation system has been investigated by Suga *et al.* who found an acyl-transferase ribozyme that was capable of transferring an *N*-biotinyl-methionyl group from a donor 6nt oligonucleotide to the 5'-hydroxyl of the ribozyme itself.<sup>[169, 170]</sup> The rate of transfer was found to be 3 orders of magnitude higher than the comparable template-directed reaction and ~10 orders of magnitude faster than the untemplated acyl-transfer reaction (Figure 37a). Further development of this idea and chemistry over several years resulted in what are named 'flexizymes' (Fx), and their development is briefly discussed here. By using *in vitro* evolution and incorporating a 70nt variable region and 20nt constant region on to the 5'-end of a tRNA (pre-tRNA), Suga and co-

workers have directed evolution of a *cis*-acting ARS-like ribozyme (Figure 37b).<sup>[171]</sup> This *cis*-ribozyme was selective for the cyanomethylester-activated *N*-biotin-Phe-CME **55** and activity was also seen with Phe-CME **56**. The *cis*-5'-ribozyme was then cleaved from the pre-tRNA by RNase P and it was shown to aminoacylate the 3'-end of tRNA in *trans* whether cleaved from the pre-tRNA or separately transcribed *in vitro*. The *trans*-5'-ribozyme was also able to aminoacylate a minihelix consisting of the acceptor:T-stem:loop region of tRNA (see Figure 33 for tRNA nomenclature). The *trans*-5'-ribozyme is regioselective as it recognises the 4nt 3'-terminus of tRNA (5'-GCCA-3' in the tRNA used in this case) in a similar manner to protein ARS. In particular, the CCA sequence is critical for activity as single mutations reduced activity by 3.3- to 5-fold. Moreover, it has also been found to selectively aminoacylate the 3'-hydroxyl of the 2'/3'-diol of the terminal A residue of the tRNA.<sup>[172]</sup> However, one issue with the *cis*- and *trans*-5'-ribozyme is that it can only charge aromatic amino acids onto the specific tRNA used in their development. A 45nt ribozyme called flexizyme3 (Fx3) was thus developed and subsequently shown to utilise asparaginyl-CME to aminoacylate with a variety of different tRNAs, and crucially with multiple turnover.<sup>[173]</sup> The Fx system has been optimised a great deal and can aminoacylate a wide variety of natural and unnatural aminoacids.<sup>[174]</sup> Although, the direction of this work travelled away from the premise of an origin of protein synthesis, it does show the power and capabilities of relatively small RNA enzymes. Moreover, it supports the idea that aminoacylating ribozymes could have existed and supported some kind of peptide bonding forming system.



### 1.7.2. A linked prebiotic origin of RNA and coded peptides

Small RNAs have been shown to catalyse or promote aminoacylation reactions and minimal substrates have been used for this purpose (Chapter 1.7.1). The RNA world hypothesis demands that RNA was the first aminoacylation catalyst. However, it is difficult to theorise how RNA passed the function of catalysis over to proteins because RNA would have to invent translation, and then in some way pass it on to coded/preformed proteins. An alternative scenario was considered, and by detailed analysis of the genetic code, Sutherland and co-workers have suggested a theory of a RNA:coded peptide subsystem. This theory is based on aminoacyl-RNA trimers that links coded protein synthesis to simultaneous RNA replication.<sup>[138, 175]</sup>

There are several key features of the genetic code that were noted soon after its resolution (Figure 38):<sup>[49]</sup>

- The genetic code is read in triplets.
- A triplet code has the capacity to code for 64 amino acids but only 20 are used.
- The 20 amino acids are not randomly distributed.
- XYU and XYC always code for the same amino acid.
- XYA and XYG usually code for the same amino acid.
- XYN (N = any base) codes for the same amino acid in half the cases.
- In some cases, there is a relationship between the second base of the codon and the chemical nature of the amino acid side chain.
- Structurally similar amino acids tend to have codon sets connected by single nucleotide changes.
- Biosynthetically related amino acids sometimes have codon sets connected by single nucleotide changes.
- The code is essentially universal.

Several theories have been suggested to explain how the genetic code was assigned. The ‘frozen accident’ hypothesis states that codon assignments were initially random and became fixed in the last common ancestor.<sup>[49]</sup> This theory is not well supported as the genetic code is not strictly universal and nor does the theory explain the ordered features of the code. The ‘adaptation theory’ attempts to explain some of these aspects and suggests that the code was assigned to minimize mutation or mistranslation. The ‘historical theory’ proposed a gradual assignment of codons as amino acids became available by biosynthesis.<sup>[176]</sup> Lastly, the ‘stereochemical theory’ suggests that chemical interaction between codons and/or anticodons with the side chains of the amino acids influenced the assignment of the genetic code.<sup>[177-179]</sup> By using the stereochemical and historical theories outlined above, Sutherland was able to propose an ancient genetic code that was simpler than the modern code.<sup>[138]</sup>

		Second Base						
		G	A	U	C			
G	Gly		<b>Asp</b>	2	Val		Ala	U
	Gly	2	<b>Asp</b>	2	Val	1	Ala	C
	Gly		<b>Glu</b>	1	Val		Ala	A
	Gly		<b>Glu</b>		Val		Ala	G
A	Ser	2	<b>Asn</b>	2	Ile		Thr	U
	Ser		<b>Asn</b>		Ile	1	Thr	C
	Arg	1	<b>Lys</b>	1/	Ile		Thr	A
	Arg		<b>Lys</b>	2	<b>Met</b>	1	Thr	G
U	<b>Cys</b>	2	<b>Tyr</b>	1	<b>Phe</b>	2	Ser	U
	<b>Cys</b>		<b>Tyr</b>		<b>Phe</b>		Ser	C
	<b>Stop</b>	1	<b>Stop</b>		Leu	1	Ser	A
	<b>Trp</b>		<b>Stop</b>		Leu		Ser	G
C	Arg		<b>His</b>	2	Leu		Pro	U
	Arg	1	<b>His</b>		Leu	1	Pro	C
	Arg		<b>Gln</b>		Leu		Pro	A
	Arg		<b>Gln</b>	1	Leu		Pro	G

Figure 38. The 'universal' genetic code. Family boxes are highlighted (bold) and allocation of the aminoacyl-tRNA synthetases (ARSs) to either class one or two (bold numbers). Amino acids highlighted in red are those that are deemed to be late additions, had low prebiotic abundance or break the class rule.

The aromatic amino acids (Trp, Try and Phe) require long and complex biosynthetic routes, as do His, Lys and Met. But since early metabolic machinery is expected to have been crude, only abiosynthesis of amino acids would have been possible so the above are thought to be late additions. Aminoacyl-tRNA synthetases (ARSs) are split into two classes, class-I and class-II (Figure 38). Across the three kingdoms each amino acid is associated with a particular class of ARS (the so called class rule) and these relationships are thought to have been established early on in evolution. Violations of the class rule are thus more likely with amino acids assigned to codons more recently. Eukarya and most bacteria possess class-II lysyl-tRNA synthetase but most archaea possess class-I enzymes. Also, most amino acids are charged directly by their cognate tRNA synthetase, but exceptions exist. Asparaginyl-tRNA and glutaminyl-tRNA can be made directly from Asn and Gln or by transamidation of Asp and Glu post-aminoacylation. Cysteinyl-tRNA can be synthesised directly or by additional activity of class-II proyl-tRNA synthetases.<sup>[180]</sup> These charging discrepancies for Gln, Asn, Lys and Cys break the class rule and are also deemed to be late codon assignments.

If these late additions and stop codons (Met being the exception) are examined, a pattern emerges where these amino acids are allowed with codons in which the second base is an A (XAZ) or the first base is a U (UYZ) (Figure 38). Thus, if XAZ and UYZ

codons are removed, AUG (Met codon) is pre-assigned to Ile, and AGN to either Ser or Arg a much simpler genetic code is revealed (Figure 39). The resulting groups of XYN that encode for the same amino acid are called family boxes (bold outline) and appear to be those that were the earliest assigned amino acids.

		Second Base				
		G	A	U	C	
First Base	G	Gly Gly <b>2</b> Gly Gly		Val Val <b>1</b> Val Val	Ala Ala <b>2</b> Ala Ala	U C A G
	A	Ser or Arg <b>1</b>		Ile Ile <b>1</b> Ile Ile	Thr Thr <b>2</b> Thr Thr	U C A G
	U					U C A G
	C	Arg Arg <b>1</b> Arg Arg		Leu Leu <b>1</b> Leu Leu	Pro Pro <b>2</b> Pro Pro	U C A G

Figure 39. The postulated simplified genetic code.

The simplified code gives much better ARS class correlation (class-I for XUZ, class-II for XCZ). It also improves relationship between codons (or anti-codons) and amino acids. XUZ now codes for hydrophobic, branched aliphatic amino acids side chains. XCZ is linked to small hydrophobic amino acids (where emphasis is placed on the Me group rather than the OH group of Thr) and XGZ is associated with Gly, Arg and Ser or Arg (depending on assignment of AGN). This retrosynthesis of the genetic code thus further points towards a stereochemical basis for its origin where the amino acid can be selected through direct chemical interaction with the first and second base of a triplet codon. This proposed early code is now based on coding by XYN (where X is –U and Y is –A).

With a simplified genetic code now proposed, Sutherland then suggested a mechanism by which templated oligomerisation of 2'/3'-aminoacyl-RNA trimers and tandem protein synthesis could be achieved. This mechanism suggests that coding was achieved by a 'folded-back' conformation where intramolecular interaction of the amino acid side chain was with the first two bases of the trimer. Base pairing of the 'folded-back' 2'/3'-aminoacyl-RNA trimers with a template would bring it into proximity with an extended

peptidyl-RNA and this would allow peptidyl transfer and phosphodiester bond formation. Once complete, the bonded ‘folded-back’ peptidyl-RNA trimer could unfold, base pair and thus enable further reaction.

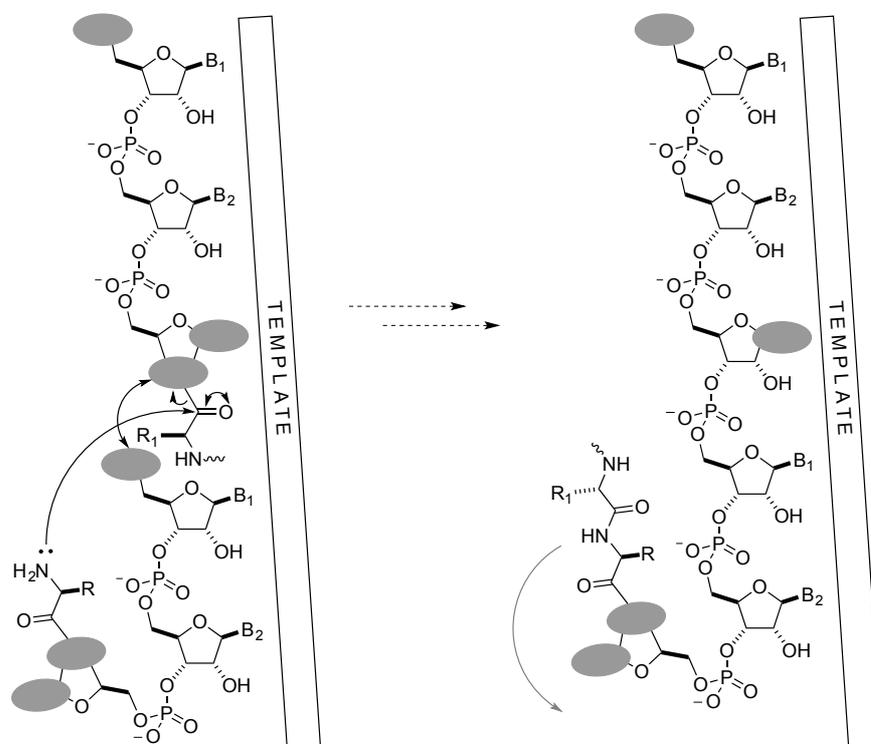


Figure 40. Postulated linked origin of RNA replication and coded peptide synthesis. Grey ovals represent unknown chemistries.

Through consideration of the possible chemistries that could work (Figure 40 grey ovals) it was the activated 2'-aminoacyl trimer **57** that was thought to be the most promising (Figure 41). The trimer **57** was proposed to have formed from cyclic trinucleotides **58**. Prebiotic formation of **58** has been demonstrated to be catalysed by montmorillonite clay, with a high yield of the natural 3',5'-linkage formed.<sup>[181]</sup> The conformation of species such as **58** have been studied in solution.<sup>[182]</sup> The lowest energy conformer was found to be where the nucleobases were axial with respect to the 18-membered ring. Modelling suggests the  $\alpha$ -amino acids could contact with bases B<sup>1</sup> and B<sup>2</sup> and also bind *via* electrostatic interactions between the phosphates and the carboxylate and ammonium groups. Nucleophilic attack by the amino acid carboxylate at the phosphate would lead to ring opening of the cyclic trinucleotide **58** to give the aminoacyl-phosphate ester trinucleotide **59**. The aminoacyl group of **59** would then undergo intramolecular transfer to the 2'-hydroxyl to give **60** and then activation of the

phosphate would give **57** the required species for oligomerisation. Thus, activation of the 3'-phosphate of 2'-aminoacyl trimer **60** would allow the attack of a 5'-hydroxyl of another trimer and give chain elongation with formation of the natural 3',5'-linkage.

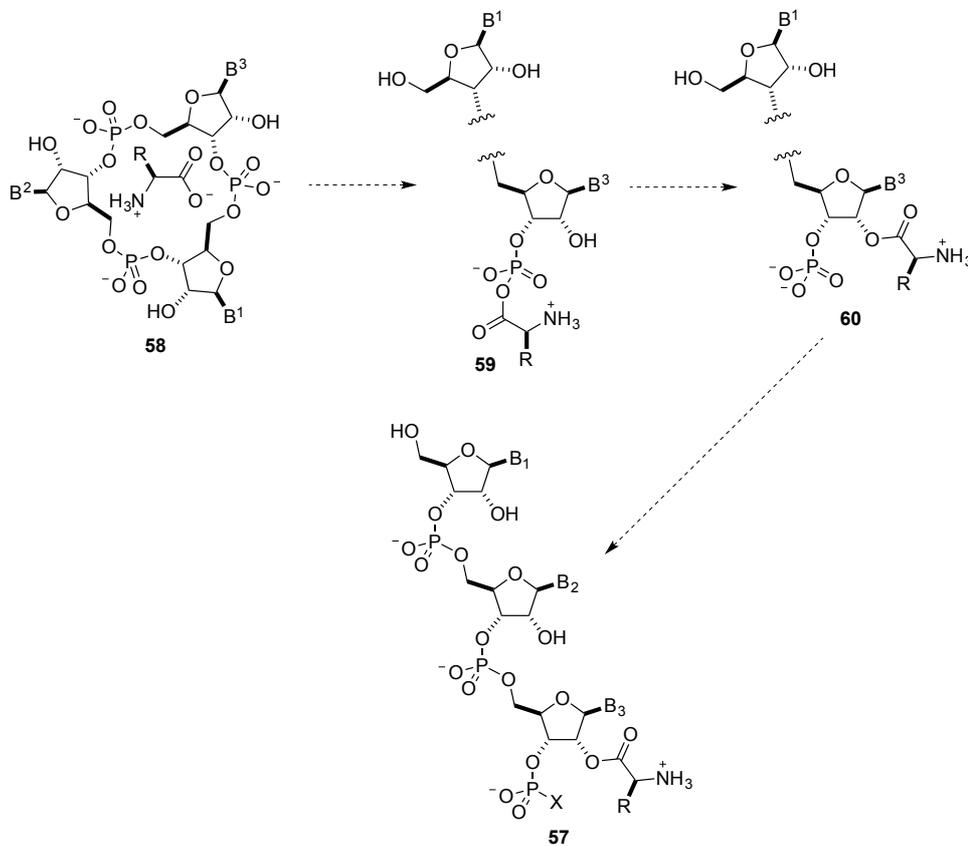


Figure 41. A proposed prebiotic synthesis of aminoacyl-RNA trimers **57** from cyclic trinucleotide **58**.

## 1.8. Project aims

1. A recent report from the Sutherland group demonstrated the successful prebiotically plausible ligation of oligonucleotides mediated by chemoselective acetylation. To support this work, synthetic standards of the partially 2'/3'-*O*-acetylated-oligonucleotide products were required. However, no commercially available synthetic precursors or procedures were available for their solid-phase synthesis. In addition, commercial starting materials, reagents and protocols were not orthogonal with a sequence possessing internal 2'- or 3'-ester groups. Thus, chapter 2 describes the design and synthesis of an acetyl-orthogonal protecting group strategy for the protection of RNA phosphoramidites. The synthesis and preparation of a photolabile solid-phase linker is described. The protocols and procedures were extensively developed to optimise the solid-phase synthesis, work-up and purification of the partially 2'/3'-*O*-acetylated-oligonucleotides.
2. The properties of partially 2'/3'-*O*-acetylated oligonucleotides were proposed to have aided replication by favouring duplex structure. With an optimised synthesis of partially acetylated-oligonucleotides available, several acetylated oligonucleotides were synthesised. The  $T_m$  and thermodynamic parameters of these oligonucleotides were investigated by UV spectroscopy to assess their potential for replication over native RNA.
3. The aminoacylation of RNA is an important process in modern biology and for a linked origin of RNA and coded peptides. Given that the activation of thioacetic acid by various electrophiles is in many cases very efficient and selective, this chemistry was applied to the activation of amino thioacids, and the subsequent aminoacylation reactions were investigated.

## 2. Solid-phase synthesis of 2'/3'-*O*-acetylated RNA oligonucleotides<sup>c</sup>

### 2.1. Background

The recent report by Bowler *et al.*<sup>[146]</sup> has shown that the ligation of short oligonucleotides can form the natural 3',5'-linkage isomerism mediated by a chemoselective acetylation. The products of this chemistry are partially 2'/3'-*O*-acetylated-RNA oligonucleotides (acetyl-RNA) and it was from this work that the need for a conventional organic synthesis of acetyl-RNA arose. The ability to synthesize acetyl-RNA would allow full control over their sequence, number and position of acetyl groups. Acetyl-RNA could then be synthesised to serve as synthetic standards to confirm the identity of the products arising from the acetylation-ligation chemistry described above. Additionally, the effect of acetyl groups on properties such as duplex stability or tertiary structure could be investigated in context of partially 2'/3'-*O*-acetylated RNA oligonucleotides as a possible precursor to extant RNA in the prebiotic world. It was this impetus that led to the desire to begin the development of chemistry to produce partially 2'/3'-*O*-acetylated-RNA oligonucleotides, which is described herein.

#### 2.1.1. Incompatibilities of conventional RNA oligonucleotide synthesis

The most common and established strategy for the automated synthesis of RNA has been the use of 2'-*O*-TBDMS (*tert*-butyldimethylsilyl) phosphoramidite chemistry and solid-phase immobilisation of the oligonucleotide.<sup>[183]</sup> Solid supported assembly of ribonucleotides occurs in a stepwise fashion (Figure 42). To begin the synthesis, the first support-bound nucleoside is deblocked by a strong acid exposing a 5'-hydroxyl and releasing a trityl cation. The next nucleoside phosphoramidite is coupled to the newly exposed 5'-hydroxyl in the presence of a weakly acidic activating agent such as 1*H*-tetrazole. Following the coupling step is a capping step that involves treatment with acetic anhydride to 'block' any unreacted 5'-hydroxyl groups and reduce the generation

---

<sup>c</sup> This chapter was conducted in collaboration with Dr. Colm D. Duffy and Dr. Jianfeng Xu. Particular contributions shall be noted within the text.

of truncated sequences. The newly formed phosphite triester is subsequently oxidised with an aqueous solution of  $I_2$  in the presence of a weak base. At this point the steps can be repeated in an iterative cycle to build the RNA oligonucleotide with the desired sequence. Once complete the final DMTr can be left intact or removed depending on the choice of oligonucleotide purification.

The fully protected solid-support bound oligonucleotides are then subjected to a solution of ammonium hydroxide to deprotect the nucleobases and cleave the oligonucleotide from the solid-support. The TBDMS protected oligonucleotide is redissolved in a suitable solvent such as DMSO and treated with a fluoride reagent most commonly triethylamine trihydrofluoride (TREAT.HF). Upon precipitation the oligonucleotide is fully deprotected and the full-length product can be purified by methods such as HPLC or polyacrylamide gel electrophoresis (PAGE).

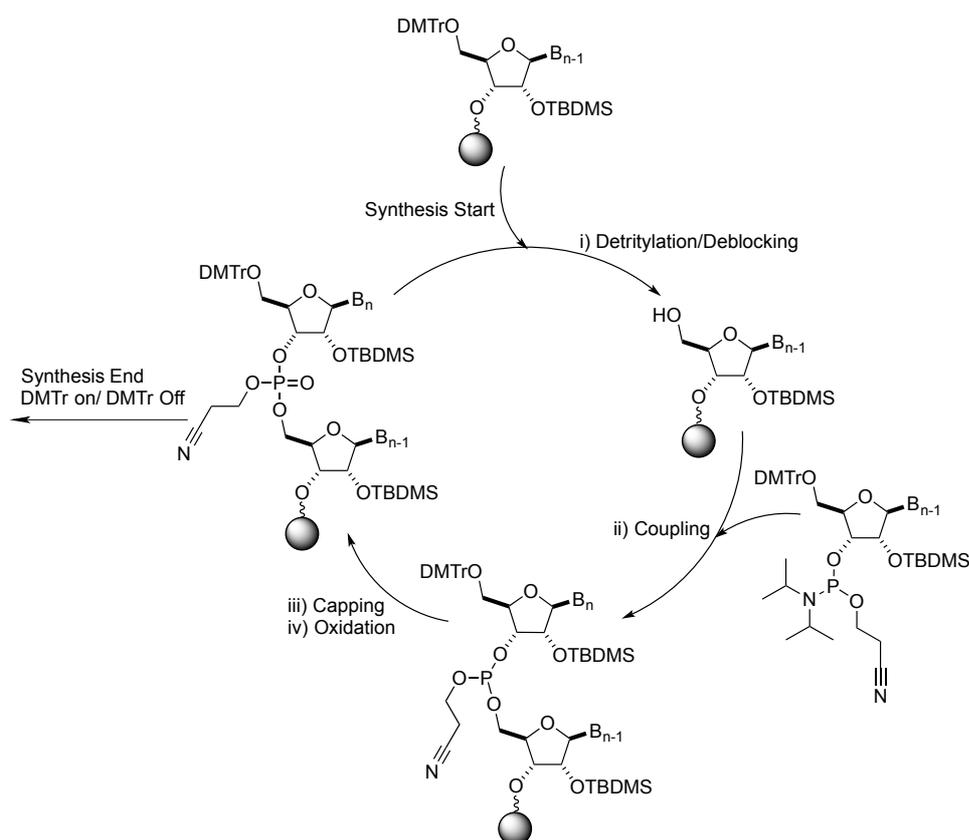


Figure 42. The steps in a typical cycle for the synthesis of ribonucleotides by the phosphite triester method.

The commercial products for RNA synthesis commonly use acyl, formamidine and the so-called 'UltraMILD' protecting group strategies for the protection of the nucleobase

exocyclic amines of A, C and G. The acyl protecting groups include acetyl, benzoyl and isobutyryl and removal of these protecting groups employ a concentrated aqueous ammonia solution with heating. However, this has seen less use with  $\text{NH}_4\text{OH}$ /methylamine (AMA) solutions becoming the standard due to the reduction of 2'-*O*-TBDMS loss during nucleobase deprotection (Figure 43a).<sup>[184-186]</sup> These conditions are also used for the deprotection of formamidine protecting groups, such as *N*-dimethylformamidine (dmf), which are used for the protection of the exocyclic amines of purine nucleobases (Figure 43b). Acetyl protected C is generally used in conjunction with formamidine protected purines.<sup>[187]</sup> The UltraMILD protecting groups are used for protection of the purine nucleobases and are differentiated by their ease of deprotection and are employed where base-sensitive nucleobase modifications are used or faster oligonucleotide deprotection times are required.<sup>[188]</sup> The UltraMILD protecting groups consist of an oxyacetyl such as phenoxyacetyl (Pac) or isopropylphenoxyacetyl (iPr-Pac) and for deprotection employ  $\text{K}_2\text{CO}_3/\text{MeOH}$  or an ammonia solution without heating (again acetyl protected C is used with UltraMILD) (Figure 43c).<sup>[189, 190]</sup> The common feature of these protecting groups is the requirement for nucleophilic bases for their removal, which is incompatible with the goal to synthesis 2'/3'-*O*-acetylated-RNA oligonucleotides (hereafter referred to as acetylated-RNA oligonucleotides).

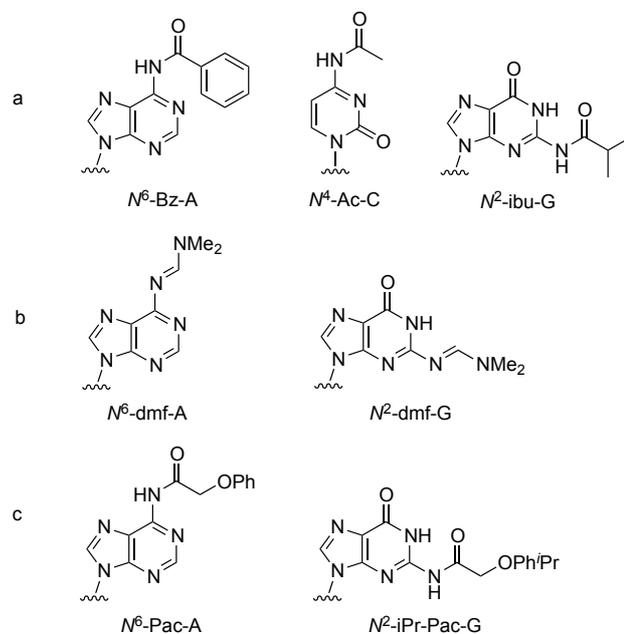


Figure 43. The a) acyl protecting groups, b) dmf protecting groups for purine nucleobases and c) UltraMILD protecting groups for purine nucleobases.

In addition to 2'/3'-*O*-acetylated phosphoramidites, the synthesis of partially acetylated-RNA oligonucleotides would also require phosphoramidites with alternative 2'/3'-hydroxyl protection that upon removal would give the free 2'/3'-hydroxyl. The extensively used TBDMS group is orthogonal to 2'/3'-acetate groups as deprotection is easily accomplished under mild conditions with the use of TREAT.HF.<sup>[191]</sup> Protection of 5'-hydroxyls is commonly with a trityl such as the 4,4'-dimethoxytrityl group (DMTr). This protecting group is ubiquitous and has the added advantage of enabling the coupling efficiency of each step to be monitored during the synthesis of the oligonucleotides. Removal of the trityl group is accomplished with a non-aqueous acid such as dichloroacetic acid (DCA) or trichloroacetic acid (TCA) and is thus compatible with a 2'/3'-acetate group (Figure 44).<sup>[192]</sup>

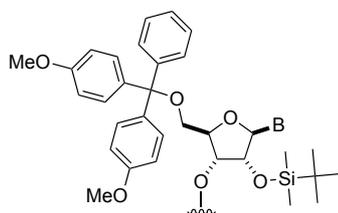


Figure 44. The 4,4'-dimethoxytrityl group used for 5'-hydroxyl protection and tert-butyldimethylsilyl ether used for 2'/3'-hydroxyl protection.

Since the report by Sinha *et al.* detailing synthesis of 2-cyanoethyl (ce) phosphoramidites, the 2-cyanoethyl protection of phosphites has found extensive use in the synthesis of RNA and DNA oligonucleotides (Figure 45). The advantages over previous phosphoramidites were ease of deprotection (within the time needed to remove the nucleobase protecting groups) and their stability as they could be stored for more than six months.<sup>[193]</sup> On completion of oligonucleotide synthesis the 2-cyanoethyl phosphate protecting groups are routinely removed under the same conditions that remove the nucleobase protecting groups. Deprotection of the 2-cyanoethyl groups proceeds by  $\beta$ -elimination<sup>[194]</sup> such that a variety of bases have been shown to be effective, including triethylamine (TEA) and methylamine. In particular a non-nucleophilic strong base, 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU), has been used in the synthesis of oligonucleotides.<sup>[195, 196]</sup> Thus, the use of a non-nucleophilic base for the deprotection of the 2-cyanoethyl groups should enable orthogonality with the 2'/3'-acetyl groups.

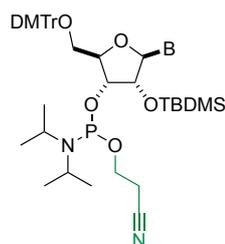


Figure 45. The 2-cyanoethyl-N,N-diisopropyl phosphoramidites most commonly used to synthesis RNA oligonucleotides, the 2-cyanoethyl phosphite protecting group is highlighted in green.

During automated synthesis the nascent oligonucleotide is attached the solid-phase support by a linker that is cleaved to during deprotection to release the oligonucleotide. The most common linker groups in commercial use are the succinate linker<sup>[197]</sup> and the Q-linker<sup>[198]</sup> which form an ester *via* the 2'- or 3'-hydroxyl of the first nucleoside. After cleavage of the oligonucleotide from the solid-support however these linkers result in a 2',3'-diol-terminated oligonucleotide. For maximum flexibility it was decided that a linker suitable for terminal 2'/3'-phosphorylation would be preferred. It was thought that if 2',3'-diols were required at a later stage the terminal phosphate could be easily removed enzymatically. Commercially available 3'-phosphate solid-supports use a  $\beta$ -eliminating sulphonyl linker and methylamine-ammonia (AMA) solutions are again recommended for the cleavage of these linkers.<sup>[199]</sup> Each of the above linker groups again require a nucleophilic base for cleavage and are not compatible with 2'/3'-acetyl groups (Figure 46).

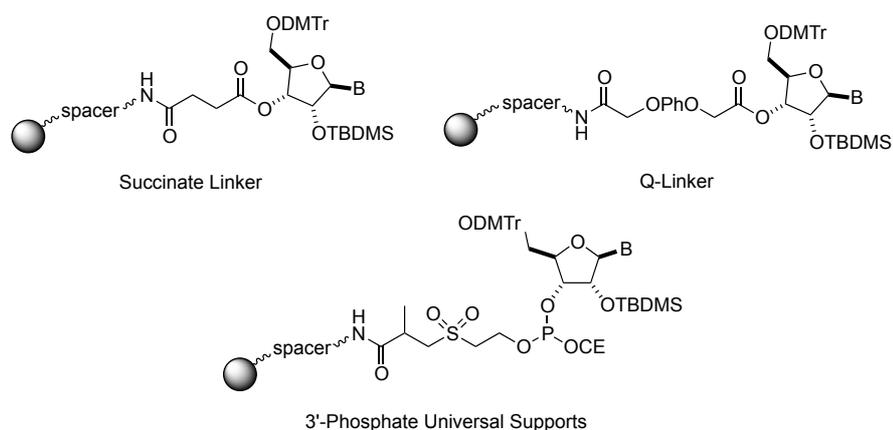


Figure 46. Three common supports used in the synthesis of oligonucleotides.

In summary, three areas of phosphoramidite protection are compatible with 2'/3'-acetyl RNA and the remaining two, upon deprotection or cleavage, will lead to removal of the desired 2',3'-acetyl groups, Figure 47 emphasises those sites of protection which require an alternative strategy:

1. 5'-Hydroxyl protection:
  - Acid-labile DMTr group.
  - Compatible if conditions are anhydrous.
2. 2'-Hydroxyl protection:
  - Fluoride-labile TBDMS.
  - Compatible if slightly acidic TREAT.HF used.
3. Phosphite/phosphate protection:
  - Base-labile 2-cyanoethyl group.
  - Compatible if base is non-nucleophilic.
4. Nucleobase protection:
  - Base-labile acyl or amidine type protecting groups.
  - Commonly requires a nucleophilic base such as methylamine for deprotection.
  - Not compatible.
5. Solid-support linker groups:
  - Succinate linker and Q-linker utilise acyl groups for linkage to oligonucleotide. Universal supports use an alkyl linkage to the oligonucleotide.
  - In most cases employ the conditions used for nucleobase deprotection for cleavage therefore nucleophilic bases are used.
  - Not compatible.

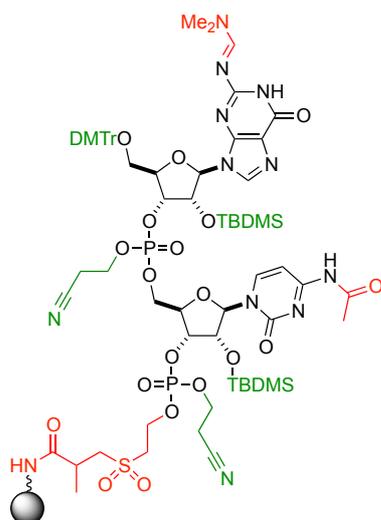


Figure 47. An oligoribonucleotide dimer synthesised by conventional chemistry prior to deprotection and cleavage. Green coloured protecting groups represent those groups that are 2'/3'-acetyl compatible and red that are not.

### 2.1.2. An acetyl compatible protecting group strategy

The analysis of current chemistry used in RNA oligonucleotide synthesis brought up two areas that needed an alternative protecting group strategy. Inspection of the literature revealed many nucleobase protecting groups but the majority require ammonia treatment for removal.<sup>[200]</sup> The chosen orthogonal protecting groups were the 2-cyanoethyl (ce), (2-cyanoethyloxy)carbonyl (ceoc) and (4-nitrophenyl)ethyl (npe) protecting groups developed by Pfeleiderer *et al.*<sup>[201, 202]</sup> These protecting groups are quantitatively removed by the strong non-nucleophilic base DBU in non-protic reaction conditions *via* a  $\beta$ -elimination process. As such deprotection conditions were thought to be compatible with the 2'/3'-acetyl groups. Additionally, the 2-cyanoethyl group is the same protection strategy that is employed for the protection of the phosphite and phosphate moieties<sup>[183]</sup> and so the nucleobase deprotection conditions can also be applied to the removal of the phosphate ce protecting groups.<sup>[195]</sup> In practice, it was envisioned the deprotection of both nucleobases and phosphates would be accomplished simultaneously, which parallels with current oligonucleotide deprotection methodologies. Trityl, TBDMS and acetyl orthogonality led to the decision that the strong-base labile nucleobase protection strategy was ideal for the synthesis of acetyl-RNA oligonucleotides.

The exocyclic amines of A and C were protected with the ceoc protecting group and the deprotections have been shown to be rapid with nucleoside half-lives of <15 minutes.<sup>[201]</sup> The exocyclic amine of G was also protected with ceoc however, the  $O^6$ -position required the npe protecting group for efficient nucleobase deprotection (Figure 48). The deprotection of  $N^2$ -ceoc-guanosine **61** was explored and the nucleoside half-life was found to be 8-9 hours (Figure 49). The unfavourable kinetics was suggested to be due to electronic effects whereby the anion formed on the nucleobase (N(1)-H deprotonation by DBU) hindered  $\beta$ -elimination of the  $N^2$ -ceoc. To solve this problem  $O^6$ -protection with the relatively more stable npe group was utilised in conjunction with  $N^2$ -ceoc. The  $O^6$ -npe group allowed  $\beta$ -elimination of the  $N^2$ -ceoc to occur first eliminating the formation of the anion on the nucleobase and resulting in a much shorter half-life of 30 minutes.

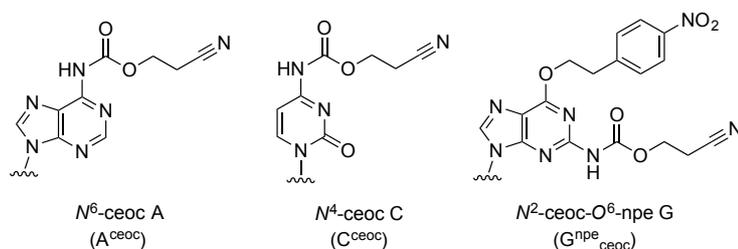


Figure 48. The exocyclic amines of A, C and G protected with the (2-cyanoethoxy)carbonyl (ceoc) protecting group. Due to kinetic reasons the  $O^6$ -position of G was also protected with the (4-nitrophenyl)ethyl (npe) protecting group.

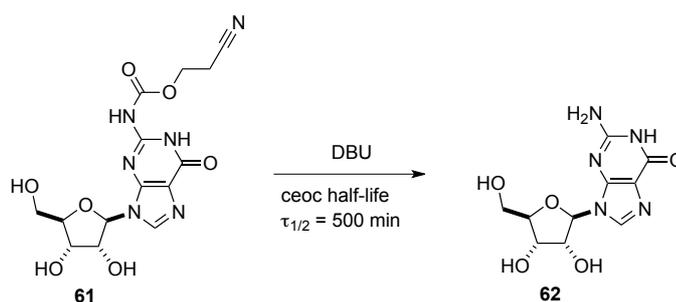


Figure 49. Slow deprotection of  $N^2$ -ceoc-guanosine **61**.

One of the issues foreseen with nucleobase and phosphate deprotection was removal of excess DBU and deprotection products (see section 2.4.2). DBU, if not fully separated from a fully 2'/3'- $O$ -protected oligonucleotide, in the presence of water would form hydroxide leading to deacetylation and possibly chain cleavage. The same could also

occur with the fully deprotected oligonucleotide where the 2'/3'-hydroxyl positions previously blocked by TBDMS would then be susceptible to chain cleavage under aqueous conditions.<sup>[203]</sup> The primary 5'-hydroxyl group would also need to remain DMTr protected to prevent processes such as acetyl migration during exposure to DBU.

Damha *et al.* had recently described a photo-labile linker **63** used for solid-phase synthesis of 2'-*O*-acetalester oligonucleotides that is orthogonal to ester and acetalesters (Figure 50). One key feature of the linker was that it was directly attached to an internucleotide phosphate which gave confidence to the proposal that it would enable synthesis of 2'/3'-phosphorylated RNA.<sup>[204]</sup> However, this linker group has  $\beta$ -protons to the linkage site such that it could also be susceptible to  $\beta$ -elimination upon exposure to DBU. A more DBU-base resistant linker group **64** was found that is structurally similar and is used predominantly in peptide synthesis.<sup>[205-207]</sup> This linker **64** has been employed for peptide synthesis on controlled-pore glass (CPG, this was the preferred solid-support, see section 2.3) and additionally its chemical stability towards DBU has been demonstrated during basic removal of Fmoc protecting groups.<sup>[207]</sup> Cleavage or photolysis of this class of *ortho*-nitrobenzyl-based groups has been previously demonstrated at wavelengths of 316-400 nm and no nucleobase modification has been reported at these wavelengths.<sup>[204, 208-210]</sup>

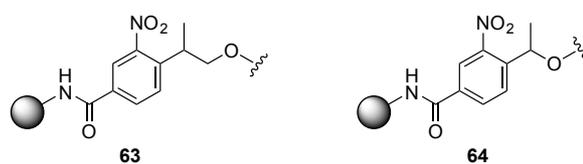


Figure 50. Photo-labile linker **63** developed by Damha and co-workers and the photolabile linker **64** that will be used in the current work.

Thus, the utilisation of the photolabile solid-support linker group **64**, which is resistant to basic conditions, allowed “on-column” deprotection of the nucleobases and phosphate groups.<sup>[196]</sup> The main advantage of the linker was to allow easy and thorough washing of the solid-support to remove all traces of DBU and deprotection by-products. After removal of the DMTr group, the TBDMS protected oligonucleotide could then be cleaved by photolysis and the TBDMS groups removed using standard procedures<sup>[211]</sup> resulting in the fully deprotected acetylated-RNA oligonucleotide.

From the discussion above, an 2'/3'-O-acetyl orthogonal protecting group strategy has been rationally designed and is summarised below (Figure 51):

1. On-column deprotection using DBU under anhydrous conditions should not lead to acetyl loss and removes:
  - Nucleobase protecting groups (ceoc and npe)
  - Phosphate protecting groups (ce)
2. On-column DMTr removal with anhydrous strong acid.
3. Light-labile linker enables oligonucleotide cleavage from the solid-support under mild conditions.
4. TBDMS groups can be removed in solution phase using standard procedures.

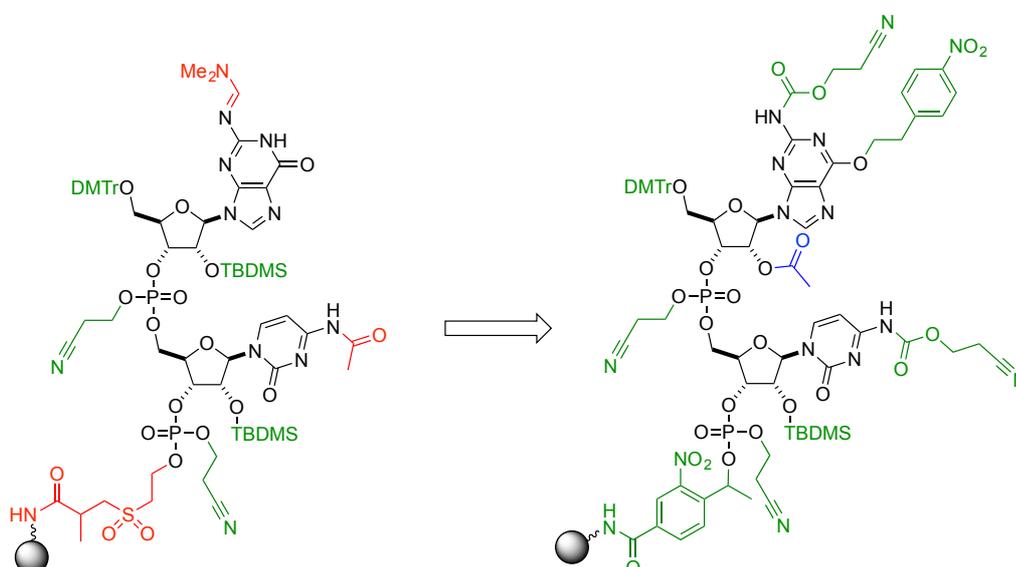


Figure 51. Protecting group and linker changes to the design of a 2'/3'-O-acetyl compatible protecting group strategy to enable the synthesis of 2'/3'-O-acetylated RNA oligonucleotides.

## 2.2. Synthesis of the phosphoramidites

### 2.2.1. Proposed synthetic route to the phosphoramidites

With a suitable protecting group strategy now designed the required phosphoramidites were established. From the four main RNA nucleobases A, C, G and U, phosphoramidites with regioisomeric 2'/3'-OAc and 2'/3'-O-TBDMS groups were required. This would allow synthesis of partially acetylated-RNA oligomers and also

enable synthesis of 3',5'- and 2',5'-linkages. These requirements give in total 16 final phosphoramidites **65-72** (Figure 52). Their proposed synthetic route was planned to proceed initially with the nucleobase protection. Once in hand the base-protected nucleosides would undergo tritylation with DMTr-Cl. At this stage the 2'/3'-diol were to be functionalised by either mono-acetylation or mono-silylation using silver salts.<sup>[212]</sup> The silylated regioisomers were to be separated and phosphitylated to yield the final TBDMS phosphoramidites. But for the 2'/3'-OAc regioisomers it is known that acyl 2'/3'-migration is quite facile so it was envisaged that separation of the two regioisomers would not be possible.<sup>[213, 214]</sup> Therefore, the phosphitylation of the acetylated-nucleosides would be conducted on the regioisomeric mixture of the mono-acetylated species prior to separation (Figure 53).

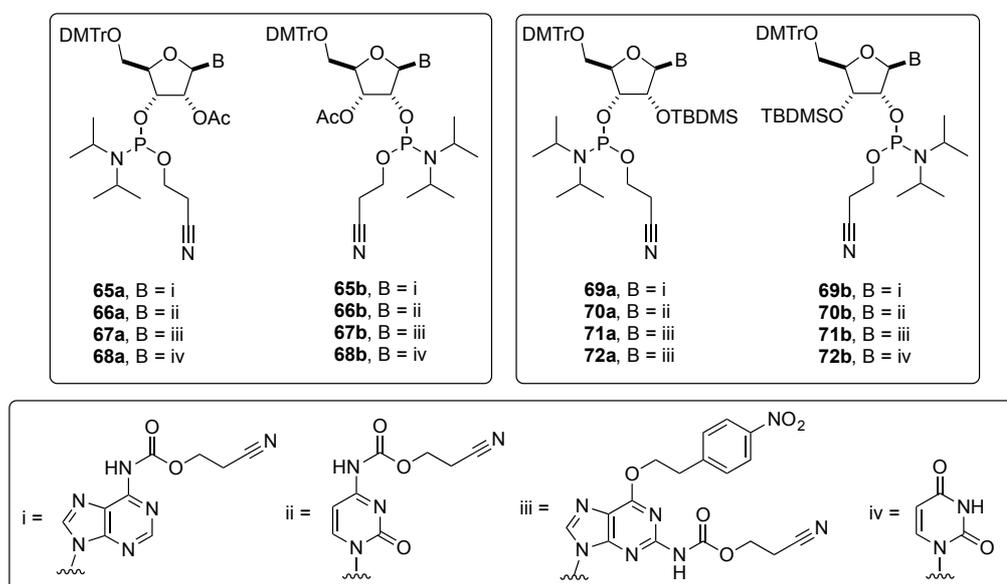


Figure 52. The 16 phosphoramidites to be synthesised.

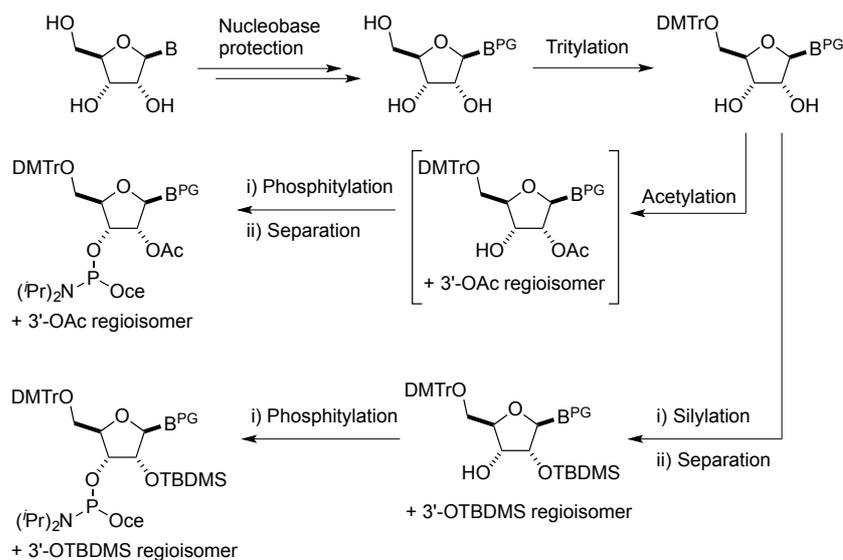


Figure 53. Proposed synthetic route to the 16 acetyl and TBDMS phosphoramidites.

## 2.2.2. Nucleobase protection

Pfleiderer *et al.* has described the protection of the nucleobases exocyclic amines of adenosine **11**, cytidine **73** and guanosine **74**.<sup>[201]</sup> Where the (2-cyanoethoxy)carbonylation reactions can be carried out with either 2-cyanoethyl carbonochloridate **75** or 1-((2-cyanoethoxy)carbonyl)-3-methyl-1*H*-imidazolium chloride **76**. The synthesis of these reagents involves the use of the highly toxic gas phosgene and as such a safer alternative was sought.

The solid reagent triphosgene, which on degradation forms three equivalents of phosgene, was chosen as the alternative to phosgene for the synthesis of **75**. As a solid, triphosgene could be handled and stored much more easily and would also negate the requirement for the condensation of a toxic gas. The synthesis of 2-cyanoethyl carbonochloridate **75** was conducted by overnight stirring of triphosgene and 3-hydroxypropanenitrile **77** according to a procedure based upon modified methods by Pfleiderer<sup>[201]</sup> and Wielser<sup>[215]</sup>. The chloridate **75** could then be used for carbonylations directly or carried over to the formation of the imidazolium salt **76**. The formation of **76** was carried out by addition of *N*-methylimidazole to a solution of **75** in CH<sub>2</sub>Cl<sub>2</sub> at 0 °C and stirring for 12 hours. The insoluble 1-[(2-cyanoethoxy)carbonyl]-3-methyl-1*H*-imidazolium chloride **76** was subsequently isolated by filtration.

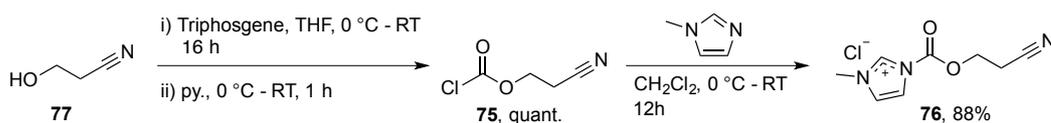


Figure 54. The synthesis of (2-cyanoethoxy)carbonylation reagents, 2-cyanoethyl carbonochloridate **75** and 1-[(2-cyanoethoxy)carbonyl]-3-methyl-1*H*-imidazolium chloride **76**.

With the (2-cyanoethoxy)carbonylation reagents at hand, the single step protection of adenosine **11** and cytidine **73** was begun by transient silylation of the hydroxyl groups of **11** and **73** by refluxing with an excess of hexamethyldisilazane (HMDS). After a change of solvent the TMS protected nucleosides were then treated for up to 48 hours with 1-[(2-cyanoethoxy)carbonyl]-3-methyl-1*H*-imidazolium chloride **76**. Workup involved hydrolysis of the TMS groups by treatment with methanol from which the base-protected nucleosides were precipitated to give good yields of **78** and **79** at 86% and 91% respectively (Figure 55).

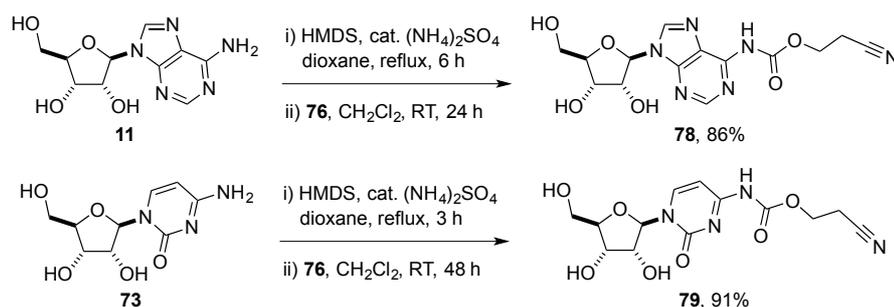


Figure 55. Synthesis of the (2-cyanoethoxy)carbonyl protected A and C, **78** and **79**.

Nucleobase protection of guanosine **74** required several steps to install both the *N*<sup>2</sup>-ceoc and the *O*<sup>6</sup>-npe protecting groups. The synthetic route as described by Pfeleiderer and coworkers<sup>[201, 216]</sup> involved first per-acylation of guanosine with isobutyryl chloride to give **80**.<sup>[217]</sup> Using Mitsunobu-type conditions, *O*<sup>6</sup>-alkylation is afforded by treating **80** with 1.5 equivalents each of diethyl azodicarboxylate (DEAD), triphenylphosphine (Ph<sub>3</sub>P) and *p*-nitrophenylethanol **81**.<sup>[218]</sup> The *O*<sup>6</sup>-alkylation product **82** is then treated with aqueous NH<sub>4</sub>OH over 6 days to remove the isobutyryl groups. In one pot the *O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-guanosine **83** hydroxyl groups are transiently silylated by treatment with trimethylsilyl chloride (TMS-Cl), followed by *N*<sup>2</sup>-carbonylation with 2-cyanoethyl carbonochloridate **76**. The TMS groups are then hydrolysed by treatment with methanol and the product **84** is isolated by precipitation (Figure 56).

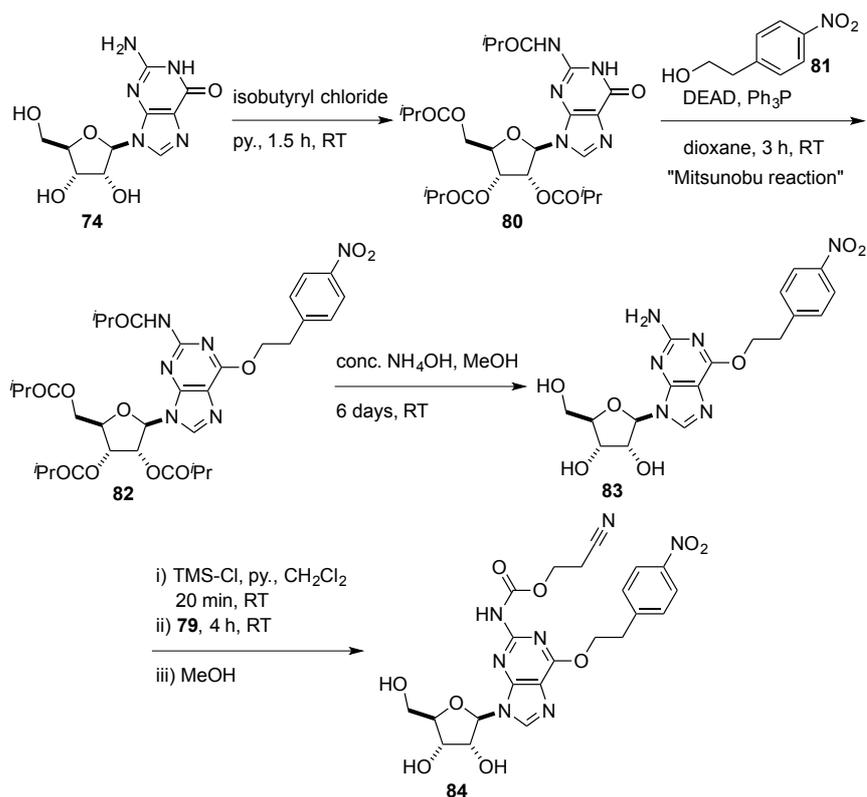


Figure 56. Literature procedure for the synthesis of **84**.<sup>[201]</sup>

Formation of **80** was high yielding (99%) and was then *O*<sup>6</sup>-alkylated to **82** under Mitsunobu reaction conditions, using the less explosive diisopropyl azodicarboxylate (DIAD), to give **82** in high yield (90%). Subjecting **82** to aqueous NH<sub>4</sub>OH lead to facile hydrolysis of the ester groups within several hours as monitored by TLC. However, hydrolysis of the *N*<sup>2</sup>-isobutyryl group proved to be difficult requiring exposure to concentrated NH<sub>4</sub>OH for more than 6 days. This long exposure to basic conditions led to slow deprotection of the *O*<sup>6</sup>-npe group as deduced by the appearance of a fast running spot on TLC assumed to correspond to *p*-nitrostyrene. Isolated yields of **83** (G<sup>npe</sup>) using isobutyryl protecting groups were no higher than 29%, giving an overall yield over three steps of 26%.

It was decided that utilising the more labile acetyl protecting group might reduce the time needed to deprotect the *N*<sup>2</sup>-position and so minimize the  $\beta$ -elimination of the *O*<sup>6</sup>-npe. Thus, the per-acetylation of **74** was preformed by treatment of a suspension of **74** in acetonitrile with acetic anhydride, DMAP and triethylamine with heating to 50 °C.<sup>[219]</sup> However, it was found that acetylation of the *N*<sup>2</sup>-position was sluggish and did to not proceed to completion. Purification of *N*<sup>2</sup>,2',3',5'-tetra-acetyl-guanosine **85** by

column chromatography was also difficult due to the closely running 2',3',5'-tri-acetyl-guanosine **86** and contamination by **86** was sometimes observed. Several variations of the reaction conditions were performed in the hope of improving the yield but maximal yields of only approximately 50% were obtained (Figure 57, Table 1).

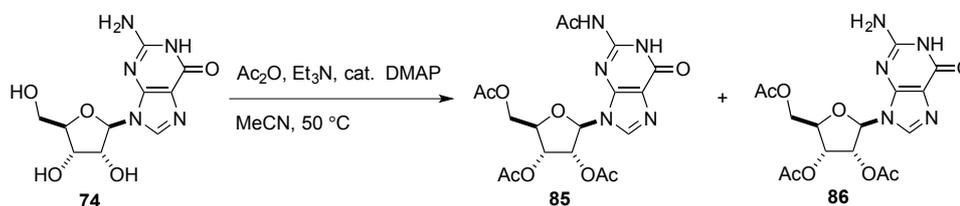


Figure 57. Reaction to per-acetylate guanosine and the incomplete acetylation to form 2',3',5'-tri-acetyl-guanosine **86**.

Entry	Equivalents to <b>74</b>		Time (h)	Isolated Yield (%)	
	$\text{Ac}_2\text{O}$	$\text{Et}_3\text{N}$		<b>85</b>	<b>86</b>
1	7	7.7	6	26	-
2	5	5.5	3	44	9
3	6	6.6	18	52	-

Table 1. Summary of yields of **85** and **86** the reaction were conducted in anhydrous acetonitrile and at  $50^\circ\text{C}$ .

Subjecting  $N^2,2',3',5'$ -tetra-acetyl-guanosine **85** to the Mitsunobu reaction yielded the  $O^6$ -alkylated compound  $N^2,2',3',5'$ -tetra-acetyl- $O^6$ -[2-(4-nitrophenyl)ethyl]-guanosine **87** cleanly and in high yield. With **87** to hand deacetylation was again conducted by treating with concentration aqueous  $\text{NH}_4\text{OH}$ . It was found that removal of the  $N^2$ -acetyl protecting group was faster than removal of the isobutyryl groups but still required exposure to aqueous  $\text{NH}_4\text{OH}$  for at least 72 hours at room temperature. Although, incubation times were shorter to remove the  $N^2$ -acetyl group, over the 72 hours dealkylation of the  $O^6$ -npe group was again observed. Thus, this led to less than ideal yields of  $O^6$ -[2-(4-nitrophenyl)ethyl]-guanosine **83** to a maximal isolated yield of 40% with an overall yield over three steps of 20%.

The problems with the previously discussed synthesis originate from the protection and deprotection of the exocyclic amine of guanosine **74**. It was found that acetylation of the exocyclic amine was difficult, which suggests a reduced nucleophilicity of the amine that is likely due to extensive delocalisation of the amine lone pair into the aromatic nucleobase (Figure 58a). This may also account for the difficulty encountered

when attempting the deacylation step by aminolysis whereby the  $N^2$ -amide bond is also further delocalised into the nucleobase. The increased delocalisation relative to a non-aromatic amide will reduce the magnitude of the dipole at the carbonyl leading a smaller  $\delta^+$  charge at the carbonyl carbon (Figure 58b). This in turn reduced the electrophilicity of the carbonyl carbon and so resulted in the slow ammonolysis reaction.

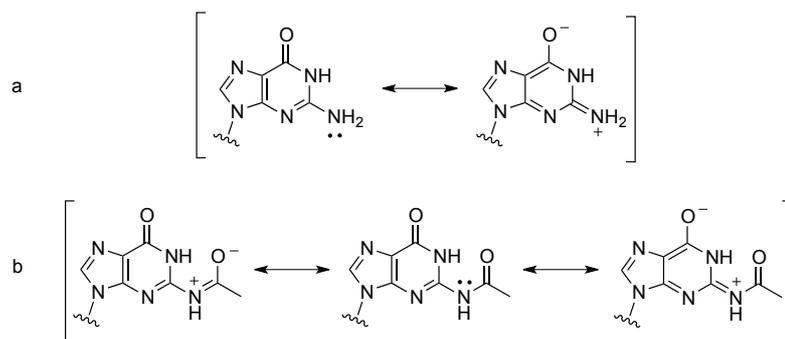


Figure 58. Delocalisation of the exocyclic amine lone pairs of a) guanosine, b)  $N^2$ -acetyl guanosine.

It was questioned whether  $N^2$ -protection was necessary for the mitigation of by-products during the Mitsunobu reaction and a literature search showed  $O^6$ -alkylation of **86** under Mitsunobu reaction conditions to be possible without extensive by-product formation.<sup>[220]</sup> An alternative route (Figure 59) to the base protected guanosine **84** began with treatment of **74** with 3.6 equivalents of  $\text{Ac}_2\text{O}$ ,  $\text{Et}_3\text{N}$  and catalytic DMAP for a maximum of 30 minutes to give 2',3',5'-tri-acetyl-guanosine **86** cleanly and in high yield by precipitation from propan-2-ol.<sup>[221]</sup> With **86** to hand,  $O^6$ -alkylation was afforded by first heating a suspension of **86**, with  $\text{Ph}_3\text{P}$  and *p*-nitrophenylethanol **81** in dioxane at 80 °C for 45 minutes. After addition of DIAD the solution was stirred at 60 °C for 4 hours to give 2',3',5'-tri-acetyl- $O^6$ -[2-(4-nitrophenyl)ethyl]-guanosine **87**.<sup>[220]</sup> Monitoring the reaction by TLC showed faint additional spots that were presumed to be by-products from side reactions due to the unprotected exocyclic amine of the nucleobase. Isolation of **87** by flash column chromatography was uncomplicated by the side products but it was found to co-run with significant amounts of triphenylphosphine oxide. The mixture was carried over to the next step where deacetylation was accomplish with concentrated aqueous  $\text{NH}_4\text{OH}$ , and because of the absence of any  $N^2$ -protection, deacetylation was complete within several hours.

Reduced exposure to basic conditions resulted in no observed loss of the  $O^6$ -npe protecting group. Precipitation of  $O^6$ -[2-(4-nitrophenyl)ethyl]-guanosine **83** from methanol enabled efficient removal of the triphenylphosphine oxide contaminant with 76% yield over two steps. Thus, synthesis of **83** with 70% yield over three steps was much improved over the previous methods.

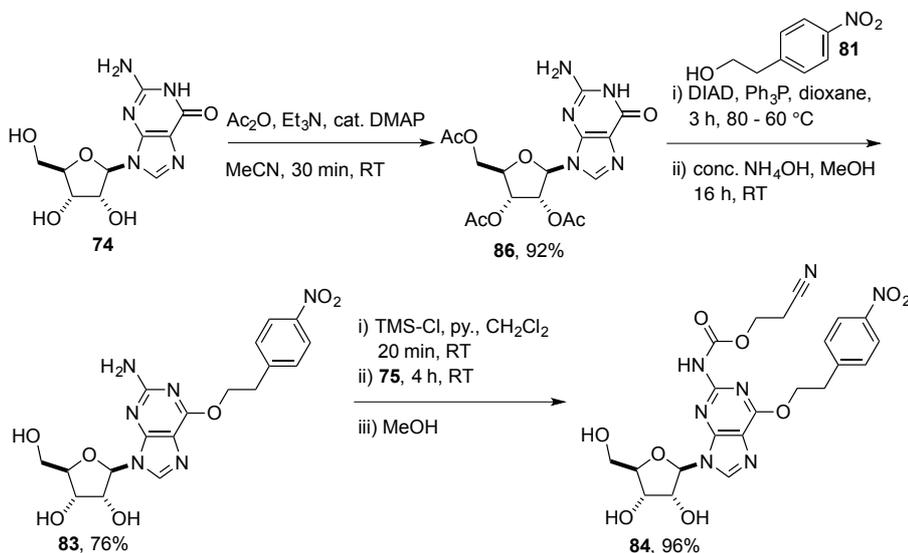


Figure 59. Optimised synthetic route to the nucleobase protection of guanosine **74** to give  $N^2$ -[(2-cyanoethoxy)carbonyl]- $O^6$ -[2-(4-nitrophenyl)ethyl]-guanosine **84**.

The final step to install the  $N^2$ ceoc group was conducted by transient protection of the hydroxyl groups with TMS-Cl in a mixture of  $\text{CH}_2\text{Cl}_2$  and pyridine, followed by addition of the 2-cyanoethyl carbonochloridate **75**. After stirring for 4 hours the reaction was quenched with methanol, which also hydrolysed the TMS groups. Literature described workup<sup>[201]</sup> involved evaporating the reaction mixture and treating the residue with water and ethyl acetate. The aqueous layer was extracted with further volumes of ethyl acetate and the combined organic layers were evaporated to dryness. The residue was resuspended in  $\text{CH}_2\text{Cl}_2$  and the insoluble material filtered, washed and dried to give **84**. On a small scale <200 mg this method was satisfactory in removing pyridinium.HCl and other contaminants as large volumes of ethyl acetate could be used to extract the product and resulted in an isolated yield of 63%. However, when used in larger scale reactions of >1 g the poor solubility of the product in either water or ethyl acetate caused the organic phase extraction to be very inefficient. In addition, the large amounts of the precipitated product **84** made extraction and physical manipulation difficult. The yield of the product extracted, although clean, was low at 33%.

The method to isolate **78** and **79** post nucleobase protection both involved precipitation from methanol and it was decided that this could be applied to the isolation of **84**, thus avoiding the problematic organic extraction. After co-evaporation of the reaction mixture with toluene and methanol to remove excess pyridine the residue was dissolved in a minimum of methanol. Addition of small amounts of water was found to aid the precipitation of the product and it was added until a precipitate began to form. The resultant mixture was placed in a fridge overnight and the precipitate collected and washed with cold methanol. This gave the product **84** but it contained pyridinium.HCl salt, which was removed by boiling a suspension of **84** in water for 15 min. This efficiently removed the salt and furnished the product **84** cleanly in an improved yield of 96% (Figure 59).

### 2.2.3. Synthesis of the 2'/3'-*O*-acetyl RNA phosphoramidites

With the nucleobase-protected nucleosides of **78**, **79** and **84** to hand, synthesis towards the 2'/3'-acetylated phosphoramidites could begin. The 5'-hydroxyls of the nucleosides **78**, **79** and **84** were first tritylated using 4,4'-dimethoxytrityl chloride (DMTr-Cl) in pyridine as base and solvent. Purification by flash column chromatography afforded the 5'-protected nucleosides **88**, **89** and **90** in high yield, with the final nucleoside 5'-*O*-(4,4'-dimethoxytrityl)-uridine **91** purchased from *ChemGenes*.

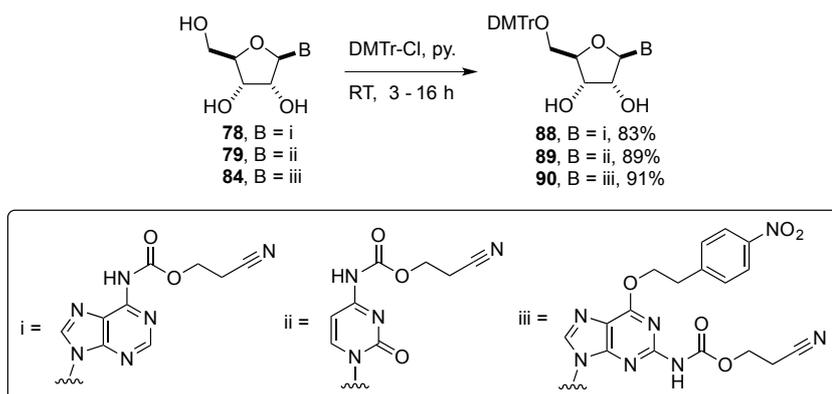


Figure 60. Dimethoxytritylation reaction of the nucleobase protected nucleosides.

Attention turned to the 2'/3'-*O*-monoacetylation of **88**, **89**, **90** and **91**. The acetylation reaction was first explored using **91** as it was commercially available. Acetylation was brought about using 1 mole equivalent of acetyl chloride to ensure no over acetylation

to the fully protected 2',3'-bis-acetylated nucleoside. The reaction was carried out in THF with 1 mole equivalent of pyridine to neutralise *in situ* generated HCl. This encouragingly furnished the inseparable regioisomeric mixture of 2'/3'-*O*-(acetyl)-5'-*O*-(4,4'-dimethoxytrityl)-uridine **92a+92b** with a ratio *ca.* 2.4:1 with respect to **92b:92a**. Acetyl groups are known to migrate between the 2'-OH and 3'-OH such that on subsequent repeat acetylation reactions some variation of the ratio was observed. Although not a significant change in the ratio it was noted that the migration tended to favour isomerisation to the 3'-*O*-acetyl isomer **92b**. Migration is known to be catalysed by base and 1-2% triethylamine is usually added to the mobile of DMTr protected compounds to neutralise the slightly acidic silica.<sup>[214]</sup> Omission of triethylamine during silica gel chromatography also results in migration indicating acetyl migration is either acid or base catalysed thus as a precaution to DMTr removal triethylamine use was continued.

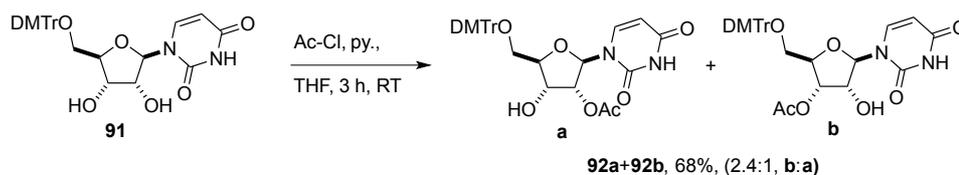


Figure 61. Reaction to form the regioisomeric mix of 2'/3'-*O*-(acetyl)-5'-*O*-(4,4'-dimethoxytrityl)-uridine **92a+92b**. Reaction carried out by Dr Colm D. Duffy.

Acetylation of the 5'-*O*-(4,4'-dimethoxytrityl)-purine nucleosides **88** and **90** was explored with attention concentrating first on **88**. Acetylation was conducted by dissolving **88** in anhydrous THF and adding 1 equivalent each of pyridine and AcCl and allowing the reaction to stir for 30 minutes. The TLC analysis of the reaction mixture indicated an incomplete reaction with a single faster running spot that was assumed to be the acetylated products. A crude NMR confirmed that the acetylation reaction was not complete and characteristic downfield shifted peaks of H-C(2') (0.99 ppm) and H-C(3') (1.08 ppm) corresponding to the 2'-*O*-acetylated **93a** and 3'-*O*-acetylated **93b** nucleosides respectively were observed. However, also observed were additional downfield shifted peaks of H-C(1'), H-C(2') and H-C(3') that corresponded to the *N*<sup>6</sup>-[(2-cyanoethoxy)carbonyl]-2',3'-*O*-(bis-acetyl)-5'-*O*-(4,4'-dimethoxytrityl)-adenosine **94** as supported by H<sup>1</sup>-H<sup>1</sup> COSY analysis. The ease of migration of the acetyl group was observed, where the ratio prior to purification was 1.8:1 with respect to **93b:93a**.

After purification isomerism is seen to favour 3'-acetylation to give a value of 2.5:1, **93b:93a** (Figure 62).

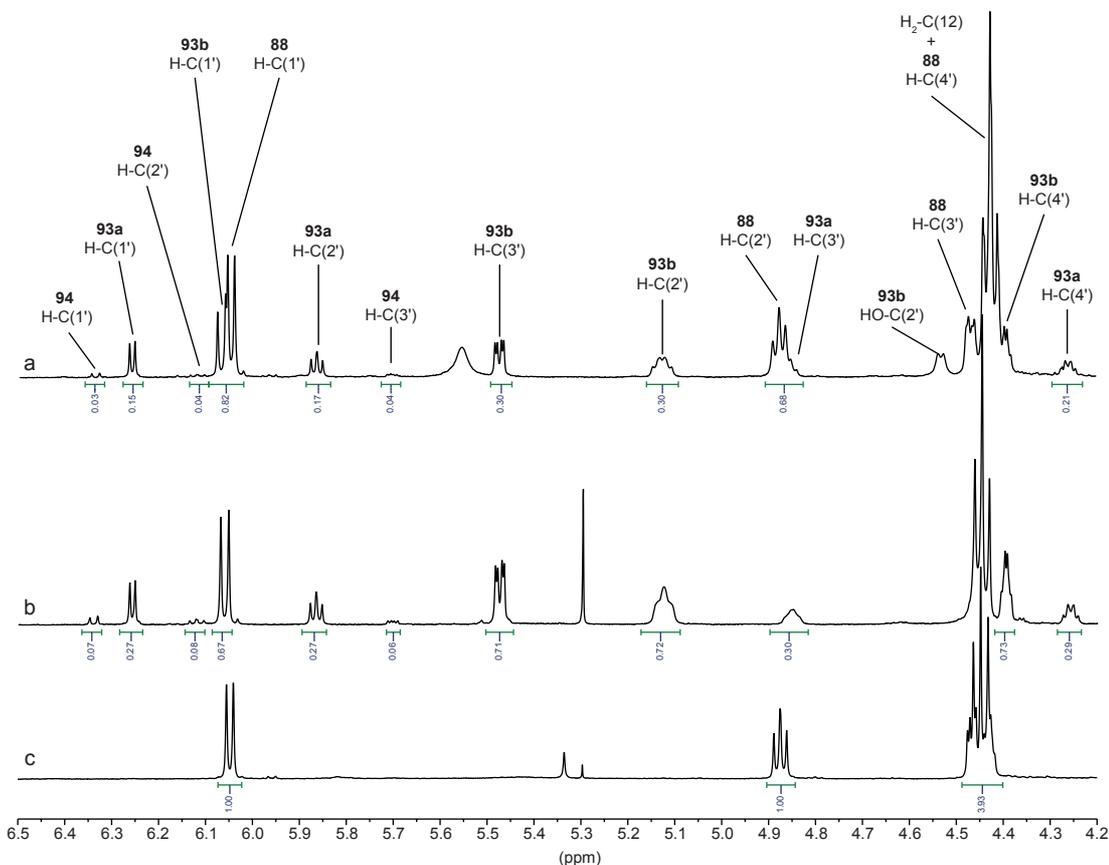


Figure 62. a) Crude  $^1\text{H-NMR}$  spectrum of the acetylation of **88**. b) NMR spectrum of the separated acetylated products showing the mixture of mono- and bis-acetylated compounds, **93a+93b** and **94**. c)  $^1\text{H-NMR}$  spectrum of the separated starting material **88**.

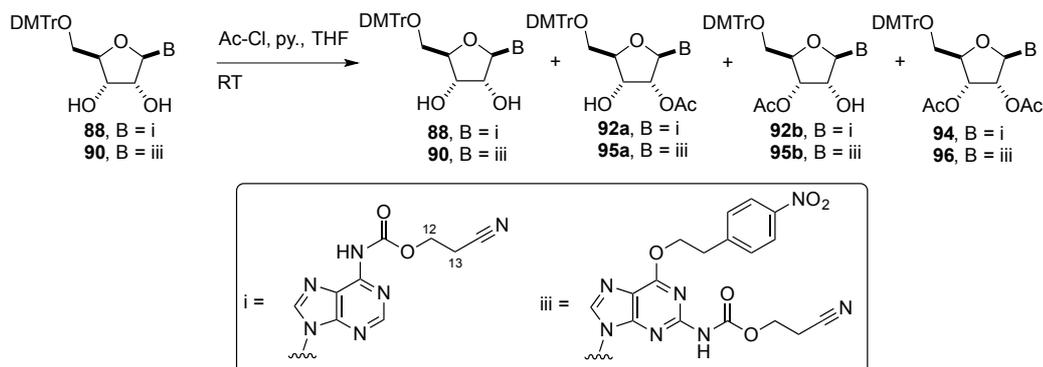


Figure 63. Acetylation of 5'-O-(4,4'-dimethoxytrityl)-purine ribonucleotides **88** and **90** with the formation of the various acetylation products.

The yields of this first reaction were quite poor with the total amount of acetylated material recovered at 19% (Table 2, entry 1). The remaining starting material was isolated in 61% yield. From these results it was thought that the reaction time was too short but also that the acetylation was slower than acetylation of **91** due to steric hindrance from the nucleobase. The reaction was repeated using 1.2 equivalents each of pyridine and AcCl with reaction carried out in THF with stirring at RT for 1 hour. The products of the reaction were then purified and separated by preparative TLC. Although yield of the mono-acetylated regioisomeric mixture was increased to 68%, both bis-acetylation and starting material were still recovered (Table 2, entry 2). The mono-acetylation of **90** was also explored but with alternative conditions (Figure 63), **90** was dissolved in anhydrous acetonitrile and treated with 1 equivalent of Ac<sub>2</sub>O, 1.1 mole equivalent of triethylamine and 0.1 mole equivalent of DMAP. After stirring for 2 hours the reaction did not proceed to completion and so the reaction mixture was subjected to purification by flash column chromatography. In this case, the starting material and mono-acetylated material were not separable and isolated as a mixture in a yield of 47%. However, it was found that a much higher amount of the bis-acetylated product was formed (Table 2, entry 3). Also of note is the ratio of the **95b:95a** within the regioisomeric mixture favoured the 3'-acetylated isomer in a ratio of 5:1 respectively.

Entry	SM	Reaction conditions (mole eq. to SM)	Time (h)	Products (% yield)			Ratio of 3'-OAc: 2'-OAc
				SM	2'/3'-OAc	2',3'-bis- OAc	
1	<b>88</b>	Ac-Cl (1), py. (1), THF	0.5	<b>88</b> (61)	<b>93a+93b+94</b> (19)		2:1
2	<b>88</b>	Ac-Cl (1.2), py.(1.2), THF	1	<b>88</b> (22)	<b>93a+93b</b> (68)	<b>94</b> (7)	3:1
3	<b>90</b>	Ac <sub>2</sub> O (1), Et <sub>3</sub> N (1.1), DMAP (0.1), MeCN	2	<b>90+95a+95b</b> (47)		<b>96</b> (27)	5:1

*Table 2. Reaction conditions and product yields from acetylation of the 5'-O-(4,4'-dimethoxytrityl)-purine ribonucleotides **88** and **90**. Reactions were carried out in anhydrous conditions, at RT and 0.1 M concentration of starting material.*

The attempted methods to mono-acetylate the 2'/3'-diol of 5'-O-(4,4'-dimethoxytrityl)-purine ribonucleotides **88** and **90** were not optimal as full reaction of the starting materials with the acetylating reagents was not observed. Despite only using one mole equivalent of acetylating reagent formation of the fully protected bis-acetylated products was also observed. It was concluded that bis-acetylation could not be controlled but also that full reaction of the starting material to produce only mono-acetylated products would not be possible.

On evaluation of the literature it was found that Fromageot and co-workers have utilised orthoesters in the 2'/3'-monoacylation of ribonucleotides and monoacylation of 1,2-*cis*-diols.<sup>[222-225]</sup> The use of orthoesters has the advantage that 2',3'-hydroxyl selectivity can be achieved by forming a cyclic 5-membered orthoester, such as 2',3'-*O*-methoxyethylidene if trimethyl orthoacetate is used. This intermediate can then be quantitatively hydrolysed to a regioisomeric mixture of 2'/3'-monoacetylated ribonucleosides. Thus, the monoacylation can be driven to completion and bisacetylation of the 2'/3'-diol can be avoided. Additionally, this method requires acidic catalysis so installation of the acid-labile 5'-DMTr group was carried out after acetylation. With a free primary hydroxyl group a small amount of 5'-orthoesterification leading to 5'-acetylation was expected, but experimentally in most cases was not observed by TLC. Where traces were observed, these 5',3'- or 5',2'-*O*-bisacetylated nucleosides were easily separated by flash column chromatography.<sup>[225]</sup>

Thus, monoacetylation of **78**, **79** and **84** was furnished by reaction with excess trimethyl orthoacetate and catalytic trifluoroacetic acid to ensure complete formation of the 2',3'-*O*-methoxyethylidene in a suitable anhydrous solvent such as dioxane. Addition of water to the reaction led to the hydrolysis of the cyclic orthoester to yield the regioisomeric mixture of 2'/3'-acetylated nucleosides in high yield after purification by flash column chromatography (Figure 64).<sup>[222]</sup> Ratio of 3'-OAc:2'-OAc regioisomers were again found to favour the 3'-OAc and were in general *ca.* 2:1 respectfully.

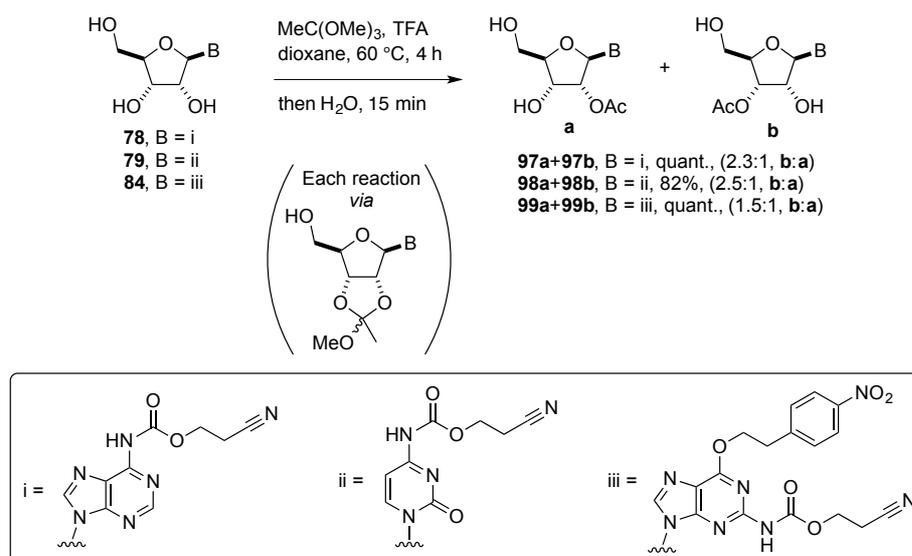


Figure 64. Monoacetylation reaction of the base-protected nucleosides **78**, **79** and **84**.

With a reliable and high yielding method for the 2'/3'-monoacetylation now established the regioisomeric mixtures **97a+97b**, **98a+98b** and **99a+99b** were tritylated under standard conditions.<sup>[201]</sup> The monoacetylated nucleosides were 5'-hydroxyl protected with DMTr-Cl in anhydrous pyridine to give the 5'-(4,4'-dimethoxytrityl) 2'/3'-monoacetylated ribonucleosides **93a+93b**, **100a+100b** and **95a+95b** (Figure 65).

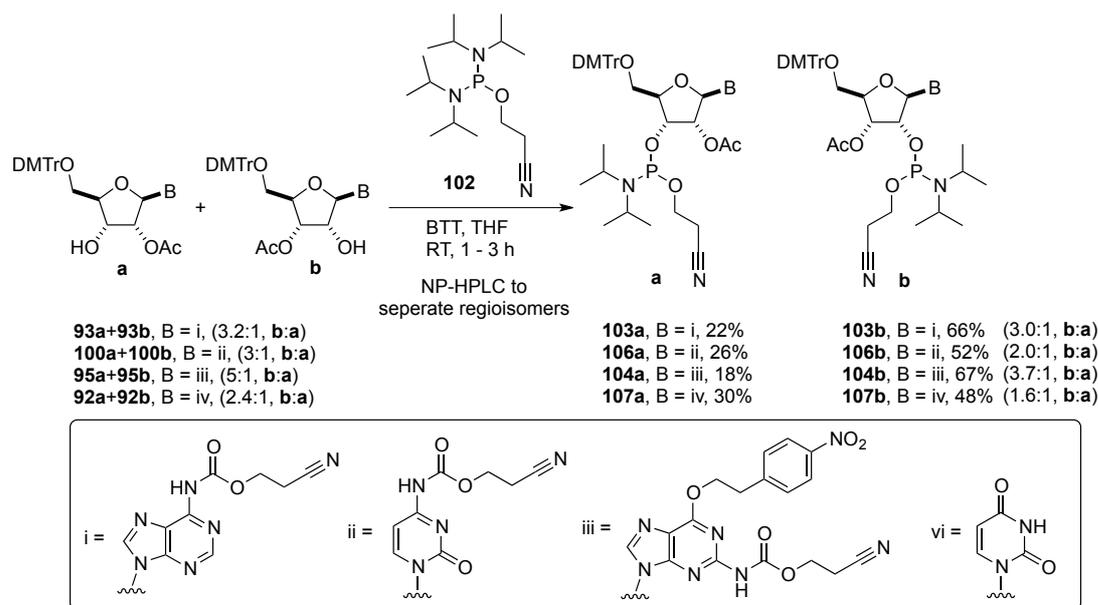


Figure 65. Tritylation reactions of the regioisomeric mixtures of **97a+97b**, **98a+98b** and **99a+99b**.

As the tritylation reaction and purification of the products required basic conditions 2'/3'-migration of the acetyl groups was again observed. With regards to the ribonucleoside derivatives **97a+97b** and **98a+98b**, the extent of the regioisomeric migration was not large. However, the derivative **99a+99b** showed a greater ratio change of 330%. It was clear the 2'/3'-OAc equilibrium favoured the 3'-OAc regioisomer, but it was envisioned that synthesis of acetylated oligonucleotides would require greater amounts of the 2'-OAc regioisomer to give the natural 3',5'-linkage.<sup>[214]</sup> Hence these high 3'-OAc:2'-OAc ratios would limit the synthesis of the 2'-O-acetyl-3'-phosphoramidites and so an evaluation of various phosphitylation reactions was conducted with the purine monoacetylated ribonucleoside derivatives **93a+93b** and **95a+95b**. It was hoped that further migration of the acetyl group the 3'-hydroxyl could be reduced or the regioisomeric ratio could be improved.

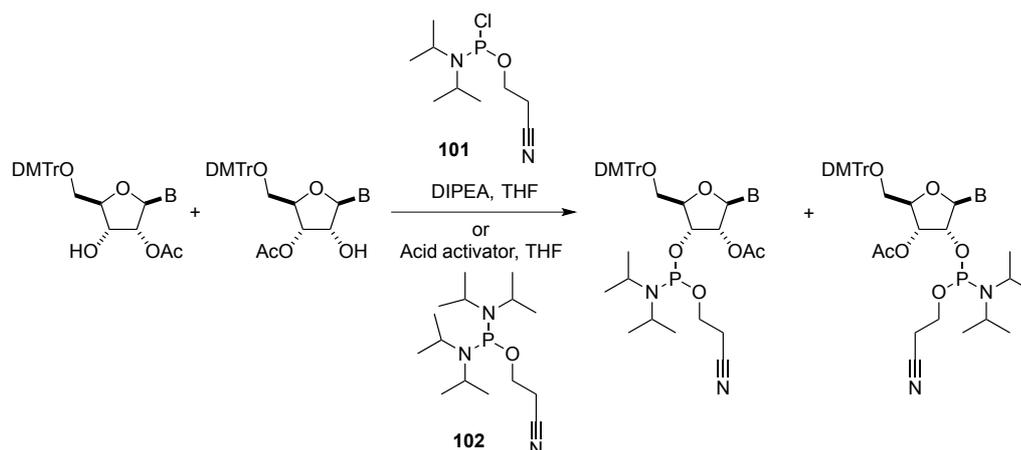


Figure 66. Generic phosphitylation reaction scheme.

The phosphitylating agents studied were 2-cyanoethyl *N,N*-diisopropyl phosphoramidochloridite **101** and 2-cyanoethyl *N,N,N',N'*-tetraisopropylphosphoramidite **102**, the former requires basic conditions in order to neutralise the in-situ generated by-product HCl whilst the later phosphitylating reagent requires acid catalysis or ‘activator’ and it was here that the most scope for improvement was foreseen as various activators of varying acidity could be used. The various reaction conditions, reagents used, ratios of the 2'/3'-OAc ribonucleoside alcohol mixtures and ratios of the isolated and separated 2'- or 3'-OAc phosphoramidites are given in Table 3.

Entry	Acetylated mixture (1 mole eq.)	Mixture ratio (3'-OAc:2'-OAc)	Reaction conditions (mole eq.)		Isolated phosphoramidite ratio	
					3'-OAc	2'-OAc
					<b>103b</b>	<b>103a</b>
1	<b>93a+93b</b>	3:1	<b>101</b> (1.3)	DIPEA (4)	5	1
2	<b>93a+93b</b>	3:1	<b>102</b> (2)	DCI (2)	4	1
3	<b>93a+93b</b>	3:1	<b>102</b> (1.2)	1 <i>H</i> -Tetrazole (1)	4	1
4	<b>93a+93b</b>	3:1	<b>102</b> (1.2)	BTT (1)	3	1
					<b>104b</b>	<b>104a</b>
5	<b>95a+95b</b>	5:1	<b>101</b> (1.3)	DIPEA (4)	7	1
6	<b>95a+95b</b>	5:1	<b>102</b> (1.2)	BTT (1)	3.7	1

Table 3. Reaction conditions used to scope the phosphitylation reaction. Reactions were conducted in anhydrous THF (starting material concentration, 0.1 M), at room temperature and reactions times within a range of 1-3 hours. 1*H*-Tetrazole (0.45 M) and 5-benzylthio-1*H*-tetrazole (BTT, 0.3 M) were added to the reaction as a solution in anhydrous MeCN. DIPEA = *N,N*-diisopropylethylamine, DCI = 4,5-dicyanoimidazole. The author would like to thank Dr Jianfeng Xu for the initial suggestion and conducting the entries 2-4.

As expected using **101** and *N,N*-diisopropylethylamine lead to undesired migration of the acetyl group and an increased formation of the 3'-OAc-2'-phosphoramidite. Where the phosphitylating agent **102** was employed, three commercially used acidic activators of varying pK<sub>a</sub> values were utilised as follows, 4,5-dicyanoimidazole, (DCI, pK<sub>a</sub> = 5.2), 1*H*-tetrazole (pK<sub>a</sub> = 4.9) and 5-benzylthio-1*H*-tetrazole (BTT, pK<sub>a</sub> = 4.1). The results show that acidic activators enabled phosphitylation to occur with reduced migration of the 2'/3'-acetyl group when compared to phosphitylation with **101** and *N,N*-diisopropylethylamine. The most effective of the acid activators was BTT and enabled phosphitylation with either no ratio change (Table 3, entry 4) or a ratio improvement towards the 2'-acetyl-3'-phosphoramidite (Table 1, entry 6).

The mechanism of activation of **102** involves protonation by BTT proceeded by nucleophilic attack at phosphite by the conjugate base of BTT and subsequent displacement of *N,N*-diisopropylamine to form a tetrazolide **105**. This tetrazolide **105** is the active phosphitylating agent and attack by a hydroxyl group at the phosphite centre with displacement of conjugated base of BTT results in phosphitylation (Figure 67).<sup>[226]</sup> Presumably, the increased acidity of BTT increased the effective concentration of the protonated phosphitylating agent and in turn increased the concentration of the tetrazolide **105**, formation of which is the slow step.<sup>[227]</sup> This increased the rate of phosphitylation reduced the reaction time and this is thought to have limited acetyl migration to the 3'-hydroxyl. Additionally, the use of acidic activators may also minimise the 2'/3'-acetyl migration by decreasing the pH and reducing the formation of a 2'- or 3'-alkoxide.<sup>[228]</sup> The improved isolated ratio of the guanosine derivative is less understood and may be due to steric demand from the protected nucleobase. This could have reduced the rate of phosphitylation of the relatively hindered 2'-hydroxyl, allowing acetyl migration to occur and leading to preferential phosphitylation of the 3'-hydroxyl.

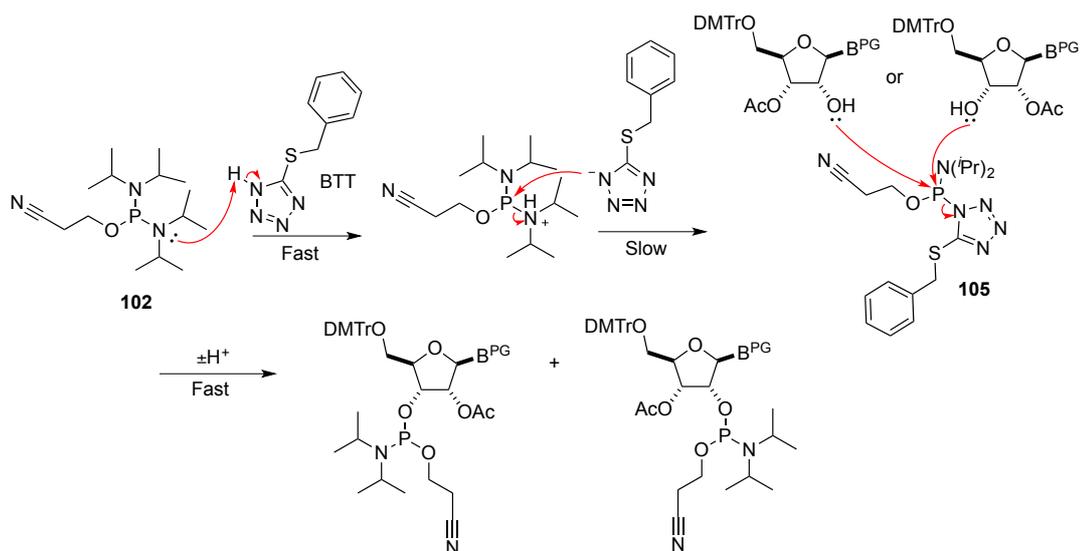


Figure 67. Mechanism of the phosphitylation reaction with **102** and the activator BTT.

It was decided that BTT was the activator of choice to improve the yields of the 2'-acetyl-3'-phosphoramidites. The four monoacetylated ribonucleoside derivatives **93a+93b**, **100a+100b**, **95a+95b** and **92a+92b** were phosphitylated as regioisomeric mixtures by treating them with an excess of **102** (1.2-2 mole equivalents) in THF and 1 mole equivalent of BTT added as a 0.35 M solution in acetonitrile. The resultant 2'/3'-regioisomeric mixture of phosphoramidites was briefly purified by flash column chromatography to remove the activator BTT. The regioisomers were cleanly separated by normal-phase HPLC to give the regioisomers in high yield (Figure 68). In each case it was found that the 3'-acetyl-2'-phosphoramidite regioisomer eluted first with typically 3-5 min separation between regioisomers. The separated regioisomers also exist as diastereoisomers and each was observed during separation but were not isolated separately (Figure 69).

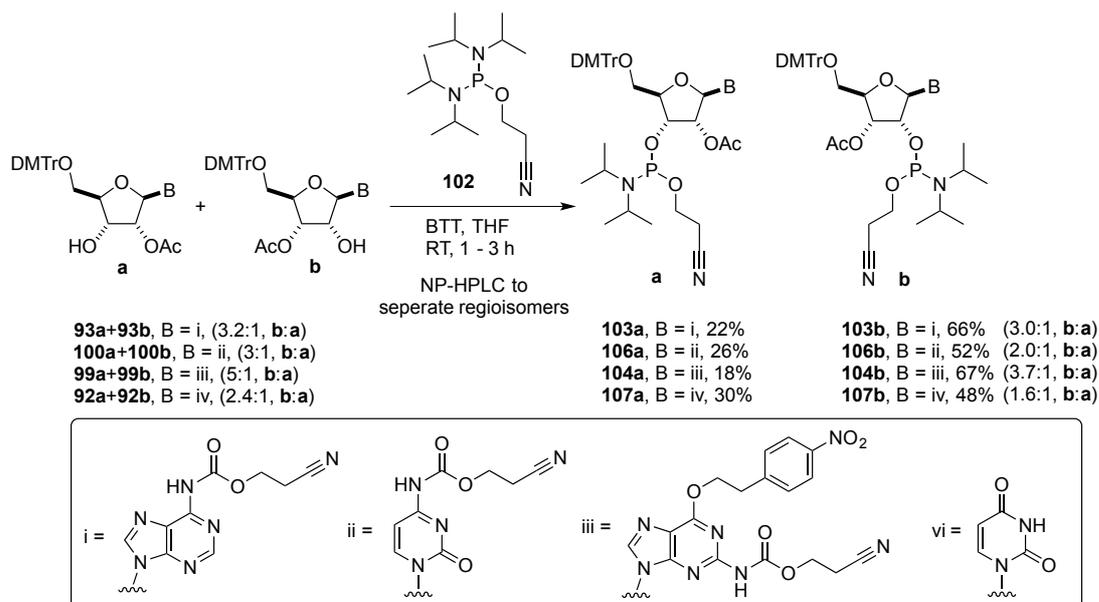


Figure 68. Phosphitylation reaction of the monoacetylated ribonucleoside derivatives. Phosphitylation reactions of **100** and **92** were carried out by either Dr Colm D. Duffy or Dr Jianfeng Xu.

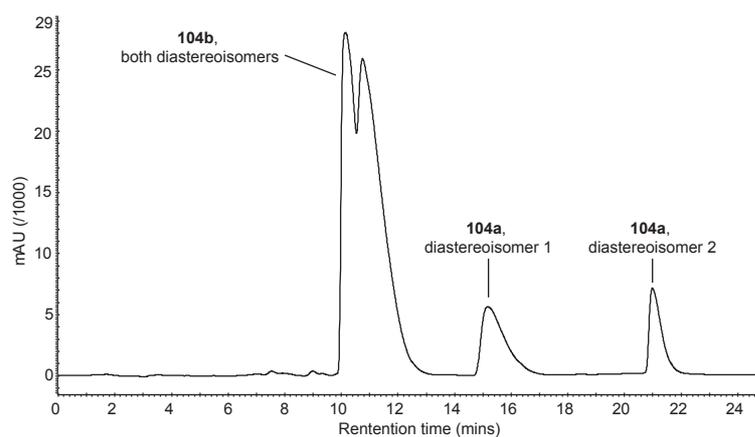


Figure 69. Preparative HPLC trace of the 2'/3'-acetylated guanosine phosphoramidites **104a** and **104b**.

To summarise, an efficient and high yielding synthetic strategy has been developed and optimised to allow the synthesis 2'/3'-monoacetylated phosphoramidites. The key step was monoacetylation *via* 2',3'-cyclic orthoester to bring about the complete monoacetylation at the 2'- or 3'-hydroxyl with no bisacetylation. This required an inversion of the initially planned steps, which meant monoacetylation was followed by the tritylation reaction. However, this led to the unavoidable and undesirable equilibration of the acetyl group to favour the 3'-OAc. This equilibrium was to some

extent remedied by utilising a more acidic activator in the final phosphitylating step, which resulted in improved 2'/3'-OAc ratio from the starting mixture. In context of the crystallisation of mixtures of 2'/3'-O-acetylated uridine derivatives, it has been suggested that the 3'-O-acetylated isomer is the more thermodynamically stable, as pure crystals of the 3'-O-acetyl isomer can be obtained in high yield.<sup>[223]</sup> The reasoning remains unclear, however this suggested that the acid catalysed phosphitylation is under greater kinetic control and hence the improved 3'-OAc:2'-OAc ratio upon use of increasingly acidic activating agents. In other words, the 3'-hydroxyl of the 2'-O-acetyl isomer is more nucleophilic than the corresponding 2'-hydroxyl of the 3'-O-acetyl isomer. Attention now turned to the synthesis of the 2'/3'-silylated phosphoramidites.

#### 2.2.4. Synthesis of the 2'/3'-O-TBDMS phosphoramidites

The synthesis of the silylated phosphoramidites follows well-developed procedures and first requires dimethoxytritylation of the base-protected ribonucleosides, which is previously described in section 2.2.3 (Figure 60). As it was felt that 3'-phosphoramidites would be required in higher quantities, to produce the natural 3',5'-linked oligonucleotides, selective 2'-silylations were considered for the next step. Ogilvie *et al.* have extensively evaluated the *tert*-butyldimethylsilyl group as a useful 2'/3'-hydroxyl protecting group and have found that the addition of silver nitrate significantly improves the selectivity of 2'-hydroxyl silylation.<sup>[212, 229-231]</sup> This method was deemed ideal as it would improve the yields of the 2'-TBDMS regioisomer yet allow synthesis of both, so allowing maximum flexibility during oligonucleotide synthesis.

The dimethoxytritylated ribonucleoside derivatives **88**, **89** and **90** were treated with TBDMS-Cl, AgNO<sub>3</sub> and pyridine in anhydrous THF for a minimum of 5 hours (Figure 70). The reaction resulted in a regioisomeric mixture of 2'/3'-O-TBDMS ribonucleosides and with the adenosine **108a** and **108b** and cytidine **109a** and **109b** derivatives these were easily separated by silica gel flash column chromatography. With silica gel chromatography of the compounds that contained the acid-labile DMTr group, 1-2% triethylamine was usually added to the mobile phase to neutralise the slightly acidic silica gel. However, it was crucial triethylamine was omitted during the



commercially available but the 3'-TBDMS uridine phosphoramidite is only available as the non-phosphitylated precursor and so this was also phosphitylated using the same conditions. It is useful to note that alternative phosphitylation reaction conditions, which utilise 2,4,6-collidine as base and *N*-methylimidazole as catalyst, are generally used due to concerns for TBDMS group migration.<sup>[232, 233]</sup> However, in this work migration of the 2'/3'-TBDMS groups were not observed within the detection limits of NMR spectroscopy (< 1%).

To summarise optimised synthetic routes have been developed to produce the required silylated phosphoramidites and the synthetic routes were applicable to multi-gram scale synthesis. This completed the synthetic work to produce the phosphoramidites for the solid-phase synthesis of partially acetylated RNA oligonucleotides. Attention now turned to the synthesis of the photolabile linker.

### **2.3. Synthesis of the photolabile linker and preparation of the solid-support**

Controlled-pore glass (CPG) solid-supports are currently the most widely used supports for automated oligonucleotide synthesis for several practical reasons.<sup>[183]</sup> Firstly, the porous surfaces of the beads are etched with either acid or base to give pores of the required mean pore size. They can also be manufactured to a uniform spherical size thus allowing a good flow through of solvent when packed into synthesis columns, without excessive back-pressure. More crucially during the synthesis of the oligonucleotides, CPG does not allow solvent swelling as opposed to other commonly used gel-type solid-supports such as cross-linked polystyrene (PS).<sup>[205]</sup> This has the advantage of enabling the by-products and excess reagents to be efficiently washed from the beads. Also the choice of solvents for each synthetic step is less crucial as swelling properties and diffusion of solvents/reagents into a solid-support does not need to be considered. This was ideal as available automated RNA oligonucleotide synthesis machines have been optimised with the consideration of utilising both anhydrous and aqueous conditions within one synthesis cycle. The surfaces of the CPG beads are derivatised by a 'spacer' that is in most cases a long-chain alkylamines (LCAA, Figure 71). These act as spacers to distance the functional end away from the surface of the solid-support to

maximise the diffusion of reagents.<sup>[205]</sup> It is to this terminal amine that the chosen photolabile linker will be attached.

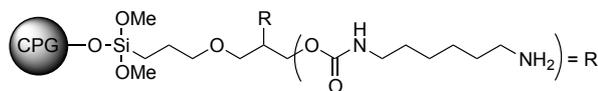


Figure 71. Structure of a commonly used long-chain alkylamine (LCAA) controlled pore glass solid support.

The chosen photolabile linker **123** is of the *ortho*-nitrobenzyl type of protecting groups and linkers that are most commonly released by UV irradiation at  $\lambda = 320\text{-}400$  nm. Photo-irradiation leads to electronic excitation of the nitro group and formation of a diradical **116**. This diradical **116** abstracts a proton from the benzylic position, which undergoes rearrangement and cyclisation to **117**. Release of the product is accompanied by the formation of the *ortho*-nitrosobenzaldehyde photoproduct **118** which can undergo further cross-linking to give azobenzene-2,2'-dicarboxylic acid **119**. This by-product, which has a deep red colour, is known to reduce cleavage yields by acting as a strong light filter.<sup>[234-237]</sup> Using  $\alpha$ -substituted *ortho*-nitrobenzyl groups such as the linker **123** chosen in this work, reduces cross-linking, so photolysis can be maximised. The linker **123** has been predominantly used in peptide chemistry and has shown overall yields (*i.e.* yields of product after synthesis and cleavage) of 40-50% indicating an efficient photocleavage and so was considered ideal for synthesis of RNA oligonucleotides on  $\mu\text{mol}$  scales.<sup>[206]</sup>

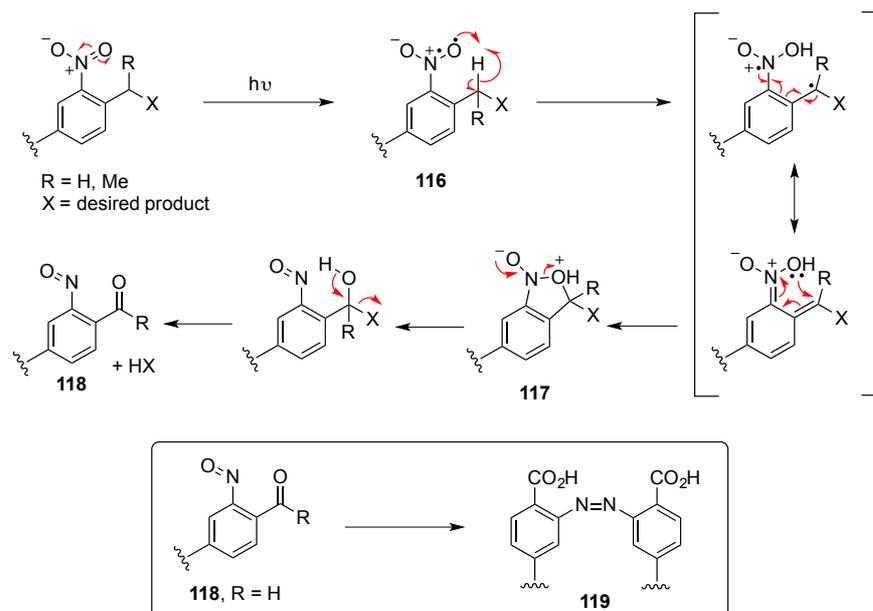


Figure 72. Cleavage mechanism of the photolabile ortho-nitrobenzyl linkers. Boxed is the dimerisation of the ortho-nitrosobenzylaldehyde photoproduct **118** to the dimer **119**.

Synthesis of the photolabile linker began with the Grignard reaction on the commercially available 3-formyl-4-nitrobenzoate **120** with methyl magnesium bromide at room temperature. This gave the  $\alpha$ -methyl alcohol **121** but the isolated product also contained the benzyl product **122**, which was present at 15% of the mixture as calculated by integration of NMR peaks (Figure 73a). It is known that Grignard reagents with  $\beta$ -hydrogens exhibit a side-reaction that involves reduction of the aldehyde by hydride transfer via a cyclic six-membered transition state (Figure 73b).<sup>[238, 239]</sup> However, since there were no  $\beta$ -hydrogens available from methyl magnesium bromide this mechanism was excluded. It was not clear by what mechanism the reduction was taking place but reduction *via* a radical pathway was possible as single electron transfer pathways for addition of Grignard reagents to aromatic aldehydes are known.<sup>[239, 240]</sup> The mixture **121+122** was taken on to the next step where it was hoped the side-product could be separated by NP-HPLC. Thus, the mixture of **121+122** was tritylated with DMTr-Cl in pyridine/ $\text{CH}_2\text{Cl}_2$  to give **123** containing **124**. The benzyl side-product **124** was cleanly separated from the  $\alpha$ -methyl dimethoxytrityl product **123** by preparative NP-HPLC to complete the synthesis of the photolabile-linker to give a yield of 9% over two steps.

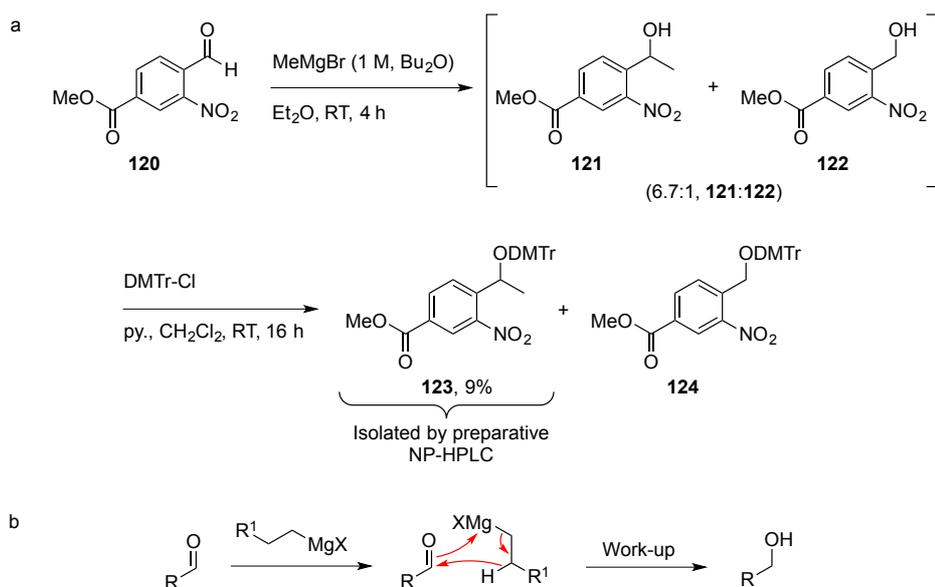


Figure 73. a) Synthesis of the photolabile linker **123**. The author thanks CDC for the development and optimisation of this synthetic route. b) Reduction of an aldehyde by ethyl magnesium bromide via  $\beta$ -hydride transfer.

The solid-support was prepared by first taking the photolabile-linker **123** and treating it with a LiOH in a mixture of THF and water to hydrolyse the methyl ester. Without further purification the lithium benzoate salt was used in the next step after fastidious removal of water by co-evaporation with anhydrous pyridine. The lithium benzoate salt was treated with isobutylchloroformate in anhydrous pyridine to afford the mixed anhydride **125**. Under anhydrous conditions the LCAA-CPG was treated with the mixed anhydride **125** (200-1000  $\mu\text{mol}$  per gram of CPG). Any unreacted amines were ‘capped’ with pivaloyl chloride (initially with  $\text{Ac}_2\text{O}$ ) and the ‘loading’ or in other words the number of linkage sites possible for oligonucleotide synthesis was determined by the trityl analysis to give a loading of 33.3-56.2  $\mu\text{molg}^{-1}$  (*vide infra*) (Figure 74). This was deemed a suitable range when compared to commercial solid-supports<sup>[241]</sup> that have loadings of 15-40  $\mu\text{molg}^{-1}$  and Johnsson *et al.* whom produce their CPG with a loading of 35-60  $\mu\text{molg}^{-1}$ .<sup>[204]</sup>

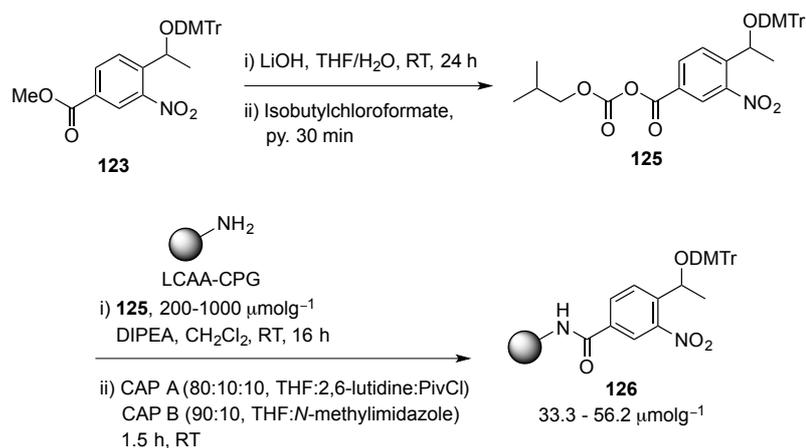


Figure 74. Preparation of the solid-phase support **126** (photolabile-CPG). The author thanks Dr Colm D. Duffy or Dr Jianfeng Xu for assistance in development of this stage of the solid-phase preparation, author contributions - optimisation of step 2.

Analytical techniques usually used in organic chemistry are usually unsuitable for the characterisation of functionalised solid-supports. For oligonucleotide synthesis the standard method is the trityl assay which takes advantage of the DMTr cation that is released on deprotection of hydroxyl group.<sup>[183]</sup> The cation is a strong chromophore and has an absorption maxima at  $\lambda$  503 nm ( $\epsilon = 76 \text{ mL cm}^{-1} \mu\text{mol}^{-1}$ ). The release of the DMTr cation during each cycle of oligonucleotide synthesis also serves to assess the yield/coupling efficiency of each phosphoramidite to the nascent oligonucleotide albeit only as an approximate calculation (Figure 75). Theoretically, the trityl assay is accurate but the assay conducted by the automated synthesis machine is only used to give an approximate yield.

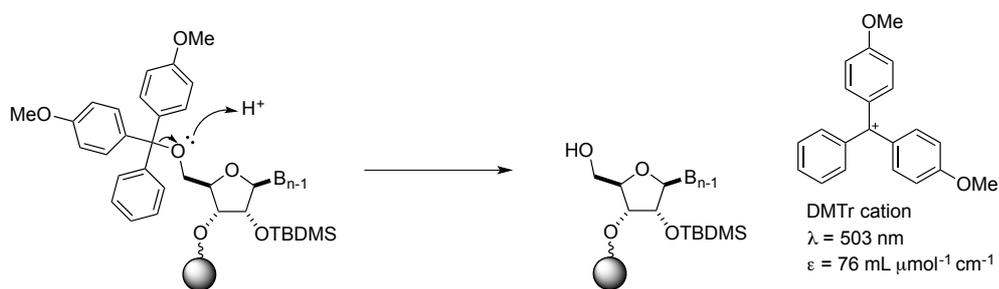


Figure 75. Acid deprotection of the DMTr group to form the DMTr cation.

The loading of solid-supports can be determined by accurately weighing an amount of solid-support (3-4 mg) directly into a 10 mL volumetric flask. The flask is filled with 3% TCA in  $\text{CH}_2\text{Cl}_2$  to afford complete release of the DMTr cation. The absorption at

$\lambda$  503 nm is measured in a UV machine and the loading determined using the equation below:

$$\text{Loading } (\mu\text{molg}^{-1}) = (A_{503} \times \frac{\text{vol}}{76}) \times (\frac{1000}{\text{wt}})$$

*Equation 1. Equation to calculate the loading.  $A_{503}$  is the absorption at  $\lambda$  503 nm, vol is the volume of the solution in mL,  $76 \mu\text{mol}^{-1} \text{mL cm}^{-1}$  is the extinction coefficient at  $\lambda$  503 nm and wt is the amount of support in mg. Path length assumed to be 1 cm.*

With the photolabile-CPG **126** prepared it was decided to test the efficiency of the photocleavage of oligonucleotides as this type of linker has only previously been used in the synthesis of peptides. An RNA oligonucleotide of sequence 5'-GCCCCCCC-3'P was synthesised for the purpose of testing the photocleavage. The synthesis used commercially available standard RNA amidites, reagents and standard synthesis programs (1  $\mu\text{mol}$  scale, see Chapter 6.2.3 for a detailed description). Trityl monitoring during the synthesis indicated that the yield of each coupling step was >95%. On completion of the trityl-on synthesis the CPG was dried under vacuum and no deprotections were carried out.

UV irradiation of the CPG was conducted in a 24 well cell culture tray using an LED light that emitted at  $\lambda$  365 nm placed at a distance of 2 cm from the bottom of the well containing the CPG. The polystyrene cell culture trays do not absorb  $\lambda$  365 nm and were ideal as they were sterile and free from RNase contamination.<sup>[242]</sup> It is useful to note that UV damage to RNA nucleobases should not be observed at these energies and it is at higher energies of 256 nm where damage such as photodimers are observed. The CPG was layered with 1 mL of acetonitrile and distributed evenly on the bottom of the vessel and irradiated for 1:20 h during which the CPG was regularly agitated and 10  $\mu\text{L}$  aliquots of solvent removed at the described time points to allow quantification of cleaved oligonucleotide by UV absorption. The results show maximum absorption is attained at 60 minutes indicating the photocleavage is effectively complete. The maximum reached O.D. = 42.3 and corresponds to 683 nmol of oligonucleotide and yield of 68%, suggesting that photocleavage was efficient. Alternatively, the UV induced cleavage of the DMTr-O<sup>-</sup> from the photolabile-CPG **126** was used to confirm this result. Thus, 20 mg of the photolabile-CPG **126** was irradiated for 1 h in 1 mL of acetonitrile, the colourless supernatant was removed and then diluted with 3% TCA in a

volumetric flask to give an orange solution. The trityl assay of this solution gave a photocleaved DMTr cation ‘loading’ of  $33.8 \mu\text{molg}^{-1}$  where the original loading of the CPG was  $56.2 \mu\text{molg}^{-1}$ , thus indicated a 60% cleavage yield of the photocleavable linker. This supported the initial finding and gave confidence that good yields of oligonucleotides could be obtained. Higher cleavage yields were likely not attainable due to the dimerisation of the *ortho*-nitrosobenzaldehyde photoproduct **119** which was observed by the change in colour of the CPG from off-white to orange.

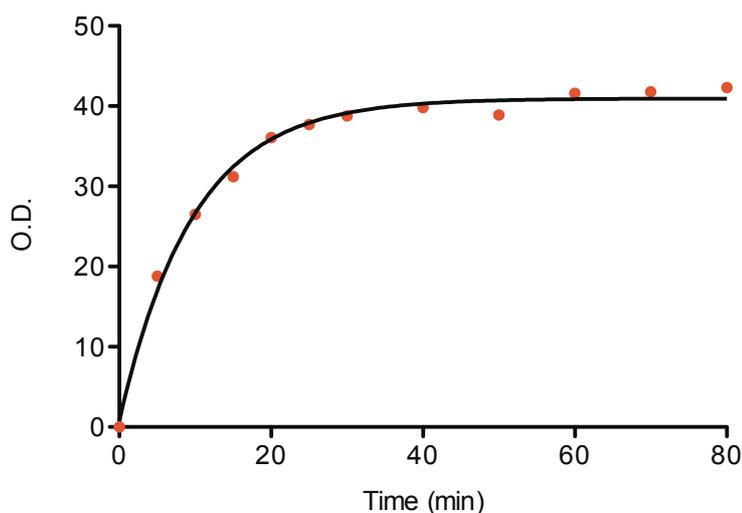


Figure 76. Rate of oligonucleotide release from the CPG when it is irradiated at  $\lambda$  365 nm. O.D. = Optical density.

Demonstrated above is a simple synthesis of a photolabile-linker **126** and preparation of the CPG solid-support. It was found to be suitable for oligonucleotide synthesis and photocleavage of the oligonucleotide was achievable in a relatively short time. With all materials and phosphoramidites prepared it was possible to begin solid-phase synthesis of acetylated oligonucleotides so development and optimisation of the synthetic methods was initiated.

## 2.4. Solid-phase synthesis of acetyl-RNA

### 2.4.1. Optimisation of the automated synthesis of acetyl-RNA oligonucleotides

Synthesis of acetylated oligonucleotides was performed using a *BioAutomation MerMade 4* automated synthesis machine. The system included preprogrammed script files to control the addition of reagents and this was used as a basis for the following work. There were several procedures and reagents that were altered and changes are described hereafter. First a brief overview of the conditions that served as a starting point for the synthesis of acetylated-RNA oligonucleotides. The synthesis cycle began with the deblock using two treatments of dichloroacetic acid (DCA) and it was at that point an approximate trityl reading was obtained by the machine. The trityl value of the elute from each deblock step is then compared to the initial deblock value and an approximate indicated percentage yield is calculated. Following on was the coupling step that added an 8-fold excess of phosphoramidite over the synthesis scale and an excess of activator ETT to the synthesis column. This step was a ‘double coupling’ meaning the CPG was treated twice for 6 minutes with the coupling mixture. Coupling of the new amidites rarely results in 100% reaction and the unreacted 5'-hydroxyls are ‘capped’ by acetylation with acetic anhydride, thus blocking them from further chain elongation. This reduces synthesis of truncated sequences with internal base deletions. The capping step is followed by oxidation of the newly formed phosphite triester to the phosphate triester using iodine, H<sub>2</sub>O and pyridine in THF. A second ‘capping’ step is then carried out but takes advantage of the acetic anhydride to remove any traces of water from the oxidation step thus ensuring anhydrous conditions for the next iteration of the synthetic cycle. In all syntheses removal of the final 5'-ODMTr group was not automatically conducted; the final trityl assay was conducted manually to enable accurate yield measurements of the full-length oligonucleotides (Figure 77).

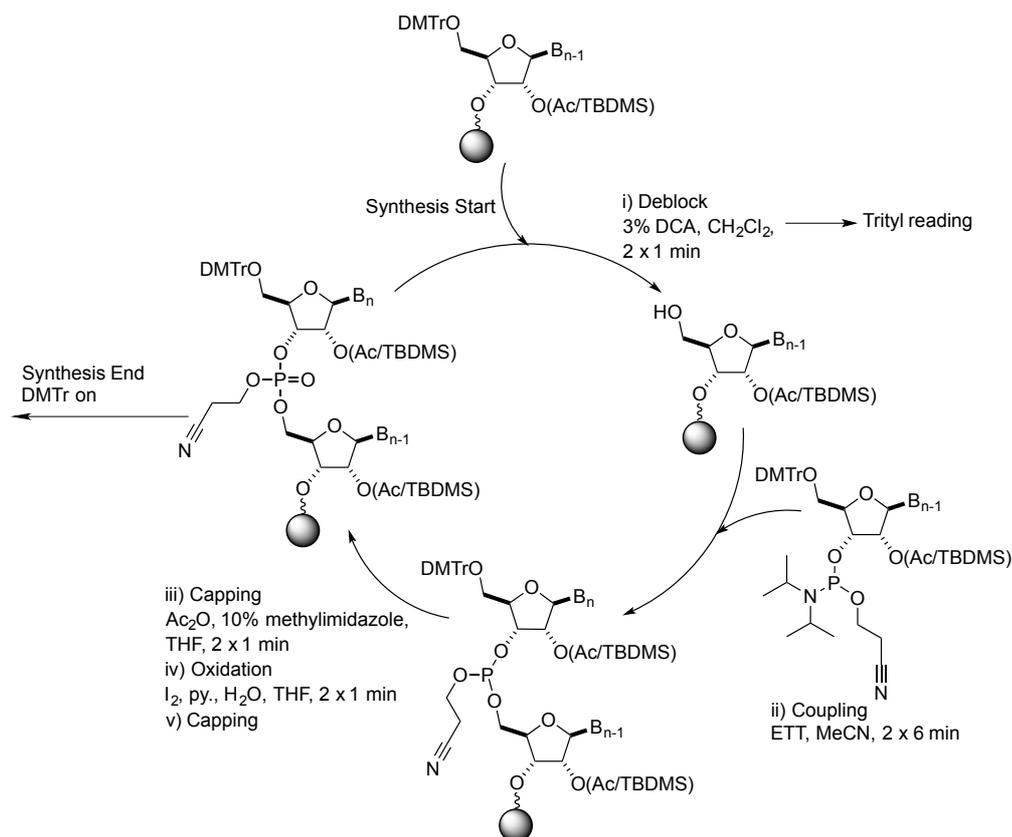


Figure 77. A standard synthesis cycle used as a basis for synthesis of acetylated-RNA oligonucleotides.

To test and develop the synthesis of acetylated-RNA oligonucleotides we designed an oligonucleotide of sequence  $5'-GCCX-3', 5'(2'-OAc)GCCX-3'P$  ( $X = C, G$ )<sup>d</sup>. This sequence was chosen as it predominantly utilises one phosphoramidite, 2'-OTBDMS C<sup>ceoc</sup> phosphoramidite **113a**, which is relatively straightforward to synthesis such that more material could be brought forward relatively swiftly if required. Inclusion of the 2'-OTBDMS G<sup>npe</sup><sub>ceoc</sub> phosphoramidite **114a** would allow deprotection of both (2-cyanoethoxy)carbonyl and the 2-(4-nitrophenyl)ethyl groups to be evaluated after an efficient synthesis was found. The use of 2'-OAc C<sup>ceoc</sup> phosphoramidite **106a** allowed evaluation of the stability of the acetyl group towards synthesis and deprotection conditions throughout the preparation of an oligonucleotide.

<sup>d</sup> The linkage nomenclature here indicates a natural 3',5'-linkage at the fourth internucleotide phosphate (5' to 3' direction) with a 2'-O-acetyl group at the same position. Where linkage isomerisation is not explicitly indicated it is assumed to be 3',5'-linked.

When the standard preprogrammed synthesis conditions were examined it was decided to alter the coupling step to a single injection *i.e.* a ‘single coupling’ in order to conserve phosphoramidites. However, reducing the reaction time and effective equivalents of phosphoramidites led to a low overall yield of 9-34%. In response to this result the reaction time of each coupling was increased to 15 minutes to give a similar overall reaction time to the 6 minute double coupling method. The knowledge gained from the synthesis of the acetylated phosphoramidites gave impetus to change the activator to BTT to improve the coupling yields by increasing the rate of reaction. These changes gave much improved yields of 64-73% of full length oligonucleotide. For the synthesis of longer oligonucleotides coupling times were increased to 20 minutes in an effort to maximise the yields further.

Once synthesis and full deprotection of an 8nt oligonucleotide with the sequence 5'-GCCX-3', 5'(2'-OAc)GCCX-3'P (X = G) had been accomplished MALDI-TOF analysis of the oligonucleotide showed a correct mass peak that corresponded to the correct singly acetylated 8nt oligonucleotide. However, peaks with masses that corresponded to bis- and tris-acetylated oligonucleotides were observed, plus 42 and 84 Da respectively. The source of the additional acetyl groups was thought to be from the capping step where the use of acetic anhydride was resulting in acetylation of the nucleobase exocyclic amines (Figure 78a). This problem has been observed by Greenberg *et al.* during the synthesis of oligonucleotides using fast-deprotecting phosphoramidites and Ultra-MILD deprotection conditions. As a solution he recommended to substitute acetic anhydride with pivalic anhydride as the capping reagent.<sup>[188]</sup> A second oligonucleotide was synthesised but with pivalic anhydride substituted as the capping agent. However, a mass peak corresponding to the bis-acetylated oligonucleotide was again observed albeit at a lower intensity. No tris-acetylated oligonucleotides were observed indicating that removal of acetic anhydride from the capping step had been partially successful (Figure 78b). During preliminary preparations of the photolabile-CPG **126** capping of the unreacted amines of the support had been carried out using acetic anhydride. It was thought that during on-column DBU deprotection of the oligonucleotides, acetyl transfer was occurring from the acetylated long-chain alkyl amines to the deprotected exocyclic amines of the nucleobases. Thus, the photolabile-CPG **126** was prepared by a modified procedure to cap the amines with pivaloyl chloride. A repeated oligonucleotide synthesis on this

pivaloyl capped CPG, and using pivalic anhydride as capping agent during automated synthesis, gave acetylated-RNA oligonucleotides with the expected number of acetyl groups (Figure 77c).

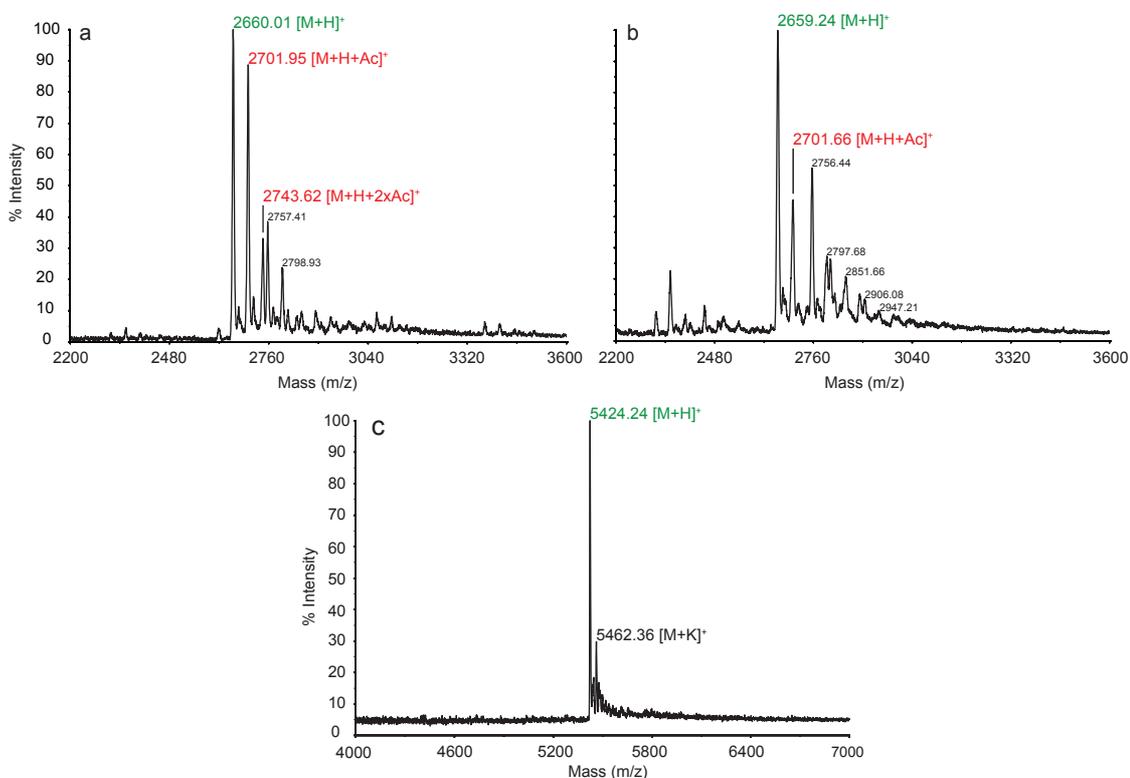


Figure 78. MALDI-TOF mass spectra of a synthesised oligonucleotide of sequence 5'-GCCG-3', 5'(2'-OAc)GCCG-3'P ( $M_r = 2662.59$ ) using a) acetyl-capped CPG and acetic anhydride used as the capping agent and b) acetyl-capped CPG and pivalic anhydride used as the capping agent. c) MALDI-TOF mass spectra of a synthesised oligonucleotide of sequence 5'-UGUGCCAGUA-3', 5'(2'OAc)-GGUUCUC-3'P ( $M_r = 5424.24$ ) using pivoyl capped CPG and pivalic anhydride as capping agent.

Use of pivalic anhydride was effective in eliminating acetylation of the exocyclic amines due its greater steric bulk. However, the increased steric bulk also decreased the rate of reaction with the 5'-hydroxyls that were not coupled with the phosphoramidite during the coupling step. As a consequence of this the capping time of 1 minute was deemed to be insufficient and was increased to 5 minutes, which was found to be effective.



*Figure 79. Reaction chamber of the MerMade 4 synthesis machine. Visible at the top right of the image are the reagent injection nozzles that move to the left to deliver reagents to the yellow synthesis column. Visible on the ends of the nozzles is build up of crystallised activator, which prevented reliable delivery of activator during synthesis of long oligonucleotides.*

The MerMade 4 oligonucleotide synthesis machine is designed such that reagents are delivered *via* injection nozzles that are separate from the synthesis columns. The ends of these nozzles are open to the inert atmosphere of the synthesis chamber, which posed some practical problems. Specifically, the activator BTT has a relatively low solubility of 0.44 M in acetonitrile and had a tendency to crystallise on the ends of the reagent delivery nozzles. As coupling and capping reaction times were far longer than standard procedures, a full synthesis cycle was approximately one hour. Depending on the sequence of the oligonucleotide some nozzles may not have been used for several hours. This led to significant precipitation of the activator at the ends of the nozzles, leading to blockage. Consequently, addition of the activating agent failed and the synthesis of  $n$ - $x$  oligonucleotides occurred (where  $n$  = the length of the desired oligonucleotide and  $x \geq 1$ ). To reduce the blockage of the nozzles an alternative activator DCI was used in place of BTT as it is more soluble and can be used at concentrations up to 1.2 M in acetonitrile (a suggested concentration is 0.5 M).<sup>[243]</sup> Despite this it was decided to take advantage of the higher solubility of DCI and use a 1.0 M solution to improve coupling yields as it was suspected that the phosphoramidites synthesised in this work were more sterically hindered than standard phosphoramidites. Although DCI is less acidic ( $pK_a = 5.2$ ) it is thought to be equally effective as the more acidic tetrazole based activators due to its greater nucleophilicity. Together with the higher concentrations of DCI it was hoped that a higher effective concentration of the activated phosphoramidites could be

achieved. Although the effect on the yields of oligonucleotides by the use of DCI was not specifically investigated no real consequences were observed. The most significant improvement was elimination of activator precipitation on the reagent delivery nozzles, which allowed for a far more consistent and reliable synthesis of full-length oligonucleotides.

In summary, conditions and reagents have been optimised for the synthesis of the partially acetylated-RNA oligonucleotides (Figure 80). The synthesis began with the unaltered deblock step using 3% DCA in CH<sub>2</sub>Cl<sub>2</sub>. Coupling of the new phosphoramidite used a single injection of reagents. To improve coupling yields the activation of phosphoramidites employed BTT and longer reaction times of at least 20 minutes were used. However, issues with solubility of BTT led to its replacement with the more soluble DCI activator used at a higher concentration of 1.0 M. Original capping conditions and acetyl blocking of the CPG amines led to *N*-acetylation of the synthesised oligonucleotides but use of pivolylyl capped CPG and pivalic anhydride as the capping reagent eliminated this issue. The capping step was also increased in time to maximise reaction of the 5'-hydroxyls with the less reactive pivalic anhydride. The oxidation step was unchanged from the original conditions.

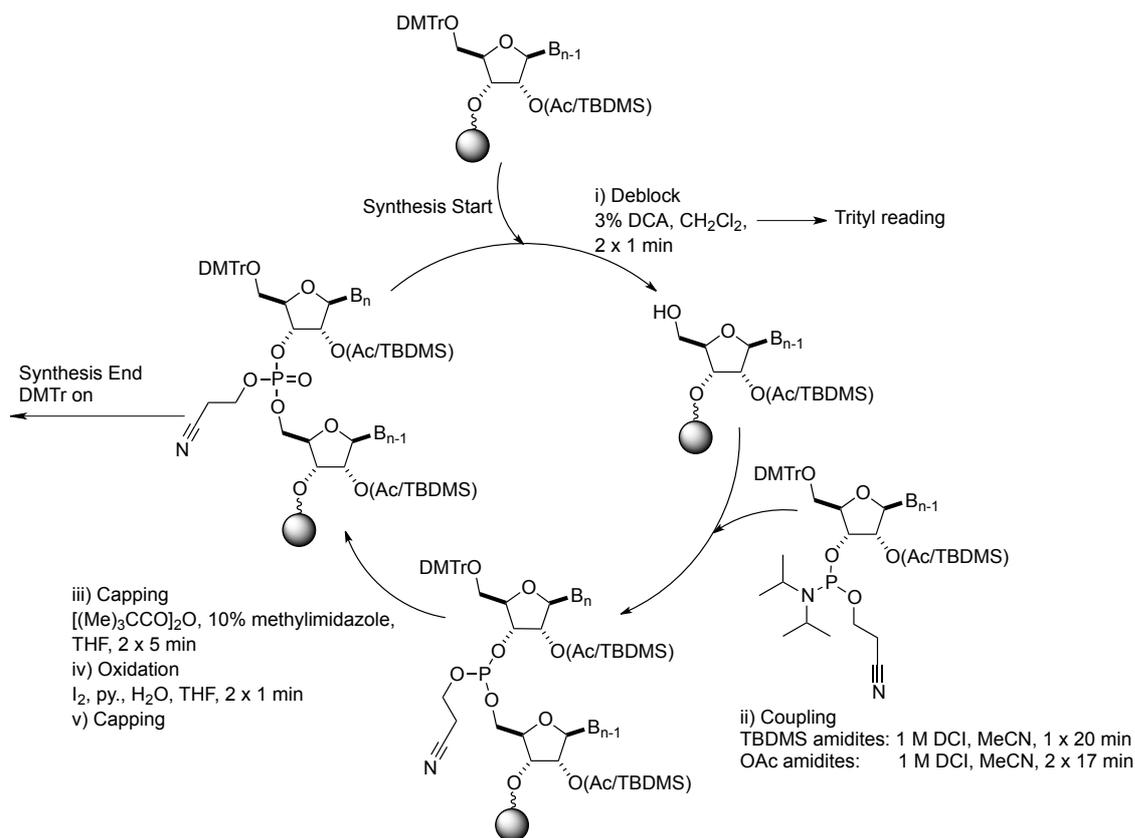


Figure 80. The optimised reaction conditions for the automated synthesis of the acetylated-RNA oligomers.

## 2.4.2. Method optimisation for the deprotection, cleavage and purification of the acetylated-RNA oligonucleotides

With an optimised automated synthesis of oligonucleotides now established it was possible to quickly synthesise oligonucleotides for the purpose of optimisation of the deprotection, cleavage and purification protocols of acetylated-RNA oligonucleotides. The general strategy was first ‘on-column’ removal of the nucleobase and phosphate protecting groups. Followed by removal of the final DMTr group and measurement of the synthesis yield by trityl assay. The next step was photocleavage of the partially protected oligonucleotide and then removal of the TBDMS protecting groups after separation from the CPG as HF reagents such as TREAT.HF are known to dissolve glass. The 2’/3’-phosphate may have also required removal, which was easily accomplished by enzymatic methods. Upon full deprotection the oligonucleotide would then be ready for purification by reverse-phase HPLC (RP-HPLC).

Nucleobase and phosphate deprotection was accomplished on-column using a combination of modified procedures by Damha *et al.*<sup>[196]</sup> and Pfeleiderer *et al.*<sup>[201]</sup> In practice, 0.5 M DBU (10 % morpholine) in anhydrous acetonitrile was initially passed over the CPG for 5 minutes and then exposed to DBU in anhydrous conditions for initially 3 hours at room temperature. However, on a few occasions presence of protected nucleobases were observed and so the CPG was exposed to 0.5 M DBU (10 % morpholine) in anhydrous acetonitrile at 40 °C for 6 hours, which effectively removed all nucleobase and phosphate protecting groups. It was crucial to carry out this deprotection under anhydrous conditions as presence of water in the strongly basic solution could lead to hydrolysis of the 2'/3'-acetyl group. On-column deprotection allowed excess DBU and by-products, notably acrylonitrile, to be easily and thoroughly washed away.

The nucleobase 2-(cyanoethoxy)carbonyl, 2-(4-nitrophenyl)ethyl and phosphate 2-cyanoethyl moieties were removed *via*  $\beta$ -elimination under DBU conditions.<sup>[244]</sup> The by-product from ceoc and ce deprotection is acrylonitrile **127** and from the npe group, *para*-nitrostyrene **128** (Figure 81a). The inclusion of the strong base labile nucleobase protecting groups meant that in particular, the relative equivalents of acrylonitrile relative to oligonucleotide were high. As such there were concerns over alkylation of the deprotected nucleobases by acrylonitrile *via* a Michael-type addition (Figure 81b). During exposure to DBU, alkylation has been seen to be very facile leading to full alkylation of thymidine in one report.<sup>[244-246]</sup> This problem is well documented in the literature and several groups have suggested scavengers of acrylonitrile to be added to the deprotection mixture. One such scavenger is nitromethane ( $pK_a = 10.2$ , H<sub>2</sub>O) but it is a good nucleophile and so it was decided that it had the potential to deacetylate a 2'/3'-acetylated oligonucleotide. In another report by Zhou *et al.*, Michael-addition of a tolyl vinyl sulphones to nucleobases were prevented by utilising morpholine that has a lower  $pK_a$  ( $pK_a = 8.4$ ).<sup>[247]</sup> The lower basicity of morpholine was attractive as a scavenger as it would be less likely to react with 2'/3'-acetyl groups. Alkylation products were not observed during the synthesis of any acetylated-RNA oligonucleotides and so the addition of morpholine was seen as a success.

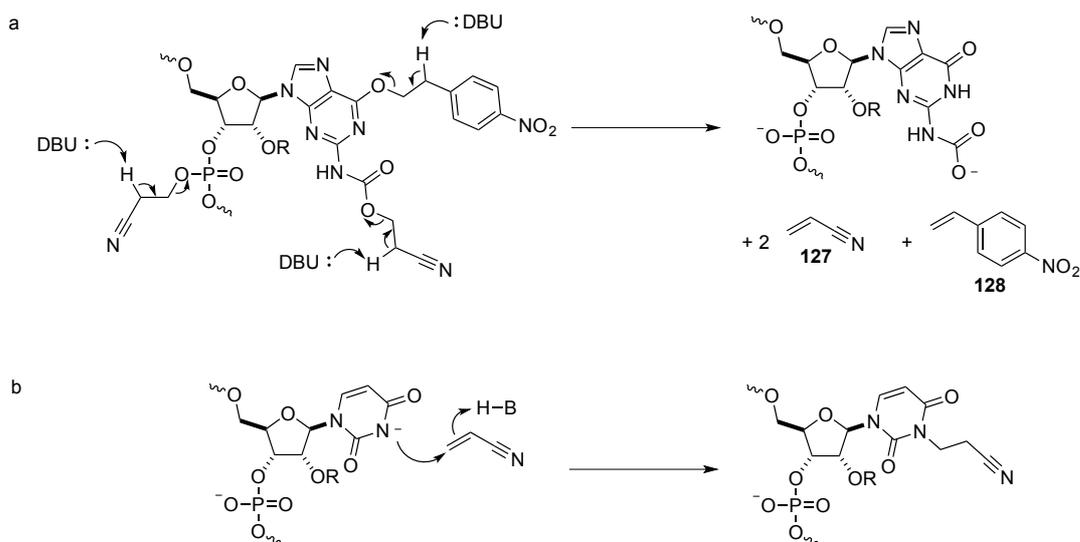


Figure 81. a)  $\beta$ -Elimination of the nucleobase and phosphate protecting groups by DBU. b) Possible  $N^3$ -alkylation of uridine nucleobase by acrylonitrile.

The next deprotection step was on-column removal of the DMTr group, which was carried out by passing 3% trichloroacetic acid (TCA) in  $\text{CH}_2\text{Cl}_2$  through the synthesis column. The TCA solution was passed over the CPG until the supernatant was colourless and the resultant orange supernatant was collected into a 50 mL volumetric flask. This enabled an accurate trityl reading and gave a CPG loading value corresponding to the yield of full-length oligonucleotides. The DMTr group was not removed before DBU treatment as a 5'-hydroxyl under basic conditions could act as a nucleophile and lead to acetyl transfer away from the 2'/3'-acetyl positions.

The acetyl-RNA oligonucleotide at this point was still attached to the CPG but was now prepared for UV irradiation to cleave the photolabile linker and release the TBDMS protected oligonucleotide. However, an unexpected and time-consuming problem was encountered, which is described below. The syntheses of several 8nt oligonucleotides, related to those described in section 2.4.1, were in general high yielding as calculated by the final trityl assay. These 8nt oligomers were used in the following photocleavage experiments and each were deprotected on-column using the DBU conditions as described above. The photocleavage was carried out according to a modified procedure by Venkatesan *et al.*, in which the CPG was layered with 3:1 mixture of  $\text{H}_2\text{O}:\text{MeCN}$  and subjected to UV irradiation for 1 hour.<sup>[248]</sup> In all cases, a 10  $\mu\text{L}$  sample of the supernatant was removed and the amount of cleavage oligonucleotide quantitated by UV absorption. The initial irradiation experiment (Table 4, entry 1) shows that although

the synthesis yield was high, the apparent cleavage yield of oligonucleotide was very poor. At this point the issue was not clear, so it was chosen to synthesis a poly-U 8nt oligonucleotide and it was found that photocleavage was efficient and high yielding (Table 4, entry 2). To rule out possible unspecified side-reactions, two poly-U 8nt oligomers were synthesised one with a 2'-acetylated U nucleoside substitution and the second with two G substitutions (Table 4, entries 3 and 4). Again the photocleavage yield for the U 8nt oligomer with one 2'-acetylated nucleoside was comparable to the poly-U photocleavage. However, it was noticed the photocleavage yield for the doubly G substituted sequence was slightly reduced (Table 4, entry 4). It was questioned whether nucleophilic attack of the free amines of G and C at the nitroso position of the *ortho*-nitrosobenzylaldehyde **118** photoproduct and formation of a diimine was the cause of the poor photocleavage yields. To ascertain whether this was the cause, glyoxylate was used to act as a nitroso scavenger in the irradiation solvent (Table 4, entries 5 and 6).<sup>[249, 250]</sup> It was also noted that with purely aqueous solvents the CPG was not wetted and required some organic solvents to reduce the surface tension of water. Again, photocleavage was carried out with the CPG exposed to two different concentrations of glyoxylate but the yields were found to be poor suggesting either the glyoxylate was not an efficient scavenger of the photoproduct **121** or covalent reattachment *via* the exocyclic amines of the oligonucleotide was not the issue.

Entry	RNA Sequence (5' to 3'P)	Trityl Yield (%)	UV Irradiation (365 nm, 1 h)		
			Irradiation Solvent (1 mL)	Yield (%)	Est. Cleavage (%)
1	GCCCC <sup>2'OAc</sup> GCCC	69	3:1 (H <sub>2</sub> O:MeCN)	1	1
2	UUUUUUUU	69	3:1 (H <sub>2</sub> O:MeCN)	55	80
3	UUU <sup>2'OAc</sup> UUUU	78	3:1 (H <sub>2</sub> O:MeCN)	62	79
4	<sup>G<sup>2',5'</sup></sup> UUU <sup>G<sup>2',5'</sup></sup> UUU	67	3:1 (H <sub>2</sub> O:MeCN)	48	71
5	GCCC <sub>3'OAc</sub> GCCC	79	20 mM glyoxylate 10 % MeCN	12	18
6	GCCC <sub>3'OAc</sub> GCCC	59	50 mM glyoxylate 25 % MeCN	4	7
7	GCCC <sup>2'OAc</sup> GCCC	62	MeCN	4	6
8	GCCG <sup>2'OAc</sup> GCCG	51	DMSO	48	95

Table 4. Photocleavage experiments, the trityl yield is given to assess the success of the synthesis and is used to calculate the percentage yield of oligonucleotide cleaved from the CPG. The oligonucleotides were all deprotected on-column prior to UV irradiation using the following conditions 0.5 M DBU (10 % morpholine) in MeCN. Superscripts/subscripts denote site of acetylation or linkage isomerism, green denotes site of 3'-5' natural linkage isomerism and red denotes site of 2'-5' unnatural linkage isomerism.

It was decided that the latter conclusion was the more likely as the CPG became orange during UV irradiation suggesting significant dimerization to **119** was occurring (see Figure 72) rather than reaction with the cleaved oligonucleotide. It was noted that sequences that were GC rich did not produce high photocleavage yields whereas U rich sequences did. All sequences except Table 4, entry 7 were deprotected under DBU conditions and the resultant oligonucleotide could have existed as a DBU salt and may have been relatively non-polar and poorly solvated by water/acetonitrile mixtures. Thus, to deduce whether solubility of the cleaved oligonucleotide was the limiting issue the organic solvents, acetonitrile and DMSO, were explored as irradiation solvents. It was found that acetonitrile did not lead to improved irradiation yields (Table 4, entry 7). Gratifyingly, using DMSO was found to be an excellent choice and resulted in improved yields of cleaved oligonucleotides (Table 4, entry 8).

With optimised conditions found for the photocleavage of acetylated-RNA oligonucleotides, the final TBDMS protecting groups could now be removed. Standard conditions were employed for the removal of the 2'/3'-*O*-TBDMS groups that involved dissolving the oligonucleotide in anhydrous DMSO and treating it with TREAT.HF at 65 °C for 3 hours.<sup>[211]</sup> The oligonucleotides were most commonly isolated by precipitation with sodium acetate after which point the oligonucleotide was quantitated by UV absorption to give a crude yield. The crude samples were then analysed by MALDI-TOF mass spectrometry in order to identify the desired oligonucleotide.

The final step before purification of the oligonucleotides was enzymatic removal of the terminal 3'-phosphate, which was accomplished by treatment with calf intestinal phosphatase (CIP) in PBS buffer (pH = 7.4) at 37 °C for 1 hour. This step was optional but in all cases during this work the terminal phosphate was removed. The fully prepared oligonucleotides were finally purified by strong anion exchange HPLC (SAX-HPLC) chromatography. The fractions containing the target oligonucleotide were combined and dialysed against 10 mM TEAA buffer to remove excess buffer and salts used in HPLC purification. Finally the purity of the oligonucleotides was confirmed by analytical SAX-HPLC and characterised by MALDI-TOF mass spectrometry.

In summary, a deprotection strategy has been optimised to allow isolation of pure partially acetylated-RNA oligonucleotides:

1. On-column nucleobase and phosphate deprotection (0.5 M DBU, MeCN 10 % morpholine).
2. On-column 5'-hydroxyl deprotection and trityl assay (3 % TCA, CH<sub>2</sub>Cl<sub>2</sub>).
3. UV irradiation – photocleavage (365 nm, 1 hour).
4. 2'/3'-*O*-TBDMS removal (DMSO, 65 °C, 3 hours).
5. Dephosphorylation, CIP (PBS buffer, pH = 7.4, 37 °C, 1 hour).
6. SAX-HPLC purification.

In particular, reaction of nucleobase and phosphate deprotection by-products resulting in alkylation of the nucleobases was considered and steps were taken to prevent alkylation by using an acrylonitrile scavenger. The poor yields of the photocleavage step were found to be due to solubility of the partially deprotected oligonucleotide and the use of DMSO was found to effectively dissolve the cleaved oligonucleotide.

## 2.5. Synthesis of acetyl-RNA oligonucleotides

The optimised synthesis conditions resulted in high yields of full length oligonucleotides with average stepwise coupling yields of >98.5 % which is comparable with standard 2'-*O*-TBDMS chemistry (Table 5, entries 1-4). However, synthesis of an acetylated-GUAA tetraloop showed a significantly lower yield with an average stepwise coupling yield of 85.1 % (Table 5, entry 5). A crude MALDI-TOF spectra of the oligonucleotide showed the major failure sequences to be due to the failed couplings of 2'-*O*-TBDMS A<sup>ccoc</sup> phosphoramidite **112a** and suggests phosphoramidite quality may have been to blame.

Entry	RNA Sequence (5' to 3')	Trityl yield (%)	Avg. coupling yield (%)	UV Irradiation	
				Yield (%)	Est. Cleavage (%)
1	UGUGCCAGUA <sup>2',OAc</sup> GGUUCUC	86	99.1	53	62
2	UGUGCCAGUA <sub>3',OAc</sub> GGUUCUC	89	99.3	56	64
3	CCAG <sup>2',OAc</sup> UAGGU <sup>2',OAc</sup> UCUC	83	98.6	62	75
4	GAGA <sup>2',OAc</sup> ACC <sup>2',OAc</sup> UACUGG	66	96.8	57	86
5	GCCG <sup>2',OAc</sup> UAAGGC	20	85.1	31	155

Table 5. Synthesised oligonucleotides, final yields are not given as only 1/2 or 1/3 of the material was purified by HPLC. Estimates of final yields range between 4-19 %, which is on par with yields of oligonucleotides from commercial sources. Superscripts/subscripts denote site of acetylation or linkage isomerism, green denotes site of 3'-5' natural linkage isomerism and red denotes site of 2'-5' unnatural linkage isomerism.

Photocleavage yields within the range of 60-80 % were obtained and are comparable to previous photocleavage experiments. The high value obtained for the percentage cleavage with entry 5 was likely due to the higher concentration of failure sequences, which are not taken into account when calculating the yield by the trityl assay. Post photocleavage, the yield is calculated by UV absorption and absorption by the 5'-OAc capped failure sequences will contribute to the calculated yield.

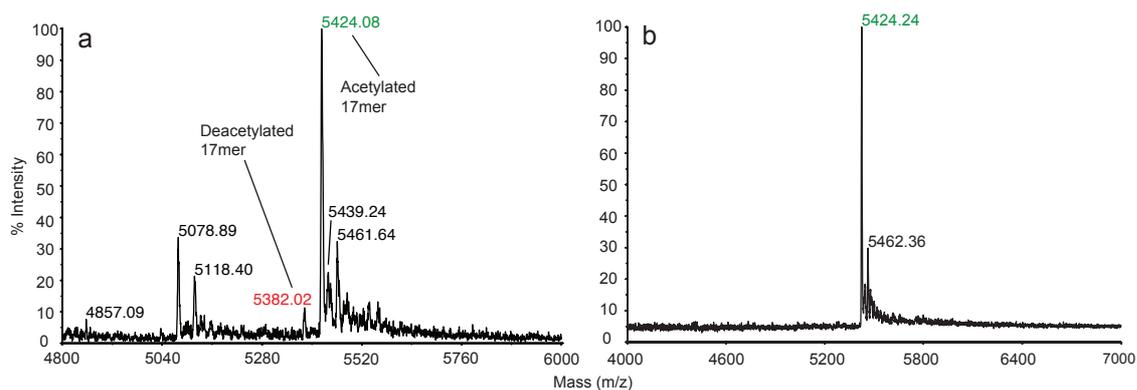


Figure 82. MALDI-TOF spectra showing that it was possible to remove the deacetylated oligonucleotides formed during deprotection: a) crude mixture before SAX-HPLC shows the mass of an acetylated 17nt oligomer (see Table 5, entry 1,  $M_r = 5423.29$ ) in green and the small amount of deacetylated 17nt oligomer in red ( $-42 \text{ Da} = \text{Ac}$ ), b) is the same material after SAX-HPLC purification and shows a mass peak for only the target oligonucleotide.

During the deprotection it was expected that a small amount of deacetylation of acetylated-RNA oligonucleotides would be observed as an unavoidable process. This indeed was the case but where deacetylation occurred only a small amount of hydrolysis was observed, up to approximately 10 % estimated by comparison of the MALDI-TOF mass spectrum peak height corresponding to the full length acetylated product. In nearly all cases it was possible to remove the deacetylated-RNA oligomers during SAX-HPLC (Figure 82).

In summary, successful optimisation of the automated synthesis cycle and the deprotection procedures has enabled synthesis of several partially 2'/3'-O-acetylated RNA oligonucleotides. The calculated average coupling yields are comparable to those obtained by commercial 2'-O-TBDMS chemistry.<sup>[196]</sup> However, average coupling yields of the final oligonucleotide (Table 5, entry 7) were found to be notably lower, as small reductions in each coupling yield has a significant effect on the yield of full length

product. Due to time constraints, reasons for the lower coupling yields with this particular sequence were not fully explored. One possibility is age of the phosphoramidites, as inadequate storage of the materials could lead to excessive absorption of atmospheric water and/or oxidation of the phosphite such that both would be detrimental to the coupling efficiency. Due to the novel nature the 2'/3'-*O*-acetylated phosphoramidites it would be useful to compare coupling efficiencies of freshly made material with material stored for different lengths of time and under varying storage conditions.

### 3. Properties of 2'/3'-*O*-acetyl RNA oligonucleotides and consequences for the non-enzymatic replication of RNA

#### 3.1. Background

As previously described in Chapter 1.6.3, chemoselective acetylation allows the rapid and efficient template-directed ligation of short RNA oligomers. Subjecting a mixture of 3'P and 2'P oligomers to the acetylation-ligation reaction conditions showed a 700-fold selectivity for ligation of the 3'P oligomers. Thus, the products of acetylation-ligation reactions are partially 2'/3'-*O*-acetylated RNA oligonucleotides (acetyl-RNA) with acetyl groups present predominantly at internal 2'-positions. Acetyl-RNA is predicted to have greater potential for replication over extant RNA due to favourable conformational changes and suppression of key hydrogen bonding interactions.

The furanose five-membered ring of the sugar-phosphate backbone of RNA is known to exist in two main conformations, usually referred to as *C2'-endo* (south conformation) and *C3'-endo* (north conformation) (Figure 83).<sup>[251]</sup> Acetylation of the 2'-hydroxyl of a 3'P oligonucleotide favours the *C3'-endo* sugar pucker because of the electronegative acetate group that stabilises a non-bonding ( $\sigma_{C-H3'} \rightarrow \sigma_{C-O2'}^*$ ) overlap.<sup>[252]</sup> The *C3'-endo* is the dominant sugar pucker in duplex RNA which forms a right-handed helix and is otherwise known as A-form RNA (Figure 83).<sup>[253]</sup> Thus, 2'-acetylation was predicted to increase the *C3'-endo* conformation and improve the propensity of acetyl-RNA to form duplex over other RNA structures such as turns and loops where the *C2'-endo* sugar pucker is preferred.

Modified RNA such as 2'-deoxy-2'-OMe-RNA and 2'-deoxy-2'-F-RNA are conformationally limited to the *C3'-endo* sugar pucker. The hybrid duplexes such as 2'-OMe-RNA:RNA and 2'-F-RNA:RNA have also been shown to possess greater thermodynamic stability when compared to their RNA:RNA hybrids.<sup>[254, 255]</sup> However, thermodynamic studies on 2'-*O*-alkyl-RNA:RNA duplexes have shown reduced  $T_m$  (defined as the temperature at which half of the oligonucleotides are folded/associated and half are unfold/dissociated) and thermodynamic stability, which decreased further

with an increase in alkyl chain length.<sup>[256]</sup> Moreover, Damha *et al.* have synthesised 2'-*O*-levulinyl modified RNA oligonucleotides and showed that with extensive modification no duplex formation was observed.<sup>[196]</sup> With this in mind, 2'-acetylation of RNA was predicted to reduce duplex stability due its greater steric bulk compared with the 2'-OMe and 2'-F modifications and functional group similarity (ester group) to the levulinyl group.

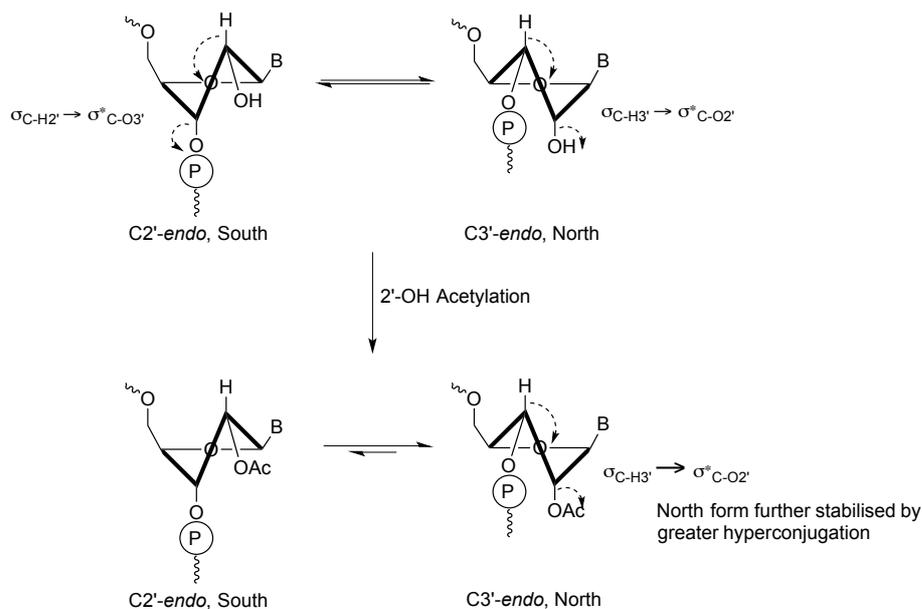


Figure 83. Conformational changes due to acetylation of RNA.

RNA has the ability to form secondary and tertiary structure to give certain sequences phenotypic properties. However, highly structured RNA, such as in the ribosome, inhibits non-enzymatic replication of RNA. An analysis of structured RNA brings to light a common and ubiquitous structural motif that has been termed the A-minor interaction. This is an important interaction that involves insertion of unpaired adenines into the minor grooves of RNA helices where hydrogen-bonding through their N(1), N(3) and 2'-hydroxyl stabilises tertiary structure (Figure 84).<sup>[257, 258]</sup> The partial acetylation of RNA oligonucleotides has the potential to block a significant number of 2'-hydroxyl hydrogen bonding interactions at A-minor sites. Additionally, steric demand and electronegativity of the acetyl group may be poorly accommodated deep within a structure. These factors are predicted to weaken the A-minor interactions, thereby reducing the formation of tertiary structure and so favour duplex RNA. Reduction of tertiary structure would allow acetyl-RNA oligonucleotides to undergo a period of genotypic behaviour (replication) and potentially overcome the problem of

product inhibition. Subsequent deacetylation, without phosphodiester backbone cleavage, of replicated acetyl-RNA will allow formation of A-minor interactions and emergence of tertiary structure.<sup>[259]</sup>

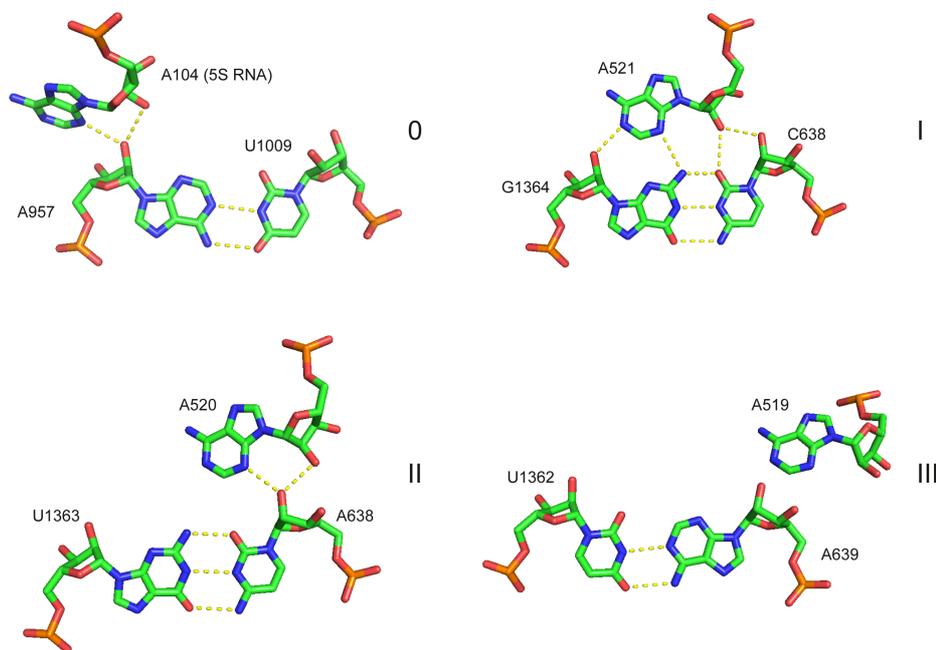


Figure 84. Examples of the four main A-minor interactions from the *H. marismortui* 50S subunit. Types I and II are the most common A-minor motifs and are highly specific for adenine bases. Type III and the much rarer type 0 motif are less specific as the bases stack against the receptor (base-paired) riboses however, adenine bases are still preferred as they maximise the stacking interaction. Redrawn from PDB file 1FFK using MacPyMOL.<sup>[257]</sup>

### 3.2. Duplex stability of acetyl-RNA assessed by UV melting analysis

Optical melting was chosen to assess the thermodynamics of acetyl-RNA secondary structure because small amounts of material are required and measurements can be obtained fairly quickly. The vast majority of thermodynamics of nucleic acids have been measured by optical melting.<sup>[260]</sup> This takes advantage of the phenomenon of stacked nucleobases such as those in duplex structures, whereby the UV absorbance of an oligonucleotide increases as the temperature increases, that is otherwise known as hyperchromism. Absorption of light causes an excitation of electrons in the nucleobases of oligonucleotides, which in turn causes an electric dipole transition moment. Where bases are stacked or within secondary structure the transition dipole induces a dipole in neighbouring bases. These induced dipoles are opposite to the transition dipole and so

reduce the effective transition dipole. As the extinction coefficient is proportional to the square of the magnitude of the transition dipole, stacked bases or structured oligonucleotides are hypochromic compared to unpaired oligonucleotides.<sup>[261]</sup> Heating and cooling a solution of oligonucleotide causes a reversible melting and annealing of oligonucleotides, which is observed as an increase and decrease respectively in absorption. In practice, a plot of absorption versus temperature gives a sigmoidal melting curve.

The  $T_m$  and thermodynamic parameters (enthalpy and entropy) can be calculated from the absorption versus temperature curves. To calculate the  $T_m$  the melting curve needs to be converted to a fraction associated/folded versus temperature curve. This is done by fitting straight lines to the lower and upper baselines of the curve that relate to the fully associated and fully dissociated states respectively (Figure 85a, see also *Equation 5*). The computed equations of these lines are used to convert each absorbance value into a fraction of the difference between the lower and upper baseline. This gives a fraction associated versus temperature curve and the  $T_m$  is the temperature at which  $\alpha = 0.5$  (Figure 85b). An important assumption is that the oligonucleotides exist in a two-state equilibrium, either duplexed/folded or single stranded/unfold. The  $\alpha$  values at the  $T_m$  transition ( $0.15 < \alpha < 0.85$ ) are used to determine  $\Delta H^\circ$ ,  $\Delta S^\circ$  and ultimately  $\Delta G^\circ$ . It is at this transition point that gives the true measure of the stability of an oligonucleotide structure. A sharp transition indicates a strongly temperature dependent affinity and a relatively more stable structure (a more favourable  $\Delta G^\circ$ ). Conversely, a transition over a wider temperature range indicates a less favourable  $\Delta G^\circ$  and a less stable structure. The association constant ( $K_a$ ) is calculated for each  $\alpha$  value at the  $T_m$  transition using Equation 2 and is related to  $\Delta G^\circ$  by Equation 3. Rearrangement of Equation 3 and plotting  $\ln(K_a)$  versus  $1/T_m$  gives the van't Hoff plot (Figure 85c), which should follow a straight line if the two-state assumption holds. From the linear regression, the slope gives  $\Delta H^\circ$  and the y-intercept gives  $\Delta S^\circ$  allowing calculation of  $\Delta G^\circ$  (see Chapter 6.3 for further details).<sup>[260, 262, 263]</sup>

$$K_a = \frac{\alpha}{\left(\frac{C_T}{n}\right)^{n-1} (1 - \alpha)^n}$$

Equation 2. Calculation of the association constant ( $K_a$ ) for non-self-complementary sequences, in terms of fraction associated ( $\alpha$ ), total oligonucleotide concentration ( $C_T$ ) and the molecularity ( $n$ , e.g.  $n = 2$  for a bimolecular interaction).

$$\Delta G^\circ = -RT \ln(K_a) = \Delta H^\circ - T \cdot \Delta S^\circ$$

$$\ln(K_a) = \frac{-\Delta H^\circ}{R} \cdot \frac{1}{T} + \frac{\Delta S^\circ}{R}$$

Equation 3. Derivation to calculate the thermodynamic parameters of the melting curves.  $R$  = Gas Constant ( $8.314 \text{ JK}^{-1} \text{ mol}^{-1}$ ),  $T$  = temperature ( $^\circ\text{C}$ ),  $\Delta G^\circ$  = Gibbs free energy ( $\text{kJmol}^{-1}$ ),  $\Delta H^\circ$  = Enthalpy ( $\text{kJmol}^{-1}$ ),  $\Delta S^\circ$  = Entropy ( $\text{JK}^{-1} \text{ mol}^{-1}$ ) and  $K_a$  = association constant.

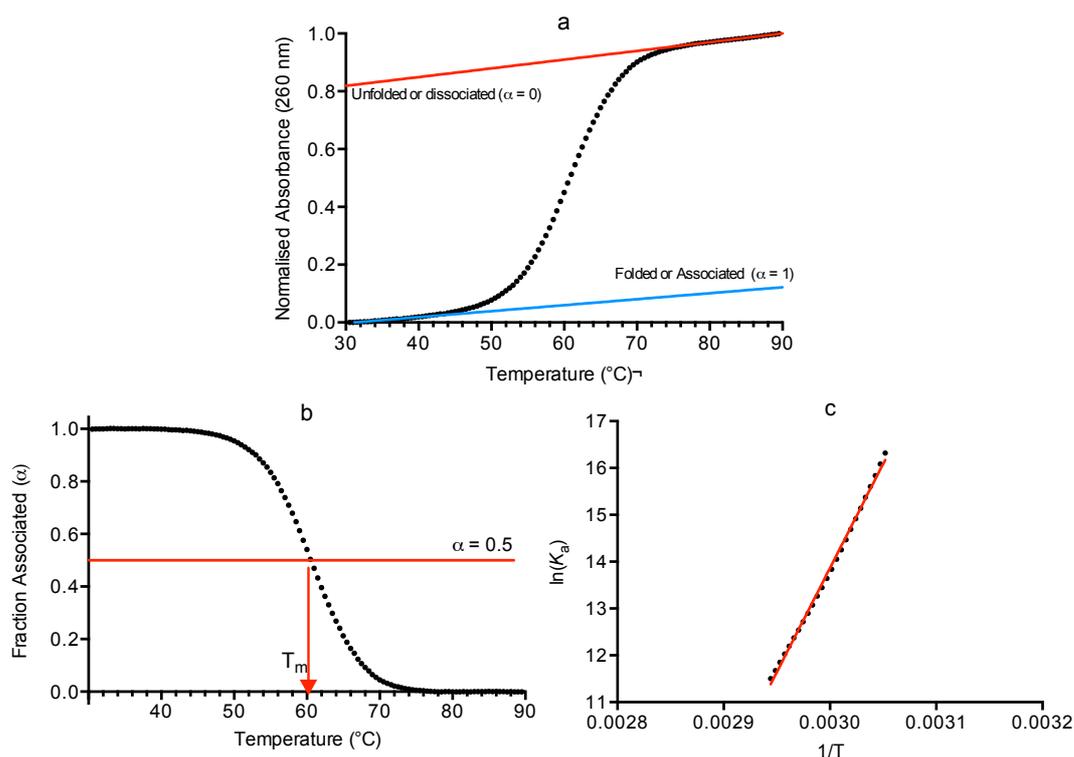
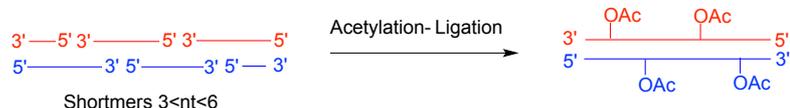


Figure 85. a) A typical melting curve with fitted upper and lower baselines. b) Fraction associated versus temperature curve with graphical representation for extraction of the  $T_m$  value. c) The van't Hoff plot with linear regression, the slope and intercept of the linear regression give the thermodynamic values for  $\Delta H^\circ$  and  $\Delta S^\circ$ . (Data and analysis of Table 6, entry 7).

To study the effect of acetylation on RNA duplex stability, sequences were selected based upon oligonucleotides originally used in the template directed non-enzymatic oligonucleotide ligation experiments by Szostak *et al.*<sup>[128, 129]</sup> The monoacetylated-RNA oligonucleotides (Table 6, entries 3-4) are based upon the acetylation-ligation products of recent work by Bowler *et al.*<sup>[146]</sup> The two complementary bisacetylated 13nt oligonucleotides (Table 6, entries 7) were designed to test the effect of multiple acetyl groups within in a duplex. The position of the acetyl groups was chosen to take into account two considerations: the first was to utilise each of the nucleoside bases; the second was so that the acetylated-RNA oligonucleotides resembled one that could have been prebiotically assembled from tiled shortmers of 3<nt<6 (Figure 86). The following results and discussion relate to data contained within Table 6.



*Figure 86. Prebiotic ligation of shortmers and considerations for the positioning the acetyl groups within the 13nt oligonucleotides for melting curve measurements.*

Entry	RNA Sequence (5' to 3')	Complement (5' to 3')	$T_m$ (°C)	$\Delta T_m$ (°C)	$\Delta T_m/OAc$ (°C)	Duplex Comparison
1	<u>UGU</u> CCAGUAGGUUCUC	GAGAACCUACUGG	74.7	+0.9	-	-
2	<u>UGU</u> CCAGUA <sup>2',5'</sup> GGUUCUC	GAGAACCUACUGG	67.8	-6.9	-	-
3	<u>UGU</u> CCAGUA <sup>2',0Ac</sup> GGUUCUC	GAGAACCUACUGG	71.6	-3.1	-3.1	-3.1
4	<u>UGU</u> CCAGUA <sup>3',0Ac</sup> GGUUCUC	GAGAACCUACUGG	72.0	+4.2	+4.2	+4.2
5	<u>UGU</u> CCAGUAGGUUCUC	GAGA <sup>2',0Ac</sup> ACC <sup>2',0Ac</sup> UACUGG	68.5	-5.9	-3.0	-3.0
6	<u>UGU</u> CCAGUA <sup>2',0Ac</sup> GGUUCUC	GAGA <sup>2',0Ac</sup> ACC <sup>2',0Ac</sup> UACUGG	65.6	-9.1	-3.0	-3.0
7	CCA <u>G</u> <sup>2',0Ac</sup> UAGGU <sup>2',0Ac</sup> UCUC	GAGA <sup>2',0Ac</sup> ACC <sup>2',0Ac</sup> UACUGG	61.5	-12.3	-3.1	-3.1
8	CCA <u>G</u> <sup>2',0Ac</sup> UAGGU <sup>2',0Ac</sup> UCUC	GAGAACCUCUACUGG	67.9	-5.9	-3.0	-3.0
9	CCA <u>G</u> UAGGUUCUC	GAGA <sup>2',0Ac</sup> ACC <sup>2',0Ac</sup> UACUGG	67.4	-6.4	-3.2	-3.2
10	CCA <u>G</u> UAGGUUCUC	GAGAACCUACUGG	73.8	-0.9	-	-

Continued  
below.

Entry	$\Delta H^\circ$ (kJmol <sup>-1</sup> )	$\Delta S^\circ$ (Jmol <sup>-1</sup> K <sup>-1</sup> )	$\Delta G^\circ_0$ (kJmol <sup>-1</sup> )	$\Delta G^\circ_{37}$ (kJmol <sup>-1</sup> )	$\Delta \Delta H^\circ$ (kJmol <sup>-1</sup> )	$\Delta \Delta S^\circ$ (Jmol <sup>-1</sup> K <sup>-1</sup> )	$\Delta \Delta G^\circ_{37}$ (kJmol <sup>-1</sup> )	$\Delta \Delta G^\circ_{37}/OAc$ (kJmol <sup>-1</sup> )
1	-530.0	-1410.5	-144.7	-92.6	-8.4	-20.0	-2.3	-
2	-421.7	-1123.7	-114.7	-73.2	+108.4	+286.8	+19.4	-
3	-420.4	-1113.3	-116.3	-75.1	+109.6	+297.2	+17.4	+17.4
4	-517.2	-1385.1	-138.9	-87.7	-95.6	-261.4	-14.5	-14.5
5	-428.5	-1140.6	-116.9	-74.7	+101.5	+269.9	+17.8	+8.9
6	-363.9	-961.0	-101.4	-65.9	+166.1	+449.5	+26.7	+8.9
7	-361.7	-967.4	-97.5	-61.7	+159.9	+423.1	+28.6	+7.2
8	-462.7	-1243.0	-123.2	-77.2	+58.8	+147.4	+13.1	+6.6
9	-461.9	-1242.2	-122.6	-76.7	+59.6	+148.3	+13.6	+6.8
10	-521.6	-1390.5	-141.8	-90.3	+8.4	+20.0	+2.3	-

Table 6. Effect of acetylation on the  $T_m$  and thermodynamic parameters of various oligonucleotide duplexes. Each measurement used 2.5  $\mu$ M of each oligomer to give a total RNA concentration of 5  $\mu$ M, in 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.5 mM Na<sub>3</sub>EDTA buffer (pH 7), 1 M NaCl and a temperature range of 30-90 °C. Underlined nucleotides denote overhanging sequence, superscripts/subscripts denote site of acetylation or linkage isomerism, green denotes site of 3'-5' natural linkage isomerism and red denotes site of 2'-5' unnatural linkage isomerism.

Comparing entries 1 and 10 it can be seen that the 4nt overhang (underlined) in the 17nt RNA oligonucleotide confers a  $T_m$  increase ( $\Delta T_m = +0.9$  °C) and a decrease in  $\Delta G^\circ_{37}$  ( $\Delta \Delta G^\circ_{37} = -2.3$  kJmol<sup>-1</sup>) indicating a slight stabilising effect. The increased stability is likely due to base stacking of the overhang.

Analysis of the  $T_m$  data showed a remarkable regularity to the effect of 2'-acetylation on the value of the  $T_m$ . With increasing number of acetyl groups the  $T_m$  decreased with a consistent  $\Delta T_m$  of between  $-3.0$ – $-3.2$  °C per acetyl group. The attenuation of the  $T_m$  did not appear to depend on the sequence location of the acetyl groups. Acetylation at the terminal positions were not explored, although it is known that other modifications such as the 2'-*O*-neopentyl located at the ends give a less significant reduction in duplex stability.<sup>[264]</sup>

It is well known that A-form RNA duplexes are well hydrated in both the major and minor grooves with regular arrangements of water molecules. In comparison to DNA duplexes, RNA duplexes are more stable partially because the 2'-hydroxyl causes the sugar to favour the C3'-*endo* pucker and form an A-form duplex. But crucially, the 2'-hydroxyl, that is both a hydrogen bond acceptor and donator, is directed into the minor groove. It is this ability that confers the largest stabilising effect as the 2'-hydroxyl allows for a more extensive and ordered hydrogen bonding network within the minor groove.<sup>[265-268]</sup> X-ray crystallographic studies show that several water molecules in the minor groove form a bridge between the 2'-hydroxyl groups of opposite strands. They are also found to hydrogen bond to the nucleobases thus providing a link between the backbone and nucleobases. This hydrogen bond network contributes to a significant enthalpy driven increase in thermodynamic stability of RNA duplex but as a result RNA also takes an inevitable entropic penalty for the increased hydration (Figure 87).<sup>[266, 267]</sup>

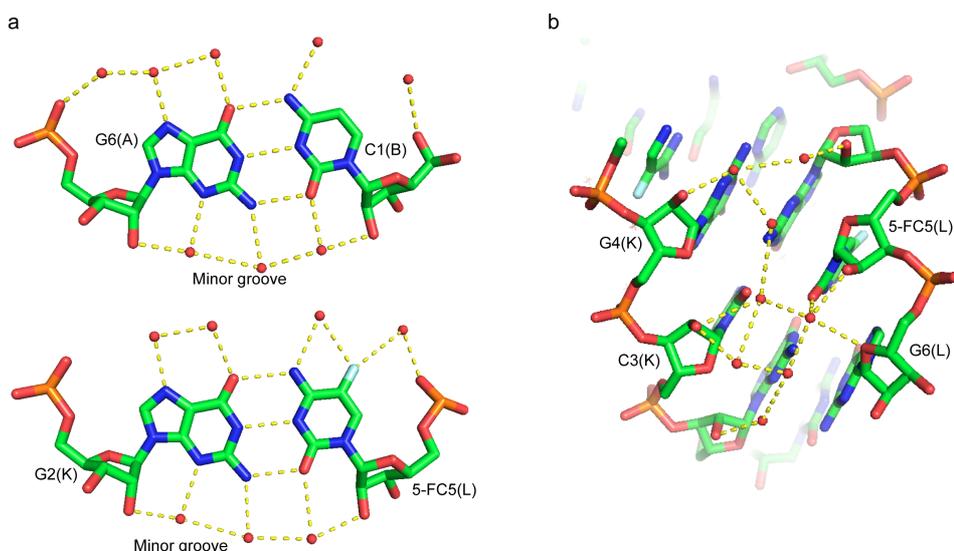


Figure 87. a) Hydration of C:G pairs illustrating the bridging of water molecule in the major and minor grooves and the hydrogen bonding to the nucleobases. b) Hydration network in the minor grooves from the crystal structure of a RNA duplex. Hydrogen bonds are indicated by dashed lines and water molecules are the red spheres. Redrawn from PDB file 3JXQ using MacPyMOL.<sup>[267]</sup>

Where an RNA oligonucleotide was 2'-O-acetylated either on one or both duplex strands it led to an increase of both  $\Delta H^\circ$  and  $\Delta S^\circ$ . The positive  $\Delta\Delta S^\circ$  values indicated a favourable entropic consequence of 2'-acetylation and preorganisation of the sugar pucker could contribute to the entropic advantage as is also suggested for other types of 2'-O-modifications such as the 2'-O-methyl modification.<sup>[255, 264, 269]</sup> However, it was believed that the effect of preorganisation was minimal as RNA duplexes generally adopt the C3'-endo sugar pucker and 2'-O-acetylation may not contribute further significant entropic increase. Rather, the more significant factor is that the 2'-O-acetyl decreased the hydrogen bonding ability at the 2'-position and is thought to result in the reduced hydration of the minor groove or entropic gain. The reduced hydration in the minor groove therefore disrupted the ordered network of water molecules causing the large and positive  $\Delta\Delta H^\circ$  or enthalpic penalty. In other words, 2'-O-acetylation is preventing the duplex RNA from anchoring water molecules (the observed favourable entropy change) *via* the 2'-hydroxyl and so preventing formation of the stabilising water bridges (the dominating observed unfavourable enthalpy change) such that the duplex is destabilised (unfavourable Gibbs energy change).

In general terms, the thermodynamic data showed that acetylation of the 2'-hydroxyls causes a decrease in the duplex stability compared to the non-acetylated parent duplex

(entries 3,5,6 vs. 1 and entries 7,8,9 vs. 10). However, in contrast to the decrease in  $T_m$ , the addition of each 2'-*O*-acetyl group does not infer equal change in  $\Delta G^\circ$ . The largest destabilisation was the addition of one acetyl group into the 17nt oligonucleotide ( $\Delta\Delta G^\circ_{37} = +17.4 \text{ kJmol}^{-1}$ , entry 3), but subsequent additions of further 2'-acetyl groups do not destabilise the duplex to the same degree per acetyl group ( $\Delta\Delta G^\circ_{37} = +8.9$ - $+6.6 \text{ kJmol}^{-1}$ , entries 5-9). This suggests that the initial disruption of the ordered hydrogen network could be quite far reaching, but the addition of more acetyl groups could not bring about the same degree of destabilisation on an already disrupted hydrogen bonding network.

The decrease in duplex stability from inclusion of 2',5'-linkages in oligonucleotide sequences is well known.<sup>[270]</sup> In this work it was observed that the inclusion of one 2',5'-linkage caused a decrease in  $T_m$  ( $\Delta T_m = -6.9 \text{ }^\circ\text{C}$ ) and an increase in  $\Delta G^\circ_{37}$  ( $\Delta\Delta G^\circ_{37} = +19.4 \text{ kJmol}^{-1}$ , entries 1 and 2). The sugar puckers of A-form duplex RNA exist as the C3'-*endo*; this north conformation leads to a compact/compressed duplex with an axial phosphate-phosphate (P...P) distance of 5.9 Å. To accommodate a 2',5'-linkage there is a strong preference for a C2'-*endo* sugar pucker presumably because this also gives a P...P distance of 5.9 Å and so allows for formation of the A-form duplex without too much deformation of the duplex (Figure 88).<sup>[271]</sup> The decrease in duplex stability was thought to be due to the C3'-*endo*/C2'-*endo* equilibrium where a significant C3'-*endo* population still existed due to stabilisation by a  $\sigma_{\text{C-H}3'} \rightarrow \sigma^*_{\text{C-O}2'}$  orbital overlap from the presence of the 2'-phosphate (Figure 89a).

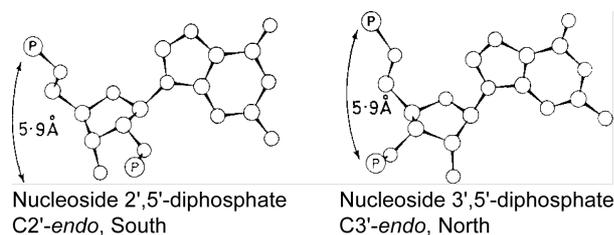


Figure 88. Duplex axial phosphate-phosphate distances and sugar puckers that allow inclusion of a 2',5'-linkage in A-form RNA duplex. Reprinted from reference<sup>[271]</sup> with permission from the copyright holder, Elsevier Limited.

Acetylation of the 3'-hydroxyl at the 2',5'-linkage was thought to have stabilised the preferred C2'-*endo* sugar pucker by facilitating a  $\sigma_{\text{C-H}2'} \rightarrow \sigma^*_{\text{C-O}3'}$  orbital overlap (Figure 89b). The UV melting studies showed an increase in  $T_m$  ( $\Delta T_m = +4.2 \text{ }^\circ\text{C}$ ) and a more

favourable  $\Delta G^\circ_{37}$  ( $\Delta\Delta G^\circ_{37} = -14.5 \text{ kJmol}^{-1}$ , entries 2 and 4) indicating a more stable duplex. In this case, stabilisation of the C2'-endo pucker reduces the pseudorotation of the furanose ring and in effect the south sugar conformation is frozen leading to the unfavourable decrease in  $\Delta S^\circ$ . The more equal P...P distance will lead to less deformation of the duplex and so maximise hydrogen bonding that would account for the favourable decrease in  $\Delta H^\circ$  ( $\Delta\Delta H^\circ = -95.6 \text{ kJmol}^{-1}$ ,  $\Delta\Delta S^\circ = -261.4 \text{ Jmol}^{-1}\text{K}^{-1}$ ). Use of nucleoside analogues to freeze the south-type sugar of 2'-5'-linked oligonucleotides is used by Erande *et al.* to improve RNA duplex formation for use in possible siRNA and miRNA applications.<sup>[272]</sup>

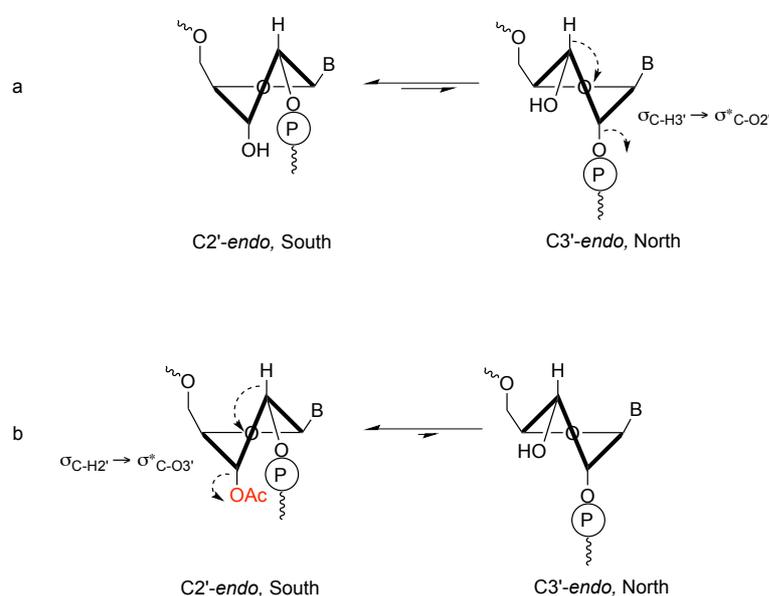


Figure 89. a) The C2'-endo sugar pucker at a 2',5'-linkage site is preferred in A-form duplex but pseudorotation to C3'-endo may be the major cause of  $T_m$  decrease. b) An increase in  $T_m$  is observed on acetylation of the 3'-hydroxyl where the electronegative acetyl group is thought to have further stabilised the preferred C2'-endo sugar pucker.

### 3.2.1. Consequences of acetyl-RNA for the non-enzymatic replication of RNA

One of the major difficulties for the replication of RNA is that once the length of the oligonucleotides reaches >30 base-pairs the  $T_m$  of the oligonucleotides become very high (close to 100 °C).<sup>[273]</sup> This means that under conditions of non-enzymatic replication of RNA a template and product strand would be very difficult to separate leading to product inhibition. Even the smallest active ribozymes are still approximately 50nt long and so it is conceivable that it would be difficult to replicate such sequences

non-enzymatically as native RNA.<sup>[274]</sup> Suggestions to lower the  $T_m$  have included introduction of copying errors or to introduce 2',5'-linkages into predominantly 3',5'-linked oligonucleotides.<sup>[275]</sup> However, both these solutions have their caveats, the former would not be suitable for inheritance of important sequences or functional RNA, while the latter 2',5'-linkages are known to hydrolyse more rapidly than 3',5'-linkages when in the context of a duplex and could lead to premature chain cleavage and degradation of oligonucleotides.<sup>[276]</sup> Additionally, the acetylation-ligation chemistry has shown that 3'-phosphate oligomers are selectively ligated and hence would lead to fewer 2',5'-linkages.<sup>[146]</sup>

The decrease of  $T_m$  and duplex stability imparted by the acetyl groups has the distinct potential to allow non-enzymatic template-directed synthesis of much longer oligonucleotides than is possible with native RNA. For example, a 33nt minizyme used by Wochner *et al.* has a calculated  $T_m = 101.3$  °C (atdbio online calculator<sup>[277]</sup>, nearest-neighbour model, 1000  $\mu$ M NaCl, 2.5  $\mu$ M oligonucleotide), which is clearly very stable and would be near impossible to denature at prebiotically plausible conditions.<sup>[278]</sup> If this minizyme was, for example, replicated from 3'-phosphate 3-5nt oligomers<sup>[133]</sup> it could contain approximately 8 internucleotide 2'-*O*-acetyl groups. Utilising this acetylated minizyme as a template for further rounds of acetylation-ligation reactions of complementary 3-5nt oligomers would result in a complementary product itself containing approximately 8 internucleotide 2'-*O*-acetyl groups. Extrapolating the  $T_m$  reduction per acetyl group for this acetylated-minizyme-product duplex would give it a predicted  $T_m$  of 51.7 °C. The significant decrease in  $T_m$  and duplex stability would enable strand separation possibly by solar heating of a small body of water during the day and reannealing of shortmers could occur gradually on cooling of the body of water during the night.

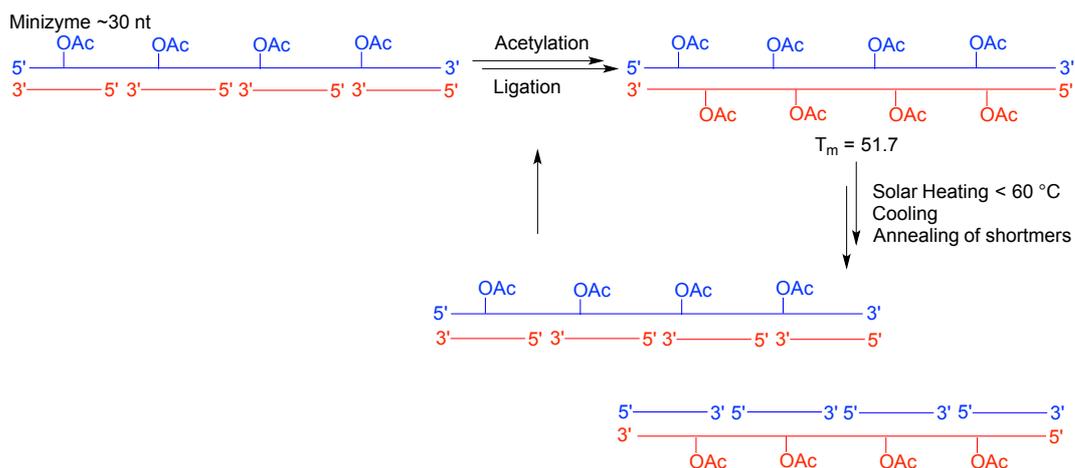


Figure 90. Replication of a long acetylated oligonucleotide with a suppressed  $T_m$  that allows duplex melting and annealing under mild conditions.

An appealing scenario for the emergence of catalytic RNA can be suggested that utilises mixtures of acetyl-RNA and native RNA (*i.e.* formed from slow deacetylation). Under the replication cycles described in Figure 90, the reduced stability of acetyl-RNA oligonucleotides would allow a significant population to undergo replication but conversely, the deacetylated RNA will remain annealed/folded (Figure 91). Utilising the pool of shortmers, continuous formation of ligated acetyl-RNA could be envisioned under acetylation and ligation conditions. Simultaneously, slow deacetylation by hydrolysis or ammonolysis of the longer oligonucleotides could reveal RNA slowly over time until catalytically superior RNA (ribozymes) emerged that could take over replication.<sup>[259]</sup> These ribozymes could then utilise the remaining RNA or other products in the local environment to carry out processes such as replication or peptide synthesis.

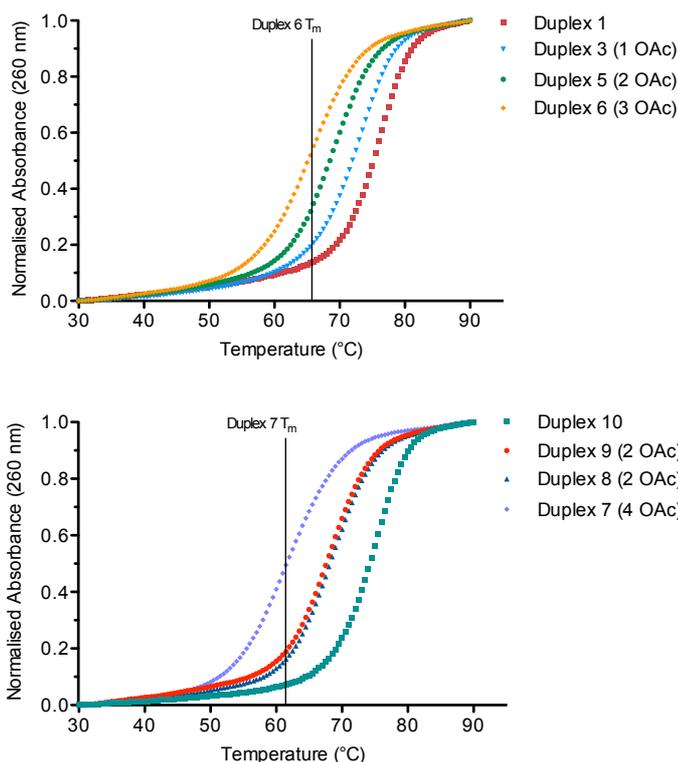
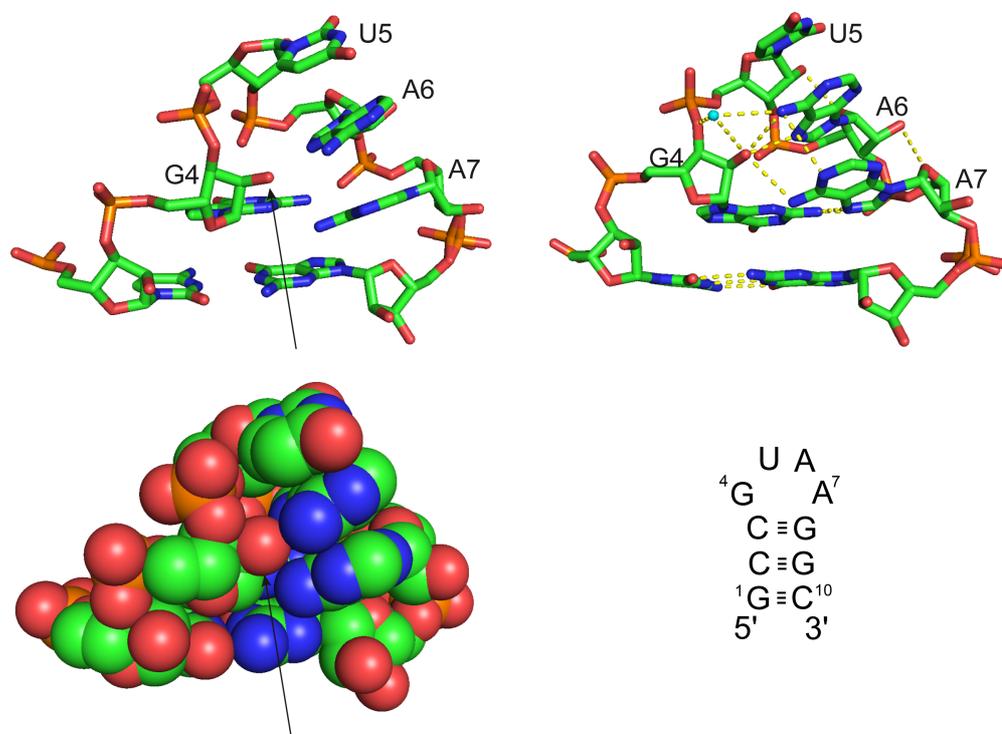


Figure 91. At the  $T_m$  (50 % dissociated) of the duplexes containing the highest number of acetyl groups the native RNA duplexes remain at or close to the fully associated state. Thus, at temperatures that denature a significant population of acetyl-RNA, native RNA is still stable, indicating the greater replication potential of acetyl-RNA.

### 3.3. Stability of an acetylated hairpin structure

To further explore the consequences of acetylated-RNA, a common secondary structure was sought that is implicated in the formation of tertiary structure. RNA hairpins loops are a type of structural motif that are some of the most common. In particular, tetranucleotide hairpin loops or tetraloops make up more than 55% of all loops.<sup>[279]</sup> These tetraloops consist of a Watson-Crick base paired A-form stem and a 4nt loop. This RNA structural motif is found throughout all large RNA structures and is implicated in the stabilisation of tertiary structure by utilising A-minor interactions to bind to a receptor.<sup>[257, 280-284]</sup> This so-called tetraloop-receptor interaction however involves a relatively large RNA that would have been difficult to synthesis by the 2'/3'-O-acetyl phosphoramidite method described previously. Therefore, as a compromise it was decided to study the effect of acetylation on the tetraloop itself. The tetraloops with the general sequence GNRA (where N = any base and R = purine base) were chosen as they make up 50% of the tetraloops, additionally their hairpin stabilities and structures

have been extensively studied.<sup>[279, 285, 286]</sup> The chosen tetraloop was a 10nt GUAA tetraloop of sequence 5' -GCCGUAAGGC-3' with a  $T_m$  value of 74.4 °C and a high-resolution crystal structure was available that allowed visualisation of the loop.<sup>[285, 287]</sup>



*Figure 92. GUAA Tetraloop. Structures show the tetraloop and the closing CG base pair of the stem. Examination of the crystal structure showed limited space at the 2'-hydroxyl (arrow) of G4, which also formed several hydrogen bonds (yellow dashed lines) to A6, A7 nucleobase amino groups and to an associated water molecule (cyan). Redrawn from PDB file 1MSY using MacPyMOL.<sup>[287]</sup>*

Inspection of the crystal structure of a GUAA tetraloop (Figure 92) and specifically its 2'-hydroxyls showed that the 2'-hydroxyl of G4 (arrow) of the loop formed several hydrogen bonds to N6, N7 of A6 and N6 of A7. In addition there was also an associated water molecule that formed hydrogen bonds to N6 of A6, O2' and O3' of G4. The remaining 2'-hydroxyls of the loop nucleosides extend out into solvent but it was not clear whether they are important for hairpin stability. The requirement of hydrogen bonding for tetraloop stability has been investigated and showed that removal of 2'-hydroxyl of G4 led to only a small decrease in stability and  $T_m$  ( $\Delta\Delta G^\circ_{70} = +1.26$  kJmol<sup>-1</sup>,  $\Delta T_m = -2.9$  °C). It was concluded that hydrogen bonds within a tetraloop contribute relatively little to its thermodynamic stability.<sup>[288]</sup> However, when a space-filling model of the tetraloop was inspected it was noted that the environment around

the 2'-hydroxyl of G4 was sterically demanding (Figure 92). It was hypothesised that addition of an acetyl group to this 2'-hydroxyl could prevent stable formation of the tetraloop and result in an unfolded hairpin that could served as a template for non-enzymatic replication.

To this end an acetylated GUAA tetraloop was synthesised, which incorporated a 2'-*O*-acetyl group at G4 of the hairpin loop. The tetraloop of sequence 5'-GCCG-3', 5' (2' OAc)-UAAGGC-3' (acetylated-tetraloop) was synthesised as described in Chapter 2.4. The stability of the acetylated-tetraloop (Table 7, entries 1) was studied by UV melting analysis. For a comparative study the melting curves of 2'-deoxy-G4 tetraloop (deoxy-tetraloop, Table 7, entries 2) and GUAA tetraloop without modification (natural tetraloop, Table 7, entries 3) were also measured.

The UV melting curves of the GUAA tetraloops (Table 7) were measured and the natural tetraloop was found to have a  $T_m$  that agreed with the literature value (avg.  $T_m = 74.0$  °C, entries 3.1-3.3. Lit.  $T_m = 74.4$  °C).<sup>[285]</sup> In agreement with previous studies on other GNRA tetraloops the loss of 2'-hydroxyl at G4 was found to have little effect, giving a small reduction in  $T_m$  (avg.  $T_m = 72.6$  °C, avg.  $\Delta T_m = -1.4$  °C, entries 2.1-2.3).<sup>[288]</sup> UV melting analysis of the acetylated-tetraloop showed reduction in  $T_m$  (avg.  $T_m = 69.0$  °C) with an average  $\Delta T_m = -5.0$  °C when compared to the natural-tetraloop. The reduction of  $T_m$  by the 2'-*O*-acetyl was greater than previous duplex studies. However, the relatively high  $T_m$  suggested that either the hairpin structure was still forming or that the unfolded hairpin was forming a 'self-complimentary' duplex (with two base pair mismatches). As hairpin formation is an intramolecular process the  $T_m$  is independent of concentration and UV melting analysis over a range of oligonucleotide concentrations gave the same  $T_m$  within error (entries 1.1-1.3,  $\pm 0.7$  °C) confirming the acetylated-tetraloop was still forming a hairpin. For comparison the same concentration independence was also observed for the deoxy-tetraloop (entries 2.1-2.3) and natural-tetraloop (entries 3.1-3.3). The calculated thermodynamic data indicated the natural tetraloop as the most stable tetraloop ( $\Delta G^\circ_{37} = -20.8$  kJmol<sup>-1</sup>). The loss of hydrogen bonding in the 2'-deoxy-G4 tetraloop gave a small decrease in hairpin stability ( $\Delta G^\circ_{37} = -19.2$  kJmol<sup>-1</sup>) with the greatest loss of stability observed in the acetylated-tetraloop ( $\Delta G^\circ_{37} = -15.5$  kJmol<sup>-1</sup>).

Entry	RNA Sequence (5' to 3')	Complement (5' to 3')	Total Conc. ( $\mu\text{M}$ )	Temp. Range ( $^{\circ}\text{C}$ )	$T_m$ ( $^{\circ}\text{C}$ )	$\Delta H^{\circ}$ ( $\text{kJmol}^{-1}$ )	$\Delta S^{\circ}$ ( $\text{Jmol}^{-1}\text{K}^{-1}$ )	$\Delta G_0^{\circ}$ ( $\text{kJmol}^{-1}$ )	$\Delta G_{37}^{\circ}$ ( $\text{kJmol}^{-1}$ )
1.1	GCCG <sup>2'</sup> OH <sup>c</sup> UAAAGGC	-	5	30-90	69.5	-212.5	-620.8	-42.9	-20.0
1.2	GCCG <sup>2'</sup> OH <sup>c</sup> UAAAGGC	-	10	30-90	68.9	-177.6	-519.6	-35.7	-16.5
1.3	GCCG <sup>2'</sup> OH <sup>c</sup> UAAAGGC	-	50	30-90	68.7	-168.3	-492.8	-33.7	-15.5
1.4	GCCG <sup>2'</sup> OH <sup>c</sup> UAAAGGC	GCCUUACGGC	10	30-95			nts		
1.5	GCCG <sup>2'</sup> OH <sup>c</sup> UAAAGGC	GCCUUACGGC	20	30-95	60.0	-198.7	-494.1	-63.7	-45.4
1.6	GCCG <sup>2'</sup> OH <sup>c</sup> UAAAGGC	GCCUUACGGC	100	30-95	70.1	-353.3	-940.5	-96.4	-61.6
2.1	GCCG <sup>2'</sup> H <sup>u</sup> UAAAGGC	-	5	30-95	72.2	-164.8	-477.6	-34.3	-16.7
2.2	GCCG <sup>2'</sup> H <sup>u</sup> UAAAGGC	-	10	30-95	72.5	-169.6	-491.2	-35.5	-17.3
2.3	GCCG <sup>2'</sup> H <sup>u</sup> UAAAGGC	-	50	30-95	73.0	-185.4	-535.9	-39.0	-19.2
2.4	GCCG <sup>2'</sup> H <sup>u</sup> UAAAGGC	GCCUUACGGC	10	30-95			nts		
2.5	GCCG <sup>2'</sup> H <sup>u</sup> UAAAGGC	GCCUUACGGC	20	30-95			nts		
2.6	GCCG <sup>2'</sup> H <sup>u</sup> UAAAGGC	GCCUUACGGC	100	30-95	63.8	-225.1	-579.8	-66.7	-45.3
3.1	GCCGUAAAGGC	-	5	30-95	73.1	-185.5	-535.8	-39.1	-19.3
3.2	GCCGUAAAGGC	-	10	30-95	74.6	-189.8	-546.0	-40.6	-20.4
3.3	GCCGUAAAGGC	-	50	30-95	74.4	-193.5	-557.0	-41.4	-20.8
3.4	GCCGUAAAGGC	GCCUUACGGC	10	30-95			nts		
3.5	GCCGUAAAGGC	GCCUUACGGC	20	30-95			nts		
3.6	GCCGUAAAGGC	GCCUUACGGC	100	30-95	66.2	-237.6	-612.0	-70.4	-47.8
4.1	-	GCCUUACGGC	5	30-90	66.7	-171.7	-505.6	-33.6	-14.9
4.2	-	GCCUUACGGC	10	30-90	68.2	-187.1	-548.3	-37.3	-17.0
4.3	-	GCCUUACGGC	50	30-90	68.6	-193.0	-565.5	-38.6	-17.7

Table 7.  $T_m$  and thermodynamic parameters used to assess the effect of acetylation on the secondary structure stability of a *GUAA* tetraloop. Where a mixture of two oligonucleotides was used each was used in equal concentration. Each measurement used 10 mM  $\text{Na}_2\text{HPO}_4$ , 0.5 mM  $\text{Na}_2\text{EDTA}$  buffer (pH 7) and with 0.1 M NaCl. nts = non-two state behaviour. Superscript denotes any 2'-modification.

The 2'-deoxy-G4 tetraloop results in reduced hydrogen bonding and a small reduction in hydration resulting in a small decrease in enthalpy ( $\Delta\Delta H^\circ = +8.1 \text{ kJmol}^{-1}$ ) and a reduced entropic penalty ( $\Delta\Delta S^\circ = +21.1 \text{ Jmol}^{-1}\text{K}^{-1}$ ) that is attributed to the loss of hydrogen bonding around the G4 2'-hydroxyl. A larger destabilisation of the acetylated-tetraloop is observed through increased reduction in the enthalpy term ( $\Delta\Delta H^\circ = +25.2 \text{ kJmol}^{-1}$ ) and a reduced entropic penalty ( $\Delta\Delta S^\circ = +64.2 \text{ Jmol}^{-1}\text{K}^{-1}$ ). Acetylation of the 2'-hydroxyl at G4 has approximately three times the destabilising effect of loss of the 2-hydroxyl. Thus, in addition to loss of hydrogen bonding and hydration in the immediate vicinity of the 2'-hydroxyl it seemed that the acetyl group conferred increased destabilisation through some other effect. The increased destabilisation could be due to steric demand required to accommodate the acetyl group by a conformational shift in the loop. This may have led to changes in atom positions resulting in non-optimal hydrogen bonding and/or  $\pi$ -stacking. Whatever the mechanism of destabilisation of the tetraloop by the acetyl group the effect is not clear without a greater understanding of the hydration of the tetraloop and structural changes, which could be solved by high-resolution X-ray crystallography.

The UV melting curves of the three GUAA tetraloops was also examined when in the presence of a complementary oligonucleotide to assess whether the acetylated-tetraloop would be suitable as a template in non-enzymatic replication. Each GUAA tetraloop was melted in the presence of the sequence complement oligonucleotide in a 1:1 ratio over a 10-fold concentration range. Ideally these experiments are recommended to be carried out over a 100-fold concentration range but limited stock of acetylated-tetraloop excluded higher concentrations that could be used and sensitivity of the spectrometer prevented the use of lower concentrations.<sup>[262]</sup> Each of the GUAA tetraloops, except the acetylated-tetraloop at a total concentration of 20  $\mu\text{M}$ , showed similar behaviour at total concentrations of 10  $\mu\text{M}$  and 20  $\mu\text{M}$  where the UV melting curves did not show two-state behaviour (Figure 93a, entries 1.4, 2.4, 2.5, 3.4 and 3.5). The UV melting curves in general had a very short baseline of few degrees followed by a transition-like absorption increase, which continued over a very wide temperature range. In each case no clear upper baseline was observed. The clearest indicator of non-two-state was the non-linearity of the data in the van't Hoff plots (Figure 93b). The results from these experiments suggested that two successive or overlapping melting curves were being

observed; in other words the tetraloops were not solely forming duplexes with the complement at these concentrations. Consequently, the UV melting curves of the complement was also measured over a range of concentrations and was found to have a relatively high concentration independent  $T_m$  (avg.  $T_m = 67.8$  °C) and comparable thermodynamic parameters (Figure 93c). This indicated that the complement was also forming a tetraloop structure that likely hindered duplex formation. The complement was later found to be a known UNAC tetraloop (N = any nucleobase).<sup>[289]</sup>

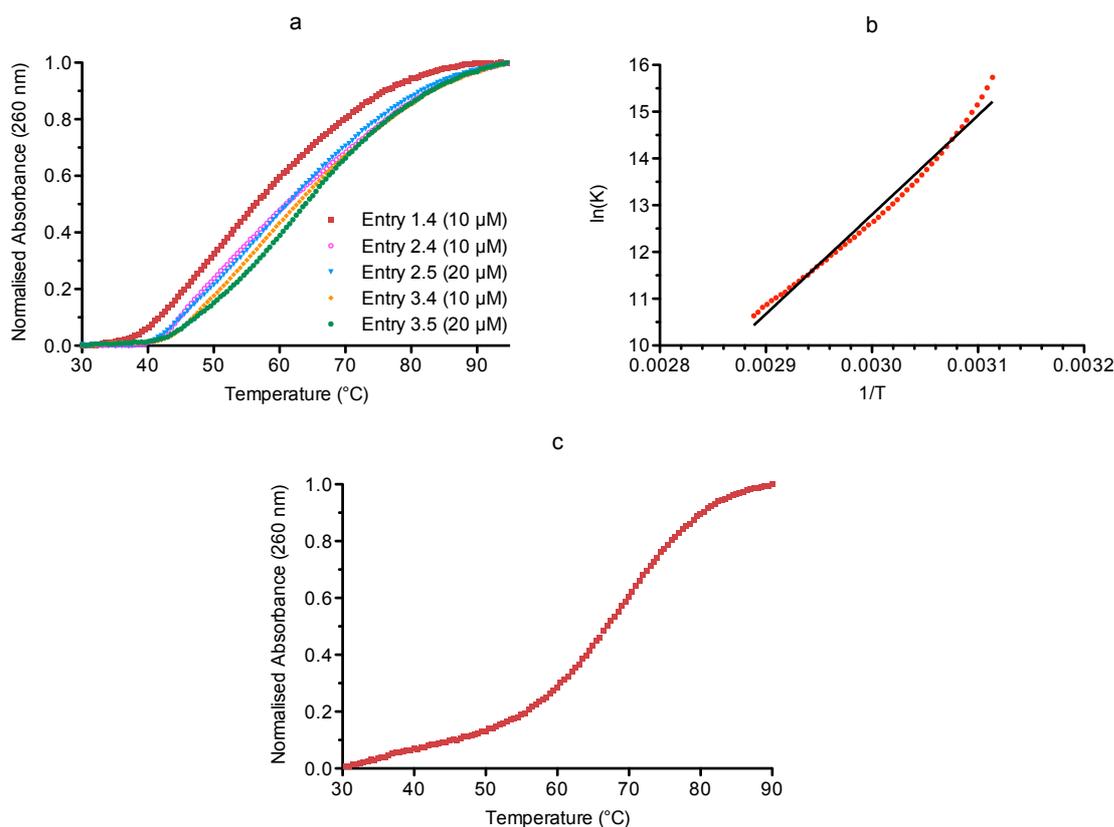


Figure 93. See Table 7 for entry references. a) Non-two-state UV melting curves of the GUA tetraloops with the complement. b) van't Hoff plot for entry 3.4, the non-linear nature of the data indicated a non-two-state behaviour. c) UV melting curve of the tetraloop complement (Entry 4.3).

The UV melting curve of the acetylated-tetraloop with the complement at 20 μM (entry 1.5) showed a very broad transition with non-distinct baselines. The melting curve plot nonetheless appeared to be a two-state transition with van't Hoff plot corroborating this observation. This observation suggested a weak duplex structure (Figure 94). The acetylated-tetraloop with the complement appeared to exhibit concentration-dependent  $T_m$  as at 100 μM an increase in  $T_m$  and  $\Delta G^\circ$  was observed ( $\Delta T_m = +10.1$  °C,  $\Delta \Delta G^\circ_{37} = -16.2$  kJmol<sup>-1</sup>, entries 1.5, 1.6). The deoxy- and natural-tetraloop with the complement

exhibited two-state behaviour at 100  $\mu\text{M}$ ; the reappearance of two-state melting curves at this higher concentration indicated duplex formation. Due to loss of the 2'-hydroxyl the deoxy-tetraloop formed a less stable duplex indicated by lower  $T_m$  and also less favourable  $\Delta G^\circ$  than the natural-tetraloop (entry 2.6 vs. 3.6).

Interestingly, the  $T_m$  and thermodynamic parameters suggested that the acetylated-tetraloop formed the most stable duplex and the deoxy-tetraloop being the weakest at 100  $\mu\text{M}$  total oligonucleotide concentration. Tetraloop destabilisation by the acetyl group at higher concentrations means that duplex structure is increasingly favoured over hairpin formation. Conversely, the stability of the hairpin in non-acetylated tetraloops competes more effectively with duplex formation (entries 1.6, 2.6 and 3.6). This result is consistent with the hypothesis that acetyl-RNA favours duplex structure and so suggests that it will enable more efficient ligation of oligonucleotides over native RNA.

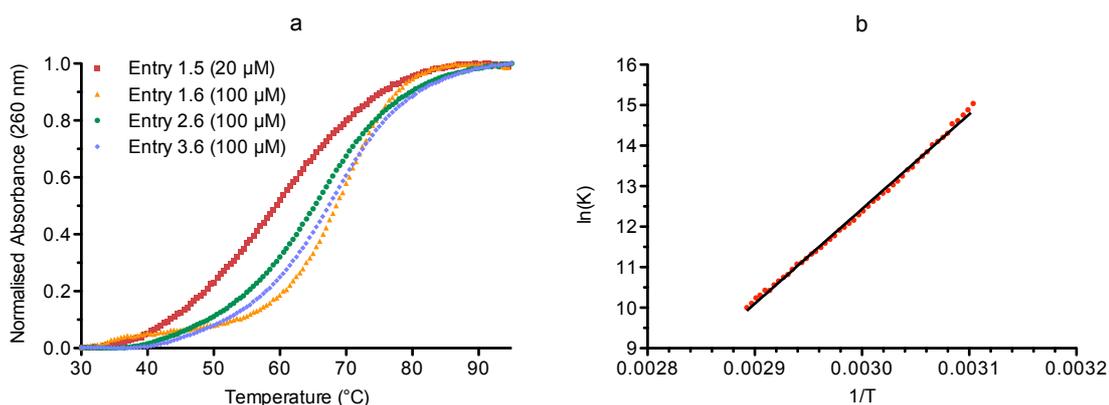


Figure 94. a) UV melting curves of GUAA tetraloops with their complementary strands at higher concentrations. b) van't Hoff plot of entry 1.5, data is very close to linear indicating two-state behaviour and duplex formation.

### 3.4. Future work and conclusions

In order to ration use of the acetylated-tetraloop small volume cuvettes (50  $\mu\text{L}$ ) were used but this had the disadvantage of low UV absorption and at low concentration instrument noise became significant. Additionally, the high stability of the parent native RNA tetraloop meant that in most cases high temperatures were required to obtain a clear upper baseline from the UV melting curves. Both these issues resulted in significant evaporation of the buffer, thus affecting the accuracy of the melting curves

especially at low concentrations. This was remedied to some extent by using relatively large volumes of mineral oil to reduce evaporation.

An alternative GNRA tetraloop should be synthesised where the native tetraloop also has a lower stability to ease melting curve measurements. A suggestion is the GAGA tetraloop of sequence 5'-GCCGAGAGGC-3', which has a lower  $T_m$  and thermodynamic parameters ( $T_m = 68.8$  °C,  $\Delta H^\circ = -148.5$  kJmol<sup>-1</sup>,  $\Delta S^\circ = -434.3$  Jmol<sup>-1</sup>K<sup>-1</sup>,  $\Delta G^\circ_{37} = -13.8$  kJmol<sup>-1</sup>; 1.0 M NaCl, 10 mM sodium cacodylate and 0.5 mM Na<sub>2</sub>EDTA, pH = 7).<sup>[285]</sup> In addition to synthesising the G4 acetylated GAGA tetraloop, a bis-acetylated GAGA tetraloop could also be made. An ideal position of secondary acetylation would be at the closing base pair. Considering a prebiotic synthesis from shorter oligomers indicates that the G8 position should be the second site of acetylation. The CG closing base pair of the hairpin loop confers a significant degree of stability to the tetraloop.<sup>[288, 290]</sup> Extra destabilisation at the CG closing base-pair of this doubly acetylated-GAGA tetraloop may mean that it is even more amenable to duplex formation and hence replication.

### 3.4.1. Disruption of A-minor and tertiary structure

Although the stability studies showed that acetyl-RNA favours duplex, the complement-free concentration independent behaviour of the acetyl-GUAA tetraloop indicated that secondary structure could still form. If secondary structure is present, tertiary structure could presumably form that could hinder duplex formation. Therefore, experiments to show that tertiary interactions can be reduced must also be carried out.

One strategy could be to use the tetraloop-receptor motif where interactions are predominately through A-minor like interactions. An example is the highly conserved '11nt motif' (R(11nt) receptor motif) that specifically recognises the GAAA tetraloop (Figure 95).<sup>[291]</sup> Blocking of the 2'-hydroxyls of either the tetraloop or receptor may weaken or completely prevent this long-range interaction. To supplement data obtained from UV melting curves additional techniques such as Differential Scanning Calorimetry (DSC)<sup>[292]</sup> or Fluorescence Resonance Energy Transfer (FRET) analysis<sup>[293]</sup> could be used to accurately calculate the thermodynamics of dissociation of the tetraloop-receptor motif and also support data from the UV melting curves.

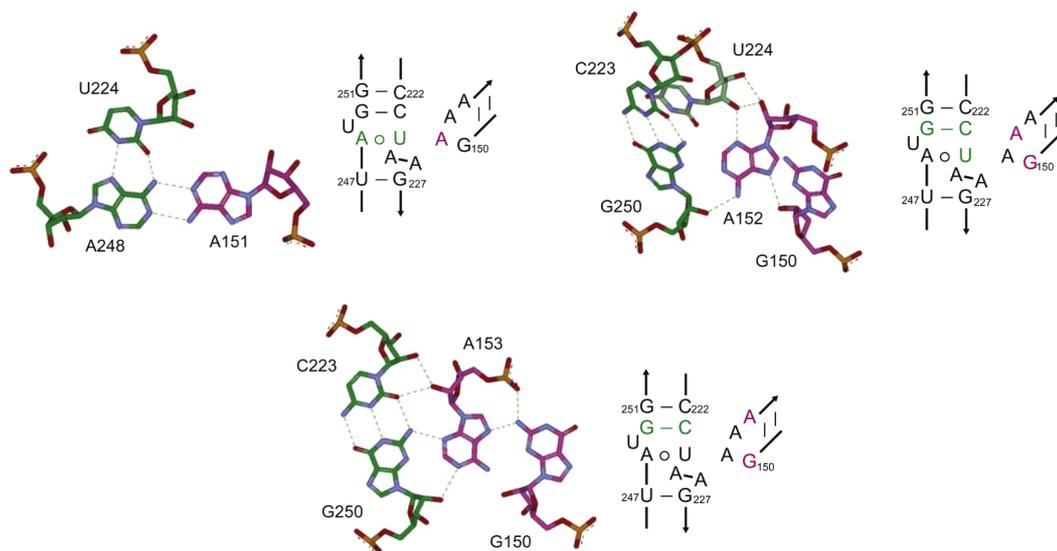


Figure 95. Intramolecular interactions involved in the recognition of a GAAA tetraloop by the R(11nt) receptor motif. Reprinted from reference <sup>[291]</sup> with permission from the copyright holder, Elsevier Limited.

The kink-turn (K-turn) secondary structure is another important structural motif that is important for RNA function such as translation and could be an ideal candidate for tertiary structure reduction by acetylation (Figure 96a).<sup>[294]</sup> A canonical K-turn is exemplified by a 3nt bulge (L1-L3) that induces a kink in RNA of approximately 60°. The bulge is flanked on its 5'-side by regular base pairing and on the 3'-side by *trans*-Hoogsteen sugar edge AG base pairs (Figure 96b). Interestingly, the importance of 2'-hydroxyls and hydrogen bonding for the stability of the K-turn has been investigated by Lilley *et al.*. Using a 25nt duplex with a K-turn at the centre, the folding of the RNA was monitored with FRET analysis. They found that removal of the 2'-hydroxyl at L1 resulted in complete ablation of folding of the K-turn (Figure 96c). Acetylation of this 2'-hydroxyl could potentially have a similar effect by blocking the hydrogen bonding ability.<sup>[295]</sup> The oligonucleotides in this case could be easily made as the length of the oligomers are not much longer than the 17nt acetyl-RNA previously synthesised. However, the fluorophore chemistry would require careful consideration to enable a compatibility with synthesis and deprotection.

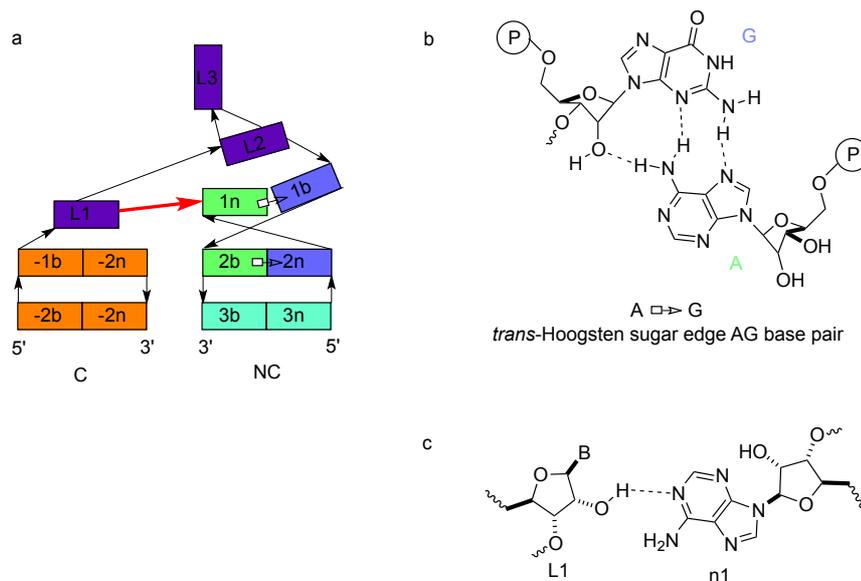


Figure 96. a) Nomenclature for the K-turn suggested by Lilley et al., the K-turn is represented using a colouring system based on the original diagram. The large red arrow indicates location of the important hydrogen bond. b) Diagram of the trans-Hoogsteen sugar edge AG base pair that is generally found between 1n-1b and 2b-2n. c) Representation of the important hydrogen bond between L1-1n. Diagrams adapted from original references.<sup>[295, 296]</sup>

### 3.4.2. Summary

In summary, UV melting studies have shown that internucleotide acetyl groups in RNA reduce the  $T_m$  consistently by approximately 3.1 °C per acetyl group in the context of duplex structure. Subsequent calculation of the thermodynamic parameters shows that the duplex stability of acetyl-RNA is also reduced. The minor groove of native RNA duplexes contain a highly ordered network of water molecules that bridge the two strands. The reduction in  $T_m$  and thermodynamic stability is attributed to blocking of 2'-hydroxyl hydrogen bonding leading to reduced hydration of the minor groove. Thus, the number of water molecules and hence hydrogen bonds are reduced and the two strands are less effectively bridged. The effect of acetylation was also investigated on the stability of a tetraloop hairpin. Acetylation of the G4 2'-hydroxyl, which is known to form hydrogen bonds within the loop, showed approximately 5 °C decrease in  $T_m$  over the natural tetraloop. Encouragingly, the acetylated-tetraloop was found to form duplex structure with its complement at lower concentrations than the equivalent deoxy- or natural-tetraloops. Additionally, at higher concentrations the acetylated-tetraloop was found to form more stable duplex structure. This supports the hypothesis that acetyl-

RNA favours duplex structure over other secondary structures such as hairpin loops. A geochemical scenario can be suggested where the concentrations of a pool(s) of acetyl-RNA could be in constant flux by action of environmental evaporation and precipitation; with these changing conditions acetyl-RNA should be able to form duplex structure over a much wider concentration range than native RNA. This could have provided an environmental pressure for the formation of a pre-RNA stage where “populations” of acetyl-RNA replicated. Subsequent slow removal of the acetyl groups by hydrolysis or ammonolysis would allow catalytic RNA to emerge over time.

## 4. Potentially prebiotic aminoacylation of RNA

### 4.1. Background

Aminoacylation is an important process for the correct translation of the genetic code and the process catalysed by aminoacyl-tRNA synthetases (ARSs) are considered the most important step.<sup>[151, 152]</sup> However, ARSs are thought to have replaced an earlier RNA-based system.<sup>[297]</sup> In current biology amino acids are first activated by reaction with adenosine-5'-triphosphate (ATP) within the ARS active site to give an aminoacyl-adenylate (aa-AMP) and pyrophosphate (PP<sub>i</sub>). The aa-AMP is now activated to nucleophilic attack by the 2'/3'-hydroxyl of the 3'-end of the corresponding tRNA with the hydrolysis of inorganic pyrophosphate providing the driving force (Figure 97a).<sup>[19]</sup> However, ATP activation of amino acids is thought to be prebiotically implausible due to thermodynamic reasons, that is the free energy contained with aa-AMP is ~37 kJmol<sup>-1</sup> higher than that available in ATP. Thus, ATP activation of amino acids is dependent on the existence of the ARS active site to allow favourable arrangement and stabilisation of the tetrahedral intermediate to provide a favourable equilibrium to the formation of aa-AMP. It has been proposed that amino acid activation was preceded by a different mechanism that was replaced once the potential energy in ATP could be exploited.<sup>[297]</sup> A notable alternative activation of amino acids is *via* *N*-carboxyanhydrides (NCA) that can be abiotically formed by reaction with the volcanic gas, carbonyl sulphide.<sup>[298, 299]</sup> Through the intermediary of NCAs it has been shown that aa-AMPs can be formed in yields of ~10% for several amino acids (Figure 97b).<sup>[300]</sup>

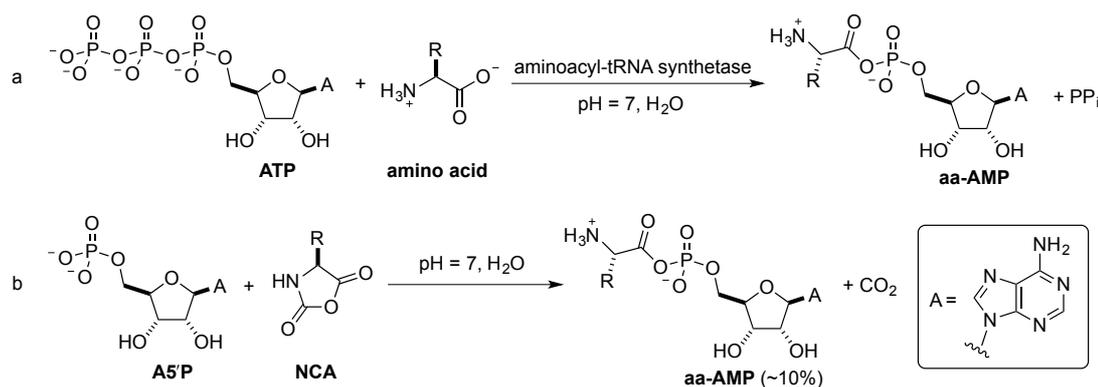


Figure 97. Formation of aminoacyl-adenylates (aa-AMP) a) in current biology by ARSs and b) reaction of adenosine-5'-phosphate A5'P with NCA.

The reaction of valyl-NCA **129** with nucleoside-2'/3'-phosphates N3'P (N = A/C) has also been investigated and found to give the corresponding 2'/3'-aminoacyl phosphates N3'P-2'val.<sup>[301]</sup> Aminoacylation to form N3'P-2'val is initiated upon transient formation of the 2'/3'-valyl phosphate, which undergoes a rapid intramolecular aminoacyl transfer to the 2'/3'-hydroxyl *via* a seven-membered transition state. Interestingly, aminoacylation of the nucleoside-3'-phosphates was more efficient than the phosphate regioisomer nucleoside-2'-phosphates. However, maximal aminoacylation yields with the nucleoside-3'-phosphates were only found to be 13.8% (Figure 98).

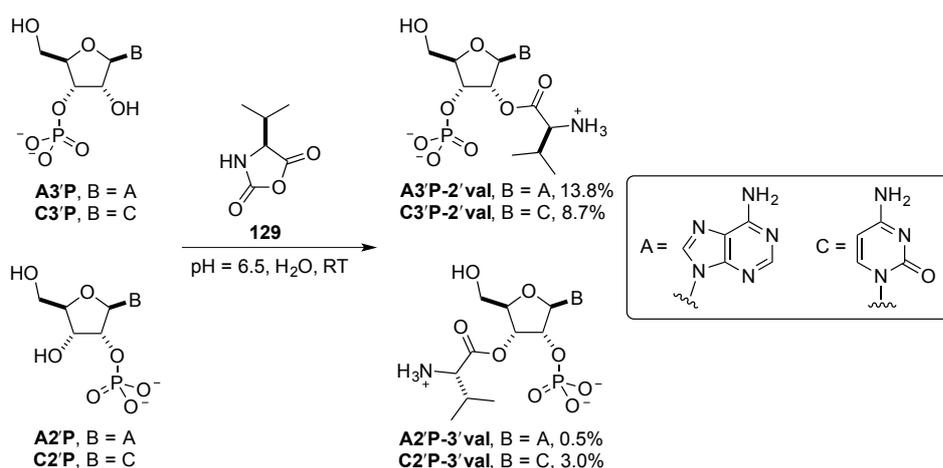


Figure 98. Aminoacylation of nucleoside-2'/3'-phosphates by valyl-NCA **129**.

The 2'-aminoacyl-3'-phosphates such as A3'P-2'val and C3'P-2'val are important structures invoked in Sutherland's theory of a linked prebiotic origin of RNA and coded

peptides.<sup>[138, 175]</sup> The aminoacylation of nucleoside-3'-phosphates is significant as it suggests that forming trimers such as **60** is possible. However, the NCA chemistry did not indicate any aminoacylations of oligonucleotides. The structure of **64** is also closely related to 2'-*O*-acetylated oligomers obtained through the acetylation-ligation chemistry (see Chapter 1.6.3).<sup>[146]</sup>

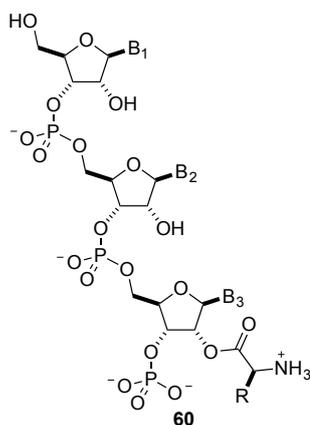


Figure 99. An amino acyl trimer candidate **60** that is related to recent studies to investigate an acetyl-mediated oligoribonucleotide ligation. *X* = phosphate activating chemistry.<sup>[138]</sup>

The acetylation-ligation chemistry showed that prebiotically available thioacetate **43** can be activated by various electrophiles to chemoselectively acetylate the 3'-terminal 2'-OH group of 3'-phosphate oligonucleotides. This allows the subsequent activation of the phosphate and then ligation to produce the native 3',5'-linked acetyl-RNA. For example, the activation of sodium thioacetate **43** with the electrophile cyanoacetylene **7** and subsequent acetylation gave the highest yields where the nucleobase was A, resulting in 2'-*O*-acetyladenosine-3'-phosphate **A3'P-2'OAc** in 52% yield (Figure 100, see Chapter 1.6.3 for further discussion).

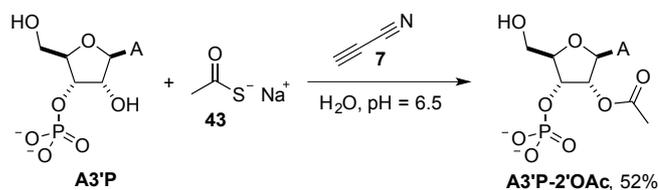


Figure 100. Chemoselective acetylation of adenosine-3'-phosphate **A3'P** by thioacetic acid **43** activated with the electrophile cyanoacetylene **7**.

It was decided to apply this chemistry to the aminoacylation of nucleoside-3'-phosphates **N3'P** using amino thioacids **130** and activation by various electrophiles (Figure 101) in the hope that an equivalent chemoselective aminoacylation could be found. Additionally, the chemistry held the potential to enable aminoacylation of oligonucleotides that was not demonstrated by NCA aminoacylation chemistries.<sup>[301]</sup>

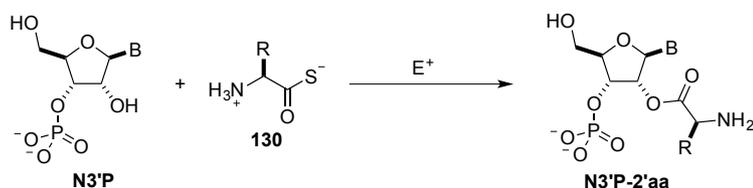


Figure 101. Aminoacylation of nucleoside-3-phosphates **N3'P** using amino thioacids **130** and electrophilic activators.

#### 4.1.1. The Iron-Sulphur World and the prebiotic plausibility of amino thioacids

In addition to thioacetate **43**, the Iron-Sulphur World, could provide possible chemistries to the prebiotic formation of amino thioacids **130**.<sup>[44]</sup> Primitive deep-sea hydrothermal vents are thought to have provided materials such as CO, H<sub>2</sub>S, Ni, Fe, Mn and Co. Under simulated hydrothermal vent conditions it has been shown that NiS catalysis utilising CO and H<sub>2</sub>S can form thioacetic acid **43** directly, in addition to other thiol derivatives and carboxylic acids.<sup>[147, 302]</sup> It is conceivable that similar chemistry could have also formed amino thioacids **130** but this has yet to be demonstrated. Despite this, amino thioacids and their derivatives such as amino thioesters are considered prebiotically plausible from the context of peptide formation.<sup>[149, 303, 304]</sup> Wieland showed that amino thioacids **130** in the presence of carbonate buffers and hence carbon dioxide (abundant in the primitive atmosphere) can form amino acid *N*-carboxyanhydrides **aa-NCA** that then undergo polymerisation with free amino acids or peptides (Figure 102).<sup>[304-307]</sup> Similarly, Wächterhäuser has shown that amino acids can be activated as NCAs with CO and H<sub>2</sub>S (or CH<sub>3</sub>SH) on the surfaces of (Ni,Fe)S minerals.<sup>[308]</sup> However, to obtain yields of up to 17% of dipeptides required relatively high pH values (pH = 8-9) and only traces of tripeptides were observed. The prebiotic plausibility is questionable as high pH values were required and low yields of peptides were obtained.

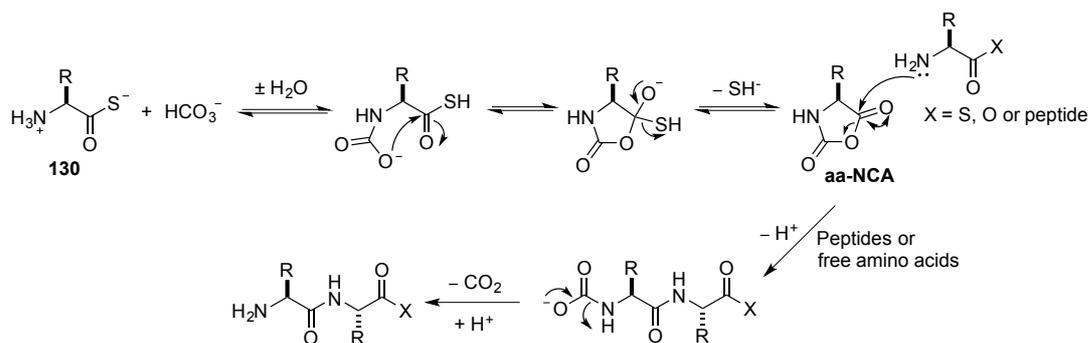


Figure 102. Polymerisation of the amino thioacids via the NCA cyclised by carbonate.

A theory related to the Iron-Sulphur World is De Duve's 'Thioester World' whereby the formation of a thioester from a carboxylic acid and thiol could be catalysed by the iron complexes invoked by Wächterhäuser.<sup>[309]</sup> The high-energy thioester bonds ( $\Delta G^{\circ} = -31.5 \text{ kJ mol}^{-1}$ ) were proposed to have represented an intermediate stage that could have provided the potential energy in a primitive metabolism in place of adenosine-5'-triphosphate (ATP,  $\Delta G^{\circ} = -32.2 \text{ kJ mol}^{-1}$ ).<sup>[310-312]</sup> This theory is supported by the fact that thioesters in the form of acetyl-coenzyme A are involved in biological processes such as fatty acid synthesis and the citric acid cycle.<sup>[313]</sup> The prebiotic considerations of amino thioesters has led Weber to develop a synthesis for amino thioesters from glycolaldehyde **22**, which is an important feedstock molecule used in the prebiotic synthesis of the activated pyrimidine nucleosides.<sup>[94, 314]</sup> The key step involves the condensation of NH<sub>3</sub> and a thiol with the  $\alpha$ -ketoaldehyde **131** to give the imine-hemithioacetal **132**, a subsequent redox rearrangement is thought to give the amino thioester **133**. As **133** is highly reactive it is not observed and subsequent hydrolysis or reaction with another amino acid gives amino acids or peptides respectively (Figure 103). Yields of amino acids gained by this synthesis are less than 0.5% indicating that formation of amino thioester is not high. Despite this bleak outlook the results of the acetylation by activated thioacetate was encouraging for the activation of amino thioacids **130** by similar electrophilic activators.

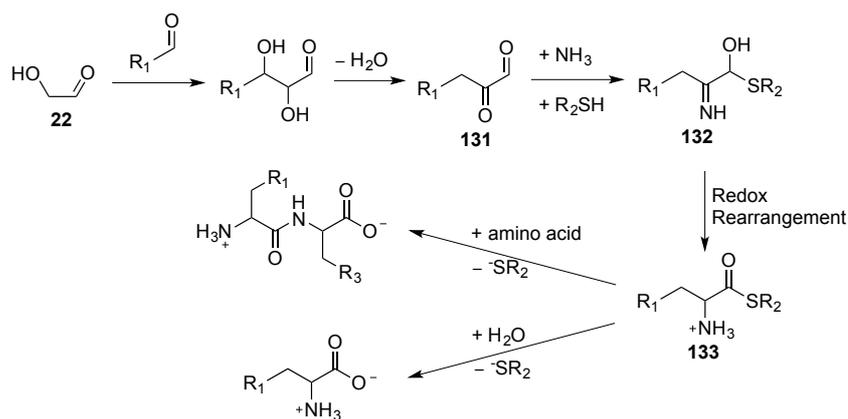


Figure 103. Prebiotic amino thioester synthesis.

## 4.2. Organic synthesis of amino thioacids

In Sutherland's theory of a linked origin of RNA and coded peptides valine **54** is identified as part of the early genetic code. To also allow comparison to previous aminoacylation studies<sup>[301]</sup> using valyl-NCA **129**, it was therefore decided to synthesise the amino thioacid, thiovaline **134**.

Commercially available Boc-Val-OH **135** was dissolved in anhydrous THF and the carboxylic acid was activated with isobutyl chloroformate (IBCF) at 0 °C. The formation of the intermediate carboxylic-carbonic anhydride renders the carboxylic carbonyl susceptible to nucleophilic attack. To this was added Li<sub>2</sub>S dissolved in anhydrous DMF that resulted in a green solution, and this reaction mixture stirred for 3 hours. Water was added to the reaction mixture, which was then adjusted to pH = 3. The Boc-protected thiovaline **136** was extracted into organic solvent. Evaporation afforded **136** cleanly, which was taken to the next step without further purification. Deprotection of the Boc-group was conducted by dissolving **136** in freshly distilled trifluoroacetic acid. Upon removal of excess TFA, thiovaline **134** was triturated with anhydrous diethyl ether, redissolved in water and lyophilised to removed traces of volatile organics (Figure 104). Thiovaline **134** was isolated as a pure solid without the need for any further purification.

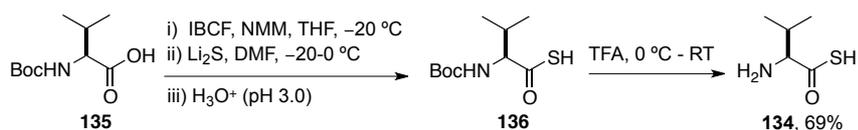


Figure 104. Synthesis of thiovaline **134**.

### 4.3. Prebiotically plausible aminoacylation of nucleoside phosphates with amino thioacids

#### 4.3.1. Aminoacylation with thiovaline **134** and cyanoacetylene **7**

Cyanoacetylene **7** is an important intermediate in the assembly of the pyrimidine ribonucleotides and has been demonstrated as a potent electrophile for the chemoselective acetylation of ribonucleotides with **43**.<sup>[94]</sup> Native amino acids exist as zwitterions at physiological pH and as **43** ( $\text{pK}_a = 3.33$ ) exists as an anion at pH = 6.5, it was expected that thiovaline **134** also existed as a zwitterion. Thus, reaction of **134** with **7** was expected to form a thioester **137** that would then react with the phosphate of a nucleoside-3'-phosphate before undergoing a *trans*-aminoacylation to the 2'-hydroxyl via a 7-membered transition state (Figure 105).

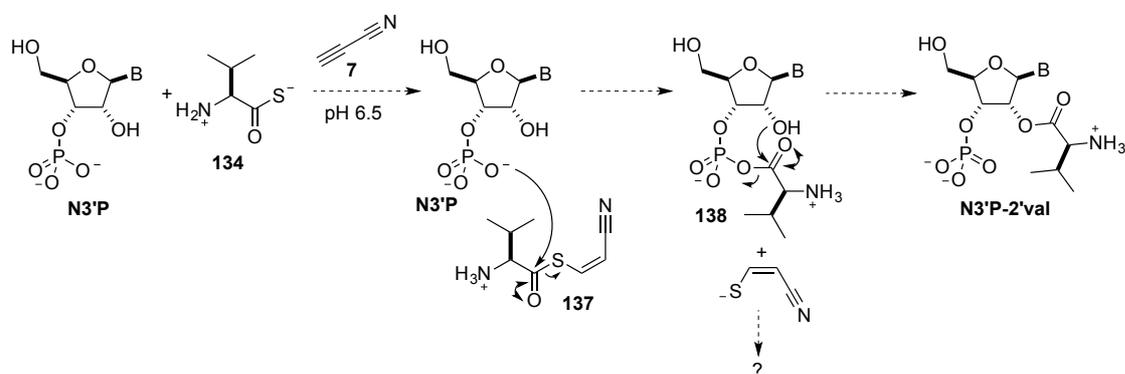


Figure 105. A potential prebiotic aminoacylation of a nucleoside-3'-phosphate **N3'P** by **134** activated with **7** via the intermediate **138** analogous to the aminoacylation by NCAs.<sup>[301]</sup>

Thus, cyanoacetylene **7** (200 mM) was added to a mixture of thiovaline **134** (100 mM) and adenosine-3'-phosphate **A3'P** (100 mM) at pH = 6.5 in D<sub>2</sub>O. Upon addition of **7** a white precipitate immediately formed and the pH rose rapidly to approximately 10. The pH was controlled by addition of 1 M DCl. The reaction resulted in formation of

adenosine-2'-valyl-3'-phosphate **A3'P-2'val** in 17% yield with 6% formation of the adenosine-2',3'-cyclic phosphate **A>P** after 1 hour (Figure 106a). A downfield shifted H-C(1') coupled to a highly downfield shifted H-C(2') (approximately 1 ppm) was observed and is characteristic of the aminoacylation at the 2'-OH (Figure 106c). The yield of **A3'P-2'val** was only a slight improvement over aminoacylation with the NCA chemistry where aminoacylation of **A3'P** was observed up to 14%.<sup>[301]</sup> In comparison, a control reaction with **43** (100 mM) showed 60% formation of only adenosine-2'-OAc-3'-phosphate **A3'P-2'OAc** and no **A>P** was observed (within the detection limit of NMR <1%). The control reaction ruled out cyanoacetylene **7** as a limiting factor, which is a gas and easily lost through diffusion suggesting other limiting factors (Figure 106b).

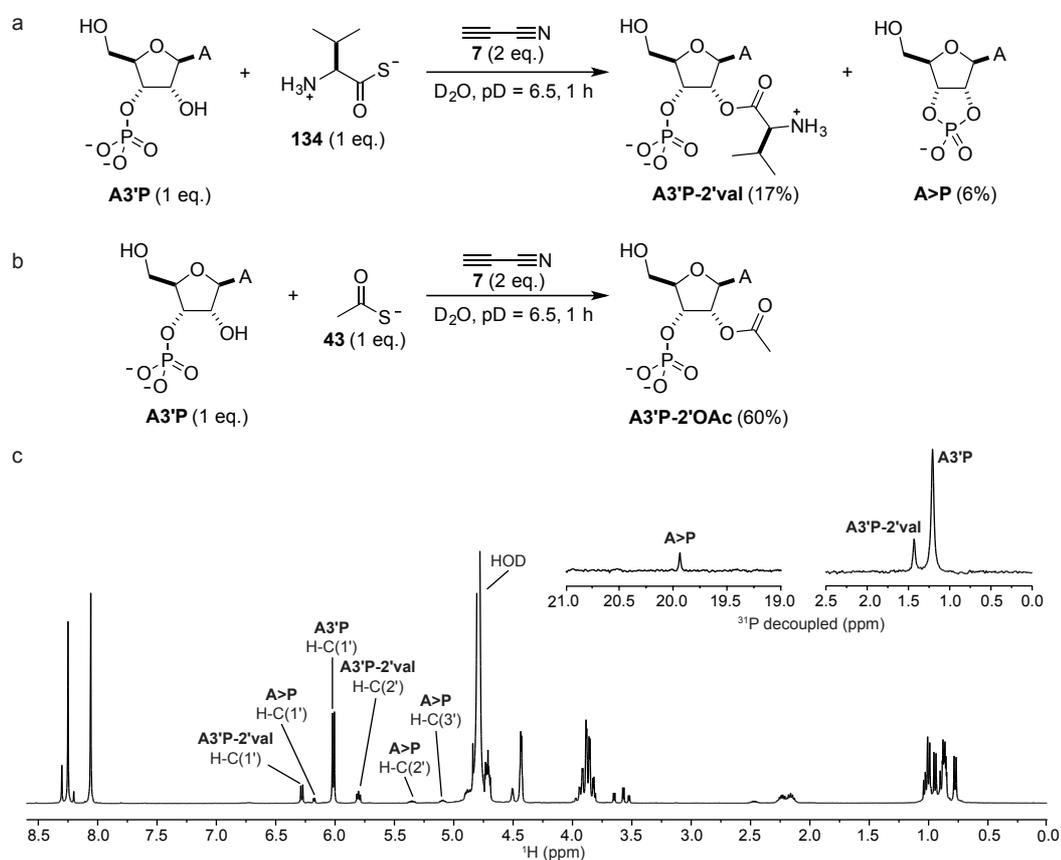


Figure 106. a) Aminoacylation reaction of **A3'P** (100 mM) with **134** (100 mM) and **7** (200 mM) in  $D_2O$  at  $pD = 6.5$ . b) Control acetylation reaction with **43** (100 mM) in place of **134**. c)  $^1H$ -NMR (400 MHz,  $D_2O$ ) and  $^{31}P$ -NMR (162 MHz,  $D_2O$ ) spectrum of the aminoacylation reaction described in a).

A repeat aminoacylation reaction was carried out in  $H_2O$  from which the white precipitate that formed was isolated. Analysis by  $^1H$  NMR spectroscopy and mass

spectrometry revealed the solid to be the  $\beta,\beta$ -bicyanovinyl-thioether **139** (Figure 107). The *cis*-geometry was assigned to the double bond according to the coupling constants between H-C(2) and H-C(3) ( $J = 10.4$  Hz).<sup>[96, 315]</sup> Although the valyl thioester **137** was not observed here it was transiently observed by NMR analysis during spiking experiments (see Figure 110).

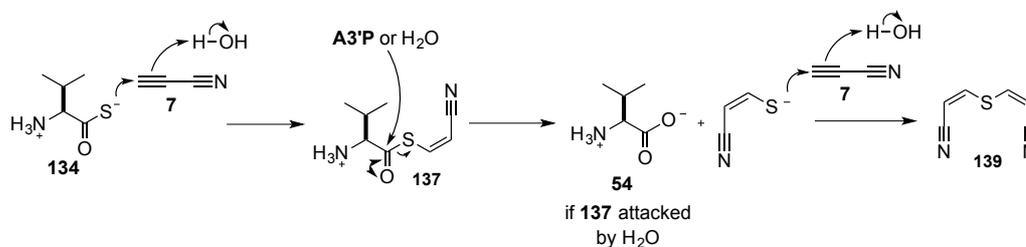


Figure 107. Formation of the solid precipitate **139** during the aminoacylation reaction with thiovaline **134** and cyanoacetylene **7**.

The low yield of **A3'P-2'val** was initially suspected to be as a consequence of the hydrolytic instability of the 2'-valyl ester due to its proximity to the 3'-phosphate. The mechanism for 2'-valyl ester hydrolysis was thought to occur in the reverse fashion to the aminoacylation (Figure 105), whereby attack of the 3'-phosphate dianion at the 2'-valyl carbonyl leads to intramolecular transfer to phosphate. Hydrolysis of the 3'-phosphate-valyl mixed anhydride **138** gives the apparent hydrolysis of the 2'-valyl group. At pH = 6.5 the phosphate of **A3'P** is mostly as the reactive dibasic form ( $pK_a = 6.16$ <sup>[150]</sup>, 70% ionised). It was thought that protonation of the phosphate to the less nucleophilic monobasic form would reduce hydrolysis of **A3'P-2'val**. Thus, the aminoacylation reaction was repeated at pH = 5.0, 6.0, and 6.5.

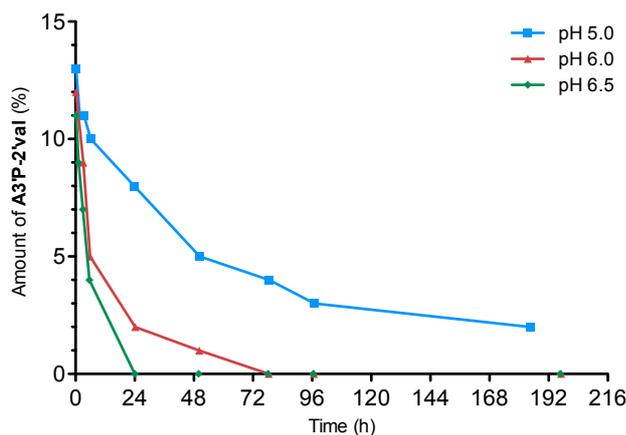


Figure 108. Time-course of the aminoacylation of **A3'P** (100 mM) with **134** (100 mM) and **9** (200 mM) in  $D_2O$  at three acidic pHs.

The initial yields **A3'P-2'val** were found to be comparable to each other and previous results with an average yield of 12% (Figure 108). The similarity in yields at the studied pHs is likely due to the rapid increase in pH on addition of **7**, which leads to a greater population of the dibasic phosphate until the pH can be controlled by addition of acid. The time course study revealed no further increase in the **A3'P-2'val** species suggesting that the aminoacylation reaction occurred immediately on addition of **7**. The clearest consequence of decreasing the pH was the reduced hydrolysis rate of the 2'-valyl group. The increased protonation of the phosphate to the less nucleophilic monobasic form reduces intramolecular attack of the 3'-phosphate at the carbonyl carbon of the 2'-valyl. At pH = 6.5 the **A3'P-2'val** had been completely hydrolysed back to **A3'P** after 24 hours, in contrast the adenosine-2'-OAc-3'-phosphate **A3'P-2'OAc** that was still present in greater than 50% yield.<sup>[146]</sup> This relative hydrolytic instability is attributed to the electron-withdrawing effect of the  $\alpha$ -amine group that renders the carbonyl carbon more  $\delta$ -positive and so more susceptible to nucleophilic attack.

Formation of **A>P** was also observed from the onset of the reaction (3-4%), in approximately equal amounts at each pH. Cyclisation of cytidine-3'-phosphate **C3'P** with cyanoacetylene **7** to cytidine-2'/3'-cyclic phosphate **C>P** (44% yield after 2 hours) has been previously studied and it is suggested that cyclisation proceeds *via* the phosphate activated adduct **140** (Figure 109a).<sup>[138]</sup> In the case of the aminoacylation reactions there also exists a second alternative cyclisation pathway. This pathway could proceed *via* the aminoacyl-mixed anhydride adduct **141** whereby the 2'-OH could attack

phosphate instead of carbon to give **A>P** (Figure 109b). The aminoacyl-mixed anhydride adduct **141** could be formed from either reaction of **A3'P** with **137** (“on-reaction”) or by hydrolysis of the 2'-valyl moiety (“off-reaction”). Cyclisation *via* this later pathway does not seem to occur as **A>P** is observed from the beginning of the reaction when the concentration of the **7** is highest. Additionally, formation of **A>P** could increase as the population of **A3'P-2'val** decreased, if hydrolysis of the 2'-valyl proceeded *via* the aminoacyl-mixed anhydride **141**, but no increase over time of **A>P** was observed.

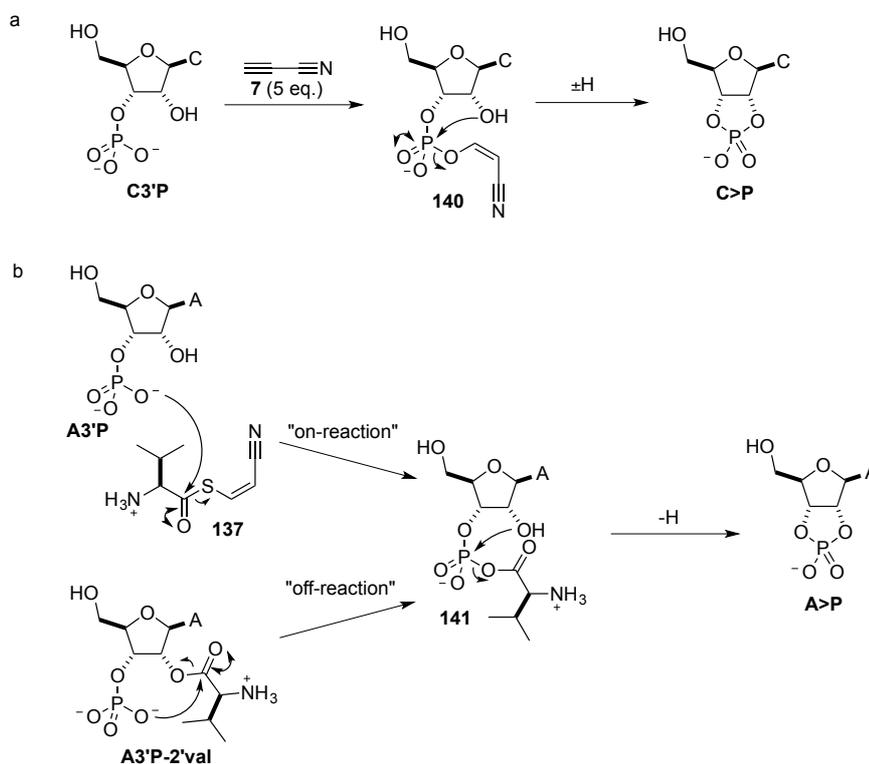
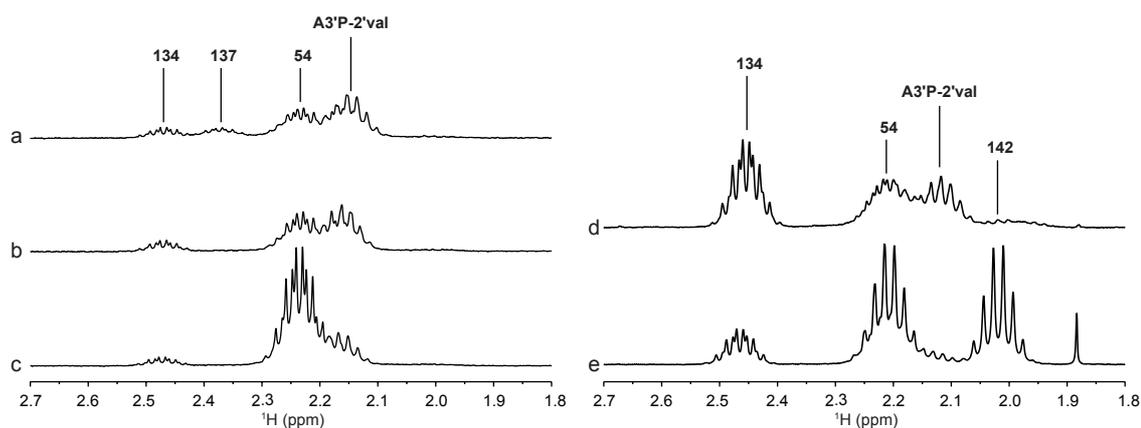


Figure 109. Cyclisation of nucleoside-3'-phosphates: a) cyclisation of **C3'P** by activation of the phosphate by **7**, b) alternative cyclisation pathway via the aminoacyl-mixed anhydride **141**.

Although the aminoacylation yields were comparable to previous work, they were lower than the acetylation reaction by activation of **43** with **7**; and attention thus turned to the fates of thiovaline **134**. However, due to overlap of the various valyl derivatives in the  $^1\text{H-NMR}$  spectrum it was difficult to obtain accurate yields (Figure 106c). The  $\beta$ -proton of the valyl species fell into a less complicated chemical shift region (1.9-2.6 ppm) so spiking experiments were conducted to identify some of the valyl products (Figure 110). The first spectrum was obtained approximately 15 minutes after an

aminoacylation reaction was initiated (Figure 110a). A multiplet at 2.37 ppm was observed that after incubation of the reaction mixture for 1 hour was no longer observed, and thus was assigned to the amino thioester **137** due to its transient nature (Figure 110b). The hydrolysis product valine **54** was also observed and so the remaining upfield multiplet at 2.16 ppm was assigned to **A3'P-2'val** (Figure 110c). In a separate experiment (Figure 110d) it was also noticed that traces of a further valyl derivative was present and it was thought to be the divaline peptide **142**, however due to overlapping peaks assignment is tentative (Figure 110e).



*Figure 110. Spiking experiments to identify the valyl derivatives formed during the aminoacylation reactions. a) Reaction 15 minutes after addition of cyanoacetylene **7**. b) Reaction 1 hour after addition of cyanoacetylene **7**. c) Spiking the reaction with commercially available valine **54**. d) Second reaction where divaline was thought to have be formed. e) Spiking with commercially available divaline **142**. (Note: intensity of spectrum d) is twice that of e) for clarity of the divaline peptide product **142**).*

The spiking experiments show that valine **54** is formed immediately after activation of thiovaline **134**, suggesting that significant amounts of **137** were consumed by hydrolysis rather than aminoacylating **A3'P**. This suggested that the intermediate is highly reactive such that the high effective concentration of water (~55 M) causes preferential hydrolysis and so limits the yield of **A3'P-2'val**.

A mixed acylation competition reaction between **134** and **43** was next conducted to investigate if the acyl thioester is a more efficient acylating agent than the amino thioester **137**. Thus, cyanoacetylene **7** (400 mM) was added to a mixture of thiovaline **134** (100 mM), thioacetate **43** (100 mM) and adenosine-3'-phosphate **A3'P** (100 mM) at pH 6.5 in D<sub>2</sub>O (Table 8). However, no selectivity was observed with the maximal 4% yield each of **A3'P-2'val** and **A3'P-2'OAc** (Table 8, entry 1). The yields of **A3'P-2'val**

and **A3'P-2'OAc** in the separate control reactions were as expected and were 21% and 56% respectively (Table 8, entries 2 and 3).

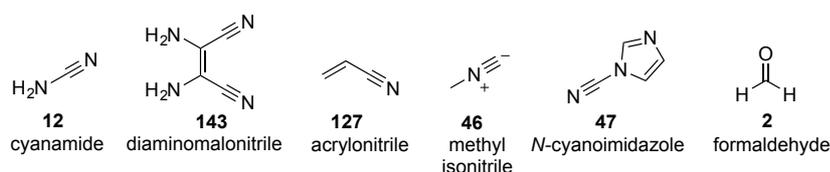
Entry	Nucleotide (100 $\mu$ M)	Val-SH ( $\mu$ M)	NaSAc ( $\mu$ M)	DCCCN ( $\mu$ M)	Products and residual starting materials (%)			
					A3'P	A3'P- 2'val	A3'P- 2'OAc	A>P
1	A3'P	100	100	400	87	4	4	< 1%
2	A3'P	100	-	200	75	21	-	5
3	A3'P	-	100	200	33	-	56	n.d.

*Table 8. Competition acylation reaction between thiovaline 134 and thioacetate 43 each entry was analysed by NMR spectroscopy after approximately 30 minutes. Entries 2 and 3 are control reaction for comparison.*

The poor acylation in the mixed reaction is likely due to the higher concentration of the nucleophilic acylating agents **134** and **43**. Such that the activated intermediates acetyl thioester and **140** preferentially reacted with unreacted **134** and **43**. On another note, the yield of **A3'P-2'val** in the control reaction was the highest observed. The variation in yields obtained so far was attributed to inconsistency in concentration of the cyanoacetylene **7** solutions. The inconsistency of **7** was because it was difficult to obtain accurate mass measurements of the condensed cyanoacetylene **7** gas before dissolution and the loss of the gas during thawing of the stored solution.

#### 4.3.2. Aminoacylation with thiovaline 134 and alternative electrophilic activators

Due to the variable yields of aminoacylation by activation with cyanoacetylene **7**, the search began for an alternative electrophilic activator in the hopes of finding an activator that would form a valyl thioester that would be more easily handled, less susceptible to hydrolysis and one that would lead to high aminoacylation yields.



*Figure 111. Selected alternative electrophiles.*

The electrophiles **12**, **143** and **2** (Figure 111) were chosen as they are components implicated in the prebiotic formation of various RNA components. Cyanamide **12** is

used in the prebiotically plausible synthesis of the activated pyrimidine nucleotides and key to the formation of 2-aminooxazole **21**.<sup>[94]</sup> Diaminomaleonitrile **143** has been found to an intermediate in the formation of adenine **10**.<sup>[316]</sup> Acrylonitrile **127** as the alkene derivative of **9** was hoped to lead to a less reactive amino thioester, so lead to less hydrolysis and hopefully preferential reaction with the phosphate. Methyl isonitrile **46** and *N*-cyanoimidazole **47** have been used to activate thioacetate **43**.<sup>[146]</sup>

When these electrophiles (200 mM) were incubated with thiovaline **134** (100 mM) overnight, **12** and **143** showed no reaction with **134**, in particular **143** did not react due to its insolubility. The reaction with electrophile **127** on the other hand only showed 15% conversion of **134** to two unassigned derivatives after 15 hours. However, this was decided to be too slow to effect any aminoacylation and so other electrophiles were investigated.

The electrophiles **46** and **47** have been shown to also bring about the selective acetylation of **A3'P** in comparable yields to activation with cyanoacetylene **7** and so their ability to furnish **A3'P-2'val** was investigated.<sup>[146]</sup> Furthermore, isonitriles have been used for the simultaneous cyclisation of nucleoside-2'/3'-phosphates and the formation of amino acid amides.<sup>[142]</sup> Aminoacylation of **A3'P** with **134** by activation with methyl isonitrile **46** was attempted. However, observation of the reaction over several days revealed no aminoacylation. Attention turned to *N*-cyanoimidazole **47** that is generally used as efficient water-soluble reagent for the formation of phosphodiester bonds especially for the ligation of DNA.<sup>[317-320]</sup> Thus, *N*-cyanoimidazole **47** (200 mM) was added to a mixture of thiovaline **134** (200 mM) and adenosine-3'-phosphate **A3'P** (100 mM) at pH = 6.5 in D<sub>2</sub>O. After 4 hours a maximal 12% yield of **A3'P-2'val** was observed with formation of 31% of **A>P**. The higher yield of the cyclic phosphate is expected considering the more common use of **47**. However, the yield of **A3'P-2'val** did not exceed those obtained by activation of **134** by **7** (Figure 112).

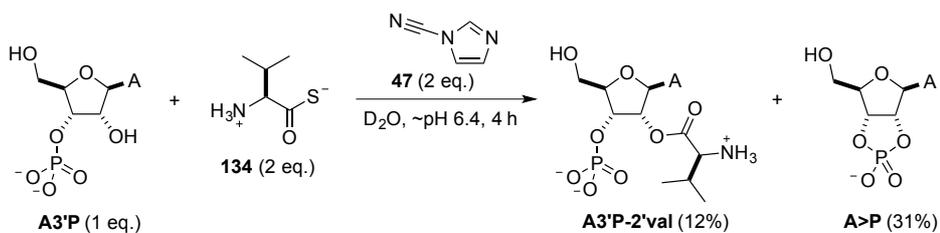


Figure 112. Aminoacylation reaction of **A3'P** (100 mM) with **134** (200 mM) and **47** (200 mM) in D<sub>2</sub>O at pH = 6.4.

Finally, formaldehyde **2** was investigated for the activation of thiovaline **137**. Initially, **2** (200 mM) was added to a solution of **134** (100 mM) at pH = 6.5 in D<sub>2</sub>O. This reaction resulted in precipitation of a white solid and <sup>1</sup>H NMR analysis of the resultant reaction mixture revealed mainly starting material and small amount of a single product. The reaction mixture was lyophilised and the residue redissolved in DMSO-*d*<sub>6</sub>. <sup>1</sup>H NMR analysis showed the formation of a single product in approximately 91% yield that corresponded to the formaldehyde-induced dimer **144**. This was proposed to have been formed by the cyclisation of thiovaline **134** with **2** to form the thiazolidin-5-one **145**, which then undergoes dimerization with another molecule of **145** through a bridging methylene derived from a third formaldehyde to give **144** (Figure 113a). The structure of **144** was confirmed by X-ray crystallography (Figure 113b).

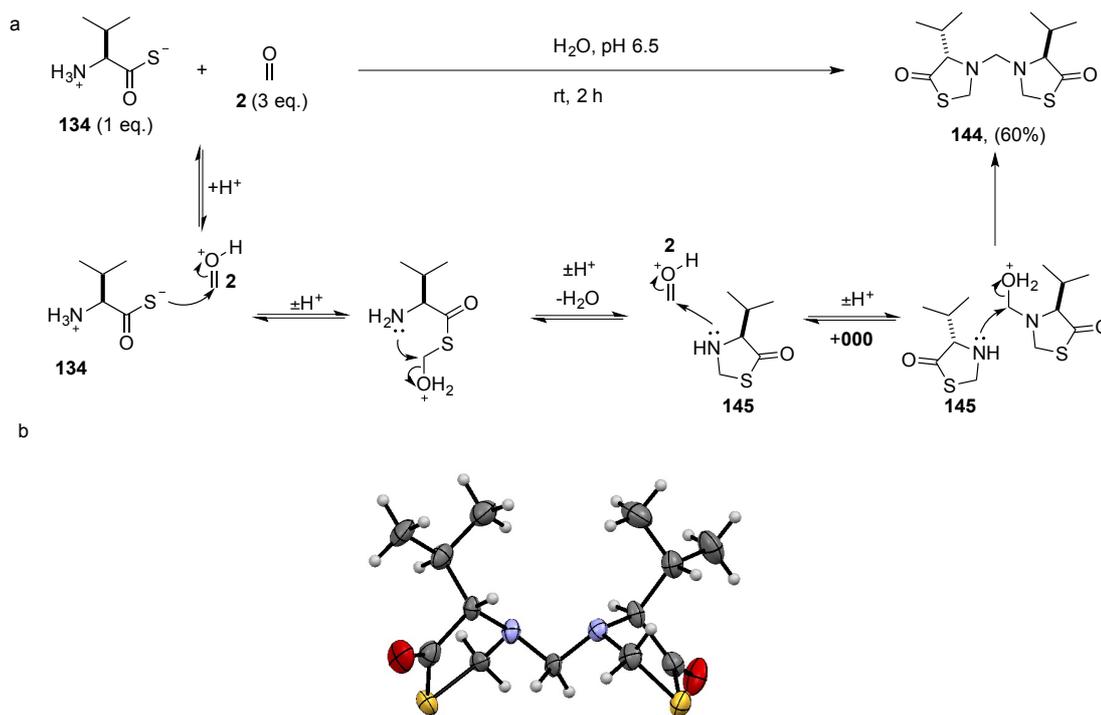


Figure 113. a) Formation of the formaldehyde-induced dimer **144**. b) X-ray crystal structure of **144**.

With the identity of the formaldehyde-induced dimer **144** deduced, it was decided to explore whether **144** could possibly be activated and act as an aminoacylating agent. It was foreseen that the insolubility of the thiovaline **134** derived dimer could hinder aminoacylation. Thus, repeating the aminoacylation reaction as previously described and using formaldehyde **2** as the electrophile resulted as expected in the precipitation of **144** with no aminoacylation observed.

To rule out solubility as a limiting factor it was decided to form a water-soluble formaldehyde-induced dimer by altering the amino acid side chain. Glutamic acid was chosen and so thioglutamic acid **146** was synthesised utilising the same procedures for the synthesis of thiovaline **134**. Boc-L-glutamic acid 5-*tert*-butyl ester **147** was treated with IBCF to activate the carboxylic acid. Li<sub>2</sub>S was then added to afford the protected-thioglutamic acid **148** and then treated with freshly distilled TFA. Thioglutamic acid **146** was again triturated with diethyl ether and stored as a 0.5 M solution at pH = 6.5 in either D<sub>2</sub>O or H<sub>2</sub>O (Figure 114).<sup>[306]</sup>

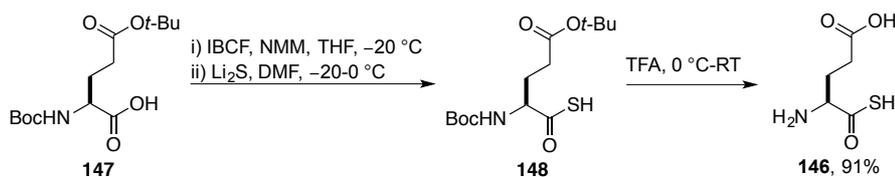


Figure 114. Synthesis of thioglutamic acid **146**.

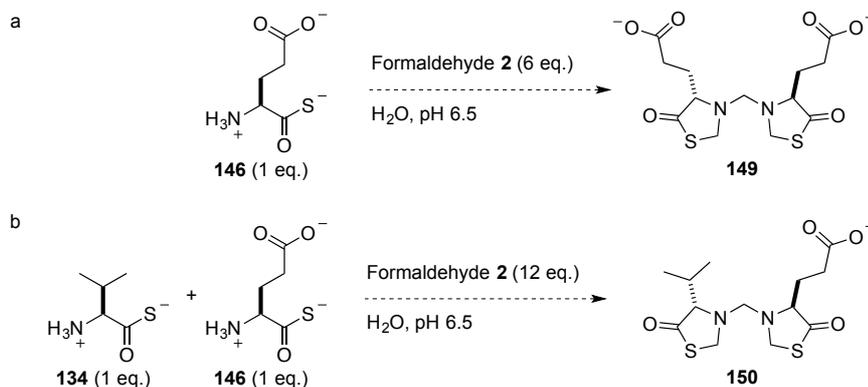


Figure 115. a) Envisioned homo-formaldehyde-induced dimerisation reaction. b) Envisioned hetero-formaldehyde-induced dimerisation reaction.

With thioglutamic acid **146** now to hand the reaction with **2** was investigated where it was hoped thioglutamic acid formaldehyde-induced dimers **149** and **150** could be formed (Figure 115). Formaldehyde **2** (600 mM) was added to a solution of thioglutamic acid **146** (100 mM) at pH = 6.5 in H<sub>2</sub>O and the mixture was allowed to stir for 3 hours upon which the solution was lyophilised and the residue redissolved in D<sub>2</sub>O for <sup>1</sup>H-NMR analysis. No peaks were observed that would correspond to a formaldehyde-induced Glu dimer **149**. However, it was found that the main by-product was glutamate **151** that was confirmed by mass spectrometry (Figure 116a).<sup>[321, 322]</sup> A mixed dimerization reaction (Figure 115b) was then conducted in which formaldehyde **2** (1200 mM) was added to a solution of thioglutamic acid **146** (100 mM) and thiovaline **134** (100 mM) at pH = 6.5 in H<sub>2</sub>O, and the mixture was allowed to stir for 3 hours. During the reaction a white solid precipitated that was removed by filtration and found to be **144** as a single product. The supernatant was lyophilised, redissolved in D<sub>2</sub>O and analysed by <sup>1</sup>H-NMR spectroscopy. This supernatant gave a mixture of predominantly glutamate **151** and valine **54** (Figure 116b). The formation of the natural amino acids indicated possible activation of the amino acids and their subsequent hydrolysis suggested aminoacylation could have been possible. A reaction with **A3'P** (100 mM),

**146** (100 mM) and **2** (300 mM) in D<sub>2</sub>O at pH = 6.5 was conducted but no aminoacylation was observed (Figure 116c).

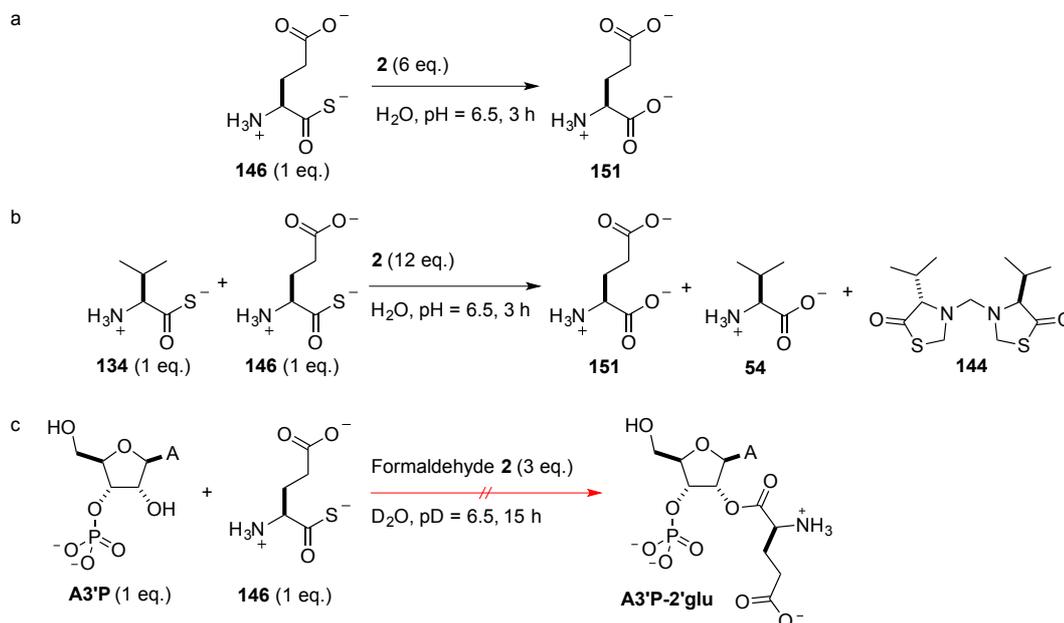


Figure 116. Reactions to form water-soluble formaldehyde-induced dimers.

The results above indicated that of the alternative activating electrophiles chosen, *N*-cyanoimidazole **47** was the most effective. However, the aminoacylation yield was not as high as those obtained by activation using **7**. Additionally, the greater cyclisation of the phosphate whilst using **47** suggests it is better suited as a phosphate activator. Utilising formaldehyde **2** to activate the amino thioacids was unsuccessful with respect to aminoacylation. In the case of **134** formation of formaldehyde-induced dimer **144** was very efficient but a similar water-soluble dimer compound could not be formed using thioglutamic acid **146**. Although, the aminoacylation reactions were not successful, the facile formation of 5-thiazolidinones maybe worthy of further study. Thiazolidinones are important structures and these compounds have been shown to have bactericidal, pesticidal, fungicidal, insecticidal properties amongst other biological activities.<sup>[323]</sup> The synthesis of 2- and 4-thiazolidinones is diverse and well established but reports of 5-thiazolidinones are less well known.<sup>[323, 324]</sup> The cyclisation of amino thioacids by formaldehyde **2** may represent a possible synthetic method to new 5-thiazolidinones, where variation of the aldehyde and amino acid side chain could result in new biologically active compounds (Figure 117).

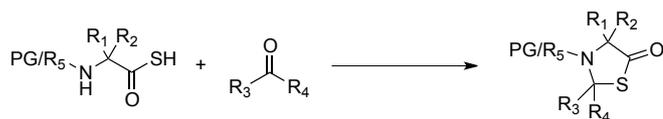


Figure 117. Potentially new synthetic route to 5-thiazolidinones.

### 4.3.3. Aminoacylation of oligonucleotides with thiovaline and cyanoacetylene

As the search for an alternative electrophile was unsuccessful it was decided to attempt aminoacylation of oligonucleotides with thiovaline **134** by activation with cyanoacetylene **7**. Thus, **7** (100 mM) was added to a mixture of thiovaline **134** (50 mM) and **CC3'P** (50 mM) at pH = 6.5 in D<sub>2</sub>O. The reaction gave a total 8% of the aminoacylated species **CC3'P-2'val**. The aminoacylation reaction appears to be highly selective for the 3'-terminal 2'-hydroxyl as there does not appear to be any other aminoacyl species. As aminoacylation is low, over aminoacylation is unlikely (Figure 118). The yield obtained in this reaction is comparable to the NCA aminoacylation of cytidine-3'-phosphate **C3'P**.<sup>[137]</sup>

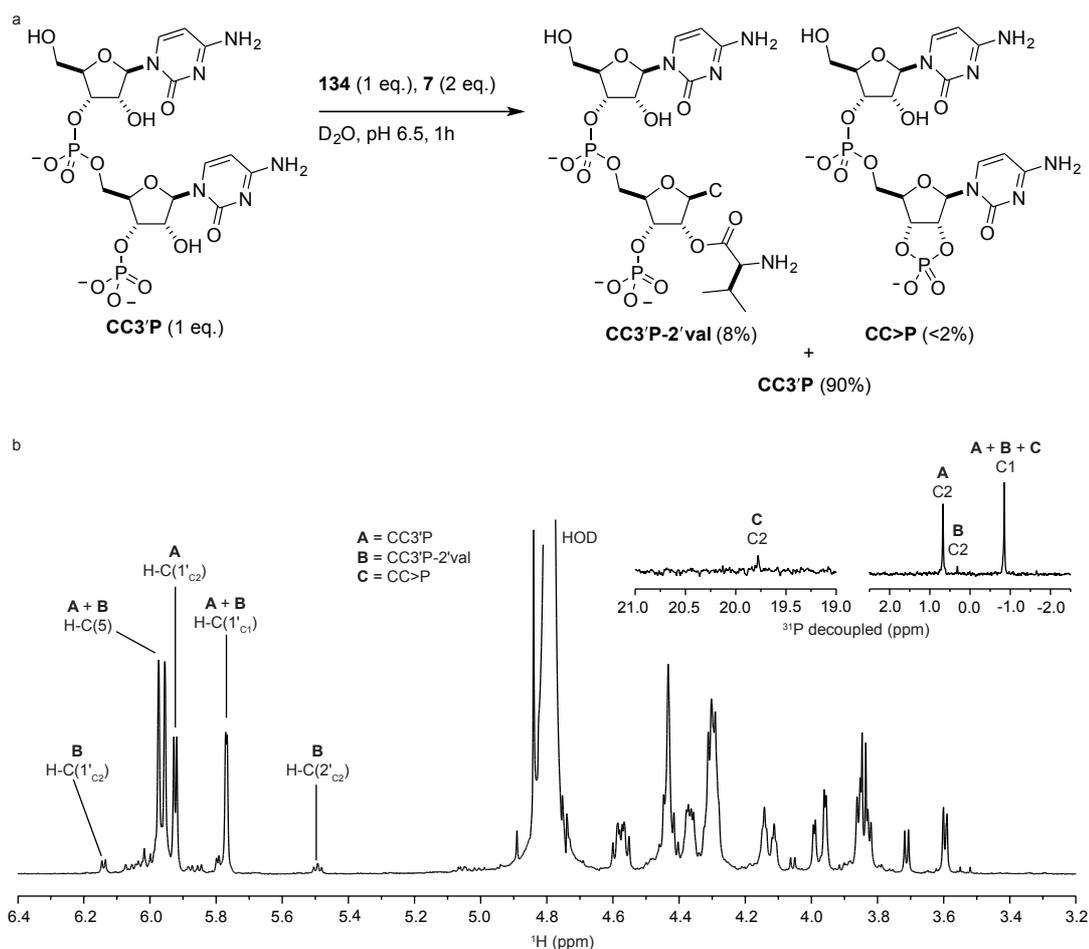


Figure 118. Aminoacylation reaction of **CC3'P** (50 mM) with **134** (50 mM) and **7** (100 mM) in D<sub>2</sub>O at pH = 6.5.

Next, the aminoacylation of a trimer **AGA3'P** was examined under the same reaction conditions as above. The <sup>1</sup>H-NMR spectroscopy revealed approximately 10% formation of the aminoacylated trimer **AGA3'P-2'val**. The yield is lower than **A3'P** monomer aminoacylation reactions but may be due to the lower concentration that the reaction was conducted (Figure 119). A trace degree of aminoacylation at the 2'-hydroxyls of the internucleotide linkages were observed in the <sup>31</sup>P NMR spectrum but these peaks were not clearly identified. Positively, this reaction showed that formation 2'-aminoacyl trimers is possible and lends support to a linked origin of RNA and coded peptides.

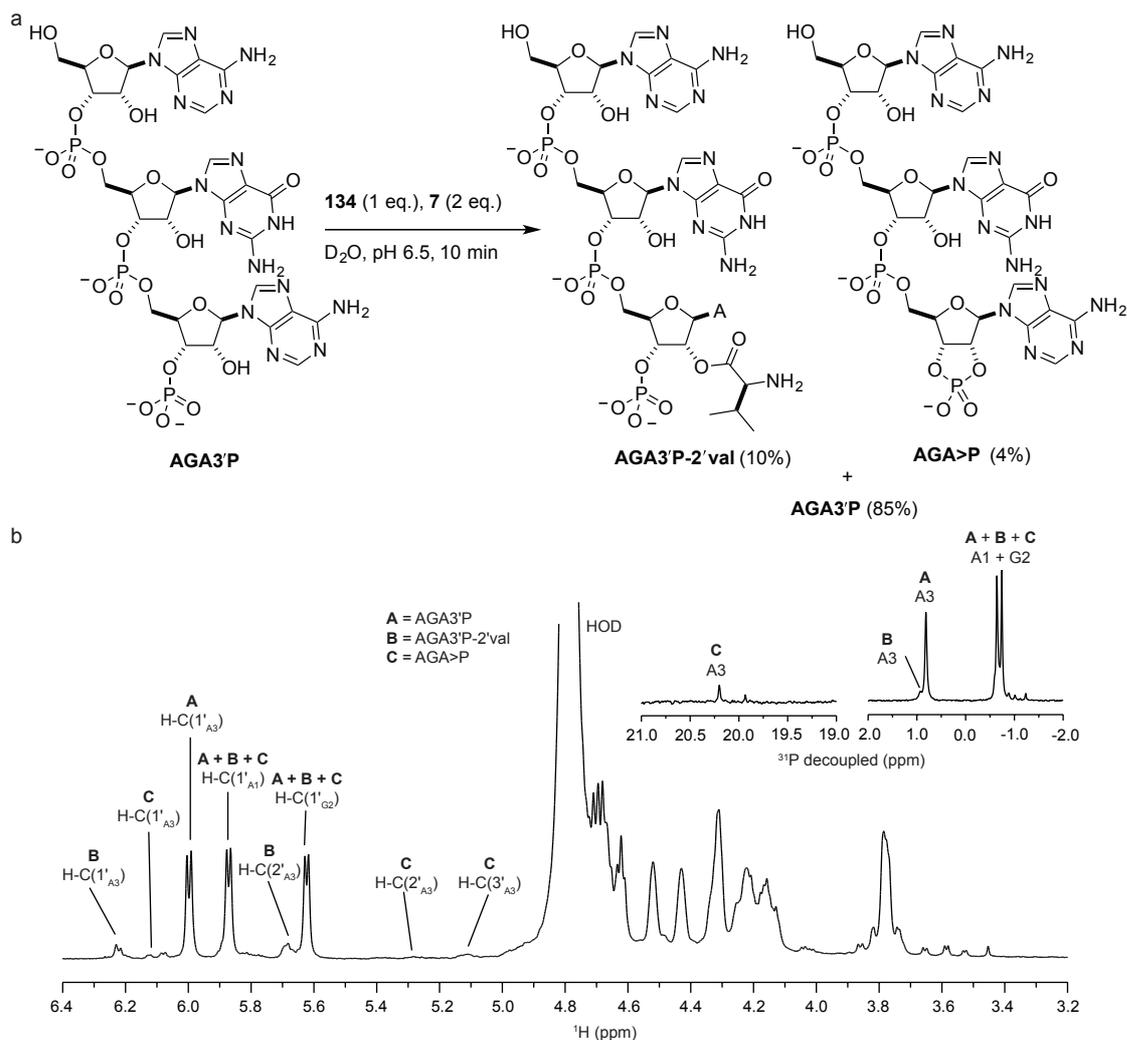


Figure 119. Aminoacylation reaction of AGA3'P (50 mM) with 134 (50 mM) and 7 (100 mM) in  $D_2O$  at pH = 6.5.

In conclusion, these results show that selective 2'-OH aminoacylation of a 3'-phosphate RNA oligonucleotide is possible. However, obtaining yields of these aminoacylated products on par with acetylation reactions, using cyanoacetylene 7 activation of amino thioacids, appears to be difficult. The limiting factors were the hydrolytic instability of both the amino thioester intermediate and the resulting aminoacyl species. Positively, the aminoacylation of the trimer gave support to Sutherland's theory of a linked prebiotic origin of RNA and coded peptides as it was demonstrated that formation of a key intermediate was possible.<sup>[138]</sup> On the other hand, the results suggest that ribozyme or enzyme-free aminoacylation by amino thioesters of a primitive tRNA may be intrinsically difficult due the high effective concentration of water. As in currently

biology, the chemistry may have required catalysis by an RNA ribozyme before proteinogenic enzymes could be formed and evolved.<sup>[325]</sup>

## 5. Conclusions

An orthogonal protecting group strategy for the solid-phase synthesis of 2'/3'-*O*-acetyl RNA oligonucleotides has been developed. Key to the synthesis of 2'- or 3'-acetylated phosphoramidites was the use of an orthoacetate that, through the intermediary of a 2',3'-cyclic orthoester, gives overall selective 2'- or 3'-acetylation. The ease of 2'/3'-migration of the acetyl group limited the synthesis of 2'-OAc phosphoramidites and their yields were maximised by use of a more acidic activator during the final phosphitylation step. Isolation of partially deprotected acetyl-RNA oligonucleotides from the solid-phase proved to be a major hurdle. After extensive investigation the problem was found to be due to solubility of the TBDMS protected oligonucleotides and utilising DMSO during the irradiation enabled isolation of the cleaved oligonucleotides. Overall, the synthetic strategy has enabled the synthesis of partially acetylated-RNA oligonucleotides with high purity and minimal loss of the acetyl groups. The synthetic strategy and methods that have been developed enabled the successful synthesis of partially 2'- or 3'-*O*-acetylated RNA oligonucleotides. However, if 2'- or 3'-*O*-acetyl-RNA were to be utilised for example, in antisense technology, several issues would first need to be addressed. Firstly, a method for 2'- or 3'-selective acetylation should be found. A selective route to 2' or 3'-*O*-TBDMS G<sup>npe</sup><sub>ceoc</sub> phosphoramidites should be sought or the 2'/3'-protecting group changed to enable greater chromatographic separation. These suggestions would completely eliminate the need for NP-HPLC separations for the purifications of all the precursors and phosphoramidites. An investigation into the stability and relative coupling rates of the phosphoramidites should be conducted and so that the feasibility of synthesising longer oligonucleotides can be assessed.

The  $T_m$  values and thermodynamic parameters of acetyl-RNA have revealed that 2'-*O*-acetyl groups destabilise the secondary structures of duplex and hairpins. The 3.1 °C reduction in  $T_m$  was found to be very consistent for each additional acetyl group within a duplex structure. Reduced stability was suggested to be an advantage for a primitive replication of acetyl-RNA by allowing the possibility for much longer oligoribonucleotides to be replicated, thus diminishing the product inhibition problem.<sup>[275]</sup> Importantly, when acetylated and non-acetylated tetraloops were melted in

presence of their complements at different oligomer concentrations, the acetylated-tetraloop was found to form duplex structure at lower concentrations. At higher concentrations where non-acetylated tetraloops also formed duplex, the acetylated-tetraloop duplex was more stable. The results indicated that acetyl-RNA favours duplex structure adding support to the hypothesis that it can be replicated in favour of native RNA. Further  $T_m$  and thermodynamic studies should be carried out with a less stable tetraloop to enable more results to be gathered that are more easily analysed. Using the tetraloops as templates, the ligation of complementary shortmers using the acetylation-ligation chemistry should be carried out to confirm acetyl-RNA's superior replication potential. RNA structures that utilise tertiary interactions should be synthesised as acetylated-RNA to confirm that interactions such as A-minor contacts can also be reduced.

An aminoacylation of nucleotides-3'-phosphates using an activated amino thioacid has been shown to selectively aminoacylate at the 2'-hydroxyl. The most effective activating electrophile was found to be cyanoacetylene **7** but yields of the 2'-aminoacyl species were low. In the hope of finding an activator that gave higher yields of the 2'-aminoacyl species a screen of prebiotically plausible electrophiles was carried out. Only *N*-cyanoimidazole **47** was found to form an activated thiovaline **134** able to aminoacylate but the yield obtained was not improved and additionally led to higher yields of the 2'/3'-cyclic phosphate. The cyanoacetylene **7** activated aminoacylation of oligomeric 3'-phosphates was also found to be selective for the terminal 2'-hydroxyl. In particular, aminoacylation of the trimer lends support to the theory of a linked prebiotic origin of RNA and coded peptides. In consideration of a possible aminoacylation and ligation of trimers, the relatively low yields of the 2'-aminoacyl species will prevent significant amounts of 3',5'-linked oligomeric material to be selectively synthesised. Alternative aminoacylating chemistries more selective for the 3'-phosphate should be sought. Aminoacylation possibly utilising an aminoacyl-ribozyme could also be investigated to enhance the yield of and stabilise the activated aminoacyl species.

## 6. Experimental

### 6.1. General procedures

Reagents were obtained from *Acros Organics*, *Alfa-Aesar*, *ChemGenes*, *Fisher Scientific*, *Glen Research*, *Link Technologies*, *New England Biolabs*, *Sigma-Aldrich*, *Santa Cruz Biotechnology*, *Thermo Scientific*, *Toronto Research Chemicals*, *Roche Diagnostics* and *VWR International* and were used without further purification. Anhydrous solvents were purchased from *Sigma-Aldrich*. Solvent pre-dried with molecular sieves were purchased from *Acros Organics*. All synthetic reactions were carried out in oven-dried glassware under an argon atmosphere unless stated otherwise. Ion-exchange resin was purchased as the Na<sup>+</sup> form (*Bio-rad* AG<sup>®</sup> 50W-X8 molecular biology grade, 200-400 mesh) and pre-washed with 1 M NaOH aq. for 1 hour then water until the filtrate was pH neutral. DCl was prepared by the addition of oxalyl chloride to D<sub>2</sub>O. NaOD was prepared by dissolving sodium metal in D<sub>2</sub>O, or by dilution of a concentrated sodium deuteroxide solution (40 wt.% in D<sub>2</sub>O) to the desired concentration. All water was purified using a *Millipore* Milli-Q Plus 185 purification system. pH and pD readings were carried out using a *Mettler Toledo* S20 SevenEasy™ pH meter with a *Thermo Scientific* glass pH electrode and pD measurements were recorded as pH values on the meter according to the following standard equation  $pD = pH + 0.41$ .<sup>[326]</sup>

Silica gel flash column chromatography was carried out using *Fluorochem* 60 Å (40-60 µm). For difficult separations and purification of the final amidites, *Silicycle* spherical silica gel 70 Å (40-75 µm) was used. TLC analysis was performed on *Merck* TLC Silica Gel 60 F<sub>254</sub> on aluminium plates. Visualisation was by UV (254 nm) irradiation, or by staining plates with the stains below followed by heating with a heat gun. Alkaline permanganate solution (KMnO<sub>4</sub> (3 g), and K<sub>2</sub>CO<sub>3</sub> (20 g) dissolved in NaOH aq. (5% w/v, 5 mL) made up to 300 mL with H<sub>2</sub>O). Vanilin (vanilin (15 g) dissolved in ethanol (250 mL) and concentrated sulphuric acid (2.5 mL) was added slowly). Phosphomolybdic acid (phosphomolybdic acid (12 g) dissolved in ethanol (250 mL)).

Normal phase-HPLC (NP-HPLC) was carried out on a *Varian* HPLC system equipped with *Varian* PrepStar Pumps modules, *Varian* ProStar UV module and a *Varian* 440-LC fraction collector. UV detection in all cases was at  $\lambda$  254 nm. For analytical separations A *YMC* YMC-Pack SIL-06 column (4.6  $\times$  250 mm) was used at a flow rate of 1 mLmin<sup>-1</sup> with a 20  $\mu$ L injection loop. For preparative separations A *YMC* YMC-Pack SIL-06 column (30  $\times$  250 mm) was used at a flow rate of 42 mLmin<sup>-1</sup> with a 1 mL injection loop. Samples were filtered through *Phenomenex* PHENX™ PTFE 0.45  $\mu$ m syringe tip filters prior to injection. Methods are described below, and solvent ratios are EtOAc:*n*-Hex:

Method A: 90:10, 0-30 min.

Method B: 90:10, 0-12 min; 90:10 to 100:0, 12-15 min; 100:0, 15-35 min.

Method C: 65:35, 1-16 min; 65:35 to 80:20, 16-18 min; 80:20, 18-30 min.

Method D: 50:50, 1-30 min.

Method E: 60:40, 1-12 min; 60:40 to 80:20, 12-15 min; 80:20, 15-35 min.

Method F: 15:85, 1-30 min.

Synthesised oligonucleotides were purified by strong anion exchange-HPLC (SAX-HPLC) using a *Varian* 940-LC liquid chromatograph with 445-LC scale-up module, a ProStar column valve module and a *Varian* 440-LC fraction collector. For preparative separations a *Dionex* DNA PAC™ PA-100 (22  $\times$  250 mm) was used at a flow rate of 15 mLmin<sup>-1</sup>. For analytical separations of oligonucleotides a *Dionex* DNA PAC™ PA-100 (4  $\times$  250 mm) was used at a flow rate of 1 mLmin<sup>-1</sup>.

Method G: detection  $\lambda$  280 nm, gradient of 0 to 0.4M NaCl in 10 mM Tris aqueous buffer (pH 8.0, 25% *v/v* formamide) over 40 min, then isocratic elution at 0.4M NaCl for 5 min.

Method H: detection  $\lambda$  260 nm, gradient of 0 to 1 M NaCl in 10 mM phosphate aqueous buffer (pH 11.5) over 30 min, then isocratic elution at 1 M NaCl for 5 min.

In all cases HPLC instrument control, data collection and analysis was performed using *Varian* Galaxie chromatography data system software.

Proton nuclear magnetic resonance ( $^1\text{H}$  NMR) spectra were recorded on a *Bruker* Avance 300 MHz spectrometer, a *Bruker* Avance III 400 MHz spectrometer or a *Bruker* Avance II+ 500 MHz spectrometer. Carbon nuclear magnetic resonance ( $^{13}\text{C}$  NMR) spectra were recorded on a *Bruker* Avance 300 MHz spectrometer at 75 MHz, a *Bruker* Avance III 400 MHz spectrometer at 101 MHz or a *Bruker* Avance II+ 500 MHz spectrometer at 125 MHz. Phosphorus nuclear magnetic resonance ( $^{31}\text{P}$  NMR) and Proton decoupled Phosphorous nuclear magnetic resonance ( $^{31}\text{P}$  NMR decoupled) were recorded on a *Bruker* Avance III spectrometer at 162 MHz. Collected spectra were referenced to TMS by way of residual non-deuterated solvent.  $\delta$  in ppm,  $J$  in Hz, assignments by COSY, HMBC, HMQC. Signal splittings are recorded as singlet (s), broad singlet (br. s), doublet (d), doublet of doublets (dd), double double doublet (ddd), triplet (t), doublet of triplets (dt), triplet of doublets (td), quartet (q), doublet of quartets (dq), triplet of quartets (tq), quintet (quin.), sextet (sex.), heptet (hept.), heptet of doublets (hept.d), doublet of heptet (dhept.) and multiplet (m). The notation (ABX) refers to a methylene spin system coupled to a unique adjacent proton.

Electrospray ionisation mass spectrometry (ESI MS) and high resolution mass spectrometry (ESI-HRMS); *Micromass* Platform II, *Waters* QTOF. ESI-HRMS; *Thermo Finnigan* MAT95XP, *Waters* LCT Premier or a *Thermo* Orbitrap instruments. MALDI-TOF spectra were obtained using an *Applied Biosystems* Voyager-DE Pro, using a matrix containing 3-hydroxypicolinic acid and diammonium hydrogen citrate (25 mg/mL and 35 mg/mL respectively) dissolved in acetonitrile in water (30% v/v). Typically 1-2  $\mu\text{L}$  of analyte solution was mixed with 8  $\mu\text{L}$  of matrix, 2  $\mu\text{L}$  of this solution was spotted in duplicate. Spectra were recorded in linear positive ionisation mode, using a minimum of 200 shots/spectrum and calibrating to internal or external synthetic RNA standards, average mass values are reported.

Infrared (IR) spectra of solid samples were recorded using a *Bruker* Equinox 55/*Bruker* FRA 106/5 with coherent 500 mW laser as Attenuated Total Reflectance (ATR) spectra with a 'golden gate' attachment and a resolution of  $2\text{ cm}^{-1}$ . Absorption maxima are quoted in wavenumbers ( $\text{cm}^{-1}$ ). Alternatively, IR spectra were recorded using a *Thermo* Nicolet iS5 with a iD5 ATR diamond attachment with resolution of  $4\text{ cm}^{-1}$ . Melting

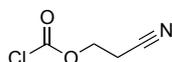
points (M.P.) were measured using a *Buchi* M-560 or a *Sanyo* Gallenkamp variable heater and values are uncorrected.

## 6.2. Experimental for Chapter 2<sup>e</sup>

### 6.2.1. Synthetic procedures for the synthesis of the phosphoramidites

#### 2-Cyanoethyl carbonochloridate 75

C<sub>4</sub>H<sub>4</sub>ClNO<sub>2</sub>; M<sub>r</sub> = 133.53

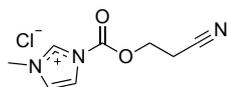


Triphosgene (5.94 g, 20.0 mmol) was dissolved in anhydrous THF (50 mL) and cooled to 0 °C. 3-Hydroxypropionitrile **77** (2.73 mL, 40.0 mmol) was diluted with anhydrous THF (17 mL) and was added dropwise to the solution of triphosgene over a 2 h period. The resultant mixture was warmed to RT and stirred overnight. Anhydrous pyridine (4.84 mL, 60.0 mmol) was diluted with anhydrous THF (5 mL) and added dropwise to the solution at 0 °C. The resultant mixture was warmed to RT and stirred for a further 1 h. The precipitate pyridinium hydrochloride salt, was filtered and the supernatant was evaporated under vacuum. The title compound was isolated as a pale yellow viscous oil in quantitative yield. The product was used immediately without further purification or stored at -30 °C under argon until required. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 4.52 (2H, t, *J* = 6.3 Hz, -CH<sub>2</sub>CH<sub>2</sub>CN), 2.84 (2H, t, *J* = 6.3 Hz, CH<sub>2</sub>CH<sub>2</sub>CN). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 151.7 (C=O), 115.3 (CN), 65.0 (CH<sub>2</sub>CH<sub>2</sub>CN), 17.8 (CH<sub>2</sub>CH<sub>2</sub>CN).

---

<sup>e</sup> Compounds not synthesised by the author are denoted with ‡ in the title and are either synthesised by Dr Colm D. Duffy or Dr Jianfeng Xu. NMR characterisation data of these compounds, in some cases, was obtained the author and is additionally denoted with *f*. Compounds are included for completeness.

## 1-[(2-Cyanoethoxy)carbonyl]-3-methyl-1*H*-imidazolium Chloride 76



$C_8H_{10}ClN_3O_2$ ;  $M_r = 215.64$

A solution of **75** (9.35g, 70.0 mmol) in anhydrous  $CH_2Cl_2$  (50 mL) was cooled to 0 °C, 1-methyl-1*H*-imidazole (5.55 mL, 70.0 mmol) diluted with anhydrous  $CH_2Cl_2$  (5 mL) was added dropwise. The mixture was stirred for 12 h and the resultant precipitate was filtered under a nitrogen atmosphere. The filtrate was washed with anhydrous  $CH_2Cl_2$  (3 × 20 mL) and dried under vacuum to yield a colourless powder (13.3 g, 88%).  $^1H$  NMR (400 MHz,  $DMSO-d_6$ )  $\delta$  10.20 (s, 1H, H-C(2)), 8.16 (t,  $J = 1.8$  Hz, 1H, H-C(4)), 8.03 (t,  $J = 1.8$  Hz, 1H, H-C(5)), 4.68 (t,  $J = 5.8$  Hz, 2H,  $-OCH_2CH_2CN$ ), 4.00 (s, 3H,  $CH_3$ ), 3.15 (t,  $J = 5.8$  Hz, 2H,  $-OCH_2CH_2CN$ ).

## *N*<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]adenosine 78

$C_{14}H_{16}N_6O_6$ ;  $M_r = 364.31$



Adenosine (10.0 g, 11.2 mmol) and a few crystals of ammonium sulphate were suspended in hexamethyldisilazane (HMDS) (140 mL) and anhydrous dioxane (140 mL). The mixture was heated to reflux for 6 h after which the solution was cooled to RT and the solvents removed under vacuum. The resultant syrup was treated with anhydrous toluene (50 mL) and the undissolved material removed by filtration. The supernatant was evaporated to dryness and the oil redissolved in anhydrous  $CH_2Cl_2$  (200 mL). The imidazolium salt **76** (2.90 g, 13.4 mmol) was added to the solution and the mixture stirred under argon until all solids had dissolved. The solvent was removed under vacuum and the residue taken up in MeOH (60 mL) and EtOH (150 mL) and stirred slowly for 24 h. The resultant precipitate was filtered, washed with ethanol (3 × 20 mL) and diethylether (3 × 20 mL). The solid was dried under vacuum to give the title product as a colourless amorphous solid (10.40 g, 86%). M.P. = 151-154 °C (lit<sup>[201]</sup> = 158-161 °C). IR ( $cm^{-1}$ ) 3422 (OH), 3264, 3176, 3111, 3055, 2966, 2941, 2919, 2872,

2251 (wk. CN), 1751 (carbamate C=O).  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  10.84 (br. s, 1H, NH), 8.71 (s, 1H H-C(8)), 8.66 (s, 1H, H-C(2)), 6.01 (d,  $J = 5.5$  Hz, 1H, H-C(1')), 5.56 (d,  $J = 6.1$  Hz, 1H, HO-C(2')), 5.26 (d,  $J = 4.8$  Hz, 1H, HO-C(3')), 5.15 (t,  $J = 5.5$  Hz, 1H, HO-C(5')), 4.62 (q,  $J = 5.8$  Hz, 1H, H-C(2')), 4.33 (t,  $J = 6.1$  Hz, 2H, H<sub>2</sub>-C(12)), 4.22-4.14 (m, 1H, H-C(3')), 3.98 (q,  $J = 3.8$  Hz, 1H, H-C(4')), 3.75–3.65 (m, 1H, H-C(5')), 3.62-3.52 (m, 1H, H-C(5'')), 2.95 (t,  $J = 6.1$  Hz, 2H, H<sub>2</sub>-C(13)).  $^{13}\text{C}$  NMR (100 MHz, DMSO- $d_6$ )  $\delta$  151.8, 151.7 (C(4)+C(6)+C(10)), 149.6 (C(2)), 143.0 (C(8)), 124.1 (C(5)), 118.7 (C(14)), 87.7 (C(1')), 85.8 (C(4')), 73.7 (C(2')), 70.4 (C(3')), 61.3 (C(5')), 60.0 (C(12)), 17.7 (C(13)).  $m/z$  ESI<sup>-</sup>: 363 ([M-H]<sup>-</sup>, 100%); ESI-HRMS (neg.) [M-H]<sup>-</sup> calculated for C<sub>14</sub>H<sub>15</sub>N<sub>6</sub>O<sub>6</sub>, 363.1058; found 363.1057.

### *N*<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]-2'/3'-*O*-acetyl-adenosine 97a+97b

C<sub>16</sub>H<sub>18</sub>N<sub>6</sub>O<sub>7</sub>; M<sub>r</sub> = 406.35

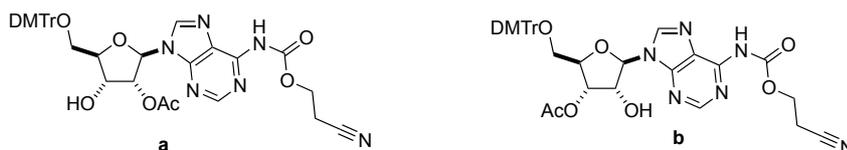


To a suspension of *N*<sup>6</sup>-[(2-cyanoethoxy)carbonyl]adenosine **78** (2.00 g, 5.48 mmol) in anhydrous dioxane (50 mL), trimethyl orthoacetate (2.07 mL, 16.4 mmol) and TFA (8.4  $\mu\text{L}$ , 0.11 mmol) were added and the mixture stirred at 50 °C for 24 h. Water (20 mL) was added to the mixture and stirred for a further 1 h at 50 °C. The solvent was removed and the residue was purified by flash column chromatography (98:2  $\rightarrow$  90:10, CH<sub>2</sub>Cl<sub>2</sub>:MeOH) to give an off-white solid of the title products as a mixture of regioisomers (quant. yield). (*Note*: the regioisomers were isolated as a mixture in a ratio of *ca.* 2.3:1, **b**:**a** calculated by integrations of both H-C(1') of **97a+97b**).  $R_f = 0.38$  (90:10, CH<sub>2</sub>Cl<sub>2</sub>:MeOH).  $^1\text{H}$  NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  9.92 (s, 0.70H, NH, **b**), 9.79 (s, 0.30H, NH, **a**), 8.62 (s, 0.30H, H-C(8), **a**), 8.40 (s, 0.70H, H-C(8), **b**), 8.27 (s, 0.30H, H-C(2), **a**), 8.21 (s, 0.70H, H-C(2), **b**), 6.17 (d,  $J = 5.7$  Hz, 0.30H, H-C(1'), **a**), 6.00 (br. s, 0.40H, 5'-OH), 5.92 (d,  $J = 7.8$  Hz, 0.70H, H-C(1'), **b**), 5.74 (t,  $J = 5.5$  Hz, 0.30H, H-C(2'), **a**), 5.58 (d,  $J = 5.2$  Hz, 0.70H, H-C(3'), **b**), 5.14 (dd,  $J = 7.9, 5.3$  Hz, 0.70H, H-C(2'), **b**), 4.85 (t,  $J = 4.3$  Hz, 0.30H, H-C(3'), **a**), 4.64 (s, 0.60H, OH), 4.49-4.38 (m, 2H, H<sub>2</sub>-C(12), **a+b**), 4.35 (s, 0.70H, H-C(4'), **b**), 4.31 (q,  $J = 2.4$  Hz, 0.30H, H-C(4'), **a**),

4.07-3.74 (m, 2H, H<sub>2</sub>-C(5'), **a+b**), 2.82 (t, *J* = 6.1 Hz, 2H, H<sub>2</sub>-C(13), **a+b**), 2.21 (s, 2.10H, CO-CH<sub>3</sub>, **b**), 2.08 (s, 0.90H, CO-CH<sub>3</sub>, **a**). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz) δ 170.7 (CO-CH<sub>3</sub>, **b**), 170.1 (CO-CH<sub>3</sub>, **a**), 152.5 (C(8), **a**), 151.9 (C(8), **b**), 150.8 (C(10), **b**), 150.7 (C(10), **a**), 150.3 (C(4), **a+b**), 149.8 (C(6), **b**), 149.6 (C(6),**a**), 143.7 (C(2), **b**), 143.1 (C(2), **a**), 123.4 (C(5), **b**), 123.0 (C(5), **a**), 117.5 (C(14), **b**), 117.4 (C(14), **a**), 90.8 (C(1'), **b**), 88.3 (C(1'), **a**), 87.0 (C(4'), **a**), 85.8 (C(4'), **b**), 75.8 (C(2'), **a**), 74.5 (C(3'), **b**), 72.7 (C(2'), **b**), 70.1 (C(3'), **a**), 62.9 (C(5'), **b**), 62.0 (C(5'), **a**), 60.4 (C(12), **b**), 60.4 (C(12), **a**), 21.1 (CO-CH<sub>3</sub>, **b**), 20.8 (CO-CH<sub>3</sub>, **a**), 18.4 (C(12), **a+b**). ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>16</sub>H<sub>19</sub>N<sub>6</sub>O<sub>7</sub><sup>+</sup>, 407.1315; found 407.1310.

***N*<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]-2'/3'-*O*-acetyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine 93a+93b**

C<sub>37</sub>H<sub>36</sub>N<sub>6</sub>O<sub>9</sub>; M<sub>r</sub> = 708.72

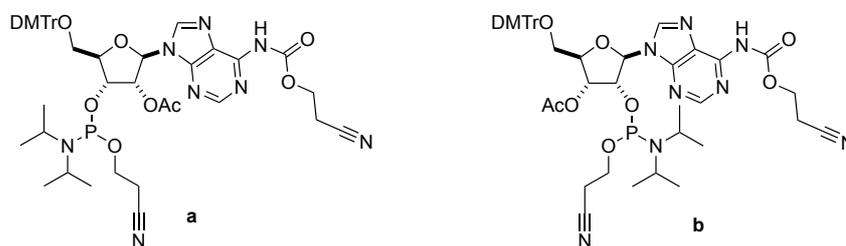


*N*<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]-2'/3'-*O*-acetyl-adenosine **97a+97b** (1.50 g, 3.69 mmol) was co-evaporated with anhydrous pyridine (3 × 20 mL). The residue was taken up in anhydrous pyridine (35 mL), DMTr-Cl (2.50 g, 7.38 mmol) was added and the mixture stirred for 3 h. MeOH (20 mL) was added and stirred for 10 min, the solvent was removed under vacuum and the residue taken up in CH<sub>2</sub>Cl<sub>2</sub> (30 mL). The organics were washed with saturated aq. NaHCO<sub>3</sub> (3 × 50 mL). The organics were separated and dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated to dryness. The residue was co-evaporated with toluene (3 × 20 mL) followed by CH<sub>2</sub>Cl<sub>2</sub> (3 × 20 mL). The crude products were isolated together by flash column chromatography (50:50:1, EtOAc:Tol:Et<sub>3</sub>N → 50:45:5:1, EtOAc:Tol:MeOH:Et<sub>3</sub>N) to give the purified mixture as an off-white solid (1.64 g, 63%). (*Note*: the regioisomers were isolated as a mixture in a ratio of *ca.* 3.2:1, **b**:**a**, calculated by integrations of both H-C(1') of **93a+93b**). *R*<sub>f</sub> = 0.15 (50:45:5:1, EtOAc:Tol:MeOH:Et<sub>3</sub>N). IR (cm<sup>-1</sup>) 3259, 3183, 3127, 2963, 2934, 2913, 2254 (wk. CN) 1742 (C=O). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 9.75, 9.63 (2 × s, 1H, NH, **b**, **a**), 8.69, 8.67 (2 × s, 1H, H-C(8), **a**, **b**), 8.26, 8.24 (2 × s, 1H, H-C(2), **b**, **a**), 7.39-7.13 (m, 9H,

DMTr, **a+b**), 6.78-6.74 (m, 4H, DMTr, **a+b**), 6.27 (d,  $J = 4.5$  Hz, 0.24H, H-C(1'), **a**), 6.10 (d,  $J = 6.5$  Hz, 0.76H, H-C(1'), **b**), 5.87 (t,  $J = 4.9$  Hz, 0.24H, H-C(2'), **a**), 5.47 (dd,  $J = 5.4, 2.4$  Hz, 0.76H, H-C(3'), **b**), 5.14 (t,  $J = 6.0$  Hz, 0.76H, H-C(2'), **b**), 4.88 (t,  $J = 5.2$  Hz, 0.24H, H-C(3'), **a**), 4.40-4.33 (m, 2.76H, (H<sub>2</sub>-C(12), **a+b**), (H-C(4'), **b**)), 4.27 (q,  $J = 3.9$  Hz, 0.24H, H-C(4'), **a**), 3.74-3.73 (m, 6H, OCH<sub>3</sub>, **a+b**), 3.56-3.35 (m, 2H, H<sub>2</sub>-C(5'), **a+b**), 2.72-2.68 (m, 2H, H<sub>2</sub>-C(13), **a+b**), 2.15 (s, 2.28H, CO-CH<sub>3</sub>, **b**), 2.08 (s, 0.72H, CO-CH<sub>3</sub>, **a**). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz)  $\delta$  170.5 (CO-CH<sub>3</sub>, **b**), 170.2 (CO-CH<sub>3</sub>, **a**), 158.6 (DMTr, **a+b**), 152.9 (C(8), **a**), 152.6 (C(8), **b**), 151.3 (C(4), **a+b**), 150.6 (C(10), **a+b**), 149.4 (C(6), **b**), 149.3 (C(6), **a**), 144.4, 144.3 (DMTr, **a+b**), 142.1 (C(2), **a**), 141.8 (C(2), **b**), 135.6, 135.5, 135.4, 135.4, 130.1, 130.1, 128.2, 128.1, 128.0, 127.1 (DMTr, **a+b**), 122.5 (C(5), **a+b**), 117.0 (C(14), **a+b**), 113.3 (DMTr, **a+b**), 89.2 (C(1'), **b**), 86.9 (DMTr-C, **b**), 86.7 (DMTr-C, **a**), 86.6 (C(1'), **a**), 83.9 (C(4'), **a**), 83.4 (C(4'), **b**), 75.8 (C(2'), **a**), 73.8, 73.7 ((C(2'), **b**), (C(3'), **b**)), 70.2 (C(3'), **a**), 63.3 (C(5'), **b**), 63.0 (C(5'), **b**), 60.1 (C(12), **a+b**), 55.3 (OCH<sub>3</sub>, **a+b**), 21.0 (CO-CH<sub>3</sub>, **b**), 20.8 (CO-CH<sub>3</sub>, **a**), 18.2 (C(13), **a+b**).  $m/z$  ESI+: 731 ([M+Na]<sup>+</sup>, 100%). ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>37</sub>H<sub>37</sub>N<sub>6</sub>O<sub>9</sub><sup>+</sup>, 709.2610; found 709.2617.

**N**<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]-2'-*O*-acetyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine 3'-*O*-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite 103a and **N**<sup>6</sup>-[(2-cyanoethoxy)carbonyl]-3'-*O*-acetyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine-2'-*O*-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite 103b<sup>‡</sup>

C<sub>46</sub>H<sub>53</sub>N<sub>8</sub>O<sub>10</sub>P; M<sub>r</sub> = 908.93



**93a+93b** (1.00g, 1.41 mmol) and 2-cyanoethyl *N,N,N',N'*-tetraisopropyl phosphoramidite (0.90 mL, 2.82 mmol) were dissolved in anhydrous THF (10 mL). A solution of 5-benzylthio-1*H*-tetrazole in anhydrous MeCN (0.35 M, 4 mL) was added dropwise and the mixture was stirred at RT for 3 h. The reaction mixture was added with stirring to saturated aq. NaHCO<sub>3</sub> (10 mL). The organics were extracted with

CH<sub>2</sub>Cl<sub>2</sub> (3 × 15 mL) and the combined organics dried over MgSO<sub>4</sub> and evaporated to dryness. The residue was passed through a short flash chromatography column (100% EtOAc) to remove the 5-benzylthio-1*H*-tetrazole. The regioisomers were dissolved in EtOAc (~100 mg/mL), purified and separated by NP-HPLC (Method A), with retention times of 9 min (**b**), 11 min (**b**), 15 min (**a**) and 24 min (**a**). The separated title regioisomers **103a** (287 mg, 22%) and **103b** (850 mg, 66%) were isolated as mixtures of two diastereomers in the form of colourless foams.

#### *Data for 103a*

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.71, 8.69 (2 × s, 2H, H-C(8), NH), 8.19 (s, 1H, H-C(2)), 7.49-7.12 (m, 9H, DMTr), 6.79 (dd, *J* = 8.6, 4.8 Hz, 4H, DMTr), 6.29 (2 × d, *J* = 5.9 Hz, 1H, H-C(1')), 6.00 (t, *J* = 5.4 Hz, 0.5H, H-C(2')), 5.90 (t, *J* = 5.7 Hz, 0.5H, H-C(2')), 5.01-4.78 (m, 1H, H-C(3')), 4.52-4.38 (m, 2.5H, H<sub>2</sub>-C(12), H-C(4')), 4.35 (q, *J* = 3.6 Hz, 0.5H, H-C(4')), 3.98-3.46 (m, 11H, ce OCH<sub>2</sub>, OCH<sub>3</sub>, <sup>i</sup>Pr CH, H-C(5')), 3.38 (ABX, *J*<sub>BA</sub> = 10.7 Hz, *J*<sub>BX</sub> = 4.0 Hz, 1H, H-C(5'')), 2.79 (t, *J* = 6.3 Hz, 2H, H<sub>2</sub>-C(13)), 2.64 (t, *J* = 6.4 Hz, 1H, ce CH<sub>2</sub>CN), 2.36 (m, 1H, ce CH<sub>2</sub>CN), 2.12, 2.08 (2 × s, 3H, CO-CH<sub>3</sub>), 1.23-1.00 (m, 12H, <sup>i</sup>Pr CH<sub>3</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz) δ 169.9, 169.9 (CO-CH<sub>3</sub>), 158.7 (DMTr), 153.0, 153.0 (C(8)), 151.6 (C(4)), 150.3 (C(10)), 149.1 (C(6)), 144.4, 144.3 (DMTr), 141.9 (C(2)), 135.6, 135.6, 135.5, 135.4, 130.3, 130.3, 130.3, 130.2, 128.4, 128.2, 128.0, 127.2, 127.1 (DMTr), 122.6 (C(5)), 117.7, 117.4 (ce CN), 116.8 (C(14)), 113.3 (DMTr), 86.9, 86.9 (DMTr-C), 86.1 (C(1')), 84.7, 84.4, 84.4 (C(4')), 74.8, 74.7, 74.7 (C(2')), 71.5, 71.3, 71.0, 70.8 (C(3')), 63.0 (C(5')), 60.3 (C(12)), 59.0, 58.8, 58.3, 58.1 (ce OCH<sub>2</sub>), 55.4, 55.4 (OCH<sub>3</sub>), 43.5, 43.4, 43.4, 43.3 (<sup>i</sup>Pr CH), 24.9, 24.8, 24.7, 24.7, 24.7, 24.6 (<sup>i</sup>Pr CH<sub>3</sub>), 21.0, 20.9 (CO-CH<sub>3</sub>), 20.3, 20.2 (ce CH<sub>2</sub>CN), 18.3 (C(13)). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 151.33-150.98 (m), 150.44-150.02 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled) δ 151.17 (s), 150.22 (s). ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>46</sub>H<sub>54</sub>N<sub>8</sub>O<sub>10</sub>P<sup>+</sup>, 909.3665; found 909.3701.

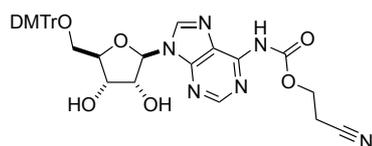
#### *Data for 103b*

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.74, 8.71, 8.69 (3 × s, 2H, NH, H-C(8)), 8.26, 8.23 (2 × s, 1H, H-C(2)), 7.47-7.13 (m, 9H, DMTr), 6.88-6.73 (m, 4H, DMTr), 6.26 (d, *J* = 5.3 Hz, 0.5H, H-C(1')), 6.22 (d, *J* = 5.7 Hz, 0.5H, H-C(1')), 5.57, 5.52 ((app. t, *J* = 4.4

Hz), (t,  $J = 4.8$  Hz, 1H, H-C(3')), 5.21, 5.15 (2 × dt,  $J = 10.4, 5.4$  Hz, 1H, H-C(2')), 4.46 (t,  $J = 6.2$  Hz, 2H, H-C(12)), 4.33 (quin.,  $J = 3.7$  Hz, 1H, H-C(4')), 3.88-3.64 (m, 7H, OCH<sub>3</sub>, ce OCH<sub>2</sub>), 3.63-3.36 (m, 5H, H-C(5')), ce OCH<sub>2</sub>, <sup>i</sup>Pr CH), 2.80 (t,  $J = 6.2$  Hz, 2H, H-C(13)), 2.55 (td,  $J = 6.4, 2.9$  Hz, 1H, ce CH<sub>2</sub>CN), 2.33 (t,  $J = 6.4$  Hz, 1H, ce CH<sub>2</sub>CN), 2.14, 2.10 (2×s, 3H, CO-CH<sub>3</sub>), 1.20-1.00 (m, 9H, <sup>i</sup>Pr CH<sub>3</sub>), 0.88 (d,  $J = 6.7$  Hz, 3H, <sup>i</sup>Pr CH<sub>3</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz) δ 169.9, 169.8 (CO-CH<sub>3</sub>), 158.7 (DMTr), 153.0, 152.8 (C(8)), 151.7, 151.6 (C(4)), 150.4, 150.3 (C(10)), 149.1, 149.0 (C(6)), 144.5, 144.5 (DMTr), 142.0, 141.8 (C(2)), 135.6, 135.5, 130.2, 130.2, 128.3, 128.2, 128.1, 128.1, 128.0, 127.2 (DMTr), 122.6 (C(5)), 117.5, 117.4 (ce CN), 116.8 (C(14)), 113.4, 113.3 (DMTr), 87.7, 87.6, 87.5 (C(1')), 87.0, 87.0 (DMTr-C), 82.5, 82.1 (C(4')), 74.7, 74.5, 74.2, 74.0 (C(2')), 72.2, 72.1 (C(3')), 63.0 (C(5')), 60.3 (C(12)), 58.8, 58.6, 58.1, 57.9 (ce OCH<sub>2</sub>), 55.4 (OCH<sub>3</sub>), 43.5, 43.4 (<sup>i</sup>Pr CH), 24.8, 24.7, 24.6, 24.6, 24.6, 24.5, 24.4, 24.3 (<sup>i</sup>Pr CH<sub>3</sub>), 21.1, 21.0 (CO-CH<sub>3</sub>), 20.3, 20.2, 20.1, 20.0 (ce CH<sub>2</sub>CN), 18.3 (C(13)). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 151.70-151.41 (m), 151.37-151.08 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled) δ 151.56 (s), 151.18 (s). ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>46</sub>H<sub>54</sub>N<sub>8</sub>O<sub>10</sub>P<sup>+</sup>, 909.3665; found 909.3701.

### **N<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]-5'-O-(4,4'-dimethoxytrityl)adenosine 88**

C<sub>35</sub>H<sub>34</sub>N<sub>6</sub>O<sub>8</sub>; M<sub>r</sub> = 666.68

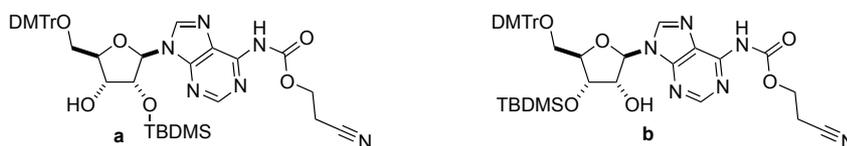


N<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]adenosine **78** (0.92 g, 2.52 mmol) was co-evaporated with anhydrous pyridine (3 × 20 ml). The residue was taken up in anhydrous pyridine (20 ml), DMTr-Cl (1.03 g, 3.4 mmol) was added and the mixture stirred for 3 h. The solvent was removed under vacuum and the residue taken up in CH<sub>2</sub>Cl<sub>2</sub> (20 ml) and the organic layer was washed with saturated aq. NaHCO<sub>3</sub> (3 × 30 ml). The organics were dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent removed under vacuum. The crude residue was co-evaporated with toluene (3 × 20 ml) followed by CH<sub>2</sub>Cl<sub>2</sub> (2 × 20 ml) to remove residual pyridine, and finally purified by flash column chromatography (75:25:2, EtOAc:Tol:Et<sub>3</sub>N → 40:2:2:2, EtOAc:Tol:MeOH:Et<sub>3</sub>N) to give the title compound as a slightly yellow foam (1.41 g, 83%). R<sub>f</sub> = 0.38 (40:2:2:1, EtOAc:Tol:MeOH:Et<sub>3</sub>N). M.P.

= 97-100 °C. IR (cm<sup>-1</sup>) 3248 (O-H), 2930, 2835 (CH), 1757 (carbamate C=O). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.67 (s, 1H, H-C(8)), 8.28 (s, 1H, H-C(2)), 7.31-7.14 (m, 9H, DMTr), 6.74 (d, *J* = 8.8 Hz, 4H, DMTr), 6.11 (d, *J* = 5.5 Hz, 1H, H-C(1')), 4.90 (t, *J* = 5.4 Hz, 1H, H-C(2')), 4.51 (dd, *J* = 5.0, 2.8 Hz, 1H, H-C(3')), 4.42-4.36 (m, 3H, H-C(4') and H<sub>2</sub>-C(11)), 3.74 (s, 6H, OCH<sub>3</sub>), 3.46 (ABX, *J*<sub>AB</sub> = 10.6, *J*<sub>AX</sub> = 3.3 Hz, 1H, H-C(5')), 3.34 (ABX, *J*<sub>BA</sub> = 10.6, *J*<sub>BX</sub> = 3.8 Hz, 1H, H-C(5'')), 2.71 (t, *J* = 6.2 Hz, 2H, H<sub>2</sub>-C(12)). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz) δ 158.5 (DMTr), 152.3 (C(8)), 150.9 (C(4)), 150.3 (C(10)), 149.1 (C(6)), 144.3 (DMTr), 141.8 (C(2)), 135.4 (DMTr), 135.4 (DMTr), 129.9 (DMTr), 127.9 (DMTr), 127.8 (DMTr), 126.9 (DMTr), 122.3 (C(5)), 116.8 (CN), 113.1 (DMTr), 90.0 (C(1')), 86.5 (DMTr-C), 85.5 (C(4')), 75.3 (C(2')), 72.1 (C(3')), 63.4 (C(5')), 60.0 (C(11)), 55.2 (OCH<sub>3</sub>), 18.1 (C(12)). *m/z* ESI-: 665 ([M-H]<sup>-</sup>, 75%), 701 ([M+Cl]<sup>-</sup>, 100%); ESI-HRMS (neg.) [M-H]<sup>-</sup> calculated for C<sub>35</sub>H<sub>35</sub>N<sub>6</sub>O<sub>8</sub><sup>-</sup>, 667.2511; found 667.2533.

***N*<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]-2'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)adenosine 108a and *N*<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]-3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)adenosine 108b**‡

C<sub>41</sub>H<sub>48</sub>N<sub>6</sub>O<sub>8</sub>Si; M<sub>r</sub> = 780.94



*N*<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]-5'-*O*-(4,4'-dimethoxytrityl)adenosine **88** (4.00 g, 6.00 mmol) was dissolved in anhydrous THF (50 mL). Anhydrous pyridine (1.80 mL, 22.2 mmol) followed by AgNO<sub>3</sub> (1.22 g, 7.20 mmol) was added, the mixture was warmed until the AgNO<sub>3</sub> had fully dissolved. Whilst the mixture was still warm, TBDMS-Cl (1.18 g, 7.80 mmol) was added resulting in a colourless precipitate. The mixture was stirred in the dark for 5 h. The solid was removed by filtration and the supernatant immediately filtered into saturated aq. NaHCO<sub>3</sub> (50 mL). The aqueous phase was extracted with EtOAc (3 × 50 mL), the combined organic phases were dried over MgSO<sub>4</sub>, and finally the solvent removed under vacuum. The crude residue was purified and the regioisomers separated by flash column chromatography (3:1 → 2:1 →

1:1, Et<sub>2</sub>O:EtOAc) to give **108a** (2.50 g, 53%) and **108b** (0.91 g, 20%) both as colourless foams.

*Data for 108a*

R<sub>f</sub> = 0.45 (1:1, Et<sub>2</sub>O:EtOAc). <sup>1</sup>H NMR (400 MHz, DMSO) δ 10.80 (s, 1H, NH), 8.58, 8.57 (2 × s, 2H, H-C(2), H-C(8)), 7.43-7.35 (m, 2H, DMTr), 7.31-7.16 (m, 7H, DMTr), 6.85 (dd, *J* = 9.0, 3.1 Hz, 4H, DMTr), 6.05 (d, *J* = 4.8 Hz, 1H, H-C(1')), 5.18 (d, *J* = 5.9 Hz, 1H, 3'-OH), 4.86 (t, *J* = 4.9 Hz, 1H, H-C(2')), 4.29 (m, 3H, H<sub>2</sub>-C(12), H-C(3')), 4.13 (q, *J* = 4.5 Hz, 1H, H-C(4')), 3.73 (s, 6H, OCH<sub>3</sub>), 3.29 (m, 2H, H-C(5'))<sup>f</sup>, 2.93 (t, *J* = 6.0 Hz, 2H, H-C(13)), 0.75 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), -0.04 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), -0.14 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (DMSO, 101 MHz) δ 158.0 (DMTr), 151.6, 151.6 (C(8), C(10)), 149.6 (C(4), C(6)), 144.8 (DMTr), 142.9 (C(2)), 135.5, 135.4, 129.7, 127.8, 127.6, 126.7 (DMTr), 124.0 (C(5)), 118.5 (C(14)), 113.1 (DMTr), 88.2 (C(1')), 85.5 (DMTr-C), 83.5 (C(4')), 74.8 (C(2')), 70.1 (C(3')), 63.4 (C(5')), 59.9 (C(12)), 55.0, 54.9 (OCH<sub>3</sub>), 25.5 (SiC(CH<sub>3</sub>)<sub>3</sub>), 17.8, 17.6 (C(13), SiC(CH<sub>3</sub>)<sub>3</sub>), -4.8 (Si(CH<sub>3</sub>)<sub>2</sub>), -5.3 (Si(CH<sub>3</sub>)<sub>2</sub>). ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>41</sub>H<sub>49</sub>N<sub>6</sub>O<sub>8</sub>Si<sup>+</sup>, 781.3381; found 781.3354.

*Data for 108b*

R<sub>f</sub> = 0.33 (1:1, Et<sub>2</sub>O:EtOAc). <sup>1</sup>H NMR (400 MHz, DMSO) δ 10.81 (s, 1H, NH), 8.63 (s, 1H, H-C(2)), 8.56 (s, 1H, H-C(8)), 7.39-7.17 (m, 9H, DMTr), 6.88-6.79 (m, 4H, DMTr), 6.00 (d, *J* = 5.1 Hz, 1H, H-C(1')), 5.47 (d, *J* = 6.0 Hz, 1H, 2'-OH), 4.88 (q, *J* = 5.4 Hz, 1H, H-C(2')), 4.49 (t, *J* = 4.6 Hz, 1H, H-C(3')), 4.32 (t, *J* = 6.0 Hz, 2H, H<sub>2</sub>-C(12)), 4.06 (q, *J* = 4.5 Hz, 1H, H-C(4')), 3.72 (s, 6H, OCH<sub>3</sub>), 3.36 (m, 1H, H-C(5'))<sup>f</sup>, 3.15 (ABX, *J*<sub>BA</sub> = 10.5, *J*<sub>BX</sub> = 4.9 Hz, 1H, H-C(5'')), 2.93 (t, *J* = 6.0 Hz, 2H, H<sub>2</sub>-C(13)), 0.84 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.08 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), 0.05 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (DMSO, 101 MHz) δ = 158.1 (DMTr), 151.7, 151.6, 151.5 (C(10), C(8), C(4)), 149.6 (C(6)), 144.7 (DMTr), 143.7 (C(2)), 135.4, 129.6, 129.6, 127.7, 127.6, 126.7 (DMTr), 124.2 (C(5)), 118.5 (C(14)), 113.1 (DMTr), 88.3 (C(1')), 85.6 (DMTr-C), 83.6 (C(4')), 72.2 (C(3')), 72.0 (C(2')), 63.0 (C(5')), 59.9 (C(12)), 55.0 (OCH<sub>3</sub>), 25.8 (SiC(CH<sub>3</sub>)<sub>3</sub>),

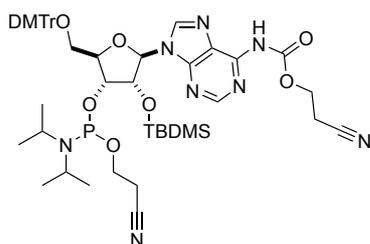
---

<sup>f</sup> Obscured by HOD peak.

18.0 (C(13)), 17.6 (SiC(CH<sub>3</sub>)<sub>3</sub>), -4.5 (Si(CH<sub>3</sub>)<sub>2</sub>), -5.1 (Si(CH<sub>3</sub>)<sub>2</sub>). ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>41</sub>H<sub>49</sub>N<sub>6</sub>O<sub>8</sub>Si<sup>+</sup>, 781.3381; found 781.3380.

***N*<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]-2'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)adenosine-3'-*O*-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite 112a‡**

C<sub>50</sub>H<sub>65</sub>N<sub>8</sub>O<sub>9</sub>PSi; M<sub>r</sub> = 981.16

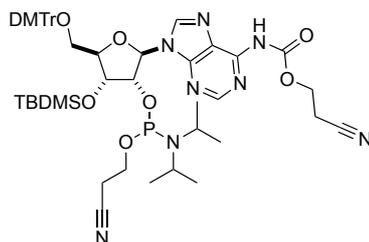


*N*<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]-2'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)adenosine **108a** (1.00 g, 1.28 mmol) was dissolved in anhydrous THF (8 mL). To this solution was added *N,N*-diisopropylethylamine (780 μL, 4.48 mmol) and 2-cyanoethyl *N,N*-diisopropyl phosphoamidochloridite (400 μL, 1.80 mmol) at 0 °C. The mixture was warmed up to RT and stirred for 5 h. Anhydrous methanol (4 ml) was added to quench the reaction and the mixture was stirred for a further 30 min. The reaction was diluted with EtOAc (10 mL) and washed with saturated aq. NaHCO<sub>3</sub> (3 × 20 mL). The combined organic layers were dried over MgSO<sub>4</sub> and the solvent evaporated under vacuum. The crude product was purified by flash column chromatography (50:50:1 → 60:40:1, EtOAc:*c*-Hex:Et<sub>3</sub>N). The title product was isolated as a mixed of two diastereoisomers in the form of a colourless foam (1.10 g, 88% yield). R<sub>f</sub> = 0.28 (60:40:1, EtOAc:*c*-Hex:Et<sub>3</sub>N). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.67, 8.65 (s, 1H, H-C(8)), 8.49 (s, 1H, NH), 8.24, 8.21 (s, 1H, H-C(2)), 7.51-7.19 (m, 9H, DMTr), 6.92-6.73 (m, 4H, DMTr), 6.08 (d, *J* = 6.3 Hz, 0.55H, H-C(1')), 6.03 (d, *J* = 6.1 Hz, 0.45H, H-C(1')), 5.12-4.98 (m, 1H, H-C(2')), 4.52-4.32 (m, 4H, H<sub>2</sub>-C(12), H-C(3'), H-C(4')), 4.03-3.52 (m, 11H, ce OCH<sub>2</sub>, OCH<sub>3</sub>, <sup>i</sup>Pr CH, H-C(5')), 3.41-3.27 (m, 1H, H-C(5')), 2.81 (m, 2H, H<sub>2</sub>-C(13)), 2.73-2.57 (m, 1H, ce CH<sub>2</sub>CN), 2.44-2.22 (m, 1H, ce CH<sub>2</sub>CN), 1.23-1.11 (m, 9H, <sup>i</sup>Pr CH<sub>3</sub>), 1.05 (d, *J* = 6.8 Hz, 3H, <sup>i</sup>Pr CH<sub>3</sub>), 0.75 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), -0.03, -0.06 (2xs, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), -0.22, -0.23 (2xs, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz) δ 158.7 (DMTr), 152.9 (C(8)), 151.5 (C(4)), 150.2

(C(10)), 148.9 (C(6)), 144.7, 144.6 (DMTr), 142.3 (C(2)), 135.8, 135.8, 135.6, 135.6, 130.3, 130.3, 130.2, 128.4, 128.3, 128.1, 128.0, 127.1 (DMTr), 122.7, 122.6 (C(5)), 117.7, 117.4 (ce CN), 116.8 (C(14)), 113.3, 113.3 (DMTr), 88.6, 88.4 (C(1')), 86.9, 86.8 (DMTr), 84.4, 84.1, 84.1 (C(4')), 75.4, 74.8, 74.8 (C(2')), 73.5, 73.4, 72.9, 72.7 (C(3')), 63.4, 63.2 (C(5')), 60.2 (C(12)), 59.0, 58.8, 57.8, 57.6 (ce OCH<sub>2</sub>), 55.4, 55.4 (OCH<sub>3</sub>), 43.6, 43.5, 43.2, 43.0 (<sup>i</sup>Pr CH), 25.7, 25.7 SiC(CH<sub>3</sub>)<sub>3</sub>, 24.9, 24.8, 24.8, 24.7 (<sup>i</sup>Pr CH<sub>3</sub>), 20.6, 20.6, 20.3, 20.2 (ce CH<sub>2</sub>CN), 18.3, 18.1, 18.0 (C(13), SiC(CH<sub>3</sub>)<sub>3</sub>), -4.5, -4.6, -5.0 (Si(CH<sub>3</sub>)<sub>2</sub>). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 151.21-150.80 (m), 149.28-148.85 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled) δ 151.04 (s), 149.06 (s). ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>50</sub>H<sub>66</sub>N<sub>8</sub>O<sub>9</sub>SiP<sup>+</sup>, 981.4460; found 981.4433.

***N*<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]-3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)adenosine-2'-*O*-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite**  
112b<sup>‡</sup>

C<sub>50</sub>H<sub>65</sub>N<sub>8</sub>O<sub>9</sub>PSi; M<sub>r</sub> = 981.16

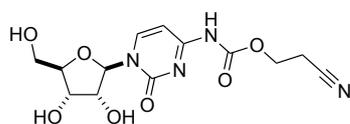


*N*<sup>6</sup>-[(2-Cyanoethoxy)carbonyl]-3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)adenosine **108b** (1.00 g, 1.28 mmol) was dissolved in anhydrous THF (8 mL). To the solution was added *N,N*-diisopropylethylamine (780 μL, 4.48 mmol) and 2-cyanoethyl *N,N*-diisopropyl phosphoramidochloridite (400 μL, 1.8 mmol) at 0 °C. The mixture was warmed to RT and stirred for 4 h. Anhydrous methanol (4 ml) was added to quench the reaction and the mixture was stirred for a further 30 min. The reaction was diluted with EtOAc (10 mL) and washed with saturated aq. NaHCO<sub>3</sub> (3 × 20 mL). The combined organic layers were dried over MgSO<sub>4</sub> and the solvent evaporated under vacuum. The crude product was purified by flash column chromatography (50:50:1 → 60:40:1 EtOAc:*c*-Hex:Et<sub>3</sub>N). The title product was isolated as a mixed of two diastereoisomers in the form of a colourless foam (1.10 g, 88% yield). R<sub>f</sub> = 0.25 (60:40:1 EtOAc:*c*-Hex:Et<sub>3</sub>N). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.70, 8.68 (s, 1H, H-C(8)),

8.56 (s, 1H, NH), 8.30, 8.27 (s, 1H, H-C(2)), 7.44-7.14 (m, 9H, DMTr), 6.79 (m, 4H, DMTr), 6.27 (d,  $J = 4.3$  Hz, 0.62H, H-C(1')), 6.18 (d,  $J = 4.7$  Hz, 0.38H, H-C(1')), 5.01 (dt,  $J = 11.3, 4.6$  Hz, 0.38H, H-C(2')), 4.82 (dt,  $J = 9.5, 4.5$  Hz, 0.62H, H-C(2')), 4.57-4.49 (m, 1H, H-C(3')), 4.45 (t,  $J = 6.2$  Hz, 2H, H<sub>2</sub>-C(12)), 4.29-4.13 (m, 1H, H-C(4')), 3.87-3.42 (m, 11H, ce OCH<sub>2</sub>, OCH<sub>3</sub>, <sup>i</sup>Pr CH, H-C(5')), 3.36-3.20 (m, 1H, H-C(5'')), 2.83-2.76 (m, 2H, H<sub>2</sub>-C(13)), 2.50 (t,  $J = 6.4$  Hz, 1H, ce CH<sub>2</sub>CN), 2.38 (t,  $J = 6.3$ , 1H, ce CH<sub>2</sub>CN), 1.15-1.02 (m, 9H, <sup>i</sup>Pr CH<sub>3</sub>), 0.95-0.79 (m, 12H, <sup>i</sup>Pr CH<sub>3</sub>, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.09, 0.06 (2xs, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), 0.01, -0.00 (2xs, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz)  $\delta$  158.7, 158.7 (DMTr), 152.7, 152.6 (C(8)), 151.4, 151.4 (C(4)), 150.3, 150.3 (C(10)), 149.0, 149.0 (C(6)), 144.6 (DMTr), 142.6, 142.4 (C(2)), 135.7, 135.7, 130.2, 130.2, 130.2, 128.3, 128.3, 128.0, 128.0, 127.1, 127.1 (DMTr), 122.7 (C(5)), 117.6 (ce CN), 116.8 (C(14)), 113.3, 113.3 (DMTr), 88.1, 88.0, 87.9 (C(1')), 86.9, 86.7 (DMTr-C), 84.8, 84.4 (C(4')), 76.0, 75.9, 75.6, 75.4 (C(2')), 71.9, 71.6, 71.5 (C(3')), 63.1, 62.8 (C(5'')), 60.3, 60.2 (C(12)), 58.6, 58.4, 57.9, 57.7 (ce OCH<sub>2</sub>), 55.4, 55.4 (OCH<sub>3</sub>), 43.5, 43.3, 43.3, 43.2 (<sup>i</sup>Pr CH), 25.9 (SiC(CH<sub>3</sub>)<sub>3</sub>), 24.8, 24.8, 24.7, 24.7, 24.4, 24.4 (<sup>i</sup>Pr CH<sub>3</sub>), 20.4, 20.3, 20.2, 20.1 (ce CH<sub>2</sub>CN), 18.3, 18.2 (C(13), SiC(CH<sub>3</sub>)<sub>3</sub>), -4.2, -4.3, -4.7, -4.8 (Si(CH<sub>3</sub>)<sub>2</sub>). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>)  $\delta$  150.74-149.95 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled)  $\delta$  150.44 (s), 150.15 (s). ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>50</sub>H<sub>66</sub>N<sub>8</sub>O<sub>9</sub>SiP<sup>+</sup>, 981.4460; found 981.4424.

#### ***N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]cytidine 79**

C<sub>13</sub>H<sub>16</sub>N<sub>4</sub>O<sub>7</sub>; M<sub>r</sub> = 340.10



A suspension of cytidine (11.8 g, 48.5 mmol) and a few crystals of ammonium sulphate in hexamethyldisilazane (HMDS) (100 mL) and anhydrous dioxane (100 mL) were heated to reflux for 3 h. The solvents were removed under vacuum and the resultant oil was treated with toluene (100 mL). The insoluble materials were removed by filtration and the supernatant evaporated to dryness. The resultant oil was taken up in anhydrous CH<sub>2</sub>Cl<sub>2</sub> (150 mL), **76** (13.6 g, 63.1 mmol) was added to the solution and the mixture stirred for 48 h. The CH<sub>2</sub>Cl<sub>2</sub> was removed under vacuum and the residue treated with

MeOH (100 mL) followed by stirring for 72 h. The resultant colourless solids were collected by filtration and washed with cold MeOH (3 × 30 mL) and Et<sub>2</sub>O (2 × 30 mL). The solids were dried in air then under high vacuum to give the title compound as a colourless amorphous solid (15.0 g, 91%). M.P. = 141-142 °C. IR (cm<sup>-1</sup>) 3493 (O-H), 3349 (NH), 3126, 3053, 2971, 2941 (CH), 1733 (carbamate C=O). <sup>1</sup>H NMR (400MHz, DMSO-*d*<sub>6</sub>) δ 10.90 (br. s., 1H, NH), 8.41 (d, *J* = 7.6 Hz, 1H, H-C(6)), 6.98 (d, *J* = 7.6 Hz, 1H, H-C(5)), 5.77 (d, *J* = 2.8 Hz, 1H, H-C(1')), 5.08 (br. s., 3H, -OH), 4.30 (t, *J* = 6.1 Hz, 2H, H<sub>2</sub>-C(9)), 3.93-4.01 (m, 2H, H-C(2'), H-C(3')), 3.89 (dt, *J* = 5.7, 2.8 Hz, 1H, H-C(4')), 3.74 (ABX, *J*<sub>AB</sub> = 12.2, *J*<sub>AX</sub> = 2.8 Hz, 1H, H-C(5')), 3.59 (ABX, *J*<sub>BA</sub> = 12.2, *J*<sub>BX</sub> = 2.8 Hz, 1H, H-C(5'')), 2.93 (t, *J* = 6.1 Hz, 2H, H<sub>2</sub>-C(10)). <sup>13</sup>C NMR (100MHz, DMSO-*d*<sub>6</sub>) δ 162.7 (C(4)), 154.4 (C(2)), 152.9 ((C(7)), 145.1 (C(6)), 118.5 (C(11)), 94.3 (C(5)), 90.1 (C(1')), 84.2 (C(4')), 74.5 (C(2')/C(3')), 68.7 (C(2')/C(3')), 60.2, 60.0 (C(9), C(5')), 17.6 (C(10)). *m/z* ESI<sup>-</sup>: 339 ([M-H]<sup>-</sup>, 100%); ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>13</sub>H<sub>17</sub>N<sub>4</sub>O<sub>7</sub><sup>+</sup>, 341.1092; found 341.1101. Elemental analysis (% calcd, % found for C<sub>13</sub>H<sub>16</sub>N<sub>4</sub>O<sub>7</sub>·½H<sub>2</sub>O): C (45.09, 45.03), H (4.85, 4.63), N (16.18, 15.97).

#### *N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]-2'/3'-*O*-(acetyl)cytidine **98a**+**98b**

C<sub>15</sub>H<sub>18</sub>N<sub>4</sub>O<sub>8</sub>; M<sub>r</sub> = 382.33



To a suspension of *N*<sup>4</sup>-[(2-cyanoethoxy)carbonyl]cytidine **79** (7.00 g, 20.6 mmol) in anhydrous MeCN (140 mL), trimethyl orthoacetate (5.82 mL, 46.4 mmol) and TFA (158 μL, 2.06 mmol) were added and the mixture stirred overnight. Water (50 mL) was added and reaction mixture stirred for a further 20 min. The reaction mixture was evaporated to dryness and the residue purified by flash column chromatography (92:8 CH<sub>2</sub>Cl<sub>2</sub>:MeOH) to give the mixture of isomers as a colourless solid (6.46 g, 82%). (Note: The products were isolated in a ratio of *ca.* 2.5:1, **b**: **a** respectively, calculated by integrations of both H-C(1') of **98a**+**98b**). R<sub>f</sub> = 0.23 (92:8 CH<sub>2</sub>Cl<sub>2</sub>:MeOH). IR (cm<sup>-1</sup>) 3327 (O-H), 3265 (NH), 3140, 2976, 2938 (CH), 2254 (wk. CN), 1752 (ester C=O), 1732 (carbamate C=O). <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) δ 10.96 (s, 1H, NH), 8.39-8.34 (m, 1H, H-C(6), **a**+**b**), 7.01 (s, 1H, H-C(5), **a**+**b**), 5.95 (d, *J* = 3.9 Hz, 0.30H, H-C(1'),

**a**), 5.86 (d,  $J = 4.9$  Hz, 0.70H, H-C(1'), **b**), 5.79 (d,  $J = 5.7$  Hz, 0.70H, 2'-OH, **b**), 5.50 (d,  $J = 5.6$ , 0.30H, 3'-OH, **a**), 5.39, 5.19 (2×br. s, 1H, 5'-OH, **a+b**), 5.17 (t,  $J = 4.5$  Hz, 0.30H, H-C(2'), **a**), 5.02 (t,  $J = 5.0$  Hz, 0.70H, H-C(3'), **b**), 4.30 (m, 3H, (H-C(2'), **b**) (H<sub>2</sub>-C(9), **a+b**)), 4.22 (q,  $J = 5.4$  Hz, 0.30H, H-C(3'), **a**), 4.13-4.10 (m, 0.70H, H-C(4'), **b**), 3.92 (dt,  $J = 5.6, 2.7$  Hz, 0.70H, H-C(4'), **a**), 3.76-3.58 (m, 2H, H<sub>2</sub>-C(5'), **a+b**), 2.93 (t,  $J = 6.0$  Hz, 2H, H<sub>2</sub>-C(10), **a+b**), 2.08 (s, 3H, CO-CH<sub>3</sub>, **a+b**). <sup>13</sup>C NMR (101 MHz, DMSO-*d*<sub>6</sub>) δ 169.8 (CO-CH<sub>3</sub>, **b**), 169.3 (CO-CH<sub>3</sub>, **a**), 162.9, 162.8 (C(4), **a+b**), 154.5, 154.2 (C(2), **a+b**), 152.9 (C(7), **a+b**), 145.0, 145.0 (C(6), **a+b**), 118.4 (C(11), **a+b**), 94.9, 94.8 (C(5), **a+b**), 89.6 (C(1'), **b**), 87.7 (C(1'), **a**), 85.0 (C(4'), **a**), 82.4 (C(4'), **b**), 75.8 (C(2'), **a**), 72.6 (C(2'), **b**), 71.9 (C(3'), **b**), 67.8 (C(3'), **a**), 60.3, 60.2, 60.0 ([C(5'), **a+b**], [C(9), **a+b**]), 20.8, 20.7 (CO-CH<sub>3</sub>, **a+b**), 17.6 (C(10), **a+b**). ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>15</sub>H<sub>19</sub>N<sub>4</sub>O<sub>8</sub><sup>+</sup>, 383.1203; found 383.1213.

***N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]-2'/3'-*O*-(acetyl)-5'-*O*-(4,4'-dimethoxytrityl)cytidine 100a+100b**

C<sub>36</sub>H<sub>36</sub>N<sub>4</sub>O<sub>10</sub>; M<sub>r</sub> = 684.69

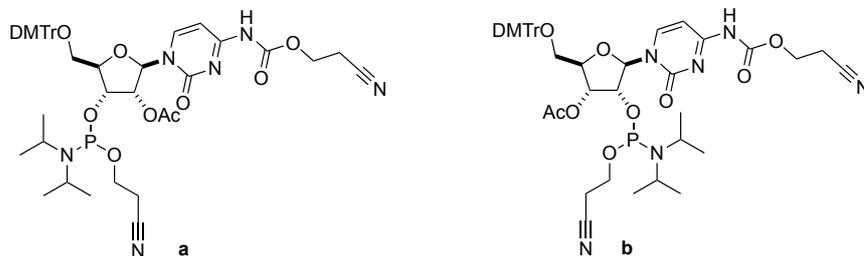


*N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]-2'/3'-*O*-(acetyl)cytidine **98a+98b** (6.46 g, 16.9 mmol) was co-evaporated with anhydrous pyridine (3 × 50 mL). The residue and DMTr-Cl (6.85 g, 20.2 mmol) were dissolved in anhydrous pyridine (65 mL) and stirred overnight at RT. MeOH (30 mL) was added to quench the reaction and the solvent completely removed. The residue was taken up in CH<sub>2</sub>Cl<sub>2</sub> and washed with saturated aq. NaHCO<sub>3</sub> (3 × 50 mL). The organic layer was separated and dried over MgSO<sub>4</sub>. The solvent was removed under vacuum and the residue co-evaporated with toluene (3 × 40 mL) then CH<sub>2</sub>Cl<sub>2</sub> (3 × 50 mL). The residue was purified by flash column chromatography (100:2, EtOAc:Et<sub>3</sub>N → 95:5:2, EtOAc:MeOH:Et<sub>3</sub>N) to yield to the mixture of regioisomers as an off-white foam (11.0g, 96%). (*Note*: the regioisomers were isolated as a mixture in a ratio of *ca.* 3:1, **b**:**a** calculated by integrations of both H-C(1') of **100a+100b**). R<sub>f</sub> = 0.29 (95:5:2, EtOAc:MeOH:Et<sub>3</sub>N). IR (cm<sup>-1</sup>) 3273 (O-H), 3001, 2965, 2934, 2837 (CH), 2359 (wk. CN), 1744, (C=O). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.34 (d,  $J = 7.5$  Hz, 1H,

H-C(6), **a+b**), 7.43-7.21 (m, 9H, DMTr, **a+b**), 7.03 (br. s, 1H, H-C(5), **a+b**), 6.86-6.83 (m, 4H, DMTr, **a+b**), 6.05 (d,  $J = 1.7$  Hz, 0.25H, H-C(1'), **a**), 5.92 (d,  $J = 2.7$  Hz, 0.75H, H-C(1'), **b**), 5.47 (dd,  $J = 4.8, 1.9$  Hz, 0.25H, H-C(2'), **a**), 5.21 (t,  $J = 5.7$  Hz, 0.75H, H-C(3'), **b**), 4.67-4.63 (m, 1H, (H-C(3'), **a**), (H-C(2'), **b**)), 4.33 (m, 2.75H, (H-C(4'), **b**), (H<sub>2</sub>-C(9), **a+b**)), 4.20-4.16 (m, 0.25H, H-C(4'), **a**), 3.79-3.78 (m, 6H, OCH<sub>3</sub>, **a+b**), 3.61-3.49 (m, 1.25H, (H<sub>2</sub>-C(5'), **a**), (H-C(5'), **b**)), 3.39 (dd,  $J = 11.2, 2.7$  Hz, 0.75H, H-C(5''), **b**), 2.72 (m, 2H, H<sub>2</sub>-C(10), **a+b**), 2.14 (s, 0.75H, CH<sub>3</sub>, **a**), 2.11 (s, 2.25H, CH<sub>3</sub>, **b**). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>)  $\delta$  170.6 (CO-CH<sub>3</sub>, **a**), 170.4 (CO-CH<sub>3</sub>, **b**), 162.9 (C(4), **a+b**), 158.8 (DMTr, **a+b**), 155.9 (C(2), **a+b**), 152.1 (C(7), **a+b**), 144.4, 144.2, 144.1 ((C(6), **a+b**), (DMTr, **a+b**)), 135.6, 135.4, 135.2 (DMTr, **a+b**), 130.1, 130.1, 128.3, 128.1, 127.2 (DMTr, **a+b**), 117.0, 116.9 (C(11), **a+b**), 113.4 (DMTr, **a+b**), 95.6 (C(5), **a+b**), 92.7 (C(1'), **b**), 89.0 (C(1'), **a**), 87.3, 87.2 (DMTr-C, **a+b**), 82.8 (C(4'), **a**), 81.8 (C(1'), **b**), 76.7 (C(2'), **a**), 74.5 (C(2'), **b**), 71.5 (C(3'), **b**), 68.5 (C(3'), **a**), 61.6 (C(5'), **b**), 61.4 (C(5'), **a**), 60.2 (C(9), **a+b**), 55.3 (OCH<sub>3</sub>, **a+b**), 20.8 (CO-CH<sub>3</sub>, **a+b**), 18.1 (C(10), **a+b**). ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>36</sub>H<sub>37</sub>N<sub>4</sub>O<sub>10</sub><sup>+</sup>, 685.2510; found 685.2528.

*N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]-2'-*O*-(acetyl)-5'-*O*-(4,4'-dimethoxytrityl)cytidine-3'-*O*-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite 106a and *N*<sup>4</sup>-[(2-cyanoethoxy)carbonyl]-3'-*O*-(acetyl)-5'-*O*-(4,4'-dimethoxytrityl)cytidine-2'-*O*-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite 106b<sup>‡,f</sup>

C<sub>45</sub>H<sub>53</sub>N<sub>6</sub>O<sub>11</sub>P; M<sub>r</sub> = 884.35



*N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]-2'/3'-*O*-(acetyl)-5'-*O*-(4,4'-dimethoxytrityl)cytidine **100a+100b** (1.50 g, 2.19 mmol) and 2-cyanoethyl *N,N,N',N'*-tetraisopropyl phosphoramidite (0.90 mL, 2.82 mmol) were dissolved in anhydrous THF (10 mL). To this solution was added dropwise 5-benzylthio-1*H*-tetrazole in anhydrous MeCN

(0.30 M, 7.3 mL). The mixture was stirred at RT for 2 h after which saturated aq. NaHCO<sub>3</sub> was added to quench the reaction. The organics were extracted with EtOAc (3 × 10 mL), combined and dried over MgSO<sub>4</sub>. The solvent was removed and the residue applied to a short flash chromatography column (100% EtOAc) to remove 5-benzylthio-1*H*-tetrazole. The regioisomers were dissolved in EtOAc (~200 mg/mL), purified and separated by NP-HPLC (Method B) with retention times of 8 min (**b**), 10.5 min (**b**), 14.5 min (**a**) and 25 min (**a**). The separated title regioisomers **106a** (500 mg, 26%, contaminated with H-phosphonate) and **106b** (1.00 g, 52%) were isolated as mixtures of two diastereomers in the form of colourless foams.

#### Data for **106a**

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.28 (s, 1H, H-C(6)), 7.51-7.21 (m, 9H, DMTr), 6.96-6.72 (m, 5H, DMTr, H-C(5)), 6.16 (2 × d, *J* = 2.9 Hz, 1H, H-C(1')), 5.56 (dd, *J* = 4.9, 3.5 Hz, 0.44H, H-C(2')), 5.52-5.45 (m, 0.56H, H-C(2')), 4.72-4.58 (m, 1H, H-C(3')), 4.37 (t, *J* = 6.4 Hz, 2H, H<sub>2</sub>-C(9)), 4.32-4.27 (m, 0.56H, H-C(4')), 4.22 (H-C(4'))<sup>§</sup>, 3.92-3.35 (m, 12H, ce OCH<sub>2</sub>, OCH<sub>3</sub>, H<sub>2</sub>-C(5'), <sup>i</sup>Pr CH), 2.77 (m, 2H, H<sub>2</sub>-C(10)), 2.64 (q, *J* = 6.4 Hz, 0.86H, ce CH<sub>2</sub>CN), 2.36 (q, *J* = 6.0 Hz, 1.14H, ce CH<sub>2</sub>CN), 2.14, 2.11 (2×s, 3H, CO-CH<sub>3</sub>), 1.34-1.00 (m, 12H, <sup>i</sup>Pr CH<sub>3</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz) δ 169.2, 169.1 (CO-CH<sub>3</sub>), 162.9 (C(4)), 158.8, 158.8 (DMTr), 154.8 (C(2)), 152.1 (C(7)), 144.7 (C(6)), 144.1, 144.0, 135.5, 135.3, 135.1, 130.3, 130.3, 128.5, 128.4, 128.4, 128.1, 128.1, 127.3, 127.3 (DMTr), 117.9, 117.5, 117.1, 116.9, 116.8 (ce CH<sub>2</sub>CN, C(11)), 113.4, 113.4 (DMTr), 95.4 (C(5)), 88.8 (C(1')), 87.2, 87.2 (DMTr-C), 83.2, 82.9 (C(4')), 75.3, 74.8 (C(2')), 69.8 (C(3')), 61.4 (C(5')), 60.1 (C(9)), 58.5, 58.3, 58.3, 58.2, 58.0 (ce OCH<sub>2</sub>), 55.3, 55.3 (OCH<sub>3</sub>), 45.7, 45.6, 45.4, 45.4, 43.4, 43.4, 43.3, 43.2 (<sup>i</sup>Pr CH), 24.7, 24.7, 24.6, 24.6, 24.5, 23.2, 23.1, 23.1, 23.0, 23.0 (<sup>i</sup>Pr CH<sub>3</sub>), 21.1, 20.9 (CO-CH<sub>3</sub>), 20.4, 20.3, 20.2, 20.2, 20.2, 20.1 (ce CH<sub>2</sub>CN), 18.1 (C(10)). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 150.69-150.47 (m), 150.21-149.92 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled) δ 150.62 (s), 150.07 (s). *m/z* ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>45</sub>H<sub>54</sub>N<sub>6</sub>O<sub>11</sub>P<sup>+</sup>, 885.3588; found 885.3617.

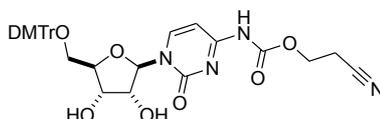
<sup>§</sup> Obscured by H-phosphonate contaminant.

Data for **106b**

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  8.52 (s, 1H, H-C(6)), 8.13 (br. s, 1H, NH), 7.42-7.21 (m, 9H, DMTr), 6.85 (m, 5H, H-C(5), DMTr), 6.14 (d,  $J = 1.8$  Hz, 0.54H, H-C(1')), 6.06 (s, 0.46H, H-C(1')), 5.25-5.06 (m, 1H, H-C(3')), 4.74-4.66 (m, 0.54H, H-C(2')), 4.65-4.56 (m, 0.46H, H-C(2')), 4.42-4.29 (m, 3H,  $\text{H}_2\text{-C}(9)$ , H-C(4')), 4.09-3.50 (m, 11H, ce  $\text{OCH}_2$ ,  $\text{OCH}_3$ ,  $^i\text{Pr CH}$ , H-C(5')), 3.41 (d,  $J = 11.4$  Hz, 1H, H-C(5')), 2.77 (t,  $J = 6.5$  Hz, 2.50H,  $\text{H}_2\text{-C}(10)$ , ce  $\text{CH}_2\text{CN}$ ), 2.64 (dt,  $J = 16.6, 5.9$  Hz, 1.50H, ce  $\text{CH}_2\text{CN}$ ), 2.06 (m, 3H,  $\text{CO-CH}_3$ ), 1.23-1.11 (m, 12H,  $^i\text{Pr CH}_3$ ).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 101 MHz)  $\delta$  169.9, 169.8 ( $\text{CO-CH}_3$ ), 162.2, 162.2 (C(4)), 158.8 (DMTr), 154.7 (C(2)), 151.6 (C7)), 144.9 (C(6)), 144.1, 135.4, 135.3, 130.2, 130.2, 128.3, 128.2, 127.3 (DMTr), 118.2, 117.8 (ce  $\text{CH}_2\text{CN}$ ), 116.5 (C(11)), 113.5 (DMTr), 94.9 (C(5)), 90.7, 90.3 (C(1')), 87.5 (DMTr-C), 80.8, 80.7, 80.5 (C(4')), 75.7, 75.5, 74.5, 74.4 (C(2')), 69.9, 69.5, 69.4 (C(3')), 61.0, 60.7 (C(5')), 60.3 (C(9)), 59.0, 58.8, 58.8, 58.6 (ce  $\text{CH}_2$ ), 55.3 ( $\text{OCH}_3$ ), 43.6, 43.5 ( $^i\text{Pr-CH}$ ), 24.8, 24.8, 24.7, 24.7, 24.6, 24.6 ( $^i\text{Pr-CH}_3$ ), 20.9, 20.9, 20.4, 20.3 ( $\text{CO-CH}_3$ ), 18.2 (C(10), ce  $\text{CH}_2\text{CN}$ ).  $^{31}\text{P}$  NMR (162 MHz,  $\text{CDCl}_3$ )  $\delta = 152.39$ - $152.23$  (m),  $150.38$ - $150.18$  (m).  $^{31}\text{P}$  NMR (162 MHz,  $\text{CDCl}_3$ , decoupled)  $\delta = 152.32$  (s),  $150.29$  (s).  $m/z$  ESI-HRMS (pos.)  $[\text{M}+\text{Na}]^+$  calculated for  $\text{C}_{45}\text{H}_{53}\text{N}_6\text{O}_{11}\text{NaP}^+$ , 907.3402; found 907.3374.

**$N^4$ -[(2-Cyanoethoxy)carbonyl]-5'- $O$ -(4,4'-dimethoxytrityl)cytidine 89**

$\text{C}_{34}\text{H}_{34}\text{N}_4\text{O}_9$ ;  $M_r = 642.66$

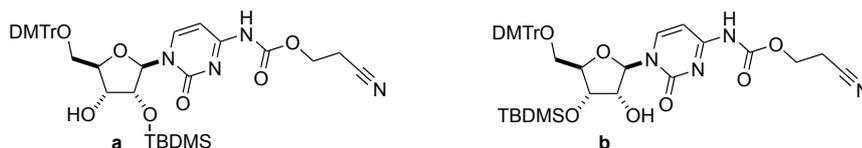


$N^4$ -[(2-Cyanoethoxy)carbonyl]cytidine **79** (9.00 g, 26.4 mmol) was co-evaporated with anhydrous pyridine ( $3 \times 60$  mL). The residue and DMTr-Cl (10.8 g, 31.7 mmol) were dissolved in anhydrous pyridine (90 mL) and stirred at RT for 4 h. MeOH (20 mL) was added and the mixture stirred for a further 30 min and the solvent was removed under vacuum. The residue was taken up in  $\text{CH}_2\text{Cl}_2$  (50 mL) and the organics washed with saturated aq.  $\text{NaHCO}_3$  ( $3 \times 100$  mL), separated and dried over  $\text{MgSO}_4$ . The crude product was co-evaporated with toluene ( $3 \times 50$  mL) and followed by  $\text{CH}_2\text{Cl}_2$  ( $3 \times$

50 mL) to remove residual pyridine. The crude product was purified by flash column chromatography (50:50:2 EtOAc:CH<sub>2</sub>Cl<sub>2</sub>:Et<sub>3</sub>N → 50:45:5:2 EtOAc:CH<sub>2</sub>Cl<sub>2</sub>:MeOH:Et<sub>3</sub>N) to give the title product as a colourless foamy solid (15.0 g, 89%).  $R_f = 0.26$  (50:45:5:2 EtOAc:CH<sub>2</sub>Cl<sub>2</sub>:MeOH:Et<sub>3</sub>N). M.P. = 93-95 °C. IR (cm<sup>-1</sup>) 3270 (br., O-H), 3066, 2965, 2933, 2837, 2255 (wk. CN), 1752 (C=O). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.32 (d,  $J = 7.5$  Hz, 1H, H-C(6)), 7.39-7.16 (m, 9H, DMTr), 6.97 (s, 1H, H-C(5)), 6.82 (dd,  $J = 9.0, 2.3$  Hz, 4H, DMTr), 5.90 (d,  $J = 1.8$  Hz, 1H, H-C(1')), 4.48-4.24 (m, 5H, H-C(2'), H-C(3'), H-C(4'), H<sub>2</sub>-C(9)), 3.77 (s, 6H, OCH<sub>3</sub>), 3.52-3.39 (m, 2H, H<sub>2</sub>-C(5')), 2.74 (t,  $J = 6.3$  Hz, 2H, H<sub>2</sub>-C(10)). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 100 MHz) δ 162.8 (C(4)), 158.7, 158.7 (DMTr), 156.3 (C(2)), 152.0 (C(7)), 144.8 (C(6)), 144.2, 135.6, 135.3, 130.2, 130.1, 128.2, 128.1, 127.2 (DMTr), 116.8 (C(11)), 113.4 (DMTr), 95.6 (C(5)), 92.9 (C(1')), 87.1 (DMTr), 84.9, 76.5, 70.7 (C(2')/C(3')/C(4')), 62.3 (C(5')), 60.3 (C(9)), 55.3 (OCH<sub>3</sub>), 18.1 (C(10)). ESI-HRMS (pos.) [M+Na]<sup>+</sup> calculated for C<sub>34</sub>H<sub>34</sub>N<sub>4</sub>O<sub>9</sub>Na<sup>+</sup>, 665.2218; found 665.2233.

***N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]-2'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)cytidine 109a** and ***N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]-3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)cytidine 109b**

C<sub>40</sub>H<sub>48</sub>N<sub>4</sub>O<sub>9</sub>Si; M<sub>r</sub> = 756.92



***N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]-5'-*O*-(4,4'-dimethoxytrityl)cytidine 89** (15.0 g, 23.3 mmol) and anhydrous pyridine (6.98 mL, 86.3 mmol) were dissolved in anhydrous THF (120 mL). To the solution was added AgNO<sub>3</sub> (4.76 g, 28.0 mmol) and warmed to encourage most of the AgNO<sub>3</sub> to dissolve. Whilst the mixture was still warm TBDMS-Cl (4.57 g, 30.3 mmol) was added upon which a colourless precipitate immediately formed. The reaction mixture was stirred at RT in the dark overnight. The solids were removed by filtration and the supernatant filtered into saturated aq. NaHCO<sub>3</sub> (50 mL). The organics were extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 × 50 mL). The combined organics were dried over MgSO<sub>4</sub> and evaporated to dryness under vacuum. The crude was purified by flash column chromatography (9:1 → 1:1, Et<sub>2</sub>O:EtOAc) to

give the separated isomers as colourless solid foams **109a** (8.12 g, 46%) and **109b** (4.26 g, 24%)

*Data for 109a*

$R_f = 0.38$  (50:50, EtOAc:Et<sub>2</sub>O) M.P. = 116-118 °C. IR (cm<sup>-1</sup>) 3542 (NH) 2951, 2929, 2856, 2255 (wk. CN), 1751 (C=O). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 9.10 (s, 1H, NH), 8.52 (s, 1H, H-C(6)), 7.49-7.21 (m, 9H, DMTr), 6.86 (m, 5H, DMTr, H-C(5)), 5.89 (s, 1H, H-C(1')), 4.45-4.25 (m, 4H, H-C(2'), H-C(3'), H<sub>2</sub>-C(9)), 4.10 (d,  $J = 7.9$  Hz, 1H, H-C(4')), 3.80 (s, 6H, OCH<sub>3</sub>), 3.65-3.50 (m, 2H, H<sub>2</sub>-C(5')), 2.83-2.67 (m, 2H, H<sub>2</sub>-C(10)), 2.43 (d,  $J = 8.1$  Hz, 1H, HO-C(3')), 0.93 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.31 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), 0.19 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 100 MHz) δ 162.6 (C(4)), 158.8 (DMTr), 154.7 (C(2)), 151.9 (C(7)), 145.1 (C(6)), 144.2, 135.6, 135.3, 130.2, 130.2, 128.3, 128.1, 127.2 (DMTr), 116.6 (C(11)), 113.4 (DMTr), 94.8 (C(5)), 90.8 (C(1')), 87.2 (DMTr), 83.1 (C(4')), 76.6 (C(2')/C(3')), 69.1 (C(2')/C(3')), 61.4 (C(5')), 60.1 (C(9)), 55.3 (OCH<sub>3</sub>), 25.9 (SiC(CH<sub>3</sub>)<sub>3</sub>), 18.1 (SiC(CH<sub>3</sub>)<sub>3</sub>/C(10)), 18.1 (SiC(CH<sub>3</sub>)<sub>3</sub>/C(10)), -4.2 (Si(CH<sub>3</sub>)<sub>2</sub>), -5.3 (Si(CH<sub>3</sub>)<sub>2</sub>).  $m/z$  ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>40</sub>H<sub>49</sub>N<sub>4</sub>O<sub>9</sub>Si<sup>+</sup>, 757.3269; found 757.3281.

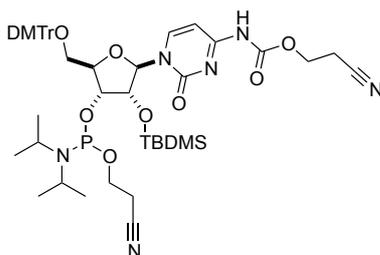
*Data for 109b*

$R_f = 0.15$  (1:1, EtOAc:Et<sub>2</sub>O) M.P. = 159-165 °C. IR (cm<sup>-1</sup>) 2951, 2929, 2855, 2254 (wk. CN), 1751 (C=O). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 9.16 (s, 1H, NH), 8.43 (s, 1H, H-C(6)), 7.43-7.20 (m, 9H, DMTr), 6.85 (d,  $J = 8.8$  Hz, 5H, DMTr, H-C(5)), 6.02 (d,  $J = 2.3$  Hz, 1H, H-C(1')), 4.41-4.26 (m, 3H, H-C(3'), H<sub>2</sub>-C(9)), 4.23-4.11 (m, 2H, H-C(2'), H-C(4')), 3.80 (s, 6H, OCH<sub>3</sub>), 3.70 (dd,  $J = 10.9, 2.0$  Hz, 1H, H-C(5')), 3.31 (m, 2H, H-C(5''), HO-C(2')), 2.77 (t,  $J = 6.5$  Hz, 2H, H<sub>2</sub>-C(10)), 0.81 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.03 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), -0.09 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 100 MHz) δ 162.7 (C(4)), 158.9 (DMTr), 155.0 (C(2)), 152.0 (C(7)), 144.8 (C(6)), 143.9, 135.2, 130.3, 128.5, 128.1, 127.4 (DMTr), 116.7 (C(11)), 113.4 (DMTr), 95.1 (C(5)), 91.4 (C(1')), 87.1 (DMTr), 83.6 (C(2')/C(4')), 76.0 (C(2')/C(4')), 70.7 (C(3')), 61.4 (C(5')), 60.2 (C(9)), 55.3 (OCH<sub>3</sub>), 25.7 (SiC(CH<sub>3</sub>)<sub>3</sub>), 18.1 (SiC(CH<sub>3</sub>)<sub>3</sub>/C(10)), -4.7 (Si(CH<sub>3</sub>)<sub>2</sub>), -4.9 (Si(CH<sub>3</sub>)<sub>2</sub>).  $m/z$  ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>40</sub>H<sub>49</sub>N<sub>4</sub>O<sub>9</sub>Si<sup>+</sup>, 757.3269; found 757.3267.

***N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]-2'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)cytidine-3'-*O*-(2-cyanoethyl-*N,N*-diisopropylphosphoramidite)**

113a

C<sub>49</sub>H<sub>65</sub>N<sub>6</sub>O<sub>10</sub>PSi; M<sub>r</sub> = 957.13

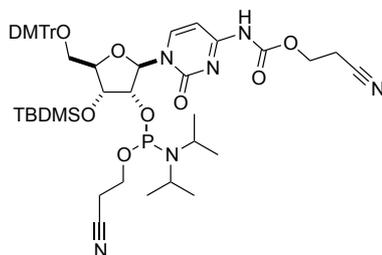


*N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]-2'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)cytidine **109a** (2.00 g, 2.64 mmol) was co-evaporated with anhydrous THF (3 × 20 mL). The residue was dissolved in anhydrous THF (20 mL) and to this solution was added DMAP (64.5 mg, 5.28 × 10<sup>-4</sup> mol) and *N,N*-diisopropylethylamine (1.84 mL, 10.6 mmol). Finally 2-cyanoethyl *N,N*-diisopropyl phosphoamidochloridite (0.88 mL, 3.96 mmol) was added dropwise and the resultant mixture stirred at RT for 16 h. The reaction was quenched with anhydrous MeOH (5 mL) and the solvent removed under vacuum. The crude product was purified by flash column chromatography (30:70:2 → 80:20:2, EtOAc:*c*-Hex:Et<sub>3</sub>N) to give the title compound as a colourless foam (2.03 g, 81%). R<sub>f</sub> = 0.38, 0.30 (80:20:2, EtOAc:*c*-Hex:Et<sub>3</sub>N). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.70-8.11 (m, 2H, H-C(6), NH), 7.50-7.21 (m, 9H, DMTr), 6.92-6.59 (m, 5H, DMTr, H-C(5)), 5.92 (d, *J* = 2.3 Hz, 0.32H, H-C(1')), 5.82 (s, 0.68H, H-C(1')), 4.45-4.21 (m, 5H, H-C(2'), H-C(3'), H-C(4'), H<sub>2</sub>-C(9)), 3.92-3.38 (m, 12H, OCH<sub>3</sub>, H<sub>2</sub>-C(5'), <sup>i</sup>Pr CH, ce OCH<sub>2</sub>), 2.76 (t, *J* = 6.5 Hz, 2H, H<sub>2</sub>-C(10)), 2.59 (t, *J* = 6.3 Hz, 0.68H, ce CH<sub>2</sub>CN), 2.41 (t, *J* = 6.4 Hz, 1.32H, ce CH<sub>2</sub>CN), 1.31-0.97 (m, 12H, <sup>i</sup>Pr CH<sub>3</sub>), 0.94-0.86 (m, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.25 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), 0.17-0.09 (m, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz) δ 162.5, 162.3 (C(4)), 158.8 (DMTr), 154.7 (C(2)), 151.9 (C(7)), 145.2 (C(6)), 144.2, 144.1, 135.5, 135.4, 135.2, 130.4, 130.3, 128.5, 128.0, 127.3 (DMTr), 117.6, 117.5, 116.6 (CN), 113.4, 113.3 (DMTr), 94.7 (C(5)), 91.5 (C(1')), 87.3, 87.2 (DMTr), 81.7, 81.5, 75.9, 75.3, 71.5, 69.6 ((C(2')/C(3')/C(4')), 61.6, 61.0 ((C(5')/ce CH<sub>2</sub>), 60.1 (C(9)), 58.4, 58.3, 58.2 (C(5')/ce

CH<sub>2</sub>), 55.3, 55.3 (OCH<sub>3</sub>), 45.5, 45.4, 43.4, 43.2, 43.1 (<sup>i</sup>Pr CH), 26.0, 25.9 (SiC(CH<sub>3</sub>)<sub>3</sub>), 25.0, 24.9, 24.9, 24.8, 24.7, 24.6, 23.1, 23.0 (<sup>i</sup>Pr CH<sub>3</sub>), 20.6, 20.5, 20.3, 20.3, 20.2 (ce CH<sub>2</sub>), 18.2, 18.1 (C(10), SiC(CH<sub>3</sub>)<sub>3</sub>), -4.2, -4.3, -4.9, -5.0 (Si(CH<sub>3</sub>)<sub>2</sub>). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 150.39-150.16 (m), 149.19-148.97 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled) δ 150.30 (s), 149.10 (s). *m/z* ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>49</sub>H<sub>66</sub>N<sub>6</sub>O<sub>10</sub>PSi<sup>+</sup>, 957.4347; found 957.4380.

***N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]-3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)cytidine-2'-*O*-(2-cyanoethyl-*N,N*-diisopropylphosphoramidite)**  
113b

C<sub>49</sub>H<sub>65</sub>N<sub>6</sub>O<sub>10</sub>PSi; M<sub>r</sub> = 957.13

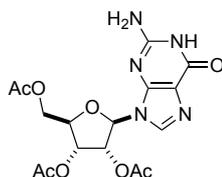


*N*<sup>4</sup>-[(2-Cyanoethoxy)carbonyl]-3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)cytidine **109b** (1.00 g, 1.32 mmol) was co-evaporated with anhydrous THF (3 × 10 mL). The residue was suspended in anhydrous THF (6 mL) and anhydrous CH<sub>2</sub>Cl<sub>2</sub> (4 mL), *N,N*-diisopropylethylamine (0.92 mL, 5.28 mmol) and DMAP (32.2 mg, 0.26 mmol) were added to the suspension. Finally, 2-cyanoethyl *N,N*-diisopropyl phosphoramidochloridite (0.44 mL, 1.98 mmol) was added dropwise and the resultant mixture stirred at RT for 4 h. The reaction was quenched with anhydrous MeOH (5 mL) and stirred for a further 5 min after which the mixture was evaporated to dryness. The crude product was purified by flash column chromatography (30:70:2 → 80:20:2, EtOAc:*c*-Hex:Et<sub>3</sub>N) to give the title compound as a colourless foam (1.13 g, 89%). R<sub>f</sub> = 0.31 (80:20:2, EtOAc:*c*-Hex:Et<sub>3</sub>N). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.68 (br. s, 1H, H-C(6)), 7.46-7.17 (m, 9H, DMTr), 6.86-6.65 (m, 5H, DMTr, H-C(5)), 6.17-6.13 (m, 1H, H-C(1')), 4.36-4.24 (m, 4H, H-C(2'), H-C(3'), H<sub>2</sub>-C(9)), 4.21-4.09 (m, 1H, H-C(4')), 4.08-3.93 (m, 1H, ce OCH<sub>2</sub>), 3.87-3.70 (m, 8H, OCH<sub>3</sub>, H-C(5'), ce OCH<sub>2</sub>), 3.69-3.57 (m, 2H, <sup>i</sup>Pr CH), 3.35-3.29 (m, 1H, H-C(5'')), 2.83-2.54 (m, 4H, H<sub>2</sub>-C(10), ce CH<sub>2</sub>CN), 1.21-1.01 (m, 12H, <sup>i</sup>Pr CH<sub>3</sub>), 0.73, 0.72 (2 × s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.03, -0.02 (2 × s, 3H,

Si(CH<sub>3</sub>)<sub>2</sub>), -0.09, -0.10 (2 × s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz) δ 162.4, 162.3 (C(4)), 158.7, 158.7 (DMTr), 154.4 (C(2)), 151.9 (C(7)), 144.9 (C(6)), 143.6, 143.6, 135.0, 135.0, 130.3, 130.2, 130.2, 130.2, 128.5, 128.5, 127.9, 127.3, 127.2 (DMTr), 118.3, 117.9 (ce CH<sub>2</sub>CN), 117.1, 116.6, 116.6 ((C(11))), 113.2 (DMTr), 94.9 (C(5)), 90.0 (C(1')), 87.1, 87.1 (DMTr-C), 82.3, 82.1 (C(4')), 76.0, 75.9 (C(2')), 69.0, 68.9 (C(3')), 60.6, 60.5 (C(5')), 60.0 (C(9)), 58.6, 58.4, 58.3, 58.2, 58.2, 57.9 (ce OCH<sub>2</sub>), 55.2 (OCH<sub>3</sub>), 43.5, 43.3, 43.2, 43.0 (<sup>i</sup>Pr CH), 25.7, 25.6, 25.5 (SiC(CH<sub>3</sub>)<sub>3</sub>), 24.7, 24.7, 24.6, 24.6, 24.6, 24.5, 24.4 (<sup>i</sup>Pr CH<sub>3</sub>), 20.1, 20.1, 20.0 (ce CH<sub>2</sub>CN), 18.0, 17.9, 17.9, 17.9 (C(10), SiC(CH<sub>3</sub>)<sub>3</sub>), -4.3, -4.3, -5.2 (Si(CH<sub>3</sub>)<sub>2</sub>). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 151.42-151.03 (m), 148.04-147.85 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled) δ 151.23 (s), 147.97 (s). *m/z* ESI-HRMS (pos.) [M+Na]<sup>+</sup> calculated for C<sub>49</sub>H<sub>65</sub>N<sub>6</sub>O<sub>10</sub>PSiNa<sup>+</sup>, 979.4161; found 979.4132.

### 2',3',5'-Tri-acetyl-guanosine 86

C<sub>16</sub>H<sub>19</sub>N<sub>5</sub>O<sub>8</sub>; M<sub>r</sub> = 409.35

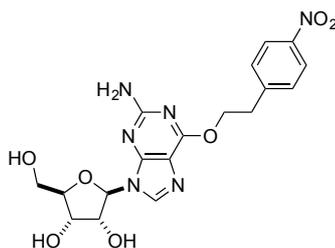


Guanosine (20.0 g, 70.4 mmol) was suspended in anhydrous acetonitrile (200 mL). To this mixture was added DMAP (647 mg, 5.29 mmol), triethylamine (39 mL, 0.28 mol) and finally dropwise acetic anhydride (24 mL, 0.25 mol). The resultant mixture was stirred for 30 min, quenched with MeOH (50 mL) and the solvent removed under vacuum. The resultant oil was treated with propan-2-ol (400 mL) and a solid precipitated. The precipitate was collected by filtration and washed with cold propan-2-ol (3 × 50 mL). The solid was air dried and then dried under high vacuum to give the title compound as a colourless amorphous solid (26.6 g, 92%). M.P. = 223-226 °C (lit.<sup>[327]</sup> = 224-227 °C). IR (cm<sup>-1</sup>) 3465 (H-N(2)), 3300 (NH<sub>2</sub>), 2727 (CH), 1770, 1745 (C=O), 1698, 1630, 1607, 1571. <sup>1</sup>H NMR (400 MHz, MeOD) δ = 7.84 (s, 1H, H-C(8)), 6.05 (d, *J* = 5.0 Hz, 1H, H-C(1')), 5.92 (t, *J* = 5.4 Hz, 1H, H-C(2')), 5.66 (app. t, *J* = 5.1 Hz, 1H, H-C(3')), 4.45-4.34 (m, 3H, H-C(4'), H<sub>2</sub>-C(5')), 2.12 (s, 3H, CO-CH<sub>3</sub>), 2.07-2.06 (m, 6H, CO-CH<sub>3</sub>). <sup>13</sup>C NMR (100 MHz, MeOD) δ 172.3, 171.5, 171.2 (CO-

CH<sub>3</sub>), 159.3 (C(6)), 155.4 (C(2)), 152.8 (C(4)), 138.2 (C(8)), 118.2 (C(5)), 87.8 (C(1')), 81.4 (C(4')), 74.3 (C(2')), 72.0 (C(3')), 64.2 (C(5')), 20.6, 20.4, 20.3 (CO-CH<sub>3</sub>). *m/z* ESI-: 408.1 ([M-H]<sup>-</sup>, 60%); ESI+: 410.1 ([M+H]<sup>+</sup>, 100%); ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>16</sub>H<sub>20</sub>N<sub>5</sub>O<sub>8</sub>, 410.1312; found 410.1314.

### ***O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]guanosine 83**

C<sub>18</sub>H<sub>20</sub>N<sub>6</sub>O<sub>7</sub>; M<sub>r</sub> = 432.39

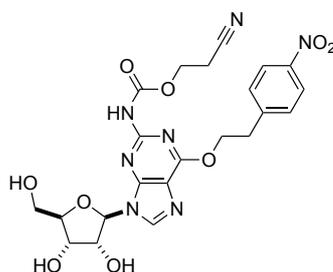


All solid reagents were dried over P<sub>2</sub>O<sub>5</sub> under vacuum for 24 h. 2',3',5'-Triacetylguanosine **86** (10.8 g, 26.3 mmol) was suspended in anhydrous dioxane (100 mL). *p*-Nitrophenylethanol (5.27 g, 31.5 mmol) and triphenylphosphine (8.27 g, 31.5 mmol) were added to this mixture and the resultant mixture heated at 80 °C for 45 min. Diisopropyl azodicarboxylate (6.20 mL, 31.5 mmol) was added dropwise upon which the solution began to boil and then the solution was stirred at 60 °C for a further 1 h. The solution was cooled to RT and evaporated to dryness under vacuum to give an oil, from which the 2',3',5'-tri-acetyl-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]guanosine was isolated by flash column chromatography (60:35:5, EtOAc:*n*-Hex:MeOH, R<sub>f</sub> = 0.25). The product which contained triphenylphosphine oxide was taken up in MeOH (200 mL) and cooled to 0 °C. To the solution was added saturated aq. NH<sub>3</sub> (200 mL) and the solution stirred in a sealed vessel at RT overnight. The solution was degassed then the solvent was removed under vacuum to give an orange oil. The oil was taken up in MeOH (~80 mL) and on concentration, by evaporation under vacuum, a yellow solid precipitated. The mixture was cooled at 4 °C overnight, the resultant solid precipitate collected by filtration and washed with cold MeOH (3 × 20 mL). The solid contained acetamide and so was suspended in H<sub>2</sub>O (100 mL) and heated at 90 °C for 15 min. Once cooled to RT, the insoluble material was collected by filtration and washed with H<sub>2</sub>O (3 × 20 mL). The solid was air dried then dried under high vacuum to give the title compound over two steps as a yellow amorphous solid (8.45 g, 76%). R<sub>f</sub> = 0.50 (9:1, CH<sub>2</sub>Cl<sub>2</sub>:MeOH).

M.P. = 173-174 °C. IR (cm<sup>-1</sup>) 3509 (H-bonded O-H), 3386 (NH), 3110 (br., O-H), 2934 (CH) 1612 (Ar), 1589, 1504 (Ar). <sup>1</sup>H NMR (400 MHz, DMSO-d<sup>6</sup>) δ 8.18 (d, *J* = 8.2 Hz, 2H, H-C(14)), 8.09 (s, 1H, H-C(8)), 7.64 (d, *J* = 8.2 Hz, 2H, H-C(13)), 6.44 (s, 2H, NH), 5.78 (d, *J* = 5.9 Hz, 1H, H-C(1')), 5.37 (d, *J* = 6.0 Hz, 1H, 2'-OH), 5.12-5.06 (m, 2H, 3'-OH and 5'-OH), 4.68 (t, *J* = 6.8 Hz, 2H, H<sub>2</sub>-C(10)), 4.46 (q, *J* = 5.7 Hz, 1H, H-C(2')), 4.10 (q, *J* = 4.3 Hz, 1H, H-C(3')), 3.89 (q, *J* = 3.8 Hz, 1H, H-C(4')), 3.65-3.50 (m, 2H, H<sub>2</sub>-C(5')), 3.25 (t, *J* = 6.8 Hz, 2H, H<sub>2</sub>-C(11)). <sup>13</sup>C NMR (100 MHz, DMSO-d<sup>6</sup>) δ 160.1 (C(6)), 159.7 (C(2)), 154.3 (C(4)), 146.7 (C(12)), 146.3 (C(15)), 138.1 (C(8)), 130.3 (C(13)), 123.4 (C(14)), 113.8 (C(5)), 86.6 (C(1')), 85.3 (C(4')), 73.5 (C(2')), 70.4 (C(3')), 65.5 (C(10)), 61.4 (C(5')), 34.4 (C(11)). *m/z* ESI<sup>-</sup>: 467.1 ([M+Cl]<sup>-</sup>, 55%), 477.1 ([M+H COO]<sup>-</sup>, 100%); ESI<sup>+</sup>: 433.2 ([M+H]<sup>+</sup>, 100%); ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>18</sub>H<sub>21</sub>N<sub>6</sub>O<sub>7</sub>, 433.1472; found 433.1462.

### *N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]guanosine **84**

C<sub>22</sub>H<sub>23</sub>N<sub>7</sub>O<sub>9</sub>; M<sub>r</sub> = 529.46

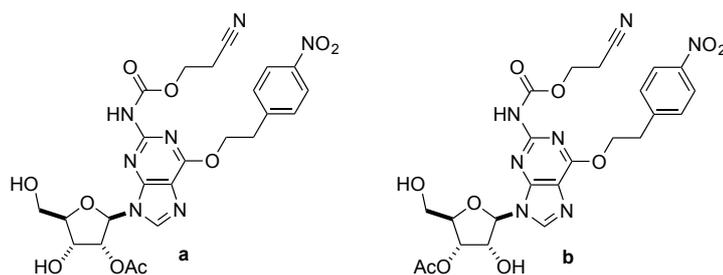


*O*<sup>6</sup>-[2-(4-Nitrophenyl)ethyl]guanosine **83** (4.32 g, 10.0 mmol) was co-evaporated with anhydrous pyridine (3 × 20 mL). The residue was dissolved in anhydrous pyridine (40 mL) and anhydrous CH<sub>2</sub>Cl<sub>2</sub> (55 mL). To the solution was added dropwise Me<sub>3</sub>Si-Cl (7.61 mL, 60.0 mmol) and the resultant mixture was stirred at RT for 20 min. 2-Cyanoethyl carbonochloridate **75** (2.00 g, 15.0 mmol) diluted in anhydrous CH<sub>2</sub>Cl<sub>2</sub> (10 mL) was added dropwise and the mixture stirred for a further 3 h. MeOH (30 mL) was added to quench the reaction and remove the Me<sub>3</sub>Si groups. The solvent was removed under vacuum and the oil co-evaporated with 1:1 MeOH:Tol. (3 × 30 mL). The residue was taken up in MeOH (15 mL) and H<sub>2</sub>O was added dropwise until precipitation of a solid began, the solution was kept at 4 °C overnight to afford a slightly pink precipitate that was collected by filtration and washed with cold MeOH (3 × 20 mL). The solid contained pyridinium.HCl and this was removed by boiling the

solid as a suspension in H<sub>2</sub>O (50 mL) for 15 min. After the suspension was cooled to RT the solid was collected, washed with H<sub>2</sub>O (3 × 20 mL) and dried under high vacuum to yield the title compound as a slightly off-white amorphous solid (5.08 g, 96%). M.P. = 173-177 °C. IR (cm<sup>-1</sup>) 3386 (NH), 3330 (O-H), 3120, 2960, 2911 (CH), 2866 (CH), 1747 (C=O). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 10.54 (s, 1H, NH), 8.43 (s, 1H, H-C(8)), 8.18 (d, *J* = 8.7 Hz, 2H, H-C(14)), 7.66 (d, *J* = 8.7 Hz, 2H, H-C(13)), 5.89 (d, *J* = 5.9 Hz, 1H, H-C(1')), 5.45 (d, *J* = 5.9 Hz, 1H, 2'-OH), 5.16 (d, *J* = 4.7 Hz, 1H, 3'-OH), 4.93 (t, *J* = 5.5 Hz, 1H, 5'-OH), 4.79 (t, *J* = 6.9 Hz, 2H, H<sub>2</sub>-C(10)), 4.61 (q, *J* = 5.7 Hz, 1H, H-C(2')), 4.31 (t, *J* = 6.0 Hz, 2H, H<sub>2</sub>-C(18)), 4.19 (td, *J* = 4.8 Hz, 3.3, 1H, H-C(3')), 3.92 (q, *J* = 4.2 Hz, 1H, H-C(4')), 3.65 (ABX, *J*<sub>AB</sub> = 11.8, *J*<sub>AX</sub> = 4.9 Hz, 1H, H-C(5')), 3.54 (ABX, *J*<sub>BA</sub> = 11.8, *J*<sub>BX</sub> = 4.9 Hz, 1H, H-C(5'')), 3.34-3.31 (m, 2H, H<sub>2</sub>-C(11)<sup>h</sup>), 2.94 (t, *J* = 6.0 Hz, 2H, H<sub>2</sub>-C(19)). <sup>13</sup>C NMR (100 MHz, DMSO-d<sub>6</sub>) δ 159.7 (C(6)), 153.1 (C(4)), 151.9 (C(2)), 151.5 (C(16)), 146.4 (C(12)), 146.3 (C(15)), 141.4 (C(8)), 130.4 (C(13)), 123.4 (C(14)), 118.6 (C(20)), 117.3 (C(5)), 87.1 (C(1')), 85.7 (C(4')), 73.4 (C(2')), 70.4 (C(3')), 66.4 (C(10)), 61.4 (C(5')), 59.4 (C(18)), 34.2 (C(11)), 17.7 (C(19)). *m/z* ESI+: 530.2 ([M+H]<sup>+</sup>, 60%); ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>22</sub>H<sub>24</sub>N<sub>7</sub>O<sub>9</sub>, 530.1636; found 530.1620.

***N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-2'/3'-*O*-(acetyl)guanosine 99a+99b**

C<sub>24</sub>H<sub>25</sub>N<sub>7</sub>O<sub>10</sub>; M<sub>r</sub> = 571.50



*N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-guanosine **84** (3.00 g, 5.67 mmol) was suspended in anhydrous dioxane (60 mL) and stirred at 60 °C for 20 min. Trimethyl orthoacetate (2.14 mL, 17.0 mmol) and trifluoroacetic acid (43 μL, 0.57 mmol) were added and the mixture stirred at 60 °C for 4 h. Full conversion to the

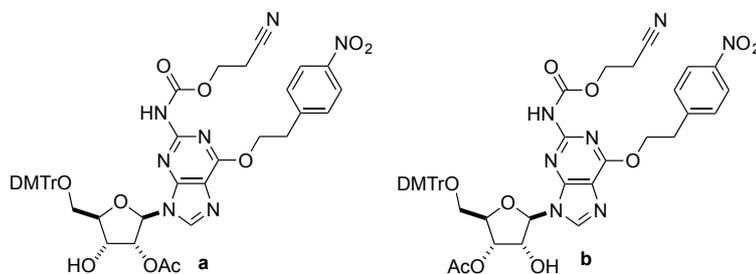
<sup>h</sup> Peak overlaps with HOD peak.

orthoester was checked by TLC before it was hydrolysed by addition of H<sub>2</sub>O (30 mL) and the mixture was stirred for a further 15 min. The mixture was evaporated to dryness under vacuum and the resultant oil purified as a mixture of regioisomers by flash column chromatography (92:8, CH<sub>2</sub>Cl<sub>2</sub>:MeOH). The title compounds were isolated as a regioisomeric mixture in the form of a colourless foam in quantitative yield (3.23 g, >99%). (*Note*: The products were isolated in a ratio of *ca.* 1.5:1 **b**:**a**, calculated by integrations of both H-C(1') of **99a+99b**). *R<sub>f</sub>* = 0.53 and 0.43 (92:8, CH<sub>2</sub>Cl<sub>2</sub>:MeOH). IR (cm<sup>-1</sup>) 3289 (O-H), 3113 (NH), 2937 (CH), 1738 (C=O), 1595 (Ar), 1515 (Ar). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 10.58-10.57 (m, 1H, NH, **a+b**), 8.46-8.45 (m, 1H, H-C(8), **a+b**), 8.18 (d, *J* = 8.6 Hz, 2H, H-C(14), **a+b**), 7.66 (d, *J* = 8.6 Hz, 2H, H-C(13), **a+b**), 6.14 (d, *J* = 5.9 Hz, 0.40H, H-C(1'), **a**), 5.89 (d, *J* = 7.0 Hz, 0.60H, H-C(1'), **b**), 5.81 (d, *J* = 6.1 Hz, 0.60H, 2'-OH, **b**), 5.59-5.55 (m, 0.80H, H-C(2'), 3'-OH, **a**), 5.31 (dd, *J* = 5.4, 2.3 Hz, 0.60H, H-C(3'), **b**), 5.12 (t, *J* = 5.6 Hz, 0.60H, 5'-OH, **b**), 5.02 (t, *J* = 5.4 Hz, 0.40H, 5'-OH, **a**), 4.91 (q, *J* = 6.2 Hz, 0.6H, H-C(2'), **b**), 4.82-4.77 (m, 2H, H<sub>2</sub>-C(10), **a+b**), 4.53 (td, *J* = 5.3 Hz, 3.7, 0.40H, H-C(3'), **a**), 4.31 (t, *J* = 6.0 Hz, 2H, H<sub>2</sub>-C(18), **a+b**), 4.10 (q, *J* = 4.1 Hz, 0.60H, H-C(4'), **b**), 3.97 (q, *J* = 4.1 Hz, 0.40H, H-C(4'), **a**), 3.73-3.56 (m, 2H, H<sub>2</sub>-C(5'), **a+b**), 3.34-3.31 (H<sub>2</sub>-C(11)<sup>i</sup>), 2.95 (t, *J* = 6.0 Hz, 2H, H<sub>2</sub>-C(19), **a+b**), 2.12 (s, 1.80H, CO-CH<sub>3</sub>, **b**), 2.03 (s, 1.20H, CO-CH<sub>3</sub>, **a**). <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>) δ 169.6 (CO-CH<sub>3</sub>, **a+b**), 159.8 (C(6), **a+b**), 153.2 (C(4), **b**), 152.9 (C(4), **a**), 152.0 (C(2), **a, b**), 151.5 (C(16), **a, b**), 146.5, 146.4, 146.3 ((C(12), **a+b**), (C(15), **a+b**)), 141.4 (C(8), **a**), 141.2 (C(8), **b**), 130.4 (C(13), **a+b**), 123.4 (C(14), **a+b**), 118.6 (C(20), **a+b**), 117.3 (C(5), **a+b**), 86.9 (C(1'), **b**), 86.1 (C(4'), **a**), 84.8 (C(1'), **a**), 83.5 (C(4'), **b**), 75.3 (C(2'), **a**), 73.2 (C(3'), **b**), 71.8 (C(2'), **b**), 68.8 (C(3'), **a**), 66.5 (C(10), **a+b**), 61.3, 61.3 (C(5'), **a+b**), 59.5, 59.3 (C(18), **a+b**), 34.3, 34.2 (C(11), **a+b**), 20.8 (CO-CH<sub>3</sub>, **b**), 20.6 (CO-CH<sub>3</sub>, **a**), 17.8, 17.7 (C(19), **a+b**). *m/z* ESI<sup>+</sup>: 572.3 ([M+H]<sup>+</sup>, 70%); ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>24</sub>H<sub>26</sub>N<sub>7</sub>O<sub>10</sub>, 572.1741; found 572.1757.

<sup>i</sup> Peak obscured by HOD peak, assignment made by 2D-COSY.

***N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-2'/3'-*O*-(acetyl)-5'-*O*-(4,4'-dimethoxytrityl)guanosine 95a+95b**

C<sub>45</sub>H<sub>43</sub>N<sub>7</sub>O<sub>12</sub>; M<sub>r</sub> = 873.86

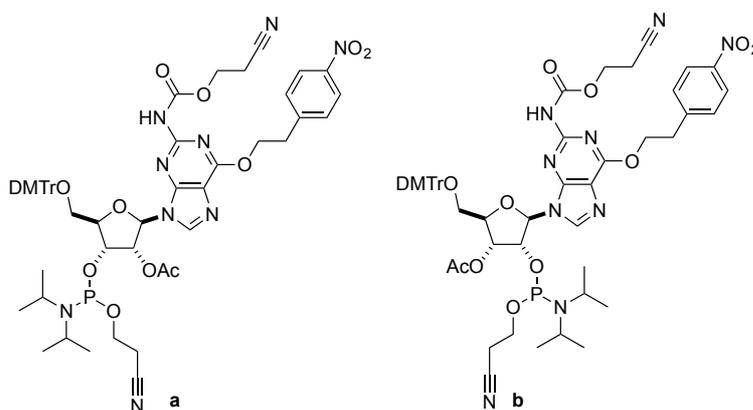


*N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-2'/3'-*O*-monoacetyl-guanosine **99a+99b** (3.23 g, 5.66 mmol) mixture was co-evaporated with anhydrous pyridine (3 × 20 mL) and the residue taken up in anhydrous pyridine (20 mL). DMTr-Cl (2.88 g, 8.49 mmol) was added and the reaction stirred at RT for 3 h. MeOH (20 mL) was added and the mixture stirred for a further 15 min before the solvent was removed under vacuum. The residue was taken up in CH<sub>2</sub>Cl<sub>2</sub> (20 mL) and the organic layer washed with saturated aq. NaHCO<sub>3</sub> (3 × 40 mL). The organics were separated, dried over MgSO<sub>4</sub>, and after evaporation the residue was co-evaporated with toluene (3 × 20 mL) followed by CH<sub>2</sub>Cl<sub>2</sub> (3 × 20 mL). The crude product was purified by flash column chromatography (30:70:2 → 70:30:2, EtOAc:Tol:Et<sub>3</sub>N) to give the title compounds as a colourless foam (3.64 g, 74%). (*Note*: the isomers were isolated as a mixture in a ratio of *ca.* 5:1, **b**:**a**, calculated by integrations of both H-C(1') of **95a+95b**). *R*<sub>f</sub> = 0.30 (70:30:2, EtOAc:Tol:Et<sub>3</sub>N). IR (cm<sup>-1</sup>) 3364 (O-H), 2933, 2837 (CH), 2360 (CN), 1742 (C=O), 1606 (Ar), 1508 (Ar). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.17-8.13 (m, 2H, H-C(14), **a+b**), 8.08 (s, 0.83H, H-C(8), **b**), 7.94 (s, 0.17H, H-C(8), **a**), 7.64 (s, 1H, NH), 7.51-7.46 (m, 2H, H-C(13), **a+b**), 7.39-7.09 (m, 9H, DMTr, **a+b**), 6.77-6.68 (m, 4H, DMTr, **a+b**), 6.42 (s, 0.75H, 2'-OH, **b**), 6.09 (d, *J* = 4.0 Hz, 0.17H, H-C(1'), **a**), 5.92-5.89 (m, 1H, (H-C(1'), **b**), (H-C(2'), **a**)), 5.48 (d, *J* = 5.5 Hz, 0.83H, H-C(3'), **b**), 5.18 (t, *J* = 5.3 Hz, 0.83H, H-C(2'), **b**), 5.11 (t, *J* = 5.6 Hz, 0.17H, H-C(3'), **a**), 4.78 (t, *J* = 6.8 Hz, 2H, H<sub>2</sub>-C(10), **a+b**), 4.47-4.32 (m, 2.83H, (H-C(4'), **b**), (H<sub>2</sub>-C(18))), 4.19 (q, *J* = 4.5 Hz, 0.17H, H-C(4'), **a**), 3.75, 3.74 (2 × s, 6H, OCH<sub>3</sub>, **a+b**), 3.48-3.23 (m, 4H, (H<sub>2</sub>-C(5'), **a+b**), (H<sub>2</sub>-C(11), **a+b**)), 2.78 (t, *J* = 6.1 Hz, 1.67H, H<sub>2</sub>-C(19), **b**), 2.69-2.66 (m, 0.33H, H<sub>2</sub>-C(19), **a**), 2.18 (s, 2.50H, CO-CH<sub>3</sub>, **b**), 2.15 (s,

0.50H, CO-CH<sub>3</sub>, **a**). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz) δ 170.5 (CO-CH<sub>3</sub>, **b**), 170.2 (CO-CH<sub>3</sub>, **a**), 161.0, 160.8 (C(6), **a+b**), 158.6 (DMTr), 152.7 (C(4), **a**), 151.8 ((C(4), **b**), (C(2), **b**)), 151.4 (C(2), **a**) 150.9 (C(16), **a**), 150.4 (C(16), **b**), 147.0 (C(15), **a+b**), 145.8 (C(12), **a**), 145.6 (C(12), **b**), 144.6, 144.2 (DMTr), 141.0 (C(8), **a**), 140.5 (C(8), **b**), 135.8, 135.4, 135.3 (DMTr), 130.2, 130.1, 130.0 ((DMTr), (C(13), **a+b**)), 128.3, 128.0, 127.9, 126.9 (DMTr), 123.9 (C(14), **a+b**), 118.5 (C(5), **a+b**), 116.8 (C(20), **a+b**), 113.2 (DMTr), 91.6 (C(1'), **b**), 86.9 (DMTr-C), 86.6 (C(1'), **a**), 85.4 (C(4'), **b**), 83.7 (C(4'), **a**), 75.8 (C(2'), **a**), 75.2, 75.1 (C(2'), C(3'), **b**), 70.1 (C(3'), **a**), 67.2 (C(10), **a+b**), 63.7 (C(5'), **b**), 63.3 (C(5'), **a**), 60.1 (C(18), **b**), 59.7 (C(18), **a**), 55.3 (OCH<sub>3</sub>), 35.1 (C(11), **a+b**), 21.2 (CO-CH<sub>3</sub>, **b**), 20.8 (CO-CH<sub>3</sub>, **a**), 18.4, 18.3 (C(19), **a+b**). *m/z* ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>45</sub>H<sub>44</sub>N<sub>7</sub>O<sub>12</sub>, 874.3048; found 874.3030.

***N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-2'-*O*-(acetyl)-5'-*O*-(4,4'-dimethoxytrityl)guanosine-3'-*O*-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite 104a and *N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-3'-*O*-(acetyl)-5'-*O*-(4,4'-dimethoxytrityl)guanosine-2'-*O*-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite 104b**

C<sub>54</sub>H<sub>60</sub>N<sub>9</sub>O<sub>13</sub>P; M<sub>r</sub> = 1074.08



***N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-5'-*O*-(4,4'-dimethoxytrityl)-2'/3'-*O*-monoacetyl-guanosine **95a+95b** (1.00 g, 1.14 mmol) was dissolved in anhydrous THF (6 mL) and stirred with 3Å molecular sieves for 20 min. To the solution was added 2-cyanoethyl *N,N,N',N'*-tetraisopropyl phosphoroamidite (0.73 mL, 2.29 mmol) then 5-benzylthio-1*H*-tetrazole (220 mg, 1.14 mmol) as a solution in anhydrous MeCN (4 mL) and the mixture was stirred for 1 h. The molecular sieves**

were removed by filtration and the supernatant was added with stirring to saturated aq. NaHCO<sub>3</sub> (30 mL) from which, the organics were extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 × 15 mL). The combined organic layers were dried over MgSO<sub>4</sub> and evaporated to dryness. The residue was purified by flash column chromatography (60:40:2, EtOAc: Pentane: Et<sub>3</sub>N, R<sub>f</sub> = 0.5) to give the purified mixture of regioisomers. The regioisomers were dissolved in EtOAc (~100 mg/mL) and separated by NP-HPLC (Method C), with retention times of 10 min (**b**), 15.5 min (**a**) and 21 min (**a**). The separated title regioisomers **104a** (227 mg, 18%) and **104b** (818 mg, 67%), were isolated as mixtures of two diastereomers in the form of colourless foams.

#### *Data for 104a*

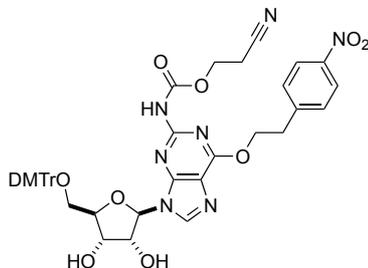
<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.16 (d, *J* = 8.6 Hz, 2H, H-C(14)), 7.96, 7.95 (2xs, 1H, H-C(8)), 7.52 (2xd, *J* = 8.6 Hz, 2H, H-C(13)), 7.47-7.15 (m, 9H, DMTr), 6.77 (d, *J* = 8.8 Hz, 4H, DMTr), 6.16 (d, *J* = 6.5 Hz, 0.40H, H-C(1')), 6.09 (d, *J* = 6.2 Hz, 0.60H, H-C(1')), 6.04 (t, *J* = 5.7 Hz, 0.45H, H-C(2')), 5.95 (t, *J* = 5.9 Hz, 0.55H, H-C(2')), 4.90-4.78 (m, 3H, H-C(3'), H<sub>2</sub>-C(10)), 4.41-4.26 (m, 3H, H-C(4'), H<sub>2</sub>-C(18)), 3.97-3.25 (m, 14H, ce OCH<sub>2</sub>, OCH<sub>3</sub>, <sup>i</sup>Pr CH, H<sub>2</sub>-C(5'), H<sub>2</sub>-C(11)), 2.76-2.62 (m, 3H, H<sub>2</sub>-C(19), ce CH<sub>2</sub>CN), 2.31 (t, *J* = 6.4 Hz, 1H, ce CH<sub>2</sub>CN), 2.09 (m, 3H, CO-CH<sub>3</sub>), 1.21-1.02 (m, 12H, <sup>i</sup>Pr CH<sub>3</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz) δ 169.8, 169.8 (CO-CH<sub>3</sub>), 160.8 (C(6)), 158.7 (DMTr), 153.1, 152.9 (C(4)), 151.6 (C(2)), 150.5, 150.4 (C(16)), 147.0 (C(15)), 145.9 (C(12)), 144.5, 144.3 (DMTr), 141.0, 140.7 (C(8)), 135.8, 135.7, 135.6, 135.6 (DMTr), 130.3, 130.2, 130.2, 130.2 (C(13), DMTr), 128.5, 128.3, 128.0, 128.0, 127.1 (DMTr), 123.9 (C(14)), 118.8, 118.7 (C(5)), 117.7, 117.4, 116.9 (CN), 113.3, 113.2 (DMTr), 86.8, 86.7 (DMTr-C), 86.3, 85.8 (C(1')), 84.8, 84.5, 84.5 (C(4')), 74.3, 74.1, 74.0 (C(2')), 71.4, 71.3, 70.9, 70.7 (C(3')), 67.1 (C(10)), 63.4, 63.3 (C(5')), 59.6, 59.5 (C(18)), 59.0, 58.8, 58.1, 57.8 (ce OCH<sub>2</sub>), 55.4 (OCH<sub>3</sub>), 43.5, 43.4, 43.2 (<sup>i</sup>Pr CH), 35.2 (C(11)), 24.8, 24.8, 24.7, 24.7, 24.6 (<sup>i</sup>Pr CH<sub>3</sub>), 21.1, 20.9 (CO-CH<sub>3</sub>), 20.5, 20.4, 20.2, 20.1 (ce CH<sub>2</sub>CN), 18.3 (C(19)). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 150.86 (m), 150.42-150.08 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled) δ 150.90 (s), 150.28 (s). *m/z* ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>54</sub>H<sub>61</sub>N<sub>9</sub>O<sub>13</sub>P, 1074.4126; found 1074.4131.

*Data for 104b*

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  8.17 (2  $\times$  d,  $J = 8.7$  Hz, 2H, H-C(14)), 8.02 (s, 0.5H, H-C(8)), 7.97 (s, 0.5H, H-C(8)), 7.53 (d,  $J = 8.7$  Hz, 2H, H-C(13)), 7.40-7.16 (m, 9H, DMTr), 6.79-6.74 (m, 4H, DMTr), 6.10 (d,  $J = 5.5$  Hz, 0.5H, H-C(1')), 6.02 (d,  $J = 5.9$  Hz, 0.5H, H-C(1')), 5.63-5.60 (m, 1H, H-C(3')), 5.37 (dt,  $J = 10.9, 5.6$  Hz, 0.5H, H-C(2')), 5.19 (dt,  $J = 10.6, 5.5$  Hz, 0.5H, H-C(2')), 4.87-4.82 (m, 2H, H<sub>2</sub>-C(10)), 4.34 (t,  $J = 6.2$  Hz, 2H, H<sub>2</sub>-C(18)), 4.29-4.25 (m, 1H, H-C(4')), 3.84-3.66 (m, 7H, OCH<sub>3</sub>, ce OCH<sub>2</sub>), 3.58-3.41 (m, 5H, H<sub>2</sub>-C(5'), ce OCH<sub>2</sub>, <sup>*i*</sup>Pr CH), 3.34 (t,  $J = 6.9$  Hz, 2H, H<sub>2</sub>-C(11)), 2.75-2.70 (m, 2H, H<sub>2</sub>-C(19)), 2.58 (t,  $J = 6.3$  Hz, 1H, ce CH<sub>2</sub>CN), 2.38-2.25 (m, 1H, ce CH<sub>2</sub>CN), 2.14-2.14, 2.11 (2  $\times$  s, 3H, CO-CH<sub>3</sub>), 1.14-1.08 (m, 9H, <sup>*i*</sup>Pr CH<sub>3</sub>), 0.90 (d,  $J = 6.8$  Hz, 3H, <sup>*i*</sup>Pr CH<sub>3</sub>).  $^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  169.9 (CO-CH<sub>3</sub>), 160.8 (C(6)), 158.7 (DMTr), 153.1, 152.9 (C(4)), 151.5, 151.5 (C(2)), 150.4, 150.3 (C(16)), 147.0 (C(12)), 145.8 (C(15)), 144.7, 144.6 (DMTr), 141.0, 140.9 (C(8)), 135.8, 135.7, 135.7, 135.6 (DMTr), 130.2, 130.1 (C(13), DMTr), 128.2, 128.2, 128.0, 128.0, 127.9, 127.1, 127.0 (DMTr), 123.8 (C(14)), 118.8, 118.6 (C(5)), 117.6, 117.3 (ce CN), 116.9 (C(20)), 113.3, 113.2 (DMTr), 87.9, 87.8 (C(1')), 86.8, 86.7 (DMTr-C), 82.6, 82.1 (C(4')), 74.4, 74.2, 73.4, 73.2 (C(2')), 72.4, 72.3, 72.3 (C(3')), 67.1 (C(10)), 63.4, 63.3 (C(5')), 59.6, 59.6 (C(18)), 58.6, 58.4, 58.1, 58.0 (ce OCH<sub>2</sub>), 55.3 (OCH<sub>3</sub>), 43.5, 43.4, 43.4, 43.3 (<sup>*i*</sup>Pr CH), 35.2 (C(11)), 24.8, 24.7, 24.6, 24.6, 24.4, 24.4 (<sup>*i*</sup>Pr CH<sub>3</sub>), 21.1, 21.0 (CO-CH<sub>3</sub>), 20.3, 20.3, 20.0, 19.9 (ce CH<sub>2</sub>CN), 18.3, 18.3 (C(19)).  $^{31}\text{P}$  NMR (162 MHz,  $\text{CDCl}_3$ )  $\delta$  151.84-151.48 (m), 151.17-150.87 (m).  $^{31}\text{P}$  NMR (162 MHz,  $\text{CDCl}_3$ , decoupled)  $\delta$  151.62, 151.01.  $m/z$  ESI-HRMS (pos.)  $[\text{M}+\text{H}]^+$  calculated for C<sub>54</sub>H<sub>61</sub>N<sub>9</sub>O<sub>13</sub>P, 1074.4126; found 1074.4169.

***N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-5'-*O*-(4,4'-dimethoxytrityl)guanosine 90**

C<sub>43</sub>H<sub>41</sub>N<sub>7</sub>O<sub>11</sub>; M<sub>r</sub> = 831.83

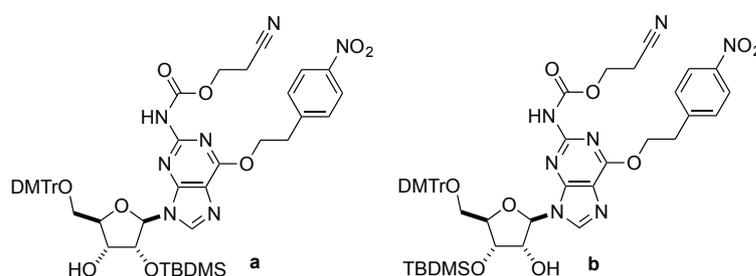


*N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]guanosine **84** (2.97 g, 5.61 mmol) was co-evaporated with anhydrous pyridine (3 × 25 mL). The residue and DMTr-Cl (2.28 g, 6.73 mmol) were dissolved in anhydrous pyridine (40 mL) and stirred at RT overnight. MeOH (10 mL) was added and the reaction mixture stirred for 15 min, before the solvent was removed under vacuum. The residue was taken up in CH<sub>2</sub>Cl<sub>2</sub> (30 mL) and the organic layer washed with saturated aq. NaHCO<sub>3</sub> (3 × 40 mL). The organic layer was separated and then dried over MgSO<sub>4</sub>, filtered and evaporated to dryness. The residue was co-evaporated with toluene (3 × 20 mL) and followed by CH<sub>2</sub>Cl<sub>2</sub> (3 × 20 mL). The resultant crude solid was purified by flash column chromatography (50:50:2, EtOAc:Tol:Et<sub>3</sub>N → 50:45:5:2, EtOAc:Tol:MeOH:Et<sub>3</sub>N) to give the title compound as an off-white foam (4.27 g, 91%). R<sub>f</sub> = 0.30 (70:30:2 EtOAc:Tol.:Et<sub>3</sub>N). M.P. = 103-106 °C. IR (cm<sup>-1</sup>) 3364 (NH), 2932 (CH), 2912 (CH), 2360 (CN), 1748 (C=O), 1606 (Ar), 1507 (Ar). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.15-8.12 (m, 3H, H-C(14), H-C(8)), 7.81 (s, 1H, NH), 7.47 (d, *J* = 8.7 Hz, 2H, H-C(13)), 7.18-7.08 (m, 9H, DMTr), 6.99 (br. s, 1H, 2'-OH), 6.69 (2×d, *J* = 8.8 Hz, 4H, DMTr), 5.90 (d, *J* = 6.2 Hz, 1H, H-C(1')), 4.93 (t, *J* = 5.7 Hz, 1H, H-C(2')), 4.79 (t, *J* = 6.7 Hz, 2H, H<sub>2</sub>-C(10)), 4.49-4.38 (m, 4H, H-C(3'), H-C(4'), H<sub>2</sub>-C(18)), 3.74 (s, 6H, OCH<sub>3</sub>), 3.38 (ABX, *J*<sub>AB</sub> = 10.6, *J*<sub>AX</sub> = 3.2 Hz, 1H, H-C(5')), 3.29 (t, *J* = 6.8 Hz, 2H, H<sub>2</sub>-C(11)), 3.17 (ABX, *J*<sub>BA</sub> = 10.6, *J*<sub>BX</sub> = 3.2 Hz, 1H, H-C(5'')), 2.77 (t, *J* = 6.1 Hz, 2H, H<sub>2</sub>-C(19)). <sup>1</sup>H NMR (100 MHz, CDCl<sub>3</sub>) δ 161.0 (C(6)), 158.6 (DMTr), 151.8 (C(4)), 151.1 (C(2)), 150.7 (C(16)), 147.0 (C(15)), 145.6 (C(12)), 144.3 (DMTr), 140.3 (C(8)), 135.5, 135.3 (DMTr), 130.0, (DMTr, C(13)), 128.0, 127.8, 126.9 (DMTr), 123.9 (C(14)), 118.6 (C(5)), 116.7 (C(20)), 113.2 (DMTr), 92.1 (C(1')), 87.1 (C(4')), 86.7 (DMTr-C), 76.7 (C(2')) 74.0

(C(3')), 67.2 (C(10)), 63.9 (C(5')), 60.1 (C(18)), 55.3 (OCH<sub>3</sub>), 35.1 (C(11)), 18.4 (C(19)). *m/z* ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>43</sub>H<sub>42</sub>N<sub>7</sub>O<sub>11</sub>, 832.2942; found 832.2941.

*N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-2'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)guanosine **110a** and *N*<sup>2</sup>-[(2-cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)guanosine **110b**

C<sub>49</sub>H<sub>55</sub>N<sub>7</sub>O<sub>11</sub>Si; M<sub>r</sub> = 946.09



*N*<sup>2</sup>-Cyanoethoxycarbonyl-*O*<sup>6</sup>-nitrophenylethyl-5'-*O*-(4,4'-dimethoxytrityl)guanosine **90** (3.20 g, 3.93 mmol) was co-evaporated with anhydrous THF (2 × 30 mL). The residue was dissolved in anhydrous THF (40 mL), to which anhydrous pyridine (1.59 mL, 19.6 mmol) and AgNO<sub>3</sub> (1.00 g, 5.89 mmol) were added. The mixture was stirred for 10 min and warmed gently to dissolve most of the AgNO<sub>3</sub>. TBDMS-Cl (1.00 g, 6.67 mmol) was added upon which a white precipitate formed immediately and the mixture was stirred in the dark at RT for 5 h. The solids were removed by filtration and the supernatant separated directly into saturated aq. NaHCO<sub>3</sub> (50 mL). The organics were extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 × 50 mL), the combined organic layers were dried over MgSO<sub>4</sub> and finally evaporated to dryness under vacuum. The mixture of 2'/3'-*O*-*tert*-butyldimethylsilyl nucleosides were purified as a mixture by flash column chromatography (9:1, Et<sub>2</sub>O:EtOAc). The purified regioisomers were dissolved in EtOAc (~300 mg/mL) and separated by NP-HPLC (Method D), with retention times of 17.8 min (**a**) and 30 min (**b**). The separated regioisomers **110a** (1.64 g, 44%) and **110b** (1.29 g, 35%) were isolated as colourless foams.

### Data for **110a**

$R_f = 0.43$  (9:1, Et<sub>2</sub>O:EtOAc) M.P. = 72-75 °C. IR (cm<sup>-1</sup>) 2953, 2927, 2855 (CH), 2253 (wk. CN), 1760 (C=O), 1606 (Ar), 1509 (Ar). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.16 (d,  $J = 8.6$  Hz, 2H, H-C(14)), 7.99 (s, 1H, H-C(8)), 7.53 (d,  $J = 8.6$  Hz, 2H, H-C(13)), 7.44 (d,  $J = 7.0$  Hz, 2H, DMTr), 7.33 (2×d,  $J = 8.9$  Hz, 4H, DMTr), 7.25-7.17 (m, 3H, DMTr), 6.78 (2×d,  $J = 8.7$  Hz, 4H, DMTr), 5.93 (d,  $J = 5.7$  Hz, 1H, H-C(1')), 5.03 (t,  $J = 5.4$  Hz, 1H, H-C(2')), 4.84 (t,  $J = 6.6$  Hz, 2H, H<sub>2</sub>-C(10)), 4.42 (q,  $J = 3.5$  Hz, 1H, H-C(3')), 4.30 (t,  $J = 6.2$  Hz, 2H, H<sub>2</sub>-C(18)), 4.23 (q,  $J = 3.1$  Hz, 1H, H-C(4')), 3.77 (m, 6H, OCH<sub>3</sub>), 3.50 (ABX,  $J_{AB} = 10.6$ ,  $J_{AX} = 2.6$  Hz, 1H, H-C(5')), 3.40-3.32 (m, 3H, H-C(5''), H<sub>2</sub>-C(11)), 2.72 (d,  $J = 3.7$  Hz, 1H, 3'-OH), 2.66 (t,  $J = 6.2$  Hz, 2H, H<sub>2</sub>-C(19)), 0.84 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.00 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), -0.18 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 160.9 (C(6)), 158.7 (DMTr), 153.0 (C(4)), 151.5 (C(2)), 150.4 (C(16)), 147.0 (C(15)), 145.8 (C(12)), 144.8 (DMTr), 141.0 (C(8)), 135.9, 135.8 (DMTr), 130.2 (C(13) and DMTr), 128.2, 128.0, 127.1 (DMTr), 123.9 (C(14)), 118.8 (C(5)), 116.8 (C(20)), 113.3 (DMTr), 88.5 (C(1')), 86.7 (DMTr-C), 84.5 (C(4')), 75.4 (C(2')), 71.6 (C(3')), 67.1 (C(10)), 63.8 (C(5')), 59.6 (C(18)), 55.4 (2 × OCH<sub>3</sub>), 35.2 (C(11)), 25.7 (SiC(CH<sub>3</sub>)<sub>3</sub>), 18.3 (C(19)), 18.0 (SiC(CH<sub>3</sub>)<sub>3</sub>), -4.9, -5.0 (Si(CH<sub>3</sub>)<sub>2</sub>).  $m/z$  ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>49</sub>H<sub>56</sub>N<sub>7</sub>O<sub>11</sub>Si, 946.3807; found 946.3785.

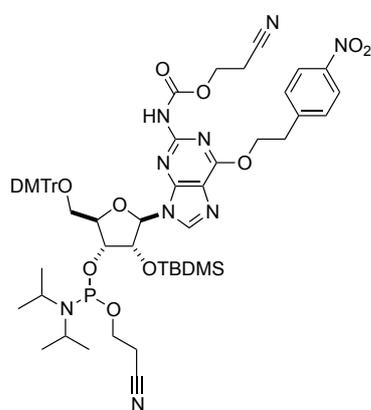
### Data for **110b**

$R_f = 0.33$  (9:1, Et<sub>2</sub>O:EtOAc) M.P. = 65-67 °C. IR (cm<sup>-1</sup>) 2953 (CH<sub>2</sub>/CH<sub>3</sub>), 2926 (CH<sub>2</sub>/CH<sub>3</sub>), 2854 (CH), 2359 (wk. CN), 1755 (C=O), 1607, 1509 (Ar). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.15 (d,  $J = 8.7$  Hz, 2H, H-C(14)), 8.05 (s, 1H, H-C(8)), 7.50 (d,  $J = 8.7$  Hz, 2H, H-C(13)), 7.46 (s, 1H, NH), 7.34-7.32 (m, 2H, DMTr), 7.24-7.14 (m, 7H, DMTr), 6.74 (d,  $J = 8.6$  Hz, 4H, DMTr), 5.93 (d,  $J = 5.2$  Hz, 1H, H-C(1')), 4.80 (t,  $J = 6.8$  Hz, 2H, H<sub>2</sub>-C(10)), 4.69 (q,  $J = 5.3$  Hz, 1H, H-C(2')), 4.56 (dd,  $J = 5.2, 3.3$  Hz, 1H, H-C(3')), 4.42-4.33 (m, 2H, H<sub>2</sub>-C(18)), 4.23-4.15 (m, 2H, H-C(4'), 2'-OH), 3.76 (s, 6H, OCH<sub>3</sub>), 3.43-3.21 (m, 4H, H<sub>2</sub>-C(5'), H-C(11)), 2.75 (t,  $J = 6.2$  Hz, 2H H-C(19)), 0.89 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.11 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), 0.04 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 160.8 (C(6)), 158.7 (DMTr), 152.5 (C(4)), 151.1 (C(2)), 150.5 (C(16)), 147.0 (C(15)), 145.8 (C(12)), 144.5 (DMTr), 140.9 (C(8)), 135.8, 135.7 (DMTr), 130.1 (C(13) and DMTr), 128.2, 127.9, 127.0 (DMTr), 123.9 (C(14)), 118.8 (C(5)), 116.8 (C(20)),

113.2 (DMTr), 90.3 (C(1')), 86.7 (DMTr-C), 85.9 (C(4')), 75.2 (C(2')), 73.1 (C(3')), 67.1 (C(10)), 63.5 (C(5')), 59.8 (C(18)), 55.3 (OCH<sub>3</sub>), 35.2 (C(11)), 25.9 (SiC(CH<sub>3</sub>)<sub>3</sub>), 18.37 (C(19)), 18.27 (SiC(CH<sub>3</sub>)<sub>3</sub>), -4.56, -4.71 (Si(CH<sub>3</sub>)<sub>2</sub>). *m/z* ESI-HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>49</sub>H<sub>56</sub>N<sub>7</sub>O<sub>11</sub>Si, 946.3807; found 946.3787.

***N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-2'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)guanosine-3'-*O*-(2-cyanoethyl-*N,N*-diisopropylphosphoramidite) 114a**

C<sub>58</sub>H<sub>72</sub>N<sub>9</sub>O<sub>12</sub>SiP; M<sub>r</sub> = 1146.30

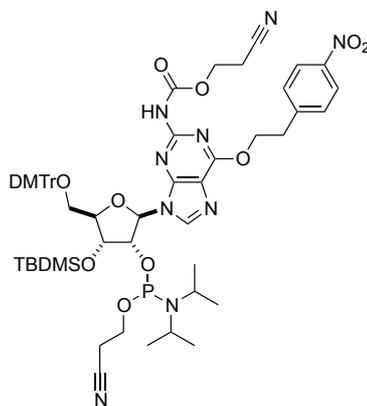


*N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-2'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)guanosine **110a** (1.00 g, 1.06 mmol) was co-evaporated with anhydrous THF (3 × 10 mL). The residue, DMAP (26.0 mg, 0.22 mmol) and *N,N*-diisopropylethylamine (0.74 mL, 4.24 mmol) were dissolved in anhydrous THF (10 mL). 2-Cyanoethyl *N,N*-diisopropyl phosphoamidochloridite (0.35 mL, 1.59 mmol) was added dropwise to the solution and the resultant mixture stirred at RT for 3 h. The reaction was quenched with anhydrous MeOH (5 mL) and the solvent removed under vacuum. The residue was purified by flash column chromatography (30:70:1 → 60:40:1, EtOAc:*c*-Hex:Et<sub>3</sub>N) to give the title compound (1.02 g, 84%) as a mixture of diastereoisomers as a colourless foam. *R*<sub>f</sub> = 0.43 (60:40:1, EtOAc:*c*-Hex:Et<sub>3</sub>N). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.16 (d, *J* = 8.7 Hz, 2H, H-C(14)), 8.04, 8.02 (2 × s, 1H, H-C(8)), 7.58-7.30 (m, 9H, H-C(13), DMTr), 7.30-7.17 (m, 2H, DMTr), 7.09 (s, 1H, NH), 6.80 (m, 4H, DMTr), 6.01 (d, *J* = 7.2 Hz, 0.40H, H-C(1')), 5.86 (d, *J* = 6.9 Hz, 0.60H, H-C(1')), 5.12 (dd, *J* = 6.9, 5.2 Hz, 0.60H, H-C(2')), 5.00 (dd, *J* = 7.2, 4.5 Hz, 0.40H, H-C(2')), 4.86 (t, *J* = 6.8 Hz, 2H, H<sub>2</sub>-C(10)), 4.43-4.19 (m, 4H, H-C(3')),

H-C(4'), H<sub>2</sub>-C(18)), 3.98-3.87 (m, 0.47H, ce OCH<sub>2</sub>), 3.98-3.87 (m, 0.53H, ce OCH<sub>2</sub>), 3.81-3.75 (m, 6H, OCH<sub>3</sub>), 3.68-3.46 (m, 4H, <sup>i</sup>Pr CH, H-C(5'), ce OCH<sub>2</sub>), 3.39-3.22 (m, 3H, H<sub>2</sub>-C(11), H-C(5'')), 2.34-2.16 (m, 3H, H<sub>2</sub>-C(19), ce CH<sub>2</sub>CN), 2.34-2.16 (m, 1H, ce CH<sub>2</sub>CN), 1.23-1.13 (m, 9H, <sup>i</sup>Pr CH<sub>3</sub>), 1.01 (d, *J* = 6.7 Hz, 3H, <sup>i</sup>Pr CH<sub>3</sub>), 0.75 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), -0.01, -0.06 (2 × s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), -0.25, -0.26 (2 × s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 160.8, 160.8 (C(6)), 158.8, 158.5 (DMTr), 153.3, 153.1 (C(4)), 151.6, 151.4 (C(2)), 150.5, 150.4 (C(16)), 147.0, 146.8 (C(15)), 145.9, 145.7 (C(12)), 144.8, 144.6 (DMTr), 141.4, 140.6 (C(8)), 136.0, 135.8, 135.7, 135.6 (DMTr), 130.2, 130.2, 130.1, 130.0, 129.9 (DMTr, C(13)), 128.3, 128.2, 128.1, 128.0, 127.2, 126.9 (DMTr), 123.8, 123.6 (C(14)), 118.9, 118.6 (C(5)), 118.0, 117.4, 116.9, 116.8 (CN), 113.4, 113.3, 113.2, 113.1 (DMTr), 88.4, 87.5 (C(1')), 86.9, 86.6 (DMTr-C), 84.7, 84.3 (C(4')), 75.8, 74.1, 74.0 (C(2')), 73.6, 73.5, 72.7, 72.6 (C(3')), 67.1, 67.0 (C(10)), 63.6, 63.5 (C(5')), 59.5, 59.5, 59.7 (ce OCH<sub>2</sub>, C(18)), 59.1, 57.8, 57.6 (ce OCH<sub>2</sub>), 55.4, 55.1 (OCH<sub>3</sub>), 43.6, 43.5, 43.1, 43.0 (<sup>i</sup>Pr CH), 35.2, 35.0 (C(11)), 25.7, 25.7, 25.5, 25.4 (SiC(CH<sub>3</sub>)<sub>3</sub>), 24.9, 24.9, 24.8, 24.8, 24.7, 24.5, 24.5 (<sup>i</sup>Pr CH<sub>3</sub>), 20.6, 20.5, 20.2, 20.1 (ce CH<sub>2</sub>CN), 18.3, 18.2, 18.1, 18.0 (ce CH<sub>2</sub>CN, SiC(CH<sub>3</sub>)<sub>3</sub>), -4.5, -4.6, -4.6, -5.1, -5.3 (Si(CH<sub>3</sub>)<sub>2</sub>). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 151.32-150.91 (m), 149.25-148.86 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled) δ 151.07 (s), 149.03 (s). *m/z* ESI- HRMS (pos.) [M+H]<sup>+</sup> calculated for C<sub>58</sub>H<sub>73</sub>N<sub>9</sub>O<sub>12</sub>SiP, 1146.4886; found 1146.4889.

***N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)guanosine-2'-*O*-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite 114b**

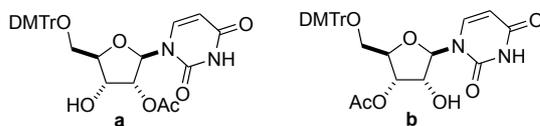
C<sub>58</sub>H<sub>72</sub>N<sub>9</sub>O<sub>12</sub>PSi; M<sub>r</sub> = 1146.30



*N*<sup>2</sup>-[(2-Cyanoethoxy)carbonyl]-*O*<sup>6</sup>-[2-(4-nitrophenyl)ethyl]-3'-*O*-*tert*-butyldimethylsilyl-5'-*O*-(4,4'-dimethoxytrityl)guanosine **110b** (1.00 g, 1.06 mmol) was co-evaporated with anhydrous THF (3 × 10 mL). The residue, DMAP (26.0 mg, 0.22 mmol) and *N,N*-diisopropylethylamine (0.74 mL, 4.24 mmol) were dissolved in anhydrous THF (10 mL). 2-Cyanoethyl *N,N*-diisopropyl phosphoramidochloridite (0.35 mL, 1.59 mmol) was added dropwise to the solution and the resultant mixture stirred at RT for 3 h. The reaction was quenched with anhydrous MeOH (5 mL) and the solvent removed under vacuum. The residue was purified by flash column chromatography (30:70:2 → 50:50:2, EtOAc:*c*-Hex:Et<sub>3</sub>N) to give the title compound (939 mg, 77%) as a mixture of diastereoisomers as a colourless foam. *R*<sub>f</sub> = 0.29 (50:50:2, EtOAc:*c*-Hex:Et<sub>3</sub>N). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.15 (d, *J* = 7.6 Hz, 2H, H-C(14)), 8.07, 8.05 (2 × s, 1H, H-C(8)), 7.52 (d, *J* = 8.6 Hz, 2H, H-C(13)), 7.45-7.17 (m, 9H, DMTr), 6.78 (d, *J* = 8.6 Hz, 4H, DMTr), 6.14 (d, *J* = 5.2 Hz, 0.51H, H-C(1')), 6.04 (d, *J* = 5.6 Hz, 0.49H, H-C(1')), 5.08 (dt, *J* = 10.8, 5.1 Hz, 0.50H, H-C(2')), 4.86-4.76 (m, 2.50H, H<sub>2</sub>-C(10), H-C(2')), 4.47 (t, *J* = 4.1 Hz, 1H, H-C(3')), 4.36 (q, *J* = 6.0 Hz, 2H, H<sub>2</sub>-C(18)), 4.17-4.13 (m, 1H, H-C(4')), 3.84-3.45 (m, 11H, ce OCH<sub>2</sub>, OCH<sub>3</sub>, <sup>*i*</sup>Pr CH, H-C(5')), 3.35-3.27 (m, 3H, H<sub>2</sub>-C(11), H-C(5'')), 2.74 (q, *J* = 6.2 Hz, 2H, H<sub>2</sub>-C(19)), 2.51 (t, *J* = 6.3 Hz, 1H, ce CH<sub>2</sub>CN), 2.40-2.32 (m, 1H, ce CH<sub>2</sub>CN), 1.14-1.03 (m, 9H, <sup>*i*</sup>Pr CH<sub>3</sub>), 0.91-0.85 (m, 12H, <sup>*i*</sup>Pr CH<sub>3</sub>, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.11-0.07 (2 × s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), 0.00--0.01 (2 × s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz) δ 160.7 (C(6)), 158.6 (DMTr), 153.1, 153.0 (C(4)), 151.4 (C(2)), 150.5, 150.4 (C(16)), 147.0 (C(15)), 145.9 (C(12)), 144.7, 144.7 (DMTr), 141.2, 141.1 (C(8)), 135.9, 135.8, 135.8, 135.8 (DMTr), 130.2 (DMTr, C(13)), 128.3, 128.0, 127.0 (DMTr), 123.8 (C(14)), 118.8, 118.7 (C(5)), 117.6, 117.6 (ce CN), 116.9 (C(20)), 113.3 (DMTr), 87.7, 87.6, 87.6 (C(1')), 86.8, 86.7 (DMTr-C), 85.3, 84.8 (C(4')), 75.8, 75.7, 75.4, 75.2 (C(2')), 72.2, 71.9 (C(3')), 67.1 (C(10)), 63.5, 63.3 (C(5')), 59.6, 59.6 (C(18)), 58.3, 58.1, 58.0, 57.8 (ce OCH<sub>2</sub>), 55.3 (OCH<sub>3</sub>), 43.4, 43.3, 43.3, 43.1 (<sup>*i*</sup>Pr CH), 35.2 (C(11)), 25.9 (SiC(CH<sub>3</sub>)<sub>3</sub>), 24.8, 24.8, 24.7, 24.6, 24.4, 24.3 (<sup>*i*</sup>Pr CH<sub>3</sub>), 20.4, 20.4, 20.1, 20.0 (ce CH<sub>2</sub>CN), 18.4 (C(19)), 18.2, 18.2 (SiC(CH<sub>3</sub>)<sub>3</sub>), -4.3, -4.7, -4.8 (Si(CH<sub>3</sub>)<sub>2</sub>). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 150.44-149.75 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled) δ 150.24 (s), 149.97 (s). *m/z* ESI-HRMS (pos.) [M+Na]<sup>+</sup> calculated for C<sub>58</sub>H<sub>72</sub>N<sub>9</sub>O<sub>12</sub>NaSiP, 1168.4700; found 1168.4669.

## 2'/3'-O-(Acetyl)-5'-O-(4,4'-dimethoxytrityl)uridine 92a+92b‡

C<sub>32</sub>H<sub>32</sub>N<sub>2</sub>O<sub>9</sub>; M<sub>r</sub> = 588.60

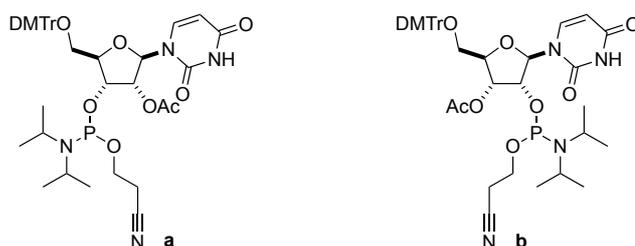


To a solution of commercially available 5'-O-(4,4'-dimethoxytrityl)uridine **91** (3.00 g, 5.49 mmol) in anhydrous THF (20 mL) was added anhydrous pyridine (0.44 mL, 5.49 mmol), followed by acetyl chloride (0.39 mL, 5.49 mmol) at 0 °C. The mixture was warmed to RT and stirred for 3 h, then quenched by addition of saturated aq. NaHCO<sub>3</sub>. The organics were extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 × 20 mL), the combined organic layers were dried over MgSO<sub>4</sub> and finally concentrated under vacuum. The residue was purified by flash column chromatography (50:50:1:1, EtOAc:Tol.:MeOH:Et<sub>3</sub>N) to give the title compounds (2.20 g, 68%) as a regioisomeric mixture and as a colourless foam. (*Note*: the isomers were isolated as a mixture in a ratio of *ca.* 2.4:1, **b**:**a** calculated by integrations of both H-C(1') of **92a+92b**). R<sub>f</sub> = 0.32 (60:40:1 CH<sub>2</sub>Cl<sub>2</sub>:EtOAc:Et<sub>3</sub>N). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 9.40 (s, 0.7H, NH, **b**), 8.66 (d, *J* = 1.4 Hz, 0.3H, NH, **a**), 7.84 (d, *J* = 8.1 Hz, 0.7H, H-C(6), **b**), 7.79 (d, *J* = 8.2 Hz, 0.3H, H-C(6), **a**), 7.44-7.18 (m, 9H, DMTr, **a+b**), 6.84 (d, *J* = 8.8 Hz, 4H, DMTr, **a+b**), 6.14 (d, *J* = 4.5 Hz, 0.3H, H-C(1'), **a**), 6.00 (d, *J* = 4.9 Hz, 0.7H, H-C(1'), **b**), 5.45-5.33 (m, 1.3H, [H-C(5), **a+b**], [H-C(2'), **a**]), 5.26 (t, *J* = 4.9 Hz, 0.7H, H-C(3'), **b**), 4.59 (q, *J* = 4.8 Hz, 0.3H, H-C(3'), **a**), 4.53 (q, *J* = 5.4 Hz, 0.7H, H-C(2'), **b**), 4.28 (d, *J* = 4.4 Hz, 0.7H, H-C(4'), **b**), 4.14 (dt, *J* = 4.9, 2.3 Hz, 0.3H, H-C(4'), **a**), 3.91 (d, *J* = 6.5 Hz, 0.7H, 2'-OH, **b**), 3.79 (s, 6H, OCH<sub>3</sub>, **a+b**), 3.55 (app. dd, *J* = 11.0, 2.3 Hz, 1H, H-C(5'), **a+b**), 3.49-3.43 (m, 1H, H-C(5''), **a+b**), 2.48 (d, *J* = 3.9 Hz, 0.3H, 3'-OH, **a**), 2.18, 2.14 (2 × s, 3H, CO-CH<sub>3</sub>, **a+b**). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz): δ 170.7, 170.5 (CO-CH<sub>3</sub>), 163.7, 163.4 (C(4)), 158.8, 158.7 (DMTr), 151.2 (C(2), **b**), 150.6 (C(2), **a**), 147.4, 144.2 (DMTr), 140.2, 140.0, 139.6 (C(6), **a+b**), 135.3, 135.2, 135.1, 130.3, 130.2, 130.2, 129.2, 129.1, 128.3, 128.3, 128.2, 128.2, 127.9, 127.9, 127.3, 127.2, 113.4, 113.2 (DMTr), 102.8, 102.8 (C(5), **a+b**), 89.5 (C(1'), **b**), 87.4, 87.3 (DMTr-C), 86.8 (C(1'), **a**), 83.7 (C(4'), **a**), 81.7, 81.5 (C(4'), **b**), 76.0, 75.8 (C(2'), **a**), 74.1 (C(2'), **b**), 71.9 (C(3'), **b**), 69.9 (C(3'), **a**), 62.4

(C(5'), **a**), 62.2 (C(5'), **a**), 55.4 (OCH<sub>3</sub>, **a+b**), 21.6 (CO-CH<sub>3</sub>, **a**), 20.9 (CO-CH<sub>3</sub>, **b**). *m/z* ESI-HRMS (pos.) [M+Na]<sup>+</sup> calculated for C<sub>32</sub>H<sub>32</sub>N<sub>2</sub>O<sub>9</sub>Na, 611.2000; found 611.1979.

**2'-O-(Acetyl)-5'-O-(4,4'-dimethoxytrityl)uridine-3'-O-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite** **107a** and **3'-O-(acetyl)-5'-O-(4,4'-dimethoxytrityl)uridine-2'-O-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite** **107b**‡

C<sub>41</sub>H<sub>49</sub>N<sub>4</sub>O<sub>10</sub>P; M<sub>r</sub> = 788.82



2'/3'-O-Acetyl-5'-O-(4,4'-dimethoxytrityl)uridine **92a+92b** (1.10 g, 1.87 mmol) was dissolved in anhydrous THF (6 mL). To this solution was added 2-cyanoethyl *N,N,N',N'*-tetraisopropyl phosphoramidite (1.20 mL, 3.74 mmol), followed by slow addition of a solution of 5-benzylthio-1*H*-tetrazole in anhydrous CH<sub>3</sub>CN (0.35 M, 5.40 mL). The mixture was stirred at RT for 2 h and quenched by addition of saturated aq. NaHCO<sub>3</sub> (6 mL). The organics were extracted with EtOAc (3 × 10 mL), the combined organic layers were dried over MgSO<sub>4</sub> and finally concentrated under vacuum. The residue was purified by a short flash column chromatography (100% EtOAc) to remove 5-benzylthio-1*H*-tetrazole. The regioisomers were dissolved in EtOAc (~200 mg/mL), and separated by NP-HPLC (Method E), with retention times of 9 min (**b**), 10.5 min (**b**), 13 min (**a**) and 19 min (**a**). The separated title regioisomers were isolated as mixtures of two diastereomers in the form of colourless foams, **107a** (440 mg, 30%, contained H-phosphonate) and **107b** (750 mg, 48%).

#### Data for **107af**

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.28 (s, 1H, NH), 7.81-7.69 (m, 1H, H-C(6)), 7.45-7.20 (m, 9H, DMTr), 6.90-6.78 (m, 4H, DMTr), 6.24-6.12 (m, 1H, H-C(1')), 5.53 (t, *J* = 5.3 Hz, 0.4H, H-C(2')), 5.38 (t, *J* = 5.7 Hz, 0.6H, H-C(2')), 5.36-5.27 (m, 1H, H-C(5)), 4.76-4.61 (m, 1H, H-C(3')), 4.30 (d, *J* = 2.6 Hz, 0.5H, H-C(4')), 4.20 (d, *J* = 3.4 Hz,

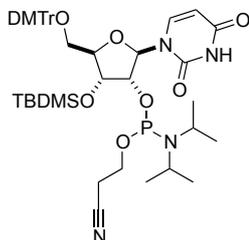
0.5H, H-C(4')), 3.98-3.85 (m, 0.5H, ce OCH<sub>2</sub>), 3.84-3.74 (m, 6H, OCH<sub>3</sub>), 3.73-3.39 (m, 5.5H, ce OCH<sub>2</sub>, H<sub>2</sub>-C(5'), <sup>i</sup>Pr CH), 2.66 (td, *J* = 6.2, 1.9 Hz, 0.8H, ce CH<sub>2</sub>CN), 2.46-2.31 (m, 1.2H, ce CH<sub>2</sub>CN), 2.20-2.07 (m, 3H, CO-CH<sub>3</sub>), 1.32-1.01 (m, 12H, <sup>i</sup>Pr CH<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 169.9, 169.7 (CO-CH<sub>3</sub>), 162.7, 162.7 (C(4)), 158.9 (DMTr), 150.3 (C(2)), 144.2, 144.1 (DMTr), 140.1 (C(6)), 135.3, 135.2, 135.1, 135.0, 134.4, 130.4, 130.3, 128.5, 128.4, 128.2, 128.2, 127.4, 127.4 (DMTr), 117.8, 117.4 (CN), 113.5, 113.4 (DMTr), 102.9, 102.9 (C(5)), 87.5, 87.4 (DMTr-C), 86.5, 86.3 (C(1')), 84.2, 83.8, 83.7 (C(4')), 74.8, 74.5, 74.5 (C(2')), 71.5, 71.3, 70.7, 70.5 (C(3')), 62.7, 62.5 (C(5')), 58.9, 58.7, 58.3, 58.1, 57.9 (ce OCH<sub>2</sub>), 55.4, 55.4 (OCH<sub>3</sub>), 43.5, 43.4, 43.3 (<sup>i</sup>Pr CH), 24.8, 24.7, 24.7, 24.7, 24.7 (<sup>i</sup>Pr CH<sub>3</sub>), 21.1, 20.9 (CO-CH<sub>3</sub>), 20.5, 20.5, 20.3, 20.3 (ce CH<sub>2</sub>CN). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 151.03 (m), 150.25 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled) δ 151.04 (s), 150.25 (s). *m/z* ESI-HRMS (pos.) [M+Na]<sup>+</sup> calculated for C<sub>41</sub>H<sub>49</sub>N<sub>4</sub>O<sub>10</sub>PNa<sup>+</sup>, 881.3084; found 811.3083.

#### *Data for 107b*

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.57 (br. s, 1H, NH), 7.86 (dd, *J* = 10.1, 8.1 Hz, 1H, H-C(6)), 7.38-7.22 (m, 9H, DMTr), 6.85-6.83 (m, 4H, DMTr), 6.13, 6.09 (2 × d, *J* = 4.4 Hz, 1H, H-C(1')), 5.39-5.27 (m, 2H, H-C(5), H-C(3')), 4.63-4.56 (m, 1H, H-C(2')) 4.23-4.22 (m, 1H, H-C(4')), 3.89-3.55 (m, 11H, ce OCH<sub>2</sub>, OCH<sub>3</sub>, <sup>i</sup>Pr CH, H-C(5')), 3.46-3.41 (m, 1H, H-C(5'')), 2.67-2.53 (m, 2H, ce CH<sub>2</sub>CN), 2.10 (2 × s, 3H, CO-CH<sub>3</sub>), 1.18-1.14 (m, 12H, <sup>i</sup>Pr CH<sub>3</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz): δ 170.0, 169.9 (CO-CH<sub>3</sub>), 162.9, 162.9 (C(4)), 158.9 (DMTr), 150.3, 150.2 (C(2)), 144.3, 144.3 (DMTr), 140.3, 140.1 (C(6)), 135.2, 135.1, 130.3, 130.2, 128.2, 128.2, 128.2, 127.4 (DMTr), 117.8, 117.6 (ce CH<sub>2</sub>CN), 113.5, 113.5 (DMTr), 102.6, 102.6 (C(5)), 88.3, 88.2 (C(1')), 87.6, 87.5 (DMTr-C), 81.5 (C(4')), 74.9, 74.8, 74.6, 74.5 (C(2')), 71.3, 71.3, 71.1 (C(3')), 62.2, 62.0 (C(5')), 58.8, 58.6, 58.6, 58.4 (ce OCH<sub>2</sub>), 55.4 (OCH<sub>3</sub>), 43.7, 43.6, 43.5, 43.5 (<sup>i</sup>Pr CH), 24.9, 24.8, 24.7, 24.7, 24.6, 24.6 (<sup>i</sup>Pr CH<sub>3</sub>), 21.1, 21.0 (CO-CH<sub>3</sub>), 20.4, 20.4, 20.3, 20.3 (ce CH<sub>2</sub>CN). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 152.01-151.60 (m), 151.28-150.88 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled) δ 151.80 (s), 151.11 (s). *m/z* ESI-HRMS (pos.) [M+Na]<sup>+</sup> calculated for C<sub>41</sub>H<sub>49</sub>N<sub>4</sub>O<sub>10</sub>PNa<sup>+</sup>, 811.3084; found 811.3072.

**3'-O-(*tert*-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)uridine-2'-O-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite 115b‡**

C<sub>45</sub>H<sub>61</sub>N<sub>4</sub>O<sub>9</sub>Psi; M<sub>r</sub> = 861.05



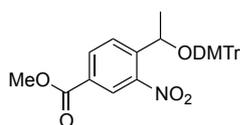
Commercially available 3'-O-(*tert*-butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)uridine **111** (1.50 g, 2.27 mmol) was co-evaporated with anhydrous THF (3 × 10 mL). The residue and *N,N*-diisopropylethylamine (1.20 mL, 6.81 mmol) were dissolved in anhydrous THF (12 mL). 2-Cyanoethyl *N,N*-diisopropyl phosphoamidochloridite (0.61 mL, 2.72 mmol) was added to the solution dropwise at 0 °C and the resultant mixture stirred at RT for 4 h. The reaction was quenched with anhydrous MeOH (2.5 mL) and the solvent removed under vacuum. The residue was taken up in EtOAc (10 mL) and the organics were washed with saturated aq. NaHCO<sub>3</sub> (3 × 10 mL). The organics were separated and dried over MgSO<sub>4</sub> and evaporated to dryness. The residue was then purified by flash column chromatography (50:50:1 EtOAc:*c*-Hex:Et<sub>3</sub>N) to give the title compound (1.80 g, 92%) as a colourless foam. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.72, 8.53 (2 × s, 1H, NH), 8.14, 8.06 (2 × d, *J* = 8.2 Hz, 1H, H-C(6)), 7.38-7.23 (m, 9H, DMTr), 6.87-6.81 (m, 4H, DMTr), 6.15 (d, *J* = 3.2 Hz, 0.4H, H-C(1')), 6.06 (d, *J* = 2.2 Hz, 0.6H, H-C(1')), 5.35, 5.27 (2 × d, *J* = 8.1 Hz, 1H, H-C(5)), 4.37-4.19 (m, 2H, H-C(2'), H-C(3')), 4.14-4.06 (m, 1H, H-C(4')), 3.99-3.85 (m, 1H, *ce* OCH<sub>2</sub>), 3.84-3.54 (m, 10H, OCH<sub>3</sub>, *ce* OCH<sub>2</sub>, <sup>*i*</sup>Pr CH, H-C(5')), 3.37-3.29 (m, 1H, H-C(5'')), 2.72-2.47 (m, 2H, *ce* CH<sub>2</sub>CN), 1.18-1.15 (m, 12H, <sup>*i*</sup>Pr CH<sub>3</sub>), 0.80-0.78 (m, 9H, Si(CH<sub>3</sub>)<sub>3</sub>), 0.09, 0.04 (2 × s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), -0.01, -0.03 (2 × s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz) δ 163.2, 163.1 (C(4)), 158.9, 158.9 (DMTr), 150.3, 150.1 (C(2)), 144.2, 144.2 (DMTr), 140.6, 140.4 (C(6)), 135.2, 135.1, 135.0, 130.4, 130.4, 128.5, 128.4, 128.1, 127.4, 127.4 (DMTr), 118.1, 117.8 (*ce* CH<sub>2</sub>CN), 113.4, 113.4 (DMTr), 102.2, 102.1 (C(5)), 88.7, 88.6, 88.5 (C(1')), 87.4, 87.3 (DMTr-

C), 83.6, 83.5 (C(4')), 76.6, 76.5 (C(2'))<sup>j</sup>, 70.5, 70.5, 70.4 (C(3')), 61.7, 61.6 (C(5')), 58.5, 58.3, 58.3, 58.1 (ce OCH<sub>2</sub>), 55.4, 55.4 (OCH<sub>3</sub>), 43.5, 43.4, 43.3 (<sup>i</sup>Pr CH), 25.9 (SiC(CH<sub>3</sub>)<sub>3</sub>), 24.9, 24.9, 24.8, 24.8, 24.6, 24.5 (<sup>i</sup>Pr CH<sub>3</sub>), 20.5, 20.4, 20.4, 20.3 (ce CH<sub>2</sub>CN), 18.2, 18.1 (SiC(CH<sub>3</sub>)<sub>3</sub>), -4.1, -4.2, -4.8, -4.9 (Si(CH<sub>3</sub>)<sub>2</sub>). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 150.60-150.25 (m), 149.59-149.25 (m). <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>, decoupled) δ 150.42 (s), 149.37 (s). *m/z* ESI-HRMS (pos.) [M+Na]<sup>+</sup> calculated for C<sub>45</sub>H<sub>61</sub>O<sub>9</sub>N<sub>4</sub>PSiNa<sup>+</sup>, 883.3838; found 883.3816.

## 6.2.2. Preparation of the Solid-Phase Support

### Methyl 4-(1-(bis(4-methoxyphenyl)(phenyl)methoxy)ethyl)-3-nitrobenzoate 123

C<sub>31</sub>H<sub>29</sub>NO<sub>7</sub>, M<sub>r</sub> = 527.56

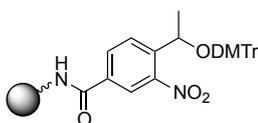


Methyl 3-formyl-4-nitrobenzoate **120** (5.00 g, 23.9 mmol) was dissolved in anhydrous diethyl ether (150 mL) and methylmagnesium bromide (1 M in dibutyl ether, 59.7 mL, 59.7 mmol) was added dropwise slowly, maintaining the reaction vessel at RT. The reaction mixture was stirred for 4 h and quenched with the addition of saturated ammonium chloride (50 mL). The organics were extracted with Et<sub>2</sub>O (3 x 50 mL), the combined organic layers were washed with water (50 mL), brine (50 mL) and dried over MgSO<sub>4</sub>. The solvent was removed under vacuum and the residue was purified by flash column chromatography (50:50, EtOAc:*c*-Hex; R<sub>f</sub> = 0.27 (90:10, CHCl<sub>3</sub>:EtOAc)). The product was found to also contain a benzyl alcohol side product that could not be separated by flash column chromatography. The material was used in the next step without further purification. Alcohol **121** (1.22 g, 5.42 mmol) was co-evaporated with anhydrous pyridine (3x10 mL). The residue and DMTr-Cl (2.75 g, 8.12 mmol) were dissolved in anhydrous pyridine (20 mL) and anhydrous CH<sub>2</sub>Cl<sub>2</sub> (10 mL) and the reaction mixture stirred at RT overnight. MeOH (20 mL) was added and the mixture stirred for a further 10 min before the solvent was removed under vacuum. The residue was dissolved in CHCl<sub>3</sub> (50 mL) and washed with saturated aq. NaHCO<sub>3</sub> (2 x 40 mL).

<sup>j</sup> Partial overlap with CDCl<sub>3</sub>

The organic layer was dried over MgSO<sub>4</sub> and evaporated under vacuum. The residue was purified by flash column chromatography (6:4, *c*-Hex:CHCl<sub>3</sub>) to afford a yellow oil. This residue contained the benzyl side product from the previous step and was further purified by NP-HLPC (Method F), with a retention time of 15.5 min. To give the title product (1.02 g, 9%) as a colourless solid.  $R_f = 0.33$  (99:1, toluene:Et<sub>3</sub>N). IR (cm<sup>-1</sup>) 1725 (C=O), 1607 (Ar), 1530, 1507 (Ar), 1282, 1247, 1149, 1029. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  8.21 (d,  $J = 1.7$  Hz, 1H, H-C(2)), 7.97 (dd,  $J = 8.2, 1.7$  Hz, 1H, H-C(6)), 7.82 (d,  $J = 8.2$  Hz, 1H, H-C(5)), 7.48-7.43 (m, 2H, DMTr), 7.29-7.10 (m, 7H, DMTr), 6.68-6.59 (m, 4H, DMTr), 5.23 (q,  $J = 6.2$  Hz, 1H, H-C(7)), 3.92 (s, 3H, COOCH<sub>3</sub>), 3.74, 3.73 (2xs, 6H, DMTr OCH<sub>3</sub>), 1.65 (d,  $J = 6.2$  Hz, 3H, H<sub>3</sub>-C(8)). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 101 MHz)  $\delta$  165.0, 158.5, 158.4, 147.6, 146.0, 145.3, 136.2, 135.7, 132.0, 130.1, 130.02, 130.0, 128.7, 127.8, 127.8, 126.8, 124.7, 113.1, 113.0, 68.0, 55.2, 55.2, 52.5, 24.8. *m/z* ESI-HRMS (pos.) [M+Na]<sup>+</sup> calculated for C<sub>31</sub>H<sub>29</sub>O<sub>7</sub>N<sub>1</sub>Na<sup>+</sup>, 550.1836; found 550.1827.

#### Procedure for the Preparation of the Photolabile-CPG 126‡



#### Methyl 4-(1-(bis(4-methoxyphenyl)(phenyl)methoxy)ethyl)-3-nitrobenzoate **123**

(500 mg, 0.94 mmol) was dissolved in THF (1.50 mL) and water (0.50 mL). LiOH (24.9 mg, 1.04 mmol) was added and the mixture was stirred at RT for 24 h, the reaction monitored by TLC until no starting material was observable. The solvent was removed to give a white solid and this material was used in the next step without further purification. The lithium salt (250 mg, 0.48 mmol) was co-evaporated with anhydrous pyridine (3 × 4 mL). The residue was dissolved in anhydrous pyridine (4 mL), isobutylchloroformate (72.0 mg, 0.53 mmol) was added and the formation of a white precipitate was observed. The reaction mixture was stirred for 30 min before the precipitate was removed by filtration under argon and the supernatant filtered directly into oven-dried glassware. The solvent was removed under vacuum and the subsequent mixed anhydride was dissolved in anhydrous CH<sub>2</sub>Cl<sub>2</sub> (4 mL) followed by the addition of *N,N*-diisopropylethylamine (68 mg, 0.53 mmol) and long chain alkylamine controlled pore glass (250 mg, 120-200 mesh, nominal diameter 500 Å, 100-175

$\mu\text{molg}^{-1}$ ). The suspension was rotated gently under argon at RT for 24 h. The CPG was filtered and washed with  $\text{CH}_2\text{Cl}_2$ , MeCN, water, MeCN and finally  $\text{CH}_2\text{Cl}_2$  (15 mL each). The modified CPG was dried overnight under vacuum then treated with CAP A (80:10:10, THF:2,6-lutidine:pivaloyl chloride) (2 mL) and commercially available CAP B (90:10, THF:*N*-methylimidazole) (2 mL) solution and the mixture was rotated gently for 1.5 h. The CPG was filtered and washed with  $\text{CH}_2\text{Cl}_2$  (15 mL) and dried under vacuum. The loading was determined by Trityl assay (see below - Section 6.2.3) and loading values of 33.3-56.2  $\mu\text{molg}^{-1}$  were obtained. The prepared CPG was stored in the dark at 4 °C.

### 6.2.3. Synthesis of oligonucleotides

#### General Methods

Trityl assays<sup>[183]</sup> were conducted using a *Varian* CARY 6000i UV-Vis spectrometer. The solid-support (3-4 mg for CPG loading calculations) was accurately weighed into a 10 or 50 mL volumetric flask. 3% TCA in  $\text{CH}_2\text{Cl}_2$  was added and the absorption of the resultant orange solution was measured on the spectrometer at  $\lambda$  503 nm. The loading was calculated using the equation below:

$$\text{Loading } (\mu\text{molg}^{-1}) = (A_{503} \times \frac{\text{vol}}{76}) \times (\frac{1000}{\text{wt}})$$

*Equation 1. Equation to calculate the loading.  $A_{503}$  is the absorption at  $\lambda$  503 nm, vol is the volume of the solution in mL,  $76 \mu\text{mol}^{-1} \text{ mL cm}^{-1}$  is the extinction coefficient at  $\lambda$  503 nm and wt is the amount of support in mg. Path length assumed to be 1 cm.*

RNA was quantitated using an *Eppendorf* Biophotometer Plus by dissolving the oligonucleotide of interest in a known volume of  $\text{H}_2\text{O}$ . A 10  $\mu\text{L}$  aliquot of this solution was diluted to a volume of 1 mL and a UV absorbance at  $\lambda$  260 nm was obtained. The concentration ( $\text{nmolmL}^{-1}$ ) calculated using the below equation:

$$\text{Oligonucleotide concentration } (\text{nmolmL}^{-1}) = \frac{A_{260} \times \text{dilution factor} \times 10^6}{\text{Molar extinction coefficient}}$$

*Equation 4. Equation for the quantification of oligonucleotides.*

RNA synthesis was typically conducted on a 1  $\mu$ mol scale and desalted using a *Waters* C18 Sep-Pak<sup>®</sup> column using the following protocol: A *Waters* C18 Sep-Pak<sup>®</sup> column (2 g, 12cc) was washed with MeOH (20 mL) then water (20 mL). The RNA sample was diluted 1:1 with 1 M triethylammonium acetate (TEAA) buffer (pH 7) and loaded slowly on to the cartridge. The column was flushed slowly with TEAA (50 mM, pH 7, 12 mL) and the oligonucleotide was eluted with 50 mM TEAA:MeOH (7:3, 10 mL) and collected in 5  $\times$  2 mL fractions. Fractions containing RNA were determined by UV absorbance at  $\lambda$  260 nm were combined, lyophilised, redissolved in water and re-lyophilised to remove traces of TEAA buffer.

### **Synthesis of the Acetylated RNA Oligonucleotides**

Oligonucleotides (Table 9, entries 3-9) were synthesised as described below. Additional non-acetylated oligonucleotides used in  $T_m$  and thermodynamic studies were purchased in HPLC-purified Na<sup>+</sup> form from *Integrated DNA Technologies*.

### **Conditions and materials for the automated synthesis of acetyl-RNA oligonucleotides**

RNA oligonucleotide synthesis was performed using a *Bioautomation* MerMade 4 on a 1  $\mu$ mol scale using monomers **112-115** (0.1 M in 1:1 CH<sub>2</sub>Cl<sub>2</sub>:MeCN), acetylated monomers **103**, **104**, **106** and **107** (0.1 M, CH<sub>2</sub>CH<sub>2</sub>) and utilised the photolabile-CPG **126**. The synthesis cycle began with detritylation using 3% dichloroacetic acid (DCA) in CH<sub>2</sub>Cl<sub>2</sub>. The coupling step utilised 80  $\mu$ L of each amidite (0.1 M, 1:1, MeCN:CH<sub>2</sub>Cl<sub>2</sub>) and 1 M 4,5-dicyanoimidazole (DCI) as the activator with a single 20 min coupling, with the exception that the acetylated amidites were given a 17 min double coupling step. A 5 min capping step was carried out with CAP A (THF:2,6-lutidine:pivalic anhydride) and CAP B (90:10, THF:*N*-methylimidazole) solutions. The oxidation of the phosphites was carried out with 0.02 M iodine (THF/pyridine/water, 7:2:1). The automation program was set as DMTr-ON such that the final DMTr group not removed.

### **Deprotection, cleavage and purification of acetyl-RNA oligonucleotides**

Without removal of the CPG from the synthesis column, the CPG was dried under vacuum for 15 min. A solution of 0.5 M DBU in anhydrous MeCN (3.5 mL, 10%

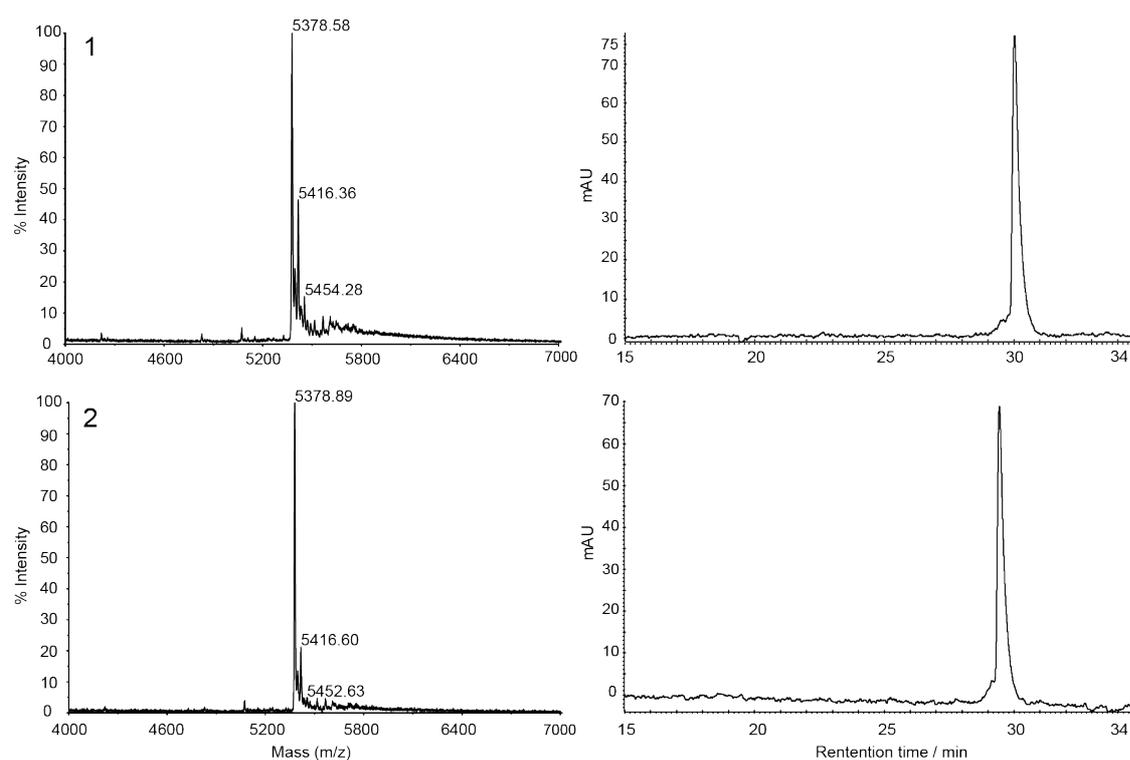
morpholine) was initially passed through the column for 5 min then the column and CPG were immersed in the DBU solution under an atmosphere of argon for 6 h at 40 °C with sonication every hour. The column was washed with anhydrous acetonitrile (10 mL) and CH<sub>2</sub>Cl<sub>2</sub> (10 mL). The final DMTr group was removed by passing 3% TCA in CH<sub>2</sub>Cl<sub>2</sub> through the column until the washings became colourless. The collected orange coloured solution was diluted to 50 mL with CH<sub>2</sub>Cl<sub>2</sub> and the yield of full length product calculated by trityl assay. At this point the CPG was extracted and placed into a 4 mL UV transparent vessel (*Corning* costar 24 Well Cell Culture Cluster 3526) and suspended in DMSO (0.5 mL). The CPG was irradiated at  $\lambda = 365$  nm (max = 34.5 mW/cm<sup>2</sup>, *Prizmatix* Mic-LED-365) for 1 h. The CPG was removed by filtration, washed with DMSO (2 × 0.5 mL) and the fractions combined. The DMSO was removed by lyophilisation and the residue redissolved in anhydrous DMSO (200  $\mu$ L). To the solution Et<sub>3</sub>N.3HF (125  $\mu$ L) was added, the mixture was thoroughly mixed and then heated at 65 °C for 3 h. The fully deprotected oligonucleotide were desalted by using a *Waters* C18 Sep-Pak<sup>®</sup> column or more commonly by the precipitation method as follows. Firstly, 3 M sodium acetate (25  $\mu$ L, pH 7) was added followed by thorough mixing. After the addition of *n*-butanol (1 mL) the mixture was cooled at -80 °C for 30 min and then centrifuged for 10 min at 13200 rpm. The *n*-butanol was decanted, followed by washing of the pellet with ethanol (2 × 0.75 mL) and finally drying the pellet in a SpeedVac at 65 °C for 1 h. The dried pellet was dissolved in RNAase-free water (0.5-1 mL) and the RNA oligomer was quantified by UV absorbance at  $\lambda$  260 nm and analysed by MALDI-TOF MS to assess the synthesis. Dephosphorylation (if required) of the acetylated RNA oligonucleotide began by dilution to a concentration of 1 $\mu$ g/10 $\mu$ L with PBS buffer (0.01 M phosphate, 0.138 M NaCl, 0.0027 M KCl, 1 M MgCl<sub>2</sub>, pH 7.4). To the dissolved RNA oligonucleotide calf intestinal alkaline phosphatase (0.5u/ $\mu$ g) was added and the mixture heated at 37 °C for 1 h. To the solution was added one volume of 1 M TEAA buffer (pH 7) and the oligonucleotide was desalted using a *Waters* C18 Sep-Pak<sup>®</sup> column prior to HPLC purification of the oligonucleotide. Acetylated RNA oligomers were purified by preparative SAX-HPLC (Method G). The target fractions were combined and desalted by dialysis (*Thermo Scientific* Slide-A-Lyzer Dialysis Cassettes, 2K MWCO) against 10 mM TEAA buffer pH 7.0 at 4 °C. Finally the purified oligonucleotides were quantified by UV absorbance at 260 nm, and characterised by MALDI-TOF MS.

## Synthesis of non-acetylated oligonucleotides

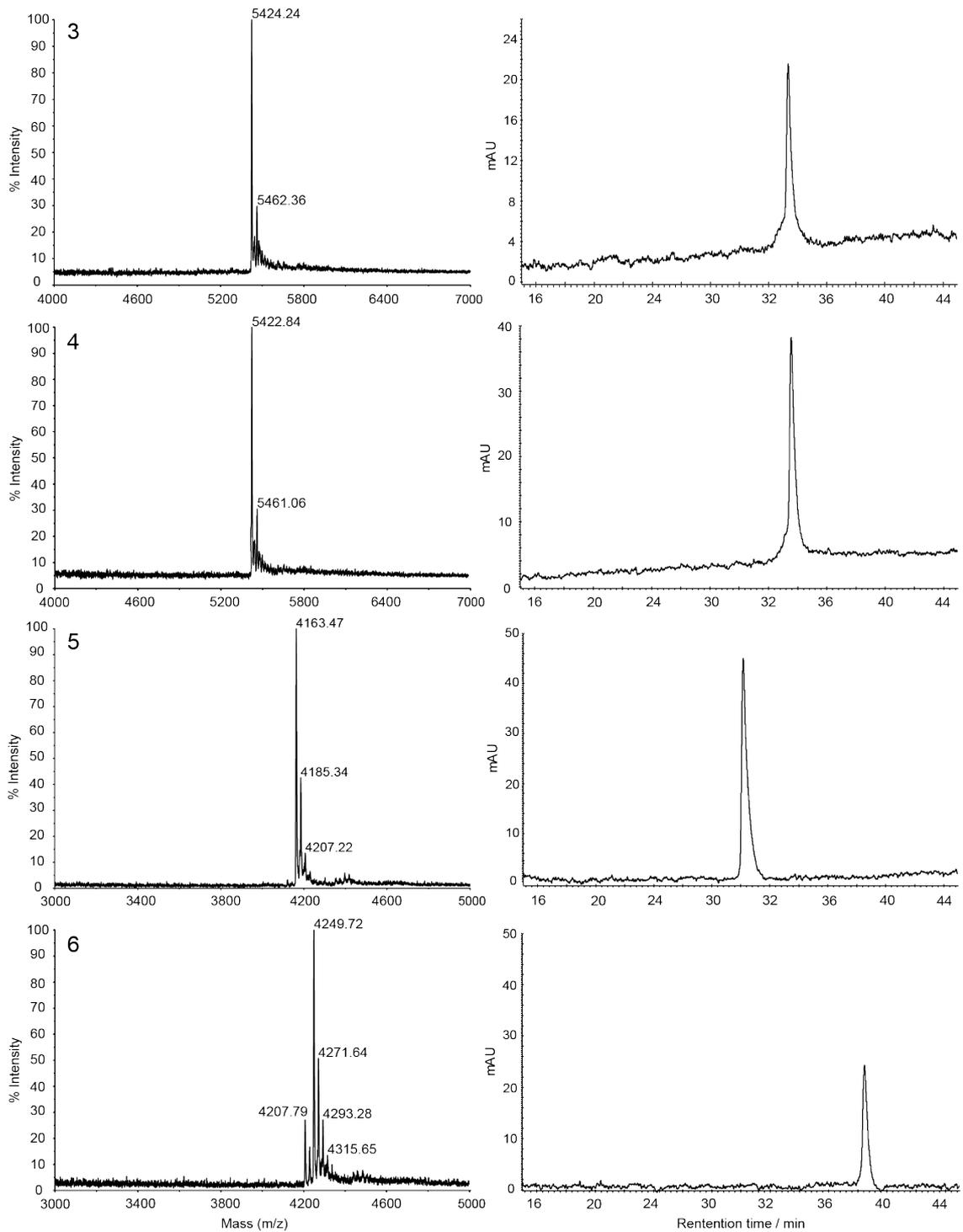
Non-acetylated oligonucleotides (Table 9, entry 1-2) were synthesised using a *BioAutomation MerMade 4* on a 1  $\mu\text{mol}$  scale utilising standard RNA phosphoramidites and reagents from *ChemGenes* or *Link Technologies*. 2'-5'-linkages were introduced using 3'-TBDMS phosphoramidites from *ChemGenes*. Oligomers were purified by SAX-HPLC (Method H), target fractions were immediately neutralised with one volume of 1M TEAA buffer (pH 7). The combined target fractions were desalted by *Waters C18 Sep-Pak*<sup>®</sup> column. The oligomers were quantified by UV absorbance at 260 nm and were characterised by MALDI-TOF MS.

Entry	RNA Sequence (5'-3')	Average Mass for [M+H] <sup>+</sup> (Da)	
		Calc.	Observed
1	UGUGCCAGUA-3', 5' -GGUUCUC	5381.26	5378.58
2	UGUGCCAGUA-2', 5' -GGUUCUC	5381.26	5378.89
3	UGUGCCAGUA-3', 5' (2'OAc) -GGUUCUC	5423.29	5424.24
4	UGUGCCAGUA-2', 5' (3'OAc) -GGUUCUC	5423.29	5422.84
5	CCAG-3', 5' (2'OAc) -UAGGU-3', 5' (2'OAc) -UCUC	4162.59	4163.47
6	GAGA-3', 5' (2'OAc) -ACC-3', 5' (2'OAc) -UACUGG	4248.70	4249.72
7	GCCG-3', 5' (2'OAc) -UAAGGC	3242.07	3242.60
8	GCCG-3', 5' (2'OAc) -AGAGGC	3281.06	3281.95
9	GCCG-3', 5' (2'OAc) -AGAG-3', 5' (2'OAc) -GC	3323.06	3324.24

Table 9. Synthesised RNA oligomers and their MALDI-TOF MS characterisation data.



*Continued overleaf.*



*Continued overleaf.*

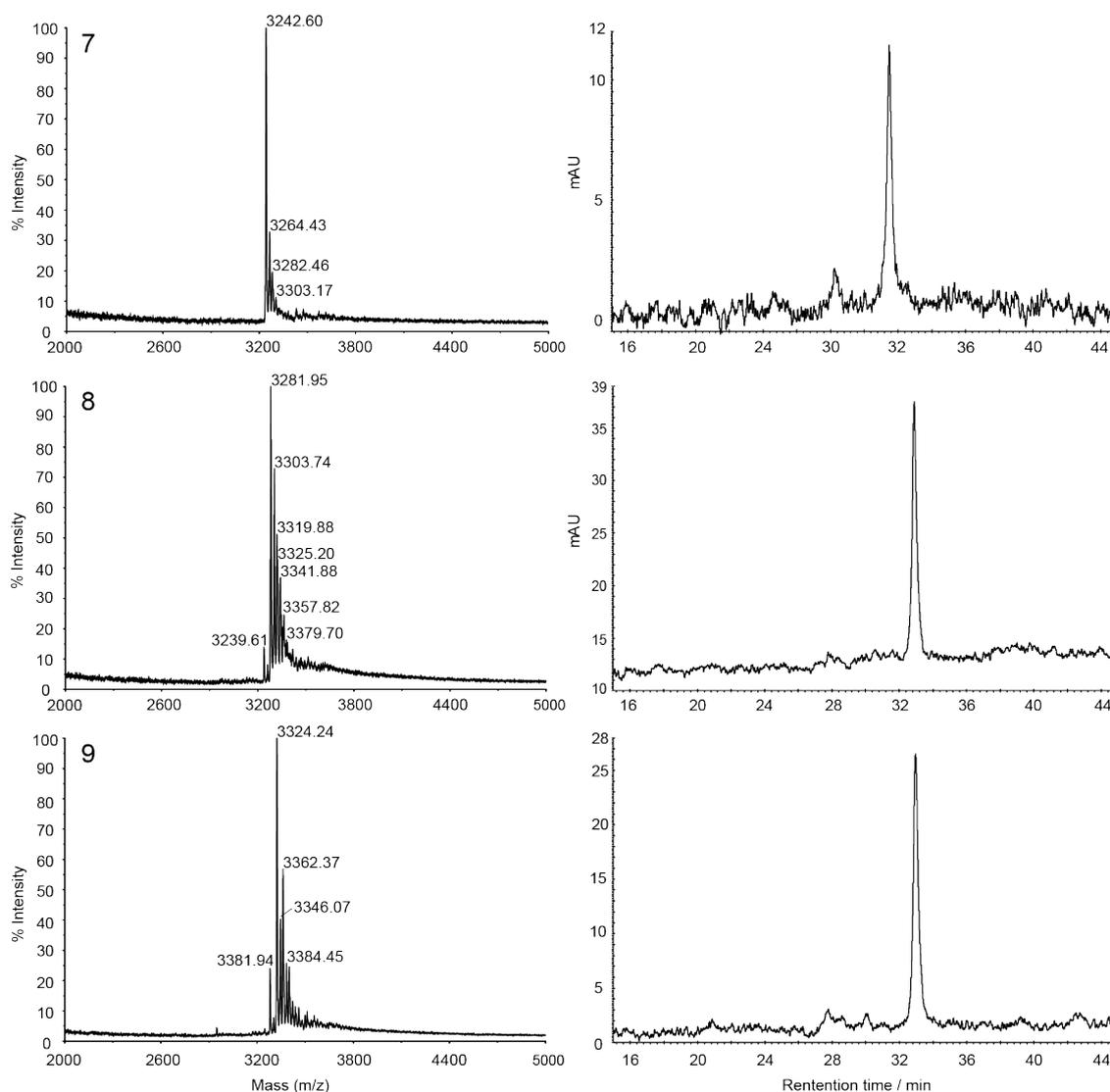


Figure 120. MADLI-TOF mass spectrum and HPLC traces of the synthesised partially acetylated RNA oligonucleotides from Table 9.

### 6.3. Procedures for Chapter 3

#### General procedure for measuring UV melting curves

Oligomers used for  $T_m$  measurements were converted to the  $\text{Na}^+$  form using prewashed *Bio-rad* AG<sup>®</sup> 50W-X8 resin. An excess of resin (100 mg) was added to an aqueous solution of RNA oligomer (0.5-1.0 mL) and the mixture agitated for a minimum of 4 h. The resin was removed by filtration and washed with water ( $2 \times 250 \mu\text{L}$ ), the oligomer was quantified by UV absorbance at 260 nm. UV thermal melting curves were acquired using a *Varian* CARY 6000i UV-Vis spectrometer equipped with a multi-sample Peltier temperature controller. All measurements were carried out in 10 mM  $\text{Na}_2\text{HPO}_4$ , 0.5 mM

Na<sub>2</sub>EDTA buffer (pH 7) and between 0.1-1 M NaCl. Prior to UV measurements, samples were degassed by heating at 95 °C for 4 min followed by brief sonication and slowly cooling to RT. Measurements were made in masked quartz cuvettes with mineral oil layered over the sample to reduce evaporation. Absorbance versus temperature spectra were measured within a range of 10-95 °C and at  $\lambda$  260 nm (280 nm was used with some GC rich oligonucleotides).<sup>[328]</sup> The temperature was ramped at a rate of 0.5 °Cmin<sup>-1</sup> with absorbance measurements taken at a 0.5 °C intervals. The oligonucleotides were annealed and equilibrated by the first heat-cool cycle and holding at the maximum temperature in the range selected for 5 min. UV melting heat-cool runs were then conducted in triplicate holding for 5 min between ramps.

### **Data analysis, T<sub>m</sub> and thermodynamic parameter calculation**

Data analysis, T<sub>m</sub> and thermodynamic parameter calculations were conducted with *GraphPad Prism 5.0d* and *Microsoft Excel:Mac 2011*.<sup>[262, 329]</sup> To calculate the T<sub>m</sub>, the absorbance versus temperature melting curves were first converted to a normalised absorbance (A<sup>n</sup>) versus temperature (T) curve. Assuming a two-state model<sup>[260]</sup> the lower (B1<sub>T</sub>) and upper (B0<sub>T</sub>) baselines (corresponding to fully associated/folded and fully dissociated/unfolded respectively) were computed by performing a linear regression on the associated and dissociated parts of the normalised melting curves. The equations of each baseline were used to transform the normalised absorbance (A<sup>n</sup>) versus temperature (T) plots into a fraction associated/folded ( $\alpha$ ) versus temperature (T) plots using *Equation 5*. Finally, the T<sub>m</sub> was extracted by reading the temperature at  $\alpha = 0.5$ . The data from three heat-cool cycles were analysed as described above and an average of the six T<sub>m</sub> values were taken as the final T<sub>m</sub>.

$$\alpha = \frac{B0_T - A_T^n}{B0_T - B1_T}$$

*Equation 5. The equation used to convert the normalised absorbance (A) versus temperature (T) plot to a fraction folded/associated ( $\alpha$ ) versus temperature (T) plot.*

Thermodynamic parameters were calculated using a van't Hoff analysis using *Microsoft Excel for Mac 2011*. The data analysis was restricted to the range  $0.15 < \alpha < 0.85$ . The association constant ( $K_a$ ) was calculated for each data point using the equations

below<sup>[263]</sup>, where the association involves non-self-complementary sequences either monomolecular (*i.e.* hairpin formation) or bimolecular (*i.e.* duplex formation):

$$K_a = \frac{\alpha}{\left(\frac{C_T}{n}\right)^{n-1} (1 - \alpha)^n}$$

*Equation 2. Calculation of the association constant ( $K_a$ ) for non-self-complementary sequences, in terms of fraction associated ( $\alpha$ ), total oligonucleotide concentration ( $C_T$ ) and the molecularity ( $n$ , e.g.  $n = 2$  for a bimolecular interaction).*

For  $T_m$  experiments in which the equilibria involve self-complementary sequences the following equation was used:

$$K_a = \frac{\alpha}{(nC_T)^{n-1} (1 - \alpha)^n}$$

*Equation 6. Calculation of the association constant ( $K_a$ ) for self-complementary sequences.*

In all cases:

$$\Delta G^\circ = -RT \ln(K_a) = \Delta H^\circ - T \cdot \Delta S^\circ$$

Therefore;

$$\ln(K_a) = \frac{-\Delta H^\circ}{R} \cdot \frac{1}{T_m} + \frac{\Delta S^\circ}{R}$$

Equation 3. Derivation to calculate the thermodynamic parameters of the melting curves.  $R$  = Gas Constant ( $8.314 \text{ JK}^{-1}\text{mol}^{-1}$ ),  $T$  = temperature ( $^\circ\text{C}$ ),  $\Delta G^\circ$  = Gibbs free energy ( $\text{kJmol}^{-1}$ ),  $\Delta H^\circ$  = Enthalpy ( $\text{kJmol}^{-1}$ ),  $\Delta S^\circ$  = Entropy ( $\text{JK}^{-1}\text{mol}^{-1}$ ) and  $K_a$  = association constant.

From the above derivation a plot of  $\ln(K_a)$  versus  $1/T_m$  was made, which should result in a straight line and is otherwise called a van't Hoff plot. A linear regression was performed in *GraphPad Prism* and from the straight line the values for the slope which gives  $-\Delta H^\circ/R$  and y-intercept which gives  $\Delta S^\circ/R$  were extracted (where  $R$  is the ideal gas constant). Thus, the Gibb's energy ( $\Delta G^\circ$ ) can be calculated using:

$$\Delta G^\circ = \Delta H^\circ - T \cdot \Delta S^\circ$$

*Equation 7. Gibbs free energy equation where  $T$  is temperature in Kelvins ( $0 \text{ }^\circ\text{C} = 273.15 \text{ K}$ ),  $\Delta H^\circ$  is the standard enthalpy and  $\Delta S^\circ$  is the standard entropy.*

Entry	RNA Sequence (5' to 3')	Complement (5' to 3')	T <sub>m</sub> (°C)	ΔH° (kJmol <sup>-1</sup> )	ΔS° (Jmol <sup>-1</sup> K <sup>-1</sup> )	ΔG° <sub>0</sub> (kJmol <sup>-1</sup> )	ΔG° <sub>37</sub> (kJmol <sup>-1</sup> )
1	UGGCCAGUA-3', 5'-GGUUCUC	GAGAACCUACUGG	74.7	-530.0	-1410.5	-144.7	-92.6
2	UGGCCAGUA-2', 5'-GGUUCUC	GAGAACCUACUGG	67.8	-421.7	-1123.7	-114.7	-73.2
3	UGGCCAGUA-3', 5' (2' OAc) -GGUUCUC	GAGAACCUACUGG	71.6	-420.4	-1113.3	-116.3	-75.1
4	UGGCCAGUA-2', 5' (3' OAc) -GGUUCUC	GAGAACCUACUGG	72.0	-517.2	-1385.1	-138.9	-87.7
5	UGGCCAGUA-3', 5'-GGUUCUC	GAGA-3', 5' (2' OAc) -ACC-3', 5' (2' OAc) -UACUGG	68.5	-428.5	-1140.6	-116.9	-74.7
6	UGGCCAGUA-3', 5' (2' OAc) -GGUUCUC	GAGA-3', 5' (2' OAc) -ACC-3', 5' (2' OAc) -UACUGG	65.6	-363.9	-961.0	-101.4	-65.9
7	CCAG-3', 5' (2' OAc) -UAGGU-3', 5' (2' OAc) -UCUC	GAGA-3', 5' (2' OAc) -ACC-3', 5' (2' OAc) -UACUGG	61.5	-361.7	-967.4	-97.5	-61.7
8	CCAG-3', 5' (2' OAc) -UAGGU-3', 5' (2' OAc) -UCUC	GAGAACCUACUGG	67.9	-462.7	-1243.0	-123.2	-77.2
9	CCAGUAGGUUCUC	GAGA-3', 5' (2' OAc) -ACC-3', 5' (2' OAc) -UACUGG	67.4	-461.9	-1242.2	-122.6	-76.7
10	CCAGUAGGUUCUC	GAGAACCUACUGG	73.8	-521.6	-1390.5	-141.8	-90.3

*Table 10. T<sub>m</sub> and thermodynamic parameters used to assess the effect of acetylation on the duplex stability of complementary sequences. Each measurement used 2.5 μM of each oligomer to give a total RNA concentration of 5 μM, in 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.5 mM Na<sub>2</sub>EDTA buffer (pH 7), 1 M NaCl and a temperature range of 30-90 °C. Data is an average of three heat-cool cycles. Error for T<sub>m</sub> values represent standard deviations of 6 values and are ±0.8 °C. Errors for thermodynamic data represent standard deviations of 6 values are ±7.4% for ΔH°, within ±8.5% for ΔS°, within ±5.0 kJmol<sup>-1</sup> for ΔG°<sub>0</sub> and within ±3.1 kJmol<sup>-1</sup> for ΔG°<sub>37</sub>.*

Entry	RNA Sequence (5' to 3')	Conc. ( $\mu\text{M}$ )	Complement (5' to 3')	Conc. ( $\mu\text{M}$ )	Total Conc. ( $\mu\text{M}$ )	Temp. Range ( $^{\circ}\text{C}$ )	$T_m$ ( $^{\circ}\text{C}$ )	$\Delta H^{\circ}$ ( $\text{kJmol}^{-1}$ )	$\Delta S^{\circ}$ ( $\text{Jmol}^{-1}\text{K}^{-1}$ )	$\Delta G^{\circ}$ ( $\text{kJmol}^{-1}$ )	$\Delta G_{37}^{\circ}$ ( $\text{kJmol}^{-1}$ )
1.1 <sup>a</sup>	GCCG-3',5' (2' OAc)-UAAAGC	5	-	-	5	30-90	69.5	-212.5	-620.8	-42.9	-20.0
1.2 <sup>a</sup>	GCCG-3',5' (2' OAc)-UAAAGC	10	-	-	10	30-90	68.9	-177.6	-519.6	-35.7	-16.5
1.3 <sup>a</sup>	GCCG-3',5' (2' OAc)-UAAAGC	50	-	-	50	30-90	68.7	-168.3	-492.8	-33.7	-15.5
1.4 <sup>b</sup>	GCCG-3',5' (2' OAc)-UAAAGC	5	GCCUUACGGC	5	10	30-95	57.8	-167.1	-396.4	-58.8	-44.1
1.5	GCCG-3',5' (2' OAc)-UAAAGC	10	GCCUUACGGC	10	20	30-95	60.0	-198.7	-494.1	-63.7	-45.4
1.6	GCCG-3',5' (2' OAc)-UAAAGC	50	GCCUUACGGC	50	100	30-95	70.1	-353.3	-940.5	-96.4	-61.6
2.1	GCCG-3',5' (2' H)-UAAAGC	5	-	-	5	30-95	72.2	-164.8	-477.6	-34.3	-16.7
2.2	GCCG-3',5' (2' H)-UAAAGC	10	-	-	10	30-95	72.5	-169.6	-491.2	-35.5	-17.3
2.3	GCCG-3',5' (2' H)-UAAAGC	50	-	-	50	30-95	73.0	-185.4	-535.9	-39.0	-19.2
2.4 <sup>b</sup>	GCCG-3',5' (2' H)-UAAAGC	5	GCCUUACGGC	5	10	30-95	58.7	-148.8	-340.0	-56.0	-43.4
2.5 <sup>b</sup>	GCCG-3',5' (2' H)-UAAAGC	10	GCCUUACGGC	10	20	30-95	59.4	-165.4	-394.8	-57.6	-43.0
2.6	GCCG-3',5' (2' H)-UAAAGC	50	GCCUUACGGC	50	100	30-95	63.8	-225.1	-579.8	-66.7	-45.3
3.1	GCCG-3',5' (2' OH)-UAAAGC	5	-	-	5	30-95	73.1	-185.5	-535.8	-39.1	-19.3
3.2	GCCG-3',5' (2' OH)-UAAAGC	10	-	-	10	30-95	74.6	-189.8	-546.0	-40.6	-20.4
3.3	GCCG-3',5' (2' OH)-UAAAGC	50	-	-	50	30-95	74.4	-193.5	-557.0	-41.4	-20.8
3.4 <sup>b</sup>	GCCG-3',5' (2' OH)-UAAAGC	5	GCCUUACGGC	5	10	30-95	59.4	-173.5	-413.1	-60.7	-45.4
3.5 <sup>b</sup>	GCCG-3',5' (2' OH)-UAAAGC	10	GCCUUACGGC	10	20	30-95	61.7	-192.2	-471.6	-63.4	-45.9
3.6	GCCG-3',5' (2' OH)-UAAAGC	50	GCCUUACGGC	50	100	30-95	66.2	-237.6	-612.0	-70.4	-47.8
4.1	-	-	GCCUUACGGC	5	5	30-90	66.7	-171.7	-505.6	-33.6	-14.9
4.2	-	-	GCCUUACGGC	10	10	30-90	68.2	-187.1	-548.3	-37.3	-17.0
4.3	-	-	GCCUUACGGC	50	50	30-90	68.6	-193.0	-565.5	-38.6	-17.7
5.0 <sup>c</sup>	GCCU-3' P	4	ACGGC	4	8	10-70	29.5	n/d			
						70-10	15.3				

Table 11.  $T_m$  and thermodynamic parameters used to assess the effect of acetylation on the secondary structure stability of a GUAA tetraloop. Each measurement used 10 mM  $\text{Na}_2\text{HPO}_4$ , 0.5 mM  $\text{Na}_2\text{EDTA}$  buffer (pH 7) and with 0.1 M NaCl. Data is an average of three heat-cool cycles.

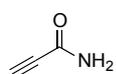
*For measurements that show two-state behavior, errors represent standard deviations of 6 values.  $T_m$  values are within  $\pm 0.7$  °C and thermodynamic data is within  $\pm 5.8\%$  for  $\Delta H^\circ$ , within  $\pm 6.1\%$  for  $\Delta S^\circ$ , within  $\pm 2.29$  kJmol<sup>-1</sup> for  $\Delta G^\circ_0$  and within  $\pm 1.39$  kJmol<sup>-1</sup> for  $\Delta G^\circ_{37}$ .<sup>a</sup> Data was obtained using absorbance at  $\lambda$  280 nm.<sup>[328]</sup> <sup>b</sup> $T_m$  curves were found not to adhere to a two-state model, these  $T_m$  values are tentatively calculated,  $T_m$  values are within  $\pm 2.2$  °C and thermodynamic data is within  $\pm 5.3\%$  for  $\Delta H^\circ$ , within  $\pm 6.6\%$  for  $\Delta S^\circ$ , within  $\pm 1.22$  kJmol<sup>-1</sup> for  $\Delta G^\circ_0$  and within  $\pm 0.95$  kJmol<sup>-1</sup> for  $\Delta G^\circ_{37}$ .<sup>c</sup> Melting curves showed hysteresis, thermodynamic data was not calculated and errors for each  $T_m$  represents standard deviations of 3 values each where the heating  $T_m$  values are within  $\pm 1.6$  °C and the cooling  $T_m$  values are within  $\pm 0.5$  °C. n/d = not determined.*

## 6.4. Procedures for Chapter 4

### 6.4.1. Synthetic procedures for materials used in the aminoacylation reactions

#### Propiolamide 152

$C_3H_3NO$ ;  $M_r = 69.02$



Methyl propiolate (20.0 g, 238 mmol) was added to liquid ammonia (150 mL) at  $-55\text{ }^\circ\text{C}$  and stirred for 24 h under an atmosphere of nitrogen. Excess ammonia was removed by warming the solution to RT and bubbling nitrogen through the solution over 2 h. This resulted in a colourless oil, which was dissolved in  $\text{Et}_2\text{O}$  (200 mL), dried over  $\text{MgSO}_4$  and evaporated to dryness. The crude product was dissolved in warm anhydrous  $\text{CH}_2\text{Cl}_2$  and cooled to  $-20\text{ }^\circ\text{C}$  overnight to give propiolamide as white needle-like crystals (15.8 g, 96%). M.P. =  $57\text{-}59\text{ }^\circ\text{C}$  (Lit.<sup>[330]</sup>  $61\text{-}62\text{ }^\circ\text{C}$ ). IR ( $\text{cm}^{-1}$ ) 3319, 3111 (br., N-H), 3291 (med, alkyne C-H), 2106 (str,  $\text{C}\equiv\text{C}$ ), 1651 and 1613 (s,  $\text{C}=\text{O}$ ).  $^1\text{H}$  NMR (Acetone- $\text{D}_6$ , 400 MHz)  $\delta$  7.44 (1H, s, CO-NH<sub>2</sub>), 7.01 (1H, s, CO-NH<sub>2</sub>), 3.48 (1H, s, **H**-CC).  $^{13}\text{C}$  NMR ( $\text{D}_2\text{O}$ , 100 MHz)  $\delta$  154.8 (CO), 78.5 (HCC-CO), 74.6 (HCC-CO).  $m/z$  ESI+: 70 ( $[\text{M}+\text{H}]^+$ , 100%).

#### Cyanoacetylene 7

$C_3\text{HN}$ ;  $M_r = 51.01$

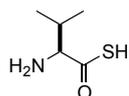


Propiolamide **152** (1.00 g, 14.4 mmol), oven dry chromatography grade sand (8 g) and  $\text{P}_2\text{O}_5$  (4.11 g, 29.0 mmol) were mixed together with a mortar and pestle. The mixture was dry distilled at  $130\text{ }^\circ\text{C}$  under a slight vacuum for 1 h. Cyanoacetylene (550 mg, 74%) was condensed at  $-78\text{ }^\circ\text{C}$  and was isolated as a white colourless solid. The white solid was dissolved immediately in required solvent (at low temperature to prevent loss

of cyanoacetylene, aliquoted and stored at  $-85\text{ }^{\circ}\text{C}$ .  $^1\text{H}$  NMR ( $\text{D}_2\text{O}$ , 300 MHz)  $\delta$  2.54 (s, 1H, H-CC).  $^{13}\text{C}$  NMR ( $\text{D}_2\text{O}$ , 100 MHz)  $\delta$  105.0 (CN), 76.1 (t,  $J = 41.1$  Hz, DCC), 55.8 (t,  $J = 7.8$  Hz, DCC-).

**(S)-2-Amino-3-methylbutanethioic S-acid (thiovaline) 134**

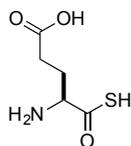
$\text{C}_5\text{H}_{11}\text{NOS}$ ;  $M_r = 133.21$



Boc-Val-OH **135** (500 mg, 2.30 mmol) and *N*-methyl morpholine (0.38 mL, 3.45 mmol) was dissolved in anhydrous THF (5 mL). Isobutyl chloroformate (0.45 mL, 3.45 mmol) was added dropwise at  $-20\text{ }^{\circ}\text{C}$ , the resultant mixture was warmed to  $0\text{ }^{\circ}\text{C}$  and stirred for 1 h. A suspension of  $\text{Li}_2\text{S}$  (210 mg, 4.60 mmol) in anhydrous DMF (10 mL) was added to the solution of activated Boc-val-OH *via* a cannula. The resultant green mixture was stirred for 1 hr at  $0\text{ }^{\circ}\text{C}$ , then water (15 mL) was added and the mixture further stirred for another 1 hr. The solution was adjusted to  $\text{pH} = 3$  with 1 M HCl and extracted with EtOAc ( $3 \times 20$  mL). The organic layer was washed with water ( $3 \times 30\text{ mL}$ ), then with brine (20 mL) and dried over  $\text{MgSO}_4$ . The solvent was removed under vacuum giving the crude Boc-Val-SH **136** (512 mg) as a yellow oil. The crude residue was dissolved in freshly distilled TFA (5 mL) and stirred for 1 h at  $0\text{ }^{\circ}\text{C}$  and at RT for a further 1 h. The mixture was evaporated to dryness under vacuum and the cream coloured residue was triturated with anhydrous  $\text{Et}_2\text{O}$  (20 mL). The precipitate was collected by filtration under a nitrogen atmosphere and further washed with anhydrous  $\text{Et}_2\text{O}$  (20 mL). The wet solid was dried under vacuum and then dissolved in water (5 mL). The resultant solution was filtered and the supernatant lyophilized to give the title compound (209 mg, 69%) as a colourless powder. M.P. =  $340\text{ }^{\circ}\text{C}$  (decomp.). IR ( $\text{cm}^{-1}$ ) 3024 (wk, N-H), 2964, 2932, 2875 (C-H), 2623 (wk, S-H), 1477 (str, C=O).  $^1\text{H}$  NMR ( $\text{D}_2\text{O}$ , 400 MHz)  $\delta$  3.77 (d,  $J = 4.6$  Hz, 1H, H-C(2)), 2.43 (hept.d,  $J = 7.0, 4.6$  Hz, 1H, H-C(3)), 0.98 (d,  $J = 7.0$  Hz, 3H,  $\text{H}_3\text{-C}(4)$ ), 0.85 (d,  $J = 7.0$  Hz, 3H,  $\text{H}_3\text{-C}(4')$ ).  $^{13}\text{C}$  NMR ( $\text{D}_2\text{O}$ , 100 MHz)  $\delta$  214.4 (CO-SH), 68.1 (C(2)), 30.2 (C(3)), 18.4, 15.5 (C(4)+C(4')). *m/z* ESI+: 133 ( $[\text{M}]^+$ , 100%); ESI-HRMS (pos.):  $[\text{M}-\text{H}]^-$  calculated for  $\text{C}_5\text{H}_{10}\text{NOS}$ , 132.0488; found 132.0492.

## $\alpha$ -Thioglutamic Acid 146

C<sub>5</sub>H<sub>9</sub>NO<sub>3</sub>S; M<sub>r</sub> = 163.19

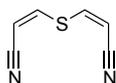


Boc-L-glutamic acid 5-*tert*-butyl ester **147** (400 mg, 1.32 mmol) and *N*-methyl morpholine (200  $\mu$ L, 1.98 mmol) were dissolved in anhydrous THF (5 mL). Isobutyl chloroformate (260  $\mu$ L, 1.97 mmol) was added dropwise at  $-20$  °C. The resultant mixture was warmed to 0 °C and stirred for 1 h. A suspension of Li<sub>2</sub>S (120 mg, 2.64 mmol) in anhydrous DMF (10 mL) was added to the solution of activated Boc-L-glutamic acid 5-*tert*-butyl ester *via* a cannula. The resultant pink mixture was stirred for 1 h at 0 °C, water (15 mL) was added and the mixture further stirred for 1 hr. The solution was adjusted to pH = 3 and extracted with EtOAc (3  $\times$  20 mL). The organic layer was washed with water (3  $\times$  30 mL), then with brine (20 mL) and dried over MgSO<sub>4</sub>. The solvent was removed under vacuum resulting in the crude Boc-L-glutamic thioacid 5-*tert*-butyl ester **148** (428 mg) as a pink crystalline solid. The pink crystalline solid was dissolved in freshly distilled TFA (10 mL) and stirred for 1 h at 0 °C and at RT for a further 1 h. The solvent was removed under vacuum and the cream coloured residue was triturated with anhydrous Et<sub>2</sub>O (20 mL). The residue was suspended in anhydrous diethyl ether (20 mL), the suspension was centrifuged and the solvent decanted and the centrifugal workup was repeated twice more. The wet solid was dried under vacuum, dissolved in water (5 mL), filtered and finally lyophilized to give the title compound (196 mg, 91%) as a hygroscopic off-white solid. After characterisation the solid was dissolved in H<sub>2</sub>O at the desired concentration, and 1 M NaOH was added to adjust the solution to pH = 6.5 and the resultant solution stored at  $-85$  °C. M.P. 58-60 °C. IR (cm<sup>-1</sup>) 2964 (br, CO-OH), 2237 (-NH<sub>3</sub><sup>+</sup>) 1691 and 1660 (str, C=O). <sup>1</sup>H NMR (D<sub>2</sub>O, 500 MHz)  $\delta$  3.93 (t, *J* = 6.3 Hz, 1H, H-C(2)), 2.47 (t, *J* = 7.6 Hz, 2H, H<sub>2</sub>-C(4)), 2.03-2.18 (m, 2H, H<sub>2</sub>-C(3)). <sup>13</sup>C NMR (D<sub>2</sub>O, 125 MHz)  $\delta$  213.8 (C(1)), 176.6 (C(5)), 61.7 (C(2)), 29.3 (C(4)), 26.8 (C(3)). *m/z* ESI<sup>-</sup>: 162 ([M-H]<sup>-</sup>, 54%), 128 ([M-H<sub>2</sub>S]<sup>-</sup>, 100%); ESI-HRMS (pos.): [M-H]<sup>-</sup> calculated for C<sub>5</sub>H<sub>8</sub>NO<sub>3</sub>S, 162.0230; found 162.0236.

*Solid precipitate isolated from aminoacylation reactions*

**(2Z,2'Z)-3,3'-Sulfanediylbisprop-2-enitrile 139**

C<sub>6</sub>H<sub>4</sub>N<sub>2</sub>S; M<sub>r</sub> = 136.17

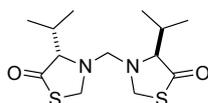


<sup>1</sup>H NMR (MeOD-d<sub>4</sub>, 300 MHz) δ 7.75 (d, *J* = 10.4 Hz, 2H, -SCHCH-), 5.82 (d, *J* = 10.4 Hz, 2H, -SCHCH-). <sup>13</sup>C NMR (MeOD-d<sub>4</sub>, 300 MHz) δ 147.9 (2 × -SCHCH-), 115.7 (CN), 97.9 (2 × -SCHCH-). *m/z* ESI: 136 ([M]<sup>+</sup>, 100%). ESI-HRMS (pos.): [M]<sup>+</sup> calculated for C<sub>6</sub>H<sub>4</sub>N<sub>2</sub>S<sub>1</sub>N, 136.0090; found 136.0087.

*Preparative synthesis of 144*

**(4S,4'S)-3,3'-Methylenebis(4-isopropylthiazolidin-5-one) 144**

C<sub>13</sub>H<sub>22</sub>N<sub>2</sub>O<sub>2</sub>S<sub>2</sub>; M<sub>r</sub> = 302.11



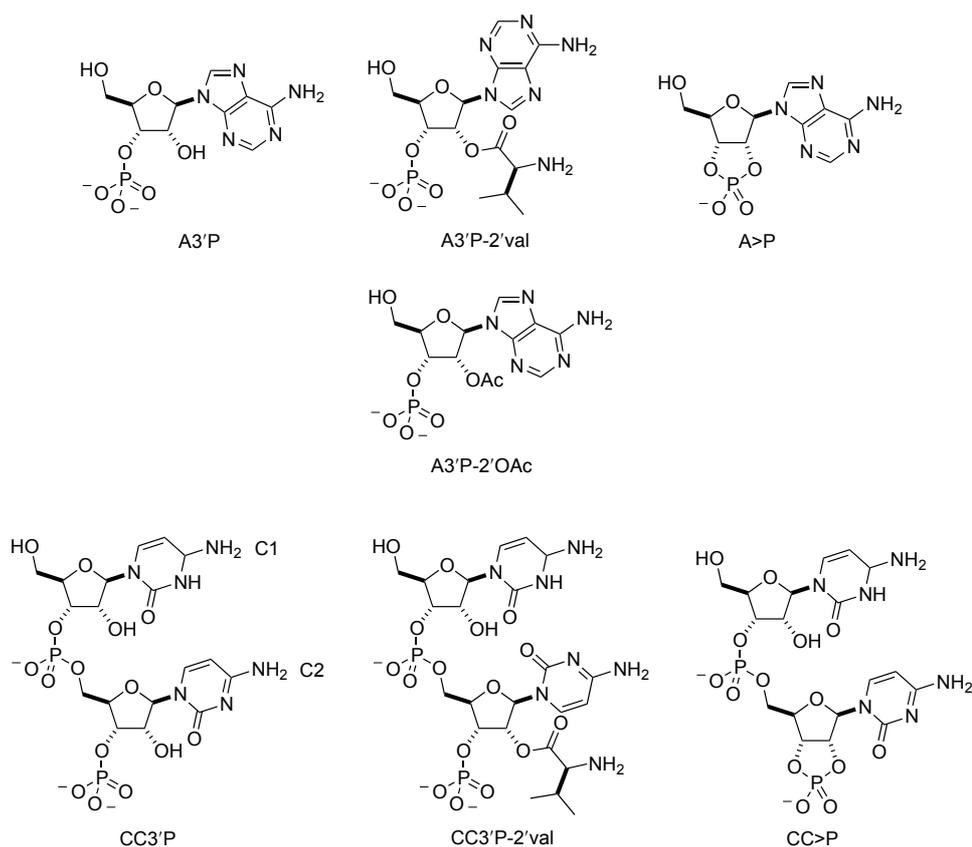
Thiovaline **134** (100 mg, 0.75 mmol) was dissolved in H<sub>2</sub>O (5 mL) and the solution adjusted to pH = 6.5 with 1 M NaOH solution. Formaldehyde **2** solution (0.22 mL, 13.4 M) was then added and the solution readjusted to pH = 6.5. The resultant mixture was allowed to stir for 2 h. The precipitate was separated by centrifugation and the pellet was washed with water (3 × 10 mL). The wet solid was dried under high vacuum to yield the title compound as a colourless powder (67.6 mg, 60%). M.P. = 123-124 °C. IR (cm<sup>-1</sup>) ν<sub>max</sub>: 2977, 2959, 2922 (C-H), 1686 (s, C=O), 2886 and 2868 (CH<sub>2</sub>-N). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 400 MHz) δ 5.02 (AB, *J*<sub>AB</sub> = 11.0 Hz, 2H, H-(C2)+H-(C2')), 4.93 (BA, *J*<sub>BA</sub> = 11.0 Hz, 2H, H-(C2)+H-(C2')), 3.48 (s, 2H, N-CH<sub>2</sub>-N), 3.13 (d, *J* = 8.9 Hz, 2H, H-C(4)+H-C(4')), 1.96 (dhept., *J* = 8.9, 6.5 Hz, 2H, -CHCH(CH<sub>3</sub>)<sub>2</sub>), 1.02 (d, *J* = 6.5 Hz, 6H, -CHCH(CH<sub>3</sub>)<sub>2</sub>), 0.98 (d, *J* = 6.5 Hz, 6H, -CHCH(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (DMSO-*d*<sub>6</sub>, 100 MHz) δ 209.8 (C(5)+C(5')), 78.2 (C(4)+C(4')), 75.3 (N-CH<sub>2</sub>-N), 58.6 ((C2)+(C2')), 28.3 (-CHCH(CH<sub>3</sub>)<sub>2</sub>), 20.2 (-CHCH(CH<sub>3</sub>)<sub>2</sub>), 18.9 (-CHCH(CH<sub>3</sub>)<sub>2</sub>). *m/z* ESI+: 325

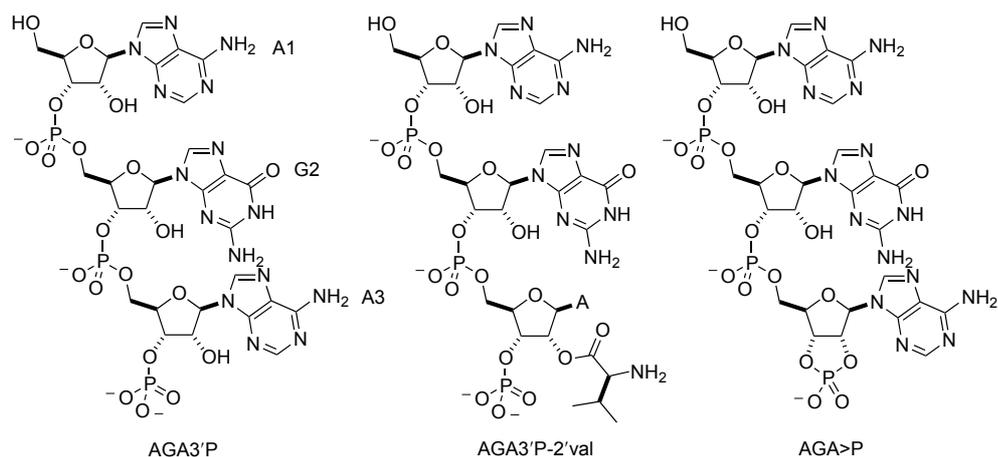
([M+Na]<sup>+</sup>, 40%), 341 ([M+K]<sup>+</sup>, 100%); ESI-HRMS (pos.): [M+Na]<sup>+</sup> calculated for C<sub>13</sub>H<sub>22</sub>N<sub>2</sub>O<sub>2</sub>S<sub>2</sub>Na, 325.1015; found 325.1020.

## 6.4.2. Procedures for aminoacylation using amino thioacids and various electrophilic activators

The formation (and quantification) of aminoacyl-nucleoside phosphate(s) was observed by NMR spectroscopic analysis. Characteristic downfield shifted H-(C2') was detected by <sup>1</sup>H-NMR and <sup>1</sup>H-<sup>1</sup>H COSY analysis, in tandem with <sup>31</sup>P NMR spectroscopy. Percentage yields of various species of nucleoside phosphate(s) were calculated by using 100% as the summation of integrals of H-C(1') from each species.

*Nomenclature for key species observed in the aminoacylation reactions*





### Table legends

*n.d.*, not detectable. *n.a.*, not assignable due to overlapping of signals/not applicable. -, not obtained. *obs.*, obscured. *part. obs.*, partially obscured.

### General procedure for the aminoacylation of nucleoside-3'-phosphates using cyanoacetylene 9

In a 1.5 mL plastic vial, nucleoside-3'-monophosphate **N3'P** (100  $\mu$ M) was dissolved in D<sub>2</sub>O (0.8 mL) by addition of 1 M NaOD to adjust the solution to pD = 6.5. To the solution was added an amino thioacid or thioacetic acid (see tables for concentration) followed by cyanoacetylene **7** (200  $\mu$ M, D<sub>2</sub>O), a rapid increase in pD was observed and returned to pD = 6.5 by addition of 1 M DCl. The solution was immediately analysed by NMR spectroscopy. Concentrations and values as above unless otherwise stated in the following tables.

### Exploratory reaction of $\beta$ -D-adenosine-3'-phosphate A3'P, thiovaline 134 and cyanoacetylene 7

Entry	Nucleotide	Products and residual starting materials/%							
		134 ( $\mu$ M)	NaSAc ( $\mu$ M)	DCCCN ( $\mu$ M)	Time (mins)	A3'P	A3'P-2'val	A3'P-2'OAc	A>P
1	A3'P	100	-	200	60	85	17	-	6
2	A3'P	-	100	200	65	40	-	60	0

Table 12. Preliminary aminoacylation experiment (entry 2 is a control for cyanoacetylene).

	A3'P	A3'P-2'val	A>P
H-C(8)	8.25 (s)	8.30 (s)	8.20 (s)
H-C(2)	8.06 (s)	8.06 (s)	8.06 (s)
H-C(1')	6.01 (d, $J = 6.3$ )	6.28 (d, $J = 6.6$ )	6.18 (d, $J = 4.4$ )
H-C(2')	obs.	5.80 (app. t, $J = 5.9$ )	5.38-5.33 (m)
H-C(3')	obs.	obs.	5.12-5.07 (m)
H-C(4')	4.43 (q, $J = 2.9$ )	4.50 (q, $J = 2.8$ )	obs.
H <sub>2</sub> -C(5')	3.98-3.81 (m)		
P	1.21 (d, $J = 8.1$ )	1.43 (d, $J = 8.8$ )	19.99-19.88 (m)
P (dec)	1.21 (s)	1.43 (s)	19.94 (s)

Table 13. Characterisation of products from the reaction described in Table 12, entry 1.

	A3'P	A3'P-2'OAc	A>P
H-C(8)	8.24 (s)	8.24 (s)	n.d.
H-C(2)	8.04 (s)	8.05 (s)	n.d.
H-C(1')	6.01 (d, $J = 6.3$ )	6.20 (d, $J = 5.6$ )	n.d.
H-C(2')	obs.	5.61 (t, $J = 5.6$ )	n.d.
H-C(3')	obs.	4.93-4.88 (m) part obs.	n.d.
H-C(4')	4.41 (q, $J = 3.3$ )	4.46 (q, $J = 3.3$ )	n.d.
H <sub>2</sub> -C(5')	3.97-3.82 (m)		n.d.
P	2.52 (d, $J = 7.6$ )	0.99 (d, $J = 8.7$ )	n.d.
P (dec)	2.52 (s)	0.99 (s)	n.d.

Table 14. Characterisation of products from the control reaction described in Table 12, entry 2.

## Aminoacylation reactions at a range of pHs

Entry	Nucleotide	134 ( $\mu$ M)	DCCCN ( $\mu$ M)	Initial pD	Equivalent pH <sup>[331]</sup>	Time (t: min)	Products and residual starting materials/%		
							A3'P	A3'P-2OAc	A>P
1	A3'P	100	200	4.6	5.0	0:00	100	0	0
						0:13	84	13	3
						1:43	86	11	3
						3:27	86	11	-
						6:10	88	10	5
						23:52	90	8	5
2	A3'P	100	200	5.6	6.0	0:00	100	0	0
						0:07	85	12	3
						1:26	-	-	a
						3:09	85	9	6
						5:52	90	5	4
						24:15	95	2	3
3	A3'P	100	200	6.1	6.5	0:00	100	0	0
						0:15	87	11	3
						1:12	88	9	2
						2:56	91	7	3
						5:39	94	4	3
						24:03	97	0	2
						49:57	97	0	2
						78:06	97	0	3
						96:30	97	0	2
						196:45	96	0	4

Table 15. Yields of product and residual starting materials of aminoacylation reaction conducted at various pDs. <sup>a</sup> Poor shim.

	A3'P	A3'P-2'val	A>P
H-C(8)	8.22 (s)	8.26 (s)	8.18 (s)
H-C(2)	8.00 (s)	8.00 (s)	8.00 (s)
H-C(1')	5.98 (d, $J = 6.2$ )	6.24 (d, $J = 5.8$ )	6.14 (d, $J = 4.7$ )
H-C(2')	obs.	5.77 (app. t, $J = 5.3$ )	5.35-5.28 (m)
H-C(3')	obs.	4.95 (dt, $J = 8.3, 4.1$ part Obs.)	5.12-5.02 (m)
H-C(4')	4.46-4.40 (m)	4.51-4.47 (m)	n.a.
H <sub>2</sub> -C(5')		3.94-3.83 (m)	
P	0.21 (d, $J = 8.4$ )	-0.18 (d, $J = 8.3$ )	19.96-19.83 (m)
P (dec)	0.20 (s)	-0.18 (s)	19.89 (s)

Table 16. Characterisation of products from the reaction described in Table 15, entry 1 at time point 0:13 h.

	A3'P	A3'P-2'val	A>P
H-C(8)	8.21 (s)	8.26 (s)	8.15 (s)
H-C(2)	7.98 (s)	8.00 (s)	7.98 (s)
H-C(1')	5.98 (d, $J = 6.3$ )	6.24 (d, $J = 6.5$ )	6.13 (d, $J = 4.3$ )
H-C(2')	obs.	5.76 (app. t, $J = 5.9$ )	5.31 (ddd, $J = 11.0, 6.8, 4.6$ )
H-C(3')	4.69 (ddd, $J = 8.1, 5.2, 3.0$ )	obs.	5.07 (td, $J = 7.3, 4.2$ )
H-C(4')	4.41 (q, $J = 2.7$ )	4.49 (q, $J = 2.6$ )	n.a.
H <sub>2</sub> -C(5')	3.97-3.81 (m)		
P	1.71 (d, $J = 6.5$ )	obs.	n.a.
P (dec)	1.69 (s)	1.77 (s)	19.91 (s)

Table 17. Characterisation of products from the reaction described in Table 15, entry 2 at time point 0:07 h.

	A3'P	A3'P-2'val	A>P
H-C(8)	8.21 (s)	8.27 (s)	8.15 (s)
H-C(2)	7.98 (s)	8.00 (s)	7.97 (s)
H-C(1')	5.98 (d, $J = 6.2$ )	6.24 (d, $J = 6.8$ )	6.13 (d, $J = 4.4$ )
H-C(2')	4.74 (t, $J = 5.7$ )	5.75 (dd, $J = 6.9, 5.1$ )	5.31 (ddd, $J = 11.1, 6.7, 4.5$ )
H-C(3')	4.69 (ddd, $J = 8.1, 5.2, 3.0$ )	obs.	5.07 (td, $J = 7.3, 4.2$ )
H-C(4')	4.40 (q, $J = 3.0$ )	4.49 (q, $J = 2.6$ )	n.a.
H <sub>2</sub> -C(5')		3.93-3.81 (m)	
P	2.49 (d, $J = 7.0$ )	n.a.	19.93 (d, $J = 8.8$ )
P (dec)	2.48 (s)	2.44 (s)	19.91 (s)

Table 18. Characterisation of products from the reaction described in Table 15, entry 3 at time point 0:15 h.

### Competition reaction between thiovaline 134 and thioacetate 43 for acylation of $\beta$ -D-adenosine-3'-phosphate A3'P

Entry	Nucleotide (100 $\mu$ M)	134 ( $\mu$ M)	NaSAc ( $\mu$ M)	DCCCN ( $\mu$ M)	pH	Time (h:min)	Products and residual starting materials/%						
							A3'P	A3'P-2'Val	A3'P-2'OAc	A>P			
Thioacetic acid 43 and thiovaline 134 competition experiment													
1	A3'P	100	100	400	6.5	0:00	100	0	0	0			
						0:22	87	4	4	<1%			
						3:19	83	4	5	6			
						7:04	89	3	4	4			
						98:54	88	2	3	5			
Thiovaline 134 control experiment													
2	A3'P	100	-	200	6.5	0:00	100	0	-	0			
						0:32	75	21	-	5			
						2:36	70	20	-	9			
						6:10	77	12	-	7			
						98:01	88	3	-	7			
Thioacetic acid 43 control													
3	A3'P	-	100	200	6.5	0:00	100	-	0	n.d.			
						0:46	33	-	56	n.d.			
						2:59	36	-	59	n.d.			
						6:44	35	-	59	n.d.			
						98:34	36	-	24	n.d.			

Table 19. Results of a competition acylation of  $\beta$ -D-adenosine-3'-phosphate A3'P between thioacetate 43 and thiovaline 134. Entries 2 and 3 are control reactions for comparison.

	A3'P	A3'P-2'val	A3'P-2'OAc	A>P	Other unassigned trace derivatives
H-C(8)			8.25 (s), 8.22 (s), 8.20 (s)		
H-C(2)			8.01 (s), 8.00 (s), 7.99 (s), 7.98 (s)		
H-C(1')	5.96 (d, $J = 6.4$ )	6.23 (d, $J = 6.6$ )	6.18 (d, $J = 5.7$ )	6.12 (d, $J = 4.1$ )	6.20-6.18 (m)
H-C(2')	4.76-4.73 obs.	5.75 (dd, $J = 6.5, 5.2$ )	5.58 (t, $J = 5.7$ )	5.34-5.27 (m)	5.68 (t, $J = 5.6$ )
H-C(3')	4.68 (ddd, $J = 8.1, 5.2, 2.9$ )		4.91-4.83 (m)	5.08-5.03 (m)	-
H-C(4')	4.40 (q, $J = 2.9$ )	n.a.	n.a.	n.a.	-
H-C(5')	3.87 (dd, $J = 13.0, 2.9$ )	n.a.	n.a.	n.a.	-
H-C(5'')	3.81 (dd, $J = 13.0, 2.9$ )	n.a.	n.a.	n.a.	-
P	1.43 (d, $J = 7.8$ )	n.a.	n.a.	n.a.	-
P (dec)	1.43 (s)	n.a.	n.a.	n.a.	0.10 (s), -0.32 (s)

Table 20. Characterisation of products from the reaction described in Table 19, entry 1 at time point 0:22 h.

	A <sup>3</sup> P	A <sup>3</sup> P-2'val	A>P
H-C(8)	8.18 (s)	8.24 (s)	8.13 (s)
H-C(2)	7.95 (s)	7.97 (s)	7.95 (s)
H-C(1')	5.95 (d, <i>J</i> = 6.3)	6.22 (d, <i>J</i> = 6.4)	6.11 (d, <i>J</i> = 4.5)
H-C(2')	obs.	5.74 (app. t, <i>J</i> = 5.9)	5.29 (ddd, <i>J</i> = 10.8, 6.7, 4.4)
H-C(3')	4.70-4.65 (m) part obs.	4.89-4.84 (m) part obs.	5.05 (td, <i>J</i> = 7.2, 3.9)
H-C(4')	4.40 (q, <i>J</i> = 3.0)	4.47 (q, <i>J</i> = 2.8)	n.a.
H <sub>2</sub> -C(5')	3.95-3.78 (m)		
P	1.09 (d, <i>J</i> = 8.4)	1.18 part obs.	20.03-19.77 (m)
P (dec)	1.09 (s)	1.16 (s)	19.91 (s)

Table 21. Characterisation of products from the reaction described in Table 19, entry 2 at time point 0:32 h.

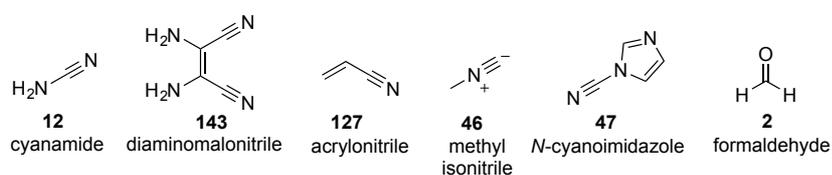
	A3'P	A3'P-2'OAc	A>P	Other unassigned trace derivatives	
$\delta$ /ppm (multiplicity, J/Hz)	H-C(8)	8.24 (s)	n.d.	n.a.	
	H-C(2)	7.98 (s)	n.d.	n.a.	
	H-C(1')	5.97 (d, $J = 6.3$ )	6.18 (d, $J = 5.9$ )	n.d.	6.24 (d, $J = 2.8$ )
	H-C(2')	obs.	5.58 (app. t, $J = 5.6$ )	n.d.	6.01-5.99 obs. (t, $J = 5.0$ )
	H-C(3')	4.68 (ddd, $J = 7.9, 5.1, 3.0$ ) part obs.	4.90 (dt, $J = 8.5, 4.3$ ) part obs.	n.d.	5.14-5.09 (m)
	H-C(4')	4.41-4.38 (m)	4.46-4.42 (m)	n.d.	n.a.
	H <sub>2</sub> -C(5')	4.00-3.78 (m)		n.d.	n.a.
	P	1.67 (d, $J = 7.9$ )	0.26 (d, $J = 8.8$ )	n.d.	n.a.
	P (dec)	1.67 (s)	0.26 (s)	n.d.	n.a.

Table 22. Characterisation of products from the reaction described in Table 19, entry 3 at time point 0:46 h.

## Investigation into aminacylation reactions with alternative activating electrophiles

Electrophiles below were mixed with thiovaline **134** only to assess whether the two would react. The electrophile and thiovaline **134** were dissolved/suspended in D<sub>2</sub>O and the solution adjusted to pD = 6.1. The resultant solution was immediately submitted for NMR analysis and again after reaction overnight. Formaldehyde **2** was further investigated (see below).

Electrophiles **46-47** are known activators of thioacetate **43** and so aminoacylation was attempted with nucleotide-3'-phosphate using the general procedure (see above).



Entry	Nucleotide (100 mM)	Val-SH (mM)	Electrophile (mM)	pH	Time (h)	Observation	Products and residual starting materials/%		
							A3'P	A3'P-2'val	A>P
1	-	100	12 (200)	6.5	19	No reaction	-	-	-
2	-	100	127 (200)	6.5	15	Slow conversion of Val-SH to at least two products	-	-	-
3	-	100	143 (200)	6.5	21	No reaction, DAMN insoluble	-	-	-
4	A3'P	200	46 (400)	6.5	3	-	97	n.d.	3
5	A3'P	200	47 (200)	6.4	4	-	57	12	31
6	-	100	2 (200)	6.5	2	Solution NMR show residual SM, with solid precipitate (see below)	-	-	-

Table 23. Entries 1-4 show reactions of various electrophiles with thiovaline **134**, and entries 5-6 are reactions with nucleotide-3'-phosphate with thiovaline **134** and various electrophiles.

	Products and residual starting materials/%			
	A3'P	A3'P-2'val	A>P	
$\delta$ /ppm (multiplicity, J/Hz)	H-C(8)	8.25 (s)	n.d.	8.25 (s)
	H-C(2)	8.03 (s)	n.d.	8.03 (s)
	H-C(1')	6.00 (d, J = 6.3)	n.d.	6.16 (d, J = 4.4)
	H-C(2')	obs.	n.d.	5.33 (ddd, J = 10.9, 6.8, 4.5)
	H-C(3')	4.70 (ddd, J = 8.1, 5.2, 2.9)	n.d.	5.08 (ddd, J = 10.8, 5.4, 2.9)
	H-C(4')	4.42 (q, J = 3.0)	n.d.	n.d.
	H-C(5')	3.90 (dd, J = 13.0, 2.8)	n.d.	n.d.
	H-C(5'')	3.85 (dd, J = 13.0, 3.3)	n.d.	n.d.
	P	2.46 (d, J = 7.8)	n.d.	19.87 (dd, J = 10.2, 7.6)
	P (dec)	2.47 (s)	n.d.	19.87 (s)

Table 24. Characterisation of products from the reaction described in Table 23, entry 4.

	A3'P	A3'P-2'val	A>P
H-C(8)	8.21 (s)	8.21 (s)	8.16 (s)
H-C(2)	8.00 (s)	8.03 (s)	8.00 (s)
H-C(1')	5.96 (d, $J = 6.1$ )	6.17 (d, $J = 6.3$ )	6.13 (d, $J = 3.8$ )
H-C(2')	4.75-4.71 obs.	5.78-5.75 (m)	5.31 (dt, $J = 10.6, 5.4$ )
H-C(3')	4.70-4.65 (m) Part obs.	obs.	5.09-5.04 (m)
H-C(4')	4.40-4.37 (m) part obs.	4.55-4.52 (s) br.	4.43-4.40 (m) part obs.
H <sub>2</sub> -C(5')	3.96-3.77 (m)		
P	2.96 (d, $J = 7.6$ )	0.92 (d, $J = 8.1$ )	19.88 (dd, $J = 10.1, 8.1$ )
P (dec)	2.96 (s)	0.91 (s)	19.88 (s)

Table 25. Characterisation of products from the reaction described in Table 23, entry 5.

*Detailed procedure for Table 23, entry 6*

Thiovaline **134** (13.8 mg, 0.1 mmol) was dissolved in D<sub>2</sub>O (1 mL) and the solution adjusted to pD = 6.1 with 1 M NaOD. Formaldehyde (15.0 μL, 0.2 mmol) was then added upon which a solid precipitate formed. The mixture was analysed NMR spectroscopy and after 2 h the mixture lyophilised and the resultant residue redissolved in *d*<sub>6</sub>-DMSO and resubmitted to NMR analysis.

Lyophilised residue was found to contain mainly a formaldehyde-thiovaline derivative that was found to be (4*S*,4'*S*)-3,3'-methylenebis(4-isopropylthiazolidin-5-one) **144** and a trace of starting material. Data as given in section 6.4.1.

*Reaction of thioglutamic acid 146 and formaldehyde 2*

Thioglutamic acid **146** (0.5 M, 0.2 mL) was diluted with H<sub>2</sub>O (0.8 mL) and the solution adjusted to pH = 6.5. Formaldehyde **2** (13.4 M, 44.8 μL) was added and pH of the solution readjusted as necessary with 1 M HCl/1 M NaOH. The solution was stirred for 3 h, lyophilised and finally taken up in D<sub>2</sub>O. The mixture was analysed by NMR spectroscopy and found to contain mainly glutamate **151**. *m/z* ESI<sup>-</sup>: 128 (20%), 146 ([Glu-H]<sup>-</sup>, 100%); ESI<sup>+</sup>: 258 (85%), 285 (100%), 286 (100%).

*Reaction between thioglutamic acid 146, thiovaline 134 and formaldehyde 2*

Thioglutamic acid **146** (0.5 M, 0.2 mL) was diluted with H<sub>2</sub>O (0.8 mL), thiovaline **134** (13.3 mg, 0.1 mmol) was added and the solution adjusted to pH = 6.5. Formaldehyde **2** (13.4 M, 89.5 μL) was added and the solution readjusted as necessary with 1 M HCl/1 M NaOH. The solution was stirred for 3 h, during which a white precipitate formed. The precipitate was filtered, washed with water and then dried under vacuum and the solid was taken up in *d*<sub>6</sub>-DMSO. The supernatant was lyophilised and the residue taken up in D<sub>2</sub>O. The solid and filtrate were analysed by NMR spectroscopy. The solid precipitate was found to be **144** and the supernatant a mixture of glutamic acid **151** derivatives and valine **54**. *m/z* ESI<sup>-</sup>: 146 ([Glu-H]<sup>-</sup>, 100%), 158 (40%); ESI<sup>+</sup>: 219 (60%), 249 (75%), 258 (100%), 279 (40%), 371 (40%).

*Reaction of  $\beta$ -D-adenosine-3'-phosphate **A3'P** with thioglutamic acid **146** and formaldehyde **2***

$\beta$ -D-Adenosine-3'-phosphate **A3'P** (34.7 mg, 0.1 mmol) was dissolved in D<sub>2</sub>O (0.6 mL) and the solution was adjusted to pD = 6.1 with 1 M NaOD, thioglutamic acid **146** (0.2 mL, 0.5 M/D<sub>2</sub>O, 0.1 mmol) was then added and the solution readjusted to pD = 6.1. Formaldehyde **2** (22.4  $\mu$ l, 0.3 mmol) was added and the pD of the resultant mixture was readjusted to pD = 6.1. The sample was immediately analysed by NMR spectroscopy.

**Aminoacylation of dimer and trimer nucleoside-3'-phosphates**

The dimer/trimer (50 mM) was dissolved in D<sub>2</sub>O (0.8 mL) and thiovaline **134** (50  $\mu$ M) was added and the solution adjusted to pD = 6.1 with 1 M DCl/1 M NaOD as necessary. Cyanoacetylene (100 mM) was added with stirring and an increase in pD was observed and the solution readjusted to pD = 6.1 with 1 M DCl. The sample was immediately analysed by NMR spectroscopy.

In the case of the CC3'P aminoacylation was determined by the distinctive downfield shift (> 1 ppm) of the H-C(2') proton of C2, that was identified by its coupling pattern and chemical shift (app. t,  $J = 4.7$  Hz). A COSY coupling was observed to a downfield shifted H-C(1'). Extent of aminoacylation was based on the integration of the downfield shifted H-C(2') and confirmed by integration of the <sup>31</sup>P CPD spectrum. Based on the chemical shifts of H-C(5) and H-C(6) no aminoacylation of the nucleobases were observed. Aminoacylation of C1 HO-C(2') was not observed because there was no a second <sup>31</sup>P CPD peak corresponding to the phosphate diester of an aminoacylated species. Other sites of aminoacylation could not be determined due to extensive peak overlap.

For the reaction AGA3'P reaction, aminoacylation was determined by the downfield shift (> 0.2 ppm) of the H-C(1') proton of A3, which was identified by its coupling pattern and chemical shift (d,  $J = 5.7$  Hz). A COSY coupling was observed to a downfield shifted H-C(2') but this peak was overlapped which hindered integration. The extent of aminoacylation was based on the integration of the downfield shifted H-C(1') and supported by integration of the <sup>31</sup>P CPD spectrum. Aminoacylation of A1 and G2 HO-C(2') was suspected due to observation of several upfield shifted <sup>31</sup>P CPD peaks corresponding to the phosphate diesters, these correspond to aminoacylation in the order

of < 2.5%. However, confirmation with  $^1\text{H}$  NMR was not possible due to extensive peak overlap and poor COSY couplings.

Entry	Oligonucleotide	Time (h)	Products and residual starting materials/%						
			CC3'P	CC3'P-2'val	CC>P	AGA3'P	AGA3'P-2'val	Unidentified aminoacyl species	AGA>P
1	CC3'P <sup>a</sup>	1	90	8	<2	-	-	-	-
2	AGA3'P <sup>b</sup>	0:10	-	-	-	~85	~12	~7	~4

*Table 26. Results of aminoacylation reaction of a dimer and a trimer with thiovaline 134 and cyanooacetylene 7. <sup>a</sup>Dimer was synthesised by S. Islam <sup>b</sup>Trimer was synthesised by F.R. Bowler.*

## 7. References

- [1] D. L. Abel, *Life* **2011**, 2, 106-134.
- [2] P. L. Luisi, *The Emergence of Life - From Chemical Origins to Synthetic Biology*, Cambridge University Press, Cambridge, **2006**.
- [3] G. F. Joyce, R. Young, S. Chang, B. Clark, D. Deamer, D. DeVincenzi, J. Ferris, W. Irvine, J. Kasting, J. Kerridge, H. Klein, A. Knoll, J. Walker, *Origins of Life: The Central Concepts*, Jones and Bartlett, Boston, **1994**.
- [4] A. Eschenmoser, *Tetrahedron* **2007**, 63, 12821-12843.
- [5] L. E. Orgel, *Trends Biochem. Sci.* **1998**, 23, 491-495.
- [6] N. H. Sleep, K. J. Zahnle, J. F. Kasting, H. J. Morowitz, *Nature* **1989**, 342, 139-142.
- [7] M. D. Brasier, O. R. Green, A. P. Jephcoat, A. K. Kleppe, M. J. Van Kranendonk, J. F. Lindsay, A. Steele, N. V. Grassineau, *Nature* **2002**, 416, 76-81.
- [8] J. W. Schopf, *Science* **1993**, 260, 640-646.
- [9] S. B. Hedges, H. Chen, S. Kumar, D. Wang, A. Thompson, H. Watanabe, *BMC Evolutionary Biology* **2001**, 1, 4.
- [10] J. F. Kasting, *Science* **1993**, 259, 920-926.
- [11] H. B. Niemann, S. K. Atreya, S. J. Bauer, G. R. Carignan, J. E. Demick, R. L. Frost, D. Gautier, J. A. Haberman, D. N. Harpold, D. M. Hunten, G. Israel, J. I. Lunine, W. T. Kasprzak, T. C. Owen, M. Paulkovich, F. Raulin, E. Raaen, S. H. Way, *Nature* **2005**, 438, 779-784.
- [12] A. Eschenmoser, E. Loewenthal, *Chem. Soc. Rev.* **1992**, 21, 1-16.
- [13] R. A. Sanchez, J. P. Ferris, L. E. Orgel, *Science* **1966**, 154, 784-785.
- [14] S. L. Miller, *Science* **1953**, 117, 528-529.
- [15] S. Pizzarello, E. Shock, *Cold Spring Harb. Perspect. Biol.* **2010**, 2.
- [16] J. R. Cronin, *Adv. Space Res.* **1989**, 9, 59-64.
- [17] N. H. Sleep, *Cold Spring Harb. Perspect. Biol.* **2010**, 2.
- [18] F. Crick, *Nature* **1970**, 227, 561-563.
- [19] J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry*, Fifth ed., W. H. Freeman and Co., New York, **2002**.

- [20] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 2179-2183.
- [21] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737-738.
- [22] [http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/2009/press.html](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2009/press.html)
- [23] G. Gamov, *Nature* **1954**, *173*, 318.
- [24] F. H. Crick, S. Brenner, R. J. Watstobi, L. Barnett, *Nature* **1961**, *192*, 1227-1232.
- [25] M. Nirenberg, J. H. Matthaei, *Proc. Natl. Acad. Sci. U. S. A.* **1961**, *47*, 1588-1602.
- [26] F. Eiserling, J. G. Levin, R. Byrne, U. Karlsson, M. W. Nirenberg, F. S. Sjostrand, *J. Mol. Biol.* **1964**, *10*, 536-540.
- [27] M. W. Nirenberg, R. G. Martin, O. W. Jones, S. H. Barondes, J. H. Mathhaei, *Federation Proceedings* **1963**, *22*, 55-&.
- [28] R. G. Martin, M. W. Nirenberg, O. W. Jones, J. H. Matthaei, *Biochem. Biophys. Res. Commun.* **1962**, *6*, 410-414.
- [29] O. W. Jones, M. W. Nirenberg, *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*, 2115-2123.
- [30] P. Leder, M. W. Nirenberg, *Proc. Natl. Acad. Sci. U. S. A.* **1964**, *52*, 420-427.
- [31] M. Nirenberg, P. Leder, M. Bernfiel, R. Brimacom, J. Trupin, F. Rottman, C. Oneal, *Proc. Natl. Acad. Sci. U. S. A.* **1965**, *53*, 1161-1168.
- [32] J. S. Trupin, F. M. Rottman, R. I. Brimacom, P. Leder, M. R. Bernfiel, M. W. Nirenber, *Proc. Natl. Acad. Sci. U. S. A.* **1965**, *53*, 807-&.
- [33] M. R. Bernfiel, M. W. Nirenber, *Science* **1965**, *147*, 479-&.
- [34] D. A. Kellogg, B. P. Doctor, J. E. Loebel, Nirenber.Mw, *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *55*, 912-&.
- [35] H. G. Khorana, *Biochem. J.* **1968**, *109*, 709-725.
- [36] C. R. Woese, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 8742-8747.
- [37] C. R. Woese, O. Kandler, M. L. Wheelis, *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 4576-4579.
- [38] F. D. Ciccarelli, T. Doerks, C. von Mering, C. J. Creevey, B. Snel, P. Bork, *Science* **2006**, *311*, 1283-1287.
- [39] W. F. Doolittle, *Science* **1999**, *284*, 2124-2128.
- [40] N. Goto, K. Kurokawa, T. Yasunaga, *Gene* **2007**, *401*, 172-180.

- [41] A. Eschenmoser, *Chem. Biodiversity* **2007**, *4*, 554-573.
- [42] A. G. Cairns-Smith, *Chem. Eur. J.* **2008**, *14*, 3830-3839.
- [43] G. Wachtershauser, *Orig. Life. Evol. Biosph.* **1990**, *20*, 173-176.
- [44] G. Wachtershauser, *Prog. Biophys. Mol. Biol.* **1992**, *58*, 85-201.
- [45] S. W. Ragsdale, *J. Inorg. Biochem.* **2007**, *101*, 1657-1666.
- [46] I. A. Berg, D. Kockelkorn, W. H. Ramos-Vera, R. F. Say, J. Zarzycki, M. Hügler, B. E. Alber, G. Fuchs, *Nat. Rev. Microbiol.* **2010**, *8*, 447-460.
- [47] A. I. Oparin, *The Origins of Life*, MacMillian, New York, **1938**.
- [48] C. R. Woese, *The Genetic Code, the Molecular Basis for Genetic Expression.*, Harper and Row, New York, **1967**.
- [49] F. H. C. Crick, *J. Mol. Biol.* **1968**, *38*, 367-&.
- [50] L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 381.
- [51] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147-157.
- [52] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849-857.
- [53] W. Gilbert, *Nature* **1986**, *319*, 618-618.
- [54] H. White, III, *J. Mol. Evol.* **1976**, *7*, 101-104.
- [55] G. M. Blackburn, M. J. Gait, D. Loakes, D. M. Williams, *Nucleic Acids in Chemistry and Biology*, 3rd ed., The Royal Society of Chemistry, Cambridge, **2006**.
- [56] P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, *Science* **2000**, *289*, 920-930.
- [57] N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, *Science* **2000**, *289*, 905-920.
- [58] M. M. Yusupov, G. Z. Yusupova, B. Albion, K. Lieberman, T. N. Earnest, J. H. D. Cate, H. F. Noller, *Science* **2001**, *292*, 883-896.
- [59] B. T. Wimberly, D. E. Brodersen, W. M. Clemons, R. J. Morgan-Warren, A. P. Carter, C. Vornrhein, T. Hartsch, V. Ramakrishnan, *Nature* **2000**, *407*, 327-339.
- [60] H. Noller, V. Hoffarth, L. Zimniak, *Science* **1992**, *256*, 1416-1419.
- [61] T. A. Steitz, P. B. Moore, *Trends Biochem. Sci.* **2003**, *28*, 411-418.

- [62] M. Krupkin, D. Matzov, H. Tang, M. Metz, R. Kalaora, M. J. Belousoff, E. Zimmerman, A. Bashan, A. Yonath, *Philos. Trans. R. Soc. B* **2011**, *366*, 2972-2978.
- [63] S. L. Miller, G. Schlesinger, *Orig. Life. Evol. Biosph.* **1984**, *14*, 83-90.
- [64] D. Ring, S. L. Miller, *Orig. Life. Evol. Biosph.* **1984**, *15*, 7-15.
- [65] G. Schlesinger, S. L. Miller, *J. Mol. Evol.* **1983**, *19*, 376-382.
- [66] A. Strecker, *Justus Liebigs Annalen der Chemie* **1854**, *91*, 349-351.
- [67] H. J. Cleaves, J. Chalmers, A. Lazcano, S. Miller, J. Bada, *Orig. Life. Evol. Biosph.* **2008**, *38*, 105-115.
- [68] J. P. Ferris, P. C. Joshi, E. H. Edelson, J. G. Lawless, *J. Mol. Evol.* **1978**, *11*, 293-311.
- [69] Y. Wolman, W. J. Haverland, S. L. Miller, *Proc. Natl. Acad. Sci. U. S. A.* **1972**, *69*, 809-811.
- [70] J. Oró, *Biochem. Biophys. Res. Commun.* **1960**, *2*, 407-412.
- [71] J. Oró, A. P. Kimball, *Arch. Biochem. Biophys.* **1962**, *96*, 293-&.
- [72] C. U. Lowe, R. Markham, M. W. Rees, *Nature* **1963**, *199*, 219-&.
- [73] R. A. Sanchez, J. P. Ferris, L. E. Orgel, *J. Mol. Biol.* **1967**, *30*, 223.
- [74] R. F. Shuman, W. E. Shearin, R. J. Tull, *J. Org. Chem.* **1979**, *44*, 4532-4536.
- [75] J. P. Ferris, L. E. Orgel, *J. Am. Chem. Soc.* **1966**, *88*, 1074-&.
- [76] A. W. Schwartz, H. Joosten, A. B. Voet, *Biosystems* **1982**, *15*, 191-193.
- [77] J. P. Ferris, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1968**, *33*, 693-704.
- [78] R. Shapiro, R. S. Klein, *Biochemistry* **1966**, *5*, 2358-2362.
- [79] A. Butlerow, *Comptes Rendus de l'Académie des Sciences* **1861**, *22*, 145.
- [80] P. Decker, H. Schweer, R. Pohlmann, *J. Chromatogr.* **1982**, *244*, 281-291.
- [81] R. Larralde, M. P. Robertson, S. L. Miller, *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 8158-8160.
- [82] G. Springsteen, G. F. Joyce, *J. Am. Chem. Soc.* **2004**, *126*, 9578-9583.
- [83] W. D. Fuller, L. E. Orgel, R. A. Sanchez, *J. Mol. Evol.* **1972**, *1*, 249-&.
- [84] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1972**, *67*, 25-&.

- [85] C. Fonseca Guerra, F. M. Bickelhaupt, S. Saha, F. Wang, *The Journal of Physical Chemistry A* **2006**, *110*, 4012-4020.
- [86] J. D. Sutherland, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*.
- [87] R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1970**, *47*, 531-543.
- [88] C. M. Tapiero, J. Nagyvary, *Nature* **1971**, *231*, 42-43.
- [89] A. A. Ingar, R. W. A. Luke, B. R. Hayter, J. D. Sutherland, *ChemBioChem* **2003**, *4*, 504-507.
- [90] G. F. Joyce, A. W. Schwartz, S. L. Miller, L. E. Orgel, *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 4398-4402.
- [91] G. F. Joyce, L. E. Orgel, in *The RNA world* (Eds.: R. F. Gesteland, T. R. Cech, J. F. Atkins), Cold Spring Harbor Laboratory Press, New York, **2006**, pp. 23-56.
- [92] C. Anastasi, M. A. Crowe, M. W. Powner, J. D. Sutherland, *Angew. Chem. Int. Ed.* **2006**, *45*, 6176-6179.
- [93] A. F. Cockerill, A. Deacon, R. G. Harrison, D. J. Osborne, D. M. Prime, W. J. Ross, A. Todd, J. P. Verge, *Synthesis-Stuttgart* **1976**, 591-593.
- [94] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239-242.
- [95] D. H. Shannahoff, R. A. Sanchez, *J. Org. Chem.* **1973**, *38*, 593-598.
- [96] J. P. Ferris, G. Goldstein, D. J. Beaulieu, *J. Am. Chem. Soc.* **1970**, *92*, 6598-6603.
- [97] R. Lohrmann, L. E. Orgel, *Science* **1971**, *171*, 490-494.
- [98] A. M. Schoffstall, *Orig. Life. Evol. Biosph.* **1976**, *7*, 399-412.
- [99] A. Choudhary, K. J. Kamer, M. W. Powner, J. D. Sutherland, R. T. Raines, *ACS Chem. Biol.* **2010**, *5*, 655-657.
- [100] H. B. Burgi, J. D. Dunitz, E. Shefter, *J. Am. Chem. Soc.* **1973**, *95*, 5065-5067.
- [101] G. F. Joyce, G. M. Visser, C. A. A. Vanboeckel, J. H. Vanboom, L. E. Orgel, J. Vanwestrenen, *Nature* **1984**, *310*, 602-604.
- [102] M. W. Powner, J. D. Sutherland, *Angew. Chem. Int. Ed.* **2010**, *49*, 4641-4643.
- [103] J. E. Hein, E. Tse, D. G. Blackmond, *Nature Chem.* **2011**, *3*, 704-706.
- [104] E. L. Shock, M. D. Schulte, *Geochim. Cosmochim. Acta* **1990**, *54*, 3159-3173.
- [105] L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99-123.
- [106] A. E. Engelhart, M. W. Powner, J. W. Szostak, *Nat Chem* **2013**, *5*, 390-394.

- [107] P. A. Frey, A. Arabshahi, *Biochemistry* **1995**, *34*, 11307-11310.
- [108] R. Lohrmann, L. E. Orgel, *Tetrahedron* **1978**, *34*, 853-855.
- [109] R. Lohrmann, *J. Mol. Evol.* **1977**, *10*, 137-154.
- [110] H. Sawai, *J. Am. Chem. Soc.* **1976**, *98*, 7037-7039.
- [111] H. Sawai, L. E. Orgel, *J. Am. Chem. Soc.* **1975**, *97*, 3532-3533.
- [112] H. L. Sleeper, L. E. Orgel, *J. Mol. Evol.* **1979**, *12*, 357-364.
- [113] H. L. Sleeper, R. Lohrmann, L. E. Orgel, *J. Mol. Evol.* **1979**, *13*, 203-214.
- [114] H. Sawai, K. Higa, K. Kuroda, *J. Chem. Soc., Perkin Trans. 1* **1992**, *0*, 505-508.
- [115] H. Sawai, K. Kuroda, T. Hojo, *Bull. Chem. Soc. Jpn.* **1989**, *62*, 2018-2023.
- [116] W. Huang, J. P. Ferris, *J. Am. Chem. Soc.* **2006**, *128*, 8914-8919.
- [117] W. Huang, J. P. Ferris, *Chem. Commun.* **2003**, *0*, 1458-1459.
- [118] K. J. Prabahar, J. P. Ferris, *J. Am. Chem. Soc.* **1997**, *119*, 4330-4337.
- [119] P. C. Joshi, M. F. Aldersley, J. W. Delano, J. P. Ferris, *J. Am. Chem. Soc.* **2009**, *131*, 13369-13374.
- [120] J. Ninio, L. E. Orgel, *J. Mol. Evol.* **1978**, *12*, 91-99.
- [121] R. Lohrmann, L. E. Orgel, *J. Mol. Biol.* **1980**, *142*, 555-567.
- [122] P. K. Bridson, L. E. Orgel, *J. Mol. Biol.* **1980**, *144*, 567-577.
- [123] T. Inoue, L. E. Orgel, *J. Am. Chem. Soc.* **1981**, *103*, 7666-7667.
- [124] H. J. Lipps, D. Rhodes, *Trends Cell. Biol.* **2009**, *19*, 414-422.
- [125] T. Inoue, L. E. Orgel, *Science* **1983**, *219*, 859-862.
- [126] G. F. Joyce, *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *52*, 41-51.
- [127] A. W. Schwartz, L. E. Orgel, *Science* **1985**, *228*, 585-587.
- [128] R. Rohatgi, D. P. Bartel, J. W. Szostak, *J. Am. Chem. Soc.* **1996**, *118*, 3340-3344.
- [129] R. Rohatgi, D. P. Bartel, J. W. Szostak, *J. Am. Chem. Soc.* **1996**, *118*, 3332-3339.
- [130] J. W. Szostak, *Nature* **2009**, *459*, 171-172.
- [131] C. Switzer, *ChemBioChem* **2009**, *10*, 2591-2593.

- [132] M. W. Powner, J. D. Sutherland, J. W. Szostak, *J. Am. Chem. Soc.* **2011**, *133*, 4149-4150.
- [133] M. S. Verlander, R. Lohrmann, L. E. Orgel, *J. Mol. Evol.* **1973**, *2*, 303-316.
- [134] M. S. Verlander, L. E. Orgel, *J. Mol. Evol.* **1974**, *3*, 115-120.
- [135] M. Renz, R. Lohrmann, L. E. Orgel, *Biochim. Biophys. Acta* **1971**, *240*, 463.
- [136] R. Lohrmann, L. E. Orgel, *Science* **1968**, *161*, 64.
- [137] M. A. Crowe, J. D. Sutherland, *ChemBioChem* **2006**, *7*, 951-956.
- [138] V. Borsenberger, M. A. Crowe, J. Lehbauer, J. Raftery, M. Helliwell, K. Bhutia, T. Cox, J. D. Sutherland, *Chem. Biodiversity* **2004**, *1*, 203-246.
- [139] Y. Furukawa, O. Miyashita, M. Honjo, *Chem. Pharm. Bull.* **1974**, *22*, 2552-2556.
- [140] I. Ugi, R. Meyr, U. Fetzer, C. Steinbrückner, *Angew. Chem.* **1959**, *71*, 373-388.
- [141] L. Banfi, R. Riva, in *Organic Reactions*, John Wiley & Sons, Inc., **2004**.
- [142] L. B. Mullen, J. D. Sutherland, *Angew. Chem. Int. Ed.* **2007**, *46*, 8063-8066.
- [143] D. A. Usher, A. H. McHale, *Science* **1976**, *192*, 53-54.
- [144] A. V. Lutay, E. L. Chernolovskaya, M. A. Zenkova, V. V. Vlassov, *Biogeosciences* **2006**, *3*, 243-249.
- [145] S. Islam, University of Manchester (Manchester), **2011**.
- [146] F. R. Bowler, C. K. W. Chan, C. D. Duffy, B. Gerland, S. Islam, M. W. Powner, J. D. Sutherland, J. Xu, *Nature Chem.* **2013**, *5*, 383-389.
- [147] C. Huber, G. Wächtershäuser, *Science* **1997**, *276*, 245-247.
- [148] W. J. Hagan, *ChemBioChem* **2010**, *11*, 383-387.
- [149] R. H. Liu, L. E. Orgel, *Nature* **1997**, *389*, 52-54.
- [150] S. C. Gupta, N. B. Islam, D. L. Whalen, H. Yagi, D. M. Jerina, *J. Org. Chem.* **1987**, *52*, 3812-3815.
- [151] P. Schimmel, *Annu. Rev. Biochem.* **1987**, *56*, 125-158.
- [152] P. Schimmel, K. Beebe, in *The RNA World*, 3rd ed. (Eds.: R. F. Gesteland, T. R. Cech, J. F. Atkins), Cold Spring Harbor Laboratory Press, New York, **2006**.
- [153] W. Freist, *Biochemistry* **1989**, *28*, 6787-6795.
- [154] C. Francklyn, P. Schimmel, *Nature* **1989**, *337*, 478-481.
- [155] K. Musierforsyth, P. Schimmel, *FASEB J.* **1993**, *7*, 282-289.

- [156] C. Francklyn, K. Musierforsyth, P. Schimmel, *Eur. J. Biochem.* **1992**, *206*, 315-321.
- [157] J. P. Shi, S. A. Martinis, P. Schimmel, *Biochemistry* **1992**, *31*, 4931-4936.
- [158] K. Tamura, P. R. Schimmel, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13750-13752.
- [159] K. Tamura, P. Schimmel, *Science* **2004**, *305*, 1253-1253.
- [160] M. Illangasekare, G. Sanchez, T. Nickles, M. Yarus, *Science* **1995**, *267*, 643 - 647.
- [161] M. Illangasekare, M. Yarus, *J. Mol. Biol.* **1997**, *268*, 631-639.
- [162] M. Illangasekare, M. Yarus, *RNA* **1999**, *5*, 1482-1489.
- [163] C. Tuerk, L. Gold, *Science* **1990**, *249*, 505-510.
- [164] D. L. Robertson, G. F. Joyce, *Nature* **1990**, *344*, 467-468.
- [165] A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818-822.
- [166] N. V. Chumachenko, Y. Novikov, M. Yarus, *J. Am. Chem. Soc.* **2009**, *131*, 5257-5263.
- [167] R. M. Turk, N. V. Chumachenko, M. Yarus, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4585-4589.
- [168] R. M. Turk, M. Illangasekare, M. Yarus, *J. Am. Chem. Soc.* **2011**, *133*, 6044-6050.
- [169] H. Suga, P. A. Lohse, J. W. Szostak, *J. Am. Chem. Soc.* **1998**, *120*, 1151-1156.
- [170] H. Suga, J. A. Cowan, J. W. Szostak, *Biochemistry* **1998**, *37*, 10118 - 10125.
- [171] H. Saito, D. Kourouklis, H. Suga, *EMBO J.* **2001**, *20*, 1797-1806.
- [172] H. Saito, H. Suga, *J. Am. Chem. Soc.* **2001**, *123*, 7178-7179.
- [173] H. Murakami, H. Saito, H. Suga, *Chem. Biol.* **2003**, *10*, 655-662.
- [174] H. Suga, G. Hayashi, N. Terasaka, *Philos. Trans. R. Soc. B* **2011**, *366*, 2959-2964.
- [175] J. D. Sutherland, J. M. Blackburn, *Chem. Biol.* **1997**, *4*, 481-488.
- [176] J. T. F. Wong, *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 1909-1912.
- [177] S. R. Pelc, M. G. E. Welton, *Nature* **1966**, *209*, 868-&.
- [178] P. Dunnill, *Nature* **1966**, *210*, 1267-&.

- [179] R. S. Root-Bernstein, *J. Theor. Biol.* **1982**, *94*, 895-904.
- [180] M. Ibba, H. D. Becker, C. Stathopoulos, D. L. Tumbula, D. Soll, *Trends Biochem. Sci.* **2000**, *25*, 311-316.
- [181] K. Kawamura, J. P. Ferris, *Orig. Life. Evol. Biosph.* **1999**, *29*, 563-591.
- [182] M. M. W. Mooren, S. S. Wijmenga, G. A. Vandermarel, J. H. Vanboom, C. W. Hilbers, *Nucleic Acids Res.* **1994**, *22*, 2658-2666.
- [183] S. Agrawal, in *Methods in Molecular Biology, Vol. 20*, Humana, Totowa, NJ, **1993**.
- [184] N. Usman, K. K. Ogilvie, M. Y. Jiang, R. J. Cedergren, *J. Am. Chem. Soc.* **1987**, *109*, 7845-7854.
- [185] T. J. Wilson, N.-S. Li, J. Lu, J. K. Frederiksen, J. A. Piccirilli, D. M. J. Lilley, *Proc. Natl. Acad. Sci. U. S. A.* **2010**.
- [186] F. Wincott, A. DiRenzo, C. Shaffer, S. Grimm, D. Tracz, C. Workman, D. Sweedler, C. Gonzalez, S. Scaringe, N. Usman, *Nucleic Acids Res.* **1995**, *23*, 2677-2684.
- [187] L. J. McBride, R. Kierzek, S. L. Beaucage, M. H. Caruthers, *J. Am. Chem. Soc.* **1986**, *108*, 2040-2048.
- [188] Q. Zhu, M. O. Delaney, M. M. Greenberg, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1105-1107.
- [189] B. Uznanski, A. Grajkowski, A. Wilk, *Nucleic Acids Res.* **1989**, *17*, 4863-4871.
- [190] J. C. Schulhof, D. Molko, R. Teoule, *Tetrahedron Lett.* **1987**, *28*, 51-54.
- [191] E. Westman, R. Stromberg, *Nucleic Acids Res.* **1994**, *22*, 2430-2431.
- [192] H. Seliger, in *Current Protocols in Nucleic Acid Chemistry Unit 2.3* (Eds.: M. Egli, P. Herdewijn, A. Matsuda, Y. S. Sanghvi), John Wiley & Sons, Inc., **2000**.
- [193] N. D. Sinha, J. Biernat, H. Köster, *Tetrahedron Lett.* **1983**, *24*, 5843-5846.
- [194] N. D. Sinha, J. Biernat, J. McManus, H. Köster, *Nucleic Acids Res.* **1984**, *12*, 4539-4557.
- [195] R. Eritja, J. Robles, A. Aviñó, F. Alberico, E. Pedroso, *Tetrahedron* **1992**, *48*, 4171-4182.
- [196] J. G. Lackey, D. Sabatino, M. J. Damha, *Org. Lett.* **2007**, *9*, 789-792.
- [197] M. J. Damha, P. A. Giannaris, S. V. Zabarylo, *Nucleic Acids Res.* **1990**, *18*, 3813-3821.
- [198] R. T. Pon, S. Y. Yu, *Nucleic Acids Res.* **1997**, *25*, 3629-3635.

- [199] W. T. Markiewicz, T. K. Wyrzykiewicz, *Nucleic Acids Res.* **1989**, *17*, 7149-7158.
- [200] R. P. Iyer, in *Current Protocols in Nucleic Acid Chemistry Unit 2.1* (Eds.: M. Egli, P. Herdewijn, A. Matsuda, Y. S. Sanghvi), John Wiley & Sons, Inc, **2000**.
- [201] C. Merk, T. Reiner, E. Kvasyuk, W. Pfeleiderer, *Helv. Chim. Acta* **2000**, *83*, 3198-3210.
- [202] H. Lang, M. Gottlieb, M. Schwarz, S. Farkas, B. S. Schulz, F. Himmelsbach, R. Charubala, W. Pfeleiderer, *Helv. Chim. Acta* **1999**, *82*, 2172-2185.
- [203] T. Wu, K. K. Ogilvie, R. T. Pon, *Nucleic Acids Res.* **1989**, *17*, 3501-3517.
- [204] R. Johnsson, J. G. Lackey, J. J. Bogojeski, M. J. Damha, *Bioorg. Med. Chem. Lett.* **2011**, *21*, 3721-3725.
- [205] F. Z. Dörwald, *Organic Synthesis on Solid Phase*, Wiley-VCH, Germany, **2002**.
- [206] A. Ajayaghosh, V. N. Rajasekharan Pillai, *Tetrahedron* **1988**, *44*, 6661-6666.
- [207] T. D. Ryba, P. G. Harran, *Org. Lett.* **2000**, *2*, 851-853.
- [208] D. Woll, J. Smirnova, M. Galetskaya, T. Prykota, J. Buhler, K. P. Stengele, W. Pfeleiderer, U. E. Steiner, *Chemistry* **2008**, *14*, 6490-6497.
- [209] S. Bühler, I. Lagoja, H. Giegrich, K.-P. Stengele, W. Pfeleiderer, *Helv. Chim. Acta* **2004**, *87*, 620-659.
- [210] S. Walbert, W. Pfeleiderer, U. E. Steiner, *Helv. Chim. Acta* **2001**, *84*, 1601-1611.
- [211] Glen Research – The Glen Report, vol. 19, no. 2, Dec 2007
- [212] G. H. Hakimelahi, Z. A. Proba, K. K. Ogilvie, *Can. J. Chem.* **1982**, *60*, 1106-1113.
- [213] C. B. Reese, D. R. Trentham, *Tetrahedron Lett.* **1965**, *6*, 2459-2465.
- [214] C. B. Reese, D. R. Trentham, *Tetrahedron Lett.* **1965**, *6*, 2467-2472.
- [215] W. T. Wiesler, M. H. Caruthers, *J. Org. Chem.* **1996**, *61*, 4272-4281.
- [216] F. Himmelsbach, B. S. Schulz, T. Trichtinger, R. Charubala, W. Pfeleiderer, *Tetrahedron* **1984**, *40*, 59-72.
- [217] S. Nishino, H. Takamura, Y. Ishido, *Tetrahedron* **1986**, *42*, 1995-2004.
- [218] O. Mitsunobu, *Synthesis-Stuttgart* **1981**, 1-28.
- [219] T. Kamimura, M. Tsuchiya, K. Urakami, K. Koura, M. Sekine, K. Shinozaki, K. Miura, T. Hata, *J. Am. Chem. Soc.* **1984**, *106*, 4552-4557.

- [220] H. Schirmesiter, F. Himmelsbach, W. Pfeleiderer, *Helv. Chim. Acta* **1993**, *76*, 385-401.
- [221] A. Matsuda, M. Shinozaki, M. Suzuki, K. Watanabe, T. Miyasaka, *Synthesis-Stuttgart* **1986**, 385-386.
- [222] A. P. Guzaev, M. Manoharan, (Ed.: U. S. P. a. T. Office), ISIS Pharmaceuticals, Inc., United States, **2004**.
- [223] H. P. M. Fromageot, B. E. Griffin, C. B. Reese, J. E. Sulston, *Tetrahedron* **1967**, *23*, 2315-2331.
- [224] M. Murata, P. Bhuta, J. Owens, J. Zemlicka, *J. Med. Chem.* **1980**, *23*, 781-786.
- [225] M. J. Robins, R. Mengel, R. A. Jones, Y. Fouron, *J. Am. Chem. Soc.* **1976**, *98*, 8204-8213.
- [226] B. H. Dahl, J. Nielsen, O. Dahl, *Nucleic Acids Res.* **1987**, *15*, 1729-1743.
- [227] S. Berner, K. Mühlegger, H. Seliger, *Nucleic Acids Res.* **1989**, *17*, 853-864.
- [228] H. P. M. Fromageot, B. E. Griffin, C. B. Reese, J. E. Sulston, D. R. Trentham, *Tetrahedron* **1966**, *22*, 705-710.
- [229] A. Somoza, *Chem. Soc. Rev.* **2008**, *37*, 2668-2675.
- [230] G. H. Hakimelahi, Z. A. Proba, K. K. Ogilvie, *Tetrahedron Lett.* **1981**, *22*, 4775-4778.
- [231] K. K. Ogilvie, D. J. Iwacha, *Tetrahedron Lett.* **1973**, *14*, 317-319.
- [232] S. A. Scaringe, C. Francklyn, N. Usman, *Nucleic Acids Res.* **1990**, *18*, 5433-5441.
- [233] J. r. Parsch, J. W. Engels, *J. Am. Chem. Soc.* **2002**, *124*, 5664-5672.
- [234] C. G. Bochet, *J. Chem. Soc., Perkin Trans. 1* **2002**.
- [235] F. Guillier, D. Orain, M. Bradley, *Chem. Rev. (Washington, DC, U. S.)* **2000**, *100*, 2091-2158.
- [236] J. A. Barltrop, P. J. Plant, P. Schofield, *Chem. Commun.* **1966**, *0*, 822-823.
- [237] A. P. Pelliccioli, J. Wirz, *Photochem. Photobiol. Sci.* **2002**, *1*, 441-458.
- [238] R. N. Haszeldine, *J. Chem. Soc.* **1953**, *0*, 1748-1757.
- [239] M. O. Smith, J. March, *March's advanced organic chemistry: reactions, mechanisms, and structure*, 6th / Michael B. Smith, Jerry March. ed., Wiley-Interscience, Hoboken, N.J., **2007**.
- [240] L. Ford, F. Atefi, R. D. Singer, P. J. Scammells, *Eur. J. Org. Chem.* **2011**, 942-950.

- [241] [http://www.linktech.co.uk/products/rna\\_synbase\\_cpg\\_solid\\_supports](http://www.linktech.co.uk/products/rna_synbase_cpg_solid_supports)
- [242] T. Li, C. Zhou, M. Jiang, *Polym. Bull. (Berlin)* **1991**, *25*, 211-216.
- [243] C. Vargeese, J. Carter, J. Yegge, S. Krivjansky, A. Settle, E. Kropp, K. Peterson, W. Pieken, *Nucleic Acids Res.* **1998**, *26*, 1046-1050.
- [244] N. D. Sinha, C07H 21/00 ed. (Ed.: W. I. P. Organization), **2001**.
- [245] A. Wilk, A. Grajkowski, L. R. Phillips, S. L. Beaucage, *J. Org. Chem.* **1999**, *64*, 7515-7522.
- [246] T. Umemoto, T. Wada, *Tetrahedron Lett.* **2005**, *46*, 4251-4253.
- [247] C. Zhou, W. Pathmasiri, D. Honcharenko, S. Chatterjee, J. Barman, J. Chattopadhyaya, *Can. J. Chem.* **2007**, *85*, 293-301.
- [248] H. Venkatesan, M. M. Greenberg, *J. Org. Chem.* **1996**, *61*, 525-529.
- [249] M. D. Corbett, B. R. Corbett, *Chem. Res. Toxicol.* **1993**, *6*, 82-90.
- [250] M. D. Corbett, B. R. Corbett, *J. Org. Chem.* **1980**, *45*, 2834-2839.
- [251] C. Altona, M. Sundaralingam, *J. Am. Chem. Soc.* **1972**, *94*, 8205-8212.
- [252] W. Guschlbauer, K. Jankowski, *Nucleic Acids Res.* **1980**, *8*, 1421-1433.
- [253] D. Elliot, M. Ladomery, *Molecular Biology of RNA*, Oxford University Press, Oxford, **2010**.
- [254] D. Venkateswarlu, K. E. Lind, V. Mohan, M. Manoharan, D. M. Ferguson, *Nucleic Acids Res.* **1999**, *27*, 2189-2195.
- [255] L. L. Cummins, S. R. Owens, L. M. Risen, E. A. Lesnik, S. M. Freier, D. McGee, C. J. Guinosso, P. D. Cook, *Nucleic Acids Res.* **1995**, *23*, 2019-2024.
- [256] E. A. Lesnik, C. J. Guinosso, A. M. Kawasaki, H. Sasmor, M. Zounes, L. L. Cummins, D. J. Ecker, P. D. Cook, S. M. Freier, *Biochemistry* **1993**, *32*, 7832-7838.
- [257] P. Nissen, J. A. Ippolito, N. Ban, P. B. Moore, T. A. Steitz, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4899-4903.
- [258] M. Selmer, C. M. Dunham, F. V. Murphy, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir, V. Ramakrishnan, *Science* **2006**, *313*, 1935-1942.
- [259] A. S. Goldsborough, *Vol. 6,867,290 B2* (Ed.: U. S. Patent), Cyclops Genome Sciences, Ltd., GB, **2005**.
- [260] S. J. Schroeder, D. H. Turner, *Methods Enzymol.* **2009**, *468*, 371-387.
- [261] V. A. Bloomfield, D. M. Crothers, I. Tinoco, *Nucleic Acids: Structures, properties and functions*, University Science Books, California **2000**.

- [262] J. L. Mergny, L. Lacroix, *Oligonucleotides* **2003**, *13*, 515-537.
- [263] L. A. Marky, K. J. Breslauer, *Biopolymers* **1987**, *26*, 1601-1620.
- [264] G. Mathis, S. Bourg, S. Aci-Seche, J.-C. Truffert, U. Asseline, *Org. Biomol. Chem.* **2013**.
- [265] E. Rozners, J. Moulder, *Nucleic Acids Res.* **2004**, *32*, 248-254.
- [266] M. Egli, S. Portmann, N. Usman, *Biochemistry* **1996**, *35*, 8489-8494.
- [267] D. A. Adamiak, J. Milecki, R. W. Adamiak, W. Rypniewski, *New J. Chem.* **2010**, *34*, 903-909.
- [268] D. A. Adamiak, J. Milecki, M. Popena, R. W. Adamiak, Z. Dauter, W. R. Rypniewski, *Nucleic Acids Res.* **1997**, *25*, 4599-4607.
- [269] E. A. Lesnik, S. M. Freier, *Biochemistry* **1998**, *37*, 6991-6997.
- [270] P. A. Giannaris, M. J. Damha, *Nucleic Acids Res.* **1993**, *21*, 4742-4749.
- [271] B. J. Premraj, P. K. Patel, E. R. Kandimalla, S. Agrawal, R. V. Hosur, N. Yathindra, *Biochem. Biophys. Res. Commun.* **2001**, *283*, 537-543.
- [272] N. Erande, A. D. Gunjal, M. Fernandes, R. Gonnade, V. A. Kumar, *Org. Biomol. Chem.* **2013**, *11*, 746-757.
- [273] T. Xia, J. SantaLucia, M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, D. H. Turner, *Biochemistry* **1998**, *37*, 14719-14735.
- [274] A. R. Ferré-D'Amare, W. G. Scott, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*.
- [275] J. W. Szostak, *J. Syst. Chem.* **2012**, *3*, 2.
- [276] D. A. Usher, A. H. McHale, *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 1149-1153.
- [277] <http://www.atdbio.com/tools/oligo-calculator>
- [278] A. Wochner, J. Attwater, A. Coulson, P. Holliger, *Science* **2011**, *332*, 209-212.
- [279] H. Heus, A. Pardi, *Science* **1991**, *253*, 191-194.
- [280] E. A. Doherty, R. T. Batey, B. Masquida, J. A. Doudna, *Nature Structural Biology* **2001**, *8*, 339-343.
- [281] M. Costa, F. Michel, *EMBO J.* **1997**, *16*, 3289-3302.
- [282] K. A. Vander Meulen, J. H. Davis, T. R. Foster, M. T. Record Jr, S. E. Butcher, *J. Mol. Biol.* **2008**, *384*, 702-717.
- [283] S. E. Butcher, A. M. Pyle, *Acc. Chem. Res.* **2011**, *44*, 1302-1311.

- [284] S. Matsumura, R. Ohmori, H. Saito, Y. Ikawa, T. Inoue, *FEBS Lett.* **2009**, *583*, 2819-2826.
- [285] J. P. Sheehy, A. R. Davis, B. M. Znosko, *RNA* **2010**, *16*, 417-429.
- [286] F. M. Jucker, H. A. Heus, P. F. Yip, E. H. M. Moors, A. Pardi, *J. Mol. Biol.* **1996**, *264*, 968-980.
- [287] C. C. Correll, K. Swinger, *RNA* **2003**, *9*, 355-363.
- [288] J. SantaLucia, Jr., R. Kierzek, D. H. Turner, *Science* **1992**, *256*, 217-219.
- [289] Q. Zhao, H.-C. Huang, U. Nagaswamy, Y. Xia, X. Gao, G. E. Fox, *Biopolymers* **2012**, *97*, 617-628.
- [290] J. M. Blose, D. J. Proctor, N. Veeraraghavan, V. K. Misra, P. C. Bevilacqua, *J. Am. Chem. Soc.* **2009**, *131*, 8474-8484.
- [291] J. Ishikawa, Y. Fujita, Y. Maeda, H. Furuta, Y. Ikawa, *Methods* **2011**, *54*, 226-238.
- [292] P. S. Pallan, E. M. Greene, P. A. Jicman, R. K. Pandey, M. Manoharan, E. Rozners, M. Egli, *Nucleic Acids Res.* **2011**, *39*, 3482-3495.
- [293] S. Preus, K. Kilså, F.-A. Miannay, B. Albinsson, L. M. Wilhelmsson, *Nucleic Acids Res.* **2013**, *41*, e18.
- [294] D. J. Klein, T. M. Schmeing, P. B. Moore, T. A. Steitz, *EMBO J.* **2001**, *20*, 4214-4221.
- [295] J. Liu, D. M. J. Lilley, *RNA* **2007**, *13*, 200-210.
- [296] N. B. Leontis, E. Westhof, *RNA* **2001**, *7*, 499-512.
- [297] R. Pascal, L. Boiteau, *Philos. Trans. R. Soc. B* **2011**, *366*, 2949-2958.
- [298] L. Leman, L. Orgel, M. R. Ghadiri, *Science* **2004**, *306*, 283-286.
- [299] G. Danger, L. Boiteau, H. Cottet, R. Pascal, *J. Am. Chem. Soc.* **2006**, *128*, 7412-7413.
- [300] L. J. Leman, L. E. Orgel, M. R. Ghadiri, *J. Am. Chem. Soc.* **2006**, *128*, 20-21.
- [301] J. P. Biron, A. L. Parkes, R. Pascal, J. D. Sutherland, *Angew. Chem. Int. Ed.* **2005**, *44*, 6731-6734.
- [302] A. Loison, S. Dubant, P. Adam, P. Albrecht, *Astrobiology* **2010**, *10*, 973-988.
- [303] M. Keller, E. Blochl, G. Wachtershauser, K. O. Stetter, *Nature* **1994**, *368*, 836-838.
- [304] A. L. Weber, L. E. Orgel, *J. Mol. Evol.* **1979**, *13*, 193-202.

- [305] T. Wieland, R. Lambert, H. U. Lang, G. Schramm, *Justus Liebigs Annalen der Chemie* **1955**, 597, 181-195.
- [306] M. C. Maurel, L. E. Orgel, *Orig. Life. Evol. Biosph.* **2000**, 30, 423-430.
- [307] A. Commeyras, H. Collet, L. Boiteau, J. Taillades, O. Vandenaabeele-Trambouze, H. Cottet, J.-P. Biron, R. Plasson, L. Mion, O. Lagrille, H. Martin, F. Selsis, M. Dobrijevic, *Polym. Int.* **2002**, 51, 661-665.
- [308] C. Huber, G. Wächtershäuser, *Science* **1998**, 281, 670-672.
- [309] H. Rauchfuss, *Chemical Evolution and the Origin of Life*, Springer, **2008**.
- [310] C. de Duve, *American Scientist* **1995**, 83, 428-437.
- [311] R. Pascal, L. Boiteau, A. Commeyras, in *Prebiotic Chemistry: From Simple Amphiphiles to Protocell Models*, Vol. 259 (Ed.: P. Walde), **2005**, pp. 69-122.
- [312] L. Boiteau, R. Pascal, *Orig. Life. Evol. Biosph.* **2011**, 41, 23-33.
- [313] J. Mann, *Chemical Aspects of Biosynthesis*, Oxford University Press, Oxford, **1994**.
- [314] D. Ritson, J. D. Sutherland, *Nature Chem.* **2012**, 4, 895-899.
- [315] D. H. Williams, F. Ian Fleming, *Spectroscopic Methods in Organic Chemistry*, McGraw-Hill, **2008**.
- [316] J. P. Ferris, L. E. Orgel, *J. Am. Chem. Soc.* **1965**, 87, 4976-&.
- [317] J. P. Ferris, C. H. Huang, W. J. Hagan, *Nucleosides Nucleotides* **1989**, 8, 407-414.
- [318] K. J. Luebke, P. B. Dervan, *J. Am. Chem. Soc.* **1991**, 113, 7447-7448.
- [319] T. H. Li, D. S. Weinstein, K. C. Nicolaou, *Chem. Biol.* **1997**, 4, 209-214.
- [320] T. Li, K. C. Nicolaou, *Nature* **1994**, 369, 218-221.
- [321] V. Govindaraju, V. J. Basus, G. B. Matson, A. A. Maudsley, *Magn. Reson. Med.* **1998**, 39, 1011-1013.
- [322] G. Graca, I. F. Duarte, B. J. Goodfellow, I. M. Carreira, A. B. Couceiro, M. d. R. Domingues, M. Spraul, L.-H. Tseng, A. M. Gil, *Anal. Chem.* **2008**, 80, 6085-6092.
- [323] S. P. Singh, S. S. Parmar, K. Raman, V. I. Stenberg, *Chem. Rev. (Washington, DC, U. S.)* **1981**, 81, 175-203.
- [324] F. C. Brown, *Chem. Rev. (Washington, DC, U. S.)* **1961**, 61, 463-521.
- [325] R. K. Kumar, M. Yarus, *Biochemistry* **2001**, 40, 6998-7004.

- [326] P. K. Glasoe, F. A. Long, *J. Phys. Chem.* **1960**, *64*, 188-190.
- [327] T. Torii, H. Shiragami, K. Yamashita, Y. Suzuki, T. Hijiya, T. Kashiwagi, K. Izawa, *Tetrahedron* **2006**, *62*, 5709-5716.
- [328] N. A. Siegfried, P. C. Bevilacqua, in *Methods Enzymol., Vol. Volume 455* (Eds.: J. M. H. Michael L. Johnson, K. A. Gary), Academic Press, **2009**, pp. 365-393.
- [329] R. N. Hannoush, M. J. Damha, *J. Am. Chem. Soc.* **2001**, *123*, 12368-12374.
- [330] S. Murahashi, T. Takizawa, S. Kurioka, S. Maekawa, *Nippon Kagaku Zasshi* **1956** *77*, 1689-1692.
- [331] A. K. Covington, M. Paabo, R. A. Robinson, R. G. Bates, *Anal. Chem.* **1968**, *40*, 700-706.