Synthetic methodology for the preparation of nucleic acid containing peptides

Heden-van Noort, G.J. van der

Chapter 1

General Introduction

The complexity of life can be appreciated when looking at the structural diversity of biopolymers involved in biological processes. A wide variety of carbohydrates, peptides and nucleic acids plays crucial roles in an even wider range of important biochemical events. To gain insight in these biological processes the need for synthetically obtained fragments of biopolymers and derivatives thereof has risen in the last half of the twentieth century. Synthetic protocols towards oligonucleic acids and peptides have been developed allowing access to well-defined material for research in medicinal chemistry, chemical biology and structural biology. An enormous leap forward in gaining accessibility to such materials was the introduction of insoluble polymeric supports for the synthesis of peptides and oligonucleic acids.[1-3] The limits of solution phase assembly, such as insolubility and difficulties in purification after each sequential step, are greatly reduced by using solid support chemistry.[2] Automated synthesis based on this approach has proven to be a useful tool to obtain protein fragments and oligonucleotides. Hybrid structures between different classes of biopolymers, such as glycoproteins, glycolipids and nucleoproteins[4,5] however are not always easily accessible using the currently used solid phase based approaches. Conditions that are used in the synthesis of one biopolymer fragment might not be compatible with the stability of the second biopolymer fragment.

The aim of the work described in this thesis is to develop synthetic methods that give access to hybrid molecules containing both a peptide part and a nucleic acid part. These methods should be applicable on solid support and ultimately open the way to the synthesis of more complex fragments derived from biomolecules.

An overview of naturally occurring hybrid molecules, consisting of a peptide and a nucleic acid and synthetic efforts to obtain fragments and derivates of these interesting molecules is presented in this chapter.

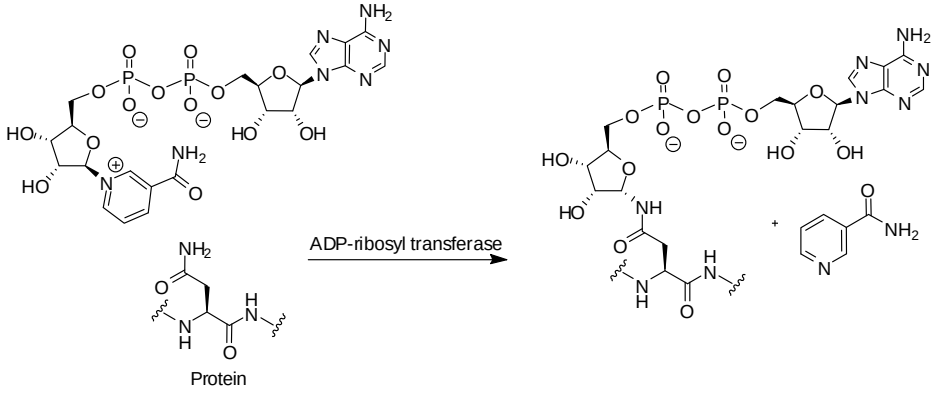
1.0 Post-translational modifications

Investigation of post-translational modifications of proteins is a major research topic since these modifications are key players in regulation of eukaryotic cellular processes.[6] Such modifications range from the introduction of a simple acetyl[7] or phosphate[8] group to attachment of small proteins such as ubiquitin.[9] Some post-translational modifications even cause alteration of the amino acid sequence of the target protein such as for instance deamidation of asparagine to aspartic acid residues.[10] Introduction of nucleic acid monomers are also found amongst the diverse post-translational modifications, in processes such as ADP-ribosylation and adenylylation.

1.1 ADP-ribosylation

Adenosine diphosphate ribosylation (ADP-ribosylation) is effected by two classes of enzymes called *poly*-ADP-ribosyl polymerases (PARPs) or *mono*-ADP-ribosyl transferases (MARTs).[11] These enzymes transfer ADP-ribose from β-NAD^+ to nucleophilic functional groups in the side chain of amino acid residues of the target protein, a process that is accompanied by the release of nicotinamide (see **Figure 1**).[12]

Figure 1



ADP-ribosylation has an influence on a variety of processes ranging from the triggering of DNA repair mechanisms to regulation of immune responses.[13,14]

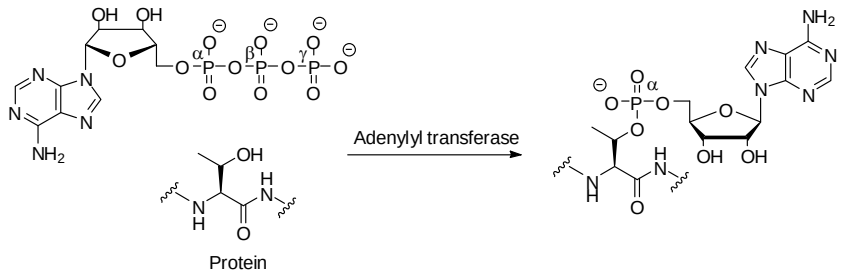
Schramm and co-workers have done extensive research on identifying the transition state of the ADP-ribosylation process effected by the bacterial pertussis toxin.[15-18] For this work they treat a synthetic peptide fragment with pertussis toxin and NAD^+ analogues to obtain ADP-ribosyl peptides. Muir and co-workers have reported the synthesis of aminooxy or *N*-methyl aminooxy

modified peptides that can be conjugated to ADP-ribose.[19] These prepared conjugates carrying a non-natural linkage to ADP-ribose are recognized by *poly*-ADP-ribosylating enzymes in enzymatic essays. Although several other reports on modified NAD^+ analogues have been published[20-24], no fully chemical synthesis of ADP-ribosyl containing peptides is known to date.

1.2 Adenylylation

Adenylylation is another post translational modification process in which adenosine monophosphate (AMP) is transferred from adenosine triphosphate (ATP) to a nucleophilic residue in the protein backbone under the influence of adenylyl transferring enzymes (see **Figure 2**).[25] If this nucleophilic residue is a hydroxyl (in the case of serine, threonine or tyrosine) a stable phosphodiester bond is formed between adenosine and the protein. Adenylylation reactions involving the introduction of such a stable adenylyl adduct usually plays a role in the regulation of protein activity.[26] When carboxylate or amine side chains are adenylylated unstable mixed anhydrides or phosphoramidates are formed. The function of adenylylation in this type of reactions is the introduction of a good leaving group. In this transient process the energy stored in the α,β- phosphoric anhydride bond in ATP is released. This overall process facilitates thermodynamically unfavorable reactions to take place. Examples of adenylyl transferases are bacterial enzymes that adenylylate GTPases.[27-29]

Figure 2



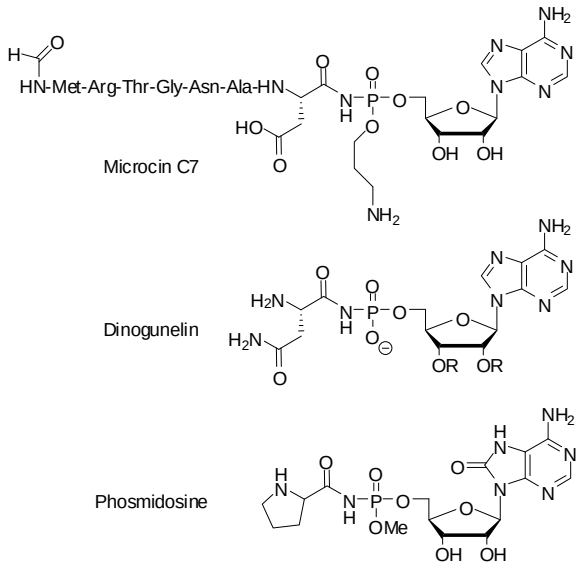
A synthetic route towards adenylylated peptides was published using a H-phosphonate approach.[30] Unprotected serine or threonine containing resin-bound peptides were transformed to their corresponding H-phosphonates and subsequently coupled to 2',3'-*iso*propyl adenosine and oxidized. Cleavage from the solid support and protective group removal using 95% trifluoroacetic acid yielded the target peptides. The yields, however, were very low, which might be due to depurination of the nucleotidyl peptides in acidic environment. Although small quantities of adenylyl serine or threonine peptides were obtained, this method might not be widely applicable since the adenylyl peptides seem to be unstable under the conditions used in this approach. Furthermore, a similar H-phosphonate approach was reported to be inefficient for the synthesis of tyrosine containing nucleotidyl peptides, thereby limiting the scope of such a H-phosphonate method to obtain adenylyl peptides.[31] Another approach towards adenylylated peptides is based on the assembly of pre-adenylylated Fmoc protected amino acids.[32,33] In this approach 2', 3'-O-*iso*propylidene protection is installed on the ribosyl part of adenosine and the exocyclic amine of adenine is protected with two Boc groups. The 5'-O-phosphoramidite of this

protected nucleoside is coupled to suitably protected serine, threonine and tyrosine. After oxidation and protective group manipulations these adenylyl amino acids are incorporated in peptides using an on-line solid phase peptide synthesis method. Acidic removal of the protective groups and simultaneous cleavage from the resin gives access to the adenylylated peptides accompanied by trace amounts of depurinated analogues.

2.0 Antibiotics

Some bacterial species are found to produce antibiotics, containing a peptide and nucleotide part, to defend themselves. *Escherichia coli* for instance produces the antibiotic Microcin C7 (see **Figure 3**).[34] The peptidyl part of this phosphoramidate mediates the transport of the antibiotic into the target bacterial cell, where processing leads to cleavage of the peptide bond between alanine and aspartic acid releasing the aspartic acid adenyl phosphoramidate which acts as an inhibitor of protein synthesis in the targeted bacterium.[35,36] A similar defense mechanism is employed by the eggs of the Japanese fish *Stichaeus grigorjewi*, that produce the Dinogunelin toxin (see **Figure 3**, R is either a proton or a unsaturated lipid residue).[37] Another example is Phosmidosine (see **Figure 3**), where 8-oxo-adenosine is attached to proline via a similar phosphoramidate linkage.[38] This antifungal antibiotic originating from *Streptomyces durhmeusie* is also found to possess anti-cancer activity. Phosmidosine and some analogues thereof have been prepared by coupling the methoxy-phosphoramidite of 8-oxo-adenosine to a protected proline amide.[39]

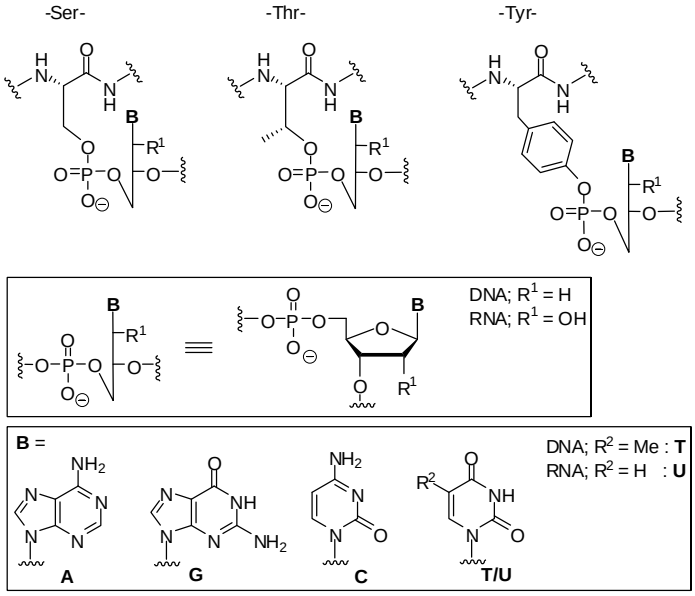
Figure 3



3.0 Nucleoproteins

Nucleoproteins are a class of hybrid biomolecules in which a protein is covalently linked to a nucleic acid chain via a phosphodiester bond between the 3'- or 5'-end of the DNA or RNA and the hydroxyl function of a serine, threonine or tyrosine residue in the protein backbone (See **Figure 4**).

Figure 4



Such nucleoproteins play important roles in the replication of RNA-viruses, DNA-viruses and bacteriophages:

RNA-viruses. In the case of poliovirus, a well investigated example of a RNA-virus, the small protein VPg (Viral Protein genome-linked), consisting of only 22 amino acids is attached via the tyrosine hydroxyl moiety to the 5'-end uridylic acid residue of the viral genome.[40] This protein acts as a cap-structure at the 5'-end of the viral RNA. Besides protecting the viral genome from exonuclease degradation VPg is believed to be a primer in progeny strand synthesis either as the free peptide or as its uridylylated form VPgpU or VPgpUpU.[41]

DNA-viruses. In adenovirus, an example of a DNA-virus, both 5'-ends of the double stranded viral genome are linked to a 55kD protein called TP (Terminal Protein) via a phosphodiester bond between serine residue 580 in the protein backbone and the 5'-deoxycytidylic acid of the genome. This nucleoprotein is believed to act as primer in DNA-replication.[42-44]

Bacteriophages. In a similar fashion bacteriophage Φ29 DNA is covalently attached via the 5'-end deoxyadenosine to a serine residue in the 27kD protein called P3.[45,46] This nucleoprotein is believed to act as primer in DNA-replication. Another example is the bacteriophage ΦX174

gene A protein, which is believed to cleave the double stranded DNA of the phage by the formation of a phosphodiester bond between a tyrosine residue and the 5'-end adenylic acid residue at the cleavage site, thereby creating a starting point for replication.[47]

Topoisomerases. Nucleoproteins also occur as intermediates in unwinding and rewinding of supercoiled DNA in higher organisms. Topoisomerase type I and type II enzymes cleave the DNA backbone, reduce tension on the intertwined DNA to facilitate access for replication enzymes and then reconnect the cleaved DNA.[48] Topoisomerase type I enzymes induce single strand breaks via the formation of an intermediate in which a tyrosine residue in the protein's backbone is connected covalently to the 3'-phosphate of the broken DNA. Topoisomerase type II enzymes induce double strand breaks forming phosphodiester bond between a tyrosine or serine and the 5'-ends of both broken DNA strands.[49] Several drugs inhibiting topoisomerases are used in antibacterial and anticancer therapies. These drugs prevent the topoisomerase to reconnect the broken DNA, resulting in a permanent protein-bound DNA lesion, which eventually leads to apoptosis.[50]

P53-protein. Another protein that is found to be covalently linked to RNA is tumor suppressor protein P53. This protein is involved in DNA repair and cell cycle regulation and the inability to express this protein leads to tumor formation.[51] It is proposed that P53 is linked to the 5'-end of the RNA via Ser-389.[52]

The synthesis of nucleoproteins or fragments thereof named nucleopeptides, is far more demanding than the synthesis of the individual nucleic acid and peptide fragments. When synthesizing such hybrid biopolymers, the inherent acid lability of the oligonucleotides prohibits the use of strong acid, which is commonly used in deprotection protocols in peptide chemistry. Likewise the strong basic conditions required for the removal of acyl protective groups at the exocyclic amines of nucleobases may be accompanied by unwanted β-elimination of phosphorylated serine and threonine residues.[53] In this process the proton on the α-position of serine or threonine is abstracted resulting in formation of dehydroalanine and the oligonucleotide 5'-phosphate. On top of this a complex protective group strategy is needed, requiring orthogonality of the permanent and temporary protective groups in the peptide and nucleic acid part of the projected nucleopeptide. One has to take all these properties of nucleopeptides and the desired protection scheme into account when devising a suitable synthetic route for fragments of nucleoproteins.

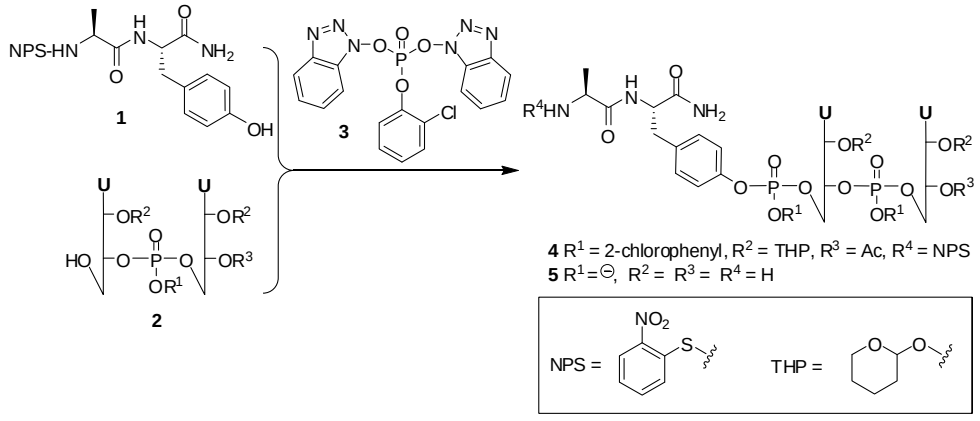
3.1 Synthesis of nucleopeptides in solution

Initial efforts to synthesize nucleopeptides take advantage of a balanced combination of solution approaches known from nucleic acid and peptide chemistry. A summary of different methods to install the crucial phosphodiester linkage between the peptide part and the nucleic acid part of nucleopeptides is described below, focusing on the phosphorus chemistry involved and the protective groups used in these strategies.

3.1.1 Phosphotriester approach

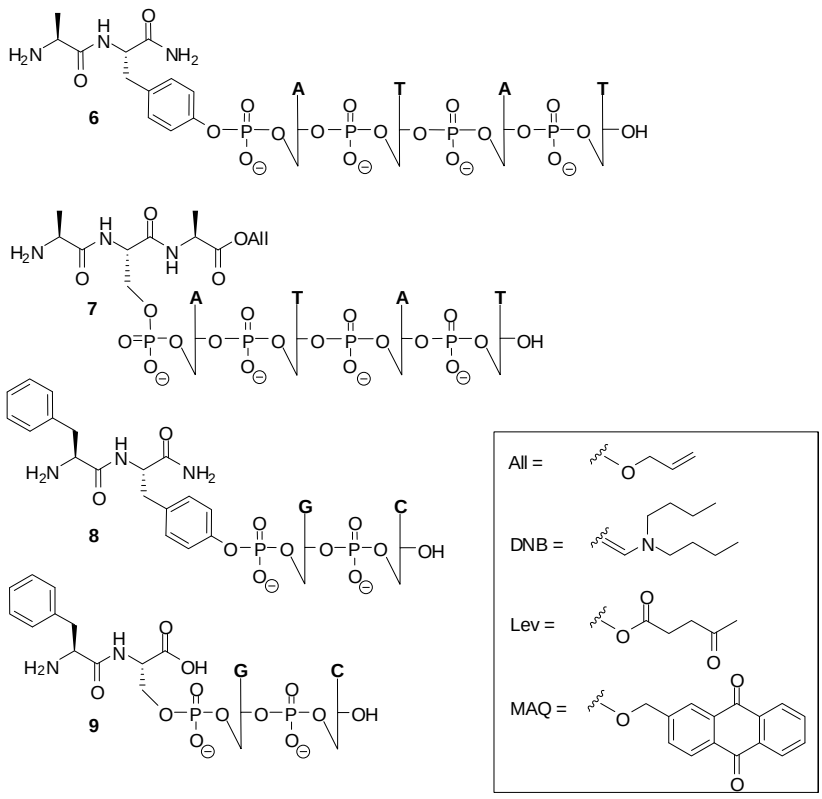
Van Boom and coworkers reported the first synthetically made ribonucleopeptide **5** (see **Figure 5**), which is prepared using a modified phosphotriester approach.[54] In this example bi-functional phosphorylating agent **3** equipped with two leaving groups (1-benzotriazolyl) and one permanent protective group (2-chlorophenyl) is first coupled to the hydroxyl functionality of dipeptide **1**, followed by coupling to dinucleotide **2** using *N*-methyl imidazole as activator. The protective groups used in this sequence of reactions are 2-nitrophenylsulfenyl (NPS) on the peptide N-terminus, 2-chlorophenyl on the phosphodiester, tetrahydropyranyl (THP) on the 2'-hydroxyl functions of the uridylic acid residues, acetyl on the 3'-hydroxyl of the terminal uridylic acid residue, while the C-terminal carboxylate of the peptide is capped as carboxamide. After construction of the protected phosphodiester linkage between the peptide part and the nucleotide part, removal of the protective groups was accomplished in a three step procedure. First oximate-reagent[55] was used to cleave the 2-chlorophenyl. Secondly a triethylamine-methanol-water treatment cleaved the acetyl group and a final acidic hydrolysis of the NPS and THP by treating with methanolic HCl (pH 2.5) for 3 hours followed by aqueous HCl (pH 2) for 14 hours. The order of removing protective groups in this sequence of reactions is carefully chosen to assure that migration or strand break in the RNA part is minimized.[56]

Figure 5



The same phosphotriester methodology was applied in the preparation of deoxynucleopeptides **6** and **7**, having a phosphodiester between tyrosine or threonine residues and tetra-deoxynucleotide [5'-pATAT-3'], respectively (see **Figure 6**).[57] The exocyclic amine of adenine was protected using the NPS group, whereas the terminal thymidylic acid 3'-OH function was protected with the levulinoyl (Lev) group. Both protective groups could be removed without problems using tri-*n*-butylphosphine and hydrazine respectively, avoiding strong acidic conditions. In the case of **7** two issues deserve attention. First, the C-terminal allyl ester introduced to give access to the free acid, could not be removed using Pd(0) based procedures, rendering a partially protected nucleopeptide fragment. Second, the choice of protective groups in the preparation of **7** avoids the use of strong base as it was reported in a preliminary study that harsh alkaline conditions led to β-elimination at the amino acid side chain in syntheses of H-Ala-Ser[5'-pTT-3']-NH_2 and H-Ala-Thr[5'-pTT-3']-NH_2.[53] The phosphotriester approach was also employed to give nucleopeptides containing the dideoxynucleotide [5'-pGC-3'] (see **Figure 6**, compound **8** and **9**).[58] The protecting group used on the guanidine base was di-*n*-butylformamidine (DNB), which is removable using hydrazinolysis, and NPS on cytidine. The 2-chlorophenyl groups, used as protection on the phosphodiester bonds, were removed using TBAF. The peptide C-terminal acid of **9** was protected as 2-(hydroxymethyl-9,10-anthraquinone (MAQ) ester, which could be removed using reductive cleavage.

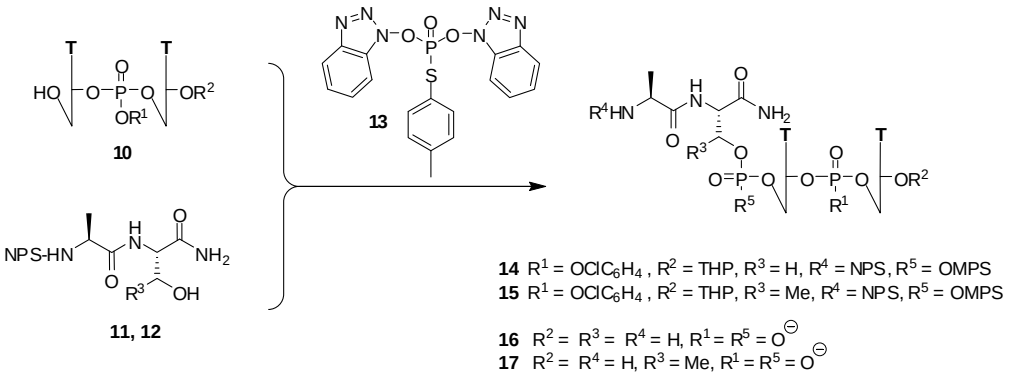
Figure 6



3.1.2 Phosphorothioate triesters

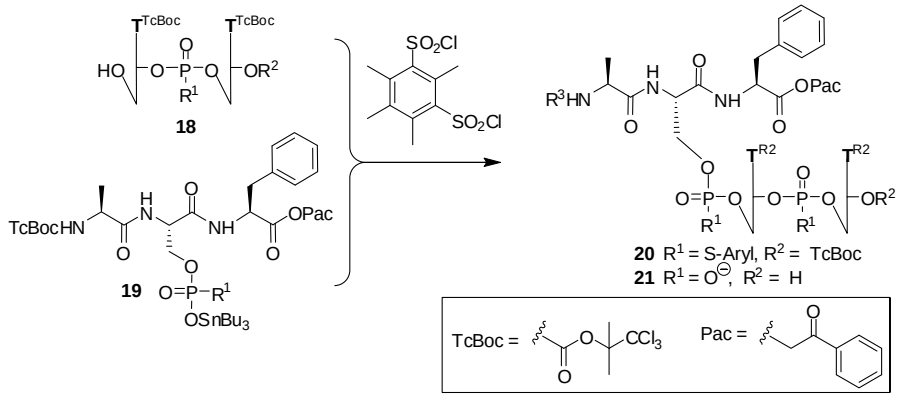
An alternative for the 2-chlorophenyl as phosphate protecting groups is a (substituted) phenylsulfenyl group.[53, 59-62] Advantage of sulfenyl groups are the relatively mild deprotection procedures such as silver acetate, iodine in pyridine followed by water or $(n\text{-}Bu_3Sn)_2O$ followed by TMSCl. These procedures avoid strong base, which is deleterious for serine and threonine containing nucleopeptides. Fully protected nucleopeptides **14** and **15** were prepared by reaction of phosphorylating agent **13** with dinucleotide **10** and subsequent condensation of the mono-hydroxybenzotriazole intermediate with dipeptides **11** or **12** (see **Figure 7**).[53,59] The sulphenyl group in nucleopeptides **14** and **15** could be easily removed using silver acetate in pyridine-water, after which further deprotection gave access to **16** and **17**.

Figure 7



Another phosphorothioate approach makes use of phosphorothioate peptides, assembled from phosphorothioate amino acids[61], that are activated and subsequently coupled to diribonucleotides to give diribonucleotide dipeptide Ala-Tyr[5'-pUU-3'] and dideoxyribonucleotide tripeptide Ala-Ser[5'-pTT-3']-Phe **21**, of which the synthesis is depicted in **Figure 8**.[60,62]

Figure 8

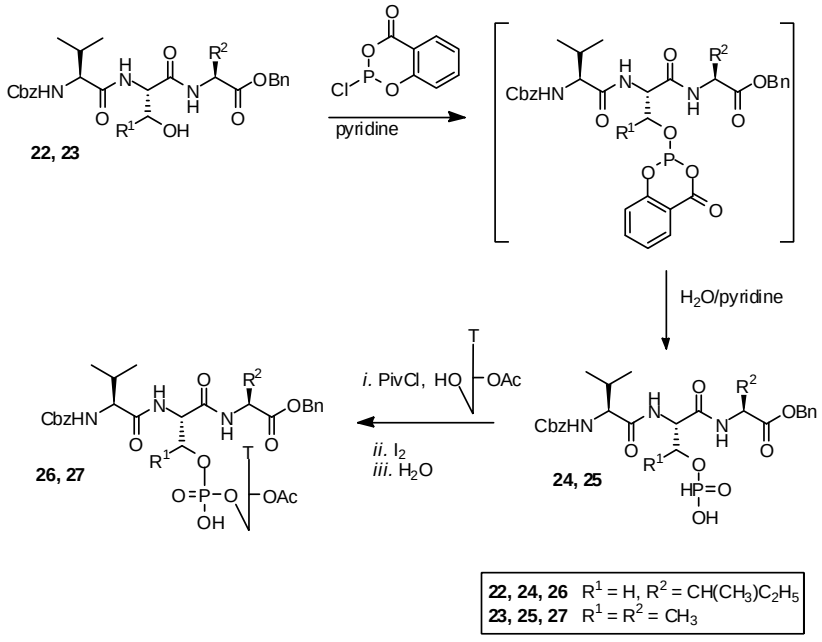


Deprotection of the phosphorothioates triesters using $(n\text{-}Bu_3Sn)_2O$ resulted in formation of the unprotected phosphodiester bonds. The (2,2,2-trichloro-*tert*-butoxycarbonyl) (TcBoc) and phenacyl (Pac) ester groups could be removed using Zn/acetylacetone in pyridine under mild conditions.

3.1.3 H-phosphonate approach

The observation that oximate based procedures to remove 2-chlorophenyl protection on the phosphotriester moiety led to β-elimination in serine nucleopeptides, encouraged the investigation of the so-called H-phosphonate approach in nucleopeptide synthesis.[31] In this approach the H-phosphonates of protected threonine and serine tripeptides (**22** and **23**, see **Figure 9**) were formed by reacting the peptide with salicylchlorophosphine to give an intermediate phosphite triester that was hydrolyzed to give H-phosphonates **24** and **25**, respectively. Subsequent reaction of H-phosphonate **24** or **25** with the free 5'-hydroxyl of 3'-*O*-acetyl thymidine under the agency of pivaloylchloride, followed by oxidation with iodine and immediate hydrolysis using water gave protected nucleopeptides **26** and **27**. Although the formation of the phosphodiester bond between thymidylic acid and peptides containing threonine or serine proved to be fruitful a similar coupling with tyrosine was not efficient, probably due to instability of the phosphite triester intermediate.

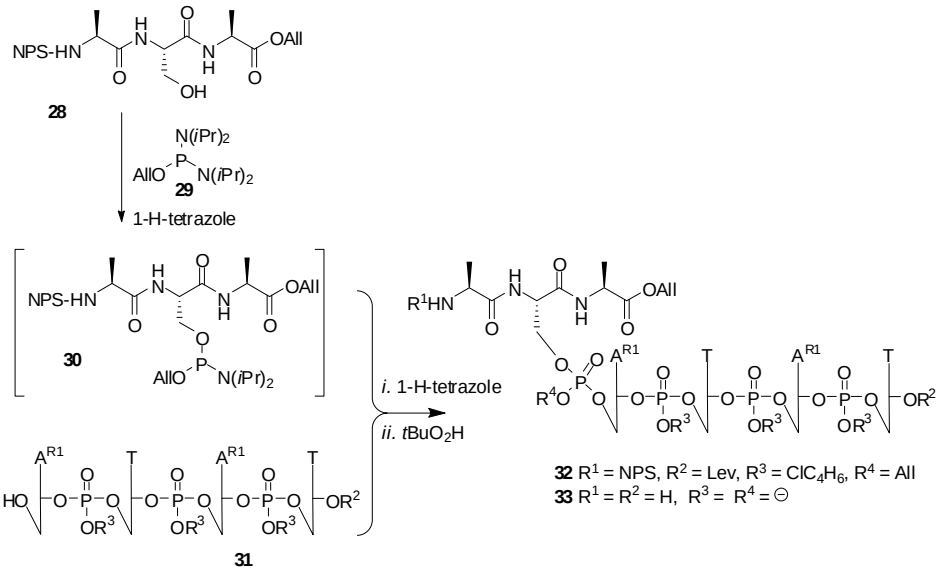
Figure 9



3.1.4 Phosphoramidite approach

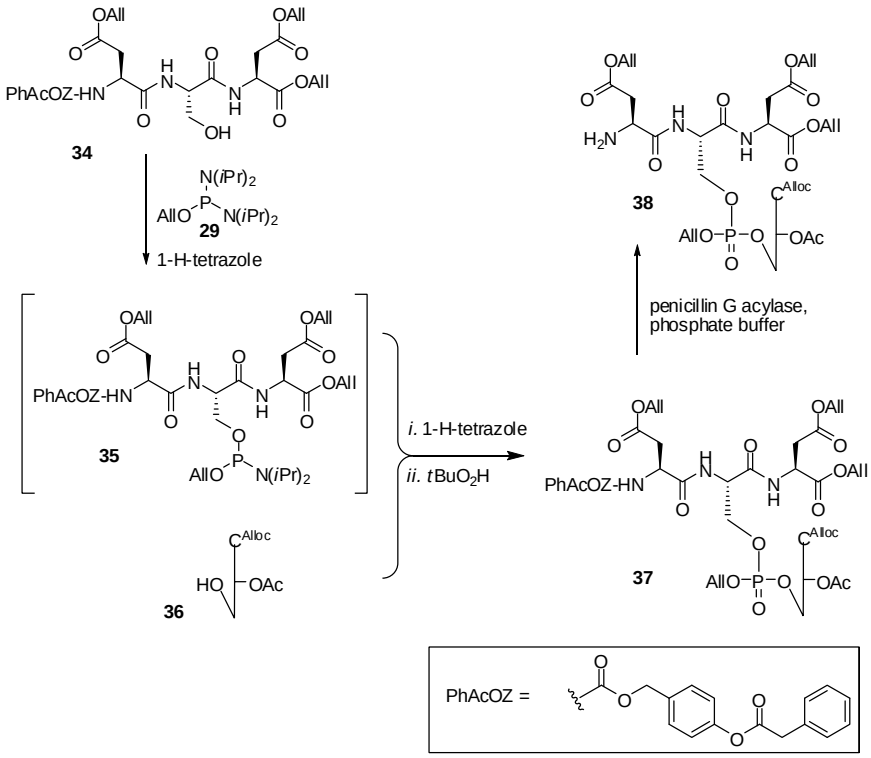
The phosphoramidite approach reported by Beaucage and Caruthers[63], based on the phosphite triester method first reported by Letsinger and Ogilvie[64], became the most used method in solid support based oligonucleotide synthesis. Therefore this method was also applied in nucleopeptide synthesis. Phosphitylation of peptide **28** using phosphordiamidite **29** gave rise to phosphoramidite **30** (see **Figure 10**).[57] The crude amidite was condensed under influence of tetrazole with separately prepared tetradeoxynucleotide **31**, followed by oxidation using peroxide to give protected nucleopeptide **32**. Using the phosphoramidite approach a higher yield (90%) is reported for the synthesis of **32** when compared to the phosphotriester method (64%).[57] Fully protected **32** was treated with TBAF in pyridine/water to cleave the phosphate protection 2-chlorophenyl and allyl groups, tri-*n*-butylphosphine in dioxane/water to cleave the NPS groups and hydrazine hydrate in pyridine/acetic acid to remove the Lev group, resulting in nucleopeptide **33**. No conditions however were found to cleave the C-terminal allyl ester.

Figure 10



Waldmann and co-workers have used a similar phosphoramidite approach to obtain nucleopeptides in which deoxycytidylic acid **36** is linked to a serine containing tripeptidyl phosphoramidite **35** (see **Figure 11**) to give fully protected nucleopeptide **37**.[65] The goal of this work was to extend the protective group repertoire used in nucleopeptide synthesis by the introduction of the phenylacetoxy benzyloxycarbonyl (PhAcOZ) group. This N-terminal PhAcOZ protective group in **37** can be removed by the enzyme penicillin G acylase under mild conditions (pH 6.8) to give **38**. No attempts were made however to remove all other protective groups on **38**.

Figure 11



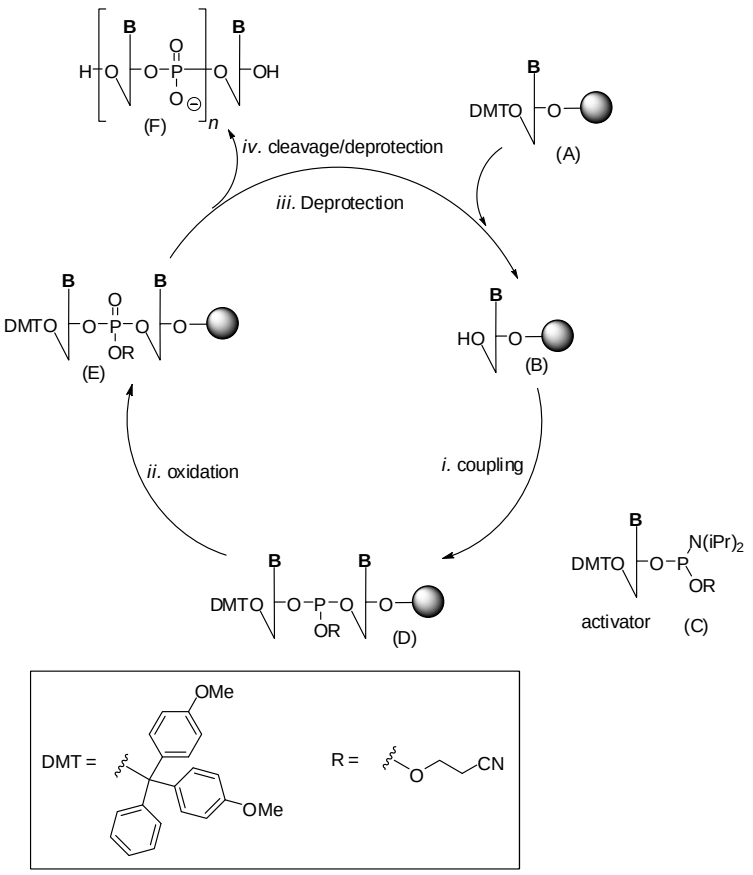
3.2 Syntheses of nucleopeptides on solid support

The (automated) solid phase synthesis of oligonucleotides and peptides eventually proved to be the best approach to obtain these biomolecules. The methodologies used in oligonucleotide and peptide chemistry were therefore adapted for the synthesis of nucleopeptides.

3.2.1 Synthesis of nucleopeptides starting from resin-bound oligonucleotides

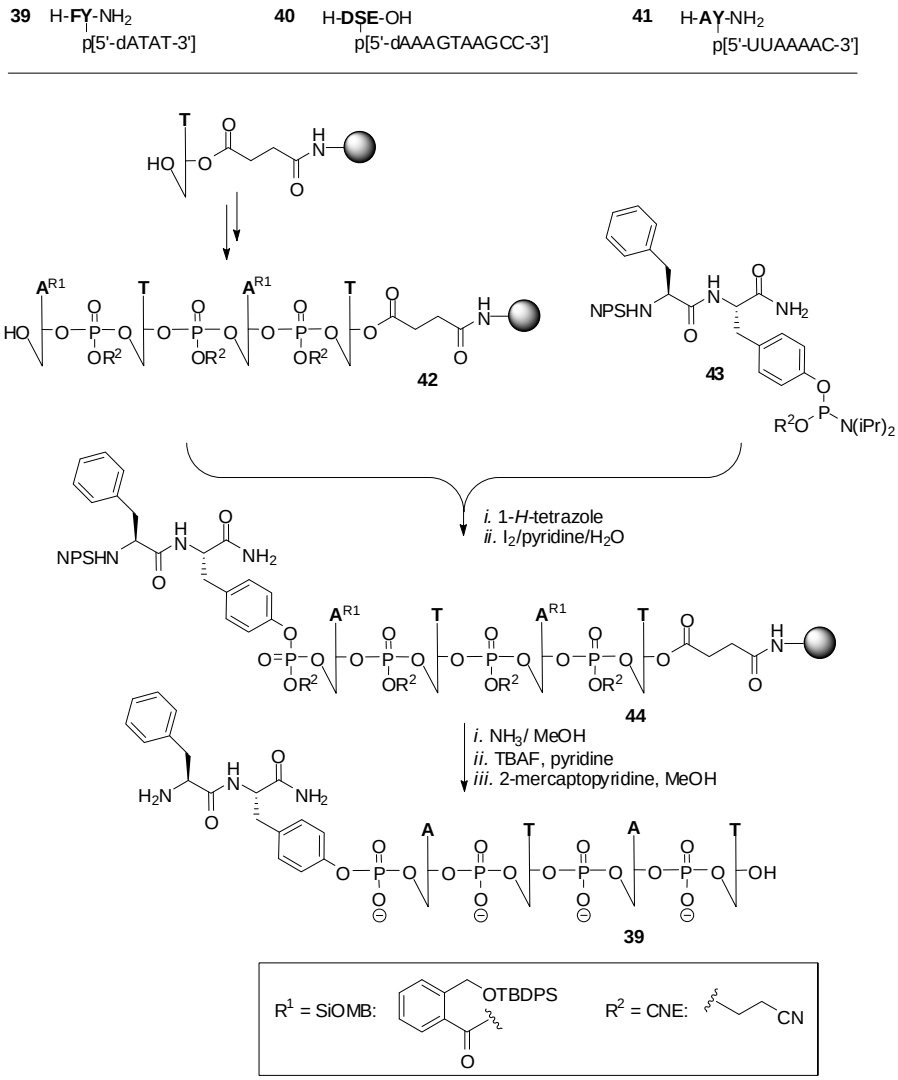
Applying the phosphoramidite methodology on a solid support was found to be most efficient in the preparation of oligonucleotides. In this method the first dimethoxytrityl (DMT) protected nucleotide is attached to the solid support (A) (see **Figure 12**) via a linker that is stable en route to the immobilized oligonucleotide (E), but can be cleaved in the final stage of the synthesis. The repetitive cycle in oligonucleotide synthesis starts with liberation of the primary hydroxyl function to give alcohol (B). Coupling with a 3'-phosphoramidite reagent (C) under agency of an activator gives phosphite triester intermediate (D), which is subsequently oxidized to give phosphotriester (E). Removal of the 5'- protecting group on the just attached nucleotide elongates the oligonucleotide with one nucleic acid residue. Repeating this cycle will elongate the resin-bound oligonucleotide chain from the 3'-end in the direction of the 5'-end, until final deprotection and release from the resin gives free oligonucleotide (F).

Figure 12



Nucleopeptides **39** - **41** were prepared via this method (exemplified for **39**, see **Figure 13**).[66-68] First oligonucleotide **42** was prepared on solid support. Next, peptidylphosphoramidite **43** was coupled and subsequently oxidized to give fully protected **44**. Deprotection and release from the solid support gave access to nucleopeptide fragment **39**. The main feature in the protective group strategy used in the construction of **39** - **41** is the use of fluoride labile 2-(*tert*-butyldiphenylsilyloxymethyl)-benzoyl (SiOMB)[69] groups on the exocyclic amines of the nucleobases. Cyanoethyl protection, and the oxalylester anchor were cleaved simultaneously with the SiOMB groups using fluoride ions in water-pyridine, proving this method to be compatible with the integrity of the nucleopeptide fragments. The 3'-hydroxyl functions of ribonucleopeptide **41** were protected with TBDMS groups, which were also cleaved by the fluoride treatment. The N-termini of **39** and **41** were masked with the NPS group, which was cleaved using 2-mercaptopyridine.

Figure 13



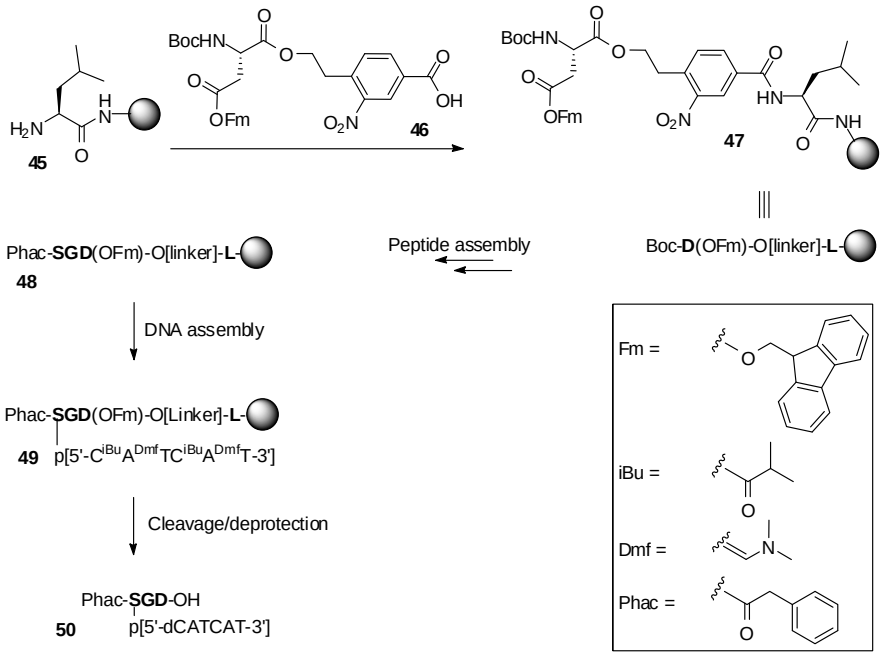
In the case of nucleopeptide **40** the peptide N-terminal amine as well as the acid functions were masked as nitro-benzyl carbamate/ester, which were removed under reductive conditions using sodium dithionite and sodium bicarbonate. Although the mild deblocking conditions needed for nitro-benzyl group removal prevented β-elimination, the deprotection was incomplete, complicating purification.

Peptidyl phosphoramidites allow access to larger nucleopeptide fragments than the solution based approaches, but there are also limitations to this technique. Low solubility of fully protected peptide fragments and the ensuing problems in preparation of peptidyl phosphoramidites[70], as well as decreased reactivity of such amidites are main disadvantages of this method.

3.2.2 Synthesis of nucleopeptides starting from resin-bound peptides

Reversing the order of nucleopeptide assembly by first preparing a resin bound peptide followed by repetitive coupling cycles of the appropriate nucleotide phosphoramidite to a hydroxyl function in the peptide backbone, might avoid the solubility and reactivity issues mentioned above. Extensive work by Grandas and co-workers has resulted in a wide repertoire of deoxynucleopeptides.[70-78] An example of this approach is represented by the synthesis of **50** (see **Figure 14**).[71] Leucyl amino methyl polystyrene resin **45** was coupled to α-(4-nitrophenyl)ethyl ester of Boc protected C-terminal aspartic acid β-fluorenemethyl ester **46**.

Figure 14



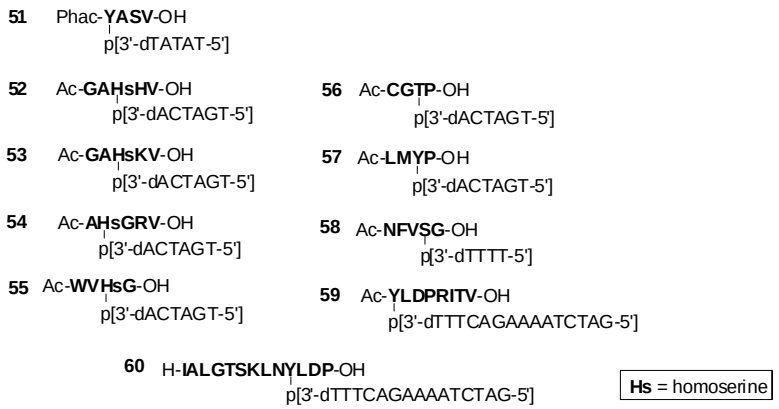
Subsequent sequential elongation of the peptide using Boc-based chemistry and final capping resulted in peptide **48**. DNA assembly on the free hydroxyl of serine was effected by repetitive coupling cycles with the appropriate 5'-phosphoramidite, capping, oxidation and liberation of the 3'-hydroxyl. In standard DNA synthesis 3'-phosphoramidites are used and the growing DNA chain is elongated from the 3'-end towards the 5'-end. In this particular case, 5'-phosphoramidites were prepared and used to elongate the growing oligonucleotide from the 5'-end towards the 3'-end. This non-standard approach proceeded without problems and final cleavage from the support by TBAF and removal of all nucleobase protective groups using ammonia in dioxane resulted in formation of the target nucleopeptide **50**.

Most of the results of the group of Grandas show examples of peptides linked to the 3'-end of DNA (in contrast to the example given in **Figure 14**, which is linked to the 5'-end). Such 3'-linked deoxyribonucleopeptides are for instance formed by topoisomerase I enzymes. Standard commercially available phosphoramidites can be used in the synthesis of such 3'-linked

nucleopeptides. Side chain protection in all naturally occurring amino acids used in this strategy are generally base-labile protective groups, of which an overview is described below.

The hydroxyl containing amino acid residues in nucleopeptide sequences, beside the linkage site for the oligonucleotide (see for instance **51**, **Figure 15**), were found to be efficiently protected as acetyl esters.[74]

Figure 15



Incorporation of a histidine residue (see for instance **52**, **Figure 15**) was reported to proceed best by using the tosyl-protection group in peptides synthesis and removing this protection with HOBt prior to assembly of the nucleic acid part.[72] It is not an option to use unprotected histidine during peptide synthesis since this causes epimerization at the α-carbon.

Lysine containing nucleopeptides (see for instance **53**, **Figure 15**) were reported to be best protected using either the fluorenemethyloxycarbonyl (Fmoc) or trifluoroacetyl (TFAc) group on the ε-amine function.[76] However, the alkaline conditions needed to cleave the TFAc group limit its use.

Arginine containing nucleopeptides (see for instance **54**, **Figure 15**) were obtained by protecting the arginine side chain using Fmoc groups.[75] Due to the difficulties in the preparation of Fmoc-Arg(Fmoc)$_2$-OH it was decided in the synthesis of **59**, (see **Figure 15**) not to use protection on arginine. Ornithine side products were found indicating cleavage of the guanidine moiety in the arginine side chain. Other side reactions found at the guanidine group, when using unprotected arginine are acylation by amino acids in subsequent peptide coupling steps or nucleotide incorporation.[78]

Tryptophane containing nucleopeptides (see for instance **55**, **Figure 15**) were obtained by protecting the indole ring with the formyl group, which could be efficiently removed by an ammonia treatment.[76]

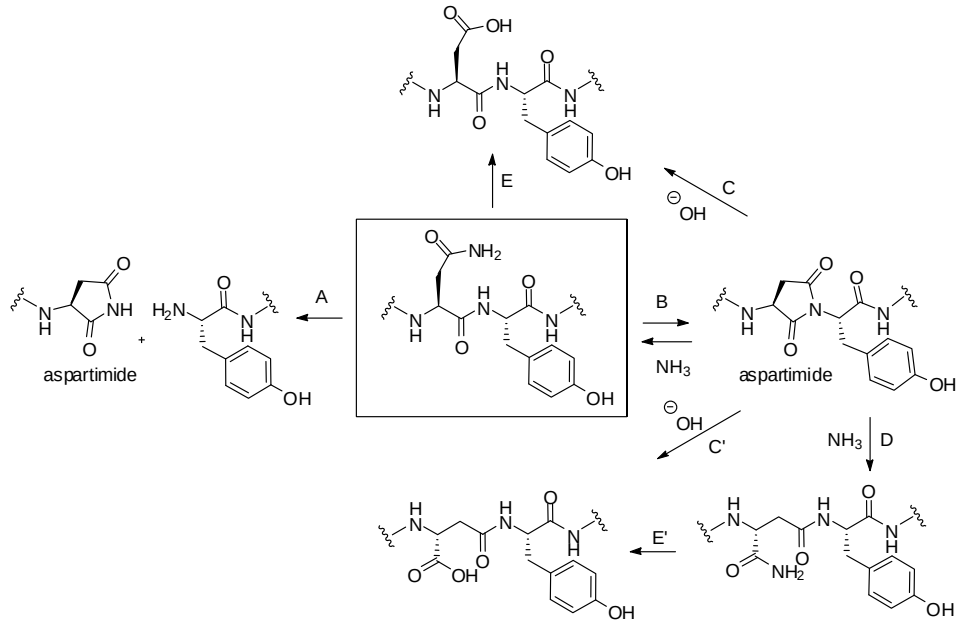
For cysteine containing peptides the standard protection of cysteine residues using the *tert*-butyl sulfenyl group is reported to give β-elimination in the final aqueous ammonia treatment.[79] The resulting dehydroalanine can be converted to serine or 2,3-diaminopropionic acid residues via conjugate addition. Prolonged heating of **56** (see **Figure 15**) in NH_3/ dioxane indeed showed

complete conversion of cysteine to serine.[78] After cleavage of the S-*t*Bu group the free thiol is prone do disulfide formation or to alkylation by the acrylonitrile generated by cleavage of the cyanoethyl phosphate protective group. To avoid above mentioned problems the best method found so far is cleaving the S-*t*Bu using NH_3/ 1M DTT in dioxane for 8 hours at room temperature.

The synthesis of methionine containing nucleopeptides (see **57**, **Figure 15**) can be accompanied by alkylation as a side reaction. Furthermore, aqueous iodine conditions used to oxidize phosphites to phosphotriesters might cause iodination and subsequent oxidation of the methionine residues. In peptide chemistry a well known method is to use methionine sulfoxide analogues, that are reduced after chain assembly, with for instance *N*-methylmercaptoacetamide. This strategy proved efficient for the synthesis of nucleopeptide **57**.[80]

Asparagine and glutamine containing nucleopeptides were synthesized using unprotected Asn or Gln building blocks.[76,78] Although primary carboxamides are reported to react with phosphoramidites[81] no significant amount of branched products were detected in the synthesis of an asparagine containing nucleopeptide (see **58**, **Figure 15**). In the synthesis of **60** the aqueous ammonia treatment to remove acyl protection groups resulted in the formation of several side products.[78] Attack of the side chain carboxamide on the α-carbon of asparagine can result in the cleavage of the peptide backbone (pathway A, **Figure 16**), which indeed was observed.

Figure 16



On the other hand, attack of the NH in the amide bond between tyrosine and asparagine can result in the formation of an aspartimide (pathway B, **Figure 16**). Hydrolytic opening of this aspartimide can give rise to the aspartic acid containing peptide (pathway C) and to the β-peptide (pathway C'). Deamidation of the nucleopeptide can give rise to the aspartic acid containing peptide as well (pathway E). Opening of the aspartimide ring by ammonia can give rise to the β-peptide isomer (pathway D). Extensive characterization of the formed products however is difficult due to their size and complexity. The only performed analysis on the formed products is MALDI-TOF analysis which cannot cast a light on the identity of the isolated product and the side-reaction(s) that have occurred in the synthesis of **60**.

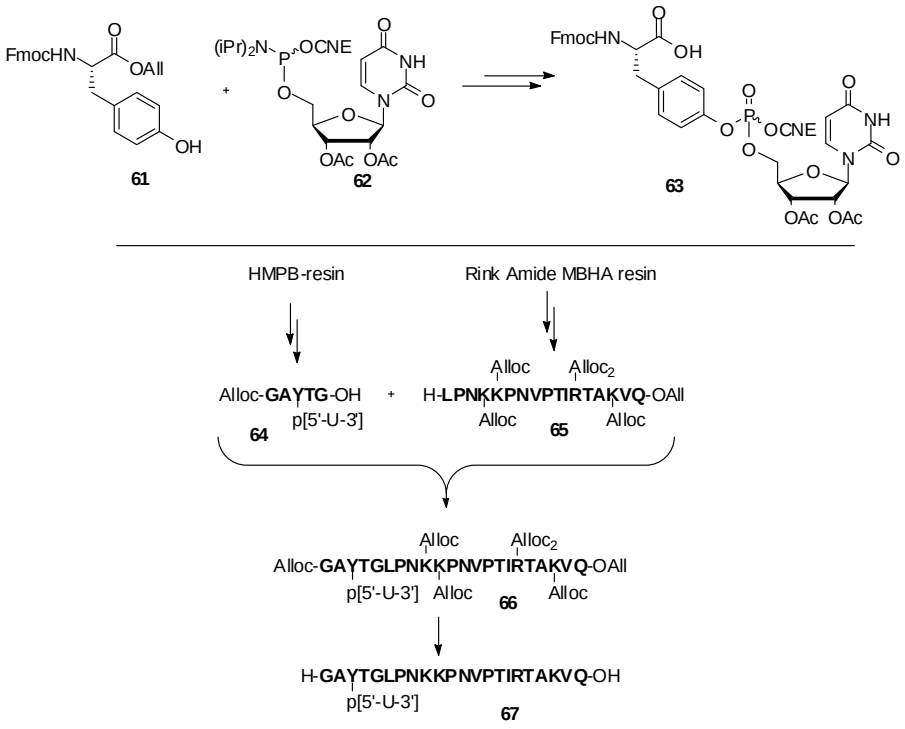
In studies on the incorporation of histidine, lysine, arginine and tryptophan in the nucleopeptides described above, it was opted to use an artificial linkage site between the peptide and deoxyoligonucleotide in order to circumvent problems such as β-elimination, associated with the labile nature of serine and threonine.[72] The use of homoserine (Hs) instead of serine not only improved the stability under basic conditions[74] but also improves the stability of nucleopeptides against 3'-exonuclease activity in biological systems.[75]

Although the above described solid phase based methodology is a powerful tool to create nucleopeptides, still the synthesis of each specific nucleopeptide is a challenge. Potential side reactions in both the peptide and nucleic acid parts and base lability are recurring problems, the severity of which varies from case to case.

3.2.3 Use of pre-uridylylated building blocks

Uridylylated VPg's are proposed to act as primers in viral replication and for this reason are interesting targets for nucleopeptide synthesis. The *mono*-uridylylated VPg (VPgpU) of polio virus type 1 (see **Figure 17**) was synthesized using a block coupling approach.[82]

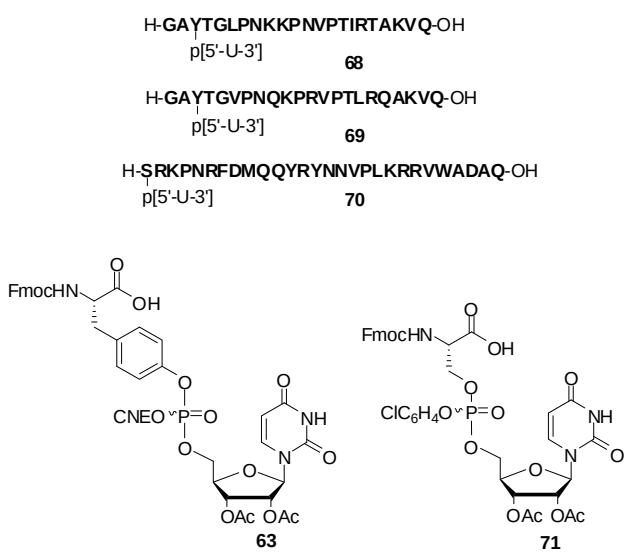
Figure 17



In this strategy pre-uridylylated tyrosine building **63** is of crucial importance. Coupling of protected tyrosine **61** and phosphoramidite **62**, followed by oxidation and protective group manipulations renders **63**. Pentameric peptide **64** was assembled on very acid-labile HMPB resin using pre-uridylylated tyrosine **65** via a Fmoc-based peptide synthesis protocol (in contrast to the Boc-based synthesis of nucleopeptides described in section *1.2.2*). After release from the solid support and removal of all protective groups except the N-terminal Alloc, nucleopeptide **64** was coupled to partially protected heptadecamer **65**. This protected peptide was prepared separately on Rink amide MBHA resin. Condensation of **64** and **65** using the coupling mixture BOP/HOBt resulted in protected uridylylated VPgpU **66**, which was deprotected using Pd(0)-catalyzed hydrostannolysis of the Alloc groups and allyl ester to give the unprotected nucleoprotein **67**. The use of pre-uridylylated building blocks was also examined in the fully sequential solid support based synthesis of the VPgpU's of polio virus (**68**), coxsackie virus (**69**) and cowpea mosaic virus (**70**, see **Figure 18**).[83] Starting from Rink Amide MBHA resin the VPgpU's were constructed in a standard peptide solid phase approach using uridylylated

tyrosine **63** or uridylylated serine **71** (see **Figure 18**). Protective groups on the amino acid side chain functionalities used in this protocol were all acid labile. Acidic cleavage of protective groups on the peptide side chains and liberation of the peptides from the solid support followed by ammonia treatment to remove the acyl protection on the nucleoside as well as the N-terminal Fmoc gave tyrosine nucleopeptides **68** and **69** without problems. In the synthesis of serine nucleopeptide **70**, the phosphodiester moiety was protected using the 2-chlorophenyl protective group instead of the cyanoethyl, since the 2-chlorophenyl can be removed using fluoride ions in aqueous solvent (water/pyridine) to avoid strong base.

Figure 18

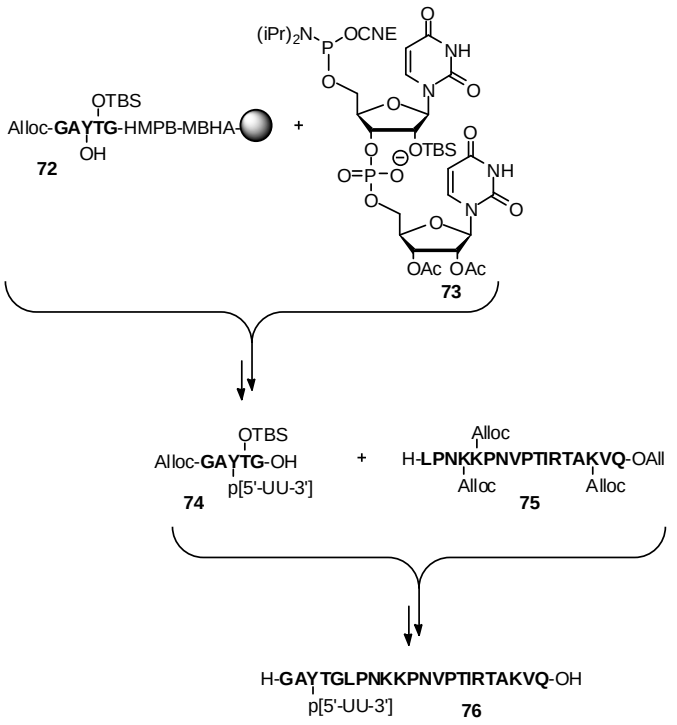


The advantage of this stepwise approach versus the block coupling method described earlier to obtain *mono*-uridylylated nucleoproteins is the ease of sequential synthesis and the use of protective groups that are cleaved using a simpler less complicated deprotection protocol. However this method does not allow the introduction of oligonucleic acid fragments since strong acid (95% trifluoroacetic acid) is used during the deprotection.

3.2.4 Solid phase convergent synthesis

Using a similar block coupling approach the double uridylylated nucleoproteins of coxsackie virus and polio virus type 1 (VPgpUpU's) were synthesized (as exemplified for polio virus VPgpUpU **76**, see **Figure 19**).[84] Pentapeptide **74** was prepared by reacting phosphoramidite **73** with resin bound **72**, followed by oxidation and cleavage of acyl protective groups and concomitant release from the solid support. Coupling of double uridylylated pentapeptide **74** to the appropriate, protected C-terminal heptadecamers of coxsackie virus or polio virus respectively followed by deprotection gives access to the respective VPgpUpU's. In a previously reported block coupling approach towards VPgpU nucleoproteins, the arginine residues were protected with Alloc groups.[82]

Figure 19



During solid phase peptide synthesis however, the δ-Alloc group on arginine was found to be partially cleaved causing side reactions to occur and thereby complicating the purification process.[84] For this reason the arginine residue in the heptadecamer fragment en route to **75** was protected with the acid labile Pbf protective group which was removed prior to the block coupling step.

Outline of this thesis

Chapter 2 of this thesis describes the properties of a phosphoramidite equipped with two *p*-methoxybenzyl groups. Arbuzov like reactions allow the preparation of phosphates, phosphoramidates, phosphorothioates, pyrophosphates and phosphofluoridates in a one-pot procedure.
Chapter 3 describes the synthesis of pre-ribosylated glutamine and asparagine building blocks and their incorporation in peptides, that are natural sites of ADP-ribosylation. The methodology described in **Chapter 2** to create phosphates and pyrophosphates is applied in the synthesis of *mono*-ADP-ribosylated peptides both in solution and solid support based approaches.
Chapter 4 deals with the ribosylation of adenosine at the 2'-OH in an α-selective fashion. This glycosylated nucleoside is the repetitive unit of *poly*-ADP-ribose. The introduction of an orthogonal protective group pattern is described that would allow the use of this building block in stepwise *poly*-ADP-ribosylated peptide synthesis.
In **Chapter 5** a general method to obtain five picornavirus VPgpU's is reported using a pre-uridylylated tyrosine building block.
In **Chapter 6** the method from **Chapter 5** is extended to generate the other three canonical pre-nucleotidylated tyrosine building blocks. These are then used to generate the corresponding nucleotidylated polio virus VPg proteins.
Chapter 7 describes a linear solid support based synthesis of the nucleopeptide derived from the N-terminal region of picornaviral VPg. This approach demonstrates RNA synthesis using 5'-phosphoramidites on solid phase bound peptides.

References and Notes

[1] Merrifield, R. B.; *J. Am. Chem. Soc.*, **1963**, *85*, 2149 – 2154.

[2] Merrifield, B.; *Science*, **1986**, *232*, 341 – 347.

[3] Caruthers, M. H.; *Science*, **1985**, *230*, 281 – 285.

[4] Beaucage, S. L.; Iyer, R. P.; *Tetrahedron*, **1993**, *49*, 10441 – 10488.

[5] Zubin, E. M.; Romanova, E. A.; Oretskaya, T. S.; *Russian Chem. Rev.*, **2002**, *71*, 239 – 264. 731.

[6] Tate, E. W.; *J. Chem. Biol.*, **2008**, *1*, 17 – 26.

[7] Collins, M. O.; Yu, L.; Choudhary, J. S.; *Proteomics,* **2007**, *7*, 2751 – 2768.

[8] Mischerikow, N.; Heck, A. J. R.; *Proteomics,* **2011**, *11*, 571 – 589.

[9] Welchman, R. L.; Gordon, C.; Mayer, R. J., *Nat. Rev. Mol. Cell Biol.*, **2005**, *6*, 599 – 609.

[10] Paleari, R.; Paglietti, E.; Mosca, A.; Mortarino, M.; Maccioni, L.; Satta, S.; Cao, A.; Galanello, *Clin. Chem.*, **1999**, *45*, 21 – 28.

[11] Corda, D.; Di Girolamo, M.; *EMBO J*., **2003**, *22*, 1953 – 1958.

[12] Hassa, P. O.; Haenni, S. S.; Elser, M.; Hottiger, M. O.; *Microbiol. Mol. Biol. Rev.,* **2006**, *70*, 789 –829.

[13] Okana, S.; Lan, L.; Caldecott, K. W.; Mori, T.; Yasui, A.; *Mol. Cell. Biol.*, **2003**, 3974 – 3981.

[14] Oliver, J.; Menissier-de Murcia, J.; Nacci, C.; Decker, P.; Andreantsitohaina, R.; Muller, S.; de la Rubia, G.; Stoclet, J. C.; de Murcia, G.; *EMBO J.,* **1999**, *18*, 4446 – 4454.

[15] Scheuring, J.; Schramm, V. L.; *J. Am. Chem. Soc.,* **1995**, *117*, 12653 – 12654.

[16] Scheuring, J.; Schramm, V. L.; *Biochemistry,* **1997**, *36*, 8215 – 8223.

[17] Scheuring, J.; Berti, P. J.; Schramm, V. L.; *Biochemistry,* **1998**, *37*, 2748 – 2758.

[18] Parikh, S. L.; Schramm, V. L.; *Biochemistry,* **2004**, *43*, 1204 – 1212.

[19] Moyle, P. M. ; Muir, T. W.; *J. Am. Chem. Soc.,* **2010**, *132*, 15878 – 15880.

[20] Comstock, L. R.; Denu, J. M.; *Org. Biomol. Chem.*, **2007**, *5*, 3087 – 3091.

[21] Szczepankiewicz, B. G.; Koppetsch, K. J., Perni, R. B.; *J. Org. Chem.*, **2011**, *76*, 6465 – 6474.

[22] Zhou, G.; Parikh, S. L.; Tyler, P. C.; Evans, G. B.; Furneaux, R. H.; Zubkova, O. V.; Benjes, P. A.; Schramm, V. L.; *J. Am. Chem. Soc.,* **2004**, *126*, 5690 – 5698.

[23] Zhang, J.; Snyder, S. H.; *Biochemistry,* **1993**, *32*, 2228 – 2233.

[24] Jiang, H.; Kim, J. H.; Frizzell, K. M.; Krause, W. L.; Lin, H.; *J. Am. Chem. Soc.,* **2010**, *132*, 9363 – 9372.

[25] Yarbrough, M. L.; Orth, K.; *Nat. Chem. Biol.*, **2009**, *5*, 378 – 379.

[26] Itzen, A.; Blankenfeldt, W.; Goody, R. S.; *Trends in Biochemical Sciences*, **2011**, *36*, 221 – 228.

[27] Yarbrough, M. L.; Li, Y.; Kinch, L. N.; Grishin, N. V.; Ball, H. L.; Orth, K.; *Science*, **2009**, *323*, 269 – 272.

[28] Worby, C. A.; Mattoo, S.; Kruger, R. P., Corbeil, L. B.; Koller, A.; Mendez, J. C.; Zekarias, B.; Lazar, D.; Dixon, J. E.; *Mol. Cell,* **2009**, *34*, 93 – 103.

[29] Muller, M. P.; Peters, H.; Blumer, J.; Blankenfeldt, W.; Goody, R. S.; Itzen, A.; *Science,* **2010**, *329*, 946 – 949.

[30] Al-Eryani, R. A.; Li. Y.; Ball, H. L.; *Tetrahedron Lett.*, **2010**, *51*, 1730 – 1731.

[31] Kuyl-Yeheskiely, E.; Tromp, C. M; Schaeffer, A. H.; van der Marel, G. A.; van Boom, J. H.; *Nucleic Acids Res.*, **1987**, *15*, 1807 – 1818.

[32] Albers, M. F.; van Vliet, B.; Hedberg, C.; *Org. Lett.*, **2011**, *13*, 6014 – 6017.

[33] Smit, C.; Blumer, J.; Eerland, M. F.; Albers, M. F.; Muller, M. P.; Goody, R. S.; Itzen, A.; Hedberg, C.; *Angew. Chem. Int. Ed.*, **2011**, *50*, 9200 – 9204.

[34] Guijarro, J. I.; Gonzales-Pastor, J. E.; Baleux, F.; San Millan, J. L.; Castilla, M.; Rico, M.; Moreno, F.; Delepierre, M.; *J. Biol. Chem.*, **1995**, *270,* 23520 – 23532.

[35] Agarwal, V.; Metlitskaya, A.; Severinov, K.; Nair, S. K.; *J. Biol. Chem.*, **2011**, *286,* 21295 – 21303.

[36] Roush, R. F.; Nolan, E. M.; Lohr, F.; Walsh, C. T.; *J. Am. Chem. Soc.*, **1998**, *130,* 3603 – 3609.

[37] Matsunaga, S.; Takahashi, N.; Fusetani, N.; *Pure Appl. Chem.*, **2009**, *81,* 1001 – 1008.

[38] Phillips, D. R.; Uramoto, M.; Isono, K.; Mc. Closkey, J. A.; *J. Org. Chem.*, **1993**, *58,* 854 – 859.

[39] Moriguchi, T.; Asai, N.; Okada, K.; Seio, K.; Sasaki, T.; Sekine, M; *J. Org. Chem.*, **2002**, *67,* 3290 – 3300.

[40] Schein, C. H.; Volk, D. E.; Oezguen, N; Paul, A.; *Proteins*, **2006**, *63*, 719 – 726.

[41] Steil, B. P.; Barton, D. J.; *Virus Res.*, **2009**, *139*, 240 – 252.

[42] Desiderio, S. V.; Kelly Jr., T. J.; *J. Mol. Biol.,* **1981**, *145*, 319 – 337.

[43] Pronk, R.; Stuiver, M. H.; van der Vliet, P. C.; *Chromosoma*, **1992**, *102*, S39 – S45.

[44] Smart, J. E.; Stillman, J. W.; *J. Mol. Biol.,* **1982**, *257*, 13499 –13506.

[45] Hermoso, J. M.; Salas, M.; *Proc. Natl. Acad. Sci. USA*, **1980**, *77*, 6425 – 6428.

[46] Salas, M.; Mellado, R. P.; Vinuela, E.; *J. Mol. Biol.,* **1978**, *119*, 269 – 291 .

[47] van Mansfeld, A. D. M.; van Teeffelen, H. A. A. M.; Baas, P. D.; Veeneman, G. H.; van Boom, J. H.; Jansz, H. S.; *FEBS Lett.*, **1984**, *173*, 351 – 356.

[48] Wang, J. C.; *Annu. Rev. Biochem.*, **1996**, *65*, 635 – 692.

[49] Connelly, J. C.; Leach, D. R. F.; *Mol. Cell*, **2004**, *13*, 307 – 316.

[50] Chen, A. Y.; Liu, L. F.; *Annu. Rev. Pharmacol. Toxicol.*, **1994**, *34*, 191 – 218.

[51] Fontouro, B. M.; Atienza, C. A.; Sorokina, E. A.; Morimoto, T.; Carroll, R. B.; *Mol. Cel. Biol.*, **1997**, *17*, 3146 – 3154.

[52] Samad, A.; Carroll, R. B.; *Mol. Cel. Biol.*, **1991**, *11*, 1598 – 1606.

[53] Kuyl-Yeheskiely, E.; van der Klein, P. A. M.; Visser, G. M.; van der Marel, G. A.; van Boom, J. H.; *Recl. Trav. Chim. Pays-Bas.*, **1986**, *105*, 67 – 70.

[54] Schattenkerk, C.; Wreesman, C. T. J.; de Graaf, M. J.; van der Marel, G. A.; van Boom, J. H.; *Tetrahedron Lett.*, **1984**, *25*, 5197 – 5200.

[55] Reese, C. B.; Zard, L.; *Nucleic Acids Res.*, **1981**, *9*, 4611 – 4626.

[56] Beaucage, S. L.; Iyear, R. P.; *Tetrahedron*, **1992**, *48*, 2223 – 2311.

[57] Kuyl-Yeheskiely, E.; Tromp, C. M.; Lefeber, A. W. M.; van der Marel, G. A.; van Boom, J. H.; *Tetrahedron*, **1988**, *20*, 6515 – 6523.

[58] Kuyl-Yeheskiely, E.; Dreef-Tromp, C. M.; Geluk, A.; van der Marel, G. A.; van Boom, J. H.; *Nucleic Acids Res.*, **1989**, *17*, 2897 – 2905.

[59] Kuyl-Yeheskiely, E.; Tromp, C. M.; Geluk, A.; van der Marel, G. A.; van Boom, J. H.; *Recl. Trav. Chim. Pays-Bas.*, **1988**, *107*, 567 – 569.

[60] Hotoda, H.; Ueno, Y.; Sekine, M.; Hata, T.; *Tetrahedron Lett.*, **1989**, *30*, 2117 – 2120.

[61] Ueno, Y.; Saito, R.; Hata, T.; *Tetrahedron Lett.*, **1991**, *32*, 1347 – 1350.

[62] Ueno, Y; Saito, R.; Hata, T.; *Nucleic Acids Res.*, **1993**, *21*, 4451 – 4457.

[63] Letsinger, R. L.; Ogilvie, K. K.; *J. Am. Chem. Soc.*, **1967**, *89*, 4801 – 4803.

[64] Beaucage, S. L.; Caruthers, M. H.; *Tetrahedron Lett.*, **1981**, *22*, 1859 – 1862.

[65] Jeyaraj, D. A.; Prinz, H.; Waldmann, H.; *Chem. Eur. J.*, **2002**, *8*, 1879 – 1887.

[66] Dreef-Tromp, C. M.; van Dam, E. M. A.; van den Elst, H.; van der Marel, G. A.; van Boom, J. H.; *Nucleic Acids Res.*, **1990**, *18*, 6491 – 6495.

[67] Dreef-Tromp, C. M.; van den Elst, H.; van den Boogaart, J. E.; van der Marel, G. A.; van Boom, J. H.; *Nucleic Acdis Res.*, **1992**, *20*, 2435 – 2439.

[68] Dreef-Tromp, C. M.; van der Maarel, J. E.; van den Elst, H.; van den Boogaart, J. E.; van der Marel, G. A.; van Boom, J. H.; *Nucleic Acids Res.*, **1992**, *20*, 4015 - 4020.

[69] Dreef-Tromp, C. M.; Hoogerhout, P.; van der Marel, G. A.; van Boom, J. H.; *Tetrahedron Lett.*, **1990**, *31*, 427 – 430.

[70] Robles, J.; Pedroso, E.; Grandas, A.; *Nucleic Acids Res.*, **1995**, *23*, 4151 – 4161.

[71] Robles, J.; Pedroso, E.; Grandas, A.; *Tetrahedron Lett.*, **1994**, *35*, 4449 – 4452.

[72] Beltran, M.; Pedroso, E.; Grandas, A.; *Tetrahedron Lett.*, **1998**, *39*, 4115 – 4118.

[73] Robles, J.; Pedroso, E.; Grandas, A.; *J. Org. Chem.*, **1994**, *59*, 2482 – 2486.

[74] Beltran, M.; Perez, Y.; Robles, J.; Pedroso, E.; Grandas, A.; *Nucl. Nucl.*, **1997**, *16*, 1487 – 1488.

[75] Robles, J.; Maseda, M.; Beltran, M.; Concernau, M.; Pedroso, E.; Grandas, A.; *Bioconjugate Chem.*, **1997**, *8*, 785 – 788.

[76] Marchan, V.; Debethune, L.; Beltran, M.; Robles, J.; Travesset, I.; Fabregas, G.; Pedroso, E.; Grandas, A.; *Nucl. Nucl.*, **1999**, *18*, 1493 – 1494.

[77] Robles, J.; Beltran, M.; Marchan, V.; Perez, Y.; Travesset, I.; Pedroso, E.; Grandas, A.; *Tetrahedron*, **1999**, *55*, 13251 – 13264.

[78] Debethune, L.; Marchan, V.; Beltran, M.; Fabregas, G.; Pedroso, E.; Grandas, A.; *Tetrahedron*, **2002**, *58*, 6965 – 6978.

[79] Schwope, I.; Bledzinski, C. F.; Richert, C. J.; *J. Org. Chem.*, **1999**, *64*, 4749 – 4761.

[80] Marchan, V.; Rodriguez-Tanty, G.; Estrada, M.; Pedroso, E.; Grandas, E.; *Eur. J. Org. Chem.*, **2000**, 2495 – 2500.

[81] Robles, J.; Pedroso, E.; Grandas, A.; *J. Org. Chem.*, **1995**, *60*, 4856 – 4861.

[82] Filippov, D. V.; Kuyl-Yeheskiely, E.; van der Marel, G. A.; Tesser, G. I.; van Boom, J. H.; *Tetrahedron Lett.*, **1998**, *39*, 3597 – 3600.

[83] Kriek, N. M. A. J.; Filippov, D. V.; van den Elst, H.; Meeuwenoord, N. J.; Tesser, G. I.; van Boom, J. H.; van der Marel, G. A.; *Tetrahedron*, **2003**, *59*, 1589 – 1597.

[84] Kriek, N. M. A. J.; Meeuwenoord, N. J.; van den Elst, H.; Heus, H. A.; van der Marel, G. A.; Filippov, D. V.; *Org. Biomol. Chem.*, **2006**, *59*, 3576 – 3586.