Università degli Studi di Padova
Dipartimento di Scienze Chimiche
Dottorato di Ricerca in Scienze Molecolari indirizzo Scienze Chimiche

Université Louis Pasteur de Strasbourg
Faculté de Chimie
Doctorat en Sciences chimiques

Ph.D. Thesis

Development of novel helical nucleopeptides for applications as nucleic acid modulators

Tutor: Prof. Fernando Formaggio (Università degli studi di Padova)

Cotutor: Dr. Alberto Bianco (ICT-IBMC. UPR 9021 CNRS, Strasbourg)

Ph.D. student: Piero Geotti-Bianchini

2th February 2009

Referees:

Prof. Paolo Maria Scrimin, Dipartimento di Scienze Chimiche, Padova.

Prof. Nicolas Winssinger, Institut de Science et d'Ingénierie Supramoléculaires, Strasbourg.

Prof. Jean-Alain Fehrentz, Institut de Biomolécules Max Mousseron, Montpellier, France.

Prof. Roberto Corradini, Dipartimento di Chimica Organica e Industriale, Università di Parma, Italia.

Coordinator of the Doctoral School in Padova: Prof. Maurizio Casarin

Università degli Studi di Padova
Dipartimento di Scienze Chimiche
Dottorato di Ricerca in Scienze Molecolari indirizzo Scienze Chimiche

Université Louis Pasteur de Strasbourg
Faculté de Chimie
Doctorat en Sciences chimiques

Ph.D. Thesis

Development of novel helical nucleopeptides for applications as nucleic acid modulators

Tutor: Prof. Fernando Formaggio (Università degli studi di Padova)

Cotutor: Dr. Alberto Bianco (ICT-IBMC. UPR 9021 CNRS, Strasbourg)

Ph.D. student: Piero Geotti-Bianchini

2th February 2009

Referees:

Prof. Paolo Maria Scrimin, Dipartimento di Scienze Chimiche, Padova.

Prof. Nicolas Winssinger, Institut de Science et d'Ingénierie Supramoléculaires, Strasbourg.

Prof. Jean-Alain Fehrentz, Institut de Biomolécules Max Mousseron, Montpellier, France.

Prof. Roberto Corradini, Dipartimento di Chimica Organica e Industriale, Università di Parma, Italia.

Coordinator of the Doctoral School in Padova: Prof. Maurizio Casarin

To Elisa,

who, by making me a better man,

has made me also a better scientist.

Foreword

Mr. Piero Geotti-Bianchini was a member of the European Doctoral College of the Universities of Strasbourg during the preparatio of his Ph.D, from 2006 to 2008, class name Virginia Woolf. He has benefited from specific financial supports offered by the College and, along with his mainstream research, has followed a special course on topics of general European interest presented by international experts.

This Ph.D research project has been led with the collaboration of two universities: the Università di Padova, Italy and the Université Louis Pasteur in Strasbourg, France.

INDEX

Abstract

Gene therapy aims at the treatment of genetic diseases at molecular level by interactions with nucleic acids (DNA and RNA). Synthetic oligonucleotides can selectively recognize complementary sequences and inhibit or modulate gene expression *in vitro*. However, it is not possible to use synthetic oligonucleotides for *in vivo* applications, due to several serious problems, particularly their rapid enzymatic hydrolysis and their extreme difficulty to cross cell membranes.

To overcome such limitations, nucleotide analogues, which should have better biological properties while retaining, or even improving, the affinity and selectivity towards complementary strands characteristic of the natural nucleotide structures, have been designed. Nucleopeptides are one of the recently developed classes of analogues; they are peptides containing nucleoamino acid residues, non proteogenic amino acids carrying nucleobases at their side chains.

The work presented in this thesis concerns the study of the properties of sequential helical nucleopeptides containing alanyl-nucleoamino acid residues (AlaB) at *i*,*i*+3 positions in view of applications as nucleic acid modulators. Indeed, if such nucleopeptides assume a 3_{10}-helical conformation, side chain nucleobases are aligned along the helical axis and this might favour cooperative interactions with complementary functionalized strands.

To better evaluate the importance of structuration on the nucleopeptide pairing properties, both rigid (based on the 3_{10} helix promoting C^{α}-tetrasubstituted residue Aib) and flexible (based in the less structuring proteogenic residue Ala) nucleopeptides have been synthesized and conformational and biological properties of both kinds of nucleopeptides have been investigated.

As regards the rigid nucleopeptides, this work reports on:

(I) The synthesis and characterization of protected nucleopeptides, each containing one of the four DNA nucleobases;

(II) The synthesis of protected hexa-nucleopeptide and of a completely deprotected, water-soluble hepta-nucleopeptide, both containing two thymines;

(III) The crystal structure of a protected thymine containing nucleo-tripeptide;

(IV) The conformational characterization in solution of the protected nucleo-tripeptides, of the nucleo-hexapeptide and of the nucleo-heptapeptide.

As regards the flexible nucleopeptides, this work reports on:

(I) The design and solid-phase parallel synthesis of model nucleopeptide libraries, for synthetic protocol optimization and conformational analysis;

(II) The design and solid-phase parallel synthesis of partially biotinylated or fluorescently derivatized nucleopeptide libraries containing each of the four DNA nucleobases, for pairing and biological tests;

(III) The conformational characterization by CD and 2D NMR of a thymine-based model nucleo-heptapeptide containing two nucleobases, and of an analogue carrying an Aib residue in the middle of its sequence.

(IV) The surface plasmon resonance analysis of nucleopeptide/nucleopeptide and of nucleopeptide/nucleotide pairing properties, through immobilization of biotinylated molecules on functionalized gold chips.

As regards the assessment of the nucleopeptide biological properties, this works reports on:

(I) Cytotoxicity tests carried on a rigid and on variously functionalized flexible thymine-based nucleopeptides;

(II) Flow-cytometric and microscopy cell penetration experiments carried on fluorescent and biotinylated thymine-based nucleopeptides, supported by colocalization tests and by time, concentration, and energy dependence analysis of cell uptake.

(III) Stability tests in mouse serum carried on flexible, mixed, and rigid thymine containing heptapeptides.

Finally, the whole of the data collected have been evaluated in order to give a global judgement on the properties of the nucleopeptides studied as precursors for the development as nucleic acid modulators.

Résumé

La thérapie génique vise à traiter des maladies en agissant au niveau moléculaire sur les acides nucléiques (ADN et ARN). Des oligonucléotides synthétiques peuvent reconnaître des séquences complémentaires et ainsi inhiber ou moduler leur expression *in vitro*. Pourtant, l'application d'oligonucléotides synthétiques *in vivo* n'est pas possible, étant donné leur dégradation rapide par les nucléases et leur faible capacité à pénétrer la membrane cellulaire. Pour remédier à ces problèmes, des analogues des nucléotides doués des meilleures propriétés biologiques, et gardant, ou même améliorant, l'affinité et la sélectivité caractéristiques des nucléotides naturels vis-à-vis des séquences complémentaires ont été proposés.

Une classe d'analogues récemment développée est constituée des nucléo-peptides, peptides qui ont dans leur séquence des nucléo-acides aminés, contenant sur les chaines latérales les bases azotées caractéristiques des acides nucléiques.

Le projet de recherche avait l'objectif d'étudier les propriétés des nucléo-peptides séquentiels ayant des alanyl nucléo-acides aminés (AlaB), en position *i,i*+3, en vue d'applications comme modulateurs d'acides nucléiques. En fait, si ces nucléo-peptides adoptent une structure hélicoïdale de type 3_{10}, les nucléobases sur les chaines latérales sont alignées et forment un côté fonctionnalisé de l'hélice, de manière à interagir d'une façon coopérative avec des séquences complémentaires.

Pour mieux étudier les effets de la stabilité conformationelle sur les propriétés de reconnaissance des nucléo-peptides, soit des nucléo-peptides rigides (basés sur l'acide aminé C^{α}-tetrasubstitué Aib, promoteur des structures hélicoïdales 3_{10}), soit des nucléo-peptides flexibles (basés sur l'acide aminé Ala, moins structurant), ont été synthétisés et leur propriétés conformationnelles et biologiques ont été étudiées.

Par rapport aux nucléo-peptides rigides, cette thèse présente:

(I) La synthèse et caractérisation de nucléo-peptides protégés contenants chacun une des quatre bases de l'ADN ;

(II) La synthèse et caractérisation d'un nucléo-hexapeptide protégé et d'un nucléo-heptapeptide complètement déprotégé renfermant deux thymines ;

(III) La structure cristalline d'un nucléo-tripeptide protégé basé sur la thymine ;

(IV) La caractérisation conformationnelle en solution des nucléo-tripeptides protégés, du nucléo-hexapeptide et du nucléo-heptapeptide.

Par rapport aux nucléo-peptides flexibles, cette thèse présente :

(I) La conception et la synthèse parallèle sur phase solide de librairies de nucléo-peptides modèles, pour l'optimisation du protocole synthétique et pour les analyses conformationnelles;

(II) La conception et la synthèse parallèle sur phase solide de librairies de nucléo-peptides partiellement biotinilés ou dérivatisés avec un fluorophore, renfermant chacune des quatre bases de l'ADN, pour des études de reconnaissance et biologiques;

(III) La caractérisation conformationnelle DC et RMN 2D d'un nucléo-heptapeptide modèle basé sur la thymine, et d'un analogue avec un résidu d'Aib au centre de sa séquence;

(IV) L'étude de la reconnaissance nucléo-peptide/nucléo-peptide et nucléo-peptide/ nucléotide mené par résonance plasmonique de surface, grâce à l'immobilisation de molécules biotinilées sur des puces d'or fonctionnalisées.

Par rapport à l'évaluation des propriétés biologiques des nucléo-peptides, cette thèse présente:

(I) Des études de cytotoxicité sur un nucléo-peptide rigide et sur plusieurs nucléo-peptides flexibles fonctionnalisés;

(II) Des essais de pénétration cellulaire suivis par cytométrie en flux et par microscopie, menés sur les nucléo-peptides biotinilés et fluorescents basés sur la thymine, supportés par des essais de co-localisation et par des mesures de pénétration en fonction du temps, de la concentration et de l'énergie cellulaire;

(III) Des mesures de stabilité en sérum murin menés sur les nucléo-heptapeptides flexible, mixte et rigide, renfermant des thymines.

Enfin, l'ensemble des données collectées a été évalué dans sa totalité afin de donner un jugement global sur les propriétés des nucléo-peptides étudiés en tant que précurseurs pour le développement de modulateurs d'acides nucléiques.

Riassunto

La terapia genica mira a curare determinate malattie operando a livello molecolare sugli acidi nucleici (DNA et RNA). Oligonucleotidi sintetici possono riconoscere selettivamente sequenze complementari ed inibire o modulare l'espressione genica *in vitro*. Non è tuttavia possible utilizzare oligonucleotidi sintetici per applicazioni *in vivo*, a causa di svariati gravi problemi, in particolare la loro rapida degradazione enzimatica e la grande difficoltà di penetrazione attraverso le membrane cellulari. Per superare tali limiti sono stati proposti analoghi di nucleotidi, che dovrebbero possedere migliori proprietà biologiche, mantenendo, o perfino migliorando, l'affinità e la selettività verso sequenze complementari caratteristiche delle strutture nucleotidiche naturali.

Una tra le classi di analoghi sviluppate di recente sono i nucleopeptidi, peptidi contenenti nucleoamminoacidi, ossia amminoacidi non proteogenici che portano una nucleobase in catena laterale.

Il presente lavoro di tesi aveva l'obiettivo di studiare le proprietà di nucleopeptidi sequenziali in conformazione elicoidale contenenti alanil-nucleoamminoacidi (AlaB) in posizione *i,i*+3 per applicazioni come modulatori di acidi nucleici. Infatti, se tali nucleopeptidi assumono una conformazione elicoidale 3_{10}, le nucleobasi in catena laterale si trovano allineate lungo l'elica e ciò potrebbe favorire interazioni cooperative con sequenze complementari.

Per meglio valutare l'importanza della strutturazione sulle proprietà di riconoscimento dei nucleopeptidi, si sono sintetizzati sia nucleopeptidi rigidi (basati sull'amminoacido C^{α}-tetrasostituito Aib, che favorisce l'assunzione di strutture elicoidali 3_{10}), sia analoghi flessibili (basati sull'amminoacido proteogenico Ala, meno strutturante), e ne sono state studiate sia le proprietà conformazionali che biologiche.

Riguardo ai nucleopeptidi rigidi sono descritte:

(I) La sintesi e la caratterizzazione di nucleopeptidi protetti contententi ciascuno una delle quattro nucleobasi del DNA;

(II) La sintesi e la caratterizzazione di un nucleoesapeptide protetto e di un nucleoeptapeptide idrosolubile completamente deprotetto, entrambi contenenti due timine;

(III) La struttura cristallina di un nucleotripeptide protetto contenente timina;

(IV) La caratterizzazione conformazionale in soluzione dei nucleotripeptidi protetti, del nucleoesapeptide e del nucleoeptapeptide.

Riguardo ai nucleopeptidi flessibili sono descritte:

(I) La progettazione e la sintesi in parallelo su fase solida di librerie di nucleopeptidi modello per l'ottimizzazione dei protocolli sintetici e per analisi conformazionali;

(II) La progettazione e la sintesi in parallelo su fase solida di librerie di nucleopeptidi contenenti ciascuna delle quattro nucleobasi del DNA, parzialmente biotinilati o derivatizzati con fluorofori, per studi di riconoscimento e biologici;

(III) La caratterizzazione conformazionale CD e 2D NMR di un nucleoeptapeptide modello contenente due timine e di un analogo con un residuo di Aib al centro della sequenza;

(IV) L'analisi per risonanza plasmonica di superficie delle proprietà di riconoscimento tra nucleopeptidi e tra nucleopeptidi e nucleotidi via immobilizzazione di molecole biotinilate su lastrine di oro funzionalizzate.

Riguardo alla valutazione delle proprietà biologiche dei nucleopeptidi, sono descritti:

(I) Test di citotossicità svolti su un nucleopeptide rigido e su vari nucleopeptidi flessibili basati sulla timina;

(II) Misure di citometria in flusso ed analisi di microscopia riguardo alle proprietà di penetrazione cellulare di nucleopeptidi fluorescenti o biotinilati basati sulla timina, supportate da prove di colocalizzazione e da un'analisi della dipendenza da tempo, concentrazione ed energia della penetrazione cellulare;

(III) Misure di stabilità in siero murino eseguite sugli eptapeptidi rigido, misto e flessibile contenenti timina.

Infine l'insieme dei dati raccolti è stato valutato nella sua totalità onde fornire un giudizio globale sulle proprietà dei nucleopeptidi studiati per lo sviluppo di modulatori di acidi nucleici.

ABBREVIATIONS

1D = monodimensional

2Cl-Z = 2-chloro-benzyloxycarbonyl

2D = bidimensional

4-MePhOH = *p*-chresol

Abs = absorbance

Ac = acetyl

Ac_2O = acetic anhydride

AcOEt = ethyl acetate

AcOH = acetic acid

A^H = 9-adeninyl

Aib = α-amino*iso*butyric acid

Ala = alanine

$AlaA^H$ = 3-9-adenyl-alanine

AlaB = β-alanyl nucleo amino acid

$AlaC^H$ = 3-9-cytosyl-alanine

$AlaG^{Cl}$ = 3-(9-2-amino-6-chloro-purinyl)-alanine

$AlaG^{OAt}$ = 3-(9-2-amino-6-oxy-*O*-(7-aza-benzotriazol-1-yl)-purinyl)-alanine

$AlaG^{OH}$ = 3-9-guanyl-alanine

AlaT = 3-1-thymyl-alanine

All = allyl

A^Z = 9-6-benzyloxycarbonyl-adenyl

B = generic nucleobase

BMB = 1,4-bis(maleimido)butane

Boc = *tert*-butyloxycarbonyl

BOP = (benzotriazol-1-yloxy) tris(dimethylamino) phosphonium hexafluorophosphate

n-BuOH = 1- butanol

Bt = biotinyl

Bz = benzoyl

Bzl = benzyl

c = concentration

C^{Boc} = 1-*N*^{4}-*tert*-butyloxycarbonyl-cytosyl
CD = circular dichroism
C^{H} = 1-cytosinyl
COSY = correlation spectroscopy
Cys = cysteine
C^{Z} = 1-*N*^{4}-benzyloxycarbonyl-cytosyl
DAPI = 4',6-diamidino-2-phenylindole
DCM = dichloromethane
DEA = *N*,*N*-diethylamine
DIEA = *N*,*N*-diisopropylethylamine
$DiOC_6$ = 3,3'-dihexyloxacarbocyanine iodide
DKP = 2,5-dioxopiperazine
DMAD = *N,N'*-dimethylazodicarboxylate
DMAP = 4-(dimethylamino)-pyridine
DMF = *N*,*N*-dimethyformamide
DMHD = *N,N'*-dimethylhydrazodicarboxylate
DMSO = dimethylsulphoxyde
DNA = deoxyribonucleic acid
EDC = *N*-ethyl-*N'*-(3-dimethylamino)propyl-carbodiimide
EDTA = ethylenediaminetetraacetic acid
EP = petroleum ether
ESI = electron-spray ionization
Et_2O = diethyl ether
EtOH = ethanol
FACS = fluorescence activated cell sorting
FBS = fetal bovine serum
FIT = fluorescein-5(6)amino-thiocarbonyl
FITC = fluorescein-5(6)*iso*thiocyanate
FM4-64 = *N*-(3-triethylammoniumpropyl)-4-(6-(4-(diethylamino) phenyl) hexatrienyl) pyridinium dibromide
Fmoc = fluorenylmethyloxycarbonyl
FT = fourier transform
G^{Cl} = 9-(2-amino-6-chloro)-purinyl

G^{OH}	= 9-guaninyl
H-A^{H}	= adenine
HATU	= *O*-(7-aza-benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate
HBTU	= *O*-(benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate
H-C^{H}	= cytosine
HEPES	= *N*-2-hydroxyethylpiperazine-*N′*-2-ethanesulfonic acid
H-G^{Cl}	= 2-amino-6-chloro-purine
H-G^{OH}	= guanine
HOAt	= 7-aza-1-hydroxy-benzotriazol
HOBt	= 1-hydroxy-benzotriazol
HOSu	= *N*-hydroxysuccinimide
HPLC	= high performance liquid chromatography
H-T	= thymine
Hz	= herz
iPrOH	= 2-propanol
IR	= infrared absorption
J	= scalar coupling constant
L	= optical path
Lys	= lysine
MALDI	= matrix assisted laser desorption ionization
MBHA	= 4-methylbenzhydrylamine
MeCN	= acetonitrile
MeOH	= methanol
min	= minute
Mp	= melting point
M_r	= molecular mass
MS	= mass spectrometry
MTS	= 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium
NMM	= *N*-methylmorpholine
NMR	= nuclear magnetic resonance

NOESY = nuclear Overhauser effect spectroscopy
OMe = methoxy
OEt = ethoxy
OSu = 1-oxysuccinimide
O^{t}Bu = tert-butoxy
Oxl = (4H)-oxazolidin-5-one or azlattone (oxazolone)
PBS = phosphate buffered saline
PhMe = toluene
PhOH = phenol
Pip = piperidine
Ppm = parts per million
py = pyridine
R_f = retention coefficient
RNA = ribonucleic acid
ROESY = rotational frame nuclear overhauser effect spectroscopy
RP = reverse phase
RPMI = Roswell Park Memorial Institute medium
rt = room temperature
RU = response unit
Ser = serine
Ser(Lactone) = serine-β-lactone
SPR = surface plasmon resonance
T = 1-thyminyl
T^{All} = 1-(3-allyl)-thymyl
TEA = *N*,*N*,*N*-triethylamine
TFA = trifluoracetic acid
Tfa = trifluoroacetyl
$(Tfa)_2O$ = trifluoroacetic anhydride
THF = tetrahydrofurane
TLC = thin layer chromatography
TMP = 2,4,6-trimethylpyridine
TMS = tetramethylsilane
TMSOTf = trimethylsilyltrifluoromethanesulphonate

TNBS = trinitrobenzensulfonic acid
TOCSY = total correlation spectroscopy
TOF = time of flight
t_r = retention time
U = uracyl
UV-Vis = ultraviolet-visible
XRD = x-ray diffraction
Z = benzyloxycarbonyl
α, $[\alpha]$ = optical rotation, specific optical rotation
δ = chemical shift
ε = molecular absorbance
Θ, Θ_T = ellipticity, molar ellipticity
λ = wavelength
ν = frequency

P.S. The chiral amino acids are in the configuration L if not otherwise specified.

1. INTRODUCTION

1.1 Nucleotides and analogues

1.1.1 Nucleic acid function

Nucleic acids (DNA and RNA) are nucleotide polymers, made of pentofuranose sugars (2-deoxyribose and ribose respectively) linked by phosphate bridges, with a nitrogenous base or *nucleobase* (adenine, cytosine, guanine, thymine or uracil, abbreviated A, C, G, T and U, respectively, Fig. 1.1) bound to the anomeric carbon of the sugar.

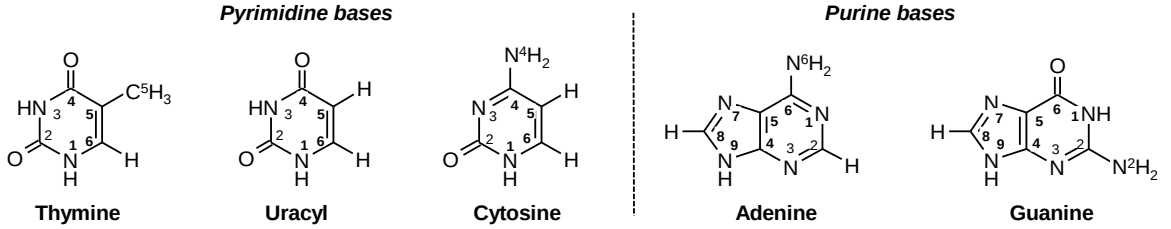


Fig. 1.1. Molecular st,ructure of the nucleobases with atom numbering. Sometimes the hexocyclic C^5H_3 group of thymine is referred to as $C(7)H_3$.

DNA has an antiparallel dimer structure (the well known *double helix*) with the two molecules (*strands*) held together by hydrogen bonds (*H-bonds*), selectively formed between a purine (A, G) and a pyrimidine (C, T) base. Watson-Crick, or canonical H-bonds, are two between adenine and thymine, three between guanine and cytosine; they are abbreviated A=T and G≡C, respectively. Watson-Crick pairs are called complementary nucleobases. Base stacking interactions between the π electron densities of adjacent bases also contribute to dimer stability. Indeed, the distance between the planes of adjacent bases in the most stable helical form (B-helix, see Fig. 1.2) is about 3.4 Å[1], which is the most favourable to π–π interactions.

It must be outlined, however, that nucleobases contain more H-bond donor and acceptor sites than the ones involved in Watson-Crick bonds, and that they can interact in different ways: the most stable interactions are those between complementary purine and pyrimidine bases (the most important are Hogsteen, reverse Watson-Crick and reverse Hogsteen interactions, see Fig. 1.3).

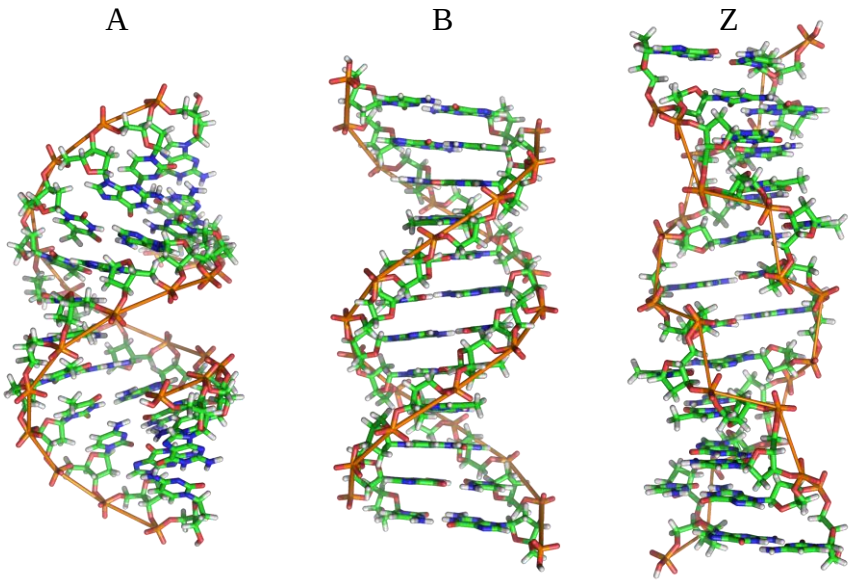


Fig. 1.2. Structures of DNA A, B, Z helices: atoms are represented with different colour: white (hydrogen), green (carbon), blue (nitrogen), red (oxygen), orange (phosphorous); the (imaginary) orange lines connect next phosphate groups. Source: Wikipedia.

Moreover, interactions with a different selectivity can be of considerable strength (by the so-called *wobbling* a guanine can form two hydrogen bonds also with adenine, thymine or with another guanine, see Fig. 1.3).[2]

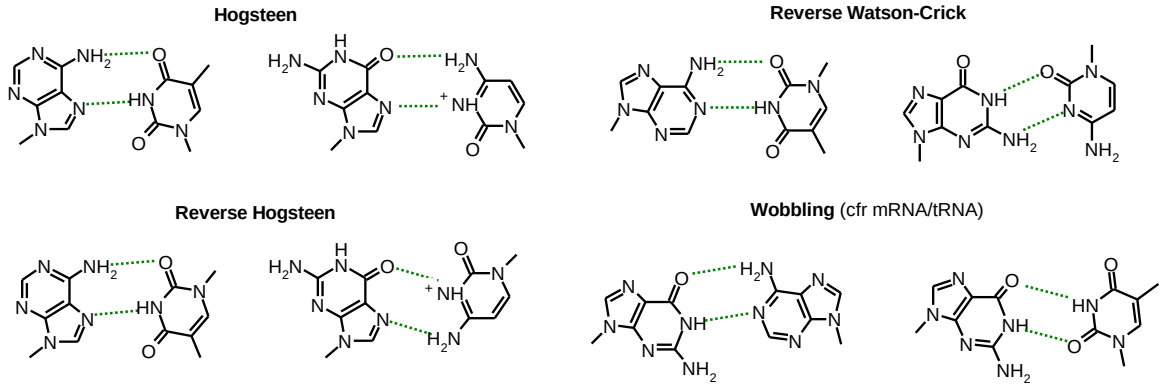


Fig. 1.3. Different non Watson-Crick nucleobase pairing modes: Hogsteen, reverse Hogsteen, reverse Watson-Crick and two examples of wobbling.

Selective A/T and G/C recognition and exclusive formation of Watson-Crick H-bonds in the DNA double helix are due to *template selectivity*. Indeed, the dimensions of the internal cavity of the helix and the relative orientation of each base pair, aligned by the two strands, stabilize preferentially the 'canonical' interactions. The importance of the double stranded template for base pairing mode selectivity is confirmed by the comparison with RNA,

which is single-stranded and enjoys a wider conformational freedom. In RNA molecules complementary sequences tend to stabilize by pairing with a much bigger variety of interactions,[3] even with selectivity different from Watson-Crick rules (for instance, wobbling allows a single amino acylated tRNA to recognize several mRNA triplets coding for the same amino acid, see below).[3b,c]

The remarkable stability of the DNA double helix and the selective pairing of the nucleobases of its two strands, enabling the univocal reconstruction of the dimer when the sequence of either strand is known, allow a reliable duplication. This in turn explains why this molecule is the repository of the genetic information,[4] in every living species on Earth, as a support for *biological information*.

Biological information includes the sequence of the RNA molecules required for cell metabolism (rRNA, tRNA, ...) and the sequence of the proteins needed for cell metabolism as catalysts, transporters, regulators... The sequence of amino acids of each protein is codified in portions of the DNA molecules called *genes*, through sequences of three nucleotides (called *triplets*). 64 different triplets exist, due to the 4^3= 64 possible nucleotide combinations, and each codes univocally for one of the 20 protein amino acids which make up proteins: this correspondence, which is almost the same for every living being, is called the *genetic code*.[5]

Protein synthesis is therefore the result of the expression of a given gene, which happens in two stages: first a gene on a DNA strand acts as a template for a complementary single-stranded RNA called messenger RNA or *mRNA* (*transcription*);[6a] then the mRNA molecule is processed in the ribosomes by binding sequentially the amino acids which it encodes, until the codified protein is synthesised (*translation*).[6b] The univocal individuation of the encoded amino acids relies on the recognition between the mRNA triplet for a given amino acid and special RNA molecules, called transfer RNA or *tRNA*, specifically bound to that amino acid, presenting a trinucleotide stretch complementary to mRNA triplet.

1.1.2 Synthetic nucleotide analogues for gene therapy

From what was described above, it clearly appears that alterations in the nucleotide sequence of the DNA double helix, such as changes in the nucleobase sequence (*mutations*) or loss of bases or nucleotides (*deletions*) can lead to the biosynthesis of proteins with an altered sequence, potentially with different tridimensional structure and

polarity and therefore with impaired functionality. Indeed, research done in the last decades has shown that the causes of a always growing number of pathologies reside in alterations of the genetic information or in defects of its expression.[7] This in turn raised a growing interest towards *gene therapy*, a therapeutic strategy operating at molecular level on genes with altered sequence or expression.

One of the first and more promising approaches of gene therapy is the control of gene expression through *RNA interference*. RNA interference (RNAi) is a natural mechanism of control of the gene expression which is found in all eukaryotes and even in bacteria:[8] if a mRNA molecule encounters another molecule (called *antisense*) able to stably pair with it (even partially), translation is prevented and the mRNA will be eventually degraded by special hydrolytic enzymes, so that the protein which it coded will not be synthesised. This means that the introduction in a cell of a synthetic antisense RNA strand complementary to the mRNA derived from a non functional or overexpressed gene could avoid the synthesis of a non functional protein or prevent a harmful protein overexpression.

Unfortunately the therapeutic use of synthetic oligonucleotides for RNA intereference therapy has several problems preventing its practical implementation.[9] The two most severe problems are:

(i) Synthetic RNAi oligonucleotides are identical in structure to biological molecules and are therefore rapidly degraded by *nucleases* (hydrolytic enzymes working on phosphodiester bonds), present in every cell;

(ii) The nucleotide large negative charge (one for each phosphate group, i.e. one for each residue) makes them so hydrophilic that they are not able to cross cell membranes, formed by hydrophobic lipid bilayers, preventing them from entering the cells.

To overcome such problems a wide range of studies has been carried on nucleotide analogues[10,11] with suitable properties for therapeutic applications, i.e.:

- Affinity to complementary strands comparable to the one of natural nucleotides;
- Pairing selectivity following Watson-Crick rules (A/T and G/C) and sensitivity to single-base mismatched sequences;
- Good chemical stability and a better biological stability in comparison to nucleotides;
- Acceptable capability of biological membrane penetration in order to enter the cells;
- Reduced aspecific toxicity.

Because nucleotides are made of three parts (phosphate bridge, pentofuranose sugar and nucleobase, Fig. 1.4) the analogues have often been developed by modification of one or two of the three portions at a time.[10b]

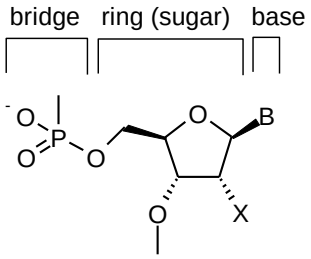


Fig. 1.4. The structure of a nucleotide (X= H for DNA, OH for RNA) and the three portions in which it can be divided.

Three groups of analogues can therefore be defined: (I) analogues with modifications involving only the phosphate bridge; (II) analogues with modifications involving the pentofuranose ring; (III) analogues featuring non natural nucleobases. Some of the most important kind of analogues of the first two groups are represented in Fig. 1.5.

Among the analogues obtained by bridge modifications, the most promising are the phosphorothioate nucleotides (PS-ODN): the sulphur atom reduces the polarity of the bridge group and makes enzymatic degradation slower. Vitravene^{TM} is the first antisense drug approved by the FDA[11a] (against cytomegalovirus induced retinitis in AIDS patients). Many other PS-ODNs are drug candidates in advanced clinical trials.[11b,c] On the other side, since a novel chiral centre is formed in each monomer and it is difficult to control its configuration during the synthesis, phosphorothioate oligonucleotides are often obtained as a mixture of isomers, with a strongly negative impact on their potential therapeutic applications. Moreover, heavy collateral effects of Vitravene and other PS-ODNs have been reported.[12]

Other analogues have been prepared by more profound bridge group modifications and often the negative charge is lost, in order to obtain neutral molecules which can more efficiently pass through cell membranes. On the other side, neutral oligomers, particularly when a significant size is reached, are often insoluble in aqueous solution and therefore not suitable for *in vivo* applications.

The most studied analogues modified at the pentofuranose ring are the morpholinonucleotides,[13] constituted by neutral monomers based on morpholine rings in which phosphate bridges are replaced by phosphorodiamide groups. Even if uncharged, such groups are very polar and thus favour the water solubility of the analogue.[13]

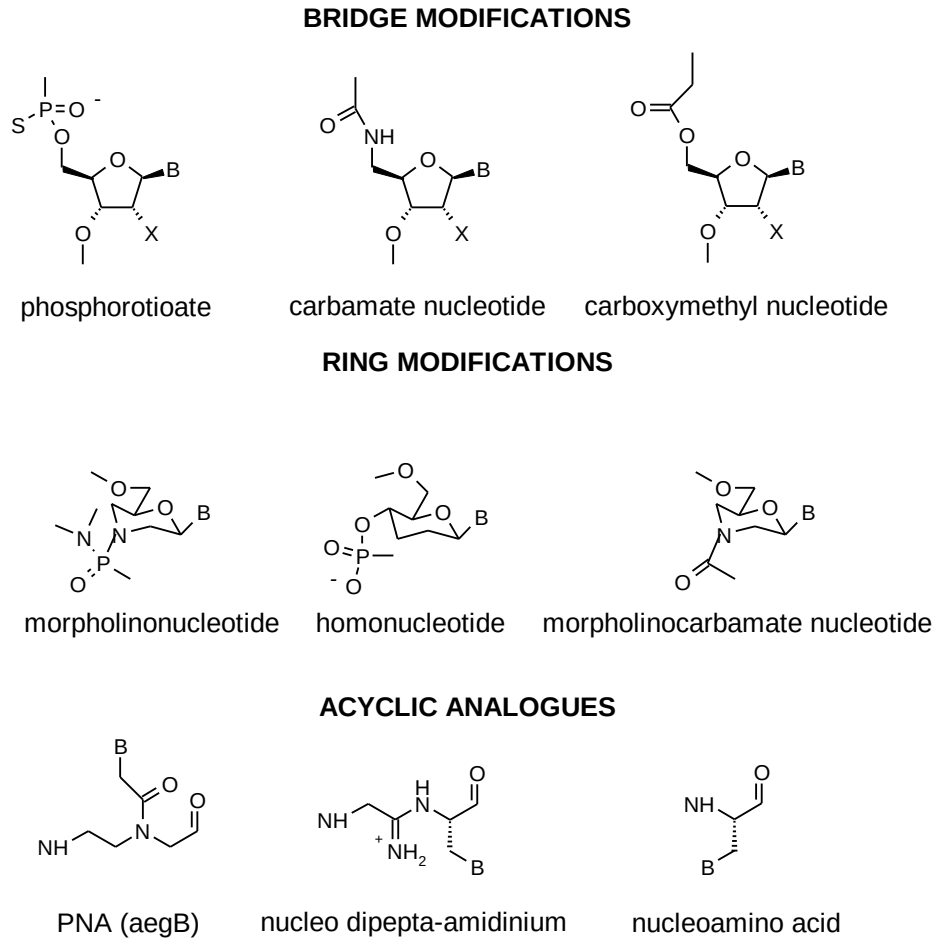


Fig. 1.5. Some of the most important oligonucleotide analogues: analogues presenting only phosphate bridge modifications (first line); cyclic analogues obtained by modification of the pentofuranose ring (second line); acyclic analogues (third line). The last group of analogues will be described in detail below.

1.1.3 Peptide nucleic acids

Since nucleotides are based on a cyclic pentofuranose structure, the vast majority of the analogues studied until the end of the '80s were also cyclic. However, at the beginning of the '90s the Danish group of P. Nielsen started investigating a novel, linear structure as a template for nucleotide analogues.[14] They hypothesized that the nucleotide chemical features required for a stable and selective pairing were a relative rigidity, as well as the number of bonds in the backbone and between nucleobase and backbone. As it can be seen in Fig. 1.6a, there are 6 bonds in the monomer backbone and 3 bonds between backbone and nucleobase. Therefore the Danish researchers designed and synthesised analogues based on *N*-aminoethyl-*N*-carboxymethylnucleobase-glycine (aegB), with the same

monomer backbone length and the same number of bonds between backbone and nucleobase (homomorphous structural replacement, see Fig.1.6b).

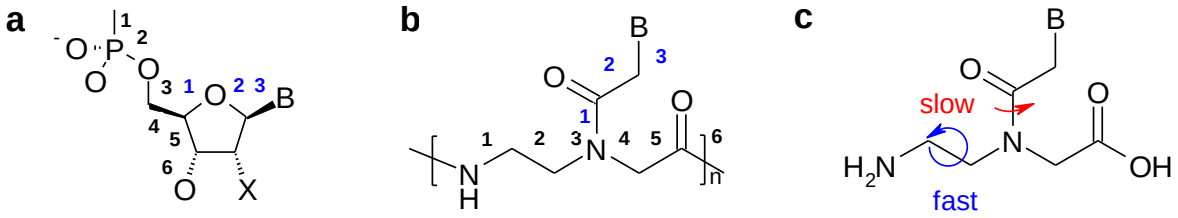


Fig. 1.6. Structure of a nucleotide (a) and of a PNA monomer (b) with number of bonds in the backbone (black) and between nucleobase and backbone (blue) highlighted. (c) Elements of rigidity (red) and flexibility (blue) in a PNA monomer.

Such structures were called *peptide nucleic acids* or *PNA* in analogy with the nucleic acids (DNA and RNA). The peptide amide in the backbone and even more the tertiary amide formed with the carboxymethyl nucleobase linker impart a certain rigidity to the backbone, which has however a flexible portion constituted by the two methylenes of the aminoethyl group (Fig. 1.6c).

The first results of their research were published in 1991[14] and confirmed their hypothesis, since stable and selective pairing was found. In fact PNA molecules are able to form antiparallel dimers with complementary sequences of single stranded PNA, RNA or DNA which are more stable than antiparallel dimers of complementary DNA strands (the affinity decreases along the sequence: PNA/PNA>PNA/RNA>PNA/DNA>DNA/DNA); this is due to the absence of charge of PNA, which avoids the electrostatic repulsion between complementary nucleotide strands. Although single-stranded PNA are generally unstructured, the structure of various PNA/DNA complexes both in solution and in the solid state has been studied.[15] The structure of the PNA/PNA dimer is called P helix,[16] in analogy to DNA A and B helices, whom it resembles: the diameter of the double helix is 28 Å, and there are 18 monomers per helical turn. The selectivity towards single base mismatched sequences is also remarkable.

PNA complementary to one stretch of double stranded DNA are even able to displace the DNA portion with the same sequence, forming a PNA/DNA heterodimer with the complementary strand (*double strand invasion* or *double strand displacement*).[17]

Double strand invasion is more efficient when homopurine or homopyrimidine DNA stretches are displaced, in particular when triplex structures PNA_2/DNA can be formed,[18] although efficient displacement of mixed purine/pyrimidine stretches is well

documented.[19] Double strand invasion is more efficient when homopurine or homopyrimidine DNA stretches are displaced, in particular when triplex structures PNA_2/DNA can be formed,[18] although efficient displacement of mixed purine/pyrimidine stretches is well documented.[19]

PNA synthesis is chemically easy and the application of automatic solid-phase peptide synthesis to their preparation is straightforward; at the beginning the Boc/Bzl strategy was employed[20] (see Fig. 1.7), while later different protection strategies[21] have been proposed. Moreover, PNA are chemically stable under neutral or acidic condition, and they are not degraded by hydrolytic enzymes *in vivo*.[17]

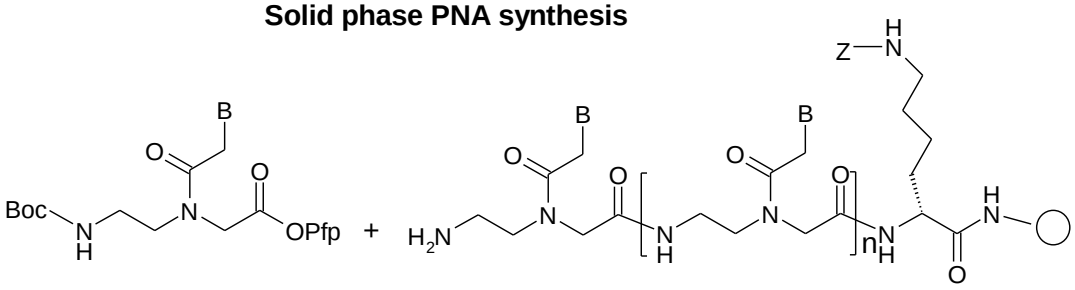


Fig. 1.7. Original solid-phase synthesis protocol for PNA (N-6 protection with Boc group, carboxyl activation as pentafluorophenyl ester, C-terminal D-lysine residue).

The discovery of this class of analogues raised therefore a strong interest in view of their pharmaceutical applications, which were the object of numerous studies[22,23] and further developments. Some of the most interesting recently developed structures are the chiral PNAs, originally proposed by the Italian group of R. Marchelli.[24] In such structures all or part of the 2-aminoethylglycine based monomers are replaced by a chiral, 2-aminoethyl-D-lysine unit (Fig. 1.8a). The α-substituents favour the structuration of the molecule and thus improve pairing selectivity. In particular, groups of three chiral monomer in the middle of a PNA sequence (*chiral boxes*) are able to achieve a pairing orientation selectivity higher than the one of conventional PNAs, without compromising the binding stability.[25] The solid-phase of chiral PNAs can be accomplished by a submonomeric approach to increase its efficiency.[26] Later development include monomers with a chiral centre in the γ position[27] (Fig. 1.8b) and with two chiral centres in the α and γ position.[28]

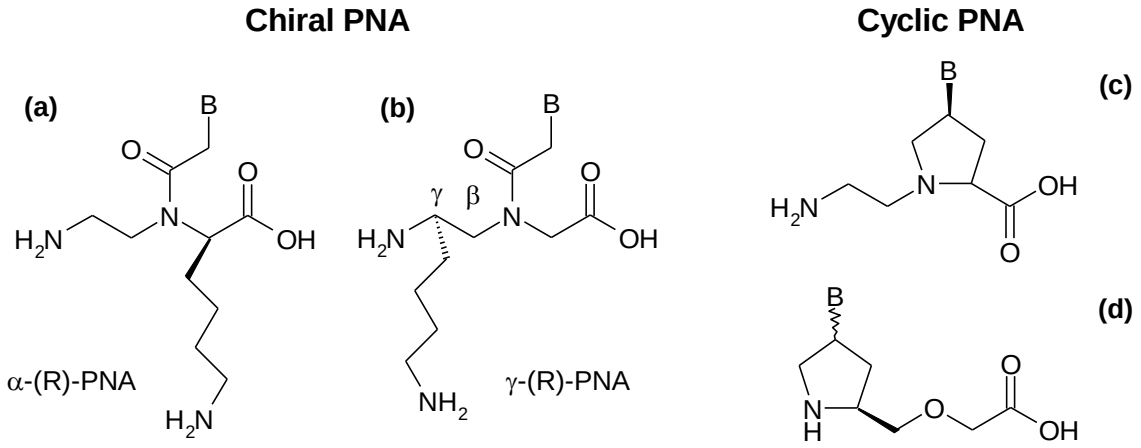


Fig. 1.8. Structures derived from the original Nielsen-type PNAs: chiral PNA with asymmetric centre in the α (a) or in the γ position (b), one of the cyclic PNAs derived from hydroxyproline (c) and oxygenated analogue or oxy-PNA (d).

Other PNA structures are cyclic[29] (Fig. 1.8c), generally based on five-membered, hydroxyproline-derived rings, whereas the higher rigidity has the same role of the side-chains in the linear chiral PNA, namely to improve binding selectivity. Cyclic PNA with an ether function imparting a higher water solubility (Fig. 1.8d) have been developed.[30]

Notwithstanding their interesting properties, PNA suffer from major drawbacks as gene therapy drug candidates, in particular:

- Reduced water solubility due to their neutrality (slightly improved by binding at the C-terminus a lysine residue, which has a positive charge in water);[18,20]
- Very low ability to cross biological membranes, with consequent poor cellular uptake, both if neutral and cationic;[18,31]
- High cytotoxicity, also as a result of their low water solubility.[18]

Even if some of these problems can be overcome by conjugating the PNAs to other water-soluble, cell-penetrating moieties,[32-34] generally polycationic peptides,[34] nowadays the most promising pharmaceutical applications of PNAs are in the diagnostics, with exclusive *in vitro* use.[18,33]

1.1.4 Nucleoamino acids and nucleopeptides

From the late ‘80s the Swiss group of A. Eschenmoser, working at the ETH of Zürich on *biochemical etiology* (the study of the reasons why biomolecules have given structures and properties), was focusing on the origin of the nucleotide structure, in particular looking for traces of a possible coevolution relation between peptides and nucleotides.[35] During such

researches, starting from L-serine, two non proteogenic alanyl amino acid derivatives were synthesised, with a nucleobase (thymine or adenine) at the β position, which were called *nucleoamino acids*.[36]

At the time, nucleoamino acids were not completely unknown. Indeed, willardiine, a nucleoamino acid carrying an uracil at the β positon, is a naturally occurring compound which was known since decades,[37] as well as a 2-aminopyrimidine-4-yl analogue called lathyrine.[38] The occurrence in nature of nucleobase containing peptides,[39] which had been called *nucleopeptides*, had raised a great interest so that research on the biological properties of the nucleoamino acids started,[40] even if without conclusive results. A thymine-based analogue of willardiine was thus synthesised as a racemate[41] by Strecker synthesis, partially resolved and polymerized,[42] however the complete resolution of its L enantiomer (with the same chirality as the natural compound) was not achieved. Similarly, ε-nucleoamino acids and their polymers were synthesised as racemates.[43] Even if interaction experiments on the resulting raceme polymers with nucleotides were attempted, no significant results were detected.[42,43] On the other side, the synthetic strategy of Eschenmoser relied on the nucleophilic ring opening of the protected L-serine-β-lactone,[44] a reaction which preserves the optical purity of the starting material.

The nucleoamino acids thus obtained were coupled with 2-aminoethylnitrile to form nucleo-dipeptanitriles, converted in nucleo-dipepta-amidines, which were oligomerized under acidic conditions to form oligo nucleo-dipepta-amidinium salts (see Fig. 1.9).[36,45]

As it can be noted in Fig. 1.9, this structure maintains the characteristic number of bonds in the monomer backbone, whereas there are only two bonds between nucleobases and backbone. The presence of a positive charge per residue in the salts of nucleo-dipeptaamidinium was a key point in the design of such analogues, as it should have imparted a higher affinity towards the polyanionic oligonucleotides. However, the oligomer salts are chemically unstable under aqueous non acidic conditions and even under acidic conditions (at pH≤ 4) they do not display significant structuration, nor any tendency to dimer formation with complementary DNA or RNA molecules.[45]

Although these rather disappointing results, soon after the work of Eschenmoser, the German group of U.Diederichsen carried on a wide research program on nucleoamino acids.

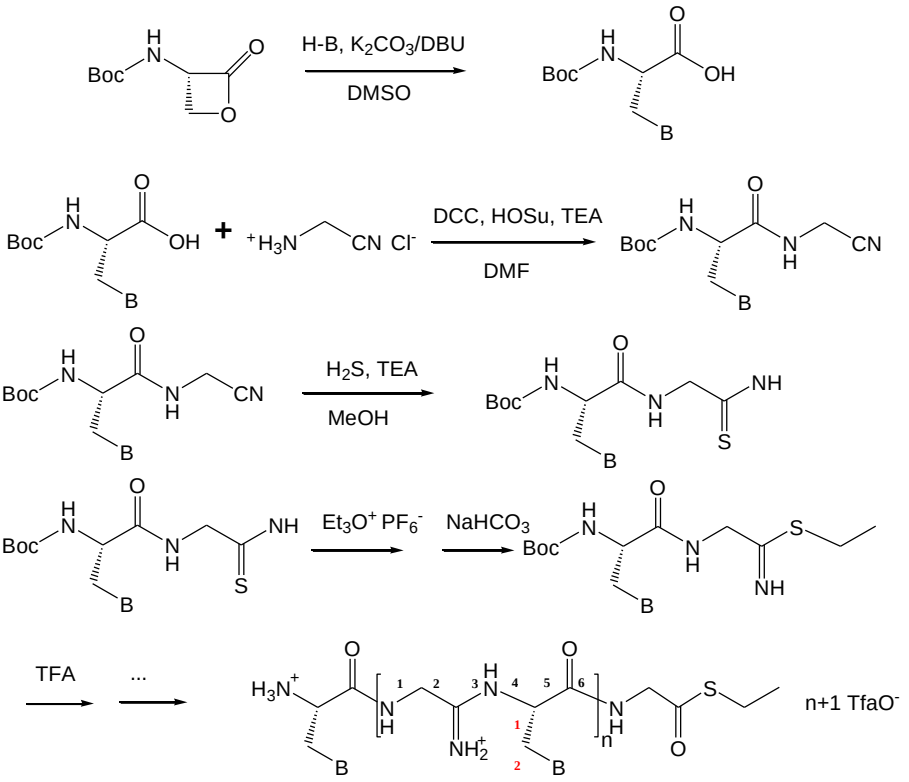


Fig. 1.9. Synthesis of oligo nucleo-dipepta-amidinium salts (B stands for 9-adeninyl and 1-thyminyl groups). Backbone and nucleobase-backbone bonds in the monomer are highlighted in black and red respectively.

Nucleoamino acids with all the four DNA nucleobases were synthesised[46] and used to form peptide oligomers called alanyl-PNA, or alanyl-nucleopeptides[47] by solid-phase peptide synthesis. Such structures, which are chemically stable, carry a nucleobase every third bond in the backbone, placed at two bonds from the backbone. For the synthesis of this kind of nucleopeptides, alanyl-nucleoamino acid residues with alternating configuration of the C^{α} were used, in order to obtain syndiotactic oligomers, which form β-sheet-like structures, in which all side-chain are protruding from the same side, as represented in Fig. 1.10. The distance between next nucleobases is about 3.6 Å, similar to that found in a DNA B-helix.[47]

When complementary functionalised nucleopeptides are mixed, they can form Watson-Crick pairs; however, different interactions and aggregation are observed when non complementary pairs or autocomplementary nucleopeptides are studied.[48,49] The stability of the dimers formed by complementary nucleopeptides is different if the dimer is homo- (e.g. if the two molecules have complementary residues with the same configuration) or hetero-chiral (e.g. if complementary residues of the two molecules have

opposite configurations). To have significant pairing stability, at least 6 complementary base pair are required.[48]

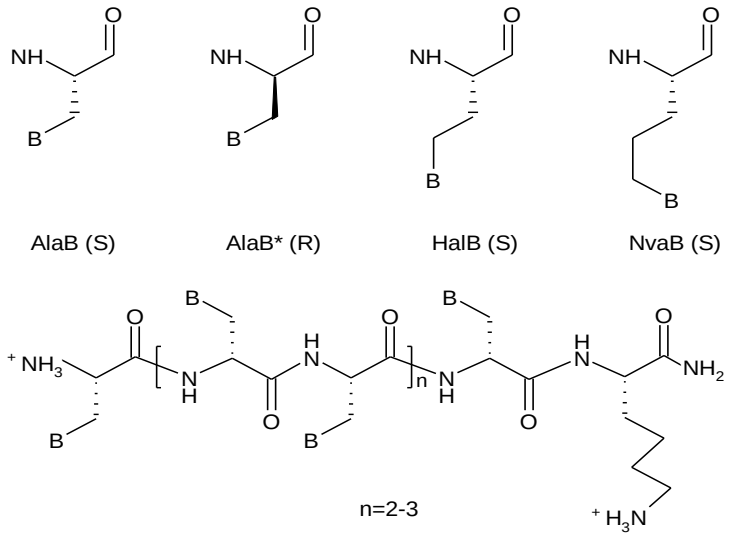


Fig. 1.10. Top: structure of monomers used for the synthesis of syndiotactic nucleopeptides (AlaB, β-alanyl nucleoamino acid; HalB, γ-homoalanyl nucleoamino acid; NvaB, δ-norvalyl nucleoamino acid); bottom: structure of a syndiotactic alanyl-nucleopeptide.

The effect of the nucleobase-backbone linker flexibility was studied by synthesizing nucleopeptides made of residues with different side-chain lengths compared to the alanyl nucleoamino acids[50,51], called homoalanyl- and norvalyl-nucleoamino acids (see Fig. 1.10, top); 'mixed' nucleopeptides made of residues with different chain length were also synthesised.[51] Homoalanyl-nucleoamino acids had been prepared decades before as a racemate,[52] as their inferior homologous, but Diederichsen synthesised them for the first time as as optically pure compounds. It was found that, among homo-nucleopeptides, the ones forming most stable dimers were the alanyl-nucleopeptides.[51]

β-nucleoamino acids (Fig. 1.11, top left) were synthesised and used to prepare β-homonucleopeptides.[53] These peptides fold in a β-sheet like structure and show higher dimer stability compared to α-nucleopeptides. To take advantage from the strong propensity of peptides made of β^3-amino acids, even with a reduced number of residues, to form stable helices,[54] sequential β-nucleopeptides were prepared.[55] They were composed by repeating tripeptide units (Fig. 1.11, middle) with a β-aminocyclohexanecarboxylic acid residue, a β-nucleoamino acid, and a β-homolysine (to enhance water-solubility).[55]

The presence of β-aminocyclohexanecarboxylic acid residue in the repeating tripeptide units was meant to favour the formation of a β-peptide 14-helix: in this structure,

residues every third position are aligned and therefore the side-chain nucleobases would be able to interact cooperatively with complementary sequences of other molecules in the same conformation. A detailed investigation[56] confirmed the assumption of a 14-helical comformation by the sequential oligomers. In general, multiple functionalization of a nucleotide analogue with the same base allows various kind of pairing interactions.[35b,47] It was found however that Watson-Crick pairing, when possible, was largely prevalent over non specific higher order interactions.

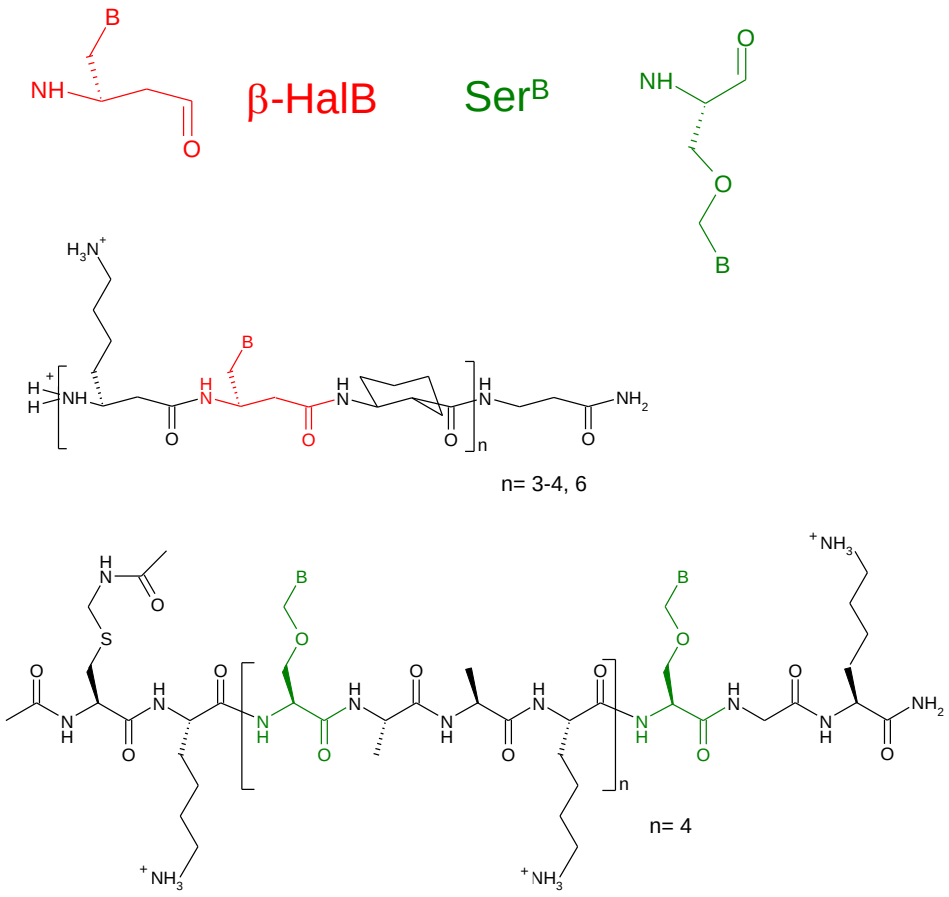


Fig. 1.11. Top: nucleoamino acids used to prepare sequential nucleopeptides. Middle: 14-helical β-nucleopeptide based on β^3-homoalanyl-nucleoamino acids at *i*,*i*+3 positions. Bottom: α-nucleopeptide based on β-basemethoxy-alanyl-α-nucleoamino acids at *i*,*i*+4 positions.

The American group of P. Garner synthesised sequential α-nucleopeptides containing nucleoamino acids based on a different structure: β-basemethoxy-alanyl nucleoamino acids[57] (*O-serinylmethyl-nucleobases*, Ser^B, Fig. 1.11, top right). Such nucleoamino acids have a longer nucleobase-backbone linker compared to nucleotides (4 bonds instead of 3), but they conserve the N-glycoside (O-C-N) substructure of nucleic acid monomers for the attachment of the nucleobases. The nucleopeptides (Fig. 1.11, bottom) carried the

nucleoamino acids at *i*,*i*+4 positions and a lysine was present in each tetrapeptide unit,[58] whose positive charge had the usual purposes both of increasing the water solubility of the nucleopeptides and favouring their interactions with nucleotides by charge complementarity.

The design of such sequential α-nucleopeptide structures with the specific aim of selective and stable nucleic acid recognition relied on two assumptions. Firstly, as eicosapeptides rich in α-helix promoting amino acids, it was hoped that the nucleopeptides would have folded in α-helices. The nucleoamino acids set every fourth position would have therefore formed a right-handed superhelix, with the same handedness of the DNA B-helix, thus favouring the the alignment of complementary nucleobases. Secondly, since the distance of next nucleoamino acids along the described superhelix is significantly larger than the distance of the planes of next nucleobases in double-stranded DNA, it was hoped that the flexible O-methyl-serine-derived linker would have allowed a synchronized tilting of the nucleobases in order to take the distance of their planes closer to the one of the nucleobases in DNA. Furthermore, Garner and his group were confident that the deformability of both α-helices and DNA[59] would have allowed an induced-fit driven by pairing stabilization.

By studying the CD spectra of their α-nucleopeptides, of complementary DNA stretches, and of their mixtures, they found that while the analogue alone was not structured, in the presence of one equivalent of complementary DNA, the CD spectrum was close to the superimposition of the contribution of a peptide α-helix[60] and of DNA in the B-form.[60a,61] As nucleopeptide/nucleic acid complex formation was confirmed by independent experiments performed with other techniques, the researchers interpreted this result as a pairing-induced structuration of their analogue.

The last two examples illustrate clearly the importance of preparing conformationally stable analogues, controlling and predicting their structures, in order to favour cooperative nucleobase-nucleobase interactions as a prerequisite for selective recognition, but also the effect that pairing forces can have on the conformation of an analogue and its stability. Therefore in the following section elements of peptide structure will be recalled and a class of amino acids known for their ability to adopt stable ordered conformations will be presented, with particular attention for its simplest member and for the helical structure its peptides generally adopt.

1.2 Peptide structures and C^{α}-tetrasubstituted α-amino acids

1.2.1. Elements of folded peptide secondary structures

Understanding the tridimensional structure of a protein or a peptide is fundamental to obtain detailed information on molecular recognition, and on peptide biological functions in general. Peptide structure can be described at different size scales and levels of complexity: the *primary* structure of a peptide is the sequence of amino acid residues along its backbone, the *secondary* structure is the description of the conformation of the peptide backbone, whereas the *tertiary* structure refers to its the global shape considering also the side-chains. The secondary structure is deeply influenced by the nature of the amino acids in the sequence and by the hydrogen bonding pattern between backbone amide protons and carbonyl oxygens. Its univocal description relies on the specification of the *dihedral*, or torsion, angles of the three backbone bonds of each residue (Fig.1.12), called ϕ, ψ, ω by the convention recommended by the IUPAC-IUB Commission on Biochemical Nomenclature.[62]

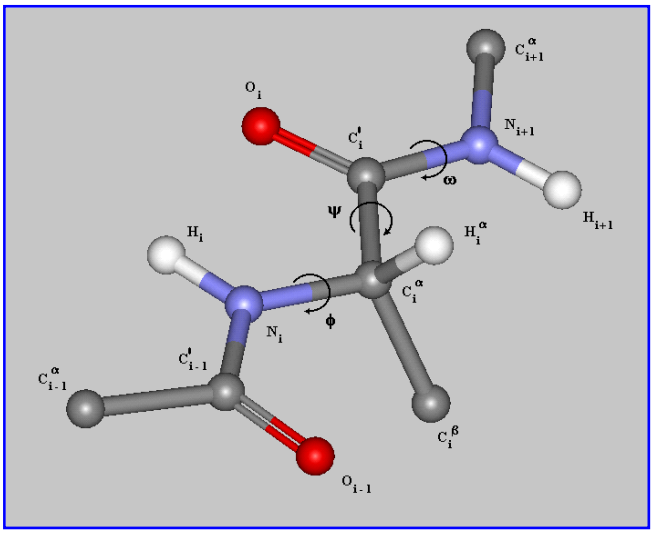


Fig. 1.12. Representation of a polypeptide chain (two peptide units). Recommended notations for the torsion angles are represented in the central residue. The chain is represented in the fully extended conformation ($\phi_i = \psi_i = \omega_i = 180°$) and the central residue is in L (S) configuration.

The most important and widespread peptide secondary structures are[63] the α-helix, the β-structures, the β-turns and the 3_{10}-helix; the most common organized secondary structures are helical. Various helical structures differ in the dihedral angles ϕ and ψ of each residue, in the number of residues per turn, in the pitch and in the number of atoms involved in the pseudocycles formed by intramolecular hydrogen bonds C=O····H-N.[64]

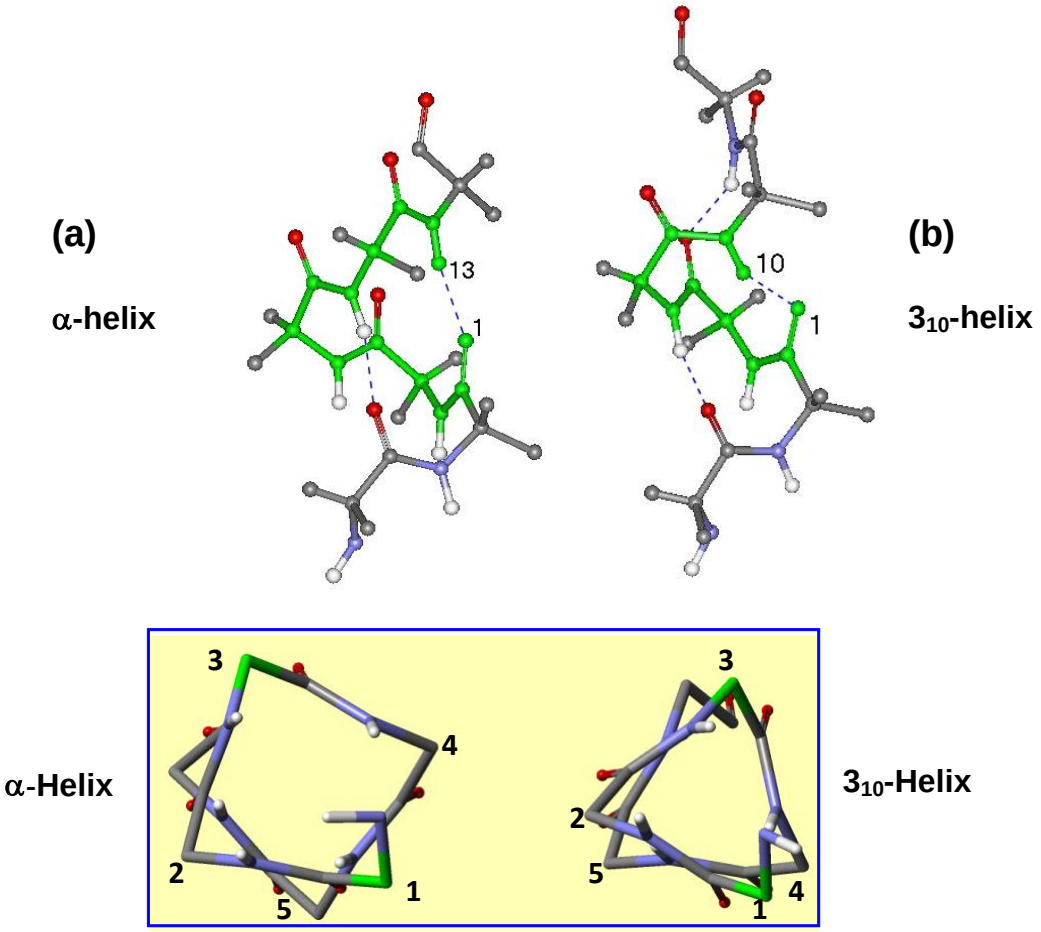


Fig.1.13. Top: side view of right-handed α-helix (a) and 3_{10}-helix (b); bottom: projections of the α-helix and of the 3_{10}-helix along the helical axis.

As mentioned above, the α-helix and the 3_{10}-helix are the most common helical structures. The α-helix (Fig. 1.13a) is characterized by 3.63 residues per turn and it is stabilized by intramolecular hydrogen bonds between the C=O group of residues in position *i* and the N-H group of residues in position *i*+4 (*i* ← *i*+4 H-bonds), thus forming 13-atoms pseudocycles (α-turns or C_{13} structures). The 3_{10}-helix (Fig. 1.13b) has 3.24 residues per turn and it is stabilized by intramolecular hydrogen bonds between the C=O group of residues in position *i* and the N-H groups of residues in position *i*+3 (*i* ← *i*+3 H-bonds),

thus forming 10-atoms pseudocycles (β-turns or C_{10} structures).[63] Hydrogen bonding patterns of both helices are compared in Fig.1.14.

Even if the difference in the torsion angle values of the two helices is not large, the 3_{10}-helix is significantly thinner (refer to Fig.1.13, bottom) and more elongated than the α-helix (the pitch is 6.3 Å for the first one and 5.7 Å for the second one). Characteristic structural parameters of the two helices are reported in Table 1.1.[65]

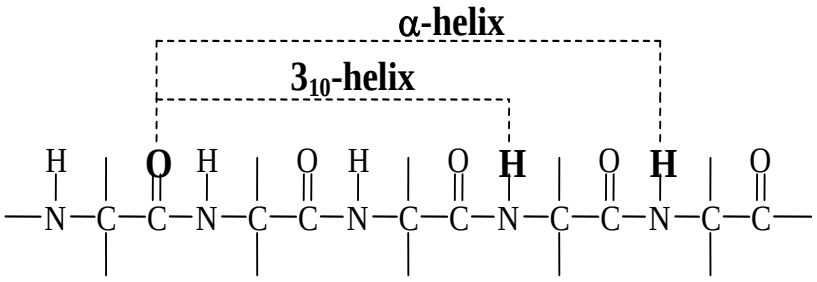


Fig. 1.14. Intramolecular hydrogen bonds in α- and 3_{10}-helices.

Parameter	**3_{10}-helix**	**α-helix**
ϕ	-57°	-63°
ψ	-30°	-42°
H-bonding angle N···O=C	128°	156°
Rotation per residue	111°	99°
Rise (axial translation per residue)	1.94 Å	1.56 Å
Residues per turn	3.24	3.63
Helical pitch	6.29 Å	5.67 Å

Table 1.1. Structural parameters for right-handed 3_{10}- and α-helices.

A 3_{10}-helix formed by C^{α}-trisubstituted (protein) α-amino acids is less stable than an α-helix, due to the larger distortion of the H-bonds and some unfavorable van der Waals interactions.[66-68] Although the 3_{10}-helical structure is thus less widespread than the α-counterpart, nevertheless it is not rare. The recent improvements in analytical techniques allowed to identify the 3_{10}-helical patterns in numerous natural proteins.[69-71] A statistical analysis of the X-ray diffraction structures of 57 globular proteins revealed the presence of

71 3_{10}-helical motifs of different length. Interestingly, in most cases such structures were found at the N- and C-termini of α-helices.

The main β-turn patterns were classified by Venkatachalam.[72] Six pairs of β-turn types (each one made of a right- and a left-handed turn) were described, depending on the values of the torsion angles of the 'internal' *i*+1 and *i*+2 residues. Table 1.2 gives ϕ, ψ torsion angle values for the right-handed β-turns with central peptide bond in trans conformation (types I, II, and III, Fig. 1.15).

β-Turn	ϕ (*i* + 1)	ψ (*i* + 1)	ϕ (*i* + 2)	ψ (*i* + 2)
Type I	-60°	-30°	-90°	0°
Type II	-60°	+120°	+80°	0°
Type III	-60°	-30°	-60°	-30°

Table 1.2. Dihedral angle values for the three most common types of right-handed β-turns. Left-handed β-turns have opposite dihedral angle values.

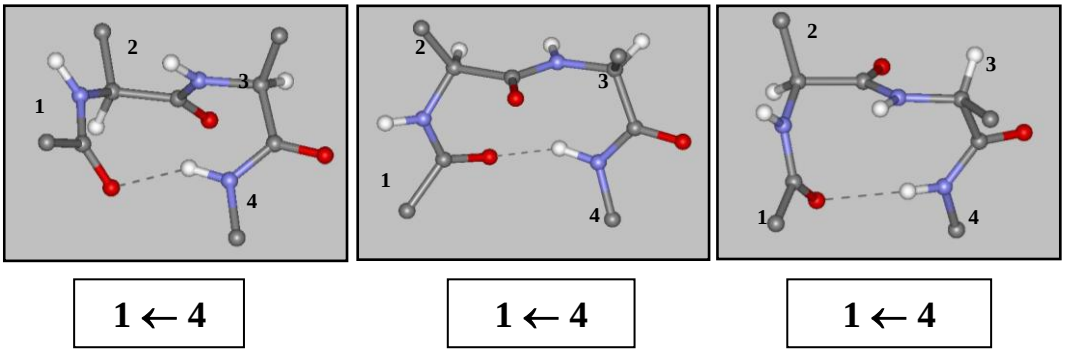


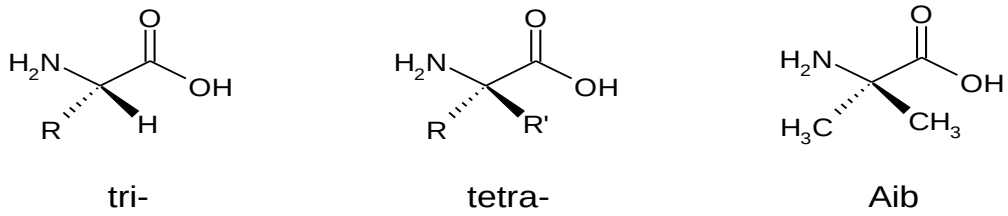
Fig. 1.15. Representation of the three ideal β-turns (I, II, III, from left to right) with *transoid* central peptide bond (ω_2= 180°).

Repeating type III β-turns lead to the formation of right-handed 3_{10}-helical structures, while type II β-turns are not helix forming and type I β-turns can be accommodated in helical structures, although with some distortions.

1.2.2 C^{α}-tetrasubstituted α-amino acids

C^{α}-*tetra*substituted α-amino acids differ from *tri*substituted α-amino acids (including protein amino acids apart from glycine) by the presence of a second aryl or alkyl

substituent at the C^{α} atom. The simplest member of this class is the α-amino*iso*butyric acid or Aib, which is achiral, since it bears two methyl groups as substituents (Fig.1.16).



tri- tetra- Aib

Fig. 1.16. A generic C^{α}-trisubstituted α-amino acid (in L configuration), a generic C^{α}-tetrasubstituted α-amino acid, and the simplest C^{α}-tetrasubstituted α-aminoacid, α-amino*iso*butyric acid (Aib).

The presence of a second substituent at the C^{α}-atom has far reaching consequences on the reactivity and the conformational properties of C^{α}-tetrasubstituted α-amino acids due to its additional steric hindrance. Firstly, C^{α}-tetrasubstituted α-amino acids are less reactive than their C^{α}-trisubstituted counterparts, and the decrease in reactivity is more pronounced for the amino than for the carboxyl group.[73] On the other side, racemisation is impossible, because of the lack of the C^{α}–H, therefore harsher reaction conditions can be used in order to compensate the lesser reactivity.

As regards the conformational properties, the increased sterical hindrance induced by the additional substituent at the C^{α} atom drastically limits the N-C^{α} and C^{α}-C' bond rotations (ϕ and ψ torsion angles, respectively),[74] thus favouring the assumption of stable structures in C^{α}-tetrasubstituted residue based peptides already in very short sequences (starting from three amide groups).[63]

It is possible to take advantage of the remarkable stability and rigidity of peptide sequences rich in C^{α}-tetrasubstituted residues in order to design and build molecular spacers[75a] and templates.[75b] It is worth reminding that a spacer is a linear system which allows the regulation of the distance between two labels in the molecule, while a template is a linear (or a cyclic) system which directs the two labels into a desired spatial separation and orientation (see Fig. 1.17). Peptide spacers bear the labels at the termini (A, B in Fig. 1.17), while peptide templates employ also, or only, the side-chains of some suitably functionalised residues in the sequence (D, E and F in Fig. 1.17).

The use of peptides as spacers and templates compared to other structures (e.g. steroids or synthetic polymers) offers several advantages. In particular, peptides can fold into ordered secondary structures stabilized by intramolecular hydrogen bonds and a rich

chemistry is available to allow site selective functionalization. Moreover, they allow a discrete variation of probe distance and a control of the relative probe orientation simply by varying the number of residues between them, or by changing the side-chain linkers employed. The use of C^{α}-tetrasubstituted residues for building templates or spacers is particularly convenient, since such peptides can form stable ordered structures already with a number of residues (about 6-8) which is much lower than the number of residues needed in the case of peptides made of C^{α}-trisubstituted residues (about 15-20).

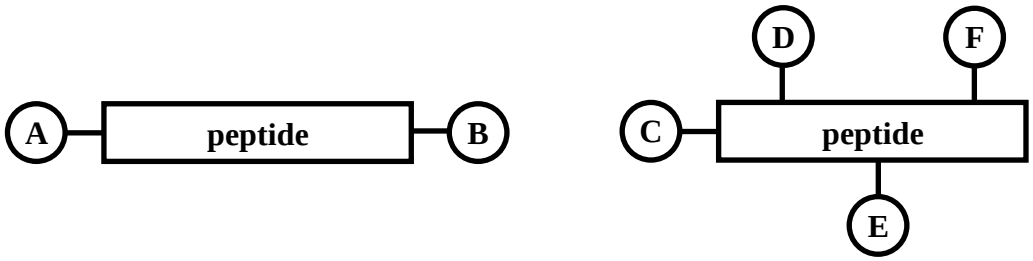


Fig. 1.17. Peptide spacer (left) and peptide template (right).

1.2.3 Conformational features of Aib-rich peptides

In the case of Aib, conformational energy calculations[76-77] highlighted that the presence of two methyl groups on the C^{α}-atom significantly restricts the conformational space accessible, which is essentially limited to the region of α- and 3_{10}-helical conformations. It is also worth recalling a recent theoretical study, from which it appears that Aib homopolymers would prefer the 3_{10}-helical structure,[78] since the α-helical structure would result very perturbed by unfavourable interchain interactions.

Because the Aib residue is achiral, right- and left-handed helices of its homopolymers are isoenergetic and the probability of each helical handedness is the same. In case together with Aib other chiral amino acid residues are present in the sequence, their chirality governs the sense of spiralization of the helix: protein L-amino acids favour right-handed helices, whereas D-amino acids favour left-handed helices.

In the peptides containing C^{α}-tetrasubstituted α-amino acids with linear, as well as with β- or δ-branched side-chains, the sense of spiralization follows the common rules of protein amino acids (L-residues induce a right-handed helix), while γ-branched side-chains tend to promote the opposite handedness. An exception is C^{α}-methyl isovaline (with one

methyl and one ethyl side-chain), which does not show any noticeable screw sense preference.

In the case of the study of the homo-octapeptide of L-*C^{α}-methyl valine*, or L-(αMe)Val, Ac-[L-(αMe)Val]$_8$-O^{t}Bu,[79] it was obtained the first CD spectrum of a 3_{10}-helix[80] (Fig. 1.18), which was close to that theoretically predicted by Woody *et al.*[81]

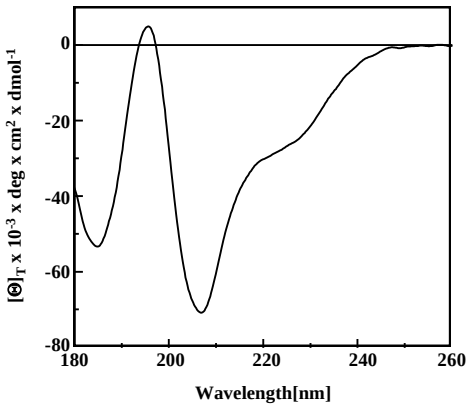


Fig. 1.18. CD spectrum of Ac-[L-(αMe)Val]$_8$-O^{t}Bu in 2,2,2-trifluoroethanol solution.

The extreme facility with which Aib containing peptides form single crystals allowed the X-ray diffraction analysis of the complete series of Aib homo-oligopeptides up to the undecamer.[29,82-92] The N-protected homo-tripeptides form a C_{10} structure; all longer peptides of the series invariably form the maximum number of consecutive C_{10} structures compatible with the length of the backbone, thus generating 3_{10}-helices. As an example the structure of the decapeptide *p*BrBz-(Aib)$_{10}$-O^{t}Bu[89,90] is represented in Fig. 1.19.

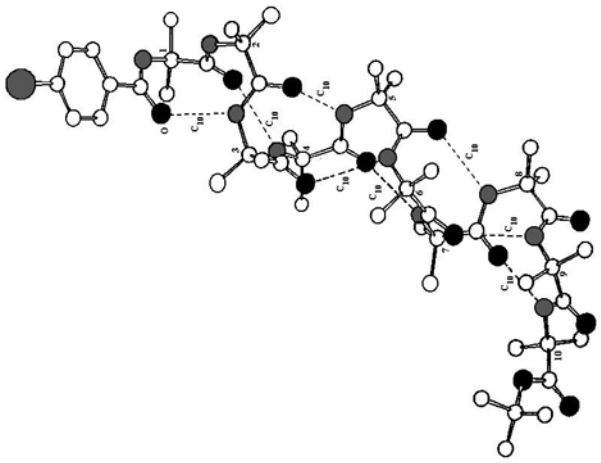


Fig.1.19. X-ray diffraction structure of *p*BrBz-(Aib)$_{10}$-O^{t}Bu, the second longest Aib homopeptide whose structure has been solved.

Conformational analyses in solution (by IR absorption and ^{1}H-NMR spectrometry) showed that this conformation prevails strongly also in solvents of reduced polarity as deuterated chloroform.[82,93]

In the case of peptides containing Aib residues and C^{α}-trisubstituted residues, in the solid state only helical structure are observed. Such helices can be of 3_{10} type, of α type, or 'mixed' (an α-helical segment preceded or followed by some C_{10} structures). More than 40 structures obtained by X-ray diffraction of peptides made of Aib and protein amino acids with length between 4 and 16 residues were reported in the literature until 1990. From their analysis it was found that chain length, Aib content and peptide sequence are among the most important factors directing the conformation towards one of the two helical structures.[94,95] In particular, the α-helix tends to be favoured as the chain length increases and as the content of Aib decreases, even if there are several exceptions and evaluation of the role played by peptide sequence is difficult (Fig. 1.20). On the other side, very short peptides (six residues or less) display an overwhelming preference for the 3_{10}-helix. In summary, the vast amount of structural data on Aib containing peptides highlights the remarkable capability of this non proteogenic residue of promoting and stabilizing β-turns as well as helical conformations, particularly 3_{10}-helices.

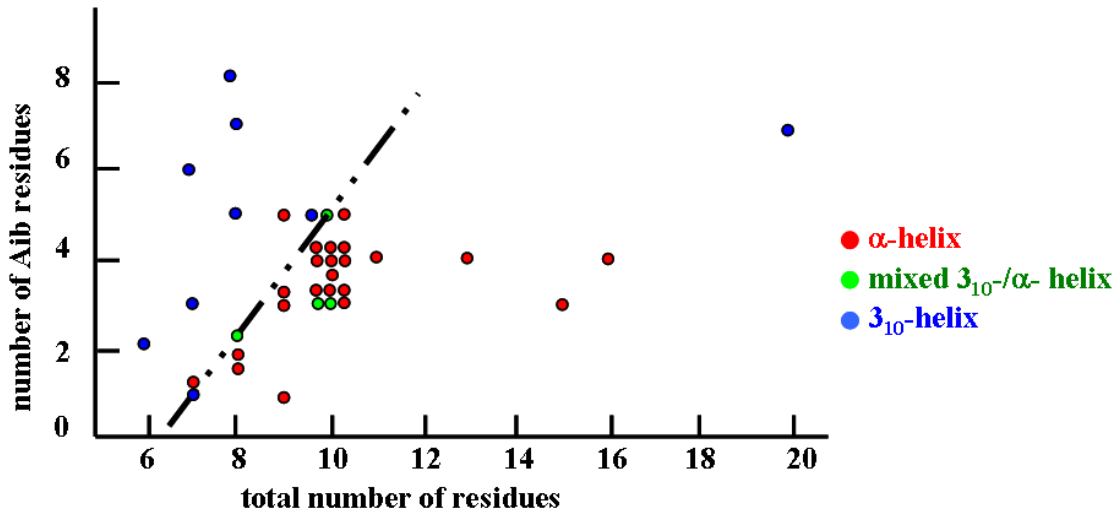


Fig. 1.20. Type of helical structure as a function of the total number of residues and of the number of Aib residues in the peptide sequence. The dashed line separing 3_{10}- from α- helices was traced arbitrarily.

1.3 Aim of the thesis

Bearing in mind the above mentioned data, it is reasonable to think that sequential peptides of limited length made of repeating tripeptide units with one C^{α}-trisubstituted residue between two C^{α}-tetrasubstituted α-amino acids, fold in a 3_{10}-helical conformation. In such helical pepides the side-chains of each third residue would be almost aligned. Therefore, if there is a functionalised residue every third position in the peptide sequence, a kind of *active side* of the helix would be formed. Two such peptides, derivatised with complementary chemical probes in the side-chains, could then recognize each other more easily, since the dimer formation would be favoured by cooperative effects.

Moreover, in the long-term perspective of designing biologically active peptides, a further advantage of employing C^{α}-tetrasubstituted α-amino acids is that peptides rich in C^{α}-tetrasubstituted residues are less easily degraded by proteolytic enzymes compared to peptides formed by protein amino acids.[96] A peptide composed completely of non natural C^{α}-trisubstituted amino acids and C^{α}-tetrasubstituted α-amino acids should have therefore a considerable stability *in vivo*.

The aim of this thesis has been to verify the hypothesis described above, in the specific case of sequential nucleopeptides as nucleic acid analogues. Therefore, the following targets were set:

- The synthesis of sequential rigid nucleopeptides made of repeating tripeptide units composed of an α-nucleoamino acid between two C^{α}-tetrasubstituted α-amino acids;
- Their conformational analysis, in order to verify the assumption of a folded conformation, likely of 3_{10}-helical type, which could favour cooperative interactions with complementary derivatised structures;
- The synthesis and conformational characterization of sequential flexible nucleopeptides, in which the C^{α}-tetrasubstituted residues would have been replaced by less structuring C^{α}-trisubstitued analogues, to evaluate the effect of backbone rigidity on the peptide structure;
- The analysis of the pairing properties of both classes of nucleopeptides in order to evaluate the effects of structuration, firstly towards complementary nucleopeptides, then in case of encouraging results, towards complementary nucleotides;

- The preliminary study of some relevant biological properties (cytotoxicity, cell penetration, *in vivo* stability) of both classes of nucleopeptides, in order to evaluate the possibility of *in vivo* applications for derivatives of nucleopeptides which should display interesting pairing properties towards oligonucleotides *in vitro*.

The Ph.D. thesis project has been developed in co-direction between the Department of Chemistry of the Università di Padova (Italy) and the "Laboratoire de Immunologie et Chimie Thérapeutiques" (UPR 9021, CNRS) of Strasbourg (France). In particular, most of the solution-phase synthesis of the nucleoamino acids and of the rigid nucleopeptides, as well as their conformational characterization, has been performed in Padova, whereas the solid-phase synthesis of the flexible nucleopeptides and the pairing and biological experiments have been performed in Strasbourg.

2. SYNTHESIS

2.1 Synthesis design

A working and efficient synthesis design is crucial for the success of any chemical synthesis. In this specific case, given the synthetic targets outlined in the preceding section, it was required:

- To identify the most convenient α-nucleoamino acid, C^{α}-tetrasubstitued α-amino acid and C^{α}-trisubstitued analogue for the preparation of rigid and flexible sequential helical nucleopeptides;
- To design synthetic targets both for obtaining model compounds, for conformational characterization and synthetic protocol evaluation, and for obtaining nucleotide analogues soluble in physiological conditions;
- To identify useful functionalisations for pairing and biological tests;
- To devise efficient synthetic strategies for the synthesis of both classes of nucleopeptides and for their functionalisation, in particular concerning to the choice of the solution or solid phase, of suitable protecting groups, and of the activation protocols.

Since the decisions taken about each point were conditioning the following steps, the discussion will address one point a time.

2.1.1 Choice of the amino acids

On the basis of the literature[51], β-alanyl α-nucleoamino acids (AlaB) were identified as the most convenient for interactions studies of complementary functionalized nucleopeptides. DNA nucleobases rather than RNA nucleobases were chosen as pendants, since uracil poses more synthetic problems than thymine.[97] Aib, or α-aminoisobutyric acid, the simplest C^{α}-tetrasubstituted α-amino acid, was chosen for the synthesis of the rigid nucleopeptides, both because of its strong 3_{10}-helical promoting properties and because the decrease in reactivity due to the sterical hindrance of the second methyl at the C^{α} is not so dramatic as in C^{α}-tetrasubstituted α-amino acids carrying bulkier substituents.[98] Consequently, the protein amino acid alanine was chosen as the C^{α}-trisubstituted analogue for the synthesis of the flexible nucleopeptides. The tripeptide units

composing the two classes of nucleopeptides were therefore Aib-AlaB-Aib and Ala-AlaB-Ala for the rigid and flexible nucleopeptides, respectively. To favour water solubility for biological and pairing measurements, the incorporation of one or two lysine residue at the termini was planned, as already seen several times in the case of peptide nucleic acids and nucleopeptides.

Concerning the amino-acid configuration, L-amino acids were employed, since they can form right-handed helices[99], with the correct spiralization sense in order to interact with nucleotides. As mentioned in the introduction, this holds true also for mixed peptides of the achiral Aib and of chiral C^{α}-trisubstituted amino acids.

2.1.2 Choice of the synthetic strategy

For the synthesis of the rigid nucleopeptides, rich in Aib residues, the solution phase synthesis seemed the most appropriate: indeed, the reduced reactivity of this hindered amino acids generally makes couplings on solid-phase sluggish. Beside that, solution phase synthesis offers the possibility of isolation, purification and characterization of all synthetic intermediates. This is valuable for the conformational studies, as well as the production of larger amounts of products by solution-phase synthesis techniques.

The sequential structure of the target nucleopeptides, terminated by an achiral residue, suggested the adoption of the *fragment condensation* approach for the synthesis, using the repeating tripeptide units as the fragments to be condensed . This approach generally allows to obtain larger quantities of final products, while speeding up the synthesis, and it can be used for parallel synthesis. This last advantage was particularly appealing, since, starting from four suitably protected nucleo-tripeptides containing each one of the DNA bases, all possible 4^2= 16 nucleo-hexapeptides or 4^3= 64 nucleo-nonapeptides could potentially be prepared.

For the synthesis of the flexible nucleopeptides, composed only of C^{α}-trisubstituted residues, the solid phase synthesis was the option of choice, since it allows the synthesis of longer sequences in short times. Because pairing and biological tests are generally performed on submicromolar scale, the reduced amount of product which can be obtained by solid-phase synthesis does not represent a problem, provided that efficient synthesis protocols are adopted.

2.1.3 Choice of suitable protecting groups

Generally speaking, the functions that require protection in peptide synthesis are the N-terminal amino and C-terminal carboxyl group, as well as functional groups at the side chains. The ideal conditions are satisfied when *orthogonal* protecting groups are employed:[100] two protecting groups are fully orthogonal if either group can be removed under conditions which do not compromise the stability of the other; partial orthogonality refers to the case that only one can be removed without causing removal of the other. Concerning the terminal functionalities, in case of the solid-phase synthesis of the flexible nucleopeptides, the C-terminal carboxyl group is bound to the resin until the cleavage at the end of the synthesis; therefore only partial orthogonality with the α-amino protecting group is required. On the contrary, the solution-phase synthesis of the rigid nucleopeptides by fragment condensation requires full orthogonality between N- and C-terminal protecting groups, while the synthesis of the single fragments can be accomplished also in the case of partial orthogonality.

Considering the side-chains of the target nucleopeptides, it appears that Aib and Ala are not an issue, whereas the nucleobase pendants of the nucleoamino acid and the ε-amino group of lysine (where present) must be taken into account. Indeed, three out of four bases (adenine, cytosine, guanine) require protection: cytosine for chemoselectivity reasons, the purine bases for solubility reasons. After their incorporation in the nucleoamino acids, nucleobase deprotection is not required until the very end of the synthesis, therefore partial orthogonality is sufficient for their protecting groups. The same is true for the ε-amino group of lysine if it is introduced simply for providing a positively charged ammonium moiety after complete deprotection; however, if this group is meant for specific derivatization, full orthogonality must be ensured.

2.1.3.1 Nucleobase protecting groups

Nucleobases are difficult to manipulate, being highly hydrophilic, highly insoluble in the most common organic solvents, and sparingly soluble and aggregation prone even in the most polar solvents, such as MeCN, DMF or DMSO.[97] The protecting groups chosen for the nucleobases, in particular for the purines, should therefore make them less hydrophilic and increase their solubility in polar organic solvents. This means that not all protecting

groups generally adopted in peptide synthesis can be successfully employed for the protection of nucleobases, either because of the different reaction conditions required for introduction or removal, or because of the insufficient increase in lipophilicity offered by the protection. Feasible options are thus significantly limited, meaning that also the choice of protecting groups that can be used for the peptide functions is limited by their compatibility with the few groups that can be efficiently used for the nucleobases.

For the protection of the N^6-amino group of adenine both the *tert*-butyloxycarbonyl (Boc) protecting group,[20,97] which is removed under acidic conditions, and the benzyloxycarbonyl (Z) protecting group,[58b,97] which is removed by catalytic hydrogenolysis, can be used (Fig. 2.1, top).

For the protection of the N^4-amino group of cytosine the same groups can be used[20,58b] (Fig. 2.1, middle). Moreover the use of the benzoyl (Bz) group has been reported.[101] The conditions for removal of the benzoyl moiety from α-amino groups are so harsh that it is generally considered a blocking, rather than protecting, group, but, in the case of the exocyclic amino groups of the nucleobases, this can be displaced under mild conditions by nucleophilic attack from hydrazine.[101b]

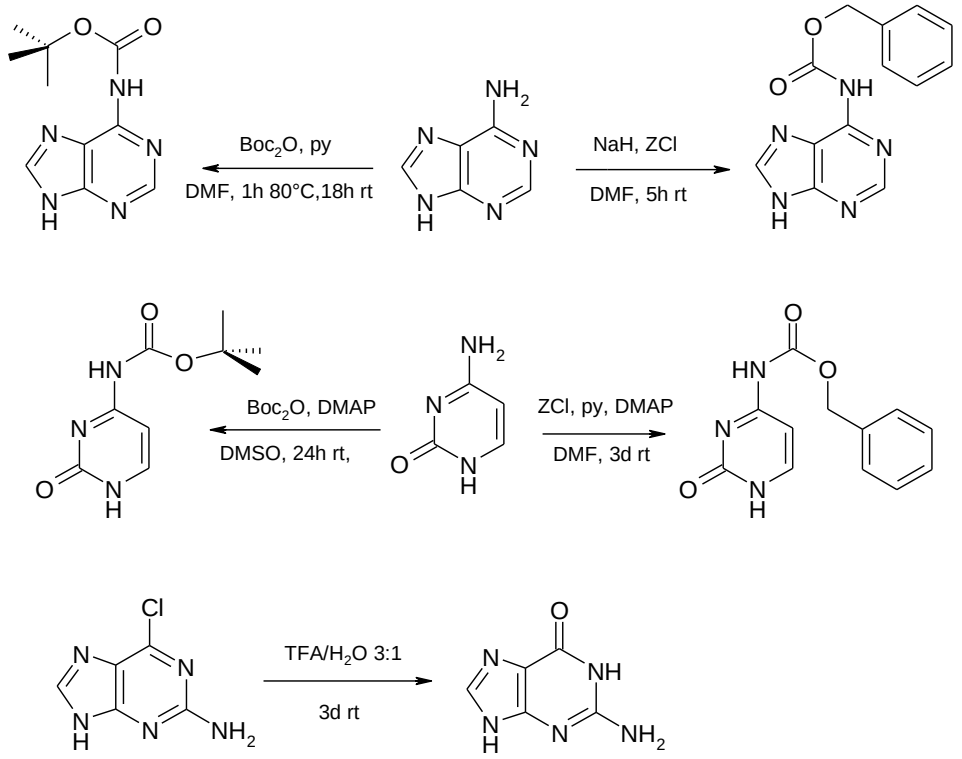


Fig. 2.1. Conditions of introduction of protecting groups employed in the present work for adenine (top) and for cytosine (middle); formation of guanine from 2-amino-6-chloro-purine (bottom).

Concerning guanine, both the O^6 protection as an ether,[102a] or as a urethane,[102b] and the N^2 protection as an amide[103] have been reported, but neither solution ensures both efficient and stable protection and removal under mild condition. In most cases dealing with nucleotide analogues,[54,104] guanine-based nucleotide analogues have been obtained starting from 2-amino-6-chloro-purine ($H\text{-}G^{Cl}$).[105] The chlorine atom favours the solubility in organic media compared to guanine and it can be replaced by an hydroxyl under strongly acidic aqueous conditions (Fig. 2.1, bottom).

2.1.3.2 N^{α}-amino- and carboxyl- protecting groups

The protecting groups chosen for the preparation of the flexible nucleopeptides had to be compatible with solid-phase synthesis conditions. The two most common strategies for the synthesis of peptides on solid-phase are the Boc/Bzl or the Fmoc/^{t}Bu.[106] In the first one, α-amino groups are protected with the acid-labile *tert*-butyloxycarbonyl (Boc)[107] protecting group, in the second one the base labile fluorenylmethyloxycarbonyl (Fmoc) protecting group is used.[108] Both Boc-protected[47] and Fmoc-protected[109] β-alanyl α-nucleoamino acids have been employed for the solid-phase synthesis of nucleopeptides. However, two more steps are required in order to obtain the (non commercially available) Fmoc-protected nucleoamino acids compared to the Boc-protected nucleoamino acids. This is because their synthesis relies on the nucleobase alkylation under strongly basic conditions (see section 2.2), which are not compatible with Fmoc-chemistry. Therefore nucleoamino acids must be prepared at first as Boc- or Z- protected derivatives, then deprotected and protected again with Fmoc. Consequently the Boc group has been chosen as α-amino protecting group for the flexible nucleopeptides, while protecting groups stable to moderate acidic conditions were needed for the side-chains (e.g. for the adenine and cytosine the Z protecting was employed).

Concerning the solution-phase synthesis of the rigid nucleopeptides, the situation was more complex, because of the need of an additional orthogonal protection for the carboxyl function. The most convenient urethane N^{α}-protecting group for solution-phase synthesis is by far the Z group.[110] The three main advantages of the use of this group are: (i) the good physical and chemical stability of Z-protected peptides, (ii) the ease of deprotection by catalytic hydrogenolysis, forming volatile co-products, which are smoothly

eliminated by evaporation and (iii) the increased tendency to precipitation and even to crystallization of pure protected peptides.

Using the Z group for α-amino protection, the C-terminal carboxyl group is generally orthogonally protected as *tert*-butyl ester (O^{t}Bu).[111] Among the advantages of this protecting group there are: (i) the increased solubility of the protected peptides in organic solvents; (ii) the reduced tendency to 2,5-dioxopiperazine formation of N^{α}-deprotected peptides; (iii) the ease of deprotection by acidolysis,[111c] with formation of volatile side products. If a third orthogonal protecting group is necessary, base-labile groups like Fmoc are generally used. The latter protecting group was not suited for the protection of the nucleobases, because it would have been removed by the basic alkylation conditions (see section 2.2). Of course simply exchanging the deprotection conditions for the N^{α}- and carboxyl- functions (by the use of the acid-labile Boc group and of the benzyl ester, OBzl,[112] removed by hydrogenolysis,[112c] respectively) would have not changed the situation.

One possibility was to use a base-labile ester, like the methyl ester (OMe[113]), for the carboxyl function, the only one not involved in the nucleobase alkylation step, thus allowing the use of Boc and Z groups together for N^{α}- and nucleobase protection. On the other side, care must be taken during the saponification of the methyl ester, since it can cause epimerization not only at the C-terminal but at other positions in the sequence. Moreover, the saponification of C^{α}-tetrasubstituted amino acid esters is particularly slow and it is generally performed only if the peptide contains exclusively achiral residues, since it requires harsher conditions compared to the saponification of protein amino acid methyl esters.

Therefore, considering that no ideal set of protecting groups valid for all four nucleobases was available, a tailored strategy was employed for the synthesis of nucleopeptides containing each base.

Since thymine does not require protection, the most convenient protecting groups (Z and O^{t}Bu) were adopted for the N^{α}- and C-terminal functions of its nucleopeptides. Adenine requires protection only for solubility reasons, since its exocyclic amino group is a quite poor nucleophile, therefore nucleobase deprotection during synthesis does not pose chemoselectivity problems. For this reason in a first time the same set of protecting groups was chosen for the α-amino and carboxyl functions of adenine-based nucleopeptides, while the Z protecting group was used for the nucleobase just to increase its solubility and

facilitate its manipulation up to the synthesis of the protected nucleo-dipeptide, then simultaneous deprotection of both Z groups was planned. Unfortunately, the use of N^{α}-Z protecting group for adenine-based nucleopeptides turned out to be not satisfactory (see section 2.3.2). Therefore, as completely orthogonal protecting groups the Boc and OMe moieties were employed for the backbone functions and the Z group was used for the nucleobase. Cytosine required an orthogonal set of protecting groups, its amino group being the most nucleophilic and basic (pKa 4.2)[97] among nucleobases, therefore the same set of orthogonal protecting groups employed in the case of adenine was used. Concerning guanine-based nucleopeptides, it was chosen to use 2-amino-6-chloro-purine as the starting material. As it was not clear whether the 6-chloro group would have resisted Boc or O^{t}Bu deprotection conditions (strong acid in aprotic solvent), the Z group was chosen for N^{α} protection and the carboxyl group was protected as OMe. In Fig. 2.2 nucleodipeptides containing the four bases are represented with the planned protecting groups.

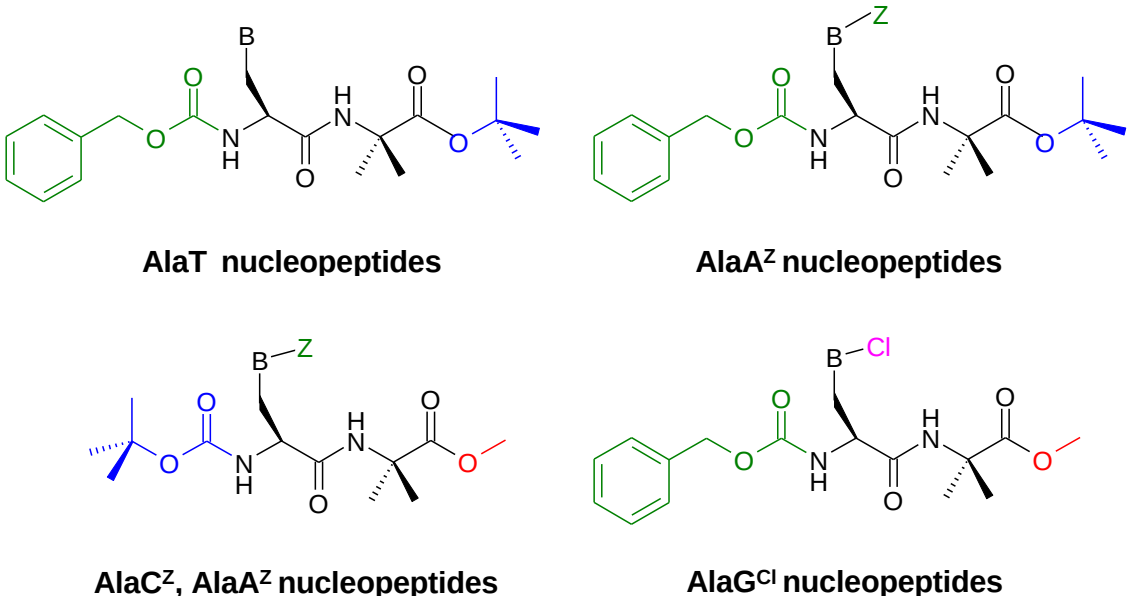


Fig. 2.2. The planned set of protecting groups for protected nucleodipeptides containing the four DNA nucleobases. Different colours are used for different deprotection conditions: hydrogenolysis (green), acidolysis (blue), saponification (red), acidic hydrolysis (purple).

2.2 Nucleoamino acid synthesis

The synthesis of the alanyl α-nucleoamino acids relies on the nucleophilic attack of the suitably protected nucleobases on the β-carbon of the N-protected serine β-lactone[44] in the presence of a strong, non-nucleophilic base. The lactone intermediate is not

commercially available in the N-protected form and therefore its synthesis deserves some comments as a prerequisite for the nucleoamino acid synthesis.

2.2.1 Synthesis of N-protected serine β-lactone

The synthesis of the β-lactone of serine was performed by Mitsunobu reaction.[114] Briefly, a phosphonium electrophile is generated by attack of triphenylphosphine on a *N,N'*-dialkyl azodicarboxylate and it binds to the carboxyl group of serine. This in turn activates the carboxylic carbon for the nucleophilic attack of the serine β-hydroxyl group, facilitated by the ability of the azo-derivative of accepting the hydroxyl proton and by the formation of the good leaving group phosphine oxide, see Fig. 2.3. The reaction is performed in tetrahydrofurane (THF) at low temperature (-78°C), in order to minimise side reactions. The use of the dimethyl azodicarboxylate (DMAD) instead of the commercially available diethyl or diisopropyl derivatives was suggested[44] in order to facilitate its chromatographic separation from the β-lactone product. However this adds another step to the preparation of the nucleoamino acids, since DMAD can not be stored longer than one or two days, even at low temperature.

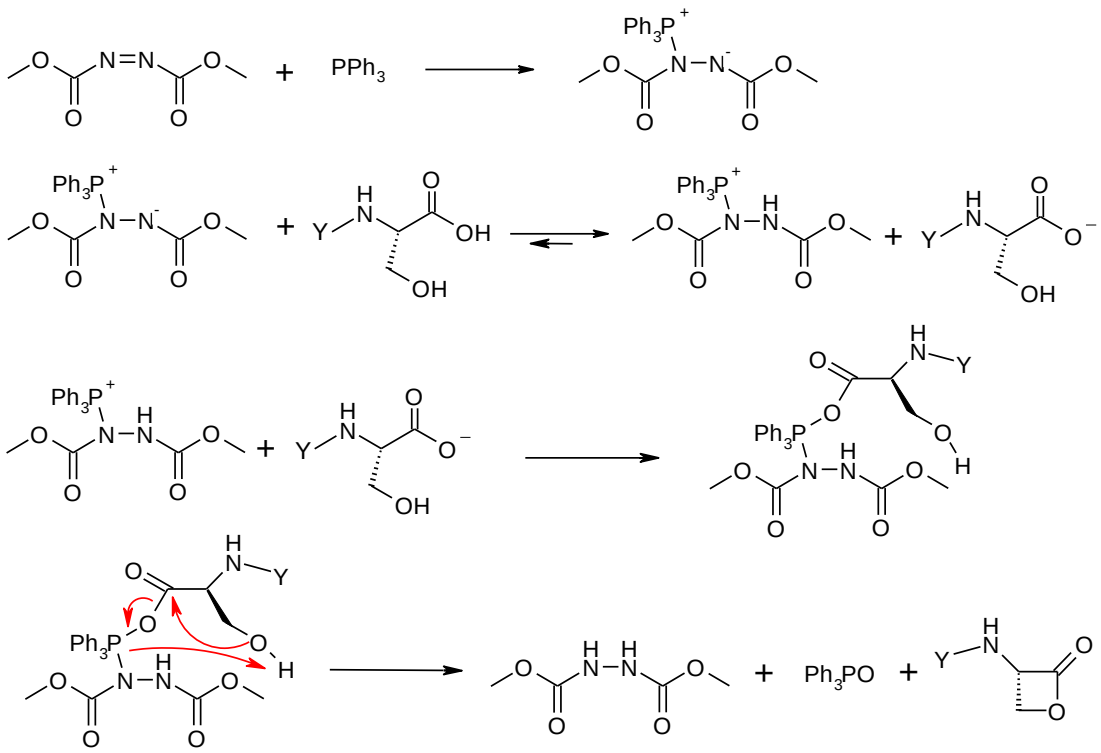


Fig. 2.3. Mechanism of formation of the N-protected serine β-lactone by Mitsunobu reaction (Y= Z, Boc). The concerted attacks in the key step are highlighted in red.

The steric tension in the four member ring directs the selectivity of nucleophilic attacks by soft nucleophiles towards the β-carbon instead of towards the carbonyl group, as expected for other esters,[44] thus allowing the following synthetic steps. This makes however the lactone extremely prone to decomposition. Indeed, the compound is not stable in the reaction mixture, as it can be attacked by unreacted phosphine or by the excess serine, and significant decomposition occurs even overnight at -20°C; therefore it must be immediately purified by flash-chromatography. Moreover, the compound is sensitive to nucleophilic impurities, particularly to moisture, so that the yield obtained in a humid summer was systematically lower than in winter. Even worse, the compound is extremely sensitive to radicals, even in traces, like the ones which are easily formed by THF or by remnants of the NBS used for DMAD formation. In this cases no product was obtained. Notwithstanding such serious problems, it was generally possible to reach yields close to literature values (55%).[44]

2.2.2 Nucleobase alkylation

Performing the nucleobase alkylation with the serine β-lactone in the presence of a strong, non-nucleophilic base exposes to the risk of loss of the α-proton of the lactone with consequent racemisation. The reaction conditions must be carefully studied in order to avoid or minimise this insidious side reaction.

The original paper of Eschenmoser[45] was not rich in details on reaction (nucleobase and non-nucleophilic base equivalents, addition order, etc.) and purification conditions, and the same details were missing in the relevant papers of Diederichsen.[47,48b,51] Optimization of the synthesis protocol was therefore necessary and thymine was chosen as the most convenient test nucleobase. However, after completing the optimization process, a different, detailed alkylation protocol was found in the literature,[109] covering all nucleobases except guanine, even if using sometimes different protecting groups. Such protocols used exactly one equivalent of non-nucleophilic base (NaH for pyrimidine nucleobases, or DBU for the purines) and required the slow dropwise addition of the serine β-lactone to a DMF solution containing a slight excess of the partially deprotonated nucleobase at -78°C. In order to compare the two methods, the alkylation was performed on a preparative scale with both and two series of protected nucleopeptides up to the hexamer were synthesised (see Section 2.3). By analyzing the

optical rotation of the nucleopeptides obtained with the two methods, it was found that the second method was more efficient in preserving the optical purity of the starting material. Therefore the latter method was adopted for the alkylation of protected cytosine and adenine and adapted to the alkylation of 2-amino-6-chloro-purine.

After the alkylation the product must be purified from the excess nucleobase and from unwanted alkylation isomers formed in the case of thymine and of the purines (see Fig.2.5).

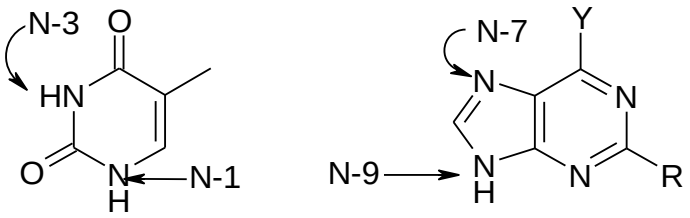


Fig. 2.5. Possible alkylation sites for thymine (left) and for the purines (right). For adenine R= H, Y=NH_2; for guanine R= NH_2, Y= OH; for 2-amino-6-chloro-purine R=NH_2, Y= Cl. Straight arrows point to target alkylation sites, while curved arrows point to undesired alkylation sites.

It results from the literature[47,115] that high regioselectivity can be obtained for thymine, whereas in the case of the unprotected purines the ratio of the N(9)-alkylated product to the N(7)-alkylated regioisomer does not exceed 3:1. In the case of adenine, protection does not change significantly the ratio; in the case of guanine, protection can even worsen the situation. Luckily enough, the use of 2-amino-6-chloro-purine allows a good selectivity towards the N(9)-alkylated product.[51,104]

Separation of most of the unreacted nucleobases from the nucleoamino acid products can be achieved either when the latter are extracted in organic phase from acidic aqueous solutions, or by careful precipitation from suitable solvent mixtures. This operation does not allow a complete elimination of the impurities, since the nucleoamino acids tend to coprecipitate, particularly when they are concentrated. Only in the case of the 2-amino-6-chloro-purine based nucleoamino acids (Boc-AlaG^{Cl}-OH and Z-AlaG^{Cl}-OH) much of the product could be efficiently purified by precipitation and recrystallization.

Chromatographic separation is therefore necessary both for the complete elimination of the excess nucleobases and for the separation of the unwanted regioisomers. The latter are particularly difficult to separate, because of their similar chromatographic behaviour compared to the products. The process is further complicated by solubility issues, and by the strong tailing in the column of both product and impurities, even in

highly polar eluents containing a significant amount of acetic acid. Generally, repeated flash-chromatographic purifications are necessary in order to obtain a satisfactory amount of acceptably pure product. Despite of the problems described, the synthesis and purification of the four Boc-protected nucleoamino acids and of the Z-protected thymine-, adenine- and 2-amino-6-chloro-purine-based nucleoamino acids was achieved. Fig. 2.6 schematically presents the products obtained and the alkylation conditions, while Graph 2.1 gives the yields for the Boc- and Z-protected derivatives containing the four nucleobases.

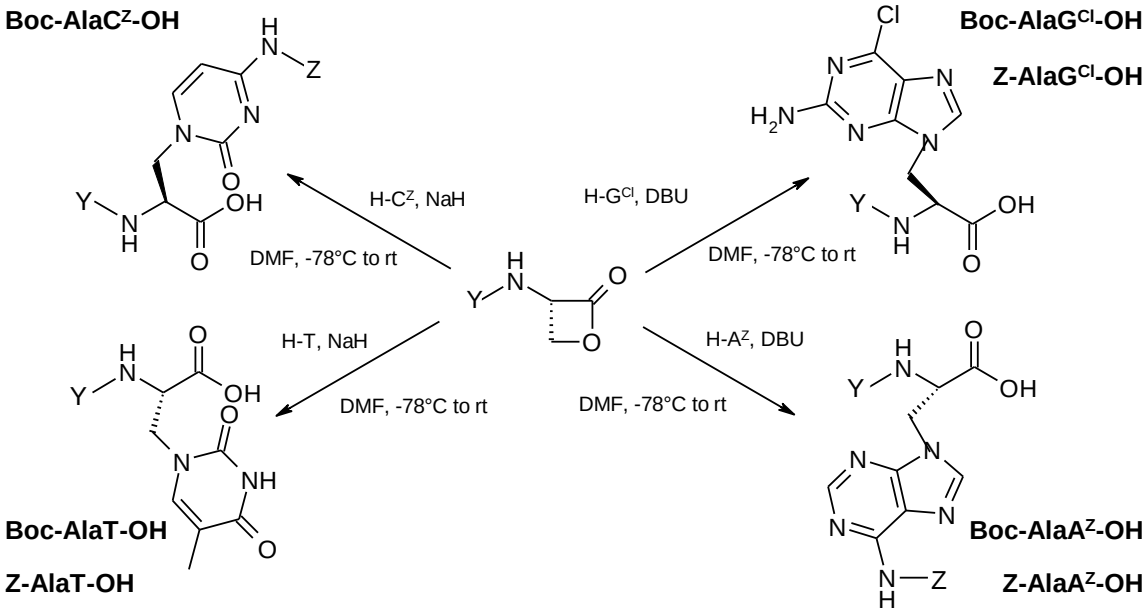
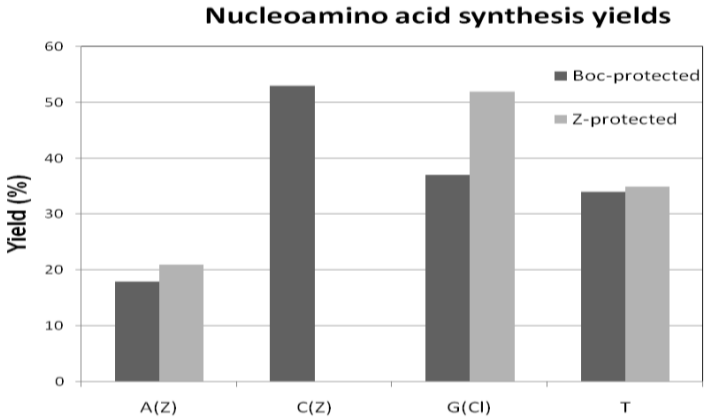


Fig. 2.6. Conditions for the synthesis of the N^{α}-protected nucleoamino acids (Y= Boc or Z).



Graph. 2.1. Yields of nucleoamino acid synthesis as a function of the nucleobase and of the N^{α}-protecting group adopted. The synthesis of the N^{α}-Z-protected cytosine-based nucleoamino acid was not performed.

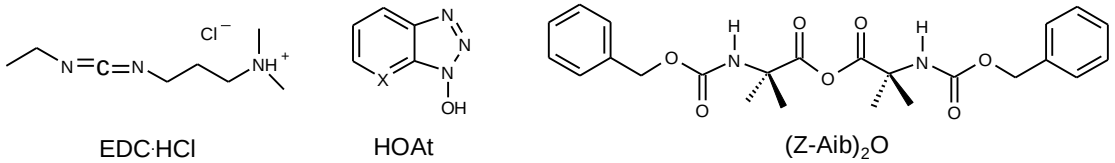
As it can be seen from the graphic, good yields were achieved with Boc-AlaC^{Z}-OH and Z-AlaG^{Cl}-OH, moderate yields were obtained with the thymine-based nucleoamino acids and with Boc-AlaG^{Cl}-OH, and poor yields were obtained with the adenine-based nucleoamino acids. In general there is little difference between Z- and Boc- protected nucleoamino acids, except for the two 2-amino-6-chloro-purine derivatives. In this case the very low solubility of the nucleobase-containing compounds in organic solvents cooperates with the general stronger ease of precipitation of the Z-protected derivatives as pure compounds compared to Boc-protected ones to allow a more efficient purification process.

2.3 Synthesis of rigid nucleopeptides

After the synthesis and purification of the nucleoamino acids, the building blocks for the solution-phase synthesis of the rigid nucleopeptides were available. The synthetic strategy chosen was to form nucleo-tripeptides by *step-by-step synthesis*, attaching one residue at time, before condensing suitably deprotected tripeptide units to obtain longer nucleopeptides. In order to perform the synthesis of the nucleo-tripeptides, each containing two hindered Aib residues, efficient methods of amino acid activation for the formation of the peptide bond (*coupling*) were required.

2.3.1 Coupling methods

Three decades of research on C$^{\alpha}$-tetrasubstituted amino acid containing peptides have provided the laboratory of Padova with a strong know-how about the activation of hindered, C$^{\alpha}$-tetrasubstituted residues.[73] The two most common and powerful activation methods are the *in situ* formation of active esters (*via* EDC/HOAt[116], Fig. 2.7, left and middle) and the use of preformed symmetric anhydrides (Fig. 2.7, right).



EDC·HCl HOAt $(Z\text{-Aib})_2O$

Fig. 2.7. *Coupling reagents* (EDC·HCl, left; HOAt, X=N, HOBt, X=CH, middle) for the formation of active esters and symmetric anhydride derived from Z-Aib-OH (right).

Active esters are easily formed, the use of HOAt efficiently suppresses racemisation due to carbodiimide overactivation,[116] and the side products deriving from the coupling reagents can be eliminated by acidic or basic washing. However, the unreacted active ester is not completely hydrolyzed or eliminated by washing and often chromatographic purification is necessary.

On the contrary, couplings *via* symmetric anhydride form a smaller number of side products, resulting in a simpler work-up, but they require more steps (anhydride synthesis and purification, coupling, sometimes N-protected amino acid recovery). Indeed, 2 equivalent of N-protected amino acid are required in order to generate 1 equivalent of acylating agent, and 1 equivalent of amino acid salt is formed after the coupling, thus making the amino acid recovery necessary when working with large amounts or valuable compounds.

The active ester coupling method was therefore always used for the activation of the nucleoamino acids, while the activation as symmetric anhydride was often employed for Aib activation. Large scale couplings involving Aib activation and couplings between protein amino acids were performed *via* active esters, using the less expensive HOBt[117] (Fig. 2.7, middle) instead of HOAt.

2.3.2 Synthesis of protected nucleo-tripeptides

The step-by-step synthesis of the thymine-based protected nucleo-tripeptide Z-Aib-AlaT-Aib-O^{t}Bu was smoothly achieved by coupling the activated nucleoamino acid with H-Aib-O^{t}Bu to form the protected nucleo-dipeptide Z-AlaT-Aib-O^{t}Bu, which was then hydrogenated and acylated with the symmetric anhydride $(Z\text{-Aib})_2$-O, as shown in Scheme 2.1.

The same strategy was not efficient in the case of the adenine-based nucleopeptides. Indeed, during the step of dipeptide deprotection, while the N^6-Z protecting group was rapidly removed by catalytic hydrogenation, the removal of the N^{α}-Z group was extremely slow, so that the reaction was not complete even after 7 hours (Fig. 2.8), and the coupling yield was negatively affected. Indeed slow hydrogenolysis of protecting groups next to the two most aromatic bases (adenine and cytosine) have been reported,[109] probably due to stacking interactions between the nucleobase and the Z phenyl ring hindering the removal of the latter.

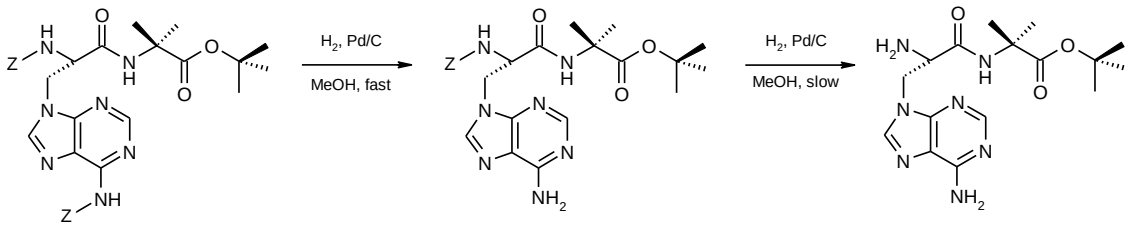
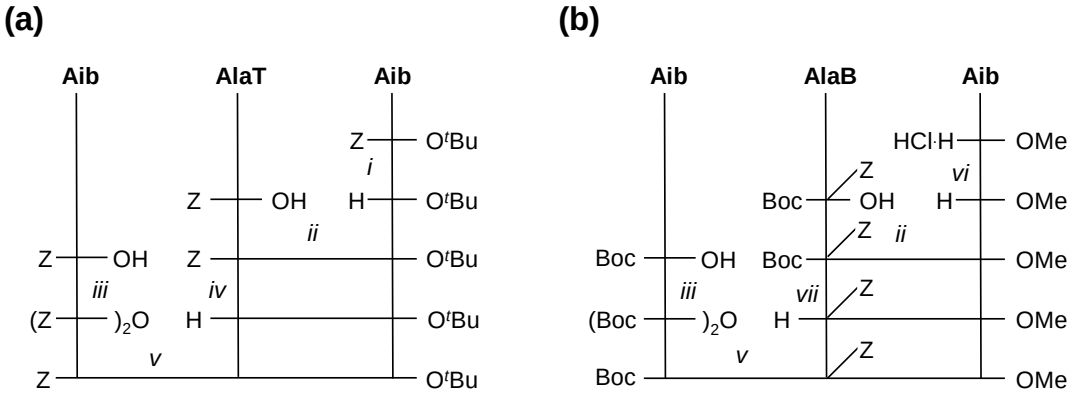


Fig. 2.8. Removal of the two Z protecting groups from Z-AlaA^{Z}-Aib-O^{t}Bu by catalytic hydrogenolysis with undesired chemoselectivity.

The N$^{\alpha}$-Z protecting group was therefore abandoned for the adenine-based nucleopeptides and not even tried for the cytosine-based ones. On the other side, the synthesis of both nucleotripeptides using the orthogonal Boc/Z protecting groups was performed without problems (Scheme 2.1b).



Scheme 2.1. Reaction conditions for the synthesis of the protected nucleo-tripeptides Z-Aib-AlaT-Aib-O^{t}Bu (a), Boc-Aib-AlaA^{Z}-Aib-OMe and Boc-Aib-AlaC^{Z}-Aib-OMe (b, AlaB= AlaA and AlaC, respectively). (i): H_2, Pd/C in DCM; (ii): EDC/HOAt, NMM, pH 8, DMF; (iii): EDC in DCM; (iv): H_2, Pd/C in MeOH; (v): NMM, pH 8, MeCN; (vi): 1 eq NMM; (vii) 20% TFA in DCM.

Protected nucleopeptides are much more lipophilic than the corresponding nucleoamino acids, therefore they are easier to handle and purify. The partial purification of the thymine-based nucleopeptides by precipitation was possible, particularly in the case of the tripeptide. Accordingly, the coupling yields were quite satisfactory (>60%) for the three groups of nucleopeptides.

Small amounts of the protected nucleo-tripeptides were used to obtain nucleopeptides with unprotected nucleobases and without other aromatic chromophores, in order to facilitate conformational analyses and pairing tests by CD. This was easily

achieved by catalytic hydrogenolysis of the Z group on cytosine and adenine, as well as on the N^{α} of the thymine-based nucleopeptide (which was replaced by a blocking acetyl moiety). Similarly, small amounts of the thymine- and cytosine-based nucleoamino acids were employed to synthesise the corresponding methyl esters using EDC/DMAP in MeOH, in order to obtain nucleoamino acid derivatives soluble in chloroform, useful for IR conformational analyses.

The synthesis of the protected guanine-based nucleo-dipeptide lead to an undesired result, since the Cl^6 on the nucleobase was replaced by HOAt used for the nucleoamino acid activation (Fig. 2.9). This was not completely a surprise, since chlorine replacement by nucleophiles during the activation of nucleotide analogues containing 2-amino-6-chloro-purine had been reported before.[118] However, since the synthesis of the guanine-based nucleoamino acid methyl ester was performed smoothly, following the same protocol adopted for the other nucleobases, it was hoped that the replacement would not have happened during the coupling. It seemed appealing to use the aromatic ether formed as an useful derivatization in order to increase the lipophilicity of the guanine-based nucleopeptides. A small scale N^{α}-Z hydrogenolysis trial showed however that a complex mixture was formed, where together with traces of the desired derivative H-AlaG^{OAt}-Aib-OMe, dipeptides containing unprotected guanine and even 2-amino-purine were formed (Fig. 2.9).

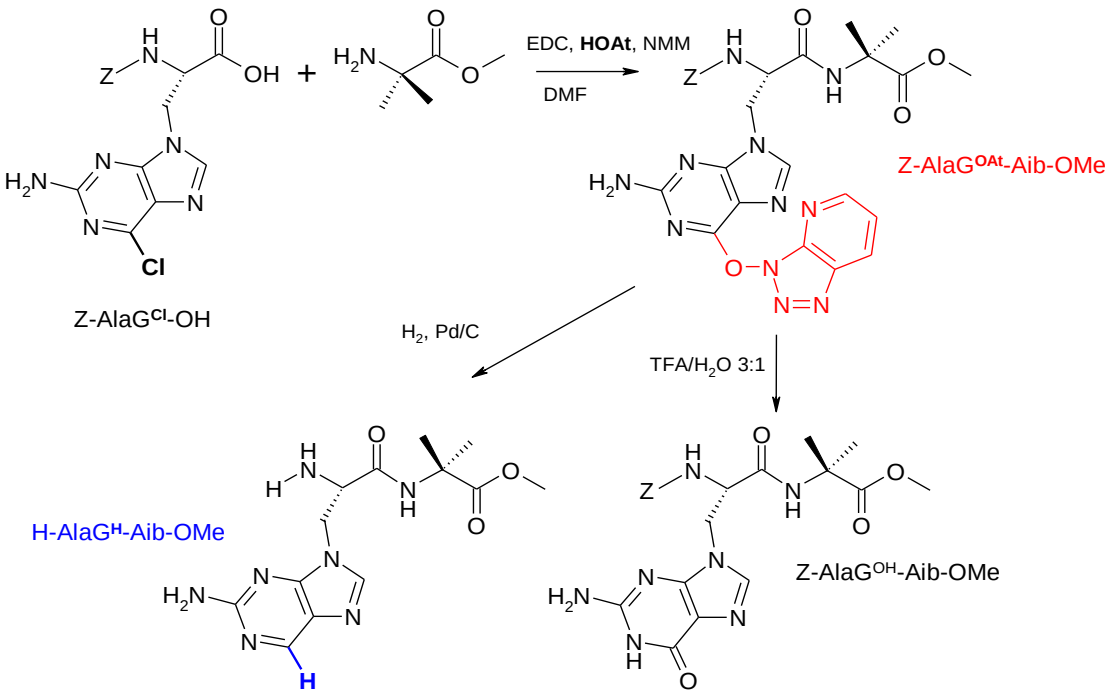
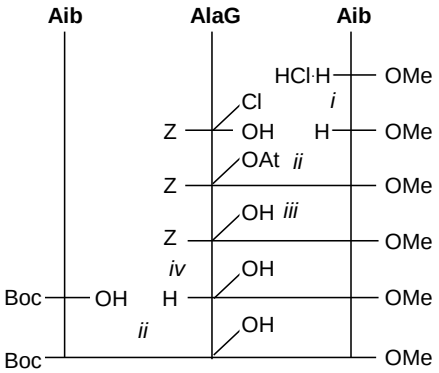


Fig. 2.9. Replacement of the Cl^6-atom by an HOAt derived group and methods of removal of the aromatic moiety.

Considering that O^6-Bzl guanine is easily deprotected under acidic treatment,[97] the complete removal of the aromatic moiety by acidic hydrolysis was decided, following the protocol reported for the deprotection of 2-amino-6-chloro-guanine to guanine.[104] The deprotection, although slow, occurred without side reactions, and the nucleo-dipeptide with free guanine Z-AlaG^{OH}-Aib-OMe was obtained. This is a extremely hydrophilic derivative, therefore it was not possible to eliminate the HOAt coproduct by acidic wash as usual, and flash-chromatographic purification with a highly polar eluent mixture was necessary.

Both to obtain a proof of concept of the viability of the N^{α}-Z deprotection of this derivative and to prepare a nucleo-tripeptide with free guanine and without other aromatic moieties for CD characterization, like the ones prepared for the other nucleobases, a small amount of Z-AlaG^{OH}-Aib-OMe was hydrogenated and coupled with excess activated Boc-Aib-OH (Scheme 2.2). The purification of the nucleo-tripeptide formed was not possible by standard work-up based on acidic and basic washes, given the high polarity, and even flash-chromatography proved not sufficient to separate it completely from some of the HOAt used for Aib activation However, by preparative HPLC an efficient purification was achieved, although with a rather poor yield. Probably the choice of the symmetric anhydride coupling method instead of the active ester formation for this step could have allowed an easier purification and consequently a higher yield.



Scheme 2.2. Reaction conditions for the synthesis of the guanine-based nucleotripeptide Boc-Aib-AlaG^{OH}-Aib-OMe. (i): 1 eq NMM; (ii) EDC/HOAt, NMM, pH 8, DMF; (iii) TFA/H_2O 3:1, 3 d; (iv): H_2, Pd/C, MeOH.

2.3.3 Synthesis of nucleopeptides containing a C-terminal lysine

To enhance the water-solubility of deprotected rigid nucleopeptides, the incorporation of a C-terminal lysine amide was planned. This was performed on the pyrimidyl based derivatives, which had proven easier to handle.

Since the C-terminal carboxyl group is blocked as an amide, orthogonal protecting groups are necessary only for the N^{α}-amino and for the side chain functional groups. In the case of the thymine-based nucleopeptides, the Z group was conserved for the N^{α}, while the acid labile Boc group was used for the ε-amino group of lysine; in the case of cytosine, the same N^{α}- and side-chain protecting groups adopted for the synthesis of its protected nucleo-tripeptides were used. Suitably protected lysine amides were thus synthesised, N^{α}-deprotected and coupled with activated N-protected Aib residues, forming protected Aib-Lys dipeptides. The synthesis of protected C-terminal lysine-amide nucleo-tetrapeptides was then performed by using the N^{α}-deprotected dipeptides as done for the Aib esters in the case of the synthesis of the nucleo-tripeptides (refer to Scheme 2.1). The presence of lysine hindered the purification by precipitation and yields were generally lower, in particular in the case of the cytosine containing nucleopeptides (40-45%).

2.3.4 Synthesis of nucleopeptides bearing two nucleobases

The synthesis of such nucleopeptides by fragment condensation was performed on the thymine-based nucleopeptides for several reasons: firstly, it was used as a way to check the preservation of the optical purity of the serine β-lactone in the alkylations performed with different protocols, then large quantities of the protected thymine containing nucleo-tripeptide were available, finally the ideal set of protecting groups was present.

On the other side, in the case of the nucleo-tripeptide methyl esters, the issue of racemisation-free saponification is raised. To devise a safe protocol, the saponification of the model peptide Z-Ala-Aib-OMe to Z-Ala-Aib-OH was studied: as the protected nucleotripeptides, this compound has a chiral residue in the middle of the sequence, while the methyl ester is bound to the hindered Aib residue. The optical purity of the acid obtained was evaluated by comparison of its specific optical rotation with the value of the same derivative obtained by acidolysis of Z-Ala-Aib-O^{t}Bu, which is considered a racemisation-free procedure. It was found that the use of 2 equivalent of LiOH 0.05 M

in a 1:1 H_2O/MeOH mixture at 0°C lead to the quantitative saponification of the dipeptide without loss of optical purity (Fig. 2.10).

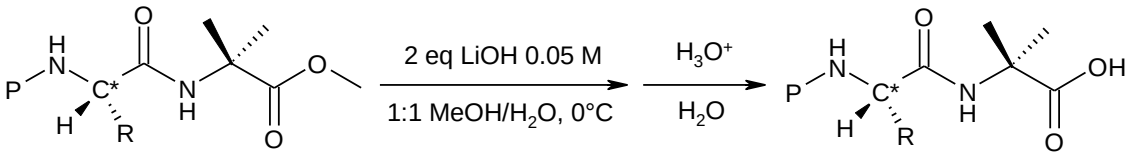


Fig. 2.10. Racemisation-free protocol for Aib methyl ester saponification in peptides containing C^{α}-trisubstituted residues.

With this result in mind, it was chosen to perform first the optimization of the fragment condensation conditions for the thymine-based nucleopeptides, and to defer the synthesis of peptides containing the other nucleobases.

When a peptide is activated, generally a 5-(4H)-oxazolone is formed (Fig. 2.11, top).[107] This species is less reactive than an active ester or an anhydride, therefore generally couplings involving peptide segments are performed under refluxing conditions. Oxazolones of C^{α}-trisubstituted amino acids racemise easily by loss of the C^{α}-proton and generation of a delocalized anion, but this is not possible for the C^{α}-tetrasubstituted Aib. In this case, however, racemisation of the residue next to the oxazolone can occur by a similar mechanism (Fig. 2.11, bottom). This side reaction can be favoured by basic conditions (pH> 8.5) and by heating.

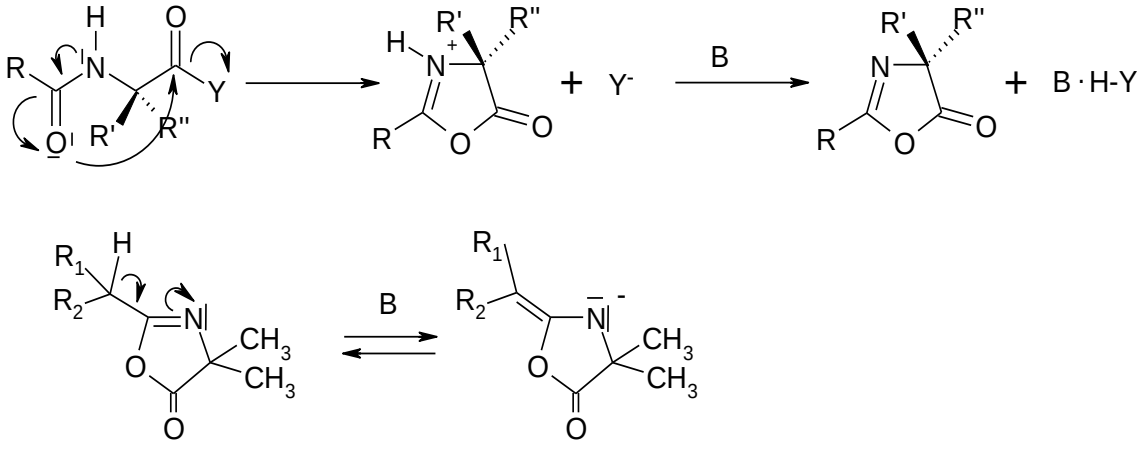


Fig. 2.11. Top: formation of the oxazolone (or *(4H)-oxazolidin-5-one*) by activation of a peptide (Y= activating group); bottom: racemisation of the residue next to the oxazolone, favoured by the generation of a delocalized anion.

Initially, the thymine containing nucleo-hexapeptide was synthesised; both the N-terminal and the C-terminal fragment were obtained by the protected thymine based nucleotripeptide, respectively by acidolysis of the *tert*-butyl ester and by catalytic hydrogenolysis of the Z protecting group. The nucleo-tripeptide acid thus formed was activated to the corresponding oxazolone with EDC and the coupling with the C-terminal fragment was performed at room temperature in DMF with the addition of the minimum amount of base to keep the pH around 8, to reduce the risk of racemisation. The reaction was slow (taking about two weeks), but a good yield of the nucleo-hexapeptide Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu was obtained (around 60%). The process was repeated three times using nucleoamino acid obtained with three different alkylation protocols (the optimized protocol derived from the method of Eschenmoser, the same protocol with a different non-nucleophilic base and the detailed protocol found in the literature).

Since in the nucleo-hexapeptide two chiral centres are present, in case of partial racemisation the four possible stereoisomers (the expected L,L, the epimerized L,D and D,L, as well as the D,D peptide) are formed. Two diastereoisomeric pairs can be therefore separated by analytical HPLC and this allows the evaluation of the raceme fraction in the starting material under the assumption that no racemisation occurred during the fragment condensation. If racemisation did occur at that stage, the optical purity of the nucleoamino acid material would be underestimated. Setting x as the fraction of D-nucleoamino acid in the starting material (or 1-x as the optical purity) and Y as the fraction of the D,L and L,D diastereoisomers in the nucleo-hexapeptides, it is straightforward to write:

$Y= 2(1-x)x= 2x-2x^2$ $x= 1/2-\sqrt{(1/4-Y/2)}$ $(1-x)= ½+\sqrt{(1/4-Y/2)}$

In the case of trace racemisation ($x<<1$):

$Y= 2(1-x)x\approx 2x$ $x= Y/2$ $(1-x)= 1-Y/2$

These equations were applied to the chromatograms of the nucleo-hexapeptides obtained using the nucleoamino acid synthesised with the three different protocols and it was found that the low-temperature protocol is the best for preserving the optical purity. Results are detailed in Table 2.1, while Fig. 2.12 shows the chromatograms for the original method by Eschenmoser and for the best method.

This result was in good agreement with the polarimetric measurements performed on the protected di- and tripeptides obtained following the different protocols. It can be inferred that the control of the alkylation conditions is crucial for preserving the optical purity and that the low-temperature protocol can significantly reduce racemisation but not suppress it completely.

Table 2.1. Diastereoisomeric ratios, epimeric fraction and molar fractions of D and L nucleoamino acid obtained for the three methods of nucleobase alkylation tested in the case of thymine evaluated by HPLC of the hexapeptides. The low-temperature protocol allowing the highest optical purity is in bold font.

Alkylation Conditions	***(D,L+L,D)/ (L,L+D,D)***	***Y***	***xD***	***xL=1-xD***
K_2CO_3, rt	*0.15*	*0.14*	*0.076*	*0.92*
DBU, rt	*0.50*	*0.33*	*0.2*	*0.8*
NaH, -78°C	***0.02***	***0.02***	***0.01***	***0.99***

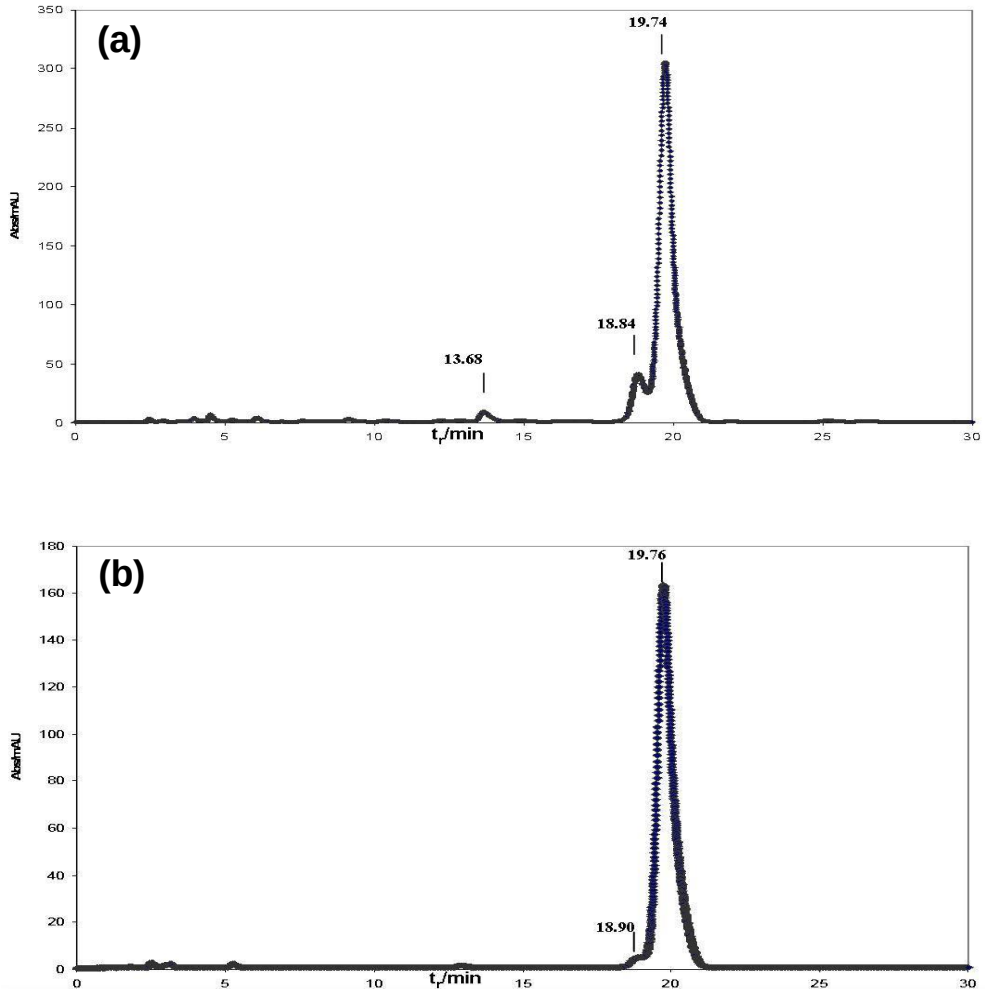
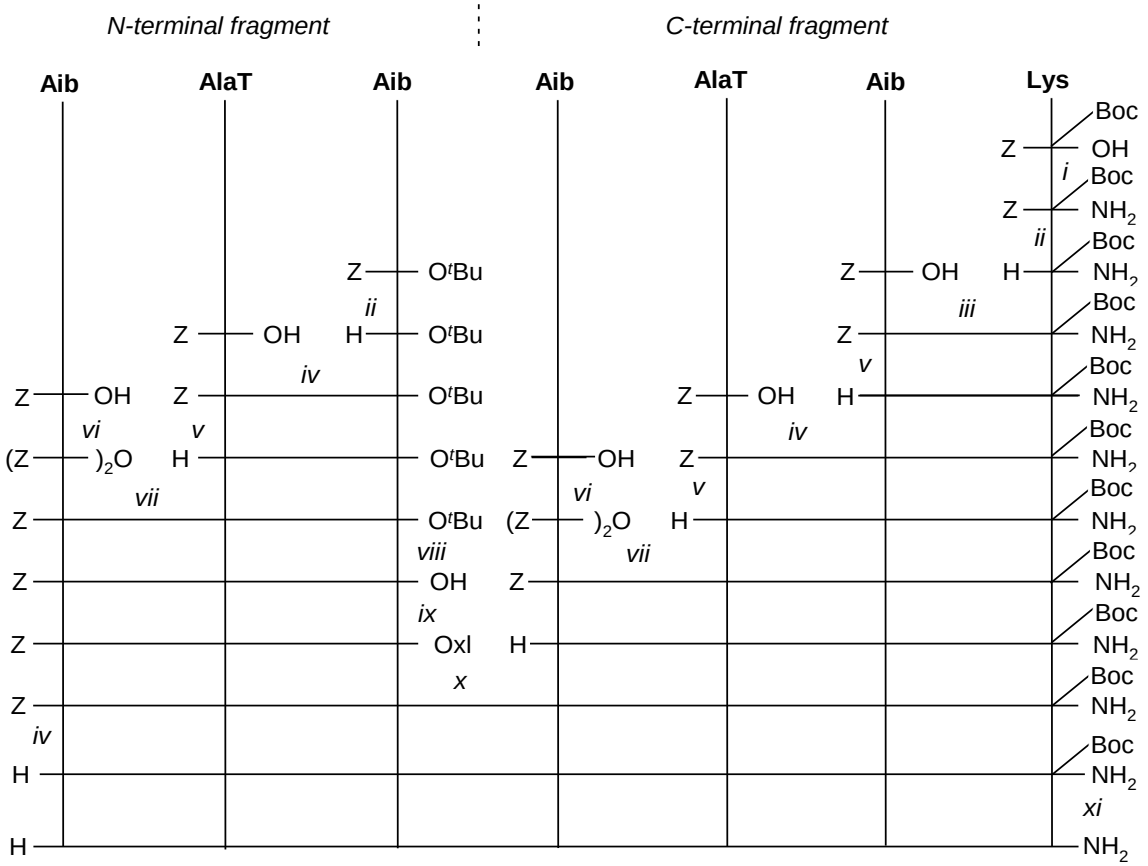


Fig. 2.12 Analytical chromatograms of the hexanucleopeptide synthesised using the nucleo amino acid obtained with the optimized protocol derived from Eschenmoser (a) and with the low temperature protocol (b). The peak at t_r 18.84 min in (a) and the shoulder at about t_r 18.90 min in (b) belong to the peptide epimers. HPLC Agilent, 44-46 % B' in 30 min.

The synthesis of a lysine amide blocked nucleo-heptapeptide was then performed. In this case the same nucleo-tripeptide acid was used as N-terminal fragment as for the last synthesis, whereas hydrogenolysis of Z-Aib-AlaT-Aib-Lys(Boc)-NH_2 afforded the C-terminal fragment (see Scheme 2.3).

The coupling was performed in MeCN under refluxing conditions because of the lower reactivity observed for the lysine containing nucleopeptides. After the flash-chromatographic treatment, the HPLC analysis evaluated the purity crude purity as around 70%, detecting minor quantities of several diastereoisomers, therefore it is likely that partial racemisation occurred. Unfortunately it was not possible to find elution conditions which allowed an efficient preparative HPLC purification of the protected peptide, nor of the N^{α}- or N^{ε}-deprotected derivatives. The crude peptide was therefore completely deprotected and at this stage preparative HPLC was performed, obtaining the deprotected nucleo-heptapeptide H-(Aib-AlaT-Aib)$_2$-Lys(H)-NH_2 as its TFA salt at 95% purity (Fig. 2.13).



Scheme 2.3. Reaction conditions for the synthesis of 2 TFA·H-(Aib-AlaT-Aib)$_2$-Lys(H)-NH_2. (i): EDC/HOBt, $NH_{3(g)}$ in DCM; (ii): H_2, Pd/C, DCM; (iii) EDC/HOBt, NMM, pH 8, DCM; (iv): EDC/HOAt, NMM, pH 8, DMF; (v): H_2, Pd/C, MeOH; (vi): EDC, DCM; (vii) NMM, pH 8, MeCN; (viii) 50 % TFA in DCM; (ix) EDC in DMF; (x) NMM, pH 8 in MeCN, reflux; (xi) 20 % TFA in DCM.

The viability of the fragment condensation approach for the synthesis of longer nucleopeptides had thus been demonstrated, even if optimization of the synthetic protocol might be necessary for application to the other nucleobases.

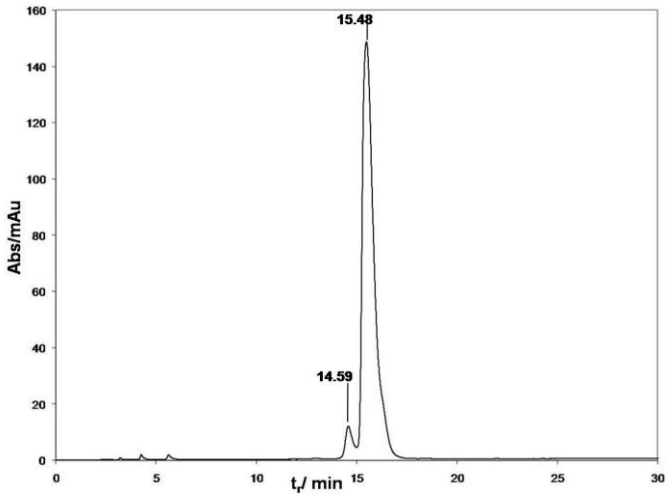


Fig. 2.13. Analytical chromatogram of TFA·H-(Aib-AlaT-Aib)$_2$-Lys(H)-NH$_2$ after preparative purification. HPLC Agilent, 1.5-6 % B' in 30 min.

2.4 Synthesis of flexible nucleopeptides

Model and functionalised flexible nucleopeptide libraries were designed and prepared. The term *library* refers to a set of compounds with similar structure synthesized by parallel synthesis.

Firstly, a convenient solid support for the solid-phase peptide synthesis (SPPS) of the flexible nucleopeptides by Boc/Bzl strategy was identified. MBHA (4-methylbenzhydrylamine) resin was used, as the peptides are not cleaved under the reaction conditions used for N$^{\alpha}$-Boc deprotection. Moreover, following cleavage from the resin, peptides are released as their C-terminal amides, thus avoiding the formation of a negative charge, which could decrease its affinity towards the polyanionic nucleotides. A semiautomatic peptide synthesizer conceived for Boc/Bzl strategy was used for the synthesis of the nucleopeptide sequences, whereas cleavage and side-chain derivatization (when required) were performed manually.

2.4.1 Model nucleopeptide libraries

The synthesis started with the structurally simpler model libraries in order to assess an optimal synthetic protocol. It was decided to prepare model nucleopeptides based on one pyrimidine and on one purine base (thymine and adenine respectively), to obtain a representative and complementary nucleobase subset without wasting too much of the nucleoamino acids in the synthesis optimization process.

The first library, used both as a proof of concept of the viability of the solid-phase synthesis and for the research of suitable activation methods and reaction condition, was made of two nucleo-hexapeptides, based on thymine and on adenine, each containing twice the same Ala-AlaB-Ala tripeptide units.

2.4.1.1. Synthetic protocol optimisation

The synthesis was performed on a 10 μmol scale by amino acid activation *via* active esters formed with BOP/HOBt[119] and DIEA was used as a base. This is a widespread protocol for solid-phase peptide synthesis using the Boc/Bzl strategy.[120] Couplings were repeated twice and 5 equivalents of proteogenic amino acid and 3 equivalents of nucleoamino acid were used per coupling. Longer coupling times in the case of the nucleoamino acids (1 hour against 20-30 minutes) were used to compensate the smaller excess added in order to spare the valuable building blocks. After completing the peptide sequences, the products were cleaved from the resin by treatment with a mixture of TFA, trimethylsilyl trifluoromethanesulfonate and *para*-chresol overnight.[121] The use of this cleavage cocktail allows simultaneous cleavage from the resin and removal of all benzyl-based protecting groups and it is much easier to handle than HF, being liquid at room temperature and less aggressive.

After recovering the peptides from the cleavage solutions by precipitation in cold, the HPLC analysis showed acceptable purity of the crudes and purification through preparative HPLC was not difficult. The yields, however, were modest (around 5 %).

A second model nucleo-heptapeptide library was designed: its nucleopeptide components were made of two tripeptide units as before, but a lysine was added at the C-terminal (Fig. 2.14, top). This was not necessary for the water-solubility of the peptides themselves, as the peptides of the first library were well soluble in water, but it was meant as a test for the synthesis of longer, potentially less soluble, nucleopeptides, to be sure that the introduction of lysine would not have caused unexpected problems.

Assuming that the not satisfactory yields obtained with the first library were due to insufficient nucleoamino acid activation, a more energetic activation protocol for the formation of the nucleoamino acid active esters was adopted. Thus HATU/HOAt[122] were employed, while the activation of the proteogenic residues was performed with the usual method. The synthesis was performed on 25 μmol scale, in order to obtain sufficient amounts of nucleopeptides for conformational analyses, not only by spectroscopic techniques, but also by 2D NMR.

2.4.1.2 Incorporation of an Aib residue by SPPS

As mentioned in the Introduction, generally short nucleopeptides made of C^{α}-trisubstitued residues are unfolded, while the presence of C^{α}-tetrasubstitued residues in a C^{α}-trisubstitued residue-based sequence can substantially improve the conformational stability of the peptides. To evaluate the ability of Aib to structure flexible nucleopeptides, two more nucleopeptides were added to the library, each one carrying an Aib residue in the middle of its sequence (Fig. 2.14, bottom). The central location was chosen in the hope that it would have exerted deeper effects on the peptide conformation than peripheric positions.

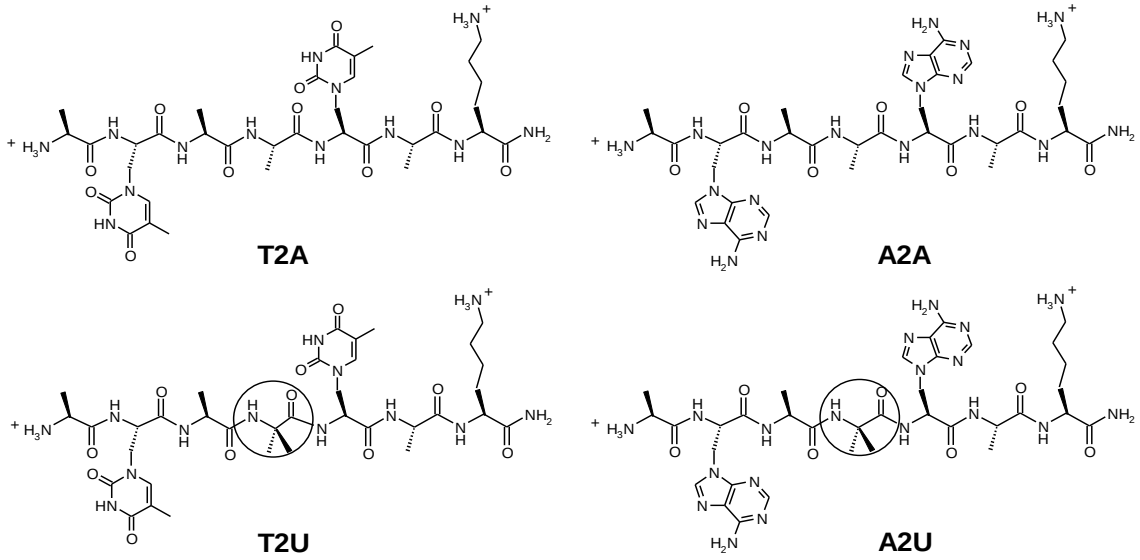


Fig. 2.14. Flexible nucleo-heptapeptides based on thymine and adenine (A2A and T2A, top), and analogues containing an Aib residue (circled) at position 4 (A2U and T2U, bottom).

Two positions were in principle available (3 and 4), both next to a nucleoamino acid ($AlaB^2$ or $AlaB^5$, respectively). Since the increased steric hindrance of Aib affects more the reactivity of the amino- than that of the carboxyl-group, it was thought that coupling the

activated Aib on a deprotected N-terminal nucleoamino acid would have been less difficult than coupling an activated nucleoamino acid on the deprotected N-terminal Aib. Position 4 rather than position 3 was therefore chosen.

The incorporation of an Aib residue in a peptide sequence by SPPS techniques is not easy.[123] To achieve an efficient activation, allowing good coupling yields, the same HATU/HOAt activation protocol used for the nucleoamino acids was applied to Aib, using a fivefold excess of the amino acid, and the coupling was repeated three times. Moreover, the coupling times for the following residue (Ala^3) were doubled.

After cleavage, recovery and purification of the nucleopeptides, very inhomogeneous results were obtained. The thymine-based nucleopeptide without Aib was obtained and purified in good yield (29 %), while the purification of the adenine-based nucleopeptide without Aib was not successful, due to impurities very close to the product peak. On the other side, it was possible to purify the two Aib containing analogues, although with a low yield (2 %). The synthesis of the thymine-based Aib containg analogue was then repeated and this time two different protocols were used in parallel, each on a 15 μmol scale. One protocol was the same used the first time, the second used BOP/HOBt activation for all amino acids apart from Aib, activated with HATU/HOAt. This was to test whether the HATU/HOAt activation was too powerful for the nucleopeptides, particularly given the longer coupling times necessary.

After cleavage, the HPLC chromatograms of both crudes were quite similar (Fig. 2.15), however the crude yield obtained with the second protocol was significantly higher.

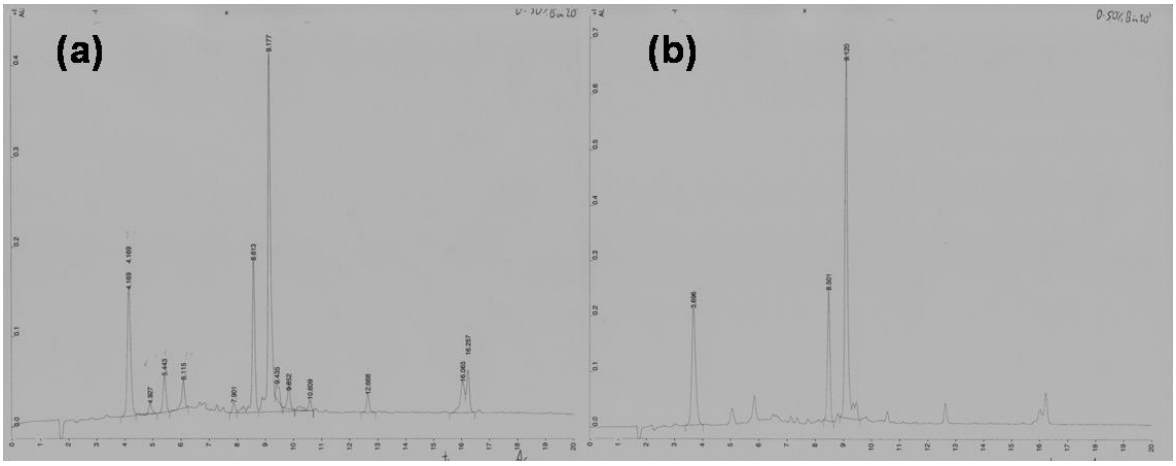


Fig. 2.15. Analytical HPLC chromatograms of the crude Aib containing thymine-based nucleo-heptapeptide T2U obtained using HATU/HOAt (a) or BOP/HOBt (b) for the nucleamino acid activation. HPLC Varian, 0-50 % B in 20 min.

After purification, the yield obtained with the first protocol was half of the yield obtained with the second (13 % and 26 %, respectively), however even the lower yield was acceptable, given the presence of Aib in the sequence.

On the other side, from this result, it is not possible to infer that the second protocol is positively better than the first one. In particular, it seemed strange that only the crude yields were different, while the crude purities were similar (about 45 % and 55 %, respectively).

2.4.2 Functionalised nucleopeptide libraries

In order to prepare guanine containing flexible nucleopeptides, it was decided to convert the 2-amino-6-chloro-purine-based nucleoamino acid Boc-AlaG^{Cl}-OH into its guanine-based derivative Boc-AlaG^{OH}-OH before use in the solid-phase synthesis, as reported in the literature for its superior homologues.[51] Since the reaction is performed under strongly acidic conditions, the N$^{\alpha}$-Boc protecting group was also removed and it was necessary to reprotect the free nucleoamino acid thus formed (Fig.2.16).

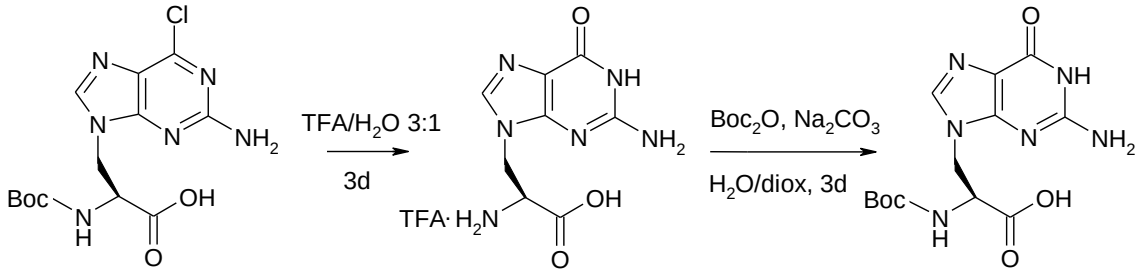


Fig. 2.16. Conversion of Boc-AlaG^{Cl}-OH to Boc-AlaG^{OH}-OH through acidic deprotection and reintroduction of the N$^{\alpha}$-Boc protecting group.

This was done by the use of Boc_2O, adapting standard procedures for α-amino acid protection.[124] Given the extreme polarity of the derivative and the presence of the free N^2H_2 on guanine, the total yield for both the deprotection and the reprotection step was acceptable (63%).

After gaining some experience on the behaviour of the nucleoamino acids on solid-phase, the design of a functionalised nucleopeptide library containing all four DNA nucleobases was addressed.

The nucleopeptides of the library were constituted of four repeating tripeptide units Ala-AlaB-Ala and of two lysines set at the C- and N-termini. While the C-terminal lysine

was introduced only for solubility reasons, the N-terminal lysine was introduced also for furhter functionalisation of its side chain. Boc-Lys(2Cl-Z)-OH was used for the introduction of the C-terminal lysine, whereas Boc-Lys(Fmoc)-OH was used for the introduction of the N-terminal lysine. It was therefore possible to selectively deprotect and derivatise the ε-amino group of the N-terminal lysine after completion of the peptide sequence by treatment with a secondary amine (20 % piperidine in DCM). Thus, by splitting the resin beads containing to each nucleopeptide, deprotecting peptide N-terminal lysine ε-amino groups and reacting them with different moieties, nucleopeptides with different functionalisations could be obtained from a single solid-phase synthesis run. The N-terminal, rather the C-terminal, lysine was functionalised because it allowed the derivatization of the last, and therefore most accessible, residue of the peptide chain after the sequence was completed.

The side-chain derivatization was performed with biotin and with a fluorophore, (fluoresceine *iso*thiocyanate, FITC). The first functionalisation was meant for obtaining biotinylated nucleopeptides suited both for surface plasmon resonance pairing experiments and for biological testing, the second mainly for biological tests.

The synthesis was planned on a 30 μmol scale for each peptide sequence. By splitting, resin beads would have been divided in three equal portions, one not further derivatised, one for biotinylation, one allowed to react with FITC. Fig. 2.17 represents the general structure of the library members and the molecules used for their functionalisation.

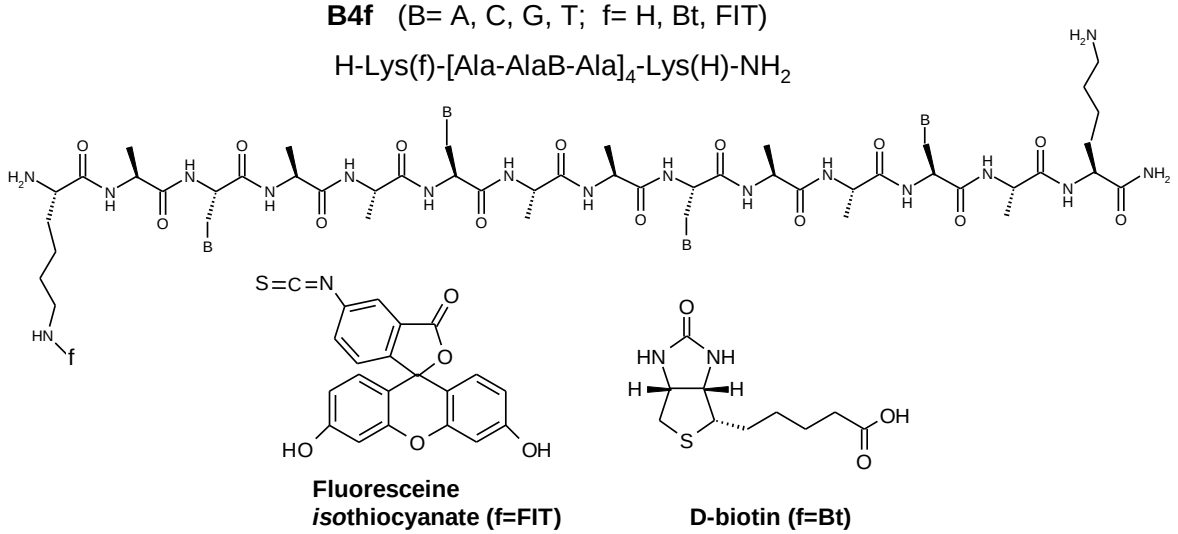


Fig. 2.17. Structure of the members of the functionalised library and of the molecules used for their functionalisation. Nucleopeptides containing each four times one of the four nucleobases were derivatised with two different moieties or left underivatised, thus forming 4*(2+1)= 12 different nucleopeptides. A fifth group of nucleopeptides carrying three cytosines and one thymine in its sequence was also prepared.

HBTU/HOBt[125] were used for active ester formation for all amino acids, since side reactions with guanine derivatives were reported by using phosphonium-based coupling reagents, like BOP.[118] This activation method has an intermediate reactivity between the HATU/HOAt and BOP, therefore it was considered a suited general purpose protocol for the synthesis. Biotin was coupled to the ε-amino group of lysine by activating its carboxyl group with HBTU/HOBt, like the carboxyl groups of the protected amino acids. FITC, being an electrophilic *iso*cyanate, does not require activation to attack free amines, therefore only pH adjustment to about 8 was necessary.

In order to prepare a single-base mismatch model, a nucleopeptide containing two different nucleobases in its sequence (one thymine in position 5 and three cytosines in position 2, 8 and 11) was designed and derivatised in the same way as the 'homobase' nucleopeptides.

The crude mixtures were rather complex, making purification difficult. Therefore the yields of the 15 functionalised nucleopeptides after purification were modest, ranging between 3 % and 11 %. The (formal) average yields per step including the purification were however comparable with the yield obtained for the best nucleoheptapeptide prepared by SPPS (78-85 % against 84 %). Only the purification of the guanine-based biotinylated peptide was not possible, because the lyophilized crude product turned insoluble in any solvent mixture. A reason could be the well-known tendency to aggregation of guanine derivatives,[97,116] maybe increased by some serendipitously strong interaction with the biotin ring. Apart from that case, the water solubility of the underivatised nucleopeptides was high and even the more lipophilic FITC-derivatised peptides were well soluble up to 1 mM.

In some cases the derivatised nucleopeptides were obtained in higher yields compared to the underivatised ones. This could be due to side reactions due to the free ε-amino group of the N-terminal lysine in basic environment before the cleavage. Yields for all the derivatives are presented in Table 2.2.

After this library, a novel library was designed, composed of thymine-based nucleopeptides containing five tripeptide units between two lysines (Fig. 2.18, top). Thymine was chosen for the preparation of this longer nucleopeptides because it had proven the less problematic nucleobase (indeed it had the highest average yield among the five nucleopeptide sequences tested, Table 2.2, last row) and because thymine-containing peptides had shown interesting pairing properties in the surface plasmon resonance experiments (see Section 4). If pairing is cooperative, a peptide with five nucleobases

should display a much stronger binding than a peptide with four. Partial biotinylation of the peptide was therefore planned to carry on pairing tests. Moreover, a cysteine containing, thymine-based (see Fig. 2.18, middle) nucleopeptide was designed to prepare a nucleopeptide dimer linked by a disulphide bridge, as already reported in the literature with other nucleotide analogues of peptide-like structure.[58]

Table 2.2. Yields of the various derivatised nucleopeptides obtained after purification.

Yield (%)	**A4**	**C4**	**G4**	**T4**	**CTCC**
H	5	3	10	8	8
Bt	9	7	-	8	4
FIT	6	11	9	7	8
Average	6.7	7	6.3	7.7	6.7

Notes. Colums refer to the nucleobase sequence of the peptides (A4: adenine-based; C4: cytosine-based; G4: guanine-based; T4: thymine-based; CTCC: mixed base peptide), rows to the derivatization (H: underivatised; Bt: biotinylated; FIT: reacted with fluoresceine *iso*thiocyanate). The last row reports the average yield for the three derivatives.

This dimer molecule would have carried ten bases (on two strands) and it could in principle have paired with complementary, adenine-based nucleopeptides or nucleotides both by independent binding of the two strands to form Watson-Crick bonds with 1:1 stoichiometry and by cooperative pairing of both strands, one by Watson-Crick, the other by Hogsteen interactions, with the same complementary sequence (triplex formation). The latter possibility is represented in Fig. 2.18, bottom.

The formation of a homodimer through a disulphide bridge can be achieved both on solid-phase and in solution. In the first case the protecting group for the thiol function is removed under oxidizing conditions when the peptides are still bound to the resin, so that a disulphide bridges are formed between next molecules. In the second case the peptide is synthesized, cleaved and purified as a monomer, then allowed to dimerize in solution. In this case, either the thiol protecting group is stable to the cleavage and removed after the purification stage, or the cleavage and the purification are performed under non-oxidizing conditions. The on-resin dimerization requires one purification step less than the dimerization in solution, on the other side the dimer is generally less pure and solubility problems can appear during the purification. Therefore, it was chosen to perform the dimerization in solution, using a protecting group that was removed during the cleavage,

and to purify the product under non-oxidizing conditions in order to allow the dimerization only of pure product.

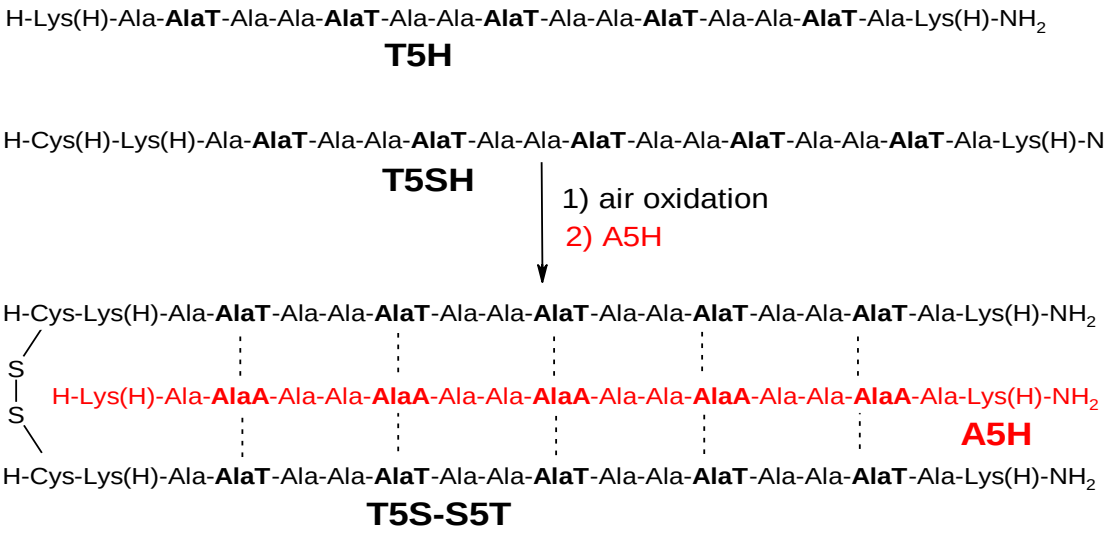


Fig. 2.18. Structure of the underivatised nucleopeptide containing five thymine bases (T5H, top) and of its cysteine containing analogue (T5SH, middle). Bottom: planned formation of a nucleopeptide dimer by disulphide bridge and triplex formation with a complementary adenine-based nucleopeptide (not succesfully synthesized).

In the Boc/Bzl strategy, cysteine is generally introduced as Boc-Cys(Acm)-OH: the acetamidomethyl protecting group (Acm) is removed by I_2[126] or other oxidising agents. Working in solution, the removal of the excess of oxidising reagent can be difficult, therefore this way was discarded. Since cysteine was introduced as the N-terminal residue, it was possible to use Fmoc-Cys(Trt)-OH, whose thiol-protecting trityl group is easily cleaved during the cleavage, while its N^{α}-protecting Fmoc group was deprotected manually prior to the cleavage from the resin. The presence of *p*-cresol as a scavenger in the cleavage cocktail protected the free thiol from oxidation. Since no derivatization of its N-terminal lysine (Lys^2) was planned, the latter was introduced as Boc-Lys(2Cl-Z)-OH, to avoid possible side-reactions due to the simultaneous formation of two free groups under basic conditions before the cleavage.

The synthesis scale was set to 10 μmol for the biotinylated derivative and to 22 μmol for the underivatised and the cysteine-containing peptide. The activation protocol employed BOP/HOBt for Boc-protected protein aminoacids and HATU/HOAt for the nucleoamino acids and for Fmoc-Cys(Trt)-OH, because it is well known that N^{α}-Fmoc-protected amino acids are less reactive than the corresponding N^{α}-Boc-protected derivatives.[107]

The crude mixtures were complex (Fig. 2.19a), as expected from the experience with the shorter nucleopeptides of analogous mixture. Purification of the three derivatives was successful (Fig. 2.19b), even if difficult, and yields were consequentially low (1-3 %).

It was tried to obtain an adenine-based nucleopeptide with similar structure to perform pairing test on the adenine-thymine pair, but the synthesis failed completely (the HPLC-MS analysis of the crude mixture showed no traces of the expected product).

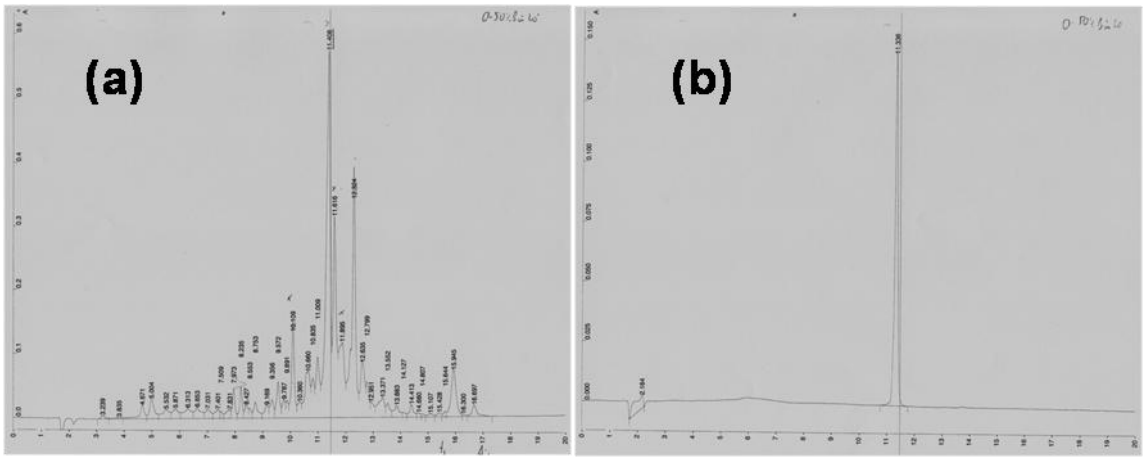


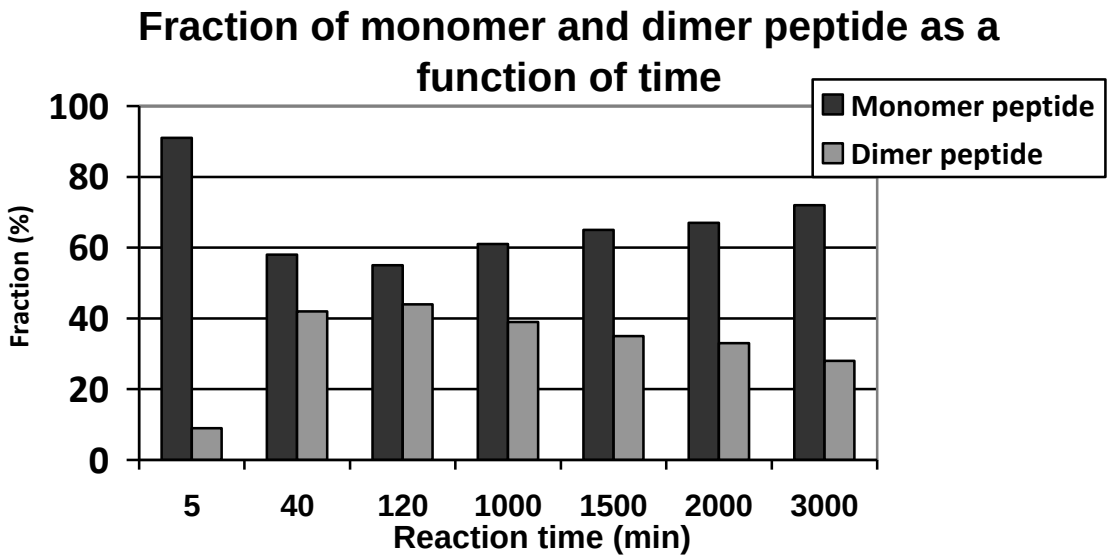
Fig. 2.19. Analytical HPLC chromatograms of the crude (a) and purified (b) underivatised thymine based nucleopeptide T5H. HPLC Varian, 0-50 % B in 20 min.

2.4.2.1 Trials of cysteine-mediated nucleopeptide dimer formation trials

After the synthesis and purification of the cysteine containing thymine-based nucleopeptide T5SH, tests of its oxidative dimerization were carried on samples with purity 80-90 % obtained during the purification process together with the best fraction (purity higher than 95 %). The monomer peptide was dissolved in H_2O/DMSO 9:1 at pH 8-9 and the solution was stirred to favour oxygen saturation, while the reaction course was monitored by HPLC.

Soon after the dissolution, close to the monomer peak and with a slightly longer retention time, as expected, the dimer peak appeared. Its relative area increased with time until about 2 hours after dissolution, as it represented about 45 % of the total area, then the increase stopped and the relative area of the dimer compared to the monomer seemed to decrease (Graph. 2.2). A closer investigation showed that on the walls of the reaction vessel a precipitate was present, which proved completely insoluble both in aqueous solutions at different pH and in every polar organic solvent mixture tried. It was assumed that when the dimer concentration passes a given threshold value, it starts to aggregate and precipitate in

an autocatalytic process, so that its actual concentration in solution (and therefore the area of its peak in the HPLC chromatograms) decreases with time.



Graph 2.2. Fractions of monomer and dimer cysteine containing peptide as a function of time evaluated by HPLC chromatograms. The higher reaction time values correspond to 16, 24, 32 and 48 hours, respectively.

Given this unexpected problem, the dimerization strategy was changed. It was therefore planned to form the dimer peptide by using a bifunctional linker, able to react with the thiol function of two molecules. As bifunctional linker, 1,4-*bis*-maleimidobutane (BMB) was chosen, as this compound is able to react selectively with free thiol functions in aqueous solution at neutral pH (see Fig.2.20).[127]

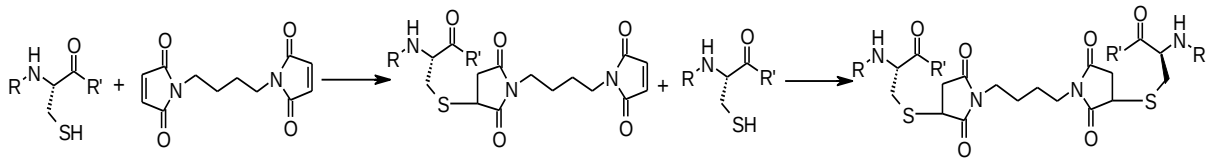


Fig. 2.20. Cysteine containing peptide dimerization through a 1,4-bis-maleimidobutane linker. In general R, R' represent the rest of the peptide sequence; in the case of the nucleopeptide T5SH R= H, R'= Lys(H)-(Ala-AlaT-Ala)$_5$-Lys(H)-NH$_2$.

The two peptide molecules would be joined by a linker composed of two rigid succinimide rings and one central flexible portion represented by the four methylene groups. With the

linker in fully extended conformation the distance between the two sulphur atoms would be about 10.9 Å.[128] This could be an advantage, since the broader gap between the two nucleopeptide strands could favour the accommodation of a third complementary strand in the middle, able to interact with both series of nucleobases by forming a triplex structure.

The reaction was performed at first following literature protocols (0.2 mM nucleopeptide in H_2O, addition of 0.5 equivalent BMB 10 mM in DMSO, pH 7, 2 hours); as the starting material appeared almost unchanged by HPLC analysis, longer reaction times (3 days) and a higher molar excess (up to 8 equivalents) were employed, resulting in the disparition of the monomer nucleopeptide peak with formation of two new peaks. However, after purification and MS analyses of both, neither resulted the expected product, only low molecular weight species being detected.

Unfortunately, the synthesis of the nucleopeptide dimer could not be tried again with a different protocol because much of the monomer had been employed for the unsuccessful trials described and what remained available would have not been sufficient for performing the biological and pairing test which were the reason why its synthesis had been undertaken.

3. CONFORMATIONAL CHARACTERISATION

The conformational preferences of the nucleopeptides were investigated both in the solid state (by X-ray diffraction of single crystals, when available) and in solution, by infrared spectroscopy (in $CDCl_3$), NMR spectrometry (in CDCl3, d_6-DMSO, or H_2O) and circular dichroism (in $CHCl_3$, MeOH or aqueous solution).

3.1 X-ray diffraction

At the moment crystallisation trials are ongoing, in order to obtain single crystals, suitable for structure determination by X-ray diffraction. The trials are performed on all the nucleoamino acid derivatives and rigid nucleopeptides synthesized; moreover, co-crystallisation trials involving complementary rigid nucleopeptides with unprotected bases are also ongoing.

Up to now, the crystallisation experiments have been successful only in the case of the protected thymine-based nucleo-tripeptide Z-Aib-AlaT-Aib-O^{t}Bu. Given the well-known propensity of Aib-rich peptides to crystallize, this result might seem quite disappointing. It must be kept in mind, however, that the nucleopeptides are more polar than most of the Aib-rich protected peptides which formed single crystals. The increased polarity and the propensity to strong intermolecular interactions of the nucleopeptides make them more prone to rapid aggregation in most of the solvents which are generally used for crystal growth, thus preventing the slow, ordered formation of the crystal nuclei necessary for obtaining single crystals.

3.1.1 Crystal structure of Z-Aib-AlaT-Aib-O^{t}Bu

The structure of this compound was solved by Dr. M. Crisma (Institute of Biomolecular Chemistry, ICB-CNR, Padova, Italy) and it proved very interesting not only from the structural point of view, but also for providing information about nucleobase-mediated intermolecular interactions.

Fig. 3.1 represents the structure of the nucleo-tripeptide. The peptide is folded in a type-I β-turn: Table 3.1 gives the relevant torsion angle values, comparing them with the

ideal values from the literature.[72] It can be seen that the observed values are within ±10° from the theoretical values, with larger deviations in the case of the nucleoamino acid.

Table 3.1 Torsion angles relative to the internal residues of Z-Aib-AlaT-Aib-O^{t}Bu compared with the ideal values for a type-I β-turn.

Angle / **Molecule**	ϕ_1	ψ_1	ϕ_2	ψ_2
Z-Aib-AlaT-Aib-O^{t}Bu	-54.6 °	-35.1°	-82.4°	-9.3 °
Ideal	-60 °	-30 °	-90 °	0 °

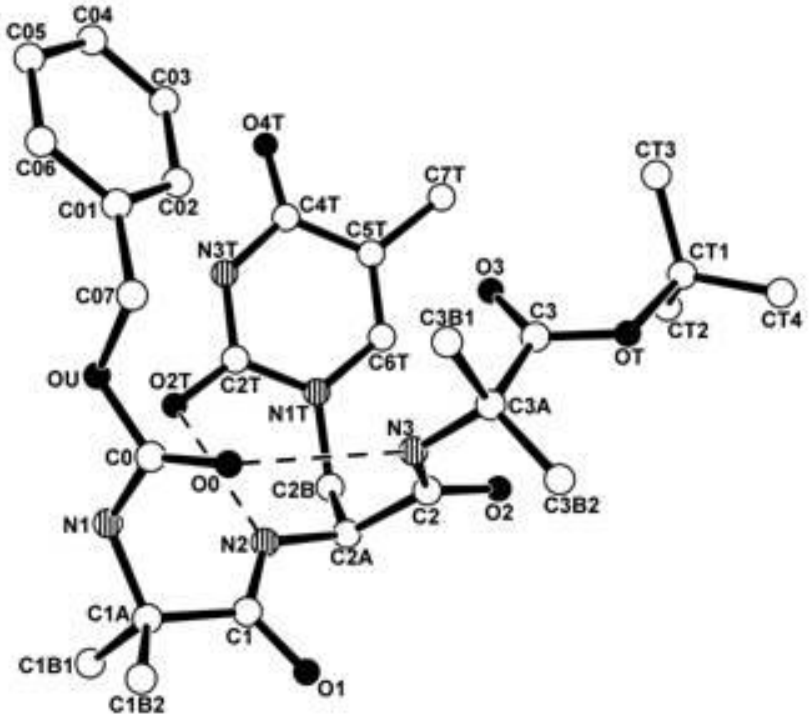


Fig. 3.1. Solid-state structure of the protected thymine-based nucleotripeptide Z-Aib-AlaT-Aib-OtBu with atom numbering. Empty circles are used for carbon atoms, striped circles for nitrogen atoms, full circles for oxygen atoms, while hydrogen atoms have been omitted for clarity. Dashed lines represent *intra*molecular H-bonds.

Two *intra*molecular hydrogen bonds are present. The first is formed between the NH of the Aib^3 and the CO of the urethane protector, forming a C_{10} structure which encompasses the β-turn, as expected for an Aib rich tripeptide. The second H-bond is an *intraresidue* bond, formed between the αNH of the nucleoamino acid and the C(2)O of the thymine ring. This was an unexpected and very interesting result, since the bond freezes the orientation of the

nucleobase with regard to the backbone, preventing rotation around the β-CH_2 of AlaT. The type-I β-turn conformation favours the formation of this bond, since the value of the ϕ_2 torsion angle (C(O) Aib^1-N^{α} $AlaT^2$-C^{α} $AlaT^2$- C(O) $AlaT^2$) is well suited to take the thymine ring next to the αNH.

The unit cell is a dimer, stabilized by two pairs of *inter*molecular hydrogen bonds (Fig. 3.2). Two H-bonds are formed between the N(3)H of the thymine rings on one molecule and the C(2)O of the thymine ring on the other one, and two are formed between the NH of the Aib^1 on one molecule and the C(4)O on the thymine ring of the other molecule. Thus all H-bond donor and acceptor sites form H-bonds within the unit cell, and interestingly the C(2)O of the thymines are *bidentate* acceptors. The dimer formation is due to thymine-thymine recognition by complementary hydrogen-bonding sites in the so called *reverse Watson-Crick* pairing mode (see Section 4.1 for further details).[56]

Table 3.2 reports the relevant properties for the intra- and intermolecular H-bonds. The intraresidue H-bond has the N-$\hat{H}^{\cdots}$O angle which deviates most from the linear geometry; this is however also the bond by which the nitrogen and the oxygen atom are closest to one another.

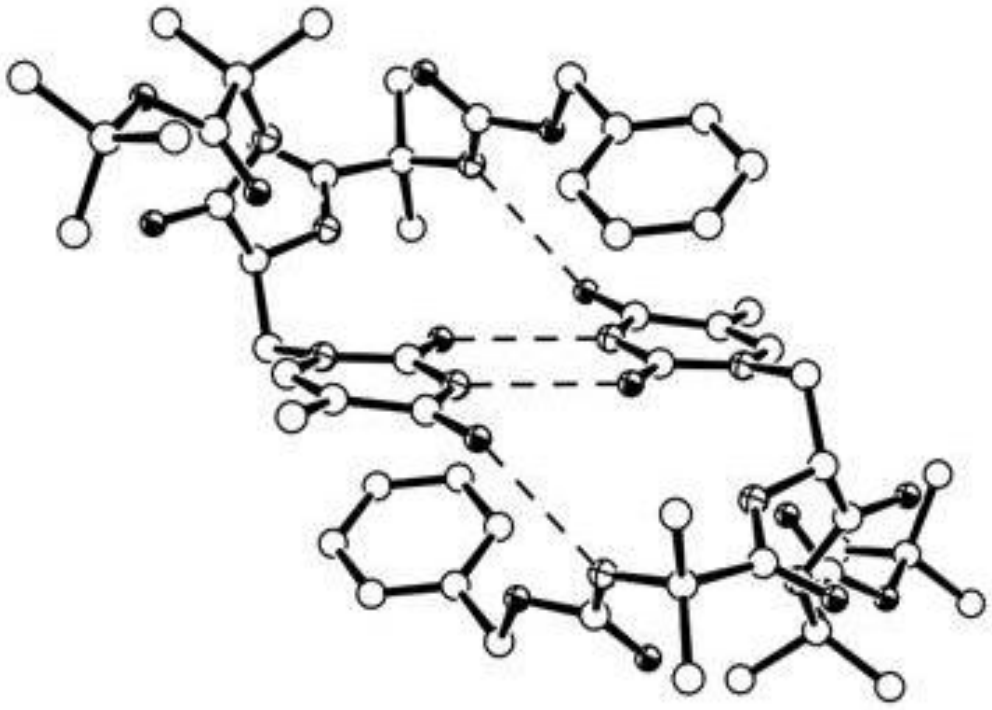


Fig. 3.2. Unit cell of the crystal of Z-Aib-(D,L)-AlaT-Aib-O^{t}Bu. Empty circles represent carbon atoms, crossed circled represent nitrogen atoms, circles with black sectors represent oxygen atoms, while hydrogen atoms have been omitted for clarity. Dashed lines represent *inter*molecular H-bonds.

Table 3.2. H-bond lengths and angles for Z-Aib-AlaT-Aib-O^{t}Bu.

D-H...A	**d(D-H)/ Å**	**d(H...A)/ Å**	**d(D...A)/ Å**	**<(DHA)/°**
N-H Aib^3...O Z	0.86	2.28	3.12	163.8
αN-H AlaT^2...O(2) T	0.86	2.03	2.80	147.9
N-H Aib^1...O(4) T' (*inter*)	0.86	2.12	2.97	168.8
N(3)-H T...O(2) T'(*inter*)	0.86	2.00	2.85	172.0

Notes: O Z refers to the carbonyl of the Z protecting group; O(2) T, O(4) T and N(3)-H T refer to the heteroatoms on the thymine ring of AlaT^2. A *prime* has been added to the atoms belonging to the second molecule in intermolecular H-bonds.

A closer examination of the unit cell allows one to realize that it is centrosymmetric, namely that the cell is a heterochiral dimer formed by one L-AlaT containing and one D-AlaT containing nucleopeptide. The fraction from which the crystal was obtained was optically active, therefore complete racemisation can be excluded. In a later crystallisation trial performed on the nucleotripeptide obtained starting from the nucleoamino acid formed with the low-temperature protocol, again a raceme crystal with the same packing was obtained. Since it was found that the raceme fraction of that sample of nucleo-tripeptide was very low, it can be inferred that the efficient packing of the raceme crystal favours crystallisation of the small raceme fraction over crystallisation of the bulk in enantiopure form.

It is important to stress that this crystal confirm the hypothesis laying at the base of this work: nucleopeptides rich in Aib assume a stable, folded conformation which allows nucleobase-nucleobase interactions. Even if in the solid state the peptide is folded in a type-I β-turn and not in the canonical 3_{10}-helix forming type-III β-turn, this should not prevent the assumption of helical structures by the longer nucleopeptides, altough likely distorted from the ideal values.

Thymine-thymine association appears to be a common tendency for the rigid nucleopeptides, not just limited to enantiomer pairs as in the crystal state, but also important in organic solution (see following Sections).

3.2 Infrared absorption spectroscopy

The IR absorption study was performed in deuterochloroform (a low polarity solvent, often employed for conformational studies on peptides). The two frequency intervals richest in conformational information are:[129]

(i) 3550-3200 cm^{-1} (*Amide A band*), related to the stretching vibrations of the N-H bonds belonging to the urethane and amide groups;

(ii) 1800-1600 cm^{-1} (*Amide I band*), related to the stretching vibrations of the C=O bonds of the ester, urethane and amide groups.

The presence of an hydrogen-bond between a donor (e.g. the amide N-H) and one acceptor (e.g. the amide C=O group) modifies the bond strength constant of both groups involved, causing an alteration of their *bending* and *stretching* vibration frequencies. Under the harmonic approximation for the oscillators, vibration frequencies are given by the Hooke's law:

$$\nu_{vibr} = \frac{1}{2\pi c} \cdot \sqrt{\frac{k}{\mu}}$$

where : ν_{vibr} = vibration frequency

c = speed of light

k = bond strength constant

μ = reduced mass $[(m_1 \cdot m_2)/(m_1+m_2)]$

More precisely, a general displacement to lower vibration frequencies (*red shift*) is observed, with formation of more intense and broader bands. From the experimental observation of this phenomenon it is possible to obtain information on the presence of NH groups engaged in hydrogen bonds or exposed to the solvent (the so-called 'free NHs').

Indeed, it is generally assumed that in deuterochloroform solvated NH groups resonate at wavenumbers higher than 3400 cm^{-1}, whereas H-bonded NHs resonate at lower wavelengths.[130,131]

Moreover, it is possible to discriminate between *intra*- and *inter*-molecular hydrogen bonds through the analysis of the variation of the relative intensity of the bands of the bonded and 'free' NHs as the concentration is changed, since *intra*molecular H-bonds are insensitive to dilution, whereas *inter*molecular H-bonds are favoured at higher

concentrations. In turn, the discrimination between *inter*- and *intra*molecular hydrogen bonds gives hints about the conformation of the peptide in solution.

The discrimination between solvated and hydrogen-bonded NH groups is possible only in solvents of low polarity, such as chloroform, which do not alter the position of the bands of the solvent exposed NHs. On the contrary, more polar solvents can form hydrogen bonds with the amide NHs and shift their absorption bands close to the region where the bands of the NHs bound to the peptide CO groups are found. Therefore, in high polarity solvents, it is very difficult to discriminate between solvated and bonded NH groups.

This means that the conformational analysis was not possible on compounds which were not soluble in deuterochloroform, like the nucleoamino acids, or the completely deprotected nucleopeptides, and sometimes were difficult to perform for the nucleopeptides with unprotected nucleobases, in particular with unprotected guanine.

The following sections present the most important information obtained with this technique first on the nucleo-tripeptides and their building-blocks, then on the lysine-containing nucleopeptides, and finally on the longer nucleopeptides.

3.2.1 Nucleopeptide building blocks

3.2.1.1 Thymine-based nucleopeptides

Thymine containing nucleopeptides have been studied in particular detail, taking advantage from the possibility of a comparison with the solid state structure of the protected tripeptide (section 3.1). This was done also in order to apply the information obtained to the interpretation of the spectra of nucleopeptides containing other bases. Table 3.3 presents the stretching frequencies in the amide A and amide I regions for nucleopeptides containing one thymine at $1 \cdot 10^{-3}$ M concentration.

In addition to the nucleoamino acid ester, to the protected nucleo-di- and -tripeptides, and to the Ac-blocked tripeptide, two more compounds are present, synthesized on purpose. The first, Z-Aib-AlaT^{All}-Aib-O^{t}Bu is a protected nucleo-tripeptide alkylated with allyl bromide at the position 3 of the thymine ring (Fig. 3.3), to suppress the contribution of the side-chain NH to the spectrum. The second is a protected nucleo-tetrapeptide, Z-Aib-Aib-AlaT-Aib-OtBu, synthesized as a higher homologue of the tripeptide by Z-deprotection and coupling with an activated Aib residue.

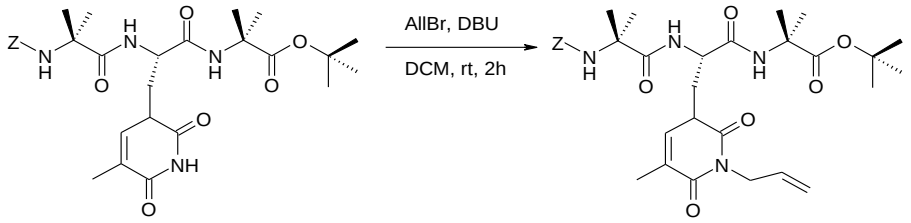


Fig. 3.3. Synthesis of Z-Aib-AlaT^{All}-Aib-O^{t}Bu by allylation of Z-Aib-AlaT-Aib-O^{t}Bu.

Table 3.3. IR absorption frequencies (cm^{-1}) in the specified intervals for nucleopeptides containing one thymine in deuterated chloroform at $1\cdot10^{-3}$ M concentration.

Peptide	3600-3200 cm^{-1}	1800-1600 cm^{-1}
Z-AlaT- OMe	3454[a], 3422, 3388	1749[a], 1714, 1684, 1648, 1622, 1600,
Z-AlaT-Aib-O^{t}Bu	3426, 3384	1726,1708[a], 1682,1646,1626,1600,
Z-Aib-AlaT-Aib-OtBu	3442, 3388[a], 3342	1732[a],1710, 1682, 1626, 1600
Z-Aib-AlaT^{All}-Aib-O^{t}Bu	3432, 3376[a], 3344	1734[a],1716, 1692, 1666, 1636, 1600
Ac-Aib-AlaT-Aib-O^{t}Bu	3436, 3398, 3340[a], 3326	1730, 1704[a], 1682, 1656[a], 1624, 1600
Z-Aib$_2$-AlaT-Aib-O^{t}Bu	3444[a], 3420, 3392, 3346,3310[a]	1736[a], 1712, 1690, 1664[a], 1624, 1600

Notes. [a]: shoulder; underlined values refer to weak (...) or strong (___) bands; values not underlined are referred to bands of intermediate intensity.

All peptides apart from Z-Aib-AlaT^{All}-Aib-O^{t}Bu display a band in the frequency range 3384-3398 cm^{-1}, thus suggesting the attribution of this band to the stretching of the N(3)H imide on the thymine ring.

The IR absorption spectra of the nucleoamino acid ester and of the nucleo-dipeptide in $CDCl_3$ at three different concentrations ($1\cdot10^{-2}$, $1\cdot10^{-3}$, $1\cdot10^{-4}$ M) in the amide A region are shown in Fig. 3.4.

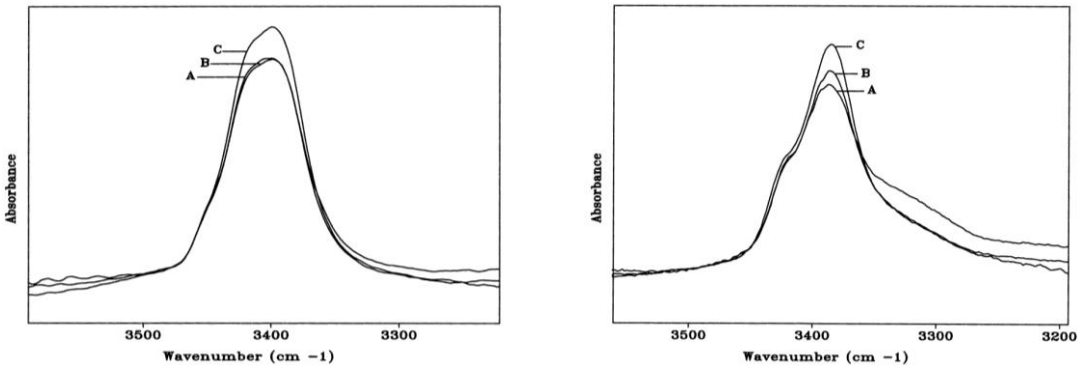


Fig. 3.4. Amide A region of the IR absorption spectra of Z-AlaT-OMe (left) and of Z-AlaT-Aib-O^{t}Bu (right) in $CDCl_3$ solution at $1\cdot10^{-2}$ (A), $1\cdot10^{-3}$ (B), $1\cdot10^{-4}$ M (C) concentration.

In Fig. 3.4 and in all other similar figures, spectra at different concentrations are normalized, since tenfold dilution are compensated by a tenfold increase of cell path length (0.1, 1.0 and 10 mm for $1\cdot10^{-2}$, $1\cdot10^{-3}$, $1\cdot10^{-4}$ M concentration, respectively).

Since no bands at frequencies shorter than 3400 cm^{-1} are observed in Fig. 3.4, it can be inferred that the NHs of the derivatives are solvated. The effect of dilution is modest.

As it can be seen from Table 3.3, all the tripeptides and the tetrapeptide display a strong band relative to bound NHs at about 3340 cm^{-1}; the low frequency points to a significant strength of the H-bonds formed.

To ascertain whether the bands of bound NHs found in the spectra of the tripeptides and of the tetrapeptide are due to intermolecular or intramolecular interactions, the effect of dilution has to be considered. There are strong dilution effects on the Z-protected tripeptide (see Fig. 3.5, left), as well as on the acetylated tripeptide and on the tetrapeptide, while dilution effects are modest for the N(3)-alkylated tripeptide (Fig. 3.5, middle). It can be therefore deduced that for first three derivatives intermolecular hydrogen bonds contribution is important, while the H-bonded NH band observed for the latter peptide is mainly due to intramolecular H-bonds.

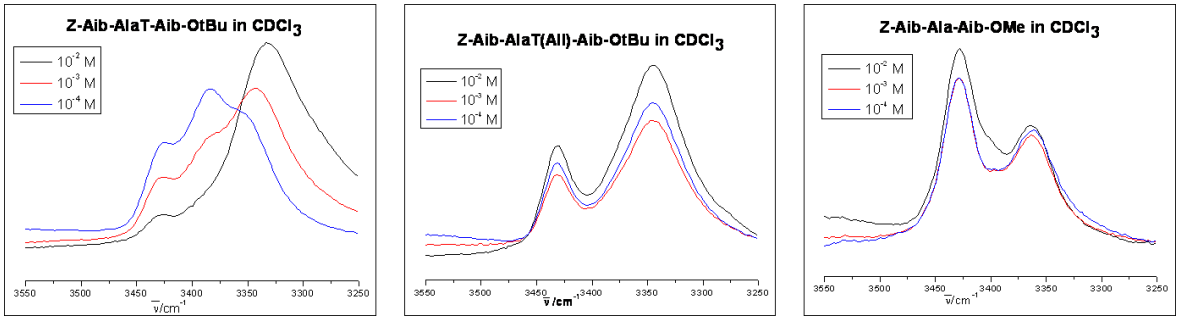


Fig. 3.5. Amide A region of the IR absorption spectra of the nucleo-tripeptides Z-Aib-AlaT-Aib-O^{t}Bu (left) and Z-Aib-AlaT^{All}-Aib-O^{t}Bu (middle), and of the model tripeptide Z-Aib-Ala-Aib-OMe (right), in $CDCl_3$ solution at $1\cdot10^{-2}$, $1\cdot10^{-3}$, $1\cdot10^{-4}$ M concentration.

The spectrum of the alkylated nucleo-tripeptide can be compared with the spectrum of the model tripeptide Z-Aib-Ala-Aib-OMe (Fig. 3.5, right). Both tripeptides have two Aib and a C^{α}-trisubstituted residue in position 2. Both tripeptides have three αNHs and in both spectra two bands are observed, one at about 3430 cm^{-1}, due to solvated NHs, the other around 3350 cm^{-1}, due to H-bonded NHs. However, for the model tripeptide the latter band is smaller than the former, while the contrary is true for the alkylated nucleo-tripeptide. In a rather simplistic way, we can attribute the smaller band to one NH group and the larger band to two NH groups: this would mean that there is only one H-bonded NH group in the

model peptide. Dilution effects are modest for both peptides, in particular going from $1 \cdot 10^{-3}$ M to $1 \cdot 10^{-4}$ M, therefore *inter*molecular stabilization at $1 \cdot 10^{-3}$ M is of little importance and *intra*molecular H-bonds must be present. The only intramolecular H-bond that is likely to be formed in the model peptide involves the NH of Aib^3, bound to the Z-CO through the formation of a C_{10} structure enclosing a β-turn, as expected for an Aib-rich tripeptide.

Similarly, there must be two intramolecularly H-bonded NHs in the alkylated tripeptide: if one is the same NH of Aib^3 forming a C_{10} structure, the second could be the αNH of the nucleoamino acid residue bonded to the C(2)O carbonyl of the thymine ring, as observed in the solid state for Z-Aib-AlaT-Aib-O^{t}Bu.

Considering the spectrum of Z-Aib-AlaT-Aib-O^{t}Bu we observe that, at the lowest concentration ($1 \cdot 10^{-4}$ M), at which intermolecular H-bonding is less important, apart from the band at 3388 cm^{-1}, due to the N(3)H of thymine, there are two bands with similar positions and relative intensities compared to the bands observed in the spectrum of the N(3)-alkylated analogue. It can be assumed therefore that the peptide is also folded in a β-turn and that the αNH of $AlaT^2$ is likely involved in an intraresidue H-bond, at least at $1 \cdot 10^{-4}$ M concentration. The strong dilution effects observed point to strong intermolecular interactions and the fact that they are not present in Z-Aib-AlaT^{All}-Aib-O^{t}Bu suggests that the free N(3)H is crucial for their strength.

The same kind of considerations supporting a folded conformation can be applied also to the spectra of the acetylated tripeptide (Fig. 3.6, left) and to the spectrum of the tetrapeptide (Fig. 3.6, right). In the case of the latter, an observation can be done that supports the presence of the intraresidue H-bond in the nucleo-*tri*peptides. At the lowest concentration ($1 \cdot 10^{-4}$ M) the ratio of the band of the 'free' NHs relative to the bound NHs is bigger than in the tripeptides, while for 'normal' Aib-rich peptides the band of the bound NHs increases relative to the band of 'free' NHs as the chain grows. This is because in 'normal' protected 3_{10}-helical peptides the first two residues are 'free' and the other are involved in the $i \leftarrow i+3$ intramolecular H-bond network. The peptides being synthesized from the C to the N terminal, each new residue is added in position 1 and the former residue in position 2, with 'free' NH, is shifted to the position 3, where its NH forms an H-bond with the CO of the protecting group. The net result is therefore that one NH more is H-bonded.

In the case of the nucleo-tetrapeptide the residue that passes from the 'unbound' position 2 to the 'bound' position 3 is the AlaT residue. If the αNH of this residue was free in the *tri*peptide and bound in the *tetra*peptide, the normal increase in the H-bonded/solvated NH ratio would be observed. On the contrary, if the αNH was already H-bonded in the *tri*peptide, the number of H-bonded residue in the *tetra*peptide would not change, while a second solvated NH would be present, allowing a relative increase of the 'free' NH band. This is exactly what can be observed.

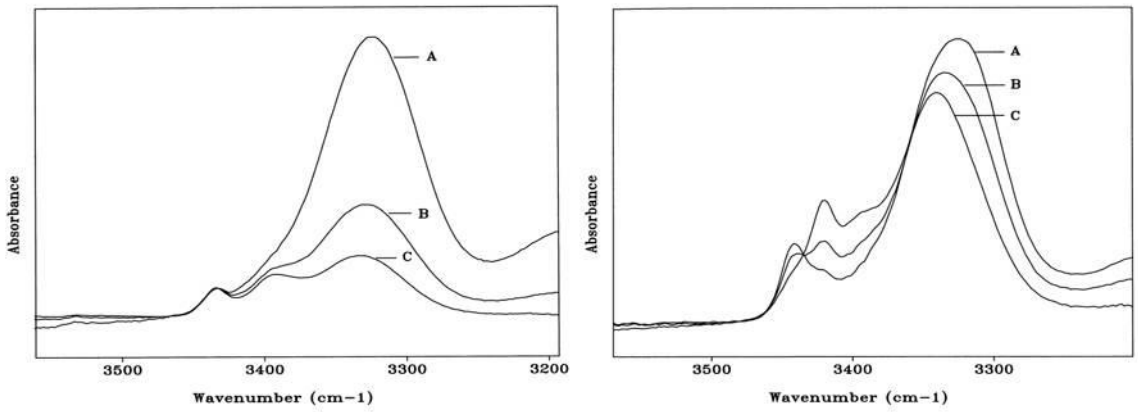


Fig. 3.6. Amide A region of the IR absorption spectra of the nucleopeptides Ac-Aib-AlaT-Aib-O^{t}Bu (left) and Z-Aib-Aib-AlaT-Aib-O^{t}Bu (right), in $CDCl_3$ solution at $1\cdot10^{-2}$ (A), $1\cdot10^{-3}$ (B), $1\cdot10^{-4}$ M (C) concentration.

The *intraresidue* H-bond seems not to be present in the shorter nucleopeptides, although weak, broad shoulder at about 3350 cm^{-1} is found in the spectrum of Z-AlaT-Aib-O^{t}Bu. Similarly, the shorter nucleopeptides display a reduced tendency to intermolecular interactions. It is intriguing to speculate that the presence of the β-turn might favour the formation of the intraresidue H-bond, and that in turn the frozen conformation of the nucleobases enhances their tendency to interact with each other. However, less structured nucleo-tripeptides being not available, it is not possible to verify this hypothesis.

3.2.1.2 Nucleopeptides containing other bases

Since thymine and cytosine, being both pyrimidyl nucleobases, have close similarities in their structure, it can be useful to consider firstly the cytosine-based nucleopeptides.

Table 3.4 presents the stretching frequencies in the amide A and amide I regions for cytosine containing nucleopeptides in $CDCl_3$ solution at $1\cdot10^{-3}$ M concentration. The amide

A region of the IR absorption spectra of the cytosine containing nucleoamino ester and of the protected nucleo-di- and -tripeptides in $CDCl_3$ solution at $1 \cdot 10^{-3}$ M concentration is shown in Fig. 3.7, left.

Table 3.4. IR absorption frequencies (cm^{-1}) in the specified intervals for cytosine containing nucleopeptides in deuterated chloroform at $1 \cdot 10^{-3}$ M concentration.

Peptide	3600-3200 cm^{-1}	1800-1600 cm^{-1}
Boc-AlaC^{Z}- OMe	3456, 3402	1749, 1709, 1666,1628,1600
Boc-AlaC^{Z}-Aib-OMe	3402, 3325	1746,1710, 1660,1626,1600,
Boc-Aib AlaC^{Z}-Aib-OMe	3436, 3402, 3316	1745, 1704, 1682, 1662
Boc-Aib-AlaC^{H}-Aib-OMe	3532, 3506[a], 3438, 3416, 3376	1736, 1700, 1678[a], 1654, 1618, 1598
All-C^{Z}	3402	1751, 1728, 1710, 1664, 1628

Notes. [a]: shoulder; underlined values refer to weak (...) or strong (___) bands; values not underlined are referred to bands of intermediate intensity.

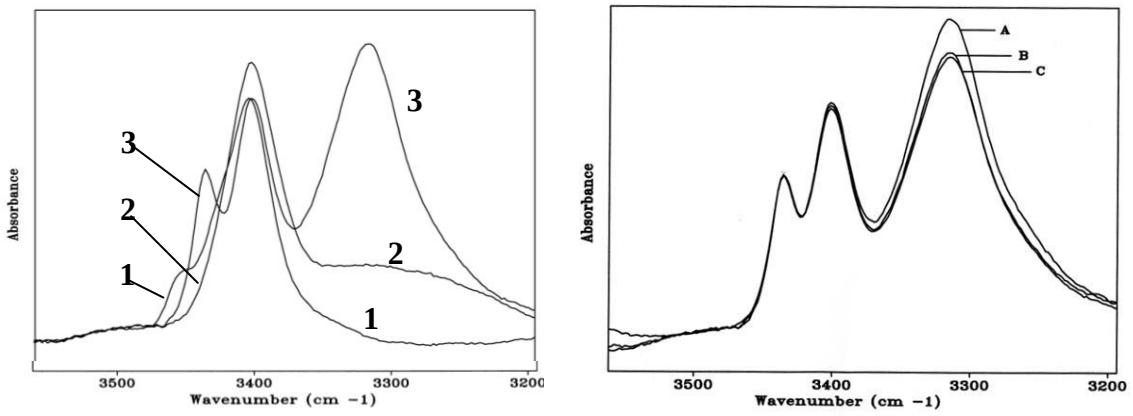


Fig. 3.7. Left: amide A region of the IR absorption spectra of the nucleoamino acid ester Boc-AlaC^{Z}-OMe (1), and of the nucleopeptides Boc-AlaC^{Z}-Aib-OMe (2) and Boc-Aib-AlaC^{Z}-Aib-OMe (3), in $CDCl_3$ solution at $1 \cdot 10^{-3}$ M concentration. Right: the same portion of the IR absorption spectra of Boc-Aib-AlaC^{Z}-Aib-OMe in $CDCl_3$ solution at $1 \cdot 10^{-2}$ (A), $1 \cdot 10^{-3}$ (B), $1 \cdot 10^{-4}$ M (C) concentration.

It can be observed that no bands due to bound NHs are present in the nucleoamino acid ester spectra, and that only a weak and broad band around 3325 cm^{-1} is present in the case of the dipeptide. On the other side, a strong band at low wavenumber (3316 cm^{-1}) is present in the tripeptide spectrum. Since dilution effects are week for this peptides, particularly going from $1 \cdot 10^{-3}$ to $1 \cdot 10^{-4}$ M (see Fig.3.7, right, as an example), the H-bonded

band observed in the di- and tripeptide spectrum is most likely due to *intra*molecular interactions.

A strong band at 3402 cm^{-1} is found in the spectra of all cytosine containing nucleopeptides with protected base, but not in the spectrum of Boc-Aib-AlaC^{H}-Aib-OMe: this suggests the attribution of the band to the stretching of the N^4H-Z anilide, which is confirmed by comparison with the spectrum of the N(1)-alkylated nucleobase All-C^{Z}, synthesized on purpose from H-C^{Z} (Fig.3.8), which has only one NH and displays one strong IR band exactly at 3402 cm^{-1}.

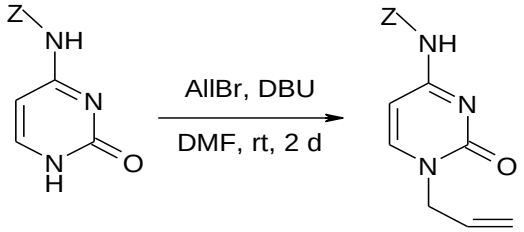


Fig. 3.8. Synthesis of All-C^{Z} by alkylation of H-C^{Z}.

The interpretation of the nucleo-tripeptide spectrum, keeping in mind what learned in the case of the thymine-based analogue, is straightforward: there two H-bonded NHs, one is again the NH of Aib^3 closing a C_{10} structure with the Boc CO group, and the other is the αNH of AlaC, bound to the C(2)O of cytosine.

The weak and broad band found in the case of the dipepide could be therefore due to the same intraresidue H-bond which is less stable than in the tripeptide because of the absence of backbone structuration. However, this can not be positively demonstrated.

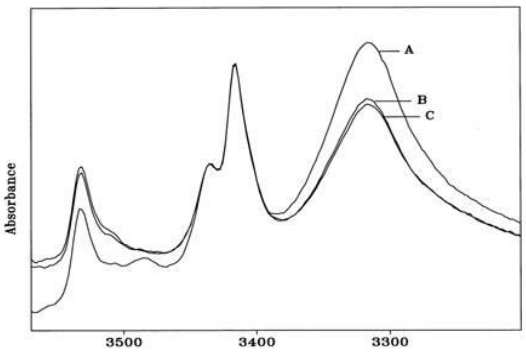


Fig. 3.9. Amide A region of the IR absorption spectra of Boc-Aib-AlaC^{H}-Aib-OMe in $CDCl_3$ solution at $1{\cdot}10^{-2}$ (A), $1{\cdot}10^{-3}$ (B), $1{\cdot}10^{-4}$ M (C) concentration. The bands around 3500 cm^{-1} and at 3416 cm^{-1} are most likely due to the free N^4H_2 of cytosine.

Considering the amide A region of the spectrum of the tripeptide with unprotected cytosine (Fig. 3.9), it appears that this derivative has a significantly stronger tendency to aggregation. As in the case of the thymine-based nucleopeptides, this finding points towards the crucial role of the nucleobase in promoting intermolecular interactions.

Table 3.5 presents the stretching frequencies in the amide A and amide I regions for adenine and guanine containing nucleopeptides at $1 \cdot 10^{-3}$ M concentration.

Table 3.5. IR absorption frequencies (cm^{-1}) in the specified intervals for adenine and guanine containing nucleopeptides in deuterated chloroform at $1 \cdot 10^{-3}$ M concentration.

Peptide	3600-3200 cm^{-1}	1800-1600 cm^{-1}
Boc-AlaA^{Z}-Aib-OMe	3420[a], 3402, 3382[a]	1736, 1710, 1678, 1612
Boc-Aib-AlaA^{Z}-Aib-OMe	3436, 3400, 3330	1736, 1704, 1680[a], 1662, 1638, 1612
Boc-Aib-AlaA^{H}-Aib-OMe	3526, 3504[a], 3482, 3440, 3414, 3330	1736, 1704, 1680[a], 1662, 1630, 1600
Z-Aib-AlaA^{H}-Aib-O^{t}Bu	3448[a], 3430, 3374[a], 3344	1714, 1682, 1660[a], 1650[a], 1626, 1600
Boc-AlaG^{OH}-Aib-OMe	3512, 3422[a], 3390, 3326	1736[a], 1712, 1690, 1668[a], 1628, 1600
Boc-Aib-AlaG^{OH}-Aib-OMe(*)	3304	1738, 1702[a], 1684, 1648, 1628, 1600

Notes. [a]: shoulder; underlined values refer to weak (. .) or strong (__) bands; values not underlined are referred to bands of intermediate intensity. (*): the real concentration of this compound was lower than 1·10-3 M due to reduced solubility in $CDCl_3$.

The amide A region of the spectra of the adenine-based nucleo-tripeptides Boc-Aib-AlaA^{Z}-Aib-OMe and Boc-Aib-AlaA^{H}-Aib-OMe in $CDCl_3$ solution is represented in Fig. 3.10, left and right, respectively.

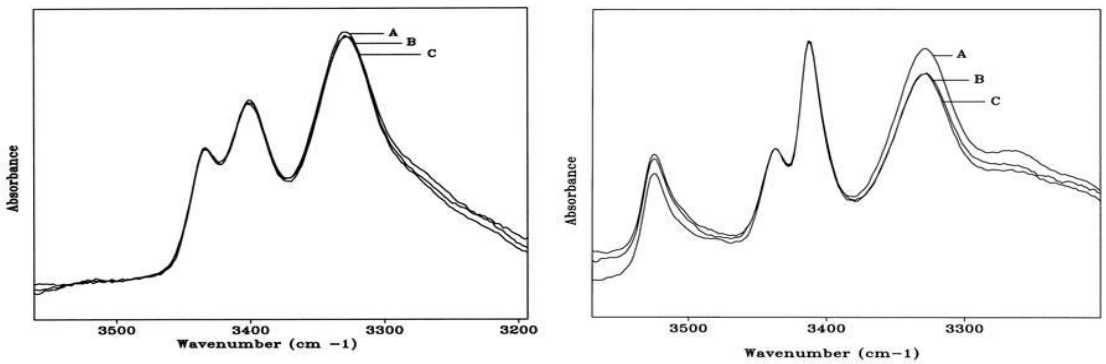


Fig. 3.10. Amide A region of the IR absorption spectra of Boc-Aib-AlaA^{Z}-Aib-OMe (left) and of Boc-Aib-AlaA^{H}-Aib-OMe (right) in $CDCl_3$ solution at $1 \cdot 10^{-2}$ (A), $1 \cdot 10^{-3}$ (B), $1 \cdot 10^{-4}$ M (C) concentration. The bands over 3500 cm^{-1} and at 3430 cm^{-1} in the right spectrum are most likely due to the free $N^{6}H_2$ of adenine.

It can be observed that both spectra are quite similar to the spectra of the corresponding cytosine based nucleopeptides (refer to Fig. 3.7, right and Fig. 3.9).

Considering also the similarity in the protecting groups and in the nucleobase NHs it seems likely that the same considerations proposed for the cyosine-based nucleopeptides hold also for the adenine-based analogues. This would mean not only that such tripeptides are folded in a β-turn, but also that the αNH of AlaA is involved in a intraresidue H-bond, probably with the N(3) of the purine ring (Fig. 3.10).

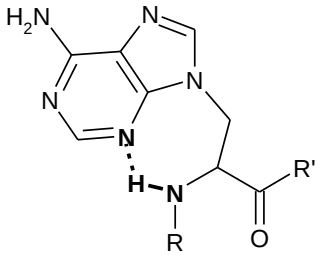


Fig. 3.10. Formation of an intraresidue H-bond in adenine-based tripeptides between the αNH of the AlaA residue and the N(3) of the nucleobase, as suggested by the IR spectra. Atoms involved in the H-bond are in bold font. R, R' represent the rest of the tripeptide.

The effects of dilution in the spectrum of Boc-Aib-AlaA^{H}-Aib-OMe are smaller than in the spectrum of the corresponding cytosine-based analogue, suggesting that nucleobase intermolecular interactions mediated by free adenine are weaker than cytosine mediated intermolecular interactions.

The analysis of the spectra of the guanine derivatives is more difficult because of their reduced solubility and their tendency to aggregation. The spectrum of the guanine-based tripeptide is very noisy (not shown) and therefore not useful for a discussion. The amide A portion of the spectrum of Z-AlaG^{OH}-Aib-OMe in $CDCl_3$ solution at $1 \cdot 10^{-3}$ M (Fig. 3.11) is better, although still noisy.

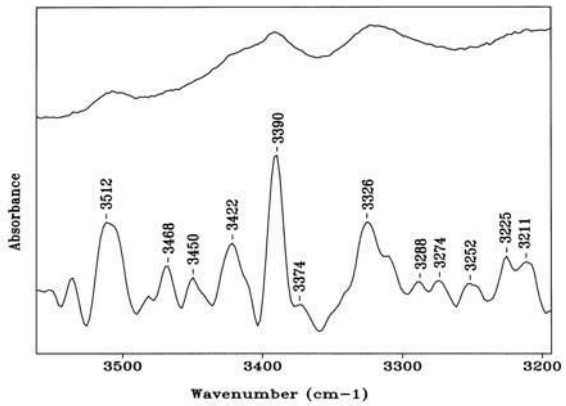


Fig. 3.11. Amide A region of the IR absorption spectrum of the guanine-based nucleodipeptide Z-AlaG^{OH}-Aib-OMe in $CDCl_3$ at $1 \cdot 10^{-3}$ M concentration and inverse second derivative elaboration.

In this spectrum a band of strongly H-bonded NHs seems to be present at 3326 cm^{-1}, together with bands due to the N^2H_2 of guanine (at 3512 cm^{-1}), to solvated peptide NHs (3422 cm^{-1}) and probably to to the N(1)H lactame (at 3390 cm^{-1}).

However, given the strong aggregation propensity of this derivative, it is possible that the first band is due exclusively to *inter*molecular interactions mediated by the free guanine and not to *intra*molecular H-bonds.

3.2.2 Lysine containing nucleopeptides

Table 3.6 presents the the stretching frequencies in the amide A and amide I regions in $CDCl_3$ solution at $1 \cdot 10^{-3}$ M concentration for the two series of thymine- and cytosine-based lysine containing nucleopeptides and for the protected dipeptide amides Z-Aib-Lys(Boc)-NH_2 and Boc-Aib-Lys(Z)-NH_2, used for the preparation of the two series.

Table 3.6. IR absorption frequencies (cm^{-1}) in the specified intervals for lysine containing peptides in deuterated chloroform at $1 \cdot 10^{-3}$ M concentration.

Peptide	3600-3200 cm^{-1}	1800-1600 cm^{-1}
Z-Aib-Lys(Boc)-NH_2	3524, 3484, 3456, 3426, 3410[a], 3360	1710, 1678, 1646, 1628, 1594
Z-AlaT-Aib-Lys(Boc)-NH_2	3524, 3484, 3454, 3414, 3392[a], 3358	1708[a], 1684. 1646[a], 1622, 1600
Z-Aib-AlaT-Aib-Lys(Boc)-NH_2	3483, 3456[a], 3428, 3394, 3360[a], 3334, 3306[a]	1708[a], 1686, 1662[a], 1624, 1600
Boc-Aib-Lys(Z)-NH_2	3522, 3480, 3450, 3424, 3354	1710, 1676, 1638, 1624, 1594
Boc-AlaC^Z-Aib-Lys(Z)-NH_2	3522, 3486, 3452, 3422[a], 3408, 3360, 3290	1751, 1720[a], 1670, 1626, 1596,
Boc-Aib-AlaC^Z-Aib-Lys(Z)-NH_2	3522, 3484, 3454[a], 3432, 3402, 3346, 3286	1755, 1720[a], 1662, 1624, 1598

Notes. [a]: shoulder; underlined values refer to weak (. . .) or strong (___) bands; values not underlined are referred to bands of intermediate intensity.

The amide A portions of the IR absorption spectra of the peptides Boc-Aib-Lys(Z)-NH_2, Boc-AlaC^Z-Aib-Lys(Z)-NH_2, Boc-Aib-AlaC^Z-Aib-Lys(Z)-NH_2 in $CDCl_3$ solution at $1 \cdot 10^{-3}$ M concentration in the amide A region are presented in Fig. 3.12. Given the presence of side chain (the εNH of Lys, and the N^4H of cytosine in the case of the nucleopeptides) and

of two C-terminal NHs, the spectra present several bands. It is therefore useful to compare the most well defined bands in the different spectra to perform a tentative attribution.

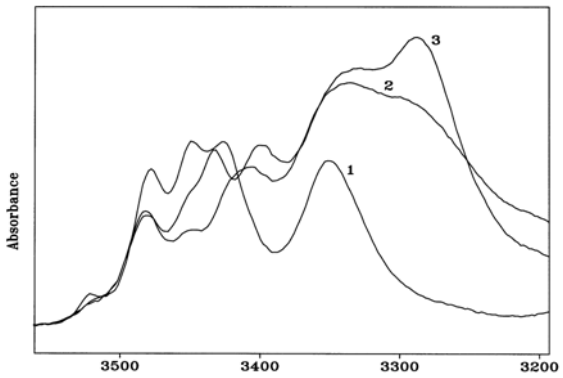


Fig. 3.11 Amide A region of the IR absorption spectra of Boc-Aib-Lys(Z)-NH_2(A), Boc-AlaC^Z-Aib-Lys(Z)-NH_2 (B), Boc-Aib-AlaC^Z-Aib-Lys(Z)-NH_2(C), in $CDCl_3$ solution at $1\cdot10^{-3}$M concentration.

In the solvated NH region one band is observed in all spectra at 3480-3486 cm^{-1}, as well as a weak band at 3522 cm^{-1}. In the H-bonded NH region, a band is present in all derivatives at 3346-3354 cm^{-1}, moving towards shorter wavelengths as the length of the peptide increases. One band is present only in the nucleopeptide spectra with similar intensity at 3402-3408 cm^{-1}; a second band in the region of H-bonded NHs is found at 3286-3290 cm^{-1} for such peptides, with higher intensity for the longer nucleopeptide.

Moreover, two bands in the solvated NH region are present in all spectra, one at 3450-3454 cm^{-1}, the other at 3422-3432 cm^{-1}, even if in the spectrum of Boc-Aib-Lys(Z)-NH_2 these bands are strong and of the same intensity, while in the nucleopeptide spectra they are weaker, in particular the higher frequency band appears just as a shoulder.

By taking also into account what stated about the other cytosine containing nucleopeptides, it can be deduced that the band at 3402-3408 cm^{-1}, present only in the nucleopeptides, is most likely due to the N^4H of cytosine. Similarly it makes sense that the εNH of lysine and one of the two C-terminal NHs, which are solvent exposed in all peptides, contribute to the bands at 3522 and 3486 cm^{-1}, present in all peptides with similar intensity.

By considering the amide A portions of the spectra of the three derivatives at $1\cdot10^{-2}$, $1\cdot10^{-3}$, $1\cdot10^{-4}$ M concentration (Fig. 3.12), it is possible to see that dilution effects are evident from $1\cdot10^{-2}$ M to $1\cdot10^{-3}$ M, particularly in the case of the nucleopeptides, highlighting a significant contribution of *inter*molecular H-bonding. However dilution effects are weaker from $1\cdot10^{-3}$ M to $1\cdot10^{-4}$ M, or at most moderate for the nucleopeptide

Boc-AlaC^{Z}-Aib-Lys(Z)-NH_2; given that the global shape of the spectra is almost unchanged at $1\cdot10^{-4}$ M concentration compared to $1\cdot10^{-3}$ M concentration, it can be deduced that the bands of bonded NHs have a prevalent contribution from *intra*molecular H-bonding interactions in the three peptides.

If the band around 3350 cm^{-1} present in all derivatives is mostly due to intramolecularly bonded NHs, this could mean that the three peptides are folded in one β-turn or in helices. Indeed, in the case of the the shortest derivative, a dipeptide amide, there are three consecutive peptide groups and a *i*←*i*+3 H-bond enclosing a β-turn can be formed between one C-terminal NH and the CO of the Boc protecting group. The longer peptides can form two or three of such H-bonds, involving also the αNH of Lys and Aib[3], respectively. This would explain also the increase in the intensity of the band and its shift towards shorter wavenumbers, as consecutive β-turns stabilize each other.

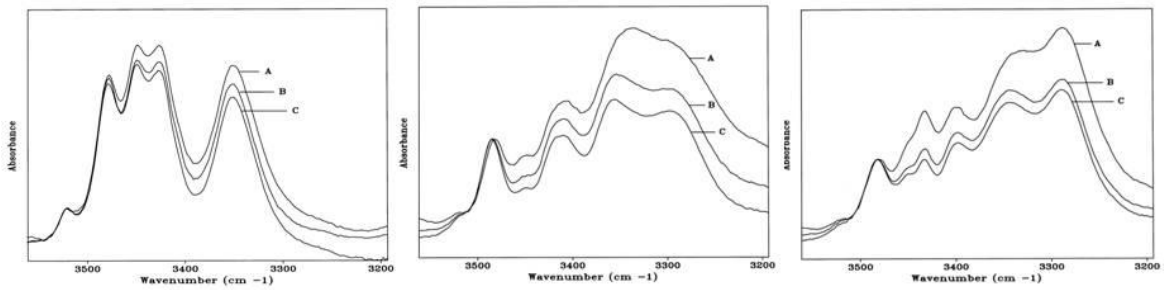


Fig. 3.12. Amide A region of the IR absorption spectra of the peptides Boc-Aib-Lys(Z)-NH_2 (left), Boc-AlaC^{Z}-Aib-Lys(Z)-NH_2 (middle), Boc-Aib-AlaC^{Z}-Aib-Lys(Z)-NH_2 (right) in $CDCl_3$ solution at $1\cdot10^{-2}$ (A), $1\cdot10^{-3}$ (B), $1\cdot10^{-4}$ M (C) concentration.

The presence of a second band in the region of H-bonded NHs for the nucleopeptides could be done to the presence of the intraresidue H-bond in the case of the αNH of AlaC. Similarly, the decrease in the relative intensity of the bands at 3420-3450 cm^{-1} could be explained by the presence of only one solvated αNH in the nucleopeptides (the NH of Aib[2] in Boc-AlaC^{Z}-*Aib*-Lys(Z)-NH_2 and the NH of Aib[1] in Boc-*Aib*-AlaC^{Z}-Aib-Lys(Z)-NH_2) against two in Boc-Aib-Lys-NH_2, since in the nucleopeptides one of the two αNHs which can not form a *i*←*i*+3 H-bond is stabilized by an intraresidue H-bond.

Fig. 3.13, left, presents the amide A region of the spectra of the peptides Z-Aib-Lys(Boc)-NH_2, Z-AlaT-Aib-Lys(Boc)-NH_2, and Z-Aib-AlaT-Aib-Lys(Boc)-NH_2 in deuterated chloroform solution at $1\cdot10^{-3}$ M concentration. The spectrum of the dipeptide amide is similar to the spectrum of the N^{α}-Boc-protected derivative already considered; the spectrum of the nucleo-tripeptide amide Z-AlaT-Aib-Lys(Boc)-NH_2 has a similar shape,

even if the H-bonded NH band is weaker compared to the cytosine-based analogue and there is a smaller contribution from intermolecularly bonded NHs (refer to Fig. 3.13, right, displaying the amide A portion of the spectrum of this derivative at different concentrations). The structuration of this peptide seems therefore to be less important than the structuration observed for its cytosine-containing analogues.

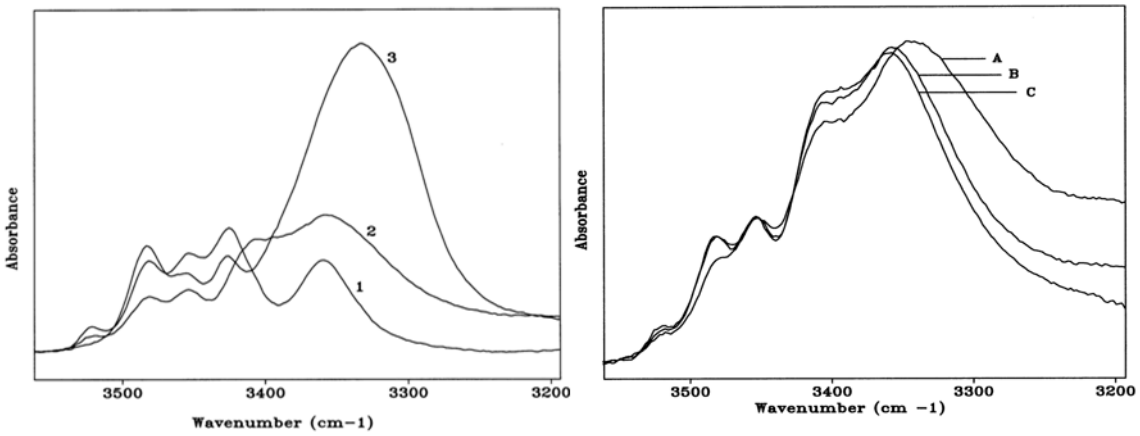


Fig. 3.13. Left: amide A region of the IR absorption spectra of the peptides Z-Aib-Lys(Boc)-NH_2 (1), Z-AlaT-Aib-Lys(Boc)-NH_2 (2), Z-Aib-AlaT-Aib-Lys(Boc)-NH_2 (3) in $CDCl_3$ solution at $1 \cdot 10^{-3}$ M concentration. Right: amide A region of the IR absorption spectra of the peptide Z-AlaT-Aib-Lys(Boc)-NH_2 in $CDCl_3$ solution at $1 \cdot 10^{-2}$ (A), $1 \cdot 10^{-3}$ (B), $1 \cdot 10^{-4}$ M (C) concentration.

In the case of the tetra-nucleopeptide, however, a very strong increase in the H-bonded NH band is observed. The increase is most likely due to strong intermolecular interactions, favoured by the presence of the free base. The spectrum of this nucleopeptide at various concentrations is shown in Fig. 3.14.

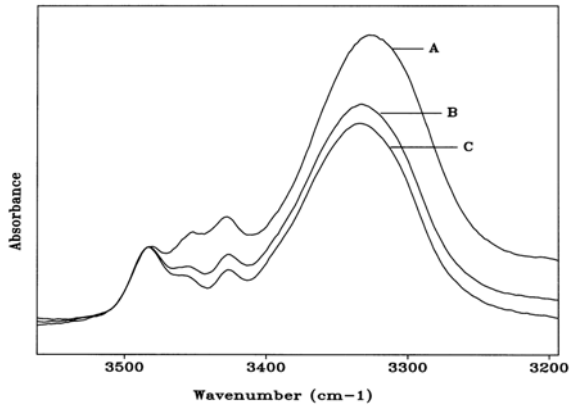


Fig. 3.14. Amide A region of the IR absorption spectrum of the nucleopeptide Z-Aib-AlaT-Aib-Lys(Boc)-NH_2 in $CDCl_3$ solution at $1 \cdot 10^{-2}$ (A), $1 \cdot 10^{-3}$ (B), $1 \cdot 10^{-4}$ M (C) concentration.

Indeed, dilution effects are evident both from $1 \cdot 10^{-2}$ to $1 \cdot 10^{-3}$ M and they are present, even if weaker, also from $1 \cdot 10^{-3}$ to $1 \cdot 10^{-4}$ M. However, even at $1 \cdot 10^{-4}$ M, the H-bonded NH band is very strong and the spectral shape does not change much by dilution. Therefore, even if strong intermolecular interactions exist, *intra*molecular H-bonds are indeed stronger in this nucleopeptide than in its shorter analogue and in the cytosine-based nucleopeptides, possibly due to a particularly stable helical structuration.

3.2.3 Nucleopeptides containing two nucleobases

Table 3.7 presents the the stretching frequencies in the amide A and amide I regions for the nucleopeptides Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu and Z-(Aib-AlaT-Aib)$_2$-Lys(Boc)-NH$_2$ in $CDCl_3$ solution at $1 \cdot 10^{-3}$ M concentration, while Fig. 3.15 presents the amide A region of the spectra of the two nucleopeptides in $CDCl_3$ solution at various concentration, ranging from $1 \cdot 10^{-2}$ M to $5 \cdot 10^{-5}$ M in the case of the hexapeptide.

In both spectra strong dilution effects are observed, even from 10^{-3} M to 10^{-4} M, showing remarkable contributions from *inter*molecular H-bonds. In particular, the shape of the spectrum of Z-(Aib-AlaT-Aib)$_2$-Lys(Boc)-NH$_2$ and its behaviour on dilution are quite similar to those observed in the case of Z-Aib-AlaT-Aib-Lys(Boc)-NH$_2$. However at the lowest concentration examined (0.1 or 0.05 mM, respectively), the strong H-bonded NH bands observed are largely due to *intra*molecularly bonded NHs. Taking also into account the low frequency of the bands (3324-3342 cm^{-1}), it can be assumed that the peptides are folded, and that they probably adopt helical structures. The strong tendency to intermolecular interaction is most likely due to the presence of two free nucleobases in such nucleopeptides.

Table 3.7. IR absorption frequencies (cm^{-1}) in the specified intervals for the protected nucleopeptides containing two thymines in deuterated chloroform at $1 \cdot 10^{-3}$ M concentration.

Peptide	3600-3200 cm^{-1}	1800-1600 cm^{-1}
Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu	3430, 3404[a], 3360[a], 3324	1726[a], 1700, 1672, 1650, 1624, 1600
Z-(Aib-AlaT-Aib)$_2$-Lys(Boc)-NH$_2$	3532, 3524, 3484, 3456, 3428, 3392[a], 3358[a], 3342, 3296[a]	1706[a], 1684, 1646[a], 1622, 1600

Notes. [a] : shoulder; underlined values refer to weak (...) or strong (__) bands; values not underlined are referred to bands of intermediate intensity.

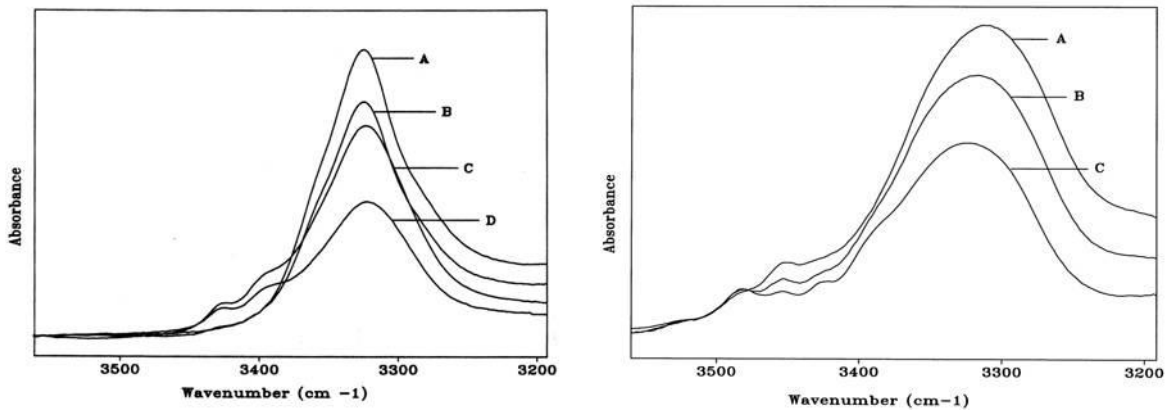


Fig. 3.15. Amide A region of the IR absorption spectra of the nucleopeptides Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu (left) and Z-(Aib-AlaT-Aib)$_2$-Lys(Boc)-NH$_2$ (right) in $CDCl_3$ solution at $1\cdot10^{-2}$ (A), $1\cdot10^{-3}$ (B), $1\cdot10^{-4}$ (C) and $5\cdot10^{-5}$ M (D) concentration.

From the analysis of the amide I portion of the IR spectra of the nucleopeptides less conformational information can be obtained. Indeed, in longer helical peptides, the maximum of the amide I absorption band is located around 1660 cm^{-1} (1650-1658 cm^{-1} for α-helical structures, 1662-1666 cm^{-1} for 3_{10}- helical structures),[131,132] while type III β-turns in Aib-rich peptides generally display maxima around 1646 and 1684 cm^{-1}.[133]

In the case of Z-(Aib-AlaT-Aib)$_2$-Lys(Boc)-NH$_2$ band maxima are found exactly at 1646 and 1684 cm^{-1}, however this is not the case of most of the peptides for which the presence of a β-turn was inferred considering the bands in the amide A region. The absence of amide I maxima around 1660 cm^{-1} even in the longest peptides is probably due to their insufficient length in order to form a fully developed 'canonical' 3_{10}-helix.

To summarise, the most important features derived from the IR analysis, all of which are in good agreement with the hypothesis at the base of this work are:

(I) Nucleo-tripeptides and longer nucleopeptides are folded, presenting intramolecular H-bonded NHs due to *i*←*i*+3 H-bonds enclosing β-turns.

(II) In thymine-, cytosine-, and adenine-based nucleo-tripeptides an intraresidue H-bond between the αNH of the nucleoamino acid and an acceptor group in the nucleobase is present, forming a 7-membered pseudoring that strongly reduces the side-chain conformational flexibility.

(III) Nucleopeptides with unprotected nucleobases have a strong tendency to nucleobase-mediated intermolecular interactions.

3.3 NMR spectrometry

The conformational analysis of the protected nucleopeptides by ^{1}H-NMR spectrometry was performed in $CDCl_3$, the same solvent used for the IR characterization. In the case of the deprotected nucleopeptides, the NMR analysis was performed in d_6-DMSO.

Working in $CDCl_3$ solution allows one to obtain useful information on the H-bonding stabilization of peptide amide NHs, which is easily related to the peptide conformation. This can be done by adding increasing amounts of the good H-bond acceptor dimethylsulfoxide (DMSO)[134] to a peptide in $CDCl_3$ solution. Indeed, the amide protons which are exposed to the solvent are stabilized by H-bond formation with DMSO molecules and this causes a downfield shift of their signals. On the other side, amide protons already engaged in stable H-bonds are solvent-shielded and their chemical shift is little affected by the variation in the solvent composition.

Given that even by the addition of small amounts of DMSO its concentration is much higher than the concentration of the peptide, *inter*molecular H-bonds between solute molecules are generally easily replaced by intermolecular H-bonds between solute and DMSO by sheer concentration effects, unless peptide-peptide interactions are remarkably strong. On the other side, *intra*molecular H-bonds are concentration insensitive and therefore generally little perturbed by DMSO addition. Consequently, it is expected that NH protons whose NMR signals do not change significantly by DMSO addition are stabilized by *intra*molecular H-bonds, unless there are reasons to believe that very strong intermolecular interactions are present.

If the NH signals in the monodimensional ^{1}H-NMR spectrum have been identified and assigned, it is therefore possible, by using the method described above, to know which NH groups are involved in *intra*molecular H-bonds. This is an advantage compared to the analysis by IR spectroscopy, which allows an estimation of the fraction of intramolecularly H-bonded NHs, but not their identification (see Section 3.2).

The assignment of 1D ^{1}H-NMR spectra of small peptides can be tentatively be performed by inspection and comparison with already assigned spectra of similar compounds. For a complete and correct assignment, however, the best way is to rely on the information obtained by two-dimensional (2D) spectra.

Three kinds of 2D NMR experiments have been performed on the nucleopeptides studied by NMR, each offering different and complementary information. COSY

experiments[135] highlight direct scalar coupling relations between nuclei (protons in case of 2D-^{1}H NMR) generally less than three chemical bonds apart. TOCSY experiments[136] allow the identification of spin systems (i.e. groups of nuclei connected by scalar coupling relations). ROESY experiments[137] are instead sensitive to the spatial proximity of the nuclei.

In general, the backbone and the alkyl chain portion of the side-chain of each amino acid residue form a spin system, whose signals can be related by TOCSY experiments. ROESY experiments can then connect residues which are next to each other along the backbone, allowing the complete assignment. In folded peptides made of protein residues a specific process, called *sequential assignment*,[138] is well established. It employs the through-space connectivity along the backbone of the kind $C^{\alpha}H(i)\rightarrow NH(i+1)$ and $NH(i)\rightarrow NH(i+1)$, derived from ROESY experiments, in order to obtain the sequence of the residues forming a peptide, whose signals have been connected by TOCSY experiments. COSY experiments are useful particularly for the assignment of the resonances within single spin systems in case this is not evident from the single chemical shifts.

Once the attribution has been completed, the information about spatial proximity obtained by ROESY experiments can be usefully employed. In particular, when nuclei which are many bonds apart appear to be next to each other, additional data on the peptide conformation are obtained. This can be used also for gaining insight in the side-chain conformation, on which it is more difficult to gather information by other techniques.

Since 2D NMR experiments can be performed in every solvent, the possibility of obtaining conformational information is not restricted by solubility in organic solvents of low polarity, as in the case of the 1D NMR or of the IR analysis. Therefore this method has been employed for studying also the conformation of the deprotected nucleopeptides.

3.3.1 Protected nucleo-tripeptides

N- and C-terminally protected thymine-, cytosine- and adenine-based tripeptides have been studied. The nucleopeptide with protected cytosine Boc-Aib-AlaC^{Z}-Aib-OMe and the nucleopeptide with unprotected adenine Boc-Aib-AlaA^{H}-Aib-OMe have been considered. The latter two nucleopeptides were selected to check whether the nucleobase protection affects the conformation of the backbone. Indeed, it was shown by the analysis of their IR spectra (Section 3.2.1.2) that adenine- and cytosine-based tripeptides, both with protected

and unprotected base, had a similar behaviour. It was therefore assumed that the analysis of the two derivatives would have given information which is also valid for the two other nucleopeptides (Boc-Aib-AlaC^{H}-Aib-OMe and Boc-Aib-AlaA^{Z}-Aib-OMe).

The data gathered by IR spectroscopy on the three nucleopeptides which needed to be supported by the NMR analysis were the assumption of a folded conformation and the presence of an intraresidue H-bond involving the αNH of the nucleoamino acid residues.

Firstly, the 1D NMR spectra of the nucleo-tripeptides were completely assigned by performing 2D TOCSY and ROESY experiments. Interestingly, in all spectra the αNH protons of the nucleoamino acid residues have the highest chemical shift among the backbone NHs (δ≈ 8.5 ppm for the pyrimidyl derivatives, δ= 9 ppm for the adenine derivative), suggesting the involvement in strong hydrogen bonds.

Fig. 3.16 presents the variation of the chemical shifts of the NH protons of the three peptides by addition of increasing amounts of d_6- DMSO.

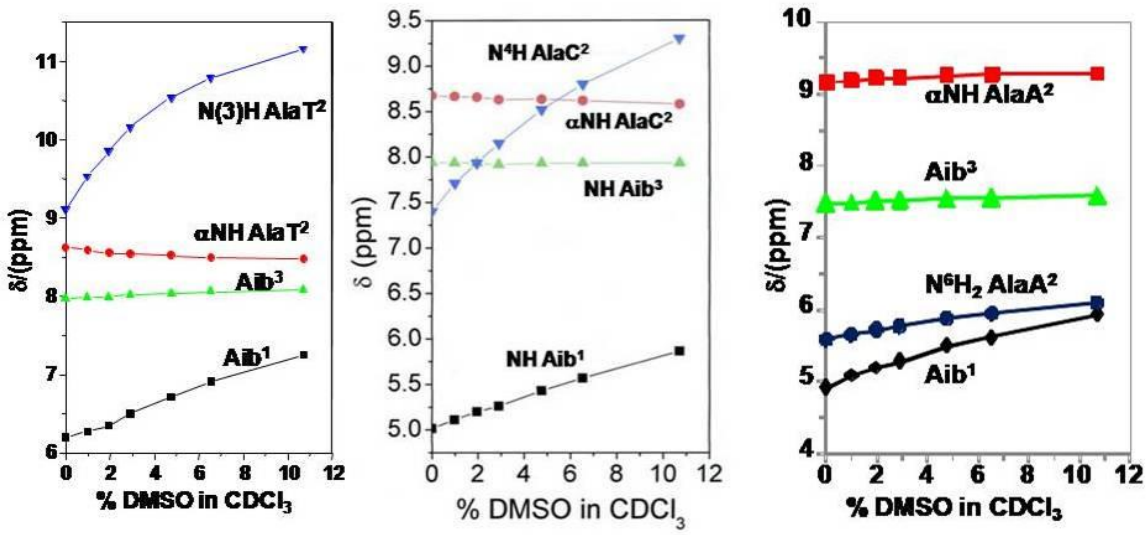


Fig. 3.16. Variation of the chemical shift of the NH protons of Z-Aib-AlaT-Aib-OtBu (left), Boc-Aib-AlaC^{Z}-Aib-OMe (middle), Boc-Aib-AlaA^{H}-Aib-OMe (right) in $CDCl_3$ solution at $1 \cdot 10^{-3}$ M concentration by addition of increasing amounts of d_6-DMSO.

All peptides have three backbone NHs (black, red and green) and one more signal (blue) due to side chain NHs (the imide N(3)H of thymine, the anilide N^{4}H-Z of the protected cytosine, the aromatic amino group N^{6}H$_2$ of the unprotected adenine) and the three graphs display parallel trends for the three nucleopeptides. Indeed, in all graphs the signals of the side chain NHs and of the NH of Aib^{1} are shifted downfields as the DMSO concentration

increases, whereas two backbone NHs are almost insensitive to DMSO addition. Therefore the latter protons are involved in intramolecular H-bonds.

The intramolecular H-bond involving Aib^3 is most likely the one enclosing a β-turn, as already explained several times in Section 3.2, while the involvement of the αNH of the nucleoamino acid residues in an intramolecular H-bond requires the presence of an intraresidue H-bond with the side-chain. The hypotheses proposed by studying the IR spectra of the nucleo-tripeptides are therefore confirmed by these experiments.

The study of ROESY spectra of the three derivatives has additionally substantiated such considerations. A portion of the ROESY spectrum of Boc-Aib-AlaC^Z-Aib-OMe is shown in Fig. 3.17, left.

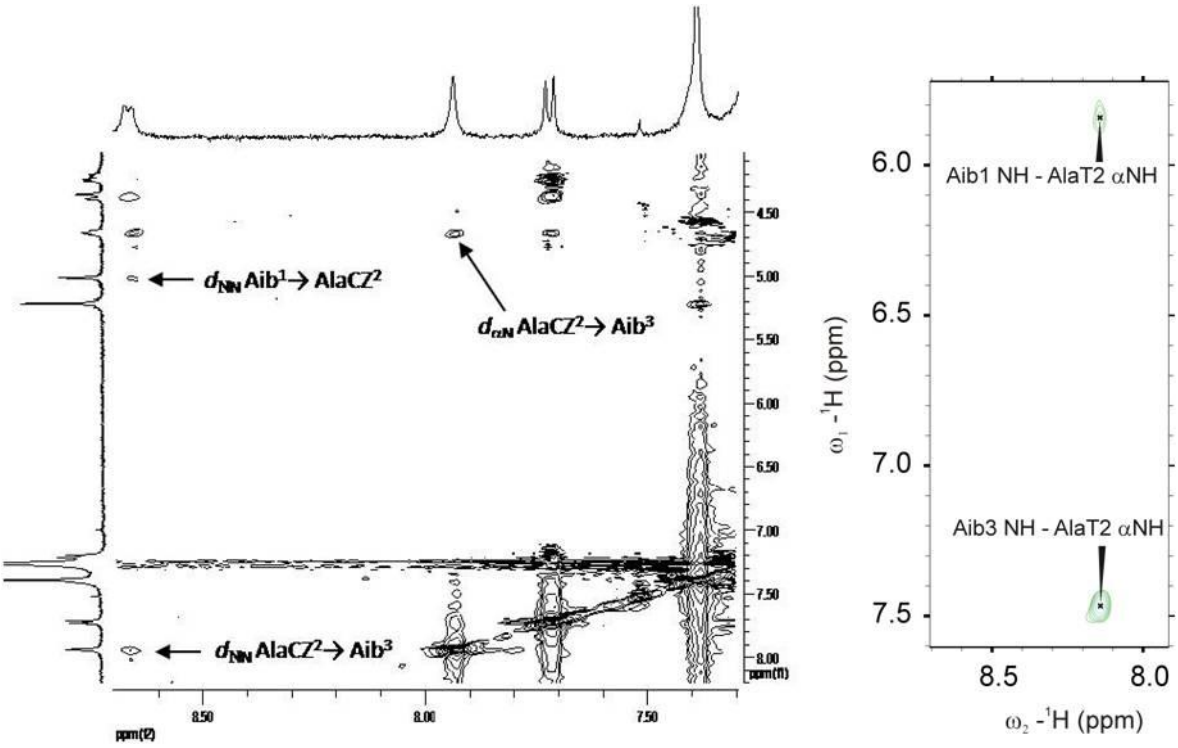


Fig. 3.17. Left: portion of the ROESY spectrum of Boc-Aib-AlaC^Z-Aib-OMe 1 mM in $CDCl_3$; right: portion of the ROESY spectrum of Z-Aib-AlaT-Aib-OtBu 1 mM in $CDCl_3$. Diagnostic cross-peaks for the adoption of folded stuctures are highlighted.

It is possible to observe the presence of both NH(*i*)→NH(*i*+1) cross-peaks and of the only possible $C^{\alpha}H(i)$→NH(*i*+1) cross-peak (since no $C^{\alpha}H$ is present on the Aib residues), all of which are diagnostic of the assumption of a folded conformation.[138] Again, this finding is in agreement with the information obtained from other experiments.

Similarly, a portion of the ROESY spectrum of Z-Aib-AlaT-Aib-OtBu is shown in Fig. 3.17, right, and both NH(*i*)→NH(*i*+1) cross-peaks are present, confirming the adoption of a folded structure also for the thymine-based nucleo-tripeptide.

Fig.3.18 displays a portion of the ROESY spectrum of Boc-Aib-AlaA^{H}-Aib-OMe, in the region where cross-peaks between αNHs and alkyl side-chains are generally found.

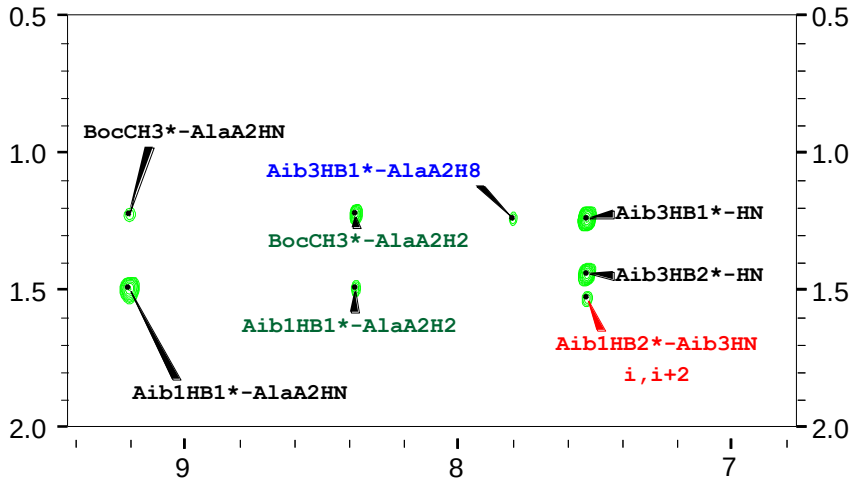


Fig. 3.18. Portion of the ROESY spectrum of the nucleopeptide Boc-Aib-AlaA^{H}-Aib-OMe 1 mM in $CDCl_3$. Cross-peaks which give evidence on the nucleobase orientation are in green (for C(2)H cross-peaks) or in blue (for C(8)H cross-peaks). The cross-peak which demonstrates the assumption of a β-turn is shown in red.

In this figure there are several cross-peaks deserving some comments.

(i) The cross-peak between the αNH of AlaA^{2} and one of the βCH$_3$ of Aib^{1} ('Aib1 HB1*-AlaA2HN) is the analogous of a C$^{\alpha}$H(*i*)→NH(*i*+1) cross-peak for a C$^{\alpha}$-tetrasubstituted residue, supporting the adoption of a folded conformation.

(ii) The cross-peak between the other βCH$_3$ of Aib^{1} and the NH of Aib^{3} (highlighted in red in Fig.3.18) is the analogous of a C$^{\alpha}$H(*i*)→NH(*i*+2) cross-peak C$^{\alpha}$-tetrasubstituted residue. This is one of the few elements allowing the discrimination between an α- and a 3_{10}-helix in longer peptides, being observed only in 3_{10}-helical conformations.[138] Of course, since in the case of a tripeptide the only folded conformation that can be formed is a β-turn, this cross-peak is just an additional, but stronger, proof of folding in a β-turn.

(iii) There are three cross peaks involving the nucleobase CH protons C(2)H and C(8)H, highlighted in green and in blue, respectively, in Fig. 3.18. One cross peaks is formed between the C(8)H and one βCH$_3$ of Aib^{3} and two cross-peaks are formed between

the C(2)H and one βCH_3 of Aib[1], on one side, and the three equivalent CH_3 of the Boc group, on the other side. From the presence of such cross peaks it is possible to deduce that the nucleobase is oriented more or less parallel to the backbone plane, with the 6-membered ring on the side of the αNH of AlaA, therefore in a suitable position for the formation of an intraresidue H-bond between the αNH and the N(3)H of adenine.

The two opposite limit orientations of the side-chain nucleobase with regard to the backbone are represented in Fig. 3.19, so that it is possible to verify that only one of the limit conformatios agrees well with the evidence obtained by the ROESY experiment.

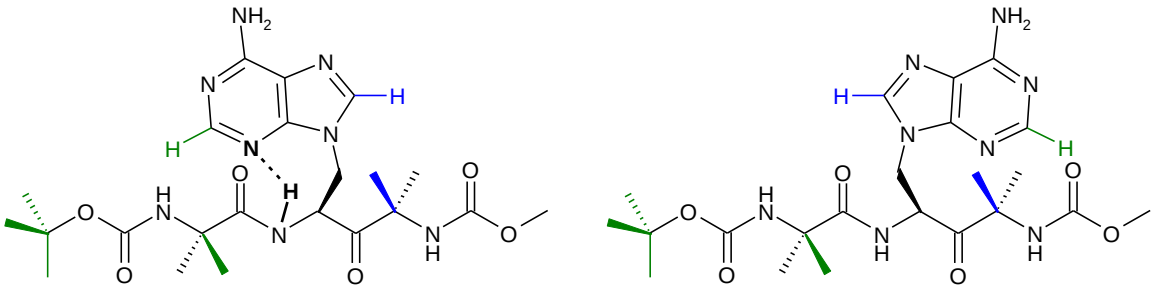


Fig. 3.19. The two possible orientations of the nucleobase when parallel to the backbone plane in Boc-Aib-AlaA^{H}-Aib-OMe. Groups connected by ROESY cross-peaks with the two adenine C(2)H and C(8)H protons have been coloured in green and blue, respectively. It appears that only the conformation presented on the left side, which allows the formation of the intraresidue H-bond (represented by a dashed line), is in agreement with the spatial proximity relations obtained by the ROESY experiment. The peptide backbone is represented in fully extended conformation for the sake of clarity.

The latter cross-peaks offer therefore a significant and independent evidence supporting the presence of the intraresidue H-bond in adenine-based nucleopeptides.

The presence of the intraresidue H-bond also in the lysine containing cytosine-based nuclepeptides, already suggested from the analysis of the IR absorption spectra, has been supported by NMR evidence in the case of one of the lysine containing cytosine-based nucleopeptides, Boc-Aib-AlaC^{Z}-Aib-Lys(Z)-NH_2. As shown in Fig. 3.20, the addition of increasing amounts of DMSO does not alter significantly the chemical shift of the αNHs of AlaC[2], Aib[3] and Lys[4]. The protons of the latter two residues can form intramolecular H-bonds if two consecutive β-turns are present; therefore this result suggests that the peptide is helical. On the contrary, the αNH of AlaC[2] can be intramolecularly H-bonded only by the formation of an intraresidue H-bond with the nucleobase.

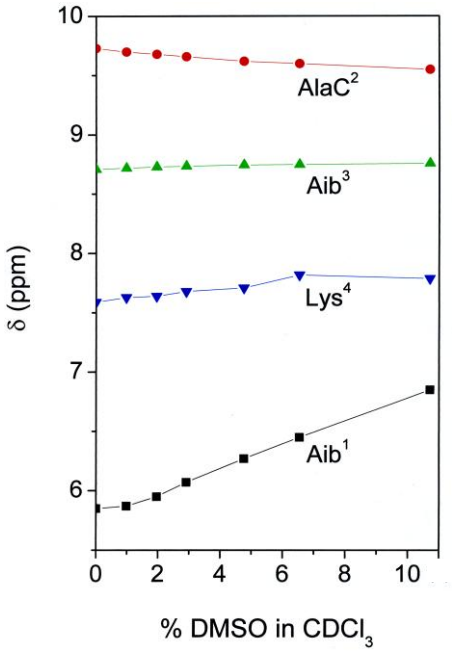


Fig. 3.20. Variation of the chemical shift of the αNH protons of Boc-Aib-AlaC^{Z}-Aib-Lys(Z)-NH_2 in $CDCl_3$ solution at $1 \cdot 10^{-3}$ M concentration by addition of increasing amounts of d_6-DMSO. For the sake of clarity only backbone NHs are plotted.

Another feature of the NMR spectrum of this nucleopeptide which deserves attention is that the resonance of the αNH of the nucleoamino acid is found at very high frequency even in pure $CDCl_3$, suggesting that the strength of the intraresidue H-bond is not only remarkable, but also greater than in the nucleo-tripeptide (δ= 9.7 and 8.7 ppm, respectively).

Also in this case, it is possible to hypothesise that there could be a relation between the increased conformational stability of the backbone and the greater strength of the intraresidue H-bond, maybe mediated by a more rigid conformation of the side-chain.

3.3.2 Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu

The 1D ^{1}H-NMR spectrum of the protected thymine-based nucleo-hexapeptide in $CDCl_3$ solution at 1 mM concentration (Fig. 3.21, trace a) presents some remarkable features. In the region where the NH signals generally appear, about twice of the expected signals are present, all with similar intensity. Moreover, three peaks are found at unusually high frequencies (δ> 11 ppm). On the contrary, the spectrum in d_6-DMSO shows the expected number of peaks. An HPLC analysis ruled out the possibility that the sample was a diastereoisomeric mixture in 1:1 proportion. It was therefore inferred that the molecule can

adopt two different conformations with the same probability in $CDCl_3$ solution and that the interconversion between the two conformations is slow, since the peaks are sharp.

Upon DMSO addition (Fig. 3.21, traces b-e) the NH signals do not shift at all; instead their intensity decreases as a new set of signals (with the expected number of peaks) grows. Part of the new peaks is shifted to lower fields as the DMSO content increases. A reasonable explanation for these results is that in neat $CDCl_3$ solution a single type of non-centrosymmetric dimeric structure is formed, whereas DMSO addition causes the dimer to dissociate into two monomeric peptides. The NH protons in the dimer are strongly H-bonded and therefore insensitive to DMSO addition, while in the monomer part of them is exposed to the solvent and shifted to lower fields when DMSO concentration increases.

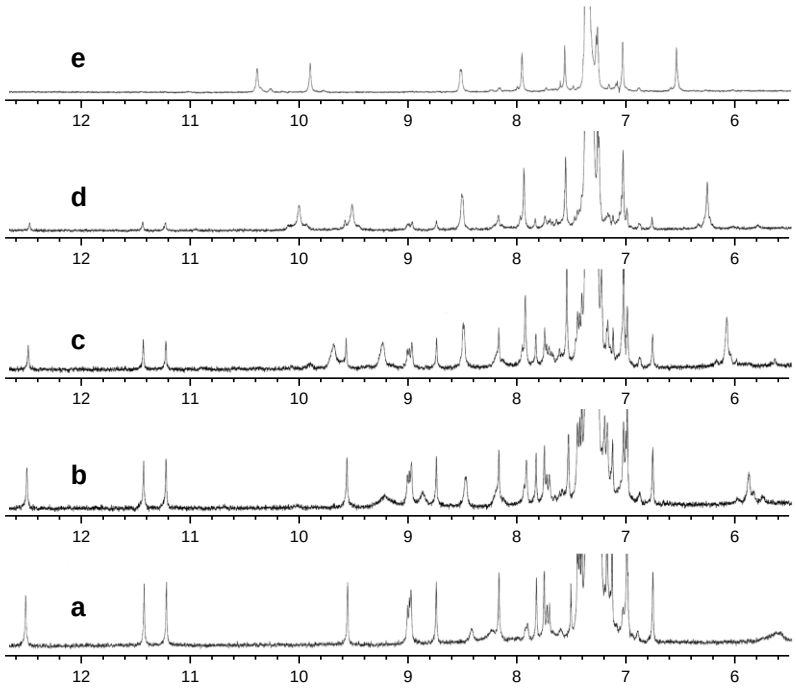


Fig. 3. 21. NH region of the NMR spectrum of the hexapeptide Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu at 1 mM concentration in $CDCl_3$ solution (**a**) and upon addition of increasing amounts of DMSO (from bottom to top): 1% (**b**), 2% (**c**), 3% (**d**), 5% (**e**).

Given the complexity of the problem (two independent molecules of the same sequential peptide rich in C^{α}-tetrasubstituted residues), the information which could be obtained from the usual 2D NMR COSY, TOCSY, ROESY experiments was not sufficient for a complete assignment. Therefore a HMBC (Heteronuclear Multiple Bond Correlation)[139] heterocorrelated 2D NMR experiment was performed. The additional data obtained enabled the assignment of all proton resonances. Given the lower sensitivity of ^{13}C-NMR

measurements compared to ^{1}H-NMR experiments, a more concentrated solution in $CDCl_3$ was used for the heteronuclear experiment (15 mM instead of 1 mM). The ^{1}H NMR spectra at the two concentrations showed no significant difference, allowing the correlation among data obtained under different experimental condition.

As an example, Fig. 3.22 shows part of a Section of the HMBC spectrum used in order to relate NHs and βCH_3 groups belonging to the same Aib residues, as well as C(6)H and C^5H_3 groups belonging to the same AlaT residues.

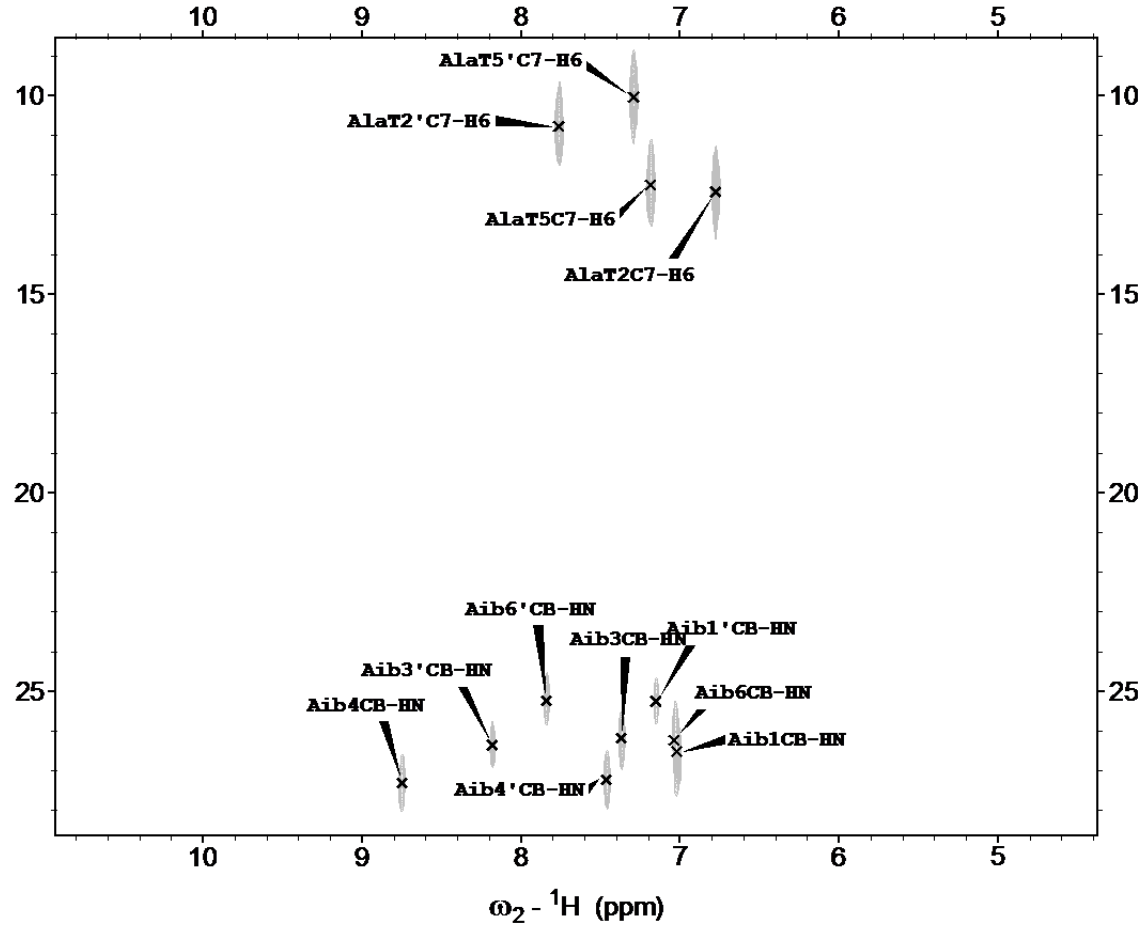


Fig. 3.22. Portion of the HMBC spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu in $CDCl_3$ solution at 15 mM concentration. Cross-peaks related to atoms of one of the two molecules in the dimer have been hyphenated for clarity. A convention for NMR analysis designs the hexocyclic C^5 of AlaT as AlaT C7.

After the assignment was completed, the hypothesis presented above was confirmed: the signals in the 1D spectrum belong to two distinct hexapeptide molecules. In addition, the three highly deshielded NHs, assigned to three N(3)Hs on three thymine rings, are likely responsible for dimer association mediated by strong H-bonds (shifting the resonances downfield).

In the ROESY spectrum many interesting cross-peaks were found, both between atoms belonging to the same molecules and between atoms belonging to the two molecules, strongly supporting the formation of a dimer. The analysis of the *intra*molecular ROESY cross-peaks focused on the individuation of the NH(*i*)→NH(*i*+1) and of the $C^{\alpha}(i)$→NH(*i*+1) (or $C^{\beta}(i)$ →NH(*i*+1) in the case of the Aib residues) cross-peaks, in order to verify the adoption of a folded structure. Table 3.8 lists the cross-peaks which have been observed.

Table 3.8. Selected intramolecular cross-peaks, useful for conformational analysis, observed in the 2D NMR ROESY spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu 1 mM in $CDCl_3$.

Cross-peak observed	**Peptide**	**Peptide'**
αNH$_i$→αNH$_{i+1}$ (NH)	NH Aib^1 → NH AlaT^2	--- → --- **(2)**
	--- → --- **(1)**	--- → --- **(2)**
	NH Aib^3 → NH Aib^4	NH Aib$^{3'}$ → NH Aib$^{4'}$
	NH Aib^4 → NH AlaT^5	NH Aib$^{4'}$ → NH AlaT$^{5'}$
	NH AlaT^5 → NH Aib^6	NH AlaT$^{5'}$ → NH Aib$^{6'}$
αCH$_i$→NH$_{i+1}$ (HA)	HA AlaT^2 → NH Aib^3	HA AlaT$^{2'}$ → NH Aib$^{3'}$
βCH$_i$→NH$_{i+1}$ (HB)	HB1 AlaT^2 → NH Aib^3	HB1 AlaT$^{2'}$ → NH Aib$^{3'}$
		HB2 AlaT$^{2'}$ → NH Aib$^{3'}$
	HA AlaT^5 → NH Aib^6	HA AlaT$^{5'}$ → NH Aib$^{6'}$
	HB1 AlaT^5 → NH Aib^6	HB1 AlaT$^{5'}$ → NH Aib$^{6'}$
		HB2 AlaT$^{5'}$ → NH Aib$^{6'}$

Notes. Residues belonging to one of the two dimer-forming molecules have been hyphenated for clarity. **(1)**: αNHs of Aib^1 and αAlaT^2 have very similar chemical shift, so that their cross-peak would be too close to the diagonal of the spectrum to be detected. **(2)**: The αNH of AlaT$^{2'}$ has a unusually low chemical shift and no cross-peaks with this proton have been observed.

As it can be observed from the table, most of the diagnostic cross-peaks are present and therefore most likely both molecules are folded. Another interesting finding obtained from the analysis of the *intra*molecular ROESY cross-peaks is that (Table 3.9) cross-peaks between the C(6)H and only one of the two βCHs are observed for all the four AlaT residues. This is an evidence that the orientation of all thymines with regard to the side-chain linker is blocked. Cross-peaks between the C(6)H and the αCH or αNH are also observed, particularly in the C-terminal AlaT^5 and AlaT$^{5'}$ residues, suggesting that the nucleobases are close to the backbone, and allowing to infer that the presence of the 'usual' intraresidue H-bond for the αNH of the AlaT residues could be possible.

Table 3.9. AlaT intraresidue cross-peaks observed in the 2D NMR ROESY spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu 1 mM in $CDCl_3$.

Molecule	**Peptide**	**Peptide'**
Intraresidue cross peaks	HA AlaT^2 → H6 AlaT^2	--- → ---
	HA AlaT^5 → H6 AlaT^5	HA AlaT$^{5'}$ → H6 AlaT$^{5'}$
	HB1 AlaT^2 → H6 AlaT^2	HB1 AlaT$^{2'}$ → H6 AlaT$^{2'}$
	HB1 AlaT^5 → H6 AlaT^5	HB1 AlaT$^{5'}$ → H6 AlaT$^{5'}$
	HN AlaT^5 → H6 AlaT^5	HN AlaT$^{5'}$ → H6 AlaT$^{5'}$

Notes. Residues belonging to one of the two dimer-forming molecules have been hyphenated for clarity.

Very strong intramolecular N(3)H-N(3)H cross-peaks between the two thymines of both molecules are observed (Fig.3.23 and 3.24). This finding suggests that the nucleobases protrude from the same side of the peptide chains, as expected for residues at *i,i+3* positions in Aib-rich 3_{10}-helical peptides. Moreover, since the cross peaks are strong, the planes of the two nuclebases on the same molecule must be closer than the backbones of the two residues, which are 6.3 Å apart a in 3_{10}-helix. This might require a particular side-chain conformation to bring the nucleobases close one to the other.

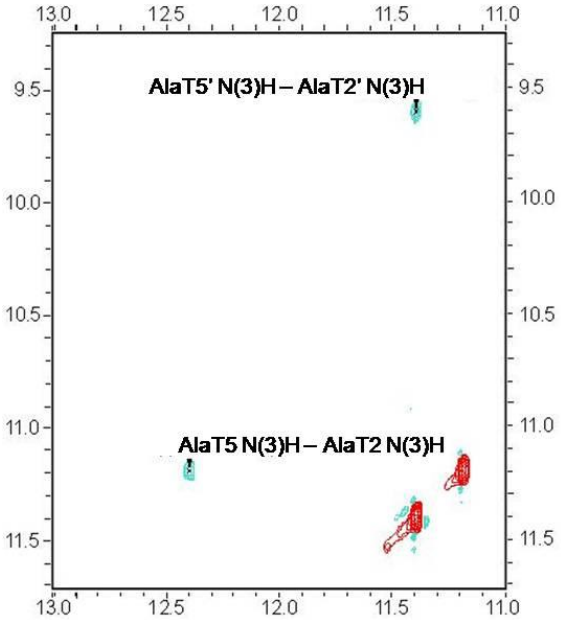


Fig. 3.23. Portion of the ROESY spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu in $CDCl_3$ solution at 1 mM concentration. Residues belonging to one of the two dimer-forming molecules have been hyphenated for clarity.

On the other side, the analysis of the *inter*molecular ROESY cross-peaks has confirmed the presence of a dimer, stabilized by interactions involving the nucleobases. Table 3.10 lists intermolecular cross-peaks due to the nucleoamino acid residues, while Fig. 3.24 shows a

portion of the ROESY spectrum with strong, both *intra*- and *inter*-molecular, thymine-thymine cross- peaks.

Table 3.10. Intermolecular cross-peaks between AlaT residues observed in the 2D NMR ROESY spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu 1 mM in $CDCl_3$.

Intermolecular cross-peaks involving AlaT-AlaT proximity	
N(3)H - thymine	<u>N(3)H AlaT$^{5'}$ → H6 AlaT^5</u> N(3)H AlaT$^{2'}$ → H6 AlaT^5 C^{5}H$_3$ AlaT$^{2'}$ → N(3)H AlaT^5 <u>C^{5}H$_3$ AlaT$^{5'}$ → N(3)H AlaT^5</u> C^{5}H$_3$ AlaT$^{5'}$ → N(3)H AlaT^2
N(3)H – alkyl portion of AlaT	N(3)H AlaT$^{2'}$ → HA AlaT^5 N(3)H AlaT$^{2'}$ → HB2 AlaT^5 <u>N(3)H AlaT$^{5'}$ → HA AlaT^5</u>

Notes. Residues belonging to one of the two dimer-forming molecules have been hyphenated for clarity. Particularly strong intermolecular cross-peaks are underlined.

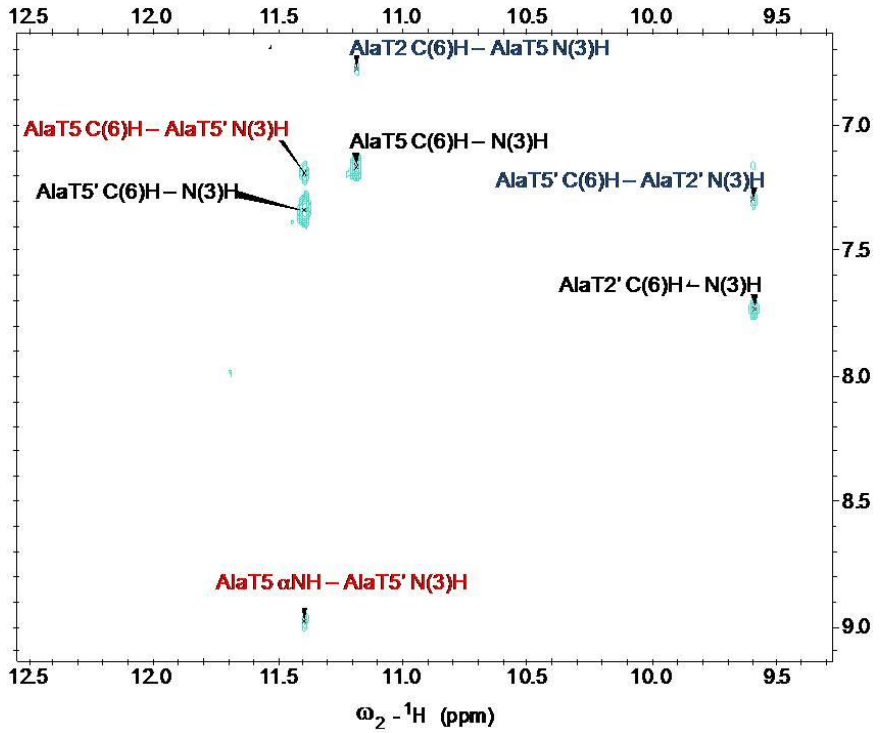


Fig. 3.24. Portion of the ROESY NMR spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu at 1 mM concentration in $CDCl_3$. Residues belonging to one of the two dimer-forming molecules have been hyphenated for clarity. Intramolecular (blue) as well as intermolecular (red) AlaT N(3)H cross peaks are present.

Apparently, the nucleobase alignment strongly favours cooperative H-bond mediated interactions, as proved by the remarkable stability of the dimer. Indeed, only a close inspection of the 1D spectrum in $CDCl_3$ at 1 mM concentration allows the

individuation of some weak signals due to the monomer (from Fig. 3.21, trace a) and about half of the nucleopeptide molecules are still forming a dimer in 2 % DMSO (when the added solvent is 250 times more abundant than the nucleopeptide).

The strongest intermolecular cross-peaks observed are those among the C-terminal nucleoamino acid residues ($AlaT^5$ and $AlaT^{5'}$), supporting the formation of a head-to-head dimer. The absence of strong $AlaT^2$-$AlaT^{2'}$ cross peaks suggests that the dimer might deviate from perfect parallelism, since the two molecules seem to be closer in their C-terminal than in their N-terminal portion.

It would be very interesting to be able to obtain the spatial arrangement of the two molecules in the dimer using the quantitative information which can be obtained by comparing the intensities of the cross-peaks observed in the ROESY experiment.

Indeed, ROESY cross-peaks are due to the *nuclear Overhauser effect* (NOE), whose intensity (η), in rigid molecules, is inversely proportional to the sixth power of the distance of the correlated nuclei ($\eta_{ij} \propto 1/r_{ij}^6$). Therefore, if the distance r_{ij} between a pair of nuclei related by a cross-peak (for instance between two non equivalent CH_2 protons) is known, the distance between a second pair of nuclei related by a second cross-peak r_{kl} can be obtained by the relative intensities of the two cross-peaks (η_{ij} and η_{kl}, respectively) simply by using the relation $r_{kl} = r_{ij} \cdot (\eta_{ij}/\eta_{kl})^{1/6}$. In the case of the dimer, such calculations have been performed by calibrating the internuclear distances on the distance between the CH_2 protons of the $AlaT^5$ residue (set at 1.78 Å), and modelling studies employing the obtained data are planned.

3.3.3 Thymine-based nucleo-heptapeptides

The availability of three thymine-based lysine containing nucleo-heptapeptides, one containing only C^α-*tri*substituted residues, one containing one Aib in the middle of its sequence, one containing four Aib residues (Fig. 3.25), allowed to compare the structuration of the Ala-rich and of the Aib-rich nucleopeptides. In turn, this would have allowed to test whether the structuration observed in the rigid nucleopeptides was is only due to the abundance of hindered C^α-*tetra*substituted residues or whether the sequential structure and the presence of the nucleoamino acid residues have some structuring properties.

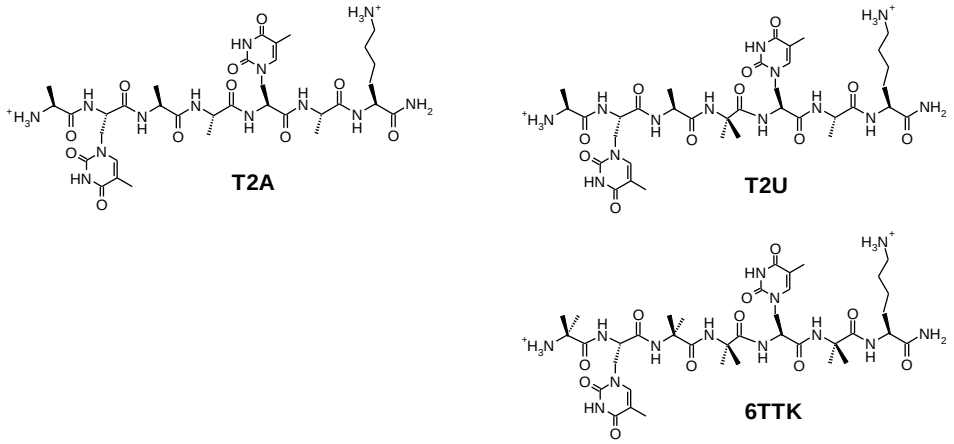


Fig. 3.25. Molecular structures of the thymine-based nucleo-heptapeptides. Top: flexible nucleopeptide containing four Ala residues (T2A); centre: 'mixed' nucleopeptide containing three Ala and one Aib residues (T2U); bottom: rigid nucleopeptide containing four Aib residues (6TTK).

Originally it was planned to perform the NMR conformational analyses in aqueous solution, in order to study the peptides in an environment close to physiological conditions. 2D NMR COSY, TOCSY and ROESY experiments were thus performed on the two Ala-rich peptides in H_2O solution with the addition of 2 % d_9-^{t}BuOH.

Unfortunately, in aqueous solution the proton exchange between the solvent and the amide NHs of the two peptides is so fast that no cross-peaks involving such protons are observed. A complete assignment therefore can not be achieved, making it impossible to obtain useful conformational information. The three nucleopeptides were therefore studied by 2D NMR in d_6-DMSO solution. After a preliminary test, comparison of the 1D ^{1}H-NMR spectra of the Ala-based nucleopeptide at 1 mM and 5 mM concentration demonstrated that no aggregation takes place. Therefore it was chosen to perform the 2D experiments at 5 mM concentration to get a better signal to noise ratio.

3.3.3.1 H-(Ala-AlaT-Ala)$_2$-Lys(H)-NH$_2$

The first nucleo-heptapeptide which was studied was the Ala-rich nucleo-heptapeptide H-(Ala-AlaT-Ala)$_2$-Lys(H)-NH$_2$. Complete assignment was achieved by the usual procedure employing TOCSY and ROESY data, then ROESY cross-peaks were used to obtain conformational information. The only peaks which it was not possible to assign belong to the two N(3)H imide protons of the thymine rings, which have very similar chemical shifts (δ 11.24 and 11.23) and which are involved in rapid exchange with the protons of traces of moisture present in the solution, so that they do not form cross-

peaks.

Table 3.11 lists all NH(*i*)→NH(*i*+1) and C^{α}(*i*)→NH(*i*+1) cross-peaks, relevant for evaluating the adoption of a folded conformation, while a portion of the ROESY spectrum is presented in Fig. 3.26.

Table 3.11. Cross-peaks providing useful information for the conformational analysis of H-(Ala-AlaT-Ala)$_2$-Lys(H)-NH$_2$ 5 mM in d_6-DMSO observed in the 2D NMR ROESY spectrum.

Cross-peak observed	
$\alpha N_i \rightarrow \alpha N_{i+1}$	**$C^{\alpha}_i \rightarrow \alpha N_{i+1}$**
(1)	
AlaT^2 - Ala^3	AlaT^2 - Ala^3
	Ala^3 - Ala^4
Ala^4 - AlaT^5	
AlaT^5 - Ala^6	AlaT^5 - Ala^6
Ala^6 - Lys^7	Ala^6 - Lys^7
Lys^7 - CONH	Lys^7 - CONH; Lys^7 - CONH' **(2)**

Notes. **(1)**: The rapid exchange of the protons of the amonium ion with the moisture present in the solution suppresses most of its cross-peaks. **(2)**: One of the two nonequivalent C-terminal amide NH protons has been hyphenated for clarity.

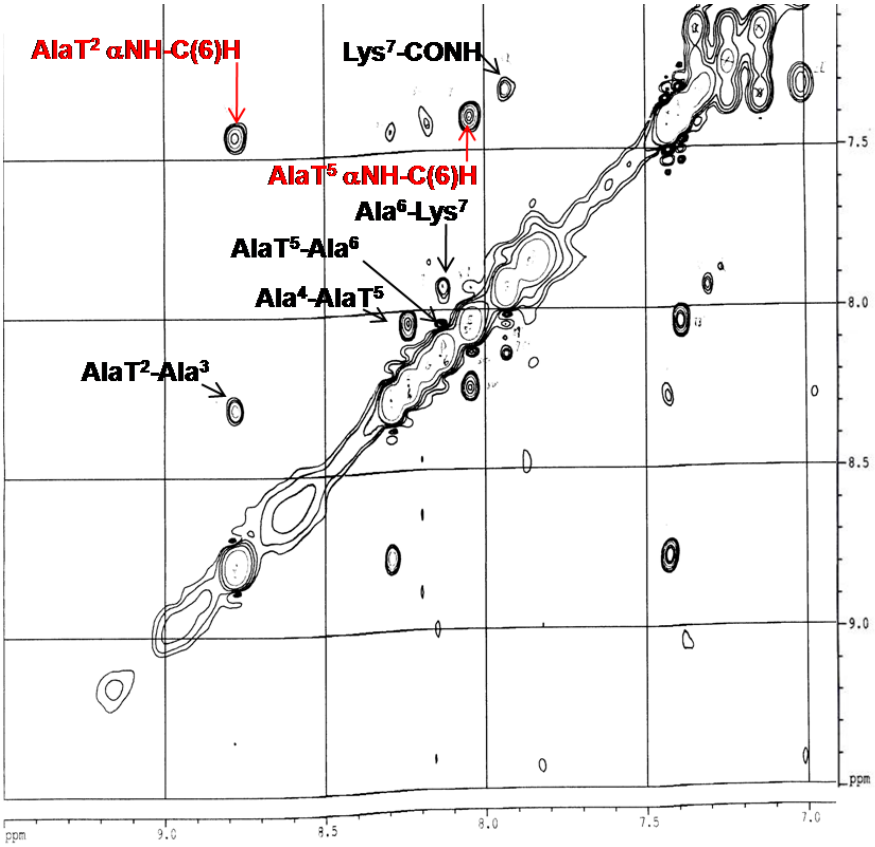


Fig. 3.26. Portion of the 2D ROESY spectrum of H-(Ala-AlaT-Ala)$_2$-Lys(H)-NH$_2$ 5 mM in d_6-DMSO. αNH(*i*)→αNH(*i*+1) cross-peaks are shown. Intraresidue AlaT cross-peaks between backbone and nucleobase are highlighted in red.

It can be seen that five of the seven possible NH(*i*)→NH(*i*+1) and C^{α}(i)→NH(*i*+1) cross-peaks are present. This suggests that the peptide is most likely folded, in particular in the C-terminal portion, where none of the cross-peaks is missing, whereas maybe the structuration is less pronounced in the N-terminal portion, where fewer cross-peaks have been observed. On the other side, no cross-peaks deriving by medium range connectivity, which could confirm the adoption of a helical structure, have been observed for this nucleopeptide.

Interestingly, two strong intraresidue cross-peaks between αNH and C(6)H protons of the AlaT residues are observed, allowing the assumption that the nucleobase is close to the backbone. This could be a sign that intraresidue H-bonds involving the nucleoamino acid residue are present for this peptide in d_6-DMSO solution, at least to a certain extent.

3.3.3.2 H-Ala-AlaT-Ala-Aib-AlaT-Ala-Lys(H)-NH_2

The nucleopeptide carrying an Aib residue was studied employing the same approaches. Again, complete assignment with the exclusion of the two N(3)H protons of the two thymines (found at δ 11.22 and 11.20) was achieved. Table 3.12 lists all NH(*i*)→NH(*i*+1) and C^{α}(*i*)→NH(*i*+1) cross-peaks as well as the C^{β}(*i*)→NH(*i*+1) cross-peak in the case of the Aib residue) cross-peaks, relevant for evaluating the adoption of a folded conformation. A section of the ROESY spectrum is presented in Fig. 3.27.

Table 3.12. Cross-peaks providing useful information for the conformational analysis of H-(Ala-AlaT-Ala)$_2$-Lys(H)-NH_2 5 mM in d_6-DMSO observed in the 2D NMR ROESY spectrum.

Cross-peak observed		
$\alpha N_i \rightarrow \alpha N_{i+1}$	**$C^{\alpha}_i \rightarrow \alpha N_{i+1}$**	**$C^{\beta}_i \rightarrow \alpha N_{i+1}$**
(1)	Ala^1 - $AlaT^2$	-----
$AlaT^2$ - Ala^3	$AlaT^2$ - Ala^3	-----
Ala^3 - Aib^4		-----
Ala^4 - $AlaT^5$	-------	Aib^4 - $AlaT^5$
$AlaT^5$ - Ala^6	$AlaT^5$ - Ala^6	-----
Ala^6 - Lys^7	Ala^6 - Lys^7	-----
Lys^7 – CONH	Lys^7 - CONH Lys^7 - CONH' **(2)**	-----

Notes. **(1)**: The rapid exchange of the protons of the amonium ion with the moisture present in the solvent suppresses most of its cross-peaks. **(2)**: One of the two nonequivalent C-terminal amide NH protons has been hyphenated for clarity. Only the C^{β}H(*i*)→NH(*i*+1) cross-peak involving Aib has been reported.

In this case, one more cross-peak is observed in each series, so that all possible NH(*i*)→NH(*i*+1) cross-peaks are present, with the exclusion of the first one, involving the N-terminal amonium ion. This suggests that folding is more stable for this nucleopeptide that for the analogue containing only Ala residues, as it could be predicted, given the presence of a rigid C^{α}-tetrasubstituted residue.

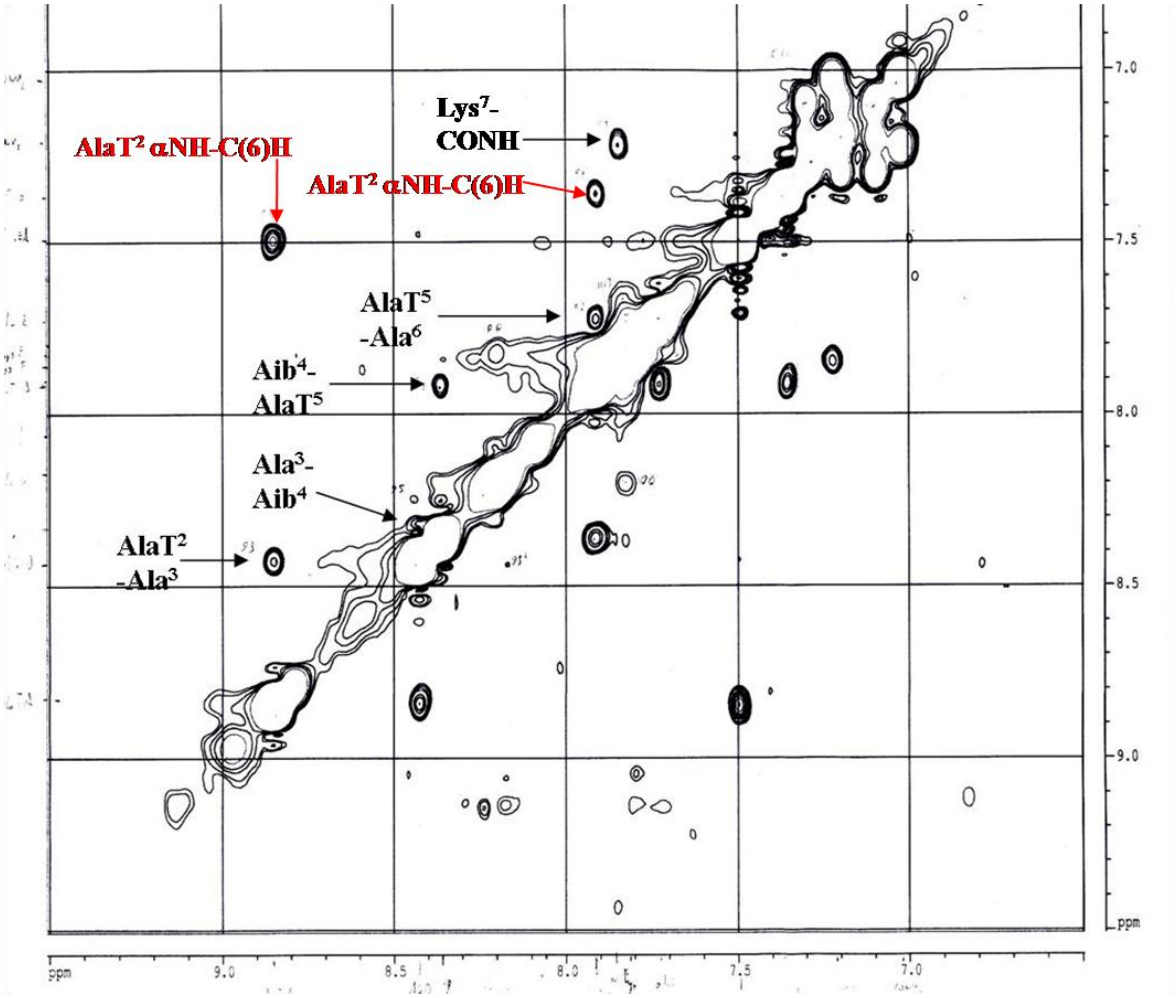


Fig. 3.26. Portion of the 2D ROESY spectrum of H-Ala-AlaT-Ala-Aib-AlaT-Aib-Lys(H)-NH_2 5 mM in d_6-DMSO. NH(*i*)→NH(*i*+1) cross-peaks are shown. Intraresidue AlaT cross-peaks between backbone and nucleobase are highlighted in red.

Despite of the more complete folding, however, only two medium range cross-peaks are observed, both quite week and between nuclei three residues apart. The one which is more interesting is formed between two βCH protons, one on each AlaT residue. This could be a hint that the two nucleobases are aligned and lie close to each other, as it was found in the spectrum of the rigid nucleo-hexapeptide in $CDCl_3$ solution. However this hypothesis needs additional evidence to be confirmed.

3.3.3.3 H-(Aib-AlaT-Aib)$_2$-Lys(H)-NH$_2$

The Aib-rich nucleo-heptapeptide was studied employing the same techniques used for the two Ala containing analogues. However, in addition to the two N(3)H protons of the two thymines, found at δ 11.33 and 11.31, it was not possible to assign the signals related to the two NH_3^+ groups (one belonging to the N-terminal Aib, the other to the side-chain of lysine), resonating at δ 8.14 and 7.64. This was probably due to the rapid exchange with protons of the moisture present in the solution.

On the other side, the information obtained by the analysis of the rest of the signals is abundant and very interesting. Fig. 3.27 shows part of a section of the ROESY spectrum, in the region where NH/NH cross-peaks can be found.

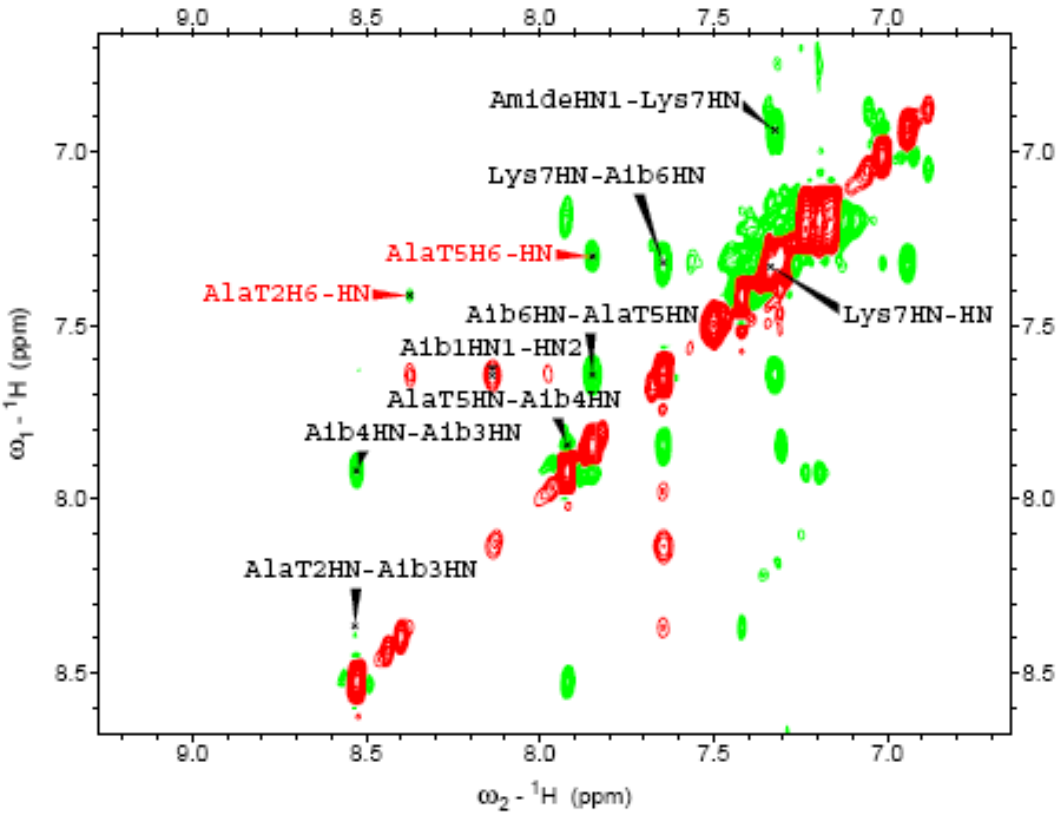


Fig. 3.27. Portion of the 2D ROESY spectrum of H-(Aib-AlaT-Aib)$_2$-Lys(H)-NH$_2$ 5 mM in d_6-DMSO. NH(*i*)→NH(*i*+1) cross-peaks are shown. The intraresidue AlaT cross-peaks between backbone and nucleobase are highlighted in red.

All NH(i)→NH(i+1) cross-peaks are present, apart the one involving the N-terminal amonium ion, as in the spectrum of the analogue with only one Aib residue. Similar to the two other analogues, the intraresidue cross-peak between the αNH and C(6)H protons of both AlaT are observed.

Fig. 3.28 presents the region of the ROESY spectrum where C^{α}/NH or AlaT C^{β}/NH cross-peaks are normally observed.

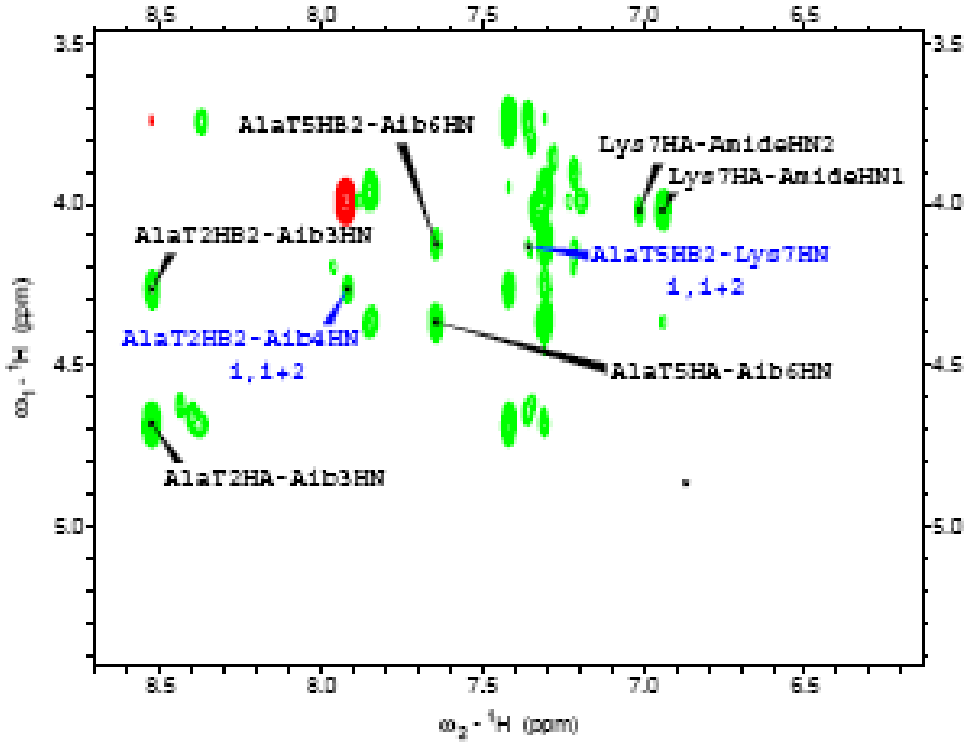


Fig. 3.28. Portion of the ROESY spectrum of H-(Aib-AlaT-Aib)$_2$-Lys(H)-NH$_2$ 5 mM in d_6-DMSO. $C^{\alpha}(i)\rightarrow NH(i+1)$ as well as $C^{\beta}(i)\rightarrow NH(i+1)$ cross-peaks are shown in black, while $C^{\beta}(i)\rightarrow NH(i+2)$ cross-peaks are highlighted in blue.

All three possible $C^{\alpha}(i)\rightarrow NH(i+1)$ cross-peaks are present, supporting the adoption a folded conformation. Only one $C^{\beta}(i)\rightarrow NH(i+1)$ cross-peak can be found for each of the two AlaT residues, suggesting that the nucleoamino acid side-chains are conformationally rigid.

Moreover, two medium range connectivity cross-peaks $C^{\beta}(i)\rightarrow NH(i+2)$, one for each AlaT residue, are observed. The presence of such cross-peaks is considered diagnostic for 3_{10}-helical structures. Therefore, it is therefore possible to conclude that this peptide is almost certainly in 3_{10}-helical conformation, at difference with its Ala-rich analogues, most probably folded, but not as surely helical.

In summary, the NMR analysis not only confirmed most of the IR-based inductions, but also further supported the general hypothesis proposed in this work. In particular, the most important results are:

(I) For adenine-, cytosine-, and thymine-based protected tripeptides in $CDCl_3$

solution, strong evidence supporting the presence of intraresidue H-bonds involving the αNH protons of the nucleoamino acid residues has been obtained.

(II) In $CDCl_3$ solution all protected nucleopeptides adopt a folded conformation, stabilized by intramolecular H-bonds, and data suggesting the formation of helical structures have been obtained in several cases.

(III) A stable parallel dimer is formed by the protected thymine-based nucleo-hexapeptide in $CDCl_3$ solution and dimerization through intermolecular H-bonds is mediated by the side-chain nucleobases, which are aligned along the backbone and can therefore bind cooperatively to the other molecule.

(IV) The thymine-based nucleo-heptapeptides are folded, and for the Aib-rich peptide the predicted adoption of a 3_{10}-helical conformation has been confirmed.

3.4 Circular Dichroism

Circular dichroism is an analytical technique based on the differential absorption of circularly polarized light by chromophores set in an asymmetric environment. The spectral region most commonly investigated for the conformational analysis of peptides lies in the far UV region (λ= 190-260 nm). The amide chromophore of peptides displays two bands in this interval, corresponding to the n→π* and π→π* transitions. The amide group is planar and therefore its transitions are optically inactive, but the presence of chiral elements (such as the C^{α} atoms of chiral amino acids) next to the cromophore, as well as its incorporation in helical secondary structures, which are intrinsically dissymmetric, make it optically active, therefore liable to CD analysis.

The interactions among consecutive amide chromophores along the backbone modulate their optical transitions in a way that depends on their relative orientation, induced by the peptide secondary structure. The global outcome is that the spectral features depend on the peptide secondary structure, so that useful information can be obtained both from the shape and from the intensity of the dichroic bands of peptides.[140] The *molar ellipticity* $[\Theta]_T$ ($deg \cdot cm^2 \cdot dmol^{-1}$) is the parameter used for comparing the intensity of the CD signal (Θ) by normalization with regard to concentration (c, mol/L) and optical path (l, cm), using the relation $[\Theta]_T = \Theta/(c \cdot l)$.

When applying circular dichroic techniques in the far UV to study the nucleopeptide conformation, however, it must be considered that chromophores other than the peptide amide groups are present. Indeed the side-chain nucleobases are powerful chromofores, absorbing already at λ< 300 nm. Like the amide groups, they are planar and optically inactive, but the proximity of chiral elements makes their transitions optically active. The superimposition of contributions from the nucleobases and from the amide groups of the backbone significantly complicates the CD spectra of the nucleopeptides, making their interpretation in terms of peptide secondary structures difficult, when not impossible.

On the other side, the nucleobase contribution to the CD spectra is largely prevalent in the region at λ> 250 nm, therefore the analysis of the bands observed in this region can be used to infer data on the side-chain nucleobase orientation. Indeed, the interaction of the nucleobase chromophore with the asymmetric centre in the backbone is modulated by the nucleobase orientation, so that different orientations can result in a different band shape.

Therefore, if all side-chains of a given compound in a sample are rigid and oriented in the same way relative to the backbone, contributions of the same kind from all molecules sum up in the sample spectrum and a strong band results. On the contrary, if the side-chains are flexible and no prevalent orientation exists, each molecule contributes in a different way to the spectrum, so that part the signals cancel each other, at least partially, and a weaker and less featured band results. Consequently, by comparing the dichroic bands observed in the region of the nucleobase chromophore in the spectra of analogous nucleopeptides acquired under similar experimental conditions, it should be possible to have an idea of the relative conformational order of their side-chains. Moreover, particularly strong dichroic bands in the nucleobase chromophore region can suggest the presence of strong intermolecular nucleobase-mediated interactions forcing the assumption of a rigid conformation of the side-chains of the groups involved.

3.4.1 Protected rigid nucleopeptides

Conformational investigations by CD on the protected nucleopeptides have been performed both in $CHCl_3$ and in MeOH. The first solvent, employed for the IR and NMR studies, is generally not suited for CD studies on peptides, since it absorbs strongly the UV radiation from $\lambda < 240$ nm, where most of the bands due to the peptide chromophore are observed. This is not a problem, however, if the nucleobase bands are of interest. MeOH, on the other side, allows the observation also of the lower wavelength portion of the spectrum. The low polarity of $CHCl_3$ favours intermolecular interactions mediated by dipolar effects (including hydrogen bonding interactons) among solute molecules. On the contrary MeOH, being a polar solvent, able to accept and to donate H-bonds, tends to disrupt such interactions.

All CD experiments on the protected nucleopeptides were performed at room temperature. Given the strong absorbance of the nucleobases in the far UV region, and since the IR experiments showed the propensity of the nucleopeptides to aggregation in $CDCl_3$, the spectra were acquired at concentrations not higher than 1 mM in both solvents.

The nucleo-tripeptides with free nucleobases and without the Z protecting group were considered more suited for the CD analysis than their corresponding Z-protected analogues, because the aromatic chromophore of the Z protector is active in the spectral region where the bands due to the nucleobases are found, and in principle this could

perturb the spectra. However, not much useful conformational information can be obtained from the CD spectra of such nucleopeptides.

In the case of the adenine-based tripeptide Boc-Aib-AlaA^{H}-Aib-OMe (Fig. 3.29) the (positive) nucleobase band with maximum at about λ= 275 nm is week in MeOH and almost absent in $CHCl_3$. No dilution effects can be observed in MeOH, while in $CHCl_3$ the spectrum measured at 0.10 mM concentration appears to be more strongly negative. Such featureless spectra in $CHCl_3$ are quite unexpected, given the amount of evidence obtained both from IR and from NMR analysis about the structuration of the peptide, and in particular about the nucleobase rigidity, in chloroform solution.

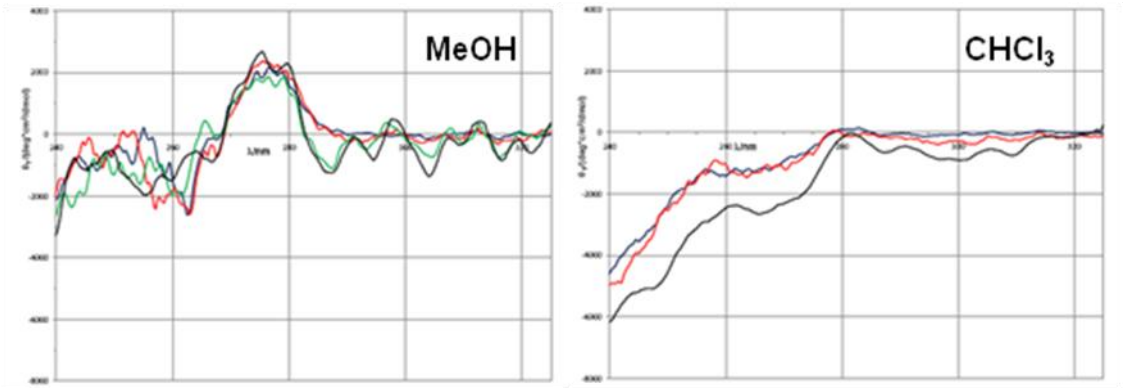


Fig. 3.29. Molar ellipticity of Boc-Aib-AlaA^{H}-Aib-OMe in MeOH (left) or $CHCl_3$ (right) at different concentrations (blue: 1.04 mM, red: 0.52 mM, green: 0.21 mM, black: 0.10 mM). The average of 8 scans (16 scans for the lowest concentration) is plotted. Spectra are drawn at the same scale ($240 \leq \lambda/nm \leq 325$, $-8000 \leq [\Theta]_T/(deg \cdot cm^2 \cdot dmol^{-1}) \leq 4000$).

In the case of Ac-Aib-AlaT-Aib-O^{t}Bu (Fig. 3.30) the negative nucleobase band with minimum around 270 nm in both solvents has a significant intensity in $CHCl_3$, and it is even stronger in MeOH. In both solvents the intensity of this band increases with dilution, although the increase in MeOH seems more regular than in $CHCl_3$, where it is scarce from 0.52 mM to 0.21 mM and pronounced from 0.21 mM to 0.10 mM.

The behaviour observed by dilution is again unexpected, because it seems to suggest that nucleobase structuration increases at lower dilution. On the contrary it was assumed that the strong intermolecular interactions observed for this nucleopeptide in chloroform solution by IR spectroscopy, which are most likely nucleobase-mediated, would have favoured the rigidity of its conformation at higher concentrations.

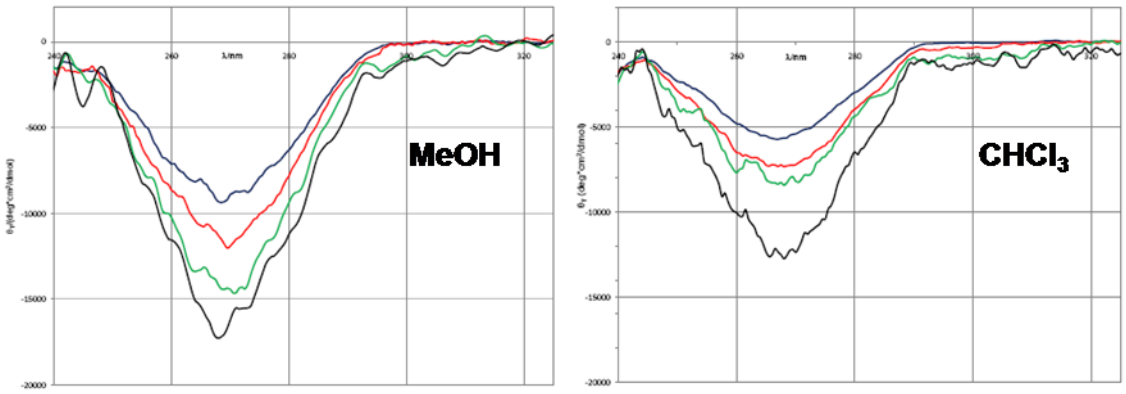


Fig. 3.30. Molar ellipticity of Ac-Aib-AlaT-Aib-O^{t}Bu in MeOH (left) or $CHCl_3$ (right) at different concentrations (blue: 1.03 mM, red: 0.51 mM, green: 0.21 mM, black: 0.10 mM). The average of 8 scans (16 scans for the most dilute sampe) is plotted. Spectra are drawn at the same scale ($240 \leq \lambda/nm \leq 325$, $-20000 \leq [\Theta]_T/(deg \cdot cm^2 \cdot dmol^{-1}) \leq 2000$).

Considering the the CD spectra of Boc-Aib-AlaC^H-Aib-OMe (Fig. 3.31, left) at 1 mM concentration it is possible to observe a quite strong negative nucleobase band with minimum at about 275 nm, whose shape and intensity are similar in MeOH and $CHCl_3$.

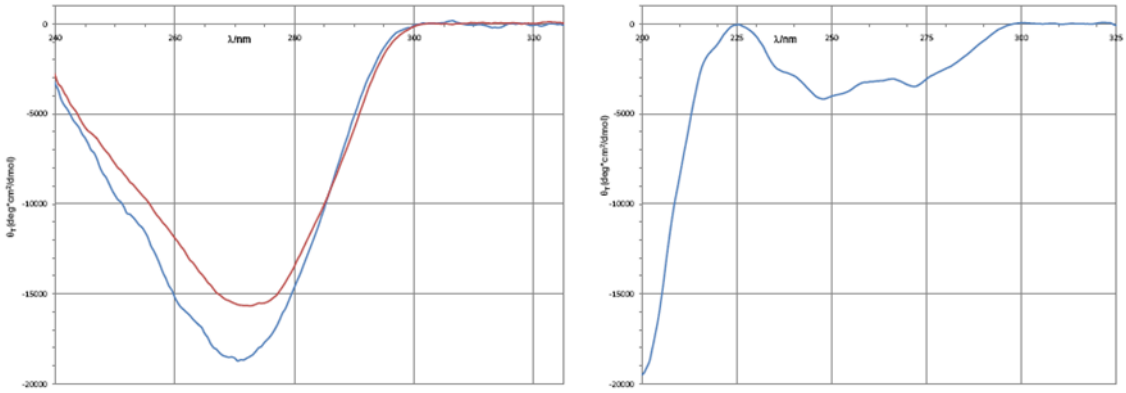


Fig. 3.31. Molar ellipticity of Boc-Aib-AlaC^H-Aib-O^{t}Bu (left) and of Boc-Aib-AlaG^{OH}-Aib-OMe (right) at 1 mM concentration in MeOH (blue) or $CHCl_3$ (red). The average of 8 scans is plotted. The same vertical scale is adopted for both spectra ($-8000 < [\Theta]_T/deg \cdot cm^2 \cdot dmol^{-1} < 4000$), while the wavelength ranges are $240 \leq \lambda/nm \leq 325$ for the cytosine-based and $200 \leq \lambda/nm \leq 325$ for the guanine-based nucleopeptide.

Due to the high polarity of this compound, the analysis of the guanine-based tripeptide Boc-Aib-AlaG^{OH}-Aib-OMe has been possible only in MeOH solution. Its CD spectrum at 1 mM concentration is shown in Fig. 3.31, right. The negative nucleobase band is broad and not very intense. As for all other protected nucleo-tripeptides, the spectrum in MeOH displays a strong negative band from $\lambda < 225$ nm, whose intensity increases towards shorter

wavelengths until the high absorbance of the solution of these compounds prevents further investigation.

3.4.2 Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu

The NMR analysis of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu had shown that a dimer exists in chloroform solution at 1 mM concentration, whereas the monomeric peptide is formed by addition of a more polar solvent like dimethylsulfoxide. It was thought that CD spectra acquired in $CHCl_3$ and MeOH would have shown some features reflecting this different behaviour. Fig. 3.32 presents the CD spectra of the nucleo-hexapeptide in MeOH (left) and in CHCl3 (right), compared with the spectra of the protected nucleo-tripeptide Z-Aib-AlaT-Aib-O^{t}Bu. This was done also to evaluate the differences between the spectra acquired in different solvents for a peptide which does not display dimer formation. Moreover, given that the presence of the N-terminal Z protecting group on the nucleo-hexapeptide could in principle perturb the spectra in the region where the nucleobase band is observed, it was useful to study a second Z-protected thymine-based nucleopeptide

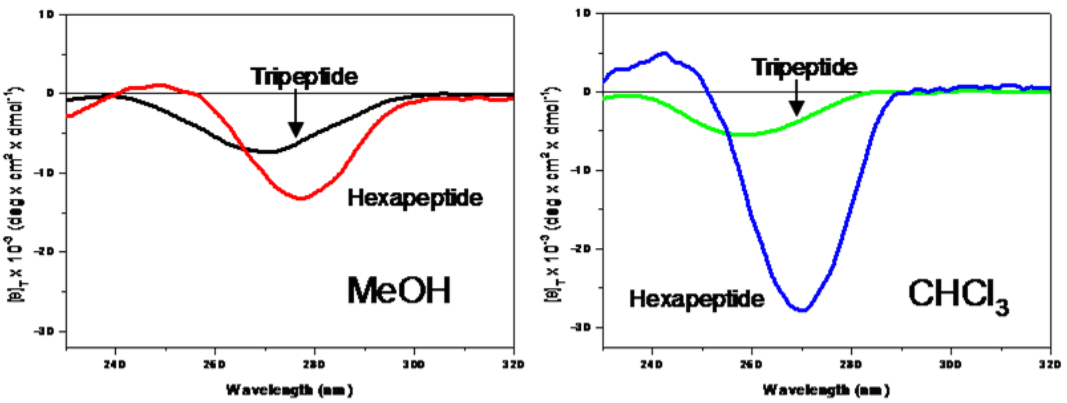


Fig. 3.32. CD spectra of the tripeptide Z-Aib-AlaT-Aib-O^{t}Bu and of the hexapeptide Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu 1 mM in MeOH (left) and in $CHCl_3$ (right) at 1 mM concentration. Spectra are drawn at the same scale ($230 \leq \lambda/nm \leq 320$, $-35000 \leq [\Theta_T]/deg \cdot cm^2 \cdot dmol^{-1} \leq 10000$). The average of 8 scans is plotted.

Considering the spectra of the tripeptide, there is not much difference between MeOH and $CHCl_3$. The negative band due to the nucleobase is slightly more intense in MeOH than in $CHCl_3$. Its minimum, found about at 267 nm in MeOH, shifts towards shorter wavelengths by about 10 nm in $CHCl_3$.

The CD spectrum of the hexapeptide in MeOH shows a negative band, with minimum shifted at longer wavelength by about 10 nm compared to the tripeptide in the same solvent. This band is about twice as intense as the band of the tripeptide. If the band is mostly due to the nucleobases, this can be rationalized by the presence of two thymines on the hexapeptide molecules, so that, at the same concentration, a sample of the latter has twice as many chromophores than a sample of the tripeptide.

If the spectrum of the hexapeptide in $CHCl_3$ is considered, the negative nucleobase band is much larger, being more than five times the band of the tripeptide in the same solvent. This can be a sign that in chloroform solution intermolecular interactions are significantly stronger in the case of the hexapeptide. To better investigate this phenomenon, CD spectra of the nucleo-hexapeptide in both solvents at different concentrations, ranging from 1 mM to 50 μM (down to 25 μM in $CHCl_3$) were acquired (Fig. 3.33).

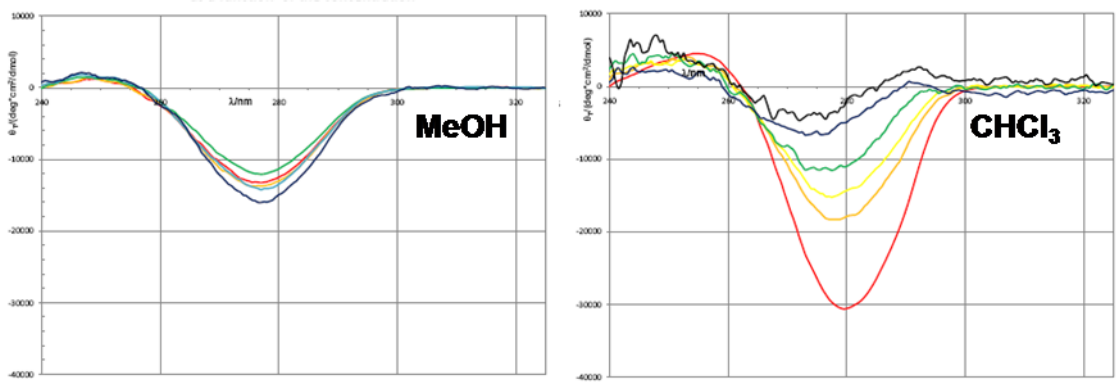


Fig. 3.33. CD spectra of hexapeptide Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu in MeOH (left) and in $CHCl_3$ (right) at various concentrations (red: 1.0 mM; orange: 0.5 mM, yellow: 0.35 mM; green: 0.20 mM; light blue: 0.10 mM; blue: 50 μM; black: 25 μM). Spectra are drawn at the same scale ($240 \leq \lambda/nm \leq 325$, $-40000 \leq [\Theta_T]/deg \cdot cm^2 \cdot dmol^{-1} \leq 10000$). The average of 8 scans (16 for the most diluted samples) is plotted.

It can be easily observed that no significant dilution effects are present in MeOH, whereas dilution effects are strong in $CHCl_3$, where the intensity of the bands decreases steadily by consecutive dilutions. Moreover the minimum of the band tends to move to shorter wavelengths at lower concentrations, closer to where it is found in MeOH solution. Apart for the noisiest and most dilute trace, two isodichroic points can be individuated at both sides of the nucleobase band (at about 265 nm and 300 nm, respectively).

This could be a sign that a conformational transition involving the nucleobases takes place as the concentration of the peptide varies, possibly related to the prevalence of the monomer over the dimeric complex at low concentration.

3.4.3 Lysine containing thymine-based nucleo-heptapeptides

Circular dichroism is a technique which can be easily applied to the conformational study of deprotected peptides since H_2O, having a weak absorbance in the far UV region, is a convenient solvent. On the other side, the high polarity of aqueous solutions and the ubiquitous presence of H-bond networks in the solvent tend to suppress intermolecular interactions among solute molecules if they are not particularly stable.

The two Ala-rich lysine containing thymine-based nucleo-heptapeptides H-(Ala-AlaT-Ala)$_2$-Lys(H)-NH_2 and H-Ala-AlaT-Ala-Aib-AlaT-Ala-Lys(H)-NH_2 have been studied in phosphate buffer (10 mM $NaH_2PO_4 \cdot H_2O/Na_2HPO_4 \cdot 2H_2O$, pH 7.4), which is a crude approximation of the physiological environment. The moderate absorbance of the medium allows to investigate the spectral region with $\lambda \geq 200$ nm. Spectra were acquired at 40 μM and 80 μM at various temperatures, in the interval $0 \leq T/°C \leq 60$. Fig. 3.34 shows the spectra of both peptides at 40 μM concentration (in the spectra acquired at 80 μM the trend is similar, but the absorbance of the solution at low wavelength is too high to obtain meaningful data).

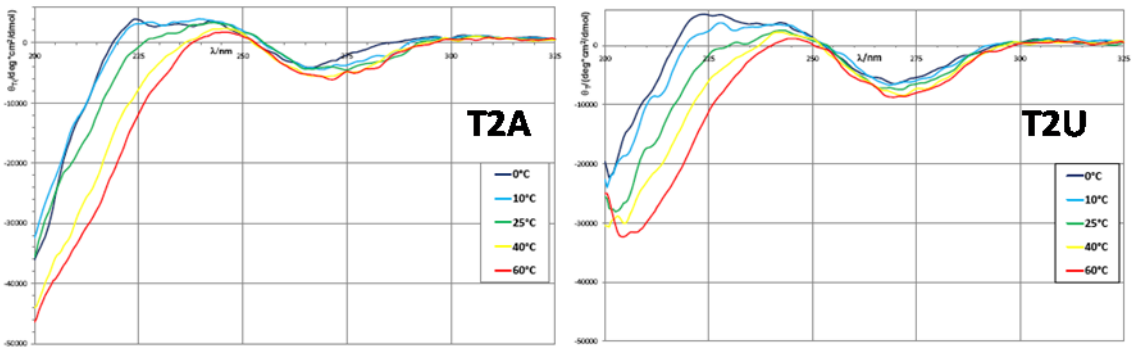


Fig. 3.34. Molar ellipticity of H-(Ala-AlaT-Ala)$_2$-Lys(H)-NH_2 (T2A, left) and H-Ala-AlaT-Ala-Aib-AlaT-Ala-Lys(H)-NH_2 40 μM (T2U, right) in phosphate buffer at different temperatures. Spectra are drawn at the same scale ($200 \leq \lambda/nm \leq 325$, $-50000 \leq [\Theta_T]/deg \cdot cm^2 \cdot dmol^{-1} \leq 6000$). The average of 8 scans is plotted.

The main features of both spectra are similar: the negative band due to the nucleobase can be observed at about 270 nm, at low temperature a weak positive band is present for 225< λ/nm< 250 while from λ< 225 nm a strong negative band is found. The variation of the spectra with the temperature is also similar, since the weak positive band tends to disappear as the temperature rises, while both negative bands increase their intensity. The increase in intensity of the nucleobase band with the temperature, which should favour the flexibility of the side-chain, is quite unexpected, like the increase of the same band with decreasing concentration in the case of the acetylated thymine-based nucleo-tripeptide. A closer closer comparison of the spectra reveals however some interesting differences: (i) at every given temperature the nucleobase band is stronger for the Aib containing nucleopeptide; (ii) while its maximum tends to move towards longer wavelengths for the Ala-based peptide, it does not change for the Aib containing peptide.
The different behaviour could be related to a different conformation of the side-chains, even if, given the variation of the nucleobase band with the temperature, it is questionable whether a stronger band can be related to a more structured side-chain.

The second difference lies in the low wavelength portion of the spectra: for the Ala-based nucleopeptide the dichroic signal decreases steadily with decreasing wavelength and no minimum can be observed until 200 nm. On the other side, for the nucleopeptide with one Aib residue, a minimum is found for all spectra in the range 201-203 nm. For peptides made only of protein amino acids it is generally accepted[140a] that unstructured molecules give negative bands with minima at λ<200 nm, whereas helical structures like the α- and the 3_{10}-helix have a minimum at about 208 nm. If the low wavelength region of the CD spectra of the nucleopeptides can be interpreted using the same criteria (as done in the literature[56,58] for other kinds of nucleopeptides) this finding could suggest that the Aib-containing nucleopeptide is more structured than the Ala-based analogue.

The Aib-rich analogue H-(Aib-AlaT-Aib)$_2$-Lys(H)-NH_2 has been studied in pure H_2O solution at different concentrations at room temperature. A reason for this choice is that the effect of temperature proved difficult to rationalize in case of the two Ala-rich analogues. On the other side, it was considered interesting to test whether dilution effects could be detected in aqueous solution for the deprotected nucleopeptides, and the need to explore a wider concentration range forced to abandon the phosphate buffer to use the less UV-absorbing pure solvent.

Fig. 3.35, left, presents the CD spectrum of the Aib-rich nucleo-heptapeptide at various concentrations in the range 0.50 mM – 17 μM, including the concentrations at which the CD analyses on the Ala-rich analogues have been performed. It can be seen that dilution effects are almost absent, apart from a slight increase in the intensity of the nucleobase band.

Fig. 3.35, right, compares the most dilute spectrum of this nucleopeptide with the spectra of the two Ala-rich analogues at the same temperature. From the comparison it is easy to note that in the spectrum of the Aib-rich peptide the nucleobase band is significantly stronger and shifted towards longer wavelengths by about 15 nm. Even more interestingly, the strong negative band found at $\lambda < 225$ nm has clearly a minimum about at 215 nm. Both features are observed in all traces of Fig.3.35, left, showing that they are not dependent from the peptide concentration.

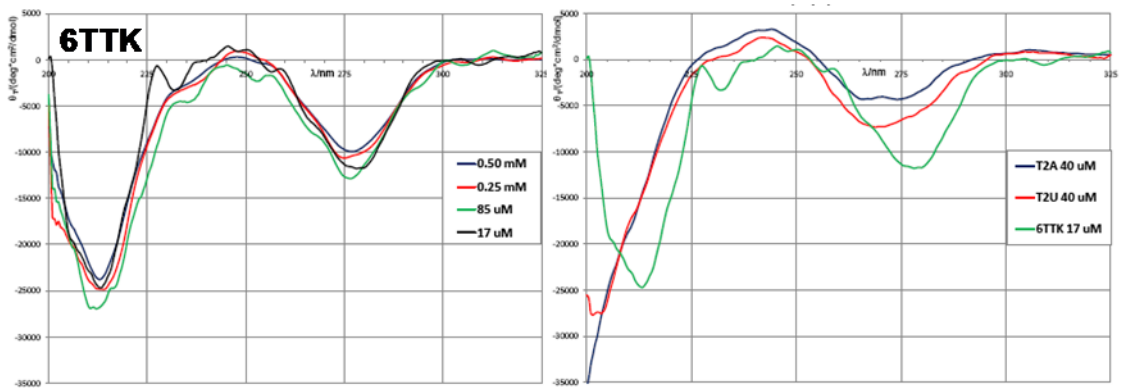


Fig. 3.35. Left: molar ellipticity of H-(Aib-AlaT-Aib)$_2$-Lys(H)-NH$_2$ (6TTK) in H$_2$O at different concentrations. Right: molar ellipticity of the same peptide 17 μM in H$_2$O and of the two Ala-rich analogues 40 μM in phosphate buffer at 25°C. Spectra are drawn at the same scale ($200 \leq \lambda/\text{nm} \leq 325$, $-35000 \leq [\Theta_T]/\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1} \leq 5000$). The average of 8 scans (16 for the most diluted sample) is plotted.

Again, under the hypothesis that bands in the low wavelength region of the CD spectra of the unprotected nucleopeptides can be interpreted as usual for peptides without nucleoamino acid residues, this seems to suggest that the Aib-rich peptide is significantly more structured than his Ala-rich analogues, and possibly that this peptide adopts an helical conformation in aqueous solution as it does in DMSO.

The different solvent used could be an issue which limits the validity of the comparison of the Ala-rich and of the Aib-rich nucleopeptides and measurements in the same experimental conditions would be required to exclude the possibility that the higher

ionic strength is responsible for the reduced structuration observed. However, the comparison between the two Ala-rich nucleopeptides and the considerations about the structuration of the Aib-rich nucleopeptide alone are not affected by this problem.

3.4.4 Flexible nucleopeptides

The underivatised flexible nucleopeptides B4H of general formula H-Lys(H)-[Ala-AlaB-Ala]$_4$-Lys(H)-NH$_2$, each containing four times one of the DNA nucleobases, have been studied by CD in PBS buffer (10 mM $NaH_2PO_4 \cdot H_2O$/$Na_2HPO_4 \cdot 2H_2O$, pH 7.4, 150 mM NaCl, 0.1 mM EDTA).[47] This buffer is a better approximation of the physiological environment than the phosphate buffer used for the Ala-rich nucleo-heptapeptides; on the other side, due to the strong absorbance of the chloride anion in the far UV, it is generally not possible to obtain meaningful data for $\lambda < 205$ nm and often, given the presence of four nucleobases per peptide, the sample absorbance was too high even at longer wavelengths. The buffer was chosen because the primary aim of the analyses of the longer nucleopeptides was the study of their nucleobase-mediated intermolecular interactions under physiological conditions rather than the analysis of their backbone conformations.

Nucleopeptide solutions at various concentrations in the range 10-40 μM were analysed at different temperatures in the range -2.5/65°C. Fig. 3.36 shows the CD spectra of the adenine-based (left) and of the cytosine-based nucleopeptides (right) at 40 μM concentration.

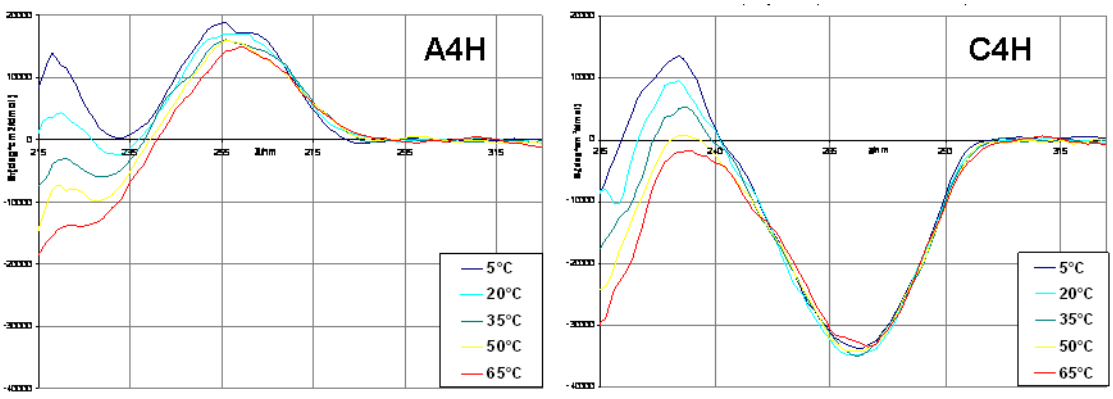


Fig. 3.36. Molar ellipticity of H-Lys(H)-(Ala-AlaA^H-Ala)$_4$-Lys(H)-NH$_2$ (A4H, left) and H-Lys(H)-(Ala-AlaC^H-Ala)$_4$-Lys(H)-NH$_2$ 40 μM (C4H, right) in PBS buffer at different temperatures. Spectra are drawn at the same scale ($215 \leq \lambda/nm \leq 325$, $-40000 \leq [\Theta_T]/deg \cdot cm^2 \cdot dmol^{-1} \leq 20000$). The average of 8 scans is plotted.

In both spectra the nucleobase bands (positive for adenine, with maximum at about 260 nm, negative for cytosine, with minimum at about 270 nm) are evident, as well as a weak positive band (with maximum at 220 nm for adenine and at 235 nm for cytosine) already observed in the thymine-based flexible nucleo-heptapeptide spectra. At lower wavelength a strong negative band is found, whose intensity increases toward shorter wavelengths, until the absorbance of the solution prevents further observation. The nucleobase band decreases slightly with increasing temperature in the case of the adenine-based nucleopeptide, whereas it does not change much in the case of the cytosine-based nucleopeptide. In both spectra the positive band decreases and the intensity of the strong negative band increases as the temperature rises. The latter features, already observed in the Ala-rich nucleo-heptapeptide spectra, are common to all spectra of the flexible nucleopeptides observed.

The spectra of the nucleopeptide carrying four thymine bases at 20 μM and at 40 μM concentration at various temperatures (Fig. 3.37, left and right, respectively) are very similar to the spectra of the flexible nucleo-heptapeptide carrying two thymines (Fig. 3.34, left).

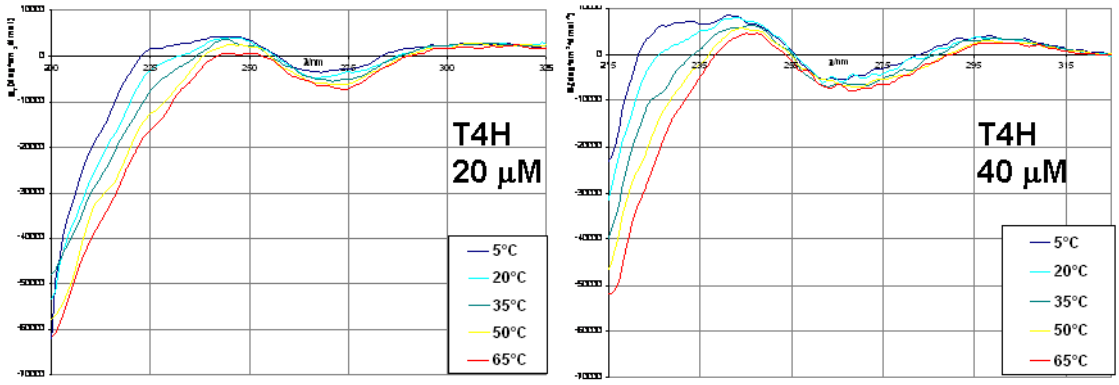


Fig. 3.37. Molar ellipticity of H-Lys(H)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH$_2$ (T4H) in PBS buffer at 20 μM (left) and 40 μM concentration (right) at different temperatures. The same vertical scale has been employed for both spectra ($-70000 \leq [\Theta_T]/\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1} \leq 10000$), while the wavelength range is $200 \leq \lambda/\text{nm} \leq 325$ for the more dilute spectrum and $215 \leq \lambda/\text{nm} \leq 325$ for the more concentrate one (due to the higher absorbance of the more concentrated sample). The average of 8 scans is plotted.

It is possible to see that the intensity of the nucleobase band does not show any significant dependence on peptide concentration, as found for the spectra of the nucleo-heptapeptides. Both spectra show the same variation with the temperature already described for the adenine- and cytosine-based analogues.

In the more diluted sample it appears that no minimum of the negative band at low wavelength can be observed at least until λ=200 nm, as for the nucleo-heptapeptides.

A different and interesting behaviour is observed in the case of the nucleopeptide carrying four guanines. Fig. 3.38 presents its CD spectrum at 20 μM (left) and 40 μM concentration (right) at various temperatures.

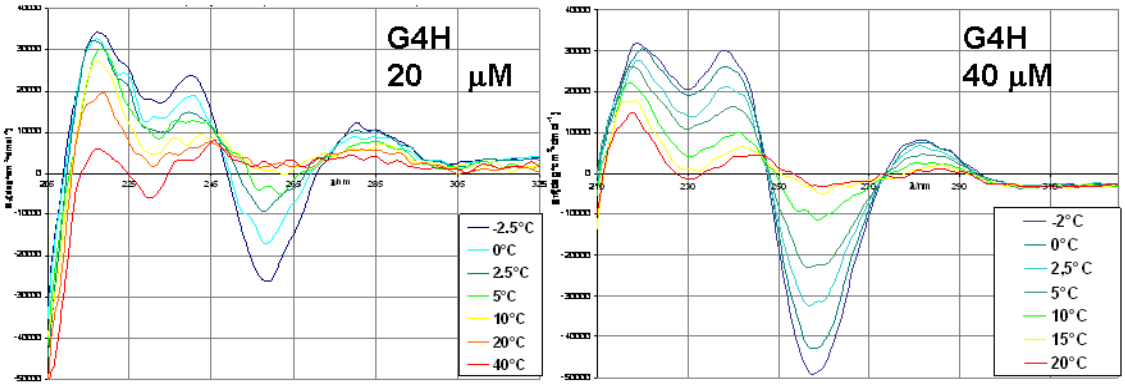


Fig. 3.38. Molar ellipticity of H-Lys(H)-(Ala-AlaG^{OH}-Ala)$_4$-Lys(H)-NH_2 (G4H) in PBS buffer at 20 μM (left) and 40 μM concentration (right) at different temperatures. The same vertical scale has been employed for both spectra (-50000≤ $[\Theta_T]/(deg \cdot cm^2 \cdot dmol^{-1})$≤ 40000), while the wavelength range is 205≤ λ/nm≤ 325 for the more dilute spectrum and 210≤ λ/nm≤ 325 for the more concentrate one (due to the higher absorbance of the more concentrated sample). The average of 8 scans is plotted.

It is easy to see that the negative nucleobase band (with minimum at 258 nm) and two positive bands at its sides are intense at low temperature and decrease rapidly as the temperature rises, until they almost disappear below room temperature. Two isodichroic points can be observed in the same position in both spectra (at about 247 and 273 nm). This finding strongly suggests that an order/disorder transition involving the side-chains bearing the nucleobases is present.

For the spectra acquired at 40 μM concentration, the variation of the intensity of the nucleobase band is unequivocally sigmoidal (Fig. 3.39, left), pointing to a cooperative transition. This is supported by the observation that, at the same concentration, the variation with the temperature of the maximum of the UV absorbance due to the nucleobase chromophore (λ= 260 nm) has a parallel sigmoidal behaviour (Fig. 3.39, right). In both graphs the inflection point can be estimated at about 3°C. Considering the spectra acquired at 20 μM, the variations in both graphs have a similar trend, but the sigmoidal curves seem to be slightly shifted towards lower temperature, so that an estimation of their

inflection point would not be reliable. On the other side, it was not possible to perform the CD measurements at temperatures lower than -2.5°C, due to the risk of freezing the solution. Consequently, given the uncertainty in the evaluation of the transition point for the nucleopeptide at lower concentration, it is not possible to state whether concentration effects are significant (pointing to the involvement of intermolecular interactions for the stabilization of the 'ordered' state at low temperature, as suggested by the cooperativity of the transition) or not.

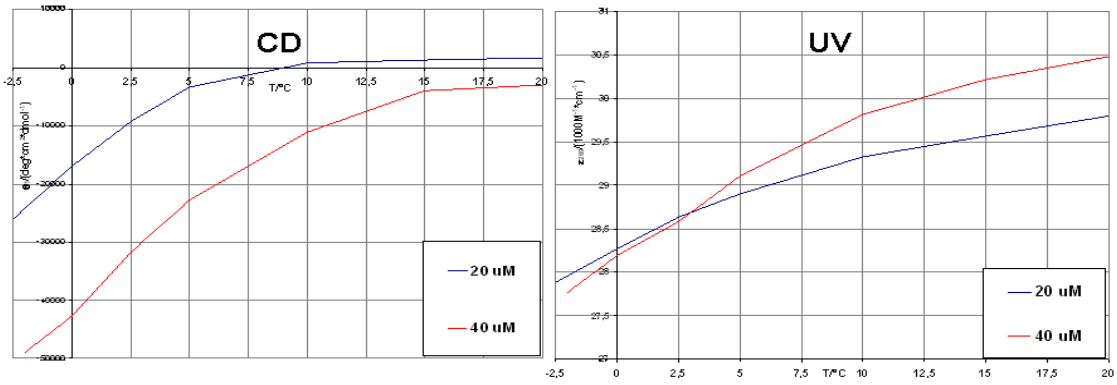


Fig. 3.39. Left: molar ellipticity at 258 nm of H-Lys(H)-(Ala-AlaG^{OH}-Ala)$_4$-Lys(H)-NH$_2$ in PBS at 20 μM and 40 μM concentration as a function of the temperature. Right: molar absorbance of the same nucleopeptide at 260 nm in PBS at 20 μM and 40 μM concentration as a function of the temperature.

The CD spectra of the underivatised nucleopeptide carrying five thymines T5H in phosphate buffer at 20 μM concentration at different temperatures in the range 0-65°C are shown in Fig. 3.40.

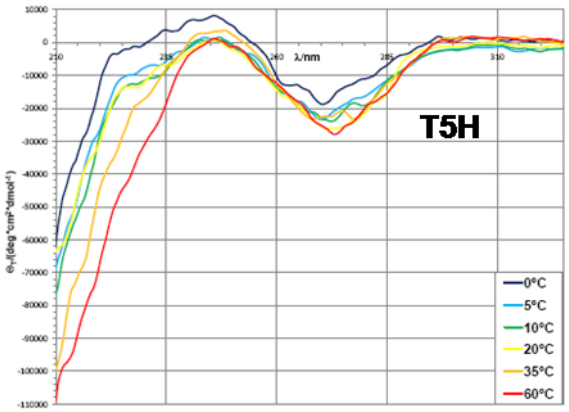


Fig. 3.40. Molar ellipticity of H-Lys(H)-(Ala-AlaT-Ala)$_5$-Lys(H)-NH$_2$ (T5H) in phosphate buffer at 20 μM (left) at different temperatures. Moalr ellipticity range is -110000≤ [Θ_T]/(deg·cm^2·dmol^{-1})≤ 10000), while the wavelength range is 205≤ λ/nm≤ 325. The average of 8 scans is plotted.

It is possible to observe that both the spectral shape and its variation with the temperature are very similar to what found for the spectra of the shorter analogue carrying four thymines, even if both the nucleobase and the strong negative band at low wavelength are significantly stronger. This could suggest that the conformational structuration of the Ala-based nucleopeptides does not change susbstantially with the backbone length.

The most important results deriving from the CD analyses can be summarised as follows:

(I) In general, it has not been observed a simple correlation between the intensity of the nucleobase band and the rigidity of the side-chains bearing the nucleobases.

(II) In the case of the nucleo-hexapeptide, the strong propensity to aggregation with formation of dimers in $CHCl_3$ solution is reflected in the nucleobase band both by intense concentration effects and by its much stronger intensity of relative to the nucleo-tripeptide bearing the same base. On the contrary, no effect has been observed in a polar solvent like methanol, where the dimer is likely not formed.

(III) The comparison of the low wavelength portion of the spectra of the thymine-based nucleo-heptapeptides in aqueous solution could suggest that the Aib-rich peptide is helical, as observed in DMSO solution by NMR, whereas the Ala-rich analogues seem to be less structured.

(IV) A cooperative order/disorder transition involving the side-chains of the nucleopeptide carrying four guanines has been observed at low temperature in aqueous solution.

4. PAIRING PROPERTIES

The pairing properties of rigid nucleopeptides with unprotected nucleobases and flexible nucleopeptides have been investigated.

4.1. Self-pairing properties

As mentioned in the introduction, the Watson-Crick pairing specificity observed in DNA is due to the template effect of the complementary strands, whereas the nucleobases by themselves could interact with a much greater freedom. All nucleobases can self-pair by forming two hydrogen bonds involving their Watson-Crick or Hogsteen binding sites (Fig. 4.1).[56] It must be recalled that 'Watson-Crick pairing mode' has a different meaning than 'Watson-Crick pairing selectivity': the first one refers simply to the relative orientation of the nucleobases and to the binding sites involved, whereas the second one refers also to the identity of the partners (implying exclusive A=T or C≡G pairing).

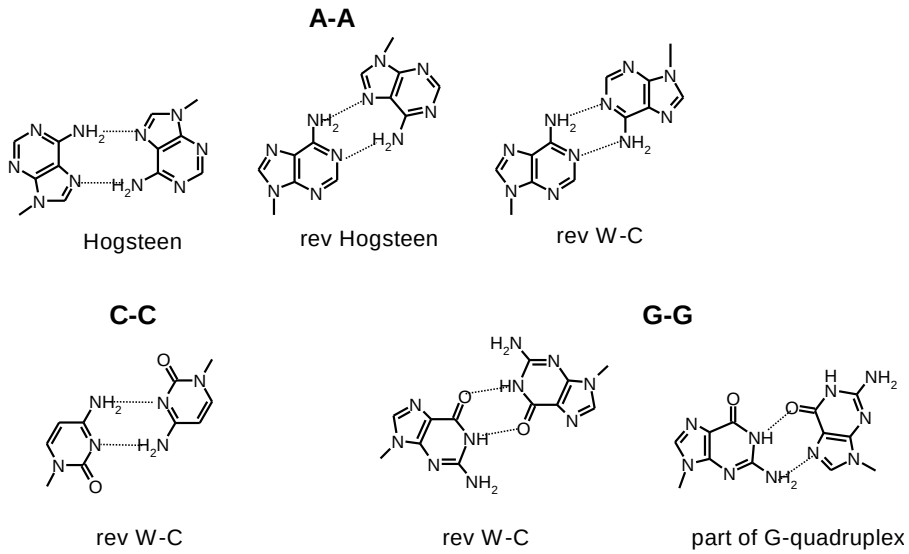


Fig. 4.1. Some of the possible self-pairing modes for adenine, cytosine and guanine; 'rev W-C' and 'rev Hogsteen' refer to the reverse Watson-Crick and reverse Hogsteen pairing modes respectively. Intermolecular hydrogen bonds are represented by dashed lines

Thymine, in particular, has four different self-pairing possibilities, so that several relative positions and orientations of the two nucleobases involved can allow self-pairing. (Fig. 4.2).

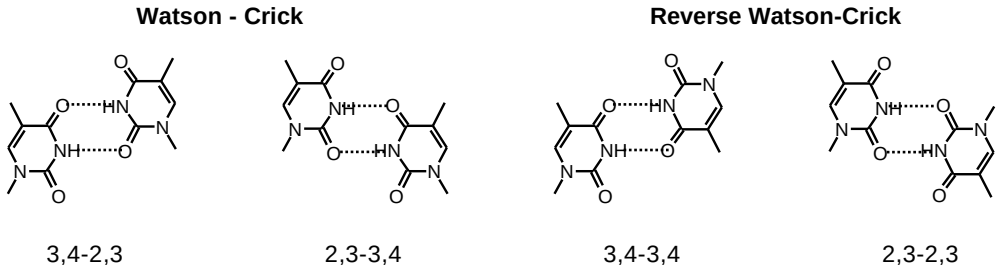


Fig. 4.2. The four possible self-pairing modes for thymine; the numbers under the different modes refer to the complementary functions involved (2: C(2)O, 3: N(3)H, 4: C(4)O). Intermolecular hydrogen bonds are represented by dashed lines

The strong propensity to intermolecular interactions observed by IR spectroscopy for the protected rigid nucleopeptides with free nucleobases in $CDCl_3$ solution (see Section 3.2.1) is likely due to nucleobase self-pairing.

In the case of the thymine-based nucleopeptides, more extensive and conclusive evidence of self-pairing exists. Firstly, the crystal of the protected thymine-based nucleo-tripeptide (see Section 3.1) has a dimer unit cell which is stabilised by thymine-thymine pairing in the reverse Watson-Crick '2,3-2,3' mode. Secondly, the NMR study of the nucleo-hexapeptide dimer formed in $CDCl_3$ solution (see Section 3.3.2) has found that the nucleobases of the two dimer-forming molecules are very close to each-other. Further evidence of thymine-thymine dimer formation is provided by the MS analysis of the protected thymine-based nucleopeptides (Fig. 4.3).

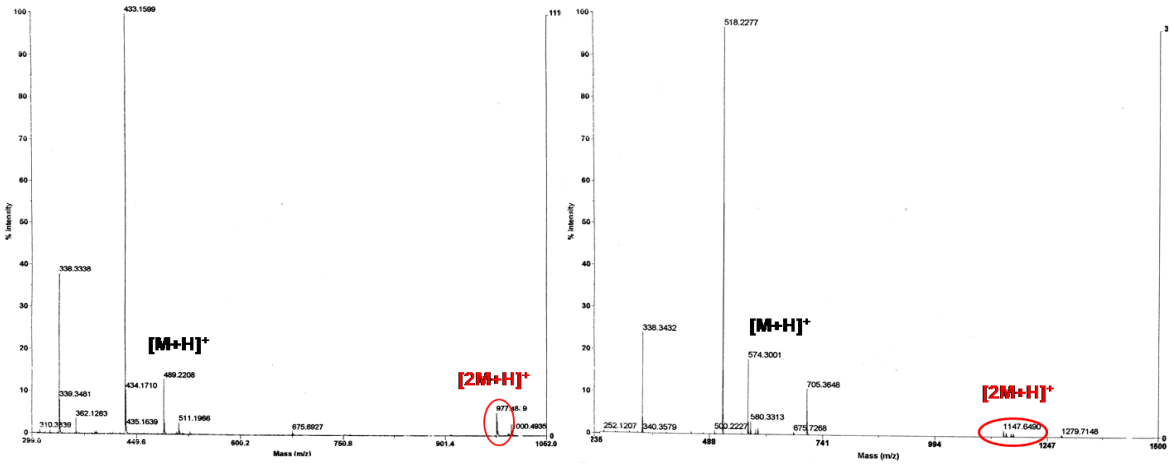


Fig. 4.3. Mass spectra of Z-AlaT-Aib-O^{t}Bu (left) and of Z-Aib-AlaT-Aib-O^{t}Bu (right), with dimer peak highlighted. Z-AlaT-Aib-O^{t}Bu: calcd $[M+H]^+$ = 489.23, $[2M+H]^+$ = 977.46, found $[M+H]^+$ = 489.22, $[2M+H]^+$ = 977.49 (circled peak). Z-Aib-AlaT-Aib-O^{t}Bu: calcd $[M+H]^+$ = 574.29, $[2M+H]^+$ = 1147.56, found $[M+H]^+$ = 574.30, $[2M+H]^+$ = 1147.55 (circled peak). The main peaks in the nucleopeptide spectra (at 433.16 and 518.23 *m/z* respectively) are due to the loss of the *tert*butyl moiety in the acidic media used for injection.

Indeed, in most of the MS spectra of such nucleopeptides a small, but significantly intense peak at m/z= $[2M+H]^+$, indicative of relevant dimer stability, has been observed. The detection of self-pairing even in a polar solvent like MeOH suggests that the interactions involved are quite strong.

4.2 Selective Watson-Crick pairing

One of the crucial requisites of every class of nucleotide analogues which are candidates for applications in the modulation of nucleic acid is a strong preference for the Watson-Crick pairing selectivity.

Different kinds of experiments have been described to test this properties, employing different techniques.[14,47,58,109] Among the instrumentally simplest methods there are spectrophotometrical investigations based on UV-Vis and CD analyses. Such methods rely on the observation that generally, on selective Watson-Crick pairing, the total absorbance of the nucleobase chromophores decreases, due to base-stacking effects, whereas the dichroic signal due to the nucleobases is more intense, due to the conformational rigidity brought about by ordered pairing.[37,56] If the sample is gradually perturbed favouring dimer dissociation, for instance by rising the temperature, the fraction of dissociated partners formed gives a different contribution to the UV or CD spectra. In case of cooperative transitions a sigmoidal variation of the spectra with the perturbation is observed and this allows the individuation of the so-called melting point of the complex (the conditions under which half of the partners are engaged in a dimer).

Therefore, the pairing properties of complementary rigid and flexible nucleopeptides were tested at first by CD measurements. In the case of the rigid nucleo-tripeptides with unprotected nucleobases, the tests were performed on the adenine-thymine pair. The cytosine-guanine nucleopeptide pair was not considered, given the insolubility of the latter in $CHCl_3$. Indeed, given that the nucleo-tripeptides carry only a nucleobase each, and that pairing is generally strong only when cooperative interactions are involved, it was feared that intermolecular interactions would have not be strong enough to be detected in a polar and H-bond forming solvent like MeOH. Analyses in $CHCl_3$ were therefore considered necessary to enhance the probability of detection of weak pairing phenomena, whereas measurements in MeOH were considered useful for a comparison of the intensity of the effects that could have been observed.

Mixtures of Ac-Aib-AlaT-Aib-O^{t}Bu and of Boc-Aib-AlaA^H-Aib-OMe in $CHCl_3$ or MeOH solution at 1:1 molar ratio with a total concentration ranging from 1 mM to 100 μM have been studied at room temperature by CD technique. However no evidence of specific pairing has been detected, since spectra are not significantly different from the sum of the contributions of the single nucleopeptide solutions at the appropriate concentration.

In the case of the underivatized flexible nucleopeptides carrying four nucleobases each, CD tests have been performed on both Watson-Crick pairs. Mixtures of equal molar amounts of complementary nucleopeptides in PBS buffer at total concentrations around 25 μM have been studied by CD at different temperatures in the range -2.5/65°C. In the case of the cytosine-guanine pair, the intensity of the dichroic bands due to the nucleobase contributions is slightly different from the sum of the contribution of the single nucleopeptide solution, and the difference is temperature-dependent (Fig. 4.4), suggesting that pairing could occur at low temperature. In this case, however, the nucleobase band of the mixture appears less intense than the sum of the contributions of the single nucleopeptides, instead of being more intense. Considering also that the CD spectra of the guanine-based nucleopeptide (see Section 3.4.4) display a temperature-dependent variation of the intensity of the nucleobase band, it can not be excluded that the observed behaviour results from a perturbation of the guanine-based nucleopeptide contribution which is not related to specific pairing, but to the increased disorder of its side chain.

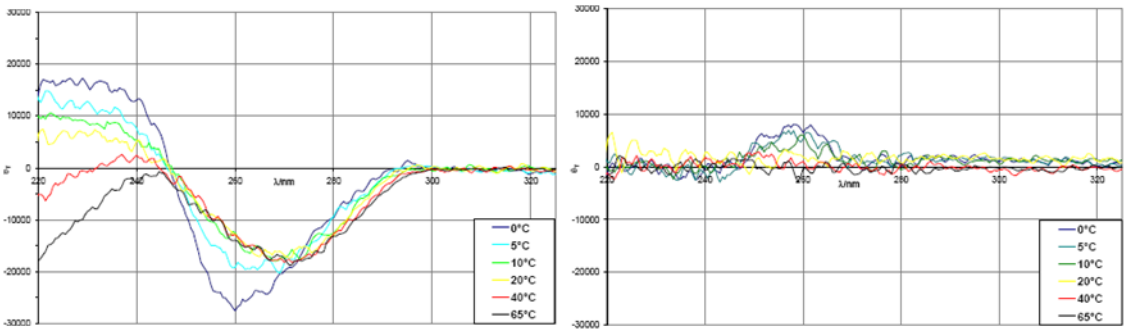


Fig. 4.4. Left: CD spectrum of a 1:1 mixture of H-Lys(H)-(Ala-AlaC^H-Ala)$_4$-Lys(H)-NH_2 and of H-Lys(H)-(Ala-AlaG^{OH}-Ala)$_4$-Lys(H)-NH_2 in PBS buffer at 20 μM total concentration as a function of the temperature. The average of 6 scans is plotted. Right: the same spectrum after subtraction of the contributions of the single nucleopeptides at 10 μM concentration as a function of the temperature. Spectra are drawn at the same scale ($220 \leq \lambda/nm \leq 325$, $-30000 \leq [\Theta_T]/(deg \cdot cm^2 \cdot dmol^{-1}) \leq 30000$). The resulting positive band is better seen as a decrease in the intensity of the negative bands of the single nucleobases and particularly given its position, of guanine.

Considering that it was possible that pairing between nucleopeptides and complementary nucleotides was stronger and therefore easier to detect than pairing among complementary nucleopeptides, CD tests of nucleopeptide/nucleotide pairing have been performed. The longest flexible nucleopeptide synthesized, carrying five thymine nucleobases, was used and a synthetic DNA oligonucleotide carrying ten adenine nucleobases (dA_{10}) was chosen as a complementary partner. The choice of a longer nucleotide was done in order to take advantage of the *flanking base effect*: in a nucleotide strand the nucleobases next to the 5'- and 3'-termini do not contribute to pairing, but stabilise the conformation of the internal bases, which are therefore capable of a more efficient binding.

Mixtures of the two compounds at various nucleopeptide/nucleotide molar ratios (1:2, 1:1, 2:1, 4:1) in phosphate buffer at a total concentration of 10 μM have been prepared and their CD spectra have been measured at different temperatures in the range 0-65°C. The use of different molar ratios could give useful information for the understanding of the pairing mode in case interactions were detected. For instance, if A:T pairing following Watson-Crick rules was prevalent, the strongest effects would be detected for a 1:1 nucleopeptide/nucleotide ratio, whereas in case of (T:A)T triplex formation, the strongest effects would be observed for a 2:1 nucleopeptide/nucleotide ratio. The spectra obtained from the mixtures, however, do not significantly differ from the weighted sums of the contributions of the nucleopeptide and of the nucleotide.

4.2.1 Surface plasmon resonance measurements

An alternative technique for detecting and evaluating pairing is surface plasmon resonance (SPR).[141] This techniques relies on the variations of the optical properties of a gold chip in response to the variation of mass brought about by binding of ligands to the suitably functionalised surface of the chip. The instrumental response can be followed against time to obtain the binding kinetics in the a*ssociation phase*, during which a solution of the ligand is allowed to flow on the chip and bind to its surface, and in the *dissociation phase*, during which a solution without ligand flows on the chip, allowing elution of the ligand. By fitting the instrumental data collected during the association and dissociation phase of the experiment it is possible to calculate the kinetic association and dissociation constant (k_a and k_d, respectively) and their ratio allows to obtain the thermodynamic dissociation constant ($K_d = k_d/k_a$). After the experiments, the chip surface can be regenerated by

washing with an energic eluent, in general aqueous HCl, in order to eliminate all substances not covalently bound, so that the chip is ready for new experiments.

Since a Biacore^{TM} SPR system was available in Strasbourg, it has been employed for pairing measurements on the flexible nucleopeptides. As mentioned above, in order to perform such SPR measurements, it is necessary to immobilise one of the two partners on solid-phase, whereas the other is allowed to flow in solution. For the immobilisation of the nucleopeptides, the biotinylated flexible nucleopeptides have been used, after preliminary activation of the dextran-functionalized gold surface to obtain covalent binding of streptavidin,[142] which in turn complexes almost irreversibly ($K_d \approx 10^{-15}$ M) the biotinylated nucleopeptides.

Two advantages of this method are that it is well-established and easy to apply. Moreover, since the biotinylated compounds are stable to the regeneration phase, several test can be performed without the need to repeat the immobilisation procedure.

On the other side, this method brings about limitations to the instrumental sensitivity. Indeed, the bulkier is the ligand and the shorter is the distance between the ligand and the gold surface, the stronger is the instrumental signal obtained Given that streptavidin is a protein with $M_r \approx 50000$ and the nucleopeptide ligands have $M_r \approx 2000$, this means that the effects of the binding of small ligands are communicated to gold surface through a bulky linker, negatively affecting the instrumental sensitivity.

Adenine-, cytosine- and thymine-based biotinylated nucleopeptides with four nucleobases were bound to the functionalised gold sensor chip. For the first series of experiments, solutions of guanine-based and thymine-based underivatised nucleopeptides in the concentration range 12.5-100 μM have been allowed to flow on the chip. Binding has been detected in the case of the adenine-thymine Watson-Crick pair (Fig. 4.5) with a dissociation constant $K_d \approx 30$ μM.

Unfortunately, pairing formation has not been detected between the cytosine- and the guanine-based nucleopeptide, whose nucleotide analogues display a stronger binding compared to adenine-thymine pairing in DNA.

On the other side, pairing between non complementary nucleopeptides has not been observed (see Fig. 4.6 as an example). Given that four non complementary pairs have been screened (T4H against T4Bt and C4Bt, G4H against A4H and A4Bt), respect of Watson-Crick selectivity, in this preliminary test, is considered encouraging.

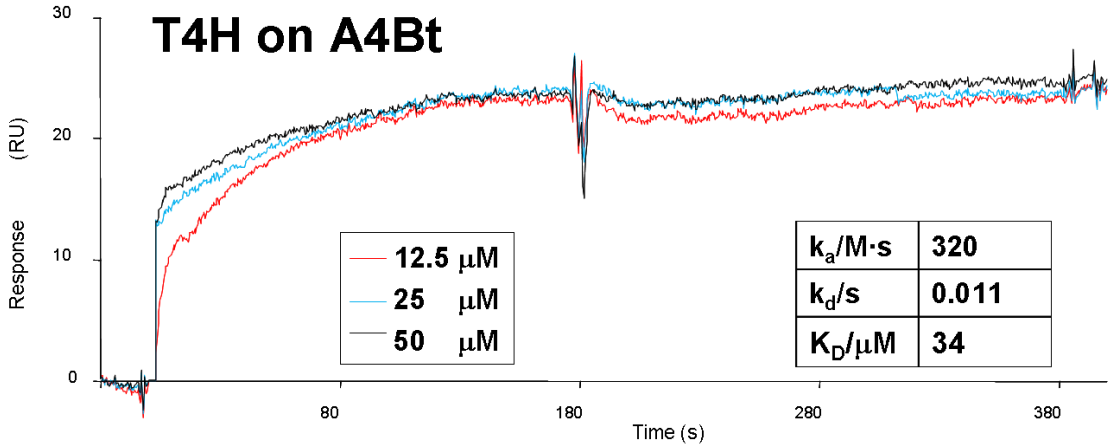


Fig. 4.5. SPR sensorgrams obtained from the pairing experiment between the biotinylated adenine-based nucleopeptide A4Bt and the underivatised thymine-based nucleopeptide T4H. The rise of the signal during the association phase and the its decrease during the dissociation phase give evidence of the pairing.

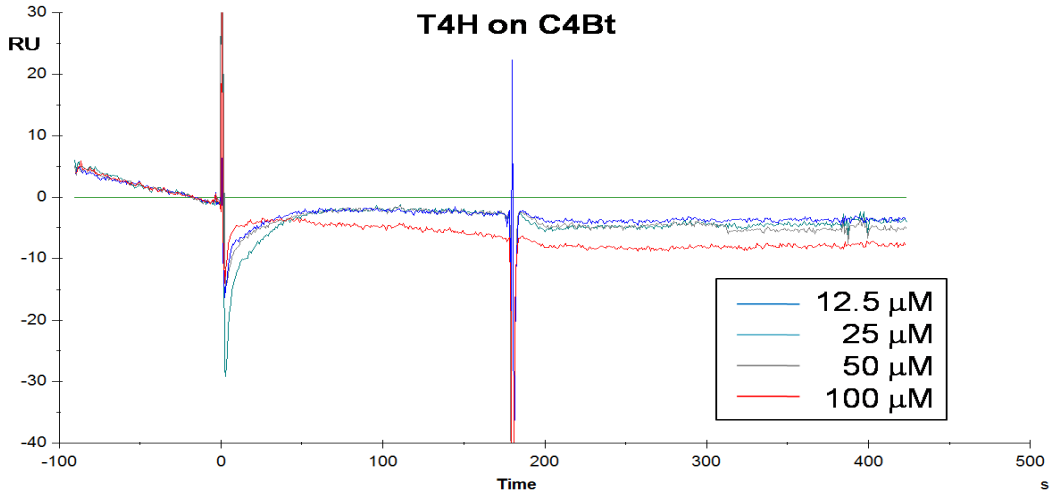


Fig. 4.6. SPR sensorgrams obtained from the pairing experiment performed between the biotinylated cytosine-based nucleopeptide C4Bt and the underivatised thymine-based nucleopeptide T4H. No pairing is detected, as confirmed by the absence of signal change the association phase.

The pairing properties of the adenine-based biotinylated nucleopeptide towards nucleotides have been evaluated by allowing complementary RNA (U_4, U_8) and DNA (dT_8, dT_{10}) oligonucleotides to flow on the functionalised gold chip. Pairing has been detected with all complementary nucleotides (see Fig. 4.7 as an example). The affinity towards the RNA nucleotides is similar to the affinity towards complementary nucleopeptides, whereas the affinity towards DNA oligonucleotides is about ten times stronger (Table 4.1).

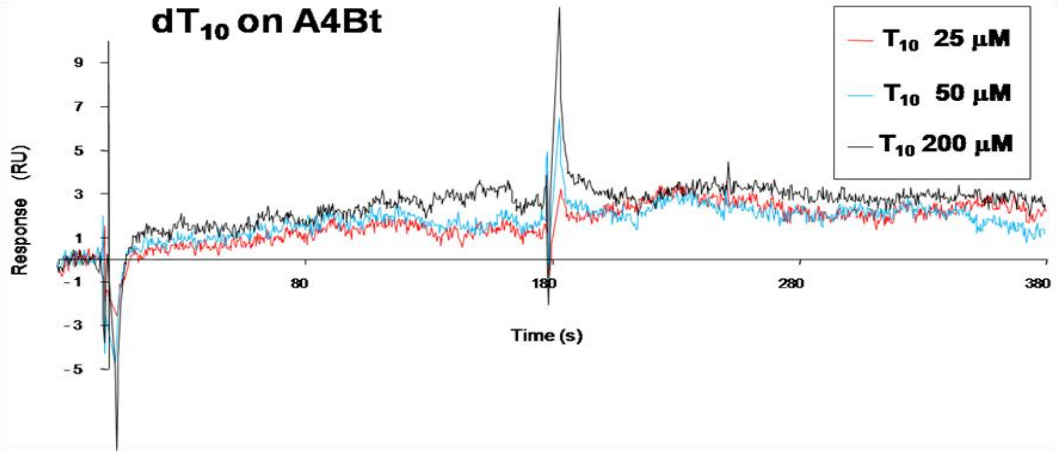


Fig. 4.7. SPR sensorgrams obtained from the pairing experiment performed between the biotinylated adenine-based nucleopeptide A4Bt and the DNA oligonucleotide dT_{10} at various concentrations. Pairing is observed even if the weak signal observed makes the spectrum noisy.

From the data presented in Table 4.1 it appears that no significant difference exists in the binding strength of the two RNA oligonucleotides. This seems in contrast with the flanking base effect, but, given the higher conformational flexibility of RNA compared to DNA, it is possible that the RNA octamer is so conformationally unstable that even the bases in the middle of its sequence are not significantly more structured than the bases of its shorter analogue.

Table 4.1. Thermodynamic dissociation constants calculated for the pairing of different RNA and DNA nucleotides with the biotinylated adenine-based nucleopeptide A4Bt compared with the dissociation constant calculated for the complementary thymine-based nucleopeptide T4H.

ABt vs	**T4H**	**U_4 (RNA)**	**U_8 (RNA)**	**dT_8 (DNA)**	**dT_{10} (DNA)**
K_d/ µM	34	36	41	3.4	3.0

Notes. Given the weak signal detected, the relative uncertainty on the dissociation constants is quite high in the case of the DNA oligonucleotides.

In this series of experiments the signals observed is weak because of the unfavourable instrumental effects brought about by the immobilisation method employed. The signal to noise ration is consequently low, and the relative uncertainty in the values of the binding constants is correspondingly rather high.

Again no nucleotide binding towards the immobilised non complementary nucleopeptides based on thymine or on cytosine has been observed (see Fig. 4.8 as an example).

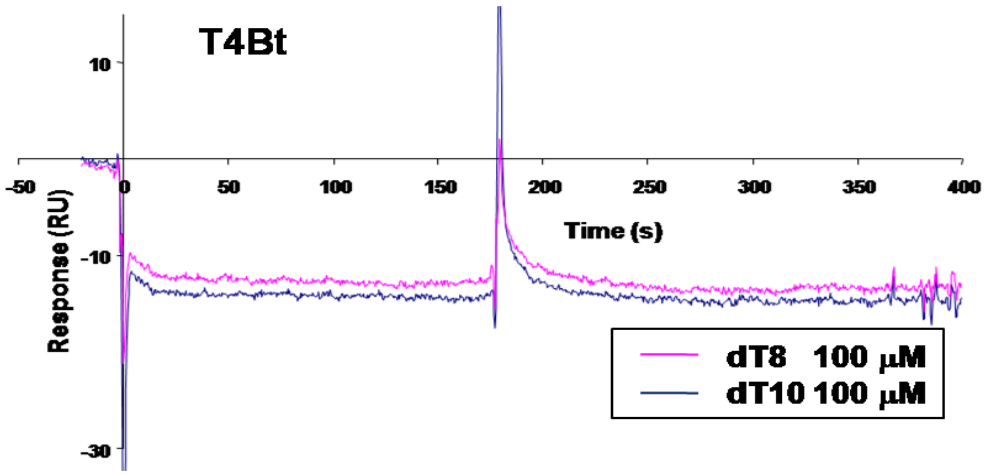


Fig. 4.8. SPR sensorgrams obtained from the pairing experiment performed between the biotinylated thymine-based nucleopeptide T4Bt and two DNA oligonucleotides (dT_8 and dT_{10}) at the maximum concentration tested. No pairing is detected.

Given that the adenine-thymine Watson-Crick pair provided interesting results, a second series of pairing experiments focusing on the same nucleobase pair has been performed. For such experiments two of the longest nucleopeptides synthesized, containing five thymines, have been used (the biotinylated nucleopeptide T5Bt has been immobilised on the chip, whereas the underivatised nucleopeptide T5H has been employed in solution).

It was hoped that in the experiments involving the nucleopeptides with five bases a significantly stronger binding would have been observed compared with the nucleopeptides with four bases, both because of the possible flanking base effects and because of binding cooperativity. Instrumentally, the analysis is also favoured, since the ligand is bulkier and therefore causes a stronger signal on binding.

Both thymine- and adenine-based DNA nucleotides (dT_{10}, dA_{10}) have been employed for the tests. Moreover, a biotinylated adenine-based nucleotide (Bt-dA_{10}) has been immobilised on the sensor chip, in order to see whether pairing properties were affected by the identity of the immobilised partners. Results of the measurements are summarized in Table 4.2.

Table 4.2. Dissociation constants calculated from SPR binding experiments

Entry	Immobilised species	Ligand	Binding	K_d (M)
1	Bt-dA_{10}	A4H	no	-
2	Bt-dA_{10}	T4H	yes	$3.53 \cdot 10^{-4}$
3	Bt-dA_{10}	T5H	yes	No good fit for KD
4	Bt-dA_{10}	dT_{10}	yes	$9.12 \cdot 10^{-5}$
5	Bt-dA_{10}	dA_{10}	no	-
6	T5Bt	A4H	yes	No good fit for KD
7	T5Bt	T4H	no	-
8	T5Bt	T5H	no	-
9	T5Bt	dT_{10}	no	-
10	T5Bt	dA_{10}	yes	No good fit for KD
11	A4Bt	A4H	yes **(*)**	$1.79 \cdot 10^{-6}$
12	A4Bt	T5H	yes	$2.43 \cdot 10^{-6}$
13	A4Bt	dA_{10}	no	-
14	T4Bt	A4H	yes	$5.01 \cdot 10^{-6}$
15	T4Bt	T5H	yes **(*)**	$2.03 \cdot 10^{-5}$
16	T4Bt	dA_{10}	Yes	(**)

Notes. (*) Pairing among non-complementary strands. (**) Pairing is so weak that it is not possible to precisely evaluate the dissociation constant ($K_d > 5 \cdot 10^{-4}$ M). Among the 20 independent experiments which are possible using four immobilised species (Bt-dA_{10}, T5Bt, A4Bt, T4Bt) and five ligands in solution (A4H, T4H, T5H, dT_{10}, dA_{10}), four have been performed in the first series of experiments (A4Bt/T4H, A4Bt/dT10, T4Bt/T4H, T4Bt/dT_{10}) and, having already been discussed above, are not reported in this table.

Table 4.2 deserves some comments. The pairing strength between complementary nucleotides dT_{10} and Bt-dA_{10} has been tested (entry 4) to have the possibility of a comparison with the values obtained for the nucleopeptides. Considering the values in the Table, it is possible to realise that their dissociation constant ($K_d \approx 90$ μM) is significantly higher (namely, binding is significantly weaker) than the values observed for nucleopeptide/nucleotide pairing ($K_d \approx 3$ μM in the case dT_{10} on A4Bt, Table 4.1).

The opposite electrostatic contributions in the case of the nucleopeptide/nucleotide pairing compared to the nucleotide/nucleotide pairing have likely a significant effect, since each nucleotide has about 10 negative charges and each nucleopeptide has 3 positive charges. However, if nucleopeptide/nucleopeptide interactions are considered, for which electrostatic forces are also repulsive, binding remains stronger than nucleotide/nucleotide

binding. Taking also into account that the complementary nucleotides have 10 nucleobases each and the complementary nucleopeptides 5 or less, it can be inferred that pairing of the latter has a considerable strength.

Among 16 entries (20 including the experiments performed before, see the footnote of Table 4.2), there are two cases of unwanted pairing of non complementary strands (A4Bt/A4H, entry 11, and T4Bt/T5H, entry 15). The case of entry 11 is indeed peculiar, since it displays the strongest affinity that it has been possible to calculate ($K_d \approx 1.8$ μM). On the other side, considering the two thymine-based partners involved in non-complementary pairing of entry 15, both display significant stronger binding of adenine-based nucleopeptides (4 times stronger for T4Bt/A4H, entry 14, and 8 times stronger for A4Bt/T5H, entry 12). Moreover, in the case of the biotinylated nucleopeptide carrying five thymine bases, T5Bt (entries 6-10), Watson-Crick selectivity is always respected.

The stronger selectivity displayed by the nucleopeptide carrying five thymines than that shown by its shorter analogue is encouraging, since it is likely an hint that for long nucleopeptide sequences Watson-Crick selectivity, only weakly favoured in the case of interactions between single nucleobases, is able to prevail due to cooperative effects, as observed for oligonucleotide pairing.

Further data supporting cooperativity in pairing can be obtained by comparing the binding strength observed for the nucleopeptides containing four or five thymines towards complementary adenine-based nucleopeptides or nucleotides. Indeed the dissociation constant observed for the binding of T4H on A4Bt (34 μM, Table 4.1) is about fifteen times higher than the one observed for T5H on A4Bt (2.4 μM, Table 4.2, entry 12), namely binding of the latter is about fifteen times stronger. If pairing of each nucleobase was completely independent, a modest increase of the binding strength would be expected (+20% if all nucleobases can bind, +50% if the nucleobases next to the termini can not contribute). Similarly, binding of the adenine-based oligonucleotide dA_{10} to T4Bt is so week that it has not been possible to obtain a reliable binding constant (entry 16), whereas the binding experiment involving dA_{10} and T5Bt has detected very strong binding (see below).

In three of the experiments involving the nucleopeptides carrying five thymines (entries 3, 6, 10) binding has been positively detected, but fitting of the instrumental data has not been possible, thus preventing the evaluation of the binding constant. This could be due to the simultaneous presence of different pairing interactions (e.g. duplex and triplex

formation). In case of entries 6 and 10 the binding strength, as evaluated by signal increase during the association phase, is comparable to the one observed in other experiments under similar conditions (see Fig. 4.9, as an example).

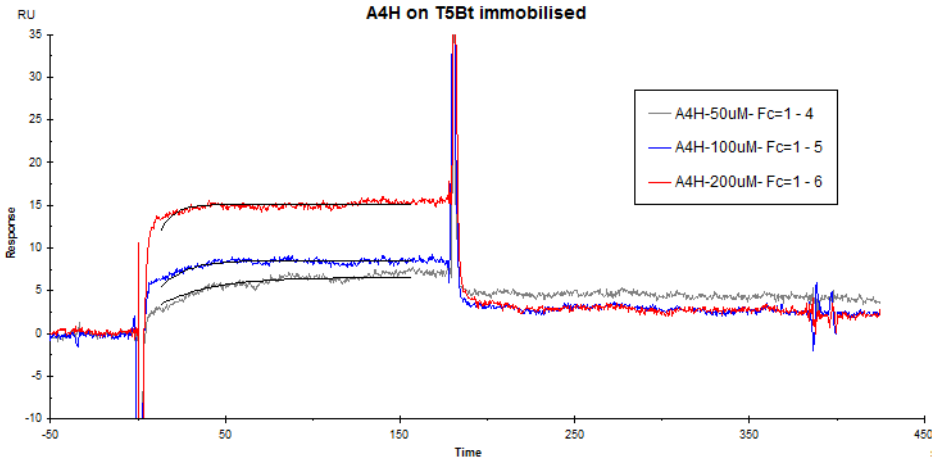


Fig. 4.9. SPR sensorgrams obtained from the pairing experiment performed between the biotinylated thymine-based nucleopeptide T5Bt and the adenine-based nucleopeptide A4H Binding is observed, but data fitting has not been possible.

On the other side, in case of the pairing experiment involving the nucleopeptide T5H to the biotinylated nucleotide Bt-dA_{10} (entry 3) binding is much stronger (Fig. 4.10).

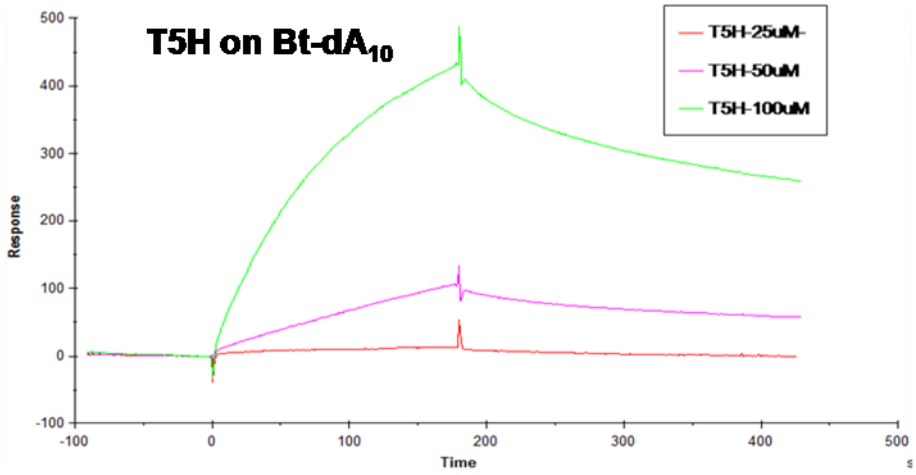


Fig. 4.10. SPR sensorgrams obtained from the pairing experiment performed between the biotinylated adenine-based nucleotide Bt-dA_{10} and the thymine-based nucleopeptide T5H Binding is observed, but data fitting has not been possible.

Indeed the signal observed for the nucleopeptide at 50 μM concentration varies of almost 100 RU (*resonance units*, 1 RU corresponds about to1 pg analyte/mm^2), against 10-20 RU usually observed in the other experiments where pairing has been detected. Given that signal variation is significantly more than proportional to the concentration, it is possible

that more than one ligand molecule interacts with each immobilised nucleotide, probably because of triplex formation.

For a better understanding of this interesting behaviour, the experiment has been repeated under slightly different experimental conditions to have the possibility of a more efficient fitting, but results of the calculations are not available yet.

The most important results obtained from the study of the nucleopeptide pairing properties can be summarised as follows:

(I) Evidence of self-pairing has been found for the rigid nucleopeptides with unprotected nucleobases. In particular self-pairing has been observed for several rigid thymine-based nucleopeptides by XRD, NMR and MS analyses.

(II) It has not been possible to detect pairing of complementary nucleopeptides, neither rigid nor flexible, by CD analyses.

(III) SPR measurements on the flexible nucleopeptides have detected binding for the adenine-thymine Watson-Crick pair both among complementary nucleopeptides and between nucleopeptides and nucleotides. Binding displays a strong preference for the Watson-Crick selectivity, and by comparing nucleopeptides with four and five nucleobases a better selectivity and a significantly stronger pairing has been found, suggesting cooperative binding.

5. BIOLOGICAL TESTS

Due to the absolute necessity of operating in conditions compatible with the metabolism of living systems, the investigation of the biological properties of the nucleopeptides has been possible only in the case of the water-soluble, deprotected peptides. Most of the tests have been therefore performed on the flexible nucleopeptides, part of which had been suitably derivatised on purpose.

Firstly the pro-apoptoticity and the cytotoxicity of the nucleopeptides have been evaluated, then their cell penetration properties have been studied and finally the results of preliminary stability tests are reported.

5.1 Apoptoticity tests

The absence of pro-apoptotic effects of the flexible nucleopeptides has been verified using the *DiOC$_6$-essay*,[143] based on the properties of the molecular probe DiOC$_6$ (3,3'-dihexyloxacarbocyanine iodide). This molecule is a fluorophore able to bind to the mitochondrial membrane as a function of the transmembrane electrochemical potential ($\Delta\Psi_m$). In apoptotic cells, the transmembrane electrochemical potential falls to zero, so that the probe does not bind to their mitochondrial membrane (Fig. 5.1).

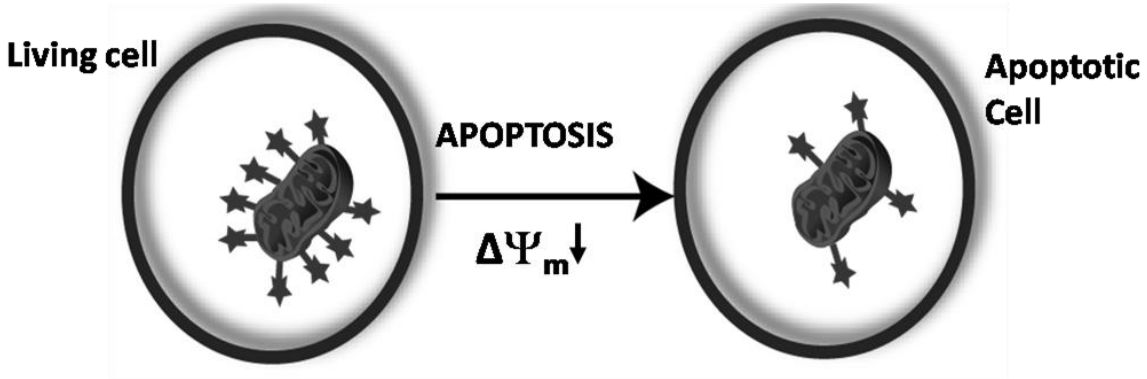


Fig. 5.1. Biochemical basis of the DiOC$_6$ essay. The circles represent the cells, the bean-shaped objects represent the mitochondria, whereas the star-shaped objects represent molecules of the fluorescent probe DiOC$_6$. The drawing is not in scale.

If, after incubation with a DiOC$_6$ solution, cells are washed and their fluorescence is analyzed by flow-cytometry techniques, it is possible to obtain a graph plotting the

distribution of the fluorescence intensity. In general two distinct populations are obtained, one with stronger fluorescence, due to living cells, and one with weaker fluorescence, due to apoptotic cells. Cells treated with a pro-apoptotic substance will display a significant increase in the apoptotic fraction compared to untreated control cells. Therefore, this test allows detection and quantification of pro-apoptotic properties.

In the specific case of the test performed on the nucleopeptides, human lymphoid Jurkat cells have been incubated for 24 hours with the underivatised nucleopeptide CTCCH, carrying three cytosines and one thymine, at 10 and 50 μM concentrations. The flow-cytometry analysis after treatment (Fig. 5.2) has demonstrated that the apoptotic fraction observed for the treated cells is not significantly different from that found for control cells.

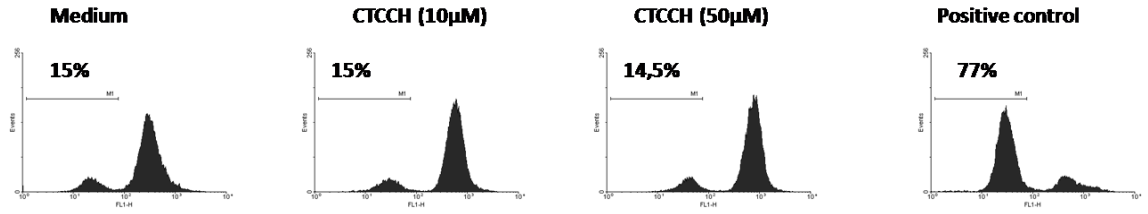


Fig. 5.2. Fluorescence intensity distribution and apoptotic cell fractions for Jurkat cells incubated for 24 hours in culture medium (left), with addition of nucleopeptide CTCCH at the given concentrations (middle) or of a known pro-apoptotic substance (doxorubicin) as a positive control.

From this experiment it is possible to infer that the nucleopeptide is not proapoptotic even at significant concentrations.

5.2 Cytotoxicity evaluation

Nucleopeptide cytotoxicity has been evaluated both by flow-cytometry analysis and by a colorimetric test. Tests have been performed on different nucleopeptides in order to gain evidence that the non cytotoxicity is not dependent on the nucleobase sequence.

Human lymphoid tumor Raji cells incubated with the nucleopeptide CTCCH for increasing times up to 16 hours at concentrations up to 50 μM have been analyzed by flow-cytometry. By considering the intensities of forward and backward scattering, it is possible to generate a graph, presenting on its axes the cell size and the cell morphological structuration.[144] The resulting distribution (*dot plot*) is dependent on the cell type and on

the cell status. Dead and viable cells generally appear as distinct populations, so that the dead cell fraction can be evaluated. The flow-cytometry analysis of the treated cells has shown no significant increase in the dead cell fraction (see Fig. 5.3).

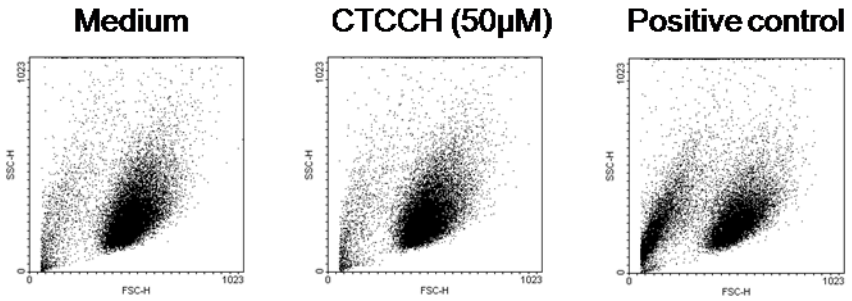


Fig. 5.3. Graphs (dot plots) representing the distributions of size and morphological structuration of Raji cells after 16 hour incubation in culture medium (left), with the addition of the nucleopeptide CTCCH 50 μM (middle) or treated with a toxic substance (doxorubicin) as a positive control (right). In all graphs the population on the left is composed by non viable cells.

The cytotoxicity of the nucleopeptides has been evaluated also by the colorimetric *MTS test*.[145] The name of the test derives from 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxyphenyl)-2-4-(sulfophenyl)-2-H-tetrazolium (MTS), a colourless molecule which, in living cells, is enzymatically reduced by a mitochondrial dehydrogenase to a coloured derivative called Formazan, with absorption maximum at λ= 492 nm (Fig. 5.4).

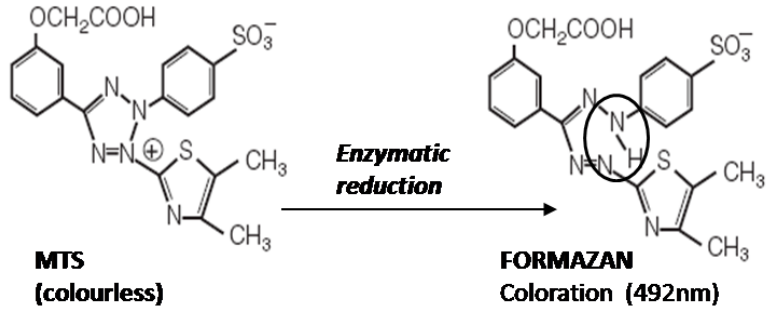


Fig. 5.4. Molecular structure of 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxyphenyl)-2-4-(sulfophenyl)-2-H-tetrazolium (MTS) and of the coloured derivative generated by enzymatic reduction in living cells. The N-N bonds which is reductively cleaved is circled.

Under certain experimental conditions, the intensity of the absorption developed after treatment with MTS is proportional to the concentration of viable cells.

Firstly, the MTS test has been performed on the nucleopeptide CTCCH at concentrations up to 50 μM and for incubation times up to 16 hours, confirming the non cytotoxicity of this nucleopeptide. After that, the test has been repeated on the

underivatised and biotinylated thymine-based nucleopeptides T4H and T4Bt. Three different human lymphoid tumor cell lines (Jurkat, Raji, BL41) have been incubated with increasing concentrations of the nucleopeptides up to 100 μM for 24 hours and no decrease in cell viability has been observed (Fig. 5.5).

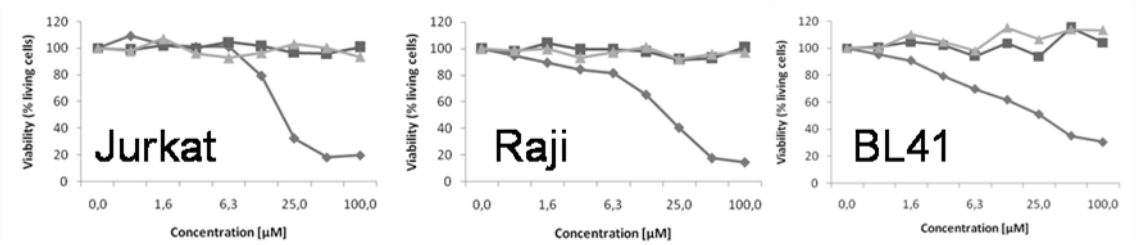


Fig. 5.5. Graphs displaying the cell viability evaluated by MTS test for Jurkat, Raji and BL41 cells treated for 24 hours with the nucleopeptides T4H (triangles) and T4Bt (squares), as well as with doxorubicin as a positive control (diamonds), at different concentrations. The average of three independent experiments is plotted. Results are normalized on the mean absorbance of cells incubated for 24 hours in the culture medium.

The test has been performed also on the rigid thymine-based nucleo-heptapeptide H-Lys(H)-(Aib-AlaT-Aib)$_2$-Lys(H)-NH$_2$ (6TTK) to check whether the more structured conformation and the presence of several C$^{\alpha}$-tetrasubstituted residues could be of harm to cell viability. Fig. 5.6 displays the results of the test on Raji cells after 24 hour treatment with increasing concentrations of nucleopeptide up to 100 μM: again no decrease in viability is observed.

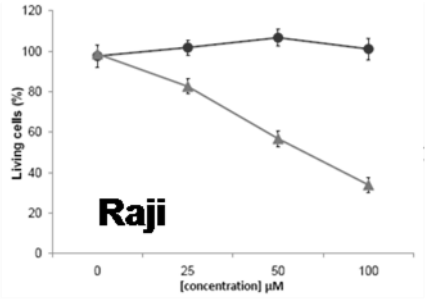


Fig. 5.6. Graph displaying the cell viability evaluated by MTS test for Raji cells treated for 24 hours with the rigid nucleopeptide 6TTK (circles) and with doxorubicin as a positive control (triangles) at different concentrations. The average of three independent experiments and the standard deviation observed are plotted. Results are normalized on the mean absorbance of cells incubated for 24 hours in the culture medium.

From the results of the tests, it can be inferred that both rigid and flexible nucleopeptides do not display cytotoxic effects even at high concentrations. This is of interest, given that other kinds of nucleotide analogues, such as PNAs, have generally remarkable toxicity.[18] Given that the performed tests evaluate the cytotoxicity as a decrease in cell viability, the observation of the absence of cytotoxic effects implies also the absence of pro-apoptotic effects of the compounds tested.

5.3 Cell penetration experiments

Cell penetration experiments have been performed employing flow-cytometry and fluorescence microscopy. Both tests are based on the detection of fluorescent probes inside the cells after nucleopeptide penetration.

The flow-cytometry test has been operated using the fluorescent thymine-based nucleopeptide T4FIT. Raji cells have been incubated for 16 hours with T4FIT at 50 μM concentration, then cells have been washed several times with culture medium before analysing the distribution of their fluorescence intensity at the characteristic emission wavelength of the fluorophore (518 nm). By comparing the observed distribution with the distribution found for untreated cells, a 30-fold increase in the fluorescence intensity of the maximum of the distribution has been detected (Fig. 5.7).

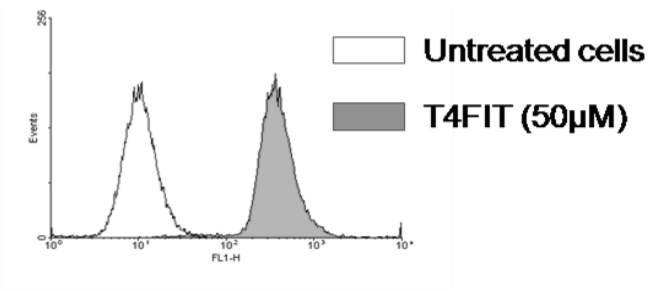


Fig. 5.7. Semilogaritmic plot of the fluorescence intensity distribution for untreated Raji cells and for Raji cells incubated for 16 hours with the nucleopeptide T4FIT at 50 μM concentration.

In a first series of cell penetration experiments evaluated by microscopy, murine renal tumor Renca cells have been incubated with the same nucleopeptide 50 μM for 16 hours, washed several times, then treated with FM4-64.[146] FM4-64, or N-(3-triethylammoniumpropyl)-4- (*p*-diethylaminophenyl-hexatrienyl) pyridinium dibromide, is a lipophilic stiryl dye with intense red fluorescence. This is useful because the dye strongly

binds to the plasma membrane and gradually penetrates into cell cytoplasm, enabling cell detection by fluorescence microscopy.

After preparation, cells have been observed with a laser confocal microscope.[147] Because of their instrumental setup, confocal microscopes focus on very thin sections of the objects observed, so that the danger of misinterpretating superimposed objects in different planes as objects lying close together in the same plane is avoided. Fig. 5.8 presents two of the images acquired, observed in the channels relative to the fluorescence of the nucleopeptide and of the FM4-64 dye, as well as the merge of the two channels, unambiguously demonstrating that nucleopeptides are found inside the cells.

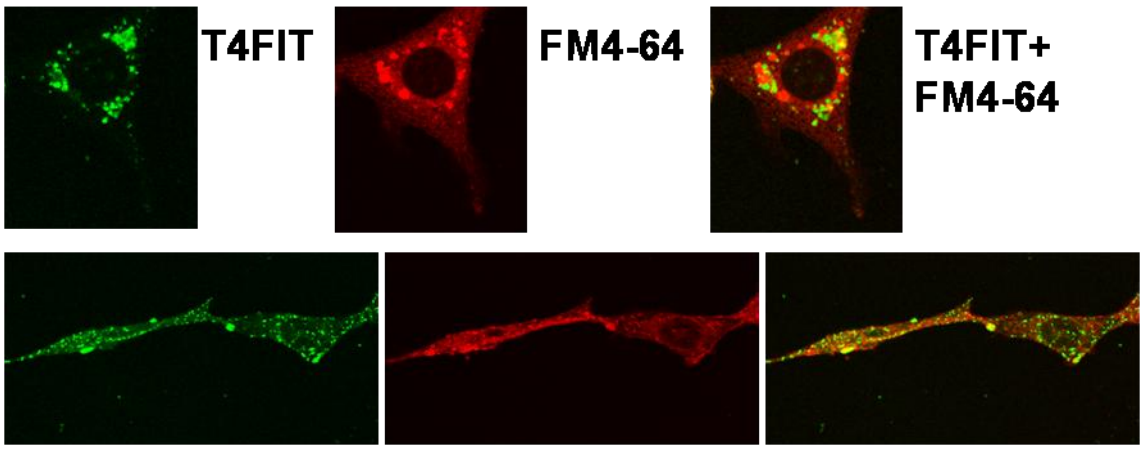


Fig. 5.8. Laser confocal microscope images of sections of single Raji cells after incubation with the nucleopeptide T4FIT 50 μM for 16 hours (detected on the green fluorescence channel, left) and treatment with the lipophilic stiryl dye FM4-64, (detected on the red fluorescence channel, middle) for staining of cell membranes and cytoplasm. Simultaneous observation of both channels (right) confirms the cytosolic localisation of the nucleopeptide in the cytoplasm. In the top images a section of the nucleus is visible as a dark circle in the middle of each image. The nucleus can be also found in the right side of the bottom images, although it is less evident.

A second series of cell penetration experiments has been performed, involving the fluorescent thymine-based nucleopeptide T4FIT and its biotinylated analogue T4Bt. Penetration of the latter has been detected by treatment of the cells with fluorescently labelled streptavidin after incubation. Renca cells have been incubated with the nucleopeptides at 0.5, 5 and 50 μM concentration for 12 hours, washed several times and treated with FM4-64 and/or with DAPI[148] prior to the observation at the confocal microscope or at the epifluorescence microscope. DAPI (4',6-diamidino-2-phenylindole) is a blue fluorescent dye able to strongly bind to double-stranded DNA that is commonly used for nucleus staining.

Biotinylated and fluorescent nucleopeptides have been detected both in the cytoplasm and in the cell nuclei (Fig. 5.9), generally in vesicular compartments; in some cases also diffuse cytosolic fluorescence has been detected. Cells are more strongly stained after incubation with more concentrated solutions of nucleopeptides, suggesting that cell penetration is dose-dependent.

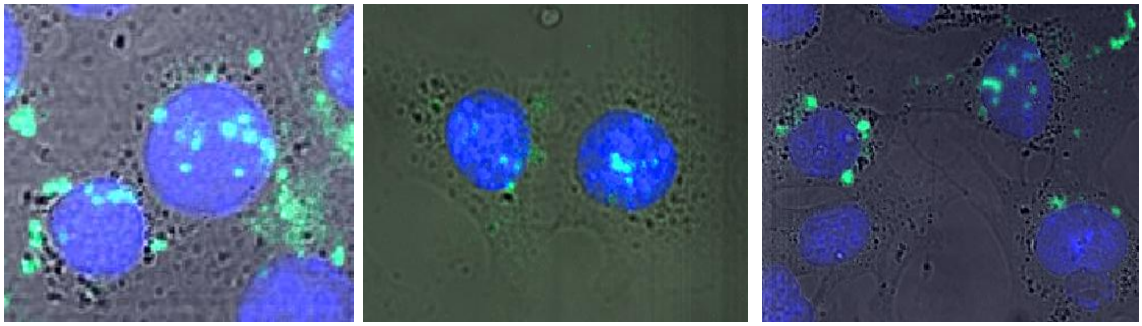


Fig. 5.9. Confocal microscope images of Renca cells incubated with 50 μM nucleopeptides (left and middle T4FIT, right T4Bt detected with fluorescently labelled streptavidin). Cell nuclei are stained with DAPI (blue), the cytoplasm is in clear field. In the middle image diffuse fluorescence in the cytosol can be observed.

5.3.1 Mechanistic data on cell uptake

The nucleopeptide capability to enter the cells, and particularly to reach the cell nucleus, is very interesting, since, as mentioned in the introduction, the oligonucleotides and many of the most successful nucleotide analogues do not display cell penetration capability, unless conjugated to suitable moieties. Nucleopeptide uptake has been therefore more deeply investigated, focusing on the fluorescent nucleopeptide T4FIT, whose detection does not require post-incubation processing.

Firstly, Raji cells have been incubated with T4FIT at 50 μM concentration and at increasing times cell samples have been drawn, suitably washed and analyzed by flow-cytometry. It is found that the fluorescence intensity increases steadily with time already for short incubation times (less than one hour). If the same experiment is performed under energy starving conditions in the presence of NaN_3 and 2-deoxy-D-glucose,[149] the fluorescence increase with time is much reduced even if it does not disappear (Fig. 5.10a). These results suggest that cell penetration is due by a significant extent to active transport mechanisms.

To further substantiate this hypothesis, Renca cells co-incubated for 12 hours with the nucleopeptide at 50 mM concentration and with fluorescently labelled transferrin (a

typical clathrin mediated endocytosis marker[150]) have been observed at the confocal microscope. As seen in Fig. 5.10b, significant but incomplete colocalisation is observed.

Considering all the information obtained by the different experiments performed, it can be stated that penetration of the nucleopeptides in the cells is due mainly, but likely not only, to endocytosis. Wider and thorough studies studies would be necessary, however, to fully understand this phenomenon.

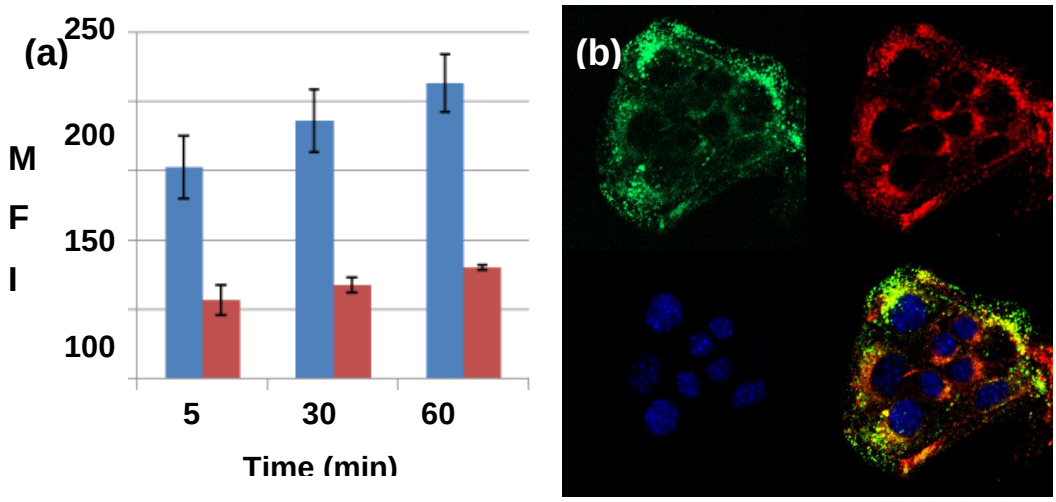


Fig. 5.10. (a) Mean fluorescence increase of Raji cells treated with the nucleopeptide T4FIT at 50 μM concentration at different time, in the absence (blue bars) and in the presence (red bars) of NaN_3 and 2-deoxy-D-glucose. (b) Confocal microscopy images of Renca cells co-incubated for 12 hours with the nucleopeptide T4FIT at 50 μM concentration (green) and with 20 mg/mL transferrin-Alexa-546 (red), after staining of the nuclei with DAPI (blue). Significant colocalisation (yellow) between peptide and fluorescently labelled transferrin is evident in the merge of the three channels.

It must be precised, however, that the study of the mechanism of cell uptake of peptides is a very complex subject. Among the main factors involved there are the chemical nature of the peptide considered, its concentration, the cell type and even the experimental conditions adopted.[151] Consequently, the exact determination of the nucleopeptide uptake mechanism is far beyond the scope of the present Ph.D. thesis project.

5.4 Stability tests

Preliminary nucleopeptide stability tests in serum have been performed on the three thymine-based nucleo-heptapeptides. Such nucleopeptides were chosen for two reasons:

(i) The availability of the significant amounts of nucleopeptide required by the stability tests;

(ii) The possibility of comparing the stability of the Ala-based nucleopeptide T2A with the stability of its analogues in which one (T2U) or all (6TTK) of the protein amino acid residues are substituted by C^{α}-tetrasubstituted Aib residues, known for imparting a higher resistance to proteolytic degradation.[96]

The test[152] requires to mix equal volumes of a peptide solution in PBS buffer and of murine serum, until a final peptide concentration of 5 mg/mL is obtained and to incubate the resulting solution at 37°C. Samples are drawn at increasing times, and after precipitation of the serum proteins by addition of TFA, the supernatant is analysed by HPLC. As a control and for normalisation, an aliquot of the solution is drawn and analysed with the same protocol immediately after mixing the peptides with the serum.

The results of the stability tests performed on the three nucleo-heptapeptides are displayed in Fig. 5.10, left.

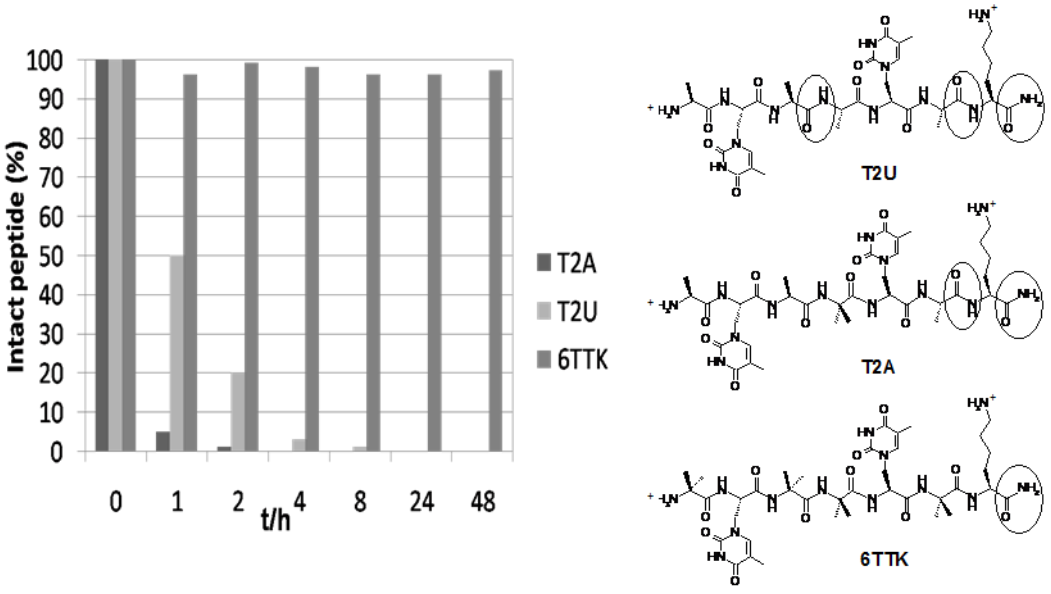


Fig. 5.10. Left: nucleo-heptapeptide stability in 50 % murine serum at 5 mg/mL concentration at different incubation times. Right: molecular structures of the three nucleopeptides tested. Circles highlight the amide bonds between protein amino acid residues and the C-terminal amide.

In the case of the Ala-based nucleopeptide T2A, extensive degradation is observed already after one hour incubation time, and after two hours the peptide peak is hardly detected. The Aib-containing nucleopeptide T2U is significantly more resistant: about one half of

original amount of peptide is observed after an hour incubation time. However, less then one fourth of the peptide is intact after two hours and after four hours the peptide is almost completely degraded. On the contrary, for the Aib-based nucleopeptide 6TTK no significant degradation is observed even after 48 hours. Figure 5.11 allows to compare the HPLC profiles for this peptide immediately after serum addition and after 48 hours incubation. The identity of the product has been confirmed also by HPLC-MS analysis.

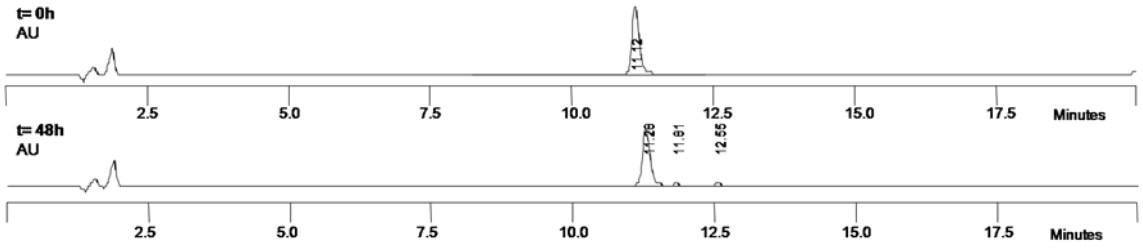


Fig. 5.11. HPLC profiles obtained from the stability test performed on the nucleopeptide 6TTK immediately after serum addition and after 48 hours incubation in 50 % murine serum. Varian HPLC, 0-50 % B in 20 min.

The reason of the marked increase in stability from the Ala-based to the Aib-based nucleopeptide is most likely that serum proteases are much more efficient in cleaving amide bonds between protein amino acid residues than amide bonds involving non proteogenic or C^{α}-tetrasubstituted residues. In the Ala-based nucleopeptide T2A there are three easily cleavable bonds, Ala^3-Ala^4, Ala^6-Lys^7 and the C-terminal Lys^7-NH_2, in the Aib containing T2U there are two, Ala^6-Lys^7 and the C-terminal Lys^7-NH_2, whereas in the Aib-based nucleopeptide 6TTK there is only C-terminal Lys^7-NH_2.

The remarkable stability of the Aib-based nucleopeptide is very interesting in view of biological applications, particularly given that cytotoxicity tests have demonstrated that this compound displays no signs of cytotoxicity up to 100 μM concentration.

Altogether, the biological tests performed on the deprotected nucleopeptides have provided evidence of interesting properties in view of biological applications. In particular:

(I) Both rigid and flexible nucleopeptides are not cytotoxic up to 100 μM concentration for incubation times up to 24 hours.

(II) Flexible fluorescent and biotinylated nucleopeptides penetrate into cell cytosol and nucleus, being found mainly in vesicular compartments. Diffuse cytosolic staining has also been observed.

(III) Cell uptake is concentration, time and energy dependent. Significant colocalisation with transferrin has been observed, suggesting that cell penetration is due mainly to endocytosis, but likely this is not the only mechanism involved.

(IV) Whereas the Ala-based nucleo-heptapeptide is rapidly degraded in serum, the introduction of a single Aib residue significantly improves resistance to proteolytic degradation, and the Aib-based analogue is stable even after 48 hour incubation.

6. EXPERIMENTAL SECTION

6.1 Materials and methods

6.1.1 Reagents and solvents

Acros-Janssen (Geel, Belgium):

Acetic anhydride, allyl bromide, benzyl chloroformate, biotine, bromocresol green, *para*-cresol, ethanolamine hydrochloride, ethylenediaminetetraacetic acid, 1-hydroxy-*1H*-benzotriazole, *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid, methyl chloroformate, phosphoric anhydride, piperidine, potassium bromide, trifluoroacetic acid,.

Alfa Aesar (Karlsruhe, Germany):

2-amino-6-chloro-purine, adenine, cytosine, thymine.

Applied Biosystems - Invitrogen (Carlsbad, USA):

4-methylbenzhydrylamine hydrochloride salt resin (loading 0.62 mmol/g), transferrin-Alexa-546, *N*- (3-triethylammoniumpropyl) -4- (6-(4-(diethylamino)-phenyl) -hexatrienyl)-pyridinium dibromide.

Bachem (Bubendorf, Switzerland):

N^{α}-*tert*butyloxycarbonyl-serine.

BD Pharmingen (Franklin Lakes, New Jersey, USA):

Fluorescein-5(6)-aminothiocarbonyl-streptavidin (streptavidin-FITC).

Biacore (Uppsala, Sweden):

N-hydroxysuccinimide, surfactant P-20.

Cambridge Isotope Laboratories (Cambridge, England):

Deuterochloroform, hexadeuterated dimethylsulphoxide, trideuterated acetonitrile.

Carlo Erba (Milan, Italy):

Acetic acid, ammonia (28% aqueous solution), 1-butanol, calcium chloride anhydrous, chloroform, dichloromethane, diethyl ether, ethanol, ethyl acetate, hydrochloric acid (37% aqueous solution), methanol, petroleum ether (boiling range 40-60°C), potassium hydrogenocarbonate, potassium hydrogenosulphate, 2-propanol, sodium carbonate dodecahydrate, sodium hydroxyde, sodium hypochlorite, sodium sulphate anhydrous, sulphuric acid (98% aqueous solution), toluene, triethylamine.

EMS-Chemie (Domat-Ems, Switzerland):

Paraformaldehyde.

*Eurogentec (*Seraing, Belgium):

Oligoribonucleotides (U_4, U_8), oligodeoxyribonucleotides (dA_{10}, 5'-biotinyl-phosphoramidite-dA_{10} or Bt-dA_{10}, dT_8, dT_{10}).

Fluka (Buchs, Switzerland):

Acetonitrile, α-amino*iso*butyric acid, chloroform (spectrophotometric grade), *N,N*-diisopropylethylamine, hydrazine hydrate, methanol, spectrophotometric grade, *N*-methylmorpholine, ninhydrin, palladium catalyst (10% on activated charcoal), serine, tetrahydrofurane.

GL Biochem Lt (Shangai, China):

7-aza-1-hydroxy-1*H*-benzotriazole, N^{α}-benzyloxycarbonyl-N^{ε}-*tert*butyloxycarbonyl-lysine, N^{α}-*tert*butyloxycarbonyl-N^{ε}-benzyloxycarbonyl-lysine, *O*-(7-aza-benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate.

Interchim (Montluçon, France):

3,3'-dihexyloxacarbocyanine iodide.

Iris Biotech (Marktredwitz, Germany):

N-benzyloxycarbonyloxysuccinimide, *N*-ethyl-*N*'-(3-dimethylamino)propyl-carbodiimide chlorohydrate.

J. T. Baker (Phillipsburg, New Jersey, USA):

N,N-dimethylformamide, pyridine.

Merck (Darmstadt, Germany):

N-bromosuccinimide, magnesium acetate, sodium azide, disodium hydrogenphosphate dihydrate, potassium carbonate, sodium azide, sodium dihydrogenphosphate monohydrate, sodium formate.

NeoMPS (Strasbourg, France):

N^{α}-*tert*butyloxycarbonyl-alanine, N^{α}-*tert*butyloxycarbonyl- N^{ε}-(2-chlorobenzyloxycarbonyl)- lysine, N^{α}-*tert*butyloxycarbonyl- N^{ε}-fluorenylmethyoxycarbonyl- lysine, *O*-(benzotriazol-1-yl)- *N,N,N',N'*-tetramethyluronium hexafluoro-phosphate.

Prolabo (Paris, France):

Sodium chloride.

Promega (Madison, Wisconsin, USA):

3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium, reagent solution.

Siad (Bergamo, Italy):

*Iso*butene.

Sigma-Aldrich (Milwaukee, USA):

2-deoxy-D-glucose, 4',6-diamidino-2-phenylindole 1,8-diazabicyclo[5.4.0]undec-7-ene, 4-(dimethylamino)-pyridine, formic acid, 2-hydroxy-cinnamic acid, ninhydrin, iodine, sodium hydride (60% dispersion in mineral oil), streptavidin, *N,N,N',N'*-tetramethyl-4,4'-diamino-diphenylmethane, trinitrobenzensulfonic acid, trimethylsilyl trifluoromethansulphonate, triphenylphosphine, Triton X-100.

6.1.2 Instruments and methods

Solid-phase synthesis

Solid-phase synthesis of the nucleopeptide sequences was accomplished using a PSP 4000 semiautomatic peptide synthesizer for Boc/Bzl strategy.

MBHA resin (loading 0.62 mmol/g) was used as solid support.

Completion of coupling or deprotection steps was checked by qualitative Kaiser test[153] or by TNBS test[154].

Melting point determination

Melting points were measured on a Leitz apparatus model Laborlux 12 and are uncorrected.

Thin layer chromatography

Silica gel 60 F_{254} (Merck) or Alugram® Sil G UV 254 (Macherey-Nagel) on aluminium foil was used to follow the reactions.

Silica gel 60 F_{254} (Merck) on glass was used for TLC characterization. Retention factors (R_f) have been measured using three different solvent mixtures as eluents.

Rf_1: $CHCl_3$/EtOH 9:1; **Rf_2**: BuOH/AcOH/H_2O 3:1:1; **Rf_3**: PhMe/EtOH 7:1.

Products were detected either by UV lamp irradiation or with exposition to I_2 vapours or by warming with a heat gun and spraying firstly with a 1.5% NaClO solution and then with a ninhydrin-TDM solution.[155]

N^{α}-protected serine-β-lactone was detected by spraying a 0.06% (w/v) solution of bromocresol green in EtOH.

Flash chromatography

For FC purifications silica gel 60 Merck (40-63 μm diameter, mesh 230-400) was used. For the various purifications, one of the following loading methods was used:

1) direct loading on top of the column of the crude mixture dissolved in a small amount of the eluent used for the purification;
2) direct loading on top of the column of the crude mixture dissolved in the minimum amount of a different solvent mixture;
3) loading on top of the column of the crude mixture adsorbed on silica gel (‘dry-loading’).

HPLC

Analytical chromatograms were recorded either on a Pharmacia LKB-LCC 2252 instrument, provided with a Kromasil Phenomenex C18 column (4.6×250 mm, 100 Å, flow 1.0 mL/min) and a UV Uvicord SD detector operating at 214 nm, or on a Agilent Technologies 1200 Series instrument, provided with a Kromasil Phenomenex C18 column (4.6×250 mm, 100 Å, flow 1.0 mL/min) or on a Varian ProStar 410 instrument, provided with a Nucleosil 100-5 C18 column (4.6×150 mm, 100 Å, flow 1.2 mL/min) and a Varian ProStar 330.71 PDA detector.

Semi-preparative scale chromatographic purifications were performed either on a Agilent Technologies 1200 Series instrument, provided with a Vydac C18 Protein and Peptides column (10×250 mm, flow 2.5 mL/min) or on a Beckmann System Gold HPLC 166 instrument, provided with a SDS C18 Macherey-Nagel column (10×250 mm, 100 Å, flow 6.0 mL/min) and a Beckmann System Gold UV detector operating at 230 nm.

Preparative scale HPLC purifications were performed on a Shimadzu SCL-6B instrument, provided with Shimadzu LC-8A pumps, a Delta Pack C18 (19×300 mm, 100 Å, flow 17 mL/min) column and a Shimadzu SPD-64 detector operating at 265 nm. As eluents following solvent mixtures were used:

- HPLC Varian and Beckmann: A: H_2O + 0.1% TFA; B: MeCN + 0.08% TFA.
- HPLC Pharmacia, Agilent and Shimadzu: A': H_2O/MeCN 9:1 + 0.05% TFA; B': MeCN/H_2O 9:1 + 0.05% TFA.

MilliQ H_2O was used for eluent preparation.

Polarimetric measurements

Optical rotations were measured on a a Perkin-Elmer model 241 polarimeter with an Haake model D8 thermostat at the sodium D line wavelength, using a cell with an optical pathlength of 10 cm. Concentrations are expressed in g/100 mL. $[\alpha]^{25}_{D}$ are calculated using the formula $[\alpha]= \alpha/(c \cdot l)$, where c is the concentration (in g/mL) and l is the optical path (in dm). Spectrophotometric grade MeOH was used as a solvent.

Circular dichroism

CD spectra were recorded either on a Jasco dichrograph model J-715 provided with a Haake model F3 thermostat or on Jasco dichrograph model J-810, provided with a JASCO ETC-505S/PTC-423S temperature controller using Hellma quartz cells (pathlenghts 0.1, 0.2, 0.5 or 1.0 cm). Instrumental parameters were as follow: response 2 sec, standard sensitivity, scan speed 100 nm/min, data pitch 0.5 nm.

Experimental values are expressed in molar ellipticity $[\Theta]_T$ (degrees·cm^2·dmol^{-1}) and were calculated, using the relation $[\Theta]_T = \Theta/(l \cdot c)$, where Θ is the observed ellipticity (in degrees), l is the optical path (in dm) and c is the peptide concentration (in mol/L).

UV spectroscopy

UV-Vis spectra were recorded on a Shimadzu UV 2501PC spectrophotometer or on a Cary 5000 spectrophotometer using Hellma 1.0 cm optical path quartz cell.

Mass spectrometry (MS)

ESI-TOF mass spectra were collected on a Mariner ESI-TOF mass spectrometer (Perseptive Biosystems). Prior to injection, samples were dissolved in a 1:1 mixture of water-MeOH containing 0.1 % formic acid. The positive ions were accelerated at 10, 15, 20 or 30 keV.

MALDI mass spectra of nucleopeptides were collected on a Bruker Protein TOF Biflex II MALDI-TOF spectrometer. 2-hydroxy-cinnamic acid (saturated solution in acetone) was used as matrix. Equal amounts (0.6 μL each) of matrix solution and of dilute aqueous nucleopeptide solution were mixed and the same volume of the resulting mixture was loaded on target disk.

HPLC-MS

HPLC-MS ESI-TOF analyses were performed on a Finnigan LCQ Advantage Thermo Scientific LC-MS, provided with a Macherey-Nagel EC 100/2 Nucleodur C18 column (2×100 mm, 100 Å, flow 300 μL/min).

As eluents, H_2O + 0.1 % formic acid (A'') and MeCN (B'') were used.

MilliQ H_2O was used for eluent preparation.

IR absorption spectroscopy

IR absorption measurements on KBr pellets were performed on a Perkin-Elmer model 580 B spectrophotometer, equipped with a Perkin-Elmer 3600 IR data station. IR absorption measurements in deuterated chloroform solution were performed on a Perkin-Elmer model 1720 X FT-IR spectrophotometer, equipped with a sample-shuttle device, at 2 cm^{-1} nominal resolution, averaging 100 scans, nitrogen flushed, connected with a IBM PS/2 model 50 Z personal computer. Solvent (baseline) spectra were obtained under the same conditions. Cells with pathlengths of 0.1, 1 and 10 mm (with CaF_2 windows) were used. Absorption maxima, shoulder and partially overlapping bands were detected with the inverse second derivative method. Spectral elaborations (solvent subtraction and derivatization) were performed using the program SpectraCalc by Galactic (Salem, MA, USA).

NMR spectrometry

^{1}H-NMR and 2D-NMR spectra were recorded on Bruker AC 200, Bruker AC 250, Bruker AC 300 or DRX 400 spectrometers. ^{13}C-NMR spectra were recorded on a Bruker AC 300 spectrometer. 2D-NMR spectra were recorded on DRX 400 Bruker or DMX 600 spectrometers. Chemical shifts (δ) are expressed in parts per million with regard to tetramethylsilane signal. Solvent residual peaks ($CHCl_3$, δ 7.26 ppm, or d_6-DMSO, δ 2.50 ppm, for ^{1}H-NMR, $H_3{}^{13}\boldsymbol{C}S(O)CH_3$, δ 39.52 ppm for ^{13}C-NMR) were used to calibrate the spectra. Peak multiplicity is described as follows: s (singlet), br s (broad singlet), d (doublet), dd (doublet of doublets), m (multiplet); coupling constants (*J*) are expressed in Hz.

For the elaboration of the spectra the programs XwinNMR e MestRe-C2.3 were applied.

X-ray diffraction (XRD)

Diffraction data were collected on a Philips PW1100 four-circle diffractometer, using graphite-monochromated CuKα radiation. Unit cell parameters were obtained by least-squares refinement of the angular settings of 48 carefully centred reflections in the 12 – 18° θ range. Intensities were corrected for Lorentz and polarisation effects, not for absorption.

The structure was solved by direct methods of the SIR 2002 program.[156] Refinement was carried out by full-matrix least-squares procedures on F^2, using all data, by application of the SHELXL-97 program,[157] with all non-H atoms anisotropic. The phenyl ring of the N-terminal Z-protecting group was constrained to the idealized geometry. H-atoms atoms were calculated at idealized positions and refined using a riding model.

Centrifuges

Hettich Zentrifugen Mikroliter Typ 2021 or Sigma 2MK centrifuges were used for Eppendorf tubes (stability tests) or 1.5 mL Starsted cryotubes (lyophilized purified products); an Eppendorf 5804 R centrifuge was used for Falcon tubes (crude peptide recovery after cleavage) .

Surface Plasmon Resonance

Surface Plasmon Resonance measurements were performed on a BIACORE 3000 system using sensor chips model CM5 (Biacore, Uppsala, Sweden).

Kinetic parameters were calculated using the BIAeval 4.1 software on a personal computer. Global analysis was performed using the simple Langmuir binding model. The specific binding profiles were obtained after subtracting the response signal from the channel control (streptavidin free). The fitting to each model was judged by the reduced chi square and randomness of residue distribution.

Microscopy imaging

Fluorescently labelled cells were observed on a Zeiss LSM 510 Meta confocal microscopy or on a Zeiss Axiovert 200M epifluorescence microscopy equipped with a Zeiss Apotome module.

Fluorescence Activated Cell Sorting

Cells were analyzed with a FACSCalibur® flow cytometer. At least 10.000 events were acquired for each experiment using the CellQuest 3.3 software developed by Becton Dickinson, Pont de Claix, France and data were processed with the WinMDI 2.8 freeware developed by Joseph Trotter, Scripps Research Institute (http://facs.scripps.edu/software.html).

Microplate biological testing

For colorimetric cell viability tests a Wallac Victor^2 microplate reader was used.

6.2 Synthesis and characterization

6.2.1 Solution phase synthesis

6.2.1.1 Synthesis of protected nucleobases and lactone precursors

H-A^{Boc} [109]

H-A^{H} (5.4 g, 40 mmol) is suspended in DMF/py 2:1 (total volume 150 mL) and Boc_2O (17.5 g, 80 mmol) is added. The mixture is heated at 70-80°C for 1 hour then stirred at room temperature overnight. The solvent is evaporated under vacuum, the remaining solid is dissolved in H_2O/MeOH 4:1 (total volume 65 mL) by heating under reflux for 15 min. The mixture is evaporated and the resulting oil is diluted in MeOH/DCM 6:1 (total volume 90 mL) then, after extensive stirring, it is filtered. The mother liquors are evaporated and the crude product is purified by flash-chromatography (eluent: DCM/MeOH 98:2).

Yield: 28 %. Mp> 250°C (dec.). Rf_1: 0.10; Rf_2: 0.80; Rf_3: 0.10.

IR (KBr): 3418, 1748, 1624, 1595, 1557 cm^{-1}.

^{1}H-NMR (d_6-DMSO, 200 MHz), δ/ppm: 12.27 (br s, 1H, N^6H-Boc), 10.55 (br s, 1H, N(9)-H), 8.57 (s, 1H, C(2)H), 8.43 (s, 1H, C(8)H), 1.51 (s, 9H, CH_3 Boc).

H-A^{Z} [20]

NaH (6.4 g, 60% in hexanes, 160 mmol) is suspended in 160 mL of anhydrous DMF and stirred under N_2 for 10 min, then H-A^{H} (5.4 g, 40 mmol) is added. Stirring continues for 2 hours until a yellow suspension is formed, then the mixture is cooled to 0°C in a water/ice bath and ZCl (13.5 mL, 100 mmol) are added dropwise in 20 min. After 4 hours the mixture is poured in 320 mL of chilled H_2O and stirred for 10 min, then it is neutralized with HCl 1N with precipitation of a solid. The precipitate is filtrated, washed with H_2O and Et_2O, then purified through double recrystallization from MeOH/$CHCl_3$.

Yield: 51 %. Mp: 222°C. Rf_1: 0.40; Rf_2: 0.80; Rf_3: 0.30.

IR (KBr):3360, 1780, 1618, 1575 cm^{-1}.

UV: ε= 15100 $M^{-1}cm^{-1}$ at λ= 275.4 nm (7.1 μM in MeOH).

^{1}H-NMR (d_6-DMSO, 200 MHz), δ/ppm: 12.31 (br s, 1H, N^6H-Z), 11.01 (br s, 1H, N(9)H), 8.63 (s, 1H, C(2)H), 8.48(s, 1H, C(8)H), 7.48 (s, 5H, C_6H_5-Z), 5.32 (s, 2H, CH_2-Z).

H-C^{Boc} [58b]

H-C^{H} (4.5 g, 40 mmol) and DMAP (0.25 g, 20 mmol) are suspended in 76 mL anhydrous DMSO under N_2 and the resulting yellow slurry is stirred for 10 min. Boc_2O (10.7 g, 48 mmol) is added under N_2 to the mixture, the flask is closed with a rubber septum pierced by a syringe needle and rapid dissolution with CO_2 development occurs. After stirring 1 day at room temperature, the yellow mixture is poured in 300 mL H_2O and vigorously stirred with decoloration and precipitation of the crude product after a while. The precipitate is filtered, washed with H_2O, AcOEt and finally EP, than it is dried overnight under vacuum. More product is recovered by concentration of the mother liquors.

Yield: 43 %. Mp: 170°C. Rf_1: 0.45; Rf_2: 0.85; Rf_3: 0.25.

IR (KBr): 3427, 1738, 1626, 1580, 1512 cm^{-1}.

UV : ε= 3600 $M^{-1}cm^{-1}$ at λ=265.0 nm (17 μM in MeOH).

NMR (d_6-DMSO, 200 MHz), δ/ppm: 7.79-7.76 (d, *J*= 7.8 Hz, 1H, C(6)H), 7.60 (br s, 1H, N^4H), 5.79-5.76 (d, *J*= 7.8 Hz, 1H, C(5)H), 1.49-1.44 (2 s, CH_3 Boc *cis-trans*).

H-C^{Bz} [101]

H-C^{H} (1.1 g, 10 mmol) is suspended in DMF, py is added (0.9 mL, 11 mmol) under stirring, then BzCl is added dropwise (2 mL, 20 mmol). A solution is suddenly formed, then the product precipitates. After stirring for 1 hour, the product is filtered and washed with plenty of H_2O, AcOEt and Et_2O, then dried under vacuum overnight.

Yield: 90 %. Mp> 250°C. Rf_1: 0.15; Rf_2: 0.80; Rf_3: 0.05.

IR (KBr): 3423, 3227, 1720, 1695, 1641, 1621, 1589, 1582, 1513 cm^{-1}.

UV: ε= 22400 $M^{-1}cm^{-1}$ at λ=250.4 nm, shoulder at λ= 294 nm (26 μM in MeOH).

^{1}H-NMR (d_6-DMSO, 200 MHz), δ/ppm: 8.01-7.96 (m, 2H, H-*orto* Bz), 7.93-7.89 (d, *J*= 7.0 Hz, 1H, C(6)H), 7.69-7.62 (m, 1H, H-*para* Bz), 7.58-7.50 (m, 2H, H-*meta* Bz), 7.26-7.22 (d, *J*= 7.0 Hz, 1H, C(5)H).

H-C^{Z} [20]

H-C^{H} (5.0 g, 45 mmol) is dissolved in 90 mL py (2 ml/mmol H-C^{H}) and DMAP (1.0 g, 8.0 mmol) are added. The solution is chilled in a water/ice bath, then Z-Cl (15 ml, 102 mmol) is added dropwise. The mixture is stirred three days at room temperature, then 180 mL chilled H_2O are added and the precipitate is filtered and washed with H_2O, then desiccated.

Yield: 40 %. Mp: 273°C. Rf_1: 0.30; Rf_2: 0.80; Rf_3: 0.30.
IR (KBr): 3475, 3416, 1744, 1688, 1627, 1587 cm^{-1}.
UV: ε= 18800 $M^{-1}cm^{-1}$ at λ=236.1 nm, ε= 7490 $M^{-1}cm^{-1}$ at λ= 289.3 nm (7.1 μM in MeOH).
^{1}H-NMR (d_6-DMSO, 200 MHz), δ/ppm: 8.31 (s, 1H, N^4H), 7.83 (d, 1H, C(6)H), 7.43 (m, 5H, C_6H_5-Z), 6.96 (d, 1H, C(5)H) , 5.21 (s, 2H, CH_2-Z).

All-C^Z

H-C^Z (0.10 g, 4.2 mmol) is suspended in 5 mL anhydrous DMF, DBU (65 μL, 4.2 mmol) is added and sudden dissolution occurs. AllBr (36 μL, 4.1 mmol) is added and the mixture is stirred at room temperature with exclusion of moisture. After 2 days the reaction is quenched by adding 1 drop H_2O, the solvent is evaporated at reduced pressure and the resulting oil is taken up in AcOEt, then neutralized by washing with $KHSO_4$ 5 %, with precipitation of some unreacted material, which is filtered off. The organic phase is desiccated on Na_2SO_4, then evaporated at reduced pressure and an oil results. The product is purified by flash-chromatography (eluent: DCM/MeOH, gradient from 40:1 to 25:1).
Yield: 31 %. Mp: 119-121°C. Rf_1: 0.85. Rf_2: 0.90. Rf_3: 0.45.
IR (KBr): 3515, 3427, 1756, 1646, 1606, 1587, 1500 cm^{-1}.
UV: ε= 16000 $M^{-1}cm^{-1}$ at λ= 240.3 nm, ε= 8630 $M^{-1}cm^{-1}$ at λ= 293.3 nm (11.0 μM in MeOH).
^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 7.58-7.51 (s and d, *J*= 7.1 Hz, 2H, C(6)H and N^4H), 7.38 (m, 5H, C_6H_5-Z), 7.23-7.20 (d, *J*= 7.1 Hz, C(5)H), 6.05-5.85 (m, 1H, C(2')H All), 5.36-5.35, 5.31-5.30 (2 d, *J*= 1.1 Hz, 2H, 2 C(3')H All), 5.22 (s, 2H, CH_2-Z), 4.52-4.49 (d, *J*= 5.8 Hz, 2H, C(1')H_2 All).

DMHD [158]

A solution of $NH_2NH_2 \cdot H_2O$ (25 mL, 380 mmol) in 100 mL EtOH is cooled in a water/ice bath and MeOC(O)Cl (29.5 mL, 380 mmol) is added dropwise. After that an aqueous solution of $Na_2CO_3 \cdot 10\ H_2O$ (109 g, 380 mmol) in 150 mL H_2O is added dropwise simultaneously to the dropwise addition of a second aliquot of MeOC(O)Cl (29.5 mL, 380 mmol). The mixture is stirred for 1 hour, then stirring is turned off and the product precipitates on standing overnight. The precipitate is filtered with cold H_2O and desiccated

under vacuum. More product is obtained by precipitation after evaporating the EtOH at reduced pressure from the solution.

Yield: 85 %. Mp: 133-135°C. Rf_1: 0.55; Rf_2: 0.90; Rf_3: 0.35.

IR (KBr): 3288, 1746, 1719 1539 cm^{-1}.

^{1}H-NMR (d_6-DMSO, 200 MHz), δ/ppm: 9.07, 8.72 (2 br s, 2H, 2 NH *transoid/ cisoid* ~5.5:1), 3.58 (s, 6H, 2 OMe).

DMAD[158, 159]

NBS (3.0 g, 16.9 mmol) is suspended in 50 mL DCM, py (1.5 mL, 18.1 mmol) is added and after stirring a solution is formed. DMHD (2,66 g, 17.9 mmol) is suspended in 50 mL DCM and the NBS solution is slowly added to the suspension. The resulting orange solution is stirred for 3 hours closed and protected from the light. The solution is then washed with $KHSO_4$ 0.3 % (pH=3) and several times with H_2O, desiccated on Na_2SO_4, and the solvent is removed at low temperature. The resulting crude product is a red oil, unstable to chromatographic treatment and which must be used quickly without further purification. Purity: 90% (extimated from TLC).

Yield: 90 % (taking into account crude purity).

IR (KBr): 2964, 1781, 1437 cm^{-1}.

6.2.1.2 Synthesis of amino acid derivatives

Boc-Aib-OH [160]

H-Aib-OH (15.3 g, 146 mmol) is dissolved in 40 mL 4 N NaOH and cooled to 0°C in a water/ice bath. Boc_2O (12.5 g, 57 mmol) is dissolved in 20 mL dioxane and slowly added to the solution, which is allowed to return to room temperature and stirred overnight. The day after, the same amount of Boc_2O is dissolved in dioxane and added to the mixture, NaOH (2.0 g, 50 mmol) dissolved in 50 mL H_2O is added and the mixture is heated at 45°C for 6 hours, then it is allowed to return to room temperature and stirred overnight. The day after, the same amount of Boc_2O is dissolved in dioxane and added to the mixture, which is heated at 45°C for 6 hours, then allowed to return to room temperature and stirred overnight. The day after the dioxane is evaporated and the remaining solution is diluted to 300 mL with H_2O, then washed with Et_2O and acidified to pH 3 with solid $KHSO_4$ until

latescence occurs. The product is extracted with AcOEt, organic extracts are washed with $KHSO_4$ 5 % and H_2O, desiccated on Na_2SO_4, and evaporated to dryness.

Yield: 63 %. Mp: 116-118°C. Rf_1: 0.65; Rf_2: 0.80; Rf_3: 0.40.

IR (KBr): 3313, 1739, 1661, 1546 cm^{-1}.

NMR ($CDCl_3$, 200 MHz), δ/ppm: 5.11 (br s, 1H, NH), 1.53 (s, 6H, 2 CH_3), 1.45 (s, 9H, CH_3 Boc).

Z-Aib-OH [73,161]

To a solution of H-Aib-OH (26.4 g, 253 mmol) and TEA (36 mL, 253 mmol) in 160 mL H_2O, Z-OSu (44.5 g, 177 mmol) dissolved in 50 mL MeCN is added under stirring. The mixture is stirred at room temperature, the pH is kept around 8.5 by addition of TEA. More Z-OSu dissolved in the minimum amount of MeCN is added after 1, 2 and 3 days (19.2 g, 76 mmol, 6.6 g, 26 mmol e 3.3 g, 14 mmol respectively), and each time the pH is adjusted by addition of TEA. After 5 days the MeCN is removed at reduced pressure, the pH is adjusted to 9 with NaOH (3 g, 75 mmol) dissolved in brine (200 mL). Unreacted Z-OSu is extracted with Et_2O and the aqueous phase is acidified to pH 2 with solid $KHSO_4$ until latescence occurs. The product is extracted with AcOEt and the organic extracts are washed with $KHSO_4$ 5 %, then several times with H_2O. The mixture is desiccated on Na_2SO_4 and concentrated under vacuum. The product precipitates from AcOEt/EP.

Yield: 78 %. Mp: 84-85 °C (from AcOEt/EP).

Rf_1: 0.65; Rf_2: 0.90. Rf_3: 0.40.

IR (KBr): 3317, 1719, 1674, 1550 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 11.38 (s, 1H, COOH), 7.34 (m, 5H, C_6H_5-Z), 5.47 (s, 1H, NH), 5.10 (m, 2H, Z CH_2), 1.57 (s, 6H, 2 CH_3).

HCl·H-Aib-OMe[162]

H-Aib-OH (20.7 g, 199 mmol) is suspended in 270 mL MeOH then cooled to -15°C in a MeOH/ice/NaCl bath. $SOCl_2$ (29.5 mL, 402 mmol) is added dropwise to the mixture under stirring, with gradual dissolution. After adding the reagent, the mixture is allowed to return to room temperature, then it is heated at gentle reflux for 5 hours. The solvent is evaporated, the residue is taken up with Et_2O and evaporated again several times. Finally the residue is dissolved in the minimum amount of MeOH and the product precipitates by Et_2O addition.

Yield: 93 %. Mp: 163-165°C. Rf_1: 0.30; Rf_2: 0.60; Rf_3: 0.20.

IR (KBr): 3386, 1738, 1586 cm^{-1}.

^{1}H-NMR (d_6-DMSO, 250 MHz), δ/ppm: 8.788 (br s, 3H, αNH_3^+), 3.74 (s, 3H, OMe), 1.48 (s, 6H, 2 CH_3).

Z-Aib-O^{t}Bu[161]

Z-Aib-OH (28.96 g, 122 mmol) is dissolved in 330 mL DCM in a pressure-proof reaction vessel, the mixture is cooled to –60 °C in an acetone/dry ice bath, *iso*butene (140 mL, 1.48 mol) is bubbled, then H_2SO_4 98 % is added (1.30 mL, 24 mmol) and the vessel is closed and allowed to return to room temperature. After 7 days, the vessel is cooled to -15°C in a NaCl/ice bath, it is opened and the solution is poured into 100 mL cold $NaHCO_3$ 5 %. Phases are separated, the organic phase is concentrated under vacuum, the aqueous phase is extracted with AcOEt. Organic phases are pooled, washed with $NaHCO_3$ 5 %, $KHSO_4$ 5 % and H_2O, then desiccated on Na_2SO_4 and evaporated to dryness.

Yield: 85 %. Mp: 63-64°C (from Et_2O/EP).

Rf_1: 0.95; Rf_2: 0.95. Rf_3: 0.40.

IR (KBr): 3370, 1712, 1519 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 7.34 (m, 5H, C_6H_5-Z), 5.45 (s, 1H, NH), 5.08 (s, 2H, Z CH_2), 1.51 (s, 6H, 2 CH_3), 1.43 (s, 9H, O^{t}Bu).

(Z-Aib)$_2$O[73,161]

To a solution of Z-Aib-OH (2.01 g, 8.5 mmol) in 5 mL of anhydrous MeCN EDC·HCl (0.83 g, 4.2 mmol) is added. After stirring for 6 hours at room temperature with the exclusion of moisture, the solvent is evaporated under vacuum. The residue is taken up with AcOEt, rapidly washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, desiccated on Na_2SO_4 and evaporated to dryness.

Yield: 77 %. Mp: 101-102°C (from AcOEt/EP).

Rf_1: 0.95; Rf_2: 0.95; Rf_3: 0.60.

IR (KBr): 3288, 1815, 1740, 1697, 1681, 1538 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 7.35 (m, 5H, C_6H_5-Z), 5.07 (s, 2H, Z CH_2), 4.95 (s, 1H, NH), 1.45 (s, 6H, 2 CH_3).

Boc-Lys(Z)-NH$_2$

To a solution of Boc-Lys(Z)-OH (6.5 g, 17 mmol) in 30 mL distilled DCM (30 ml) cooled to 0°C in a water/ice bath, HOBt (2.4 g, 17.4 mmol) and EDC·HCl (3.50 g, 17.9 mmol) are added, then NH_3 (generated by adding excess solid NaOH to 80 mL NH_3 30 % aqueous solution, 1.25 mol) is bubbled in the solution. After stirring for 1 hour, the solvent is evaporated, the residue is taken up with AcOEt and washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, desiccated on Na_2SO_4, and evaporated to dryness. The resulting product is a waxy solid.

Yield: 85 %. $[\alpha]^{25}_D$ = +0.92°, $[\alpha]^{25}_{436}$ = +1.5° (c= 0.75 in MeOH).

Rf_1: 0.75; Rf_2: 0.95; Rf_3: 0.60.

IR (KBr): 3377, 3335, 1693, 1657 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 250 MHz), δ/ppm: 7.35 (m, 5H, C_6H_5-Z), 6.20 (s, 1H, 1 $CONH_2$), 5.53 (s, 1H, 1 $CONH_2$), 5.22 (d, 1H, αNH), 5.09 (s, 2H, CH_2-Z), 4.93 (t, 1H, εNH), 4.1 (m, 1H, αCH), 3.20 (q, 2H, εCH_2), 1.89 (m, 1H, 1 βCH_2), 1.62 (m, 1H, 1 βCH_2), 1.57 (m, 2H, δCH_2), 1.43 (s, 9H, Boc CH_3), 1.37 (m, 2H, γCH_2).

Z-Lys(Boc)-NH_2

Z-Lys(Boc)-OH (6.2 g, 16.1 mmol) is dissolved in distilled DCM, then cooled in a water/ice bath. HOBt (2.40 g, 17.6 mmol) and EDC·HCl (3.6 g, 18.3 mmol) are added, then NH_3 (formed by addition of excess solid NaOH to 80 mL NH_3 30 % aqueous solution, 1.25 mol) is bubbled in the solution. After stirring for 1 hour, the solvent is evaporated, the residue is taken up with AcOEt, washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 %, H_2O, desiccated on Na_2SO_4 and evaporated to dryness.

Yield: 98 %. Mp: 141-143°C. $[\alpha]^{25}_D$ = -4.6° (c= 0.4 in MeOH).

Rf_1: 0.60; Rf_2: 0.90; Rf_3: 0.40.

IR (KBr): 3379, 3320, 3203, 1684, 1656, 1538 cm^{-1}.

NMR ($CDCl_3$, 250 MHz), δ/ppm: 7.35 (m, 5H, C_6H_5-Z), 6.28 (br s, 1H, CONH), 5.63-5.62 (br s, 2H, 1 CONH and αNH), 5.10 (s, 2H, CH_2-Z), 4.66 (t, 1H, εNH), 4.19-4.17 (m, 1H, αCH), 3.12-3.09 (m, 2H, εCH_2), 1.83-1.65 (m, 6H, βCH_2, γCH_2 and δCH2), 1.42 (s, 9H, CH_3 Boc).

Z-Ser-OH

H-Ser-OH (5.3 g, 50 mmol) is dissolved in H_2O after addition of TEA (7.0 mL, 50 mmol). Z-OSu (11.3 g, 45 mmol) is dissolved in the minimum amount of MeCN and slowly added

to the mixture. The pH is adjusted to about 8 with TEA and the reaction is stirred at room temperature overnight. The MeCN is evaporated, the pH is adjusted again with Na_2CO_3 10 % and the solution is washed with Et_2O. The mixture is acidified to pH 3 with solid $KHSO_4$, until latescence occurs. The product is extracted with AcOEt, washed with $KHSO_4$ and several times with H_2O, desiccated on Na_2SO_4 and the concentrated. The product precipitates from AcOEt/EP.

Yield: 62 %. Mp:116-118 °C $[\alpha]^{25}_D = 5.9$ ° (c= 0.5 in MeOH).

Rf_1: 0.35; Rf_2: 0.65; Rf_3: 0.25.

IR (KBr): 3400, 1768, 1693, 1520 cm^{-1}.

NMR (d_6-DMSO, 200 MHz), δ/ppm: 7.40 (m, 5H, C_6H_5-Z), 5.08 (s, 2H, CH_2-Z), 4.96-4.93 (d, *J*= 7.2 Hz, 1H αNH), 4.14-4.05 (m, 1H, αCH), 3.70-3.68 (m, 2H, βCH_2).

6.2.1.3 Synthesis of nucleoamino acids

Boc-Ser(Lactone) [44a]

PPh_3 (3.80 g, 14.5 mmol) is dissolved in 58 mL of distilled THF under N_2 and the solution is cooled to -78°C in an acetone/dry ice bath. DMAD (2.43 g, 15.0 mmol) diluted in 15 mL of anhydrous THF is added dropwise in 10 min and after 10 min a solution of Boc-Ser-OH (3.12 g, 15.2 mmol) in 26 mL of distilled THF is added dropwise in 20 min. The temperature is kept at -78° for additional 20 min, and then the mixture is allowed to return to room temperature in 2.5 hours. The solution is concentrated to a pale yellow oil and the crude product is immediately purified by flash-chromatography (eluent: EP/AcOEt, gradient from 3:1 to 3:2).

Yield: 55 %. Mp: 117-119°C. $[\alpha]^{25}_D = -26.6$° (c= 0.5 in MeOH).

Rf_1: 0.70; Rf_2: 0.80; Rf_3: 0.40.

IR (KBr): 3360, 1844, 1752, 1678, 1643, 1532 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 5.19 (br s, 1H, αNH), 5.13-5.05 (m, 1H, αCH), 4.49-4.40 (m, 2H, βCH_2), 1.46 (s, 9H 3 CH_3 Boc).

Z-Ser(Lactone) [44a]

PPh_3 (3.90 g, 14.8 mmol) is dissolved in 60 mL anhydrous THF and the solution is cooled to -78°C in an acetone/dry ice bath. A solution of DMAD (2.20 g, 15.2 mmol) in THF (4 ml) is added dropwise in 10 min. After 10 min a solution of Z-Ser-OH (3.80 g, 15.8 mmol)

in THF (15 ml) is added dropwise in 20 min. The temperature is kept at -78°C for additional 20 min, then the mixture is allowed to return to room temperature in 2,5 hours. The THF is evaporated at reduced temperature and the mixture is immediately purified by flash-chromatography (eluent: EP:AcOEt, gradient from 7:3 to 11:9).

Yield: 54 %. Mp: 129°C. $[\alpha]_D^{25}$ = -26.8° (c = 0.5, MeOH).

Rf_1: 0.90; Rf_2: 0.95; Rf_3: 0.50.

IR (KBr): 3366, 1845, 1830, 1685, 1530 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 250 MHz), δ/ppm: 7.36 (s, 5H, C_6H_5-Z), 5.45-5.42 (d, *J*= 5.9 Hz, 1H, αNH), 5.15 (s, 2H, CH_2-Z), 5.11 (s, 1H, α-CH), 4.58-4.38 (2m, 2H, β-CH_2).

Boc-AlaA^Z-OH [163]

H-A^Z (1.63 g, 6.1 mmol) is suspended in 61 mL anhydrous DMF, DBU (0.84 mL, 5.5 mmol) are added and the yellow solution is stirred for 2 hours. After cooling to -78°C in an acetone/dry ice bath, a solution of Boc-Ser(Lactone) (1.04 g, 5.5 mmol) in 11 mL of anhydrous DMF is added dropwise in 1 hour. The mixture is allowed to return to room temperature and stirred overnight. The solvent is evaporated and the resulting oil is taken up with 90 mL H_2O, with precipitation of some excess H-A^Z, which is filtered off. The solution is cooled in a water/ice bath and acidified carefully with HCl 2N until latescence is observed. The resulting suspension is extracted with AcOEt, NaCl is added to the acidic phase until latescence and the suspension is extracted four times with AcOEt. The combined organic extracts are washed with $KHSO_4$ 5 % saturated with NaCl and brine, desiccated on Na_2SO_4 and the solvent is evaporated to obtain a wet solid which is purified by flash-chromatography (eluent: $CHCl_3$/MeOH/AcOH).

Yield: 18 %. Mp> 250°C (dec.). $[\alpha]^{25}_D$ = -17.6° (c= 0.3 in MeOH).

Rf_1: 0.05; Rf_2: 0.50; Rf_3: 0.05. MS calcd $[M+H]^+$: 457.18, found: 457.12.

IR (KBr): 3391, 3330, 1758, 1752, 1716, 1678, 1514 cm^{-1}.

UV: ε= 17200 $M^{-1}cm^{-1}$ at λ= 268.8 nm (17 μM in MeOH).

HPLC (Pharmacia, 10 - 70% B' in 30 min) t_r: 16.1 min .

^{1}H-NMR (*d_6*-DMSO, 200 MHz), δ/ppm: 13.07 (br s, 1H, COOH), 10.65 (br s, 1H, N^{6}H adenine), 8.63 (s, 1H, C(2)H adenine), 8.31 (s, 1H, C(8)H adenine), 7.48-7.32 (m, 6H, C_6H_5-Z and αNH AlaA), 5.22 (s, 2H, CH_2-Z), 4.67-4.64 (m, 1H, αCH AlaA), 4.47-4.44 (m, 2H βCH_2 AlaA), 1.27-1.13 (2 s, 9H, 3CH_3 Boc *cis-trans*).

^{13}C-NMR (*d*$_6$-DMSO), δ/ppm: 172.01 (COOH), 155.97, 152.92, 152.24, 150.11 (C(2) adenine), 145.37 (C(8) adenine), 137.20 (C-*ipso* C_6H_5), 129.19, 128.76, 128.61, 124.02 (C(5) adenine), 79.29 (*C*$(CH_3)_3$ Boc), 67.00 (CH_2-Z), 53.62 (αCH AlaA), 44.50 (βCH_2 AlaA), 28.80-28.37 (CH_3 Boc *cis-trans*).

Z-AlaA^Z-OH

H-A^Z (2.88 g 10.7 mmol) is suspended in 160 mL anhydrous DMF. Z-Ser(Lactone) (1.17 g 5.3 mmol) is added, then K_2CO_3 (1.48 g, 10.7 mmol) is added and the resulting suspension is stirred at room temperature with the exclusion of moisture. After 3 days it is filtered and evaporated to an oil, from which some excess base precipitates, which is filtered off. The oil is taken up in MeOH/AcOEt and some more H-A^Z precipitates on standing overnight. The crude mixture is purified by repeated flash-chromatography (eluent: DCM/MeOH/AcOH).

Yield: 21 %. Mp> 250°C (dec.). $[\alpha]_D^{25}$ = -1.2°, $[\alpha]_{436}^{25}$= -1.3° (c= 0.2, MeOH).

Rf_1: 0.05; Rf_2: 0.55; Rf_3: 0.05.

IR (KBr): 3419, 3262, 1721, 1685, 1607, 1585, 1522 cm^{-1}.

^{1}H-NMR (*d*$_6$-DMSO, 200 MHz), δ/ppm: 12.93 (s br, 1H, N^{6}H adenine), 8.76 (s, 1H, C(2)H adenine), 8.61 (s, 1H, C(8)H adenine), 7.60-7.56 (d, *J*= 6.0 Hz, 1H, αNH AlaA), 7.32 (m, 10H, 2 C_6H_5-Z), 5.26, 4.91 (2 s, 4H, 2 CH_2-Z), 4.57-4.35 (m, 3H, αCH and βCH_2).

Boc-AlaC^Z-OH

H-C^Z (2 g, 8.4 mmol) is suspended in 100 mL DMF under N_2 and NaH (0.29 g, 7.15 mmol) is added. After stirring 2 hours at room temperature, the suspension is cooled to -78°C in an acetone/dry ice bath and a solution of Boc-Ser(Lactone) (1.5 g, 8 mmol) in DMF is added dropwise in 1 hour. The mixture is allowed to return to room temperature and stirred overnight with exclusion of moisture. The solvent is removed at reduced pressure and 120 mL H_2O are added. The solution is cooled in a water/ice bath and acidified with careful addition of HCl 2 N to pH 3, when latescence occurs. The mixture is extracted with AcOEt, saturated with NaCl and extracted several times with AcOEt. The organic extracts are pooled, washed with $KHSO_4$ 5 % and with brine, desiccated on Na_2SO_4 and evaporated to dryness. The crude product is purified by flash-chromatography (eluent: DCM/MeOH/AcOH 15:1:0.2).

Yield: 53 %. Mp:167-180°C. $[\alpha]_D^{25}$ = -85° (c = 0.5, MeOH).

Rf_1: 0.60; Rf_2: 0.70; Rf_3: 0.55.

IR (KBr): 3416, 1746, 1700, 1627, 1564 cm^{-1}.

UV: ε= 18100 $M^{-1}cm^{-1}$ at λ= 297.6 nm (18.5 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 10.81 (br,1H, OH) 7.83 (d, 1H, C(6)H cytosine), 7.38 (m, 5H, C_6H_5-Z), 7.20 (d, 1H, αNH), 6.97 (d, 1H, C(5)H cytosine), 5.17 (s, 2H, AlaC(Z) CH_2-Z), 4.30 (m, 2H, 1H AlaC(Z) αCH and 1 βCH_2), 3.67 (m, 1H, 1 βCH_2), 1.19 (s, 9H, Boc CH_3).

Boc-AlaG^{Cl}-OH

H-G^{Cl} (1.42 g, 8.3 mmol) is suspended in 91 mL anhydrous DMF, DBU (1.15 mL, 7.5 mmol) is added and a pale yellow solution is formed. After stirring for 2 hours, the solution is cooled to -78°C in an acetone/dry ice bath and a solution of Boc-Ser(Lactone) (1.44 g, 7.5 mmol) in DMF is added dropwise in 1 hour. The mixture is allowed to return to room temperature and stirred overnight with exclusion of moisture. The solvent is removed at reduced pressure and 100 mL H_2O are added until a clear solution is formed, which is cooled in a water/ice bath and carefully acidified with HCl 2 N to pH 2, when latescence occurs. The mixture is extracted with AcOEt, saturated with NaCl and extracted several times. The organic extracts are washed with $KHSO_4$ 5 % added with plenty of NaCl and with brine, desiccated on Na_2SO_4 and evaporated to dryness. Part of the product precipitates from $CHCl_3$/MeOH/AcOEt, the rest is purified by flash-chromatography (eluent: $CHCl_3$/MeOH 12:1 + 0.3% AcOH).

Yield: 37%. Mp> 250°C. $[\alpha]^{25}_D$ = -70° (c= 0.17 in MeOH).

Rf_1: 0.00; Rf_2: 0.70; Rf_3: 0.00.

IR (KBr): 3441, 3402, 3325, 3203, 1692, 1637, 1609, 1572, 1506 cm^{-1}.

HPLC (Pharmacia, 15-55 % B' in 30 min) t_r: 14.7 min.

UV: ε= 22900 $M^{-1}cm^{-1}$ at λ= 223.8, ε= 6800 $M^{-1}cm^{-1}$ at λ= 310.2 nm (15 μM in MeOH).

^{1}H-NMR (d_6-DMSO, 200 MHz), δ/ppm: 13.10 (s, 1H, COOH) , 7.93 (s, 1H, C(8)H guanine), 7.34-7.30 (d, *J*= 7.6 Hz, 1H, αNH), 6.97 (br s, 2H, N^2H_2 guanine), 4.48-4.35 (m, 2H, 1H αCH and 1H βCH_2), 4.25-4.15 (m, 1H, 1 βCH_2), 1.28-1.11 (2 s, 9H, CH_3 Boc *cis-trans*).

Z-AlaG^{Cl}-OH

H-G^{Cl} (1.2 g, 7 mmol) is suspended in 80 mL anhydrous DMF, DBU (1 mL, 6.3 mmol) is added and the resulting solution is stirred for 2 hours at room temperature with the exclusion of moisture, then it is cooled to -78°C in an acetone/dry ice bath. A solution of Z-Ser(Lactone) (1.4 g, 6.3 mmol) dissolved in 30 mL DMF is added dropwise in 1 hour. The mixture is allowed to return to room temperature and stirred overnight. The solvent is evaporated under vacuum and the residue is taken up with 200 mL H_2O, cooled to 0°C, then acidified with HCl 2 N until pH 3, when latescence occurs. The resulting suspension is extracted several times with AcOEt. Organic extracts are pooled, washed with $KHSO_4$ 5 % and with brine, desiccated on Na_2SO_4, then evaporated to dryness. The crude product is purified partially by recrystallization from MeOH (hot/cold), partially by flash-chromatography (eluent: DCM/MeOH/AcOH 15:1:0.3).

Yield: 52 %. Mp: 138-140°C. $[\alpha]^{25}_D$ = -32.6° (c= 0.23 in MeOH).

Rf_1: 0.00; Rf_2: 0.65; Rf_3: 0.00.

IR (KBr): 3472, 3320, 3202, 1717, 1687, 1630, 1572, 1518 cm^{-1}.

UV: ε= 7550 $M^{-1}cm^{-1}$ at λ= 243.4, ε= 4900 $M^{-1}cm^{-1}$ at λ= 305.3 nm (47 μM in MeOH).

^{1}H-NMR (*d_6*-DMSO, 200 MHz), δ/ppm: 7.96 (s, 1H, C(8)H guanine) , 7.82 (d, 1H, αNH), 7.35 (m, 5H, C_6H_5-Z), 6.98 (s, 2H, N^2H_2 guanine), 4.97 (s, 2H, Z-CH_2), 4.46 (m, 1H, 1 βCH), 4.27 (m, 2H, αCH and 1 βCH).

Boc-AlaG^{OH}-OH

Boc-AlaG^{Cl}-OH (0.82 g, 2.3 mmol) is dissolved in 8.0 mL TFA/H_2O 3:1 and the mixture is stirred at room temperature for 60 hours. The solution is evaporated, taken up with PhMe and evaporated again until a solid results, then taken up again with AcOEt and evaporated to dryness. It is then redissolved in 20 mL NaOH 2 % (pH 9-10), cooled to 0°C in a water/ice bath and Boc_2O (0.50 g, 2.3 mmol) dissolved in 12 mL MeCN is slowly added. The mixture is allowed to return to room temperature and stirred for 3 days. The MeCN is evaporated at reduced pressure, the mixture is diluted with 20 mL H_2O and washed with Et_2O, then cooled to 0°C in a water/ice bath and acidified carefully with HCl 1 N to pH 3, until latescence and precipitation of a solid is observed. 180 mL AcOEt are added and the mixture is stirred for 5 min, then it is allowed to stand, the part of the product which is not dissolved is filtered and washed with HCl 1 mM, acetone, $CHCl_3$, MeOH, and Et_2O then desiccated under vacuum overnight. The organic layer is concentrated at reduced pressure.

From the concentrated organic layer and from the acidic aqueous solution more product precipitates on standing in the cold overnight.

Yield: 63 %. Mp> 250 °C (dec.). $[\alpha]^{25}_{D}$= -73°

Rf_1: 0.00. Rf_2: 0.30. Rf_3: 0.00.

IR (KBr): 3422, 3220, 1706, 1634, 1595, 1539 cm^{-1}.

UV: ε= 12900 $M^{-1}cm^{-1}$ at λ= 255.5 nm (10 μM in MeOH).

HPLC (Varian, 1-31 % B in 20 min): t_r 15.81 min.

^{1}H-NMR (*d*$_6$-DMSO, 300 MHz), δ/ppm: 13.03 (br s, 1H, COOH), 10.56 (s, 1H, N(1)H guanine), 7.51 (s, 1H, C(8)H guanine), 7.31-7.28 (d, *J*= 8.4 Hz, αNH), 6.49 (s, 2H, N^2H_2 guanine), 4.36-4.28 (m, 2H, αCH and 1 βCH), 4.11-4.08 (m, 1H, 1 βCH), 1.32-1.19 (2 s, 9H, CH_3 Boc *cis-trans*).

^{13}C-NMR (*d*$_6$-DMSO, 76 MHz), δ/ppm: 171.90 (COOH), 157.21, 155.77, 154.06, 151.62, 138.23 (C(8) guanine), 116.91 (C(5) guanine), 78.94 (O*C*$(CH_3)_3$ Boc), 53.42 (αCH), 43.52 (βCH2), 28.53-28.07 (CH_3 Boc *cis-trans*).

Boc-AlaT-OH [109]

H-T (0.99 g, 7.6 mmol) is suspended in 68 mL anhydrous DMF, NaH (0.27 g 60 % in mineral oil, 6.7 mmol) is added, and the resulting suspension is stirred for 2 hours. The mixture is then cooled to -78°C in an acetone/dry ice bath, and a solution of Boc-Ser(Lactone) (1.30 g, 6.8 mmol) in 20 mL DMF is added dropwise in 1 hour. The mixture is allowed to return to room temperature and stirred overnight. After evaporation of the solvent, the resulting yellowish oil is dissolved in the minimum amount of H_2O, cooled in a water/ice bath and acidified carefully with HCl 2 N until latescence is observed. The emulsion is then extracted four times with AcOEt. The combined organic phases are then washed with $KHSO_4$ 5 % and brine, desiccated on Na_2SO_4 and concentrated to an oil. The product is purified by flash-chromatography (eluent: DCM/MeOH/AcOH).

Yield: 34%. Mp> 250°C (dec.). $[\alpha]^{25}_{D}$ = -55.8° (c= 0.3 in MeOH).

Rf_1: 0.0; Rf_2: 0.55; Rf_3: 0.0. MS calcd $[M+Na]^+$: 336.12, found: 335.98.

IR (KBr): 3426, 1681, 1515 cm^{-1}.

UV: ε= 8960 $M^{-1}cm^{-1}$ at λ= 271.4 nm (14 μM in MeOH).

HPLC (Varian, 0-50 % B in 20 min): t_r 12.1 min.

^{1}H-NMR (*d*$_6$-DMSO, 200 MHz), δ/ppm: 11.16 (br s, 1H, N(3)H thymine), 7.33 (s, 1H, C(6)H thymine), 6.76 (d, *J*= 8.0 Hz, 1H, αNH), 4.27-4.18 (m, 2H, αCH and 1 βCH), 3.48-

3.39 (dd, *J*= 14.2, 11.0 Hz, 1H, 1 βCH), 1.71 (s, 3H, C^5H_3 thymine), 1.38-1.30 (2 s, 9H, 3 CH_3 Boc, *cis-trans*).

^{13}C-NMR (*d_6*-DMSO): 171.79 (COOH), 164.82 (C(4)O thymine), 155.60 (CO Boc), 151.36 (C(2)O thymine), 142.56 (C(6)H thymine) 107.95 (C(5) thymine), 78.59 (*C*$(CH_3)_3$ Boc), 52.48 (αCH), 49.96 (βCH_2), 28.48-28.19 (CH_3 Boc, *cis-trans*), 12.48 (C^5H_3 thymine).

Z-AlaT-OH

Method A [45]: H-T (1.51 g, 11.6 mmol) is dissolved in 230 mL anhydrous DMF. Z-Ser(Lactone) (1.22 g, 5.5 mmol) is added, then K_2CO_3 (1.60 g, 11.6 mmol) is added and the resulting suspension is stirred with the exclusion of moisture for 2 days at room temperature, then it is filtered and evaporated to an oil, from which some excess H-T precipitates, which is filtered off. The oil is taken up in MeOH/AcOEt and some more H-T precipitates on standing overnight. The crude mixture is purified through repeated flash-chromatography (eluent:DCM/MeOH/AcOH).

Yield: 27 %. Mp> 250°C (dec.). $[\alpha]_D^{25}$ = -33° (c= 0.5, MeOH).

Rf_1: 0.0; Rf_2: 0.55; Rf_3: 0.0. MS calcd $[M+H]^+$: 348.12; found: 348.14.

IR (KBr): 3415,1684, 1610, 1521 cm^{-1}.

UV: ε= 8960 $M^{-1}cm^{-1}$ at λ= 271.4 nm (14 μM in MeOH).

^{1}H-NMR (*d_6*-DMSO, 200 MHz), δ/ppm: 11.12 (s br, 1H, N(3)H thymine), 7.42 (s, 1H, C(6)H thymine), 7.34-7.32 (m, 5H, C_6H_5-Z), 6.70-6.66 (d, *J*= 6.6 Hz, 1H, αNH), 5.05-4.88 (2 d, *J*= 12.0 Hz, 2H, CH_2-Z), 4.36-4.29 (d, *J*= 14 Hz, 1H, αCH), 4.08 (m, 1H, 1 βCH), 3.85 (m, 1H, 1 βCH), 1.79 (s, 3 H, C^5H_3 thymine).

Method B [109]: H-T (1.07 g, 8.2 mmol) is dissolved in 75 mL anhydrous DMF under N_2. NaH (0.30 g, 60 % dispersion in mineral oil, 7.6 mmol) is added and the heterogeneous reaction mixture is stirred for 2 hours, then cooled to -78°C in an acetone/dry ice bath. Z-Ser(Lactone) (1.66 g, 7.5 mmol) is dissolved in 20 mL anhydrous DMF and added dropwise in 1 hour, while the temperature is kept at -78°C, then the reaction is allowed to come back to room temperature and it is stirred overnight. The solvent is evaporated at reduced pressure to a yellow oil, it is taken up in 100 mL H_2O, then acidified with 2N HCl up to pH=3, when strong latescence occurs. The emulsion is extracted with AcOEt four times until the acidic phase is clear. Organic extracts are pooled and washed with $KHSO_4$ 5

% added with plenty of NaCl and then with brine. They are concentrated to a deliquescent yellowish solid, which is purified by repeated flash-chromatography (eluent: $CHCl_3$/MeOH/AcOH).

Yield: 35%. Mp> 250°C (dec.). $[\alpha]_D^{25} = -91°$ (c= 0.2, MeOH).

Rf_1: 0.0; Rf_2: 0.55; Rf_3: 0.0.

IR (KBr): 3416, 1681, 1626 cm^{-1}.

UV: ε= 8960 $M^{-1}cm^{-1}$ at λ= 271.4 nm (14 μM in MeOH).

^{1}H-NMR (*d_6*-DMSO, 200 MHz), δ/ppm: 11.30 (br s, 1H, N(3)H thymine), 7.57-7.53 (d, *J*= 8.2 Hz, αNH), 7.39 (s, 1H, C(6)H thymine), 7.35 (m, 5H, C_6H_5-Z), 5.04 (s, 2H, CH_2-Z), 4.37-4.22 (m, 2H, αCH and 1 βCH), 3.67-3.55 (dd, *J*= 13.2, 9.7 Hz, 1H, 1 βCH), 1.71 (s, 3H, C^5H_3 thymine) .

6.2.1.4 Synthesis of model peptides and peptide building blocks

Boc-Aib-Lys(Z)-NH_2

The active ester formed by dissolving in the minimum amount of distilled DCM Boc-Aib-OH (3.0 g, 14.5 mmol) and adding HOAt (2.0 g, 15 mmol), HCl·EDC (3.1 g, 15.4 mmol) and NMM (4.6 mL, 41 mmol) is added to a solution of TFA·H-Aib-Lys(Z)-NH_2 (obtained by acidolysis with TFA in DCM of 5 g, 13.0 mmol of the corresponding Boc-protected derivative) in distilled DCM. After stirring 1 day at room temperature with the exclusion of moisture, the solvent is evaporated under vacuum and the residue is taken up with AcOEt. The mixture is washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, desiccated on Na_2SO_4 and evaporated to dryness. The crude product is purified by flash-chromatography (eluent: DCM/MeOH, gradient from 30:1 to 18:1) to give a waxy solid.

Yield: 71%. $[\alpha]_D^{25} = -15°$ (c= 1.2, MeOH).

Rf_1: 0.70; Rf_2: 0.90; Rf_3: 0.55.

IR (KBr): 3404, 3334, 1694, 1672.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 7.35 (m, 5H, C_6H_5-Z), 7.08 (s, 1H, NH Aib), 6.66 (d, 1H, αNH Lys), 5.35 (br s,1H, εNH Lys), 5.10 (d, 2H, CH_2-Z), 4.98 (br s, 2H, $CONH_2$), 4.40 (m,1H, αCH Lys), 3.20 (m, 2H, εCH_2 Lys), 2.00 (m, 1H, 1 βCH_2 Lys), 1.69 (m, 3H, 1 βCH_2 and δCH_2 Lys), 1.49 (s, 4H, 1 CH_3 Aib and 1 γCH_2 Lys), 1.43 (s, 10H, 3 CH_3 Boc and 1 γCH_2 Lys), 1.40 (s, 3H, 1 CH_3 Aib).

Z-Aib-Lys(Boc)-NH_2

The active ester formed by dissolving Z-Aib-OH (1.89 g, 8.0 mmol) in 15 mL distilled DCM cooled to 0°C in a water/ice bath and adding HOBt (1.11 g, 8.1 mmol) and EDC·HCl (1.61 g, 8.2 mmol) is added to a H-Lys(Boc)-NH_2 solution (obtained by catalytic hydrogenolysis of 2.68 g, 7.1 mmol, of the corresponding Z-protected peptide). NMM (1.0 mL, 9.2 mmol) is added to adjust the pH to about 8.5 and the mixture is stirred at room temperature for 3 days. The solvent is evaporated, the residue is taken up in AcOEt, the mixture is washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, then desiccated on Na_2SO_4 and evaporated to dryness. The product is an oil.

Yield: 85 %. $[\alpha]_D^{25}$ = -12.5° (c= 1.0, MeOH).

Rf_1: 0.55; Rf_2: 0.85; Rf_3: 0.35.

IR (KBr): 3340, 1685, 1525 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 250 MHz), δ/ppm: 7.34 (m, 5H, C_6H_5-Z), 6.86 (br s, 1H, NH Aib), 6.64-6.61 (d, *J*= 7.7 Hz, 1H, αNH Lys), 5.44, 5.38 (2s, 2H, $CONH_2$), 5.13-5.03 (2 d, 2H, CH_2-Z), 4.74-4.70 (t, 1H, εNH Lys), 4.42-4.31 (m, 1H, αCH Lys), 3.11-3.03 (q, *J*= 6.2 Hz, 2H, εCH_2 Lys), 1.97-1.95 (m, 1H, 1 βCH Lys) 1.67-1.58 (m, 1H, 1 βCH Lys), 1.52, 1.49 (2 s, 7H, 2 CH_3 Aib and 1 δCH Lys), 1.42 (s, 10H, 3 CH_3 Boc and 1 δCH Lys), 1.34-1.31 (m, 2H, γCH_2 Lys).

Z-Ala-Aib-OMe

Z-Ala-OH (0.39 g, 1.74 mmol) is dissolved in the minimum amount of distilled DCM, it is cooled to 0°C in a water/ice bath and HOBt (0.24 g, 1.8 mmol) and EDC·HCl (0.36 g, 1.8 mmol) are added. After 10 min HCl·H-Aib-OMe (0.57 g, 3.7 mmol) is added to the mixture and the pH is adjusted to 8 with NMM (0.73 mL, 6.5 mmol). After 6 hours the solvent is evaporated, the residue is taken up with AcOEt, the mixture is washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, then desiccated on Na_2SO_4 and evaporated to dryness. The product is an oil.

Yield: 74 %. $[\alpha]_D^{25}$ = -30° (c= 0.45, MeOH).

Rf_1: 0.85; Rf_2: 0.90; Rf_3: 0.60.

IR (KBr): 3321, 1741, 1726, 1666, 1587, 1532 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 7.35 (m, 5H, C_6H_5-Z), 6.87 (br s, 1H, NH Aib), 5.41-5.38 (d, *J*= 6.6 Hz, 1H, NH Ala), 5.12 (s, 2H, CH_2-Z), 4.30-4.19 (m, 1H, αCH Ala), 3.72 (s, 3H, OMe), 1.52 (s, 6H, 2 CH_3 Aib), 1.38-1.35 (d, *J*= 7.0 Hz, CH_3 Ala).

Z-Ala-Aib-O^{t}Bu

Z-Ala-OH (0.25 g, 1.12 mmol) is dissolved in the minimum amount of distilled DCM, it is cooled to 0°C in a water/ice bath and HOBt (0.16 g, 1.2 mmol) and EDC·HCl (0.23 g, 1.3 mmol) are added. After 10 min H-Aib-O^{t}Bu (obtained by catalytic hydrogenolysis of 0.91 g, 3.1 mmol, of the corresponding Z-protected derivative) is added to the mixture and the pH is adjusted to 8 with NMM (0.25 mL, 2.2 mmol). After stirring overnight at room temperature with the exclusion of moisture, the solvent is evaporated, the residue is taken up with AcOEt, the mixture is washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, then desiccated on Na_2SO_4 and evaporated to dryness. The product precipitates from AcOEt/EP.

Yield: 74 %. Mp: 97-98°C. $[\alpha]_D^{25}$ = -30° (c= 0.45, MeOH).

Rf_1: 0.90; Rf_2: 0.90; Rf_3: 0.55.

IR (KBr): 3343, 3306, 1738, 1715, 1691, 1678, 1653, 1551, 1536 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 7.35 (m, 5H, C_6H_5-Z), 6.59 (br s, 1H, NH Aib), 5.34-5.31 (d, *J*= 5.5 Hz, 1H, NH Ala), 5.12 (s, 2H, CH_2-Z), 4.22-4.15 (m, 1H, αCH Ala), , 1.52 (s, 6H, 2 CH_3 Aib), 1,45 (s, 9H, CH_3 O^{t}Bu), 1.39-1.35 (d, *J*= 7.0 Hz, CH_3 Ala).

Z-Ala-Aib-OH

Method A: Z-Ala-Aib-OMe (0.057 g, 0.18 mmol) is dissolved in 3.6 mL MeOH, cooled to 0°C in a water/ice bath and the same volume of 0.1 M solution of LiOH in H_2O (prepared by dissolving LiOH·H2O, 0.21 g, 0.50 mmol, in 5.0 mL H_2O) is slowly added under stirring. The reaction monitored by TLC (DCM/MeOH 20:1) and after 2 hours the reaction is complete. The solution is neutralized by addition of $KHSO_4$ 5 %, then the MeOH is evaporated at reduced pressure, the solution is acidified to pH 3 by addition of solid $KHSO_4$ with turbidity and the acidic phase is extracted with AcOEt. The organic extracts are pooled and washed with $KHSO_4$ 5 % and H_2O, then desiccated on Na_2SO_4 and evaporated to dryness. The product precipitates from DCM/Et_2O.

Yield: 92 %. Mp: 126°C. $[\alpha]_D^{25}$ = -18.8° (c= 0.80, MeOH).

Rf_1: 0.30; Rf_2: 0.90; Rf_3: 0.20.

IR (KBr): 3352, 3317, 1719, 1695, 1663, 1622, 1528 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 7.34 (m, 5H, C_6H_5-Z), 6.95 (br s, 1H, NH Aib), 6.74 (br s, 1H, OH), 5.69-5.65 (d, *J*= 7.4 Hz, 1H, NH Ala), 5.11 (s, 2H, CH_2-Z), 4.35-4.23 (m, 1H, αCH Ala), 1.54, 1.51 (2s, 6H, 2 CH_3 Aib), 1.38-1.34 (d, *J*= 7.0 Hz, CH_3 Ala).

Method B: Z-Ala-Aib-O^{t}Bu (0.102 g, 0.28 mmol) is dissolved in the minimum amount of DCM, the solution is cooled, and TFA (0.42 mL, 5.6 mmol) is added, then the mixture is allowed to return to room temperature and stirred for 2 hours. N_2 is bubbled in the mixture in order to strip the excess TFA. The residue is taken up in DCM twice and again N_2 is bubbled in the mixture. The residue is taken up in DCM and evaporated at reduced pressure, then taken up with Et_2O and evaporated several times, until a colourless solid is obtained.

Yield: 98 %. $[\alpha]_D^{25} =$ -18.6° (c= 0.60, MeOH).

All other characterization data are the same as those found for Method A.

Z-Aib-Ala-Aib-OMe [164]

Z-Aib-OH (0.33 g, 1.4 mmol) is dissolved in the minimum amount of distilled DCM, they are cooled to 0°C, then HOBt (0.19 g, 1.4 mmol) and EDC·HCl (0.31 g, 1.6 mmol) are added. After 10 min HCl·H-Ala-Aib-OMe (1.1 mmol, obtained by idrogenolysis of 0.39 g Z-Ala-Aib-OMe in the presence of 1 equivalent of HCl to avoid diketopiperazine formation) is dissolved in 1 mL anhydrous MeCN and transferred into the reaction vessel. NMM (0.40 mL, 3.6 mmol) is added in order to adjust the pH to 8-9, the heterogeneous reaction mixture starts to dissolve and it is allowed to return to room temperature. After 6 days the solvent is evaporated, the mixture is taken up in AcOEt and washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, then desiccated on Na_2SO_4 and evaporated to dryness. The crude product is purified by flash-chromatography (eluent: DCM/MeOH 30:1).

Yield: 44 %. Mp: 47-49°C. $[\alpha]_D^{25} =$ -16° (c= 0.7, MeOH).

Rf_1: 0.65. Rf_2: 0.80. Rf_3: 0.40.

IR (KBr): 3517, 3468, 3321, 3257, 1719, 1694, 1653, 1540, 1503 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 7.34 (m, 5H, C_6H_5-Z), 7.07 (br s, 1H, αNH Aib^3), 6.53-6.49 (d, *J*= 8.0 Hz, 1H, αNH Ala), 5.23 (br s, 1H, αNH Aib^1), 5.08 (s, 2H, CH_2-Z), 4.49-4.38 (q, *J*= 7.2, αCH Ala), 3.70 (s, 3H, OMe), 1.52 (2 s, 6H, 2 CH_3 Aib), 1.49 (2 s, 6H, 2 CH_3 Aib), 1.34-1.31 (d, *J*= 7.2 Hz, 3H, CH_3 Ala).

6.2.1.5 Synthesis of nucleoamino acid derivatives and of nucleopeptides

Boc-AlaA^Z-Aib-OMe

Boc-AlaA^{Z}-OH (0.91 g, 2.0 mmol) is dissolved in anhydrous DMF under N_2, cooled to 0° in a water/ice bath and HOAt (0.32 g, 2.3 mmol) and EDC·HCl (0.45 g, 2.3 mmol) are added. After 10 min, HCl·H-Aib-OMe (0.61 gm 4.0 mmol) is added and NMM (0.9 mL, 8.1 mmol) is added until a homogeneous mixture is formed, at pH 8, then it is allowed to return to room temperature. After 7 days the solvent is evaporated at reduced pressure, the remaining yellow oil is taken up with AcOEt, washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, then desiccated on Na_2SO_4 and evaporated to pale yellow oil, which is diluted with a little AcOEt. Pure colourless product precipitates on standing, more product can be obtained from the mother liquors by flash-chromatography (eluent: DCM/MeOH 35:1).

Yield: 67 %. Mp: 184-187°C (from AcOEt).

$[\alpha]_D^{25}$ = -4.1°, $[\alpha]_{436}^{25}$ = -5.0° (c= 0.7, MeOH). Rf_1: 0.75. Rf_2: 0.90. Rf_3: 0.45.

MS calcd $[M+H]^+$: 556.25; found: 556.24.

IR (KBr): 3416, 3266, 3231, 1747, 1716, 1689, 1672, 1610, 1577, 1538 cm^{-1}.

UV: ε= 25000 $M^{-1}cm^{-1}$ at λ= 268.4 nm, shoulder at λ= 278 nm (12.6 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 250 MHz), δ/ppm: 8.75 (s, 1H, C(2)H adenine), 8.36 (s, 1H, N^{6}H adenine), 7.97 (s, 1H, C(8)H adenine), 7.43-7.39 (m, 5H, C_6H_5-Z), 7.12 (br s, 1H, NH Aib), 6.10 (br s, 1H, αNH AlaA), 5.30 (s, 2H, CH_2-Z), 4.61-4.59 (m, 3H, αCH and βCH_2 AlaA), 3.68 (s, 3H, OMe), 1.45 (s, 6H, 2 CH_3 Aib), 1.40 (s, 9H, CH_3 Boc).

Z-AlaA^{Z}-Aib-O^{t}Bu

Z-AlaA^{Z}-OH (0.21 g, 0.43 mmol) is dissolved in the minimum amount of anhydrous DMF, cooled to 0°C in a water/ice bath and HOAt (0.07 g, 0.50 mmol) and EDC·HCl (0.11 g, 0.55 mmol) are added. After 10 min H-Aib-O^{t}Bu (obtained by catalytic hydrogenolysis of 0.32 g, 1.1 mmol of the corresponding Z-protected derivative in DCM) is added and the mixture is allowed to return to room temperature. After 1 day the pH is adjusted at 8 by addition of NMM (90 μL, 0.8 mmol). After 5 days the solvent is evaporated at reduced pressure, the residue is taken up with AcOEt, washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, then desiccated on Na_2SO_4 and evaporated to an oil. The crude product is purified by flash-chromatography (eluent: DCM/MeOH 30:1).

Yield: 45 %. Mp: 191-193°C. $[\alpha]_D^{25}$ = -12.9° (c= 0.5, MeOH).

Rf_1: 0.70. Rf_2: 0.85. Rf_3: 0.40.

IR (KBr): 3425, 3267, 1731, 1709, 1693, 1648, 1601, 1577, 1538 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 8.74 (s, 1H, C(2)H adenine), 8.24 (s, 1H, N^6H adenine), 7.95 (s, 1H, C(8)H adenine), 7.43-7.33 (m, 10H, 2 C_6H_5-Z), 7.16 (br s, 1H, NH Aib), 6.56-6.53 (d, *J*= 6.0 Hz, 1H, αNH AlaA), 5.30 (s, 2H, CH_2-Z AlaA), 5.10 (s, 2H, CH_2-Z N^6H adenine), 4.69-4.60 (m, 3H, αCH and βCH_2 AlaA), 1.40 (s, 9H, CH_3 Boc), 1.36 (s, 6H, 2 CH_3 Aib).

Boc-Aib-AlaA^Z-Aib-OMe

Boc-Aib-OH (0.38 g , 1.9 mmol) is dissolved in the minimum amont of distilled DCM, cooled to 0°C in a water/ice bath and HOAt (0.28 g, 2.1 mmol) and EDC·HCl (0.43 g, 2.2 mmol) are added. The active ester formed is added to TFA·H-AlaA^Z-Aib-OMe (obtained by acidolysis with TFA in DCM of 0.68 g, 1.16 mmol, of the corresponding Boc-protected peptide) and NMM (0.33 mL, 2.9 mmol) is added to adjust the pH to 9. After 9 days the solvent is evaporated, the residue is taken up with AcOEt, washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, then desiccated on Na_2SO_4 and evaporated to an oil, from which the product is obtained by flash-chromatography (eluent: DCM/MeOH, gradient from 45:1 to 22:1).

Yield: 63 %. Mp: 109-111°C. $[\alpha]_D^{25}$ = -69° (c= 0.4, MeOH).

Rf_1: 0.75. Rf_2: 0.90. Rf_3: 0.35.

IR (KBr): 3445, 3327, 1742, 1691, 1689, 1612, 1589, 1527 cm^{-1}.

UV: ε= 15300 $M^{-1}cm^{-1}$ at λ= 268.2 nm, shoulder at λ= 279 nm (10.1 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 8.81-8.78 (s and d, *J*= 6.4 Hz, 2H, C(2)H adenine and αNH AlaA), 8.16 (br s, 1H, N^6H adenine), 7.95 (s, 1H, C(8)H adenine), 7.54 (br s, 1H, NH Aib[3]), 7.43-7.36 (m, 5H, C_6H_5-Z), 5.31 (s, 2H, CH_2-Z), 4.95-4.90 (m, 3H, αCH and 1 βCH AlaA, NH Aib[1]), 4.46-4.40 (d, *J*= 12.6 Hz, 1H, 1 βCH AlaA), 3.54 (s, 3H, OMe), 1.51, 1.47, 1.44, 1.24 (4 s, 12H, 4 CH_3 2 Aib), 1.22 (s, 9H, CH_3 Boc).

Boc-Aib-AlaA^H-Aib-OMe

Boc-Aib-AlaA^Z-Aib-OMe (0.056 g, 0.087 mmol) is dissolved in abundant MeOH and N_2 is fluxed in the mixture, Pd catalyst (0.030 g, 10 % on activated charcoal) is added under N_2, then after 10 min H_2 is bubbled. After 2,5 hours H_2 bubbling is stopped, N_2 is bubbled for 10 min and the catalyst is filtered on a celite pad. The mother liquors are evaporated at reduced temperature, then taken up with AcOEt and evaporated again until a solid is obtained, which is desiccated under vacuum.

Yield: 95 %. Mp: 203-205°C. $[\alpha]_D^{25}$ = -72° (c= 0.33, MeOH).

Rf_1: 0.30. Rf_2: 0.80. Rf_3: 0.10.

IR (KBr): 3402, 1739, 1694, 1645, 1603, 1578, 1533 cm^{-1}.

UV: ε= 13300 $M^{-1}cm^{-1}$ at λ= 259.3 nm (26 μM in MeOH).

^{1}H-NMR ($CDCl_3$/DMSO 3:1, 250 MHz), δ/ppm: 9.38-9.35 (d, *J*= 6.4 Hz, 1H, αNH AlaA), 8.32 (s, 1H, C(2)H adenine), 7.65 (s, 1H, C(8)H adenine), 7.63 (br s, 1H, NH Aib^3), 6.63 (br s, 1H, NH Aib^1), 6.49 (br s, 2H, N^6H_2 adenine), 4.80-4.72 (dd, *J*= 14.0, 5.0 Hz, 1H, 1 βCH AlaA), 4.63 (m, 1H, αCH AlaA), 4.31-4.25 (d, *J*= 14.0, 1H, 1 βCH AlaA), 3.41 (s, 3H, OMe), 1.36 (s, 6H, 2 CH_3 Aib), 1.30 (s, 3H, 1 CH_3 Aib), 1.23 (s, 12H, 1 CH3 Aib and CH_3 Boc).

Z-Aib-AlaA^H-Aib-O^{t}Bu

$(Z\text{-Aib})_2O$ (0.215 g, 0.47 mmol) is dissolved in the minimum amount of distilled DCM and added to H-AlaA^H-Aib-O^{t}Bu (obtained by catalytic hydrogenolysis of 0.118 g, 0.19 mmol of the corresponding N^α,N^6-*bis*-Z-protected peptide) dissolved in anhydrous DMF and cooled to 0°C in a water/ice bath, then NMM (60 μL, 0.53 mmol) is added and the reaction is allowed to return to room temperature and stirred with exclusion of moisture. After 6 days the solvent is evaporated at reduced pressure, the residue is taken up with AcOEt, washed with $KHSO_4$ 5 % and H_2O. The organic phase is stirred for 30 min in $NaHCO_3$ 5 % in order to hydrolyze the excess anhydride, then phases are separated, the organic phase is washed again with $NaHCO_3$ 5 % and with H_2O, desiccated on Na_2SO_4 and concentrated under vacuum. The product is purified by flash-chromatography (eluent: $CHCl_3$/MeOH).

Yield: 31 %. Mp: 183-185°C. $[\alpha]_D^{25}$ = -19° (c= 0.3, MeOH).

Rf_1: 0.30. Rf_2: 0.50. Rf_3: 0.20.

IR (KBr): 3390, 3270, 1733, 1709, 1650, 1602, 1576, 1539 cm^{-1}.

UV: ε= 10600 $M^{-1}cm^{-1}$ at λ= 259.6 nm (10.1 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 9.10-9.06 (d, *J*= 6.4 Hz, 1H, αNH AlaA), 8.32 (s, 1H, C(2)H adenine), 7.83 (s, 1H, C(8)H adenine), 7.37 (br s, 1H, NH Aib^3), 7.30 (m, 5H, C_6H_5-Z), 5.67 (s, 2H, N^6H_2 adenine), 5.27 (br s, 1H, NH Aib^1) 5.04-4.82 (2 d and m, 4H, CH_2-Z, αCH and 1 βCH AlaA), 4.39-4.33 (d, *J*= 12.0 Hz, 1H, 1 βCH AlaA), 1.54 (s, 3H, 1 CH_3 1 Aib), 1.50 (s, 3H, 1 CH_3 1 Aib), 1.38 (s, 9H, CH_3 O^{t}Bu), 1.30 (s, 3H, 1 CH_3 Aib), 1.21 (s, 3H, 1 CH_3 1 Aib).

Boc-AlaC^{Z}-OMe

After dissolving Boc-AlaC^{Z}-OH (80 mg, 0.17 mmol) in the minimum amount of anhydrous DMF, DMAP (15 mg, 0.12 mmol) is added. The mixture is cooled to 0°C and EDC·HCl (46 mg, 0.24 mmol) then MeOH (0.22 ml, 5.1 mmol) are added. After stirring at room temperature for 3 hours, the solvent is evaporated. The residue is taken up in AcOEt and washed with $KHSO_4$ 5 %, desiccated on Na_2SO_4, and evaporated to dryness.

Yield: 65 %. Mp: 95-98°C. $[\alpha]_D^{25} = -74°$ (c= 0.6, MeOH).

Rf_1: 0.80; Rf_2: 0.90; Rf_3: 0.40.

IR (KBr): 3385, 3274, 1748, 1700, 1665, 1627, 1556 cm^{-1}.

UV: ε= 13400 $M^{-1}cm^{-1}$ at λ= 242.4 nm, ε= 6520 $M^{-1}cm^{-1}$ at λ= 291.9 nm (44 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 250 MHz), δ/ppm: 7.80 (br, 1H, NH-Z), 7.53 (d, 1H, C(6)H cytosine), 7.40 (m, 5H, C_6H_5-Z), 7.20 (d, 1H, C(5)H cytosine), 5.76 (d, 1H αNH), 5.23 (s, 2H, CH_2-Z), 4.57 (m, 1H, 1 βCH), 4.32 (m, 2H, αCH and 1 βCH), 3.79 (s, 3H, OMe), 1.44 (s, 9H, CH_3 Boc).

Boc-AlaC^{Z}-Aib-OMe

The active ester formed by dissolving in the minimum amount of anhydrous DMF Boc-AlaC^{Z}-OH (1.4 g, 3.0 mmol) and adding HOBt (0.45 g, 3.35 mmol) and EDC·HCl (0.70 g, 3.6 mmol) is added to a solution of HCl·H-Aib-OMe in DMF (0.7 g, 4.5 mmol). The pH is adjusted to about 8 by adding NMM (0.9 mL, 8.9 mmol) and the mixture is stirred overnight at room temperature with the exclusion of moisture. The solvent is evaporated, the residue is taken up with AcOEt, washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, desiccated on Na_2SO_4 and evaporated to dryness. The crude product is purified by flash-chromatography (eluent: DCM/MeOH 15:1).

Yield: 60 %. Mp: 173-176°C. $[\alpha]^{25}_D = -35.1°$ (c = 0.13, MeOH).

Rf_1: 0.80; Rf_2: 0.90; Rf_3: 0.50.

IR (KBr): 3425, 3259, 1744, 1704, 1676, 1627, 1554 cm^{-1}.

UV: ε= 13400 $M^{-1}cm^{-1}$ at λ= 242.5 nm, ε= 6180 $M^{-1}cm^{-1}$ at λ= 292.4 nm (24 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 7.99 (br, 1H, αNH AlaC), 7.91 (s, 1H, N^4H cytosine), 7.69 (d, 1H, C(6)H cytosine), 7.37 (m, 5H, C_6H_5-Z), 7.21 (d, 1H, C(5)H

cytosine), 6.36 (br, 1H, NH Aib), 5.20 (s, 2H, CH_2-Z), 4.57 (m, 1H, 1 βCH AlaC), 4.44 (m, 2H, αCH and 1 βCH AlaC), 1.50 (s, 6H, 2 CH_3 Aib), 1.38 (s, 9H, CH_3 Boc).

Boc-Aib-AlaC^{Z}-Aib-OMe

To a chilled solution of TFA· H-AlaC^{Z}-Aib-OMe (obtained by treating with TFA in DCM 1.0 g, 1.7 mmol the corresponding Boc-protected peptide) in 5 mL distilled DCM, (Boc-Aib)$_2$O (1.0 g, 2.5 mmol) dissolved in 5 mL distilled DCM is added and the pH is adjusted to about 8 with NMM. After 2 days, more (Boc-Aib)$_2$O (0.4 g, 1 mmol) is added and the solution is stirred overnight. The solvent is evaporated under vacuum, the residue is taken up with AcOEt, washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, desiccated on Na_2SO_4 and evaporated to dryness. The crude product is purified by flash-chromatography (eluent: DCM/MeOH 30:1).

Yield: 85 %. Mp: 173-176 °C. $[\alpha]^{25}_{D}$= -53.9° (c = 0.5, MeOH).

Rf_1 = 0.80; Rf_2 = 0.95; Rf_3 = 0.70.

IR (KBr): 3313, 1744, 1664, 1626, 1552 cm^{-1}.

UV: ε= 13400 $M^{-1}cm^{-1}$ at λ= 240.4 nm, ε= 6000 $M^{-1}cm^{-1}$ at λ= 294.3 nm (18 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 8.66 (d, 1H, αNH AlaC), 7.97 (s, 1H, NH Aib^3), 7.75 (d, 1H, C(6)H cytosine), 7.53 (s, 1H, N^{4}H cytosine), 7.38 (m, 5H, C_6H_5-Z), 7.21 (d, 1H, C(5)H cytosine), 5.25 (s, 2H, CH_2-Z), 5.10 (s, 1H, NH Aib^1), 4.68 (m, 1H, 1 βCH AlaC), 4.23 (m, 2H, αCH and 1 βCH AlaC), 3.62 (s, 3H, OMe), 1.48 (s, 6H, 2 CH_3 Aib), 1.40 (2 s, 15H, 3 CH_3 Boc and 2 CH_3 Aib).

Boc-Aib-AlaC^{H}-Aib-OMe

Boc-Aib-AlaC^{Z}-Aib-OMe (0.047 g, 0.076 mmol) is dissolved in abundant MeOH, N_2 is fluxed in the mixture, Pd catalyst (0.018 g, 10 % on activated charcoal) is added under N_2, then after 10 min H_2 is bubbled. After 2,5 hours H_2 bubbling is stopped, N_2 is bubbled for 10 min and the catalyst is filtered on a celite pad. The mother liquors are evaporated at reduced temperature, taken up with AcOEt and evaporated again, obtaining a solid, then taken up with Et_2O/EP and evaporated to dryness, finally desiccated under vacuum overnight.

Yield: 95 %. Mp: 205-207°C. $[\alpha]_D^{25}$ = -70° (c= 0.6, MeOH).

Rf_1: 0.10. Rf_2: 0.65. Rf_3: 0.05.

IR (KBr): 3423, 1744, 1685, 1647, 1525 cm^{-1}.

UV: ε= 7200 $M^{-1}cm^{-1}$ at λ= 273.2 nm (24.5 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 250 MHz), δ/ppm: 8.83-8.81 (d, *J*= 5.5 Hz, 1H, αNH AlaC), 7.98 (br s, 1H, NH Aib^3), 7.42-7.39 (d, *J*= 7.0 Hz, 1H, C(6)H cytosine), 5.72-5.69 (d, *J*= 7.0 Hz, 1H, C(5)H cytosine), 5.30 (br s, 1H, NH Aib^1), 4.59 (m, 1H, αCH AlaC), 4.31-4.25 (d, *J*= 13.8 Hz, 1 βCH AlaC), 4.17-4.09 (dd, *J*= 13.8, 6.0 Hz, 1 βCH AlaC), 3.65 (s, 3H, OMe), 2.20-1.90 (br s, 2H, N^4H_2 cytosine), 1.49 (s, 6H, 2 CH_3 Aib), 1.41 (s, 15H, 2 CH_3 Aib and CH_3 Boc).

Boc-AlaC^Z-Aib-Lys(Z)-NH_2

A solution in 5 mL distilled DCM of TFA· H-Aib-Lys(Z)NH_2 (obtained by treatment with TFA in DCM of 1.0 g, 1.85 mmol of the corresponding Boc-protected peptide) is cooled in a water/ice bath and basified to pH 8 with NMM (0.61 mL, 5.6 mmol). The active ester formed by dissolving in the minimum amount of anhydrous DMF Boc-AlaC^Z-OH (0.42 g, 0.98 mmol) and adding HOAt (0.15 g, 1.13 mmol) and EDC·HCl (0.23 g, 1.19 mmol) is added to the solution and the pH is adjusted again to 8 with NMM (0.2 mL, 1.19 mmol). The mixture is stirred overnight with the exclusion of moisture. The solvent is evaporated under vacuum, the residue is taken up with AcOEt, washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, desiccated on Na_2SO_4, and evaporated to dryness. The crude product is purified by flash-chromatography (eluent: DCM/EtOH, gradient from 20:1 to 10:1).

Yield: 46 %. Mp: 118-121°C. $[\alpha]^{25}_D$= -37.6° (c = 0.35, MeOH).

Rf_1: 0.75; Rf_2: 0.95; Rf_3: 0.70.

IR (KBr): 3322, 1751, 1658, 1626, 1554 cm^{-1}.

UV: ε= 17400 $M^{-1}cm^{-1}$ at λ= 241.5 nm, ε= 8230 $M^{-1}cm^{-1}$ at λ= 290.2 nm (17 μM in MeOH).

^{1}H-NMR (d_6-DMSO, 200 MHz), δ/ppm: 10.75 (br s, 1H, αNH AlaC), 8.21 (s, 1H, N^4H cytosine), 7.83 (d, 1H, C(6)H cytosine), 7.53 (d, 1H, αNH Lys), 7.38 (m, 5H, C_6H_5-Z N^4H cytosine), 7.33 (m, 5H, C_6H_5-Z εNH Lys), 7.22 (d, 1H, C(5)H cytosine), 7.20-6.90 (3 br s, 4H, $CONH_2$, NH Aib, εNH Lys), 5.18 (s, 2H, CH_2-Z N^4H cytosine), 4.97 (s, 2H, CH_2-Z εNH Lys), 4.35 (m, 1H, αCH AlaC), 4.28 (m, 1H, αCH Lys), 4.03, 3.68 (2 m, 2H, βCH_2 AlaC), 2.96 (m, 2H, εCH_2 Lys), 1.76, 1.54 (2m, 2H, βCH_2 Lys), 1.30 (m, 19H, 3 CH_3 Boc, 2 CH_3 Aib, γCH_2 and δCH_2 Lys).

Boc-Aib-AlaC^{Z}-Aib-Lys(Z)-NH$_2$

A solution of TFA·H-AlaC^{Z}-Aib-Lys(Z)NH$_2$ (obtained by treatment with TFA in DCM of 0.3 g, 0.39 mmol of the corresponding Boc-protected peptide) in 5 mL distilled DCM is cooled in a water/ice bath and NMM (0.86 mL, 0.78 mmol) is added to adjust the pH to about 8. The active ester formed by dissolving Boc-Aib-OH (0.16 g, 0.78 mmol) in the minimum amount of anhydrous DCM and adding HOAt (55 mg, 0.41 mmol) and EDC·HCl (89 mg, 0.46 mmol) is added to the solution and more NMM (50 μL, 0.46 mmol) is added to keep the pH around 8. After stirring at room temperature overnight with the exclusion of moisture, the solvent is evaporated under vacuum, the residue is taken up with AcOEt, washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, desiccated on Na_2SO_4 and evaporated to dryness. The product is purified by flash-chromatography (eluent: DCM/MeOH 15:1).

Yield: 40 %. Mp: 127-134 °C. $[\alpha]^{25}_{D}$= -34.5° (c= 0.45, MeOH).

Rf_1: 0.50; Rf_2: 0.90; Rf_3: 0.25.

IR (KBr): 3315, 1748, 1663, 1626, 1525 cm^{-1}.

UV: ε= 17000 M^{-1}cm^{-1} at λ= 241.3 nm, ε= 8010 M^{-1}cm^{-1} at λ= 289.9 nm (17 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 9.23 (d, 1H, αNH AlaC), 8.21 (s, 1H, NH Aib^3), 7.55 (d, 1H, C(6)H cytosine), 7.65 (d, 1H, αNH Lys), 7.55-7.20 (m, 11H, 2 C_6H_5-Z and N^{4}H cytosine), 7.09 (d, 1H, C(5)H cytosine), 5.35 (s, 1H, NH Aib^1), 5.22 (s, 2H, CH_2-Z N^{4}H cytosine), 5.17 (t, 1H, εNH Lys), 5.05 (s, 2H, CH_2-Z εNH Lys), 4.40-4.00 (m, 4H, αCH Lys, αCH and βCH_2 AlaC), 3.19 (m, 2H, εCH_2 Lys), 1.60-1.25 (m, 27H, 3 CH_3 Boc, 4 CH_3 2 Aib, βCH_2, γCH_2 and δCH_2 Lys).

Z-AlaG^{Cl}-OMe

After dissolving Z-AlaG^{Cl}-OH (0.60 g, 0.17 mmol) in the minimum amount of anhydrous DMF, DMAP (15 mg, 0.12 mmol) is added. The solution is cooled to 0 °C and EDC·HCl (45 mg, 0.24 mmol), then MeOH (0.25 mL, 5.1 mmol) are added. The mixture is allowed to retun to room temperature and stirred overnight, then evaporated to dryness and the crude product is purified by flash-chromatography (eluent: DCM/MeOH/AcOH 20:1:1).

Yield: 30 %. Mp: 125-130 °C. $[\alpha]_D^{25}$ = -42.7° (c= 0.2, MeOH).

Rf_1: 0.65; Rf_2: 0.90; Rf_3: 0.35.

IR (KBr): 3388, 3324, 3210, 1713, 1699, 1611, 1566, 1524 cm^{-1}.

UV: ε= 9800 $M^{-1}cm^{-1}$ at λ= 246.7 nm, ε= 8350 $M^{-1}cm^{-1}$ at λ= 305.1 nm (44 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 7.36 (s, 1H, C(8)H guanine), 7.36 (m, 5H, C_6H_5-Z), 5.88 (d, 1H, αNH), 5.10 (m, 2H, Z-CH_2), 5.02 (s, 2H, N^2H_2 guanine), 4.75 (m, 1H, 1 βCH), 4.54 (m, 2H, αCH and 1 βCH), 3.80 (s, 3H, OMe).

Z-AlaG^{OAt}-Aib-OMe

The active ester obtained by dissolving Z-AlaG^{Cl}-OH (0.49 g, 1.25 mmol) in the minimum amount of anhydrous DMF and adding HOAt (0.19 g, 1.4 mol) and EDC·HCl (0.29 g, 1.48 mmol) is added to a solution of HCl·H-Aib-OMe (0.28 g, 1.8 mmol) in distilled DCM. NMM (0.3 mL, 3.28 mmol) is added to adjust the pH to about 8 and the mixture is stirred at room temperature with the exclusion of moisture for 7 days. The solvent is evaporated under vacuum, the residue is taken up with DCM, washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 % and H_2O, desiccated on Na_2SO_4 and evaporated to dryness. The product is purified by flash-chromatography (eluent: DCM/MeOH, gradient from 30:1 to 25:1).

Yield: 53 %. Mp: 113-118°C. $[\alpha]_D^{25}$= - 48.8° (c = 0.35, MeOH).

Rf_1: 0.60; Rf_2: 0.90; Rf_3 = 0.35.

IR (KBr): 3387, 1730, 1681, 1639, 1570, 1520 cm^{-1}.

UV: ε= 23600 $M^{-1}cm^{-1}$ at λ= 290.7 nm, ε= 22500 $M^{-1}cm^{-1}$ at λ= 245.8 nm (48 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 8.68 (dd, 1H, H-6' OAt), 8.45 (dd, 1H, H-4' OAt), 7.68 (s, 1H, C(8)H guanine), 7.43 (q, 1H, H-5' OAt), 7.30 (m, 5H, C_6H_5-Z), 6.99 (br s, 1H, Aib NH), 6.49 (d, 1H, αNH AlaG), 5.09 (s, 2H, CH_2-Z), 4.80 (s, 2H, N^2H_2 guanine), 4.61 (m, 1H, αCH AlaG), 4.40 (m, 2H, βCH_2 AlaG), 3.66 (s, 3H, OMe), 1.42 (2 s, 6H, CH_3 Aib).

Z-AlaG^{OH}-Aib-OMe

Z-AlaG^{OAt}-Aib-OMe (0.17 g, 2.9 mmol) is dissolved in 4 mL TFA/H_2O 3:1 with formation of a strongly red solution and stirred at room temperature for 3 days, with partial decoloration of the mixture. The excess TFA is stripped with N_2, the solution is diluted with H_2O and neutralized with $NaHCO_3$ 5%, then extracted with AcOEt several times after addition of NaCl. Organic extracts are pooled, washed with $NaHCO_3$ 5%, brine, $KHSO_4$ 5% and brine, then desiccated on Na_2SO_4 and evaporated to dryness to yield reddish crude

product and HOAt (about 1:2). The acidic aqueous phase is neutralized, added with plenty of NaCl and extracted several times with AcOEt, the organic extracts are washed with brine, then desiccated on Na_2SO_4 and evaporated to dryness to yield colourless pure product. The crude product is dissolved in $CHCl_3$ and purified by flash-chromatography (eluent: DCM/MeOH 10:1).

Yield: 67%. Mp: 135-137°C. $[\alpha]_D^{25}$ = -20.5° (c= 0.4, MeOH).

Rf_1: 0.10. Rf_2: 0.75. Rf_3: 0.05.

IR (KBr): 3424, 1683, 1627, 1594, 1574, 1533 cm^{-1}.

UV: ε=15500 $M^{-1}cm^{-1}$ at λ= 255.7 nm, shoulder at λ= 278 nm (47 μM in MeOH).

HPLC (Agilent, 10-70 % B' in 20 min): t_r 16.0 min.

^{1}H-NMR (*d_6*-DMSO, 250 MHz), δ/ppm: 10.55 (br s, 1H, N(1)H guanine), 8.39 (br s, 1H, NH Aib), 7.64 (d, *J*= 9.0 Hz, 1H, αNH AlaG), 7.51 (s, 1H, C(8)H guanine), 7.35-7.24 (m, 5H, C_6H_5-Z), 6.47 (s, 2H, N^2H_2 guanine), 4.98 (s, 2H, CH_2-Z), 4.46-4.41 (m, 1H, αCH AlaG), 4.29-4.23 (dd, *J*= 15.3, 13.6 Hz, 1H, 1 βCH AlaG), 4.07-3.98 (dd, *J*= 13.6, 8.8 Hz, 1H, 1 βCH AlaG), 3.55 (s, 3H, OMe), 1.33 (s, 3H, 1 CH_3 Aib), 1.30 (s, 3H, 1 CH_3 Aib).

Boc-Aib-AlaG^{OH}-Aib-OMe

Boc-Aib-OH (0.05 g, 0.25 mmol) is dissolved in the minimum amount of distilled DCM, cooled to 0°C in a water/ice bath and HOAt (0.042 g, 0.30 mmol) and EDC·HCl (0.061 g, 0.31 mmol) are added. The active ester formed is added to H-AlaG^{OH}-Aib-OMe (obtained by catalytic hydrogenolysis of 0.030 g, 0.064 mmol, of the corresponding Z-protected peptide) suspended in 1 mL of anhydrous DMF and NMM (25 μL, 0.22 mmol) is added to take the pH to about 9. Further 3 mL anhydrous DMF are added and the slowly dissolving the mixture is stirred at room temperature with exclusion of moisture. After 7 days the solvent is evaporated at reduced pressure. The resulting oil is directly purified by flash-chromatography (eluent: $CHCl_3$/MeOH, gradient from 10:1 to 7:1, dry loading). Tubes containing the product are pooled and evaporated to yield pinkish crude product, which is purified by preparative HPLC (Shimadzu, isocratic 25 % B') and lyophilized to yield a colourless solid.

Yield: 15 %. Mp: 168-170°C. $[\alpha]_D^{25}$ = -28.5° (c= 0.165, H_2O).

Rf_1: 0.05. Rf_2: 0.70. Rf_3: 0.00. MS calcd$[M+H]^+$: 523.26; found: 523.26.

IR (KBr): 3426, 1685, 1635, 1538, 1534 cm^{-1}.

UV: ε= 12900 $M^{-1}cm^{-1}$ at λ= 255.3 nm, shoulder at λ= 274.4 nm (32.7 μM in MeOH).

HPLC (Agilent, 10-70 % B' in 20 min): t_r 10.7 min.

^{1}H-NMR (d_6-DMSO, 250 MHz), δ/ppm: 10.69 (br s, 1H, N(1)H guanine), 8.29-8.26 (d, *J*= 6.6 Hz, 1H, αNH AlaG), 7.77 (br s, 1H, NH Aib), 7.54 (s, 1H, C(8)H AlaG), 7.36 (s, 1H, NH Aib), 6.54 (s, 2H, N^2H_2 guanine), 4.48-4.43 (m, 1H, αCH AlaG), 4.31-4.16 (m, 2H, βCH_2 AlaG), 3.54 (s, 3H, OMe), 1.34 (2 s, 12H, 1 CH_3 1 Aib and CH_3 Boc), 1.28 (s, 3H, 1 CH_3 1 Aib), 1.20 (s, 6H, 2 CH_3 Aib).

Z-AlaT-OMe

Z-AlaT-OH (0.05 g, 0.14 mmol) is dissolved in the minimum amount of MeOH, DMAP (0.01 g, 0.08 mmol) is added, and the solution is cooled to 0°C in a water/ice bath. EDC·HCl (0.10 g, 0.50 mmol) is dissolved in the minimum amount of distilled DCM and added dropwise to the mixture, which is allowed to return to room temperature and stirred with exclusion of moisture for 2 hours. AcOEt is added to the mixture, which is thereafter concentrated to an oil and purified by flash-chromatography (eluent: DCM/MeOH 20:1).

Yield: 48 %. Mp: 159-161°C. $[\alpha]_D^{25}$ = -57.6° (c= 0.5, MeOH).

Rf_1: 0.70. Rf_2: 0.85. Rf_3: 0.45. MS calcd$[M+H]^+$: 362.14; found: 362.15.

IR (KBr): 3354, 1744, 1718, 1698, 1682, 1534 cm^{-1}.

UV: ε= 4490 $M^{-1}cm^{-1}$ at λ= 267.4 nm (13.3 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 250 MHz), δ/ppm: 8.64 (br s, 1H, N(3)H thymine), 7.35-7.32 (m, 5H, C_6H_5-Z), 6.90 (s, 1H, C(6)H thymine), 5.80-5.77 (d, *J*= 6.8 Hz, 1H, αNH), 5.17-5.05 (2 d, *J*= 12.2 Hz, 2H, CH_2-Z), 4.59-4.52 (m, 1H, αCH), 4.13-4.08 (m, 2H, βCH_2), 3.78 (s, 3H, OMe), 1.82 (s, 3H, C^5H_3 thymine).

Z-AlaT-Aib-O^{t}Bu

Z-AlaT-OH (0.50 g, 1.4 mmol) is dissolved in 6 mL anhydrous DMF, the solution is cooled to 0°C in a water/ice bath then HOAt (0.224 g, 1.6 mmol) and EDC·HCl (0.327 g, 1.7 mmol) are added. H-Aib-O^{t}Bu (obtained by catalytic hydrogenolysis of 0.73 g, 2.4 mmol, of the corresponding Z-protected derivative) is added and the pH is adjusted to 8.5 with NMM (0.15 mL, 1.3 mmol). After 5 days the solvent is evaporated, the mixture is taken up in AcOEt, then washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 %/brine 1:2 and H_2O. The solvent is evaporated and some product precipitates from MeOH. The rest of the crude is purified by flash-chromatography (eluent: 13:9 AcOEt/EP + 1% MeOH).

Yield: 64 %. Mp: 148-150°C. $[\alpha]_D^{25}$ = -38° (c= 0.2, MeOH).

Rf_1: 0.60; Rf_2: 1.00; Rf_3: 0.45. MS calcd $[M+H]^+$: 489.18; found: 489.22.

IR (KBr): 3400, 3354, 1730, 1684, 1653, 1530 cm^{-1}.

UV: ε= 6650 $M^{-1}cm^{-1}$ at λ= 268.0 nm (13 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 8.22 (br s, 1H, N(3)H thymine), 7.34 (m, 5H, C_6H_5-Z), 7.16 (br s, 1H, NH Aib), 7.04 (s, 1H, C(6)H thymine), 6.20-6.16 (d, *J*= 7.2 Hz, 1H, αNH AlaT), 5.11 (s, 2H, CH_2-Z), 4.54-4.45 (m, 1H, αCH AlaT), 4.22-4.13 (dd, *J*= 14.2, 3.7 Hz, 1H, 1 βCH AlaT), 3.96-3.85 (dd, *J*= 14.2, 3.7 Hz, 1H, 1 βCH AlaT), 1.85 (s, 3H, C^5H_3 thymine), 1.49-1.47 (2 s, 6H, 2 CH_3 Aib), 1.45 (s, 9H, O^{t}Bu).

Z-Aib-AlaT-Aib-O^{t}Bu

$(Z\text{-}Aib)_2O$ (0.52 g, 1.2 mmol) is dissolved in 2 mL anhydrous MeCN, then added to H-AlaT-Aib-O^{t}Bu (obtained by hydrogenolysis of 0.330 g, 0.68 mmol, of the corresponding Z-protected peptide). NMM (0.125 mL, 1.1 mol) is added, then the mixture is stirred at room temperature with the exclusion of moisture for 3 days. The solvent is evaporated, the mixture is taken up with AcOEt and washed with $KHSO_4$ 5 %, H_2O. The organic phase is stirred for 30 min in $NaHCO_3$ 5 %/H_2O 1:2 in order to hydrolyze the excess anhydride, then phases are separated, the organic phase is washed again with $NaHCO_3$ 5 %/H_2O 1:2 and with H_2O, desiccated on Na_2SO_4 and concentrated under vacuum. Pure product precipitates from AcOEt/EP, more product is obtained from the mother liquors by flash-chromatography (eluent: DCM/EtOH 20:1) .

Yield: 73 %. Mp 114-116°C (from AcOEt/EP). $[\alpha]_D^{25}$ = -43° (c= 0.4, MeOH).

Rf_1: 0.55 Rf_2: 0.95 Rf_3: 0.40. MS calcd$[M+H]^+$: 574.29; found: 574.30.

IR (KBr): 3423, 3336, 1734, 1682, 1673, 1529 cm^{-1} .

UV: ε= 10200 $M^{-1}cm^{-1}$ at λ= 268.2 nm (13 μM in MeOH).

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 9.71 (br s, 1H, N(3)H thymine), 8.23-8.20 (d, *J*= 7.2 Hz, 1H, αNH AlaT), 7.49 (s, 1H, C(6)H thymine), 7.30 (m, 5H, C_6H_5-Z), 7.04 (br s, 1H, NH Aib^3), 6.23 (br s, 1H, NH Aib^1), 5.02 (s, 2H, CH_2-Z), 4.64 (m, 1H, αCH AlaT), 4.03-4.01 (m, 2H, βCH_2 AlaT), 1.88 (s, 3H, C^5H_3 thymine) , 1.54 (s, 3H, 1 CH_3 Aib^3), 1.45-1.44 (2s, 15H, 2 CH_3 Aib^1 and 3 CH_3 O^{t}Bu), 1.38 (s, 3H, 1 CH_3 Aib^3).

Z-Aib-AlaT-Aib-OH

Z-Aib-AlaT-Aib-O^{t}Bu (0.125 g, 0.22 mmol) is dissolved in the minimum amount of DCM, an equal volume of deprotecting mixture (TFA/H_2O 25:1) is added, then the solution is

stirred at room temperature for 1,5 hours. The solvent mixture is stripped by bubbling N_2 in the solution, then the residue is taken up with DCM and stripped again with N_2 several times. The residue is taken up with Et_2O and evaporated at reduced pressure several times, until a dry odourless solid is obtained, which is desiccated under vacuum overnight.

Yield: 98%. Mp> 250°C. $[\alpha]_D^{25}$ = -38° (c= 0.6, MeOH).

Rf_1: 0.05. Rf_2: 0.80. Rf_3: 0.05.

IR (KBr): 3403, 1782, 1697,1676, 1528 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 9.98 (s, 1H, N(3)H thymine), 8.30-8.27 (d, *J*= 6.8 Hz, 1H, αNH AlaT), 7.68 (br s, 1H, NH Aib^3), 7.29 (s, 1H, C(6)H thymine), 6.04 (br s, 1H, NH Aib^1), 5.11-4.95 (2 d, *J*= 12.4 Hz, 2H, CH_2-Z), 4.72 (m, 1H, αCH AlaT), 4.07 (m, 2H, βCH_2 AlaT), 1.85 (s, 3H, C^5H_3 AlaT), 1.54 (s, 3H, 1 CH_3 1 Aib), 1.50 (s, 3H, 1 CH_3 1 Aib), 1.47 (s, 3H, 1 CH_3 1 Aib), 1.45 (s, 3H, 1 CH_3 1 Aib).

Z-Aib-AlaT^{All}-Aib-O^{t}Bu

Z-Aib-AlaT-Aib-O^{t}Bu (0.059 g, 0.10 mmol) is dissolved in 2.0 mL DCM, DBU (16 μL, 0.10 mmol) and AllBr (16 μL, 0.18 mmol) are added. The mixture is stirred 3 days at room temperature with the exclusion of moisture. After quenching the reaction with the addition of 0.05 mL MeOH, the mixture is directly purified by flash-chromatography (eluent: DCM/MeOH 60:1 to 40:1) to yield an oil.

Yield: 27%. $[\alpha]_D^{25}$ = -23° (c 0.3, MeOH).

Rf_1: 0.70. Rf_2: 0.90. Rf_3: 0.45. MS calcd $[M\text{-}tBu]^+$: 556.24, found: 556.23.

IR (KBr): 3397, 1732, 1700, 1666, 1639, 1529 cm^{-1}.

HPLC (Agilent, 10-70 % B' in 30 min): t_r 16.37 min.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 8.06-8.03 (d, *J*= 7.0 Hz, 1H, αNH AlaT), 7.45 (s, 1H, C(6)H thymine), 7.34 (m, 5H, C_6H_5-Z), 7.27 (s, 1H, NH Aib^3), 5.90-5.75 (m, 1H, C(2')H All), 5.25 (s, 1H, NH Aib^1), 5.23 (m, 1H, 1 C(3')H All), 5.16 (m, 1H, 1 C(3')H All), 5.11-5.00 (m, 2H, CH_2-Z), 4.67-4.61 (m, 1H, αCH AlaT), 4.51-4.49 (2 s, 2H, βCH_2 AlaT), 4.20-4.09 (m, 2H, C(1')H_2 All), 1.90 (s, 3H, C^5H_3 AlaT), 1.51 (s, 3H, 1 CH_3 1 Aib), 1.47 (s, 3H, 1 CH_3 1 Aib), 1.45 (s, 3H, 1 CH_3 1 Aib), 1.43 (s, 9H, CH_3 O^{t}Bu), 1.39 (s, 3H, 1 CH_3 1 Aib).

Ac-Aib-AlaT-Aib-O^{t}Bu

To H-Aib-AlaT-Aib-O^{t}Bu (obtained by catalytic hydrogenolysis of 0.055 g, 0.096 mmol, of the corresponding Z-protected tripeptide) dissolved in anhydrous MeCN, Ac_2O (0.20 mL, 2.0 mmol) is added and the reaction is stirred at room temperature with exclusion of moisture for 3 hours, then it is quenched by addition of MeOH. The mixture is evaporated at reduced pressure and the resulting oil is taken up with PhMe and evaporated again, until a solid is obtained, from which the product is obtained by flash-chromatography (eluent: $CHCl_3$/MeOH, gradient from 22:1 to 15:1, dry loading).

Yield: 63 %. Mp> 250°C. $[\alpha]_D^{25}$ = -26.3° (c= 0.38, MeOH).

Rf_1: 0.40. Rf_2: 0.80. Rf_3: 0.20.

IR (KBr): 3421, 3315, 1671, 1533, 1515 cm^{-1}.

UV: ε= 10000 $M^{-1}cm^{-1}$ at λ= 267.5 (31.6 μM in MeOH).

HPLC (Agilent, 10-70 % B' in 20 min) t_r 11.1 min.

^{1}H-NMR ($CDCl_3$, 250 MHz), δ/ppm: 10.73 (br s, 1H, N(3)H thymine), 8.31-8.28 (d, *J*= 7.2 Hz, 1H, αNH AlaT), 7.68 (br s, 1H, NH Aib), 7.60 (s, 1H, C(6)H thymine), 7.20 (br s, 1H, NH Aib), 4.76-4.71 (m, 1H, αCH AlaT), 4.24-4.07 (m, 2H, βCH_2 AlaT), 1.91 (s, 6H, CH_3 Ac and C^5H_3 thymine), 1.55 (s, 3H, 1 CH_3 1 Aib), 1.47 (s, 3H, 1 CH_3 1 Aib), 1.44 (s, 3H, 1 CH_3 1 Aib), 1.41 (s, 9H, CH_3 O^{t}Bu) 1.36 (s, 3H, 1 CH_3 1 Aib).

Z-AlaT-Aib-Lys(Boc)-NH_2

Z-AlaT-OH (0.41 g, 1.2 mmol) is dissolved in anhydrous DMF, cooled to 0°C in a water/ice bath and HOAt (0.19 g, 1.4 mmol) and EDC·HCl (0.28 g, 1.5 mmol) are added. H-Aib-Lys(Boc)-NH_2 (obtained by catalytic hydrogenolysis of 0.89 g, 1.7 mmol, of the corresponding Z-protected derivative) is added and the pH is adjusted to about 8.5 with NMM (0.09 mL, 0.8 mmol). The mixture allowed to return to room temperature and stirred with exclusion of moisture. After 6 days, the solvent is removed at reduced pressure, the resulting oil is taken up with AcOEt, washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 %/H_2O 1:2 and H_2O, then desiccated on Na_2SO_4 and evaporated to an oil. The crude product is purified by flash-chromatography (eluent: DCM/MeOH, gradient from 18:1 to 9:1).

Yield: 58 %. Mp: 142-145°C (from EP). $[\alpha]_D^{25}$ = -35.9° (c= 0.34, MeOH).

Rf_1: 0.30. Rf_2: 0.85. Rf_3: 0.20.

IR (KBr): 3410, 1675, 1521 cm^{-1}.

UV: ε= 15000 $M^{-1}cm^{-1}$ at λ= 267.3 nm (16.5 μM in MeOH).

HPLC (Agilent, 10-50 % B' in 25 min): t_r 20.0 min.

^{1}H-NMR (d_6-DMSO, 200 MHz), δ/ppm: 8.32 (br s, 1H, N(3)H thymine), 7.64-7.60, 7.47-7.43 (2 d, *J*= 6.2 Hz, 2H αNH AlaT and αNH Lys), 7.35-7.30 (m, 7H, C_6H_5-Z, C(6)H thymine and NH Aib), 7.11 (br s, 1H, 1 CONH), 7.00 (br s, 1H, 1 CONH), 6.76-6.72 (t, *J*= 4.3 Hz, 1H, εNH Lys), 5.06-4.89 (m, 2H, CH_2-Z), 4.41-4.33 (m, 1H, 1 αCH), 4.23-4.19 (m, 2H, 1 αCH and 1 βCH AlaT), 3.64-3.55 (m, 1H, 1 βCH AlaT), 2.88-2.80 (m, 2H, εCH_2 Lys), 1.67 (s, 4H, C^5H_3 thymine and 1 βCH Lys), 1.59-1.49 (m, 1H, 1 βCH Lys), 1.34 (s and m, 16H, 1 CH_3 Aib, 3 CH_3 O^{t}Bu, γCH_2 and δCH_2 Lys), 1.30 (s, 3H, 1 CH_3 Aib).

Z-Aib-Aib-AlaT-Aib-O^{t}Bu

$(Z\text{-Aib})_2O$ (0.31 g, 0.61 mmol) is dissolved in anhydrous MeCN and added to H-Aib-AlaT-Aib-O^{t}Bu (obtained by catalytic hydrogenolysis of 0.145 g, 0.26 mmol, of the corresponding Z-protected derivative) dissolved in anhydrous MeCN. NMM (0.135 mL, 0.12 mmol) is added to adjust the pH at about 8.5 and the mixture is stirred at room temperature with the exclusion of moisture. After 5 days the solvent is evaporated at reduced pressure, the residue is taken up with AcOEt, washed with $KHSO_4$ 5 % and H_2O, stirred with $NaHCO_3$ 5 % for 1 hour, then H_2O is added and the organic layer is washed with $NaHCO_3$ 5 %/H_2O 1:2 and H_2O, desiccated on Na_2SO_4 and evaporated to an oil. The product is purified by flash-chromatography (eluent: DCM/MeOH, gradient from 35:1 to 25:1).

Yield: 25 %. Mp: 131-133°C. Rf_1: 0.65. Rf_2: 0.90. Rf_3: 0.30.

IR (KBr): 3421, 1732, 1689, 1674, 1532 cm^{-1}.

UV: ε= 10500 $M^{-1}cm^{-1}$ at λ= 267.9 nm (47 μM in MeOH).

HPLC (Agilent, 10-70 % B' in 20 min): t_r 19.4 min.

^{1}H-NMR ($CDCl_3$, 250 MHz), δ/ppm: 10.38 (br s, 1H, N(3)H thymine), 8.10 (br s, 1H, NH Aib), 7.77-7.74 (d, *J*= 8.0 Hz, 1H, αNH AlaT), 7.51 (s, 1H, C(6)H thymine), 7.19 (m, 5H, C_6H_5-Z), 7.16 (br s, 1H, NH Aib), 5.36 (br s, 1H, NH Aib[1]), 5.17-5.04 (2 d, *J*= 12.5 Hz, 2H, CH_2-Z), 4.63-4.56 (m, 1H, αNH AlaT), 4.50-4.46 (m, 1H, 1 βCH AlaT), 4.05-4.00 (m, 1H, 1 βCH AlaT), 1.87 (s, 3H, C^5H_3 thymine), 1.73 (s, 3H, 1 CH_3 1 Aib), 1.53 (s, 3H, 1 CH_3 1 Aib), 1.52 (s, 3H, 1 CH_3 1 Aib), 1.51 (s, 3H, 1 CH_3 1 Aib), 1.49 (s, 3H, 1 CH_3 1 Aib), 1.46 (s, 9H, CH_3 O^{t}Bu), 1.26 (s, 3H, 1 CH_3 1 Aib).

Z-Aib-AlaT-Aib-Lys(Boc)-NH_2

$(Z\text{-Aib})_2O$ (0.58 g, 1.26 mmol) dissolved in anhydrous MeCN is added to H-AlaT-Aib-Lys(Boc)-NH_2 (obtained by catalytic hydrogenolysis of 0.46 g, 0.65 mmol, of the corresponding Z-protected peptide), cooled to 0°C in a water/ice bath, then 1.5 mL anhydrous DMF are added and a solution is obtained. The pH is adjusted to 8 with NMM (0.08 mL, 0.71 mmol), the mixture is allowed to return to room temperature and stirred with the exclusion of moisture for 7 days. The solvent is evaporated at reduced pressure, the residue is taken up with AcOEt, washed with $KHSO_4$ 5 % and H_2O, stirred with $NaHCO_3$ 5 %/H_2O 1:2 for 1 hour, then washed with $NaHCO_3$ 5 %/H_2O 1:2 and H_2O, desiccated on Na_2SO_4 and evaporated to a solid foam, which is purified by flash-chromatography (eluent: DCM/MeOH, gradient from 15:1 to 13:1).

Yield: 75 %. Mp: 124-126°C. $[\alpha]_D^{25}$ = -55.6° (c= 0.25, MeOH).

Rf_1: 0.25. Rf_2: 0.80. Rf_3: 0.15. MS calcd$[M+H]^+$: 745.39; found: 745.38.

IR (KBr): 3329, 1675, 1525 cm^{-1}.

UV: ε= 13400 $M^{-1}cm^{-1}$ at λ= 267.3 nm (17.8 μM in MeOH).

HPLC (Agilent, 20-60 % B' in 20 min): t_r 19.4 min.

^{1}H-NMR ($CDCl_3$, 250 MHz), δ/ppm: 9.22 (br s, 1H, N(3)H thymine), 8.63-8.62 (d, *J*= 3.5 Hz, 1H, αNH AlaT), 7.92 (s, 1H, C(6)H thymine), 7.34-7.33 (m, 5H, C_6H_5-Z), 7.17-7.13 (d, *J*= 8.0 Hz, 1H, αNH Lys), 7.03 (br s, 2H, $CONH_2$), 6.59 (br s, 1H, NH Aib^3), 5.78 (br s, 1H, NH Aib^1), 5.05 (s, 2H, CH_2-Z), 4.62 (br s, 1H, εNH Lys), 4.28-4.15 (m, 3H, αCH Lys, αCH and 1 βCH AlaT), 4.06-3.98 (dd, *J*= 15.0, 6.3 Hz, 1H, βCH AlaT), 3.14-3.09 (m, 1H, 1 εCH Lys), 3.03-2.95 (m, 1H, 1 εCH Lys), 2.05-1.95 (m, 1H, 1 βCH Lys), 1.88 (s, 3H, C^5H_3 thymine), 1.87-1.80 (m, 1H, 1 βCH Lys), 1.52-1.43 (2 s and m, 13H, 3 CH_3 Aib, γCH_2 and δCH_2 Lys), 1.42-1.41 (2 s, 12H, 1 CH_3 1 Aib and 3 CH_3 Boc).

Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu

H-Aib-AlaT-Aib-O^{t}Bu (obtained by catalytic hydrogenolysis of 0.105 g, 0.18 mmol, of the corresponding Z-protected peptide) is dissolved in 1 mL anhydrous DMF and added to a solution of Z-Aib-AlaT-Aib·Oxl (0.088g, 0.18 mmol, obtained by the activation of Z-Aib-AlaT-Aib-OH, 0.093 g, 0.18 mmol, with EDC·HCl, 0.045 g, 0.23 mmol) in 1.5 mL DMF. NMM (0.11 mL, 0.10 mmol) is added and the mixture is stirred at room temperature with the exclusion of moisture for 10 days. The solvent is evaporated and the resulting oil is taken up in AcOEt, washed with $KHSO_4$ 5 %, H_2O, $NaHCO_3$ 5 %/H_2O 1:2 and H_2O, then

desiccated on Na_2SO_4 and concentrated under vacuum. The product precipitates from AcOEt/EP.

Yield: 61 %. Mp: 156-157°C. $[\alpha]_D^{25}$ = -22° (c= 0.4, MeOH).

Rf_1: 0.40; Rf_2: 0.80; Rf_3: 0.25. MS calcd$[M+H]^+$: 939.46; found: 939.42.

IR (KBr): 3417, 1674, 1528 cm^{-1}.

UV: ε= 12100 $M^{-1}cm^{-1}$ at λ= 268.8 nm (11 μM in MeOH).

HPLC (Agilent, 44-46 % B' in 30 min): *method A*, t_r 18.84 min (D,L + L,D) 9.5 %, t_r 18.75 min (L,L + D,D) 88.7 %; *method B*: t_r 18.90 min (D,L + L,D) 1.0 %, t_r 18.76 min (L,L + D,D) 99.0 %.

^{1}H-NMR (d_6-DMSO, 200 MHz), δ/ppm: 11.32-11.30 (2 br s, 2H, 2 N(3)H of 2 thymines), 8.60-8.57 (d, *J*= 6.8 Hz, 1H, αNH 1 AlaT), 7.99 (s, 1H, NH 1 Aib), 7.86 (s, 1H, NH Aib), 7.72-7.60 (2 s and d, *J*= 8 Hz, 3H, 2 C(6)H of 2 thymines and αNH 1 AlaT), 7.37 (m, 5H, C_6H_5-Z), 7.26 (s, 1H, NH 1 Aib), 7.19 (s, 1H, NH 1 Aib), 5.15-4.99 (2 d, *J*= 12.4 Hz, 2H, CH_2-Z), 4.51-4.25 (m, 4H, 2 αCH and 2 βCH of 2 AlaT), 3.89-3.77 (dd, *J*= 13.0, 10.4 Hz, 1H, 1 βCH 1 AlaT), 3.67-3.55 (m, 1H, 1 βCH 1 AlaT), 1.72 (s, 3H, 1 C^5H_3 1 AlaT), 1.70 (s, 3H, 1 C^5H_3 1 AlaT), 1.37-1.23 (m, 33H, 8 CH_3 of 4 Aib and 3 CH_3 O^{t}Bu).

Z-(Aib-AlaT-Aib)$_2$-Lys(Boc)-NH$_2$

Z-Aib-AlaT-Aib·Oxl (0.17 g, 0.34 mmol, obtained by activation of Z-Aib-AlaT-Aib-OH, 0.30 g, 0.58 mmol, with EDC·HCl, 0.18 g, 0.94 mmol) is suspended in anhydrous MeCN and added to H-Aib-AlaT-Aib-Lys(Boc)-NH_2 (obtained by catalytic hydrogenolysis of 0.33 g, 0.40 mmol, of the corresponding Z-protected peptide) and the resulting suspension is heated at gentle reflux, with gradual dissolution. After 2 days more Z-Aib-AlaT-Aib·Oxl (0.086 g, 0.17 mmol, obtained by activation of Z-Aib-AlaT-Aib-OH, 0.10 g, 0.19 mmol, with EDC·HCl, 0.065 g, 0.32 mmol) is added and the pH is adjusted to about 8 with TEA (0.07 mL, 0.50 mmol). The reaction is carried on 2 days at room temperature and 2 days at gentle reflux. After 6 days the solvent is evaporated at reduced pressure to a reddish oil, which is purified by flash-chromatography (eluent: $CHCl_3$/MeOH, gradient 15:1 to 7:1, dry loading).

The product contains minor quantities of several diastereoisomers (HPLC analysis) which can not be easily separated by preparative techniques. Purity (estimated by HPLC): 70%.

Crude yield: 34 %. Yield: 24 % (taking into account product purity).

Rf_1: 0.10. Rf_2: 0.85. Rf_3: 0.05. MS calcd$[M+H]^+$: 1110.56; found: 1110.55.

IR (KBr): 3422, 1674, 1526 cm^{-1}.

UV: ε= 20100 $M^{-1}cm^{-1}$ at λ= 270.5 nm (13.7 μM in MeOH).

HPLC (Agilent, 20-50 % B' in 30 min): t_r 26.45 min, 12 % , t_r 27.07 min, 69 % (product), t_r 27.84 min, 10 % .

^{1}H-NMR (d_3-MeCN/$CDCl_3$ 3:1, 250 MHz), δ/ppm: 9.94 (br s, 1H, N(3)H 1 AlaT), 9.61 (br s, 1H, N(3)H 1 AlaT), 7.98-7.59 (m, 3H, 2 αNH 2 AlaT and 1 αNH), 7.34 (m, 6H, C_6H_5-Z and 1 C(6)H 1 thymine), 7.21-6.82 (m, 4H, $CONH_2$, 1 C(6)H 1 thymine and 1 αNH), 6.39 (br s, 1H, NH 1 Aib), 5.75 (br s, 1H, NH Aib[1]), 5.32 (t, 1H, εNH Lys), 5.15-5.01 (m, 2H, Z-CH_2), 4.42-4.37 (m, 1H, 1 αCH), 4.22-4.04 (m, 4H, 2 αCH and 2 βCH AlaT), 3.98-3.93 (m, 1H, 1 βCH 1 AlaT), 3.81-3.75 (m, 1H, 1 βCH 1 AlaT), 2.99 (m, 2H, εCH_2 Lys), 1.78 (2 s, 6H, 2 C^5H_3 2 thymines), 1.52-1.46 (m, 8H, 2 CH_3 Aib and δCH_2 Lys), 1.41 (s, 8H, 2 CH_3 Aib and γCH_2 Lys), 1.36 (s, 18H, 3 CH_3 Aib and 3 CH_3 Boc), 1.32 (s, 3H, 1 CH_3 1 Aib).

H-(Aib-AlaT-Aib)$_2$-Lys(Boc)-NH_2

Z-(Aib-AlaT-Aib)$_2$-Lys(Boc)-NH_2 (0.10 g, 0.09 mmol crude) is dissolved in abundant MeOH, N_2 is fluxed in the mixture, Pd catalyst (0.038 g, 10% on activated charcoal) is added under N_2, then after 10 min H_2 is bubbled. After 5,5 hours H_2 bubbling is stopped, N_2 is bubbled for 10 min and the catalyst is filtered on a celite pad. The mother liquors are evaporated at reduced temperature, taken up with MeCN, then with AcOEt until a solid is obtained, which is desiccated under vacuum overnight. The minor diastereoisomers present can not be easily separated by preparative techniques. Purity (extimated by HPLC): 70%.

Yield: 98 % (crude).

MS calcd$[M+H]^+$: 976.52; found: 976.51.

IR (KBr): 3421, 1663, 1523 cm^{-1}.

HPLC (Agilent, 10-25 % B' in 30 min): t_r 22.65 min, 72 % (product), t_r 23.78 min, 8 %, t_r 25.20 min, 17 %, t_r 26.34 min, 2 % .

TFA·Z-(Aib-AlaT-Aib)$_2$-Lys(H)-NH_2

Z-(Aib-AlaT-Aib)2-Lys(Boc)-NH_2 (2.5 mg, 2.3 μmol) is dissolved in the minimum amount of DCM, then an equal volume of TFA is added and the mixture is stirred for 10 min. The solvents are stripped with N_2, the residue is taken up with DCM and stripped again with N_2, finally taken up with MeOH and evaporated to a solid, which is desiccated

under vacuum overnight. The minor diastereoisomers present can not be easily separated by preparative techniques. Purity (estimated by HPLC): 70%.

Yield: 97 % (crude).

MS calcd$[M+H]^+$: 1010.50; found: 1010.46.

HPLC (Agilent, 10-30% B' in 30 min): t_r 27.23 min, 6 %, t_r 28.00 min, 15 %, t_r 29.18 min, 72 % (product), t_r 30.1 min, 5 % .

2 TFA·H-(Aib-AlaT-Aib)$_2$-Lys(H)-NH$_2$

To H-(Aib-AlaT-Aib)$_2$-Lys(Boc)-NH$_2$ (0.08 g, 0.08 mmol) 0.52 mL deprotecting mixture TFA/H_2O 25:1 is added with sudden dissolution and pale pink coloration. After 1 hour the mixture is diluted with DCM, solvents are stripped with N_2, the residue is taken up with DCM and stripped again with N_2 2 times, then taken up with MeOH. The crude product is purified by preparative HPLC (Shimadzu, 1.5-4.8 % B' in 22 min), then lyophilized.

Yield: 95 % (crude). Yield: 18 % (purified). Mp: 176-178°C

$[\alpha]_D^{25}$ = -27° (c= 0.18, MeOH). Rf_1: 0.00. Rf_2: 0.05. Rf_3: 0.00.

MS calcd$[M+H]^+$: 876.47; found: 876.52.

IR (KBr): 3418, 1674, 1532 cm^{-1}.

UV: ε= 15800 $M^{-1}cm^{-1}$ at λ= 269.3 nm (17.1 μM in H_2O).

HPLC (Agilent, 1.5-6 % B' in 30 min), crude: t_r 17.44 min, 7 %, t_r 18.52 min (product), 71 %, t_r 19.49 min, 11 %; purified: t_r 17.44 min, 4.6 %, t_r 18.52 min (product), 95.4 %.

^{1}H-NMR (*d_6*-DMSO, 400 MHz), δ/ppm: 11.33 (br s, 1H, N(3)H 1 AlaT), 11.31 (br s, 1H, N(3)H 1 AlaT), 8.56, (br s, 1H, NH Aib^3), 8.38 (d, 1H, αNH AlaT^2), 8.14 (br s, 3 NH, 1 NH_3^+), 7.94 (br s, 1H, NH Aib^4), 7.85 (d, 1H, αNH AlaT^5), 7.64 (br s, 4H, NH Aib^6 and 1 NH_3^+), 7.42 (s, 1H, C(6)H AlaT^2), 7.33 (d, 1H, αNH Lys7), 7.31 (s, 1H, C(6)H AlaT^5), 7.04 (br s, 1H, 1 CONH), 6.94 (br s, 1H, 1 CONH), 4.68 (m, 1H, αCH AlaT^2), 4.36 (m, 1H, αCH AlaT^5), 4.28-4.26 (m, 1H, 1 βCH AlaT^2), 4.11 (m, 1H, αCH Lys^7), 4.01-3.96 (m, 2H, βCH_2 AlaT^5), 3.73 (m, 1H, 1 βCH AlaT^2), 2.73 (m, 2H, εCH2 Lys^7), 1.80 (m, 1H, 1 βCH Lys^7), 1.72 (s, 3H, C^5H_3 AlaT^2), 1.72-1.70 (s, 3H, C^5H_3 AlaT^5), 1.59 (m, 1H, 1 βCH Lys^7), 1.48 (m, 2H, δCH_2 Lys^7), 1.42 (s, 3H, 1 CH_3 Aib^1), 1.40 (s, 6H, 1 CH_3 Aib^1 and 1 CH_3 Aib^6), 1.36 (s, 3H, 1 CH_3 Aib^3), 1.34 (s, 7H, 1 γCH Lys^7, 1 CH_3 Aib^3 and 1 CH_3 Aib^6), 1.28 (s, 7H, 1 γCH Lys^7 and 2 CH_3 Aib^4).

6.2.2 Solid phase synthesis

6.2.2.1 General methodology and remarks

General synthesis protocol

Nucleopeptide sequences are synthesized using the semiautomatic peptide synthesizer.[165] for the Boc/Bzl strategy; splitting, Fmoc group removal, ε-NH_2-functionalisation (where necessary) and cleavage are performed manually.

As a solid support MBHA·HCl resin (loading 0.62 mmol/g) is used.

Swelling: Before starting the synthesis, the resin is swollen by washing with DMF and DCM several times.

Deprotection: N^{α}-Boc protecting groups are removed by treating the resin twice with 2 mL 100% TFA (firstly for 1 min, then for 3 min). After Boc-deprotection the resins are washed with iPrOH, DMF and DCM.

Coupling: Activated N^{α}-Boc-protected amino acids in DMF solution are transferred into the reactor, then DIEA is added. Preactivation is avoided, since it is known to favor racemisation.[166]

Couplings are allowed to proceed for 20 min for proteogenic activated amino acids on N-terminal proteogenic amino acids, 30 min for proteogenic activated amino acid on N-terminal nucleoamino acids or 60 min for activated nucleoamino acids. Each coupling is repeated twice and subsequently qualitative Kaiser test[153] or TNBS test[154] are used to verify the completion of the reaction. When a negative response to the test is obtained, the coupling is repeated once more, generally allowing a longer coupling time.

After completing the peptide sequences, the resin beads are transferred to 20 mL syringes provided with sintered glass filters in order to operate the remaining steps manually.

Cleavage: A mixture of TFA/TMSOTf/4-MePhOH (80:25:10, v/v/w)[167] is prepared and 0.8 mL of this mixture are transferred to each syringe. Cleavage is allowed to proceed overnight at room temperature.

Recovery: The crude nucleopeptides are precipitated from cleavage solutions using cold Et_2O and after centrifugation (Eppendorf 5804 R centrifuge, 5 min at 4000 rpm) the supernatant is eliminated. The precipitate is resuspended with cold Et_2O, allowed to stand 1 hour in the cold and after centrifugation (5 min at 4000 rpm) the supernatant is eliminated again. After waiting 1 hour to allow the remaining Et_2O to evaporate, the crude

nucleopeptides are dissolved in 1 mL of H_2O milliQ, analyzed with HPLC (Varian, 1-100 % B in 20 min) and lyophilized.

Purification: The lyophilized crude nucleopeptides are dissolved in H_2O milliQ or in H_2O milliQ/MeCN (total volume 2.5-3.0 mL) and AcOH (0.1-0.2 mL) is added. 0.5-1.0 mL of the resulting solution are injected in the semipreparative HPLC (Beckmann) per run and 0.5-2.5 mL fractions are collected. Gradient: 1-41 % B in 30 min for FIT-nucleopeptides, 1-16 %, 1-21 % or 1-31 % B in 30 min for all other nucleopeptides. The purity of the collected fractions is checked by analytical HPLC (Varian) after 1:1 dilution with H_2O milliQ. Gradients: 0-40 %, 0-50% or 5-65 % B in 20 min. The pure fractions are screened by MALDI-MS spectrometry in order to confirm the presence of the desired product. Pure fractions are finally pooled, lyophilized, redissolved in 1.0 mL of a 1 mM solution of HCl in H_2O milliQ and relyophilized.

6.2.2.2 Synthesis of model nucleopeptides libraries

6BB library (6AA and 6TT peptides)

General sequence: H-(Ala-AlaB-Ala)$_2$-NH_2, where AlaB= $AlaA^H$, AlaT.

Synthesis scale: 10 μmol resin.

Activation protocol: 5 eq each of Boc-protected proteogenic amino acid (Boc-Ala-OH), BOP, HOBt and 15 eq DIEA; 3 eq each of N^{α}-Boc-protected nucleoamino acid (Boc-$AlaA^Z$-OH, Boc-AlaT-OH), BOP, HOBt, 10 eq DIEA.

6AA [H-(Ala-$AlaA^H$-Ala)$_2$-NH_2] Yield: 5 % (0.30 mg).
MS calcd $[M+H]^+$: 692.31, found: 692.2 [MALDI-MS].
HPLC (Varian, 0-40 % B in 20 min): t_r 8.10

6TT [H-(Ala-AlaT-Ala)$_2$-NH_2] Yield: 5 % (0.35 mg).
MS calcd $[M+H]^+$: 710.34 found: 710.5 [MALDI-MS].
HPLC (Varian, 0-40 % B in 20 min): t_r 8.41 min.

B2f library (T2A, T2U, A2A and A2U peptides)

General sequence: H-Ala-AlaB-Ala-Xxx-AlaB-Ala-Lys(H)-NH_2, where AlaB= $AlaA^H$, AlaT, Xxx= Ala, Aib.

Synthesis scale: 25 μmol resin (for T2A, A2A, A2U); 15 μmol resin in each of two reactors used for a parallel resynthesis of T2U following two different activation protocols (protocol x and y, respectively).

Activation protocol: 5 eq each of Boc-protected proteogenic amino acid (Boc-Ala-OH, Boc-Lys(2Cl-Z)-OH), BOP, HOBt and 15 eq DIEA; 5 eq each of Boc-protected C^{α}-tetrasubstituted amino acid (Boc-Aib-OH), HATU, HOAt and 15 eq DIEA; 3 eq each of N^{α}-Boc-protected nucleoamino acid (Boc-AlaA^{Z}-OH, Boc-AlaT-OH), HATU, HOAt, 10 eq DIEA (T2A, A2U, A2A, T2U reactor x) or 3 eq each of N^{α}-Boc-protected nucleoamino acid (Boc-AlaT-OH), BOP, HOBt (T2U reactor y) and 10 eq DIEA.

Special coupling procedures: Triple coupling are used for Aib introduction. Coupling times of 40 min are used for activated proteogenic amino acids on N-terminal nucleoamino acids, coupling times of 60 min are used for Aib introduction and coupling times of 60 min are used for activated proteogenic amino acids on N-terminal Aib.

A2A [H-(Ala-AlaA^{H}-Ala)$_2$-Lys(H)-NH_2]

MS calcd $[M+H]^+$: 838.43, found: 838.5 [Finnigan HPLC-MS, 1-31 % B in 10 min].

HPLC (Varian, 0-100 % B in 20 min): t_r 6.27 min.

It is not possible to obtain sufficiently pure fractions. Crude yield: 88 % (7.4 mg)

A2U [H-Ala-AlaA^{H}-Ala-Aib-AlaA^{H}-Ala-Lys(H)-NH_2] Yield: 2 % (0.4 mg).

MS calcd $[M+H]^+$: 852.45, found: 852.5 [Finnigan HPLC-MS, 1-31 % B in 10 min].

HPLC (Varian, 0-40 % B in 20 min): t_r 8.80 min.

T2A [H-(Ala-AlaT-Ala)$_2$-Lys(H)-NH_2] Yield: 29 % (5.9 mg).

MS calcd $[M+H]^+$: 820.41, found: 820.5 [Finnigan HPLC-MS, 1-31 % B in 10 min].

HPLC (Varian, 0-50 % B in 20 min): t_r 8.41 min.

NMR (3 mM, d_6-DMSO, 500 MHz), δ/ppm: 11.22, 11.20 (2 br s, 2H, N(3)H AlaT^2 and AlaT^5), 8.80-8.79 (d, 1H, αNH AlaT^2), 8.30-8.28 (d, 1H, NH Ala^3), 8.25-8.24 (d, 1H, NH Ala^3), 8.19 (br s, 3H, NH_3^+ Ala^1), 8.14-8.13 (d, 1H, NH Ala^6), 8.06-8.04 (d, 1H ,αNH AlaT^5), 7.95-7.93 (d, 1H, αNH Lys^7), 7.87 (br s, 3H, NH_3^+ Lys^7), 7.44 (s, 1H, C(6)H thymine AlaT^2), 7.40 (s, 1H, C(6)H thymine AlaT^5), 7.31 (br s, 1H, 1 CONH), 7.01 (br s, 1H, 1 CONH), 4.66-4.63 (m, 1H, αCH AlaT^2), 4.61-4.57 (m, 1H, αCH AlaT^5), 4.27-4.24 (m, 1H, αCH Ala^3), 4.24-4.21 (m, 1H, αCH Ala^6), 4.17-4.14 (m, 1H, αCH Ala^4), 4.13-

4.11 (m, 1H, αCH Lys^7), 4.09-4.06 (m, 1H, 1 βCH AlaT^5), 4.04-4.03 (m, 1H, 1 βCH AlaT^2), 3.84-3.79 (m, 2H, 1 αCH Ala^1 and 1 βCH AlaT^2), 3.74-3.69 (m, 1H, 1 βCH AlaT^5), 2.77-2.74 (m, 2H, εCH$_2$ Lys^7), 1.72 (m, 3H, C^{5}H$_3$ thymine AlaT^2), 1.71 (m, 3H, C^{5}H$_3$ thymine AlaT^5), 1.68-1.64 (m, 2H, βCH$_2$ Lys^7), 1.55-1.52 (m, 2H, δCH$_2$ Lys^7), 1.34-1.33 (d, 3H, CH$_3$ Ala^1), 1.31-1.25 (m, 2H, γCH$_2$ Lys^7),1.23-1.21 (2 d, 6H, CH$_3$ Ala^3 and Ala^6), 1.19-1.18 (d, 3H, CH$_3$ Ala^4).

T2U [H-Ala-AlaT-Ala-Aib-AlaT-Ala-Lys(H)-NH$_2$]
Yield: 13 % (1.7 mg), method x; yield: 26 % (3.4 mg), method y.
MS calcd [M+H]$^+$: 834.42, found: 834.5 [Finnigan HPLC-MS, 1-31 % B in 10 min].
HPLC (Varian, 0-50 % B in 20 min): t$_r$ 9.10 min.
NMR (5 mM, *d*$_6$-DMSO, 500 MHz), δ/ppm: 11.24, 11.23 (2 br s, 2H, N(3)H AlaT^2 and AlaT^5), 8.86-8.84 (d, 1H, αNH AlaT^2), 8.43-8.42 (d, 1H, NH Ala^3), 8.36 (br s, 1H, NH Aib^3), 8.21 (br s, 3H, NH$_3^+$ Ala^1), 7.92-7.91 (d, 1H, αNH AlaT^5), 7.86-7.84 (d, 1H, αNH Lys^7), 7.84-7.82 (br s, 3H, NH$_3^+$ Lys^7), 7.74-7.72 (d, 1H, NH Ala^6), 7.50 (d, 1H, C(6)H thymine AlaT^2), 7.36 (s, 1H, C(6)H thymine AlaT^5), 7.22 (br s, 1H, 1 CONH), 7.02 (br s, 1H, 1 CONH), 4.66-4.62 (m, 1H, αCH AlaT^2), 4.49-4.44 (m, 1H, αCH AlaT^5), 4.28-4.24 (m, 1H, 1 βCH AlaT^5), 4.22-4.21 (m, 1H, αCH Ala^3), 4.21-4.19 (m, 1H, αCH Ala^6), 4.15-4.11 (m, 1H, αCH Lys^7), 4.02-3.98 (m, 1H, 1 βCH AlaT^2), 3.92-3.87 (m, 1H, 1 βCH AlaT^2), 3.86-3.80 (m, 2H, αCH Ala^1 and 1 βCH AlaT^5), 2.78-2.72 (m, 2H, εCH$_2$ Lys^7), 1.73 (d, 3H, C^{5}H$_3$ thymine AlaT^2), 1.71 (d, 3H, C^{5}H$_3$ thymine AlaT^5), 1.68-1.65 (m, 2H, βCH$_2$ Lys^7), 1.58-1.52 (m, 2H, δCH$_2$ Lys^7), 1.36-1.34 (m, 2H, γCH$_2$ Lys^7), 1.33-1.32 (d, 3H, CH$_3$ Ala^1), 1.29 (s, 6H, 2 CH$_3$ Aib), 1.27-1.25 (d, 3H, CH$_3$ Ala^6), 1.25-1.24 (d, 3H, CH$_3$ Ala^3).

6.2.2.3 Synthesis of functionalised nucleopeptide libraries

In order to allow selective derivatisation of N-terminal Lys-ε-NH$_2$ (B4f and CTCCF libraries, T5Bt peptide), C-terminal Lys are introduced as Boc-Lys(2Cl-Z)-OH, while N-terminal Lys are introduced as Boc-Lys(Fmoc)-OH. To allow the side-chain functionalisation of the nucleopeptides with different moieties (B4f and CTCCf libraries), after completion of the peptide sequence resin beads are split in order to perform the different treatments required.

In case dimerisation through disulfide bridge formation is planned (T5SH peptide), a Cys residue is introduced as Fmoc-Cys(Trt)-OH at the N-terminal position of the sequence. For this peptide the N-terminal lysine is introduced as Boc-Lys(2Cl-Z)-OH, as ε-derivatisation is not required. The Fmoc group is removed manually prior to cleavage procedure, which causes the contextual Trt group removal. Unwanted oxidation of the thiol function during purification is avoided by operating always under acidic conditions.

Additional steps

Resin splitting: The resin beads are split by suspending them in a DMF/DCM mixture and transferring parts of the suspension into 2 mL syringes provided with sintered glass filters. The volumes transferred into different syringes are proportional to the fraction of the peptide to be derivatised with the specific treatment planned for each syringe.

Fmoc deprotection: Fmoc group on Lys-ε-NH_2 or on Cys-α-NH_2 is removed by treating the resin with 2 mL of a 20% (v/v) solution of piperidine in DCM for 30 min, then washed four times with DCM. The procedure is repeated twice.

Derivatisation with FITC**:** 5 eq FITC are dissolved in 1 mL DMF and transferred into the syringe, then 15 eq DIEA are added. The coupling is allowed to proceed for 60 min. Subsequently, qualitative Kaiser test is used to verify the completion of the reaction. When a negative response to the test is obtained, the coupling is repeated a second time.

Biotinylation**:** 5 eq each of biotin, HBTU, HOBt are dissolved in 1 mL of a 1:1 DMF/DMSO mixture and transferred into the syringe, then 15 eq DIEA are added. The coupling is allowed to proceed for 30 min and it is repeated twice. Subsequently, qualitative Kaiser test is used to verify the completion of the reaction. When a negative response to the test is obtained, the coupling is repeated a third time.

B4f library (A4H, A4Bt, A4FIT; C4H, C4Bt, C4FIT; G4H, G4Bt, G4FIT; T4H, T4Bt, T4FIT peptides)

General sequence: H-Lys(f)-(Ala-AlaB-Ala)$_4$-Lys(H)-NH_2, where f= H, FIT, Bt and AlaB= $AlaA^H$, $AlaC^H$, $AlaG^{OH}$, AlaT.

Synthesis scale: 30 μmol resin per peptide sequence, 10 μmol resin for each derivative after splitting.

Activation protocol: 5 eq each of Boc-protected proteogenic amino acid (Boc-Ala-OH, Boc-Lys(2Cl-Z)-OH, Boc-Lys(Fmoc)-OH), HBTU, HOBt and 15 eq DIEA; 3 eq each of

N^{α}-Boc-protected nucleoamino acid (Boc-AlaA^{Z}-OH, Boc-AlaC^{Z}-OH, Boc-AlaG^{OH}-OH, Boc-AlaT-OH), HBTU, HOBt and 10 eq DIEA.

A4H [H-Lys(H)-(Ala-AlaA^{H}-Ala)$_4$-Lys(H)-NH$_2$] Yield: 5 % (0.9 mg).
MS calcd $[M+H]^+$: 1658.82, found: 1658.8 [MALDI MS].
HPLC (Varian, 0-40 % B in 20 min): t_r 9.10 min.

A4Bt [H-Lys(Bt)-(Ala-AlaA^{H}-Ala)$_4$-Lys(H)-NH$_2$] Yield: 9 % (1.7 mg).
MS calcd $[M]^+$: 1883.88, found: 1884.1 [MALDI MS].
HPLC (Varian, 5-65 % B in 20 min): t_r 13.48 min.

A4FIT [H-Lys(FIT)-(Ala-AlaA^{H}-Ala)$_4$-Lys(H)-NH$_2$] Yield: 6 % (1.2 mg).
MS calcd $[M+H]^+$: 2047.86, found: 2048.3 [MALDI MS].
HPLC (Varian, 5-65 % B in 20 min): t_r 10.19 min.

C4H [H-Lys(H)-(Ala-AlaC^{H}-Ala)$_4$-Lys(H)-NH$_2$] Yield: 3 % (0.4 mg).
MS calcd $[M]^+$: 1561.77, found: 1561.4 [MALDI MS].
HPLC (Varian, 5-65 % B in 20 min): t_r 5.63 min.

C4Bt [H-Lys(Bt)-(Ala-AlaC^{H}-Ala)$_4$-Lys(H)-NH$_2$]: Yield: 7 % (1.3 mg).
MS calcd $[M+H]^+$: 1788.86, found: 1789.5 [MALDI MS].
HPLC (Varian, 5-65 % B in 20 min): t_r 6.83 min.

C4FIT [H-Lys(FIT)-(Ala-AlaC^{H}-Ala)$_4$-Lys(H)-NH$_2$] Yield: 11 % (2.1 mg).
MS calcd $[M+H]^+$: 1951.82, found: 1951.9 [MALDI MS].
HPLC (Varian, 5-65 % B in 20 min): t_r 9.90 min.

G4H [H-Lys(H)-(Ala-AlaG^{OH}-Ala)$_4$-Lys(H)-NH$_2$] Yield: 10 % (1.8 mg).
MS calcd $[M+H]^+$: 1722.80, found: 1723.2 [MALDI MS].
HPLC (Varian, 5-65 % B in 20 min): t_r 6.51 min.

G4Bt [H-Lys(Bt)-(Ala-AlaG^{OH}-Ala)$_4$-Lys(H)-NH$_2$]
After the first lyophilization the crude product can not be redissolved in any solvent mixture. Crude yield: 94 % (18 mg).

HPLC (Varian, 0-50 % B in 20 min): crude, t_r 10.63 min, 53 %.

G4FIT [H-Lys(FIT)-(Ala-AlaG^{OH}-Ala)$_4$-Lys(H)-NH_2] Yield: 9 % (2.0 mg).
MS calcd $[M+H]^+$: 2111.83, found: 2111.6 [MALDI MS].
HPLC (Varian, 5-65 % B in 20 min): t_r 10.24 min.

T4H [H-Lys(H)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2]: Yield: 8 % (1.3 mg).
MS calcd $[M+H]^+$: 1622.78, found: 1622.8 [MALDI MS].
HPLC (Varian, 5-65 % B in 20 min): t_r 7.36 min.

T4Bt [H-Lys(Bt)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2]: Yield: 8 % (1.5 mg).
MS calcd $[M+H]^+$: 1848.86, found: 1848.8 [MALDI MS].
HPLC (Varian, 5-65 % B in 20 min): t_r 8.78 min.

T4FIT [H-Lys(FIT)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2]: Yield: 7 % (1.4 mg).
MS calcd $[M+H]^+$: 2011.81, found: 2011.7 [MALDI MS].
HPLC (Varian, 5-65 % B in 20 min): t_r 11.20 min.

CTCCf library (CTCCH, CTCCBt, CTCCFIT peptides)

General sequence: H-Lys(f)-Ala-AlaC^H-Ala-Ala-AlaT-Ala-(Ala-AlaC^H-Ala)$_2$-Lys(H)-NH_2, where f= H, FIT, Bt.

Synthesis scale: 30 μmol resin per peptide sequence, 10 μmol resin for each derivative after splitting.

Activation protocol: 5 eq each of Boc-protected proteogenic amino acid (Boc-Ala-OH, Boc-Lys(2Cl-Z)-OH, Boc-Lys(Fmoc)-OH), HBTU, HOBt and 15 eq DIEA; 3 eq each of N^{α}-Boc-protected nucleoamino acid (Boc-AlaC^Z-OH, Boc-AlaT-OH), HBTU, HOBt and 10 eq DIEA.

CTCCH [H-Lys(H)-Ala-AlaC-Ala-Ala-AlaT-Ala-(Ala-AlaC-Ala)$_2$-Lys(H)-NH_2]
Yield: 8 % (1.2 mg). MS calcd $[M+H]^+$: 1577.77, found: 1577.9 [MALDI MS].
HPLC (Varian, 5-65 % B in 20 min): t_r 6.09 min.

CTCCBt [H-Lys(Bt)-Ala-AlaC-Ala-Ala-AlaT-Ala-(Ala-AlaC-Ala)$_2$-Lys(H)-NH_2]
Yield: 4 % (0.8 mg). MS calcd $[M+H]^+$: 1803.82, found: 1803.7 [MALDI MS].

HPLC (Varian, 5-65 % B in 20 min): t_r 7.63 min.

CTCCFIT [H-Lys(FIT)-Ala-AlaC-Ala-Ala-AlaT-Ala-(Ala-AlaC-Ala)$_2$-Lys(H)-NH_2]
Yield: 8 % (1.6 mg). MS calcd $[M+H]^+$: 1966.82, found: 1966.7 [MALDI MS].
HPLC (Varian, 5-65 % B in 20 min): t_r 10.43 min.

T5f library (T5H, T5Bt, T5SH peptides)
General sequence: X-Lys(f)-[Ala-AlaT-Ala]$_5$-Lys(H)-NH_2 where X=H, H-Cys(H), f= H, Bt.
Synthesis scale: 22 μmol resin for T5H and T5SH, 10 μmol resin for T5Bt.
Activation protocol: 5 eq each of Boc-protected proteogenic amino acid (Boc-Ala-OH, Boc-Lys(2Cl-Z)-OH, Boc-Lys(Fmoc)-OH, Fmoc-Cys(Trt)-OH), BOP, HOBt and 15 eq DIEA; 3 eq each of N^{α}-Boc-protected nucleoamino acid (Boc-AlaT-OH), HATU, HOAt and 10 eq DIEA.

T5H [H-Lys(H)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2] Yield: 5 % (2.3 mg).
MS calcd $[M+2H]^{2+}$: 980.46, found: 980.8 [Finnigan HPLC-MS, 0-100 % B in 10 min].
HPLC (Varian, 0-50 % B in 20 min): t_r 11.30 min.

T5Bt [H-Lys(Bt)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2] Yield: 3 % (0.7 mg).
MS calcd $[M+H]^+$: 2185.99, found: 2187.1 [MALDI MS].
HPLC (Varian, 0-50 % B in 20 min): t_r 10.21 min.

T5SH [H-Cys(H)-Lys(H)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2] Yield: 3 % (1.5 mg).
MS calcd $[M+2H]^{2+}$: 1031.96, found: 1032.3 [Finnigan HPLC-MS, 0-100 % B in 10 min].
HPLC (Varian, 0-50 % B in 20 min): t_r 12.07 min.

6.2.2.4 Cysteine-mediated nucleopeptide dimerisation trials

(T5S)$_2$ [H-Cys(┐)-Lys(H)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2
H-Cys(┘)-Lys(H)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2]
T5SH (0.5 mg, 0.25 μmol) is dissolved in 1.5 mL H_2O/DMSO 9:1 added with TEA to set the pH at 9 and stirred in an open reaction vessel in order to allow atmospheric oxidation. The reaction is monitored by HPLC (Varian, 0-50 % B in 20 min). After reaching a critical

concentration, the compound precipitates, while the unreacted starting material stays in solution, therefore HPLC conversion measurements give decreasing values with time. It is not possible to redissolve the dimer in any solvent mixture in order to purify and fully characterize it

HPLC (Varian, 0-50% B in 20 min): reaction mixture after 2 hours (maximum apparent conversion), t_r 12.07 min, 54.8 % [T5SH], t_r 12.59 min, 43.8% [$(T5S)_2$].

T5S-BMB-S5T [H-Cys(┐)-Lys(H)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2
BMB
H-Cys(┘)-Lys(H)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2]

T5SH (0.4 mg, 0.21 μmol) is dissolved in 1.0 mL milliQ H_2O and the pH is set to 7 by addition of a 0.01 M solution of TEA in H_2O (10 μL, 1 μmol). A solution of BMB (2.5 mg, 0.01 mmol) in 1.0 mL DMSO is prepared (sonication and heating are necessary) and 10 μL of this solution (0.5 equivalents) are added to the mixture, which is stirred in a closed vessel in order to avoid atmospheric oxidation of the thiol functions. After 2 hours the HPLC analysis (Agilent, 0-50 % B' in 20 min) doesn't show significant changes in the product peak. After 1 hour 10 μL of the BMB solution (0.5 equivalents) are added. After 8 hours the product peak starts to decrease and a new peak ('B') is observed. 40 μL of BMB solution are added (2.0 equivalents) and the mixture is stirred overnight. The day after the new peak has grown, but the increase doesn't continue during the day, and a new peak is observed. 100 μL BMB (5 equivalents) are added and the mixture is stirred overnight. The day after the new peak is prevalent and the second has grown. The two peaks are separed semipreparative HPLC (Agilent, 1-10 % B' in 40 min, peak 'A' t_r 15.8-17.6 min, peak 'B' t_r 25.3-26.5 min) and analyzed by HPLC and MS (Mariner). The HPLC analysis of the content of the second peak shows two well separed peaks with a different retention time from the original, pointing to substantial degradation. Both seem to be low-molecular weight impurities.

Peak 'A'

HPLC (Agilent, 0-40 % B in 20 min): t_r 10.5 min.

MS: observed m/z: 278.58, 518.23, 540.15.

Peak 'B'

HPLC (Agilent, 0-40 % B in 20 min), before purification: t_r 13.1 min.

HPLC (Agilent, 0-40 % B in 20 min), after purification: t_r 5.02 min, t_r 19.7 min.

MS: observed m/z: 185.1, 201.12.

6.3 Pairing and biological tests

6.3.1 Surface plasmon resonance measurements

Experimental conditions

Biosensor assays are performed with HEPES-buffered saline (HBS-EP) as running buffer (10 mM HEPES, 150 mM NaCl, 3 mM $Mg(AcO)_2$, 3 mM EDTA, 0.005% surfactant P20, pH 7.4).

Two series of experiments are performed, requiring three sensor chips:

- For the first series of experiments, A4Bt, T4Bt and C4Bt are immobilised on the sensor chip surface and solutions of G4H, T4H, U_4, U_8, dT_8 and dT_{10} are injected;
- For the second series of experiments, T5Bt, Bt-dA_{10} and biotin (as a control) are immobilised on a first sensor chip surface and solutions of A4H, T4H, T5H, dT_{10} or dA_{10} are injected; similarly, A4Bt, T4Bt and biotin (as a control) are immobilised on a second sensor chip and solutions of A4H, T5H, dT_{10} or dA_{10} are injected.

All nucleopeptide and oligonucleotide solutions are prepared in running buffer.

Immobilisation on streptavidin coated chips

For each experiment a single four channel sensor chip CM5 is used.

Immobilisation of streptavidin is performed by injecting onto the surface of the sensor chip, activated with EDC/HOSu, 35 µL of streptavidin (100 µg/mL in acetate buffer, pH 4.9), which gives a signal of approximately 4000-5000 RU, followed by 20 µL of ethanolamine hydrochloride, pH 8.5, to saturate the free activated sites of the matrix. Biotinyled ligands (1 µM in HEPES buffer) are allowed to interact with streptavidin until a response of approximately 600-700 RU is obtained (60 RU for biotin).

Kinetic experiments

Binding experiments are carried out at 25°C with a constant flow rate of 20 µL/min. Nucleopeptides and DNA oligonucleotides are dissolved in running buffer to form stock solutions (about 1 mM), from which samples at different concentrations (from 12.5 to 200 µM for T4H, T5H, dA_{10}, dT_{10}, from 25 to 200 µM for A4H, dT_8) are prepared by

appropriate dilutions with running buffer and injected. RNA oligonucleotides are dissolved in running buffer to form 100 μM solutions which are injected.
The association phase is allowed to run for 3 min, followed by a dissociation phase of 3 min. After each run, the channels are completely regenerated with 10 μM HCl.

Data analysis
The kinetic parameters are calculated using the BIAeval 3.1 software on a personal computer. The analysis is performed using the simple Langmuir binding model with/without separate k_a/k_d. The specific binding profiles are obtained after subtracting the response signal from the control channel and from blank-buffer injection. The fitting to each model is judged by the chi square value and randomness of residue distribution compared to the theoretical model.

6.3.2 $DiOC_6$ test

Apoptotic effects of the flexible nucleopeptides are evaluated by the 3,3'-dihexyloxacarbocyanine iodide ($DiOC_6$) uptake test.[143] Jurkat lymphoid cells are incubated in 10% FBS/RPMI for 24 hours at 37°C in 5 % CO_2 in the presence of the nucleopeptide CTCCH (10 μM or 50 μM). The same cells incubated in 10% FBS/RPMI for 24 hours at 37°C in 5 % CO_2 with or without doxorubicin are used as a positive and negative control, respectively.
To evaluate apoptosis induction approximately $1 \cdot 10^6$ cells washed in PBS are resuspended in 300 μL of PBS containing 40 nM $DiOC_6$ and incubated at 37°C in 5 % CO_2 for 30 min. Cells are then washed 3 times with PBS and analyzed with a FACSCalibur® flow-cytometer (at least 10'000 events for each experiment are acquired) by flow cytometry and the apoptotic fraction is compared with the apoptotic fraction found in negative control cells.

6.3.3 MTS cytotoxicity tests

Lymphoid Raji and Jurkat and tumor BL41 cells are used for nucleopeptide cytotoxicity evaluation by the 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphen-yl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) test.[145]

Cells are seeded at a density of 50.000 cells per well in 96-well plates, incubated in 10% FBS/RPMI and treated for 24 hours at 37°C in 5% CO_2 with flexible (CTCCH, T4H, T4Bt) or rigid (2 HCl·H-(Aib-AlaT-Aib)$_2$-Lys(H)-NH_2) nucleopeptides at concentrations up to 100 μM. Doxorubicin at the same concentrations is used as a positive control. After the treatment, the cells are incubated for 2 hours at 37°C with 20 μL of MTS reagent solution. Absorbance is measured at 490 nm wavelength using a microplate reader.
Results are the average of three independent experiments.

6.3.4 Cell penetration experiments

Raji cells are used for flow-cyotometric nucleopeptide uptake tests, while murine renal cancer Renca cells are used for the microscopy uptake analysis.

6.3.4.1 Flow cytometry uptake analysis

For preliminary uptake assessment, cells are seeded at a density of $1 \cdot 10^5$ cells per well in 96-well plates, incubated in 10% FBS/RPMI at 37°C in 5 % CO_2 and treated with T4FIT (50 μM) for 16 hours. Following incubation, cells are washed 3 times with PBS and analyzed with a FACSCalibur® flow-cytometer (at least 10'000 events for each experiment are acquired). Raji cells incubated in the same conditions in the absence of nucleopeptide are used as a control.
To evaluate uptake dependence on time, concentration and energy, cells are seeded at a density of $1 \cdot 10^5$ cells per well in 96-well plates, incubated in 10% FBS/RPMI at 37°C in 5 % CO_2 and treated with T4FIT (5 or 50 μM) for 5, 30 and 60 min in the presence or in the absence of 2-deoxy-D-glucose (50 mM) and NaN_3 (10 mM). Following incubation, cells are washed 3 times with PBS and analyzed with a FACSCalibur® flow-cytometer (at least 10'000 events for each experiment are acquired).

6.3.4.2 Epifluorescence and confocal microscopy uptake analysis

Cells are seeded onto glass coverslips at 20.000 cells per well in 24-well plates, grown overnight in 10% FBS/RPMI medium at 37°C in 5 % CO_2, then treated for 12 hours with increasing concentration (0.5, 5.0 and 50 μM) of T4Bt or T4FIT at 37°C in 5% CO_2. The

coverslips are then washed 3 times with cold PBS and fixed with 4% *para*formaldehyde for 10 min at 4°C.

In the case of T4Bt, after fixation, the cells are washed 3 times with cold PBS, and then permeabilized at room temperature with PBS containing 0.1% Triton X-100 for 10 min. Following another ish, the cells are incubated with streptavidin-FITC for 1 hour at room temperature.

The cells are stained with 0.5 μg/mL DAPI solution for 10 min at room temperature to identify nuclei. In some experiments the cell are also stained with the lipophilic dye FM4-64 (25 μM) for 10 min in order to label membranes. The coverslips are then transferred to slides and cells are visualized by confocal microscopy or epifluorescence microscopy.

In order to perform co-localization experiments, cells are seeded onto glass coverslips at 20.000 cells per well in 24-well plates, grown overnight in 10% FBS/RPMI medium at 37°C in 5 % CO_2, then co-treated with transferrin-Alexa-546 at 20 μg/mL and T4FIT (50 μM) for 30 min at 37°C in 5% CO_2. The coverslips are then washed 3 times with cold PBS and fixed with 4% *para*formaldehyde for 10 min at 4°C. The cells are also stained with 0.5 μg/mL DAPI solution for 10 min at room temperature to identify nuclei and their cell membranes labeled with the lipophilic dye FM4-64 before visualization by confocal microscopy.

6.3.5 Stability tests

Flexible nucleo-heptapeptides T2A and T2U (about 2.5 mg) and the rigid nucleo-heptapeptide 6TTK (1.0 mg) are dissolved in PBS buffer at 10 mg/mL concentration, an equal volume of freshly prepared mouse serum is added and time count starts. The solutions are kept in closed vials at 37°C in a thermostatic bath and sampled immediately after serum addition (as a control) and 1, 2, 4, 8, 24 and 48 hours thereafter.

For flexible nucleopeptide solutions, 50 μL solution are transferred into an Eppendorf vial, 5 μL TFA are added in order to quench and precipitate serum enzymes, the suspension is diluted with 200 μL PBS, then centrifugated (Sigma centrifuge, 10 min at 5000 rpm) and the supernatant is diluted with 300 μL H_2O milliQ. In the case of the rigid nucleopeptide solution, 20 μL solution are used, 2 μL TFA and 80 μL PBS are added, then after centrifugation the supernatant is diluted with 120 μL H_2O milliQ.

The prepared samples are analyzed by HPLC (Varian, 0-50 % B in 20 min).

7. CONCLUSIONS

This project aimed at studying the properties of sequential helical nucleopeptides containing alanyl nucleoamino acid residues (AlaB) at *i,i+3* positions in view of applications as nucleic acid modulators. The working hypothesis to be tested was that, if such nucleopeptides assume a 3_{10}-helical conformation, side chain nucleobases are aligned along the helical axis and this might favour cooperative interactions with complementary functionalised strands.

To better evaluate the importance of structuration on the nucleopeptide pairing properties, both rigid (based on the 3_{10}-helix promoting C^{α}-tetrasubstituted residue Aib) and flexible (based on the less structuring proteogenic residue Ala) nucleopeptides have been synthesised.

Protected rigid tripeptides containing each of the four DNA nucleobases have been synthesised and characterized. A protected hexapeptide and a completely deprotected, water-soluble heptapeptide, both containing two thymines, have been synthesised and studied. The solid state structure of a protected thymine-based tripeptide has been studied by X-ray diffraction. The tripeptide is folded in a β-turn and shows a backbone-to-side-chain intraresidue H-bond freezing the nucleobase orientation. The unit cell is a dimer stabilized by intermolecular thymine involving H-bonds.

The conformational analysis by IR spectroscopy and NMR spectrometry in $CDCl_3$ solution has confirmed that protected nucleo-tripeptides are folded and an intraresidue backbone-to-side-chain H-bond, stabilizing the αNHs of adenine-, cytosine- and thymine-based nucleoamino acid residues, has been detected in such peptides. It has been found that nucleopeptides with free nucleobases have a strong tendency to intermolecular interactions; in case of the thymine-based nucleopeptides, this favours the formation of dimers, as observed in the crystal state. In particular the thymine containing nucleo-hexapeptide forms a well defined and very stable dimeric structure in $CDCl_3$, which has been studied by 2D-NMR.

The completely deprotected lysine containing thymine-based nucleo-heptapeptide has been studied by 2D NMR in DMSO solution, providing very strong evidence in support of the adoption of a 3_{10}-helical structure. CD data suggesting the adoption of an helical structure in aqueous solution H_2O have also been obtained.

Flexible model nucleopeptide libraries have designed and synthesised by solid-phase parallel synthesis, for synthetic protocol optimization and conformational analysis. Library members are made based on repeating Ala-AlaB-Ala tripeptide units containing the same base (adenine or thymine), in some cases with one lysine at the C-terminus to enhance their water solubility. A thymine-based nucleo-heptapeptide containing two tripeptide units and an analogue, carrying an Aib residue in the middle of its sequence, have been conformationally characterized by CD in aqueous solution and by 2D NMR in DMSO solution. In aqueous solution data suggesting some degree of folding have been obtained for the Aib containing nucleopeptide, while in DMSO both adopt a folded conformation.

Functionalised nucleopeptide libraries, for pairing and biological tests, have been designed and synthesised by solid-phase parallel synthesis. Library members are made of repeating Ala-AlaB-Ala tripeptide units containing the same base, with two lysines at the termini in order to enhance their water solubility and to allow for selective partial derivatisation of the N-terminal lysine ε-amino group with biotin or with a fluorophore. Functionalised nucleopeptides containing the four DNA nucleobases with four tripeptide units and thymine-based nucleopeptides with five tripeptide units have been prepared.

Flexible functionalised nucleopeptide pairing properties have been studied by surface plasmon resonance techniques through immobilization of biotinylated nucleopeptides or nucleotides on a functionalised gold chip. Selective adenine/thymine recognition has been detected both in the case of nucleopeptide/nucleopeptide interactions and in the case of nucleopeptide/nucleotide interactions. Pairing is significantly stronger for the thymine-based nucleopeptide with five tripeptide units than for the nucleopeptide with four tripeptide units, thus suggesting cooperative binding.

Aiming at future applications as novel biomedical tools, cytotoxicity tests have been performed on a rigid and on variously functionalised flexible thymine-based nucleopeptides. No decrease in the viability of different tumor and lymphoid cell lines has been detected after cells were treated with the nucleopeptides at concentrations up to 100 μM for 24 hours.

Cell penetration experiments, carried on fluorescent and biotinylated thymine-based nucleopeptides at concentrations 0.5-50 μM, with 12 hours incubation time, have shown that the nucleopeptides enter the cells and are found both in the cytosol and in the nucleus, generally in vesicular compartments. A deeper investigation has demonstrated that cell

uptake is time, concentration and energy dependent, since it is significantly reduced when cells are energy-starved by treatment with NaN_3 and 2-deoxy-D-glucose. Significant colocalization with transferrin, a typical clathrin mediated endocytosis marker, has been observed, thus suggesting that nucleopeptide uptake is due mainly, but likely not only, to endocytosis.

Stability tests in mouse serum on the thymine-based nucleo-heptapeptides have shown that the Ala-based peptide is rapidly degraded, the peptide containing one Aib residue in its sequence is much more resistant and the Aib-based analogue is stable even after 48 hours.

Altogether, this work shows that sequential nucleopeptides adopt a stable conformation that favours intermolecular interactions among the nucleobases. In particular the rigid nucleopeptides, rich in Aib residues, tend to fold into 3_{10}-helical structures and display interesting biological properties, making them promising for the development of analogues with applications as nucleic acid modulators.

8. REFERENCES

1. (a) Mandelkern, M., Elias, J., Eden, D., and Crothers, D., *J. Mol. Biol.*, **152**, 153 (1981); (b) Ghosh, A. and Bansal, M., *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 620 (2003).
2. Sponer, J., Leszczinski, J., and Hobza, P., *Biopolymers*, **61**, 3 (2001).
3. (a) Lee, J.C. and Gutell, R.R., *J. Mol. Biol.*, **344**, 1225 (2004); (b) Crick, F., *J. Mol. Biol.*, **19**, 548, (1966); (c) Varani, G. and McClain, W., *EMBO Rep.*, **1**, 18 (2000).
4. Fox, S.W., *Mol. Cell. Biochem.*, **15**, 129 (1974).
5. (a) Zubay, G., *Nature*, **182**, 112 (1958); (b) Knight, R.D., Freeland, S.J., and Landweber, L.F., *Trends Biochem. Sci.*, **24**, 241 (1999).
6. (a) Jukes, T.H., *Adv. Biol. Med. Phys.*, **9**, 1 (1963); (b) Schreiber, G., *Angew. Chem. Int. Ed. Engl.*, **10**, 638 (1971).
7. Pagon, R.A., Bird, T.C., Dolan, C.R., Smith, R.H. and Stephens, K. (Eds.), *GeneReviews,* University of Washington, Seattle (WA, USA), 1993-2008.
8. (a) Mizuno, T., Chou, M.-Y., and Inouye, M., *Proc. Natl Acad. Sci. USA*, **81**, 1966 (1984); (b) Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C., *Nature*, **391**, 806 (1998).
9. Uhlmann, E. and Peyman, A., *Chem. Rev.*, **90**, 543 (1991).
10. (a) Halford, M.H., and Jones, A.S., *Nature*, **217**, 638 (1968); (b) Miller, P.S., Cushman, C.D., Levis, J.T., and Eckstein, F. (Eds.), *Oligonucleotides and analogues. A practical approach.* JRC Press (European Union), 1991.
11. (a) Detrick, B., Nagineni, C.N., Grillone, L.R., Anderson, K.P., Henry S.P., and Hooks, J.J., *Invest. Ophthalmol. Vis. Sci.*, **42**, 163 (2001); (b) Takemoto, H. and Inahi, Y., *Adv. in Polymer Science*, **41,** 123 (1981); (c) Ravikumar, V.T., Lima, W.F., Van Sooy, K., and Turney, B., *Nucleosides Nucleotides Nucl. Ac.*, **23**, 149 (2004); (d) Ho, P.T. and, Parkinson, D.R., *Semin. Oncol.*, **24**, 187 (1997).
12. (a) Waters, J.S., Webb, A., Cunningham, D, Clarke, P.A., Raynaud, F., di Stefano, F., and Cotter, F.E., *J. Clin. Oncol.*, **18**, 1812 (2000); (b) Chi, K.N., Gleave, M.E., Klasa, R., Murray, J., Bryce, C., Lopes de Menezes, D.E., D'Aloisio, S., and Tolcher, A.W., *Clin. Cancer Res.*, **7**, 3920 (2001).
13. Summerton, J. And Weller, D., *Antisense Nucleic Acid Drug Dev.*, **7**, 187 (1995); (b) Amantano, A. and Iversen, P.L. *Curr. Opin. Pharm.*, **5**, 550 (2005); (c) Moulton, J.D. and Yan, Y.-L., *Curr. Protoc. Mol. Biol.*, **83**, 26.8.1 (2008).
14. Nielsen, P.E., Egholm, M., Berg, R.H., and Buchardt, O., *Science*, **254**, 1497 (1991).

15. (a) Betts, L., Josey, S.A., Veal, J.M., Jordan, S.R., *Science*, **270**, 1838 (1995); (b) Eriksson, M., Nielsen, P.E., *Nature Struct. Biol.*, **3**, 5, 410 (1996); (c) Brown,S.C., Thomnson S.A., Veal, J.M., and Davis, D.G., *Science*, **265**, 777 (1994).

16. Rasmussen, H., Sandholm, O., Kastrup, J., and Nielsen, P.E., *Nature Struct. Biol.*, **4**, 2, 98 (1997).

17. Peffer, N.J., Hanvey, J.C., Biss, J.E., Thomson, S.A., Hassman, C.F., Noble, S.A., and Babiss, L.E., *Proc. Natl Acad. Sci. USA*, **90**, 10649 (1993).

18. Uhlman, E., Peyman, A., Breipohl, G., and Will, D.W., *Angew. Chem. Int. Ed.*, **37**, 2796 (1998).

19. (a) Boffa, L.C., Carpaneto, E.M., Allfrey, V.G., *Proc. Natl. Acad. Sci. USA*, **92**, 1901 (1995); (b) Boffa, L.C., Morris, P.L., Carpaneto, E.M., Louissaunt, M., and Allfrey, V.G., *J. Biol. Chem.*, **271**, 13228 (1996); (c) Ishihara, T., and Corey, D.R., *J. Am. Chem. Soc.*, **121**, 2012 (1999); (d) Zhang, X., Ishihara, T., and Corey, D.R., *Nucleic Acids Res.*, **28**, 3332 (2000); (e) Nielsen, P.E., Egholm, M., *Bioorg. Med. Chem.*, **9**, 2429 (2001).

20. Egholm, M., Buchardt, O., Nielsen, P.E., and Berg, R.H., *J. Am. Chem. Soc.*, **114**, 1895 (1992).

21. (a) Will, D.W., Breipohl, G., Langner, D., Knolle, J., and Uhlmann, E., *Tetrahedron*, **51**, 12069 (1995); (b) Planas, M., Bardaji, E., Jensen, K.J., and Barany, G., *J. Org. Chem.*, **64**, 7281 (1999); (c) Debraen, F.W. and Winssinger, N., *Org. Lett.*, **5**, 4445 (2003).

22. (a) Noble, S.A., Bonham, M.A., Bisi, J.E., Bruckenstein, D.A., Brown, P.H., Brown, S.C., Cadilla, R., Gaul, M.D., Hanvey, J.C., Hassman, C.F., Josey, J.A., Luzzio, M.J., Myers, P.M., Pipe, A.J., Ricca, D.J., Su, C.W., Stevenson, C.L., Thomson, S.A., Wiethe, R.W., Babiss, L.E., *Drug Dev. Res.*, **34**, 184 (1995); (b) Nielsen, P.E., *Mol. Biotechol.*, **26**, 233 (2004).

23. (a) Wang, L., Zhang, Q.-F., Wang, X.-S., Xue, Y.-W., Pang, D., and Fu, S.-B., *Chin. Med. J.*, **117**, 566 (2004); (b) Kaihatsu, K., Shah, R.H., Zhao, X., and Corey, D.R., *Biochemistry* **42**, 13396 (2003).

24. Sforza, S., Haaima, G., Marchelli, R., and Nielsen, P. E., *Eur. J. Org. Chem.*, **1999**, 197.

25. (a) Sforza, S., Corradini, R., Ghirardi, S., Dossena, A., and Marchelli, R., *Eur. J. Org. Chem.*, **2000**, 2905; (b) Sforza, S., Tedeschi, T., Corradini, R., Dossena, A., and Marchelli, R., *Chem. Commun.*, **2003**, 1102.

26. Sforza, S., Tedeschi, T., Corradini, R., Ciavardelli D., Dossena, A., and Marchelli, R., *Eur. J. Org. Chem.*, **1999**, 197.

27. Dragulescu-Andrasi, A., Rapireddy, S., Frezza, B.M., Gayathri, C., Gil, R.R., and Ly, D.H., *J. Am. Chem. Soc.*, **128**, 10258 (2006).

28. Tedeschi, T., Sforza, S., Corradini, R., and Marchelli, R., *Tetrahedron Lett.*, **46**, 8935 (2005).

29. (a) Gangamani, B.P., Kumar, V.A., Ganesh, K.N., *Tetrahedron*, **52**, 15017 (1996); (b) D'Costa, M., Kumar, V.A., and Ganesh, K.N., *Org. Lett.*, **3**, 1281 (2001); (c) Vilaivan T. and Srisuwannaket, C., *Org. Lett.*, **8**, 1897 (2006).

30. (a) Kitamatsu, M., Shigeyasu, M., Saitoh, M., and Sisido, M., *Biopolymers (Pept. Sci.)*, **84**, 267 (2006); (b) Kitamatsu, M., Kashiwagi, T., Matsuzaki, R. and Sisido, M., *Chem. Lett.*, **35**, 300 (2006).

31. (a) Koppelhus, U., Awasthi, S., Zachar, V., Holst, H.U., Ebbesen, P., and Nielsen, P.E., *Antisense Nucleic Acid Drug Dev.*, **12**, 1236 (2002); (b) Eriksson, M., Nielsen, P. E., and Gond, L. B., *Biol. Chem.*, **277**, 7144 (2002).

32. (a) Boffa, L.C., Scarfì, S., Mariani, M.R., Damonte, G., Allfrey, V.G., Benatti, U., and Morris, P.L., *Cancer Res.*, **60**, 2258 (2000); (b) Cutrona, G., Carpaneto, E.M., Ulivi, M. Roncella, S., Landt, O., Ferrarini, M., and Boffa, L.C., *Nat. Biotechnol.*, **18**, 300 (2000).

33. Eriksson, M., Nielsen, P.E., and Good, L., *J. Biol. Chem.*, **277**, 7144 (2002).

34. (a) Oehlke, J., Wallukat, T.G., Wolf, Y., Ehrlich, A., Wiesner, B. , Berger, K., and Bienert, M., *Eur. J. Biochem.*, **271**, 3043 (2004); (b) Braun, K., Peschke, P., Pipkorn, R., Lampel, S., Wachsmuth, M., Waldeck, W., Friedrich, E., and Debus, J., *J. Mol. Biol.*, **318**, 237 (2002); (c) Pelleston, F., Paulasova, P., *Chromosome*, **112**, 375 (2004).

35. (a) Eschenmoser, A., *Nachr. Chem.. Tec. Lab.*, **39**, 795 (1991); (b) Eschenmoser, A. and Dobler, M., *Helv. Chim. Acta*, **25**, 258 (1992); (c) Eschenmoser, A., *Pure Appl. Chem.*, **65**, 1199 (1993); (d) Böhringer, M, Roth, H.-J., Hunziker, J., and Eschenmoser, A., Helv. Chim. Acta, 75, 1416 (1992); (e) Eschenmoser, A. and Krishnamurthy, R., *Pure Appl. Chem.*, **72**, 343 (2000).

36. Lohse P., *Ph.D. thesis*, ETH-Zürich (Switzeland), 1992.

37. Gmelin, R., *Z. Physiol. Chem.*, **316**, 164 (1959).

38. (a) Bell, E.A. and Foster, R.G., *Nature*, **194**, 91 (1962); (b) Giller, S.A., Yu Lidak, M., Shluke, Y.Y., and Shvachkin, T., *Chem. Heterocyclic Compd.*, **2**, 124 (1968); (c) Yu Lidak, M., Paegle, R.A., Petz, K.Y., and Shvachktin, T., *Ibid.*, **2**, 193 (1968); (d) Yu Lidak, M., Shluke, Y.Y., and Shvachkin, T., *Ibid.*, 2, 955 (1968).

39. Bonnen, J., Dahmus, M.E., Fainbrough, D., Huang, R.C., Marushige, K., and Tuan D.H.T., *Science*, **159**, 47 (1968).

40. Beiser, S.M and Erlanger, B.F., *Cancer Res.*, **26**, 2012 (1966).

41. (a) Doel, M.T., Jones, A.S., and Taylor, N., *Tetrahedron Lett.*, **10**, 2285 (1969); (b) Doel, M.T., Jones, A.S., and Walker, R.T., *Tetrahedron*, **30**, 2755 (1974).

42. Battrey, J.O, Jones, A.S., and Walker, R.T., *Tetrahedron*, **31**, 73 (1975).

43. (a) De Koning, H. and Pandit, U.K., *Recueil*, **91**, 875 (1971); (b) De Koning, H. and Pandit, U.K., *Ibid.*, 1069.

44. (a) Arnold, L.D., Kalantar, T.H., Vederas, J.C., *J. Am. Chem. Soc.*, **107**, 7105 (1985); (b) Arnold, L.D., May, R.G., and Vederas, J.C., *J. Am. Chem. Soc.*, **110**, 2237 (1988); (c) Pansare, S., Huyer, G., Arnold, L.D., Vederas, J.C., *Organic Syntheses,* **Coll. Vol. 9** p. 58 (1998), **Vol. 7** p. 70 (1992).

45. Oberhauser, B., Oberhauser, H., Hofbauer, B., Baschang, G., and Eschenmoser, A., *Croat. Chem. Acta*, **69**, 2, 535 (1996).

46. Diederichsen, U., *Angew. Chem. Int. Ed.*, **36**, 1886 (1997).

47. Diederichsen, U., *Angew. Chem. Int. Ed.*, **35**, 445 (1996).

48. (a) Diederichsen, U., *Bioorg. Med. Chem. Lett.*, **7**, 1743 (1997); (b) Diederichsen, U., *Angew. Chem. Int. Ed.*, **37**, 2273 (1998).

49. Hoffmann, M.F.H., Brückner, A.M., Hupp, T., Engels, B., and Diederichsen, U., *Helv. Chim. Acta*, **83**, 2580 (2000).

50. Diederichsen, U. and Schmitt, H.W., *Tetrahedron Lett.*, **37**, 475 (1996).

51. Diederichsen, U., Weicherding, D., and Diezemann, N., *Org. Biomol. Chem.*, **3**, 1058 (2005).

52. (a) Nollet, A.J.H., Huting, C.M., and Pandit, U.K., *Tetrahedron*, **25**, 5971 (1969) (b-c) Nollet, A.J.H. and Pandit, U.K., *Ibid.*, 5983 and 5989.

53. (a) Diederichsen, U. and Schmitt, H.W., *Angew. Chem. Int. Ed.*, **37**, 302 (1998); (b) Diederichsen, U. and Schmitt, H.W., *Eur. J. Org. Chem.* **1998**, 827.

54. (a) Seebach, D. and Matthews, J.L, *Chem. Commun.*, **1997**, 2015; (b) Gellman, S.H., *Acc. Chem. Res.*, **31**, 173 (1998); (c) Chen, R.P., Gellman, S.H., and DeGrado, W.F., *Chem. Rev.*, **101**, 3219 (2001).

55. (a) Brückner, A.M., Schmitt, H.W., and Diederichsen, U., *Helv. Chim. Acta*, **85**, 3855 (2002); (b) Brückner, A.M., Garcia, M. , Marsh, A., Gellman, S.H., and Diederichsen, U., *Eur. J. Org. Chem.*, **2003**, 3555.

56. Weiß, A. and Diederichsen, U., *Eur. J. Org. Chem.*, **2007**, 5531.

57. Garner, P., Dey, S., Huang, Y., and Zhang, X., *Org. Lett.*, **1**, 403 (1999).

58. (a) Garner, P., Dey, S., and Huang Y., *J. Am. Chem. Soc.*, **122**, 2405 (2000); (b) Huang, Y., Dey, S., Zhang, X., Sönnichsen, F., and Garner, P., *J. Am. Chem. Soc.*, **126**, 4626 (2004).

59. Curran, T., Kerppola, T.K., *Science*, **254**, 1210 (1991).

60. (a) Fasman, G.D., *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenum Press, New York (NY, USA), 1996; (b) Woody, A.W. and Tinoco, I.jr., *J. Chem. Phys.*, **46**, 4927 (1967).

61. Usatyi, A.F. and Shlyakhtenko, L.S., *Biopolymers*, **12**,45 (1973).

62. IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry*, **9**, 3471 (1970).

63. Toniolo, C. and Benedetti, E., *Trends Biochem. Sci.*, **16**, 350 (1991).

64. Ramachandran, G.N.and Sasisekharan, V., *Adv. Protein Chem.*, **23**, 283 (1968).

65. Pavone, V., Di Blasio, B., Pedone, C., Santini, A., Benedetti, E., Formaggio, F., Crisma, M. and Toniolo, C., *Gazz. Chim. Ital.*, **121**, 21 (1991).

66. Donohue, J., *Proc. Natl. Acad. Sci. USA*, **39**, 470 (1953).

67. Ramachandran, G.N., Venkatachalam, C.M., and Krimm, S., *Biophys. J.*, **6**, 849 (1966).

68. Baker, E.N. and Hubbard, E., *Prog. Biophys. Mol. Biol.*, **44**, 97 (1984).

69. Robbins, A.H. and Stout, C.D., *Protein Struct. Funct. Genet.*, **5**, 289 (1989).

70. Pathak, D. and Ollis, D., *J. Mol. Biol.*, **214**, 497 (1990).

71. Kostrikis, L.G., Liu, D.J., and Day, L.A., *Biochemistry*, **33** (1994).

72. Venkatachalam, C.M., *Biopolymers*, **6**, 1425 (1968).

73. Valle, G., Formaggio, F., Crisma, M., Bonora, G.M., Toniolo, C., Bavoso, A., Benedetti, E., Di Blasio, B., Pavone, V., and Pedone, C., *J. Chem. Soc., Perkin Trans. 2*, 1371 (1986).

74. (a) Ramachandran, G.N. and Chandrasekharan, R., *Indian J. Biochem. Biophys.*, **9**, 1 (1972); (b) Burgess, A. W. and Leach, S.J., *Biopolymers.*, **12**, 2599 (1973).

75. (a) Toniolo, C., Crisma, M., Formaggio, F., and Peggion, C., *Biopolymers (Pept. Sci.).*, **60**, 396 (2001); (b) Toniolo, C., Crisma, M., Formaggio, F., Peggion, C., Broxterman, Q.B., and Kaptein, B., *Biopolymers (Pept. Sci.)*, **76**, 162 (2004).

76. (a) Marshall, G.R., in *Intra-Science Chemistry Reports*, Kharasch, N. (Ed.), Gordon and Breach, New York, (NY, USA), **Vol. 5**, pp. 305-316 (1971); (b) Venkataram Prasad, B.V.and Sasisekharan, V., *Macromolecules*, **12**, 1107 (1979).

77. (a) Paterson, Y., Rumsey, S.M., Benedetti, E., Némethy, G., and Scheraga, H.A., *J. Am. Chem.Soc.*, **103**, 2947 (1981); (b) Improta, R., Rega, N., Aleman, C. and Barone, V., *Macromolecules*, **34**, 7550 (2001).

78. Improta, R., Barone, V., Kudin, K.N., and Scuseria, G.E., *J. Am. Chem. Soc.*, **123**, 3311 (2002).

79. Polese, A., Formaggio, F., Crisma, M., Valle, G., Toniolo, C., Bonora, G.M., Broxtermann, Q.B. and Kamphuis, J., *Chem. Eur. J.*, **2**, 1104 (1996).

80. Toniolo, C., Polese, A., Formaggio, F., Crisma, M., and Kamphuis, J., *J. Am. Chem. Soc.*, **118**, 2744 (1996).

81. Manning, M. and Woody, R., *Biopolymers*, **31**, 569 (1991).

82. Benedetti, E., Bavoso, A., Di Blasio, B., Pavone, V., Pedone, C., Crisma, M., Bonora, G.M., and Toniolo, C., *J. Am. Chem. Soc.*, **104**, 2437 (1982).

83. Toniolo, C., Valle, G., Bonora, G.M., Crisma, M., Formaggio, F., Bavoso, A., Benedetti, E., Di Blasio, B., Pavone, V., and Pedone, C., *Biopolymers*, **25**, 2237 (1986).

84. Valle, G., Crisma, M., Toniolo, C., Beißwenger, R., Rieker, A. and Jung, G., *Liebigs Ann. Chem.*, **1989**, 337.

85. Valle, G., Crisma, M., and Toniolo, C., *Zeit. Kristallogr.*, **188**, 261 (1989).

86. Souhassou, M., Smith, G.D., Leplawy, M.T., and Marshall, G.R., *Acta Crystallogr.*, ***A46***, suppl., C-140, PS-04.01.16 (1990).

87. Di Blasio, B., Santini, A., Pavone, V., Pedone, C., Benedetti, E., Moretto, V., Crisma, M., and Toniolo, C., *Struct. Chem.*, **2**, 523 (1991).

88. Bavoso, A., Benedetti, E., Di Blasio, B., Pavone, V., Pedone, C., Toniolo, C., and Bonora G.M., *Proc. Natl. Acad. Sci. USA*, **83**, 1988 (1986).

89. Pavone, V., Di Blasio, B., Santini, A., Benedetti, E., Pedone, C., Toniolo, C., and Crisma, M., *J. Mol. Biol.*, **214**, 633 (1990).

90. Toniolo, C., Crisma, M., Bonora, G.M., Benedetti, E., Di Blasio, B., Pavone, V., Pedone, C., and Santini, A., *Biopolymers*, **31**, 129 (1991).

91. Geßmann, R., Brückner, H. e Kokkinidis, M., *Acta Crystallogr.*, **B54**, 300 (1998).

92. Geßmann, R., Brückner, H. e Petratos, K., *J. Pept. Sci.*, **9**, 753 (2003).

93. (a) Paterson, Y., Stimson, E.R., Evans, D.J., Leach, S.J. and Scheraga, H.A., *Int. J. Pept. Protein Res.*, **20**, 468 (1982); (b) Toniolo, C., Bonora, G.M., Barone, V., Bavoso, A., Benedetti, E., Di Blasio, B., Grimaldi, P., Lelj, F., Pavone, V. and Pedone, C., *Macromolecules*, **18**, 895 (1985).

94. Karle, I.L. and Balaram, P., *Biochemistry*, **29**, 6747 (1990).

95. Benedetti, E., Di Blasio, B., Pavone, V., Pedone, C., Santini, A., Crisma, M. and Toniolo, C., in *Molecular Conformation and Biological Interactions*, Balaram, P. and Ramaseshan, S. (Eds.), Indian Academy of Science, Bangalore (India), 1998, pp. 497-502.

96. Zikou, S., Koukkou, A.I., Mastora, P., Sakarellos-Daitsiotis, M., Sakarellos, C., Drainas, C., and Panou-Pomonis, E., *J. Pept. Sci.*, **13**, 481 (2007).

97. Agrawal, S. (Ed.), *Methods in Molecular Biology., Vol. 26*, Humana Press Inc., Totowa (NJ, USA), 1994.

98. (a) Benedetti, E., Barone, V., Bavoso, A., Di Blasio, B., Lelj, F., Pavone, V., Pedone, C., Bonora, G.M., Toniolo, C., Leplawy, M.T., Kaczmarek, K. and Redlinski, A., *Biopolymers*, **27**, 357 (1988); (b) Toniolo, C., Bonora, G. M., Bavoso, A., Benedetti, E., Di Blasio, B., Pavone, V., Pedone, C., Barone, V., Lelj,

F., Leplawy, M.T., Kaczmarek, K.E. and Redlinski, A., *Biopolymers*, **27**, 373 (1988).

99. Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V., *J. Mol. Biol.*, **7**, 95 (1963).

100. (a) Merrifield, R.B., Barany, G., Cosand, W.L., Engelhard, M., and Mojsov, S., *Pept.: Proc. Am. Pept. Symp.*, 5th, **1977**; (b) Green, T.W. and Wuts, P.G., *Protective Groups in Organic Synthesis*, 3rd edn. Wiley, Hoboken (NJ, USA), 1999.

101. (a) Schaller, H., Weimann, G., Lerch, W.B., and Khorana, H.G., *J. Am. Chem. Soc.*, **85**, 3821 (1963); (b) Urdea, M.S. and Won, T., *Tetrahedron Lett.*, **27**, 2933 (1986).

102. (a) Zhou, X.-X., Sandström, A., and Chattopadyhaya, J., *Chem. Scr.*, **26**, 241 (1986); (b) Kamimura, T., Tsuchiya, M., Koura, K., Sekine, M., and Hata, T., *Tetrahedron Lett.*, **24**, 2775 (1983).

103. (a) Xu, Y.-Z. and Swann, P.F., *Nucleic Acids Res.*, **18**, 4061; (b) Büchi, H. and Khorana, H.G., *J. Mol. Biol.*, **72**, 251 (1972).

104. Dueholm, K.L., Egholm, M., Behrens, C., Christensen, L., and Nielsen, P.E., *J. Org. Chem.*, **59**, 5767 (1994).

105. Von Saltza, M.H., *Ph.D. Thesis*, University of Madison (WI, USA), 1959.

106. Benoiton, N. L., *Chemistry of peptide synthesis*, CRC Press, Boca Raton (FL, USA), 2006, chap. 5, pp.125-156.

107. Bodanszky, M., *Principles of Peptide Chemistry*, Springer-Verlag, New York (NY, USA), 1984, p.99.

108. Carpino, L.A. and Han, G.Y., *J. Am. Chem. Soc.*, **93**, 5748 (1970).

109. Chaltin, P., Lescrinier, E., Lescrinier, T., Rozanski, J., Hendrix, C., Rosemeyer, H., Busson, R., Van Aershot, A., and Herdewijn, T., *Helv. Chim. Acta*, **85**, 2258 (2002).

110. Bergmann, M., Zervas, L., *Ber. Dtsch. Chem. Ges.*, **65**, 1192 (1932).

111. (a) Anderson, G.W., and Callahan, F.M., *J. Am. Chem. Soc.*, **82**, 3359 (1960); (b) McGahren, W.J. and Goodman, M., *Tetrahedron*, **23**, 2017 (1967); (c) Bryan, D.B., Hall, R.F., Holden, K.G., Huffman, W.F., and Gleason, J.G., *J. Am. Chem. Soc.*, **99**, 2353 (1977).

112. (a) Nemes, B., Andries, T., and Steglich, W., *J. Chem. Soc., Chem. Commun.*, **1982**, 1132; (b) Vedejs, E., Bennatt, N.S., Conn, L.M., Owen, S.T., Gingras, M., Lèis, S., Oliver, P.A., and Paterson, M.J., *J. Org. Chem.*, **58**, 7286 (1993); (c) Hartung, W.H., and Simoneff, R., *Org. React.*, **VII**, 263 (1953).

113. (a) Leplawy, M.T., Jones, D.S., Kenner, G.W., and Sheppard, R.C., *Tetrahedron*, **11**, 39 (1960); (b) Boissonnas, R.A., Guttmann, S., Jaquenoud, P.-A., and Waller, J.-P., *Helv. Chim. Acta*, **39**, 1421 (1956); Corey, E.J., Székely, I., and Shiner, C.S., *Tetrahedron Lett.*, **16**, 3529 (1977).

114. (a) Mitsunobu, O. *Synthesis*, **1981**, 1; (b) Grochowski, E., Hilton, B.D., De Kupper, R.J., and Michejda, C.J., *J. Am. Chem. Soc.*, **104**, 6876 (1982).

115. (a) Vorbrüggen, H. and Bennua, B., *Chem. Ber.*, **114**, 1279 (1981); (b) Zou, R. and Robins, M.J., *Can. J. Chem.*, **65**, 1436 (1987).

116. Carpino, L.A., *J. Am. Chem. Soc*, **115**, 4397 (1993).

117. König, W. and Geiger, R., *Ber. Dtsch. Chem. Ges.,* **103,** 788; **103**, 2024 (1970).

118. Pritz, S., Wolf, Y., Klemm, C., and Bienert, M., *Tetrahedron Lett.*, **47**, 5893 (2006).

119. Castro, B., Dormay, J.R., Evin, R., and Selve, C., *Tetrahedron Lett.*, **14**, 1219 (1975).

120. Hudson, D., *J. Org. Chem.*, **53**, 617 (1988).

121. Fujii, N., Otaka, A., Kamura, O., Akaji, K., Funakoshi, S., Hayashi, Y., Kuroda, Y., and Yajima, H., *J. Chem. Soc., Chem. Commun.*, **1987**, 274.

122. (a) Carpino, L.A., *J. Am. Chem. Soc.*, **115**, 4397 (1993); (b) Miranda, L.P. and Alewood, P.F., *Procl. Nat. Acad. Sci. USA*, **96**, 1181 (1999).

123. (a) Bellon, P., Cotton, R., Ciler, M.B., Atherdon, E., Horton, J., and Richards, J.D., *Peptides 1988*, Jung., G., Bayer, E. (Eds.), Walter de Gruyer, Berlin-New York, 1989, p. 699; (b) Spencer, J.R., Antonenko, V.V., Delael, N.G.J., and Goodman, M., *Int. J. Pept. Prot. Res.*, **40**, 282 (1992); (c) Wenschuh, H., Beyermann, M., Krause, E., Brudel, M., Winter, R., Schümann, M., Carpino, L.A., Bienert, M., *J. Org. Chem.*, **59**, 3275 (1994); (d) Kamiński, Z.J., Kolesińska, B., Kolesińska, J., Sabatino, G., Chelli, M., Rovero, P., Błaszczyk, M., Główka, M.L., and Papini, A.M., *J. Am. Chem. Soc.*, **127**, 16912 (2005).

124. Tarbell, D.S., Yamamoto, Y., and Pope, B.M., *Procl. Nat. Acad. Sci. USA*, **69**, 730 (1972).

125. (a) Knorr, R, Trzeciak, A, Bannawarth, W, and Gillessen, D. *Tetrahedron Lett.***, 30**, 1927 (1989); (b) Fields, G.C., Lloyd, D.H., Macdonald, R.C., Otteson, K.M., and Noble, R.L., *Pept. Res.***, 4**, 95 (1991).

126. (a) Kamber, B., *Helv. Chim. Acta*, **54**, 927 (1971); (b) Kamber, B., Hartmann, A., Eisler, K., Reiniker, B., Rink, H., Sieber, P., and Rittle, W., *Helv. Chim. Acta*, **63**, 899 (1980).

127. (a) Schmid, S.L. and Carter, L.L, *J. Cell. Biol.*, **111**, 2307 (1990); *(b)* Smith, B.A., Daniels, D.S., Coplin, A.E., Jordan, G.E., McGregor, L.M., and Schepartz, A., *J. Am. Chem. Soc.*, **130**, 2948 (2008).

128. Green, N.S., Reisler, E., and Houk, K.N., *Protein Sci*., **10**, 1293 (2001).

129. Bellamy, M., *The Infra-Red Spectra of Complex Molecules*, Methuen, London (U.K.), 1956.

130. Palumbo, M., Da Rin, S., Bonora, G.M., and Toniolo, C., *Makromol. Chem.*, **177**, 1477 (1976).

131. Bonora, G.M., Mapelli, C., Toniolo, C., Wilkening, R.R. and Stevens, E.S., *Int. J. Biol. Macromol.*, **6**, 179 (1984).

132. Kennedy, D.F, Crisma, M., Toniolo, C., and Chapman, D., *Biochemistry*, **30**, 654 (1991).

133. Dwivedi, A.M., Krimm, S., and Malcom, B.R., *Biopolymers*, **23**, 2305 (1984).

134. (a) Martin, R. and Hauthal, G., *Dimethyl Sulphoxide*, Van Nostrand-Reinhold, Wokingham (U.K.), 1975; (b) Pitner, T.P. and Urry, D.W., *J. Am. Chem. Soc*., **94**, 1399 (1972).

135. (a) Griesinger, C., Otting, G., Ernst, R.R., and Wütrich, K., *J. Am. Chem. Soc.*, **110**, 7870 (1988); (b) Williams, K.R. and King, R.W., *J. Chem. Ed.*, **67**, A125 (1990).

136. (a) Braunschweiler, L. and Ernst, R.R., *J. Magn. Reson.*, **53**, 521 (1985); (b) Bazzo, R., Boyd, J., Campbell, I.D., and Soffe, N., *J. Magn. Res.*, **73**, 369 (1987).

137. (a) Bothner-By, A.A., Stephens, R.L., Lee, J., Warren, C.D., and Jeanloz, R.W., *J. Am. Chem. Soc.*, **106**, 811 (1984); (b) Bax, D. and Davis, D.J., *J. Magn. Reson.*, **63**, 207 (1985).

138. Wüthrich, K., *NMR of Proteins and Nucleic Acids*, Wiley, New York (NY, USA), 1986.

139. Bernassau, N.M. and Nuzillard, J.M., *J. Magn. Reson. Ser. B*, **103**, 77 (1994).

140. (a) Fasman, G.D., *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenum Press, New York (NY, USA), 1996; (b) Woody, A.W. and Tinoco, I.jr., *J. Chem. Phys.*, **46**, 4927 (1967); (c) Sudha, T.S., Vijayakumar, E.K.S., and Balaram, P., *Int. J. Pept. Prot. Res.*, **22**, 464 (1983).

141. *Surface plasmons on smooth and rough surfaces and on gratings*, Springer Verlag, Berlin (Germany), 1988.

142. Wilchek, M. and Bayer, E.A., *Protein Recognition of Immobilized Ligands.* Hutchins, T.W., ed. Alan R. Liss, Inc., New York (USA), 1989, pp. 83-90.

143. Koning, A.J., Lum, P.Y., Willliams, J.M., and Wright, R., *Cell. Motil. Cytoskeleton*, **25**, 111 (1993).

144. Fathman, C.G., Cone, J.L., Sharrow, S.O., Tyrer, H., and Sachs, D.H., *J. Immunol.*, **115**, 584 (1975).

145. Cory, A.H., Owen, T.C., Barltrop, J.A., and Cory, J.G., *Cancer Commun.*, **3**, 207 (1991).

146. Vida, T. A. and Emr, S.D., *J. Cell Biol.*, **128**, 779 (1995).

147. (a) Sheppard, C.J. and Wilson, T., *J. Microsc.*, **124**, 107 (1981); (b) Pawley, J.B. (Ed.), *Handbook of Biological Confocal Microscopy*, 3th ed., Springer, Berlin (Germany), 2006.

148. Kapuscinski, J., *Biotech. Histochem.*, **70**, 220 (1995).

149. (a) Schmid, S.L. and Carter, L.L., *J. Cell. Biol.*, **111**, 2307 (1990); (b) Smith, B.A., Daniels, D.S., Coplin, A.E., Jordan, G.E., McGregor, L.M., and Schepartz, A., *J. Am. Chem. Soc.*, **130**, 2948 (2008).

150. Rolfs, A. and Hediger, M.A., *J. Physiol.*, **518**,1 (1999).

151. (a) Duchardt, F., Fotin-Mleczek, M., Schwarz, H., Fischer, R., and Brock, R., *Traffic* 2007, **8**, 848; (b) Futaki, S., Nakase, I., Tadokoro, A., Takeuchi, T., and Jones, A.T., *Biochem. Soc. Trans.*, **35**, 784 (2007); (c) Bareford, L.M., Swaan, P.W., *Adv. Drug Del. Rev.*, **59**, 748 (2007).

152. Guichard, G., Benkirane, N., Zeder-Lutz, G., van Regenmortel, M.H., Briand, J.-P., and Muller, S., *Proc. Nat. Ac. Sci.(USA)*, **91**, 9765 (1994).

153. Kaiser, E. Colescott, R.L., Bossinger, C.D., and Cook, P.I., *Anal. Biochem.* **34**, 595 (1970).

154. Hancock, W.S. and Battersby, J.E., *Anal. Biochem.*, **71**, 260 (1976).

155. Von Arx, M. Faupel, M., and Brugger, M., *J. Chromatogr.*, **120**, 224 (1976).

156. Burla, M.C., Camalli, M., Carrozzini, B., Cascarano, G.L., Giacovazzo, C. Polidori, G., and Spagna, R., *J. Appl. Crystallogr.*, **36**, 1103 (2003).

157. Sheldrick, G.M., *Acta Crystallogr.* **2008**, *A64*, 112.

158. Tognon, S., *Master thesis*, Departement of Organic Chemistry, Padova (Italy), 2001-2002.

159. MacKay, D., McIntyre, D.D., *Can. J. Chem.*, **62**, 355 (1984).

160. Valle, G., Bonora, G.M., and Toniolo, C., *Can. J. Chem.*, **62**, 2661 (1984).

161. McGahren, W.J. and Goodman, M., *Tetrahedron*, **23**, 2017 (1967).

162. Leplawy, M.T., Jones, D.S., Kenner, G.W. and Sheppard, R.C., *Tetrahedron*, **11**, 39 (1960).

163. Geotti-Bianchini, P., Beyrath, J., Chaloin, O., Formaggio, F., and Bianco, A., *Org. Biomol. Chem.*, **6**, 3661 (2008).

164. Benedetti, E., Bavoso, A., Di Blasio, B., Pavone, V., Pedone, C., Toniolo, C., Bonora, G.M., and Crisma, M., *Int. J. Pept. Protein Res.*, **22**, 385 (1983).

165. Neimark, J. and Briand, J.-P., *Pept. Res.*, **6**, 219 (1993).

166. Corradini, R., Sforza, S., Dossena, A., Palla, G., Rocchi, R., Filira, F., Nastri, F., and Marchelli, R., *J. Chem. Soc., Perkin Trans. 1*, 2690 (2001).

167. Fujii, N., Otaka, A., Kamura, O., Akaji, K., Funakoshi, S., Hayashi, Y., Kuroda, Y., and Yajima, H., *J. Chem. Soc., Chem. Commun.*, 274 (1987).

Publications et communications derived from the thesis

Publications:

Geotti-Bianchini, P., Beyrath, J., Chaloin, O., Formaggio, F., Bianco, A., **Design and Synthesis of Intrinsically Cell-Penetrating Nucleopeptides**, *Org. Biomol. Chem.*, **6**, 3661 (2008) [DOI: 10.1039/b811639c].

Geotti-Bianchini, P., Crisma, M., Peggion, C., Bianco, A., Formaggio, F., **Conformationally controlled, thymine-based α-nucleopeptides**, *Chem. Commun.*, submitted.

Geotti-Bianchini, P., Crisma, M., Peggion, C., Bianco, A., Formaggio, F., **Synthesis and 3D-structure of conformationally controlled nucleopeptides,** proceedings of the 20th American Peptide Symposium, E. Escher and W. Lubell Eds. 2008, in press.

Geotti-Bianchini, P., Beyrath, J., Chaloin, O., Bianco, A., Formaggio, F., **Design and synthesis of cell-penetrating nucleopeptides**, proceeding of the 30th European Peptide Symposium in *Peptides 2008*, in press.

Posters :

Geotti-Bianchini, P., Crisma, M., Peggion, C., Bianco, A., Formaggio, F., **Synthesis of helical, conformationally controlled nucleopeptides**, Poster P 58 at the 15th Réunion du Groupement Français de Peptides et Protéines. Dinard (France), 20-25 May 2007.

Geotti-Bianchini, P., Crisma, M., Bianco, A., Formaggio, F., **Synthesis and 3D-structure of conformationally controlled nucleopeptides**, Poster 138 at the 20th American Peptide Society Symposium. Montréal (Canada), 26-30 June 2007.

Geotti-Bianchini, P., Beyrath, J., Chaloin, O., Decossas, M., Bianco, A., Formaggio, F., **Design and synthesis of cell-penetrating nucleopeptides,** poster 14 at the National Meeting of Chemistry of Biological Systems - Italian Chemical Society (SCI- Section CSB). Riccione (Italy) 23-24 September 2008.

Oral communications:

Geotti-Bianchini, P., Beyrath, J., Chaloin, O., Decossas, M., Bianco, A., Formaggio, F., **Conformationally controlled nucleopeptides: synthesis, 3D-structure and biological studies.** National Meeting of Chemistry of Biological Systems - Italian Chemical Society (SCI- Section CSB). Montagnana (Italy), 8-9 September 2007.

Geotti-Bianchini, P., Beyrath, J., Chaloin, O., Bianco, A., Formaggio, F., **Design and synthesis of cell-penetrating nucleopeptides**, abstract Y12 at the Young Investigator Mini Symposium at the 30th European Peptide Symposium, Helsinki (Finland), 30 August - 5 September 2008.

Acknowledgements

This Ph.D. thesis project was funded by the French-Italian University (UIF-UFI) through a Vinci Ph.D. scholarship and a 'bourse d'accompagnement'.

I wish to thank my tutor, professor Fernando Formaggio and my cotutor, doctor Alberto Bianco, for having supervised my thesis project, for all what they have taught me, professionally and, maybe more importantly, as persons, and for their unlimited readiness to give me help, advice, and confort every time I was in need (i.e. just too often).

I would like to thank professor Claudio Toniolo for having disclosed to me the charme of peptide research and for allowing me to work in his group for my master and Ph.D. thesis. I would like to thank doctor Sylviane Muller for allowing me to work in her laboratory during my stay in Strasbourg, and for the high opinion of me she has displayed. I thank doctor Jean-Paul Briand for having taught me very much about semi-automated peptide synthesis and peptide purification through preparative HPLC.

I thank doctor Marco Crisma for the resolution of the structure of the only nucleopeptide crystal I was able to obtain, and for the patience in considering many failed crystallization trials, as well as for what I learned from him about peptide structure. I thank doctor Cristina Peggion for the 2D NMR spectra, in particular for the huge amount of experiments and inductive work she has done in order to understand the arrangement of the hexapeptide dimer. I owe enormous gratitude to doctor Alessandro Moretto, who has given me precious help and advice, but also moral assistance, from the beginning to the end of my synthetic work in Padova, representing for me, with his words and even more with his example, the model of the successful junior researcher.

I thank doctor Barbara Biondi for performing MS analyses and trying to decipher my orrible handwriting. I thank Salvatore Spatola for the IR spectra. I thank doctor Andrea Caporale for introducing me to the use of the preparative HPLC and for his valuable synthetic expertise, as well as for the friendly talks in Padova and Helsinki.

I thank doctor Olivier Chaloin for the SPR measurements on my nucleopeptides, for his patience in understanding which kind of experiments I wanted to perform and in explaining their results, as well as for his sympathy. I would like to thank doctor Marion Decossas for her expertise about the microscopy experiments. I thank Julien Beyrath for the biological tests he performed and for the thousand questions he answered me about them, but also for having shared my emotions and disappointments in the six refusals of our first manuscript. I thank doctor engineer Jean Neimark for explaining me about his peptide synthesizer and for his kindness and nice attitude. I would like to thank also the specialists of the NMR facility in Strasbourg, and specially doctor Lionel Allouche, for performing the 2D NMR experiments and teaching me about spectral processing, starting from the most elementary operations.

I would like to thank two graduate students who shared a part of this project: Enrica Venturelli fought with me against the serine β-lactone sensitivity during her BA thesis, while Elena Piva had to deal with the lability of the 6-chloro group of the guanine-based nucleopeptides and with the quest for a working orthogonal set of protectors for cytosine-based nucleopeptides during her MA thesis. I think I was really lucky to work with such motivated, skilled and hard-working students, which are also kind and nice persons, friends as well as coworkers.

I thank the members of my doctoral commission in Padova, in particular its president, professor Stefano Mammi, for the attention given to my research and for their stimulating criticism and suggestions.

I wish to thank the European Doctoral College of Strasbourg for accepting my application, providing support for the accommodation during my stay in Strasbourg and for offering me the possibility of attending conferences about various aspects of European integration, history and culture. I would like to thank particularly Ms. Céline Montibeller for her incredible ability to know everything, remember everything and provide everything I could need and doctor Michael Vorbeck for his generous engagement in finding interesting speakers about interesting subjects, for his humour and sympathy.

Terminati i ringraziamenti istituzionali, mi ritengo libero di proseguire nella mia lingua: le persone che vorrei ringraziare sono davvero tantissime, e purtroppo lo spazio di queste pagine non basta. Per primi non posso non citare i dottorandi che mi hanno preceduto nel nostro laboratorio: Ivan, sempre simpatico, e capace di sorridere, e ridere, pur nelle situazioni più stressanti, Chiara, che nella sua saggezza mi ha insegnato che occorre sempre avere pronto un 'piano B', Marta, di cui in questi anni ho imparato a conoscere la generosità, la disponibilità, e l'umiltà, ma anche la competenza professionale, e che ringrazio anche per avermi sempre ricordato di 'togliere la trave' al momento giusto… Vorrei poi salutare con un grande 'in bocca al lupo' i dottorandi che rimangono: Morelle, di cui ammiro sempre la saggezza e la pazienza (che nello stress degli ultimi mesi ho a volte messo a dura prova), e Matteo, sempre efficiente e simpatico.

Grazie alle due 'laureande giovani', Claudia e Vanessa (o 'Vane&Clau'), che hanno portato in laboratorio una ventata di freschezza e allegria, ed hanno sofferto tante volte la fame per attendere 'ancora cinque minuti' che fossi pronto per andare a mensa e lì atteggiarmi a 'vecchio saggio del paese'. In bocca al lupo a Edoardo, il nuovo arrivato.

I thank the lab mates, but also the friends, I had the opportunity to know in Strasbourg: Shou-Ping, always nice and smiling, Raquel, never losing her sympathy, even when her dwarfs were growing, Prabhpreet, so kind and ready to help, and Cristian 'l'Argentin', for his fascinating climbing stories and photos.

Merci beaucoup à Claire, qui à toujours été prête à m'aider et qui est une des personnes plus sympathiques que je connais, à Lucile, qui m'a appris beaucoup, à Marie-Charlotte, qui à lutté avec moi contre l'HPLC, à Paul, qui a partagé l'expérience Finlandaise, et à Cristophe, pour son dessin. Ringrazio Joël per le lezioni di calcio (teoria e pratica) e le risate al congresso di Dinard. Grazie molte a Cristian, per la sua generosità e simpatia.

Vorrei ringraziare mia madre per il continuo supporto morale (divenuto in realtà 'sopporto' nelle ultime settimane di redazione della tesi), nonché per avermi sempre invitato a dare il giusto valore agli impegni professionali rispetto a tutto il resto. In questi tre anni mio padre è stato forse chi ha ascoltato di più l'evoluzione delle mie ricerche, delle mie speranze, delusioni e soddisfazioni; anche senza poterlo esprimere a parole, è riuscito ad incoraggiarmi ed a farmi comprendere il suo interesse per il mio lavoro. Costanza ha sempre cercato di aiutarmi a sdrammatizzare e vivere con più ironia ed autoironia, e mi ha aiutato anche concretamente nella revisione delle bozze. Ringrazio tutti i parenti, e soprattutto i nonni, per la fiducia e gli incoraggiamenti, nonché per il loro affetto.

Per ultimo il ringraziamento più importante, a Elisa. Non so se senza di lei avrei avuto la forza di imboccare il cammino di dottorato e di percorrerlo fino alla fine, mentre sono certo di avere vissuto diversamente il mio percorso di crescita scientifica grazie ai suoi suggerimenti ed alle sue critiche sempre costruttive, anche se spesso volutamente provocatorie. Il mio livello di consapevolezza nello studio delle proprietà biologiche dei nucleopeptidi è migliorato molto grazie ai suoi stimoli ed alla sua disponibilità a chiarire molti dei miei dubbi da rozzo chimico. Questa tesi, che le deve così tanto in qualità e quantità, è dunque dedicata a lei.