Conformationally controlled, thymine-based alpha-nucleopeptides

Conformationally controlled, thymine-based α-nucleopeptides†

Piero Geotti-Bianchini,[ab] Marco Crisma,[a] Cristina Peggion,[a] Alberto Bianco*[b] and Fernando Formaggio*[a]

Received (in Cambridge, UK) 18th December 2008, Accepted 30th March 2009
First published as an Advance Article on the web 30th April 2009
DOI: 10.1039/b822789f

Rigid peptide backbones and backbone-to-side chain H-bonds permit the design of α-nucleopeptides with known 3D-structure; thymine–thymine base pairing is also observed.

RNA interference (RNAi) approaches are currently widely investigated for gene therapy applications.[1] The goal is to design molecules able to prevent the translation of undesirable proteins upon binding to specific *m*RNA. To this aim, synthetic oligonucleotides are of little use because of their limited resistance to enzymatic hydrolysis and their poor ability to cross biological membranes.

To overcome these drawbacks, different solutions have been proposed, mostly based on phosphate and/or sugar modifications, while preserving the natural nucleobases. Examples are the recently proposed homo-oligomers of nucleoside-derived amino acids[2] or the classic peptide nucleic acids (PNAs).[3] The latter have gained much attention as they are insensitive to phosphatase hydrolytic actions, while retaining (or even improving) polynucleotide complexation properties. Chiral PNA[4] and oxy-PNA[5] enjoy interesting applications, although their low ability to permeate membranes enables so far only *in vitro* uses. A polyamide-based alternative to PNAs is represented by nucleopeptides,[6,7] *i.e.* peptides constituted by nucleobase-containing amino acids.[8] In the case of α-nucleopeptides, the backbone perfectly matches that of a natural α-peptide, thus allowing for an easier prediction of their 3D-structure.[7,9] This observation stimulated us to undertake a research program aimed at building α-nucleopeptides with predictable and well-defined spatial distributions of their nucleobases.

By taking advantage of our long-standing experience in the synthesis and 3D-characterisation of conformationally constrained α-amino acids and peptides, particularly those based on C^{α}-tetrasubstituted α-amino acids,[10] we designed and synthesised short, β-turn- or helix-forming, α-nucleopeptides. Specifically, alanyl–thymine (AlaT, Fig. 1) residues were inserted into host, homo-peptide stretches of α-aminoisobutyric acid (Aib, Fig. 1). The primary structure of the longest peptide, a hexamer (Fig. 1), was designed to allow the alignment of two nucleobases on the same face of the helical molecule.

[a] *Institute of Biomolecular Chemistry, Padova Unit, CNR, Department of Chemistry, University of Padova, 35131, Padova, Italy. E-mail: fernando.formaggio@unipd.it*
[b] *CNRS, Institut de Biologie Moléculaire et Cellulaire, Laboratoire d'Immunologie et Chimie Thérapeutiques, 67084, Strasbourg, France. E-mail: a.bianco@ibmc.u-strasbg.fr*

† Electronic supplementary information (ESI) available: Synthesis, spectroscopic and X-ray data. CCDC 714103. For ESI and crystallographic data, see DOI: 10.1039/b822789f

The synthesis of *Z*-L-AlaT-OH was achieved according to literature procedures,[6,11] starting from the N^{α}-protected L-Ser β-lactone. In our hands, the method described in ref. 11 gave less racemisation, as proved by the higher optical rotation values measured. This finding was confirmed by comparing the HPLC chromatograms of two samples of *Z*-(Aib-L-AlaT-Aib)$_2$-O^{t}Bu (Fig. 1), synthesised from nucleoamino acids prepared according to two different methods (ESI, Fig. S1 and S2).†

Due to the low reactivity of Aib in peptide bond formation, particularly when two or more residues of this type occur consecutively, the nucleopeptides were synthesised in solution. The sterically hindered peptide bond between *Z*-L-AlaT-OH and H-Aib-O^{t}Bu was successfully formed by means of the very effective 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide (EDC)/7-aza-1-hydroxy-1,2,3-benzotriazole (HOAt) coupling method.[12] Catalytic hydrogenolysis of *Z*-L-AlaT-Aib-O^{t}Bu, followed by acylation with the symmetrical anhydride from *Z*-Aib-OH yielded the protected tripeptide *Z*-Aib-L-AlaT-Aib-O^{t}Bu. The hexapeptide *Z*-(Aib-L-AlaT-Aib)$_2$-O^{t}Bu was prepared in satisfactory yield by condensing the 5(4*H*)-oxazolone from *Z*-Aib-L-AlaT-Aib-OH (generated in turn by acidolysis of its *tert*-butyl ester) with H-Aib-L-AlaT-Aib-O^{t}Bu (obtained by catalytic hydrogenolysis of its N^{α}-protected precursor) in DMF at room temperature.

A conformational NMR investigation was performed on *Z*-Aib-L-AlaT-Aib-O^{t}Bu in $CDCl_3$ solution. The presence of sequential NH–NH cross-peaks in the ROESY spectrum (Fig. 2, left) suggests that this nucleopeptide adopts a folded conformation in this solvent. The chemical shift behaviour of the four NH protons upon DMSO addition (Fig. 2, right) indicates solvent-shielding for the α-NHs of AlaT2 and Aib3. Whereas this result is expected for the Aib3 NH if involved in a β-turn, it is somehow surprising for the AlaT α-NH.

To better understand this behaviour we carried out an IR absorption analysis in $CDCl_3$ solution in the concentration range 10–0.1 mM. The NH stretching region (3500–3200 cm^{-1}, amide A) allows one to discriminate between free (solvated) and H-bonded amide NHs (vibration frequency above and below 3400 cm^{-1}, respectively).[13] Self-association is observed at high peptide concentration. At the lowest concentration, the

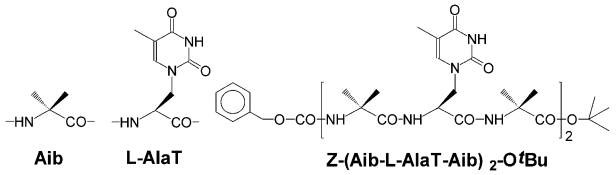

Fig. 1 Chemical structures of Aib, L-AlaT and the longest nucleopeptide synthesised.

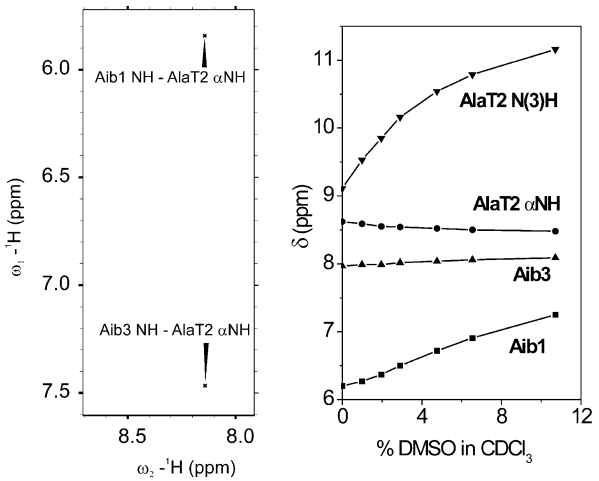

Fig. 2 Portion of the ROESY spectrum (left) of *Z*-Aib-L-AlaT-Aib-O^{t}Bu (1 mM in $CDCl_3$ solution) and chemical shift dependence (right) of its NH protons upon addition of increasing amounts of DMSO.

spectrum of the nucleo-tripeptide shows three bands centred at about 3430, 3385 and 3350 cm^{-1} (ESI, Fig. S3).† To facilitate data interpretation, we also synthesised the corresponding nucleo-tripeptide allylated at the N(3)H of the thymine ring. Its spectrum (ESI, Fig. S4)† displays only two bands, at about 3430 cm^{-1} and 3350 cm^{-1}, respectively. Therefore, the band at about 3385 cm^{-1} in the spectrum of *Z*-Aib-L-AlaT-Aib-O^{t}Bu is likewise due to the nucleobase NH. In addition, the absence of aggregation in the allylated tripeptide suggests a fundamental role of the nucleobase in the self-association process.

The band at about 3350 cm^{-1} in the spectrum of *Z*-Aib-L-AlaT-Aib-O^{t}Bu and its allylated analogue might be associated, at least in part, with the Aib3 NH, H-bonded in a β-turn, as it is present also in the spectrum of *Z*-Aib-L-Ala-Aib-OMe[14] (ESI, Fig. S4).† Furthermore, the H-bonded band is much more intense for the allylated AlaT tripeptide than for the Ala analogue. As inferred from the NMR analysis, this observation suggests that the α-NH of AlaT2 as well is intramolecularly H-bonded to a significant extent.

Luckily enough, we were able to obtain a single crystal of *Z*-Aib-(D,L)-AlaT-Aib-O^{t}Bu, suitable for X-ray diffraction analysis.†‡[15] The peptide backbone is folded into a type-I (I′) β-turn (Fig. 3, top). This outcome, not unexpected for Aib-rich peptides,[10] is in agreement with the IR absorption and NMR analyses in solution. However, besides the intramolecular H-bond (from the Aib3 NH to the *Z* carbonyl) stabilizing the β-turn, a second intramolecular H-bond between the α-NH of AlaT2 and the C(2)═O of its thymine moiety is observed. This finding, explaining also the IR and NMR results described above, is of particular significance because such a backbone-to-side chain H-bond possibly reduces the nucleobase mobility, thus allowing for a fine tuning of its spatial arrangement.

The crystal state packing mode (Fig. 3, bottom) reveals an additional interesting feature: dimerisation through thymine–thymine base pairing. This behaviour is probably responsible for the remarkable tendency of these nucleopeptides to self-associate, as observed also in our IR absorption analysis. Unfortunately, only for the hexamer was a dimeric structure

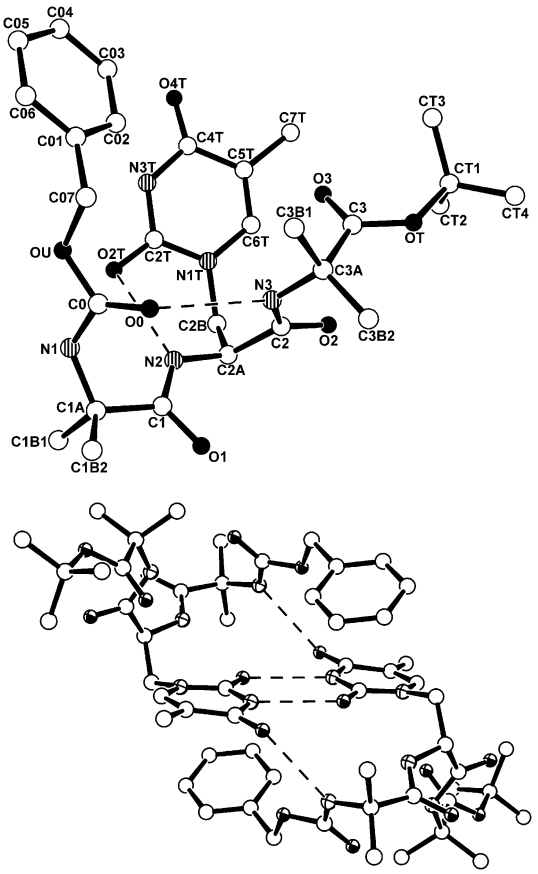

Fig. 3 Top: X-ray diffraction structure of *Z*-Aib-(D,L)-AlaT-Aib-O^{t}Bu with atom numbering. The L-enantiomer is shown. H-atoms have been omitted for clarity. Bottom: crystal packing mode.[15] Intra- and intermolecular H-bonds are represented by dashed lines.

clearly assigned by NMR in $CHCl_3$ solution (see below), whereas we were unable to unambiguously draw the same conclusion for the tripeptide.[16]

Thymine–thymine association appears to be a common tendency in these nucleopeptides, not just limited to enantiomeric pairs as in the crystal state. Indeed, we often observed in the mass spectra of AlaT-containing nucleopeptides small, but significantly intense, peaks at $m/z = [2M + H]^+$, indicative of relevant dimer stability (ESI, Fig. S6).†

The IR spectra of the hexapeptide *Z*-(Aib-AlaT-Aib)$_2$-O^{t}Bu at different concentrations show remarkable contributions from intermolecular H-bonds (ESI, Fig. S5).† However at the lowest concentration examined (0.05 mM), the strong H-bonded N–H band is largely due to intramolecularly bonded NHs. Taking also into account the position of this band (3324 cm^{-1}), we can tentatively conclude that the hexapeptide folds into a helical structure.

Preliminary CD data show that in MeOH solution the thymine absorption band (~270 nm, induced CD) is about twice as intense in the hexapeptide (two thymines) as compared to the tripeptide (one thymine). However, in $CHCl_3$, the solvent used for the IR absorption and NMR analyses, the same band in the hexapeptide is about five times as intense as that of the tripeptide (ESI, Fig. S7).† We are inclined to attribute the observed phenomenon to a stronger propensity

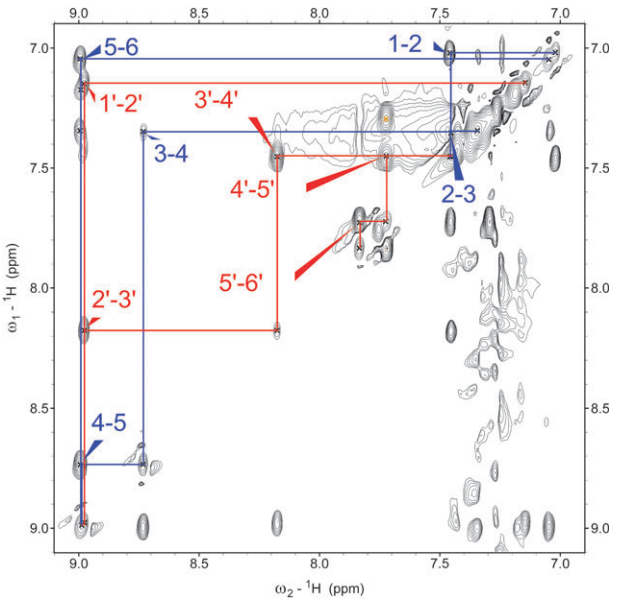

Fig. 4 $NH_i \rightarrow NH_{i+1}$ sequential NOEs in the ROESY spectrum of *Z*-(Aib-L-AlaT-Aib)$_2$-O^{t}Bu (5 mM in $CDCl_3$). For clarity, only the residue numbers of the two molecules (one is primed) are indicated (see ESI for details).†

to nucleobase-mediated self-association[17] of the hexapeptide, occurring in solvents of low polarity. However, more extensive experimental and theoretical investigations are needed to fully understand this phenomenon.

The ^{1}H NMR analysis of the hexa-nucleopeptide in $CDCl_3$ solution not only confirms the IR and CD findings on self-aggregation, but also suggests the formation of a dimer. At 1 mM concentration, the spectrum shows twice as many NH signals (sharp peaks) as expected (ESI, Fig. S8, trace a),† three of them resonating at unusually high frequency ($\delta > 11$ ppm). Interestingly, upon DMSO addition (ESI, Fig. S8, 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. 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, while in the monomer part of them are exposed to the solvent and shift to lower fields when DMSO concentration increases.

2D-NMR homo- and hetero-correlated experiments, carried out in $CDCl_3$ solution, allowed us to assign all proton resonances (ESI, Table S4)† and to confirm this hypothesis: the signals in the 1D spectrum belong indeed to two distinct hexapeptide molecules. In the ROESY spectrum two sets of sequential $NH_i \rightarrow NH_{i+1}$ NOEs are observed (Fig. 4), thus indicating that both molecules adopt a helical structure.[18]

This finding suggests that the nucleobases protrude from the same side of the peptide helix, as expected for residues at $i, i + 3$ positions in Aib-rich 3_{10}-helical peptides. Indeed, intra-chain thymine–thymine connectivities are found (ESI, Table S5).† Such nucleobase alignment strongly favours cooperative H-bond mediated interactions, as proved by the stability of the dimer and the highly deshielded NHs (ESI, Table S4)† of the thymine rings. By combining this information with the presence of inter-chain, thymine–thymine cross peaks (ESI, Table S5),† it appears that our helical hexa-nucleopeptide forms a thymine-mediated parallel dimer.

To summarise, interesting features of these new nucleopeptides are: (i) the markedly rigid peptide backbones forming β-turn or helices, depending on main-chain length, (ii) the backbone-to-side chain H-bond reducing the nucleobase mobility and (iii) the strong tendency to form nucleobase-mediated dimers. Altogether, we believe that these features will allow us to design nucleopeptides with a precise spatial distribution of their nucleobases.[2] Such structured nucleopeptides might be able to force complementary polynucleotide chains to adopt unusual conformations, thus influencing their biological functions.[8]

We are currently extending to the other nucleobases the approach here illustrated for thymine.

This work was funded by a grant of the "Università italo-francese/Université franco-italienne" (VINCI 2005) to P.G.-B.

Notes and references

‡ Crystal data for *Z*-Aib-(D,L)-AlaT-Aib-O^{t}Bu: $C_{28}H_{39}N_5O_8$; M = 573.64, monoclinic, space group $P2_1/n$, a = 12.341(3), b = 17.198(4), c = 14.434(3) Å, β = 93.36(6)°, V = 3058.2(12) Å^3, T = 293(2) K. 5075 reflections measured, 4537 independent reflections (R_{int} = 0.031), R_1 = 0.0515 [$F \geq 4\sigma(F)$], wR_2 = 0.1756 (F^2 all data), goodness-of-fit on F^2 = 1.131.

1 C. G. Hebert, J. J. Valdes and W. E. Bentley, *Curr. Opin. Biotech.*, 2008, **19**, 500.
2 S. Chandrasekhar, G. P. K. Reddy, M. U. Kiran, Ch. Nagesh and B. Jagadeesh, *Tetrahedron Lett.*, 2008, **49**, 2969.
3 P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497.
4 T. Tedeschi, M. Chiari, G. Galaverna, S. Sforza, M. Cretich, R. Corradini and R. Marchelli, *Electrophoresis*, 2005, **26**, 4310.
5 M. Kitamatsu, M. Shigeyasu, T. Okada and M. Sisido, *Chem. Commun.*, 2004, 1208.
6 P. Lohse, B. Oberhauser, B. Oberhauser-Hofbauer, G. Baschang and A. Eschenmoser, *Croat. Chem. Acta*, 1996, **69**, 535.
7 U. Diederichsen, *Angew. Chem., Int. Ed. Engl.*, 1996, **35**, 445.
8 P. Geotti-Bianchini, J. Beyrath, O. Chaloin, F. Formaggio and A. Bianco, *Org. Biomol. Chem.*, 2008, **6**, 3661.
9 U. Diederichsen, D. Weicherding and N. Diesemann, *Org. Biomol. Chem.*, 2005, **3**, 1058.
10 C. Toniolo, M. Crisma, F. Formaggio and C. Peggion, *Biopolymers (Pept. Sci.)*, 2001, **60**, 396.
11 P. Chaltin, E. Lescrinier, T. Lescrinier, J. Rozanski, C. Hendrix, H. Rosemeyer, R. Busson, A. Van Aershot and T. Herdewijn, *Helv. Chim. Acta*, 2002, **85**, 2258.
12 L. A. Carpino, *J. Am. Chem. Soc.*, 1993, **115**, 4397.
13 G. M. Bonora, C. Mapelli, C. Toniolo, R. R. Wilkening and E. S. Stevens, *Int. J. Biol. Macromol.*, 1984, **6**, 179.
14 E. Benedetti, A. Bavoso, B. Di Blasio, V. Pavone, C. Pedone, C. Toniolo, G. M. Bonora and M. Crisma, *Int. J. Pept. Protein Res.*, 1983, **22**, 385.
15 Despite numerous attempts, we could grow single crystals only from a sample containing about 6% of *Z*-Aib-D-AlaT-Aib-O^{t}Bu. For this reason, two enantiomeric tripeptides (one containing L-AlaT and the other D-AlaT) are present in the crystal cell.
16 400 MHz NMR spectra were recorded, at 298 K, for various $CDCl_3$ solutions of *Z*-Aib-L-AlaT-Aib-O^{t}Bu, at concentrations up to the solubility limit (about 100 mg mL^{-1}).
17 A. Weiß and U. Diederichsen, *Eur. J. Org. Chem.*, 2007, 5531.
18 K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986, p. 162.