Supplementary Material (ESI) for Chemical Communications

Conformationally controlled, thymine-based α-nucleopeptides

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

[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*

SUPPORTING INFORMATION

Content

Reagents and materials

Solvents

Reagent-grade dimethyl formamide (DMF) purchased from Baker was stored on 3 Å molecular sieves in a closed flask under N_2; reagent-grade dichloromethane (DCM) purchased from Carlo Erba was distilled on phosphoric anhydride. Acetic acid (AcOH), 1-butanol, chloroform ($CHCl_3$), dichloromethane (DCM), diethyl ether (Et_2O), ethanol (EtOH), ethyl acetate (AcOEt), hydrochloric acid, 37% aqueous solution (HCl), methanol (MeOH), petroleum ether, boiling range 40-60°C (EP), 2-propanol (iPrOH), toluene (PhMe), trifluoroacetic acid (TFA) were reagent grade and used without further purification. Acetonitrile (MeCN) was HPLC grade and used without further purification.

Deuterated solvents (deuterochloroform, $CDCl_3$, d_6-dimethylsulphoxide, d_6-DMSO) were purchased from Cambridge Isotope Laboratories. Spectrophotometric grade methanol for polarimetric, circular dichroic and spectrophotometric characterization was purchased from Fluka.

Reagents

1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU), formic acid, iodine, *N*-methyl-morpholine (NMM), ninhydrin, phosphoric anhydride, sodium hydride, 60% dispersion in mineral oil, *N,N,N',N'*-tetramethyl-4,4'-diamino-diphenylmethane (TDM), were purchased from Aldrich. Calcium chloride anhydrous, potassium bromide, potassium carbonate, potassium hydrogenocarbonate, potassium hydrogenosulphate, sodium chloride, sodium sulphate anhydrous were purchased from Carlo Erba. α-aminoisobutyric acid (H-Aib-OH), was purchased from Fluka. Thymine was purchased from Lancaster. 1-hydroxy-7-aza-*1H*-benzotriazole (HOAt) was purchased by GL Biochem. *N*-benzyloxycarbonyl-serine (Z-Ser-OH), *N*-ethyl-*N'*-3-dimethylamino-propylcarbodiimide hydrochloride (EDC) were purchased from Iris Biotech. Silica gel for flash-chromatography (diameter 40-63 μm, mesh 230-400) was purchased from Merck.

N-benzyloxycarbonyl-serine β-lactone [Z-Ser(Lactone)] was prepared according to the literature [L. D. Arnold, T. H. Kalantar and J. C. Vederas, *J. Am. Chem. Soc.*, **107**, 7105 (1985)].

Instruments and characterization techniques

Melting points: Melting points (Mp) were measured on a Laborlux 12 Leitz apparatus and are uncorrected.

Polarimetry: Optical rotations were measured on a Perkin-Elmer 241 spectropolarimeter. 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).

Thin layer chromatography (TLC): Silica gel 60 F_{254} (Merck) on aluminum 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: 1-butanol/AcOH/H_2O 3:1:1; Rf_3: PhMe/EtOH 7:1.

Peptides were detected either by UV lamp irradiation or with exposition to I_2 vapors or by warming with a heat gun and spraying firstly with a 1.5% NaClO solution and then with a ninhydrin-TDM solution prepared according to literature procedures [E. Von Arx, M. Faupel and M. Brugger *J. Chromatogr.*, **120**, 224 (1976)].

Infrared spectroscopy (IR): Absorption infrared spectra (KBr pellets or $CDCl_3$ solutions) were performed on a Perkin-Elmer 1720 X FT-IR spectrophotometer, nitrogen flushed, equipped with a sample-shuttle device, at 2 cm^{-1} nominal resolution, averaging 100 scans. 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.

UV spectroscopy: UV-Vis spectra were recorded on a Shimadzu UV 2501PC spectrophotometer using a 1.0 cm optical path quartz cell.

HPLC: Analytical chromatograms were recorded on a Agilent Technologies 1200 Series instrument, provided with a Kromasil C18 Phenomenex (4.6×250 mm, 100 Å) column.

MilliQ H_2O has been used for eluent preparation.

Eluents: A: H_2O/MeCN 9:1 + 0.05% TFA; B: MeCN/H_2O 9:1 + 0.05% TFA.

Circular dichroism: Spectra were recorded on a Jasco J-715 dicrograph using Hellma quarz cells with an optical path length of 0.01 cm at room temperature.

NMR spectrometry: ^{1}H-NMR spectra for peptide characterization were recorded on a Bruker AC 200 spectrometer. ^{1}H-NMR spectra for NMR titrations and 2D-NMR spectra were recorded on DRX 400 or DMX 600 spectrometers at room temperature. Collected data were processed by the XwinNMR program with a Bruker X-32 elaborator.

Chemical shifts (δ) are expressed in parts per million (ppm) with respect to tetramethylsilane (TMS). Solvent residual peaks ($CHCl_3$, δ 7.26 ppm, or d_6-DMSO, δ 2.50 ppm) 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.

Mass spectrometry (MS): Mass spectra were recorded on 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.

Synthesis and characterization of peptides and nucleopeptides

Z-L-AlaT-OH

Method A (*Croat. Chem. Acta*, 1996, **69**, 535, ref. 5 in the main text)

1.51 g thymine (11.6 mmol) are dissolved in 230 mL anhydrous DMF. 1.22 g Z-Ser(Lactone) (5.5 mmol) are added, then 1.60 g K_2CO_3 are 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 thymine precipitates, which is filtered off. The oil is taken up in MeOH/AcOEt and on standing overnight some thymine precipitates. The crude mixture is purified through repeated flash-chromatography (eluent: DCM/MeOH/AcOH).

Yield: 0.51 g (25 %). Mp> 250°C (dec.) $[\alpha]_D^{25}$ = -33° (c 0.5, MeOH). Rf_1: 0.0; Rf_2: 0.55; Rf_3: 0.0.

IR (KBr) 3415,1684, 1610, 1521 cm^{-1}. UV: ε=8960 $M^{-1}cm^{-1}$ at λ=271.4 nm (14 μM in MeOH).

NMR (d_6-DMSO): 11.12 (s br, 1 H, N(3)H thymine), 7.42 (s, 1 H, C(6)H thymine), 7.34-7.32 (m, 5 H, C_6H_5-Z), 6.70-6.66 (d, J=6.6 Hz, 1 H, αNH AlaT), 5.05-4.88 (dd, J=10, 12 Hz, CH_2-Z), 4.36-4.29 (d, J=14 Hz, 1 H, αCH), 4.08 (s br, 1 H, 1 βCH), 3.85 (s br, 1 H, 1 βCH), 1.79 (s, 3 H, C^5H_3 thymine) .

MS calcd $[M+H]^+$: 348.12; found: 348.14.

Method B (*Helv. Chim. Acta*, 2002, **85**, 2258, ref 10 in the main text)

1.07 g thymine (8.2 mmol) are dissolved in 75 mL anhydrous DMF under N_2. 0.30 g NaH (60% dispersion in mineral oil, 7.6 mmol) are added and the heterogeneous reaction mixture is stirred for two hours, then cooled to -78°C in an acetone/dry ice bath. 1.66 g Z-Ser(Lactone) (7.5 mmol) are 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 5% $KHSO_4$ added with plenty of NaCl and then with brine. They are concentrated to a deliquescent yellowish solid which is purified by repeated flash-chromatography ($CHCl_3$/MeOH/AcOH).

Yield: 25%. 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, 1 H, N(3)H thymine), 7.57-7.53 (d, J=8.2 Hz, αNH AlaT), 7.39 (s, 1 H, C(6)H thymine), 7.35 (m, 5 H, C_6H_5-Z), 5.04 (s, 2 H, CH_2-Z), 4.37-4.22

(m, 2 H, αCH + 1 βCH AlaT), 3.67-3.55 (dd, J=9.7, 13.2 Hz, 1 H, 1 βCH AlaT), 1.71 (s, 3 H, C^5H_3 thymine).

Z-L-AlaT-Aib-O^{t}Bu

0.50 g Z-AlaT-OH (1.4 mmol) are dissolved in ~6 mL anhydrous DMF, they are cooled to 0°C in a water/ice bath then 0.224 g HOAt (1.6 mmol) and 0.327 g EDC·HCl (1.7 mmol) are added. 2.4 mmol H-Aib-O^{t}Bu (obtained by catalytic hydrogenolysis of 0.73 g Z-Aib-O^{t}Bu) are 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 5% $KHSO_4$, H_2O, 1:2 5% $NaHCO_3$/brine and H_2O. The solvent is evaporated and some product precipitates from MeOH. The rest of the crude is purified by flash-chromatography (13:9 AcOEt/EP +1% MeOH).

Yield: 40 %. Mp: 148-150°C. Rf_1: 0.60; Rf_2: 1.00; Rf_3: 0.45.

$[\alpha]_D^{25}$ = -38° (c 0.2, MeOH), method B. 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, 1 H, N(3)H thymine), 7.34 (m, 5 H, C_6H_5-Z), 7.16 (br s, 1 H, NH Aib), 7.04 (s, 1 H, C(6)H thymine), 6.20-6.16 (d, J=7.2 Hz, αNH AlaT), 5.11 (s, 2 H, CH_2-Z), 4.54-4.45 (m, 1 H, αCH AlaT), 4.22-4.13 (dd, J=3.7, 14.2 Hz, 1 H, 1βCH AlaT), 3.96-3.85 (dd, J=3.7, 14.2 Hz, 1 H, 1βCH AlaT), 1.85 (s, 3 H, C^5H_3 thymine), 1.49-1.47 (2 s, 6 H, 2 β-CH_3 Aib), 1.45 (s, 9 H, O^{t}Bu).

MS calcd $[M+H]^+$: 489.18; found: 489.22.

Method A $[\alpha]_D^{25}$ = -10° (c 0.5, MeOH)

Z-Aib-L-AlaT-Aib-O^{t}Bu

0.52 g $(Z\text{-Aib})_2O$ (1.2 mmol) are dissolved in 2 mL anhydrous MeCN, then added to 0.239 g (0.68 mmol) H-AlaT-Aib-O^{t}Bu (obtained by hydrogenolisis of 0.330 g Z-AlaT-Aib-O^{t}Bu). 0.125 g NMM (1.1 mol) are 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 H_2O, then desiccated on Na_2SO_4 and concentrated under vacuum. Pure product precipitates from AcOEt/petroleum ether, more product is obtained from the mother liquors by flash-chromatography (eluent: DCM/EtOH 20:1) .

Yield: 0.249 g (64 %). Mp 114-116°C (dec.) Rf_1: 0.55 Rf_2: 0.95 Rf_3: 0.40.

$[\alpha]_D^{25}$ = -43° (c 0.4, MeOH), method B. 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).

NMR ($CDCl_3$): 9.71 (s br, 1 H, N(3)H thymine), 8.23-8.20 (d, J=7.2 Hz, 1 H, αNH AlaT), 7.49 (s, 1 H, C(6)H thymine), 7.30 (m, 5 H, C_6H_5-Z), 7.04 (s br, 1 H, αNH Aib^3), 6.23 (s br, 1 H, αNH Aib^1), 5.02 (s, 2 H, CH_2-Z), 4.64 (m, 1 H, αCH AlaT), 4.03-4.01 (m, 2 H, βCH_2 AlaT), 1.88 (s, 3 H, C^5H_3 thymine) , 1.54 (s, 3H, 1 CH_3 Aib^3), 1.45-1.44 (s+s, 15 H, 2 CH_3 Aib^1 and 3 CH_3 O^{t}Bu), 1.38 (s, 3 H, 1 CH_3 Aib^3).

MS calcd[M+H]: 574.29; found: 574.30.

Method A $[\alpha]_D^{25}$ = -12° (c 0.5, MeOH)

Z-Aib-L-AlaT-Aib-OH

0.125 g Z-Aib-AlaT-Aib-O^{t}Bu (0.22 mmol) are 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. Rf_1: 0.05. Rf_2: 0.80. Rf_3: 0.05.

$[\alpha]_D^{25}$ = -38° (c= 0.6, MeOH), method B. IR (KBr): 3403, 1782, 1697,1676, 1528 cm^{-1}.

^{1}H-NMR ($CDCl_3$, 200 MHz), δ/ppm: 9.98 (s, 1 H, N(3)H thymine), 8.30-8.27 (d, J=6.8 Hz, 1 H, αNH AlaT), 7.68 (br s, 1 H, NH Aib^3), 7.46 (s, 1 H, C(6)H thymine), 7.29 (m, 5 H, C_6H_5-Z), 6.04 (br s, 1 H, NH Aib^1), 5.11-4.95 (2 d, J=12.4 Hz, 2 H, CH_2-Z), 4.72 (m, 1 H, αCH AlaT), 4.07 (m, 2 H, βCH_2 AlaT), 1.85 (s, 3 H, C^5H_3 AlaT), 1.54 (s, 3 H, 1 CH_3 1 Aib), 1.50 (s, 3 H, 1 CH_3 1 Aib), 1.47 (s, 3 H, 1 CH_3 1 Aib), 1.45 (s, 3H, 1 CH_3 1 Aib).

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

0.80 g H-Aib-AlaT-Aib-O^{t}Bu (0.18 mmol) (obtained by hydrogenolisis of 0.105 g Z-Aib-AlaT-Aib-O^{t}Bu) are dissolved in 1 mL DMF and added to a solution of 0.088 g (0.18 mmol) Z-Aib-AlaT-Aib·Oxl (obtained by the activation of 0.093 g Z-Aib-AlaT-Aib-OH with 0.045 g (0.23 mmol) EDC·HCl) in 1.5 mL DMF. 0.011 mL NMM (0.11 mmol) are 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/petroleum ether.

Yield: 0.102 g (61%). Mp: 156-157°C. Rf_1: 0.40. Rf_2: 0.80. Rf_3: 0.25.

$[\alpha]_D^{25}$ = -22° (c 0.4, MeOH), method B. IR (KBr): 3417, 1674, 1528 cm^{-1}.

UV: ε=12100 $M^{-1}cm^{-1}$ at λ=268.8 nm (11 μM in MeOH).

^{1}H-NMR (*d_6*-DMSO, 200 MHz), δ/ppm: 11.32-11.30 (2 br s, 2 H, 2 N(3)H of 2 thymines), 8.60-8.57 (d, J=6.8 Hz, 1 H, αNH AlaT), 7.99 (s, 1 H, αNH Aib), 7.86 (s, 1 H, αNH Aib), 7.72-7.60 (2 s + d, J=8 Hz, 3 NH, 2 C(6)H of 2 thymines and αNH AlaT), 7.37 (m, 5 H, C_6H_5-Z), 7.26 (s, 1 H, αNH Aib), 7.19 (s, 1 H, αNH Aib), 5.15-4.99 (2 d, J=12.4 Hz, 2 H, CH_2-Z), 4.51-4.25 (m, 4 H, 2 αCH and 2 βCH of 2 AlaT), 3.89-3.77 (dd, J=10.4, 13 Hz, 1 H, 1 βCH AlaT), 3.67-3.55 (m, 1 H, 1 βCH AlaT), 1.72 (s, 3 H, 1 C^5H_3 AlaT), 1.70 (s, 3 H, 1 C^5H_3 AlaT), 1.37-1.23 (m, 33 H, 8 CH_3 of 4 Aib and 3 CH_3 O^{t}Bu).

Method A HPLC (44-46% B in 30 min): t_r 18.84 min (D,L+L,D) 9.5 %, t_r 18.75 min (L,L + D,D) 88.7 %.

Method B HPLC (44-46% B in 30 min): tr 18.90 min (D,L+L,D) 1.0 %, t_r 18.76 min (L,L + D,D) 99.0 %.

MS calcd[M+H]: 939.46; found: 939.42.

Z-Aib-L-AlaT(N3-allyl)-Aib-O^{t}Bu

0.059 g Z-Aib-AlaT-Aib-O^{t}Bu (0.10 mmol) are dissolved in 2.0 mL DCM, 16 μL DBU (0.10 mmol) and 16 μL AllBr (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 (DCM/MeOH 60:1 to 40:1) to yield an oil.

Yield: 0.017 g (27%). $[\alpha]_D^{25}$ = -23° (c 0.3, MeOH) Rf_1: 0.70. Rf_2: 0.90. Rf_3: 0.45.

IR (KBr): 3397, 1732, 1700, 1666, 1639, 1529 cm^{-1}.

NMR ($CDCl_3$), δ/ppm: 8.06-8.03 (d, J=7.0 Hz, 1 H, αNH AlaT), 7.45 (s, 1 H, C(6)H thymine), 7.34 (m, 5 H, 5 H, C_6H_5-Z), 7.27 (s, 1 H, αNH Aib), 5.90-5.75 (m, 1 H, C(2)H allyl), 5.25 (s, 1 H, αNH Aib), 5.23 (m, 1 H, 1 C(3)H allyl), 5.16 (m, 1 H, 1 C(3)H allyl), 5.11-5.00 (m, 2 H, CH_2-Z), 4.67-4.61 (m, 1 H, αCH AlaT), 4.51-4.49 (2 s, 2 H, βCH2 AlaT), 4.20-4.09 (m, 2 H, C(1)H_2 allyl), 1.90 (s, 3 H, C^5H_3 AlaT), 1.51 (s, 3 H, 1 CH_3 Aib), 1.47 (s, 3 H, 1 CH_3 Aib), 1.45 (s, 3 H, 1 CH_3 Aib), 1.43 (s, 9 H, 3 CH_3 O^{t}Bu), 1.39 ((s, 3 H, 1 CH_3 Aib).

HPLC: 10-70% B in 30 min: t_r 16.37 min.

MS calcd $[M\text{-}tBu]^+$: 556.24, found: 556.23.

Z-Aib-L-Ala-Aib-OMe [*Int. J. Pept. Protein Res.*, 1983, **22**, 385, ref. 13 in the main text].

0.33 g Z-Aib-OH (1.4 mmol) are dissolved in the minimum amount of distilled DCM, they are cooled to 0°C, then 0.19 g HOBt (1.4 mmol) and 0.31 g EDC·HCl (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 MeCN anhydrous and transferred into the reaction vessel. 0.40 mL NMM (3.6 mmol) are added in order to adjust the pH to 8-9 and the heterogeneous reaction mixture starts to dissolve. After 6 days the solvent is evaporated, the mixture is taken up in AcOEt and washed with 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 (DCM/MeOH 30:1).

Yield: 0.195 g (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}.

NMR ($CDCl_3$), δ/ppm: 7.34 (m, 5 H, C_6H_5-Z), 7.07 (s br, 1 H, αNH Aib[3]), 6.53-6.49 (d, J=8.0 Hz, 1 H, αNH Ala), 5.23 (s br, 1 H, αNH Aib[1]), 5.08 (s, 2 H, CH_2-Z), 4.49-4.38 (q, J=7.2, αCH Ala), 3.70 (s, 3 H, OMe), 1.52 (2 s, 6 H, 2 CH_3 Aib), 1.49 (2 s, 6 H, 2 CH_3 Aib), 1.34-1.31 (d, J=7.2 Hz, 3 H, CH_3 Ala).

HPLC analysis of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu diastereoisomeric purity

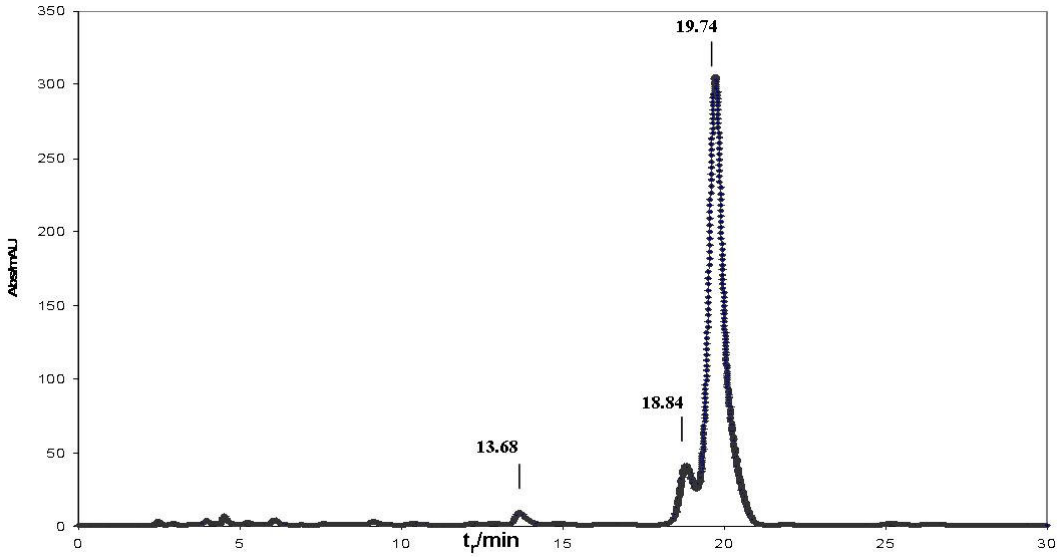


Fig. S1 Analytical chromatogram of the hexanucleopeptide synthesized using the nucleo amino acid obtained with protocol A. The peak at t_r 18.84 min belongs to the peptide epimers. Eluition gradient: 44-46% B in 30 min. Detection at λ=265 nm.

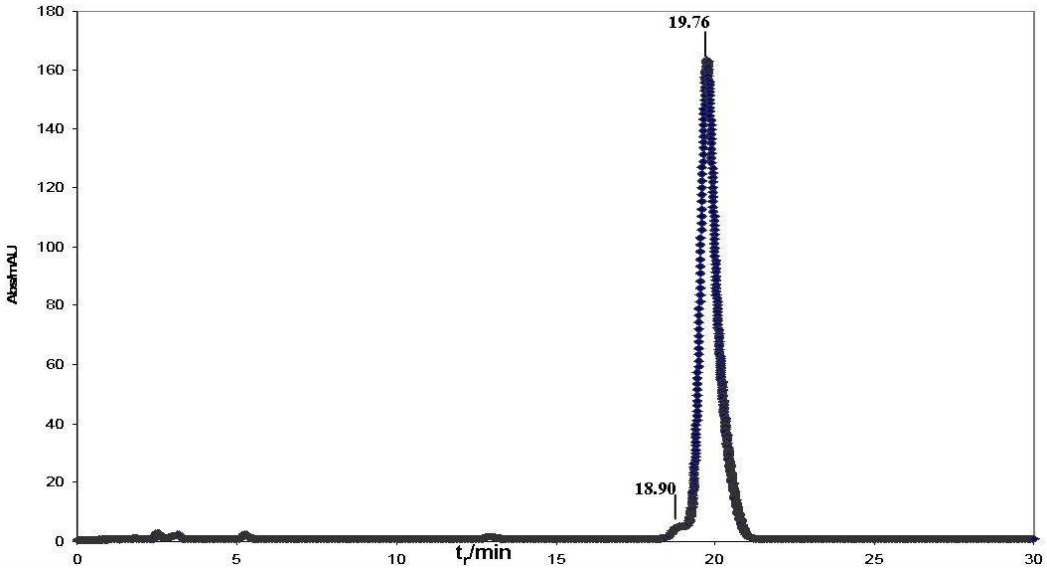


Fig. S2 Analytical chromatogram of the hexanucleopeptide synthesized using the nucleo amino acid obtained with protocol B. The peak belonging to the peptide epimers appears as a foot of the main peak at about t_r 18.90 min. Eluition gradient: 44-46% B in 30 min. Detection at λ=265 nm.

X-ray diffraction

Single crystals of Z-Aib-(D,L)-AlaT-Aib-O^{t}Bu were grown by slow evaporation from a methanol solution. 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 polarization effects, not for absorption.

The structure was solved by direct methods of the SIR 2002 program [M. C. Burla, M. Camalli, B. Carrozzini, G. L. Cascarano, C. Giacovazzo, G. Polidori, and R. Spagna, *J. Appl. Crystallogr.* **2003**, *36*, 1103.]. Refinement was carried out by full-matrix least-squares procedures on F^2, using all data, by application of the SHELXL-97 program [G. M. Sheldrick, *Acta Crystallogr.* **2008**, *A64*, 112-122.], 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.

Crystal data and structure refinement parameters are listed in Table S1. Relevant torsion angles and H-bond parameters are reported in Tables S2 and S3, respectively.

Table S1. Crystal data and structure refinement for Z-Aib-(D,L)-AlaT-Aib-O^{t}Bu.

Identification code	mc110	
Empirical formula	C28 H39 N5 O8	
Formula weight	573.64	
Temperature	293(2) K	
Wavelength	1.54178 Å	
Crystal system	Monoclinic	
Space group	$P2_1/n$	
Unit cell dimensions	a = 12.341(3) Å	α= 90°.
	b = 17.198(4) Å	β= 93.36(6)°.
	c = 14.434(3) Å	γ = 90°.
Volume	3058.2(12) Å^3	
Z	4	
Density (calculated)	1.246 Mg/m^3	
Absorption coefficient	0.764 mm^{-1}	
F(000)	1224	
Crystal size	0.45 x 0.35 x 0.20 mm^3	
Theta range for data collection	4.00 to 60.05°.	
Index ranges	-13<=h<=13, -1<=k<=19, 0<=l<=16	
Reflections collected	5075	
Independent reflections	4537 [R(int) = 0.0312]	
Completeness to theta = 60.05°	99.9 %	
Absorption correction	None	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	4537 / 0 / 367	
Goodness-of-fit on F^2	1.131	
Final R indices [I>2sigma(I)]	R1 = 0.0515, wR2 = 0.1524	
R indices (all data)	R1 = 0.0631, wR2 = 0.1756	
Extinction coefficient	0.0038(4)	
Largest diff. peak and hole	0.201 and -0.270 e.Å^{-3}	

Table S2. Selected torsion angles [°] for Z-Aib-(D,L)-AlaT-Aib-O^{t}Bu.

C02-C01-C07-OU	-78.4(3)
C06-C01-C07-OU	103.1(3)
C01-C07-OU-C0	140.9(3)
C07-OU-C0-N1	175.9(3)
OU-C0-N1-C1A	-176.6(2)
C0-N1-C1A-C1	-54.6(3)
N1-C1A-C1-N2	-35.0(3)
C1A-C1-N2-C2A	-171.7(2)
C1-N2-C2A-C2	-82.4(3)
N2-C2A-C2B-N1T	74.3(3)
C2A-C2B-N1T-C2T	-84.2(3)
C2A-C2B-N1T-C6T	103.8(3)
N2-C2A-C2-N3	-9.3(3)
C2A-C2-N3-C3A	179.7(2)
C2-N3-C3A-C3	52.6(3)
N3-C3A-C3-OT	-142.7(2)
C3A-C3-OT-CT1	-179.5(2)
C3-OT-CT1-CT4	-179.7(3)
C3-OT-CT1-CT3	60.8(4)
C3-OT-CT1-CT2	-63.0(4)

Table S3. Hydrogen bonds for Z-Aib-(D,L)-AlaT-Aib-O^{t}Bu [Å and °].

D-H...A	d(D-H)	d(H...A)	d(D...A)	<(DHA)
N3-H3...O0	0.86	2.28	3.117(3)	163.8
N2-H2...O2T	0.86	2.03	2.800(3)	147.9
N1-H1...O4T#1	0.86	2.12	2.971(3)	168.8
N3T-H3T...O2T#1	0.86	2.00	2.853(3)	172.0

Symmetry transformations used to generate equivalent atoms:
#1 -x,-y,-z+1

IR absorption spectra

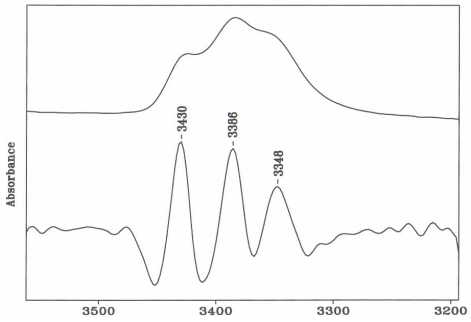


Fig. S3 IR-Absorption spectra in the amide A region and inverted second derivatives of Z-Aib-L-AlaT-Aib-O^{t}Bu at $1\cdot10^{-4}$ M concentration in $CDCl_3$ solution. Three bands at about 3430, 3385, 3350 cm^{-1} are present.

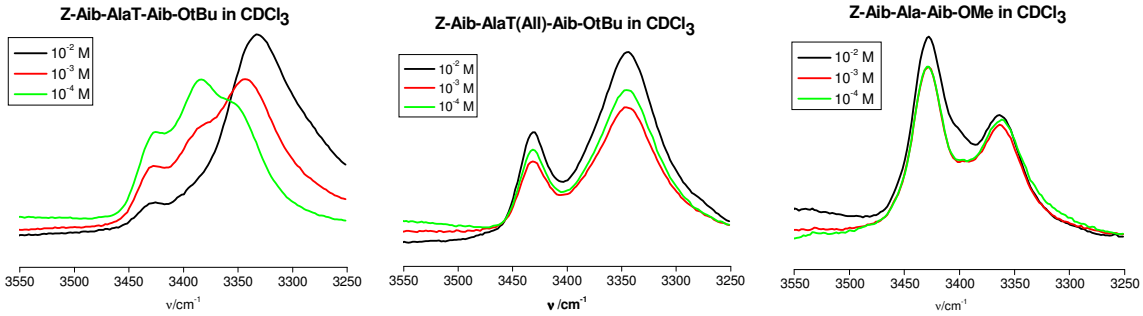


Fig. S4 IR-absorption spectra in the amide A region of Z-Aib-L-AlaT-Aib-O^{t}Bu (left), Z-Aib-L-AlaT(N3-allyl)-Aib-O^{t}Bu (middle) and Z-Aib-L-Ala-Aib-OMe (right) at $1\cdot10^{-2}$, $1\cdot10^{-3}$ and $1\cdot10^{-4}$ M concentrations in $CDCl_3$ solution. An increase of the cell path length compensates for every dilution. Comparing the left and middle spectra, the absence of the band at about 3385 cm^{-1} (see Fig. SI-3) and the strong decrease in the aggregation propensity are evident. Comparing the middle and right spectra, it appears that the relative intensity of the bands of the H-bonded NHs (at about 3350 cm^{-1}) with respect to the bands of the free NHs (at about 3430 cm^{-1}) is much larger in the former.

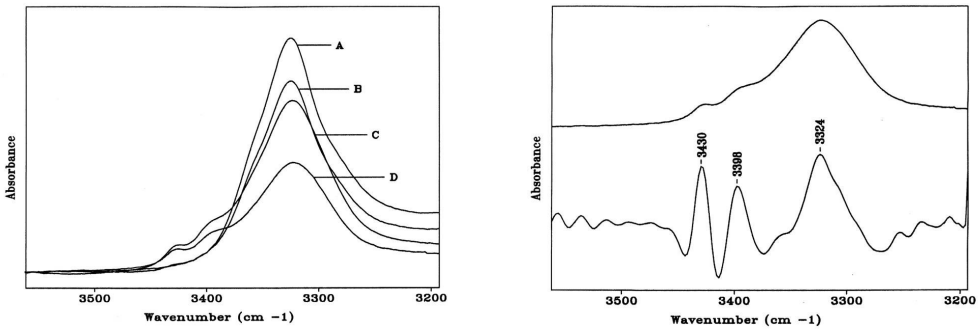


Fig. S5 Left: IR-absorption spectra in the amide A region of Z-(Aib-L-AlaT-Aib)$_2$-O^{t}Bu at $1\cdot10^{-2}$ M (**A**), $1\cdot10^{-3}$ M (**B**), $1\cdot10^{-4}$ M (**C**) and $5\cdot10^{-5}$ M (**D**) concentrations in $CDCl_3$ solution. Right: IR-absorption spectra and inverted second derivative at the highest dilution.

Mass spectrometry

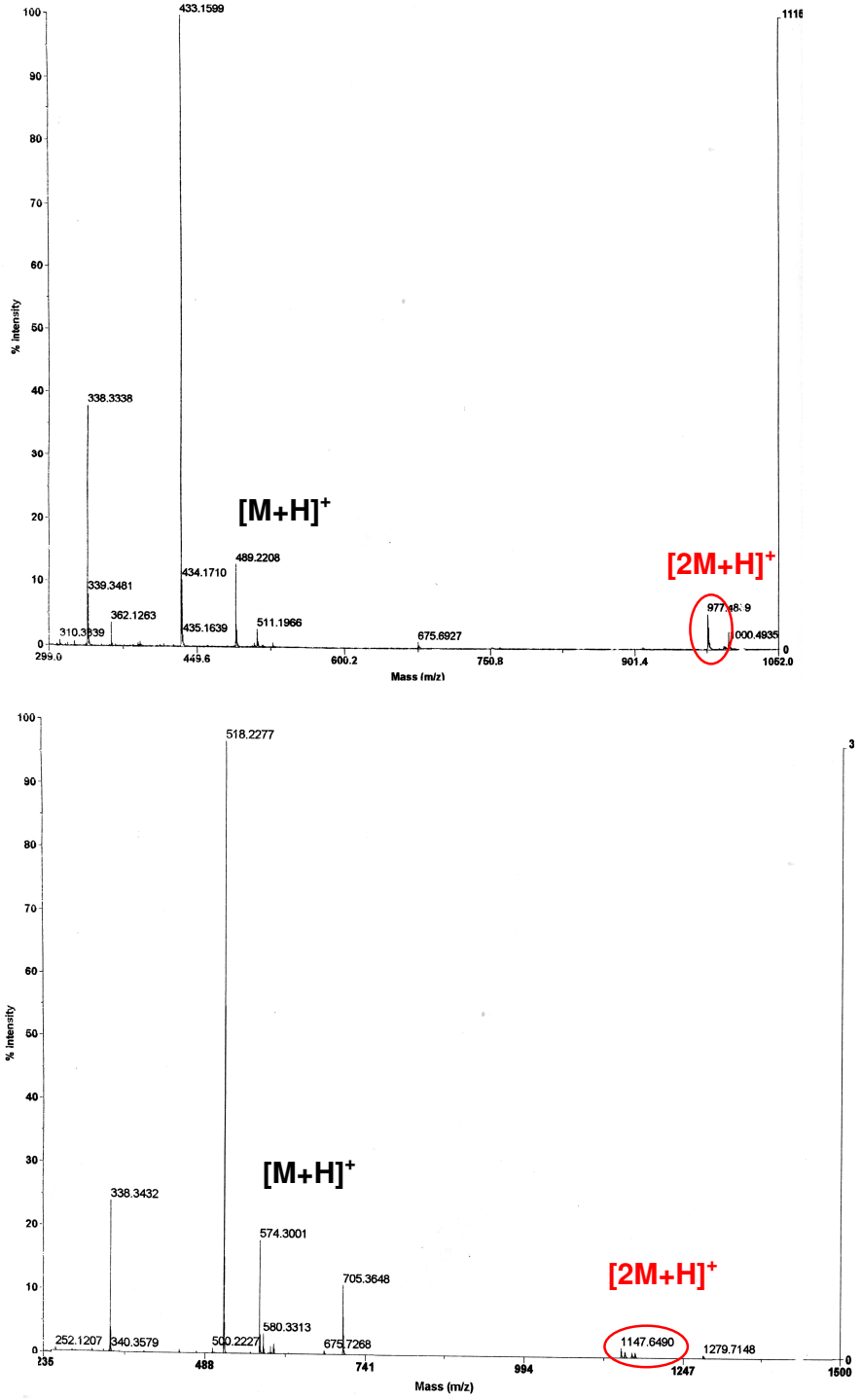


Fig. S6 Mass spectra of Z-L-AlaT-Aib-O^{t}Bu (top) and of Z-Aib-L-AlaT-Aib-O^{t}Bu (bottom), with dimer peak highlighted. Z-L-AlaT-Aib-O^{t}Bu: calcd $[M+H]^+ = 489.23$, $[2M+H]^+ = 977.46$, found $[M+H]^+ = 489.22$, $[2M+H]^+ = 977.49$ (circled peak). Z-Aib-L-AlaT-Aib-O^{t}Bu: calcd $[M+H]^+ = 574.29$, $[2M+H]^+ = 1147.56$, 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.

Circular dichroism spectra

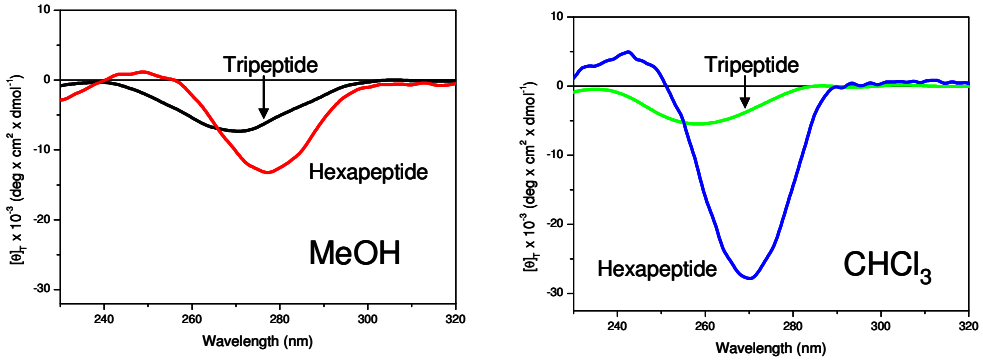


Fig. S7 CD spectra of Z-Aib-L-AlaT-Aib-O^{t}Bu and Z-(Aib-L-AlaT-Aib)$_2$-O^{t}Bu in the solvent indicated. Peptide concentration: 1 mM.

NMR ANALYSIS OF Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu

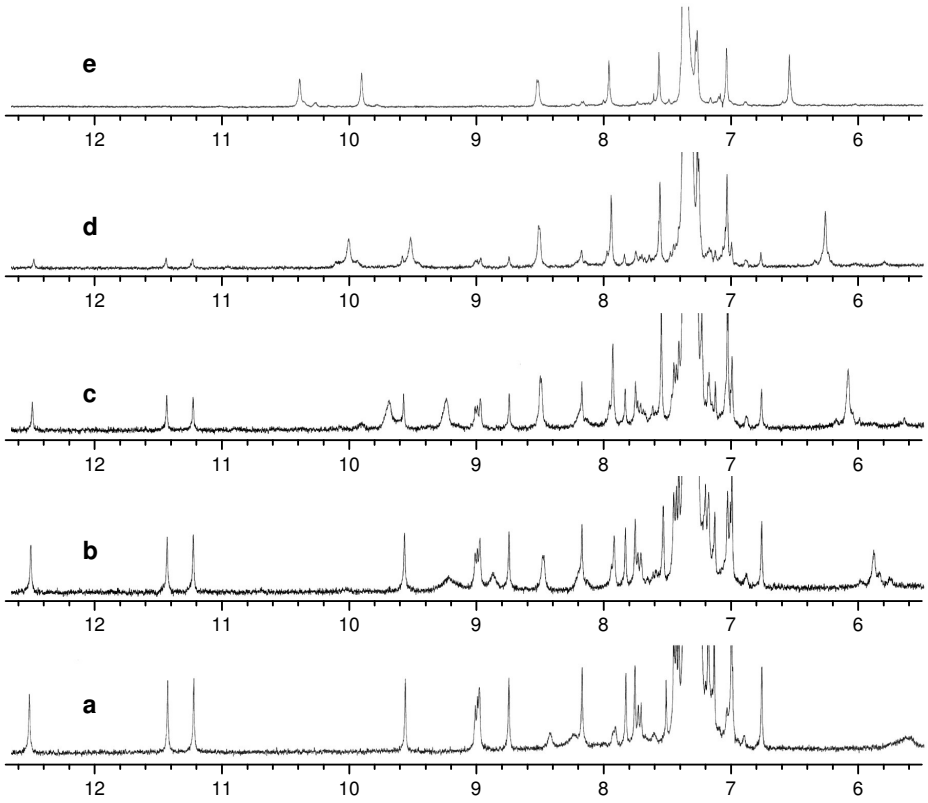


Fig. S8 NH region of the NMR spectrum of the hexapeptide Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu at 1 mM concentration in $CDCl_3$ solution (**a**) and upon addition of increasing amounts of DMSO (from bottom to top): 1% (**b**), 2% (**c**), 3% (**d**), 5% (**e**).

2D-NMR spectra used for assigning the Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu resonances

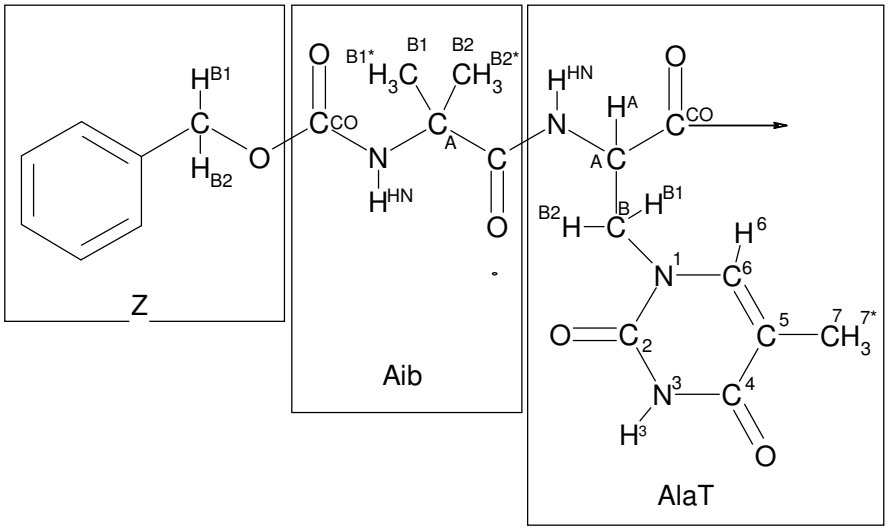


Fig. S9: Chemical structure showing the proton denominations used in the following Tables and spectra sections.

Table S4. Chemical shift assignments for the dimer of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu (5 mM in $CDCl_3$). The two dimer-forming molecules are, from now on, differently coloured. One of them is also primed.

	^{1}H chemical shifts (ppm)									^{13}C chemical shifts (ppm)					
Residue	**H6**	**H7***	**HA**	**HB1**	**HB2**	**HB1***	**HB2***	**HN**	**H3**	**C2**	**C4**	**C5**	**C6**	**C7**	**CA**
Z	-	-	-	4.843	5.29	-	-	-	-	-	-	-	-	-	-
Aib1	-	-	-	-	-	1.443	1.509	7.017	-	-	-	-	-	-	55.86
AlaT2	6.777	1.869	4.149	4.148	4.837	-	-	7.458	11.41	152.3	163.9	110.8	141.3	12.15	51.06
Aib3	-	-	-	-	-	1.499	1.499	7.356	-	-	-	-	-	-	56.1
Aib4	-	-	-	-	-	1.427	1.476	8.739	-	-	-	-	-	-	56.58
AlaT5	7.18	1.866	4.088	3.241	3.782	-	-	8.997	9.571	151.8	163.6	108.7	144.7	11.94	51.61
Aib6	-	-	-	-	-	1.148	1.325	7.047	-	-	-	-	-	-	55.59
Z'	-	-	-	5.089	5.153	-	-	-	-	-	-	-	-	-	-
Aib1'	-	-	-	-	-	1.38	1.521	7.151	-	-	-	-	-	-	55.9
AlaT2'	7.756	1.284	4.099	3.696	4.378	-	-	8.981	12.48	152.9	166.2	110.2	143.4	10.45	60.91
Aib3'	-	-	-	-	-	1.447	1.447	8.175	-	-	-	-	-	-	57.14
Aib4'	-	-	-	-	-	1.513	1.625	7.45	-	-	-	-	-	-	57.41
AlaT5'	7.297	1.539	4.353	3.737	4.473	-	-	7.726	11.21	150.2	166.3	110.7	140.2	9.799	54.54
Aib6'	-	-	-	-	-	1.444	1.571	7.834	-	-	-	-	-	-	56.42

* Three hydrogen atoms have the same chemical shift.

TOCSY experiment

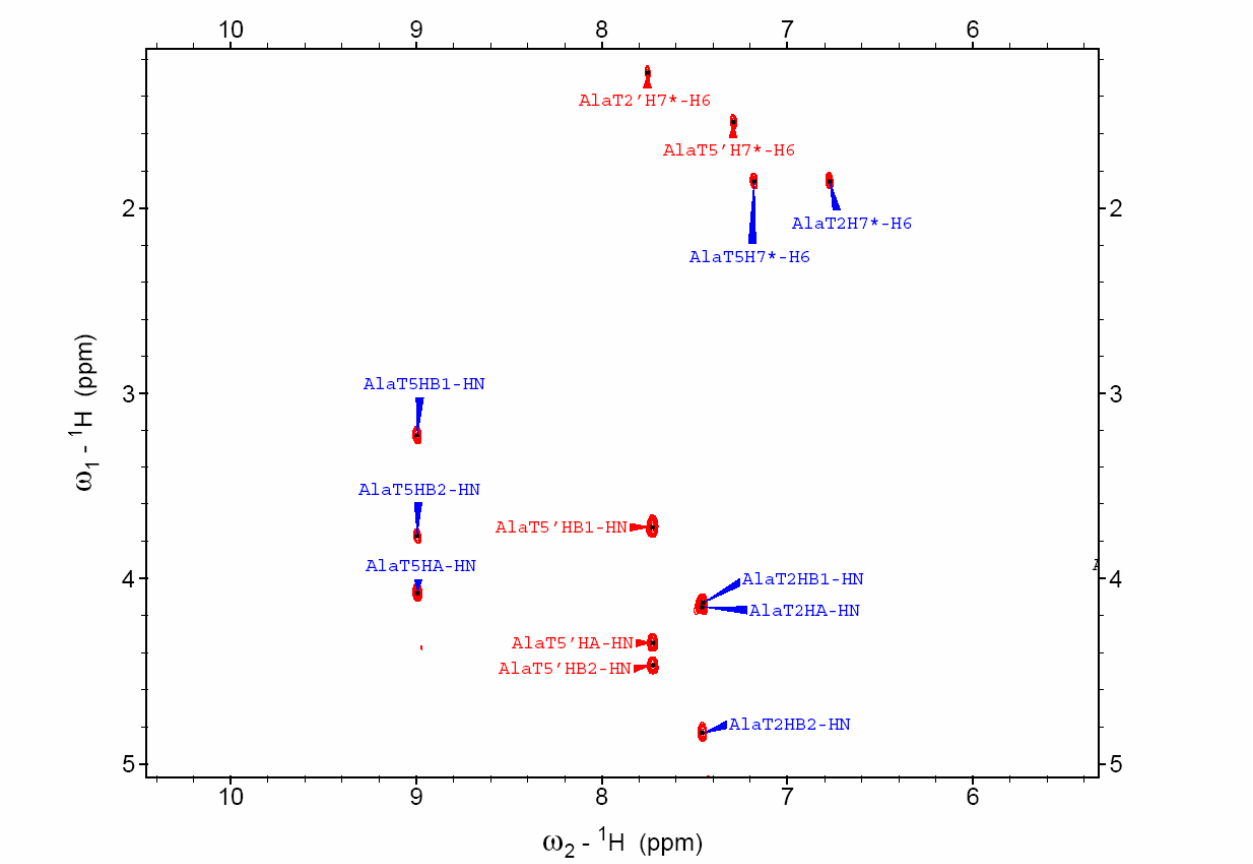
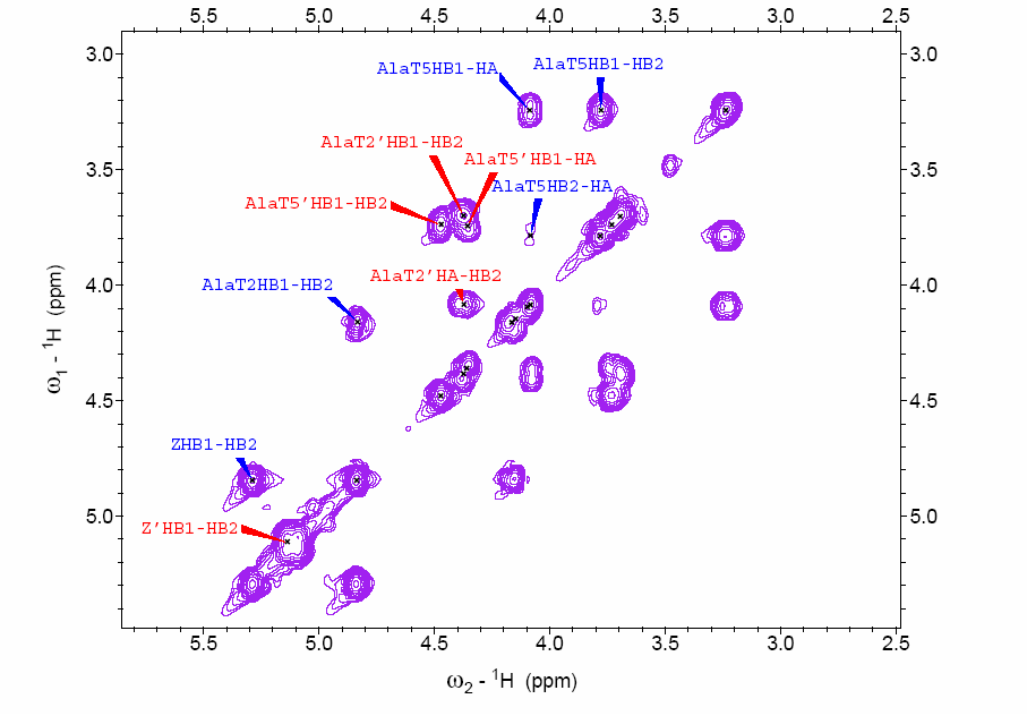
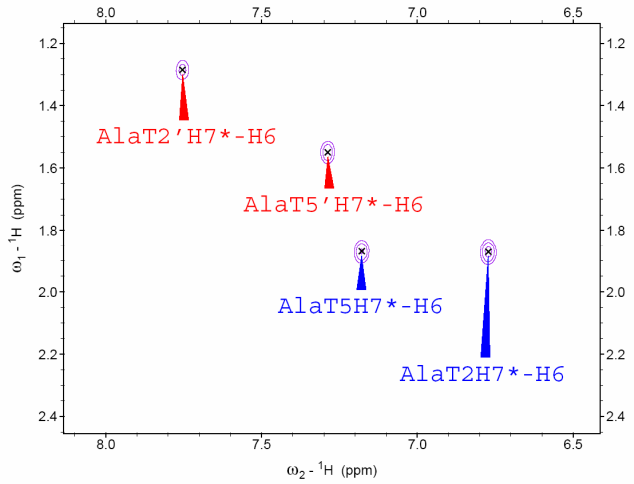


Fig. S10 TOCSY spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu (5 mM in $CDCl_3$). The NH-αCH, NH-βCH and H6-H7* regions are shown. Cross peaks are indicated in blue for "non-primed" peptide and in red for "primed peptide".

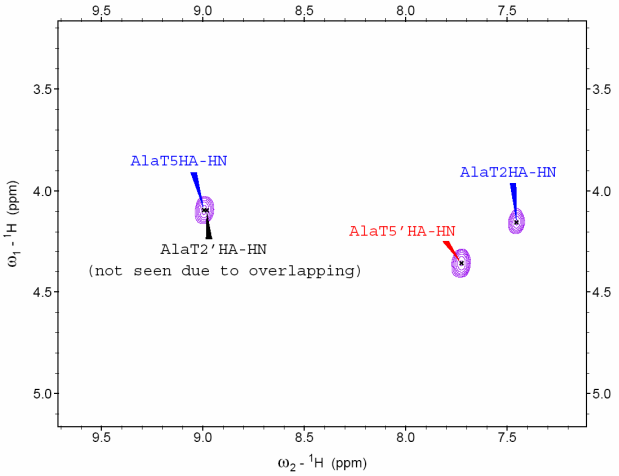
COSY experiment



COSY spectrum-A: αCH-βCH region.



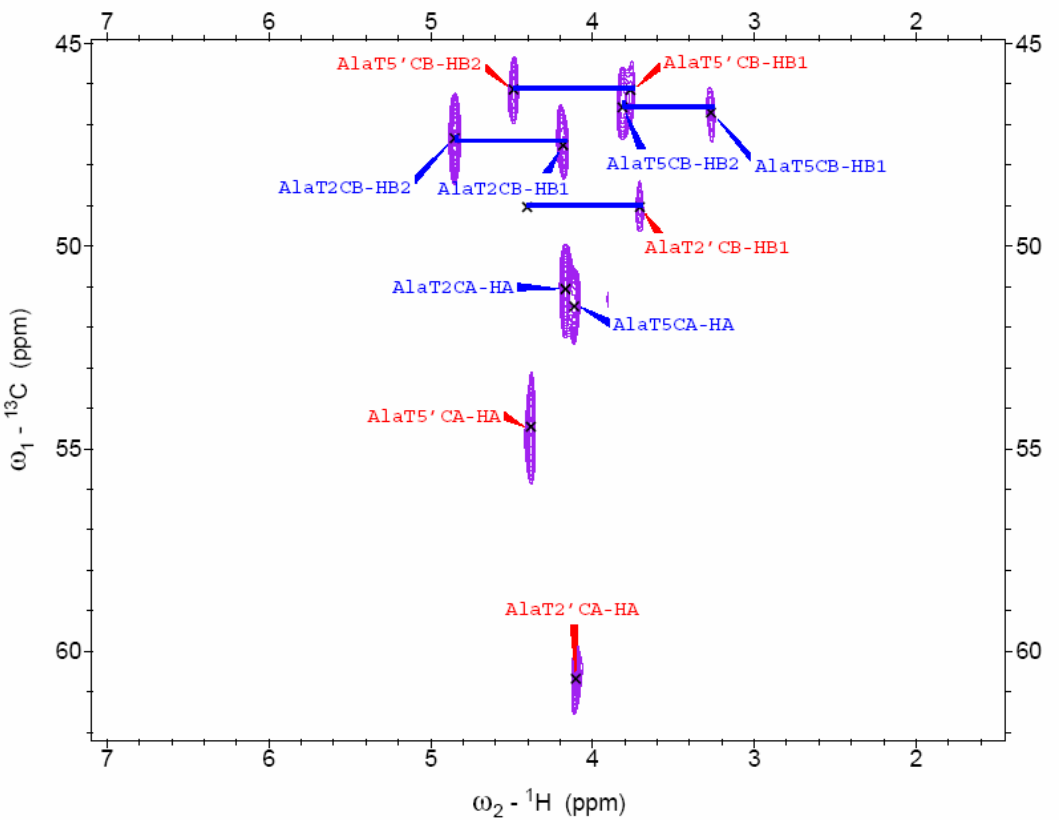
COSY spectrum-B: H6-H7* region.



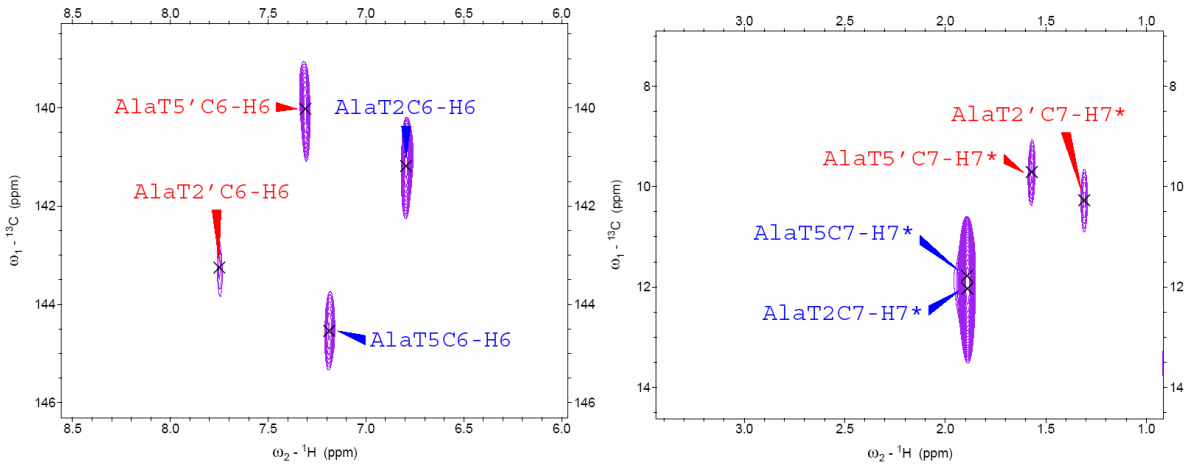
COSY spectrum-C: αCH-NH region.

Fig. S11 Sections of the COSY spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu (5 mM in $CDCl_3$). Cross peaks are indicated in blue for "non-primed" peptide and in red for "primed peptide"

HMQC experiment



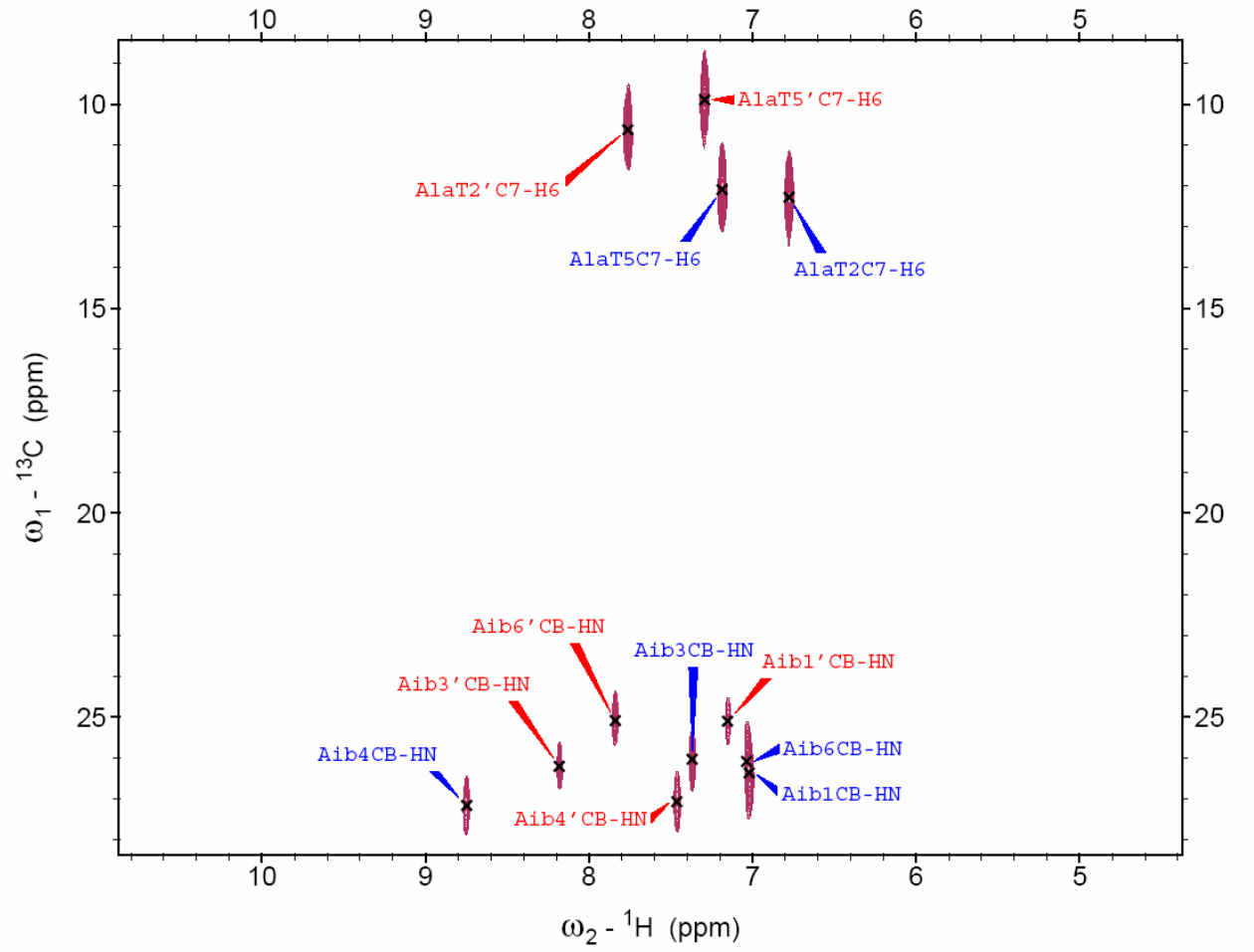
HMQC-A: αCH-C$^{\alpha}$ and βCH-C$^{\beta}$ region.



HMQC-B: H6-C6 region. **HMQC-C:** H7*-C7 region.

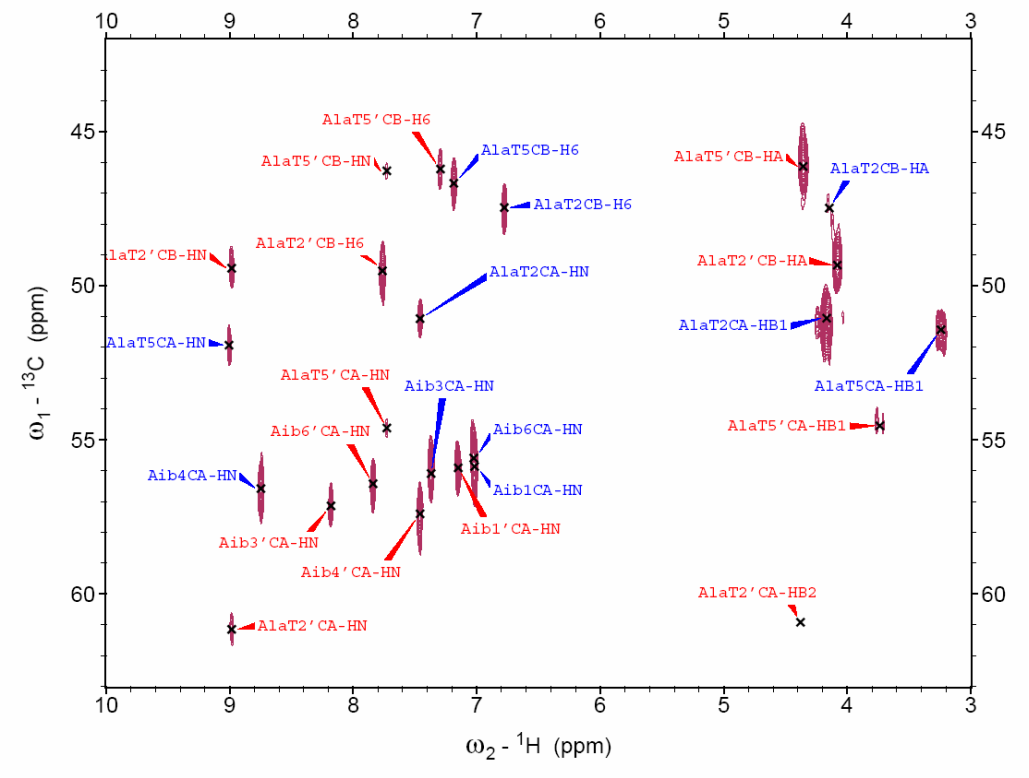
Fig. S12 Sections of the HMQC spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu (5 mM in $CDCl_3$). Cross peaks are indicated in blue for "non-primed" peptide and in red for "primed peptide".

HMBC experiment

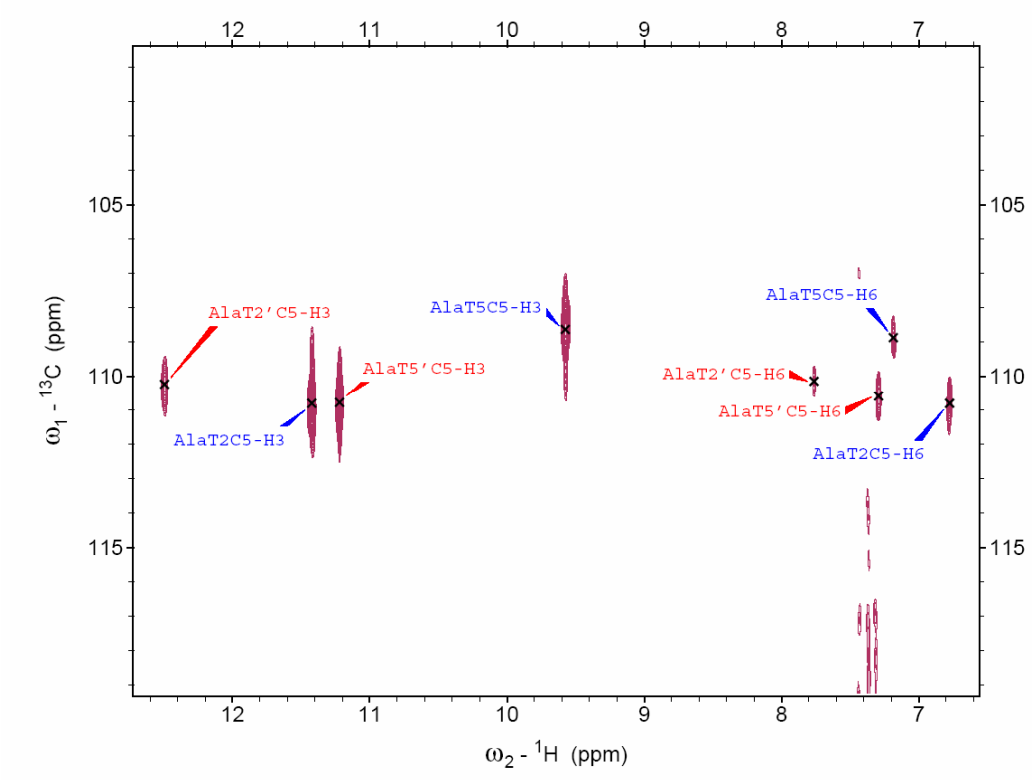


HMBC-A: NH-C^{β} and H6-C7 regions.

Fig. S13 Section of the HMBC spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu (5 mM in $CDCl_3$). Cross peaks are indicated in blue for "non-primed" peptide and in red for "primed peptide".

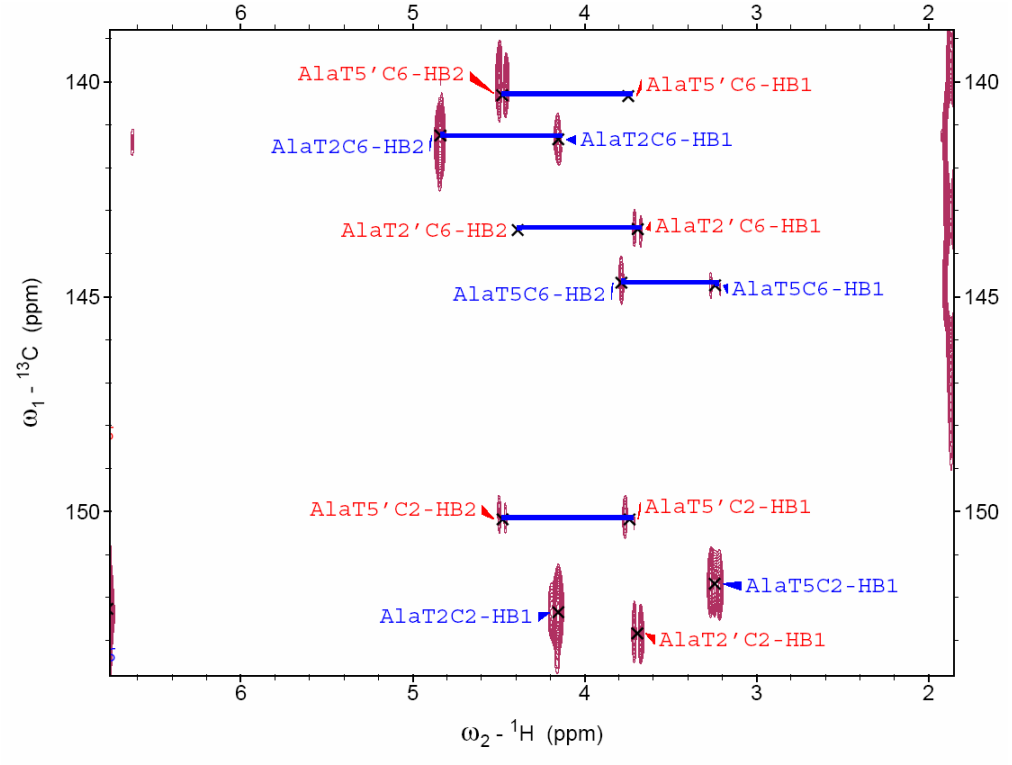


HMBC-B: NH-C$^{\alpha}$, NH-C$^{\beta}$, αCH-C$^{\beta}$, βCH-C$^{\alpha}$ and H6-C$^{\beta}$ regions.

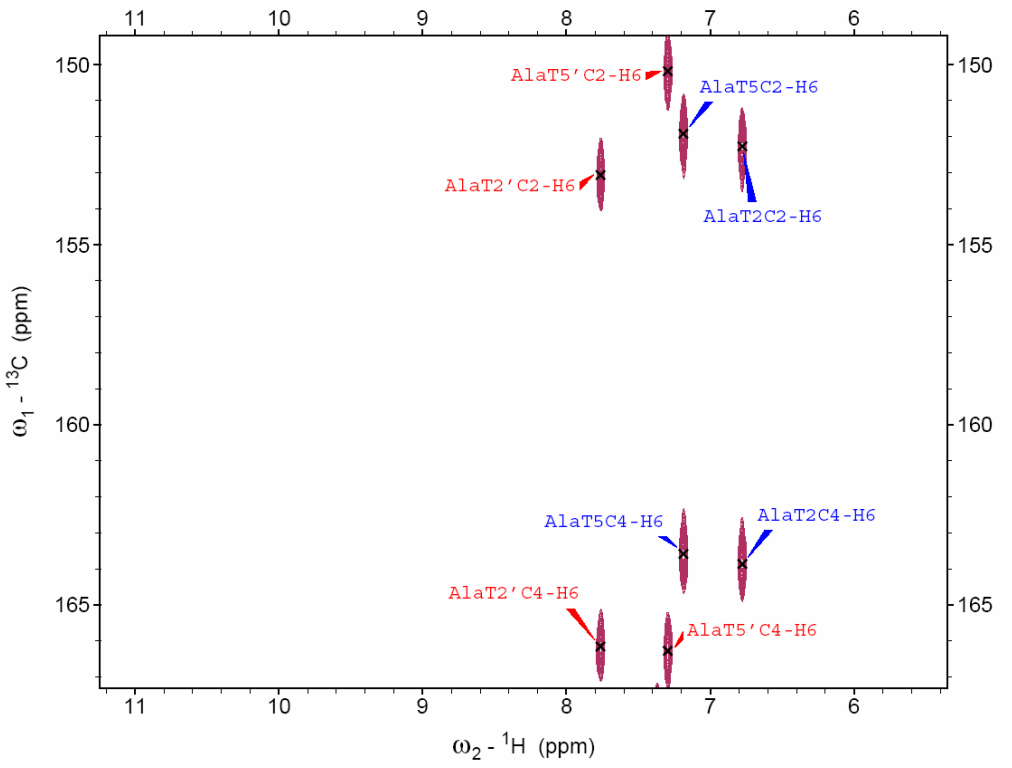


HMBC-C: H3-C5 and H6-C5 regions.

Fig. S14 Sections of the HMBC spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu (5 mM in $CDCl_3$). Cross peaks are indicated in blue for "non-primed" peptide and in red for "primed peptide".



HMBC-D: βCH-C2 and βCH-C6 regions.



HMBC-E: H6-C2 and H6-C4 regions.

Fig. S15 Sections of the HMBC spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu (5 mM in $CDCl_3$). Cross peaks are indicated in blue for "non-primed" peptide and in red for "primed peptide".

2D-NMR spectra used for structure analysis in solution

Table S5. NOE signals for the dimer of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu (5 mM in $CDCl_3$) extracted from the ROESY spectrum. Indications for one dimer-forming molecule are in blue, for the second molecule in red and primed. Additional colours are used to indicate different types of NOEs (as described in the table).

NOE signals* from ROESY spectrum		
Type of NOE	**Peptide**	**Peptide'**
sequential **$NH_i \rightarrow NH_{i+1}$** In Figure ROESY-A: blue for "non-primed" peptide red for "primed-peptide"	NH Aib1 → NH AlaT2 --- → --- (close to diagonal) NH Aib3 → NH Aib4 NH Aib4 → NH AlaT5 NH AlaT5 → NH Aib6	NH Aib1' → NH AlaT2' NH AlaT2' → NH Aib3' NH Aib3' → NH Aib4' NH Aib4' → NH AlaT5' NH AlaT5' → NH Aib6'
sequential **$\alpha CH_i \rightarrow NH_{i+1}$** **$\beta CH_i \rightarrow NH_{i+1}$** In Figure ROESY-B: turquoise for "non-primed" peptide magenta for "primed-peptide"	HA AlaT2 → NH Aib3 HB1 AlaT2 → NH Aib3 HA AlaT5 → NH Aib6 HB1 AlaT5 → NH Aib6	HA AlaT2' → NH Aib3' HB1 AlaT2' → NH Aib3' HB2 AlaT2' → NH Aib3' HA AlaT5' → NH Aib6' HB1 AlaT5' → NH Aib6' HB2 AlaT5' → NH Aib6'
intra-residue **αCH→H6** **βCH→H6** **NH→H6** **H3→H6** In Figure ROESY-D: orange for both peptides	HA AlaT2 → H6 AlaT2 HA AlaT5 → H6 AlaT5 HB1 AlaT2 → H6 AlaT2 HB1 AlaT5 → H6 AlaT5 HN AlaT5 → H6 AlaT5 H3 AlaT5 → H6 AlaT5	HA AlaT2' → H6 AlaT2' HA AlaT5' → H6 AlaT5' HB1 AlaT2' → H6 AlaT2' HB2 AlaT2' → H6 AlaT2' HB1 AlaT5' → H6 AlaT5' HN AlaT5' → H6 AlaT5'
intra-chain (medium-long-range) **$H3_i \rightarrow H3_{i+3}$** **$H3_i \rightarrow H6_{i+3}$** **$H3_i \rightarrow NH_{i+1}$** In Figure ROESY-C,D,E: cyan for both peptides	H3 AlaT2 → NH AlaT2' H3 AlaT2 → HB2 AlaT2' H3 AlaT2 → NH Aib3	H3 AlaT2' → H3 AlaT5' H3 AlaT5' → H7* AlaT2'
inter-chain In Figure ROESY-D,E: green	H3 AlaT2 → NH AlaT2' H3 AlaT2 → HB2 AlaT2' H3 AlaT2 → HB2 AlaT5' H3 AlaT2 → H7* AlaT5' H3 AlaT5 → H6 AlaT2' H3 AlaT5 → H6 AlaT5' H3 AlaT5 → H7* AlaT5' H3 AlaT5 → H7* AlaT2'	H3 AlaT5' → H6 AlaT5 H3 AlaT5' → HB1 AlaT5 H3 AlaT5' → H7* AlaT5

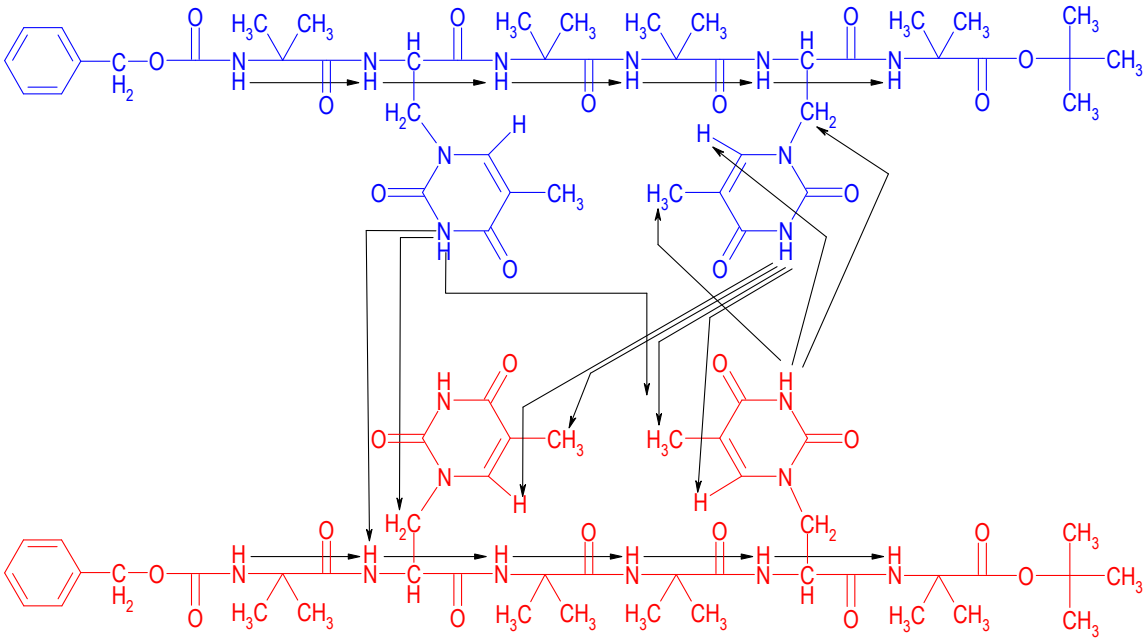


Fig. S16 Schematic representation of the most conformationally informative NOE signals observed in the ROESY spectrum of Z-(Aib-AlaT-Aib)$_2$-O^{t}Bu (5 mM in $CDCl_3$). The ***intra***-molecular $NH_i \rightarrow NH_{i+1}$ sequential NOEs indicate that both molecules adopt an helical structure (likewise of the 3_{10}-type) [Wüthrich, K., *NMR of Proteins and Nucleic Acids*, Wiley, New York, p. 162 (1986)], whereas the ***inter***-molecular cross peaks between the four nucleobases suggest a thymine-mediated, parallel, dimeric aggregation.

ROESY EXPERIMENT

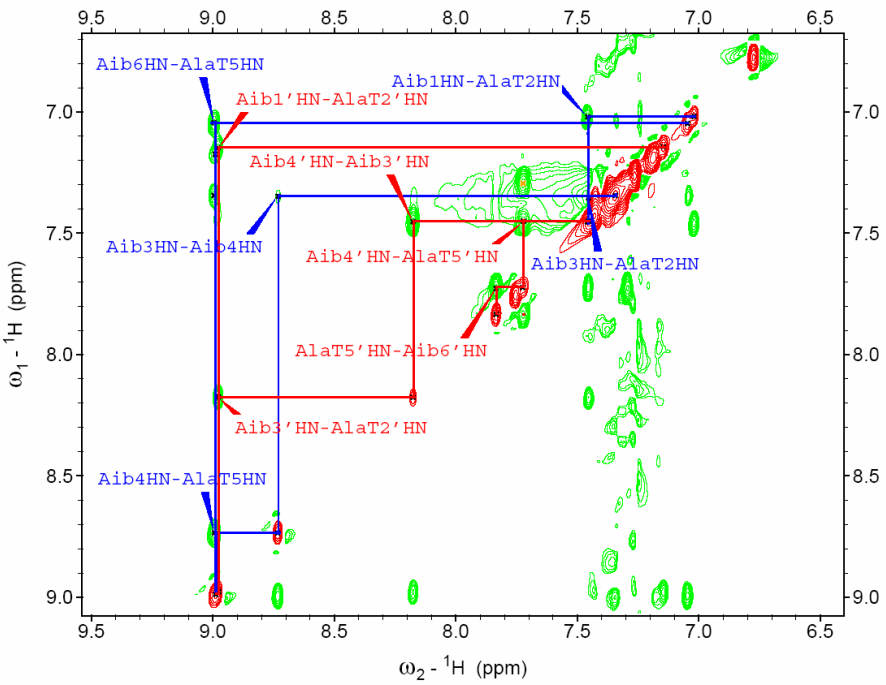


Fig. S17-A: $NH_i \rightarrow NH_{i+1}$ connectivities are shown (blue for "non-primed" peptide and red for "primed peptide").

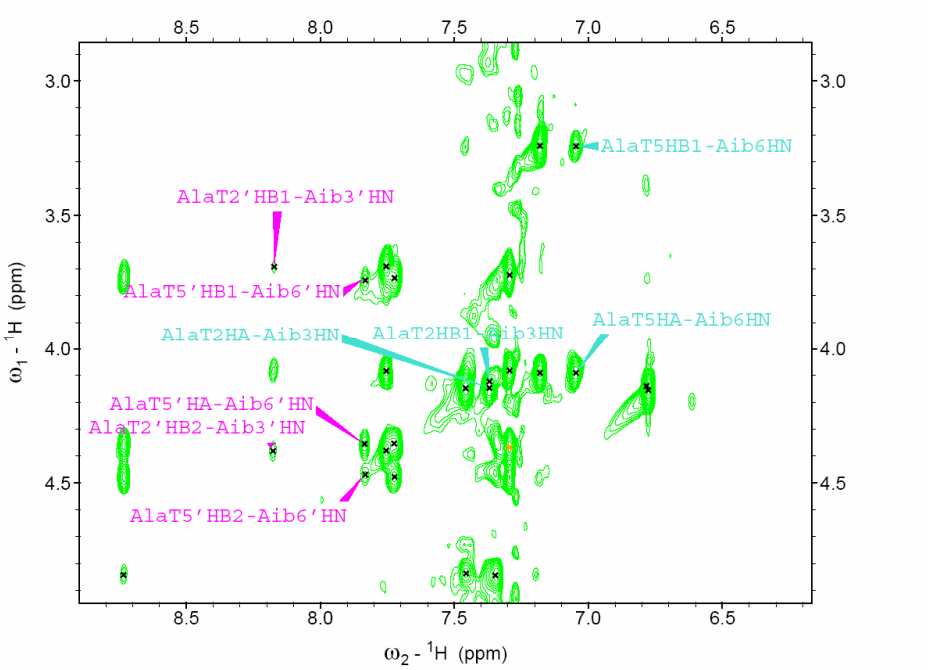


Fig. S17-B: $\alpha CH_i \rightarrow NH_{i+1}$, $\beta CH_i \rightarrow NH_{i+1}$ sequential connectivities are shown (turquoise for "non-primed" peptide and magenta for "primed peptide").

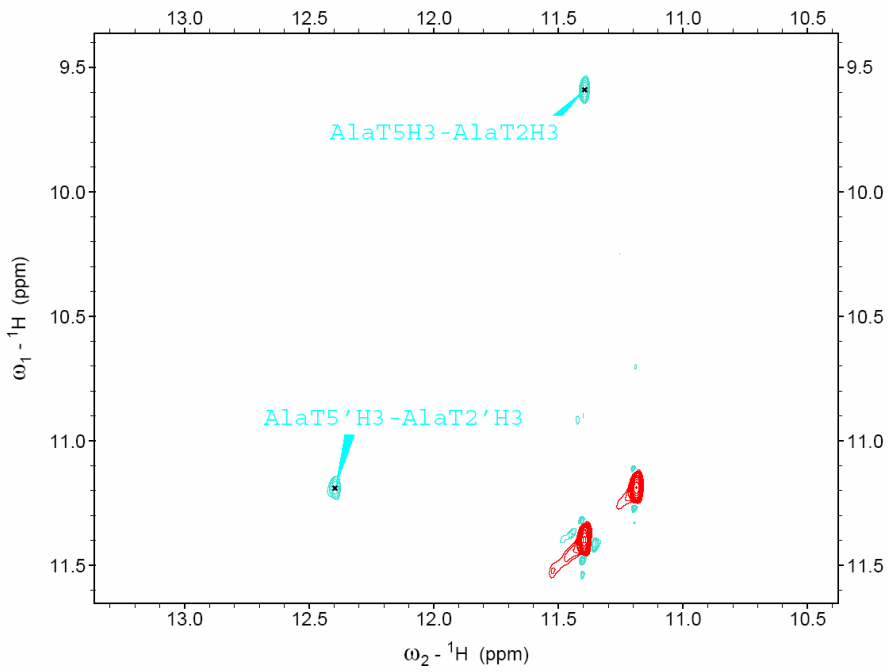


Fig. S17-C: $H3_i \rightarrow H3_{i+3}$, intra-chain connectivities are shown (cyan for both molecules).

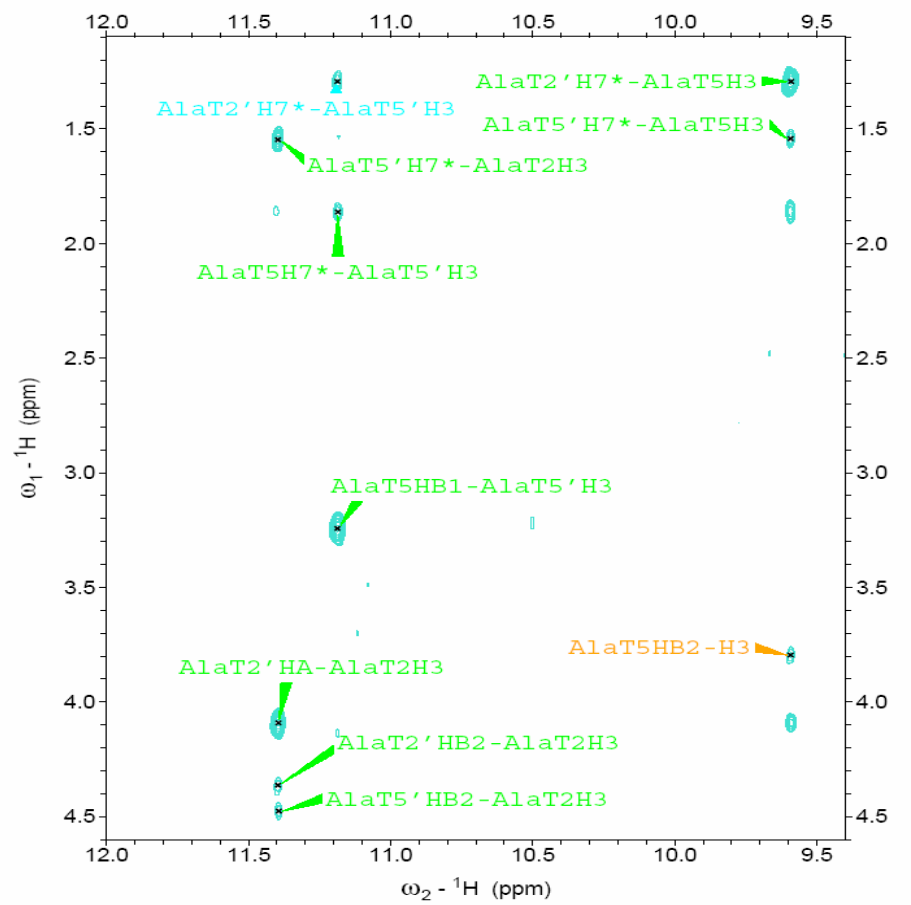


Fig. S17-D: **Inter**-chain connectivities are shown in green; **intra**-chain connectivities are shown in cyan; **intra**-residue connectivities are shown in orange.

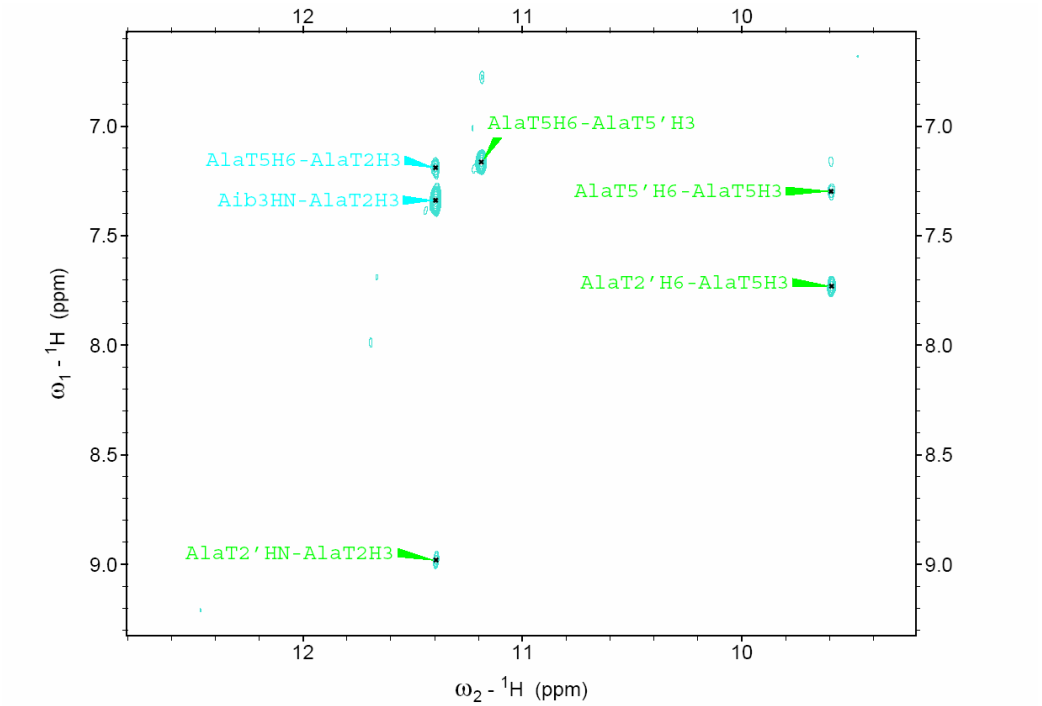


Fig. S17-E: **Inter**-chain connectivities in green; **intra**-chain connectivities in cyan.