Design and Synthesis of Intrinsically Cell-Penetrating Nucleopeptides

Piero Geotti-Bianchini,[a,b] Julien Beyrath,[a] Olivier Chaloin,[a] Fernando Formaggio[b]* and Alberto Bianco[a]*

Supporting Information

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 tetrahydrofurane (THF) purchased from Fluka was distilled under N_2 from a purple solution of sodium-benzophenone ketile. Acetic acid (AcOH), 1-butanol, chloroform ($CHCl_3$), dichloromethane (CH_2Cl_2), 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), pyridine, 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, CDCl3, d_6-dimethylsulphoxide, d_6-DMSO) were purchased from Cambridge Isotope Laboratories.

Reagents

Benzyl chloroformate (ZCl), biotine, bromocresol green, *para*-cresol (4-MePhOH), piperidine were purchased from Acros. *N,N*-

dimethylamino pyridine (DMAP), 2-hydroxy-cinnamic acid, ninhydrin, sodium hydride, 60% dispersion in mineral oil, *N,N,N',N'*-tetramethyl-4,4'-diamino-diphenylmethane (TDM), trimethylsilyl trifluoromethansulphonate (TMSOTf), triphenylphosphine (PPh_3), were purchased from Aldrich. N^{α}-*tert*-butyloxycarbonyl-serine (Boc-Ser-OH) was purchased from Bachem. Calcium chloride anhydrous, potassium bromide, potassium hydrogenocarbonate, potassium hydrogenosulphate, sodium sulphate anhydrous were purchased from Carlo Erba. Paraformaldehyde was purchased from EMS. Diisopropyl-ethyl amine (DIEA) was purchased from Fluka. Adenine, cytosine, thymine were purchased from Lancaster. DNA oligonucleotides (dT_8 and dT_{10}) were purchased from Eurogentec (Seraing, Belgium). Boc-protected amino acids for solid phase synthesis (Boc-Ala-OH, Boc-Lys(2-Cl-Z)-OH, Boc-Lys(Fmoc)-OH), *O*-(benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluoro-phosphate (HBTU), 1-hydroxy-*1H*-benzotriazole (HOBt) were purchased from NeoMPS. Silica gel for flash-chromatography (diameter 40-63 μm, mesh 230-400) and sodium azide were purchased from Merck. 2-Deoxy-D-glucose was obtained from Sigma.

N,N'-dimethyl-azo-dicarboxilate (DMAD) was prepared according to the literature.[1]

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). Spectrophotometric grade MeOH (Fluka) was used as a solvent.

Thin layer chromatography (TLC): 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 eluants.

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

Nucleoamino acids and protected nucleobases 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.[2]

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

Infrared spectroscopy (IR): Absorption infrared spectra (KBr pellets) were collected with a Perkin-Elmer 1720 X FT-IR spectrophotometer.

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 either on a Pharmacia LKB-LCC 2252 instrument, provided with a C18 Phenomenex (4.6×250 mm, 100 Å) column and a UV Uvicord SD detector operating at 214 nm, or on a Varian ProStar 410, provided with a Nucleosil 100-5 C18 column (4.6×150 mm) and a Varian ProStar 330.71 PDA detector.

Preparative chromatograms were operated on a Beckmann System Gold HPLC 166 instrument, provided with a SDS C18 Macherey-Nagel (10×250 mm, 100 Å) column and a Beckmann System Gold UV detector operating at 230 nm.

Eluants: A: H_2O + 0.1% TFA; B: MeCN + 0.08% TFA (HPLC Varian and Beckmann).

C: H_2O/MeCN 9:1 + 0.05% TFA; D: MeCN/H_2O 9:1 + 0.05% TFA (HPLC Pharmacia).

MilliQ H_2O has been used for eluant preparation.

NMR spectrometry: ^{1}H-NMR spectra were recorded either on a Bruker AC 200 spectrometer at 200 MHz or on a Bruker AC 300 spectrometer at 300 MHz. ^{13}C-NMR spectra were recorded on a Bruker AC 300 at 76 MHz. 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, 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.
Mass spectrometry (MS): ESI-TOF mass spectra of nucleoamino acids were collected on a Finnigan LCQ Advantage Thermo Scientific LC-MS, while 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 dilute aqueous nucleopeptide solution were mixed and the same volume of the resulting mixture was loaded on target disk.

Solution phase synthesis of protected nucleobases and nucleoamino acids

N$^{\alpha}$-*tert*-butyloxycarbonyl-serine-β-lactone[1] [Boc-Ser(lactone)]: 3.80 g PPh_3 (14.5 mmol) are dissolved in 58 mL of distilled THF under N_2 and the solution is cooled to -78°C in an acetone/dry ice bath. 2.43 g DMAD (15.0 mmol) diluted in 15 mL of anhydrous THF are added dropwise in 10 min and after 10 min a solution of 3.12 g Boc-Ser-OH (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 pale yellow oil and the crude product is immediately purified through flash-chromatography (EP/AcOEt, 7:3 to 11:9). Yield: 48%. 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).

N^{α}-*tert*-butyloxycarbonyl-β-thymin-1-yl-alanine[3] (Boc-AlaT-OH): 0.99 g thymine (7.6 mmol) are suspended in 68 mL anhydrous DMF, 0.27 g of 60% NaH in mineral oil (6.7 mmol) are 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 1.30 g of Boc-Ser(lactone) (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 2N 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 (CH_2Cl_2/MeOH/AcOH). Yield: 34%. Mp: >250°C (dec.) $[\alpha]^{25}_D$ = -55.8° (c=0.3 in MeOH). Rf_1: 0.0; Rf_2: 0.57; 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) t_r: 12.1 min (0-50% B in 20 min). ^{1}H-NMR (d_6-DMSO, 200 MHz), δ/ppm:11.16 (br s, 1 H, N(3)H thymine), 7.33 (s, 1 H, C(6)H thymine), 6.76 (d, 1 H, *J*=8.0 Hz, αNH AlaT), 4.27-4.18 (m, 2 H, αCH + 1 βCH AlaT), 3.48-3.39 (dd, *J*=11.0, 14.2 Hz, 1 H, 1 βCH AlaT), 1.71 (s, 3 H, C^5H_3 thymine), 1.38-1.30 (s + s, 9 H, 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 AlaT), 49.96 (βCH_2 AlaT), 28.48-28.19 (CH_3 Boc, *cis-trans*), 12.48 (C^5H_3 thymine).

N^6-benzyloxycarbonyl-adenine[4] (HA^Z): 6.4 g of 60% NaH in hexanes (160 mmol) are suspended in 160 mL of anhydrous DMF and stirred under N_2 for 10 min, then 5.4 g adenine (40 mmol) are 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

13.5 mL of ZCl (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, and 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).

N^{α}-*tert*-butyloxycarbonyl-β-(N^6-benzyloxycarbonyl)-adenin-1-yl-alanine (BocAlaA^{Z}OH): 1.63 g HA^Z (6.1 mmol) are suspended in 61 mL of anhydrous DMF, 0.84 mL DBU (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 1.04 g Boc-Ser(lactone) (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 HA^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 ($CHCl_3$/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) t_r: 16.1 min (10 - 70% D in 30 min). ^{1}H-NMR (d_6-DMSO), δ/ppm: 13.07 (br s, 1H, COOH),

10.65 (br s, 1H, N^6H 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 + 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 (s + 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*).

N^6-benzyloxycarbonyl-cytosine[4] (HC^Z): 5.2 g of cytosine (46 mmol) are suspended in 83 mL of py and 0.58 g DMAP (4.7 mmol). The mixture is cooled in a water/ice bath and 15.3 mL ZCl (102 mmol) are added dropwise. The mixture becomes homogeneous, then precipitation of the product begins. Stirring continues for 3 days at room temperature, while the suspension turns yellow-orange. 180 mL chilled H_2O are added and the resulting almost colourless suspension is stirred for 10 min in a water/ice bath. The precipitate is filtrated and washed with plenty of H_2O, then it is desiccated under vacuum. Yield: 49%. Mp:>200°C (dec).
Rf_1: 0.35; Rf_2: 0.80; Rf_3: 0.30. IR (KBr):3423, 1743, 1688, 1629, 1588, 1513 cm^{-1}. UV: ε= 4430 $M^{-1}cm^{-1}$ at λ=288.6 nm, ε= 12800 $M^{-1}cm^{-1}$ at λ=238.6 nm (40.0 μM in MeOH).
^{1}H-NMR (d_6-DMSO, 200 MHz), δ/ppm: 11.07 (br s, 1 H, N^4H cytosine), 7.84 (d, J=7.2 Hz, 1H, C(5)H cytosine), 7.44 (m, 5H, C_6H_5-Z), 6.96 (d, 1H, J=7.2 Hz, C(5)H cytosine), 5.21 (s, 2H, CH_2-Z).

N^α-*tert*-butyloxycarbonyl-β-N^4-benzyloxycarbonyl-cytosin-1-yl-alanine (Boc-$AlaC^Z$-OH): 1.36 g HC^Z (5.4 mmol) are suspended in 80 mL anhydrous DMF, 0.20 g of 60% NaH in hexanes (5.0 mmol) are added, and the resulting yellow solution is stirred for 2 hours.

The mixture is then cooled to -78°C in an acetone/dry ice bath, and a solution of 0.96 g of Boc-Ser(lactone) (5.0 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 foam is dissolved in the minimum amount of H_2O, cooled in a water/ice bath and acidified carefully with HCl 2N until latescence is observed. AcOEt is added, after redissolution of the product some unreacted HC^Z is filtered off, and the phases are separated. The acid phase is extracted three times with AcOEt, the combined organic phases are washed with brine, desiccated on Na_2SO_4 and concentrated to a pale yellow solid. The remaining HC^Z is filtered off after redissolution and precipitation in MeOH (hot/cold) and the resulting oil is purified by flash-chromatography (CH_2Cl_2/MeOH/AcOH).

Yield: 51%. Mp: 167-170°C (from AcOEt/PhMe).

$[\alpha]^{25}_D$ =-85° (c=0.3 in MeOH). Rf_1: 0.0 Rf_2: 0.70 Rf_3: 0.0.

IR (KBr): 3416, 3250, 1751, 1704, 1650, 1628, 1562 cm^{-1}.

UV: ε=18100 $M^{-1}cm^{-1}$ at λ=297.6 nm, ε=33600 $M^{-1}cm^{-1}$ at λ=240.1 nm (18 μM in MeOH).

^{1}H-NMR (d_6-DMSO, 200 MHz), δ/ppm: 10.78 (br s, 1 H, N^4H-Z cytosine), 7.90 (d, J=7.40 Hz, 1 H, C(5)H cytosine), 7.42 (s, 5 H, C_6H_5-Z), 6.97 (d, J=7.40 Hz, 1 H, C(6)H cytosine), 6.85 (d, 1 H, αNH AlaC), 5.21 (s, 2 H, CH_2-Z), 4.49-4.41 (dd, J=3.9, 11.4 Hz, 1 H, 1 αCH),

4.35-4.26 (m, 1 H, 1 βCH), 3.67-3.35 (m, 1 H, 1 βCH), 1.32-1.26 (s + s, 9 H, 3 CH_3 Boc *cis-trans*).

^{13}C-NMR (d_6-DMSO), δ/ppm: 172.39 (COOH), 163.21 (C(4) cytosine), 155.60-155.41 (CO-Boc *cis-trans*), 153.67 (CO-Z), 150.77 (CO (2) cytosine), 136.46 (C-*ipso* C_6H_5), 128.90 (C-*meta* C_6H_5), 128.57 (C-*para* C_6H_5), 128.31 (C--*ortho* C_6H_5), 119.00 (C(6) cytosine), 93.97 (C(5) cytosine), 78.55 ($C(CH_3)_3$ Boc), 66.84 (CH_2-Z), 52.80 (αCH AlaC), 51.96 (βCH_2 AlaC), 28.49 (CH_3 Boc).

Solid-phase synthesis of nucleopeptides

Solid-phase synthesis of the nucleopeptide sequences was accomplished using a PSP 4000 semiautomatic peptide synthesizer for Boc/Bzl strategy. The general sequence of the peptides is H-Lys(f)-(Ala-AlaB-Ala)$_4$-Lys-NH_2, where AlaB= AlaA or AlaT or AlaC and f refers to underivatized peptides (f=H) or to peptides derivatized with fluoresceine isothyocyanate (f=FIT) or biotin (f=Bt) respectively (see Figure 1 in main text or below for the molecular structures). In order to allow selective derivatization of Lys^1-ε-NH_2, Lys^{14} was introduced as Boc-Lys(2-Cl-Z)-OH , while Lys^1 was introduced as Boc-Lys(Fmoc)-OH. Fmoc-deprotection, splitting and subsequent Lys^1-ε-NH_2 derivatization were accomplished manually. MBHA·HCl resin (loading 0.62 mmol/g) from Applied Biosystems was used as a solid support. Synthesis scale: 30 μmol resin for each sequence before splitting in three parts, 10 μmol for each final derivative.

General synthesis protocol: Nucleopeptide sequences were synthesized using the semiautomatic peptide synthesizer.[5] Before starting the synthesis, the resin was swollen by washing with DMF and CH_2Cl_2. N^{α}-Boc protecting groups were removed treating the resin twice with 2 mL 100% TFA (firstly for 1 min, then for 3 min). After Boc-deprotection the resins were washed with iPrOH, DMF and CH_2Cl_2. Amino acids were activated as follows: 5 eq each of

N^{α}-Boc-protected proteogenic amino acid (Ala, Lys(2-Cl-Z), Lys(Fmoc)), HOBt, HBTU, were dissolved in 2 mL DMF and transferred into the reactor, then 15 eq DIEA are added; 3 eq each of Boc-protected nucleoamino acid (AlaT, $AlaA^{Z}$, $AlaC^{Z}$), HOBt, HBTU were dissolved in 2 mL DMF and transferred into the reactor, then 10 eq DIEA were added. Preactivation was avoided, since it is known to favor racemization.[6]

Couplings were 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 was repeated twice and subsequently qualitative Kaiser test[7] was used to verify the completion of the reaction. When a negative response to the test was obtained, the coupling was repeated a third time. After completing the peptide sequences, the resin beads were transferred to 20 mL syringes provided with sintered glass filters in order to operate manually the remaining steps. Fmoc group on Lys^{1}-ε-NH_2 was removed treating the resin with 2 mL of a 20% (v/v) solution of piperidine in CH_2Cl_2 for 30 min, and then washed four times with CH_2Cl_2. The procedure was repeated twice. The resin beads were split by suspending them in a DMF/CH_2Cl_2 mixture and transferring equal volumes of the suspension into three 2 mL syringes provided with sintered glass filters.

Nucleopeptide derivatization with FITC: 5 eq FITC were dissolved in 1 mL DMF and transferred into the syringe, then 15 eq DIEA were added. The coupling was allowed to proceed for 60 min. Subsequently, qualitative Kaiser test was used to verify the completion of the reaction. When a negative response to the test was obtained, the coupling was repeated a second time.

Nucleopeptide biotinylation: 5 eq each of biotin, HBTU, HOBt were dissolved in 1mL of a 1:1 DMF/DMSO mixture and transferred into the syringe, then 15 eq DIEA were added. The coupling was allowed to proceed for 30 min and it was repeated twice. Subsequently, qualitative Kaiser test is used to verify the completion of the

reaction. When a negative response to the test was obtained, the coupling was repeated a third time.

Nucleopeptide cleavage: A mixture of TFA/TMSOTf/4-MePhOH (80:25:10, v/v/w)[8] was prepared and 0.8 mL of this mixture were transferred to each syringe. Cleavage was allowed to proceed overnight at room temperature.

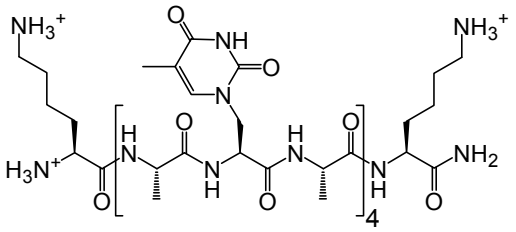
Recovery, analysis and purification of nucleopeptides

Nucleopeptide recovery: The crude nucleopeptides were precipitated from cleavage solutions using cold Et_2O and after centrifugation (5 min at 4000 rpm) the supernatant was eliminated. The precipitate was washed with cold Et_2O and after centrifugation (5 min at 4000 rpm) the supernatant was again eliminated. After waiting 1 hour to allow the remaining Et_2O to evaporate, the crude nucleopeptides were dissolved in 1 mL of H_2O milliQ, analyzed with HPLC and lyophilized.

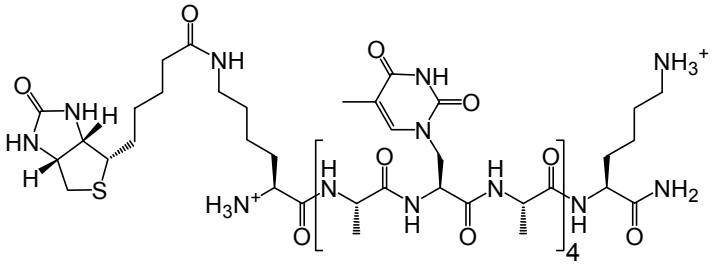
Nucleopeptide purification: The lyophilized crude nucleopeptides were dissolved in H_2O milliQ or in H_2O milliQ/MeCN (total volume 2.5-3.0 mL) and AcOH (0.1-0.2 mL) was added. 0.6-1.0 mL of the resulting solution were injected in the semipreparative HPLC per run and 0.5-2.5 mL fractions were collected. Gradient: 1-41% B in 30 min for FIT-nucleopeptides, 1-31% B in 30 min for all other nucleopeptides. The purity of the collected fractions was checked by analytical HPLC (Varian) after 1:1 dilution with H_2O milliQ. Gradient: 5-65% B in 20 min. The pure fractions were screened by MALDI-MS spectrometry in order to confirm the presence of the desired product. Pure fractions were finally pooled, lyophilized, redissolved in 1.0 mL of a 1 mM solution of HCl in H_2O milliQ and relyophilized.

Nucleopeptides characterization

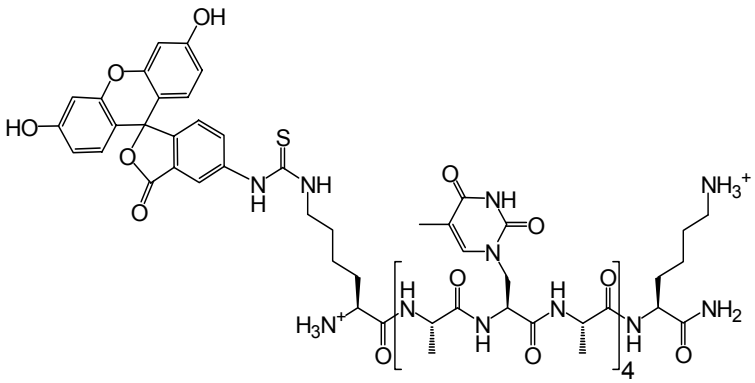
T4H [H-Lys(H)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2]: Yield: 5% (0.8 mg). HPLC: t_r 7.36 min (5-65% B in 20 min). MS calcd $[M+H]^+$: 1622.78, found: 1622.8.



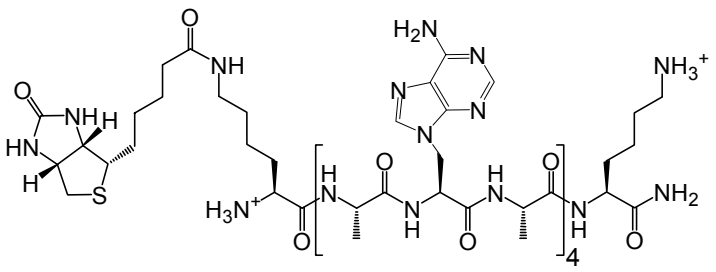
T4Bt [H-Lys(Bt)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2]: Yield: 7% (1.3 mg). HPLC: t_r 8.78 min (5-65% B in 20 min). MS calcd $[M+H]^+$: 1848.86, found: 1848.8.



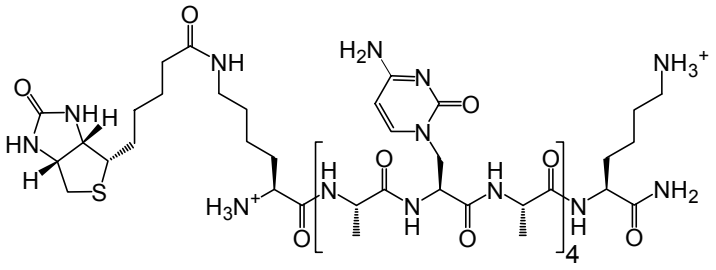
T4FIT [H-Lys(FIT)-(Ala-AlaT-Ala)$_4$-Lys(H)-NH_2]: Yield: 7% (1.4 mg). HPLC: t_r 11.20 min (5-65% B in 20 min). MS calcd $[M+H]^+$: 2011.81, found: 2011.7.



A4Bt [H-Lys(Bt)-(Ala-AlaA-Ala)$_4$-Lys(H)-NH_2]: Yield: 9% (1.7 mg). HPLC: t_r 13.48 min (5-65% B in 20 min). MS calcd $[M]^+$: 1883.88, found: 1884.1.



C4Bt [H-Lys(FIT)-(Ala-AlaC-Ala)$_4$-Lys(H)-NH_2]: Yield: 7% (1.3 mg). HPLC: t_r 7.12 min (5-65% B in 20 min). MS calcd $[M+H]^+$: 1788.85, found: 1789.47.



Surface plasmon resonance

The BIACORE 3000 system, sensor chip CM5, surfactant P20, amine coupling kit containing N-hydroxysuccinimide (NHS) and *N*-Ethyl-*N'*-dimethylaminopropyl carbodiimide (EDC), were from BIACORE (Upsala, Sweden). Streptavidin was obtained from Sigma. All biosensor assays were performed with Hepes-buffered saline (HBS-EP) as running buffer (10mM Hepes, 150 mM sodium chloride, 3 mM magnesium acetate, 0.005% surfactant P20, pH 7.4). DNA oligonucleotides were dissolved in the running buffer. Immobilization of streptavidin was performed by injecting, onto the activated surface by EDC/NHS of a sensor chip CM5, 35 μL of streptavidin (100 μg/mL in formate buffer, pH 4.3), which gave a signal of approximately 5000 RU, followed by 20 μL of ethanolamine hydrochloride, pH 8.5, to saturate the free activated sites of the matrix. The nucleopeptides **A4Bt**, **T4Bt**, **C4Bt** (1 μM in HBS-EP buffer) were allowed to interact with streptavidin until a response of approximately 700-1000 RU was obtained. All the binding

experiments were carried out at 25°C with a constant flow rate of 20 μL/min. Different concentrations of nucleopeptide **T4H** (12.5 to 100 μM), oligonucleotide dT_8 (25 to 100 μM) and oligonucleotide dT_{10} (25 to 200 μM) were injected. The association phase was allowed to run for 3 min, followed by a dissociation phase of 3 min. After each run, the channels were completely regenerated with 10 μM HCl. 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.

Fluorescent microscopy imaging

Fluorescence microscopy was used to examine the uptake of nucleopeptides in RENCA cells. Cells were seeded onto glass coverslips at 20.000 cells per well in 24-well plates, grown overnight in 10% FBS/RPMI medium, 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 cover-slips were then washed three 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 were washed three times with cold PBS, and then permeabilized at room temperature with PBS containing 0.1% Triton X-100 for 10 min. Following another wash, the cells were incubated with streptavidin-FITC (BD Pharmigen) for 1 hour at room temperature. The cells were also stained with 0.5 μg/mL DAPI (Sigma) solution for 10 min at room temperature to identify nuclei. In some experiments the cell membranes were labeled with the lipophilic dye FM4-64 (Molecular Probes). The coverslips were then transferred to slides and cells were visualized by confocal microscopy (LSM 510 Meta, Zeiss) or epifluorescence microscopy (Axiovert 200M, Zeiss, equipped with a Zeiss Apotome module).

To verify the co-localization into the endosomes following the uptake of the nucleopeptides, cells were co-treated with transferrin-Alexa-546 (Molecular Probes) at 20 μg/mL and **T4FIT** (50 μM) for 30 min.

Cell Viability

Viability of Raji, Jurkat and BL41 cells was assessed by the 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphen-yl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) test. Cells were seeded at a density of 50.000 cells per well in 96-well plates, incubated in 10% FBS/RPMI and treated for 24 hours with varying doses of nucleopeptides (**T4H** and **T4Bt**). Doxorubicin was used as a positive control. Following treatment, the cells were incubated for 2 hours at 37°C with 20 μL of MTS reagent solution (Promega). Absorbance was measured at 490 nm wavelength using a microplate reader (Wallac, Victor^2).

Flow cytometry

The uptake of the nucleopeptides was also assessed by flow cytometry. Raji cells were seeded at a density of 100.000 cells per well in 96-well plates, incubated in 10% FBS/RPMI and treated with **T4FIT** (5 or 50 μM) for 5, 30 and 60 min. To assess the mechanism of uptake, some experiments were conducted with the medium supplemented with 2-deoxy-D-glucose (50 mM) and NaN_3 (10 mM). Cells were analyzed with a FACSCalibur®. At least 10.000 events were acquired for each experiment using the CellQuest 3.3 software (Becton Dickinson, Pont de Claix, France) and the data were processed with the WinMDI 2.8 freeware (Joseph Trotter, Scripps Research Institute, http://facs.scripps.edu/software.html).

References

[1] Arnold, L. D., Kalantar, T. H., Vederas, J . C., *J. Am. Chem. Soc.*, **107**, 7105 (1985).

[2] Von Arx, E., Faupel, M., Brugger, M., *J. Chromatogr.*, **120**, 224 (1976).

[3] Chaltin, P., Lescrinier, E., Lescrinier, T., Rozanski, J., Hendrix, C., Rosemeyer, H., Busson, R., Van Aershot, A., Herdewijn, T., *Helv. Chim. Acta*, **85**, 2258 (2002).

[4] Egholm, M. B., Buchardt, O., Nielsen, P. E., Berg, R. H., *J. Am Chem. Soc.*, **114**, 1895 (1992).

[5] Neimark, J., Briand, J.-P., *Pept. Res.*, **6**, 219 (1993).

[6] Corradini, R., Sforza, S., Dossena, A., Palla, G., Rocchi, R., Filira, F., Nastri, F., Marchelli, R., *J. Chem. Soc., Perkin Trans. 1*, 2690 (2001).

[7] Kaiser, E. Colescott, R. L. Bossinger, C. D. Cook, P. I. *Anal. Biochem.* **34**, 595 (1970).

[8] Fujii, N., Otaka, A., Kamura, O., Akaji, K., Funakoshi, S., Hayashi, Y., Kuroda, Y., Yajima, H., *J. Chem. Soc., Chem. Commun.*, 274 (1987).

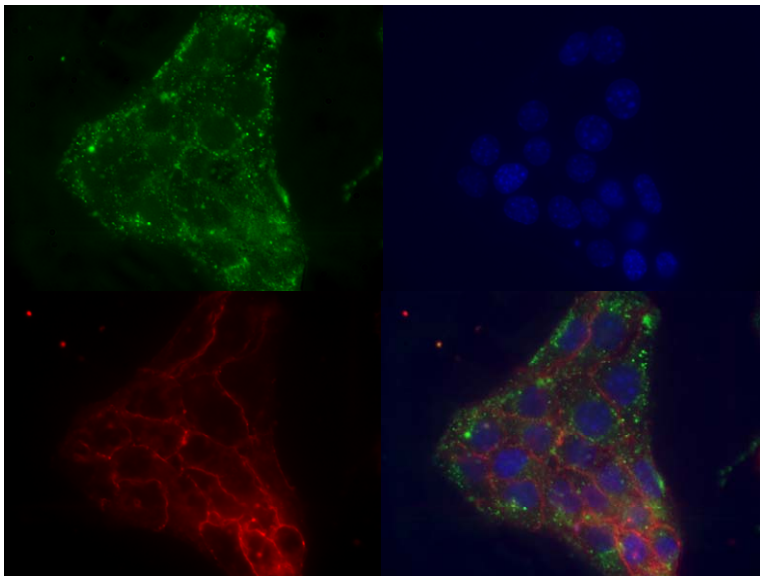


Figure S1. Epifluorescence microscopy images of RENCA cells incubated for 16 hours with 50 μM of **T4FIT** at 37°C in 5% CO_2 (green). The plasma membranes were stained with the lypophilic dye FM4-64 (red). The cell nucleus was labeled with DAPI (blue).

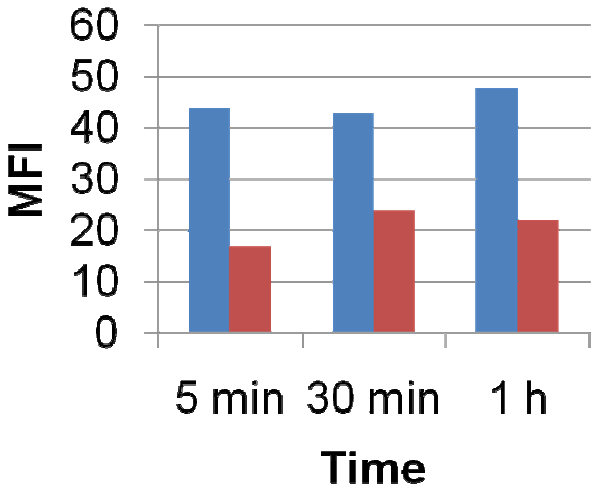


Figure S2. Time dependent uptake of 5 μM of **T4FIT** by Raji cells at 37°C in the absence (cyan bars) or in the presence (red bars) of endocytosis inhibitors NaN_3 and 2-deoxy-D-glucose, quantified by flow cytometry. Plots display the mean fluorescence signal.

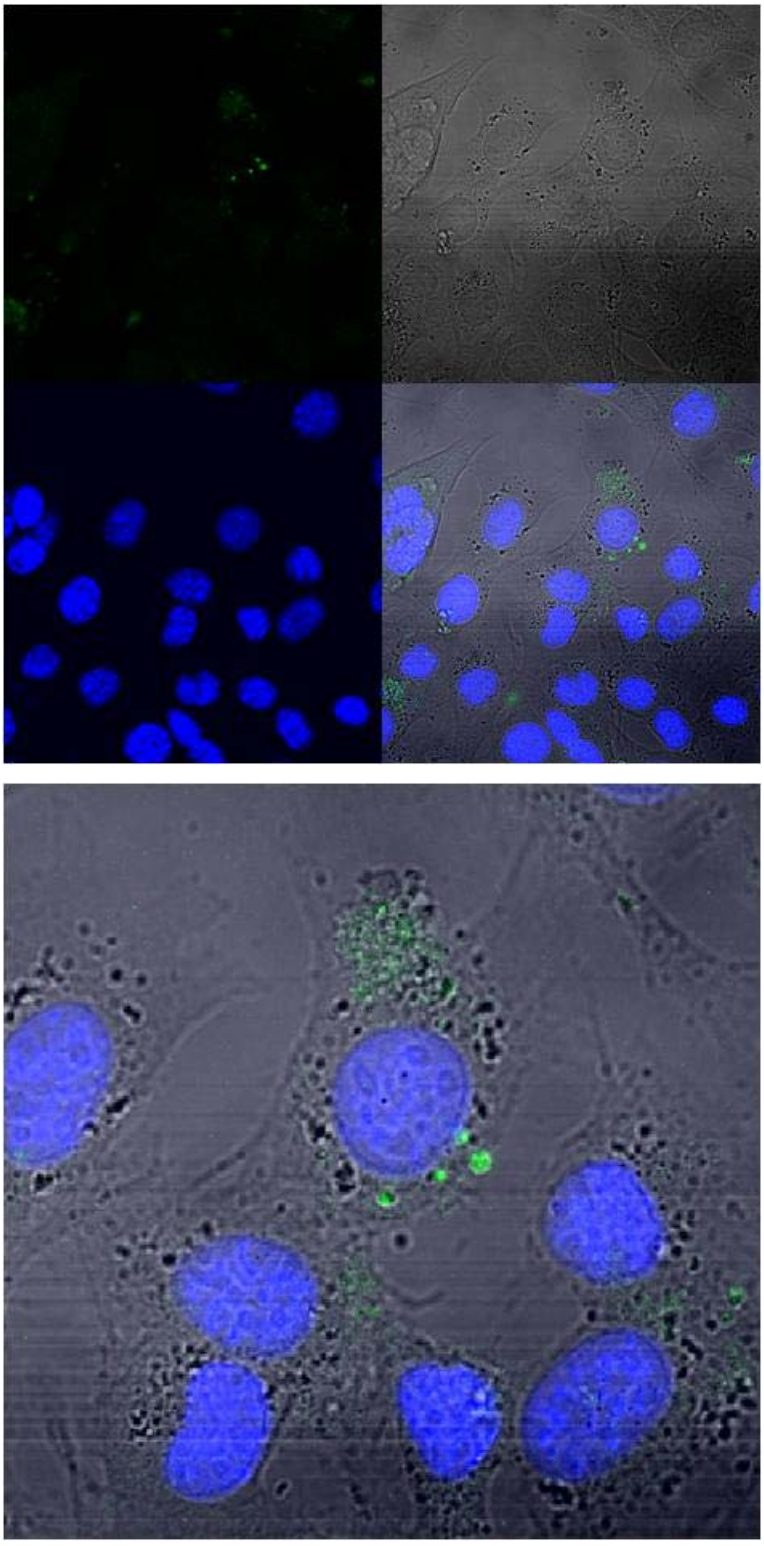


Figure S3. Confocal microscopy images of RENCA cells incubated for 12 h with 5 μM of **T4FIT** at 37°C in 5% CO_2. The cell nucleus was labeled with DAPI (blue). For the sake of clarity, the image below is a magnification of the lower right panel of the top figure.

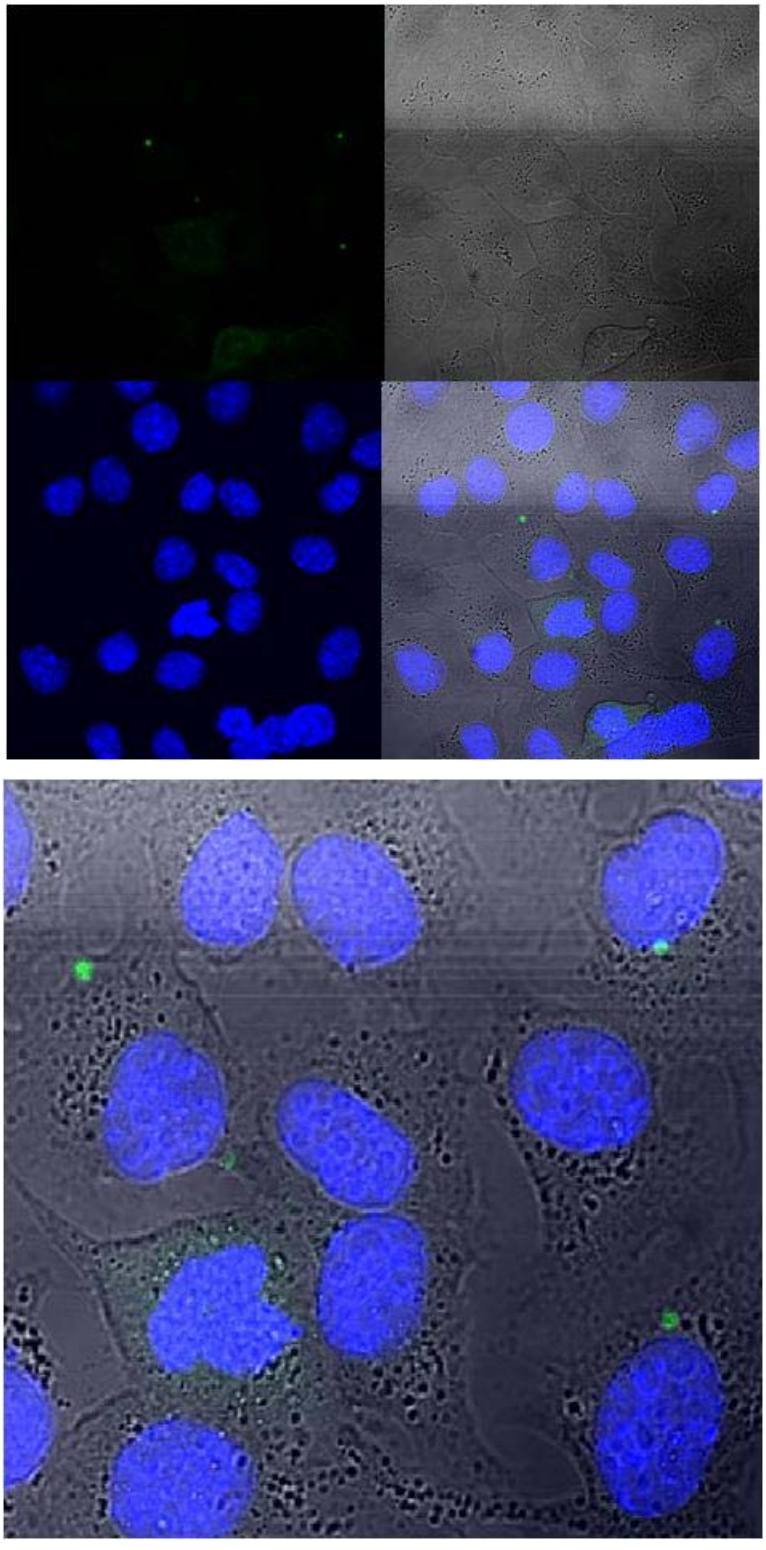


Figure S4. Confocal microscopy images of RENCA cells incubated for 12 h with 5 μM of **T4Bt** at 37°C in 5% CO_2, revealed with streptavidin-FITC (green). The cell nucleus was labeled with DAPI (blue). For the sake of clarity, the image below is a magnification of the lower right panel of the top figure.

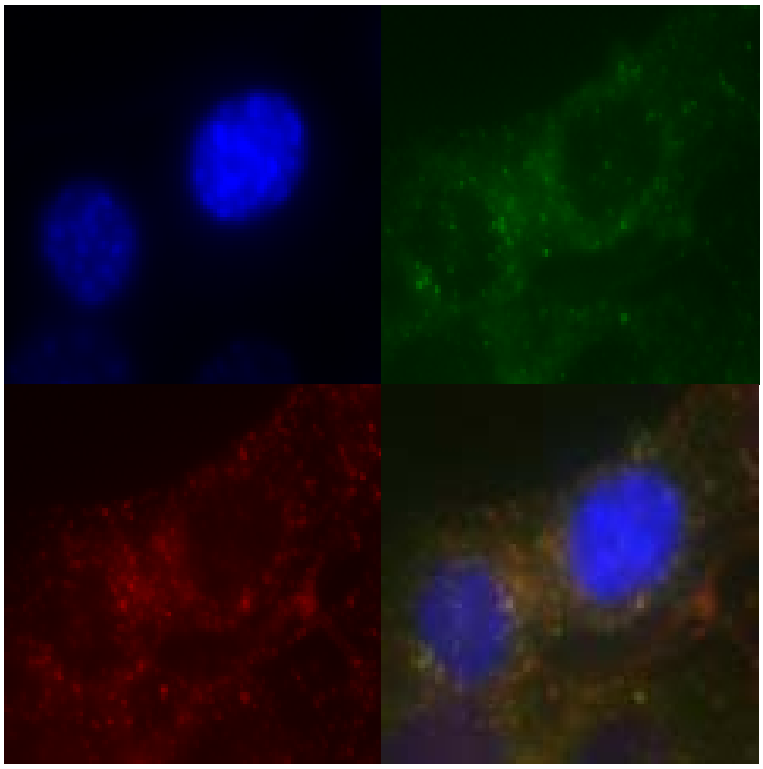


Figure S5. Epifluorescence microscopy images of RENCA cells co-incubated with 50 μM of **T4FIT** (green) and 20 μg/mL of transferrin-Alexa-546 (red) at 37°C for 30 min. Superposition of the images evidences endosomal co-localization (orange). The cell nucleus was labeled with DAPI (blue).

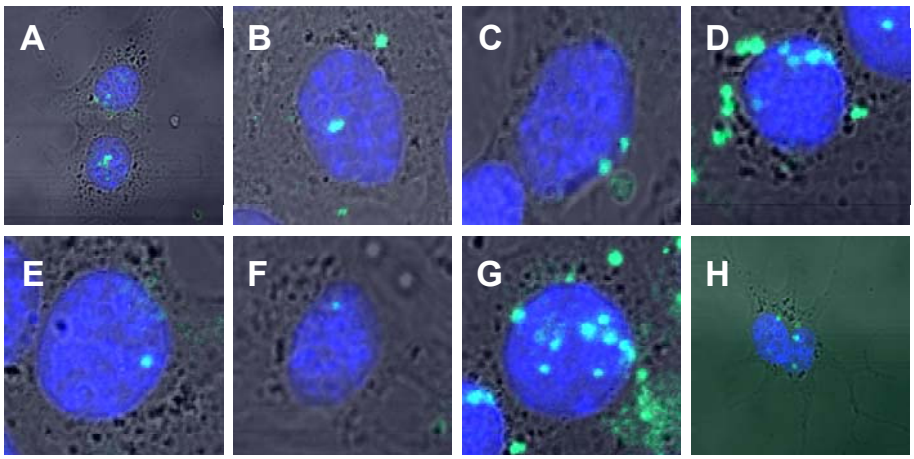


Figure S6. Confocal microscopy images of RENCA cells incubated with 50 μM of **T4FIT** (A-G) and **T4Bt** (H) at 37°C for 12 h. Nucleopeptide **T4Bt** was revealed with streptavidin-FITC (green). The cell nucleus was labeled with DAPI (blue).

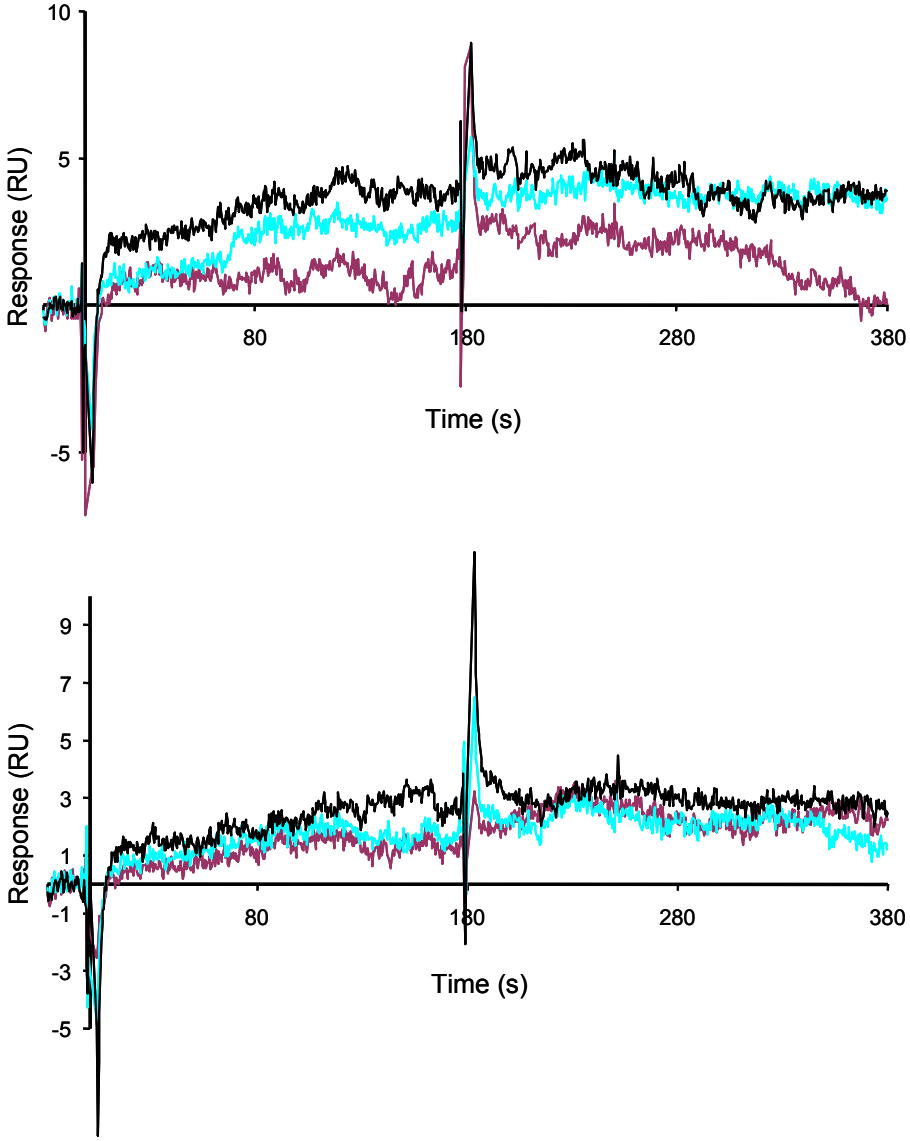


Figure S7. Sensorgrams obtained by allowing various concentrations of oligonucleotides dT_8 (top) [25 μM (magenta), 50 μM (cyan), 100 μM (black)] and dT_{10} (bottom) [25 μM (magenta), 100 μM (cyan), 200 μM (black)] to interact with the nucleopeptide **A4Bt** immobilized onto sensor chip. Nucleopeptide **A4Bt** was immobilized on the sensor chip as described in Figure S7. After this coupling, increasing concentrations of dT_8 and dT_{10} from 25 to 200 μM were injected to evaluate the capacity of interaction of the nucleopeptide with the two complementary oligonucleotides. Both dT_8 and dT_{10} gradually complexed **A4Bt** as evidenced by an increase of the signal during the association phase. During the dissociation phase, both oligomers were very slowly released. The (apparent)

association rate constant (ka) and the (apparent) dissociation rate constant (kd) corresponded to 7.66×10^{2} $M^{-1}s^{-1}$ and 2.63×10^{-3} s^{-1} for dT_8, and to 2.29×10^{3} $M^{-1}s^{-1}$ and 6.94×10^{-3} s^{-1} for dT_{10}. The resulting dissociation constants at the equilibrium (K_D) are 3.44 μM and 3.04 μM, respectively, which represent the affinity between the complementary chains.

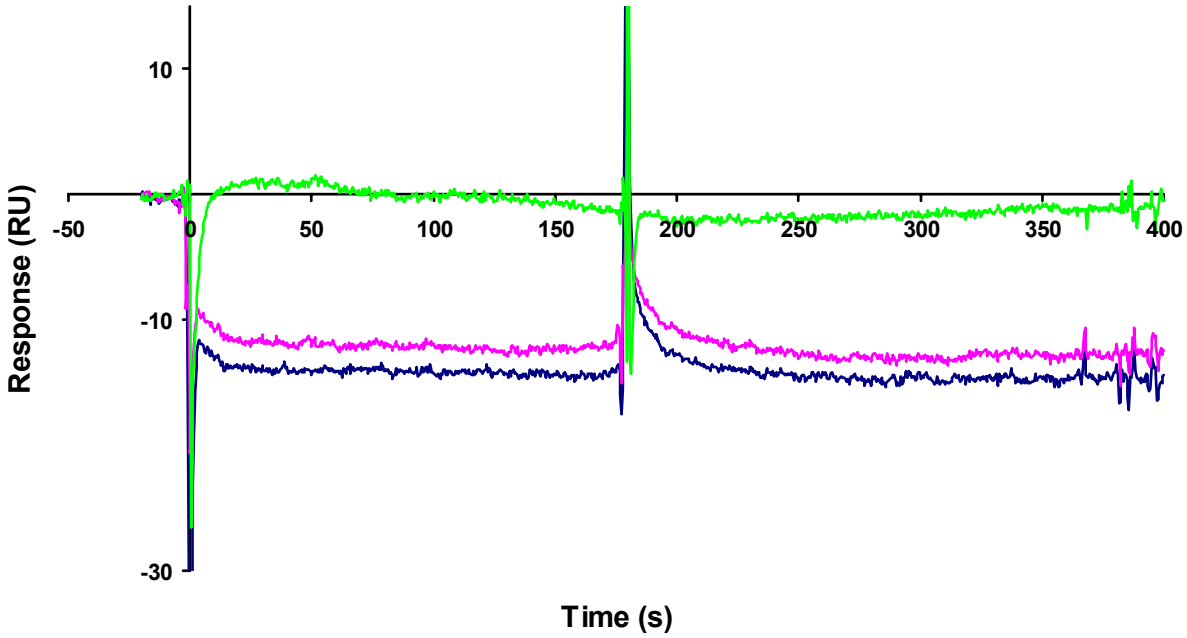


Figure S8. Sensorgrams obtained by allowing 100 μM solutions of nucleopeptide **T4H** (green curve), oligonucleotides dT_8 (blue curve) and dT_{10} (magenta curve) to interact with the nucleopeptide **T4Bt** immobilized onto sensor chip. Nucleopeptide **T4Bt** was immobilized on the sensor chip as described in Figure S7 for **A4Bt**. After this coupling, **T4H**, dT_8 and dT_{10} at 100 μM concentration were injected to evaluate the capacity of interaction of the nucleopeptide with the non complementary nucleopeptide and oligonucleotide sequences. None of the oligomers form a complex with **T4Bt**, as evidenced by the absence of signal variation.

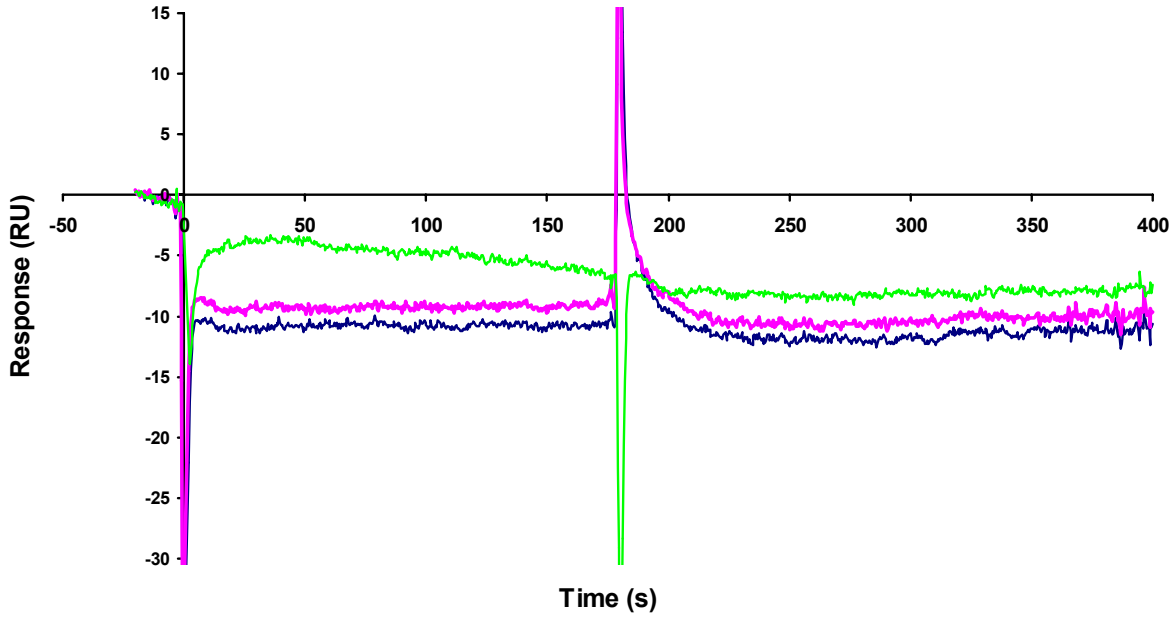


Figure S9. Sensorgrams obtained by allowing 100 μM solutions of nucleopeptide **T4H** (green curve), oligonucleotides dT_8 (blue curve) and dT_{10} (magenta curve) to interact with the nucleopeptide **C4Bt** immobilized onto sensor chip. Nucleopeptide **C4Bt** was immobilized on the sensor chip as described in Figure S7 for **A4Bt**. After this coupling, **T4H**, dT_8 and dT_{10} at 100 μM concentration were injected to evaluate the capacity of interaction of the nucleopeptide with the non complementary nucleopeptide and oligonucleotide sequences. None of the oligomers form a complex with **C4Bt**, as evidenced by the absence of signal variation.

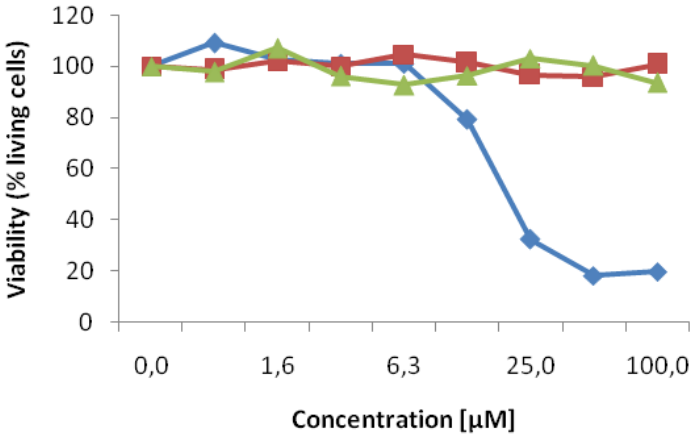


Figure S10. Effect of **T4H** and **T4Bt** on the viability of Jurkat cell line. Concentration-response curves: Jurkat cells were treated with increasing concentrations of nucleopeptides for 24 h, followed by MTS test. For each experiment, the percentages are expressed relative to the mean value of the untreated cells. Square: **T4H**; Triangle: **T4Bt**; Diamond: Doxorubicin.

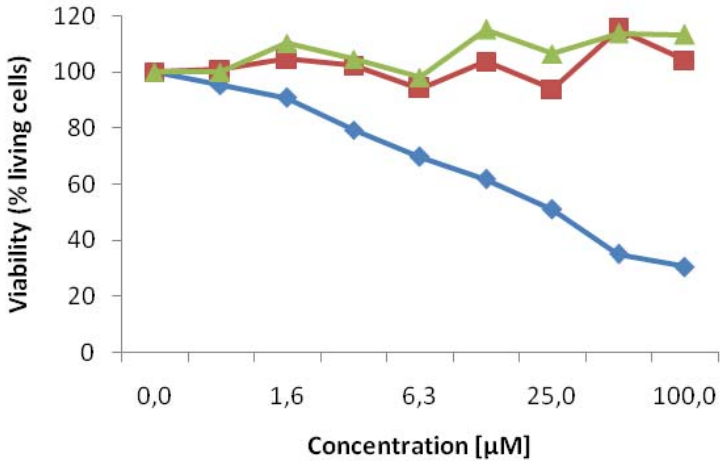


Figure S11. Effect of **T4H** and **T4Bt** on the viability of BL41 cell line. Concentration-response curves: BL41 cells were treated with increasing concentrations of nucleopeptides for 24 h, followed by MTS test. For each experiment, the percentages are expressed relative to the mean value of the untreated cells. Square: **T4H**; Triangle: **T4Bt**; Diamond: Doxorubicin.