Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Exploring the impact of side-chains length on peptide/RNA binding events

Lola Sbicca[1], Alejandro López González[2], Alexandra Gresika[1], Audrey Di Giorgio[1], Jordi Teixido Closa[2], Roger Estrada Tejedor[2], Marie-Line Andréola[3], Stéphane Azoulay[1], and Nadia Patino[1,*]

[1] Université Côte d'Azur, CNRS, Institut de Chimie de Nice, 06108 Nice, France
[2] Molecular Design Lab., IQS School of Engineering, Universitat Ramon Llull, 08017 Barcelona, Spain
[3] Laboratoire MFP, UMR 5234 CNRS, Université Bordeaux, FR Transbiomed, 146 Rue Léo Saignat, F-33076 Bordeaux, France

* To whom correspondence should be addressed. +33 (0)4 92 07 61 46; Fax: +33 (0)4 92 07 61 51; Email: patino@unice.fr

Chemistry

Abbreviations:

Ac_2O (Acetic anhydride), Boc_2O (terbutyloxy anhydride), Boc (terbutyloxycarbonyl), $CDCl_3$ (Deuterium Chloroform), cyhex (Cyclohexane), DCM (Dichloromethane), DIPEA (N,N-Diisopropylethylamine), DEA (Diethylamine), DMF (Dimethylformamide), DMSO (Dimethylsulfoxide), D_2O (Deuterium oxide), EtOAc (Ethyl acetate), EtOH (Ethanol), Fmoc (Fluorenylmethoxycarbonyl), HBTU (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluroniumhexafluorophosphate), MeOH (Methanol), Pbf (2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl), Pyr (Pyridine), TEA (Triethylamine), TFA (Trifluoroacetic acid), TFMSA (Trifluoromethanesulfonic acid), THF (Tetrahydrofurane), TIS (Triisopropylsilane), Trt (Trityl), Z (Benzyloxycarbonyl).

Materials and equipment

Solvents and reagents were obtained from commercial sources and used without further purification. Protected amino acids Fmoc-Lys(Boc)-OH, Fmoc-Gln-OH, Fmoc-Orn(Boc)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Agb(Pbf)-OH, Fmoc-HomoArg(Pbf)-OH and Fmoc-Rink-Amide MBHA resin were purchased from Sigma-Aldrich and Novabiochem. Fmoc-HomoLys(Boc)-OH and Fmoc-HomoGln-OH were synthezised as described below, according to slightly modified published procedures(26, 27).
Analytical thin-layer chromatography (TLC) was conducted on Sigma-Aldrich precoated silica gel 60F254 plates and compounds were visualized with ninhydrin test and/or under ultraviolet light (254 nm). Column chromatographies were carried out on silica gel (Merck Geduran® Si 60 (40-63) µm)). Analytic and semi-preparative reverse-phase HPLC were performed using a Waters apparatus, including a Waters™ 2695 Separation Module, a Waters™ 600 pump and a 996 photodiode array detector. Solvent A and solvent B, respectively 0.1% TFA in water and 0.1% TFA in acetonitrile, were used for HPLC studies, at a flow rate of 1 mL/min for analytic HPLC and at 3.5mL/min for semi-preparative purifications, using respectively a Thermo Scientific BETABASIL C18 column (250 x 4.6 mm, 5 µm, 150 Å) and a Thermo Scientific BETABASIC-18 column (250 x 10 mm, 5 µm, 300 Å). Data were monitored using a Waters Millenium software. All HPLC analyses were run at room temperature. ^{1}H, ^{13}C, COSY, TOCSY, HSQC and DEPT NMR experiments were performed on Brucker AC 200 or AV 500 spectrometers, operating respectively at 200MHz and 500MHz, using deuterated MeOD, D_2O or $CDCl_3$ purchased from Eurisotop. Chemical shift (δ) were reported in parts

per million (ppm). ^{1}H NMR splitting patterns were designated as singlet (s), doublet (d), triplet (t), quartet (q). Splitting patterns that could not be interpreted or easily visualized were recorded as multiplet (m) or broad resonance (br). For the atom numbering, see figures S1 and S2 in Supporting Information. ESI mass spectra were recorded with a Bruker Esquire 3000 plus, equipped with an atmospheric pressure ionization source. This method used in either positive mode or a negative one gives respectively either $(M+H)^+$ and/or $(M+Na)^+$ signals or $(M-H)^-$ signals. HRMS analyses were carried out on an LTQ Orbitrap hybrid mass spectrometer with an electrospray ionization probe (Thermo Scientific, San Jose, CA) by direct infusion from a pump syringe, to confirm correct molar mass and high purity of Tat(+1) and Tat(-1) peptides.

Synthesis and characterization of unnatural amino acid residues

Fmoc-HomoLys(Boc)-OH:

Fmoc-Gly(All)-OMe (2)

To a solution of commercially available Fmoc-Gly(All)-OH **(1)** (1.38 g, 4.09 mmol) in DMF were added cesium carbonate (1.33 g, 4.09 mmol) and methyl iodide (382 µL, 6.14 mmol). The mixture was allowed to stir at rt for 1h then extracted with EtOAc (3×30 mL). The combined organic fractions were washed with brine (75 mL), dried ($MgSO_4$) and concentrated under vacuum to yield compound **(2)** (1.27g, 85%) as a white solid.

Rf (cyhex//EtOAc 7:3): 0.46
1H NMR (200 MHz, $CDCl_3$) δ (ppm): 7.79-7.26 (m, 8H, H_{Ar}(Fmoc)); 5.77-5.60 (m, 1H, CHγ); 5.36-5.11 (m, 3H, $(CH_2)\delta$ + NH); 4.42-4.23 (m, 4H, CHα + CH (Fmoc) + CH_2 (Fmoc); 3.77 (s, 3H, OCH_3); 2.59-2.55 (q, 2H, $(CH_2)\beta$).
MS (ESI$^+$) (m/z) calculated for $C_{21}H_{21}NO_4$ $[M+H]^+$: 352.1; found $[M+H]^+$: 352.2.

Compounds (4) and (4'):

To a solution of N-Boc-allylamine (**3**) (730 mg, 4.26 mmol) in CH_2Cl_2 (60 mL) was added **(2)** (749 mg, 2.13 mmol). $RuCl_2(PCy_3)_2$ (351 mg, 0.43 mmol) was added in two times (0.1 eq. at first then 0.1 eq after 10h). The mixture was heated at reflux for 18h and the solvent was removed by rotary evaporation. The crude product was purified by silica gel flash column chromatography (cyhex/EtOAc 9:1) to yield compound **(4)** (421 mg, 40%) as a brown oil and by-product **(4')** as a brown solid (312 mg, 46%).

***Compound (4)*:**

Rf (cyhex//EtOAc 7:3): 0.27
^{1}H NMR (200 MHz, $CDCl_3$) δ (ppm): 7.79-7.26 (m, 8H, H_{Ar}(Fmoc)); 5.45-5.30 (m, 3H, NHα + CHγ + CHδ); 4.43-4.22 (m, 5H, CH(Fmoc) + CH_2(Fmoc) + NHε [Boc] + CHα); 3.76 (s, 3H, OCH_3); 3.125 (m, 2H, $(CH_2)\zeta$); 2.55-2.50 (m, 2H, $(CH_2)\beta$); 2.25-2.15 (m, 2H, $(CH_2)\varepsilon$); 1.43 (s, 9H, Boc).
MS (ESI$^+$) (m/z) calculated for $C_{28}H_{34}N_2O_6$ $[M+H]^+$: 494.2; found $[M+H]^+$: 495.3

Compound (4'):

Rf (cyhex//EtOAc 7:3): 0.21

^{1}H NMR (200 MHz, $CDCl_3$) δ (ppm): 5.53-5.42 (m, 2H, 2CHδ); 4.60 (s, 2H, 2NHε); 3.20-3.10 (m, 4H, 2(CH_2)ζ); 2.28-2.14 (m, 4H, 2(CH_2)ε); 1.45 (s, 18H, 2(Boc)).

MS (ESI$^+$) (m/z) calculated for $C_{16}H_{30}N_2O_4$ $[M+H]^+$: 314.2; found $[M+H]^+$: 314.1

Fmoc-HomoLys(Boc)-OMe (5)

To a solution of **(4)** (550 mg, 1.11 mmol) in THF (75 mL) was added 10% Pd/C (284 mg, 0.267 mmol). The resulting mixture was stirred for 16h at 25 °C under H_2 (1 atm, balloon) and then filtered and concentrated under reduced pressure. The crude product was purified by silica gel flash column chromatography (cyhex/EtOAc 8:2) to yield compound **(5)** (230 mg, 41%) as a brown oil.

HPLC (A/B: from 50/50 to 0/100 in 30 min): Rt = 15.8 min

^{1}H NMR (200 MHz, $CDCl_3$) δ (ppm): 7.78-7.31 (m, 8H, H_{Ar}(Fmoc)); 5.36-5.31 (d, 1H, NHα□; 4.54-4.19 (m, 5H, CH(Fmoc) + CH_2(Fmoc) + NHε + CHα□; 3.76 (s, 3H, OCH_3); 3.11-3.09 (m, 4H, (CH_2)ζ+ (CH_2)β□; 1.89-1.23 (m, 15H, (CH_2)$_\gamma$ + (CH_2)δ + (CH_2)$_\varepsilon$ + Boc).

MS (ESI$^+$) (m/z) calculated for $C_{28}H_{36}N_2O_6$ $[M+H]^+$: 497.3; found $[M+H]^+$: 497.2

Fmoc-HomoLys(Boc)-OH

To a solution of **(5)** (190 mg, 0.38 mmol) in 1 mL of dioxane was added an aqueous solution of LiOH 1N (460 μmol, 0.46 mmol). The mixture was stirred at room temperature for 2h, acidified with HCl 1M until pH= 3 then extracted with EtOAc (3×30 mL). The combined organic fractions were washed with brine, dried ($MgSO_4$) then concentrated in vacuum to yield the title compound (156 mg, 85%) as a white solid.

HPLC (A/B: from 40/60 to 0/100 in 30 min): Rt = 29.2 min

^{1}H NMR (200 MHz, MeOD) δ (ppm): 7.79-7.24 (m, 8H, H_{Ar} (Fmoc)); 4.53- 3.96 (m, 4H, CH (Fmoc) + CH_2 (Fmoc) + CHα□; 3.00 (m, 2H, (CH_2)ζ]); 1.98-1.03 (m, 17H, (CH_2)β + (CH_2)γ + (CH_2)δ + (CH_2)ε + Boc).

^{13}C NMR (50 MHz, MeOD) δ (ppm): 158.6 (2C, CO(O) (Fmoc) + CO(O)NH (Boc)); 145.4, 142.6, 128.8, 128.2, 126.3, 121.0 (12C, Fmoc); 79.8 (1C, Boc); 67.9 (1C, CH_2 (Fmoc)); 66.9 (1C, CHα); 48.7 (1C, CH (Fmoc)); 41.3 (1C, (CH_2)ζ); 33.5 (1C, (CH_2)β); 30.9 (1C, (CH_2)ε); 28.9 (3C, Boc); 27.6 (1C, (CH_2)γ); 26.6 (1C, (CH_2)δ).

MS (ESI$^-$) (m/z) calculated for $C_{27}H_{33}N_2O_6$ $[M-H]^-$: 481.2 ; found $[M-H]^-$: 481.00.

Fmoc-HomoGln-OH:

Commercially available H-Lys(Z)-OH **(6)** (2.8 g, 10 mmol) was dissolved in 1.8 M H_2SO_4 solution in aqueous 50% AcOH (125 mL). After cooling at 10°C, powdered $KMnO_4$ was added portion wise (5g, 31.8 mmol) over a period of 20min. The mixture was stirred at 10°C for 1h, then a saturated solution of Na_2SO_3 was added followed by a 25% NH_3OH solution (until pH = 5.5). The formed precipitate was filtered off and washed with cold water, affording 1.5 g of a white solid that was used in the next step without further purification. The above solid was dissolved in TFA (trifluoroacetic acid) (20 mL) then 2.5 mL of TIS and 2.5 mL of TFMSA were added at 0 ° C. The mixture was stirred at room temperature for 2h then concentrated under *vacuum*. The residue was triturated in Et_2O and filtered off, giving 1.2 g of a white solid that was used in the next step without further purification. The solid was dissolved in 40 mL of a solution of CH_3CN/H_2O (3:1), and DIEA (5

mL, 30 mmol) then Fmoc-OSu (5.06 g, 15 mmol) were added. The mixture was stirred at room temperature for 2h then concentrated under *vacuum*. The residue was taken up in 20 mL of water and extracted with EtOAc (3 x 20 mL). The combined organic phases were washed with brine. The aqueous phases were collected, acidified at 0°C with a 1M HCl solution until pH= 3, then extracted with EtOAc (3 x 20 mL). The combined organic phases were washed with brine, dried over Mg_2SO_4, then evaporated to give the title compound as a white solid, in 31% overall yield (1.2 g, 3.14 mmol).

HPLC (A/B: from 50/50 to 0/100 in 30 min): Rt= 17.8 min

^{1}H NMR (200 MHz, MeOD) δ (ppm): 7.82-7.27 (m, 8H, H_{Ar} (Fmoc)); 4.36-4.34 (d, 2H, CH_2 (Fmoc)); 4.26-4.08 (m, 2H, CH(Fmoc) + CHα); 2.24 (m, 2H, (CH_2)δ); 1.97-1.46 (m, 4H, (CH_2)β□□□ (CH_2)γ).

^{13}C NMR (50 MHz, MeOD) δ (ppm): 178.1 (1C, C(O)OH); 175.5 (1C, C(O)NH_2); 158.6 (1C, OC(O) (Fmoc); 145.7, 142.8, 129.2, 128.6, 126.8, 121.4 (12C, Fmoc); 68.1 (1C, CH_2(Fmoc)); 55.3 (1C, CHα); 48.7 (1C, CH(Fmoc)); 36.1 (1C, CHδ); 32.5 (1C, CHβ□; 23.5 (1C, CHγ).

MS (ESI$^+$) (m/z) calculated for $C_{21}H_{22}N_2O_5$ $[M+Na]^+$: 405.1; found $[M+Na]^+$: 405.0; **(ESI$^-$)** (m/z) calculated for $C_{21}H_{22}N_2O_5$ $[M-H]^-$: 381.15; found $[M-H]^-$: 381.00.

General method for the solid-phase synthesis of Tat(+1) and Tat(-1)

Tat(+1) and Tat(-1) peptides were prepared following a standard solid-phase strategy, on a Fmoc-Rink-Amide MBHA (100-200 mesh, 0.79 mmol/g, Novabiochem). Elongation was performed using 100 mg of resin, following standard Fmoc protocols:

Fmoc cleavage*:* DMF/piperidine (8:2, v:v), 2 x 10 min, DCM wash, DMF wash.

Coupling conditions*:* The couplings were performed for 1h with a volume of 1.5 mL of a preactivated (3 min) mixture of Fmoc-amino acid residue (0.32 mmol), DIEA (1.26 mmol), HOBt (0.32 mmol) and HBTU (0.32 mmol) in DMF. The couplings were monitored by a Kaiser test and repeated twice.

Capping: Ac_2O/Pyr/DMF 15:15:70 (v:v:v), 2 x 5 min, DMF wash, DCM wash.

Cleavage from the resin*:* The compounds were cleaved (and totally deprotected) using a TFA/H_2O (97:3, v:v) mixture for 2h. The resulting solution combined with 1 mL of TFA (used to wash the resin) was added to cold anhydrous diethyl ether (20 mL). Crude peptide was isolated by centrifugation (3,000 min^{-1}, 10 min), washed with diethyl ether (30 mL) then centrifuged (3 times). After the last washing step, the solid was dried by lyophilisation on a Flexy-Dry FTS system, to give a white powder.

<u>Peptide Tat(+1):</u>

Yield: 28% (50 mg)

HPLC (A/B: from 95/5 to 60/40 in 40 min): Rt = 12.6 min

^{1}H NMR (500 MHz, D_2O) δ (ppm): 4.30-4.22 (s, 9H, CH[α_{1-9}]); 3.23-3.19 (t, 12H, CH_2[ε_{1-3}] + CH_2[ε_{5-6}] + CH_2[ε_9]); 3.00 (t, 4H, CH_2[ζ_{7-8}]); 2.32 (t, 2H, CH_2[δ_4]); 2.06 (s, 3H, Ac); 1.80-1.77 (m, 30H, CH_2[β_{1-9}] + CH_2[δ_{1-3}] + CH_2[δ_{5-6}] + CH_2[δ_9]); 1.68-1.62 (m, 6H, CH_2[γ_4] + CH_2[ε_{7-8}]); 1.41-1.35 (m, 20H, CH_2[γ_{1-3}] + CH_2[γ_{5-9}] + CH_2[δ_{7-8}]).

^{13}C NMR (125 MHz, D_2O) δ (ppm): 179.2 (1C, **C**O-CH_2[δ_4]); 176.8 (1C, **C**O-CH_2[α_1]); 175.0 (1C, **C**O-CH_3); 174.4-173.8 (8C, CO-NH); 157.1 (6C, NH-**C**(=NH)-NH_3^+); 54.7-53.8 (9C, CH[α_{1-9}]); 41.3 (6C, CH_2[ε_{1-3}] + CH_2[ε_{5-6}] + CH_2[ε_9]); 39.7 (1C, CH_2[δ_4]); 30.9 (9C, CH_2[β_{1-9}]); 27.8 (6C, CH_2[δ_{1-3}] + CH_2[δ_{5-6}] + CH_2[δ_9]); 26.9 (2C, CH_2[ε_{7-8}]); 25.7-25.0 (2C, CH_2[δ_{7-8}]); 22.7-22.5 (8C, CH_2[γ_{1-3}] + CH_2[γ_{5-9}]); 22.0 (1C, CH_3[Ac]); 21.9 (1C,

$CH_2[\gamma_4]$).

HRMS (ESI$^+$) (m/z) calculated for $C_{64}H_{128}N_{31}O_{11}$ $[M+H]^+$: 1507.03, found $[M+H]^+$: 1507.04

Peptide Tat(-1):

Yield: 20% (33 mg)

HPLC (A/B : from 95/5 to 85/15 in 20 min): Rt = 6.3 min

^{1}H NMR (500 MHz, D_2O) δ (ppm): 4.63-4.60 (t, 1H, $CH[\alpha_4]$); 4.38-4.25 (m, 8H, $CH[\alpha_{1-3}]$ + $CH[\alpha_{5-9}]$); 3.24-3.17 (m, 12H, $CH_2[\gamma_{1-3}]$ + $CH_2[\gamma_{5-6}]$ + $CH_2[\gamma_9]$); 2.96-2.95 (t, 4H, $CH_2[\delta_{7-8}]$); 2.77-2.70 (m, 2H, $CH_2[\beta_4]$); 2.08-1.89 (m, 17H, $CH_2[\beta_{1-3}]$ + $CH_2[\beta_{5-6}]$ + $CH_2[\beta_{8-9}]$ + CH_3[Ac]); 1.73-1.69 (m, 6H, $CH_2[\beta_7]$ + $CH_2[\gamma_{7-8}]$).

^{13}C NMR (125 MHz, D_2O) δ (ppm): 175.0 (1C, CO-CH_3 [Ac]); 174.1 (1C, **C**O-$CH_2[\beta_4]$); 173.0 (1C, CO-$CH[\alpha_4]$); 157.2 (6C, NH-C(=NH)-NH_3^+); 51.9-51.5 (8C, $CH[\alpha_{1-3}]$ + $CH[\alpha_{5-9}]$); 50.9 (1C, $CH[\alpha_4]$); 39.2 (2C, $CH_2[\delta_{7-8}]$); 38.1 (6C, $CH_2[\gamma_{1-3}]$ + $CH_2[\gamma_{5-6}]$ + $CH_2[\gamma_9]$); 36.5 (1C, $CH_2[\beta_4]$); 30.5-30.3 (8C, $CH_2[\beta_{1-3}]$ + $CH_2[\beta_{5-9}]$); 28.3-28.2 (2C, $CH_2[\beta_{7-8}]$); 23.7 (2C, $CH_2[\gamma_{7-8}]$); 22.1 (1C, CH_3[Ac]).

HRMS (ESI$^+$) (m/z) calculated for $C_{46}H_{92}N_{31}O_{11}$ $[M+H]^+$: 1254.75, found $[M+H]^+$: 1254.76

Peptide Tat(+1) :

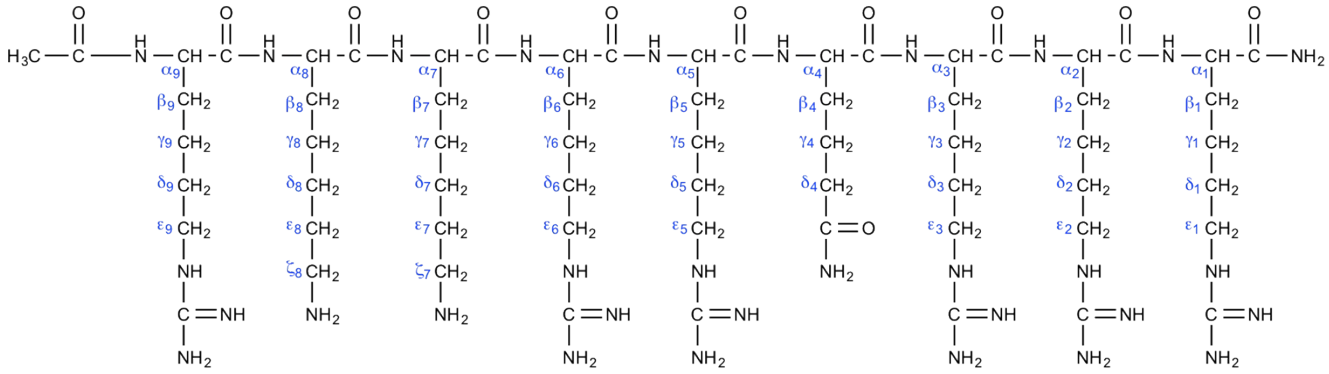

Figure S1. Structure of Tat(+1) and atom numbering for NMR attribution.

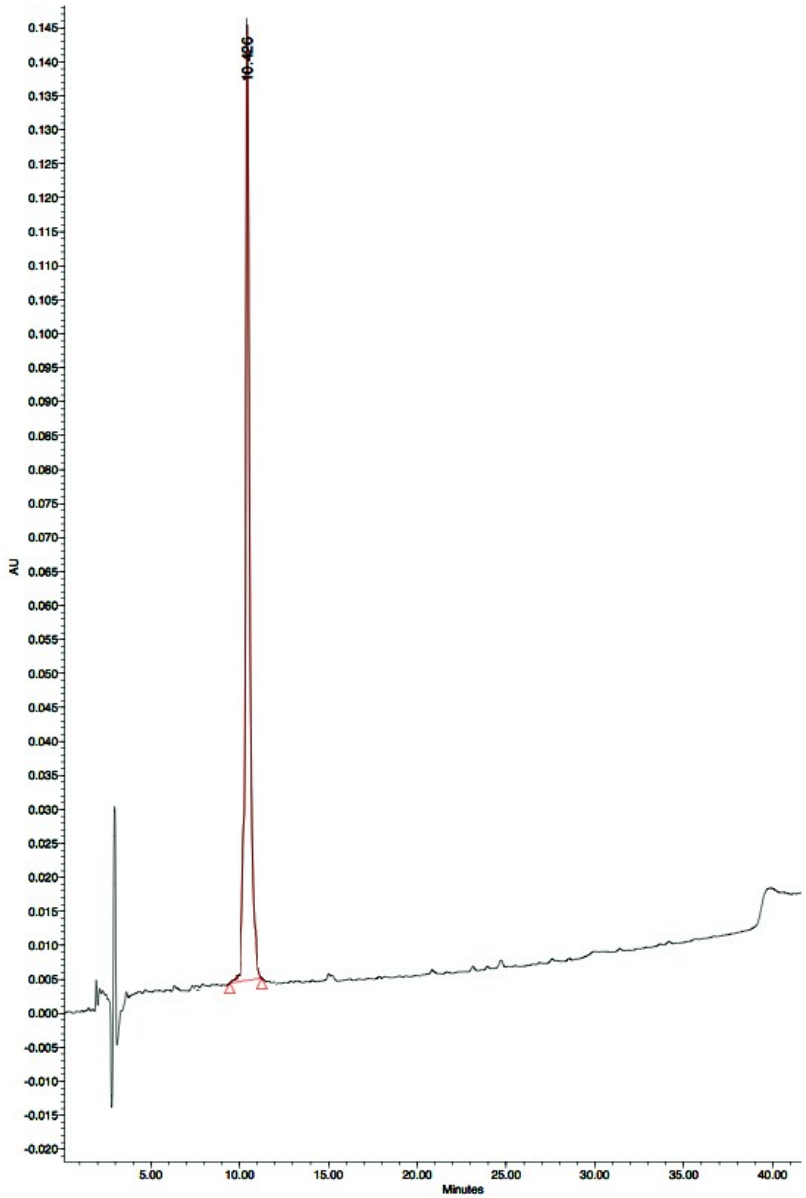

Figure S2. HPLC Spectra of Tat(+1). 0.1% TFA in water/ 0.1% TFA in acetonitrile: from 95/5 to 60/40 in 40 min; flow: 1 mL/min.

Tat(+1)

SM (ESI+) (m/z) : calculée pour $C_{64}H_{128}N_{31}O_{11}$ [M+H$^+$] = 1507,03, trouvée [M+H$^+$] = 1507,04

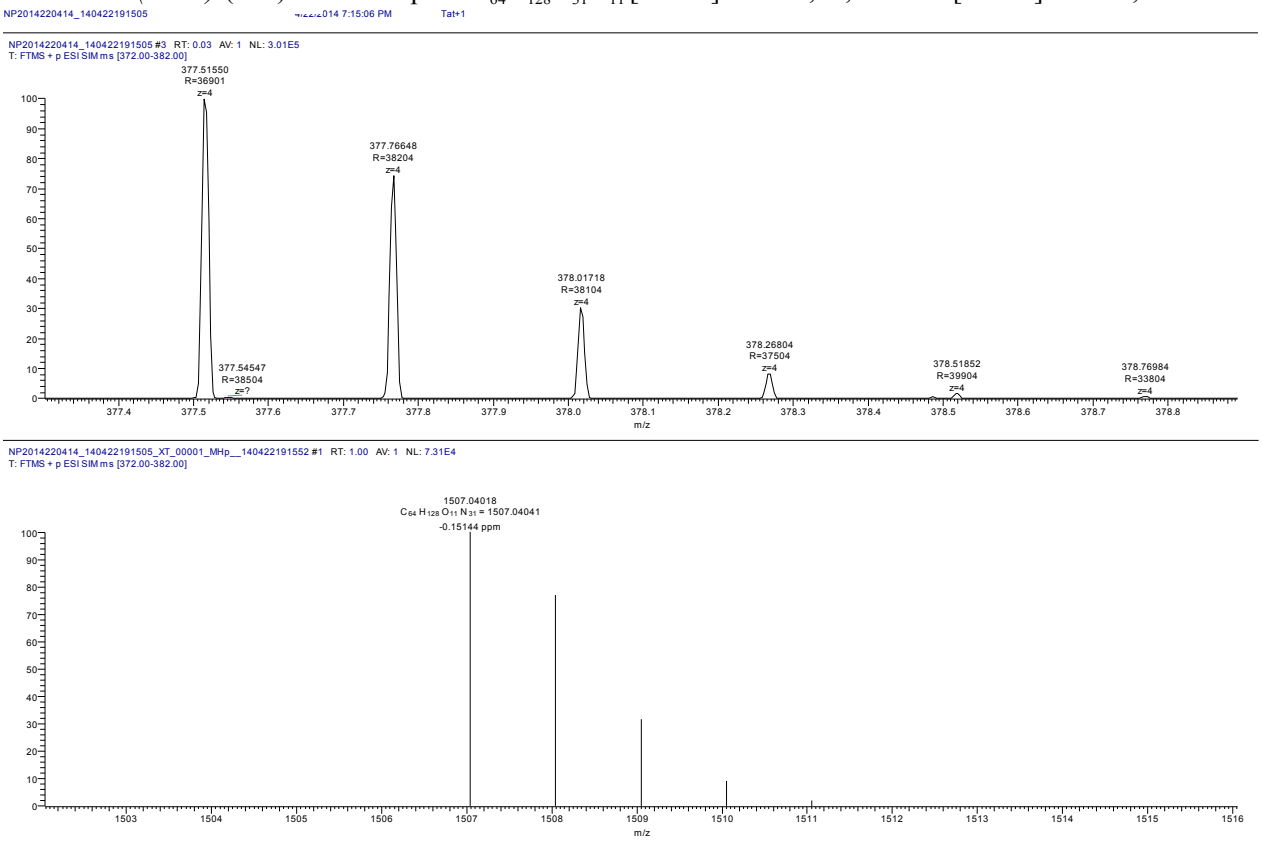

Figure S3. HRMS spectra of Tat(+1)

Peptide Tat(-1):

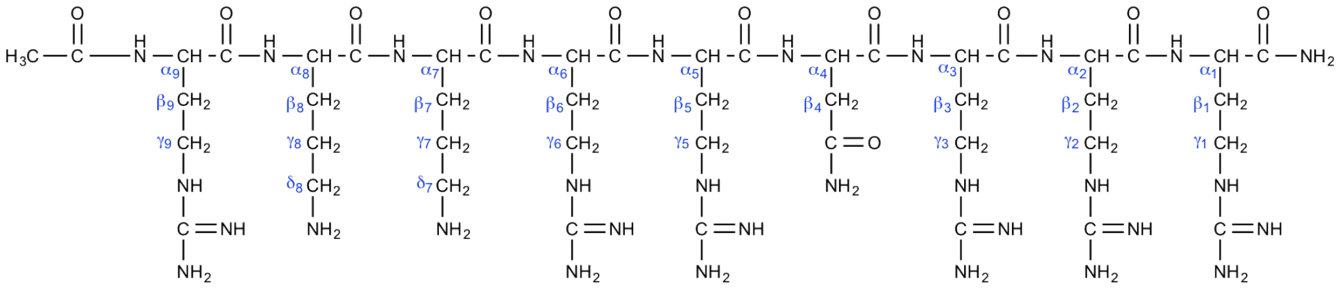

Figure S4. Structure of Tat(-1) and atom numbering for NMR attribution.

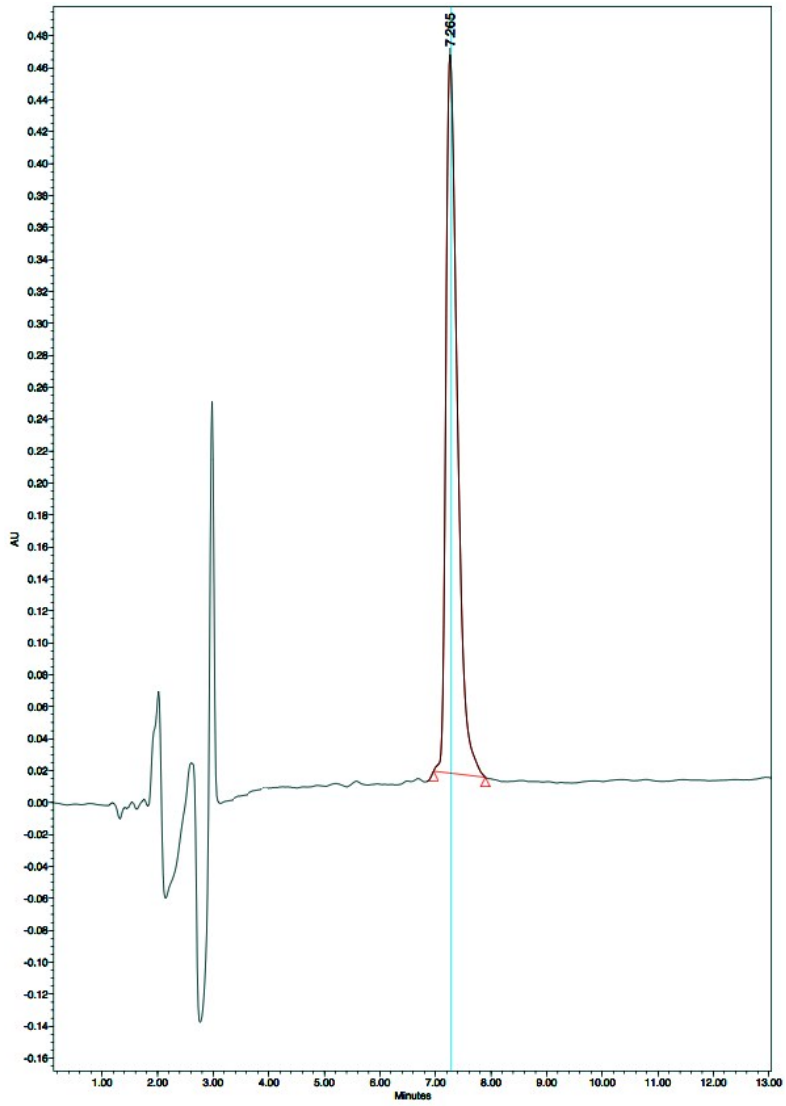

Figure S5. HPLC Spectra of Tat(-1). 0.1% TFA in water/ 0.1% TFA in acetonitrile: from 95/5 to 85/15 in 20 min; flow: 1 mL/min.

Tat(-1)

SM (ESI+) (m/z) : calculée pour $C_{46}H_{92}N_{31}O_{11}$ [M+H$^+$] = 1254,75, trouvée [M+H$^+$] = 1254,76

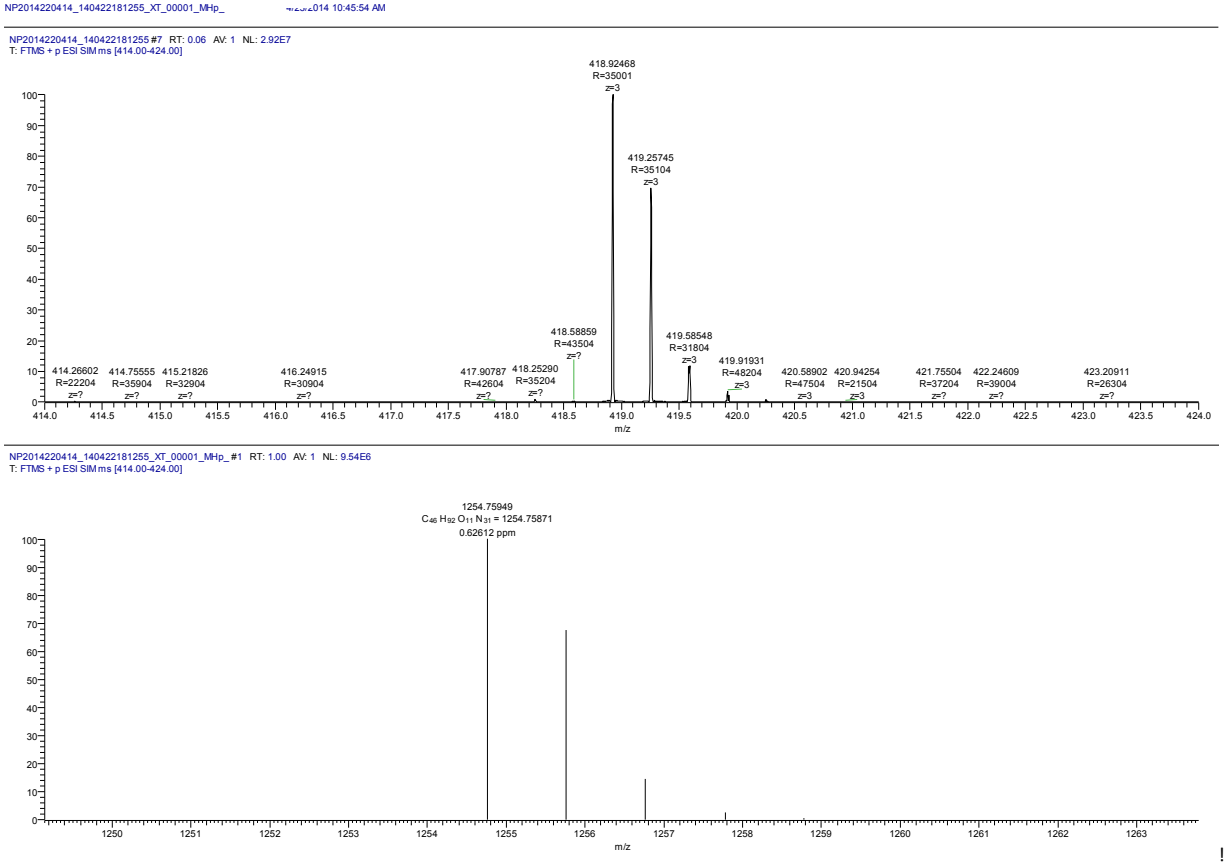

Figure S6. HRMS spectra of Tat(-1)