Supporting Information for:

Deciphering structure-activity relationships in a series of Tat/TAR inhibitors

Lise Pascale,[a] Alejandro López González,[b] Audrey Di Giorgio,[a] Marc Gaysinski,[a] Jordi Teixido Closa,[b] Roger Estrada Tejedor,[b] Stéphane Azoulay,[a] and Nadia Patino*[a]

[a] Institut de Chimie de Nice UMR7272, Université Nice Sophia Antipolis, 06108 Nice Cedex, France. [b] Molecular Design Lab., IQS School of Engineering, Universitat Ramon Llull, 08017 Barcelona, Spain.

*Corresponding author: patino@unice.fr; Tel: +33 (0)4 92 07 61 46; Fax: +33 (0)4 92 07 61 51
1 - CHEMISTRY

Characterization of PAA pentamers

Table S1: HPLC and HRMS data for individual penta-PAA

| Penta-AA | HPLC | HRMS m/z [M+H]^+ |
|----------|------|------------------|
|          | Retention time (min) | Purity* (%) | Molecular Formula | Calculated | experimental |
| FFFRF Ia | 11.6 | 68 | C_{72}H_{108}N_{20}O_{12} | 1445.85284 | 1445.85280 |
| KFFRF Ib | 10.6 | 92 | C_{69}H_{112}N_{21}O_{12} | 1426.87938 | 1426.87876 |
| RFFRF Ic | 10.7 | 76 | C_{69}H_{112}N_{23}O_{12} | 1454.88553 | 1454.88476 |
| FKFRF IIa | 10.5 | 89 | C_{69}H_{112}N_{21}O_{12} | 1426.87938 | 1426.87775 |
| KFFRF IIb | 9.7 | 82 | C_{66}H_{115}N_{22}O_{12} | 1407.90593 | 1407.89865 |
| RKFRF IIc | 9.8 | 90 | C_{66}H_{115}N_{24}O_{12} | 1435.91208 | 1435.90468 |
| FRFRF IIIa | 10.7 | 88 | C_{66}H_{112}N_{23}O_{12} | 1454.88553 | 1454.88222 |
| KRFRF IIIb | 9.9 | 86 | C_{66}H_{115}N_{24}O_{12} | 1435.91208 | 1435.91166 |
| RRFRF IIIc | 9.8 | 89 | C_{66}H_{114}N_{26}O_{12} | 1463.91823 | 1463.91788 |

*HPLC conditions: \( C_{18} \), H_{2}O (0.1% TFA)/Acetonitrile (0.1% TFA): 90/10 to 0/100 for 30 min; flow rate: 1 mL/min.

**HPLC purity of the crude product. All compounds were purified before testing.
HPLC analyses after semi-preparative HPLC purification

- PAA Ia

**Sample Information**

| SampleName       | Unknown       | Sample Type   | Unknown       |
|------------------|---------------|---------------|---------------|
| Vial             | 1             | Date Acquired | 17/02/11 13:43:13 |
| Injection        | 2             | Acq Method Set| Analytique Monomere |
| Injection Volume | 20.00 ul      | Processing Method| Analytique Monomere |
| Channel          | 996           | Date Processed | 21/02/11 18:35:47 |
| Run Time         | 47.0 Minutes  |               |               |

![Chromatogram](image)

**Instrument Method: Analytique Monomere**

*Stored: 10/10/10 16:09:01*

**W600 Gradient Table**

| Time (min) | Flow (ml/min) | %A | %B | %C | %D | Curve |
|------------|---------------|----|----|----|----|-------|
| 1          | 1.00          | 0.0| 0.0| 60.0| 10.0| 1     |
| 2          | 30.00         | 1.00| 0.0| 0.0| 100.0| 6     |
| 3          | 35.00         | 1.00| 0.0| 0.0| 100.0| 0     |
| 4          | 36.00         | 1.00| 0.0| 0.0| 90.0| 10.0| 6     |
• PAA Ib

Sample Information

| Parameter          | Value                  |
|--------------------|------------------------|
| SampleName         | L-Pentyl-PFKK 10 comm  |
| Vial               | 1                      |
| Injection          | 3                      |
| Injection Volume   | 20,00 µl               |
| Channel            | 996                    |
| Run Time           | 47.0 Minutes           |
| Sample Type        | Unknown                |
| Data Acquired      | 18/03/11 16:40:28     |
| Acq Method Set     | Analytique Monomere    |
| Processing Method  | Analytique Monomere    |
| Date Processed     | 18/03/11 16:22:15      |

Chromatogram

Retention Time (min) | Area | % Area | Height |
1                    | 10.697 | 8923493 | 100.0 | 1315224 |

Instrument Method: Analytique Monomere

W900 Gradient Table

| Time | Flow | %A | %B | %C | %D | Curve |
|------|------|----|----|----|----|-------|
| 1    | 1.00 | 0.0| 0.0| 99.0| 1.0|       |
| 2    | 38.00| 1.00| 0.0| 0.0| 99.0| 1.0 |   |
• PAA Ic

Sample Information

Sample Name: C-PentaL-FRFRF n 34 somi
Vial: 1
Injection: 2
Injection Volume: 20.00 ul
Channel: 999
Run Time: 47.0 Minutes

Sample Type: Unknown
Date Acquired: 11/05/11 14:59:24
Acq Method Set: Analytique Monomere
Processing Method: Analytique Monomere
Date Processed: 11/05/11 15:59:57

Chromatogram

| Retention Time (min) | Area   | % Area  | Height |
|----------------------|--------|---------|--------|
| 1                    | 10.681 | 13779685| 100.00 | 1989597 |

Instrument Method: Analytique Monomere

Stored: 02/05/11 11:58:06

W600 Gradient Table

| Time   | Flow | %A | %B | %C | %D | Curve |
|--------|------|----|----|----|----|-------|
| 1      | 1.00 | 0.0| 0.0| 60.0| 10.0| 6     |
| 2      | 30.00| 1.00| 0.0| 0.0| 60.0| 10.0| 6     |
| 3      | 35.00| 1.00| 0.0| 0.0| 0.0| 100.0| 6 |
| 4      | 36.00| 1.00| 0.0| 0.0| 0.0| 100.0| 6 |


• PAA IIa

Sample Information

Sample Name: L-C-Penta-L-FRFFK n 23 somr
Vial: 1
Injection: 9
Injection Volume: 20.00 ul
Channel: 996
Run Time: 47.0 Minutes

Sample Type: Unknown
Date Acquired: 21/02/11 17:40:59
Acq Method Set: Analytique Monomere
Processing Method: Analytique Monomere
Date Processed: 21/02/11 18:34:51

Chromatogram

Instrument Method: Analytique Monomere
Stored: 19/10/10 16:09:01

W600 Gradient Table

| Time | Flow | % A | % B | % C | % D | Curve |
|------|------|-----|-----|-----|-----|-------|
| 1    | 1.00 | 0.0 | 0.0 | 100.0 | 10.0 | 0     |
| 2    | 30.00| 1.00| 0.0 | 0.0 | 100.0 | 0     |

| Time | Flow | % A | % B | % C | % D | Curve |
|------|------|-----|-----|-----|-----|-------|
| 3    | 35.00| 1.00| 0.0 | 0.0 | 100.0| 0     |
| 4    | 30.00| 1.00| 0.0 | 0.0 | 100.0| 10.0  |
• PAA IIb

Sample Information

| SampleName          | L-PentaC-FRFK n 8 somnv       | Sample Type       | Unknown      |
|---------------------|--------------------------------|-------------------|--------------|
| Vial                | 1                              | Date Acquired     | 18/03/11 11:21:41 |
| Injection           | 1                              | Acq Method Set    | Analytique Monomere |
| Injection Volume    | 20,00 µl                       | Processing Method | Analytique Monomere |
| Channel             | 996                            | Date Processed    | 18/03/11 12:15:42 |
| Run Time            | 47,0 Minutes                   |                   |              |

Chromatogram

Instrument Method: Analytique Monomere

Stored: 19/10/10 16:08:01

W600 Gradient Table

| Time  | Flow | %A  | %B  | %C  | %D  | Curve |
|-------|------|-----|-----|-----|-----|-------|
| 1     | 1.00 | 0.0 | 0.0 | 90.0| 10.0|       |
| 2     | 30.00| 1.00| 0.0 | 0.0 | 100.0| 6     |
| 3     | 35.00| 1.00| 0.0 | 0.0 | 0.0  | 100.0 | 5     |
| 4     | 38.00| 1.00| 0.0 | 0.0 | 0.0  | 100.0 | 5     |

| Retention Time (min) | Area | % Area | Height |
|----------------------|------|--------|--------|
| 1                    | 9,902| 100.00 | 87,4092|
• PAA IIc

Sample Information

| Sample Name          | Unknown                                      |
|----------------------|----------------------------------------------|
| Vial                 | 1                                            |
| Injection            | 4                                            |
| Injection Volume     | 20.00 ul                                     |
| Channel              | 996                                          |
| Run Time             | 47.0 Minutes                                 |
| Date Acquired        | 10/05/11 16:20:09                            |
| Acq Method Sel       | Analytique Monomere                          |
| Processing Method    | Analytique Monomere                          |
| Date Processed       | 10/05/11 17:22:00                            |

Chromatogram

Instrument Method: Analytique Monomere

Stored: 02/05/11 11:58:06

W600 Gradient Table

| Time   | Flow | %A | %B | %C | %D | Curve |
|--------|------|----|----|----|----|-------|
| 1      | 35.00| 0  | 0  | 0  | 0  | 0     |
| 2      | 36.00| 0  | 0  | 0  | 0  | 0     |

| Time   | Flow | %A | %B | %C | %D | Curve |
|--------|------|----|----|----|----|-------|
| 3      | 35.00| 0  | 0  | 0  | 0  | 0     |
| 4      | 36.00| 0  | 0  | 0  | 0  | 0     |
- PAA IIIa

Sample Information

| Sample Name       | -C-Penta-L-FRFR n 24 smi |
|-------------------|--------------------------|
| Vial              | 1                        |
| Injection         | 12                       |
| Injection Volume  | 20,00 ul                 |
| Channel           | 996                      |
| Run Time          | 47.0 Minutes             |

Sample Type: Unknown
Date Acquired: 21/02/11 20:02:30
Acq Method Set: Analytique Monomere
Processing Method: Analytique Monomere
Date Processed: 21/02/11 21:04:20

Chromatogram

| Retention Time (min) | Area  | % Area | Height |
|----------------------|-------|--------|--------|
| 1                    | 10.735| 10433468 | 100.00 | 1788648 |

Instrument Method: Analytique Monomere
Stored: 19/10/10 16:09:01

W600 Gradient Table

| Time   | Flow | %A  | %B  | %C  | %D  | Curve |
|--------|------|-----|-----|-----|-----|-------|
| 1      | 1,00 | 0.0 | 0.0 | 60.0| 10.0|       |
| 2      | 30.00| 1,00| 0.0 | 0.0 | 100.0| 8     |
| 3      | 35.00| 1,00| 0.0 | 0.0 | 100.0| 8     |
| 4      | 38.00| 1,00| 0.0 | 0.0 | 90.0 | 10.0 | 8     |
• PAA IIIb

Sample Information

| SampleName       | L-Penta-C-FFFA N 9 cognition | Sample Type | Unknown |
|------------------|------------------------------|-------------|---------|
| Vial             | 1                            | Date Acquired | 16/03/11 15:42:49 |
| Injection        | 2                            | Acq Method Set | Analytique Monomere |
| Injection Volume | 20.00 ul                     | Processing Method | Analytique Monomere |
| Channel          | 995                          | Date Processed | 16/03/11 16:36:22 |
| Run Time         | 47.0 Minutes                 |              |         |

Chromatogram

Instrument Method: Analytique Monomere

Stored: 19/10/10 16:09:01

W600 Gradient Table

| Time  | Flow | %A  | %B  | %C  | %D  | Curve | Time  | Flow | %A  | %B  | %C  | %D  | Curve |
|-------|------|-----|-----|-----|-----|-------|-------|------|-----|-----|-----|-----|-------|
| 1     | 1.00 | 0.0 | 0.0 | 90.0| 10.0|       | 3     | 35.00| 1.00| 0.0 | 0.0 | 100.0| 6     |
| 2     | 30.00| 1.00| 0.0 | 0.0 | 100.0| 6     | 4     | 35.00| 1.00| 0.0 | 0.0 | 100.0| 6     |
• PAA IIIc

Sample Information

| SampleName          | -C-Penta-L-FRFR n 30 somi |
|---------------------|--------------------------|
| Vial                | 1                        |
| Injection           | 3                        |
| Injection Volume    | 20.00 ul                 |
| Channel             | 996                      |
| Run Time            | 47.0 Minutes             |
| Sample Type         | Unknown                  |
| Date Acquired       | 10/05/11 15:27:07        |
| Acq Method Set      | Analytique Monomere      |
| Processing Method   | Analytique Monomere      |
| Date Processed      | 10/05/11 17:36:04        |

Chromatogram

Instrument Method: Analytique Monomere

Stored: 02/05/11 11:58:06

| W/500 Gradient Table |
|-----------------------|
| Time  | Flow | %A | %B | %C | %D | Curve |
|-------|------|----|----|----|----|-------|
| 1     | 35.00| 1.00| 0.0| 0.0| 100.0| 6     |
| 2     | 30.00| 1.00| 0.0| 0.0| 100.0| 6     |
HRMS analyses

• PAA Ia

C:\Xcalibur\data\310511NP22  5/30/2011 4:40:18 PM

310511NP22 #48  RT: 1.11  AV: 1  NL: 3.19E5
F: FTMS + p ESI d Full ms2 723.43@cid21.00 [185.00-1460.00]

310511NP22_XT_00001_MHp_ #1  RT: 1.00  AV: 1  NL: 3.84E7
T: FTMS + p ESI Full ms [50.00-2000.00]
• PAA Ib
• PAA Ic

310511NP24 #62 RT: 1.43 AV: 1 NL: 1.68ES
F: FTMS + p ESI d Full ms2 727.94@cid21.00 [190.00-1470.00]

310511NP24_XT_00001_MHp_ #1 RT: 1.00 AV: 1 NL: 3.83E7
T: FTMS + p ESI Full ms [50.00-2000.00]
• PAA Ila

310511NP28 #61 RT: 1.41 AV: 1 NL: 1.11E5
F: FTMS + p ESI d Full ms2 713.94@cid21.00 [185.00-1440.00]

310511NP28_XT_00001_MHp_ #1 RT: 1.00 AV: 1 NL: 3.98E7
T: FTMS + p ESI Full ms [50.00-2000.00]
• PAA IIb
- PAA IIc
• PAA IIIa-MS2 on 4+ m/z 364.48
PAA IIIb

![Mass Spectrum](image1.png)

![Mass Spectrum](image2.png)
• PAA IIIc

310511NP36 #32 RT: 0.70 AV: 1 NL: 9.77E4
F: FTMS + p ESI d Full ms2 732.46@cel21.00 [190.00-1475.00]

310511NP36_XT_00001_MHp_ #1 RT: 1.00 AV: 1 NL: 2.95E7
T: FTMS + p ESI Full ms [50.00-2000.00]

C66H115O12N26 = 1463.91823
-0.24118 ppm
2 - TAR BINDING STUDIES

2.1 - NMR experiments: material and methods

High resolution NMR experiments were recorded on a BRUKER AVANCE Ultra shield DRX 500 spectrometer operating at 500.13 MHz for 1H, equipped with a temperature control unit (BCU 6.0, BVT 3000), and an inverse probe head (5mm PHTXI 1H-13C/15N Z-GRD). Proton chemical shift was referenced internally by setting the carrier frequency on water at the center of the spectrum (4.71 ppm at 13°C and 4.70 ppm at 35°C). Chemical shifts (δ) are expressed in parts per million (ppm). All NMR experiments were carried out using standard pulse sequences supplied by the spectrometer manufacturer (BRUKER). 1D and 2D spectra were processed using TOPSPIN 2.1 NMR Software (BRUKER).

For the preparation of all NMR samples, a folding of TAR RNA (100 µM) was performed in the appropriate buffer as described in material and methods section of the paper. After refolding, the NMR sample (alone or with the appropriate amount of PAA) is incorporated into a Shigemi NMR tube.

1H NMR imino proton spectra were recorded in a H2O/D2O (90/10) buffer containing 20 mM phosphate and 50 mM NaCl at 286°K (13°C) by using a WATERGATE 3-9-19 water suppression. Each proton NMR spectrum was acquired using 10.964 KHz Spectral Width (SW), 64K complex data point, acquisition time (aq) of 2.98 s, relaxation delay (D1) of 1s, number of scan (ns) between 1000 and 2000, number of dummy scan (ds) 4 and a 90° flip angle pulse width. Water suppression was achieved using WATERGATE pulse sequence. Gradient pulse were sine shape (SINE.100), 1.5 ms long (P16) with 100 µs gradient recovery delay (D16) and strengths set to 8.44 Gauss.cm-1 (20%). A 45.6 µs delay (D19) was used for binomial water suppression. Prior to Fourier transformation, the fids were multiplied by an exponential line broadening function of 3 Hz.

gs-TOCSY Phase sensitive (States – TPPI mode) experiments using MLEV 17 pulse sequence for spin lock were recorded in a D2O buffer (50 mM NaCl, 20 mM phosphate, pH 7.4) at 308K (35°C) by using a WATERGATE 3-9-19 water suppression. Each TOCSY 2D NMR spectrum was acquired with a spectral width of 5 KHz in both dimension, 2K complex data point in F2, 256 t1 increments (between 32 and 64 scans by increment ) in F1, 0.20 s for aq and D1 of 2 s. MLEV 17 pulse sequence for spin lock was set to 60 ms. Water suppression was achieved using WATERGATE pulse sequence. Gradient pulse were sine shape (SINE.100), 1.5 ms long (P16) with 100 µs gradient recovery delay (D16) and strengths set to 8.44 Gauss.cm-1 (20%). A 100 µs delay (D19) was used for binomial water suppression. Prior to Fourier transformation a QSINE window function (SSB =2) was applied in both dimension and the data were zero filled and linear predicted (NC=32) to 1K data points in F1.

2.2 - Circular Dichroism and UV melting studies
2.3 - Thermodynamic analyses

Comment on statistical validity for EEC phenomena: It has been argued that the apparent compensation may be in many cases a statistical artifact resulting from experimental uncertainties in the measured thermodynamic data collected using the van't Hoff analysis. Krug proposed a simple statistical test to determine whether the observed compensation is relevant or not by determining the confidence interval for the compensation temperature (Tc).\(^1\) If the experimental temperature Texp lies outside this confidence interval, the correlation would be significant at the 95% confidence level. In our case, Texp falls always outside the confidence limits, which could indicate that the observed phenomenon has a biological relevance.

3 - MOLECULAR MODELING

3.1 - Protocols and inputs

AutoDock Vina docking protocol

---

\(^1\) Sharp, K. (2001) Entropy-enthalpy compensation: fact or artifact? Protein Sci, 10, 661.
Docking of the PAAs on HIV-1 TAR X-ray structure (PDB ID 2kx5(16)) was performed with AutoDock Vina software. First, the TAR structure was prepared by trimming the peptide from the original pdb file. Then, the RNA was prepared for docking by using the automated preparation protocol of AutoDock Tools (ADT), which includes protonation of the structure and Gasteiger partial charges assignment. The following input file was used for running all docking calculations:

```
AutoDock Vina input
---------------------------------
receptor = tar.pdbqt
ligand = paa.pdbqt
center_x = -28.779
center_y = 63.327
center_z = -53.986
size_x = 36.75
size_y = 36.75
size_z = 28.5
out = vina.pdbqt
---------------------------------
```

Autodock Vina was run 10 times for each PAA and only the 40 most energetically favorable conformations were selected and clustered with ADT. A representative conformation from the best scoring docking was selected and subjected to molecular dynamics.

**Molecular dynamics protocol**

The molecular dynamics simulations were run with pmemd.MPI. The results were visualized and analyzed with VMD and Pymol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.). The `closest` command from `cpptraj` module was used for stripping all except the closest N waters from the trajectory. All the MM/GBSA calculations were run in a standard basis.

---

2 Trott, O. and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, 31, 455.

3 Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J Mol. Graph.*, 14, 33.