

## Supporting Information

### Self-aggregation propensity of Tat peptide revealed by UV-Vis, NMR and MD analyses

Sara Macchi<sup>a</sup>, Riccardo Nifosi<sup>a</sup>, Giovanni Signore<sup>a,b</sup>, Sebastiano Di Pietro<sup>b</sup>, Claudia Boccardi<sup>b</sup>,  
Francesca D'Autilia<sup>b</sup>, Fabio Beltram<sup>a</sup>, and Francesco Cardarelli<sup>b,\*†</sup>

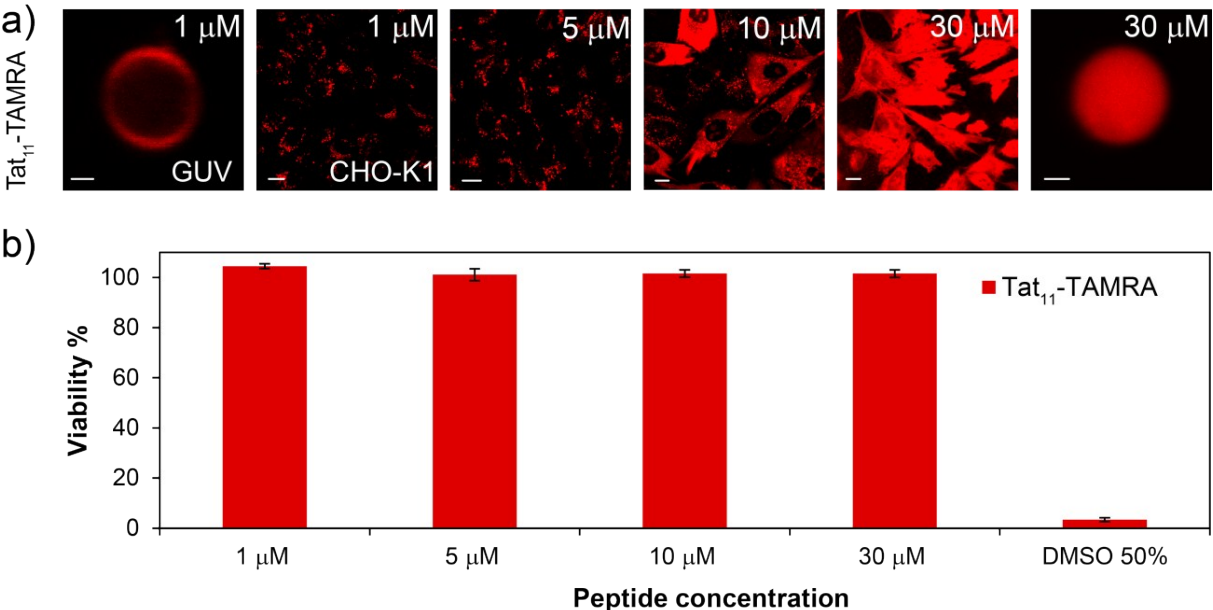
<sup>a</sup>NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, Piazza San Silvestro 12 - 56127  
Pisa, Italy

<sup>b</sup>Center for Nanotechnology Innovation @NEST, Istituto Italiano di Tecnologia, Piazza San  
Silvestro 12-56127 Pisa, Italy

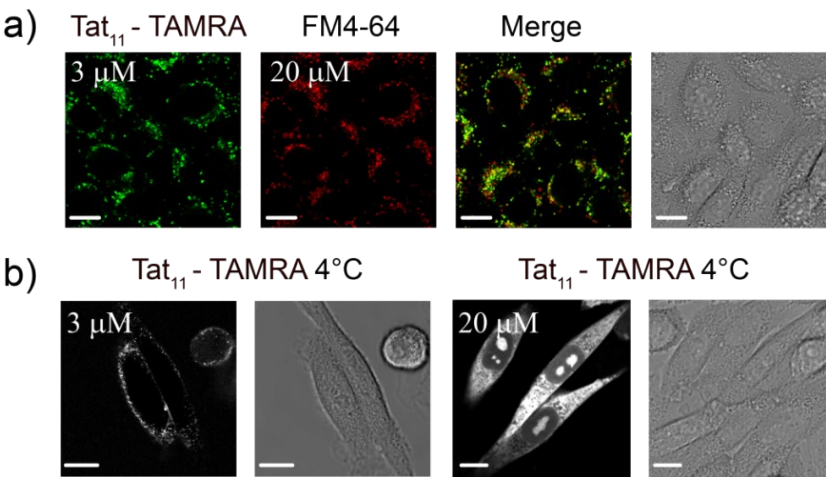
\* To whom correspondence should be addressed: francesco.cardarelli@nano.cnr.it

†present address: NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, Piazza San  
Silvestro 12 - 56127 Pisa, Italy

# Section S1: Figures and Tables of CPPs



**Fig. S1:** Cell behavior and viability test for Tat<sub>11</sub>-TAMRA (a) GUV and cell uptake of Tat<sub>11</sub>-TAMRA at increasing concentrations (1 and 5  $\mu$ M: endocytosis; 10 and 30  $\mu$ M: transduction). Cells scale bar: 10  $\mu$ m; GUVs scale bar: 1  $\mu$ m. (b) Cell metabolic activity of Tat<sub>11</sub>-TAMRA measured with WST-8 assay. In general, cells remain viable in all the concentration range tested upon 30 min of incubation time. Here untreated cells are defined as 100% viable (not shown), while cells exposed to 50% dimethyl sulfoxide (DMSO) are used as positive control for a decreased metabolic activity.

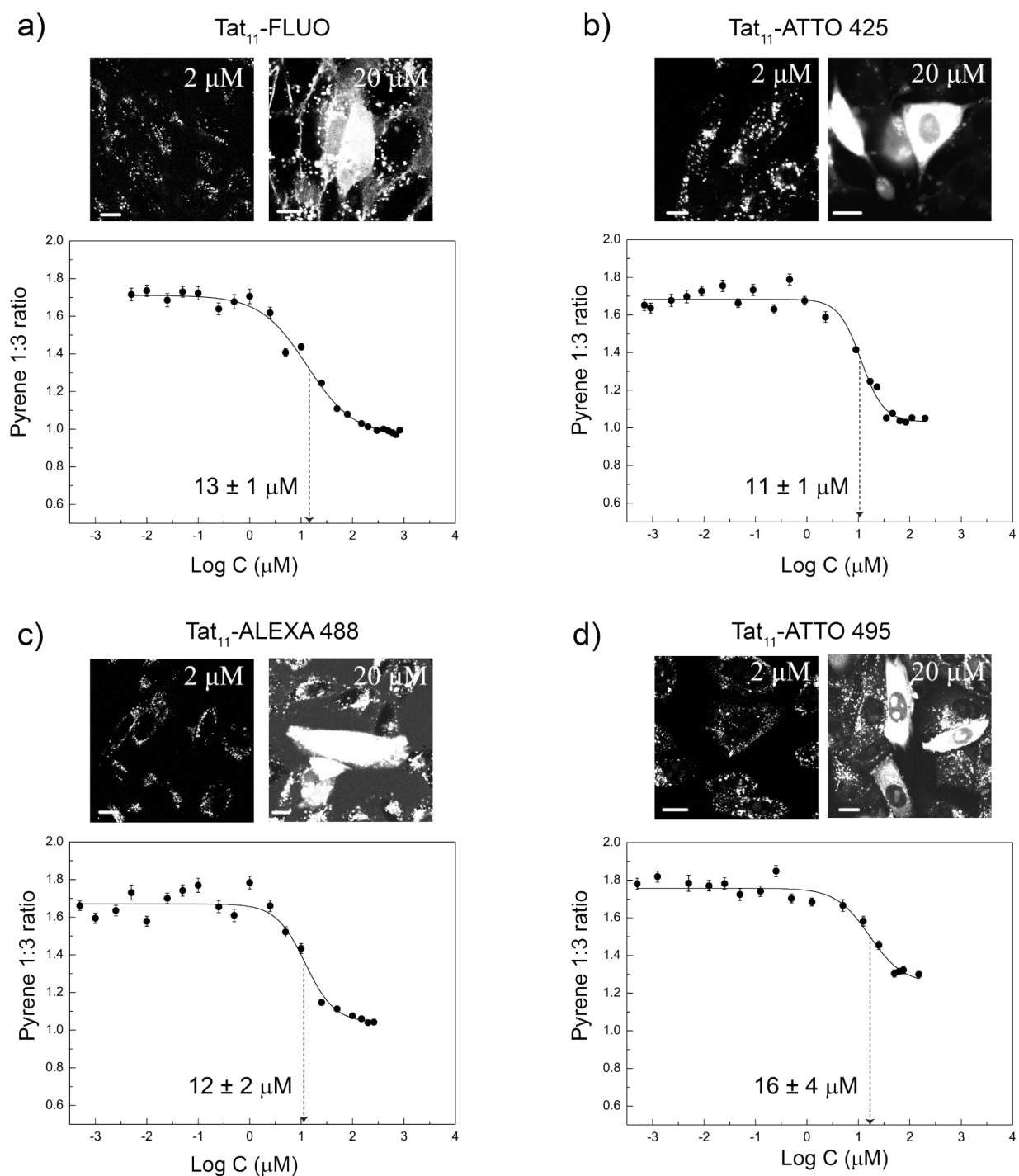


**Fig. S2:** Endocytosis assay and cell uptake in CHO-K1 cells at 4°C of TAMRA-labeled Tat<sub>11</sub>. (a) Confocal images of Tat<sub>11</sub>-TAMRA at 3  $\mu$ M co-treated with FM4-64 at 20  $\mu$ M. In both channels only bright spots are detected which are perfectly superimposed in the merge channel, thus confirming that cell uptake mechanism is endocytosis. Scale bars: 10  $\mu$ m (b) Confocal images of

TAMRA-labeled peptide administered to cells at 3  $\mu$ M and 20  $\mu$ M (left and right columns, respectively) at 4°C. The experiment confirms that at this temperature endocytosis has been inhibited, as in the first case endocytosis is strongly reduced (less number and brightness of endocytic bright spots), while in the second case direct translocation is the dominant uptake mechanism (cell cytoplasm is full of fluorescent peptide). Scale bars: 10  $\mu$ m.

Peptide (-dye)	m/z (charge)
<b>Tat<sub>11</sub></b>	313.0 (5); 391.0 (4); 521.0 (3); 559.0 (3+K <sup>+</sup> )
<b>Tat<sub>11</sub>-Cys</b>	333.6 (5); 416.7 (4); 555.3 (3)
<b>Tat<sub>11</sub>-TAMRA</b>	358.3 (6); 429.6 (5); 537.0 (4); 715.4 (3)
<b>Tat<sub>11</sub>-FLUO</b>	349.6 (6); 495.5 (5); 524.1 (4); 698.3 (3)
<b>Tat<sub>11</sub>-ATTO 425</b>	438.2 (5); 547.2 (4); 729.6 (3)
<b>Tat<sub>11</sub>-ALEXA 488</b>	591.4 (4); 788.2 (3); 826.2 (3+K <sup>+</sup> )
<b>Tat<sub>11</sub>-ATTO 495</b>	357.7 (6); 428.3 (5); 534.8 (4); 713.2 (3)
<b>R9</b>	285.7 (5); 356.9 (4); 475.6 (3)
<b>R9-Cys</b>	382.7 (4); 510.0 (3); 548.0 (3+K <sup>+</sup> )
<b>Ant</b>	450.4 (5); 562.7 (4); 750.0 (3)
<b>Ant-Cys</b>	392.5 (6); 470.8 (5); 588.3 (4); 784.9 (3)
<b>R9-TAMRA</b>	336.2 (6); 402.6 (5); 502.8 (4)
<b>R9-FLUO</b>	326.5 (5); 489.2 (4); 652.3 (3)
<b>R9-ATTO 495</b>	334.2 (6); 401.0 (5); 500.8 (4)
<b>Ant-TAMRA:</b>	405.6 (7); 472.9 (6); 567.2 (5); 708.5 (4); 944.2 (3)
<b>Ant-FLUO</b>	463.7 (6); 556.3 (5); 695.0 (4); 926.3 (3)
<b>Ant-ATTO 495</b>	471.5 (6); 565.4 (5); 706.6 (4); 941.4 (3)

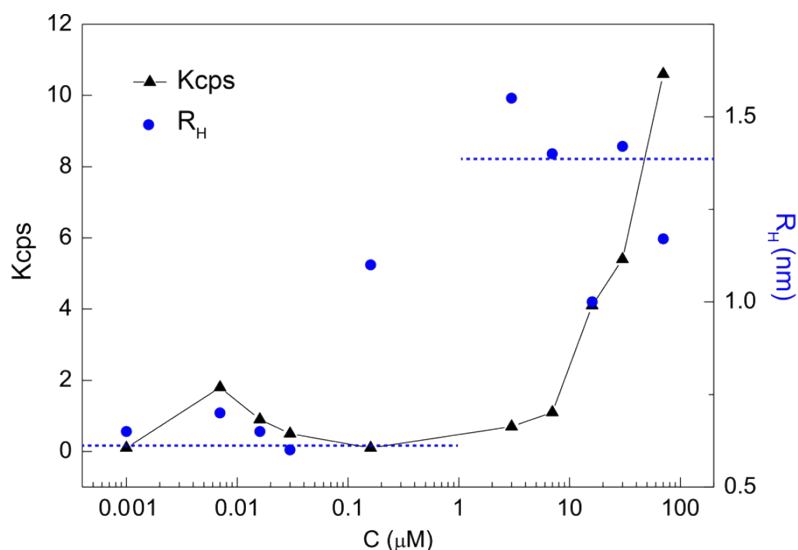
**Table S1:** Mass spectrum ions of all peptides used (peptide sequences are indicated in Section S2).



**e)**

Dye (charge)	FLUO (q=-2)	ATTO 425 (q=0)	ALEXA 488 (q=-2)	ATTO 495 (q=+1)
Dye structure				

**Fig. S3:** Cell uptake behaviour and self-aggregation tendency of dye-labeled Tat<sub>11</sub> peptide together with dyes' main features. (a – d) Upper panel: confocal microscopy images of dye-labeled Tat<sub>11</sub> in CHO-K1 cells at concentration below (2  $\mu$ M) and above (20  $\mu$ M) the center of the sigmoid. Cell scale bars: 10  $\mu$ m. Lower panel: pyrene 1:3 ratio data (points) and Boltzmann fitting curve (solid line). Dashed arrows point out the center of the sigmoid (indicated in each figure with standard error). Vertical bars: standard errors. (e) Net charges and structures of the dyes used at pH 7.

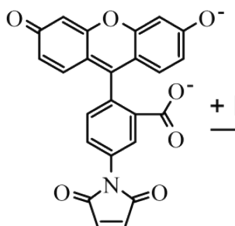
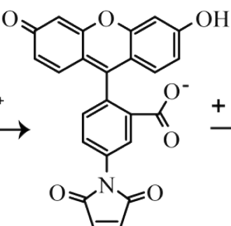
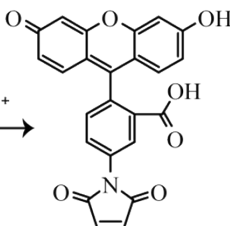


**Fig. S4:** Dynamic light scattering results for Tat<sub>11</sub>-TAMRA in 0.001-100  $\mu$ M concentration range. Hydrodynamic radius values ( $R_H$ , blue points) are shown with Kilo counts per seconds (Kcps, black triangles). The average of  $R_H$  values below 1  $\mu$ M (about 0.7 nm, dashed line on the left) is less than the average above this concentration (about 1.3 nm, dashed line on the right), thus indicating an aggregation step. Moreover the  $R_H$  value doesn't increase at high concentrations, underlying that only a single species is present (dimer, see the main text). At the same time, a linear increase of Kcps values from 1  $\mu$ M on indicates that the number of elements of dimer species is becoming higher.

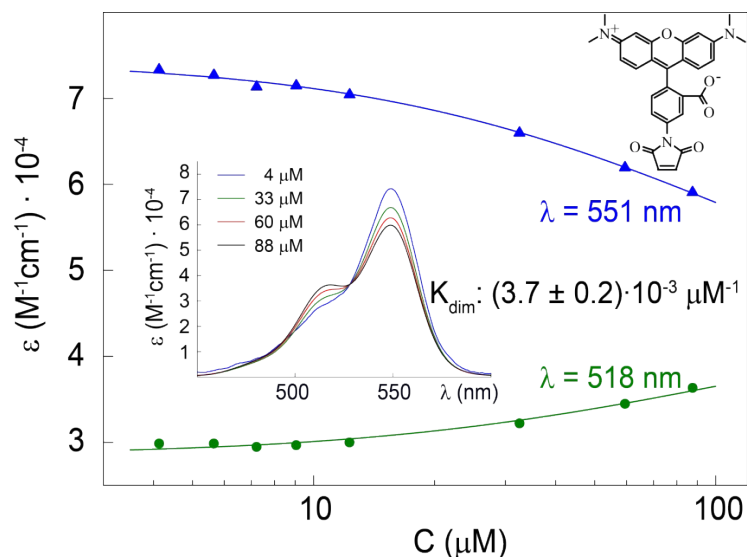
Tat <sub>11</sub> - Dye	- FLUO	- ATTO 425	- ALEXA 488	- ATTO 495
$R_H$ (nm)	180 $\pm$ 20 *	1 - 10	1 - 10	1 - 10

\*This value is explained in detail in Fig. S5

**Table S2:** Hydrodynamic radius  $R_H$  (nm  $\pm$  standard errors, from DLS) of dye-labeled Tat<sub>11</sub>.

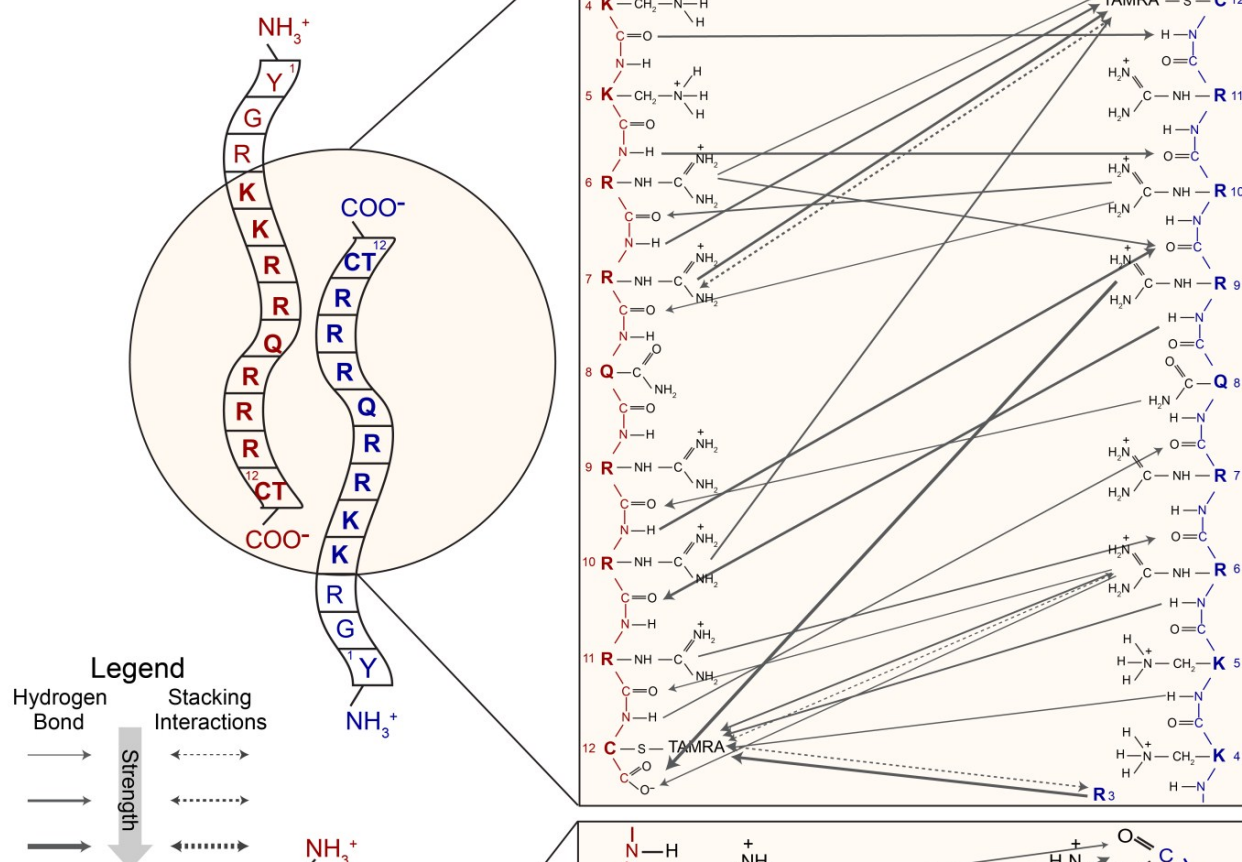
FLUO:			
	$+ H^+$	$+ H^+$	
pH:	7	5	3
Charge:	-2	-1	0
Structure:	Dianion	Anion	Neutral
Tat <sub>11</sub> - FLUO R <sub>H</sub> (nm):	180 ± 6	1 - 10	1 - 10

**Fig. S5:** Different features of FLUO dye and nanoparticle radius of Tat<sub>11</sub>-FLUO (from DLS) for decreasing pH values. The structure of the dye shows two net negative charges at pH 7, one net negative charge at pH 5 and no negative charges at pH 3. Interestingly, at both pH 5 and 3 Tat<sub>11</sub> aggregates come back to smaller values (R<sub>H</sub>: 1-10 nm), thus indicating that the loss of the first and then of the second acidic proton of FLUO dye strictly regulates the aggregate size<sup>1</sup>.

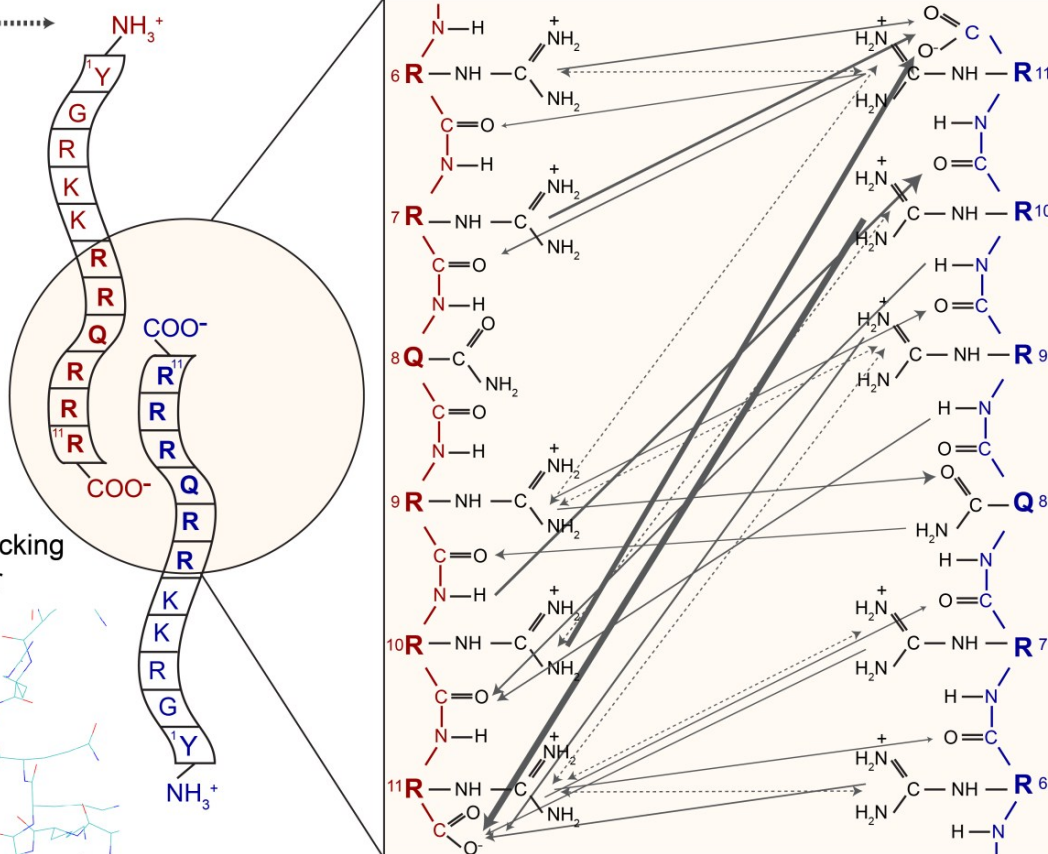


**Fig. S6:** TAMRA extinction coefficient data (points and triangles) and fitting curves (solid lines) for increasing Tat<sub>11</sub>-TAMRA concentrations are shown for  $\lambda=518$  nm (green) and  $\lambda=551$  nm (blue). The dimerization constant  $K_{dim}$  is obtained by the global analysis (see the main text and Section S2). Right top panel: TAMRA dye chemical structure. Central panel: extinction coefficients of TAMRA at 4  $\mu$ M (blue line), 33  $\mu$ M (green line), 60  $\mu$ M (red line) and 88  $\mu$ M (black line) for increasing  $\lambda$  values (nm).

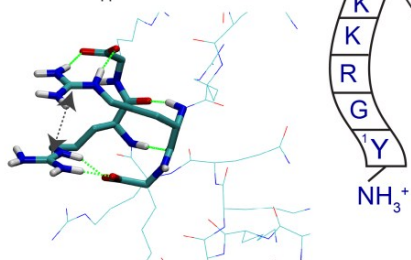
**a) Tat<sub>11</sub>-TAMRA dimer**



**b) Tat<sub>11</sub> dimer**



**c) R10-R10 stacking in Tat<sub>11</sub> dimer**





**Fig. S7:** Scheme of the most relevant interactions holding the peptides together in the dimeric structures. Left side: representation of the dimers of Tat<sub>11</sub>-TAMRA (a) and of Tat<sub>11</sub>(b). N- and C-terminal charged regions are indicated for each peptide. The first ('1' for both peptides) and the last ('11' for Tat<sub>11</sub> and '12' for Tat<sub>11</sub>-TAMRA) residues are numbered. Right side: zoom on the C-terminal regions of the dimeric structures of Tat<sub>11</sub>-TAMRA (a) and of Tat<sub>11</sub>(b), in which hydrogen bonds (solid arrows) and Arginine-Arginine/Arginine-TAMRA stacking interactions (dotted arrows) are indicated. In the legend, "strength" indicates the percentage of the simulation time for which the interaction is present (see Table S3). (c) Representation of stacking interaction (dotted arrow), lasting for ~30% of the simulated time, between the Arginine residues at position 10 on each Tat<sub>11</sub> peptide. Hydrogen bonds (inter backbone and between R10 guanidinium groups and R11 carboxyl C-terminal groups) are in green. For further explanation, see Section S3.

**a) Tat<sub>11</sub> – dim a**

		<b>R6</b>		<b>R7</b>		<b>Q8</b>		<b>R9</b>		<b>R10</b>		<b>R11</b>	
		bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc
<b>R6</b>	bb											0.20	
	sc											0.36	
<b>R7</b>	bb											0.06	0.20
	sc											0.17	
<b>Q8</b>	bb							0.16				0.05	
	sc							0.10				0.07	
<b>R9</b>	bb							0.18	0.34				
	sc											0.59	
<b>R10</b>	bb									1.45			
	sc											1.80	
<b>R11</b>	bb		0.36		0.92					1.57			
	sc	0.12		0.44									

**b) Tat<sub>11</sub> – dim a\***

		<b>R3</b>		<b>K4</b>		<b>K5</b>		<b>R6</b>		<b>R7</b>		<b>Q8</b>		<b>R9</b>		<b>R10</b>		<b>R11</b>	
		bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc
<b>K4</b>	bb																	0.06	
	sc																	0.10	
<b>K5</b>	bb																		
	sc																		
<b>R6</b>	bb																	0.14	
	sc												0.09					0.55	
<b>R7</b>	bb																	0.35	
	sc																	0.91	
<b>Q8</b>	bb																		
	sc																	0.06	
<b>R9</b>	bb											0.17		0.10					
	sc											0.06		0.15					
<b>R10</b>	bb													0.40		1.38			
	sc																	1.34	
<b>R11</b>	bb		0.29					0.66		0.27				0.36		1.92			
	sc							0.37		0.27									



c) Tat<sub>11</sub> – *dim b*

		R3		K4		K5		R6		R7		Q8		R9		R10		R11	
		bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc
R3	bb									0.09						0.21			
	sc																	0.30	
K4	bb									0.20									
	sc																		
K5	bb											0.36		0.13					
	sc																	0.19	
R6	bb															0.11			
	sc					0.38													
R7	bb							0.58											
	sc							0.07		0.28		0.13				0.09		0.23	
Q8	bb					0.16				0.14		0.14							
	sc							0.19				0.06				0.31			
R9	bb																		
	sc																		
R10	bb							0.20											
	sc																	0.46	
R11	bb		<b>0.65</b>					0.16	1.16										
	sc	0.25																	

d) Tat<sub>11</sub> - TAMRA – *dim a*

		G2		R3		K4		K5		R6		R7		Q8		R9		R10		R11		CT12	
		bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc
Y1	bb																						
	sc																						0.06
G2	bb																						
	sc																						
R3	bb																						
	sc																						<b>0.95</b>
K4	bb																						0.05
	sc																						
K5	bb																			0.07		0.16	
	sc																				0.08		
R6	bb																	0.06		0.37		0.48	
	sc																		0.08	0.12	0.48		
R7	bb																				0.25		
	sc																				0.06		
Q8	bb																						
	sc														0.11		0.10						
R9	bb									0.40								<b>1.56</b>					
	sc																			<b>1.11</b>			
R10	bb									<b>0.94</b>													
	sc									0.48		0.10											
R11	bb									0.09													
	sc																						
CT12	bb			0.24	<b>0.63</b>					0.22													
	sc	0.05		0.05	0.28						<b>0.67</b>	<b>0.86</b>						0.54					

e) Tat<sub>11</sub> - TAMRA – *dim b*

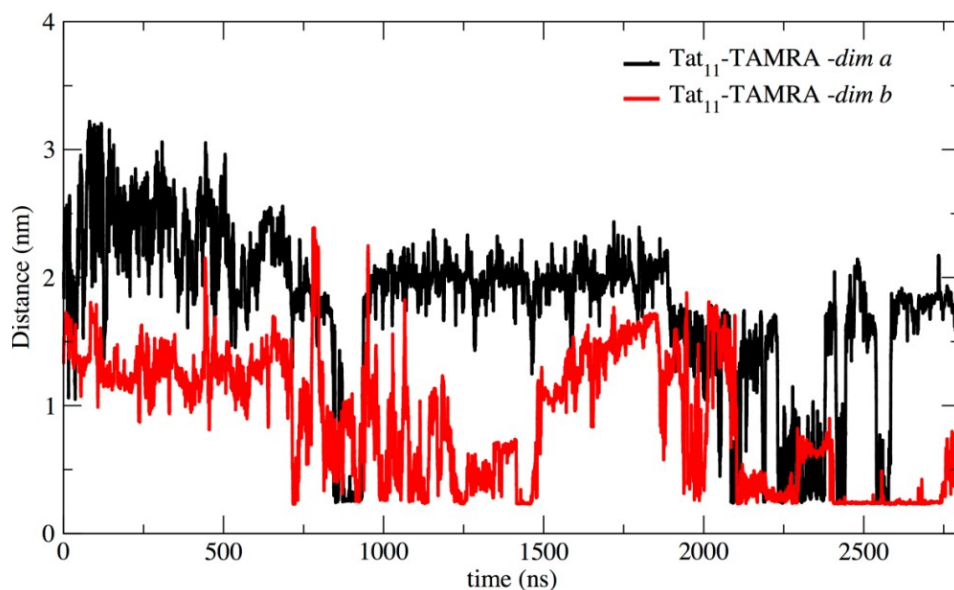
		Y1		G2		R3		K4		K5		R6		R7		Q8		R9		R10		R11		CT12	
		bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc	bb	sc
R3	bb																								
	sc																							0.07	0.09
K4	bb																								
	sc																								
K5	bb																								
	sc																								
R6	bb																							0.05	
	sc																							0.19	0.12
R7	bb																								
	sc																							0.21	0.44
Q8	bb																								0.11
	sc																				0.05	0.08			
R9	bb																							0.11	
	sc																							0.37	
R10	bb																				0.06			0.30	
	sc																						0.14	0.17	
R11	bb																				0.58	0.05		0.30	
	sc																				0.27		0.28	0.27	
CT12	bb					0.09							0.10								0.57			0.09	
	sc	0.08				0.05	0.10	0.11				<b>0.76</b>	0.09	0.09				<b>1.01</b>	0.18	0.83					

f) Arginine-Arginine Stacking

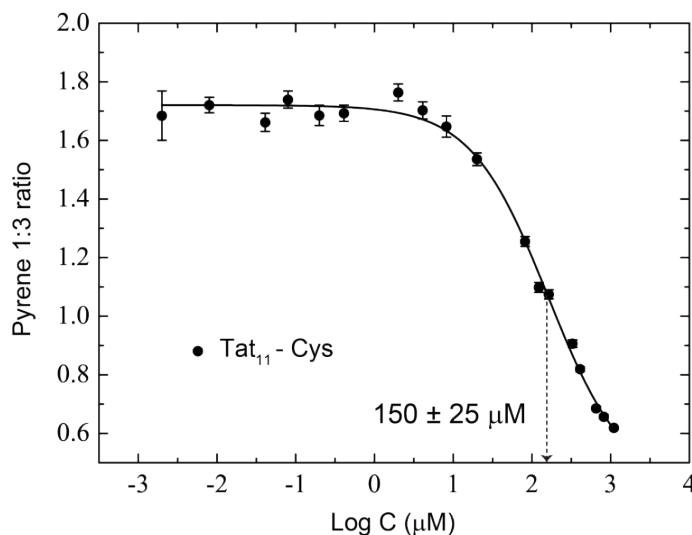
Tat <sub>11</sub> – <i>dim a</i>			Tat <sub>11</sub> – <i>dim a*</i>			Tat <sub>11</sub> – <i>dim b</i>			Tat <sub>11</sub> -TAMRA - <i>dim a</i>			Tat <sub>11</sub> -TAMRA - <i>dim b</i>		
Involved residues		Occ.	Involved residues		Occ.	Involved residues		Occ.	Involved residues		Occ.	Involved residues		Occ.
R10	R10	0.29	R10	R10	0.37	R7	R6	0.24	R7	TMR	0.65	R6	TMR	0.27
R11	R9	0.18	R11	R9	0.28	R6	R11	0.11	TMR	R3	0.31	R10	R9	0.20
R6	R11	0.17	R11	R6	0.11	R10	R3	0.11	R3	TMR	0.10	R9	TMR	0.14
R9	R11	0.16	R11	R7	0.08	R3	R11	0.07	TMR	R6	0.09	R10	R11	0.12
R11	R6	0.09	R7	R11	0.07	R7	R3	0.06				R11	R10	0.07
R11	R7	0.09	R6	R11	0.06							TMR	R9	0.05
R9	R9	0.07	R9	R9	0.05							TMR	R3	0.05

**Table S3:** (a-e) Number of inter peptide H-bonds, averaged during the trajectory. For each pair of residues, the table reports the H-bonds occurring within the backbone (bb), within the side chains (sc) and sc to bb or viceversa. Only H-bonds with occupation number higher than 0.05 are reported, and values higher than 0.6 are in bold. (f) List of inter peptide Arginine-Arginine and Arginine-TAMRA stacking (Occ. stands for occupancy during the simulation, see Section S4 for the geometric criteria used). The amino acid numbering in the tables follows the scheme represented in

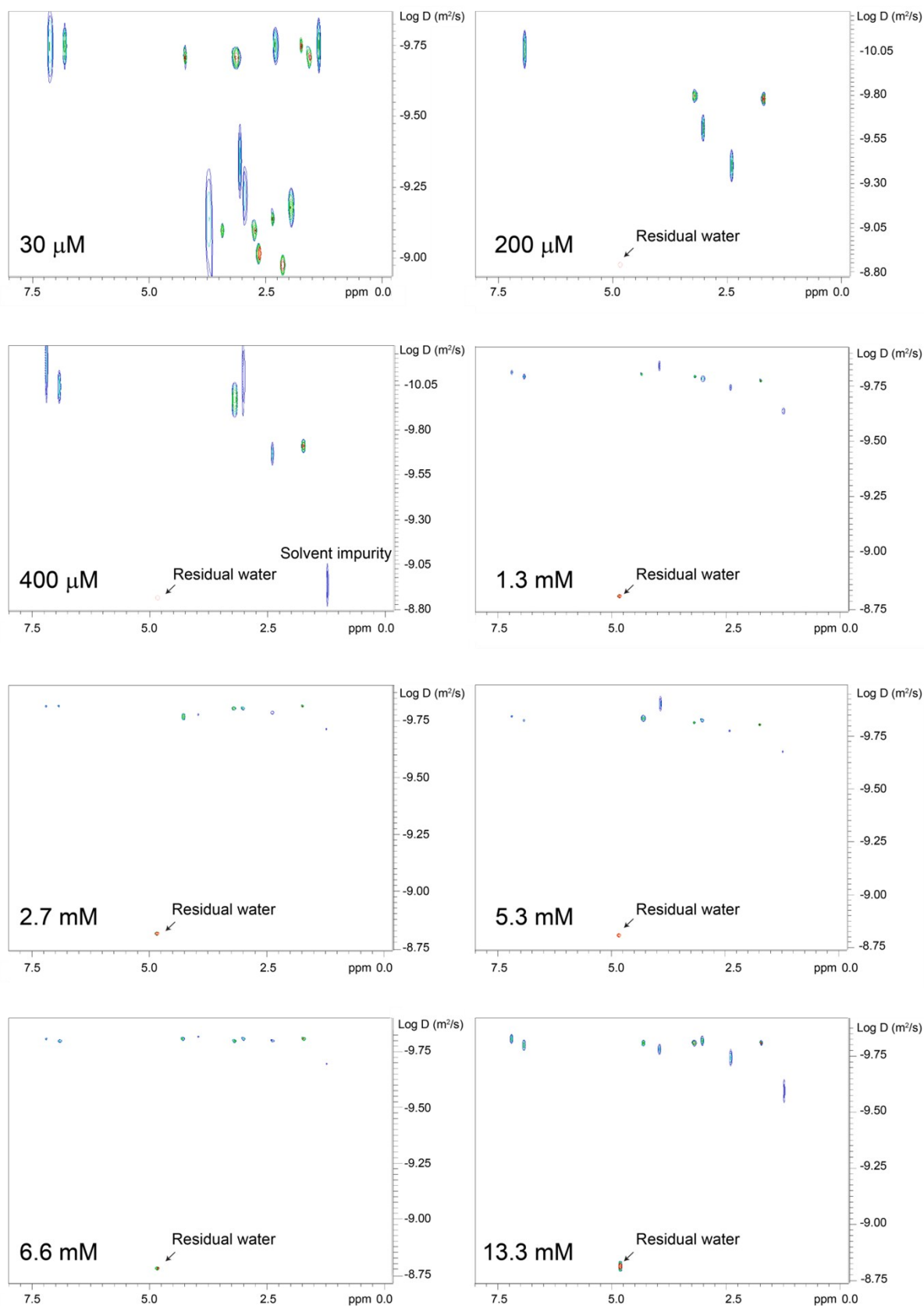
Fig. S7. Tat<sub>11</sub>-*dim a* and Tat<sub>11</sub>-TAMRA-*dim a* are the simulations discussed in the main text. The additional simulations (*dim a\**, *dim b*) are discussed in Section S3.



**Fig. S8:** Minimum distance (100-ps running average) between the TAMRA fluorophores in the two Tat<sub>11</sub>-TAMRA dimer simulations. The additional simulations (*dim a\**, *dim b*) are discussed in Section S4.

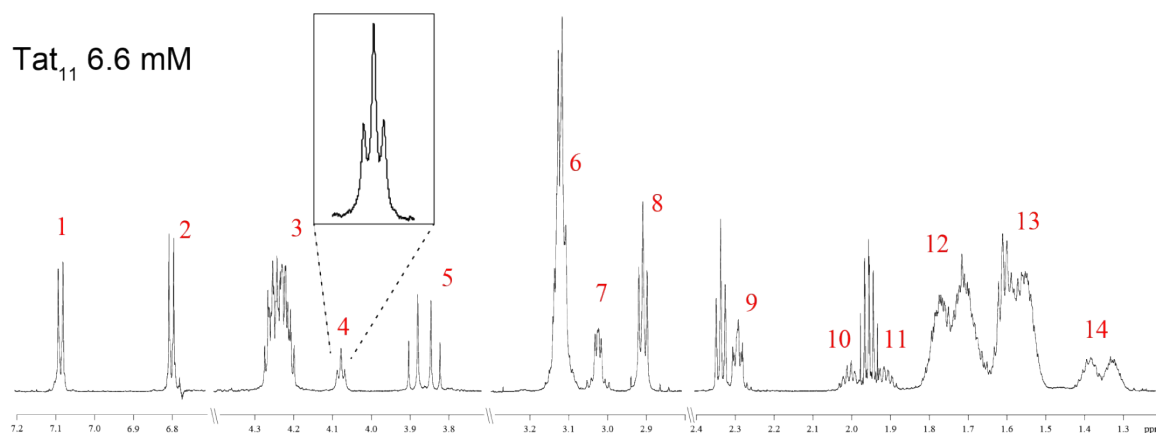


**Fig. S9:** Plot of pyrene 1:3 ratio at increasing Tat<sub>11</sub>-Cys concentrations. Experimental data (points) are shown together with Boltzmann fitting curve (solid line). Dashed arrow points out the center of the sigmoid (indicated in the figure with standard error).

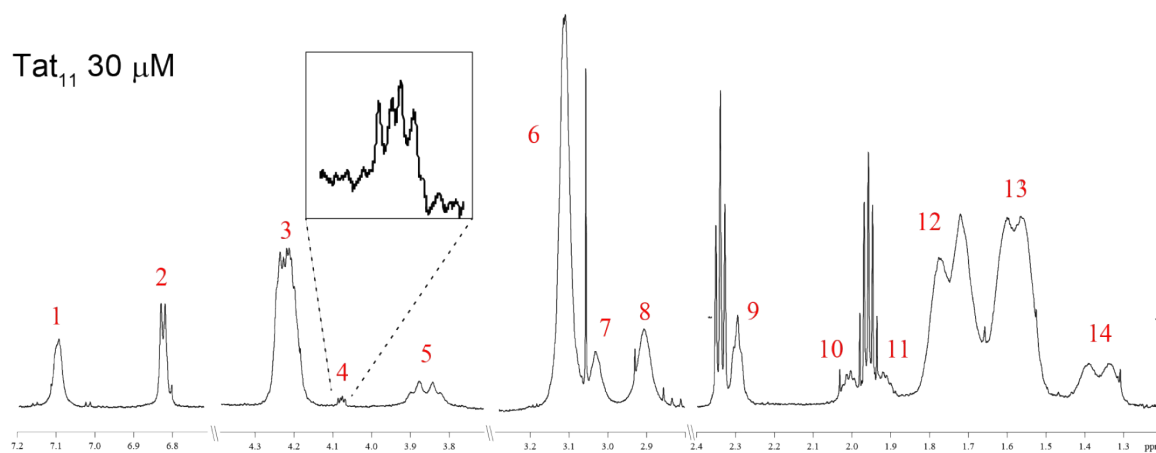


**Fig. S10** DOSY - NMR maps for Tat<sub>11</sub> peptide solutions in phosphate buffer 20 mM, pH 7.4. Diffusion coefficients obtained from each map are reported in the main text. For further technical details, see Section S2

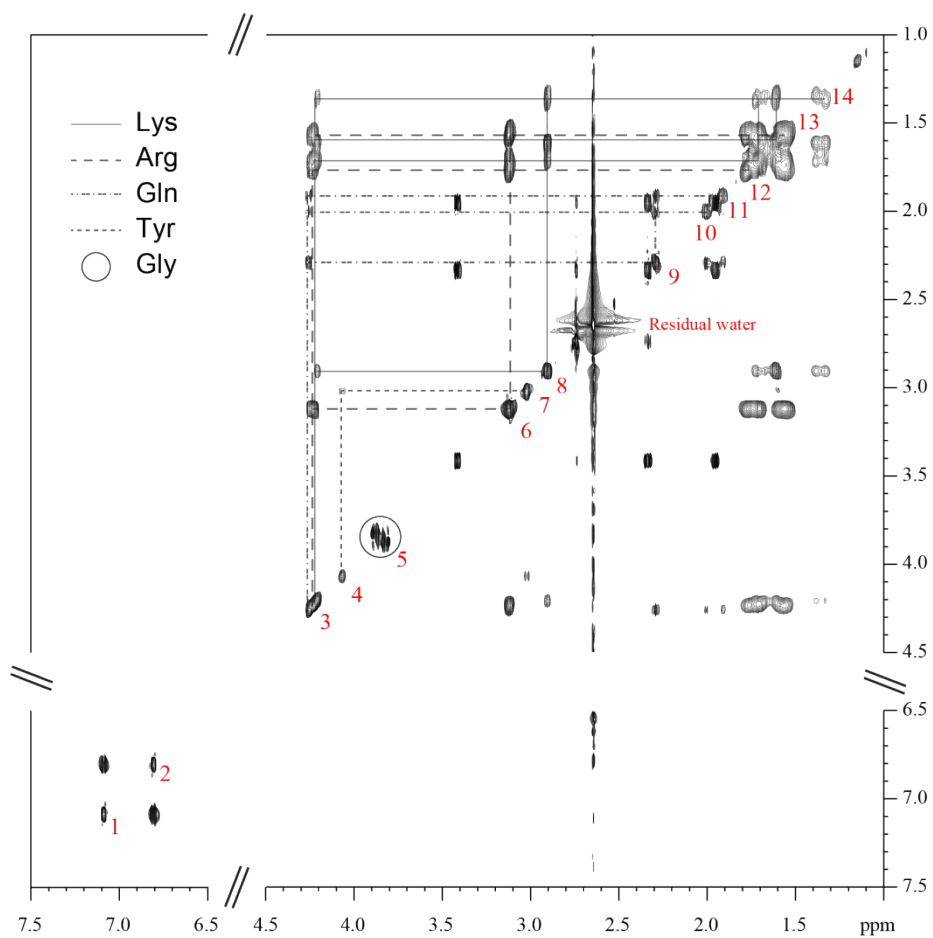
a) Tat<sub>11</sub> 6.6 mM



b) Tat<sub>11</sub> 30  $\mu$ M



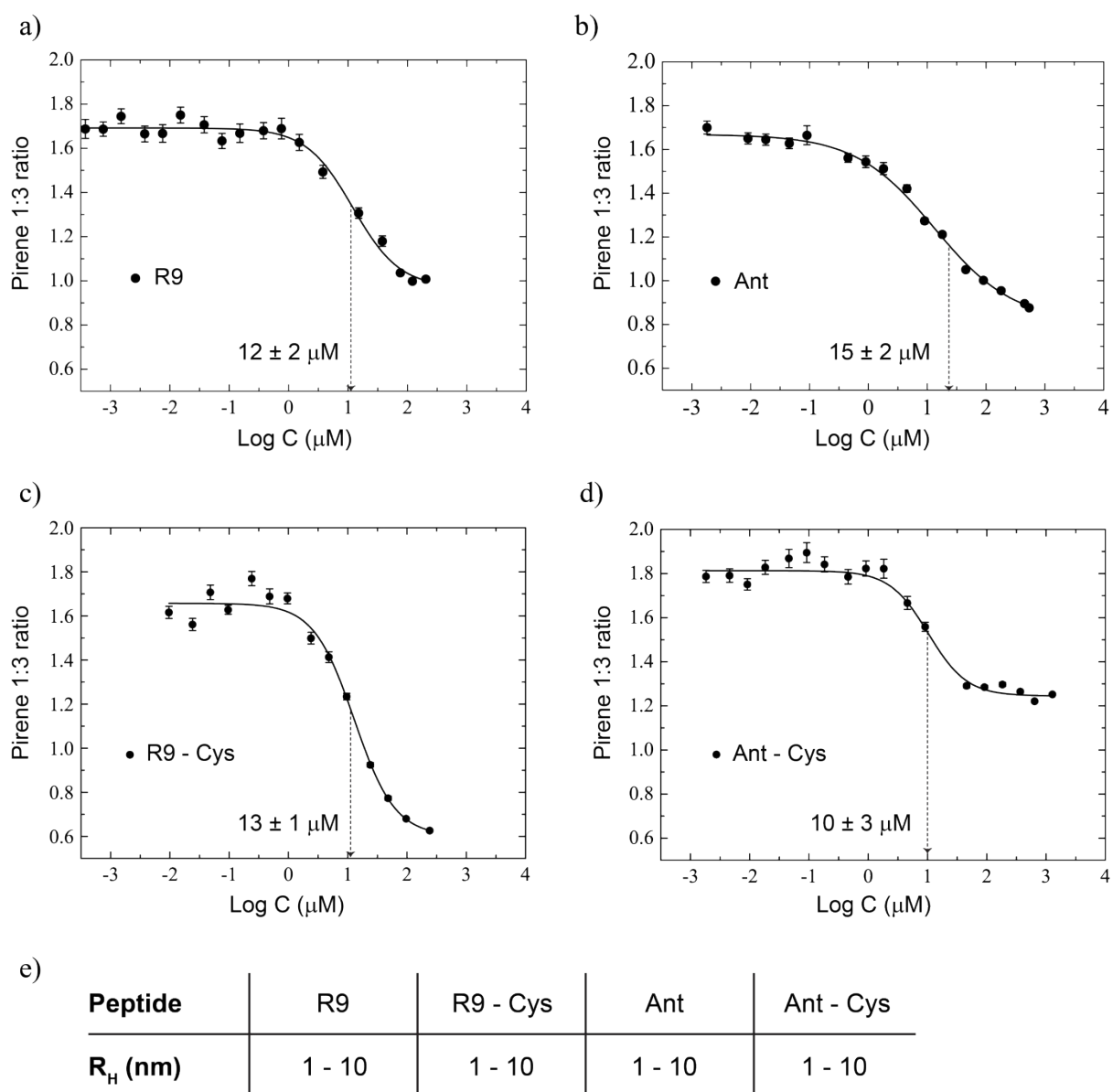
c)



d)

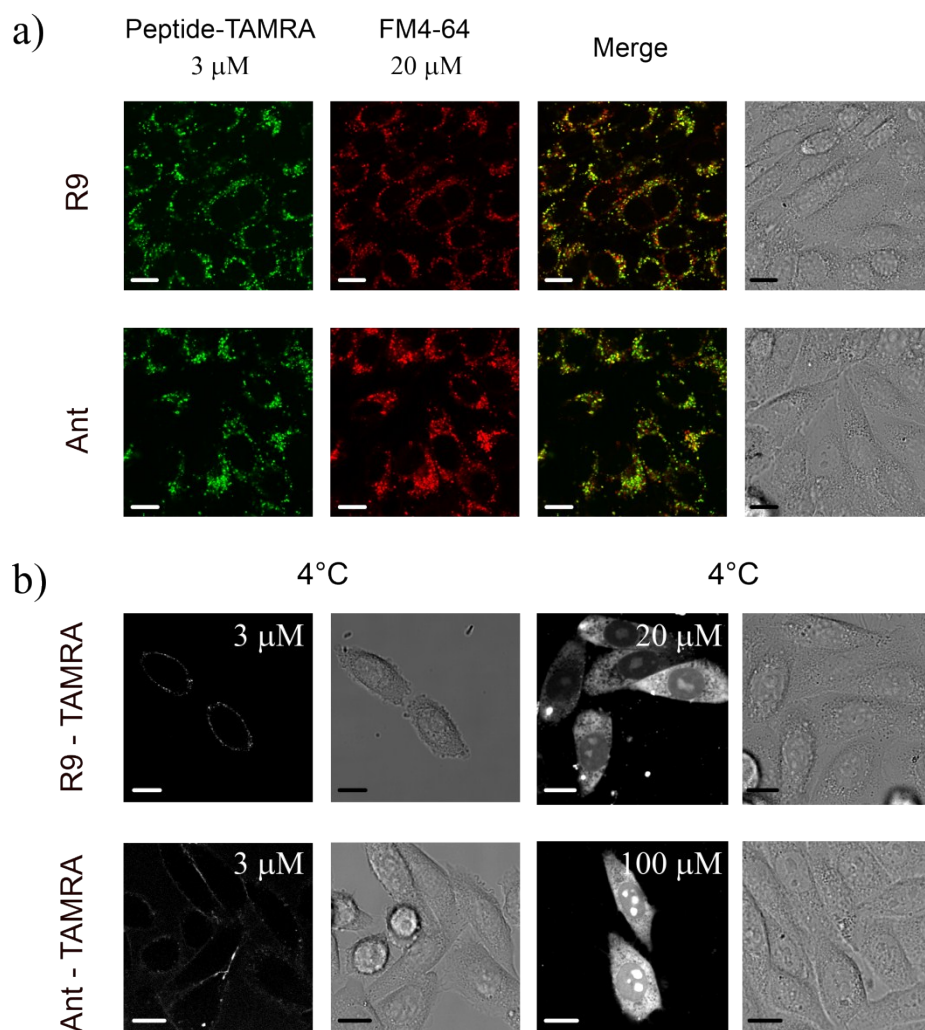
I) Peak n°	1	2	3	4	5	6	7	8	9	10	11	12	13	14
II) Assignment	Tyr $\gamma$	Tyr $\delta$	Arg $\alpha$ Gln $\alpha$ Lys $\alpha$	Tyr $\alpha$	Gly $\alpha$	Arg $\delta$	Tyr $\beta$	Lys $\varepsilon$	Gln $\gamma$	Gln $\beta$	Gln $\beta$	Lys $\beta$ Arg $\beta$	Lys $\delta$ Arg $\gamma$	Lys $\gamma$
III) N° of proton (integration)	2	2	9	1	2	12	2	4	2	1	1	16+16		4

**Fig. S11:** Comparison of  $^1\text{H}$ NMR spectra for 6.6 mM and 30  $\mu\text{M}$  Tat<sub>11</sub> peptide solutions in 20 mM deuterated phosphate buffer, pH 7.4, together with TOCSY - NMR experiment of Tat<sub>11</sub> 6.6 mM and signals' assignment. (a) Tat<sub>11</sub>  $^1\text{H}$ NMR spectrum of 6.6 mM sample and (b) Tat<sub>11</sub>  $^1\text{H}$ NMR spectrum of 30  $\mu\text{M}$  sample, in which peptide signals have been indicated by numbers (1-14). (c) TOCSY - NMR bidimensional experiment of Tat<sub>11</sub> (6.6 mM sample). On the main diagonal we find the projection of  $^1\text{H}$ -NMR peptide spectrum, while in the rest of the map interactions between protons of the same amino acid are pointed out by symmetrical off diagonal peaks. The diagonal peak between peaks 5 and 6 is a sample impurity. Each peptide signal has been assigned to the corresponding amino acidic proton. As indicated in the legend on the left of the figure, each line type represents the pattern of a different amino acid. (d) I) Peak numbers in the spectra. II) Peak assignment from NMR - TOCSY experiment. III) Number of protons giving rise to each peak intensity, obtained from  $^1\text{H}$ NMR spectra integration. For further technical details, see Section S2.

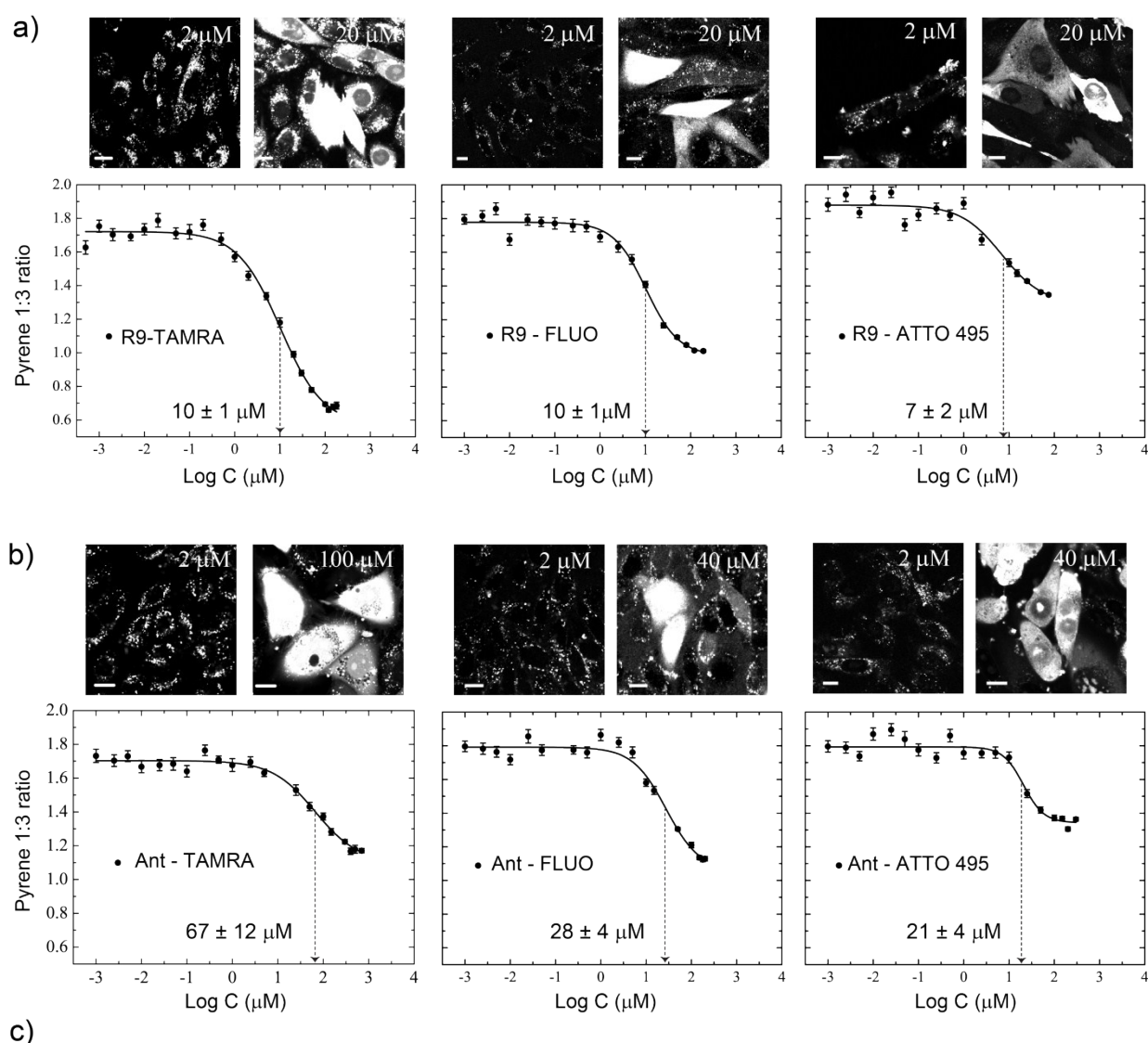


**Fig. S12:** CPPs' self-aggregation tendency and hydrodynamic radii. (a-d) Pyrene 1:3 ratio results for CPPs' sequences. Experimental data (points) are shown with Boltzmann fitting curve (solid line). For both R9 and Ant peptides the center of the sigmoid (pointed out by dashed arrows and indicated in each figure with standard error) is substantially unchanged by adding the Cysteine residue. Vertical bars: standard errors. (g) Hydrodynamic radius  $R_H$  (nm) from DLS for the two R9 and Ant peptide variants.





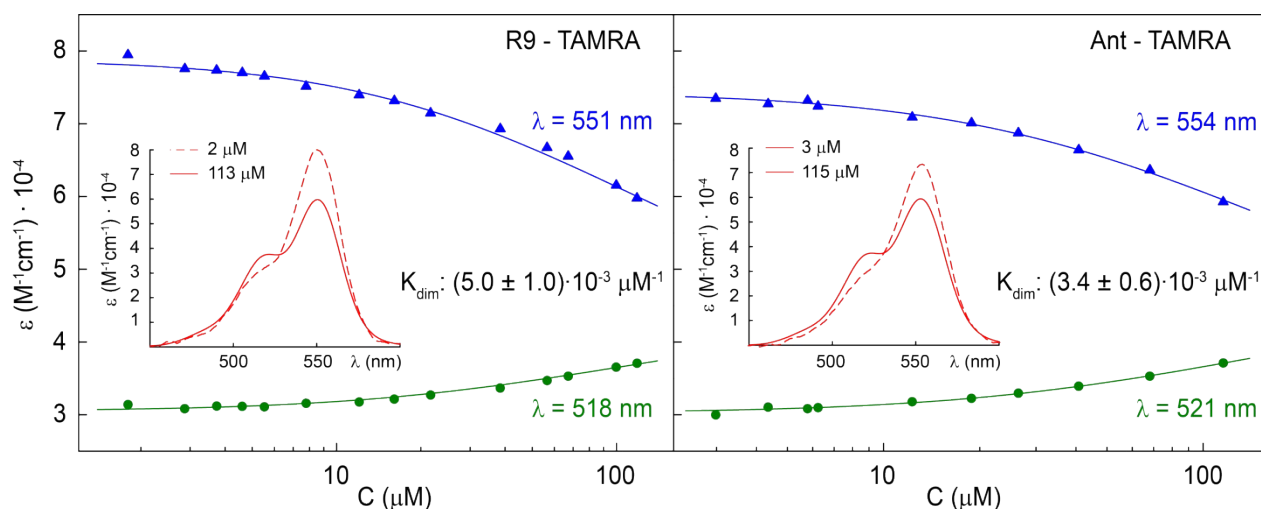
**Fig. S13:** Endocytosis assay and cell uptake in CHO-K1 cells at 4°C of TAMRA-labeled peptides. (a) Confocal images of R9-TAMRA and Ant-TAMRA at 3 μM co-treated with FM4-64 at 20 μM. In both channels only bright spots are detected which are perfectly superimposed in the merge channel, thus confirming that cell uptake is endocytosis. Scale bars: 10 μm (b) Confocal images of TAMRA-labeled peptides administered to cells at low concentration (3 μM) and high concentrations (20 μM and 100 μM for R9-TAMRA and Ant-TAMRA) at 4°C. The experiment confirms that at this temperature endocytosis has been inhibited, as in the first case endocytosis is strongly reduced (less number and brightness of endocytic bright spots), while in the second case direct translocation is the dominant uptake mechanism (cell cytoplasm is full of fluorescent peptide). Scale bars: 10 μm.



\*This value is detailed explained by referring to Table S4

**Fig. S14:** Cell-uptake, aggregation propensity and hydrodynamic radius of dye-labeled R9 and Ant peptides. (a) Upper panels: cell uptake of TAMRA, FLUO and ATTO 495 – labeled R9 peptide in CHO-K1 cells at a concentration below ( $2 \mu\text{M}$ ) and above ( $20 \mu\text{M}$ ) the center of the sigmoid. Endocytic bright spots are detectable at the lower concentration, while direct translocation is dominant at the higher concentration. Scale bars:  $10 \mu\text{m}$ . Lower panels: pyrene 1:3 ratio data (points) and Boltzmann fitting curve (solid line) of TAMRA, FLUO and ATTO 495 – labeled R9 peptide. Dashed arrows point out the center of the sigmoid (indicated in each figure with standard error). Vertical bars: standard errors. (b) Upper panels: cell uptake of TAMRA, FLUO and ATTO 495 –

labeled Ant peptide in CHO-K1 cells at a concentration below (2  $\mu\text{M}$ ) and above (100  $\mu\text{M}$  for Ant-TAMRA, 40  $\mu\text{M}$  for Ant-FLUO and Ant-ATTO 495) the center of the sigmoid. Endocytic bright spots are detectable at the lower concentration, while direct translocation is dominant at the higher concentration. Scale bars: 10  $\mu\text{m}$ . Lower panels: pyrene 1:3 ratio data (points) and Boltzmann fitting curve (solid line) of TAMRA, FLUO and ATTO 495 – labeled Ant peptide. Vertical bars: standard errors. The addition of a dye doesn't change significantly the self-aggregation tendency for R9. Instead, differently from both Tat<sub>11</sub> and R9, a higher increase in this tendency is detected passing from wild type Ant peptide to its dye-labelled variants. (c) Hydrodynamic radius  $R_H$  (nm) from DLS for the two R9 and Ant peptide variants. In all cases DLS measurements point out only small aggregates, with radius ranging between 1 and 10 nm, except for R9-FLUO peptide, in which FLUO dye triggers the formation of bigger aggregates ( $\sim 200$  nm radius; see Table S4 for detailed explanation).



**Fig. S15:** Extinction coefficient data (points and triangles) and fitting curves (solid lines) at increasing (a) R9-TAMRA and (b) Ant-TAMRA concentrations for the two maxima wavelengths ( $\lambda=518$  nm and  $\lambda=551$  nm for R9-TAMRA;  $\lambda=521$  nm and  $\lambda=554$  nm for Ant-TAMRA). In both cases, the dimerization constant  $K_{\text{dim}}$  is obtained by the global analysis (see the main text and Section S2). An increase of about 1.5 folds in the  $K_{\text{dim}}$  is detected passing from Ant-TAMRA to R9-TAMRA, coherently with pyrene-based measurements. In fact, in pyrene 1:3 ratio experiments the center of the sigmoid is about seven-fold lower passing from 67  $\mu\text{M}$ , in the case of Ant-TAMRA, to 10  $\mu\text{M}$ , in the case of R9-TAMRA (see Fig. S14). Central panel of (a): extinction coefficients of R9-TAMRA at 2  $\mu\text{M}$  (dashed line) and 113  $\mu\text{M}$  (solid line) for increasing  $\lambda$  value (nm). Central panel of (b): extinction coefficients of Ant-TAMRA at 3  $\mu\text{M}$  (dashed line) and 115  $\mu\text{M}$  (solid line)

for increasing  $\lambda$  value (nm). A red shift of 3 nm is present passing from R9 (Tat<sub>11</sub> and TAMRA dye) to Ant peptide extinction coefficients.

pH:	7	5	3
Charge:	-2	-1	0
Structure:	Dianion	Anion	Neutral
R9 - FLUO R <sub>H</sub> (nm):	230 ± 8	1 - 10	1 - 10

**Table S4:** Different features of FLUO dye (pH, charge, structure) and nanoparticle radius of R9-FLUO for decreasing pH values. Analogously to Tat<sub>11</sub>-FLUO (see Fig. S5), the structure of the dye shows two net negative charges at pH 7, one net negative charge at pH 5 and no negative charges at pH 3. At both pH 5 and 3, R9 aggregates come back to smaller values (R<sub>H</sub>: 1-10 nm), thus indicating that the loss of the first and then of the second acidic proton of FLUO dye strictly regulates the aggregate size<sup>1</sup>.

## Section S2: Experimental Procedures

### *Peptide synthesis, purification and labeling with fluorescent tags*

All peptides (Tat<sub>11</sub>: YGRKKRRQRRR, R9: RRRRRRRRRR, Ant: RQIKIWFQNRRMKWKK, and their Cysteine-terminal variants) were prepared by solid-phase synthesis using Fmoc chemistry on an automatic Liberty Blue Peptide Synthesizer with an integrated microwave system (CEM, North Carolina, USA). Crude peptides were purified by RP-HPLC (Dionex Ultimate 3000 PLC system with autosampler) on a Jupiter 4 $\mu$ m Proteo 90 A column (250 $\times$ 10 mm; Phenomenex) using these solvents: water:TFA 100:0.01 v/v (eluent A)/ acetonitrile:water:TFA 95:5:0.01 v/v (eluent B), flux 5 ml/min. The purified product was confirmed by electrospray mass spectroscopy using an API3200QTRAP aHybrid Triple Quadrupole/Linear Ion Trap (ABSciex, Foster City, California, USA). The cysteine residue added to the C-terminus of each peptide provided a sulfhydryl group for further ligation to the fluorophores. TAMRA-maleimide and Fluorescein-maleimide (FLUO) were provided from Sigma Aldrich; ATTO 425-maleimide and ATTO 495-maleimide were purchased from ATTO-TEC GmbH, Germany; ALEXA 488-maleimide was purchased from Life Technologies. A suitable amount of peptide (10  $\mu$ mol) was dissolved in 0.2 mL of freshly degassed 10 mM PBS buffer (pH 7.4). A 1.5-fold molar excess of maleimide fluorophores dissolved in dry DMF (except for ATTO 495 maleimide, which was dissolved in PBS buffer) was added to the solution, and finally the mixture was incubated for 2 h at 25°C. Crude products were purified by HPLC and the identity of labeled peptides was confirmed by LC-MS. The purified products were freeze-dried and stored at -80 °C. The day of the experiment they were dissolved in 25-mM phosphate buffer (0.22  $\mu$ m filtered; pH 7.4). Each peptide stock solution concentration was verified by UV-Vis absorbance (Jasco 550 spectrometer, Jasco, Tokyo, Japan). Unlabeled peptides were quantified by amino acidic digestion obtained by the mineralization of a known amount of peptide (5-50  $\mu$ g) dissolved in 200 – 500  $\mu$ l of HCl 6 M, with the following conditions: temperature: 170 °C; time: 20 minutes. Derivatized amino acids were then detected by RP-HPLC (Jupiter 4 $\mu$ m Proteo 90 Å column 250 $\times$ 4.6 mm; Phenomenex) with these solvents: water:TFA 100:0.01 v/v (eluent A)/ acetonitrile:water:TFA 95:5:0.01 v/v (eluent B), flux 1 ml/min, and their identity was confirmed by electrospray mass spectroscopy.

### *Preparation of peptide solution for translocation experiments*

Peptide solutions used in translocation experiments were always freshly prepared. Purified dye-labeled peptides were pre-dissolved in 100 – 200  $\mu$ l of 25 mM phosphate buffer (0.22  $\mu$ m filtered; pH 7.4) and sonicated for 15 minutes at 37°C, then quantified by UV-Vis absorbance (concentration

ranging between 100  $\mu\text{M}$  and 1 mM). The extinction coefficients used ( $\text{cm}^{-1}\text{M}^{-1}$ ) are the following: TAMRA:  $\epsilon_{543}=80000$ ; ATTO 425:  $\epsilon_{436}=45000$ ; FLUO:  $\epsilon_{492}=82700$ ; ALEXA 488:  $\epsilon_{490}=72000$ ; ATTO 495:  $\epsilon_{500}=80000$ . Depending on the final concentration needed for the translocation experiment, the proper amount of stock solution was diluted in serum-free DMEM F12 (1X) growth medium and then administered to cells. For endocytosis assay, FM4-64 dye was purchased from Molecular Probes and it was dissolved in water at 300  $\mu\text{M}$ . The day of the experiment, a diluted solution (20  $\mu\text{M}$ ) in serum-free medium was prepared and it was mixed with the TAMRA-labeled peptide solution (3  $\mu\text{M}$ ) in the same medium. The obtained solution was administered to cells for 45 minutes, then it was removed and, before performing confocal microscopy experiments, cells were incubated with 10% serum DMEM F12 (1X) growth medium for 15 minutes at 37°C.

### ***Cell culture and translocation experiments***

Chinese hamster ovary (CHO-K1) cells were provided by ATCC (reference numbers: CCL-61) and grown in 10% serum containing DMEM F-12 (1X) medium at 37°C with 5%  $\text{CO}_2$ , according to manufacturer's instructions. In order to perform translocation experiments cells were plated onto 35 mm glass-bottom petri dishes (WillCo-dish GWSt-3512) 24h before the experiment. The day of the experiment cells were removed from the incubator and washed with PBS at room temperature. Upon peptide administration, cells were incubated for 20 minutes at 37°C and then washed with PBS at room temperature. Finally 1 ml of 10%-serum DMEM-F12 medium was added on the cells.

### ***WST-8 cell viability assay***

Cytotoxicity of Tat<sub>11</sub>-TAMRA peptide was evaluated by the WST-8 assay. CHO-K1 cell proliferation was evaluated by plating  $5 \times 10^3$  cells per well in 96-well plates. After 24 hours in serum-containing DMEM-F12 medium, the cells were incubated with a serum-free DMEM-F12 solution of peptide for 30 minutes at 37 °C. After the incubation, the medium was removed, cells were washed with PBS 1X and incubated for 2 hours at 37 °C with 100  $\mu\text{l}$  of a 10%-solution of WST-8 dissolved in serum-containing DMEM-F12. Absorbance at 450 nm was measured using a multiplate reader (Microplate Reader, GloMax® -Discover and Explorer Systems, Promega). Cell viability was quantitatively determined by comparing peptide-treated cells with untreated cells (as a reference of 100% viability) and cells treated with dimethyl sulfoxide (DMSO) 50% v/v. Reported data represent the average of three independent experiments, in which three wells for each concentration were measured. Error bars represent the standard errors.

### ***Confocal microscopy***

Laser scanning confocal microscopy experiments were performed with a Leica TCS SP5 SMD inverted confocal microscope (Leica MicrosystemsAG) interfaced with a diode laser (Picoquant) for excitation at 405 nm. Glass-bottom Petri dishes containing plated cells were mounted in a temperature-controlled chamber at 37°C and 5% CO<sub>2</sub> (Leica Microsystems) and viewed with a 63×1.2 numerical aperture (NA) water-immersion objective (Leica Microsystems). The excitation wavelengths (nm) and the collection ranges (nm) adopted for each dye are the following: TAMRA:  $\lambda_{\text{exc}} = 561$ ; range: 580-680; ATTO 425:  $\lambda_{\text{exc}} = 405$ ; range: 430-530; FLUO, ALEXA 488, ATTO 495:  $\lambda_{\text{exc}} = 488$ ; range: 500-600; FM4-64:  $\lambda_{\text{exc}} = 561$ ; range: 700-800. All images and videos collected were analyzed by ImageJ software version 1.440 (NIH Image; <http://rsbweb.nih.gov/ij/http://rsbweb.nih.gov/ij/>).

### ***Pyrene 1:3 ratio assay***

Before performing the pyrene 1:3 ratio assay, all *cuvettes* used for this experiment were silanized by immersion first in Silanization Solution I (Sigma Aldrich) for 90 minutes and then in Methanol (Sigma Aldrich) for 90 minutes; finally they were washed with water and acetone. Fluorescence emission spectra at 25°C of free and labeled-peptide solutions in 25-mM phosphate buffer (pH=7.4, 0.22  $\mu\text{m}$  filtered) containing 0.5  $\mu\text{M}$  of pyrene were recorded using an excitation wavelength of 335 nm. The intensities  $I_1$  and  $I_3$  were measured at the wavelengths corresponding to the first and third vibronic bands of pyrene (373 nm, emission range: 368-378 nm, and 385 nm, emission range: 380-390 nm, respectively). According to literature the plot of the *pyrene 1:3 ratio* index as a function of the surfactant or peptide<sup>2</sup> concentration can be described by a decreasing sigmoid of the Boltzmann type. Experimental raw data were fitted using the OriginPro8 software.

### ***Absorbance measurements***

UV-Vis measurements were performed on Jasco 550 spectrometer (Jasco, Tokyo, Japan). In order to have absorbance values in the range of 0.1-1.0, the optical path lengths were adjusted between 1.0 and  $3 \cdot 10^{-1}$  cm for the phosphate buffer solutions of dye-labeled peptides at concentrations in the range  $10^{-6}$ - $10^{-4}$  M. In order to quantify more precisely the sample concentration we used the extinction coefficient value of TAMRA at the isosbestic point ( $\epsilon_{528} = 37500 \text{ cm}^{-1}\text{M}^{-1}$  from Ref.<sup>3</sup>). We verified that changing the wavelength of the isosbestic point by 2 nanometers and the relative eps of 10% led to changes within the estimated errors of the  $K_{\text{dim}}$  obtained from the fit described below.

The dimerization reaction  $M + M \leftrightarrow D$  is described by (see also Refs.<sup>3,4</sup>):



$$K_{dim} = \frac{[D]}{[M]^2} \quad (1)$$

Where  $K_{dim}$  is the equilibrium constant for dimerization. Calling  $C_T$  the total analytical concentration of the peptide, one gets:

$$C_T = [M] + 2 \cdot [D] \quad (2)$$

and, by dividing for  $C_T$ :

$$1 = \frac{[M]}{C_T} + \frac{2 \cdot [D]}{C_T} = f_M + f_D \quad (3)$$

This gives:

$$K_{dim} = \frac{[D]}{[M]^2} = \frac{\frac{[D]}{[C_T]^2}}{\frac{[M]^2}{[C_T]^2}} = \frac{\frac{f_D}{2} \cdot \frac{1}{C_T}}{f_M^2} \quad (4)$$

$$2 \cdot K_{dim} \cdot C_T = \frac{f_D}{f_M^2} \quad (5)$$

$$2 \cdot K_{dim} \cdot C_T = \frac{f_D}{(1 - f_D)^2} \quad (6)$$

Solving for  $f_D$  and taking the  $f_D \leq 1$ , solution results in

$$f_D = \frac{4 \cdot K_{dim} \cdot C_T + 1 - \sqrt{8 \cdot K_{dim} \cdot C_T + 1}}{4 \cdot K_{dim} \cdot C_T} \quad (7)$$

$$f_M = \frac{\sqrt{8 \cdot K_{dim} \cdot C_T + 1} - 1}{4 \cdot K_{dim} \cdot C_T} \quad (8)$$

Then the UV spectra can be analyzed according to:

$$\varepsilon(\lambda, C_T) = f_M(C_T) \cdot \varepsilon_M(\lambda) + f_D(C_T) \cdot \varepsilon_D(\lambda) \quad (9)$$

or:

$$\varepsilon(\lambda, C_T) = f_M(C_T) \cdot (\varepsilon_M(\lambda) - \varepsilon_D(\lambda)) + \varepsilon_D(\lambda) \quad (10)$$

We performed a global fit of the epsilon at the various concentrations in the 490-580 nm window, yielding the  $K_{\text{dim}}$  of the reaction and the  $\varepsilon_M$  and  $\varepsilon_D$ . By performing the global fitting described in the text we could observe that no significant variations on  $\varepsilon$  value was verified by changing the isosbestic point of some nanometers (data not shown) at different  $\lambda$ .

### ***Dynamic Light Scattering (DLS)***

DLS measurements were performed at 25°C in a 50- $\mu$ L quartz silanized *cuvette* on a Zetasizer nano ZS DLS (Malvern Instrument) following the manufacturer's instructions. Solutions of free and dye-labeled peptide in 25 mM phosphate buffer (0.22  $\mu$ m filtered; pH 7.4) were analyzed with a single scattering angle of 90°. Each value reported is the average of seven consecutive measurements that were repeated in three different experiments. Tat<sub>11</sub>-FLUO and R9-FLUO solutions at pH=3 and pH=5 were prepared in citrate-phosphate buffer and pH was adjusted to the desired value by adding 1M HCl solution.

### ***Nuclear Magnetic Resonance (NMR)***

NMR diffusion experiments were performed on a Bruker AVANCE 300 spectrometer operating at 300.13 MHz, equipped with a 3 mm probe. Experiments were performed at 298 K and TMS (tetramethylsilane) was used as an internal reference for chemical shift. A stock Tat<sub>11</sub> peptide solution (80 mM) was prepared in deuterated 20 mM phosphate buffer, pH 7.4. An appropriate amount from this stock solution was diluted in 250-300  $\mu$ L of the same buffer to obtain samples at different concentrations. DOSY maps (sequences *stebpgp1s*) were acquired using 900 and 700 scans were carried out for 200  $\mu$ M and 400  $\mu$ M samples, while 150 scans were enough for the other samples (between 1300  $\mu$ M and 13000  $\mu$ M). Gradient strength  $G$  applied was between from 0.9 to 45.7 G cm<sup>-1</sup> (from 2 to 95%), the diffusion time ( $\Delta$ ) was 100 ms and the gradient pulse time ( $\delta$  \* 0.5) was 3 ms. These last two parameter were optimized (or verified) recording two proton spectra at fixed field gradient (2 and 95 %), verifying that the overall intensity of the spectrum at 95 % field gradient decreased to the 10% intensity with respect to the 2 % field gradient experiment. The 30  $\mu$ M sample DOSY experiment (sequences *ledbpgppr2s*) was recorded on a Bruker AVANCE 700 spectrometer operating at 700.13 MHz, equipped with a 3 mm proton cryo-cooled probe (temperature: 298K, internal reference: TMS). 800 scans were carried out, gradient

strength G went from 0.7 to 28.6 G cm<sup>-1</sup>, the diffusion time was 199 ms and the gradient pulse time was 4 ms.

The equation used for fitting diffusion data (y = peak area intensity vs x = gradient strength) is:

$$y = A \cdot e^{\{-D \cdot x^2 \cdot \gamma^2 \cdot LD^2 \cdot [BD - (LD:3)] \cdot 10^4\}} \quad (11)$$

in which A is the function amplitude, D is the diffusion coefficient in m<sup>2</sup>/s,  $\gamma$  the gyromagnetic ratio and LD (Little Delta) and BD (Big Delta) are experiment parameters. Once knew the D values, the  $R_H$  values were calculated from the Stokes-Einstein equation:

$$R_H = \frac{k_B \cdot T}{6 \cdot \pi \cdot \eta \cdot D} \quad (12)$$

where  $K_B$  is the Boltzmann constant; T is 298 K;  $\eta$  is  $8.96 \cdot 10^{-4}$  Pa·s.

Static 1D <sup>1</sup>H experiments of two selected Tat<sub>11</sub> samples (6.6 mM, for defining the first TOCSY dimension, and 30  $\mu$ M, for improving its signal to noise ratio) were performed on the Bruker AVANCE 700 spectrometer by using standard pulse sequences and parameters. The number of averaged scans was 256 and 32 for the 30  $\mu$ M and for the 6.6 mM samples, respectively. The 90° pulse duration and the recycle delay were 10.25  $\mu$ s - 10 s, and 11.75  $\mu$ s - 2.5 s for 30  $\mu$ M and 6.6 mM solutions, respectively. In order to assign proton signals, for Tat<sub>11</sub> peptide 6.6 mM solution a TOCSY experiment (2D-TOCSY, sequences name: mlevphpr.2) was acquired with 8 scans. The complex time-domain points were 4096 while 256 incremental delays were used in the  $t_1$  direction. 80 ms was mixing time. The sample temperature was 298 K and a radio-frequency saturation pulse was added to suppress the signal from H<sub>2</sub>O during the relaxation delay. All the NMR experiments were performed with a spectral window of 8389 MHz and processed with Topspin 3.2. For what concerns the diffusion coefficient fitting, by following the equations (1) – (8), we obtained the following equations:

$$D_T = f_M(C_T) \cdot D_M + f_D(C_T) \cdot D_D \quad (13)$$

$$D_T = f_M(C_T) \cdot (D_M - D_D) + D_D \quad (14)$$

### ***GUV preparation***

The lipid DOPC, 1,2-dioleoyl-sn-glycero-3-phosphocholine (10mg/mL in chloroform) was purchased from Avanti Polar Lipids (Alabaster, AL, USA) and used without further purification.

Low gelling temperature agarose, BioReagent, for molecular biology, was purchased from Sigma Aldrich (St. Louis, MO, USA). Liposomes of DOPC were prepared using the standard method<sup>5</sup>. A thin film of lipid was obtained by evaporating 100  $\mu$ L of chloroform solution containing 1mg of DOPC by placing the sample in a centrifugal evaporator under vacuum for 2 h. The lipid film was hydrating by adding 250  $\mu$ L phosphate-buffered saline (PBS) at pH 7.45 at room temperature. The final DOPC concentration was 5 mM. The vesicles were frozen in liquid nitrogen and then thawed at 50 °C in a water bath. The freeze-thaw cycle was repeated five times<sup>6</sup>. Tat<sub>11</sub>-TAMRA was dissolved in PBS and added to 25  $\mu$ L of liposomes dispersion in different concentration (1  $\mu$ M, 3  $\mu$ M and 30  $\mu$ M). The samples were incubated for 30 minutes, then were centrifuged at 13,200 rpm for 60 minutes at 4 °C. After centrifugation, the supernatant was removed and replaced with 25  $\mu$ L of PBS. The procedure was repeated another time. Agarose gel was used to immobilize liposomes as described in Ref.<sup>7</sup> Agarose was dissolved in PBS at a concentration of 0.75% w/v. Liposomes were mixed in gel while the agarose was in the fluid state. After mixing, the solution was placed on a glass bottom petri dish and was left at room temperature for jellification. The vesicles were observed on an Olympus FluoView FV1000 confocal microscope with a 60x NA 1.20 water immersion objective. Tat<sub>11</sub>-TAMRA was excited using a laser at 543 nm, the emission signal was collected between 550 and 650 nm.

## Section S3: Additional Molecular Dynamics Simulations analysis

Molecular dynamics simulations were performed on fully solvated Tat<sub>11</sub> and Tat<sub>11</sub>-TAMRA monomers, and in dimeric forms. The list of the various simulations with all useful details is given in Table S5, including the simulations described in the main text, Tat<sub>11</sub>-*dim a* and Tat<sub>11</sub>-TAMRA – *dim a*. Below we describe the monomer and dimer simulations in more detail.

### *Monomers*

To achieve accelerated exploration of the monomer configurations (simulations Tat<sub>11</sub>-*mon a* and Tat<sub>11</sub>-TAMRA –*mon a*) we employed the Hamiltonian Replica Exchange method, where, in order to minimize the number of replica needed, only the solute is “heated” by suitably scaling its intra and inter molecular interactions. We also performed  $\mu$ s-timescale normal simulations for comparison (Tat<sub>11</sub> –*mon c* Tat<sub>11</sub>-TAMRA –*mon b*). In the case of Tat<sub>11</sub> we compared the results of two different force fields, Amber99SB\*-ILDN and Charmm22\*, both shown to yield satisfactory agreement with NMR measurements on peptides<sup>8</sup>.

In Fig. S16 the result of a secondary structure analysis (using STRIDE<sup>9</sup>) on these simulations is reported. Amber99SB\*-ILDN and Charmm22\* give overall similar results for Tat<sub>11</sub>, revealing only residual presence of secondary structure such as helices and  $\beta$ -strands/hairpin. The Amber simulation (Tat<sub>11</sub> –*mon a*) contains more turn sections than the one with Charmm (Tat<sub>11</sub> –*mon b*). The labelling with TAMRA (Tat<sub>11</sub>-TAMRA –*mon a*, with Amber) does not seem to perturb the peptide in terms of secondary structure content. The comparison between the HREX (Tat<sub>11</sub>-TAMRA –*mon a*) and normal MD simulation (Tat<sub>11</sub>-TAMRA –*mon b*) reveals a still not fully complete exploration of the configurational space in the 1.8  $\mu$ s time span, although the general features are reproduced (see average radius of gyration reported in Table S6).

### *Dimers*

Starting structures for the dimers were obtained by two methods. The first consisted in forcing the inter-monomer contact by applying a harmonic restraint on their centres of mass (with a force constant  $k = 1000$  kJ/mol) during a 10 ns simulation and then gradually releasing the restraints in a series of 5 simulations (2ns each). While this resulted in a stable dimer for Tat<sub>11</sub>-TAMRA (Tat<sub>11</sub>-TAMRA – *dim b*), in the Tat<sub>11</sub> case the structure dissociated within the first 200 ns of unrestrained MD. Other attempts with longer restrained simulations (up to 100 ns) all resulted in unstable dimer simulations for Tat<sub>11</sub> (i.e. dissociation was observed within the first 200 ns). This is a first indication

that the Tat<sub>11</sub> dimer is less stable than that of Tat<sub>11</sub>-TAMRA, in keeping with the lower measured  $K_{\text{dim}}$  of the former with respect to the latter.

The second method involved a more thorough exploration of the potential dimer configurations by metadynamics, with the reaction coordinate defined as the inter-peptide contact number (see methods). From these simulations we extracted two high-contact structure for the Tat<sub>11</sub> dimer, NC119 (starting structure for Tat<sub>11</sub>-*dim a*) and NC136 (Tat<sub>11</sub>-*dim b*), and one for Tat<sub>11</sub>-TAMRA dimer (NC130, Tat<sub>11</sub>-TAMRA -*dim a*). Before the production runs, 5ns simulations with restraints on solute non-hydrogen atoms were performed to gradually release possible strains due to the bias potential. Additionally, in the Tat<sub>11</sub> NC119 case, we run two other MD simulations with different starting random velocities (Tat<sub>11</sub>-*dim a* ' and Tat<sub>11</sub>-*dim a* ' ') and one with the Charmm22\* force field (Tat<sub>11</sub>-*dim a* \*).

While both Tat<sub>11</sub>-TAMRA dimer structures are stable on the explored timescale, only two Tat<sub>11</sub>-*dim* simulations led to dimer lifetime longer than 2  $\mu\text{s}$ , again an indication of the lower stability of the Tat<sub>11</sub> dimer.

The various simulations were analysed in terms of the contact map (Fig. S17), secondary structure (Fig. S18) and gyration radius (Fig. S19 and Table S6), contact surface and number of inter-peptide hydrogen bonds (Table S7). In the case of Tat<sub>11</sub>-TAMRA dimer also the TAMRA-TAMRA distance is reported (Fig. S8). The contacts in the Tat<sub>11</sub>-*dim a* simulations take place mostly between each peptide C-terminal region (Fig. S17, residues 8 to 11). The secondary structure displays mostly coil and some turn motives, and the association between the two peptides includes a small  $\beta$ -bridge segment at amino acid 10 on each monomer (Fig. S18). The simulation with Charmm22\* force field (Tat<sub>11</sub>-*dim a* \*) confirms the same overall picture, both in terms of contact map and secondary structure content. In Tat<sub>11</sub>-*dim b* the contact region (Fig. S18) is wider, leaning almost towards a completely antiparallel pairing (this would show as a diagonal perpendicular to the red one of “self-interaction”), with some fraction of  $\beta$ -sheet pairing occurring between amino acid 7-10 on monomer A and 3-7 of monomer B. This mode of association gives rise to a more compact structure, with a shorter radius of gyration (1.03 vs 1.1-1.2 nm of the Tat<sub>11</sub>-*dim a* simulations). It should be observed however that this secondary structure pairing is less stable than the  $\beta$ -bridge of Tat<sub>11</sub>-*dim a* (the former occurs for less than 30% of the total simulation time, the latter for more than 60%) and that in the Tat<sub>11</sub>-*dim b* simulation the dimer dissociates after  $\sim 1 \mu\text{s}$ .

Similar features are observed in the Tat<sub>11</sub>-TAMRA dimer simulations, this time the Tat<sub>11</sub>-TAMRA-*dim* showing a more antiparallel-like peptide pairing, and a considerably larger contact surface

(Table S7). The Tat<sub>11</sub>-TAMRA-*dim a* features a very stable  $\beta$ -sheet pairing segment between amino acids 5-10 of peptide A and 9-11 of peptide B. This motif is absent in Tat<sub>11</sub>-TAMRA-*dim b*, which overall looks like a less tight dimer in terms of contact surface area. Despite these differences the average gyration radius from the two simulations is almost identical.

Table S3, the scheme in Fig. S7, Movie1 and Movie2 report a detailed analysis of the molecular interactions contributing to the dimer stability, in terms of the hydrogen bond/salt bridge network and Arginine/Arginine stacking. Each simulation displays a different network, and also within the same simulation most hydrogen bonds are transient. This hints at different possible dimer-association modes. Nonetheless, some common features emerge from the analysis. First, salt-bridges are formed between the terminal COO<sup>-</sup> and mostly Arginine (occasionally Lysine) side chains. When TAMRA is present, salt-bridges form between Arginine and its COO<sup>-</sup>. Also inter-backbone H-bonds (Fig. S20) are established, in particular when  $\beta$ -sheet pairing is present. Second, another common motif is the stacking between 2 or more guanidinium groups of Arginine residues (Table S3), in some cases with a sizable occurrence during the simulation (for example in Tat<sub>11</sub>-*dim a* Arg at position 10 on each peptide are stacked for ~30% of the simulated time, see Fig. S7c). Given the electrostatic repulsion of the two positively charged guanidinium groups, these motives presumably do not give a net contribution to the stability of the dimer. Rather, they reveal the presence of extended (though possibly transient). Given the electrostatic repulsion of the two positively charged guanidinium groups, these motives presumably do not contribute directly to the stability of the dimer. Rather, they reveal the presence of extended (though always transient) H-bond networks that are able to maintain such local structures<sup>10</sup>. In particular, the previously mentioned stacking of the two Arginine at position 10 in Tat<sub>11</sub>-*dim a* is maintained by the salt bridges with the C-terminal carboxylate groups. In the case of Tat<sub>11</sub>-TAMRA some Arginine residues are also stacked with TAMRA, and this arrangement is favoured by the salt bridge between the guanidinium group and TAMRA carboxylate.

Name	Method/ Force field <sup>b</sup>	Duration	Starting configuration	Analysis	Notes
<b>Monomers</b>					
Tat <sub>11</sub> – <i>mon a</i>	H-REX <sup>a</sup>	300ns	Random coil	50-300ns	
Tat <sub>11</sub> – <i>mon b</i>	H-REX Charmm 22*	300ns	Random coil	50-300ns	
Tat <sub>11</sub> – <i>mon c</i>	Normal MD	1.3 $\mu$ s	Random coil	0.1-1.3 $\mu$ s	
Tat <sub>11</sub> -TAMRA – <i>mon a</i>	H-REX	300ns	Random coil	50-300ns	
Tat <sub>11</sub> -TAMRA – <i>mon b</i>	Normal MD	2.2 $\mu$ s	Random coil	0.3-2.2 $\mu$ s	
<b>Dimers</b>					
Tat <sub>11</sub> – <i>pre dim</i>	Metadynamics	300ns	Random coils		

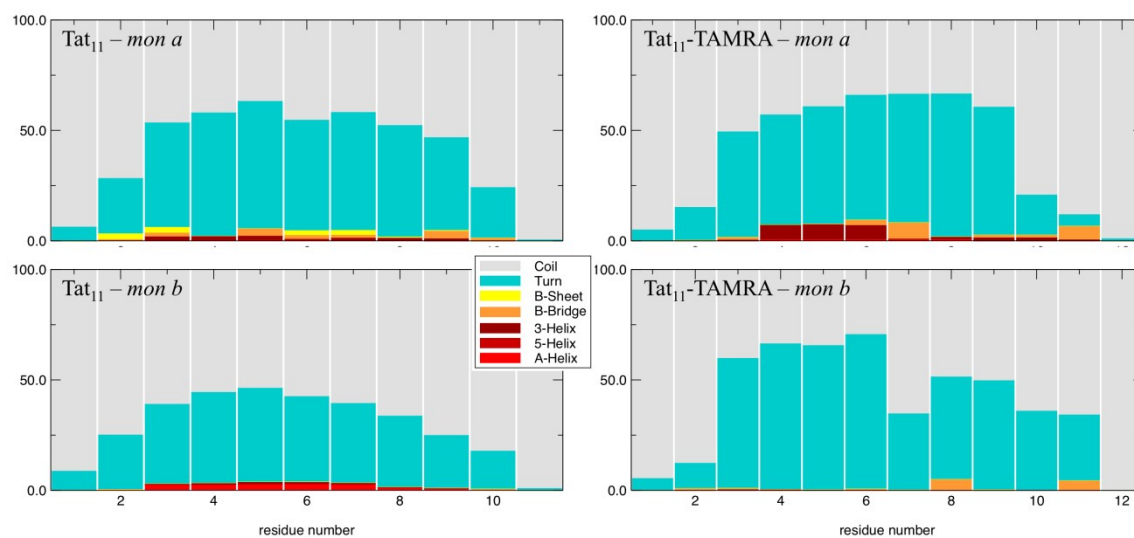


Tat <sub>11</sub> – <i>dim a</i>	Normal MD	2.8 $\mu$ s	NC119 from Tat <sub>11</sub> – <i>pre dim</i>	0.3-3.2 $\mu$ s	
Tat <sub>11</sub> – <i>dim a'</i>	Normal MD	0.8 $\mu$ s	NC119 from Tat <sub>11</sub> – <i>pre dim</i> Random vel 1	0.1-0.7 $\mu$ s	dissociates at 0.75 $\mu$ s
Tat <sub>11</sub> – <i>dim a''</i>	Normal MD	0.9 $\mu$ s	NC119 from Tat <sub>11</sub> – <i>pre dim</i> Random vel 2	0.1-0.8 $\mu$ s	dissociates at 0.83 $\mu$ s
Tat <sub>11</sub> – <i>dim a*</i>	Normal MD Charmm 22*	2.6 $\mu$ s	NC119 from Tat <sub>11</sub> – <i>pre dim</i>	0.3-2.6 $\mu$ s	
Tat <sub>11</sub> – <i>dim b</i>	Normal MD	1.2 $\mu$ s	NC136 from Tat <sub>11</sub> – <i>pre dim</i>	0.1-1 $\mu$ s	dissociates at 1.015 $\mu$ s
Tat <sub>11</sub> -TAMRA – <i>pre dim</i>	Metadynamics	130ns	Random coils		
Tat <sub>11</sub> -TAMRA – <i>dim a</i>	Normal MD	2.8 $\mu$ s	NC130 from Tat <sub>11</sub> -TAMRA – <i>pre dim</i>	0.3-3.2 $\mu$ s	
Tat <sub>11</sub> -TAMRA – <i>dim b</i>	Normal MD	2.8 $\mu$ s	From 20ns forced simulation	0.3-3.2 $\mu$ s	

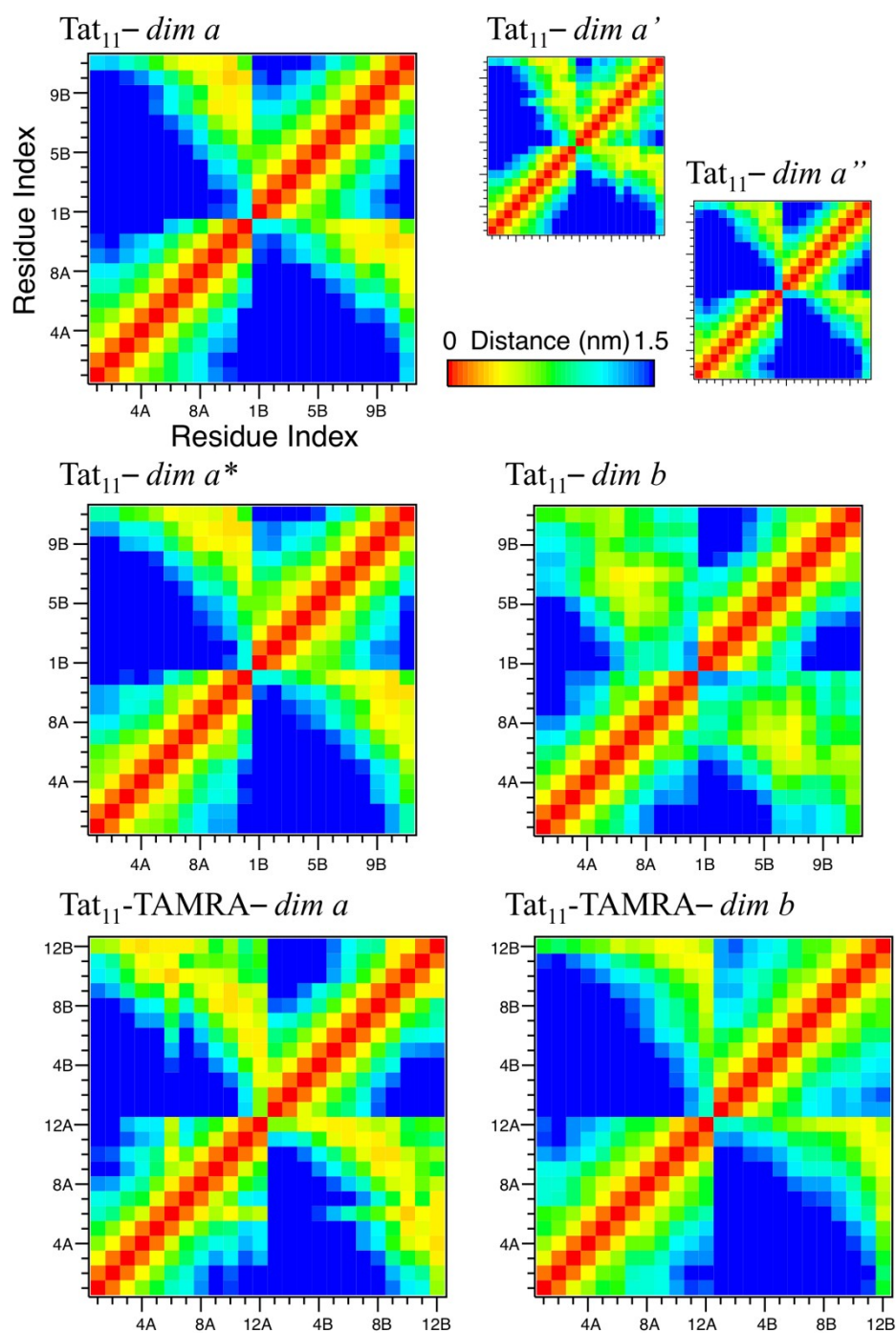
<sup>a</sup> HREX=Hamiltonian Replica Exchange

<sup>b</sup> Force field is Amber99SB\*-ILDN unless otherwise specified

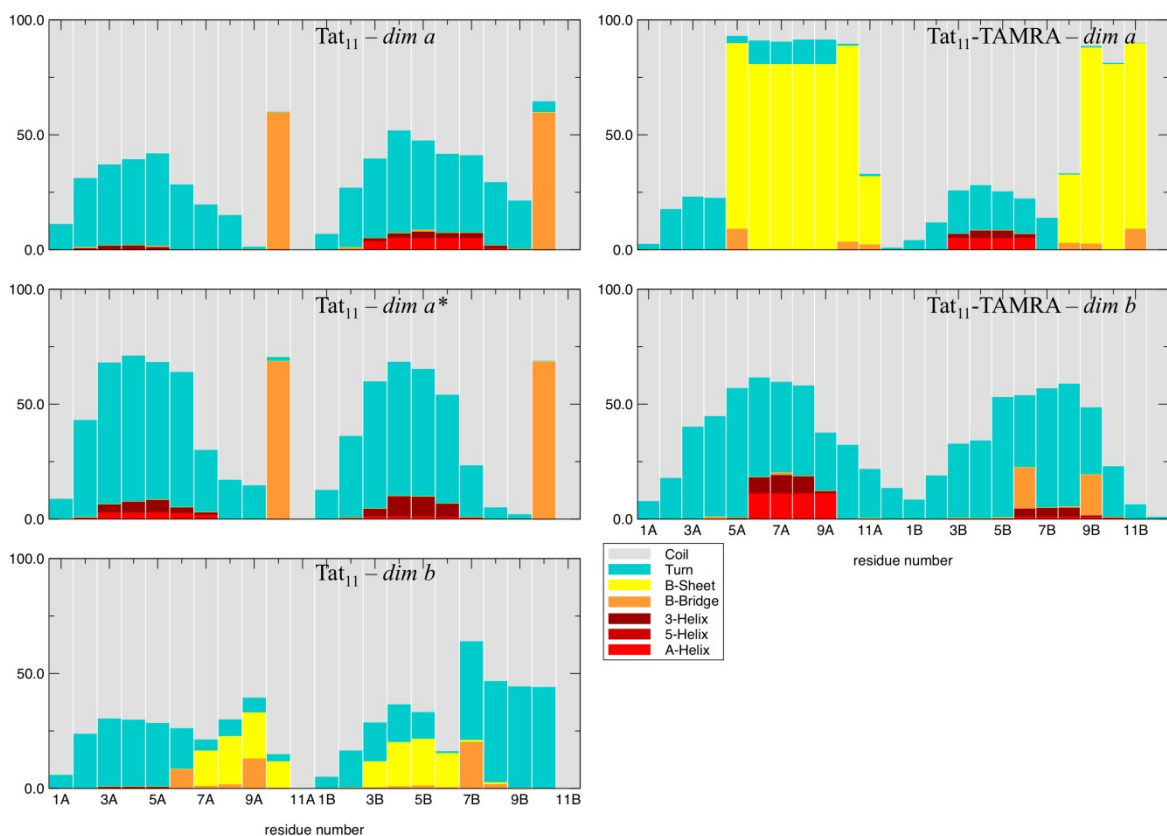
**Table S5:** Summary of the simulations



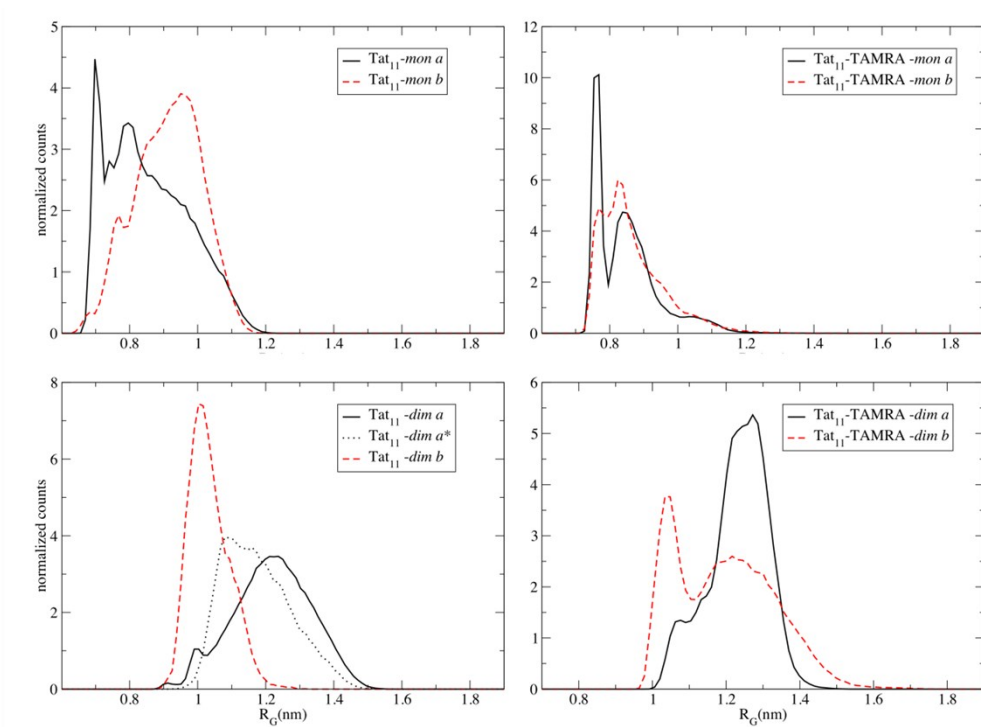
**Fig. S16:** Histograms of secondary structure counts (the percentage is reported in the y axis) during the MD simulations of the monomers.



**Fig. S17:** Inter and intra peptide contact maps. The reported distance values are the average of inter-residue minimum distances excluding hydrogen atoms.



**Fig. S18:** Histograms of secondary-structure counts during the MD simulations of the dimers.



**Fig. S19:** Histograms of radius of gyration ( $R_G$ ) during the MD simulations.

	<b>R<sub>G</sub> (nm)</b>	<b>R<sub>H</sub> (nm)<sup>§</sup></b>	<b>D<sub>RH</sub> (μm<sup>2</sup>/s)<sup>¶</sup></b>	<b>D<sub>MSD</sub> (μm<sup>2</sup>/s)<sup>#</sup></b>
Tat <sub>11</sub> – <i>mon a</i>	0.85 (0.04)	1.10 (0.05)	223 (10)	-
Tat <sub>11</sub> – <i>mon b</i>	0.90 (0.03)	1.16 (0.04)	211 (7)	-
Tat <sub>11</sub> – <i>mon c</i>	0.86 (0.07)	1.11 (0.08)	221 (15)	195 (11)
Tat <sub>11</sub> -TAMRA – <i>mon a</i>	0.84 (0.02)	1.08 (0.03)	226 (6)	-
Tat <sub>11</sub> -TAMRA – <i>mon b</i>	0.87 (0.06)	1.12 (0.08)	218 (15)	193 (10)

<sup>§</sup> estimate from  $R_H = R_G * (5/3)^{1/2}$ .

<sup>¶</sup> estimate from Eq. 12, Section S2.

<sup>#</sup> from the mean square displacement during MD, see Methods

**Table S6:** Radius of gyration and estimated diffusion coefficient from 300ns Hamiltonian replica exchange simulations (HREX) and normal MD of the Tat<sub>11</sub> and Tat<sub>11</sub> –TAMRA monomers.

	<b>contact surface area (nm<sup>2</sup>)</b>	<b>number of inter-peptide H-bonds</b>	<b>number of inter-peptide backbone H-bonds</b>	<b>R<sub>G</sub> (nm)</b>	<b>R<sub>H</sub> (nm)</b>	<b>D<sub>RH</sub> (μm<sup>2</sup>/s)</b>	<b>D<sub>MSD</sub> (μm<sup>2</sup>/s)</b>
Tat <sub>11</sub> – <i>dim a</i>	4.6 (0.6)	10.0 (2.3)	1.9 (0.9)	1.21 (.04)	1.56 (.05)	157 (5)	156 (8)
Tat <sub>11</sub> – <i>dim a'</i>	5.4 (0.5)	10.7 (3.1)	1.5 (0.7)	1.13 (.03)	1.46 (.04)	168 (4)	-
Tat <sub>11</sub> – <i>dim a''</i>	4.8 (0.6)	8.1 (2.7)	1.6 (0.6)	1.25 (.09)	1.61 (.12)	152 (11)	-
Tat <sub>11</sub> – <i>dim a*</i>	4.8 (0.5)	10.5 (2.4)	1.9 (0.9)	1.17 (.04)	1.51 (.05)	162 (6)	155 (8)
Tat <sub>11</sub> – <i>dim b</i>	6.0 (0.9)	8.6 (3.0)	1.8 (1.1)	1.03 (.04)	1.33 (.05)	184 (7)	157 (8)
Tat <sub>11</sub> -TAMRA – <i>dim a</i>	9.5 (0.9)	12.4 (2.9)	3.7 (0.8)	1.22 (.06)	1.57 (.08)	156 (8)	155 (8)
Tat <sub>11</sub> -TAMRA – <i>dim b</i>	6.6 (1.2)	9.7 (2.9)	0.9 (0.6)	1.22 (.05)	1.57 (.07)	156 (6)	153 (8)

**Table S7:** Contact surface area, number of inter-peptide and inter-peptide backbone H-bonds, radius of gyration and estimated diffusion coefficient from MD simulations of the Tat<sub>11</sub> and Tat<sub>11</sub> – TAMRA dimers. Average values are reported (estimated errors in brackets).

## Section S4: Computational Details

### *Systems*

Starting configurations for Tat<sub>11</sub> were obtained by minimizing linear peptide structures build by VMD<sup>8</sup>, according to the sequence: **YGRKKRRQRRR**. For the Tat<sub>11</sub>-TAMRA system, the additional Cysteine residue was included in the C-terminus, and TAMRA was linked to the Cys side chain. The TAMRA molecule contains a maleimide terminal group. Upon reaction with the thiol group of the C-terminal Cysteine residue, this maleimide group converts into a succinimide segment linking the dye to the peptide (Fig. S20). The Amber99SB\*-ILDN force field was used for the simulations<sup>11,12</sup>. The parameterization of TAMRA and the succinimide segment is described below. For control simulations also the Charmm22\* was employed<sup>13</sup>, for the unlabelled Tat<sub>11</sub> peptide only. The systems were solvated in a truncated octahedron box with water molecules and 0.1 -0.15 M NaCl (for a total of 12 Na<sup>+</sup>, 20 Cl<sup>-</sup>, 6238 water molecules for monomer simulations; 24 Na<sup>+</sup>, 40 Cl<sup>-</sup>, 8563 water molecules for Tat<sub>11</sub>-*dim*; 26 Na<sup>+</sup>, 42 Cl<sup>-</sup>, 9239 water molecules for Tat<sub>11</sub>-TAMRA-*dim*). Periodic boundary conditions were applied and electrostatics was treated using the smooth particle mesh Ewald method with a grid spacing of 1.35 Å and a 11.0 Å real-space cut-off. TIP3P water model was used for the simulations. In all MD runs a 2-fs time step was used, LINCS was applied to constrains covalent bond lengths involving hydrogen atoms<sup>14</sup>, and the neighbour list is updated every 10 steps. Constant 300 K temperature and 1 bar pressure were maintained, respectively, by v-rescale thermostat<sup>15</sup> (with a coupling of  $\tau_T = 0.2$  ps) and Parrinello-Rahman barostat ( $\tau_P = 5$  ps). The simulations and analysis were performed with the Gromacs 5.0.5 package<sup>16</sup>, apart from secondary structure analysis, performed with STRIDE<sup>9</sup>. The errors in MD-averaged quantities are estimated by binning analysis.

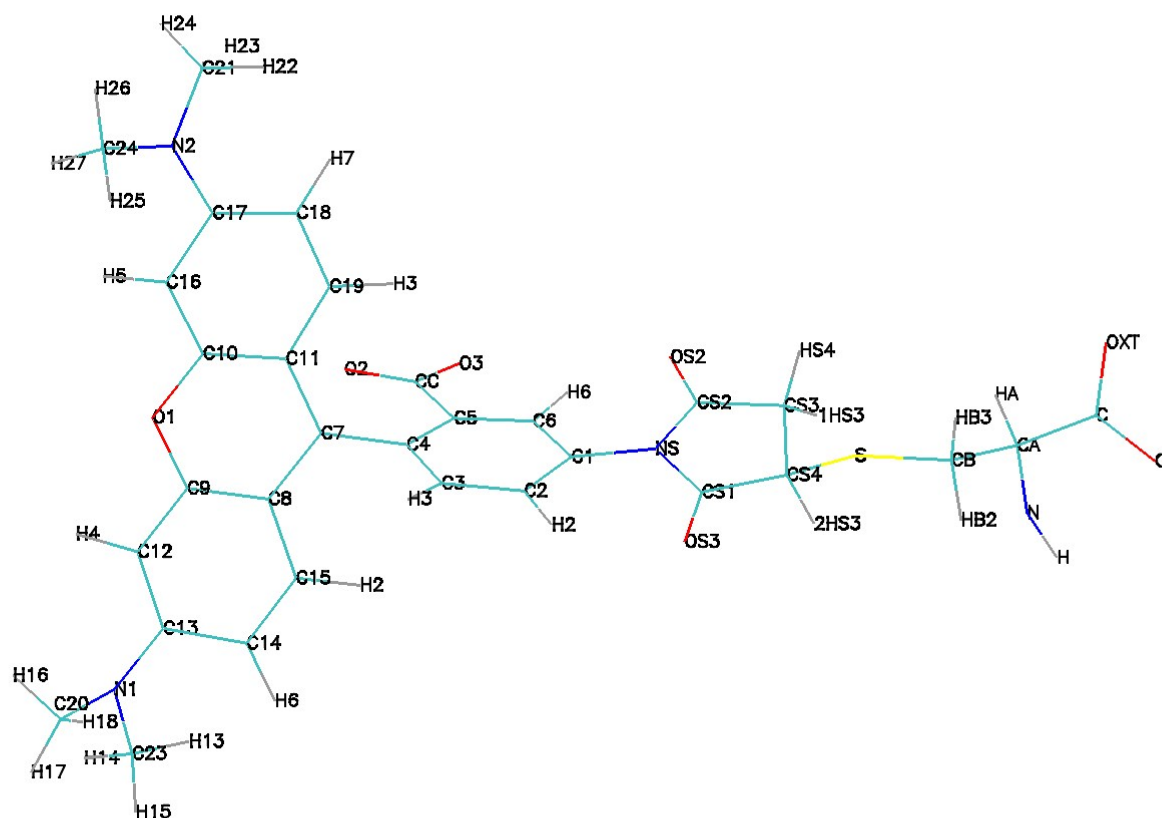
### *Parameterization of the TAMRA-succinimide-cysteine segment*

The atom types and partial charges of the TAMRA and succinimide were assigned in analogy to the Antechamber GAFF suggestions<sup>17,18</sup>, using RESP HF6-31G\* for partial charges and comparing previous parameterization of similar chemical groups. In detail, C-terminal Cysteine part was parameterized as in Amber99SB\*-ILDN with a modification of the partial charge on S (from -0.3102 to -0.2564). For the succinimide ring and the xanthene ring we followed the Amber atom type assignment in Ref.<sup>19</sup> with the definition of a new OA atom type. The other partial charges were obtained by a HF/6-31G\* RESP fitting on TAMRA+succinimide molecule, on a B3LYP 6-31G\*/PCM(water) geometry (the gas-phase optimized geometry resulted in the lactone form, with

the closing C4-C5-CO-O2-C7 ring, while the open zwitterionic form is known to be more stable in water<sup>20</sup>. The final charges and Amber/GAFF atom types are reported in Table S8.

Name	Type	Partial charge
C-term Cysteine		
O, OXT	O2	-0.7981
C	C	0.7497
CA	CT	-0.1635
N	N	-0.3821
H	H	0.2681
HA	H1	0.1396
CB	CT	-0.1196
HB2, HB3	H1	0.1437
S	S	-0.2564
Succinimide		
CS3,CS4	c3	-0.120
CS1,CS2	C	0.470
OS2,OS3	O	-0.521
HS4,1HS3	HC	0.090
2HS3	H1	0.190
NS	N	-0.030
TAMRA		
C1	CA	-0.017
C2	CA	-0.083
C3	CA	-0.265
C4	CA	-0.039
C5	CA	0.148
C6	CA	-0.180
CC	C	0.594
O2,O3	O	-0.663
H1	HA	0.151
H8	HA	0.129
H9	HA	0.163
C7	cc	0.182
C8, C11	CA	0.046
C15, C19	CA	-0.138
C14,C18	CA	-0.227
C13,C17	CA	0.162
N1,N2	NH	-0.052
C20,C21,C23,C24	c3	-0.188
C16,C12	CA	-0.301
O1	OA	-0.244
C9,C10	CA	0.245
H13-H18 H22-H27	H1	0.097
H2, H3	HA	0.181
H6,H7	HA	0.157
H4,H5	HA	0.152

**Table S8:** Atom types and partial charges for TAMRA- succinimide-cysteine (atom names are given in Fig. S20).



**Fig. S20:** Atom names of the TAMRA-succinimide-cysteine.

### ***Hamiltonian Replica Exchange and metadynamics simulations***

Metadynamics and Hamiltonian Replica Exchange (HREX) simulations were carried out using PLUMED2.0<sup>21</sup>. 12 replica were used with the following pseudo-temperature ranges: 300-1000K *Tat<sub>11</sub> mon a*, 300-800K *Tat<sub>11</sub>-TAMRA mon a*. The intermediate temperatures were chosen in order to guarantee at least a 10% success exchanges. The collective variable for metadynamics simulations, again using PLUMED, was chosen as the inter-peptide coordination number (defined as the number of inter-peptide pairs of non-hydrogen atoms within a 0.38 nm cutoff distance) and the Gaussian functions (sigma=0.5 and height=0.05 kJ/mol) were deposited each 2ps.

### ***Calculation of diffusion coefficient***

Diffusion coefficients for the various species were determined from the molecular dynamics simulations either using Eq. 12 with  $R_H = R_G \cdot (5/3)^{1/2}$  in the spherical particle approximation ( $D_{RH}$ ) or from the mean square displacement (MSD) of the center of mass during the molecular dynamics trajectory.



Corrections for finite simulation box and TIP3P water viscosity were included according to Ref.<sup>22</sup>. In details, the diffusion coefficient ( $D_0$ ) extrapolated for infinite box was obtained by:

$$D_0 = D_{PBC} + \frac{k_B T}{6\pi\eta_{TIP3P}L} \left( \xi - \frac{4\pi R_H^2}{3L^2} \right) \quad (15)$$

where  $L$  is the box length,  $\eta_{TIP3P}$  is the TIP3P water viscosity ( $3.08 \cdot 10^{-4} \text{ kg m}^{-1} \text{ s}^{-1}$ ),  $\xi$  is the numerical factor for self-correction. In the case of truncated octahedron boxes it was calculated to be  $\xi \approx 3.15166862$ . To correct for the experimental water viscosity the extrapolated  $D_0$  is scaled according to  $D_{MSD} = \eta_{exp}/\eta_{TIP3P} D_0$ , with  $\eta_{exp} = 8.96 \cdot 10^{-4} \text{ kg m}^{-1} \text{ s}^{-1}$ .

$D_{PBC}$  was obtained from the linear coefficient of the MSD regression in a 1-10 ns window, according to the relation  $MSD = 6 D t$ . This procedure was performed only for normal MD runs and for simulations longer than 1  $\mu\text{s}$ .

### ***Hydrogen bond analysis***

The number of hydrogen bonds (salt bridges are treated as hydrogen bonds) was estimated by the *gmx hbond* routine of Gromacs. As common in MD simulation analysis, the presence of a H-bond is deduced from the local geometry of the donor-hydrogen-acceptor triplet. A H-bond was considered as occupied when the donor-acceptor distance was smaller than 3.5 Å and the donor-hydrogen-acceptor angle smaller than 30°. Both values are the standard choice in this kind of analysis.

The stacking between Arginine guanidinium groups with other guanidinium groups or with TAMRA was evaluated based on the distance between the involved groups and the angle formed by the normal to the plane containing them. To evaluate the distance, for Arginine we used the central carbon atom of the guanidinium group and for TAMRA the C7 carbon atom (Fig. S20); to define the planes we used the three nitrogen of guanidinium and the C7 C9 C10 carbon atoms of TAMRA. The stacking was considered as present when the distance was less than 5.5 Å and the angle less than 30°.

## Supporting references

- 1 Sjoback, R., Nygren, J. & Kubista, M. Absorption and fluorescence properties of fluorescein. *Spectrochimica Acta Part A* **51**, L7-L21 (1995).
- 2 Macchi, S. *et al.* Spontaneous membrane-translocating peptides: influence of peptide self-aggregation and cargo polarity. *Sci Rep* **5**, 16914, doi:10.1038/srep16914 (2015).
- 3 Corsepius, N. C. & Lorimer, G. H. Measuring how much work the chaperone GroEL can do. *Proc Natl Acad Sci U S A* **110**, E2451-2459, doi:10.1073/pnas.1307837110 (2013).
- 4 Ajtai, K. *et al.* Stereospecific reaction of muscle fiber proteins with the 5' or 6' isomer of (iodoacetamido)tetramethylrhodamine. *Biochemistry* **31**, 12431-12440 (1992).
- 5 Bangham, A. D., Standish, M. M. & Weissmann, G. The action of steroids and streptolysin S on the permeability of phospholipid structures to cations. *J Mol Biol* **13**, 253-259 (1965).
- 6 Mayer, L. D., Hope, M. J., Cullis, P. R. & Janoff, A. S. Solute distributions and trapping efficiencies observed in freeze-thawed multilamellar vesicles. *Biochim Biophys Acta* **817**, 193-196 (1985).
- 7 Lira, R. B., Steinkuhler, J., Knorr, R. L., Dimova, R. & Riske, K. A. Posing for a picture: vesicle immobilization in agarose gel. *Sci Rep* **6**, 25254, doi:10.1038/srep25254 (2016).
- 8 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38 (1996).
- 9 Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566-579, doi:10.1002/prot.340230412 (1995).
- 10 Lee, D., Lee, J. & Seok, C. What stabilizes close arginine pairing in proteins? *Phys Chem Chem Phys* **15**, 5844-5853, doi:10.1039/c3cp00160a (2013).
- 11 Best, R. B. & Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* **113**, 9004-9015, doi:10.1021/jp901540t (2009).
- 12 Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958, doi:10.1002/prot.22711 (2010).
- 13 Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys J* **100**, L47-49, doi:10.1016/j.bpj.2011.03.051 (2011).
- 14 Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472 (1997).

- 15 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101, doi:10.1063/1.2408420 (2007).
- 16 Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19-25 (2015).
- 17 Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* **25**, 247-260, doi:10.1016/j.jmglm.2005.12.005 (2006).
- 18 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157-1174, doi:10.1002/jcc.20035 (2004).
- 19 VanBeek, D. B., Zwier, M. C., Shorb, J. M. & Krueger, B. P. Fretting about FRET: correlation between kappa and R. *Biophys J* **92**, 4168-4178, doi:10.1529/biophysj.106.092650 (2007).
- 20 Pedone, A. & Barone, V. Unraveling solvent effects on the electronic absorption spectra of TRITC fluorophore in solution: a theoretical TD-DFT/PCM study. *Phys Chem Chem Phys* **12**, 2722-2729, doi:10.1039/b923419e (2010).
- 21 Bussi, G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Molecular Physics* **112**, 379-384 (2014).
- 22 Yeh, I. C. & Hummer, G. Diffusion and electrophoretic mobility of single-stranded RNA from molecular dynamics simulations. *Biophys J* **86**, 681-689, doi:10.1016/S0006-3495(04)74147-8 (2004).

**Movie 1:** Extract from the Tat<sub>11</sub>-TAMRA - *dim a* simulation (2 to 2.5  $\mu$ s). The two peptides are shown with different colours (yellow and orange) for the cartoon representation. The blue and red spheres correspond to the C $\alpha$  of the N- and C-terminal residues respectively. Hydrogen bonds are shown as dotted green lines. Arginine 10 of each peptide is highlighted (thicker sticks). Snapshots are taken each 50 ps and a running average over 2 snapshots is taken to smooth the molecular motion. The trajectory is fitted on the two terminal residues of each residue (R9 and R10).

**Movie 2:** Extract from the Tat<sub>11</sub> - *dim a* simulation (2 to 2.5  $\mu$ s). See caption of Movie 1 for further details.