Electronic Supplementary Information for:

P-sulfonatocalix[4]arene Turns Peptide Aggregates into an Efficient Cell-penetrating Peptide

Mahsima Heydari [1]; Najmeh Salehi [2]; Reza Zadmard [3]; Werner M. Nau[4]; Khosro Khajeh [5]; Zahra Azizi [6] and Amir Norouzy [*1]

[1] Bioprocess Engineering Department, National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran

[2] School of Biology, College of Science, University of Tehran, Tehran, Iran.

[3] Department of Organic Chemistry, Chemistry and Chemical Engineering Research Center of Iran, Tehran, Iran

[4] School of Science, Constructor University, Bremen, Germany

[5] Department of Biochemistry, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran

[6] Department of Molecular Medicine, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences, Tehran, Iran

Corresponding author: Amir Norouzy (a.norouzy@nigeb.ac.ir)

Peptide synthesis, purification, and identification

Peptide synthesis was carried out in a solid-phase peptide synthesis vessel. 2-chlorotrityl chloride resin (CTC) resin (300 mg, 0.4 mmol/g) was swollen in 3 ml of DMF for 30 min. The first amino acid from the C-terminus was attached to the resin with a solution of Fmoc-Arg (Pbf)-OH (86 mg, 0.119 mmol) in 4.5 ml DCM:DMF (9:1) containing 5 equivalent of DIPEA relative to Arg and incubated for 60 min. The unreacted functional groups of the resin were capped with methanol. The Fmoc of Arg was removed with a solution of 20% piperidine in DMF in 20 min. The freed α-amine of Arg was then treated with a solution of the second amino acid (343 mg, 0.476 mmol), HBTU (0.95 equivalent), DIPEA (2.5 equivalent,) in DMF for 20 min. The amino acid coupling and Fmoc deprotection was repeated to the last amino acid. The Kaiser test was used to check the completeness of both Fmoc removal and coupling reactions. Finally, 5-carboxyflourescein (FAM) was doubly couped to the N-terminus of the peptide. The labeled-peptide/resin was filtered and

washed three times with DCM, then with DMF and finally with methanol, and left to dry under vacuum. The cleavage process was carried out with a cleavage cocktail contained TFA 90 %, phenol 5%, water 2.5 % and diisopropylsilane 2.5% (as the scavenger) for 7 hr. As a result, the peptide was cleaved from the resin and the protecting groups of Arg and Tyr sidechains were removed. The cleaved peptide was partially purified by precipitation with cold diethyl ether. The precipitate was dissolved in 50% of acetonitrile in water and subjected to purification with HPLC. The impure peptide solution was injected into a C-18 preparative HLPC column (150 × 2.2 mm). The mobile phase gradient was changed from 5% of acetonitrile in water to 5% of water in acetonitrile in 60 min. The pick containing the peptide was harvested by a fraction collector and lyophilized. Purity of the purified peptide was assayed by an analytical HPLC (Figure S1).

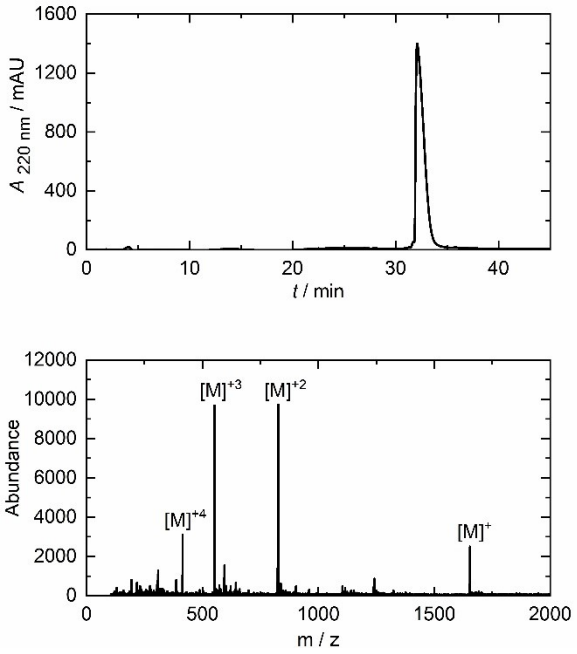


Figure S1. Chromatogram (on the top) and mass spectrum (at the bottom) of the pure *FAM*-Y_4R_4 peptide.

To confirm the accuracy of the peptide synthesis, the molucular weight of the purified peptide was measured by a mass spectrometer. The molecular weight of the peptide is theoretically 1653.75 $gr.mol^{-1}$. The corresponding m/z values of the peptide ion that obtained 1-4 proton is observed in the spectrum (Figure S1); therefore, the peptide was synthesized accurately.

Circular dichroism (CD)

A solution of 35 μM peptide was prepared in 10 mM phosphate buffer saline in D_2O and the pD was adjusted to 7. The peptide solution was subjected to CD measurement. D_2O was preferred over water as the peptide solvent because the hydroxide group of water enormously absorbs UV light from 190 to 205 nm consequently, less circular light reaches to the detector that makes the obtained data inaccurate [1]. The CD spectrum was recorded with Jasco J-810 CD Spectrometer using 0.1 cm quartz cuvette. The spectrum was recorded from 190–260 nm. The obtained CD spectrum did not exhibit any well-known α-helix or β-sheet pattern (Figure S2). Consequently, *FAM*-Y_4R_4 aggregates is not a self-assembly of any common secondary structures. However, small positive bands at 225 and 205 nm together with a negative band at 190 nm resemble turns structure.[2]

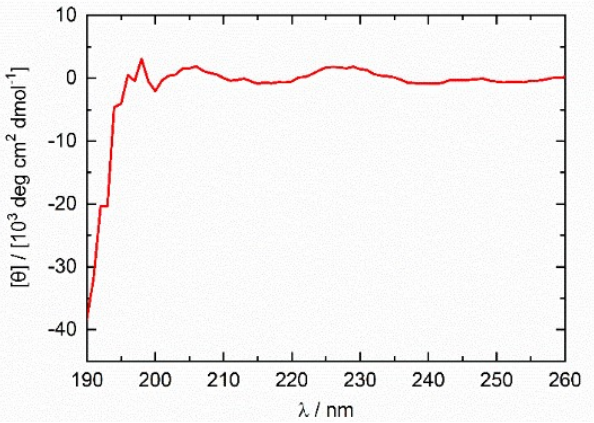


Figure S2. Circular dichroism spectrum of *FAM*-Y_4R_4.

Quantifying the cation-π interaction

The side chains of Arg and Tyr contain guanidino and phenol (Ph) respectively. To illustrate the cation-π interaction between the side chains, guanidine hydrochloride (Gn.HCl, as representative of guanidino) was spectrally titrated with phenol. The Gn.HCl/Ph complex formation was detected by increase in the OD spectrum with λ_{max} = 270 nm (the inset of Figure S3). By plotting the OD at 270 nm against the concentration of Gn.HCl a binding isotherm was obtained (Figure S3). By using a nonlinear regression algorithm, the K_a value of the cation-π interaction between Gn.HCl and Ph (1:1 model) was calculated to be 9.7×10^5 M^{-1}. The obtained K_a value was used in ΔG =

$-RT\ln K_a$ equation (in which R = 1.98×10^{-3} kcal.K^{-1}.mol^{-1}, and K = to 298 Kelvin) and ΔG = -8.13 kcal.mol^{-1} was obtained.

The guanidino and phenol (Ph) groups are present in the side chains of Arg and Tyr, respectively. To demonstrate the cation-π interaction between these side chains, phenol was subjected to spectral titration with guanidine hydrochloride (Gn.HCl, serving as a representative of guanidino). The formation of the Gn.HCl/Ph complex was observed through the increase in the OD spectrum, with a maximum wavelength (λ_{max}) of 270 nm (Figure S3, inset). A binding isotherm was obtained by plotting the OD at 270 nm against the concentration of Gn.HCl (Figure S3), using a nonlinear regression algorithm. The resulting K_a value for the cation-π interaction between Gn.HCl and Ph (1:1 model) was calculated to be 9.7×10^5 M^{-1}. This K_a value was then utilized in the equation ΔG = –RTlnKa (where R = 1.98×10^{-3} kcal.K^{-1}.mol^{-1}, and T = 298 Kelvin), yielding ΔG = -8.13 kcal.

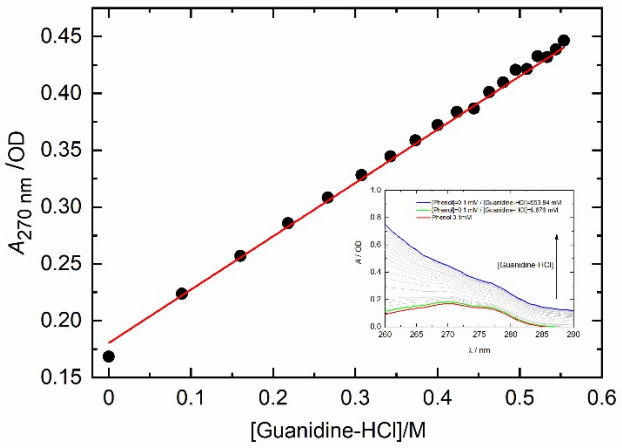


Figure S3. Binding isotherm of guanidine hydrochloride to phenol, fitted using a nonlinear regression algorithm. The inset depicts the spectral titration.

Molecular dynamics simulation

The toppar and parameters file of *FAM* was generated with CHARMM General Force Field (CGenFF) program version 2.5. All MD simulations were performed with CHARMM36m for proteins, CHARMM36 for cgenff, and *FAM* force field in NAMD2.13. The psfgen plugin of VMD 1.9.3 was used to generate the psf file. The system was solvated using TIP3P water and ionized to neutralize the system with NaCl. The *FAM*-Y_4R_4 structure was minimized with the conjugate gradient method for 50,000 steps. In the equilibration run, the system was heated from 0 to 300 K

gradually with the constant Berendsen pressure for 3 ns. Then the equilibration run was continued with constant pressure at 300 K for 2 ns. After that, three same independent production runs were carried out for a totally of 150ns under NPT conditions. The pressure and temperature were controlled with Nose-Hoover Langevin piston pressure and Langevin dynamics. The time step of the equilibration and production runs were set at 1 and 2 fs, respectively. The periodic boundary conditions were applied to the system. The particle mesh Ewald (PME) algorithm was used to calculate the long-range electrostatics interaction. The electrostatic and van der Waals interactions cutoff were set to 12 Å, and the switch distance cutoff of 10 Å was used for the smoothing function. The long-range electrostatic and short-range nonbonded interactions were measured every 2 and 1 fs, respectively.

Structural analysis

The structural analysis of the *FAM*-Y_4R_4 monomer and aggregate structure trajectories such as root mean square deviations (RMSDs), the radius of gyration (Rg), the solvent accessible surface area (SASA), the number of hydrogen bonds, Cα coordinates, clustering, and graphical representations were performed by VMD 1.9.3. The clustering method is based on the quality threshold (QT) algorithm and the distance function of "fitrmsd" was chosen to measure the root mean squared atom-to-atom distance after alignment. The number of clusters and maximal distance value between two frames that are considered similar were set to 4 and 5, respectively. The non-covalent interactions were shown with LigPlot+v.1.4.

Sampling of *FAM*-Y_4R_4 conformations

The 3D model of *FAM*-Y_4R_4 was built and the MD simulation was used to refine the model and sample a structural ensemble for this peptide. After minimizing and equilibrating, three independent 50 ns MD simulations were carried out for *FAM*-Y_4R_4. Figure S4A shows a root mean square deviation (RMSD) time-course pattern of this peptide with respect to the initial structure.

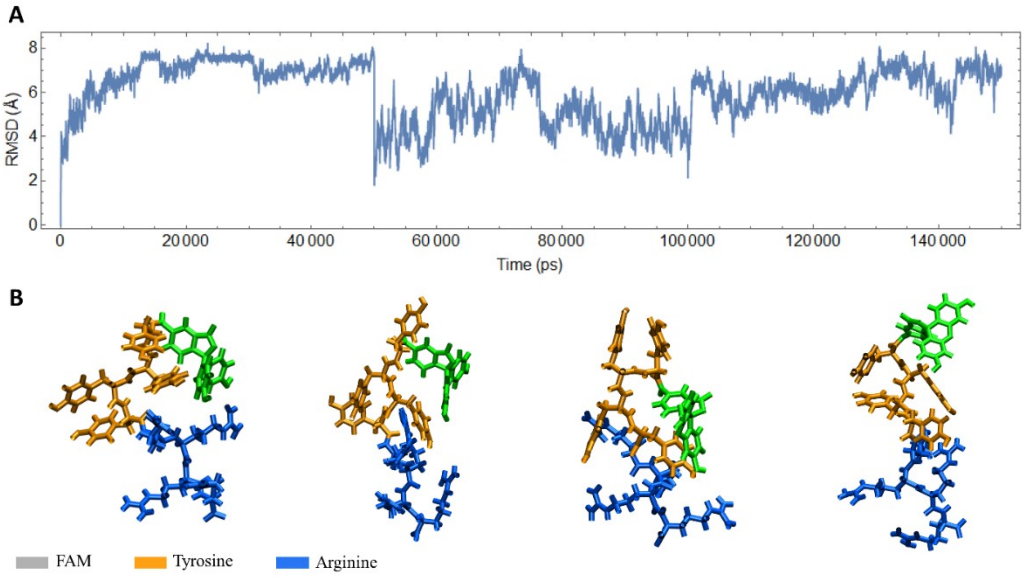


Figure S4. Root mean square deviation (RMSD) values of *FAM*-Y_4R_4 monomer during the three 50 ns MD simulations with regard to the initial structure (A), The representative structure of the four most stable clusters of *FAM*-Y_4R_4 monomers (B).

The RMSD value is a standard measure that calculates the average distance between atoms of two superimposed conformations, which changed from 1.25 to 8.24 Å in these trajectories. This extended RMSD values range showed that these three MD simulations provided a diverse structural ensemble for the *FAM*-Y_4R_4 peptide. Clustering was applied to reveal the most stable clusters of *FAM*-Y_4R_4 whose representative structures were displayed in figure S4B.

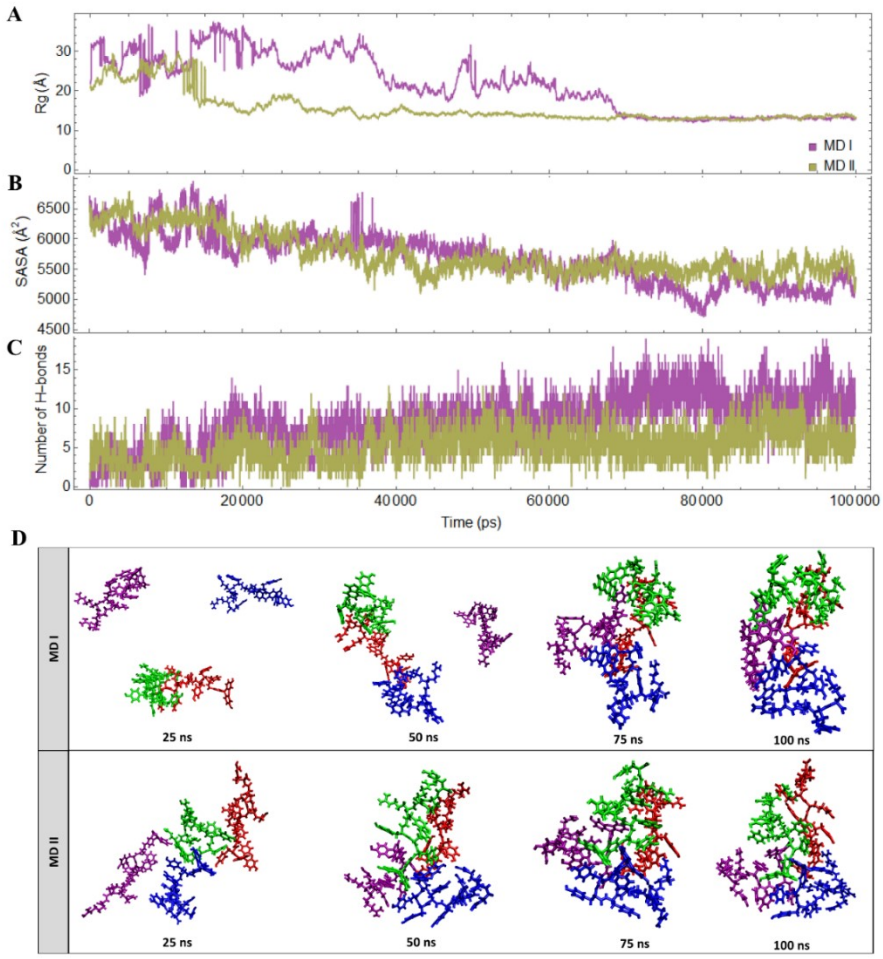


Figure S5. Post-processing analysis of MD I and II trajectories along 100 ns simulations: Rg values (A), SASA values (B), and the number of hydrogen bonds (C). Time evolution of peptide aggregation: The representation of MD I and II simulations of the four most stable *FAM*-Y_4R_4 conformations after 25, 50, 75, and 100 ns of production runs. The four different *FAM*-Y_4R_4 conformations were shown in blue, red, green, and purple (D).

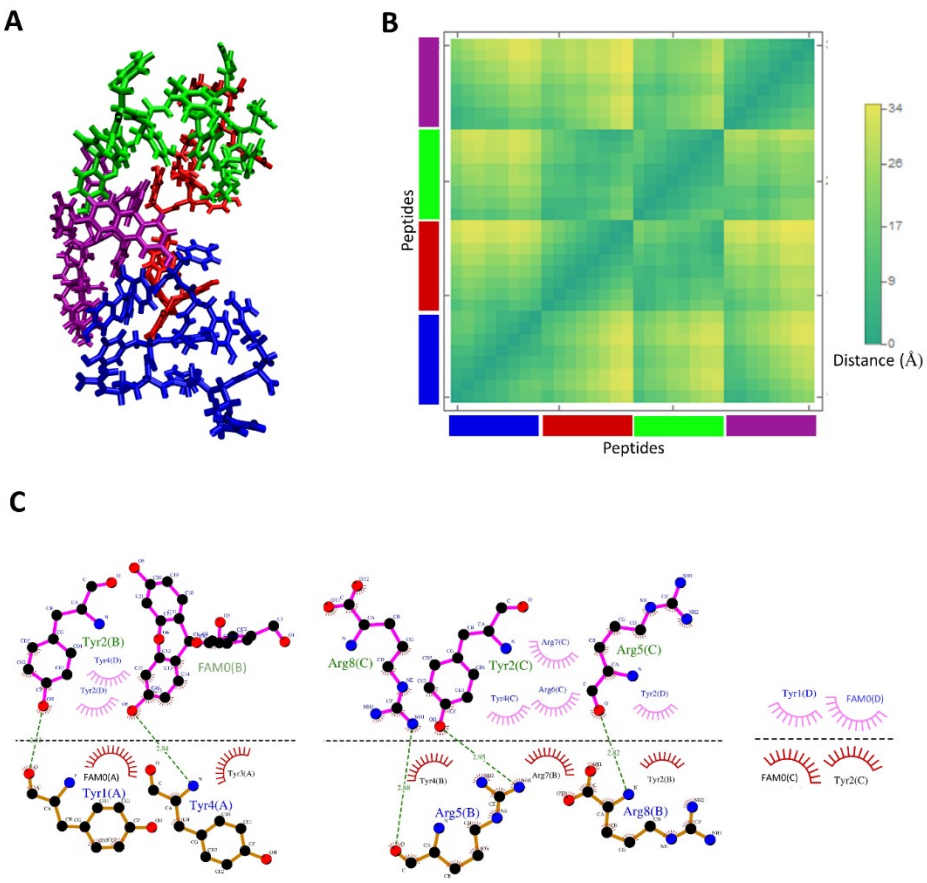
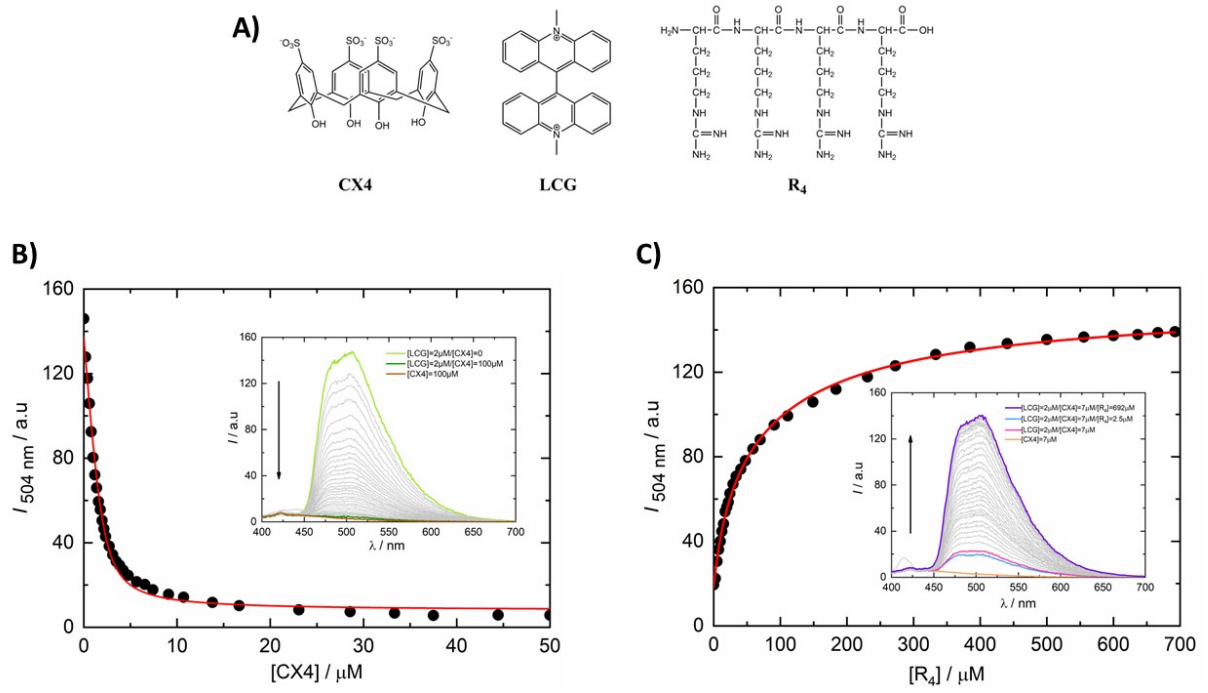


Figure S6. The representation of the MD I system at 100 ns (A). The distance map of the four *FAM*-Y_4R_4 peptides in the structure of panel A (B), The hydrogen bonds and hydrophobic contacts between four *FAM*-Y_4R_4 peptides in the structure of panel A (C).

Indicator displacement assay

The interaction between R_4 peptide and *p*-sulphonato-calix[4]arene (CX4) was investigated using Indicator Displacement Assay (IDA). In this method, CX4 was incubated with the fluorescent dye N,N'-dimethyl-9,9'-biacridinium dinitrate (lucigenin, LCG), which entered the CX4 cavity to form the CX4/LCG complex. While inside CX4, LCG fluorescence was quenched. The binding constant of LCG/CX4 was determined through fluorescence spectral titration. The LCG concentration was fixed at 2 µM, while the CX4 concentration (host) was increased from 0.2 to 100 µM in a 20 mM PBS solution at pH 7. Emission spectra of LCG at each CX4 concentration were recorded using a Cary Eclipse fluorescence spectrometer, exciting LCG at 369 nm and measuring emission intensity from 400 to 700 nm. The binding isotherm curve of LCG/CX4 was plotted at the maximum

emission wavelength (λ_{max} = 504 nm), yielding $K_{a\ (LCG/CX4)} = 3 \times 10^6\ M^{-1}$ (Figure S7). Previous studies have suggested a binding constant for LCG/CX4 of $K_{a\ (LCG/CX4)} \approx 10^7\ M^{-1}$ [3-5].



D)

Figure S7. A) Chemical structures of *p*-sulfonatocalix[4]arene (CX4), Lucigenin (LCG), and R_4 peptide. B) Binding isotherm of LCG to CX4 calculated using a one host to one guest model. C) Binding isotherm of R_4 to CX4 obtained by displacing R_4 with LCG in the CX4/LCG complex. The insets in the isotherms display the fluorescence titrations.

To investigate the binding constant of R_4 with CX4 ($K_{a\ (CX4/R4)}$) various amounts of R_4 peptide were added to a fixed concentration of LCG/CX4 solution. Each arginine side chain in R_4 could potentially compete with LCG for occupancy within the CX4 cavity:

$$n\ CX4/LCG + R_4 \leftrightharpoons n\ CX4/R_4 + n\ LCG \qquad (n \leq 4)$$

The release of LCG from CX4 resulted in fluorescence turn-on of the solution. Emission spectra were obtained by exciting LCG at 369 nm and recording emission intensities from 400 to 700 nm. The binding isotherm of peptide to CX4 was constructed by plotting fluorescence intensity values at λ_{max} = 504 nm against R_4 concentration (Figure S7). The binding isotherm curve was fitted using non-linear regression analysis with OriginLab software, employing a one host, one guest, and one competitor algorithm model. The average binding constant of CX4 to R_4 was determined as K_a

$_{(CX4/R4)}$ = 4.5 × 10^5 M^{-1}. Previous research has shown that the binding constant of CX4 with free arginine amino acid at pH 6 is 2800 M^{-1} [6]. This indicates that CX4 exhibits a much higher affinity for R_4 than for a single Arg amino acid due to a phenomenon known as multivalency in which multiple ligand molecules bind to a receptor molecule [7].

Fluorescence microscopy

MCF-7 and A549 cells were seeded at a density of 1×10^4 cells/well in 96-well plates. The cells were then incubated with *FAM*-Y_4R_4 or *FAM*-Y_4R_4/CX4 for 18 hours. After incubation, the medium was removed and the cells were washed three times with PBS.

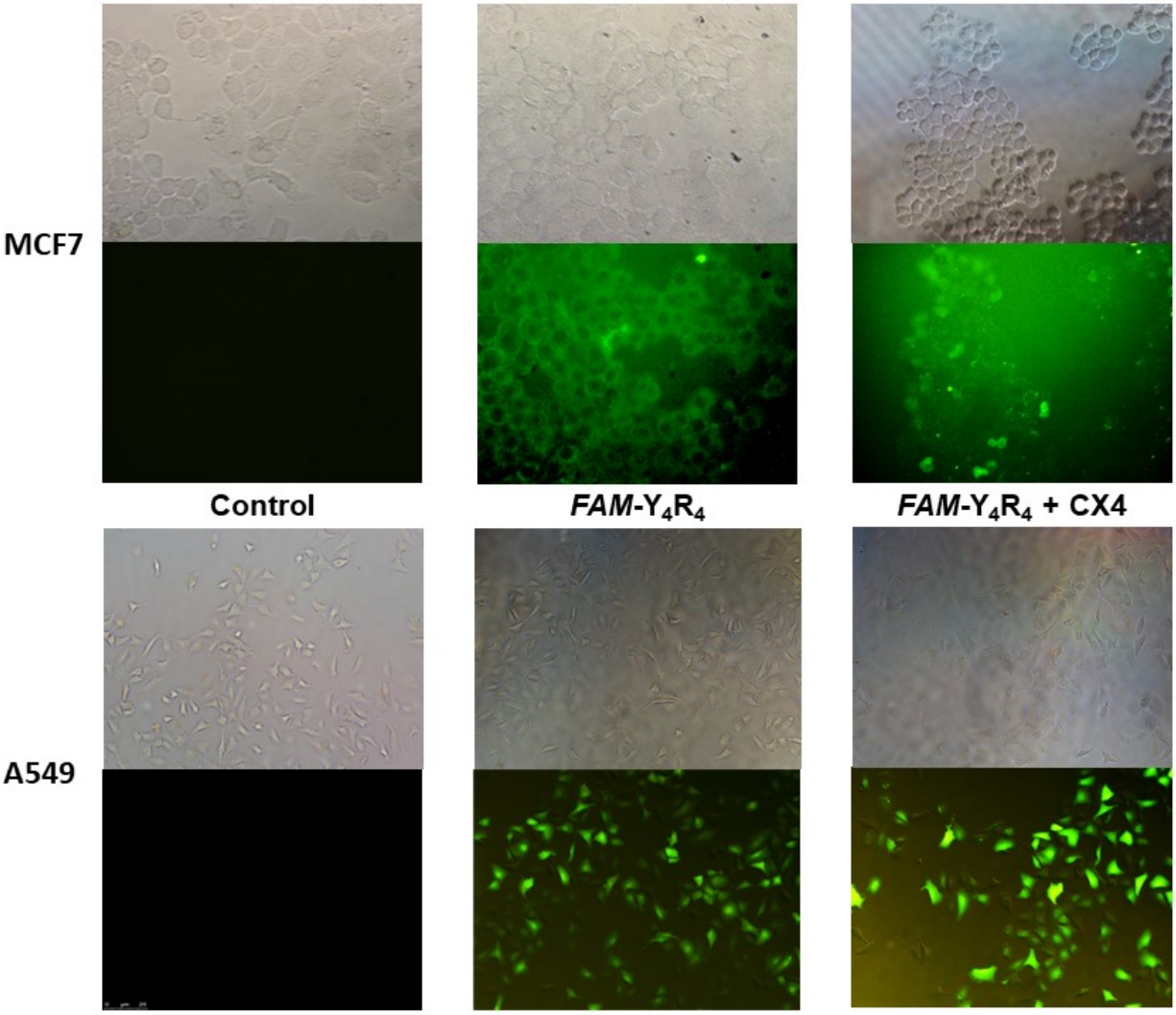


Figure S8. Fluorescence and phase contrast images of MCF-7 (on the top) and A549 cells (at the bottom) incubated with *FAM*-Y_4R_4 in the presence and absence of CX4 for 18 hours.

The resulting images were captured using a fluorescent microscope. For the control group, cells were not incubated with any substances, and images were captured using the same experimental setup (Figure S8).

References

1. P. Zuman and W. Szafranski, *Analytical Chemistry*, 1976, **48**, 2162-2163.
2. A. Rodger, in *Encyclopedia of Biophysics*, eds. G. Roberts and A. Watts, Springer Berlin Heidelberg, Berlin, Heidelberg, 2018, DOI: 10.1007/978-3-642-35943-9_634-1, pp. 1-6.
3. G. Ghale, A. G. Lanctôt, H. T. Kreissl, M. H. Jacob, H. Weingart, M. Winterhalter and W. M. Nau, *Angewandte Chemie International Edition*, 2014, **53**, 2762-2765.
4. D.-S. Guo and Y. Liu, *Acc. Chem. Res.*, 2014, **47**, 1925-1934.
5. A. Norouzy, Z. Azizi and W. M. Nau, *Angew. Chem. Int. Ed.*, 2015, **54**, 792-795.
6. A. Hennig, H. Bakirci and W. M. Nau, *Nature methods*, 2007, **4**, 629-632.
7. A. Mulder, J. Huskens and D. N. Reinhoudt, *Org. Biomol. Chem.*, 2004, **2**, 3409-3424.