ELECTRONIC SUPPLEMENTARY INFORMATION

A designed DNA binding motif that recognizes extended sites and spans two adjacent major grooves

Jéssica Rodríguez, Jesús Mosquera, Rebeca García-Fandiño, M. Eugenio Vázquez,* and José L. Mascareñas*

Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares (CIQUS), Departamento de Química Orgánica. Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain.

Abbreviations

DEDTC: sodium diethyldithiocarbamate

DIC: *N,N′*-Diisopropylcarbodiimide

EDT: ethanedithiol

HATU: 2-(1H-7-aza- benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate

HBTU: 2-[(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate

TMR: tetramethylrhodamine dye

TFA: trifluoroacetic acid

TIS: triisopropylsilane

General peptide synthesis procedures

All peptide synthesis reagents and amino acid derivatives were purchased from *GL Biochem* (Shanghai) and *NovaBiochem*; amino acids were purchased as protected Fmoc amino acids with the standard side chain protecting scheme: Fmoc-Ala-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Ser(*t*-Bu)-OH, Fmoc-Glu(O*t*-Bu)-OH, Fmoc-Trp(Boc)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Ile-OH, Fmoc-Thr(*t*-Bu)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-His(Trt)-OH and Fmoc-Asp(O*t*-Bu)-OH except for the orthogonally protected Fmoc-Lys(Alloc)-OH, which was purchased from *Bachem*. All other chemicals were purchased from *Aldrich* or *Fluka*. All solvents were dry and synthesis grade, unless specifically noted.

Peptides were synthesized using an automatic peptide synthesizer from *Protein Tecnologies PS3 PeptideSynthesizer*. Peptide synthesis was performed using standard Fmoc solid-phase methods on a PAL–PEG–PS resin (0.19 mmol/g) using HBTU/HOBt (4 equiv) as coupling agent, DIEA as base (6 equiv) and DMF as solvent. The deprotection of the temporal Fmoc protecting group was performed by treating the resin with 20% piperidine in DMF solution for 20 min. For **BR**, after the N-terminal aspartic acid, we coupled p-ABA chromophore.[1]

Lys(Alloc) side chain deprotection: The resin containing the alloc peptide (75 mg, approx. 0.015 mmol) was suspended in CH_2Cl_2 and shaken for 1 h to ensure a good swelling. Then $Pd(OAc)_2$ (1 mg, 0.3 equiv), 4-methylmorpholine (16 µL, 10 equiv), PPh_3 (6 mg, 1.5 equiv), $PhSiH_3$ (18 µL, 10 equiv) and 1 mL CH_2Cl_2 were added, and the mixture shaken overnight. The resin was then filtered and washed with DMF, DEDTC, DMF and CH_2Cl_2 and dried under a nitrogen stream.

Coupling of acetic anhydride: The resin containing the peptide (50 mg, aprox. 0.01 mmol) was suspended in DMF and shaken for 1 h to ensure good swelling. A solution of acetic anhydride (20 uL, 0.2 mmol, 20 equiv), and DIEA (400 µL, 0.195 M in DMF, 0.08 mmol, 8 equiv)

[1] I. Coin, M. Beyermann, M. Bienert, *Nat. Protoc.,* **2007**, *2*, 3247–3256.

was added. The reaction mixture was shaken for 0.5 h. The resin was washed with DMF (3 ×, 5 min) and CH_2Cl_2 (2 ×, 5 min).

Bromoacetic acid coupling: The resin containing the peptide (50 mg, aprox. 0.01 mmol) was suspended in CH_2Cl_2 and shaken for 1 h to ensure good swelling. On the other hand, bromoacetic acid (28 mg, 20 equiv) was dissolved in 630 µL CH_2Cl_2 and cooled to 0 ºC. Diisopropylcarbodiimide (DIC, 16 µL, 10 equiv) was added, and the mixture stirred at 0 ºC for 20 min. The solid was filtered off and the filtrate was added over the resin. The suspension was shaken for 30 min. The resin was then filtered and washed with CH_2Cl_2.

TMR coupling: For TMR-**BR(Hk)**-Br, after the N-terminal aminoacid, we coupled Fmoc-6-aminohexanoic acid as spacer between the peptide and the fluorophore, and finally TMR fluorophore: 5(6)-carboxytetramethylrhodamine was coupled using 3 equiv of the rhodamine (0.15 mmol, 64.5 mg), 3 equiv of HATU and 5 equiv of DIEA 0.2 M in DMF for 60 min.

Peptide cleavage/deprotection: The resin-bound peptide was treated for approximately 2 h with the following cleavage cocktails: 940 µL TFA, 25 µL EDT, 25 µL H_2O and 10 µL TIS (1 mL of cocktail / 40 mg resin) for cysteine-containing peptides, and 900 µL TFA, 50 µL CH_2Cl_2, 25 µL H_2O and 25 µL TIS (1 mL of cocktail / 40 mg resin) for the rest of the peptides.

High-Performance Liquid Chromatography (HPLC): HPLC was performed using an *Agilent 1100* series Liquid Chromatograph Mass Spectrometer system. Analytical HPLC was carried out using a *Eclipse XDB-C18 analytical column (4.6 x 150 mm, 5 µm)*, 1 mL/min, gradient 5 to 75% B over 30 min. Purification of the peptides was performed on a semipreparative *Phenomenex Luna-C18* (250 × 10 mm) reverse-phase column. The crude products were purified by reverse-phase (RP) HPLC, 4 mL/min, gradient 10 to 50% B over 40 min. (A: H_2O 0.1% TFA, B: CH_3CN 0.1% TFA) and identified as the desired peptides.

Electrospray Ionization Mass Spectrometry (ESI/MS): ESI was performed with an *Agilent 1100* Series LC/MSD VL G1956A model in positive scan mode.

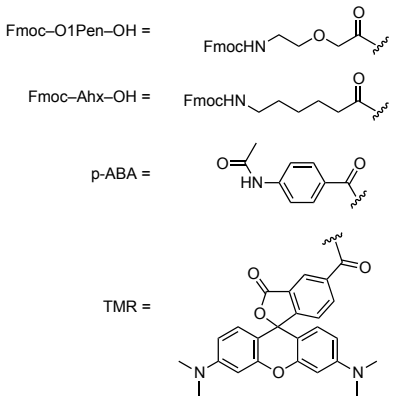


Fig. S1 Structures of Fmoc-O1Pen-OH, Fmoc-Ahx-OH, p-ABA and TMR.

BR(Hk)-Br: was isolated with an approx. yield of 22%.

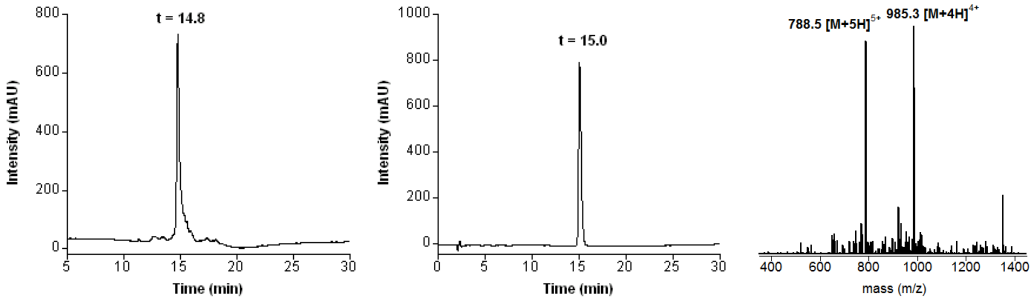


Fig. S2 HPLC chromatogram of: Left) crude residue after the synthesis; Center) purified peptide. Gradient 5 to 75% B over 30 min. Right) Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. for $C_{165}H_{287}BrN_{62}O_{45}$: 3936.13. Found: 985.3 $[M+4H]^{4+}$; 788.5 $[M+5H]^{5+}$.

GAGA-SH was isolated with an approx. yield of 17%.

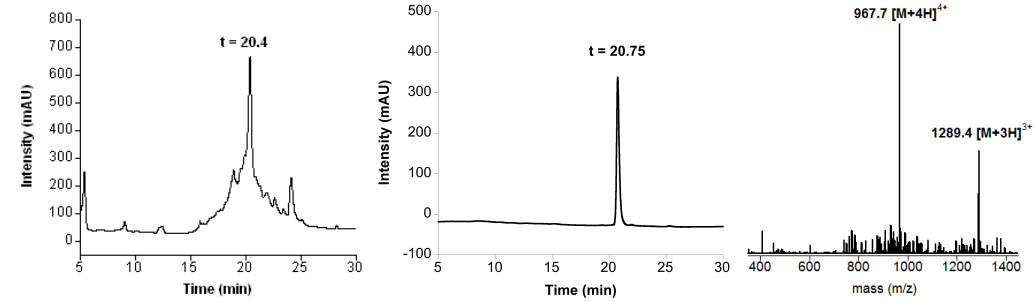


Fig. S3 HPLC chromatogram of: Left) crude residue after the synthesis; Center) purified peptide. Gradient 5 to 75% B over 30 min. Right) Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. for $C_{166}H_{267}N_{53}O_{48}S_3$: 3869.43. Found: 1289.4 $[M+3H]^{3+}$; 967.7 $[M+4H]^{4+}$.

BR(Gly$_9$)-Br: was isolated with an approx. yield of 20%.

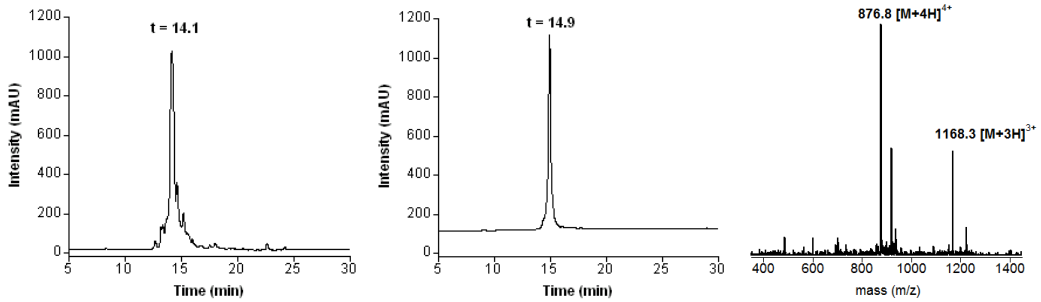


Fig. S4 HPLC chromatogram of: Left) crude residue after the synthesis; Center) purified peptide. Gradient 5 to 75% B over 30 min. Right) Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. for $C_{141}H_{238}BrN_{55}O_{45}$: 3500.72. Found: 1168.3 $[M+3H]^{3+}$; 876.8 $[M+4H]^{4+}$.

TMR-BR(Hk)-Br: was isolated with an approx. yield of 26%.

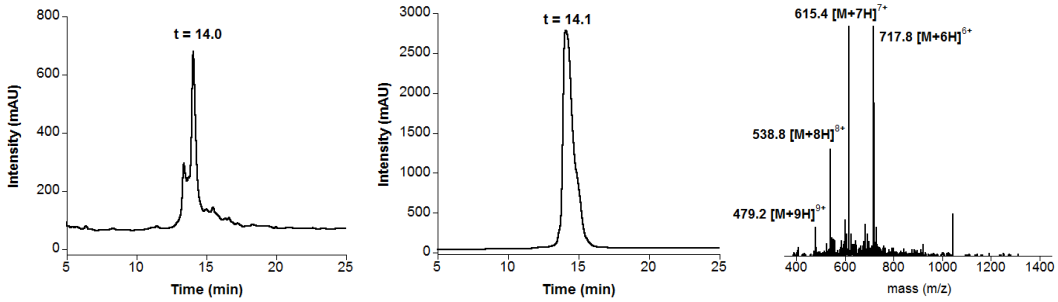


Fig. S5 HPLC chromatogram of: Left) crude residue after the synthesis; Center) purified peptide. Gradient 15 to 55% B over 40 min. Right) Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. for $C_{187}H_{313}BrN_{65}O_{47}$: 4301.31. Found: 717.8 $[M+6H]^{6+}$; 615.4 $[M+7H]^{7+}$; 538.8 $[M+8H]^{8+}$; 479.2 $[M+9H]^{9+}$.

Synthesis of the peptide BR(Gly$_9$)GAGA:

GAGA-SH (0.1 mg, 3 × 10^{-5} mmol) was dissolved in deoxygenated phosphate buffer (130 µL, 10 mM, pH = 7.5). A $ZnSO_4$ solution (4.5 × 10^{-5} mmol, 9 µL of a 5 mM solution in Milli-Q water, 1.5 eq.) was added, and the mixture was stirred at rt for 10 min under Ar. **BR(Gly$_9$)**-Br (6 × 10^{-5} mmol, 18 µL of a 3.4 mM solution in deoxygenated phosphate buffer, 2 eq.) was added, and the solution was stirred again at rt for 1 h under Ar. The product was purified by RP–HPLC, 4 mL/min, gradient 10 to 50% B over 40 min. (A: H_2O 0.1% TFA, B: CH_3CN 0.1% TFA, retention time = 24.3 min), and identified as **BR(Gly$_9$)GAGA** by mass spectrometry.

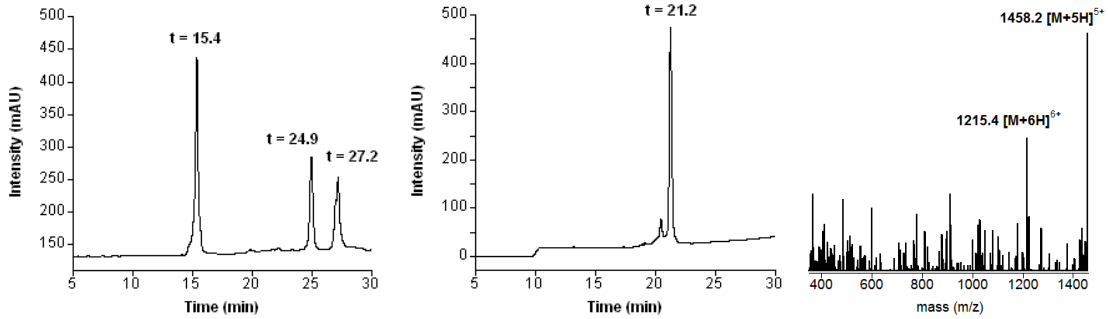


Fig. S6 HPLC chromatograms of the crude (left, gradient 10 to 50% B over 40 min) and the purified **BR(Gly$_9$)GAGA** (center, gradient 5 to 75% B over 30 min). Right) Mass spectrum of the purified peptide.

Left: t = 15.4 starting **BR(Gly$_9$)**-Br; t = 24.9 product **BR(Gly$_9$)GAGA**. **EM–ESI$^+$ (m/z):** Calcd. for $C_{307}H_{504}N_{108}O_{93}S_3$: 7287.72. Found: 1458.2 $[M+5H]^{5+}$; 1215.4 $[M+6H]^{6+}$; t = 27.2 starting **GAGA**-SH. Center: t = 21.2 product **BR(Gly$_9$)GAGA**. **EM–ESI$^+$ (m/z):** Calcd. for $C_{307}H_{504}N_{108}O_{93}S_3$: 7287.72. Found: 1458.2 $[M+5H]^{5+}$; 1215.4 $[M+6H]^{6+}$.

Synthesis of BR(Hk)-GAGA:

GAGA-SH (0.1 mg, 3 × 10^{-5} mmol) was dissolved in deoxygenated phosphate buffer (130 µL, 10 mM, pH = 7.5). A $ZnSO_4$ solution (4.5 × 10^{-5} mmol, 9 µL of a 5 mM solution in Milli-Q water, 1.5 eq.) was added, and the mixture was stirred at rt for 10 min under Ar. **BR(Hk)**-Br (6 × 10^{-5} mmol, 18 µL of a 3.4 mM solution in deoxygenated phosphate buffer, 2 eq.) was added, and the solution was stirred again at rt for 1 h under Ar. The product was purified by RP–HPLC, 4 mL/min, gradient 10 to 50% B over 40 min. (A: H_2O 0.1% TFA, B: CH_3CN 0.1% TFA, retention time = 24.3 min), and identified as product **BR(Hk)GAGA** by mass spectrometry.

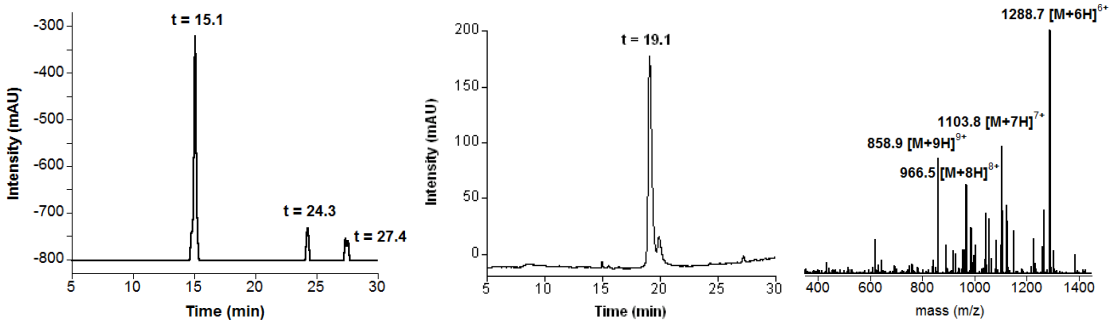


Fig S7. HPLC chromatograms of the crude (left, gradient 10 to 50% B over 40 min) and the purified **BR(Hk)GAGA** (center, gradient 5 to 75% B over 30 min). Right) Mass spectrum of the purified peptide.

Left: t = 15.1 starting **BR(Hk)**-Br; t = 24.3 product **BR(Hk)GAGA**. **EM–ESI$^+$ (m/z):** Calcd. for $C_{331}H_{553}N_{115}O_{93}S_3$: 7727.96. Found: 1288.7 $[M+6H]^{6+}$; 1103.8 $[M+7H]^{7+}$; 966.5 $[M+8H]^{8+}$; 858.9 $[M+9H]^{9+}$; t = 27.4 starting **GAGA**-SH. Center: t = 19.1 product **BR(Hk)GAGA**. **EM–ESI$^+$ (m/z):** Calcd. for $C_{331}H_{553}N_{115}O_{93}S_3$: 7727.96. Found: 1288.7 $[M+6H]^{6+}$; 1103.8 $[M+7H]^{7+}$; 966.5 $[M+8H]^{8+}$; 858.9 $[M+9H]^{9+}$.

Synthesis of the peptide TMR-**BR(Hk)GAGA**

GAGA-SH (0.1 mg, 3×10^{-5} mmol) was dissolved in deoxygenated phosphate buffer (130 µL, 10 mM, pH = 7.5). A $ZnSO_4$ solution (4.5×10^{-5} mmol, 9 µL of a 5 mM solution in Milli-Q water, 1.5 eq.) was added, and the mixture was stirred at rt for 10 min under Ar. TMR-**BR(Hk)**-Br (6×10^{-5} mmol, 18 μL of a 3.4 mM solution in deoxygenated phosphate buffer:CH_3CN 1:1, 2 eq.) was added, and the solution was stirred again at rt for 1 h under Ar. The product was purified by RP–HPLC, 4 mL/min, gradient 15 to 55% B over 40 min. (A: H_2O 0.1% TFA, B: CH_3CN 0.1% TFA, retention time = 14.7 min), and identified as TMR-**BR(Hk)GAGA** by mass spectrometry.

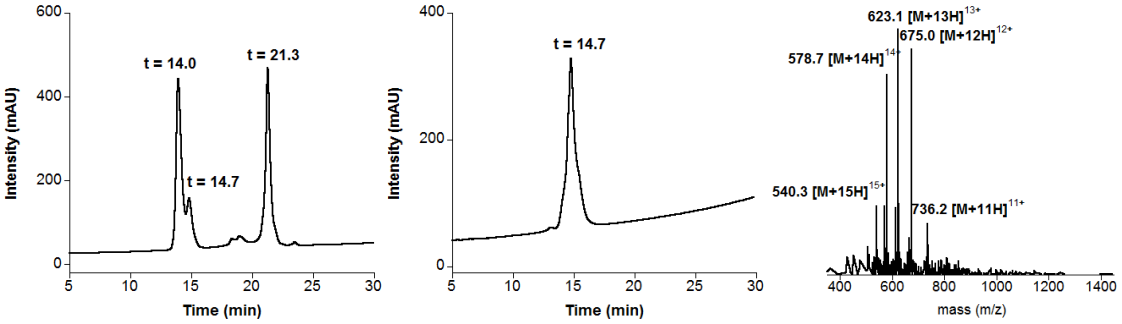
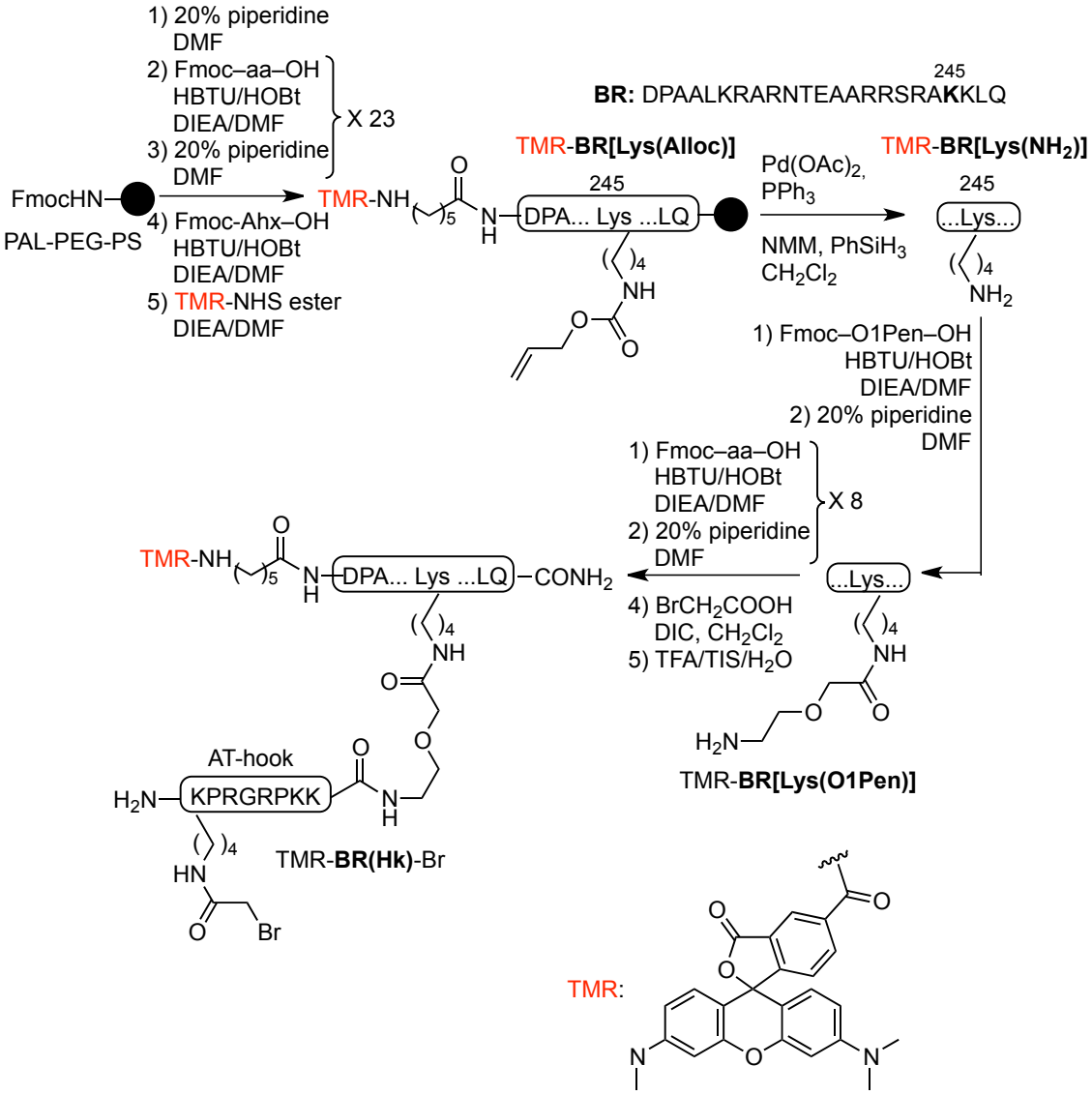


Fig S8. HPLC chromatograms of the crude (left, gradient 15 to 55% B over 40 min) and the purified TMR-**BR(Hk)GAGA** (center, gradient 15 to 55% B over 30 min). Right) Mass spectrum of the purified peptide.

Left: t = 14.0 starting TMR-**BR(Hk)**-Br; t = 14.7 product **BR(Hk)GAGA**. **EM–ESI$^+$ (m/z):** Calcd. for $C_{353}H_{579}N_{118}O_{95}S_3$: 8087.3. Found: 736.2 $[M+11H]^{11+}$; 675.0 $[M+12H]^{12+}$; 623.1 $[M+13H]^{13+}$; 578.7 $[M+14H]^{14+}$; 540.3 $[M+15H]^{15+}$; t = 21.3 starting **GAGA**-SH. Center: t = 14.7 product TMR-**BR(Hk)GAGA**. **EM–ESI$^+$ (m/z):** Calcd. for $C_{353}H_{579}N_{118}O_{95}S_3$: 8087.3. Found: 736.2 $[M+11H]^{11+}$; 675.0 $[M+12H]^{12+}$; 623.1 $[M+13H]^{13+}$; 578.7 $[M+14H]^{14+}$; 540.3 $[M+15H]^{15+}$.


1) 20% piperidine DMF
2) Fmoc–aa–OH HBTU/HOBt DIEA/DMF
3) 20% piperidine DMF
X 23
4) Fmoc-Ahx–OH HBTU/HOBt DIEA/DMF
5) TMR-NHS ester DIEA/DMF
FmocHN
PAL-PEG-PS
TMR-NH
DPA... Lys ...LQ
245
BR: DPAALKRARNTEAARRSRAKKLQ
TMR-BR[Lys(Alloc)]
Pd(OAc)2, PPh3
NMM, PhSiH3
CH2Cl2
TMR-BR[Lys(NH2)]
...Lys...
1) Fmoc–O1Pen–OH HBTU/HOBt DIEA/DMF
2) 20% piperidine DMF
1) Fmoc–aa–OH HBTU/HOBt DIEA/DMF
2) 20% piperidine DMF
X 8
4) BrCH2COOH DIC, CH2Cl2
5) TFA/TIS/H2O
CONH2
AT-hook
KPRGRPKK
TMR-BR(Hk)-Br
TMR-BR[Lys(O1Pen)]
TMR:

Miniprotein design

These models were built with MacPymol (v1.7.6.2 Enhanced for Mac OS X) using the PDB deposited structures (PDB IDs: 1YSA for GCN4, 3UXW for AT-hook and 1YUI for GAGA).

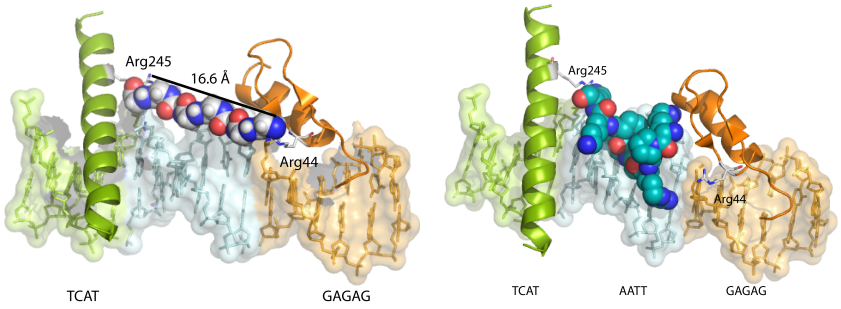


Fig. S9 Model of the simultaneous interaction of the GCN4 basic region and the GAGA motif located at adjacent major grooves: Left: with a Gly_9 linker without minor groove contacts; Right: with an AT-hook moiety inserted in the minor groove.

Oligonucleotide sequences

Double stranded (only one strand is shown) oligonucleotides used for EMSA experiments with conjugates **BR(Gly$_9$)GAGA** and **BR(Hk)GAGA** were supplied by *Thermo Fischer* and their sequences were:

A 5´–CGCG**TCAT***AATT*GAGAGCGC–3´
B 5´–CGCG**TCAT**CAGCGAGAGCGC–3´
C 5´–CGCG**TCAT**AAATTGAGAGCGC–3´
D 5´–CGCG**TCAT**AATTCGCGACGC–3´
E 5´–CGCGTGCTAATTGAGAGCGC–3´

EMSA experiments

EMSAs were performed with a *BioRad* Mini Protean gel system, powered by an electrophoresis power supplies PowerPac Basic model, maximum power 150 V, frequency 50-60 Hz at 140 V (constant V). Binding reactions were performed over 30 min in 18 mM Tris-HCl (pH 7.5), 90 mM KCl, 1.8 mM $MgCl_2$, 0.2 mM TCEP, 9% glycerol, 0.11 mg/mL BSA, 2.2% NP-40 and 0.02 mM of $ZnCl_2$. In the experiments we used 75 nM of the ds–DNAs and a total incubation volume of 20 µL. After incubation for 30 min products were resolved by PAGE using a 10% non-denaturing polyacrylamide gel and 0.5× TBE buffer for 40 min at 20 ºC, and analyzed by staining with SyBrGold (Molecular Probes: 5 µL in 50 mL of 1× TBE) for 10 min. and visualized by fluorescence. 5x TBE buffer (0.445M Tris, 0.445 M Boric acid).

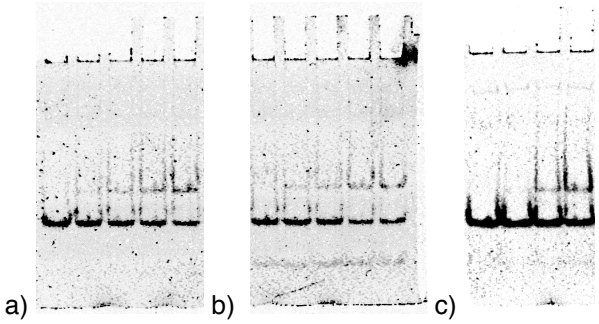


Fig S10. Complete gel pictures of the DNA binding studies of **BR(Gly$_9$)GAGA** by EMSA. (a) [**BR(Gly$_9$)GAGA**] = 0, 200, 400, 600, 800 nM with 75 nM of dsDNA ***A***. (b) [**BR(Gly$_9$)GAGA**] = 0, 200, 400, 600, 800 nM with 75 nM of dsDNA ***B***. (c) [**BR(Gly$_9$)GAGA**] = 0, 400, 600, 800 nM with with 75 nM of dsDNA ***C***.

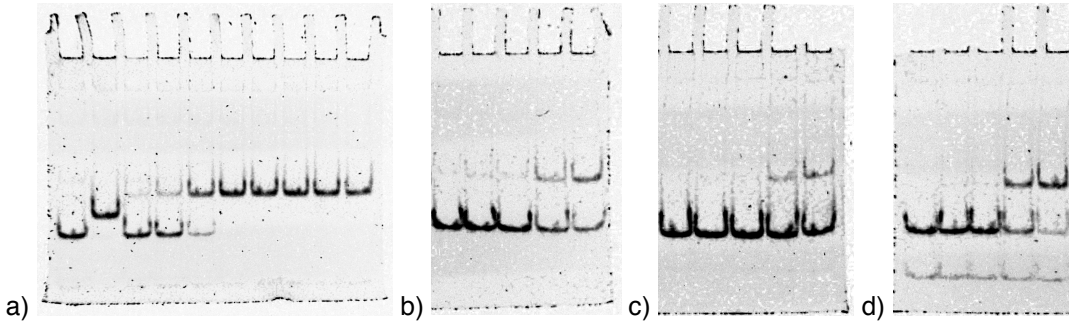


Fig S11. Complete gel pictures of the DNA binding studies of **BR(Hk)GAGA** by EMSA. (a) [**BR(Hk)GAGA**] = 0, control, 100, 200, 400, 600, 700, 800, 900, 1000 nM with 75 nM of dsDNA ***A***; (b) [**BR(Hk)GAGA**] = 0, 200, 400, 600, 800 nM with 75 nM of dsDNA ***D***; (c) [**BR(Hk)GAGA**] = 0, 200, 400, 600, 800 nM with75 nM of dsDNA ***E***; (d) [**BR(Hk)GAGA**] = 0, 200, 400, 600, 800 nM with 75 nM of dsDNA ***B***.

Fluorescence Anisotropy

Measurements were made with a *Jobin-Yvon Fluoromax-3*, (DataMax 2.20) coupled to a *Wavelength Electronics* LFI–3751 temperature controller, using the following settings: integration time: 2.0 s; excitation slit width: 5.0 nm; emission slit width: 20.0 nm; excitation wavelength 559 nm; emission wavelength 585 nm.

TMR-***A*** or TMR-***B*** (TMR-5'–GCG CTCTC*AATT***ATGA** CGCG–3' or TMR-5'–CGCG **TCAT***CAGC*GAGAG CGC–3', only one strand shown) (5µL, 5 µM), and calf thymus DNA (50 µM in base pairs) were added to 995 µL of Tris-HCl buffer 20 mM pH 7.5, 100 mM NaCl, 0.02 mM of $ZnCl_2$, and the anisotropy was measured. Aliquots of a stock solution in water of the peptide **BR(Hk)GAGA** (12.5 µM) were successively added to this solution, and the anisotropic value was recorded after each addition.

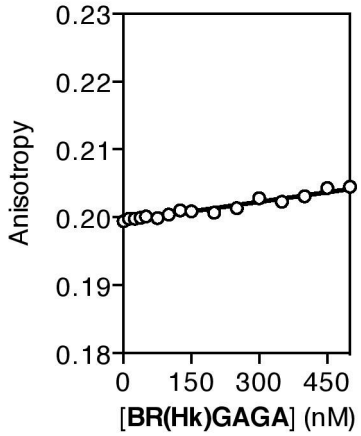


Fig S12. Fluorescence anisotropy titration of a 25 nM solution of non target TMR-***B*** in the presence of excess of competing calf thymus and with increasing concentrations of **BR(Hk)GAGA**.

CD measurements

Circular Dichroism experiments were made with a *Jasco-715* coupled with a thermostat *Nestlab* RTE-111. The settings used were: Acquisition range: 300-195nm; band width: 2.0 nm; resolution: 0.2 nm; accumulation: 5 scans; sensitivity 10 mdeg; response time: 0.25 s, speed: 100 nm/min. Measurements were made in a 2 mm cell at 20 ºC. Samples contained 10 mM phosphate buffer pH 7.5 and 100 mM of NaCl, 5 µM peptide and 5 µM of corresponding dsDNA (when present).

The mixtures were incubated for 5 min before registering. The CD spectra of the peptides (when measured in the presence of DNA) were calculated as the difference between the spectrum of the peptide/DNA mixture and the measured spectrum of a sample of the DNA oligonucleotide. The spectra are the average of 5 scans and were processed using the "smooth" macro implemented in the program *Kaleidagraph* (v 3.5 by Synergy Software).

Cellular internalization studies

Vero cells growing on glass coverslips were incubated in PBS containing 5 µM of TMR-**BR(Hk)GAGA** for 30 min. Then they were washed twice with PBS and observed *in vivo* in a fluorescence microscope equipped with adequate filters and differential interference contrast (DIC) microscopy. Digital pictures of the different samples were taken under identical conditions of gain and exposure.