Electronic Supplementary Material (ESI) for RSC Chemical Biology.

Molecular dynamic modelling of the interaction of a synthetic zinc-finger miniprotein with DNA

Jessica Rodriguez,*[a,b] Federica Battistini,[b,c] Soraya Learte-Aymamí,[a] Modesto Orozco*[b,c] and José L. Mascareñas*[a]

[a] *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.*
[b] *Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), Baldiri Reixac 10-12, 08028 Barcelona, Spain.*
[c] *Department of Biochemistry and Molecular Biology. University of Barcelona, 08028 Barcelona, Spain.*

Contents

1. General peptide synthesis procedures and characterization

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. All other chemicals were purchased from *Aldrich* or *Fluka*. All solvents were dry and synthesis grade, unless specifically noted.

Electrospray Ionization Mass Spectrometry (ESI/MS) was performed with an *Agilent 6120 Quadrupole LC/MS model* in positive scan mode.

gaga-Hk: H-**SQSEQPATCP ICYAVIRQSR NLRRHLELRH FGGGGKPRGR PKK**-NH_2

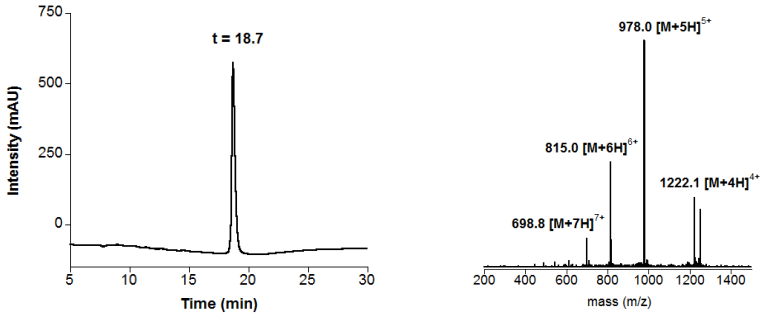

Figure S1. Left) HPLC chromatogram of purified peptide. Gradient 5 to 75% B over 30 min. Right) Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. for $C_{209}H_{347}N_{75}O_{57}S_2$: 4883.6. Found: 1222.1 $[M+4H]^{4+}$; 978.0 $[M+5H]^{5+}$; 815.0 $[M+6H]^{6+}$; 698.8 $[M+7H]^{7+}$.

gaga-Hk(N61R): H-**SQSEQPATCP ICYAVIRQSR <u>R</u>LRRHLELRH FGGGGKPRGR PKK**-NH_2

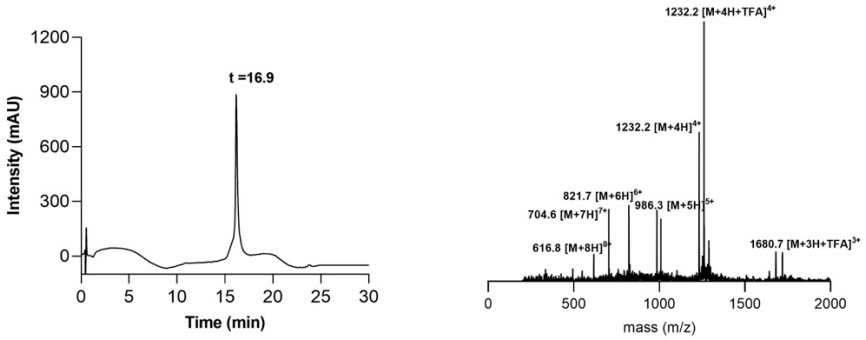

Figure S2. Left) HPLC chromatogram of purified peptide. Gradient 5 to 75% B over 30 min. Right) Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. for $C_{211}H_{354}N_{78}O_{55}S_2$: 4924.67 Found: 1680.7 $[M+3H+TFA]^{3+}$; 1232.2 $[M+4H]^{4+}$; 986.3 $[M+5H]^{5+}$; 821.7 $[M+6H]^{6+}$; 704.6 $[M+7H]^{7+}$; 616.8 $[M+8H]^{8+}$

2. Oligonucleotide sequences

Double stranded (only one strand is shown) oligonucleotides used for EMSA experiments with peptides **gaga-Hk** and **gaga-Hk(N61R)** were supplied by *Thermo Fischer* and their sequences were (putative binding site is underlined):

GAGAG	5´-CGCGTCAT*AATT*GAGAGCGC-3´
GA*A*AG	5´-CGCGTCAT*AATT*GAAAGCGC-3´
***C*AGAG**	5´-CGCGTCAT*AATT*CAGAGCGC-3´
G*C*GAG	5´-CGCGTCAT*AATT*GCGAGCGC-3´
G*G*GAG	5´-CGCGTCAT*AATT*GGGAGCGC-3´

3. Molecular dynamics simulations

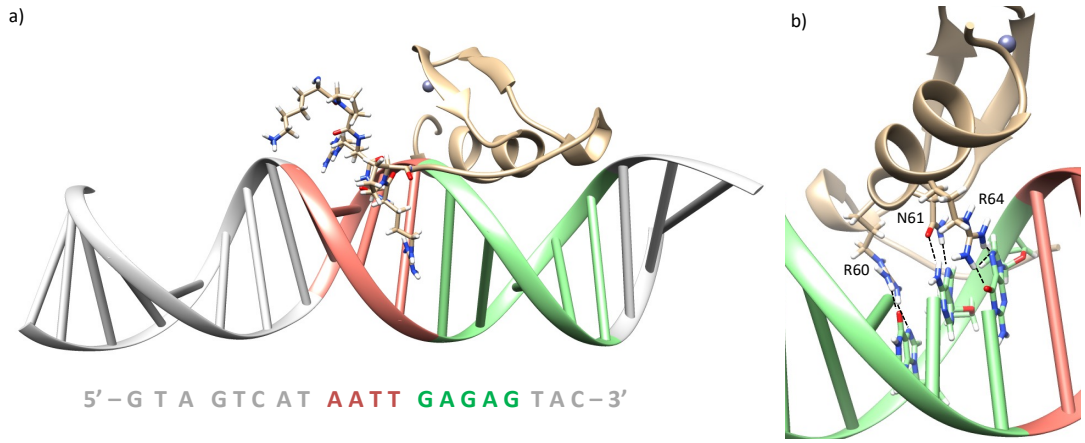

Figure S3. a) Snapshot from the last frame of the 500 ns MD simulation of miniprotein **Hk-gaga** bound to the target DNA sequence: AATT-GAGAG. b) Key hydrogen bonding interactions in the major groove.

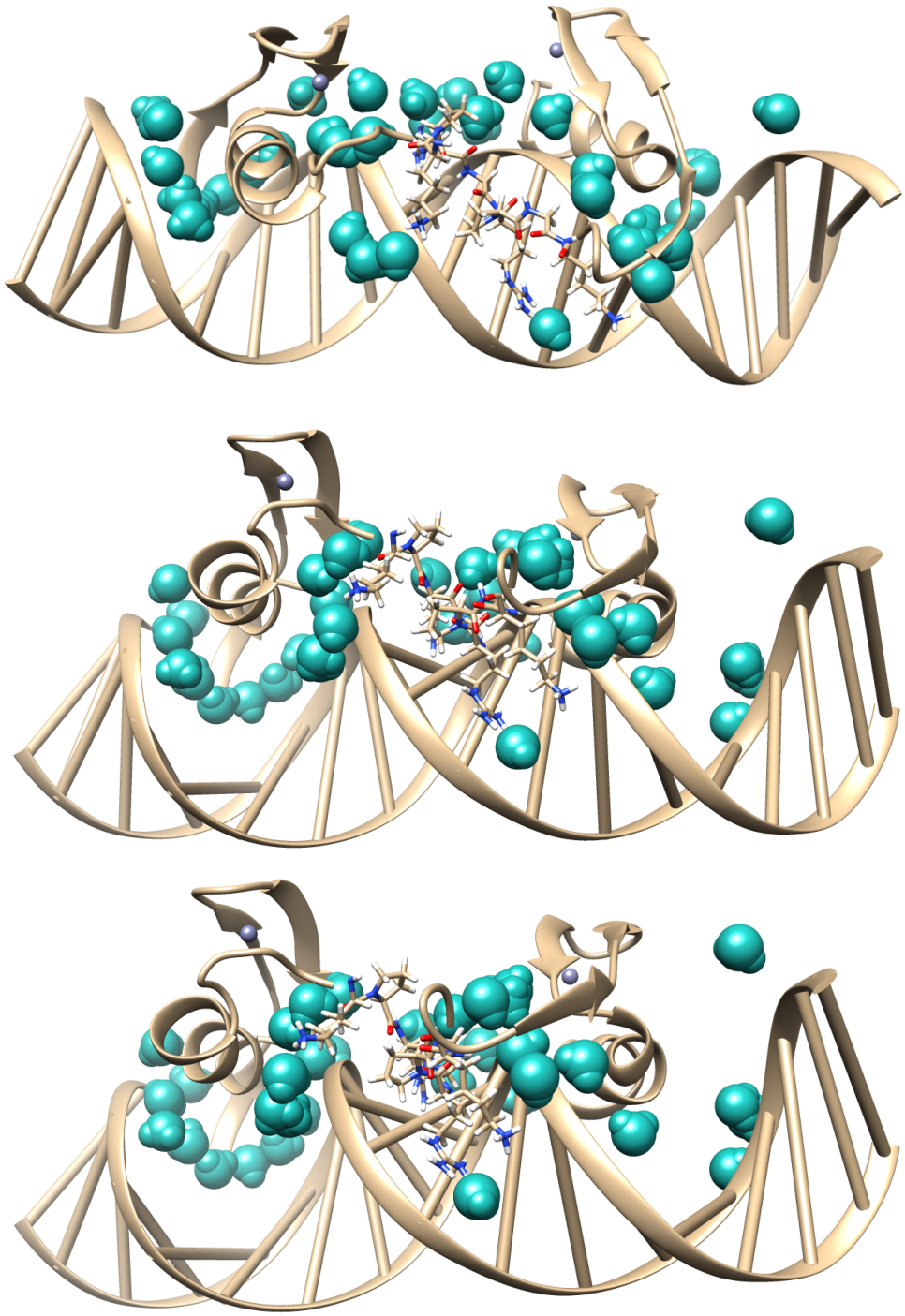

Figure S4. Snapshot from the final frame of 500 ns MD simulation of miniprotein **gaga-Hk-gaga** bound to the target DNA sequence: CTCTC-AATT-GAGAG showing water molecules (drawn as green spheres) at the protein-DNA interface from three different viewpoints.

RMSD analysis

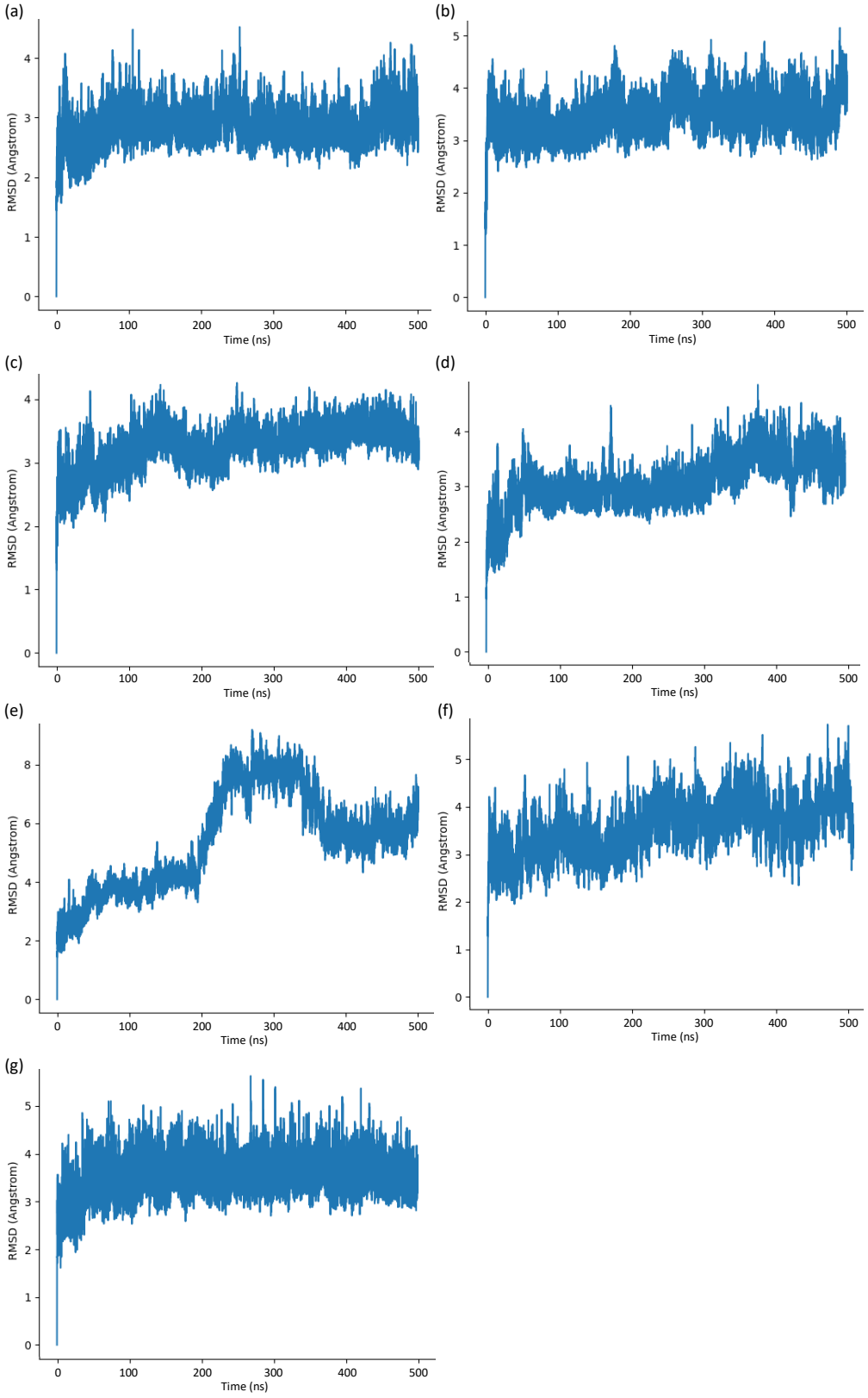

Figure S5. Time-dependent RMSD of all system atoms for the MD simulation of **gaga-Hk-gaga** with a DNA containing the (a) **GAGAG**, (b) ***C*AGAG**, (c) **G*C*GAG**, (d) **G*G*GAG** and (e) **GA*A*AG** sequences; (f) for the MD simulation of **gaga-Hk-gaga(N61R)** with a DNA containing the **G*G*GAG** sequence; and (g) for the MD simulation of **Hk-gaga** with a DNA containing the target sequence AATT-GAGAG. The RMSD were computed taking as reference the corresponding initial structures after equilibration.

Hydrogen bond analysis

Note that Arg106, Asn107 and Arg110 are equivalent to Arg60, Asn61 and Arg64 but on the other ZF.

Major grooves interactions					
Acceptor	**DonorH**	**Donor**	**Frames (%)**	**AvgDist**	**AvgAng**
DG_15@O6	ARG_60@HH12	ARG_60@NH1	94	2.7707	160.4881
DG_33@O6	ARG_110@HH12	ARG_110@NH1	84	2.8162	159.0033
DG_35@O6	ARG_106@HH12	ARG_106@NH1	70	2.7729	159.9845
DG_33@N7	ARG_110@HH22	ARG_110@NH2	64	2.8820	157.1983
DA_34@N7	ASN_107@HD21	ASN_107@ND2	56	2.9036	163.0831
ASN_107@OD1	DA_34@H62	DA_34@N6	47	2.8702	161.3636
DG_13@O6	ARG_64@HH12	ARG_64@NH1	42	2.8228	155.4851
ASN_61@OD1	DA_14@H62	DA_14@N6	42	2.8698	160.9038
DA_14@N7	ASN_61@HD21	ASN_61@ND2	39	2.9115	162.4801
DG_13@N7	ARG_64@HH22	ARG_64@NH2	38	2.8871	160.0813
DG_15@N7	ARG_60@HH22	ARG_60@NH2	30	2.9099	154.7495
DG_35@N7	ARG_106@HH22	ARG_106@NH2	19	2.9112	153.8144
DG_15@O6	ARG_60@HH22	ARG_60@NH2	6	2.8856	145.2355
Minor groove interactions					
Acceptor	**DonorH**	**Donor**	**Frames (%)**	**AvgDist**	**AvgAng**
DT_12@O2	ARG_78@HE	ARG_78@NE	52	2.8325	157.4144
DT_11@O2	GLY_79@H	GLY_79@N	27	2.8814	158.8433
DA_9@N3	ARG_80@HH11	ARG_80@NH1	17	2.9062	153.6572
DT_12@O2	ARG_78@HH21	ARG_78@NH2	8	2.8477	149.7130

Table S1. Selected hydrogen bonds from the MD simulation of **gaga-Hk-gaga** with the target DNA sequence CTCTC-AATT-GAGAG. Hydrogen bonds corresponding to the interaction of the AT-hook moiety with the DNA are highlighted in green. Hydrogen bonds corresponding to the interaction of the first ZF unit with the DNA are highlighted in blue.

Acceptor	**DonorH**	**Donor**	**Frames (%)**	**AvgDist**	**AvgAng**
DG_35@O6	ARG_106@HH12	ARG_106@NH1	90	2.7694	158.7358
DG_33@O6	ARG_110@HH12	ARG_110@NH1	88	2.8129	159.6743
DG_15@O6	ARG_60@HH12	ARG_60@NH1	81	2.7880	156.3465
DA_34@N7	ASN_107@HD21	ASN_107@ND2	65	2.8935	160.9129
DG_33@N7	ARG_110@HH22	ARG_110@NH2	64	2.8910	156.1910
DG_15@N7	ARG_60@HH22	ARG_60@NH2	50	2.8998	158.3646
ASN_107@OD1	DA_34@H62	DA_34@N6	35	2.8723	160.2125
DG_35@N7	ARG_106@HH22	ARG_106@NH2	30	2.9088	154.1841

Table S2. Selected hydrogen bonds on the DNA major grooves from the MD simulation of **gaga-Hk-gaga** with a DNA containing the mutated DNA ***C*AGAG**. Hydrogen bonds corresponding to the interaction of the first ZF unit with the DNA are highlighted in blue.

Acceptor	DonorH	Donor	Frames (%)	AvgDist	AvgAng
DG_33@O6	ARG_110@HH12	ARG_110@NH1	91	2.8004	159.5689
DG_35@O6	ARG_106@HH12	ARG_106@NH1	85	2.7770	159.0432
ASN_107@OD1	DA_34@H62	DA_34@N6	55	2.8671	161.6743
DA_34@N7	ASN_107@HD21	ASN_107@ND2	53	2.8955	161.8855
DG_33@N7	ARG_110@HH22	ARG_110@NH2	45	2.9065	153.2353
DG_15@O6	ARG_60@HH11	ARG_60@NH1	27	2.8344	152.5052
DG_13@O6	ARG_64@HH12	ARG_64@NH1	20	2.8346	156.1969
ASN_61@OD1	DC_14@H42	DC_14@N4	19	2.8419	161.0667
DG_15@O6	ARG_60@HH12	ARG_60@NH1	19	2.7847	156.2874
DG_35@N7	ARG_106@HH22	ARG_106@NH2	16	2.9145	153.9513
DG_15@N7	ARG_60@HH22	ARG_60@NH2	8	2.9080	154.0211
DG_13@N7	ARG_64@HH22	ARG_64@NH2	8	2.8981	154.0466

Table S3. Selected hydrogen bonds on the DNA major grooves from the MD simulation of **gaga-Hk-gaga** with a DNA containing the mutated DNA **G*C*GAG**. Hydrogen bonds corresponding to the interaction of the first ZF unit with the DNA are highlighted in blue.

Acceptor	DonorH	Donor	Frames (%)	AvgDist	AvgAng
DG_35@O6	ARG_106@HH12	ARG_106@NH1	94	2.7742	159.7599
DG_33@O6	ARG_110@HH12	ARG_110@NH1	88	2.8179	160.5142
DG_15@O6	ARG_60@HH12	ARG_60@NH1	64	2.7834	157.7788
ASN_107@OD1	DA_34@H62	DA_34@N6	63	2.8641	162.8155
DG_33@N7	ARG_110@HH22	ARG_110@NH2	48	2.8985	153.4175
DA_34@N7	ASN_107@HD21	ASN_107@ND2	35	2.9159	163.9048
DG_35@N7	ARG_106@HH22	ARG_106@NH2	27	2.9111	153.7900
DG_13@O6	ARG_64@HH12	ARG_64@NH1	21	2.8436	157.0673
DG_13@N7	ARG_64@HH22	ARG_64@NH2	21	2.8934	157.4287
DG_14@N7	ASN_61@HD21	ASN_61@ND2	17	2.9120	161.2525
DG_15@O6	ARG_60@HH22	ARG_60@NH2	14	2.8755	146.8598
DG_15@N7	ARG_60@HH21	ARG_60@NH2	9	2.8773	151.8553

Table S4. Selected hydrogen bonds on the DNA major grooves from the MD simulation of **gaga-Hk-gaga** with a DNA containing the mutated DNA **G*G*GAG**. Hydrogen bonds corresponding to the interaction of the first ZF unit with the DNA are highlighted in blue.

Acceptor	DonorH	Donor	Frames (%)	AvgDist	AvgAng
DG_35@O6	ARG_106@HH12	ARG_106@NH1	88	2.7729	159.6006
DG_33@O6	ARG_110@HH12	ARG_110@NH1	84	2.8163	159.3698
DG_33@N7	ARG_110@HH22	ARG_110@NH2	71	2.8820	157.4429
DA_34@N7	ASN_107@HD21	ASN_107@ND2	62	2.8994	163.4050
ASN_107@OD1	DA_34@H62	DA_34@N6	43	2.8747	160.8400
DG_35@N7	ARG_106@HH22	ARG_106@NH2	23	2.9103	152.3228

Table S5. Selected hydrogen bonds on the DNA major grooves from the MD simulation of **gaga-Hk-gaga** with a DNA containing the mutated DNA **GA*A*AG**. Hydrogen bonds corresponding to the interaction of the first ZF unit with the DNA are highlighted in blue.

Acceptor	DonorH	Donor	Frames (%)	AvgDist	AvgAng
DG_33@O6	ARG_110@HH12	ARG_110@NH1	90	2.8112	160.1762
DG_35@O6	ARG_106@HH12	ARG_106@NH1	87	2.7729	158.6988
DG_15@O6	ARG_60@HH12	ARG_60@NH1	85	2.7865	157.6290
DG_33@N7	ARG_110@HH22	ARG_110@NH2	54	2.8958	154.5046
DA_34@N7	ASN_107@HD21	ASN_107@ND2	52	2.8960	162.8308
DG_14@O6	ARG_61@HH22	ARG_61@NH2	49	2.8152	155.7491
ASN_107@OD1	DA_34@H62	DA_34@N6	47	2.8738	161.9905
DG_15@N7	ARG_60@HH22	ARG_60@NH2	46	2.8909	154.2609
DG_14@O6	ARG_61@HH12	ARG_61@NH1	37	2.8371	156.9705
DG_35@N7	ARG_106@HH22	ARG_106@NH2	25	2.9128	153.4190
DG_14@N7	ARG_61@HH22	ARG_61@NH1	22	2.8860	153.0125
DG_15@O6	ARG_60@HH22	ARG_60@NH2	6	2.8828	147.1970

Table S6. Selected hydrogen bonds on the DNA major grooves from the MD simulation of **gaga-Hk-gaga(N61R)** with a DNA containing the mutated DNA **G*G*GAG**. Hydrogen bonds corresponding to the interaction of the first ZF unit with the DNA are highlighted in blue.

Acceptor	DonorH	Donor	Frames (%)	AvgDist	AvgAng
DG_35@O6	ARG_71@HH12	ARG_71@NH1	93	2.7768	157.8610
DG_33@O6	ARG_75@HH12	ARG_75@NH1	76	2.8308	158.4658
DG_33@N7	ARG_75@HH22	ARG_75@NH2	70	2.8866	158.3265
DA_34@N7	ASN_72@HD21	ASN_72@ND2	62	2.8946	158.7258
DG_35@N7	ARG_71@HH22	ARG_71@NH2	38	2.9074	156.0187
ASN_72@OD1	DA_34@H62	DA_34@N6	29	2.8752	159.6233

Table S7. Selected hydrogen bonds on the DNA major grooves from the MD simulation of **Hk-gaga** with the target DNA sequence AATT-GAGAG. Hydrogen bonds corresponding to the interaction of the first ZF unit with the DNA are highlighted in blue. Arg71 corresponds to Arg60 in the manuscript; Asn72 corresponds to Asn61 in the manuscript; Arg75 corresponds to Arg64 in the manuscript.

Module motions

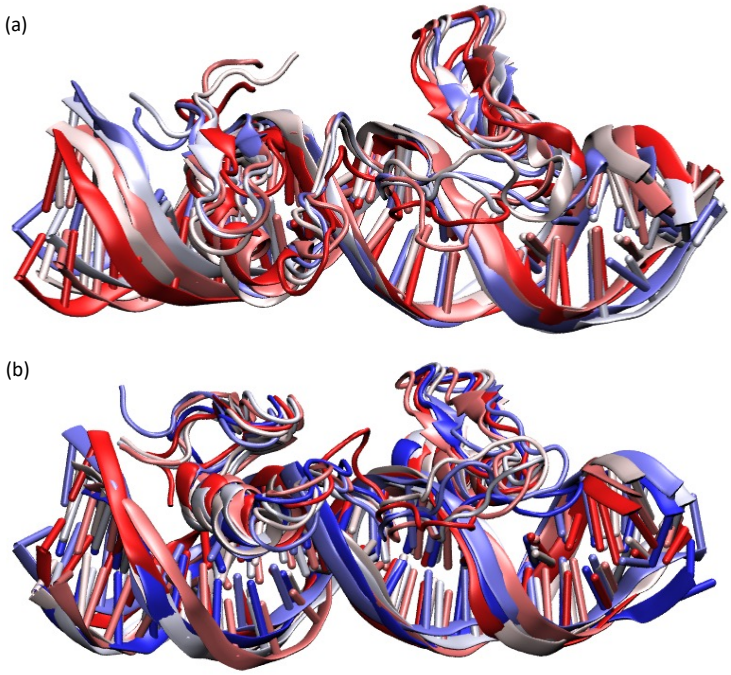

Figure S6. Lateral views of the superimposition of snapshots from the MD simulation of **gaga-Hk-gaga** with a DNA containing the a) **G<u>C</u>GAG** and b) **G<u>G</u>GAG** sequences, going from 0 to 500 ns every 100 ns of simulation (color scale from red to blue, being white in the middle of the simulation)

Base pair parameters for DNA

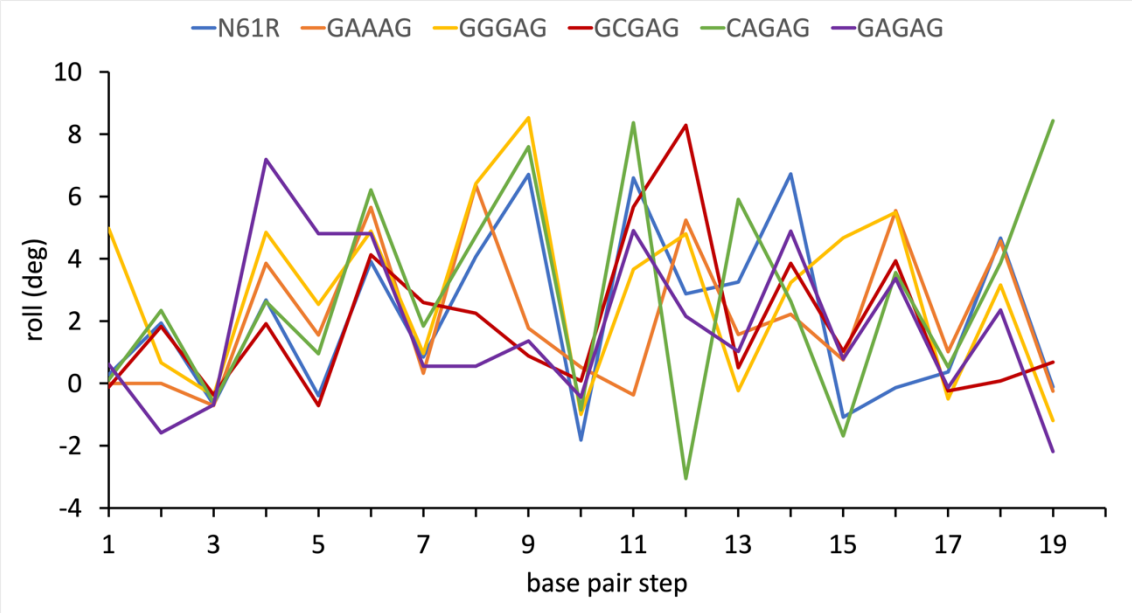

Figure S7. Comparison of base pair parameters (roll) averaged over the last 250 ns of the MD simulations for consensus and mutated DNA sequences.