SUPPORTING INFORMATION

Nucleolin discriminates drastically between long-loop and short-loop Quadruplexes.

Abhijit Saha,[1,2] Patricia Duchambon,[1,2] Vanessa Masson,[3] Damarys Loew,[3] Sophie Bombard[1,2], Marie-Paule Teulade-Fichou[1,2*]

[1]CNRS UMR9187, INSERM U1196, Institut Curie, PSL Research University, 91405 Orsay, France

[2]CNRS UMR9187, INSERM U1196, Université Paris Sud, Université Paris Saclay, 91405 Orsay, France.

[3]Institut Curie, PSL Research University, Centre de Recherche, Laboratoire de Spectrométrie de Masse Protéomique, 26 rue d'Ulm, Paris 75248 Cedex 05, France

***Corresponding Author:** Dr. Marie-Paule Teulade-Fichou, email: **mp.teulade-fichou@curie.fr**

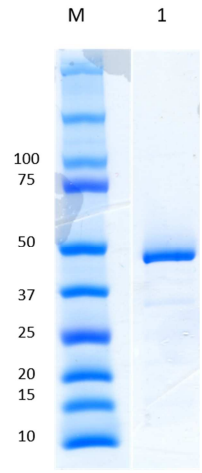


Figure S1. SDS PAGE (4-20%) for purity check of the recombinant NCL used in this study (size 46.1 KDa). Lane M: marker, lane 1: NCL alone.

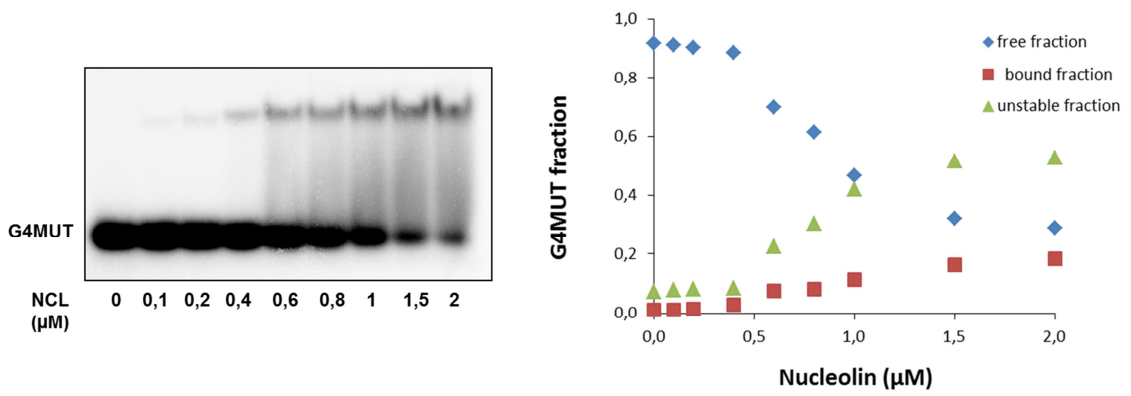
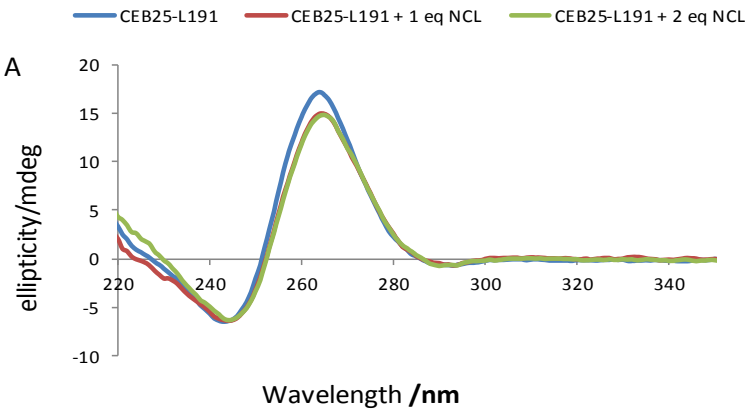
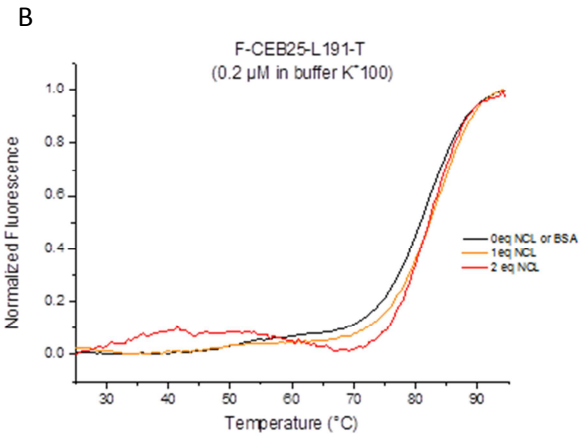
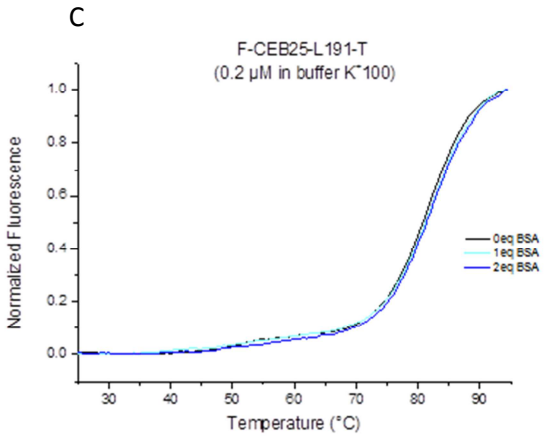
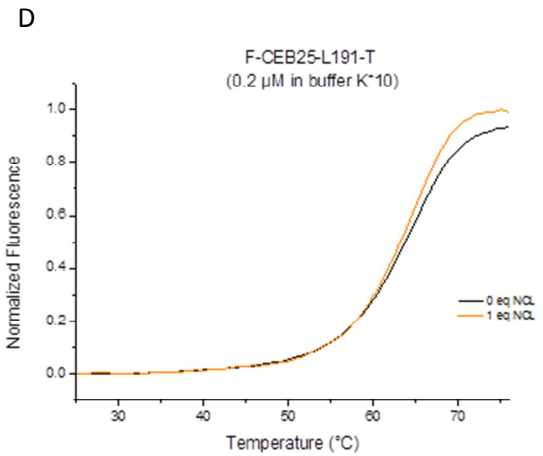


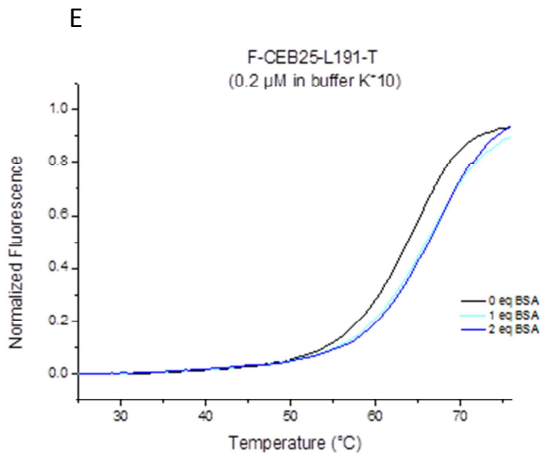
Figure S2. Left) EMSA for monitoring binding of NCL to the G4Mut sequence. NCL concentration ranged from 100 to 2000 nM, corresponding to 0.1, 0.2, 0.5, 0.7, 0.9, 1.1, 1.7, 2.3 µg of recombinant protein in a 20 µL reaction. Right) Quantification of bands: Unstable fraction corresponds to the smearing band.











F

Tm (°C)	K100	K10
0eq	81.1	64.2
1eqNCL	81.5	63.4
2eqNCL	82.6	ND
1eqBSA	81.7	65.6
2eqBSA	81.4	66.0

Figure S3 : A) CD spectra of CEB25-L191 (3µM) alone (blue line, min 245 and max 264 nm indicate parallel topology see also Fig S5) and after incubation with NCL (1 and 2 eq red and green lines) in buffer containing 10mM Tris (pH7.5) 100mM KCl and 1mM EDTA in 1% glycerol. Spectra have been substracted as appropriate for NCL and buffer signals. B, C, D and E) FRET melting curves of Fam/Tamra labelled CEB25 L191 (F-CEB25-L191-T) (0.2 µM) alone and in incubation with (1 and 2 eq) NCL or BSA in buffer containing B and C)10mM Tris (pH7.5) 100mM KCl and 1mM EDTA in 1% glycerol (K100) or D and E) 10mM Tris (pH7.5) 10mM KCl and 1mM EDTA in 1% glycerol (K10). F) Tm values. ND: not determined due to low fluorescence signal.

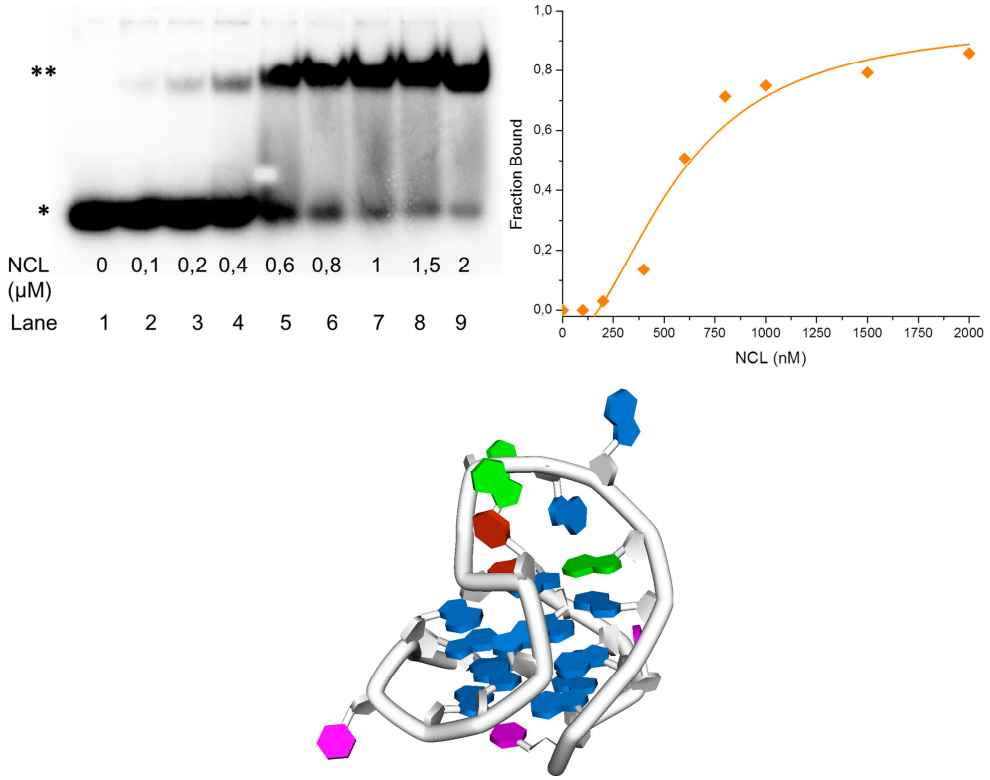


Figure S4. EMSA for monitoring NCL binding to the BCL2mid sequence (NMR structure shown below, PDB 2F8U, G blue, A green, T orange, C violet). NCL concentration ranged from 100 to 2000 nM, corresponding to 0.1, 0.2, 0.5, 0.7, 0.9, 1.1, 1.7, 2.3 µg of recombinant protein in a 20 µL reaction. 1:1 fitting of the curve gives K_D= 558 nM.

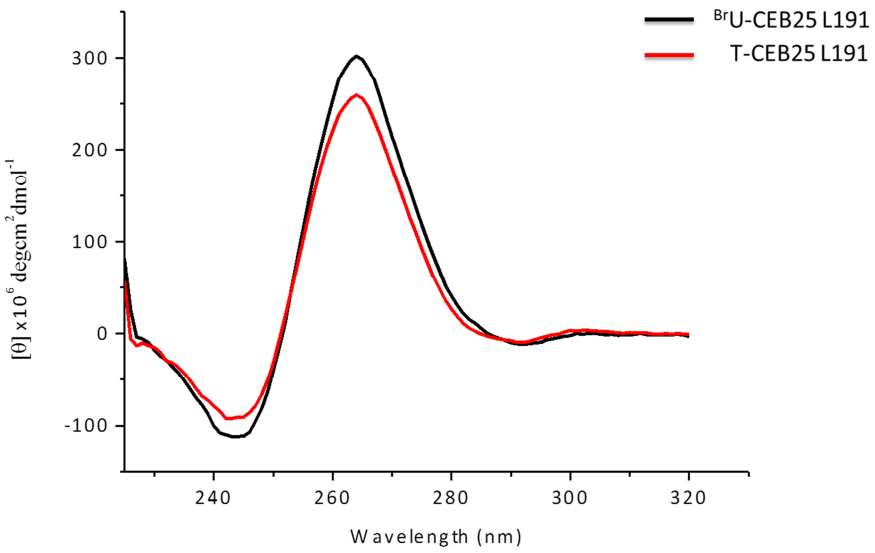


Figure S5: CD spectra of WT-CEB25-L191 and of its analogue sequence modified in position 16 by BrU (oligonucleotide 16BrU, Fig 4C). Maxima at 264 nm and minima at 245 nm are typical of parallel topology. For other sequences see Saha, A., Bombard, S., Granzhan, A., Teulade-Fichou, M. P. (2018) Probing of G-Quadruplex Structures via Ligand-Sensitized Photochemical Reactions in BrU-Substituted DNA. *Scientific Reports.* 8, 1-14.

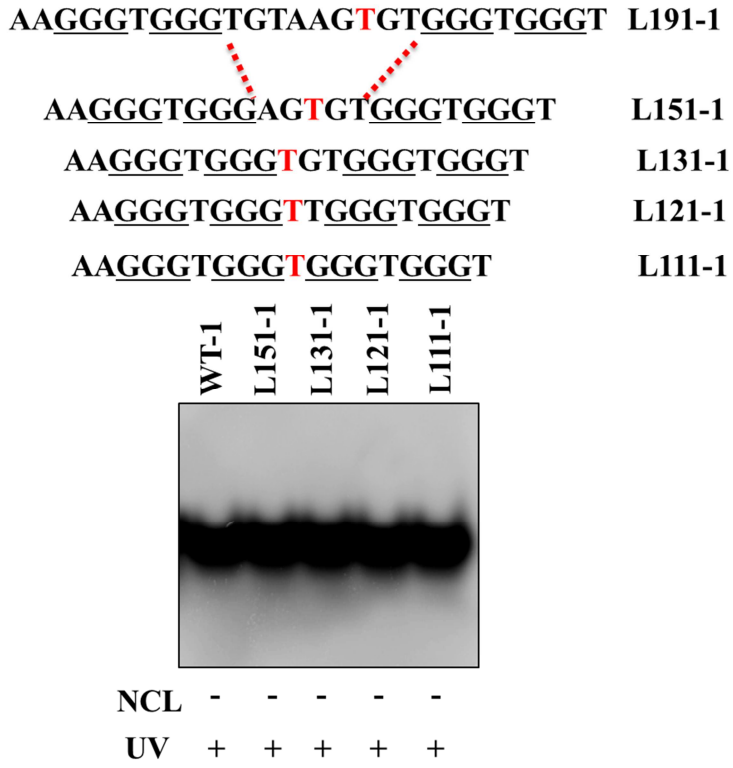


Figure S6 SDS PAGE analysis of the control experiment of photo-crosslinking reaction from Fig 3 using single BrU substituted CEB25 sequence variants named CEB25-L1X1-1, X indicates central loop length X= 9,5,3,2,1, -1 stands for 1 BrU residue. [DNA] = 20 nM, [NCL] = 1 µM, UV irradiation time 10 min, Exc: 300 nm

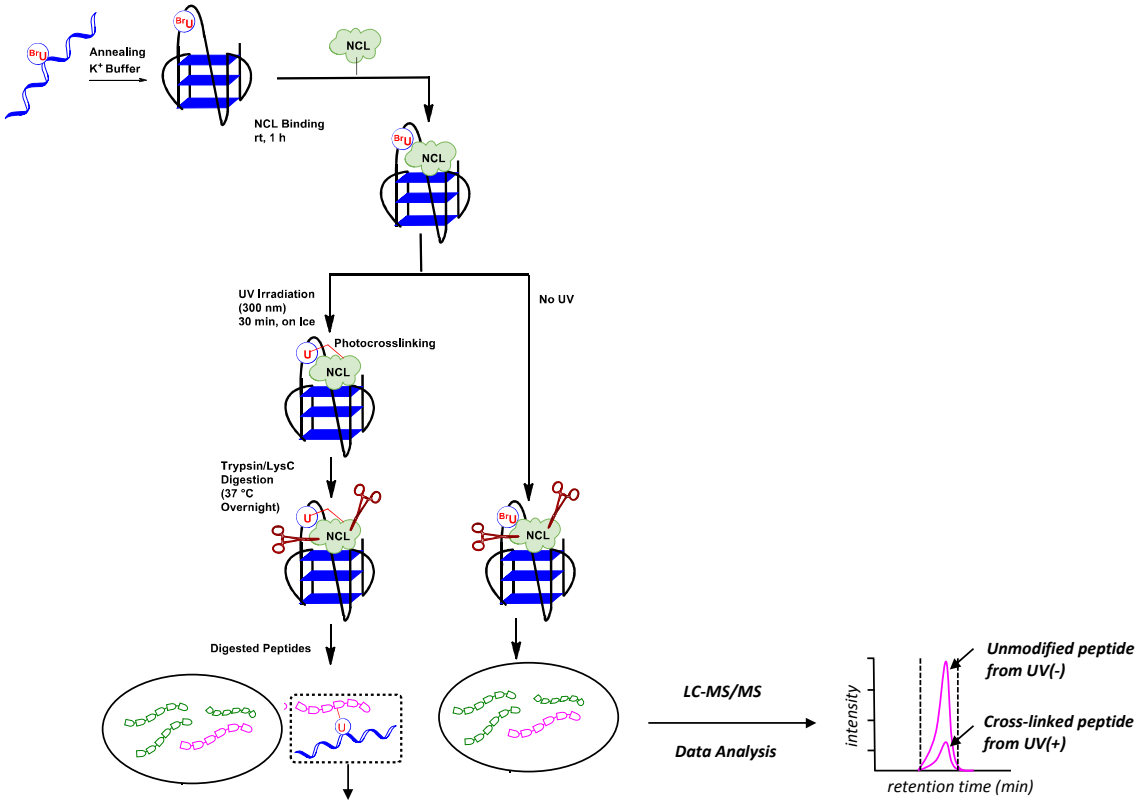


Figure S7. Schematic representation of the LC-MS/MS analysis workflow for the identification of NCL peptides interacting with G4 DNA. The UV-irradiated sample, (i.e. covalent binding of NCL with DNA) and non UV-irradiated sample (i.e. non-covalent complex of NCL with DNA) were enzymatically digested with trypsin/lysC and the resulting peptides were characterized by LC-MS/MS. Identification of putatively bound peptides will be indicated by the ratio of peak area of each peptide in the UV+ and UV- samples.

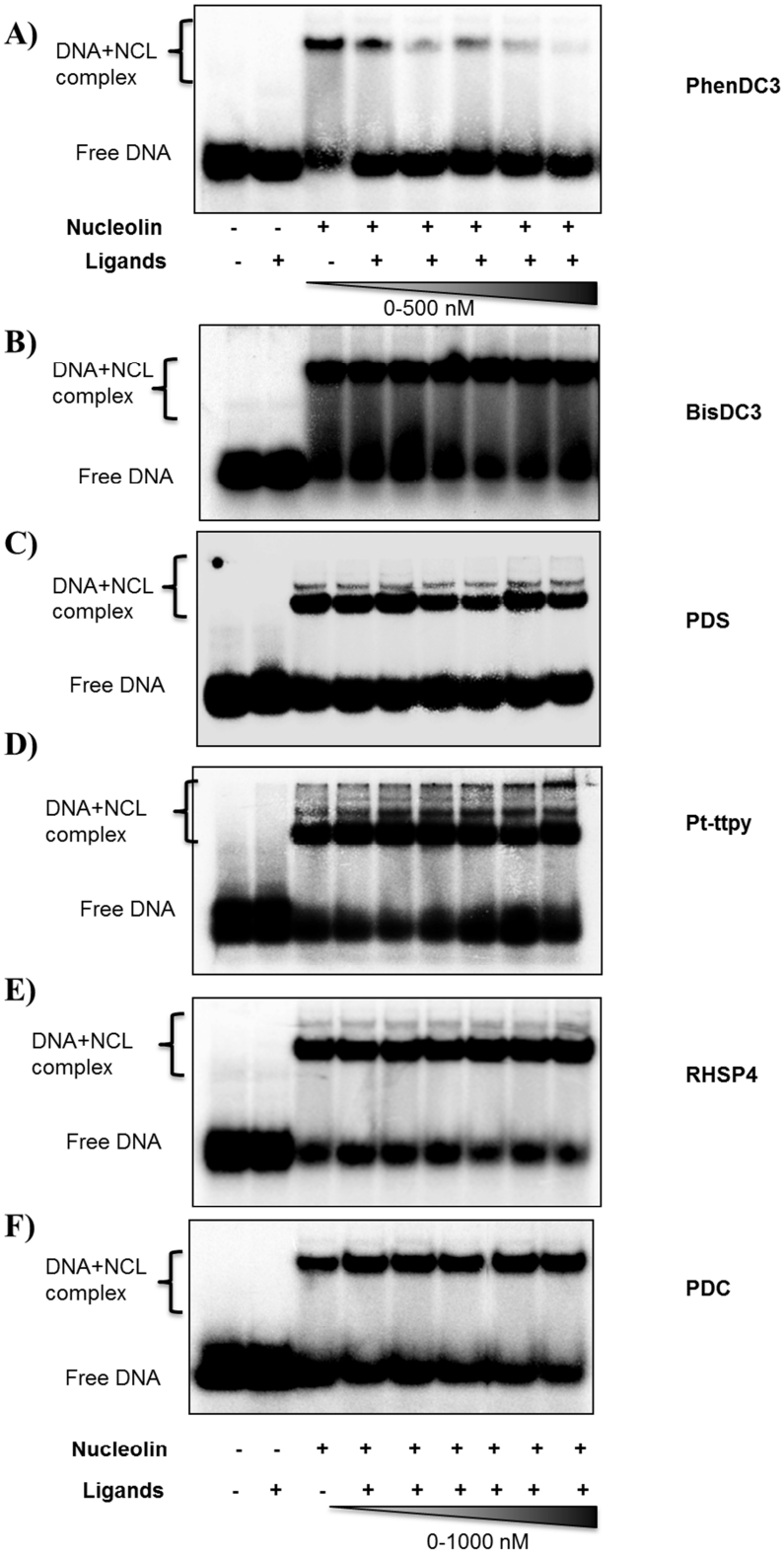
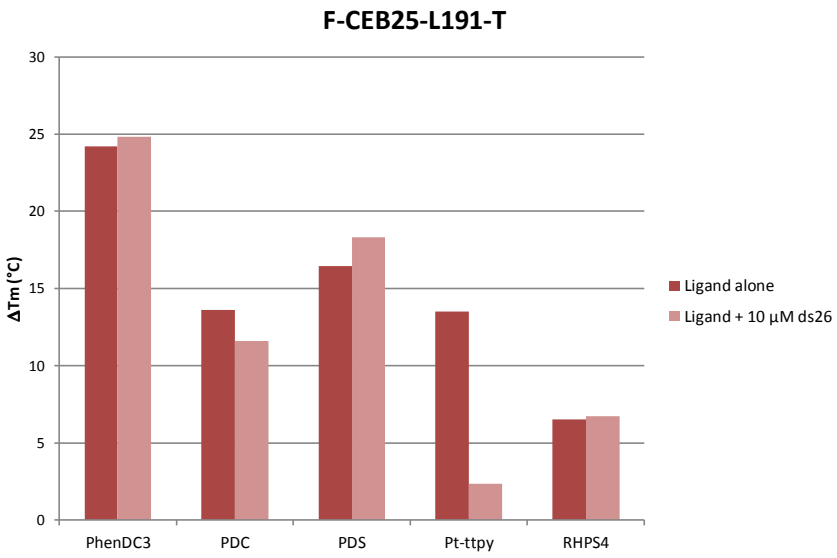


Figure S8. Displacement of NCL from CEB25-L191 followed by EMSA gel (10%) in the presence of a) PhenDC3 b) BisDC3 c) PDS d) Pt-ttpy e) RHSP4 f) 360A-PDC. [DNA]= 20 nM, [NCL] = 250 nM, [Ligand] = 0-500 nM (a); 0-1000 nM (b-f)



Ligand	$\Delta T_{1/2}$ (°C) (ligand alone)	$\Delta T_{1/2}$ (°C) (+ ds 26)
PhenDC3	24.6	24.8
PDC	13.6	11.6
PDS	16.4	18.3
Pt-ttpy	13.5	2.3
RHPS4*	6.6	6.7

Figure S9: Stabilisation of G4-CEB25-191 by ligands measured by FRET-melting ($\Delta T_{1/2}$ / °C) using doubly labelled F-CEB25-191-T. [oligonucleotide] = 0.2µM, [ligand]= 1µM, without (red) and with (pink) duplex competitor [ds26] = 10µM. All ligands are selective ($\Delta T_{1/2}$ is poorly affected by the duplex competitor) at the exception of Pt-ttpy. *The values determined for RHPS4 are given as indication but could be biased as the fluorescence of this ligand interferes with the FRET signal.

Table S1: Thermal stability of the CEB25 variants measured by UV-melting in 1mM K^+ buffer (Li Cacodylate l0 mM, LiCl 99 mM, KCl 1 mM). Thermal stability is expected to be even higher in our conditions (100mM K^+). Data taken from ref 9 (EMBO J., 2015, 34 1718) and Nucleic Acid Research 2006,34, 5133 (BCL2-mid).

.

G4 name	**T_m °C**
CEB25-L191	55.1
CEB25-L151	59.7
CEB25-L131	61.9
CEB25-L121	67.9
CEB25-L111	73.4
BCL2-mid	66.0

Table S2. List of peptides identified following the workflow of Fig S5 with corresponding UV+/UV- ratios of extracted areas and p values. The 15- aa-peptide on the top (line 1) is the only one to be significantly depleted (ratio <0.5, p values < 0.05). Of note the resulting tryptic peptides (8-aa fragment GIAYIEFK, line 18) and (SKGIAYIEFK, line 7, miscleaved) show no significant variation in ratio. However, the second TEADAEK 7-aa fragment has never been detected in the control and in the UV irradiated conditions indicating that this part might be tightly protected by the protein and could thus be the preferred crosslinking site.

Peptide	Precursor Mz	RT	Begin Pos	End Pos	+UV/-UV	P-value
GIAYIEFKTEADAEK	**563.285798**	**42.32**	**251**	**265**	**0.45492881**	**0.008188063**
EAMEDGEIDGNKVTLDWAKPK	588.287508	48.17	449	469	0.474168303	0.06090465
GFGFVDFNSEEDAKAAK	612.288091	51.33	432	448	1.688455459	0.114163818
ALELTGLKVFGNEIK	545.650412	69.01	184	198	0.69581864	0.16409645
VTLDWAKPK	530.305666	26.28	461	469	0.686531433	0.165291277
GFGFVDFNSEEDAK	782.343903	66.42	432	445	0.773293039	0.201777732
SKGIAYIEFK	579.324056	31.77	249	258	1.581714147	0.234953752
ETGSSKGFGFVDFNSEEDAK	718.655279	49.7	426	445	1.621130712	0.391724244
SISLYYTGEKGQNQDYR	1012.481793	30.34	279	295	0.780469869	0.522825416
SAPELKTGISDVFAK	522.619001	46.14	140	154	1.135953077	0.547015549
ALELTGLK	423.76056	36.9	184	191	1.116380638	0.584138579
GYAFIEFASFEDAK	798.874839	86.88	345	358	1.13774566	0.598987053
IVTDRETGSSKGFGFVDFNSEEDAK	913.431349	46.68	421	445	0.867698327	0.601222436
KFGYVDFESAEDLEK	593.949075	56.36	169	183	1.176903508	0.636493219
AIRLELQGPR	577.846021	28.09	373	382	0.845054772	0.63995955
VTLDWAKPKGEGGFGGR	445.48633	32.87	461	477	1.288923708	0.663272678
TGISDVFAKNDLAVVDVR	641.34594	72.96	146	163	0.717791443	0.699798221
GIAYIEFK	471.76056	48.58	251	258	0.93572101	0.721750201
FGYVDFESAEDLEK	825.872493	71.86	170	183	0.934034301	0.737398345
VEGTEPTTAFNLFVGNLNFNKSAPELK	736.130273	93.6	119	145	1.108359159	0.744128083
VTQDELKEVFEDAAEIR	997.499651	78.47	225	241	1.090472103	0.789647877
VSDGSSEIFFK	609.298235	44.4	35	45	1.025025084	0.876253049
NLPYKVTQDELKEVFEDAAEIR	653.587975	85.02	220	241	1.051834361	0.877705999
EVFEDAAEIR	590.788035	39.28	232	241	1.022458879	0.891477809
GLSEDTTEETLKESFDGSVR	735.013243	52.21	399	418	1.028120215	0.892488445
TGISDVFAK	470.2531	40.45	146	154	1.01785008	0.919304582
SKGYAFIEFASFEDAK	604.961315	67.85	343	358	0.965470774	0.956801899
VEGTEPTTAFNLFVGNLNFNK	772.390097	102.69	119	139	1.010858945	0.959180766
SISLYYTGEK	581.795329	36.15	279	288	1.008079381	0.966120408