Supplementary Materials

1. Supplemental Material and Methods

   Calculation of CBL0137 binding to DNA

   Supplemental figures S1-S6

2. Table S1. MACS analysis of ChIP-seq data

3. Table S2. biClustering analysis of ChIP-seq data

4. Table S3. ColoWeb analysis of overlap of SSRP1 binding sites and non-B DNA

5. Table S4. LoLa analysis of overlap of SSRP1 binding sites and non-B DNA

Supplemental Material and Methods

Cloning of GFP-tagged SSRP1 mutants with nuclear localization signal

To generate GFP-tagged SSRP1 deletion mutant first nuclear localization signal (NLS) was cloned into pTZV3-CMV-eGFP vector (Tranzyme) instead of GFP through BamHI and XhoI using annealed oligonucleotides: NLS-fw: 5' AGATCT ATG CCG CCT AAA AAG AAA CGT AAG GTA CTCGAG and NLS-rv: 3' TCTAGA TAC GGC GGA TTT TTC TTT GCA TTC CAT GAGCTC. GFP lacking first codon for methionine was PCR amplified from original pTZV3-CMVeGFP vector using the following primers: GFP-fw (no Met): 5' AAA TTT CTC GAG GTG AGC AAG GGC GAG GA and GFP-rev:3' AAA TTT TCT AGA CTT GTA CAG CTC GT. PCR products were cut with XhoI and XbaI and cloned into pTZV3-CMV-NLS vector. SSRP1 mutants, in the form of PCR products amplified from pEZ-LV105-SSRP1 plasmid (GeneCopoeia, Rockville, MD), were cloned into pTZV-CMV-NLS-GFP via XbaI and AgeI. The following primers were used:

C-term-Stop: 5'-CGACGTGCTCTAGAGCCAGCTCCTCCAGT-3' and

5'-CGACGTGCACCGGTCTACTCATCGGATCCTG-3';

HMG-Stop: 5'-ACGCGTCTTCTAGACCCAAGAGGCCCATG-3' and

5'-CGACGTGCACCGGTCTACTCATCGGATCCTG-3';

HMG: 5'-ACGCGTCTTCTAGACCCAAGAGGCCCATG-3' and

5'-CGACGTGCACCGGTTCATTTCATGGCTTTTTC-3';

IDD-HMG: 5'-CGACGTGCTCTAGAGCCAGCTCCTCCAGT-3' and

5'-CGACGTGCACCGGTTCATTTCATGGCTTTTTC-3';

CID: 5'-ACGCGTCTTCTAGAAAAGAATATGAAGGG-3' and

5’-CGACGTGCACCGGTCTACTCATCGGATCCTG-3’.

DNA-ase I footprinting

The sequence of 200 bp DNA fragment used in footprinting (linear DNA fragment): 3'-GAGATCTCAGCTGGACGTCCGTACGTTCGAACCGCATTAGTACCAG TATCGACAAAGGACACACTTTAACAATAGGCGAGTGTTAAGGTGTGTTGTATGCTC GGCCTTCGTATTTCACATTTCGGACCCCACGGATTACTCACTCGATTGAGTGTAATT AACGCAACGCGAGTGACGGGCGAAAGGTCAGCCCTTTTATA -5’. Fluorescently labeled PCR product was obtained using pUC19 as a template and the following primers: 5’-TAMRA-atattttcccgactggaaagcggg-3' and 5'-ctctagtgtcgacctgcaggcatg-3'. After purification via PAAG electrophoresis DNA was dissolved in buffered solution containing 10 mM Tris-HCl, 0,1 mM CaCl2, 2,5 mM MgCl2, pH 7,5 up to concentration of ~0.2 μM. Then samples of 18 μl of DNA solution were mixed with 2 μl of the ligand solution at appropriate concentration. After 15 min of incubation at 37°C to ensure equilibration of the binding reaction, the digestion was initiated by the addition of 4 ml of a DNase I solution which concentration was adjusted to yield a final enzyme concentration of ~0.05 U/ml in the reaction mixture. After 15 min, the reaction was stopped by heating at 90oC for 5 min in the presence of 50% formamide. Electrophoresis was performed on 10% polyacrylamide gel under denaturing conditions (7 M urea in 20% formamide). Typhoon 9410 Variable Mode Imager was used for data collecting from the gel.

The tBiClustering procedure

For identification of patterns of sequences associated with SSRP1 in curaxin-treated cells, we used SAM files of HT1080 cell samples. Read coverage (read per position counts) was calculated per interval of 30 bp across all chromosomes. The intervals with difference between sums of counts from three replicates of CBL0137-treated samples and sum of counts in three replicates of control samples more than 300 were selected. The sequences of the selected intervals were clustered by the bi-clustering procedure, tBiClustering. We developed this approach to find repeats in raw reads based on unsupervised detecting dense associations between k-mers (k-tuples in sequences, k = 12, 15) and the CHiP-seq sequenced fragments of the genome (reads). In general the task can be defined as a search for densely connected subsets of vertices of two disjoint types in a bi-partite graph G {V1, V2, E}. A bipartite graph is an undirected graph whose vertex set V can be partitioned in two disjoint subsets V1 and V2, such that every edge of E connects a vertex from V1 and one from V2 (Diestel 2005). Here two types of vertices are k-mers and the CHiP-seq reads, and a k-mer and a read are linked by an edge if the k-mer is a sub-sequence of the read.

Our tBiClustering approach is exploiting the co-clustering idea: find two subsets, $Vs \subset V1$ and $Vt \subset V2$, simultaneously, such that they are densely connected. The density of a bipartite sub-graph Gst is defined as:

$$dGst = |Est|/(|Vs| \bullet |Vt|)$$

According to this definition, $dGst \in [0, 1]$, and a subgraph has the density one if and only if it is a biclique.

Our tBiClustering approach is using the Markov Chain Monte Carlo Gibbs-sampling like algorithm (KADIYALA and KARLSSON 1997) for finding dense sub-clusters (better that a threshold for density) in a total bi-partite graph: all CHiP-seq reads vs. all their k-mers. The algorithm is described by the following steps:

1. Initialize randomly a set of reads as vertices Vs and generate a set of all k-mers (vertices Vt) that are sub-sequences of the Vs reads.

2. By greedy algorithm select k-mers from Vt with highest counts as sub-sequences in the reads of Vs: the set of vertices Vt1 that provide maximum density of the Gst1 {Vs, Vt1, Est1 } graph.

3. Based on the Vt1 set of kmers update the set Vs of reads to Vs1 by removing the reads from Vs with lowest links to Vt1 in order to provide maximum density of the Gs2t1 {Vs1, Vt1, Es2t1 } graph.

4. Based on the Vt1 set of kmers update the set Vs1 of reads to Vs2 by adding the reads from the V-Vs set of reads with the highest counts of links to Vt1 in order to provide maximum density of the Gs2t1 {Vs2, Vt1, Es2t1 } graph.

5. If Vs2 set not equal to the Vs set then go to Step 2.

6. Else finish building of the current tBiCluster Gsitj {Vsi, Vtj, Esitj }: check if density of the created tBiCluster is better that the predefined threshold.

7. Remove edges of the Esitj set from the set of all edges E if tBiCluster is accepted.

8. Go to Step 1 and start a new tBiCluster.

9. The procedure stops when the set E is exhausted.

Thus, on every cycle starting with randomly selected reads the algorithm converges to a dense bi-partite sub-graph of mutually linked reads and kmers. The assembled kmers inside every tBiCluster are repetitive sub-sequences of the CHiP-seq reads.

Calculations of K_d of CBL0137

Let us consider binding of ligands L with DNA,

$$\mathrm{DNA} + \mathrm{L} \leftrightarrow \mathrm{DNA} \cdot \mathrm{L}$$

The dissociation constant for the process is defined as

$$K_{\mathrm{d}} = \frac{[\mathrm{DNA}][\mathrm{L}]}{[\mathrm{DNA} \cdot \mathrm{L}]}, \qquad (1)$$

where DNA concentration is measured in M of bp. Introducing the ratio of bound ligands to DNA base pairs as ν, we obtain that

$$K_{\mathrm{d}} = \frac{1-\nu}{\nu}[\mathrm{L}], \quad \text{or} \qquad \frac{\nu}{[\mathrm{L}]} = \frac{1-\nu}{K_{\mathrm{d}}} \tag{2}$$

The latter form of the equation is usually called the Scatchard plot. To find K_{d} we need to find v from experimental data. We can do it by analyzing the obtained gels.

The mobility of supercoiled DNA in the gel depends on the linking number difference, ΔLk, which is defined as

$$\Delta Lk = Lk - Lk_0, \tag{3}$$

where *Lk* is the number of turns one strand makes around the other strand in the circular DNA molecule, and Lk_0 is the number of turns in the same DNA in linear form. *Lk* is a topological feature of the molecule, it cannot be changed without breaking at least one strand. Lk_0, on the other hand, is not an invariant, and depends on solution condition. Lk_0 increases with adding more ions in the solution and decreases when the temperature is rising. It is important that mobility in gel depend only on the absolute value of ΔLk.

When DNA is ligated in solution without any added ligand, its average *Lk* is equal to Lk_0. However, when your circular DNA is transferred from ligation buffer to the electrophoresis buffer, Lk_0 is reduced and average ΔLk becomes positive, +2 in our case. During ligation in the presence of an intercalating ligand the average *Lk* becomes smaller than in absence of the ligand, due to DNA unwinding by the ligand intercalation. So, DNA becomes negatively supercoiled after the ligand is washed out from the solution, and its supercoiling keeps the information about the unwinding. On the gel it looks like average ΔLk is ≈ 0 for 1 μM of CBL and (-3.3) for 2 μM of CBL. To get actual ΔLk due to the ligand intercalation we have to account for the change of Lk_0 due to the buffer change, which is +2. Thus, the actual ΔLk introduced by the CBL equals -2 and -5.3, correspondingly.

To convert the values of ΔLk into the number of bound molecules of a ligand, *n*, we need to know the unwinding angle associated with the intercalation. The average value of the angle over known intercalating ligands is (-18º). Assuming that this is the unwinding for CBL, we obtain that

$$n = \Delta Lk \times 360^{\circ} / 18^{\circ} = 20\Delta Lk \, .$$

This gives us 40 and 106 bound molecules per DNA molecule in 1 and 2 μM of CBL, correspondingly. Converting this into v we need to divide this on DNA length. So, we obtain $v_1 = 0.015$ and $v_2 = 0.04$. To calculate K_d we also need to know [L]. We can calculate it as

$$[L] = [L]_0 - v[DNA]_0, \quad (4)$$

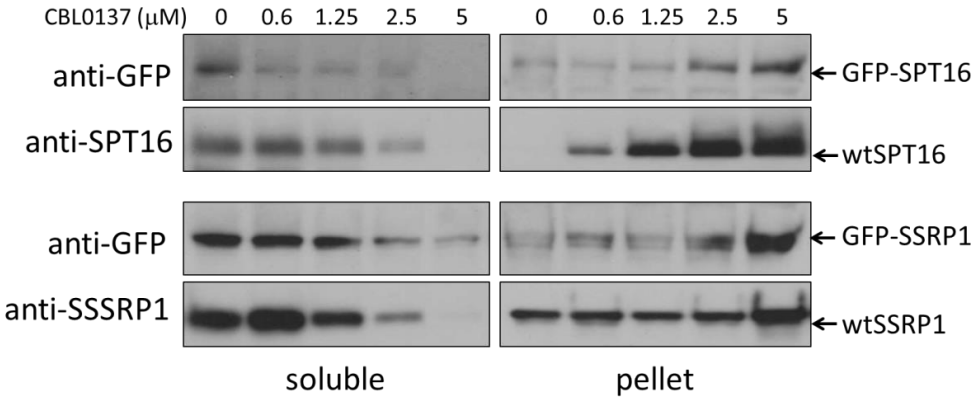
where $[DNA]_0$ is the total DNA concentration in μM of bp. So, $[L]_1$ = 0.89 μM and $[L]_2$ = 1.7 μM. Substituting this into Eq. (2) we obtain two estimations for K_d, 58 μM and 41 μM. The value of ΔLk was estimated more accurately for 2 μM of CBL, so the second estimation is more accurate. Still, the error in K_d is rather large, since we do not know the unwinding angle precisely. Therefore, it seems reasonable to say that

$$K_d = (40 \pm 20)\ \mu M.$$

This dissociation constant is nearly 3 orders of the magnitude lower than K_d for chloroquine under ionic conditions of the ligation buffer.

Supplemental Figures

A



B

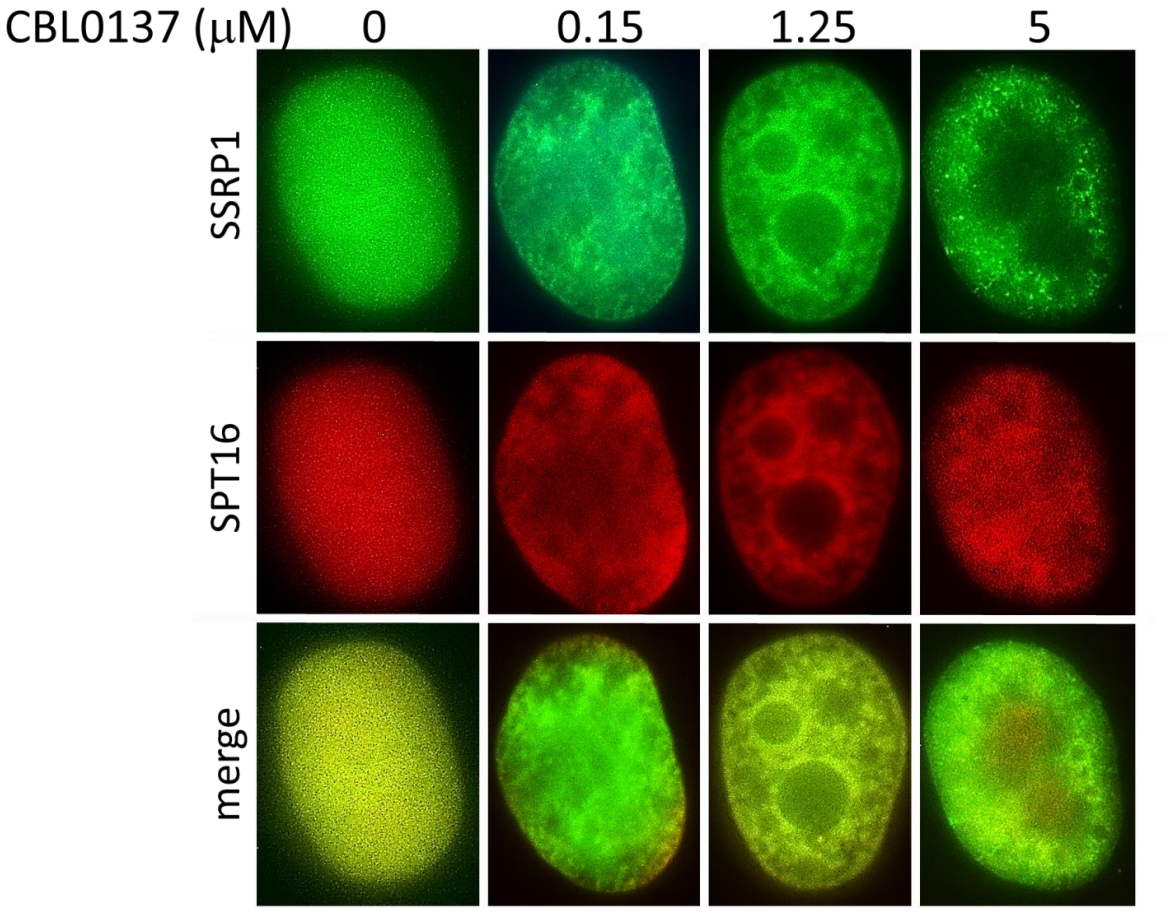


Figure S1. *C-trapping* of FACT. A. Similar pattern of *c-trapping* between endogenous wild type (wt) and GFP- tagged FACT subunits. Immunoblotting of the lysates of A293 cells transfected with either GFP-tagged SSRP1 or SPT16 48 hours prior to treatment and treated with the indicated doses of CBL0137 for 1 hour. B. Comparison of *c-trapping* of SSRP1 and SPT16. Fluorescent images of HT1080 cells transduced with GFP-tagged SSRP1 and mCherry tagged SPT16 and treated with CBL0137 for 1 hour.

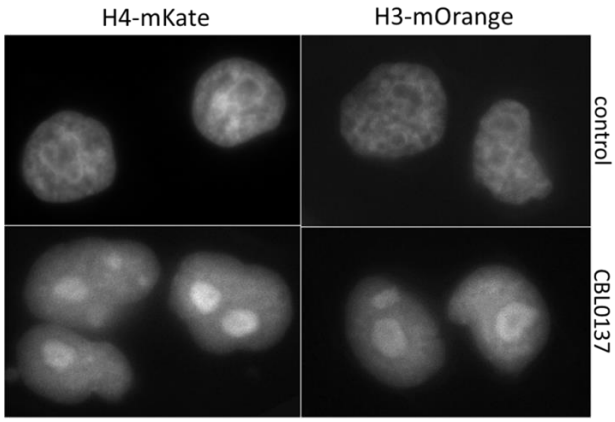


Figure S2. Effect of CBL0137 on chromatin. A. Fluorescent imaging of nuclei of HeLa cells transfected with histones tagged with fluorescent proteins before and after treatment with 5 μM CBL0137 for 15 minutes.

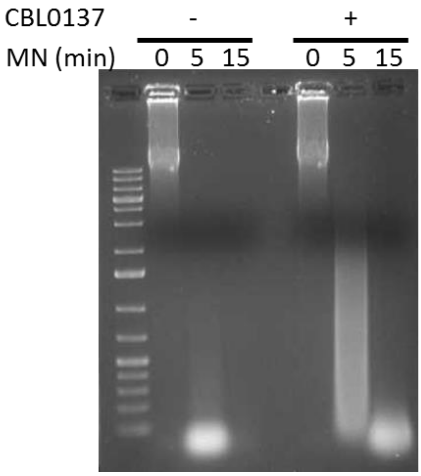
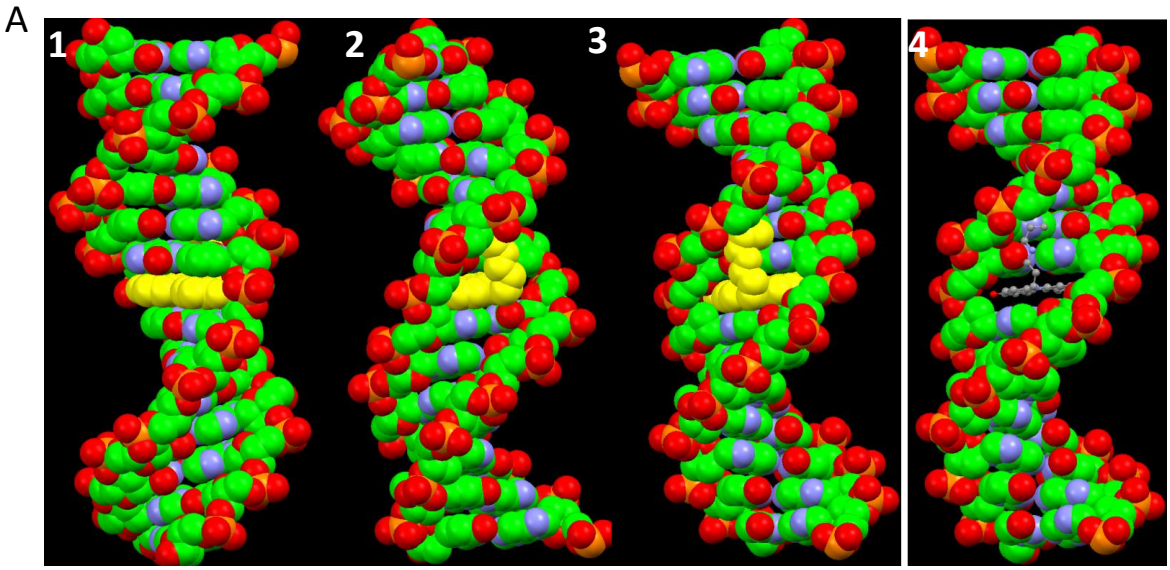
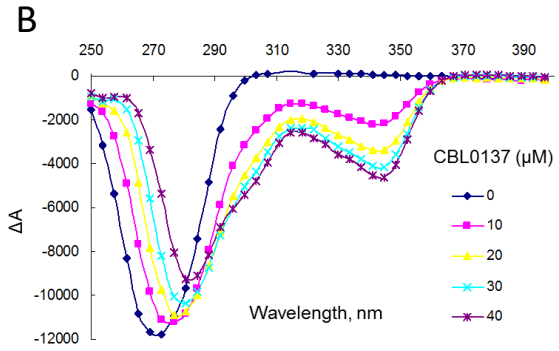
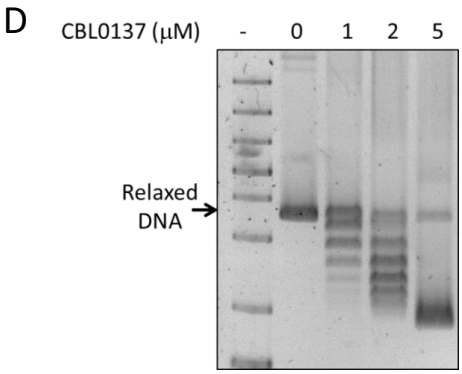


Figure S3. Gel electrophoresis of genomic DNA incubated with vehicle control or 10 μM CBL0137 (~1 molecule per 10bp) for 15 min followed by digestion with Micrococcal nuclease (MN) for the indicated periods of time.







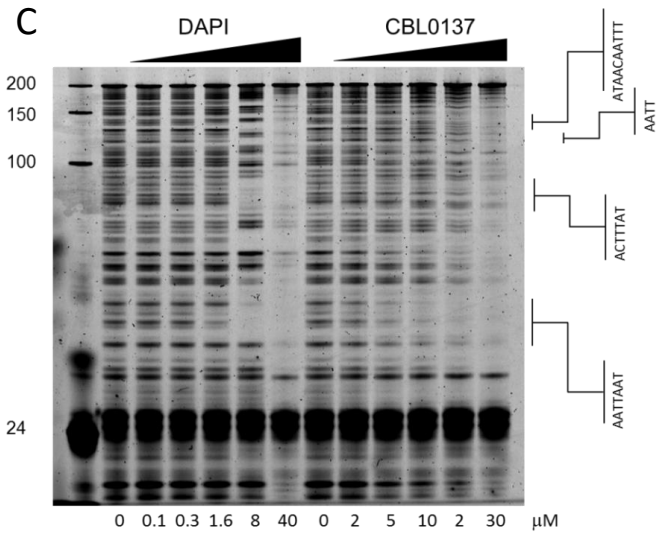


Figure S4. Characterization of CBL0137 binding to DNA. A. Modeling of CBL0137 binding to DNA demonstrates the location of carbazole side chains in the major groove (1) and the minor groove (2-4). 1-3 – space filling models, 4 – space filling model for DNA and stick model for curaxin CBL0137. B. Circular dichroism of DNA alone or in the presence of increasing amount of CBL0137. The appearance of a new negative band with λmax = 345 nm in the presence of CBL0137 indicates that the chromophore fragment (namely carbazole) is located in the DNA complex with an angle value ranged ranging from 54 to 90° with respect to the DNA helix axis, which is possible only upon CBL0137 intercalation. C. DNAse I footprinting of a 200bp DNA fragment in the presence of CBL0137 and DAPI used as positive control. D. Gel electrophoresis of circular nicked DNA incubated with different concentrations of CBL0137 followed by ligation and DNA purification.

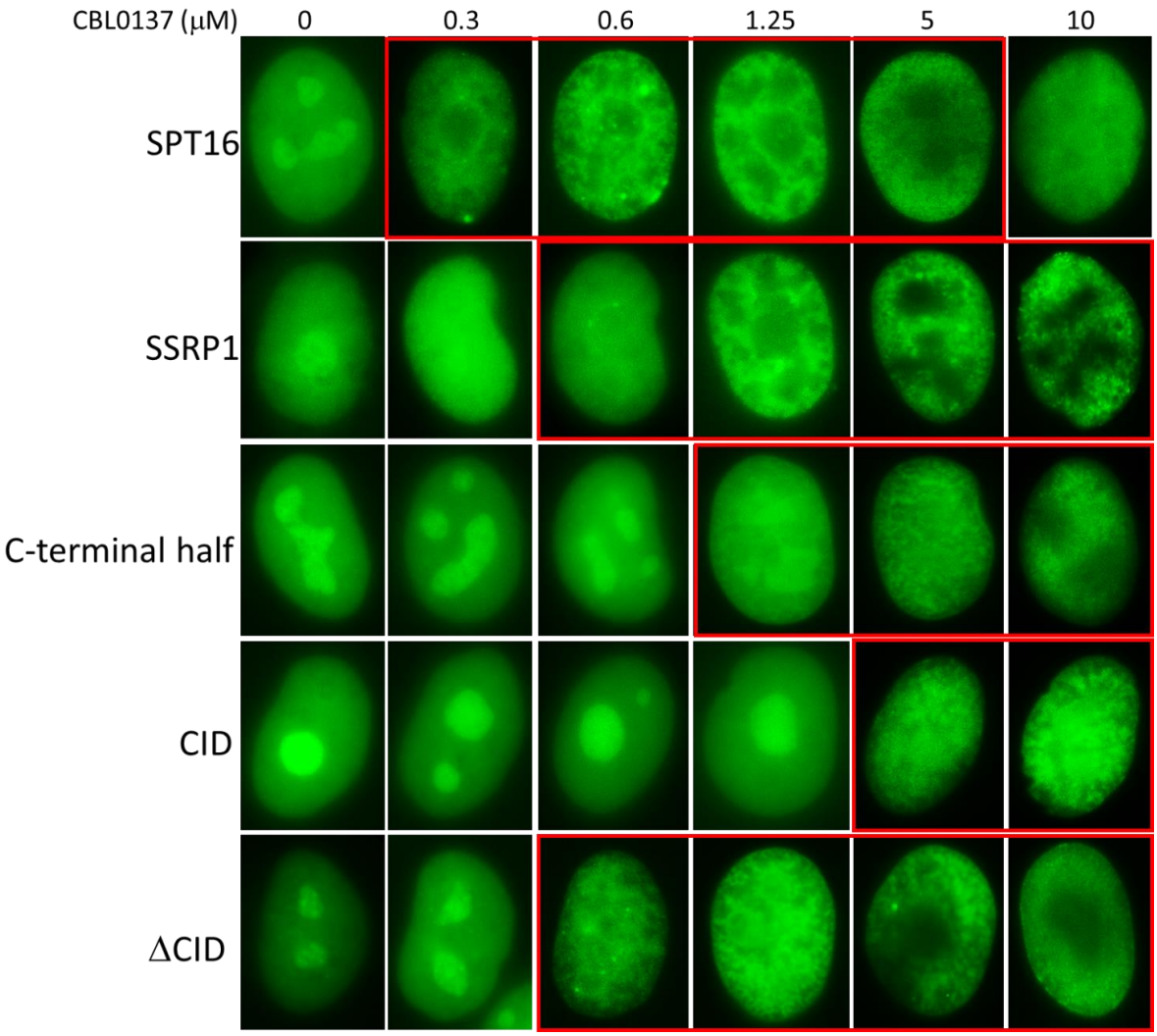


Figure S5. Comparison of *c-trapping* between GFP-tagged SPT16, SSRP1 and SSRP1 deletion mutants. Fluorescent images of HT1080 cells stably expressing the indicated constructs. Cells were treated with CBL0137 for 1 hour. Red rectangles indicate cells with evidence of *c-trapping*.

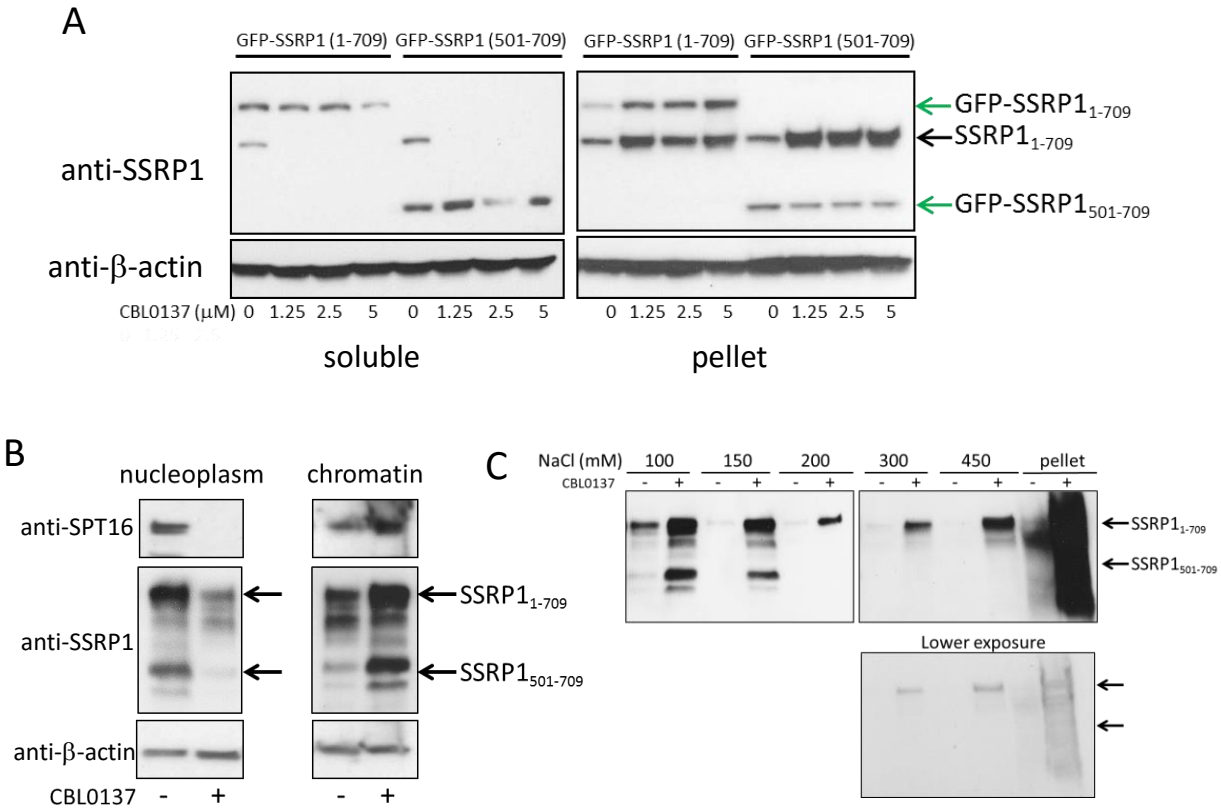


Figure S6. C-terminal half of SSRP1 binds chromatin with lower affinity than full length SSRP1. A. Immunoblotting of soluble cell extract and chromatin pellet from cells transduced with GFP-tagged SSRP1 (1-709aa) or C-terminal half of SSRP1 (501-709aa) and treated with different concentrations of CBL0137 for 1 hour. Green arrows – ectopic SSRP1 variants, black arrow - endogenous SSRP1. B. Immunoblotting of nucleoplasm and chromatin fraction of HT1080 cells transduced with GFP-tagged SSRP1 (1-709aa) and C-terminal half of SSRP1 (501-709aa) and treated with 3 μM of CBL0137 for 1 hour. C. Western blotting of fractions obtained via sequential washing of the chromatin pellet from B with buffer containing increasing concentrations of NaCl.

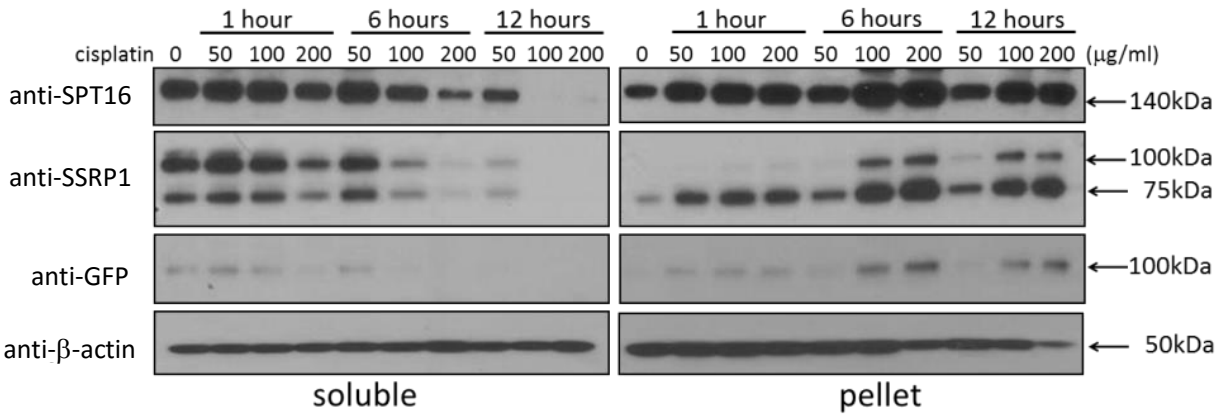


Figure S7. *C-trapping* in cisplatin-treated cells. Immunoblotting of the lysates of HT1080 cells transduced with GFP-tagged SSRP1 and treated with different doses of cisplatin for the indicated period of time. SSRP1 antibody detects both endogenous SSRP1 (~80kDa) and GFP-tagged SSRP1 (~100kDa).

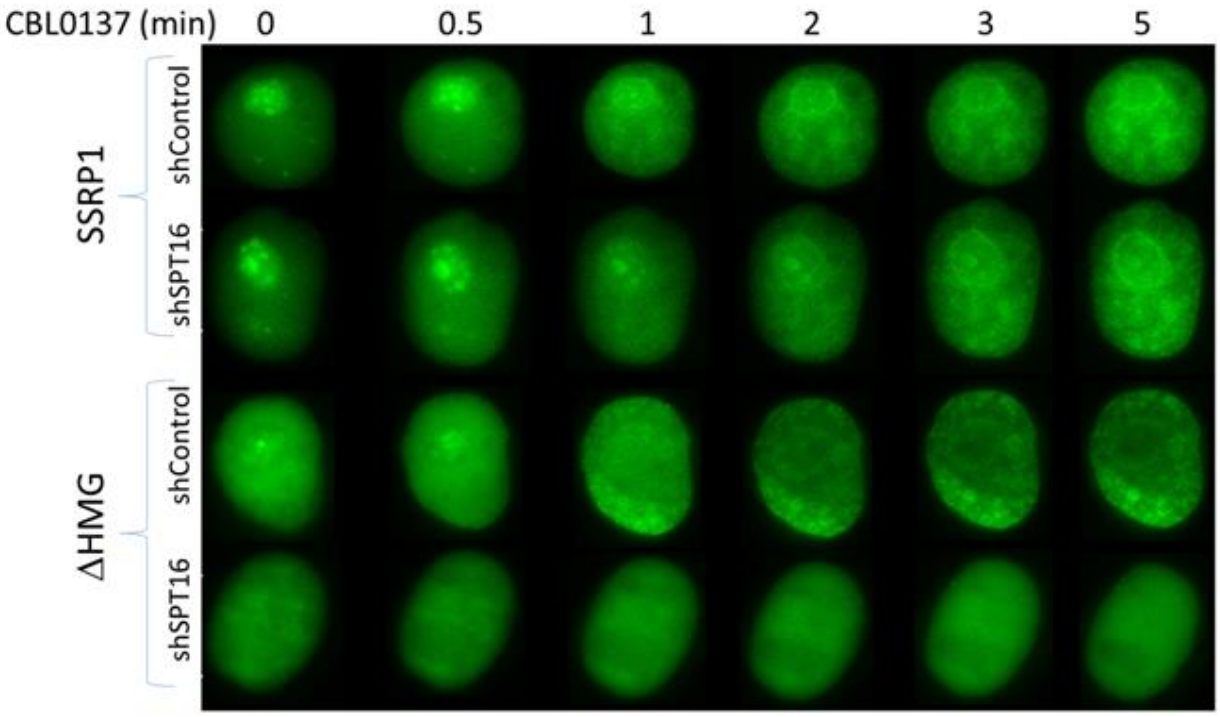
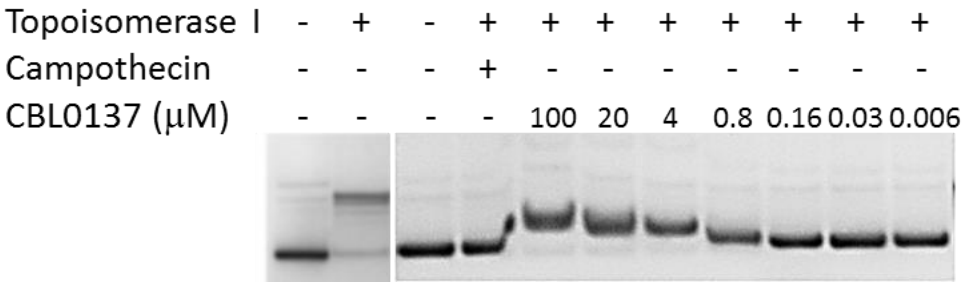


Figure S8. SSRP1 lacking HMG and CID domains undergoes c-trapping only in the presence of SPT16. HT1080 cells stably expressing either full length GFP tagged SSRP1 or SSRP1 lacking the HMG and CID domains (ΔHMG) were transduced with lentiviral shRNA constructs, control or SPT16. Immunofluorescence images of the same cells for each genetic background before and at different time points after start of treatment with 2 μM of CBL0137.

A



B

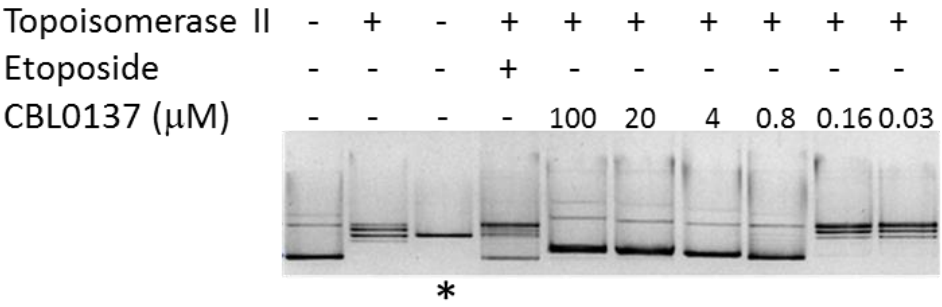


Figure S9. Inhibition of topoisomerase I (A) and topoisomerase IIa (B) activity by CBL0137. Agarose gel electrophoresis of the products of in vitro assays (see details in Material and Methods) with supercoiled plasmid DNA. Campothecin – topoisomerase I inhibitor. Etoposide – topoisomerase II inhibitor. * - this lane contains linearized plasmid control.

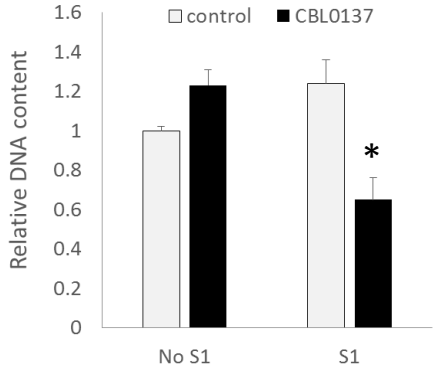


Figure S10. Increased sensitivity of DNA in nuclei of HeLa cells treated with CBL0137 (50μM, cell equivalent of 2.5μM, 15 min) to S1 digestion. Relative signal intensity of propidium iodide stained DNA per G1 nuclei incubated with CBL0137 followed by treatment with S1 (see details in Material and Methods). Mean of two replicates in three independent experiments. Error bars are s. d. * - $p<0.05$. S1 digestion and or CBL0137 treatment separately both increase accessibility of DNA to propidium iodide (PI) in control cells, but together when combined, they reduce the total amount of DNA stained with PI.

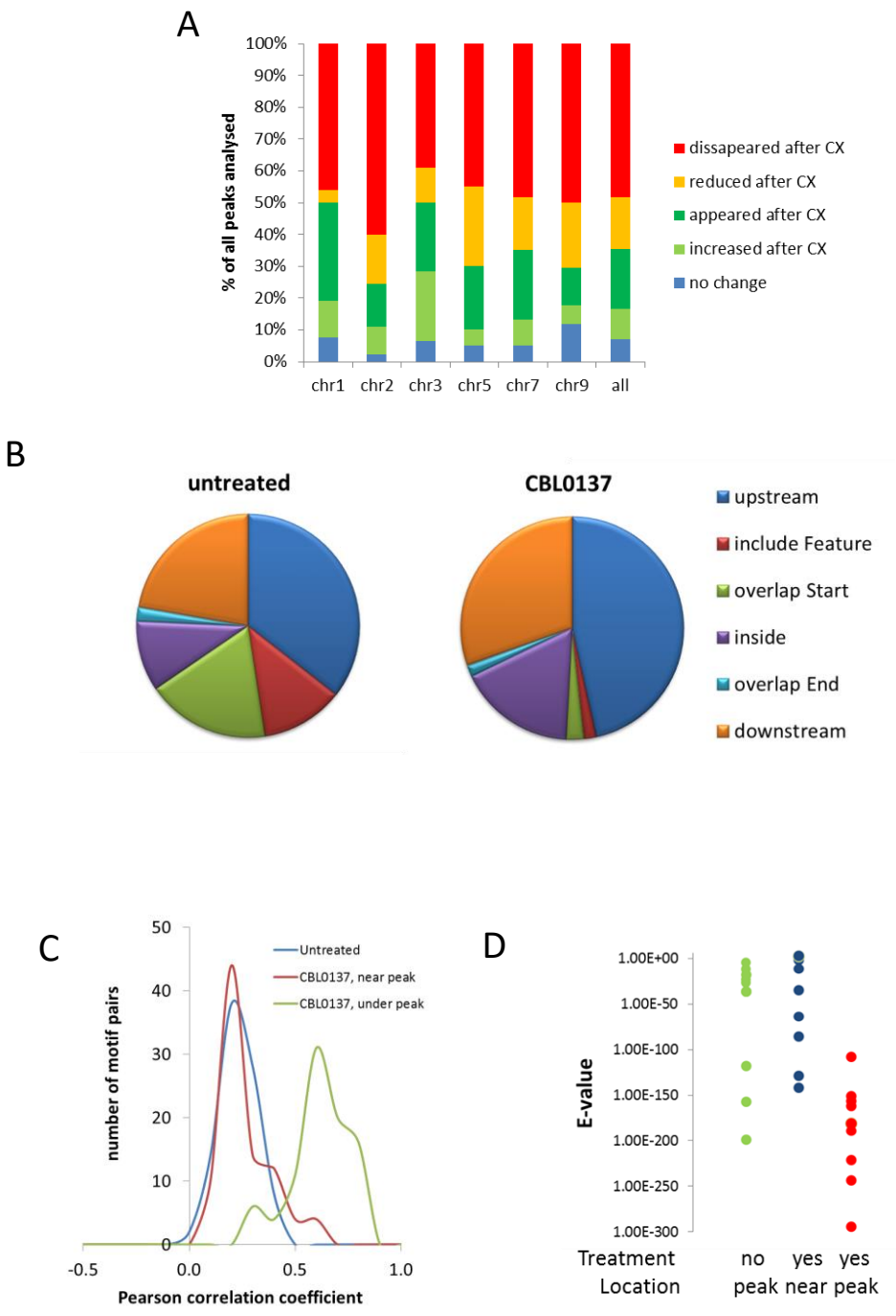


Figure S11. Curaxin CBL0137 treatment leads to redistribution of FACT in chromatin of cells. A. Analysis of changes in distribution of peaks identified using ChIP-sequencing in HT1080 cells before and after curaxin CBL0137 treatment. B. Distribution of SSRP1 binding in relation to genome features assessed using MACS. C, D. Identification of unknown motifs using MEME among sequences underlying SSRP1 bound peaks in untreated cells, cells treated with CBL0137, and adjacent to SSRP1 bound peak regions in cells treated with CBL0137. C. Similarity of 10 motifs with the best e-values identified in each pool of sequences. Histogram of Pearson correlation coefficients of each pair of motifs. D. Comparison of distribution of e-values between 10 best motifs from each pool of sequences.

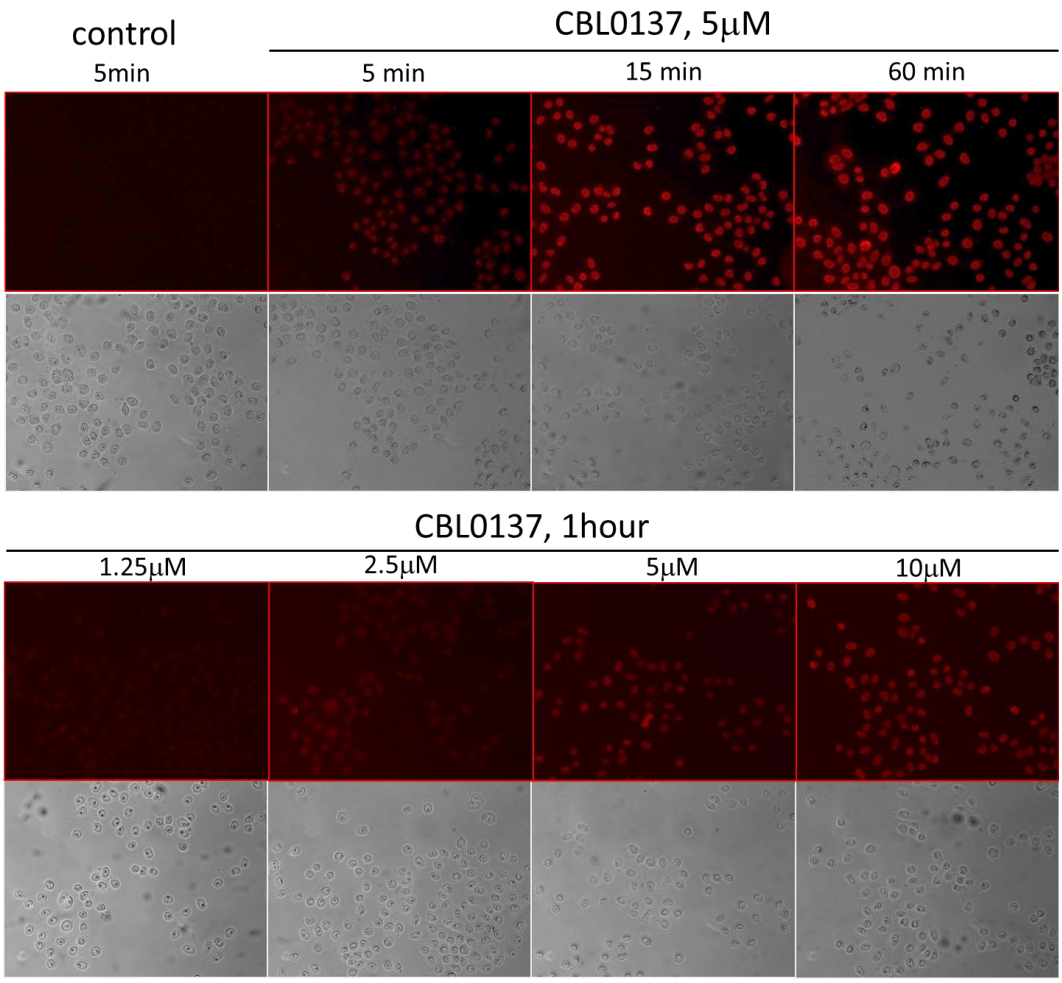
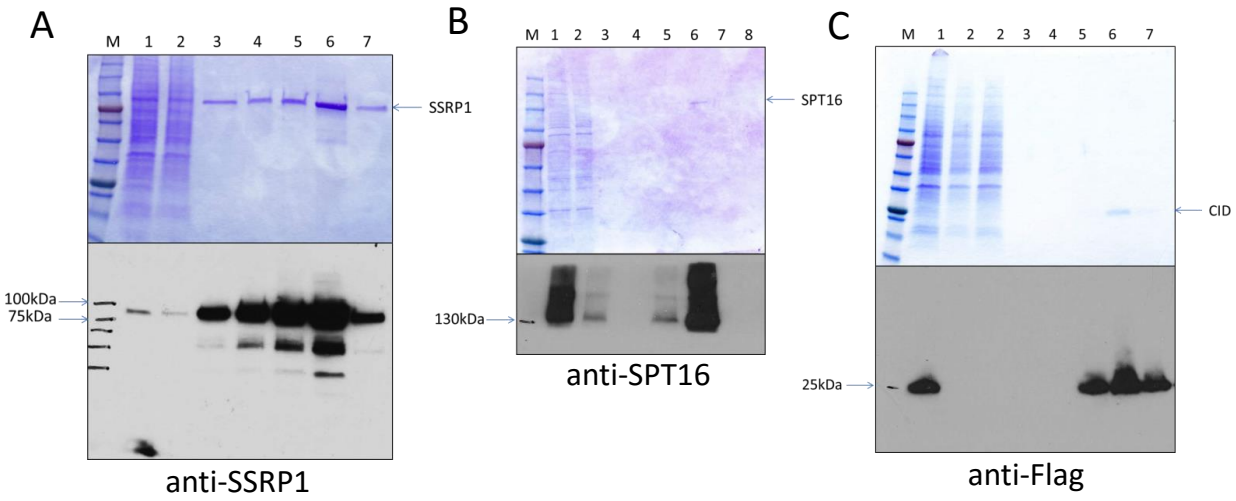


Figure S12. Staining of HeLa cells with antibody to Z-DNA.
Immunofluorescence imaging with corresponding phase contrast.



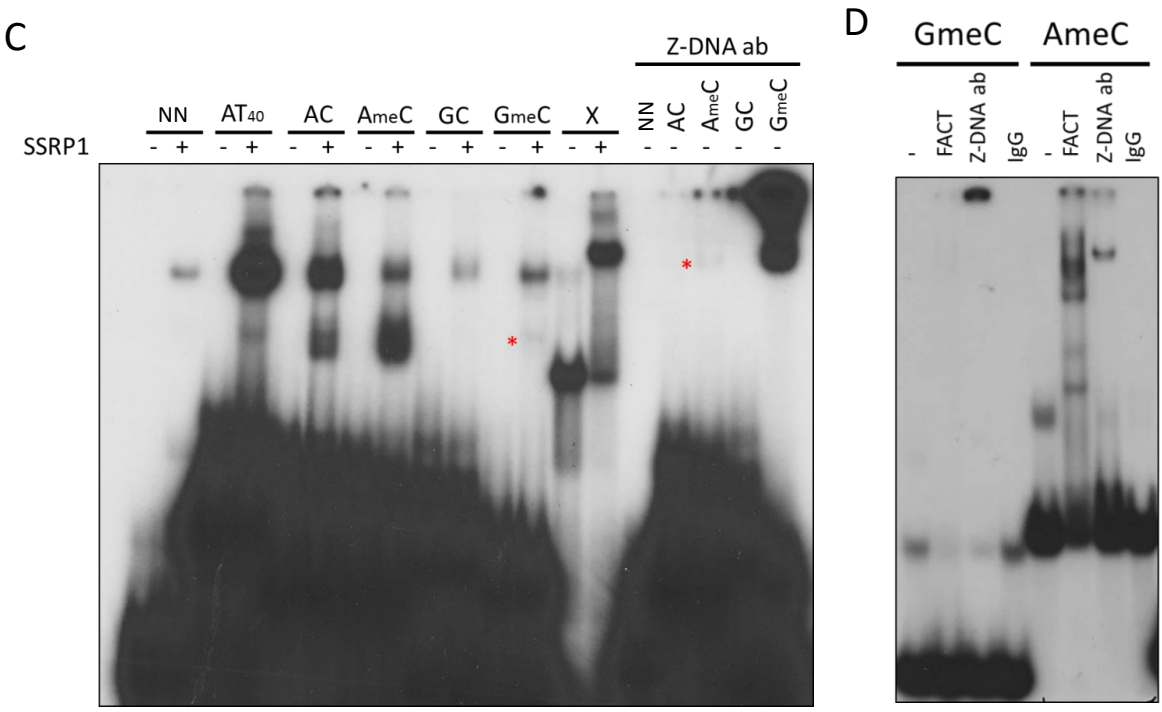


Figure S13. Binding of SSRP1 to different types of DNA oligonucleotides in cell-free conditions. A-C. Results of affinity purification of Flag-tagged SSRP1 (A), SPT16 (B) and CID domain of SSRP1 (C) from Hela cells. Coomassie staining and western blotting of lysate (1), flow-through fractions (2) and eluates with Flag peptide (3-8). C-D. Gel shift assays of (C) SSRP1 or Z-DNA antibody incubated for 20 min at RT with different types of ^{32}P-labeled double stranded linear oligonucleotides: 17 bp non-repetitive DNA fragment (NN), repetitive fragments of the indicated composition (exact composition is given in Materials and Methods) or cruciform DNA probe (X). (D) of FACT, Z-DNA antibody or control IgG with ^{32}P-labeled GmeC or AmeC probes.