



**HAL**  
open science

## Structural analysis of the *Candida albicans* mitochondrial DNA maintenance factor Gcf1p reveals a dynamic DNA-bridging mechanism

Aleix Tarrés-Solé, Federica Battistini, Joachim Gerhold, Olivier Piétrement, Belén Martínez-García, Elena Ruiz-López, Sébastien Lyonnais, Pau Bernadó, Joaquim Roca, Modesto Orozco, et al.

### ► To cite this version:

Aleix Tarrés-Solé, Federica Battistini, Joachim Gerhold, Olivier Piétrement, Belén Martínez-García, et al.. Structural analysis of the *Candida albicans* mitochondrial DNA maintenance factor Gcf1p reveals a dynamic DNA-bridging mechanism. *Nucleic Acids Research*, 2023, 10.1093/nar/gkad397. hal-04120278

**HAL Id: hal-04120278**

**<https://hal.science/hal-04120278>**

Submitted on 28 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Structural analysis of the *Candida albicans* mitochondrial DNA maintenance factor Gcf1p reveals a dynamic DNA-bridging mechanism

Aleix Tarrés-Solé<sup>1</sup>, Federica Battistini<sup>2,3</sup>, Joachim M. Gerhold<sup>4</sup>, Olivier Piétrement<sup>5</sup>, Belén Martínez-García<sup>6</sup>, Elena Ruiz-López<sup>1</sup>, Sébastien Lyonnais<sup>1</sup>, Pau Bernadó<sup>7</sup>, Joaquim Roca<sup>6</sup>, Modesto Orozco<sup>2,3</sup>, Eric Le Cam<sup>5</sup>, Juhan Sedman<sup>4</sup> and Maria Solà<sup>1,\*</sup>

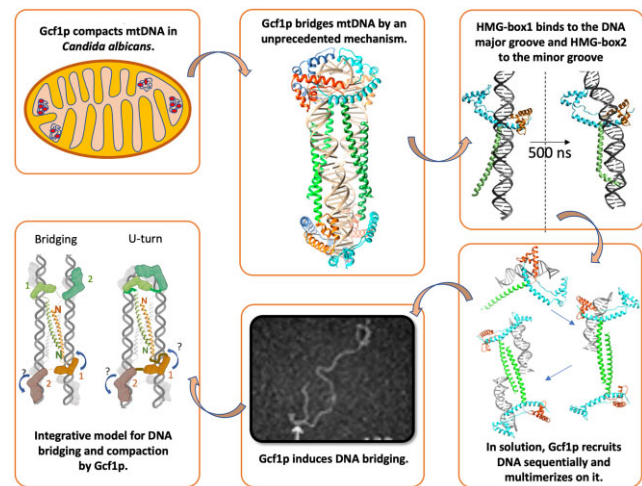
<sup>1</sup>Structural MitoLab, Molecular Biology Institute Barcelona (IBMB-CSIC), Barcelona Science Park, Barcelona 08028, Spain, <sup>2</sup>Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona 08028, Spain, <sup>3</sup>Department of Biochemistry, University of Barcelona, Barcelona 08028, Spain, <sup>4</sup>Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia, <sup>5</sup>Genome Integrity and Cancer UMR 9019 CNRS, Université Paris Saclay, Gustave Roussy Campus, 114 rue Edouard Vaillant 94805 Villejuif Cedex, France, <sup>6</sup>Molecular Biology Institute Barcelona (IBMB-CSIC), Barcelona 08028, Spain and <sup>7</sup>Centre de Biologie Structurale (CBS), Inserm, CNRS and Université de Montpellier, France, Sébastien Lyonnais, UAR 3725 CNRS, Université de Montpellier, 34000 Montpellier, France

Received July 19, 2022; Revised April 01, 2023; Editorial Decision April 27, 2023; Accepted May 03, 2023

## ABSTRACT

The compaction of mitochondrial DNA (mtDNA) is regulated by architectural HMG-box proteins whose limited cross-species similarity suggests diverse underlying mechanisms. Viability of *Candida albicans*, a human antibiotic-resistant mucosal pathogen, is compromised by altering mtDNA regulators. Among them, there is the mtDNA maintenance factor Gcf1p, which differs in sequence and structure from its human and *Saccharomyces cerevisiae* counterparts, TFAM and Abf2p. Our crystallographic, biophysical, biochemical and computational analysis showed that Gcf1p forms dynamic protein/DNA multimers by a combined action of an N-terminal unstructured tail and a long helix. Furthermore, an HMG-box domain canonically binds the minor groove and dramatically bends the DNA while, unprecedentedly, a second HMG-box binds the major groove without imposing distortions. This architectural protein thus uses its multiple domains to bridge co-aligned DNA segments without altering the DNA topology, revealing a new mechanism of mtDNA condensation.

## GRAPHICAL ABSTRACT



## INTRODUCTION

*Candida* species are major components of the healthy mammalian mucosa mycobiome, and are present in the oral cavity (1), gastrointestinal (GI) (2) and reproductive (3) tracts, and on the external epithelium (4). *C. albicans* confers several benefits in mammals, including the stimulation of mucosal immunity in the GI-tract mucosa (5), the direct

\*To whom correspondence should be addressed. Tel: +34 934034950; Fax: +34 934034979; Email: maria.sola@ibmb.csic.es  
Present addresses:

Aleix Tarrés-Solé, ALBA synchrotron, 08290 Cerdanyola del Vallès, Barcelona, Spain.

Joachim Gerhold, Icosagen Cell Factory OÜ, Tartu County 61713, Estonia.

Olivier Piétrement, Laboratoire Interdisciplinaire Carnot de Bourgogne, CNRS UMR 6303, Université de Bourgogne, 21078 Dijon Cedex, France.

inhibition of pathogens (6), and the triggering of trained immunity (7). However, *Candida* spp. can transition from beneficial unicellular forms to filamentous hypha that cause infections ranging from superficial to invasive and potentially life-threatening (8–12). In hospitals in the developed world, invasive candidiasis is the most widespread fungal disease and, unfortunately, candida species are becoming increasingly resistant to treatments (10).

The search for hypha-specific regulators in *C. albicans* identified the Glom-like DNA condensation factor (Gcf1p), which displays dynamic oligomerization behavior, forming monomers and dimers in hyphal extracts but higher-order species in yeast cells (13). Genomic and functional comparisons indicated a role for Gcf1p in mitochondrial nucleoids (14), which contain mtDNA and associated regulatory proteins. *C. albicans* mtDNA is a linear 40.4 kb molecule that encodes 14 subunits of the oxidative phosphorylation (OXPHOS) pathway, the main source of ATP in aerobic conditions (15). Impairment of *Candida* mtDNA or OXPHOS starves the cell of ATP and reduces viability (16–19). Gcf1p overexpression was shown to increase mtDNA levels by > 50%, whereas gene repression decreased mtDNA levels by 80%, reduced cell growth, and delayed cell-cycle (20). Deletion of the GCF1 gene reduced the number of recombination structures and Holliday junctions significantly, suggesting that Gcf1p maintains mtDNA by participating in recombination-dependent mtDNA replication (21,22). Gcf1p sequence analysis and structural prediction indicated the presence of two high-mobility group (HMG)-box domains (23), <https://alphafold.ebi.ac.uk/search/text/gcf1p>. This resembles the well-characterized mtDNA compactors from *S. cerevisiae* (Abf2p) and mammals (TFAM), which bind non-specifically to mtDNA and thereby induce bending and compaction (24–28). Gcf1p may therefore fulfil a similar architectural function (23). However, functional studies revealed some divergencies. In addition to its architectural function, TFAM also binds to mtDNA promoters (29), a function not described for Gcf1p. Functional complementation experiments showed that TFAM can partially complement the loss of Abf2p in *S. cerevisiae* (30), whereas Gcf1p cannot (20). Moreover, Gcf1p is predicted to have an additional N-terminal helical extension that is probably involved in coiled-coil interactions (20,23). Furthermore, Abf2p and TFAM bind efficiently to 25 bp fragments of DNA whereas Gcf1p requires at least 50 bp (23). This data suggest that Abf2p and TFAM use different DNA compaction mechanisms compared to Gcf1p.

To shed light on the regulation of mtDNA by Gcf1p, we prepared Gcf1p/DNA complexes and analyzed them by an integrative approach including complementary techniques such as X-ray diffraction, size exclusion chromatography combined with multi-angle laser light scattering analysis (SEC-MALLS), small-angle X-ray scattering (SAXS), electrophoretic mobility shift assays of mutants (EMSAs), transmission electron microscopy (TEM) topology analysis and molecular dynamics (MD) simulations. The results show a highly dynamic DNA bridging mechanism by Gcf1p, which implies non-canonical structural features for an HMG-box protein and which does not alter DNA topology.

## MATERIALS AND METHODS

### Cloning, mutagenesis, protein production and complex preparation

Mature Gcf1p (Gcf1p) comprises residues 25–245, lacking the Mitochondrial Targeting Sequence (MTS) (13). The N-terminal deletion (mutant Gcf1p53-245) was generated by round-the-horn PCR (31) by using oligonucleotides 5'-TCTACCAAACCTCCTAAAGTCGATACCAAG-3' and 5'-TCCTTGGAATACAGGTTTTCCAGATCCGATT TTGGAGGATGGTCGCC-3' with Phusion polymerase (Thermo Fisher Scientific). The mutant Gcf1pCCmut (V58A, L65A, I69A, A75K, L79K) was prepared by Genscript Biotech. All proteins constructs were fused to glutathione-S-transferase (GST) in vector pGEX-4T1 re-engineered to contain a tobacco etch virus (TEV) cleavage site. The recombinant protein was expressed in non-auxotrophic *Escherichia coli* expression strain Rosetta 2 cultured in LB medium. Seleno-methionine (Se-Met; Molecular Dimensions) derivatives were produced in minimal medium and the endogenous synthesis of L-methionine was inhibited by adding an amino acid mix (Sigma-Aldrich) when the culture reached an OD at 600 nm of 0.5. All cultures were grown at 37°C to an OD<sub>600</sub> of 0.6, and protein expression was induced by adding IPTG (final concentration 1 mM) followed by incubation at 24°C for 4 h. Cell pellets were resuspended in lysis buffer A (750 mM NaCl, 100 mM HEPES-Na pH 7.25) supplemented with 5 mM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol, 1 mM EDTA, 1× protease inhibitor cocktail (PIC, Roche), 25 μg/ml of either RNase A and DNase I, and lysed using a CF-1 cell disruptor (Constant Systems). Lysates were supplemented with 1 mM PMSF and 0.01% Triton X-100, centrifuged at 74 000×g for 30' using a JA25-50 rotor in a JX20 Avanti centrifuge (BD Biosciences) and loaded onto a GSTrap FF 1 ml column (Cytiva) coupled to an Äkta FPLC device (GE Healthcare) at a flow rate of 0.5 ml/min. The column was washed with 10 column volumes (CV) of washing buffer (1 M NaCl, 100 mM HEPES-Na pH 7.25, 5 mM MgCl<sub>2</sub>, 1 mM β-mercaptoethanol, 1 mM EDTA, 1× PIC, 10 μg/ml of either RNase A and DNase I, 0.01% Triton X-100) followed by 2 CV of digestion buffer (buffer A supplemented with PIC 1× and 1mM PMSF). The column was injected with 50 μg TEV protease, incubated O/N at RT in the presence of 1 mM PMSF, and the protein was eluted in 10 CV of fresh digestion buffer. Protein samples were diluted with 50 mM HEPES-Na pH 7.5 to a final NaCl concentration of 200 mM (protein concentration, 0.2 mg/ml), and loaded onto a 1 ml MonoS 5/50 column (Cytiva) at a constant flow of 0.5 ml/min. Gcf1p was eluted by applying a linear gradient with buffer B containing HEPES-Na (pH 7.5) and 1 M NaCl, in 30 CVs. Eluted samples were flash frozen in liquid nitrogen before storage at –80°C.

Double-stranded DNA substrates were prepared from an equimolar mix of complementary oligonucleotides (Biomers) annealed O/N from 98°C (5 min) to RT, in the presence of 20 mM NaCl, 20 mM Tris-HCl pH 8.0 and 5 mM MgCl<sub>2</sub>, and stored at –20°C. DNA20 (5'-ATAATAAATTATATAATATA-3' and complementary

sequence) and DNA50 (5'-GGAGCAGGTATTGGTATTGCTATCGTATTCGCAGCTTTAATTAATGGTGT-3' and complementary sequence) were derived from the sequences previously crystallized with Abf2p (26), based on the original 64 bp sequence (32). The complexes were prepared by, first, mixing Gcf1p and DNA at protein:DNA ratios of 1:1.2 and 2:1.2 at a final protein concentration of 0.2mg/ml. Second, protein/DNA complexes were stabilized by a step-wise dialysis for 2h against buffer 1 (350 mM NaCl, 50 mM Tris-HCl pH 8.0) and buffer 2 (180 mM NaCl, 35 mM Tris-HCl pH 8.0), then O/N against buffer 3 (20 mM NaCl, 20 mM Tris-HCl pH 8.0). Complexes were used directly for crystallization and SAXS, or were flash frozen in liquid nitrogen and stored at -80°C for SEC-MALLS.

### EMSA

We labelled 25 pmol of a synthetic oligonucleotide with 3.34 pmol of  $\gamma$ -(32P)-ATP (7000 Ci/mmol, Hartmann) using five units of T4 polynucleotide kinase (Thermo Fisher Scientific) in 10  $\mu$ l of standard kinase buffer A for 30 min at 37°C. The reaction was stopped by heating to 75°C for 10 min. The reaction product was annealed directly with an equimolar amount of the complementary strand oligonucleotide at 95°C for 10 min followed by slow cooling to RT for 3–5 h. The probe DNA was purified by 10% non-denaturing polyacrylamide gel electrophoresis (1:30 ratio of bisacrylamide:acrylamide in 0.5  $\times$  TBE buffer at 15 V/cm for 60 min at 4°C. The DNA fragments were excised from the gel and eluted with 10–20 volumes of TEN (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM NaCl) for 16–20hr at RT. DNA was precipitated with 2.5 volumes of ethanol and redissolved in TEN. Prior to the EMSA, we mixed 1 nM <sup>32</sup>P-end-labeled DNA and the indicated concentrations of Gcf1p in 10  $\mu$ l buffer containing 65 mM Tris-HCl pH 7.5, 150 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 1 mM dithiothreitol (DTT), 0.13 mg/ml BSA, 3% glycerol. Before mixing, Gcf1p was diluted in 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 0.1 mg/ml BSA, 10% glycerol. The gels were run 15 V/cm for 1 h at 4°C, dried and exposed overnight to a phosphorimager screen, scanned using a Typhoon Trio device (Amersham Biosciences), and quantified using ImageQuant TL toolset (Amersham Biosciences). Sequences of used DNAs are shown in Supplementary Table S1.

### Gcf1p-DNA20 crystallization and structure solution

Both native and SeMet-derivatized Gcf1p/DNA20 crystals were obtained at a protein:DNA ratio of 2:1.2, with a protein concentration of 8 mg/ml in 20% PEGLMW (Hampson Research), 0.1 M Na/K-phosphate buffer (pH 6.2), 0.2 M NaCl, 3% glycerol and 3% L-trehalose. Crystals were cryo-protected with 28% PEG-LMW, 6% glycerol, 0.12 M Na/K-phosphate (pH 6.2) and 0.24 M NaCl. X-ray diffraction datasets were collected at the XALOC beamline of the ALBA Synchrotron (Cerdanyola del Vallès, Spain). SeMet crystals were subjected to a fluorescence scan for data collection at the Se-edge (12.66 keV, or 0.979312 Å, Supplementary Table S2), with Friedel pairs collected within the

same frame for anomalous signal maximization. Datasets were processed with XDS (33) and aimless (34). Native and SeMet-derivatized crystals belonged to space group  $P2_12_12$  and were 99% isomorphous. A partial molecular replacement (MR) solution was found with Phaser-MR (35) using two ideal poly-ala  $\alpha$ -helices of 20 and 25 residues, together with two double-stranded DNA fragments corresponding to positions 11–19 from the Abf2p crystal structure (PDB 5JH0). This was fixed as a partial solution in MR-SAD with Phaser-EP (35). Four Se atom positions were found, related by a two-fold non-crystallographic symmetry (NCS) axis, thus indicating the presence of two copies of Gcf1p in the asymmetric unit (a.u.). The map was improved by density modification with Phenix-Resolve (36) which included solvent flattening and NCS averaging. Automated model building was carried out with Phenix-Autobuild (36) keeping the DNA as a ligand, followed by map inspection using COOT (37). Cycles of manual model building alternated with automatic refinement with Phenix. As ligands, two glycerols related by a two-fold NCS are found at a cavity formed by the N-terminal helix, HMG-box1 and the long HMG-box2 helix3. The interfaces and crystal contacts were analyzed with the PISA server (38).

### SEC-MALLS

We injected 100  $\mu$ l samples of Gcf1p, Gcf1p/DNA20, Gcf1p/DNA50 or Gcf1p\_CC/DNA50 (Supplementary Table S3) at a protein concentration of 2 mg/ml onto a Superdex 200 10/300 column (Cytiva). Complexes were prepared at protein:DNA ratios of 1:1.2 (Gcf1p/DNA20) and 2:1.2 (Gcf1p/DNA50 and Gcf1p\_CC/DNA50). Column equilibration and running buffer for isolated Gcf1p was 750 mM NaCl, 50 mM Tris-HCl pH 8.0, and 20 mM NaCl, 50 mM Tris-HCl pH 8.0 for the complexes. For both cases the flow rate was 0.5 ml/min at RT. The column was coupled to a MALLS system including a DAWN-HELEOS-II detector (Wyatt Technology) with a laser emission wavelength of 664.3 nm. Peak concentrations were measured with an Optilab T-rEX differential refractive index detector (Wyatt Technology) assuming a dn/dc of 0.185 ml/g. Molecular weights (MWs) and complex ratios were determined using conjugate calculations in ASTRA 6.0.5.3 (Wyatt Technology). We also measured absorbance at 280nm. Experiments were carried out at the Automated Crystallography Platform at the Barcelona Science Park.

### SAXS

Serial dilutions of Gcf1p at 10, 5, 2.5 and 1.25 mg/ml in 750 mM NaCl, 50 mM Tris-HCl pH 8.0, were analyzed at BM29 beamline of the ESRF synchrotron (Grenoble, France). Gcf1p and Gcf1p53-245 complexed with DNA20 or DNA50 were prepared at a 1:1.2 protein:DNA ratio followed by three-step dialysis as described above, due to the partial aggregation of the 2:1 ratio used for crystallization. Concentration series were prepared at 2.0, 1.0, 0.5 and 0.25 mg/ml. Free Gcf1p53–245 was prepared at 2.0 mg/ml and measured together with the complexes at

the P12 beamline (DESY, Hamburg, Germany). Data reduction, Guinier analysis, MW estimations and the calculation of pair-wise distance distribution functions ( $P(r)$ ) were carried out with PRIMUS (39). We extrapolated to zero concentration for the free Gcf1p and Gcf1p53-245/DNA20 samples, which did not show concentration-dependent differences. For these samples, data were merged from low and high  $q$  values measured at low and high concentration, respectively. The MW was obtained by the consensus Bayesian estimation of the molecular mass (40) (see Supplementary Table S4). The Gcf1p/DNA20 samples showed a concentration-dependent variation of the scattering curves, and further analysis of the Guinier region and overall curve suggested oligomerization. The Ensemble Optimization Method EOM (41) was used to model the flexible regions of free Gcf1p with data in the range  $0.2 < q < 5 \text{ nm}^{-1}$  and by assigning random coil structure to the N-terminal tail (residues 25–53), the inter-domain linkers and the extended regions of both HMG-boxes (residues 109–117 and 157–175), and the C-terminal tail of HMG-box2 (residues 238–245). Gcf1p53–245/DNA20 was modeled using CRY SOL (42) with the crystal structure as the starting conformation. For Gcf1p/DNA20, we generated different protein/DNA subunits combinations from the supra-assembly, and fitted the models with OLIGOMER (39). The volume fraction of each model was optimized to describe the experimental data. In this case, the theoretical curves used by OLIGOMER were computed with WAXSiS (43), which uses explicit-solvent all-atom MD simulations for the calculations. The radius of gyration  $R_g$  was estimated using the Flory equation, where  $N$  is the number of residues:  $R_g = 0.3 \cdot N^{0.33} \text{ nm}$ .

## TEM

Negatively supercoiled pUC191 (sc-pUC191, 2686bp) was extracted from *E. coli* cultures using the *QIAPrep Spin MiniPrep kit* (Qiagen). The plasmid was relaxed (r-pUC191) using topoisomerase I, which yielded a Gaussian distribution of topoisomers. The plasmid was linearized (l-pUC191) using the single-cutter SmaI. The three substrates were purified using the GFX PCR and gel band purification kit (Cytiva). In addition, a 1131 bp fragment of pBR322 between positions 2576 and 3707 (DNA1000) was amplified using Phusion DNA polymerase, and with forward and reverse primers 5'-CGACGCTCAAGTCAGAGG-3' and 5'-GATAACTGCGGCCACC-3'. The amplicon was purified on a MiniQ column (Cytiva) using a SMART chromatography system (Amersham Biosciences). DNA was then precipitated by adding 100% ethanol and 0.3 M sodium acetate (pH 4.5). The DNA pellet was washed three times in 70% ethanol and stored at  $-20^\circ\text{C}$  in 10 mM Tris (pH 7.5), 1 mM EDTA.

Serial dilutions of 20 mM Gcf1p were prepared in 750 mM NaCl, 50 mM Tris-HCl (pH 8.0) to find an ideal initial protein concentration for preparation of nucleoprotein complexes. The complexes were first incubated in 50 mM Tris-HCl (pH 8.0), 400 mM NaCl and 3% glycerol, followed by progressive exchange to the final incubation buffer 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 3% glycerol. After 30' at RT, the nucleoprotein complexes were purified

from free unbound protein by SEC, loading each sample onto a Superose 6 column (Cytiva) in incubation buffer at a flow rate of 0.1 ml/min. Eluted fractions were diluted 3-fold in incubation buffer and 5  $\mu\text{l}$  were immediately deposited on a 600-mesh copper grid coated with a thin carbon film, preactivated by glow-discharge in the presence of amylamine (Sigma-Aldrich) in a homemade device (ELC team). The grids were washed with 0.2% aqueous uranyl acetate (Merck) and dried with ashless filter paper (VWR) (44). TEM images were captured in annular or crystallographic dark-field mode (Zeiss 902 and 912 microscopes, respectively) using Veletta or Tengra high-resolution CCD cameras respectively and iTEM software (Olympus, Soft Imaging Solutions). We measured the length of the DNA fragments in the absence or presence of Gcf1p (50 nM), with iTEM software. These measurements were performed with 1131 bp linear fragments from pBR322 (from positions 2576–3707 'DNA 1000'), and with relaxed topo1-closed and supercoiled pUC19 plasmids. In the presence of Gcf1p different events coexisted: DNA kinks, bridging and compaction. However, the contour length of the latter could not be measured because of the impossibility to follow the DNA chain within the compacted DNA molecule. For the linear fragments of 1131 bp, we made series of 100 measurements in duplicate. For the experiments with pUC19, we measured series of 50 samples.

## Topology analysis

Negatively supercoiled plasmid DNA (4.5 kb, 0.35nM) was relaxed with topoisomerase I in 40  $\mu\text{l}$  of buffer R (50 mM Tris-HCl pH 7.9, 100 mM NaCl, 10 mM MgCl<sub>2</sub>, 100  $\mu\text{g}/\text{ml}$  BSA) in the presence of TFAM, Gcf1p or Gcf1p53-245 at the indicated concentrations (0, 20 or 100 nM). The resulting Lk distributions were resolved by 2D electrophoresis in 0.8% agarose gels at  $25^\circ\text{C}$ . The first dimension (top to bottom) separation was carried out using TBE buffer containing 0.6  $\mu\text{g}/\text{ml}$  chloroquine at 50 V for 14 h, and the second dimension (left to right) separation was carried out using TBE buffer containing 3  $\mu\text{g}/\text{ml}$  chloroquine at 60 V for 8 h in. The gel was blotted and probed for the plasmid DNA, and changes in Lk were calculated by counting the distance from the center of the Lk distribution in the presence and absence of bound protein. For the EMSA, 1  $\mu\text{g}$  of negatively supercoiled 4.5kb DNA plasmid relaxed with topoisomerase I was mixed with increasing amounts (0–2  $\mu\text{g}$ ) of Gcf1p in a final volume of 20  $\mu\text{l}$ . Following incubation at  $30^\circ\text{C}$  for 10 min, the sample was supplemented with 20% glycerol and analyzed by 0.8% gel electrophoresis at  $25^\circ\text{C}$  in TBE buffer containing 0.6  $\mu\text{g}/\text{ml}$  of chloroquine at 50 V for 14 h. The gel was stained with ethidium bromide and the shifted bands quantified using ImageJ. For DNA knotting analyses, a supercoiled 4.5 kb plasmid (YRp4) was enzymatically nicked with endonuclease BstNBI (NEB). 1  $\mu\text{g}$  of purified, nicked plasmid was incubated with increasing amounts (0–2  $\mu\text{g}$ ) of Gcf1p in 20  $\mu\text{l}$  of buffer R for 10 min at  $30^\circ\text{C}$ . The mixtures were supplemented with topoisomerase II and ATP before further incubation for 10 min at  $30^\circ\text{C}$ . The samples were analyzed by agarose gel electrophoresis as described above, and the bands stained with ethidium bromide to detect knotted DNA circles.

## MD simulations

Five systems were simulated: (i) 1:1 protein:DNA; (ii) 2:2 proteins:DNA with the DNAs parallel to each other in the pseudo-U-turn arrangement; (iii) 2 proteins:1 long DNA with the two proteins interacting via the coiled-coil; (iv) the supra-assembly by joining two asymmetric units of the X-ray crystal structure and (v) the supra-assembly with continuous DNA at the U-turn, by imposing a link between the 3' end of chain Y (Thy<sub>20</sub>) and the 5' end of chain X (Ade<sub>1</sub>), and the same between chain W and chain Z. The numbering of the sequence was modified accordingly. Starting geometries were derived from X-ray structures, but the DNAs were enlarged to avoid effects with DNA terminal bases. For complexes from the first three systems, the DNA sequence was extended using the construct 5'ATA TAT ATA TAT ATA AAATAATAATTATATAATATAAT ATA TAT ATA TAT AAT AAT AAA TTA T 3', where the underlined sequence is the one crystallized with Gcf1p, following the sequence from *S. cerevisiae* mtDNA (45). The starting structures with elongated DNA were built with PYMOL (<https://www.schrodinger.com/products/pymol>) with optimized parameters (46). The complexes were placed at the center of a truncated octahedral box of TIP3P (47) water molecules, neutralized by K<sup>+</sup> ions, resulting from KCl addition (48) to reach a salt concentration of 150 mM. Each system was minimized using a multi-step procedure (49), thermalized to 298 K (in the NVT ensemble), and then equilibrated and simulated without restraints collecting 500 ns for each simulation as previously described (49) (Supplementary methods for full procedures). We used the newly revised force field parmBSC1 for the DNA (50) and the amber14sb force field for the protein. Each simulation was extended for at least 500 ns of production time using AMBER 18, followed by analysis with the CPPTRAJ (51) (*B*-factor of the heavy atoms, which is the fluctuation of atomic positions, squared and weighted (r.m.s.f.)); and hydrogen bonds by using the 'hbond' function, in which only the interactions formed at least more than 30% of the time were considered as stable), Curves (52) and BIGNASIM (53) suite of programs for the base-pair parameters. The systems were visualized with VMD v1.9.4 (54) and Chimera (55).

## Miscellaneous

The alignment of Gcf1p HMG-box1 and HMG-box2 was based on the structures. The *Candida* sequences were aligned using BLAST and presented with ESPRIPT (56) after modification. The structures were represented with Chimera (55).

## RESULTS

### Analysis of Gcf1p binding to different DNAs and crystallization of the complex

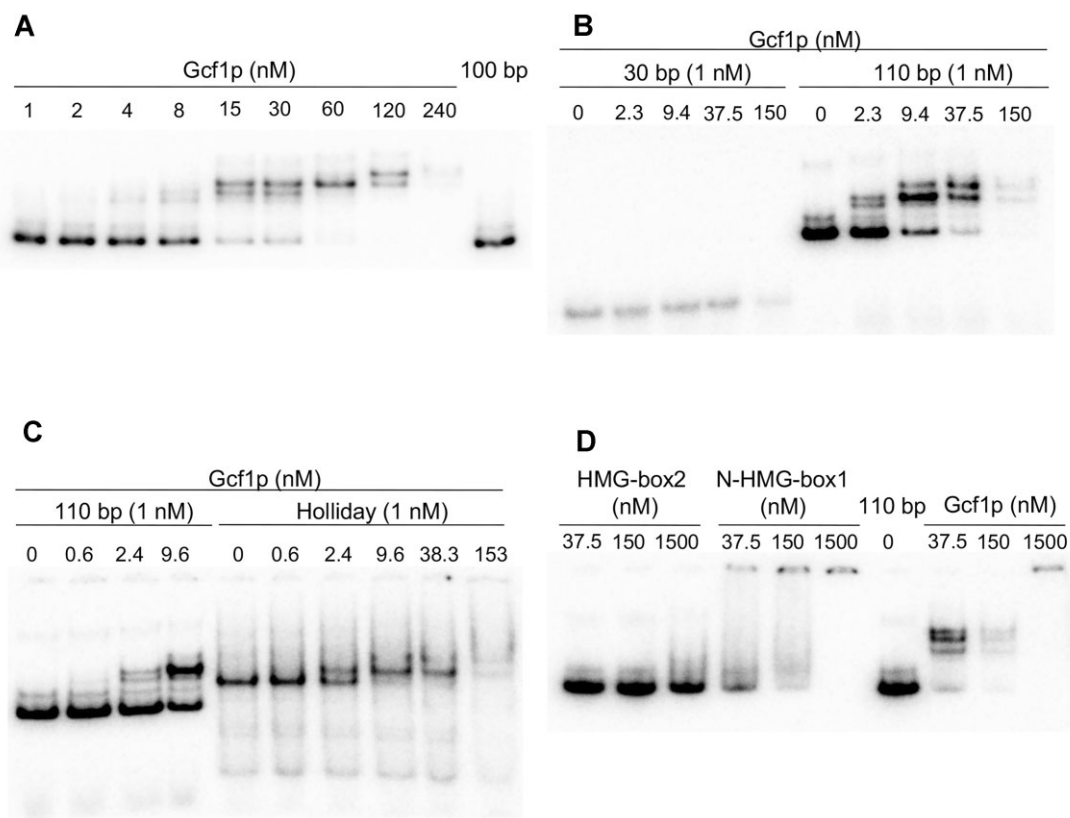
The ability of Gcf1p (UniProt ID Q59QB8-1, residues 25–245) to bind DNA was initially analyzed by EMSA. We incubated Gcf1p with labeled DNAs of 30–110 bp (DNA30–110) or with a Holliday junction (HJ) (Figure 1, Supplementary Table 1). Increasing the amount of Gcf1p shifted

DNA100 to progressively slower but well-defined bands, suggesting the presence of Gcf1p/DNA multimers of increasing size that eventually did not enter the gel probably due to aggregation (Figure 1A). In contrast, no shifts were observed for DNA30 (Figure 1B), which resembles earlier results with DNAs of 15 and 25 bp (23), whereas DNA110 under the same conditions formed progressively larger complexes, as also shown for DNAs of 50 bp (23). The HJ resulted in an initial shift that progressed to a second band and a smear (Figure 1C), as also previously shown (23). Next, we analyzed the ability of the different Gcf1p domains to bind the DNA by comparing constructs Gcfp25-155 (comprising the N-terminal flexible region, the N-terminal  $\alpha$ -helix and HMG-box1, thus N-HMG-box1), and Gcf1p151-245 (HMG-box2) with WT Gcf1p, by EMSA. Increasing amounts of Gcf1p progressively shifted DNA110 to two bands and, at higher concentrations, to a third upper band of aggregated protein that accumulated at the base of the loading well (Figure 1D). In contrast, the N-HMG-box1 fragment did not yield well-defined bands, but the free DNA became progressively fainter at higher protein concentrations. The aggregation of complexes probably impaired migration, as supported by the increasing signal at the base of the loading wells. Fragment HMG-box2 did not bind DNA efficiently, as indicated by the negligible change in intensity of free DNA at even very high protein concentrations. Similar results have been described for TFAM and Abf2p HMG-box2 domains, which show a much lower affinity for DNA than HMG-box1 (26,57,58).

Crystallization trials with HJ and DNA50 produced fragile, poorly diffracting crystals that we were unable to optimize. Other trials were based on a 22 bp sequence previously crystallized with Abf2p (Abf2p/Af22) (26). This DNA contained an adenosine tract whose rigid and narrow minor groove prevented binding of Abf2p, thus the protein bound the remaining sequence leading to a more homogeneous complex (26). Length variants of this sequence from (18–28 bp) formed profuse crystals but only those of DNA20 were suitable for X-ray diffraction and data collection.

### The crystal structure of the Gcf1p/DNA complex reveals non-canonical HMG-boxes

The crystal structure of Gcf1p/DNA20 features two proteins bound to two DNAs. The two Gcf1p molecules are related by a 2-fold axis (molecules molA and molB, RMSD 0.821 Å) and sandwich two DNA20 molecules (strand pairs XW and ZY). The DNA20 molecules make contacts at their ends, which are kinked (Figure 2A). Gcf1p comprises four domains, namely a non-traced disordered region (residues 25–56, which probably contribute to the high *B*-factor, Supplementary Table S2), a long N-terminal helix (57–104), HMG-box1 (106–156) and HMG-box2 (159–239) (Figure 2B). The structure resembles a hammer, in which the N-terminal helix is the handle and the HMG-boxes, at one end of the vertical axis, form the head (Figure 2B). The HMG-boxes show the classical three-helix L-shape. Helices 1 and 2 form a small helical bundle, and helix3 packs against an N-terminal elongated segment (Supplementary

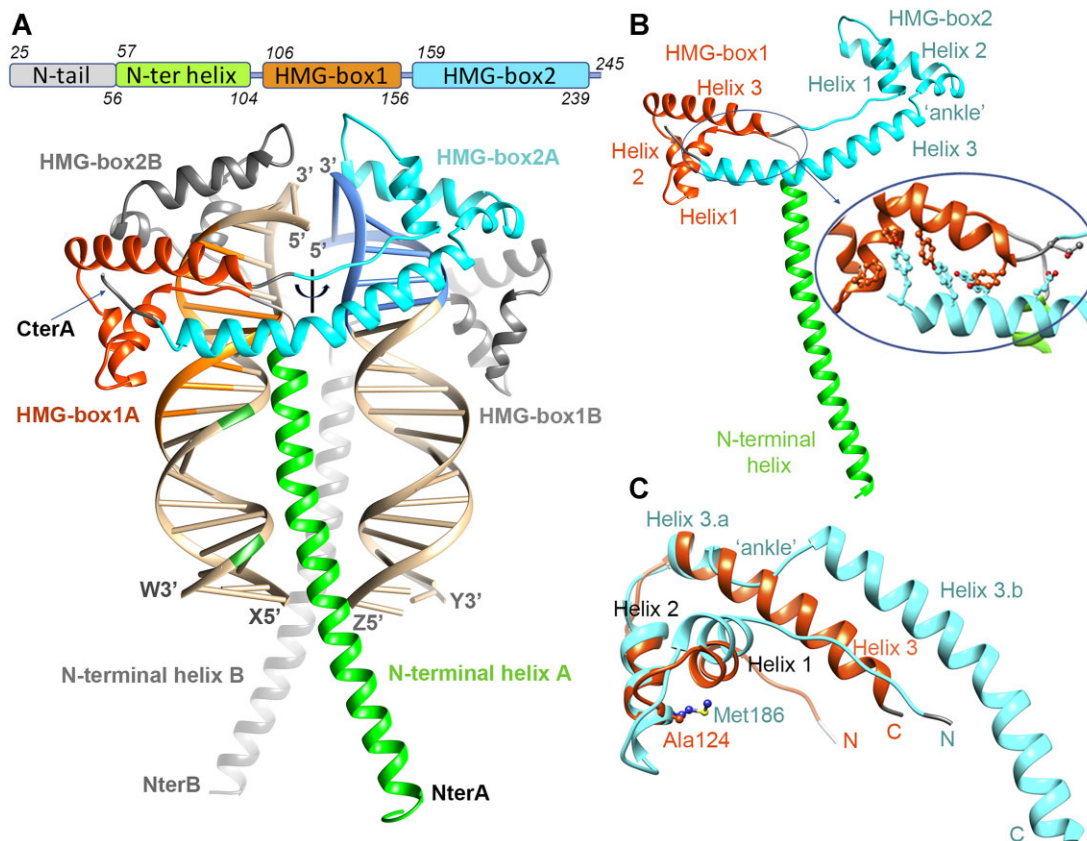


**Figure 1.** Electrophoretic mobility shift assays (EMSAs) showing the binding of different  $^{32}\text{P}$ -labelled DNAs to Gcf1p variants. (A) EMSA of  $^{32}\text{P}$ -labelled 100 bp DNA incubated with increasing concentrations (from left to right) of Gcf1p (indicated in nM). (B) EMSA of  $^{32}\text{P}$ -labelled 30 bp or 110 bp DNA as in (A). (C) EMSA of 110 bp or a 197 bp Holliday junction as in (A). (D) EMSA of a 110bp DNA binding to Gcf1p fragment 151–245 (HMG-box2), fragment 25–155 (N-terminal disordered region, N-terminal helix and HMG-box1: N-HMG1), and the mature 25–245 protein (Gcf1p). Lane ‘0’ contains free  $^{32}\text{P}$ -DNA. DNA concentration is 1 nM in all cases.

Figure S1A, B). The Gcf1p HMG-boxes possess non-canonical features. HMG-box1 (48 residues) is much smaller than any other known, which usually contain 75 residues (59). The superposition of the HMG-boxes shows that HMG-box1 helices 1 and 2 are much shorter than those of HMG-box2 (Figure 2C), but the residues of the small hydrophobic core between helices are conserved in both domains, and also in TFAM and Abf2p (Supplementary Figure S1A, B, D). This hydrophobic core between helix 1 and 2 progresses towards the interface between helix3 and the N-terminal elongated segment, in which also participate conserved residues (shown as stars in Supplementary Figure S1D), except in Gcf1p HMG-box1 in which the hydrophobic core is shifted to a region that contacts HMG-box2 helix3 (Figure 2B and Supplementary Figure S1C, D). As for HMG-box2, helix3 is interrupted at the highly-conserved Tyr204 residue by a Gcf1p-specific short  $\beta$ -strand (residues 204–206) that interacts with the N-terminal elongated region of this domain (Supplementary Figure S1B), creating an ‘ankle’ that interrupts both helix3 and HMG-box2 hydrophobic core (Figure 2C, Supplementary Figure S1B). Thereafter, HMG-box2 helix3 reaches the junction between the ‘hammer head’ and ‘handle’, and runs antiparallel along HMG-box1 helix3, with whom makes the aforementioned hydrophobic contacts that seal both domains (Figure 2B, Supplementary Figure S1C).

### HMG-box1 and HMG-box2 contact the major and minor grooves of independent DNAs

Surprisingly, the two HMG-box domains contact the DNA by using different mechanisms. None of them make polar interactions with base atoms, only with the backbone, which indicates no direct recognition of the DNA sequence (Figures 2A and 3; Supplementary Figure S2A). HMG-box2 from molA canonically binds and bends the 5' end of DNA ZY (HMG-box2 from molB imposes the same effect on DNA XW). The concave surface of the HMG-box L-shape contacts the minor groove, separates the strands and induces a sharp DNA bend by inserting Met186 (HMG-box2 helix 2) at base-pair step  $\text{T}_2\text{A}_3/\text{T}_{18}\text{A}_{19}$  (strands ZY), which disrupts the stacking and increases the roll angle to  $+60^\circ$  (Figure 3A, B, Supplementary Figure S2A). Met186 thus behaves as the DNA insertion wedge typically found in HMG-boxes (59) but surprisingly belongs to helix 2 (also unexpectedly found in TFAM HMG-box2 (24)) rather than helix 1 as typically observed in HMG-box domains (including TFAM HMG-box1 and Abf2p, Supplementary Figure S1D, green boxes). Phe168 and Ala169 (HMG-box2 helix 1) induce shearing to following base pair step ( $\text{A}_3\text{A}_4/\text{T}_{17}\text{T}_{18}$ ), the whole protein domain imposing a bend of  $90^\circ$  over the entire DNA region (Figure 3A; Supplementary Figure S2A). In contrast, HMG-box1 from molA contacts



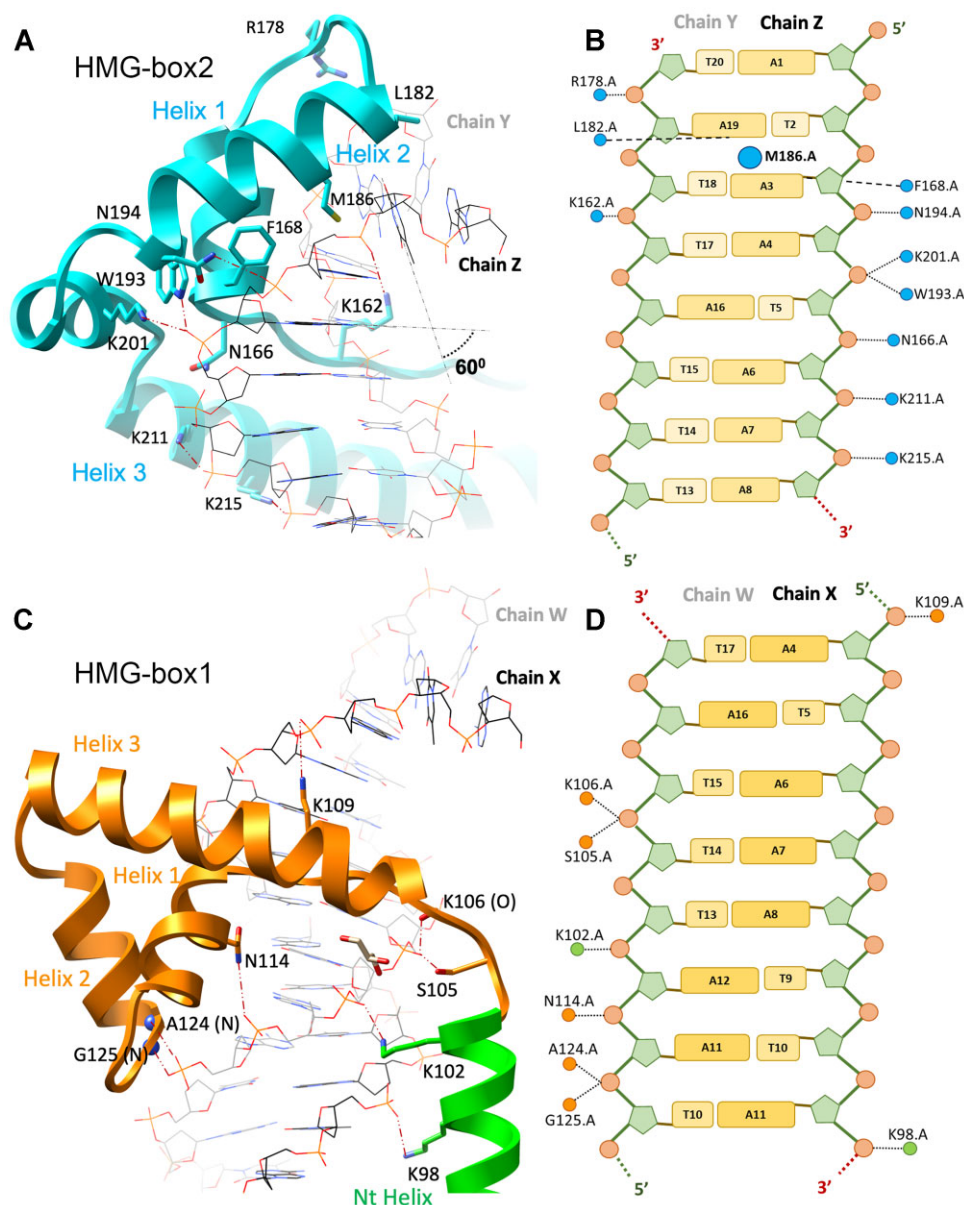
**Figure 2.** Overall structure of the Gcf1p/DNA20 complex. (A) The crystal structure of the Gcf1p/DNA complex shows two proteins (molecules molA, in color and molB in gray) sandwiching two DNAs (strands WX and YZ). The two proteins are related by a 2-fold axis (vertical black line with an arrow) as are the two DNAs. The N-terminal helix (in green), HMG-box1 (in red) and HMG-box2 (in blue) domains of Gcf1p chain A contact DNA regions in matching colors. Note the bending of the 3' and 5' DNA ends contacted by the HMG-box2 domain of either protein. Labels follow the same color code as the domains. The protein termini are indicated as N-ter and C-ter. The linear map of the domains is shown above. (B) Gcf1p chain A from panel (A) is shown in more detail. HMG-box1 and HMG-box2, colored as in panel (A), lay at each side of the N-terminal helix axis. The inset illustrates some of the residues involved in the hydrophobic contact between helix 3 from either box. (C) Superimposition of HMG-box1 (orange) and HMG-box2 (sky blue). Ala124 and Met186 are indicated in the corresponding colors. Helices 1, 2, 3 and the ankle region of HMG-box2 are indicated. The protein N- and C-termini are indicated as N and C.

the major groove of DNA molecule XW and barely distorts the B-DNA conformation (the same occurs between molB and strands ZY, Figure 3C; Supplementary Figure S2A). The contacted DNA region includes the symmetric A-tract (A<sub>6</sub>)A<sub>7</sub>A<sub>8</sub>T<sub>9</sub>T<sub>10</sub>, which is rather rigid, as we previously showed (see Discussion and (26)). Binding to the major groove is an unprecedented interaction for an HMG-box. Specifically, two regions of HMG-box1 establish electrostatic contacts with three DNA sites, as a clamp (Figure 3C, D; Supplementary Figure S2A): The N-terminal elongated region fits into the major groove, in which Ser105 O<sup>γ</sup> and Lys106 C = O contact Thy<sub>14</sub> phosphate oxygens from strand W (site 1), whereas Lys109 extends towards the other edge of the major groove, and its N<sup>ε</sup> contacts the Ade<sub>4</sub> phosphate from strand X at next DNA turn (site 2). Thereafter, the polypeptide chain snugs along strand W and the Ala124 and Gly125 amides, at the tip of the HMG-box1, contact strand W Ade<sub>11</sub> phosphate group (site 3). This contact is further stabilized by the end of Gcf1p N-terminal helix, Lys98 N<sup>ε</sup> contacts strand X Thy<sub>12</sub> phosphate, whereas Lys102 brings its positive N<sup>ε</sup> into the minor groove between X Thy<sub>12</sub> and W Thy<sub>13</sub> phosphates (Figure 3B). In molB,

the overall arrangement is the same but some contacts vary due to occasional reorientation of side chains, indicating the slight flexibility of the interaction. No residues from the concave surface of HMG-box1 contact the DNA bases. Indeed, at the equivalent position of the inserted residue in HMG-box2 (Met186), HMG-box1 carries residue Ala124, whose small side chain cannot insert between base pairs (Figure 2C, Supplementary Figure S2A).

Given the twofold axis, HMG-box domains 1 and 2 from Gcf1p molecule B likewise contact DNA chains ZY and XW, respectively (Figure 2A). Thus, four HMG-boxes completely cover base pairs 1–11 of either DNA molecule (22 bp in total), like a cap (Figure 2A, Supplementary Figure S2B). The strong electronegative charge of the DNA phosphate backbone is compensated by positively charged side chains (Supplementary Figure S2C). At the top of the *cap*, the loops at the tip of the A and B HMG-box2 domains show different conformations and interact weakly with each other (Figure 2A, Supplementary Figure S2B). Underneath, the DNA ends are stacked in a distorted manner, in which Thy<sub>20</sub> (strand Y) interacts with the pair A<sub>1</sub>/T<sub>20</sub> (strands X and W, respective). The two DNAs are quite





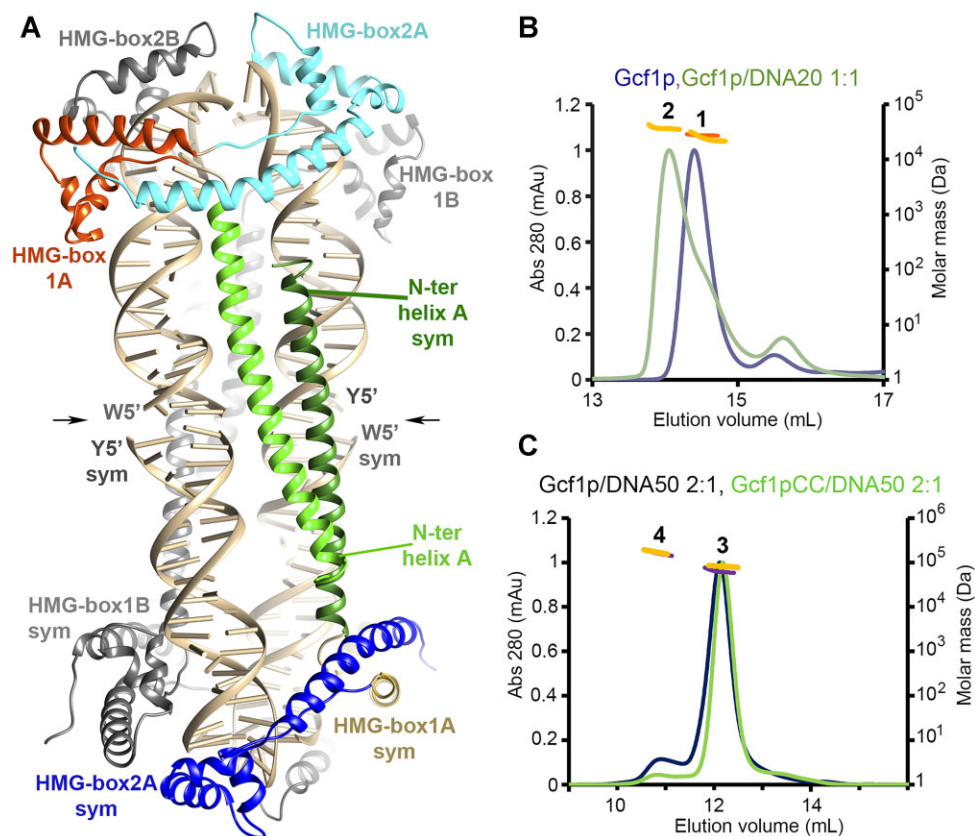
**Figure 3.** Protein-DNA contacts in the Gcf1p/DNA20 complex. (A) Contact of HMG-box2 with DNA molecule ZY. The angle between base pairs of the inserted DNA step, or the overall bending angle are indicated. (B) Schematic representation, ZY DNA contacts with HMG-box2 residues (in blue), in (A). The central circle represents the inserting residue M186. (C) Contact of HMG-box1 with DNA molecule XW. Note that Lys109 (K109) belongs to the N-terminal helix (Nt helix). Main chain atoms (N, O, from amides and carbonyl) are indicated in brackets. (D) Schematic representation of the DNA contacted by HMG-box1 residues (in orange). Colors are the same as in Figure 2A.

compressed, with Met186 insertion sites only 4 bp apart, in contrast to the 11 bp (one DNA-turn) in Abf2p complexes, and the 12 bp in TFAM complex. Therefore, the two Gcf1p molecules induce a strongly distorted DNA U-turn, in which the DNA ends are brought into contact by imperfect stacking.

#### The N-terminal helix creates a gcf1p/DNA20 supra-assembly in the crystal and in solution

Outside the cap-covered region, the two DNA helices (XW and ZY) extend parallel to the N-terminal  $\alpha$ -helix towards a symmetry mate (Figure 2A), with which the  $\alpha$ -helix forms

an antiparallel coiled-coil by burying valines, isoleucines, leucines, and the methylene groups of lysine residues (Figure 4A, Supplementary Figure S2). In this way, the two proteins involved in the distorted U-turn contact a symmetry pair, giving rise to a protein/DNA super-assembly. The coiled-coil is the most stable interface in the crystal ( $\Delta G = -18.5$  kcal/mol, probability of complex formation = 1, according to PISA (38)), and the high conservation of the N-terminal  $\alpha$ -helix in *Candida* species suggests similar coiled-coil interactions (Supplementary Figure S3). The XW and ZY DNA ends also interact with symmetry partners, but the grooves are not continuous between crystallographic unit cells (arrows in Figure 4A). Therefore, symmetry



**Figure 4.** Assembly of symmetry mates in the crystal through the N-terminal  $\alpha$ -helix coiled-coil. (A) Arrangement of the two symmetrically related complexes mediated by the coiled-coil between the N-terminal  $\alpha$ -helices. The upper complex corresponds to that in Figure 2A (same color code). The lower complex (sym) corresponds to the symmetry partner. The buried area is 29 180  $\text{\AA}^2$  from a total surface of 67 630  $\text{\AA}^2$ , and the free-energy gain during complex formation is  $\Delta G^{\text{int}} = -140.1$  kcal/mol according to PISA, see Materials and Methods). Arrows indicate the contact between DNA symmetry partners, where the minor groove interacts with the symmetrical major groove. (B) Estimation of the absolute molecular weight (MW) of Free Gcf1p (10 mg/ml, blue curve) eluted as peak 1, as shown by the corresponding MW (deep orange, peak Pk1 WT in Supplementary Table 3). Gcf1p/DNA20 complex prepared at a 1:1 ratio (protein = 4 mg/ml, green curve) eluted as a major peak 2 and a smaller peak 1 (shoulder), corresponding to the 1:1 complex and free Gcf1p, respectively (corresponding MWs in light orange, Pk1 and Pk2 WT + 20 in Supplementary Table 3). The third smaller peak represents an impurity. (C) Estimation of the MW of Gcf1p/DNA50 complex prepared at a 2:1 ratio (black curve) eluted as peak 3 corresponding to a 1:1 complex, and peak 4 corresponding to a 4:2 complex (MWs shown in magenta, Pk3 and Pk4 WT + 50 in Supplementary Table 3). Mutation of residues at the N-terminal coiled coil resulted in a smaller peak 4 (green curve, MWs in orange, CC + 50 in Supplementary Table 3). Note that both the wild-type and coiled-coil mutant eluted identically as peak 3 in a 1:1 complex.

partners form stable interactions through the N-terminal coiled-coil, but *via* an irrelevant contact between DNA ends.

To confirm the oligomerization of Gcf1p/DNA complex in solution, we measured the absolute MW by SEC-MALLS. Free Gcf1p eluted as a monomer ( $27.50 \pm 0.05$  kDa, theoretical MW = 25.9 kDa, peak 1 in Figure 4B; Supplementary Table 3). Gcf1p/DNA20 complexes were prepared at protein:DNA ratio of 1:1, which resulted in single Gcf1p/DNA complexes (41.6 kDa, theoretical MW = 38 kDa for peak 2 in Figure 4B; Supplementary Table 3), but not the dimers or tetramers suggested by the crystal. The discontinuous DNA between crystalline symmetry mates may have destabilized the multimerization of complexes in solution, so we also tested DNA50 (the shortest reported DNA shifted by Gcf1p in EMSA experiments (23)) which could represent a continuous DNA in a coiled-coil dimer. At a 1:1 ratio, SEC-MALLS showed a major peak corresponding to a 1:1 complex (56.8 kDa for peak 3 in Supplementary Figure S4 versus theoretical 50.9 kDa; Supplementary Table 3) preceded by a minor peak consistent with

a complex of four proteins and two DNAs (162.4 kDa for peak 4 versus theoretical 153.6 kDa), consistent with the supra-assembly induced by the coiled-coil. We then generated a disruptive mutant (Gcf1pCC) with substitutions that weakened van der Waals interactions or induced repulsions between the N-terminal helices (A75R, L79A, I83R and L93K). Gcf1p/DNA50 and Gcf1pCC/DNA50 prepared at a 2:1 ratio (as used for crystallization), and under identical measurement conditions, showed the 1:1 complex (peak 3 in Figure 4C; Supplementary Table 3) preceded by the 4:2 peak (peak 4). Remarkably, however, the height of Gcf1pCC peak 4 was only 40% of that of WT Gcf1p, and it was much flatter, suggesting a mixture of stoichiometries and highlighting the importance of the coiled-coil in Gcf1p multimerization.

#### Gcf1p/DNA complexes form dynamic heterogeneous multimers in solution

We then analyzed the structure of Gcf1p alone or in a complex with DNA in solution by SAXS. Free Gcf1p measured

at different concentrations (1.25–10 mg/ml) was present as a monomer, consistent with the SEC-MALLS data (Figure 5A, Guinier plot in Supplementary Figure S4, Supplementary Table 4). The conformation was extended ( $R_g = 3.76$  nm versus theoretical 1.8 nm), and highly flexible (flat Kratky profile in Supplementary Figure S4). The coexistence of different conformations was confirmed by a wide distribution of atom pairs distances,  $P(r)$  (Supplementary Figure S4). Flexibility was not restricted to the N-terminal tail, given that the deletion mutant Gcf1p53–245 showed a similar behavior (Supplementary Figure S4). To model Gcf1p in solution, we used the EOM (41) to generate an initial pool of 10 000 models with different conformations based on the crystal structure. The EOM selected a subensemble of four molecules with important variations in the relative orientation of domains, whose averaged theoretical curve fitted with the experimental data ( $\chi^2 = 1.033$ ), confirming the wide conformational space sampled by Gcf1p (Figure 5A).

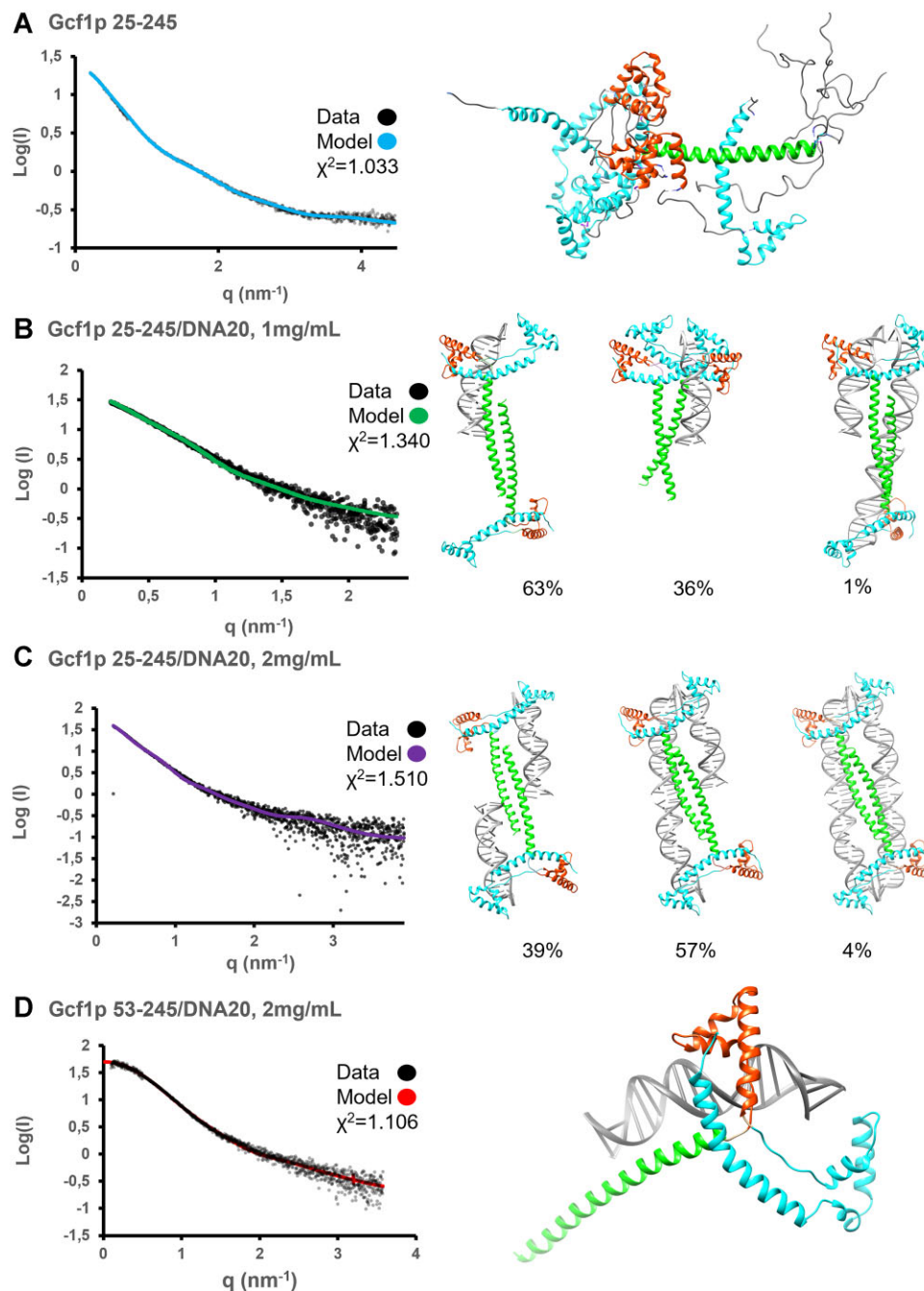
The Gcf1p/DNA20 complex (0.25–2.0 mg/ml, Figure 5B) was slightly less flexible and more globular and compact than the free protein ( $R_g$  decreased from 3.75 to 3.58 nm, Supplementary Table 4; bell-shaped Kratky profile and slightly narrower  $P(r)$  plots in Supplementary Figure S5) and, surprisingly, showed a concentration-dependent increase in the MW, suggesting complex multimerization (Supplementary Figure S5; Supplementary Table 4). To model the Gcf1p/DNA20 data, we used all possible protein/DNA subunits combinations that appeared in the crystal, from 1:1 complex to the supra-assembly. The best fit to the 1.0 mg/ml data ( $\chi^2 = 1.340$ ; Figure 5B; Supplementary Table 4) combined the coiled-coil dimer (63% volume fraction) and the dimer from the sandwiched pseudo-U-turn (36% volume fraction) (Figure 5B), together with a third minor species (1%) of the coiled-coil dimer bound to two additional DNA20 molecules. At 2.0 mg/ml, all models featured only the coiled-coil dimer bound to two or more DNA molecules ( $\chi^2 = 1.510$ ; Figure 5C; Supplementary Table 4). Surprisingly, mutant Gcf1p53–245/DNA20 showed a stable MW at all concentrations in a 1:1 ratio (Supplementary Figure S5; Supplementary Table 4). The latter was fitted using a model directly derived from the X-ray structure, which showed the best fit was just a 1:1 complex of DNA20 bound to HMG-box1 ( $\chi^2 = 1.106$ , at 2 mg/ml; Figure 5D).

These data show that free Gcf1p is characterized by intrinsic flexibility and an extended conformation but this is limited by DNA binding. Full-length Gcf1p/DNA20 complexes formed multimers in solution via the coiled-coil, in agreement with our SEC-MALLS and substantiating the crystal structure, which generated suitable protein/DNA models. The complexes tended to form a heterogeneous sample featuring multiple protein/DNA stoichiometries, involving more HMG-box domains at higher concentrations, indicating a highly dynamic interaction. In contrast, multimerization was prevented by the deletion of the disordered N-terminal tail (33% lysine content  $pI = 10.1$ ), suggesting it recruited Gcf1p/DNA complexes by electrostatic interactions with the DNA.

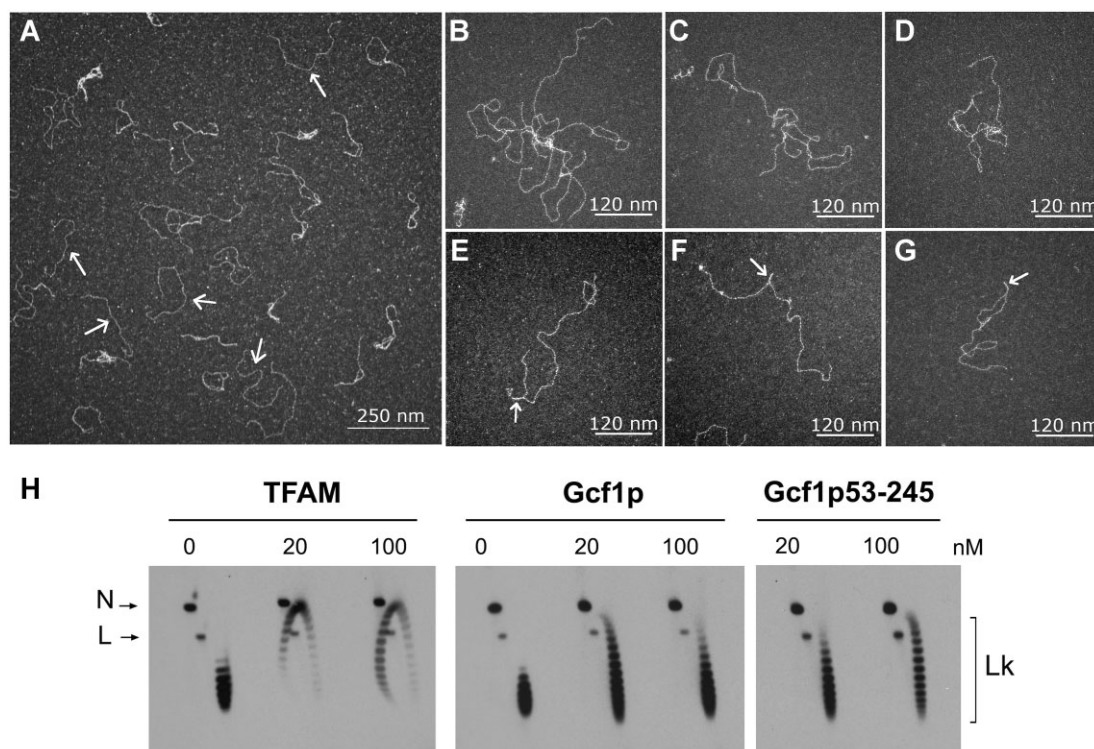
### Gcf1p polymerizes, bridges and compacts the DNA without modifying DNA topology

To determine how Gcf1p binds to long DNA molecules, we analyzed its formation of complexes with linear, relaxed and supercoiled DNA substrates > 1000 bp. TEM showed that Gcf1p distorts the DNA locally (compare Figure 6A with B–G, and Supplementary Figure S6), and formed extensive oligomers while other DNA molecules remained naked, indicating cooperative binding (Figure 6; Supplementary Figure S6). Various configurations were observed, including Gcf1p polymerization along the DNA molecule, or intramolecular and intermolecular DNA bridging, which are events that lead to DNA folding and local DNA condensation. DNA kinking, protein polymerization and DNA bridging were induced by the protein rather than spontaneous DNA supercoiling, given that they occurred on linear and relaxed closed circular DNAs (Figure 6; Supplementary Figure S6). Local sharp kinks and tight DNA hairpins were found among these bridging configurations (white arrows in Figure 6E–G; Supplementary Figure S6), like those previously observed for the *E. coli* histone-like nucleoid structuring protein (H-NS) (60) but differing from the DNA loops observed in TFAM and Abf2p (61,62). DNA bridging thus led to various levels of DNA folding and condensation (Figure 6B–D Supplementary Figure S6). Interestingly, more regular bridging was observed on negatively supercoiled DNA, in which the DNA topography itself favors oriented assembly and hence alignment (Supplementary Figure S6), as previously reported for H-NS (63) and transcription factor LrpC (64). The N-terminal deletion mutant Gcf1p53–245 likewise showed cooperative binding, DNA bridging and tight hairpins (Supplementary Figures S6 and S7), indicating that this region did not stabilize the DNA distortions. The TEM micrographs did not show relevant changes in DNA length nor DNA wrapping compared to the free molecules (for quantitative measurements, see Supplementary Figures S8, S9), in contrast with LrpC (64) and TFAM (65), suggesting that Gcf1p has a different effect on DNA topology. However, regarding highly compacted molecules, these results should be taken with caution given that the contour length cannot be measured in these specimens.

The changes in DNA topology were studied in more detail by 2D electrophoresis to determine whether Gcf1p changed Lk. The effect of TFAM on DNA topology has been studied widely (66,67), so we compared TFAM and Gcf1p. We relaxed a 4.5 kb plasmid with topoisomerase I in the presence of each protein, and visualized the resulting distributions of Lk topoisomers (Figure 6). TFAM strongly reduced the Lk, because each TFAM:DNA complex constricted approximately  $-0.2$  units of Lk. In contrast, Gcf1p had a negligible effect on the Lk ( $-0.05$  units). EMSA confirmed that Gcf1p bound to the DNA cooperatively, as denoted by the sigmoidal curve of shifted DNA at increasing protein concentrations (Supplementary Figure S7). Therefore, in sharp contrast to TFAM, Gcf1p did not significantly deform the DNA. This was in accordance with the EM images, in which there was no evidence of DNA shortening or extension. Yet, a negligible change in



**Figure 5.** Molecular modelling of Gcf1p in solution by SAXS, unbound and in a complex with DNA. (A) Experimental scattering curve of free Gcf1p represented on a logarithmic scale as a function of the momentum transfer,  $q = 4\pi \sin(\theta) / \lambda - 1$  ( $2\theta$ , scattering angle;  $\lambda = 1.5$  Å, X-ray wavelength). The fitted ensemble optimization method curve (in blue) describes the complete  $q$ -range by the subset of four models shown on the right, which are superposed by the N-terminal helix. The domain colors match those in Figure 2A. (B) SAXS curve of Gcf1p25–245/DNA20 at 1 mg/ml with the fitted theoretical averaged curve corresponding to the three models (right panel) selected by Oligomer (see Methods section). (C) SAXS curve of Gcf1p23–245/DNA20 at 2.0 mg/ml, otherwise as in (B). (D) SAXS curve of Gcf1p53–245/DNA20 with the fitted theoretical curve (in red) of the crystallographic model (on the right). In all cases SAXS data are represented as black dots, and the quality of fitting ( $\chi$ ) is indicated.



**Figure 6.** Characterization of Gcf1p/DNA interactions by TEM and DNA topology analysis. (A) The incubation of linear DNA (1000 bp) with Gcf1p showed molecules covered with the protein and coexisting bare DNA molecules (white arrows), which is consistent with cooperative binding. (B–D) Examples of intermolecular bridging of a linearized pUC19 plasmid, induced by Gcf1p. (E–G) Examples of intramolecular DNA bridging induced by Gcf1p (white arrows). (H) Linking number (Lk) distributions resulting from topoisomerase I relaxation of a 4.5 kb DNA plasmid incubated with increasing concentrations (nM) of TFAM, Gcf1p, and Gcf1p53–245 and resolved by 2D agarose gel electrophoresis. The position of nicked circles (N), linear molecules (L) and the arch of Lk topoisomers are indicated.

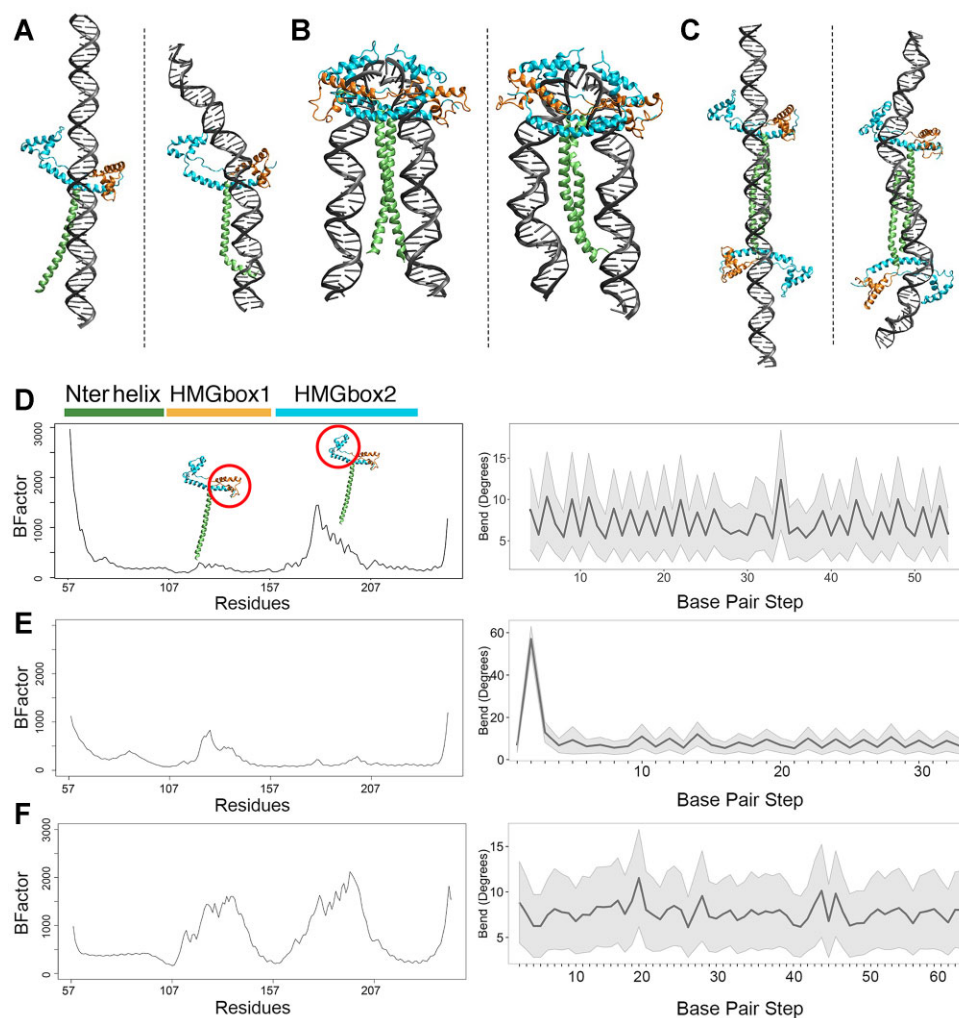
Lk could be due to changes in wrapping ( $W_r$ ) compensated by changes in unwinding (or twist,  $T_w$ ). Interestingly, in the crystal structure, the two pairs of HMG-boxes at each end of the supra-assembly are rotated by  $\sim 70^\circ$  with respect to each other. Thus, the DNAs axes along the N-terminal helix coiled coil are not totally parallel on a plane, but deviated by  $\sim 25^\circ$  (Supplementary Figure S7), creating a slightly positive  $W_r$ . This is compensated by a negative  $T_w$  induced by the four HMG-box2 domains, which unwind and bend the DNA so that the DNA ends join and create the pseudo-continuous, biologically irrelevant, crystallographic DNA circle (Supplementary Figure S7). Therefore, our data shows that  $W_r$ - $T_w$  compensating events could occur during DNA compaction, not necessarily following the pattern in the crystal but yielding a constant Lk (see Discussion). These results were further confirmed by MD (see below). In any case, the observed DNA compaction appears as DNA folding without major topological changes. Likewise, Gcf1p53-245 did not change the Lk distribution (Figure 6), further confirming that the N-terminal disordered region has a negligible effect on DNA topology.

To further confirm that Gcf1p bridges intramolecular segments of duplex DNA, we mixed it with a relaxed, nicked plasmid and added topoisomerase II and ATP. Under these conditions, low ratios of Gcf1p:DNA increased the DNA knotting activity of topoisomerase II, which indicated DNA looping by Gcf1p (Supplementary Figure S7).

Remarkably, there was no further increase in DNA knotting at higher Gcf1p:DNA ratios, consistent with the cooperative polymerization of the protein along the pre-formed loops.

### Molecular dynamics of Gcf1p/DNA complexes

To determine the mechanism of the Gcf1p/DNA interaction we first ran MD simulations of three models based on the crystal structure, at different Gcf1p:DNA ratios, namely 1:1, 2:1 and 2:2 (Figure 7A–C). The first model was a 1:1 complex with a 56 bps B-DNA (Figure 7A, left panel). All-atom MD simulation showed that the initially straight B-DNA segment was bent to increase the number of interactions with the protein *via* stable hydrogen bonds (present in  $> 30\%$  of the simulation time; residues shown in Supplementary Figure S10; H-bonds listed in Supplementary Table S5). Meanwhile, the protein underwent a remarkable change in the conformation of its main and side chains, especially HMG-box2 (reflected in the B-factors, Figure 7D, left panel). The N-terminal helix kinked and inserted its N-terminus into the major groove close to one end of the DNA (Figure 7A, right panel). In the central DNA region, HMG-box1 remained stably bound to the major groove as a belt. Close to the other DNA end, HMG-box2 was initially not contacting the DNA. During the MD simulation, this domain showed high mobility (Supplementary Figure S11) and occasionally reached the DNA minor



**Figure 7.** Dynamic behavior of Gcf1p on DNA at different protein:DNA ratios. (A) Gcf1p in a 1:1 complex with a 45 bp DNA at the beginning (left) and end (right) of the simulation, separated by a dashed line. The coloring of the domains is as in Figure 2. (B) Gcf1p in a 2:2 complex with 35 bp DNAs, resulting in a pseudo-U-turn arrangement (left and right figures as in panel A). Note the formation of a parallel coiled coil between the two DNAs. (C) Complex of the coiled-coil dimer with a 69 bp DNA at a ratio 2:1 (left and right figures as in panel A). (D) The left graph shows the B-factors of the protein backbone of panel (A), reflecting the displacement of the atoms along the simulation. Above, schematic representation of the protein domains colored as in Figure 2. The structures on the top of the peaks encircle the HMGbox1 (left) or HMGbox2 (right), which correspond to the residues numbered along the X-axis. Note the longer displacements in HMG-box2. The right graph shows the average bend (black line), in degrees, of the DNA along the simulation, with standard deviation shown in gray. (E) *B*-factors and average bend of the pseudo-U-turn in (B) along the simulation, represented as in (D). (F) *B*-factors and average bend of the coiled-coil dimer in (C), represented as in (D).

groove (Figure 7A, left panel). Remarkably, at the end of the simulation, both HMG-box1 and HMG-box2 showed binding to the DNA major and minor grooves, respectively, as in the crystal structure (Supplementary Figure S12). The complex was highly dynamic, the DNA bends reaching  $\sim 20^\circ$  at the HMG-box1 binding site (major peak in Figure 7D, right panel). In this swing, not all segments behaved equally: the loops between the domains showed lower mobility than the domains that waved to reach the DNA (compare B-factor values of the different regions in Figure 7D, left panel). The protein/DNA complex remained stable, and major variations were detected at the backbone of the HMG boxes, specifically HMG-box1 helices 2 and 3, and HMG-box2 helix 2, which increased the number of interactions with the DNA grooves (Supplementary Figure S10 and Supplementary Table S5).

The second model was a 2:2 distorted U-turn complex, as in the crystal, with the DNAs enlarged to 35 bps (Figure 7B). During the MD simulation, the two proteins remained attached to the corresponding side of the DNA U-turn, in a highly stable complex. Only small local rearrangements occurred between the DNA ends within the U-turn, which resulted in  $\pi$ - $\pi$  stacking interactions between bases but this did not create a continuous B-DNA-like stacking (Supplementary Figure S12). The greatest variation in DNA bending occurred at the inserted DNA step (Figure 7E). These local changes increased the number of interactions with the protein (Supplementary Figures S10, S12, and Supplementary Table S5), especially HMG-box1, which showed more rearrangements without detaching from the major groove (Figure 7B, Supplementary Figure S12). In contrast, the N-terminal helices of either monomer approached to each

other and formed a stable parallel coiled-coil that became dominant (Supplementary Figure S13), and improved the interaction energy by forming 33 new interactions between the two helices. Along the whole simulation, the most stable contacts with the DNA included the N-terminal helix with the DNA parallel to the ‘new’ coiled-coil, HMG-box1 helix1 and HMG-box2 helix1.

The third system was a 2:1 complex in which a long B-DNA (69-bp, straight conformation) was aligned along the main axis of a crystallographic coiled-coil dimer (Figure 7C). In this arrangement, the DNA is located between HMG-box1 and 2 of either protein. This starting model did not implicitly contain the interactions from the crystal, given that the DNA was not in contact with the HMG-boxes. During the MD simulation, the DNA and proteins approached each other, and after 500 ns HMG-box1 and HMG-box2 interacted with the major and minor grooves, respectively, similar to the X-ray structure, which involved important displacements as reflected by the high B-factors (Figure 7F, left panel; Supplementary Figure S12). As for the 1:1 complex above, the secondary structure of neither proteins showed major deformation, with only HMG-box2 helix2 needing to adapt to make better contact with the DNA minor groove. This induced the major deformation of the DNA along the MD (see bending at DNA steps close to 20, and 40–50, due to contacts with HMG-box2 from A and B molecules, respectively, in Figure 7F, right panel). Once the protein/DNA contact occurred, the DNA maintained contact with the HMG-boxes but mainly with the coiled-coil.

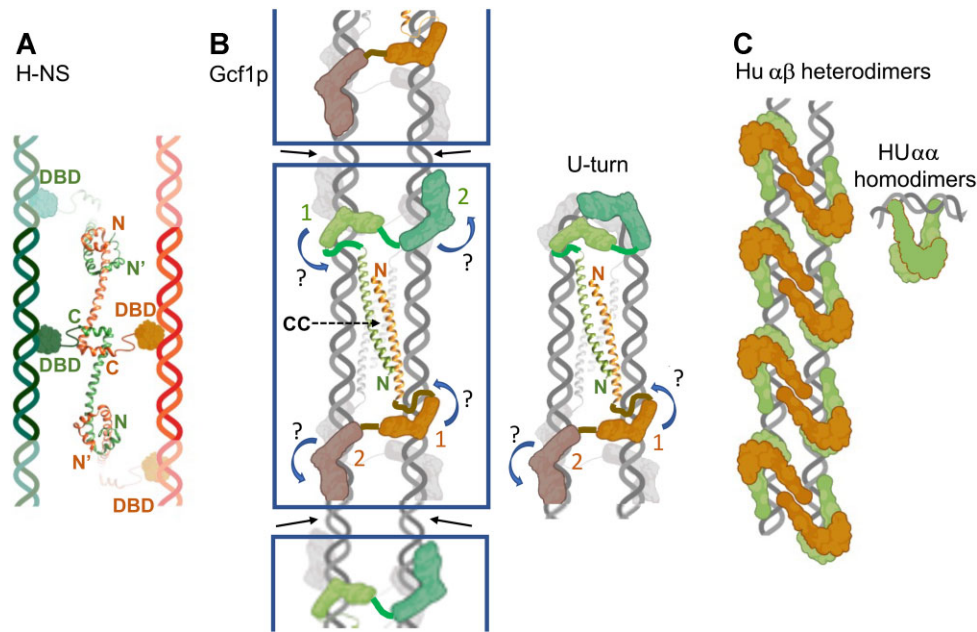
Finally, we run a simulation of the 4:4 supra-assembly and, even at extended simulations (800 ns), this model showed only very minor local rearrangements, mainly improving the interactions between the terminal bases of the DNA ends from the interrupted U-turn (Supplementary Figure S13). However, given that, at the U-turn the DNA, is discontinuous, we wondered if the tensions due to the tight U-bend had been dissipated by the free DNA ends. Therefore, we tested whether the U-turn was stable with a continuous DNA or, on the contrary, strong tensions would break stacking and Watson-Crick interactions, and/or generate other distortions. Thus, we imposed a 3'-5' link between chains Y (Thy<sub>20</sub>) and X (Ade<sub>1</sub>), renumbered the sequence (the same for chains W and Z), and run another simulation of 500 ns. The 4:4 supra-assembly remained stable and the DNA did not reveal major changes nor topological variations (Supplementary Figure S13), suggesting that (i) a tight U-turn induced by simultaneous binding of two HMG-box-2 domains is plausible and (ii) the DNA wrapping was rather an intrinsic feature of this complex than just due to crystal packing (Supplementary Figure S13). We also wondered if a continuous straight DNA extended from the DNA ends at the U-turn could be simulated. However, this entailed manual manipulations based on highly arbitrary decisions. As an alternative, we searched a timepoint during the simulation showing reorientation of the DNA ends to be able to extend them, but such a remodeling did not occur, further substantiating the stability of the U-turn during the calculations.

In conclusion, the MD simulations of the complexes confirmed that the HMG-boxes 1 and 2 have different site

preferences for DNA interactions, specifically binding at the major and minor grooves, respectively, confirming the recognition mode found in the X-ray structure. The N-terminal helix formed stable coils and, together with HMG-box residues, made stable contacts with the DNA. By virtue of the protein flexibility and the different modes of recognition by the different domains, the protein showed great ability to bend the DNA. Yet, the direction of DNA bending differed due to the different features of starting models, such as different DNA conformations (U-turn or linear), and the number of proteins (one or two) bound in different relative orientations.

## DISCUSSION

The crystal structure of Gcf1p/DNA shows that one Gcf1p molecule binds two independent DNAs via its HMG-boxes. HMG-box2 canonically binds to the minor groove, separates both strands and inserts a residue between base pairs and thus induces a sharp bend. It should be noted that we used the same DNA sequence as for the Abf2p crystal structure and, remarkably in both cases the inserted step is T<sub>2</sub>A<sub>3</sub>/T<sub>18</sub>A<sub>19</sub> (26). This step is followed by a symmetric adenine tract (A<sub>6</sub>)A<sub>7</sub>A<sub>8</sub>T<sub>9</sub>T<sub>10</sub>, a rather rigid and difficult to distort sequence with a narrow minor groove (26). Crystal structures of Abf2p in complex with this and with a second sequence containing an asymmetric A-tract (AAAA) showed that the Abf2p HMG-boxes avoided binding A-tracts (26). In Gcf1p, the small HMG-box1 fits into the major groove of the linear and rigid A-tract, whereas canonical HMG-box2 distorts the bendable region of the DNA molecule. In both cases, no contacts with base-specific atoms are observed, supporting the notion that Gcf1p does not recognize the DNA sequence. Instead, during crystallization, Gcf1p recognized the structural DNA features. Such a non-sequence specific recognition is consistent with the architectural role of proteins with HMG-boxes in tandem (59). On the other hand, despite both Abf2p and Gcf1p structures sharing the DNA sequence and mechanism of bending (insertion at the same step), the binding is protein-dependent. Whereas Gcf1p induces a U-turn with the HMG-boxes2 from two proteins that sandwich the DNA, a single molecule of Abf2p binds the DNA from one side, as a staple, and induces the U-turn by means of its HMG-boxes1 and 2 (26). The finding that HMG-box1 fits into the major groove is unprecedented. Such a behavior was confirmed by SAXS models in solution, and by MD simulations that showed the protein spontaneously directing HMG-box1 to the major groove, while HMG-box2 bound to the minor groove. Another surprising feature was that the coiled-coil predicted for the N-terminal helix was apparent only in the presence of DNA, creating dimers of the complexes but also higher-order multimers, which were impaired by mutations that disrupted the coiled-coil stable interaction. Higher concentration of the Gcf1p/DNA complexes favored coiled-coil dimers and increased the number of HMG-boxes bound to DNA, suggesting a highly dynamic mechanism that may underlie the local concentration effects of mtDNA packaging. Previous results showed that the formation of Gcf1p multimers on *C. albicans* DNA is linked to metabolism. Hyphal extracts contain monomers



**Figure 8.** DNA bridging by DNA condensing proteins. (A) Schematic representation of DNA bridging by the histone-like nucleoid structuring protein H-NS from *E. coli*, based on (74). In the figure, two different N-terminal domains, in green and orange respectively, interlock by their C-terminal end (C). The N-terminus (N) of each domain also interlocks with neighboring N-terminal domains (N'). The C-terminal end is followed by the DNA binding domain (DBD). Due to the C-terminal end interlock, the two DBDs of a dimer coincide in space but are proposed to bridge different, parallel DNA molecules. The same contact occurs at the preceding and following dimers, which faint towards the depth due to the helicoidal overall arrangement of the protein/DNA complex. (B) Working model of DNA binding by Gcf1p. Left, successive Gcf1p coiled coil dimers, with respective monomers colored in green and orange, are framed. The tetramer is represented by the extra two molecules in grey. The N-terminus of the helix (N) and the coiled coil (CC) are indicated. The HMG-box 1 and 2 (labelled 1 and 2, respectively) are indicated for each monomer in the central frame. Blue bent arrows indicate possible domain reorientation, and the question mark the uncertainty about this. Black arrows indicate potential distorted DNA points by HMG-box 2, which attract neighboring molecules. The compensation between wrapping ( $Wr$ ) and unwinding ( $Tw$ ) (see results) is symbolized by a linear DNA that apparently does not change the length. Right, schematic representation of a U-turn induced by one end of the Gcf1p dimer. (C) Left, schematic representation of the assembly of consecutive  $\alpha\beta$  heterodimers of the HU protein that bridge two parallel, linear DNA molecules. A similar but less tight arrangement is observed for HU  $\alpha\alpha$  homodimers. Right, representation of the HU  $\alpha\alpha$  homodimers binding to to the minor groove of distorted DNA. Based on (80).

and dimers, whereas higher-order complexes are found in yeast cells (13). Given our results, Gcf1p levels in mitochondria could induce the different oligomers, suggesting that Gcf1p mitochondrial translocation could act as a switch for *C. albicans* metabolism. TFAM phosphorylation inhibits DNA binding (68), thus regulatory mechanisms could also tune Gcf1p availability.

Our EM data indicate that Gcf1p condenses DNA *via* a bridging mechanism, imposing either parallel or antiparallel arrangements of DNA strands, but the resolution of the technique does not reveal the molecular details. *S. cerevisiae* Abf2p (26,61) and human TFAM (24,25,27,58,65,69–72) compact DNA by their two HMG-boxes, without additional domains. For TFAM, DNA bridging may involve protein-protein interactions, or cross-strand binding, or both mechanisms simultaneously (27,69–71,73). DNA bridging by Gcf1p may also involve the coiled-coil contact, which places a pair of HMG-boxes 1 and 2 at each end of the dimer. A comparable case is the *S. typhimurium* H-NS protein, which forms tight hairpins like Gcf1p on the DNA (60). The crystal structure of an H-NS fragment lacking the C-terminal DNA-binding domain (DBD) features a superhelix in which the N-terminal dimerization region of consecutive fragments contact each other face to face, whereas C-terminal dimerization regions contact tail-to-tail (74). The N- and C-terminal contacts alternate along the su-

perhelix, and the DBDs at the C-terminus are expected to be oriented towards the solvent to bind DNA (Figure 8A). Similarly, ParB, from *Bacillus subtilis*, involved in DNA segregation and sporulation, induces DNA condensation by also alternating homodimers of the N- and C-terminal domains on the nucleic acid, which bind non-specifically (75).

The extensively bridged DNA fragments observed in our EM micrographs should combine both coiled-coil dimerization at one end of Gcf1p, and DNA bridging by HMG-boxes at the other end (Figure 8B). HMG-boxes between consecutive dimers are not expected to interact but create local DNA distortions recognized and stabilized by additional Gcf1p proteins, thereby promoting cooperative binding and protein nucleation on the DNA, similar to TFAM (23,26,59,76–78) (black arrows in Figure 8B). Unlike H-NS, Gcf1p does not form a DNA superhelix, as confirmed by our topology assays indicating almost no change in Lk (Figure 8B). However, no change in Lk could also be due to DNA wrapping compensated by local changes in twist (unwinding). As shown by the crystal structure, overall wrapping of the DNA is compensated by DNA unwinding at regions contacted by HMG-box2. In the crystal, yet, the DNA fragments form a biologically-irrelevant minicircle, and our MD calculations suggest that, on a linear continuous DNA, the HMG-boxes can rearrange their orientation (bent arrows in Figure 8B). On the other hand, MD



showed that the U-turn found in the crystal is highly stable, thus it possibly reflects a special binding mode of Gcf1p that not necessarily occur during bridging but conceivably in more compacted states of the nucleoid. Drastic different binding modes were also suggested for TFAM, which forms a DNA U-turn in the transcription initiation complex that is difficult to conceal with fast TFAM sliding on the DNA (77) but plausible during compaction (71). In this line, remarkably different DNA binding modes have been described for either the *E. coli* Integration Host Factor (IHF) or the nucleoid-packaging protein HU. IHF shows three binding modes, (i) the protein dimerizes and positions two  $\beta$ -hairpins within the DNA minor groove which, together with neighboring protein regions, bend the DNA by 65°; (ii) an additional region of IHF participates in more extensive contacts with the DNA, which then bends by 115° and (iii) a second additional surface further participates in binding, inducing a DNA bending of 160°. Finally, IHF bound to DNA by the  $\beta$ -hairpins can, in addition, bridge a second DNA molecule by a far fourth surface and, notably, both bridged DNAs are not bent but parallel in an extended linear conformation due to repulsions between them (79). Regarding HU, dimers of the two variants  $\alpha$  and  $\beta$  induce DNA condensation by bridging linear DNA molecules in parallel (80) (Figure 8C). The HU $\alpha\beta$  heterodimers bridge parallel DNA fragments considerably more tightly than HU $\alpha\alpha$  homodimers, but in both cases contacts between neighboring dimers involve a long  $\beta$ -hairpin arm. Strikingly, a HU dimer can also bind distorted DNA (bent by 100–140°) by snugging the two very same  $\beta$ -hairpins into the DNA minor groove (80), showing once again high variability in binding modes. According to our results, we can also consider different binding mechanisms for Gcf1p. Our current working model is as it follows. During Gcf1p assembly on the DNA, Gcf1p dimerizes by the N-terminal coiled coil and the HMG-boxes are reoriented and bridge the DNA (likely in a much lesser extent than in our micrographs), in a process in which DNA unwinding compensates DNA wrapping (Figure 8B). The highly basic N-terminal disordered region participates in the dynamic attraction of complexes, but it is not involved in the bridging mechanism itself. According to our micrographs, a succession of bridging events could eventually lead to DNA condensation. However, for highly compacted DNA, we remain cautious regarding its topological features (wrapping and unwinding), since we could not measure the contour length in these specimens and thus cannot conclude if DNA shortening occurred or not. Given the results from SAXS and MD, we cannot discard that DNA U-turns appear in such a tightly compacted DNA, as also proposed for TFAM and Abf2p (24,26,58,71).

Deleting the *Gcf1* gene in *C. albicans* was previously shown to reduce mtDNA copy number and recombination intermediates (20), suggesting that Gcf1p stabilizes such structures in the recombination-driven replication mechanism of *Candida* mtDNA (21). This could also reflect the stabilization of DNA pairs by Gcf1p during recombination, although such functions should be exquisitely regulated as over-compaction of mtDNA inhibits transcription and regulation (65). Further molecular analysis is required to provide insight into this potential function of Gcf1p.

In the Gcf1p crystal structure, 2-fold rotation places a second coiled-coil-dimer at the other side of the parallel DNAs, generating the supra-assembly. This places two HMG-box2 domains face to face, while bending the DNA such that the two DNA ends engage in a highly distorted stacking that, during MD simulations, does not rearrange into a regular 5'-3' DNA stacking, suggesting a crystallographic contact. Each HMG-box2 bends the DNA by inserting Met186 at DNA steps spaced by 4 bp, creating a dramatically twisted U-turn. Both Abf2p/DNA and TFAM/DNA crystal structures also show partitioned U-turns, created by HMG-box1 and HMG-box2 of a single protein that contact different DNA molecules (26,73). However, in these two cases the DNA ends join with perfect stacking, resulting in a smooth, pseudo-continuous U-turn in which the two insertion sites are separated by 11 and 12 bp (a DNA helical turn). In principle, a similar pseudo-continuous U-turn could be generated from the Gcf1p/DNA crystal complex by adding extra bases between the two DNA ends. This would imply the major reorientations of all Gcf1p domains, which is conceivable in principle considering the high flexibility of Gcf1p.

In conclusion, our data show that Gcf1p is a flexible protein, with both HMG-boxes binding independent DNAs. The N-terminal tail of different complexes forms a coiled-coil, whereas the N-terminal disordered region dynamically recruits additional complexes. These interactions lead the formation of dimers and/or larger multimers, and eventually to supra-assemblies. Gcf1p cooperatively accumulates on the DNA, leading to bridging, formation of hairpins and DNA aggregations, aligning dsDNA segments in parallel without distortions in twist or writhe. The observed dynamic binding should modulate the compaction and regulation of mtDNA, which in turn underpins mitochondrial activity and cell survival.

## DATA AVAILABILITY

PDB code of the X-ray crystal structure coordinates is 7ZIE. SAXS data from Gcf1p alone or in complex with DNA and models are accessible at the Small Angle Scattering Biological Data Bank with codes: SASDPY5–Gcf1p protein (corresponding to Figure 6A); SASDPZ5 - Gcf1p protein bound to DNA at 1.0 mg/ml (Figure 6B). SASDP26–Gcf1p protein bound to DNA at 2.0 mg/ml (Figure 6C). SASDP36–N-terminal truncated Gcf1p (amino acids 59–245) bound to DNA (Figure 6D). Molecular dynamics simulations of Gcf1p in complex with DNA are at the server <https://mmb.irbbarcelona.org/BIGNASim/>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Neus Gual and Nicolás Ortiz for their help with protein production and DNA binding experiments. We thank Dr Richard M. Twyman for writing assistance. We thank the personnel at synchrotrons ESRF (Grenoble,

France), ALBA (Cerdanyola del Vallès, Spain) and DESY (Hamburg, Germany), and from the Automated Crystallography Platform (IBMB-CSIC) for their highly valuable support. We thank Dr Massimo D. Sammito, Prof. Airlie McCoy and Prof. Randy Read for their kind help in structure resolution.

**Author contributions:** A.T.S., E.R.L. prepared the proteins; A.T.S. did the structural studies; A.T.S. and S.L. did the biophysical studies; F.B. did the M.D. simulations; J.S. and J.M.G. designed and prepared the genetic construct and EMSA; A.T.S., O.P. and E.L.C. did the EM studies; B.M. and J.R. did the topology studies; A.T.S. and P.B. did SAXS analysis; all authors participated in manuscript writing, M.S. designed and supervised the project.

## FUNDING

Spanish Ministry of Science, Innovation and Universities MCIN/AEI/10.13039/501100011033 ERDF 'A way to make Europe' [BFU2015-70645-R, 2018-RTI2018-101015-B-100, PID2021-129038NB-I00 to M.S., RTI2018-096704-B-100 to M.O., PID2019-109482GB-I00 to J.R.]; Generalitat de Catalunya [2014-SGR-997, 2017-SGR-1192, 2021-SGR-00425 to M.S., SGR2017-134 to M.O.]; Instituto Nacional de Bioinformática, the European Research Council [FP7-HEALTH-2012-306029-2 to M.S.]; H2020 European Commission [BioExcel-2. Centre of Excellence for Computational Biomolecular Research' 823830, to M.O.]; MINECO Severo Ochoa Award of Excellence from the Government of Spain (to IRB Barcelona); M.O. is an ICREA (Institutió Catalana de Recerca i Estudis Avançats) academia researcher; A.T.-S. received a PhD fellowship from the Ministry of Education, Professional Formation (MEFP); A.T.-S. worked on his PhD thesis under the Doctorate Program Biochemistry, Molecular Biology and Biomedicine of the Autonomous University of Barcelona (UAB); the Structural Biology Unit at IBMB-CSIC was awarded with a 'Maria de Maeztu' Unit of Excellence mention by MINECO [MDM-2014-0435]; IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from MINECO; the CBS-Montpellier is a member of France-BioImaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), two national infrastructures supported by the French National Research Agency [ANR-10-INBS-04-01, ANR-10-INBS-05]; J.S. and J.M.G. were supported by Estonian Institutional Funding Grant [IUT14021]. Funding for open access charge: Spanish Ministry of Science and Innovation.

**Conflict of interest statement.** None declared.

## REFERENCES

- Ghannoum, M.A., Jurevic, R.J., Mukherjee, P.K., Cui, F., Sikaroodi, M., Naqvi, A. and Gillevet, P.M. (2010) Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog.*, **6**, e1000713.
- Hallen-Adams, H.E. and Suhr, M.J. (2017) Fungi in the healthy human gastrointestinal tract. *Virulence*, **8**, 352–358.
- Kalia, N., Singh, J. and Kaur, M. (2020) Microbiota in vaginal health and pathogenesis of recurrent vulvovaginal infections: a critical review. *Ann Clin. Microbiol. Antimicrob.*, **19**, 5.
- Kuhbacher, A., Burger-Kentischer, A. and Rupp, S. (2017) Interaction of *Candida* species with the skin. *Microorganisms*, **5**, 32.
- Atarashi, K., Tanoue, T., Ando, M., Kamada, N., Nagano, Y., Narushima, S., Suda, W., Imaoka, A., Setoyama, H., Nagamori, T. *et al.* (2015) Th17 cell induction by adhesion of microbes to intestinal epithelial cells. *Cell*, **163**, 367–380.
- Markey, L., Shaban, L., Green, E.R., Lemon, K.P., Meccas, J. and Kumamoto, C.A. (2018) Pre-colonization with the commensal fungus *Candida albicans* reduces murine susceptibility to *Clostridium difficile* infection. *Gut Microbes*, **9**, 497–509.
- Quintin, J., Saeed, S., Martens, J.H.A., Giamarellos-Bourboulis, E.J., Ifrim, D.C., Logie, C., Jacobs, L., Jansen, T., Kullberg, B.J., Wijmenga, C. *et al.* (2012) *Candida albicans* infection affords protection against reinfection via functional reprogramming of monocytes. *Cell Host Microbe*, **12**, 223–232.
- Chen, H., Zhou, X., Ren, B. and Cheng, L. (2020) The regulation of hyphae growth in *Candida albicans*. *Virulence*, **11**, 337–348.
- Kauffman, C.A. (2006) Fungal infections. *Proc. Am. Thorac. Soc.*, **3**, 35–40.
- Kullberg, B.J. and Arendrup, M.C. (2015) Invasive Candidiasis. *N. Engl. J. Med.*, **373**, 1445–1456.
- Gudlaugsson, O., Gillespie, S., Lee, K., Vande Berg, J., Hu, J., Messer, S., Herwaldt, L., Pfaller, M. and Diekema, D. (2003) Attributable mortality of nosocomial candidemia, revisited. *Clin. Infect. Dis.*, **37**, 1172–1177.
- Wisplinghoff, H., Bischoff, T., Tallent, S.M., Seifert, H., Wenzel, R.P. and Edmond, M.B. (2004) Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. *Clin. Infect. Dis.*, **39**, 309–317.
- Kim, S., Wolyniak, M.J., Staab, J.F. and Sundstrom, P. (2007) A 368-base-pair cis-acting HWP1 promoter region, HCR, of *Candida albicans* confers hypha-specific gene regulation and binds architectural transcription factors Nhp6 and Gcf1p. *Eukaryot. Cell*, **6**, 693–709.
- Nosek, J., Tomaska, L., Bolotin-Fukuhara, M. and Miyakawa, I. (2006) Mitochondrial chromosome structure: an insight from analysis of complete yeast genomes. *FEMS Yeast Res.*, **6**, 356–370.
- Kolondra, A., Labedzka-Dmoch, K., Wenda, J.M., Drzewicka, K. and Golik, P. (2015) The transcriptome of *Candida albicans* mitochondria and the evolution of organellar transcription units in yeasts. *Bmc Genomics [Electronic Resource]*, **16**, 827.
- Chabrier-Rosello, Y., Foster, T.H., Mitra, S. and Haidaris, C.G. (2008) Respiratory deficiency enhances the sensitivity of the pathogenic fungus *Candida* to photodynamic treatment. *Photochem. Photobiol.*, **84**, 1141–1148.
- Chabrier-Rosello, Y., Giesselman, B.R., De Jesus-Andino, F.J., Foster, T.H., Mitra, S. and Haidaris, C.G. (2010) Inhibition of electron transport chain assembly and function promotes photodynamic killing of *Candida*. *J. Photochem. Photobiol. B*, **99**, 117–125.
- Joers, P., Gerhold, J.M., Sedman, T., Kuusk, S. and Sedman, J. (2007) The helicase CaHmi1p is required for wild-type mitochondrial DNA organization in *Candida albicans*. *FEMS Yeast Res.*, **7**, 118–130.
- Richard, M.L., Nobile, C.J., Bruno, V.M. and Mitchell, A.P. (2005) *Candida albicans* biofilm-defective mutants. *Eukaryot. Cell*, **4**, 1493–1502.
- Visacka, K., Gerhold, J.M., Petrovicova, J., Kinsky, S., Joers, P., Nosek, J., Sedman, J. and Tomaska, L. (2009) Novel subfamily of mitochondrial HMG box-containing proteins: functional analysis of Gcf1p from *Candida albicans*. *Microbiology*, **155**, 1226–1240.
- Gerhold, J.M., Aun, A., Sedman, T., Joers, P. and Sedman, J. (2010) Strand invasion structures in the inverted repeat of *Candida albicans* mitochondrial DNA reveal a role for homologous recombination in replication. *Mol. Cell*, **39**, 851–861.
- Gerhold, J.M., Sedman, T., Visacka, K., Slezakova, J., Tomaska, L., Nosek, J. and Sedman, J. (2014) Replication intermediates of the linear mitochondrial DNA of *Candida parapsilosis* suggest a common recombination based mechanism for yeast mitochondria. *J. Biol. Chem.*, **289**, 22659–22670.
- Bakkaiova, J., Marini, V., Willcox, S., Nosek, J., Griffith, J.D., Krejci, L. and Tomaska, L. (2015) Yeast mitochondrial HMG proteins: dNA-binding properties of the most evolutionarily divergent component of mitochondrial nucleoids. *Biosci. Rep.*, **36**, e00288.
- Rubio-Cosials, A., Sidow, J.F., Jimenez-Menendez, N., Fernandez-Millan, P., Montoya, J., Jacobs, H.T., Coll, M., Bernado, P. and Sola, M. (2011) Human mitochondrial transcription factor A

- induces a U-turn structure in the light strand promoter. *Nat. Struct. Mol. Biol.*, **18**, 1281–1289.
25. Ngo, H.B., Kaiser, J.T. and Chan, D.C. (2011) The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nat. Struct. Mol. Biol.*, **18**, 1290–1296.
  26. Chakraborty, A., Lyonnais, S., Battistini, F., Hospital, A., Medici, G., Prohens, R., Orozco, M., Vilardell, J. and Sola, M. (2016) DNA structure directs positioning of the mitochondrial genome packaging protein Abf2p. *Nucleic Acids Res.*, **45**, 951–967.
  27. Ngo, H.B., Lovely, G.A., Phillips, R. and Chan, D.C. (2014) Distinct structural features of TFAM drive mitochondrial DNA packaging versus transcriptional activation. *Nat. Commun.*, **5**, 3077.
  28. Hillen, H.S., Morozov, Y.I., Sarfallah, A., Temiakov, D. and Cramer, P. (2017) Structural basis of mitochondrial transcription initiation. *Cell*, **171**, 1072–1081.
  29. Fisher, R.P., Parisi, M.A. and Clayton, D.A. (1989) Flexible recognition of rapidly evolving promoter sequences by mitochondrial transcription factor 1. *Genes Dev.*, **3**, 2202–2217.
  30. Yoon, Y.G., Koob, M.D. and Yoo, Y.H. (2011) Mitochondrial genome-maintaining activity of mouse mitochondrial transcription factor A and its transcript isoform in *Saccharomyces cerevisiae*. *Gene*, **484**, 52–60.
  31. Munteanu, B., Braun, M. and Boonrod, K. (2012) Improvement of PCR reaction conditions for site-directed mutagenesis of big plasmids. *J. Zhejiang Univ. Sci. B*, **13**, 244–247.
  32. Difley, J.F. and Stillman, B. (1992) DNA binding properties of an HMG1-related protein from yeast mitochondria. *J. Biol. Chem.*, **267**, 3368–3374.
  33. Kabsch, W. (2010) Xds. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 125–132.
  34. Evans, P.R. and Murshudov, G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D, Biol. Crystallogr.*, **69**, 1204–1214.
  35. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
  36. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Cross, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J. *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.*, **75**, 861–877.
  37. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 486–501.
  38. Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
  39. Konarev, P.V., Volkov, V.V., Sokolova, A.V., M.H.J., K. and Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.*, **36**, 1277–1282.
  40. Hajizadeh, N.R., Franke, D., Jeffries, C.M. and Svergun, D.I. (2018) Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. *Sci. Rep.*, **8**, 7204.
  41. Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. and Svergun, D.I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.*, **129**, 5656–5664.
  42. Svergun, D.I., Barberato, C. and Koch, M.H.J. (1995) CRY SOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, **28**, 768–773.
  43. Knight, C.J. and Hub, J.S. (2015) WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. *Nucleic Acids Res.*, **43**, W225–W230.
  44. Dupaigne, P., Tavares, E.M., Pietrement, O. and Le Cam, E. (2018) Recombinases and Related Proteins in the Context of Homologous Recombination Analyzed by Molecular Microscopy. *Methods Mol. Biol.*, **1805**, 251–270.
  45. Fangman, W.L., Henly, J.W., Churchill, G. and Brewer, B.J. (1989) Stable maintenance of a 35-base-pair yeast mitochondrial genome. *Mol. Cell. Biol.*, **9**, 1917–1921.
  46. Arnott, S. and Hukins, D.W. (1972) Optimised parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Commun.*, **47**, 1504–1509.
  47. Jorgensen, W., Chandrasekhar, J. and Madura, J.D. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–.
  48. Smith, D.E. and Liem, X.D. (1994) Computer simulations of NaCl association in polarizable water. *J. Chem. Phys.*, **100**, 3757.
  49. Pasi, M., Maddocks, J.H., Beveridge, D., Bishop, T.C., Case, D.A., Cheatham, T. III, Dans, P.D., Jayaram, B., Lankas, F., Laughton, C. *et al.* (2014) muABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.*, **42**, 12272–12283.
  50. Ivani, I., Dans, P.D., Noy, A., Perez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goni, R., Balaceanu, A. *et al.* (2016) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, **13**, 55–58.
  51. Roe, D.R. and Cheatham, T.E. III (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.*, **9**, 3084–3095.
  52. Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D. and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: curves+. *Nucl. Acids Res.*, **37**, 5917–5929.
  53. Hospital, A., Andrio, P., Cugnasco, C., Codo, L., Becerra, Y., Dans, P.D., Battistini, F., Torres, J., Goni, R., Orozco, M. *et al.* (2016) BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Res.*, **44**, D272–D278.
  54. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
  55. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
  56. Gouet, P., Robert, X. and Courcelle, E. (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.*, **31**, 3320–3323.
  57. Gangelhoff, T.A., Mungalachetty, P.S., Nix, J.C. and Churchill, M.E. (2009) Structural analysis and DNA binding of the HMG domains of the human mitochondrial transcription factor A. *Nucleic Acids Res.*, **37**, 3153–3164.
  58. Rubio-Cosials, A., Battistini, F., Gansen, A., Cuppari, A., Bernado, P., Orozco, M., Langowski, J., Toth, K. and Sola, M. (2018) Protein flexibility and synergy of HMG domains underlie U-turn bending of DNA by TFAM in solution. *Biophys. J.*, **114**, 2386–2396.
  59. Malarkey, C.S. and Churchill, M.E. (2012) The high mobility group box: the ultimate utility player of a cell. *Trends Biochem. Sci.*, **37**, 553–562.
  60. Dame, R.T., Wyman, C. and Goosen, N. (2001) Structural basis for preferential binding of H-NS to curved DNA. *Biochimie*, **83**, 231–234.
  61. Friddle, R.W., Klare, J.E., Martin, S.S., Corzett, M., Balhorn, R., Baldwin, E.P., Baskin, R.J. and Noy, A. (2004) Mechanism of DNA compaction by yeast mitochondrial protein Abf2p. *Biophys. J.*, **86**, 1632–1639.
  62. Kaufman, B.A., Durisic, N., Mativetsky, J.M., Costantino, S., Hancock, M.A., Grutter, P. and Shoubridge, E.A. (2007) The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. *Mol. Biol. Cell*, **18**, 3225–3236.
  63. Dame, R.T., Wyman, C. and Goosen, N. (2000) H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res.*, **28**, 3504–3510.
  64. Beloin, C., Jeusset, J., Revet, B., Mirambeau, G., Le Hegarat, F. and Le Cam, E. (2003) Contribution of DNA conformation and topology in right-handed DNA wrapping by the *Bacillus subtilis* LrpC protein. *J. Biol. Chem.*, **278**, 5333–5342.
  65. Farge, G., Mehmedovic, M., Baclayon, M., van den Wildenberg, S.M., Roos, W.H., Gustafsson, C.M., Wuite, G.J. and Falkenberg, M. (2014) In vitro-reconstituted nucleoids can block mitochondrial DNA replication and transcription. *Cell Rep.*, **8**, 66–74.
  66. Fang, Y., Akimoto, M., Mayanagi, K., Hatano, A., Matsumoto, M., Matsuda, S., Yasukawa, T. and Kang, D. (2020) Chemical acetylation of mitochondrial transcription factor A occurs on specific lysine residues and affects its ability to change global DNA topology. *Mitochondrion*, **53**, 99–108.
  67. Menger, K.E., Rodriguez-Luis, A., Chapman, J. and Nicholls, T.J. (2021) Controlling the topology of mammalian mitochondrial DNA. *Open Biol.*, **11**, 210168.

68. Lu, B., Lee, J., Nie, X., Li, M., Morozov, Y.I., Venkatesh, S., Bogenhagen, D.F., Temiakov, D. and Suzuki, C.K. (2013) Phosphorylation of human TFAM in mitochondria impairs DNA binding and promotes degradation by the AAA+ Lon protease. *Mol. Cell*, **49**, 121–132.
69. Kukat, C., Davies, K.M., Wurm, C.A., Spahr, H., Bonekamp, N.A., Kuhl, I., Joos, F., Polosa, P.L., Park, C.B., Posse, V. *et al.* (2015) Cross-strand binding of TFAM to a single mtDNA molecule forms the mitochondrial nucleoid. *Proc. Nat. Acad. Sci. U.S.A.*, **112**, 11288–11293.
70. Uchida, A., Murugesapillai, D., Kastner, M., Wang, Y., Lodeiro, M.F., Prabhakar, S., Oliver, G.V., Arnold, J.J., Maher, L.J., Williams, M.C. *et al.* (2017) Unexpected sequences and structures of mtDNA required for efficient transcription from the first heavy-strand promoter. *Elife*, **6**, e27283.
71. Cuppari, A., Fernandez-Millan, P., Battistini, F., Tarres-Sole, A., Lyonais, S., Iruela, G., Ruiz-Lopez, E., Enciso, Y., Rubio-Cosials, A., Prohens, R. *et al.* (2019) DNA specificities modulate the binding of human transcription factor A to mitochondrial DNA control region. *Nucleic Acids Res.*, **47**, 6519–6537.
72. Rubio-Cosials, A. and Sola, M. (2013) U-turn DNA bending by human mitochondrial transcription factor A. *Curr. Opin. Struct. Biol.*, **23**, 116–124.
73. Choi, W.S. and Garcia-Diaz, M. (2021) A minimal motif for sequence recognition by mitochondrial transcription factor A (TFAM). *Nucl. Acids Res.*, **50**, 322–332.
74. Arold, S.T., Leonard, P.G., Parkinson, G.N. and Ladbury, J.E. (2010) H-NS forms a superhelical protein scaffold for DNA condensation. *Proc. Nat. Acad. Sci. U.S.A.*, **107**, 15728–15732.
75. Madariaga-Marcos, J., Pastrana, C.L., Fisher, G.L., Dillingham, M.S. and Moreno-Herrero, F. (2019) ParB dynamics and the critical role of the CTD in DNA condensation unveiled by combined force-fluorescence measurements. *Elife*, **8**, e43812.
76. Yoshida, Y., Izumi, H., Ise, T., Uramoto, H., Torigoe, T., Ishiguchi, H., Murakami, T., Tanabe, M., Nakayama, Y., Itoh, H. *et al.* (2002) Human mitochondrial transcription factor A binds preferentially to oxidatively damaged DNA. *Biochem. Biophys. Res. Commun.*, **295**, 945–951.
77. Farge, G., Laurens, N., Broekmans, O.D., van den Wildenberg, S.M., Dekker, L.C., Gaspari, M., Gustafsson, C.M., Peterman, E.J., Falkenberg, M. and Wuite, G.J. (2012) Protein sliding and DNA denaturation are essential for DNA organization by human mitochondrial transcription factor A. *Nat. Commun.*, **3**, 1013.
78. Lyonais, S., Tarres-Sole, A., Rubio-Cosials, A., Cuppari, A., Brito, R., Jaumot, J., Gargallo, R., Vilaseca, M., Silva, C., Granzhan, A. *et al.* (2017) The human mitochondrial transcription factor A is a versatile G-quadruplex binding protein. *Sci. Rep.*, **7**, 43992.
79. Yoshua, S.B., Watson, G.D., Howard, J.A.L., Velasco-Berrelleza, V., Leake, M.C. and Noy, A. (2021) Integration host factor bends and bridges DNA in a multiplicity of binding modes with varying specificity. *Nucleic Acids Res.*, **49**, 8684–8698.
80. Hammel, M., Amlanjyoti, D., Reyes, F.E., Chen, J.H., Parpana, R., Tang, H.Y., Larabell, C.A., Tainer, J.A. and Adhya, S. (2016) HU multimerization shift controls nucleoid compaction. *Sci. Adv.*, **2**, e1600650.