Structural and DNA binding properties of mycobacterial integration host factor mIHF

Nina T Odermatt, Moreno Lelli, Torsten Herrmann, Luciano A Abriata, Aleksandre Japaridze, Hubert Voilquin, Rajkumar Singh, Jérémie Piton, Lyndon Emsley, Giovanni Dietler, et al.

Structural and DNA Binding Properties of Mycobacterial Integration Host Factor mIHF

Authors

Nina T. Odermatt[a,1], Moreno Lelli[b,2/3], Torsten Herrmann[b,4], Luciano A. Abriata[a,5], Aleksandre Japaridze[c,6], Hubert Voilquin[c,7], Rajkumar Singh[a,8], Jérémie Piton[a,9], Lyndon Emsley[d,10], Giovanni Dietler[c,11], Stewart T. Cole[a,12]

a. École Polytechnique Fédérale de Lausanne, School of Life Sciences, Station 19, 1015 Lausanne, Switzerland
b. University of Lyon, Institute for Analytical Sciences – Center for High Field NMR, CNRS UMR 5280, ENS Lyon, UCB Lyon 1, France
c. École Polytechnique Fédérale de Lausanne, School of Basic Sciences, Route de la Sorge, 1015 Lausanne, Switzerland
d. École Polytechnique Fédérale de Lausanne, School of Basic Sciences, Av. F.-A. Forel 2, 1015 Lausanne, Switzerland

Corresponding author: Prof. Stewart T. Cole, École Polytechnique Fédérale de Lausanne, Global Health Institute, Station 19, 1015 Lausanne, Switzerland.
Tel. +33145688001

[1] Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse 10, 35043 Marburg, Germany. nina.odermatt@mpi-marburg.mpg.de
[2] Department of Chemistry, University of Florence, Via della Lastruccia, 3-13, 50019 Sesto Fiorentino (FI), Italy. moreno.lelli@unifi.it
[3] Centro di Ricerca di Risonanze Magnetiche (CERM), University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino (FI), Italy. Torsten.Herrmann@ibs.fr
[4] Univ. Grenoble Alpes, CNRS, CEA, IBS, 71 avnue des Marturs, CS 10090, 38044 Grenoble Cedex 9, France. Torsten.Herrmann@ibs.fr
[5] luciano.abriata@epfl.ch
[6] A.Japaridze@tudelft.nl
[7] hubertvoilquin@gmail.com
[8] rajkumarbiotech@gmail.com
[9] jeremie.piton@gmail.com
[10] lyndon.emsley@epfl.ch
[11] giovanni.dietler@epfl.ch
[12] Institut Pasteur, 25-28 rue du Docteur Roux, 75724 Paris Cedex 15, France

Abstract

In bacteria, nucleoid associated proteins (NAPs) take part in active chromosome organization by supercoil management, three-dimensional DNA looping and direct transcriptional control. Mycobacterial integration host factor (mIHF, *rv1388*) is a NAP restricted to Actinobacteria and essential for survival of the human pathogen *Mycobacterium tuberculosis*. We show *in vitro* that DNA binding by mIHF strongly stabilizes the protein and increases its melting temperature. The structure obtained by Nuclear Magnetic Resonance (NMR) spectroscopy characterizes mIHF as a globular protein with a protruding alpha helix and a disordered N-terminus, similar to *Streptomyces coelicolor* IHF (sIHF). NMR revealed no residues of high flexibility, suggesting that mIHF is a rigid protein overall that does not undergo structural rearrangements. We show that mIHF only binds to double stranded DNA in solution, through two DNA binding sites (DBSs) similar to those identified in the x-ray structure of sIHF. According to Atomic Force Microscopy, mIHF is able to introduce left-handed loops of ca. 100 nm size (~300 bp) in supercoiled cosmids, thereby unwinding and relaxing the DNA.

Highlights

- The structure of mIHF was solved by solution NMR spectroscopy
- mIHF is an alpha-helical protein with two DNA-binding sites
- mIHF introduces left-handed loops into DNA
- DNA unwinds and re-compacts in a mIHF-level dependent manner

Keywords

Mycobacterial Integration Host Factor, mIHF, Nucleoid Associated Protein, DNA binding, NMR structural determination, atomic force microscopy

Abbreviations

mIHF, mycobacterial Integration Host Factor; DBS, DNA-binding site; dsDNA, double stranded DNA; NAP, nucleoid associated protein; ssDNA, single stranded DNA

1. Introduction

Tuberculosis, caused by *Mycobacterium tuberculosis*, is the leading cause of death by a single infectious agent, claiming over 1.5 million lives each year (WHO, 2016). The pathogen mainly causes lung infections (Farer et al., 1979) that are hard to diagnose, difficult to treat and can lead to death when undetected (Zumla et al., 2013). *M. tuberculosis* can enter a dormant phase (Boshoff and Barry, 2005), characterized by phenotypic drug resistance (Gomez and McKinney, 2004) and has evolved a sophisticated virulence strategy employing type VII secretion systems known as ESX-1 to ESX-5 (Gröschel et al., 2016). Control of virulence factor expression is crucial and has to be carefully adapted to active growth or the latent phase (Zondervan et al., 2018). Transcription is controlled over interlayered levels involving the DNA supercoiling state and by local transcription factors, such as repressors and activators, sigma factors and nucleoid associated proteins (NAPs) (Balleza et al., 2009).

NAPs are small, highly abundant proteins in bacteria with hundreds of binding sites on the chromosome. Their ability to bridge DNA, assisted by supercoiling and macromolecular crowding, leads to compaction of the chromosome (Luijsterburg et al., 2006) and its organization into 1 Mbp macrodomains (Song and Loparo, 2015). In addition to their architectural role, NAPs act as global gene regulators, often involved in induction of virulence expression and genetic adaptation to environmental cues (Dorman, 2004). In *E. coli*, at least 12 NAPs have been identified that regulate a vast number of genes (Dillon and Dorman, 2010). Although *M. tuberculosis* is exposed to different stresses during host entry and the accompanying immune response (Manganelli et al., 2004), and can also enter a non-replicating state, NAPs seem to be underrepresented in this bacterium. To date, only four proteins were confirmed to act as NAPs; Lsr2, HupB, EspR and mIHF. Lsr2 protects *M. tuberculosis* from reactive oxygen species (Colangeli et al., 2009), and acts as a dimer (Summers et al., 2012) in repressing gene expression (Gordon et al., 2010). HupB is the only conserved NAP and shows sequence similarity to HU of Gram-negative bacteria. HupB is involved in iron homeostasis, cell wall synthesis and adhesion (Pandey et al., 2014). EspR is essential for the expression of the ESX-1 virulence factors and several other virulence associated genes besides (Blasco et al., 2012). The mycobacterial integration host factor, mIHF, was shown to be essential for bacterial survival and regulates genes involved in lipid metabolism, metabolic pathways, translation and virulence (Odermatt et al., 2018).

Most NAPs act as dimers, with each monomer binding to a distinct DNA locus and bringing these loci together upon dimerization, as exemplified by EspR (Blasco et al., 2011), Lsr2 (Summers et al., 2012) or HupB (Bhowmick et al., 2014) in *M. tuberculosis*. On the contrary, the mIHF homologue from *Streptomyces coelicolor*, sIHF, has two separate DNA-binding sites (DBS) and was proposed to bind DNA as a monomer (Swiercz et al., 2013). mIHF was shown to bind and bend DNA non-specifically (Mishra et al., 2013), but the exact DNA-binding mechanism of mIHF and its structure are unknown. Here, we complement the genetic analysis of mIHF (Odermatt et al., 2018) by providing structural insights into its biological function. We determined the solution state structure of mIHF and measured its dynamics by Nuclear Magnetic Resonance (NMR) spectroscopy, and studied its DNA binding properties by NMR, Isothermal Titration Calorimetry (ITC), and Atomic Force Microscopy (AFM).

2. Methods

2.1 Sequence analysis and alignment

Amino acid sequences were obtained from NCBI, aligned with T-Coffee (Notredame et al., 2000) and visualized with BioEdit (Thomas A. Hall, 1999). Secondary structure was predicted by PSIPRED (Jones et al., 2017).

2.2 Cloning, expression and purification of mIHF

Oligonucleotides were synthesized by MICROSYNTH (Switzerland). Primers *mihF*-exp-F and *mihF*-exp-R (all primers are listed in Table S1) were used to PCR-amplify the *mihF* gene beginning at the translational start site. Restriction sites NdeI and BamHI were included in the primers to clone the resulting fragment into pET28a vector to give the expression vector pNO72 producing N-terminally 6x His-tagged mIHF protein. Recombinant His-tagged protein was expressed by growing *E. coli* BL21 DE3 (ThermoFisher) cells transformed with pNO72 in Luria-Bertani (LB) broth in the presence of 50 µg ml^{-1} kanamycin. Protein expression was induced at an OD_{600} = 0.6 with 500 µM IPTG overnight at 16° C with constant shaking at 180 rpm. Cells were harvested by centrifugation and stored at -80°C until further processing. The pellet was thawed in lysis buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM Imidazole, 10% (w/v) glycerol and 1% TWEEN 20) supplemented with 1 tablet cOMPLETE mini protease inhibitor (Roche) and 10 µl DNase I (Roche) and lysed in a cell disrupter (EmulsiFlex, Avestin, Canada). Insoluble debris was removed after centrifugation and supernatant was loaded on a His-trap column pre-equilibrated with buffer A (500 mM NaCl, 50 mM Tris-HCl pH 7.5). Increasing concentrations of buffer B (500 mM NaCl, 50 mM Tris-HCl pH 7.5, 500 mM Imidazol) eluted the protein from the column and fractions were visualized by COOMASSIE blue staining of a sodium dodecyl sulphate (SDS) gel (NuPAGE 4 – 12% Bis-Tris Gel, INVITROGEN). Pooled fractions containing mIHF were dialyzed into Tris-buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.5) overnight at 4°C. To remove the His-tag, 500 µl thrombin (Sigma) was added to the protein before dialysis. The cut mIHF protein was subsequently purified by passage over Ni-NTA Agarose (Machery-Nagel) and Benzamidine Sepharose (GE Healthcare) to remove uncut protein, His-tags and thrombin. A size-selection purification step was applied to remove unfolded protein and increase sample purity. mIHF was passed over a HiLoad 16/600 column and fractions were pooled and concentrated in centrifugal filters

(Ultracel, Amicon) with 3 kDa cut-off. Purified mIHF was stored at -80°C in small aliquots until further use. Concentration of mIHF was measured by Qubit (INVITROGEN).

Primers *mihF*-R90E-F and *mihF*-R90E-R were used in the QuickChange II Site-Directed Mutagenesis kit (Agilent) as recommended by the manufacturer. The single amino acid mutation R90E in pNO72 gave rise to pNO84.

For the production of isotope labelled mIHF (^{15}N- and ^{13}C-ammonium chloride and D-glucose from Cambridge Isotope Laboratories, USA), the exponentially growing culture was resuspended in minimal medium containing 25 mM KH_2PO_4, 50 mM $Na_2HPO_4 \cdot 2\ H_2O$, 1 g l^{-1} ^{15}N-labelled ammonium chloride, 10 mM NaCl, 0.2 mM $CaCl_2$, 1 mM $MgSO_4$, 50 µM iron solution, 10 ml trace metal solution and 3 g l^{-1} ^{13}C labelled glucose. 1000 x iron solution contained 100 µM citric acid and 50 µM $FeCl_3$. 100 x trace metal solution contained 68.2 µM $MnCl_2 \cdot 4\ H_2O$, 3.7 µM $ZnSO_4 \cdot 7\ H_2O$, 0.4 µM $CoCl_2 \cdot 6\ H_2O$, 0.6 µM $CuCl_2 \cdot 2\ H_2O$, 1.6 µM H_3BO_3, 2.1 µM $NiCl_2 \cdot 6\ H_2O$, 2.1 µM $Na_2MoO_4 \cdot 2\ H_2O$, 1.9 µM $Na_2SeO_3 \cdot 5\ H_2O$. *E. coli* was grown overnight in minimal medium at 16°C and protein was purified as described above.

Due to the remainder of the His-tag (GSHM), the actual mIHF protein starts with amino acid 5 (underlined). The sequence of the protein used throughout this investigation is thus: GSHMVALPQLTDEQRAAALEKAAAARRARAELKDRLKRGGTNLTQVLKDAESDEVLGKMKVSALLEALP KVGKVKAQEIMTELEIAPTRRLRGLGDRQRKALLEKFGSA.

2.3 Circular dichroism (CD) measurement

Circular dichroism of mIHF was measured at a concentration of 10 – 20 µM protein in a quartz cuvette of 1 mm path length and analysed using a Jasco J-815 CD spectrometer. Data were converted to mean residual ellipticity and corrected for difference in concentration (Greenfield, 2007). Far UV (195 – 250 nm) was used for protein, and near UV (250 – 340 nm) for DNA absorbance. Thermal unfolding was monitored over the whole spectra. Molar ellipticity changes at 222 nm were used to fit a sigmoidal model and calculate the melting temperature (GraphPad Prism). DNA titration was carried out at 20°C with 0, 0.003, 0.006, 0.340, 0.500 and 0.700 equivalents of DNA. A 40 bp dsDNA oligonucleotide (5' – CTGGAGGAGCTGGCAGCAGCGTTTCCGGGTGATGGCTGGT – 3') was used. As mIHF does not contain amino acids that absorb at 280 nm, the signal was undisturbed in the transition between near and far UV.

2.4 Gel retardation assay

100 ng linear DNA was incubated with increasing amounts of recombinant mIHF (1 pmol – 2 nmol) and mIHF R90E (0.1 nmol – 2 nmol) for 10 min at room temperature and electrophoresed on a 0.75% agarose gel in 0.5x TBS (25 mM Tris-HCl pH 7.5, 50 mM NaCl). Post-run staining was performed with GelRed (Biotium) for 20 min at a 1:10´000 ratio of GelRed to 0.5x TBS.

2.5 Nuclear Magnetic Resonance (NMR) spectroscopy and solution structure determination

NMR measurements for protein assignment and structure determination were carried out on 1.34 mM uniformly ^{13}C- and ^{15}N-labelled mIHF samples in 100 µl, using Shigemi NMR tubes. Protein solution was prepared in 90% H_2O (50 mM phosphate buffer pH 7.5, 100 mM NaCl) and 10% $^{2}H_2O$. 0.2% sodium azide was added to prevent sample degradation. The NMR spectra were acquired at 14.1 T, 18.8 T and 23.5 T (600.55, 800.13 and 1000.30 MHz of proton Larmor frequency, respectively) on Bruker Avance III spectrometers, all equipped with a cryogenic cooled probehead. The experiments for determining the ^{1}H, ^{13}C and ^{15}N protein backbone resonance assignment were performed at 14.1 T, using the standard J-UNIO protocol (Serrano et al., 2012) automated assignment APSY routine. Standard CBCACONH (Hiller et al., 2008), HACACONH (Fiorito et al., 2006), and HACANH (Hiller et al., 2005) APSY experiments were used; the $\pi/2$ hard pulses were 11.18 µs 14.45 µs and 37.2 µs for ^{1}H, ^{13}C and ^{15}N, respectively. The number of projections acquired were 56, 56, and 41 with 16, 8, and 8 scans for CBCACONH, HACACONH, and HACANH, respectively, and with a direct acquisition time of 91.7 ms and a recycle delay of 0.9 s. The ^{1}H-^{15}N 3D and ^{1}H-^{13}C 3D HSQC NOESY spectra used to obtain side-chain resonance assignments and structural restraints for the solution structural determination were performed at 18.8 T and 288 K using standard NMR sequences. The $\pi/2$ hard pulses were 10.52 µs 13.6 µs and 35.0 µs for ^{1}H, ^{13}C and ^{15}N, respectively. The mixing times were 100 ms and 120 ms for ^{1}H-^{15}N 3D and ^{1}H-^{13}C 3D HSQC NOESY, respectively. Up to 1536 x 72 x 100 complex points were acquired for direct ^{1}H, indirect ^{15}N and ^{1}H dimensions, respectively; for ^{1}H-^{13}C 3D HSQC NOESY 1088 x 80 x 90 complex points were acquired for direct ^{1}H, indirect ^{13}C and ^{1}H dimensions, respectively. The number of scans were 8, with a direct acquisition time of 84.86 ms and a

recycle delay of 1.20 s. Data were processed with 1024 x 128 x 256 complex points matrix and square cosine window functions.

The analysis of the APSY experiments was performed with the software UNIO-MATCH for automated backbone resonance assignment (Volk et al., 2008) and UNIO-ATNOS/ASCAN (Fiorito et al., 2008) for automated side-chain chemical shift assignment. The input for automated NOESY peak picking and NOE assignments with UNIO-ATNOS/CANDID (Herrmann et al., 2002a, 2002b) consisted of the UNIO-MATCH chemical shift assignments for the polypeptide backbone, the UNIO-ATNOS/ASCAN output of side-chain chemical shift assignments, and the NOESY datasets. Table 1 reports NMR structural statistics from NOE assignment and structure calculation using UNIO-ATNOS/CANDID.

2.6 DNA titration experiments monitored by NMR

The ^{15}N-labeled protein was prepared at 200 μM in 50 mM phosphate buffer pH 7.5, 100 mM NaCl, with 10% $^{2}H_2O$. Titration of this sample with dsDNA (5'-AGCTCGTCAACGCCTT-3', 5 mM stock solution, purchased from MICROSYNTH) was monitored through ^{1}H,^{15}N HSQC spectra collected at 298 K on an AVANCEII-800 MHz spectrometer equipped with a CPTC ^{1}H,^{13}C,^{15}N 5 mm cryoprobe. The HSQC spectra were collected with 128 increments in the ^{15}N dimension. Since broadening was observed for most crosspeaks along the titration, the number of scans was increased throughout the titration as required to keep a workable signal-to-noise ratio (from 8 scans without DNA to 64 scans at 2.2 equivalents of added DNA).

All data for titrations were acquired and processed with a Bruker TOPSPIN 2.0 instrument and analysed with NMRFAM-SPARKY. For data analysis, the backbone ^{1}H and ^{15}N assignments derived from the automatic structure calculation procedure were first verified against 3D CBCANH and HN(CA)CO spectra. All assignments except that for Arg27 were confirmed, while assignments for Gly40, Arg92 and Asp96 were also retrieved. After this procedure the only non-proline residues missing H,N assignments are Gly1 through Met4, Leu19, Arg27 through Ala30 and Arg97. The final table of backbone H,N assignments is given in Fig. S1, CBCANH spectra in Fig. S2.

Chemical shift perturbations (CSP) were calculated from weighed differences in ^{1}H and ^{15}N chemical shifts (ΔδH and ΔδN, (Williamson, 2013)) as:

$$CSP = \sqrt{\frac{(\Delta\delta H)^2 + (\Delta\delta N/5)^2}{2}}$$

2.7 ^{15}N-Relaxation data

The backbone dynamics of mIHF was investigated through ^{15}N-relaxation measurements, acquired on the same sample described above for structure determination at 298 K, corrected for the different temperature. Standard experiments (Cavanagh et al., 2007) were used to measure the longitudinal and transverse ^{15}N relaxation times (^{1}H-^{15}N HSQC T_1 and ^{1}H-^{15}N HSQC T_2) (Kay et al., 1992) and the {^{1}H}-^{15}N NOE HSQC at 23.5 T (1.0 GHz of ^{1}H Larmor frequency). For these experiments, the $\pi/2$ hard pulses were 11.93 μs 13.75 μs and 26.5 μs for ^{1}H, ^{13}C and ^{15}N, respectively. The build-up times for the ^{15}N T_1 were 0.010 s, 0.010 s, 0.100 s, 0.300 s, 0.500 s, 0.700 s, 1.00 s, 1.30 s, 1.60 s, 1.60 s, with two experiments acquired twice. To measure ^{15}N T_2 a CPMG sequence was used with a τ-π-τ sequence of 450 μs for τ delay and 160 μs for the π-pulse. The CMPG build-up times were 17 ms, 34 ms, 51 ms, 68 ms, 68 ms, 102 ms, 136 ms, 170 ms, 204 ms with one experiment acquired twice. For each build-up point the number of scans were 32 with 1.50 s of recycle delay for the ^{1}H-^{15}N HSQC T_1, and 2.50 s of recycle delay for the ^{1}H-^{15}N HSQC T_2 to allow for the probe duty cycle. For the {^{1}H}-^{15}N NOE HSQC 64 scans were acquired for each experiment, with a saturation time of 6.0 s. Each point in the HSQC plane relaxation data was processed with a 2048 x 512 complex points matrix and square cosine windows function for both direct and indirect dimensions. Relaxation data were analysed with the PROTEIN DYNAMICS CENTER 2.4.6 (2016 Nov/14) software provided by BRUKER BIOSPIN.

2.8 Isothermal titration calorimetry

ITC experiments were performed using a Malvern MicroCal PEAQ-ITC calorimeter (Malvern Instruments) at 25°C with 19 injections of substrate addition. ITC cells were loaded with 300 µl of 5–10 µM mIHF in buffer A (10 mM Tris, pH 8.0, 50 mM NaCl). Different length DNA oligonucleotides were used for binding measurements: 7bp ssDNA (3´-CTCACGC-5´, 100-500 µM), 7bp dsDNA (100-250 µM), 16bp ssDNA (5'-AGCTCGTCAACGCCTT-3', 100-250 µM), 16bp dsDNA (100-500 µM), 40bp ssDNA (5´-CTGGAGGAGCTGGCAGCAGCGTTTCCGGGTGATGGCTGGT-3´, 100-250 µM), and 40bp dsDNA (100-500 µM). Data were analysed using Malvern PEAQ-ITC analysis software provided by Malvern instruments.

2.9 Atomic force microscopy

The *mihF* gene and promoter region were obtained by amplification with primers *mihF*-prom-F and *mihF*-prom-R. Cloning of the product into pTOPO (ThermoFisher) gave rise to pNO75. pNO75 was cut with SpeI and BamHI and the resulting linear fragment was purified with the PCR purification kit (Sigma-Aldrich). I95 cosmid DNA consists of a 42.6 kb region, coordinates 4,292,000 – 4,326,000 of the *M. tuberculosis* chromosome, cloned into pYUB412 and amplified in *E. coli* (Bange et al., 1999). Cosmid DNA was extracted using a Midi Prep kit (Promega) and eluted in TE buffer (10 mM Tris, 1 mM EDTA).

DNA at 0.5 ng μl^{-1} (final concentration) was mixed with varying concentrations of mIHF, brought to a final volume of 20 µl in buffer II (5 mM $MgCl_2$, 5 mM Tris-HCl pH 7.5, 50 mM NaCl, for linear fragment) or in ultrapure water (for cosmid I95). The mixture was deposited on freshly cleaved mica and incubated for 5 min at 25°C. For experiments with I95 cosmid, the mica surface was functionalized with APTES in a separate step prior to DNA deposition as described previously (Japaridze et al., 2016). AFM images were collected using a MultiMode SPM with a Nanoscope III controller (VEECO Instruments, Santa Barbara, CA, USA) operated in tapping-mode in air. The AFM cantilevers had a spring constant of 5 N/m (Bruker cantilevers, TAP150A) with resonance frequencies ranging between 120 and 160 kHz. All AFM images consist of 512 × 512 pixels with scan frequency ≤1 Hz. Each experiment was performed at least in duplicate and AFM images were obtained at several separate mica locations. Only DNA complexes that were completely visible in an AFM image were considered for statistical analysis. Images were simply flattened using the GWYDDION software (Version 2.22), and no further image processing was carried out (Nečas and Klapetek, 2012).

DNA molecules were traced using DNA TRACE software previously described (Mikhaylov et al., 2013). Based on the statistics of tens-to-hundreds of individual molecules the contour length and the radius of gyration were calculated by the software.

2.10 LC-MS/MS

Bands of interest were excised from SDS-PAGE gels and in-gel digested with trypsin or chymotrypsin. After gel extraction, samples were dried by vacuum centrifugation and analysed by LC-MS/MS. Peptides were separated on a Dionex Ultimate 3000nano UPLC system coupled in-line to a high-resolution mass spectrometer (ThermoFisher Scientific). A pre-column (ACCLAIM PEPMAP C18 Trap) was used to capture the samples and separation was

performed over a 90-min gradient using a capillary column (ACCLAIM PEPMAP C18 RSLC) at 250 nl min^{-1}. Positive data-dependent acquisition mode was used and precursor peptides with charge > 2 were fragmented. Database search was performed with Proteome Discoverer 1.4 on Mascot, Sequest and MS-Amanda against Tuberculist R27 database (Lew et al., 2011). Met oxidation, Ser-Thr-Tyr phosphorylation and peptide N-acetylation were set as variable modifications while Cys carbamidomethylation was set as fixed modification. Scaffold 4.75 was used for data inspection.

2.11 Accession ID

The atomic coordinates and structure of mIHF (Rv1388) have been deposited in the Protein Data Bank under the accession number xxxx.

3. Results

3.1 mIHF DNA binding domain prediction and mutant generation

The 105-amino acid mIHF protein has a calculated molecular weight of 11.5 kDa and contains no tryptophan, tyrosine or cysteine residues, rendering this protein much harder to detect by UV-Vis spectrophotometry. Amino acid sequence alignments (Fig. 1) show that mIHF is highly conserved among Actinobacteria. Conversely, mIHF does not share any sequence similarity with its orthologues from Gram-negative bacteria.

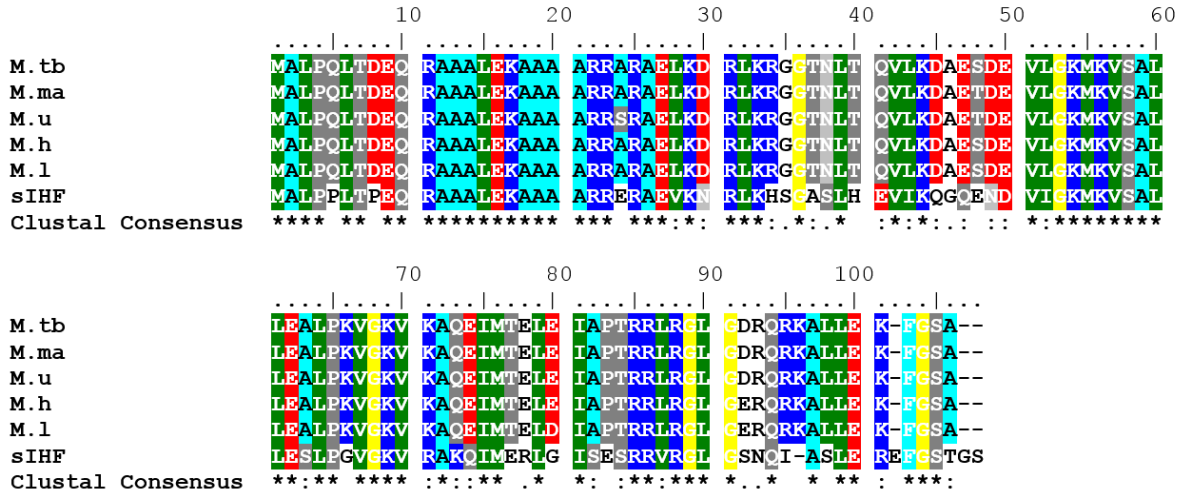

Fig. 1. Amino acid sequence alignment of IHF proteins. Amino acids are similarity coded using the BLOSUM62 matrix at 80% cutoff. mIHF from *M. tuberculosis* (M.tb) is aligned with protein sequences from *M. marinum* (M.ma), *M. ulcerans* (M.u), *M. haemophilum* (M.h), *M. leprae* (M.l) and sIHF from *S. coelicolor* (sIHF, SCO1480).

Secondary structure prediction revealed mIHF to be an alpha-helical protein, with the first helix (residues 8 – 34) predicted to form a coiled coil ((Lupas et al., 1991), Fig. S3). sIHF contains a helix-two turns-helix (H2TH) motif, where the two turns between the helices contact DNA at DBS-I of the protein. The second DBS lies on the opposite side of the protein, in a turn connecting two helices (Swiercz et al., 2013) and the proposed DNA binding mode suggests a new category of NAPs, characterized by two distinct DNA binding domains in the same protein rather than by dimerization of two proteins with one DBS each. sIHF is thus unique and not related to other NAPs like IHF or H-NS from *E. coli* or Lsr2 from *M. tuberculosis* (Swiercz et al., 2013).

Alignment of mIHF with sIHF showed high homology levels between the two sequences, with 62% identity and 79% similarity ((Altschul SF et al., 1990), BLOSUM62 matrix). These data suggested that the two proteins might have a very similar structure. We therefore exploited

the sIHF structure with the predicted DBSs (Swiercz et al., 2013) to identify the putative DBSs of mIHF. Overall, both domains were conserved but several interesting differences between sIHF and mIHF were noticed. For instance, positions G67 and G80 in sIHF were occupied by the positively charged lysine and the negatively charged glutamic acid residues, respectively, in mIHF. These changes might reflect the adaptation to specific DNA binding patterns in the respective genomes or modify sequence specificity. Furthermore, several positively charged arginine residues are present in the mIHF and sIHF DBS and are presumably responsible for the direct contact with the negatively charged DNA backbone. We chose R90, R92 and G93 conserved in both proteins, as targets of site-directed mutagenesis to investigate the DNA binding properties of the mutant proteins. Arginine was mutated to glutamic acid (R90E, R92E), and glycine to tryptophan (G93W). The mutant proteins containing R92E or G93W were not stable after purification, while mIHF R90E was as stable as mIHF wild type (wt) and therefore used for future experiments.

3.2 mIHF is a highly soluble, alpha helical DNA-binding protein

The mIHF protein was purified to homogeneity with a concentration of 60 mg ml^{-1} and displayed elevated stability at room temperature. The structural features of mIHF were characterized via circular dichroism (CD) spectroscopy. Fig. 2a illustrates the mean residue ellipticity of mIHF wt and its R90E mutant.

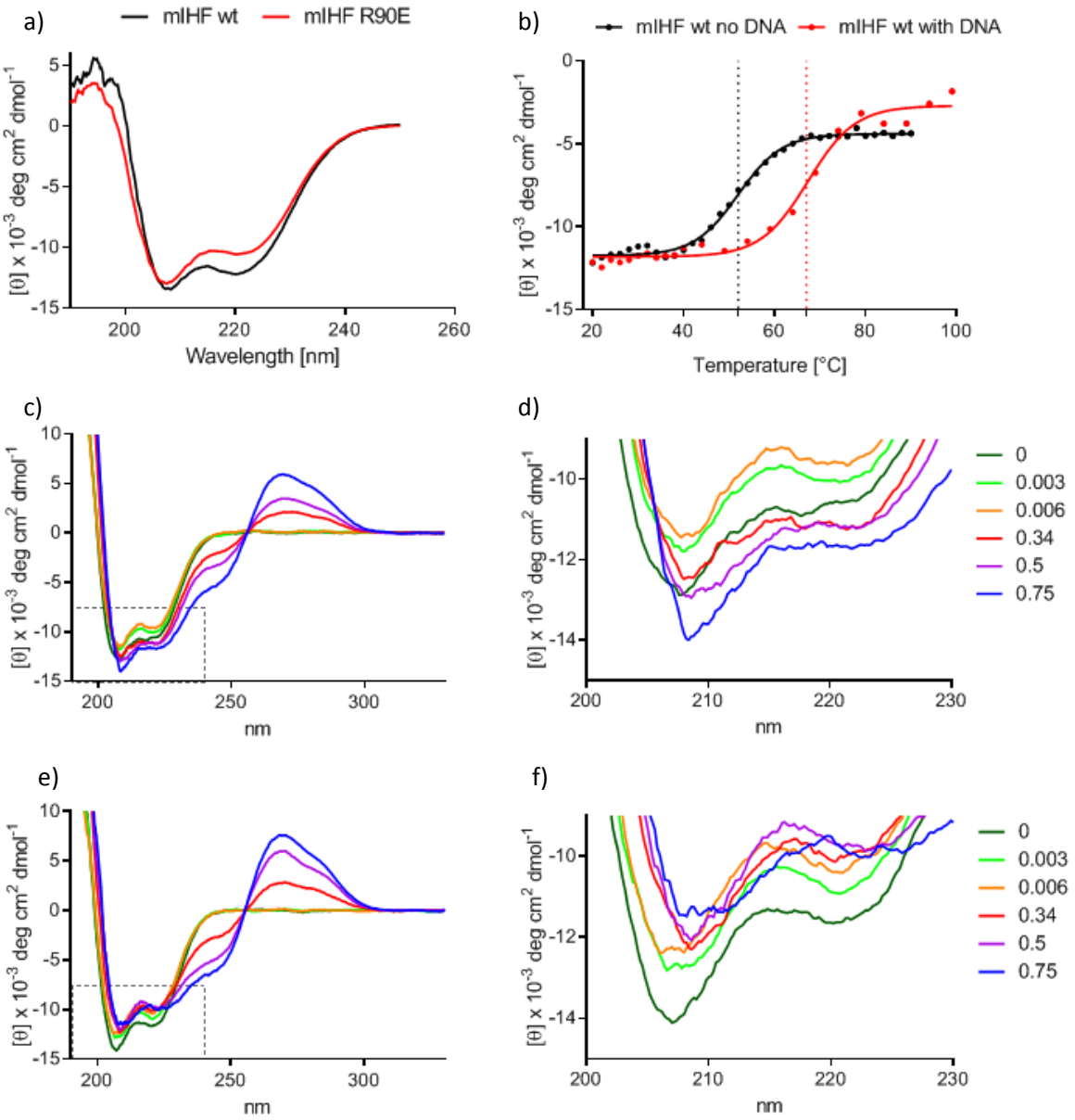

Fig. 2. Circular dichroism (CD) spectra of mIHF protein mutants. a) CD spectra of mIHF wt and mIHF R90E. b) Thermal unfolding of mIHF wt in absence and presence of dsDNA at 222 nm. Dotted lines indicate melting temperature with (red) and without (black) DNA. c) CD spectra of mIHF wt with DNA titration and zoom of minima representing alpha helices (d). Different DNA equivalents were used as indicated. e) CD spectra of mIHF R90E with DNA titration and zoom (f), as for mIHF wt.

Both proteins showed a folded structure dominated by alpha helices, as reflected in the two pronounced negative peaks at 208 and 222 nm. mIHF R90E displayed a minor deviation from mHIF wt, suggesting that their overall secondary structures are very similar but not completely identical. The CD signal at 222 nm, as a function of temperature, was fitted to a

sigmoidal non-linear model and the melting temperature was calculated as 52°C (Fig. 2b). The protein did not refold into its alpha helical structure at 20°C after heating to 95°C. The melting temperature increased to 67°C in the presence of DNA, suggesting higher stability of mIHF when bound to DNA (Fig. 2b). To screen for conformational changes in the mIHF protein upon DNA binding, the wt and mutant proteins were incubated with very low (0.003 and 0.006 equivalents) and increasingly higher amounts (0.34, 0.5, 0.75 equivalents) of a 40 bp dsDNA oligonucleotide. The minima at 208 nm and at 222 nm became less pronounced in mIHF wt at low amounts of DNA, indicating a small conformational change with no emergence of a different secondary structure. At higher DNA/protein ratios, a deeper minima at 208 nm appeared (Fig. 2c and zoom in Fig. 2d). mIHF R90E showed a similar pattern with a decrease in both 208 nm and 220 nm minima, but no subsequent increase at high DNA concentrations. To assess how the mIHF protein shifted linear DNA, a 700 bp DNA fragment was incubated with mIHF wt and mIHF R90E in a gel retardation assay. While mIHF wt shifted the DNA effectively, mIHF R90E had a markedly lower retardation effect (Fig. 3).

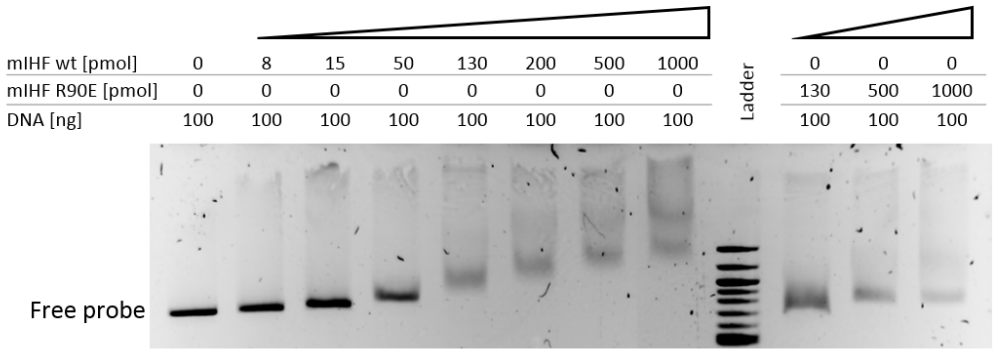

Fig. 3. Gel retardation assay obtained upon incubation of a 700 bp DNA fragment with mIHF or with mIHF R90E. DNA was incubated with increasing concentrations of mIHF as detailed in the image and run on a 0.75% agarose gel. Post-run staining was performed with GelRed. Ladder: 100 bp molecular weight marker.

This confirmed that R90 is important for DNA binding by mIHF. The minor DNA shift caused by mIHF R90E suggested that the mutant protein is still capable of binding DNA, confirming the CD results, but with a decreased DNA binding strength compared to mIHF wt. Thus, mIHF R90E represented a suitable control for other DNA-mIHF interaction experiments.

3.3 Solution structure of mIHF and protein dynamics

In order to define the exact binding mechanism and interaction with DNA, structure determination of mIHF was carried out in solution by NMR spectroscopy. High-dimensional automated projection spectroscopy (APSY) spectra for backbone assignment and 3D $^{15}N/^{13}C$-HSQC-NOESY spectroscopy to obtain side-chain resonance assignments and conformational restraints were acquired (Table 1).

Table 1. NMR structural statistics.

NMR constraints	
Distance	1818
Intra-residue (|*i-j*| = 0)	429
Inter-residue	
Sequential (|*i-j*|=1)	613
Medium range (1 ≤ |*i-j*| ≤ 4)	542
Long-range (|*i-j*| > 5)	234
TALOS+ dihedral angle restraints	181
Backbone RMSD from mean structure (residue 18 – 107) [Å]	0.99
Ramachandran analysis	
most favoured	85.2%
additional allowed	13.6%
generously allowed	1.0%
Disallowed	0.2%

The backbone and total atom assignments were completed to 90.24% and 76.71%, respectively. An additional four amino acids (GSHM) remained after removal of the His-tag; the mIHF protein therefore starts at residue number 5. Note that the NMR statistics (Table 1) concern the entire construct from residues 1 to 109, and the residue numbering in the following section refers to the protein construct used for the NMR study. mIHF was identified as a globular protein (Fig. 4a) for residues 12 – 107 with an RMSD (root mean square deviation of atomic positions) of 1.45 Å or 1.11 Å for residues 16 – 107, excluding the first four N-terminal helical residues which show increased conformational spread. The N-terminal domain of the mIHF protein (residues V5-T11) did not form a defined secondary

structure and was disordered (Fig. 4b) which is consistent with the NMR chemical shift-based secondary prediction.

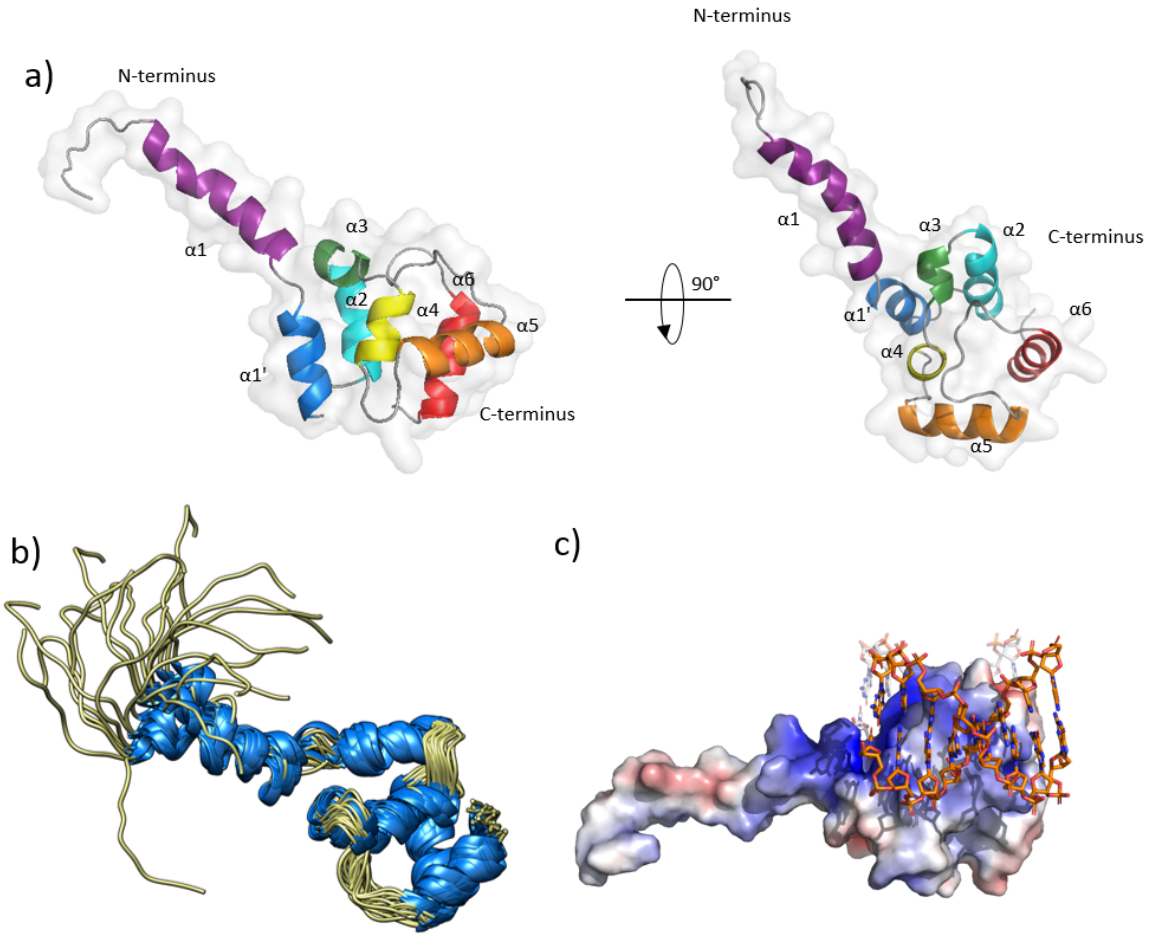

Fig. 4 Structure of mIHF. a) Ribbon drawing of mIHF (90° rotated around z-axis). b) Overlay of the 20 lowest energy structures of mIHF. c) Surface charge of mIHF, blue representing positively and red negatively charged surface of mIHF. DNA (red) is superimposed from PDPB ID 4ITQ after alignment of the sIHF protein to mIHF.

Overall, the NMR structure of mIHF resembled closely the sIHF X-ray structure with an overlay backbone RMSD of 1.62 Å and a similar topology with six α-helices. The flexible N-terminal end of mIHF is constituted by residues V5 to R15 followed by the first α-helix (α1) comprising A16 to R38. This helix was predicted to form a coiled coil and might be the site of dimerization for mIHF. A minor kink at residues A28 and R29 caused a slight divergence from an optimal helix. α1 is connected by a five-residue loop to α2, a short helix from L47 to E54. α3 only spans one turn from K57 to V61 and is followed by α4 from L64 – K70, α5 (Q77 – P87) and finally α6 (A101 – A109) (Fig. 4b).

Backbone assignments by UNIO-MATCH were completed to 90.24%. To probe the internal dynamics of mIHF and its aggregation state, three relaxation parameters for each backbone amide were characterized: the longitudinal and transverse ^{15}N relaxation rates $R_1 = 1/T_1$ and $R_2 = 1/T_2$, respectively, as well as the steady state heteronuclear Overhauser effect ($\{^1H\}$-^{15}N NOE). ^{15}N relaxation data are related to the correlation time of the reorientation motion of the H-N bonds due to protein dynamic in solution. For globular proteins in solution this is primarily due to the molecular rotational correlation time, which in turn is related to the molecular hydrodynamic radius. We can derive the molecular rotational correlation time from the T_1/T_2 ratio for the 44 most rigid residues (with $\{^1H\}$-^{15}N NOE >0.65, and neglecting the internal dynamics) that provided a rotational correlation time of 6.9 ± 1.7 ns (at 298 K), which is in line with a molecular weight of 10-12 kDa. In the supporting information the measured values of ^{15}N R_1, R_2 and $\{^1H\}$-^{15}N NOE were reported in table S2 and plotted in Fig. S4.

3.4 mIHF binds to double-stranded DNA and has two distinct DNA-binding domains

The ligand binding affinity of purified mIHF protein with dsDNA was determined by isothermal titration calorimetry (ITC). mIHF can bind dsDNA of different lengths, with affinity constants in the low micromolar range. The dissociation constant (K_D) values for 7, 16 and 40 bp dsDNA were 6.08, 2.39 and 9.16 µM, respectively (Fig. S5 a, b, c). Binding was only observed for dsDNA, but not for ssDNA oligonucleotides of the same lengths (Fig. S5 d, e, f).

To investigate the interaction of mIHF with dsDNA in structural terms, we performed a titration of 200 µM ^{15}N-labelled protein with 16 bp dsDNA and monitored the effects of DNA addition on mIHF by solution NMR (Figure 5). Specifically, we titrated a dsDNA molecule of sequence 5'-AGCTCGTCAACGCCTT-3' into ^{15}N-labelled protein and acquired ^1H–^{15}N HSQC spectra at 0, 0.25, 0.50, 1.0 and 2.20 equivalents (dsDNA/protein ratios). Analysis of the titration data relied on the backbone ^1H and ^{15}N assignments obtained from the automated assignment and structure calculation procedures, which we verified and completed by inspecting CBCANH and HN(CA)CO 3D spectra (See fig. S1 / S2 for confirmation of assignments). Of the non-proline residues, only the four N-terminal residues plus L19, R27 through A30 and R97 lack backbone H and N shifts in our final assignment, and hence were not included in the analysis.

Effects on mIHF HSQC spectra were evident already at 0.25 equivalents, as we observed the shifting and broadening of several crosspeaks. Figure 5A illustrates this on a region of the HSQC spectrum, with well-resolved crosspeaks, that exemplifies K70 broadening beyond detection and several other crosspeaks experiencing large shifts. Resonance broadening suggests exchange kinetics in the millisecond-microsecond timescale, which is consistent with an affinity in the micromolar range as determined by ITC. To better define the chemical shift perturbations, we continued the titration recording spectra at 0.50, 1.0 and 2.2 equivalents, increasing the number of scans as required to compensate for the broader signals. Figure 5B presents a plot of intensity ratios at 0.25 equivalent (where effects are already strong but most signals are still intense enough for quantification) and Figure 5C presents the chemical shift perturbations at 2.2 equivalents (where they are largest). Figure 5D in turn maps these intensity drops and chemical shift perturbations onto the mIHF structure. We observe clustering of the residues most affected by intensity loss or shifts in two sites, which is consistent with dsDNA bound to the two sites identified in the X-ray structure of sIHF bound to DNA (Swiercz et al., 2013). This is relevant because it means that the two sites identified for sIHF across the crystallographic unit cell are not merely an artefact of crystallization, but actual binding sites functional in solution. Residues K33, K37, E66, L68, K70, G72, K75 and A76 are all impacted by broadening (red in Figure 5D) and thus correspond to DBS-I of sIHF, together with K73, which experiences a large shift (blue). Residues K60, S62, L68, V71, M80, T81, T88, R89, L91, R92, G93, D96, Q98, were largely affected by chemical shift perturbation (blue in Figure 5D) and correspond to DBS-II proposed for sIHF together with R90 which experiences broadening (red). Qualitatively, it appears that binding to DBS-II proceeds at a faster exchange timescale, i.e. is of lower affinity than DBS-I.

Finally, from the results of our titration, we also observed some residues undergoing chemical shift perturbations and broadening outside of the DNA interaction sites, especially on one face of the N-terminal helix: A17, K21, A23, A24. Notably, this helix has some predicted coiled coil propensity (fig. S3), which could imply weak dimerization upon DNA binding.

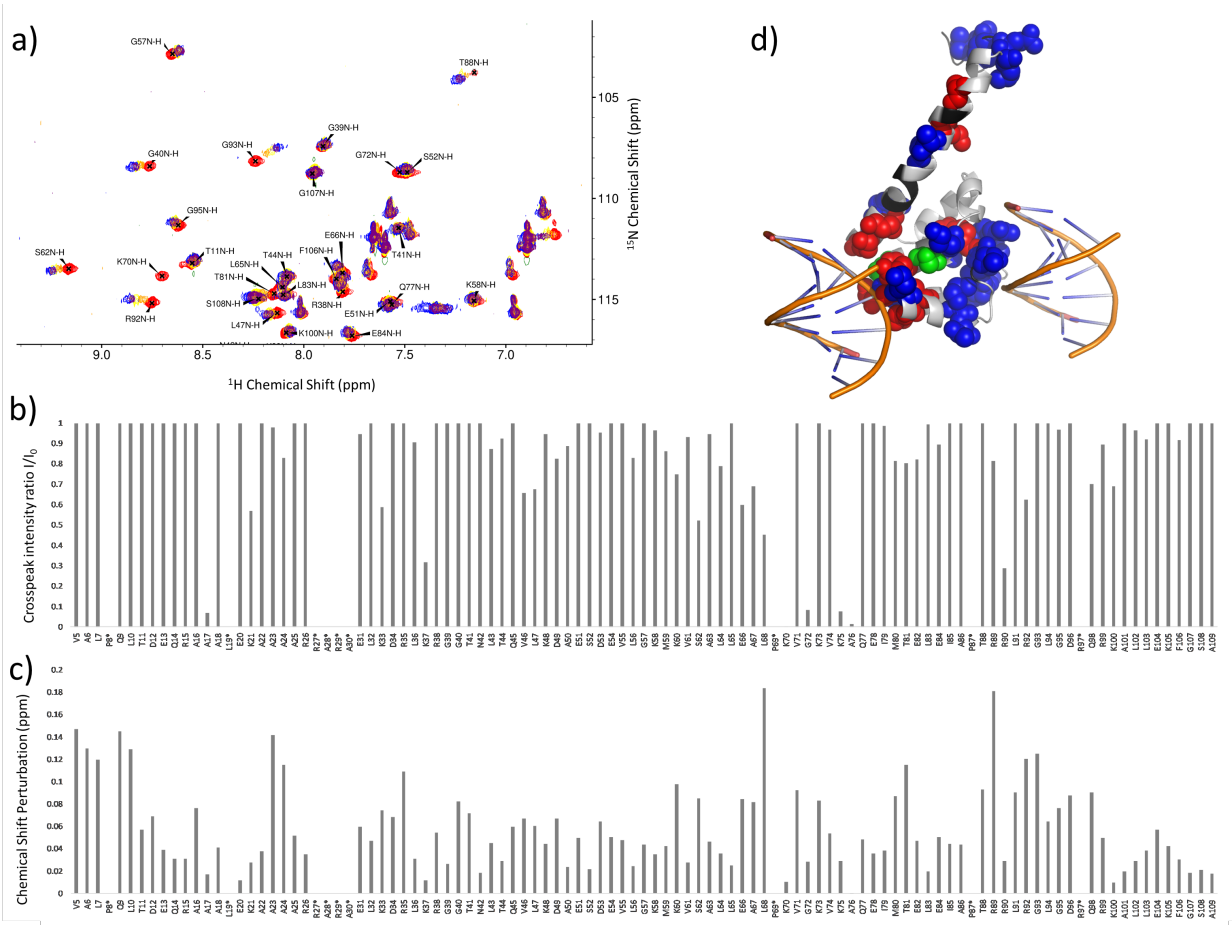

Fig. 5. Titration of double-strand DNA onto mIHF, monitored by ^{1}H,^{15}N HSQC spectra. a) Titration of ^{15}N-mIHF with DNA, followed through NMR spectroscopy at 800 MHz ^{1}H frequency. Overlay of ^{1}H-^{15}N HSQC spectra without DNA (red) and with 0.25 (orange), 0.50 (green), 1 (purple) and 2.2 equivalents of 16 bp dsDNA (blue). b) Residue-wise ratio of crosspeak intensities at 0.25 equivalents (DNA/protein) relative to the intensities in the spectrum of the free protein. (c) Residue-wise ^{1}H,^{15}N chemical shift perturbations at 2.2 DNA/protein equivalents relative to the free protein. In panels b and c, residues labeled with an asterisk lack H,N assignments or are prolines. (d) Mapping of residues displaying a sharp drop in the intensity ratios at 0.25 equivalents (red spheres, I/I0 < 0.6) and of residues experiencing large chemical shift perturbations (blue spheres, CSP > 0.08 ppm) on a cartoon representation of mIHF NMR structure. Green spheres are for residues that experienced both large intensity drop and large chemical shift perturbations. The two dsDNA molecules are overlaid from PDB ID 4ITQ after alignment of the sIHF protein to mIHF. Cartoon portions in black correspond to residues missing H,N assignments, marked with an asterisk in panels (b) and (c). A 3D interactive version of panel b is available at http://lucianoabriata.altervista.org/papersdata/mihfdnatitration.html

3.5 mIHF introduces defined loops into linear DNA fragments

To study mIHF binding to DNA, purified mIHF protein was incubated with DNA substrates of various length, topology and GC content, and imaged using AFM. The first substrate tested

was a 1 kb linear fragment containing the *mihF* gene and the upstream promoter region with a GC content of 65%, representing the mean GC content in *M. tuberculosis*. The DNA fragment alone showed a looping probability of 10%, which increased upon addition of mIHF and reached 53% when the maximum amount was added (Table S3, Fig. 6).

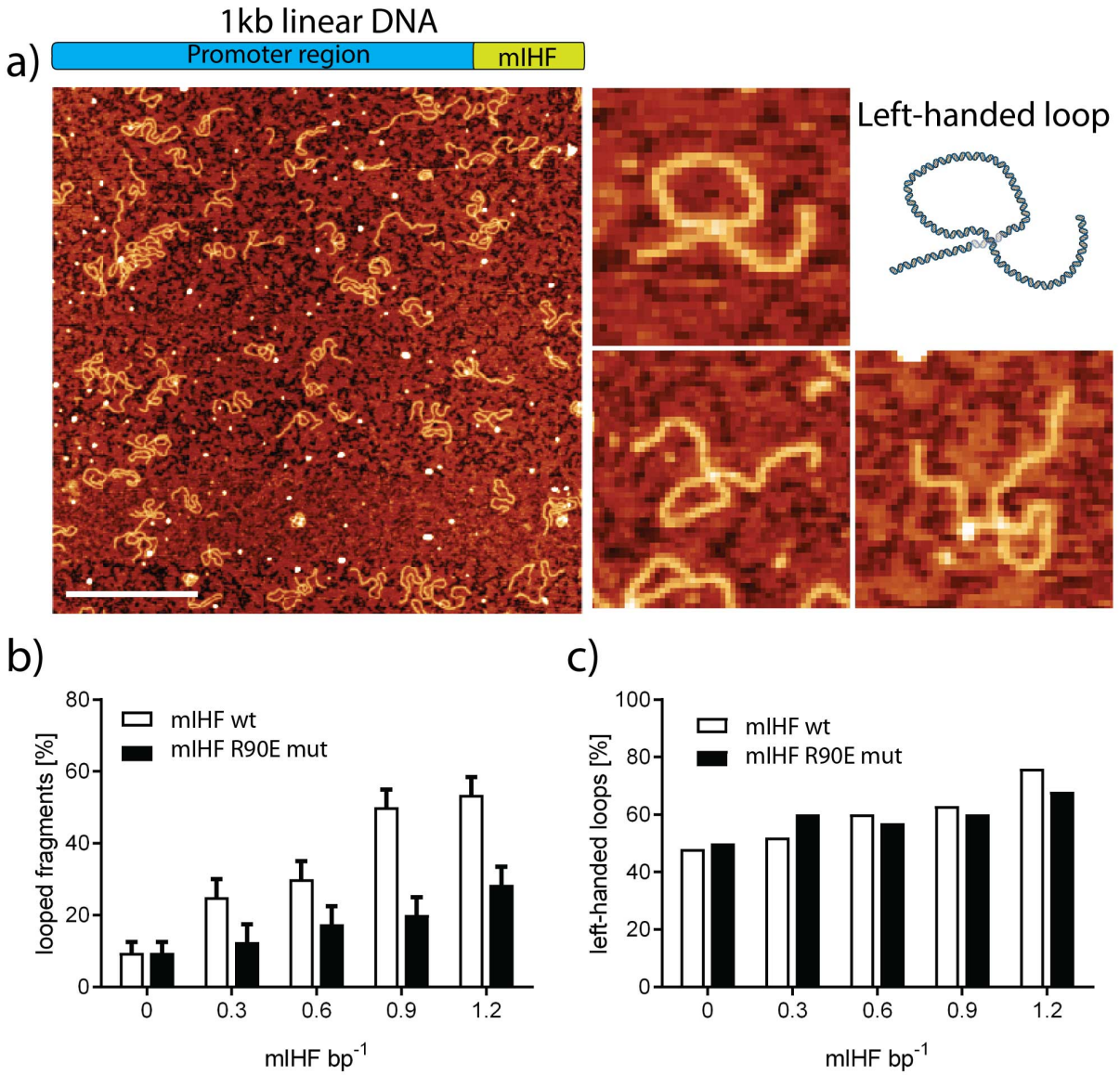

Fig. 6. Atomic force microscopy experiments. A 1 kb linear DNA fragment harbouring the *mihF* gene and its promoter region was incubated with the mIHF protein. a) Image of DNA fragments and zoomed images of looped structures at 0.21 mIHF bp^{-1}. Scale bar represents 500 nm. b) and c) Graphs representing the mean percentage ± standard deviation of looped fragments observed upon addition of increasing concentrations of mIHF wt and mIHF R90E and the percentage of left-handed loops.

The size of the loops was well defined with a mean length of 110 ± 10 nm, which corresponds to ca. 300 bp (Table S4). When the same experiment was repeated with mIHF R90E, loops in the linear DNA were observed as well but with a lower ratio compared to the wt protein. Only 28% looped fragments were detected upon binding of mIHF R90E at the same maximal concentration (Table S3, Fig. 6b). The average loop length was similar for mIHF and mIHF R90E with 110.5 ± 10 nm and 106.75 ± 10 nm, respectively (Table S4). These results indicated that mIHF R90E did exhibit DNA binding activity, but to a lower extent compared to the wt protein, confirming the gel retardation assay results. Naked DNA had approximately 50% left- and right-handed loops, corresponding to positive and negative supercoiling, respectively. mIHF wt, as well as mIHF R90E, increased the number of left-handed loops in a concentration-dependent manner to 76% for the wt protein and to 68% for the mutant (Fig. 6c, Table S5).

3.6 Large cosmids unwind and compact in an mIHF concentration dependent manner

The *M. tuberculosis* chromosome is a circular, supercoiled macromolecule of 4.4 Mb in length. Topology and size are not correctly represented by short linear DNA fragments or plasmids. Furthermore, bacterial chromatin is organized into microdomains of ca. 10 kb (Dame and Tark-Dame, 2016), a compartmentalization that cannot be mimicked by plasmids. Therefore, we used the 42.6 kb I95 supercoiled cosmid (Bange et al., 1999) to reproduce more closely the features of the genomic DNA. Supercoiled I95 forms highly ordered structures, the so-called hyperplectonemes, only present in DNA molecules above 30 kb in size (Japaridze et al., 2017). Incubation of I95 with mIHF wt at low concentrations (0.6 mIHF bp^{-1}) led to unwinding of the cosmid from the hyperplectonemic form and reduced its complexity (Fig. 7).

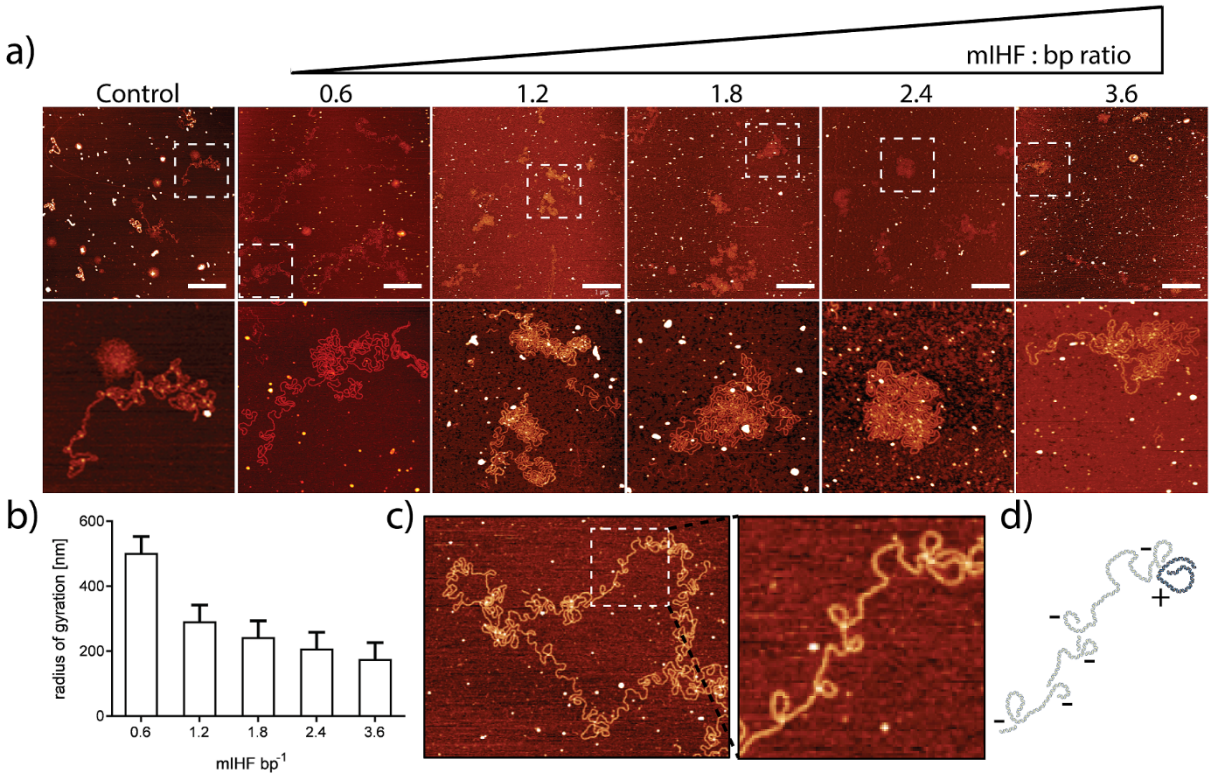

Fig. 7. Atomic force microscopy experiments with cosmid I95. a) Cosmid I95 alone and incubated with increasing concentrations of mIHF wt. The lower panels are zoomed images of the squares defined in the upper panels. Scale bar represents 1 µm and mIHF bp^{-1} ratios are indicated above images. b) Radius of gyration measured for cosmids in the presence of mIHF. Bars represent mean ± standard deviation. c) Cosmid I95 incubated with 1.2 mIHF bp^{-1}, zoom to an example structure and handedness of loops ("-" = left-handed, "+" = right-handed, (d)).

At higher mIHF ratios (1.2 to 3.6 mIHF bp^{-1}), the cosmid collapsed and condensed again, as illustrated by the decrease in radius of gyration (smallest circle that contains the DNA polymer, Fig. 7b, Table S6). Re-compaction of the cosmid at high mIHF ratios only reached first-order plectonemes, but did not lead to the formation of hyperplectonemes like those observed in the absence of protein. Similar to the loops observed on the linear DNA fragment, mIHF also introduced mostly left-handed loops in the cosmid, as illustrated in Fig. 7c and d. The R90E mutant simplified the topology of I95, but no collapse was visible at higher ratios of the protein, in agreement with the reduced DNA-binding observed with linear DNA (Fig. S6).

4. Discussion and conclusion

The importance of mIHF in *M. tuberculosis* physiology was suggested by Schubert and colleagues, whose mass spectrometry experiments predicted the protein to be among the ten most abundant proteins in the bacterium (Schubert et al., 2015). This initial indication was confirmed by our genetic studies, where we proved the essentiality of the gene and the role of the protein in controlling transcription of a variety of virulence factors and housekeeping genes (Odermatt et al., 2018). Here, we investigated the structure of mIHF and its DNA binding properties.

High similarity to the homologue sIHF from *S. coelicolor* was already evident from the primary structure alignment and further validated by the NMR structure, which differed for only 1.3 Å compared to the sIHF crystal structure. mIHF formed a globular protein with a protruding α-helix at the N-terminal end and the core of the protein consisted of five short α-helices.

ITC confirmed a K_D in the low micromolar range, corresponding to a medium-high affinity of the transcription factor for dsDNA. Further, the difference in K_D for different substrate lengths ranged only four-fold, suggesting no preference for DNA length. No binding was observed to ssDNA, and similar results were obtained by Swiercz *et al.*, who showed a much higher affinity of sIHF for dsDNA than for ssDNA (Swiercz et al., 2013).

ITC and NMR titrations of mIHF with DNA, and temperature stabilization of mIHF in presence of DNA, all clearly indicate that mIHF binds specifically double stranded DNA molecules. In particular, mapping of binding sites by NMR spectroscopy indicates binding on two sites that are consistent with those identified in the X-ray structure of sIHF. High-resolution AFM images showed that incubation of mIHF with a 1 kb linear DNA fragment introduced loops of a defined size. Such DNA loops are commonly associated with regulation of gene expression in bacteria, best described in the *lac* operon of *E. coli*, where LacI activates transcription by looping the promoter region and by constraining supercoils (Fulcrand et al., 2016). Although mIHF does appear to bind DNA in a sequence-independent manner, it affects the expression of many genes in *M. tuberculosis* (Odermatt et al., 2018). The mIHF protein might therefore have a direct effect on gene expression by looping the DNA and by bringing other regulatory elements closer to the target gene. Support for this is provided by the large number of loci occupied by mIHF and other NAPs, such as EspR (Odermatt et al., 2018). 76% of the loops

observed at the highest protein:DNA ratio (1.22 mIHF bp^{-1}) were of left-handed orientation. Naturally, DNA in bacterial cells is tightly packed in negative supercoils as exemplified by the I95 cosmid. Introduction of left-handed loops by mIHF relaxed and opened the cosmid, thus making it potentially accessible to replication and transcription machineries. Indeed, mIHF was found to act mainly as a transcriptional activator (Odermatt et al., 2018). This pattern is similar to that caused by *E. coli* HU on cosmid I95, where ordered, regular, left-handed loops were observed (Japaridze et al., 2017). Similarly to mIHF, *E. coli* HU binds without sequence specificity to DNA, but prefers gapped and cruciform DNA (Pinson et al., 1999). On the contrary, *M. tuberculosis* EspR introduced loops and additional twists into the DNA (Blasco et al., 2012). Therefore, it appears that mIHF and EspR have divergent effects on the DNA topology of the tubercle bacillus.

In conclusion, we showed here that mIHF has a globular structure, comprising six alpha helices, that is similar overall to that of sIHF despite the divergent sequences. The protein has two DNA binding sites that are functional in solution. DNA binding plays an important role in shaping the DNA structure by introducing left-handed loops. We propose these DNA binding capabilities could possibly be involved in promoting transcription.

5. Acknowledgments

We would like to thank the mass spectrometry facility at EPFL for their technical support and Stefanie Boy-Röttger for technical support. This work was supported by the Swiss National Science Foundation (grant 31003A-162641 to STC) and by H2020-INFRAIA iNEXT (grant 653706) for supporting the NMR instrument time in the Lyon CRMN platform.

6. Author contributions

NTO conceived the experiments, carried out the wet lab experiments, ML and TH determined the NMR structure of mIHF, LAA verified resonance assignments and performed and analysed the NMR titration, AJ and HV did the AFM studies, RS performed the ICT experiment, JP suggested the mutations, LE, GV and STC supervised the study. NTO, LAA and STC wrote the manuscript.

7. References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Balleza, E., López-Bojorquez, L.N., Martínez-Antonio, A., Resendis-Antonio, O., Lozada-Chávez, I., Balderas-Martínez, Y.I., Encarnación, S., Collado-Vides, J., 2009. Regulation by transcription factors in bacteria: Beyond description. FEMS Microbiol. Rev. 33, 133–151. https://doi.org/10.1111/j.1574-6976.2008.00145.x

Bange, F.C., Collins, F.M., Jacobs, W.R., 1999. Survival of mice infected with *Mycobacterium smegmatis* containing large DNA fragments from *Mycobacterium tuberculosis*. Tuber. Lung Dis. 79, 171–180. https://doi.org/10.1054/tuld.1998.0201

Bhowmick, Bhowmick, T., Ghosh, S., Dixit, K., Ganesan, V., Ramagopal, U. a, Dey, D., Sarma, S.P., Ramakumar, S., Nagaraja, V., 2014. Targeting *Mycobacterium tuberculosis* nucleoid-associated protein HU with structure-based inhibitors. Nat. Commun. 5, 4124. https://doi.org/10.1038/ncomms5124

Blasco, B., Chen, J.M., Hartkoorn, R., Sala, C., Uplekar, S., Rougemont, J., Pojer, F., Cole, S.T., 2012. Virulence regulator EspR of *Mycobacterium tuberculosis* is a nucleoid-associated protein. PLoS Pathog. 8, e1002621. https://doi.org/10.1371/journal.ppat.1002621

Blasco, B., Stenta, M., Alonso-Sarduy, L., Dietler, G., Peraro, M.D., Cole, S.T., Pojer, F., 2011. Atypical DNA recognition mechanism used by the EspR virulence regulator of *Mycobacterium tuberculosis*. Mol. Microbiol. 82, 251–64. https://doi.org/10.1111/j.1365-2958.2011.07813.x

Boshoff, H.I.M., Barry, C.E., 2005. Tuberculosis - Metabolism and respiration in the absence of growth. Nat. Rev. Microbiol. 3, 70–80. https://doi.org/10.1038/nrmicro1065

Cavanagh, J., Fairbrother, W., III, A.P., Skelton, N., 2007. Protein NMR Spectroscopy, 2nd Edition Principles and Practice, Current biology. https://doi.org/10.1016/B978-012164491-8/50005-1

Colangeli, R., Haq, a, Arcus, V.L., Summers, E., Magliozzo, R.S., McBride, a, Mitra, a K., Radjainia, M., Khajo, a, Jacobs, W.R., Salgame, P., Alland, D., 2009. The multifunctional

607 histone-like protein Lsr2 protects mycobacteria against reactive oxygen intermediates.
608 Proc. Natl. Acad. Sci. U. S. A. 106, 4414–8. https://doi.org/10.1073/pnas.0810126106

609 Dame, R.T., Tark-Dame, M., 2016. Bacterial chromatin: converging views at different scales.
610 Curr. Opin. Cell Biol. 40, 60–65. https://doi.org/10.1016/j.ceb.2016.02.015

611 Dillon, S.C., Dorman, C.J., 2010. Bacterial nucleoid-associated proteins, nucleoid structure
612 and gene expression. Nat. Rev. Microbiol. 8, 185–95.
613 https://doi.org/10.1038/nrmicro2261

614 Dorman, C.J., 2004. H-NS: a universal regulator for a dynamic genome. Nat. Rev. Microbiol.
615 2, 391–400. https://doi.org/10.1038/nrmicro883

616 Farer, L., Lowell, A., Meador, M., 1979. Extrapulmonary tuberculosis in the United States. Am
617 J Epidemiol 109, 205–217.
618 https://doi.org/https://doi.org/10.1093/oxfordjournals.aje.a112675

619 Fiorito, F., Herrmann, T., Damberger, F.F., Wüthrich, K., 2008. Automated amino acid side-
620 chain NMR assignment of proteins using 13C- and 15N-resolved 3D [1H, 1H]-NOESY. J.
621 Biomol. NMR 42, 23–33. https://doi.org/10.1007/s10858-008-9259-x

622 Fiorito, F., Hiller, S., Wider, G., Wüthrich, K., 2006. Automated resonance assignment of
623 proteins: 6D APSY-NMR. J. Biomol. NMR 35, 27–37. https://doi.org/10.1007/s10858-
624 006-0030-x

625 Fulcrand, G., Dages, S., Zhi, X., Chapagain, P., Gerstman, B.S., Dunlap, D., Leng, F., 2016. DNA
626 supercoiling, a critical signal regulating the basal expression of the lac operon in
627 *Escherichia coli*. Sci. Rep. 6, 19243. https://doi.org/10.1038/srep19243

628 Gomez, J.E., McKinney, J.D., 2004. *M. tuberculosis* persistence, latency, and drug tolerance.
629 Tuberculosis 84, 29–44. https://doi.org/10.1016/j.tube.2003.08.003

630 Gordon, B.R.G., Li, Y., Wang, L., Sintsova, A., van Bakel, H., Tian, S., Navarre, W.W., Xia, B.,
631 Liu, J., 2010. Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and
632 virulence genes in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U. S. A. 107, 5154–
633 9. https://doi.org/10.1073/pnas.0913551107

634 Greenfield, N.J., 2007. Using circular dichroism collected as a function of temperature to

635 determine the thermodynamics of protein unfolding and binding interactions. Nat.
636 Protoc. 1, 2527–2535. https://doi.org/10.1038/nprot.2006.204

637 Gröschel, M.I., Sayes, F., Simeone, R., Majlessi, L., Brosch, R., 2016. ESX secretion systems:
638 mycobacterial evolution to counter host immunity. Nat. Rev. Microbiol. 14, 677–691.
639 https://doi.org/10.1038/nrmicro.2016.131

640 Herrmann, T., Güntert, P., Wüthrich, K., 2002a. Protein NMR structure determination with
641 automated NOE-identification in the NOESY spectra using the new software ATNOS. J.
642 Biomol. NMR 24, 171–189. https://doi.org/10.1016/S0022-2836(02)00241-3

643 Herrmann, T., Güntert, P., Wüthrich, K., 2002b. Protein NMR structure determination with
644 automated NOE assignment using the new software CANDID and the torsion angle
645 dynamics algorithm DYANA. J. Mol. Biol. 319, 209–227. https://doi.org/10.1016/S0022-
646 2836(02)00241-3

647 Hiller, S., Fiorito, F., Wuthrich, K., Wider, G., 2005. Automated projection spectroscopy
648 (APSY). Proc. Natl. Acad. Sci. 102, 10876–10881.
649 https://doi.org/10.1073/pnas.0504818102

650 Hiller, S., Wider, G., Wüthrich, K., 2008. APSY-NMR with proteins: Practical aspects and
651 backbone assignment. J. Biomol. NMR 42, 179–195. https://doi.org/10.1007/s10858-
652 008-9266-y

653 Japaridze, A., Muskhelishvili, G., Benedetti, F., Gavriilidou, A.F.M., Zenobi, R., De Los Rios, P.,
654 Longo, G., Dietler, G., 2017. Hyperplectonemes: A Higher Order Compact and Dynamic
655 DNA Self-Organization. Nano Lett. acs.nanolett.6b05294.
656 https://doi.org/10.1021/acs.nanolett.6b05294

657 Japaridze, A., Vobornik, D., Lipiec, E., Cerreta, A., Szczerbinski, J., Zenobi, R., Dietler, G., 2016.
658 Toward an Effective Control of DNA's Submolecular Conformation on a Surface.
659 Macromolecules 49, 643–652. https://doi.org/10.1021/acs.macromol.5b01827

660 Jones, D.T., Buchan, D., Cozzetto, D., Bryson, K., 2017. PSIPRED [WWW Document]. URL
661 http://bioinf.cs.ucl.ac.uk/psipred/

662 Kay, L.E., Bull, T.E., Nicholson, L.K., Griesinger, C., Schwalbe, H., Bax, A., Torchia, D.A., 1992.

663 The measurement of heteronuclear transverse relaxation times in ax3 spin systems via
664 polarization-transfer techniques. J. Magn. Reson. 100, 538–558.
665 https://doi.org/10.1016/0022-2364(92)90058-F

666 Lew, J.M., Kapopoulou, A., Jones, L.M., Cole, S.T., 2011. TubercuList - 10 years after.
667 Tuberculosis 91, 1–7. https://doi.org/10.1016/j.tube.2010.09.008

668 Luijsterburg, M.S., Noom, M.C., Wuite, G.J.L., Dame, R.T., 2006. The architectural role of
669 nucleoid-associated proteins in the organization of bacterial chromatin: a molecular
670 perspective. J. Struct. Biol. 156, 262–72. https://doi.org/10.1016/j.jsb.2006.05.006

671 Lupas, A., Van Dyke, M., Stock, J., 1991. Predicting coiled coils from protein sequences.
672 Science (80-.). 252, 1162–1164. https://doi.org/10.1126/science.252.5009.1162

673 Manganelli, R., Proveddi, R., Rodrigue, S., Beaucher, J., Gaudreau, L., Smith, I., 2004. Sigma
674 Factors and Global Gene Regulation in *Mycobacterium tuberculosis*. Society 186, 895–
675 902. https://doi.org/10.1128/JB.186.4.895

676 Mikhaylov, A., Sekatskii, S.K., Dietler, G., 2013. DNA trace: A comprehensive software for
677 polymer image processing. J. Adv. Microsc. Res. 8, 241–245.
678 https://doi.org/10.1166/jamr.2013.1164

679 Mishra, A., Vij, M., Kumar, D., Taneja, V., Mondal, A.K., Bothra, A., Rao, V., Ganguli, M.,
680 Taneja, B., 2013. Integration Host Factor of *Mycobacterium tuberculosis*, mIHF,
681 Compacts DNA by a Bending Mechanism. PLoS One 8, e69985.
682 https://doi.org/10.1371/journal.pone.0069985

683 Nečas, D., Klapetek, P., 2012. Gwyddion: an open-source software for SPM data analysis.
684 Open Phys. 10, 181–188. https://doi.org/10.2478/s11534-011-0096-2

685 Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: A novel method for fast and
686 accurate multiple sequence alignment. J. Mol. Biol 302.

687 Odermatt, N.T., Sala, C., Benjak, A., Cole, S.T., 2018. Essential Nucleoid Associated Protein
688 mIHF (Rv1388) Controls Virulence and Housekeeping Genes in *Mycobacterium*
689 *tuberculosis*. Sci. Rep. 8, 1–14. https://doi.org/10.1038/s41598-018-32340-2

690 Pandey, S.D., Choudhury, M., Sritharan, M., 2014. Transcriptional regulation of

691 *Mycobacterium tuberculosis hupB* gene expression. Microbiology 160, 1637–1647.
692 https://doi.org/10.1099/mic.0.079640-0

693 Pinson, V., Takahashi, M., Rouviere-Yaniv, J., 1999. Differential binding of the *Escherichia coli*
694 HU, homodimeric forms and heterodimeric form to linear, gapped and cruciform DNA. J
695 Mol Biol 287, 485–497. https://doi.org/10.1006/jmbi.1999.2631

696 Schubert, O.T., Ludwig, C., Kogadeeva, M., Zimmermann, M., Rosenberger, G.,
697 Gengenbacher, M., Gillet, L.C.C., Collins, B.C.C., Röst, H.L., Kaufmann, S.H.E., Sauer, U.,
698 Aebersold, R., 2015. Absolute Proteome Composition and Dynamics during Dormancy
699 and Resuscitation of *Mycobacterium tuberculosis*. Cell Host Microbe 18, 96–108.
700 https://doi.org/10.1016/j.chom.2015.06.001

701 Serrano, P., Pedrini, B., Mohanty, B., Geralt, M., Herrmann, T., Wüthrich, K., 2012. The J-
702 UNIO protocol for automated protein structure determination by NMR in solution. J
703 Biomol NMR 53, 341–354. https://doi.org/10.1007/s10858-012-9645-2

704 Song, D., Loparo, J.J., 2015. Building bridges within the bacterial chromosome. Trends Genet.
705 31, 164–173. https://doi.org/10.1016/j.tig.2015.01.003

706 Summers, E.L., Meindl, K., Usón, I., Mitra, A.K., Radjainia, M., Colangeli, R., Alland, D., Arcus,
707 V.L., 2012. The structure of the oligomerization domain of Lsr2 from *Mycobacterium*
708 *tuberculosis* reveals a mechanism for chromosome organization and protection. PLoS
709 One 7, e38542. https://doi.org/10.1371/journal.pone.0038542

710 Swiercz, J.P., Nanji, T., Gloyd, M., Guarné, A., Elliot, M. a, 2013. A novel nucleoid-associated
711 protein specific to the actinobacteria. Nucleic Acids Res. 41, 4171–84.
712 https://doi.org/10.1093/nar/gkt095

713 Thomas A. Hall, 1999. BioEdit: a user-friendly biological sequence alignment editor and
714 analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
715 https://doi.org/10.1002/jps

716 Volk, J., Herrmann, T., Wüthrich, K., 2008. Automated sequence-specific protein NMR
717 assignment using the memetic algorithm MATCH. J. Biomol. NMR 41, 127–138.
718 https://doi.org/10.1007/s10858-008-9243-5

WHO, 2016. Global Tuberculosis Report 2016. Cdc 2016 214. https://doi.org/ISBN 978 92 4 156539 4

Williamson, M.P., 2013. Using chemical shift perturbation to characterise ligand binding. Prog. Nucl. Magn. Reson. Spectrosc. 73, 1–16. https://doi.org/10.1016/j.pnmrs.2013.02.001

Zondervan, N.A., Van Dam, J.C.J., Schaap, P.J., Dos Santos, V.A.P.M., Suarez-Diez, M., 2018. Regulation of three virulence strategies of *Mycobacterium tuberculosis*: A success story. Int. J. Mol. Sci. 19, 9–12. https://doi.org/10.3390/ijms19020347

Zumla, A., Nahid, P., Cole, S.T., 2013. Advances in the development of new tuberculosis drugs and treatment regimens. Nat. Rev. Drug Discov. 12, 388–404. https://doi.org/10.1038/nrd4001