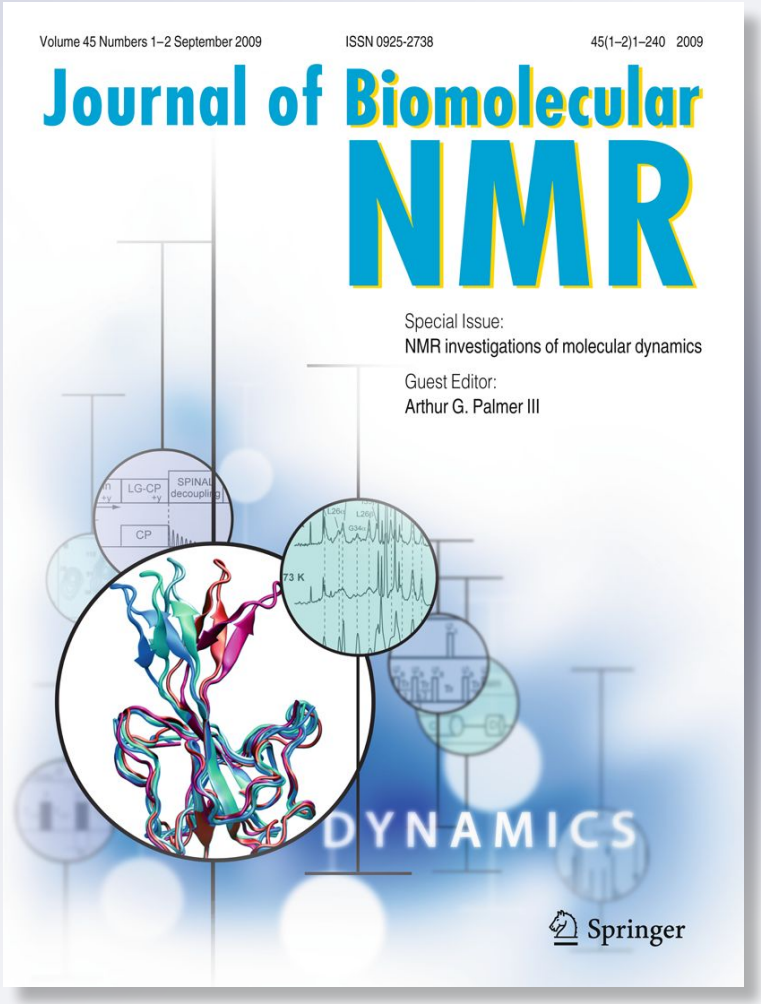
NMR structure note: N-terminal domain of Thermus thermophilus *CdnL*

Aranzazu Gallego-García, Yasmina Mirassou, Montserrat Elías-Arnanz, S. Padmanabhan & M. Angeles Jiménez

Journal of Biomolecular NMR

ISSN 0925-2738
Volume 53
Number 4

J Biomol NMR (2012) 53:355-363
DOI 10.1007/s10858-012-9648-z



Springer

J Biomol NMR (2012) 53:355–363
DOI 10.1007/s10858-012-9648-z

NMR STRUCTURE NOTE

NMR structure note: N-terminal domain of *Thermus thermophilus* CdnL

Aranzazu Gallego-García · Yasmina Mirassou · Montserrat Elías-Arnanz · S. Padmanabhan · M. Angeles Jiménez

Received: 3 April 2012 / Accepted: 19 June 2012 / Published online: 11 July 2012

Biological context

The CarD_CdnL_TRCF protein family (PF02559 in the protein family database; http://pfam.sanger.ac.uk) consists of members that are widely (and exclusively) distributed in bacteria, and for many of these proteins the functions remain to be fully characterized. The family is defined by the ~180-residue N-terminal domain of CarD, a global transcriptional regulator in the Gram-negative soil bacterium *Myxococcus xanthus* required in the activation of light- and starvation-induced genes as well as in other processes (Padmanabhan et al. 2001; Cayuela et al. 2003; García-Moreno et al. 2010). This domain in CarD interacts with CarG, a zinc-associated factor essential in every CarD-dependent process (Penalver-Mellado et al. 2006), and with a specific domain in the RNA polymerase (RNAP) β-subunit (García-Moreno et al. 2010). CarD can also bind to DNA via an intrinsically unfolded, ~140-residue C-terminal domain resembling eukaryotic high mobility group A (HMGA) proteins (Padmanabhan et al. 2001). Besides CarD and its orthologs (found so far only in myxobacteria; Elías-Arnanz et al. 2010), the CarD_CdnL_TRCF family consists of two other classes of proteins. One class corresponds to that formed by the RNAP-interacting domain, RID, that spans an ~70-residue segment of the much larger (>1,000 residues) transcription-repair coupling factor or TRCF, a widely conserved multidomain protein that mediates transcription-coupled repair of DNA lesions encountered by the transcribing complex in bacteria (Selby and Sancar 1995; Deaconescu et al. 2006). The other class within the CarD_CdnL_TRCF family includes several proteins that are typically 150–200 residues long (and so considerably smaller than CarD or TRCF orthologs) and have been denoted as CdnL (for CarD N-terminus-Like) to distinguish them from CarD (Cayuela et al. 2003; García-Moreno et al. 2010). CdnL has been found to be essential for cell viability and growth in nearly all of the few bacterial species where it has been examined (García-Moreno et al. 2010; Stallings et al. 2009). However, the functions of CdnL and its mode of action remain enigmatic. In *M. xanthus*, CdnL has been shown to be functionally distinct from CarD, and to interact with the same domain of RNAP-β as CarD but not with CarG. Despite lacking the ability to directly bind DNA, CdnL localizes to the nucleoid in vivo and directly or indirectly affects cell division (García-Moreno et al. 2010). Mycobacterial CdnL has been reported to also interact with RNAP-β and has been implicated in regulating ribosomal RNA transcription (Stallings et al. 2009). In bacteria where it is present, CdnL often coexists with TRCF and, in myxobacteria, with CarD as well. Since all three proteins interact with the same RNAP-β domain, understanding the molecular details for these interactions would provide significant insights into the interplay among them and their functional differences.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-012-9648-z) contains supplementary material, which is available to authorized users.

A. Gallego-García · M. Elías-Arnanz
Departamento de Genética y Microbiología, Área de Genética (Unidad Asociada al IQFR-CSIC), Facultad de Biología, Universidad de Murcia, Regional Campus of International Excellence "Campus Mare Nostrum", 30100 Murcia, Spain

Y. Mirassou · S. Padmanabhan (✉) · M. A. Jiménez (✉)
Instituto de Química Física Rocasolano, CSIC, Serrano 119, 28006 Madrid, Spain
e-mail: padhu@iqfr.csic.es

M. A. Jiménez
e-mail: majimenez@iqfr.csic.es
URL: http://rmnpro.iqfr.csic.es/

High-resolution structural data are available for *E. coli* TRCF (Deaconescu et al. 2006) and for a complex of the 71-residue *Thermus thermophilus* TRCF-RID with an RNAP-β domain from *Thermus aquaticus* that is near-identical to that in *T. thermophilus* (Westblade et al. 2010). *E. coli* lacks a CdnL homolog whereas *T. thermophilus*, in contrast, possesses one that we hereafter denote as TtCdnL. The function of TtCdnL in *T. thermophilus* remains to be examined, but it has been shown to interact with RNAP-β via its N-terminal region, like *M. xanthus* CdnL (Stallings et al. 2009; our unpublished data). In this study, we determined the solution structure of this 67-residue N-terminal domain, TtCdnLNt, by NMR to compare it with the structures reported for TRCF-RID of *T. thermophilus* (in its complex with the RNAP-β domain) and of *E. coli* in full-length TRCF. We also analyzed using a bacterial two-hybrid system how mutating specific residues to Ala in TtCdnLNt or in the RNAP-β domain affects the interaction between them, enabling a comparison with the contacts observed in the structure of the complex formed by *T. thermophilus* TRCF-RID and the RNAP-β domain. The data thus provide structural descriptions of a crucial module in CarD_CdnL_TRCF family and its interactions with RNAP.

Methods and results

Protein expression and purification

Escherichia coli strain DH5α was used for plasmid constructions and BL21-DE3 for protein overexpression. The coding region for the 67 N-terminal residues of TtCdnL followed by a TAA stop codon was PCR-amplified with an NdeI site at the 5′ end and an EcoRI site at the 3′ end, purified, and cloned into these sites in pTYB12 (New England Biolabs) to generate the construct used to overexpress intein-tagged TtCdnLNt. For this, a 10 mL starter culture of freshly transformed *E. coli* BL21(DE3) containing the pTYB12 construct was grown at 37 °C in Luria Broth (LB) medium with 100 μg/mL of ampicillin (Amp) to an OD_{600} of 0.6–1.0. It was added to 1 L of fresh LB/Amp, grown at 37 °C to an OD_{600} of 0.6, and after an hour incubation at 18 °C, overexpression of intein-tagged TtCdnLNt was induced overnight at 18 °C with 0.5 mM isopropyl ß-D-thiogalactoside (IPTG). To overexpress [^{13}C,^{15}N]-labeled intein-tagged TtCdnLNt, the cells from the 1 L culture (on reaching an OD_{600} of 0.6) were pelleted by centrifugation for 15 min at 5000×*g*, washed twice with 125 mL MOPS medium, and suspended in 1 L MOPS medium supplied with 1 g/L $^{15}NH_4Cl$ and 2.5 g/L $^{13}C_6$-glucose as the sole nitrogen and carbon sources (together with Amp, trace amounts of metal salts, biotin and thiamine), grown for 1 h at 18 °C, and induced overnight at 18 °C with 0.5 mM IPTG. After overnight induction with IPTG, cells were harvested by centrifugation (15 min at 5000×*g*) and the pellet was stored at −70 °C until further use. Intein-tagged unlabeled or [^{13}C,^{15}N]-labeled TtCdnLNt was purified using chitin resin and the intein was removed by on-column intramolecular cleavage in the presence of 50 mM dithiothreitol using the IMPACT kit protocols (New England Biolabs). The cleaved protein was passed through a small amount of chitin resin a second time to remove residual intein and dialyzed extensively against 100 mM NaCl, 50 mM phosphate buffer, pH 7.0, and 0.05 % NaN_3, and concentrated using Amicon Ultra (molecular weight cut-off 3,000 Da). The identity of purified TtCdnLNt (with an additional N-terminal AGH resulting after intein tag removal) was confirmed by mass spectrometry, and its concentration was determined using the BioRad protein assay kit and/or the absorbance at 280 nm (ε_{280} = 5,960 $M^{-1}cm^{-1}$).

Bacterial two-hybrid analysis

Interaction of TtCdnLNt with the RNAP-β fragment between residues 17 and 139 (Tt$\beta_{17–139}$) was tested in vivo using an *E. coli* two-hybrid system, in which interaction between two test proteins leads to functional complementation of the T25 and T18 fragments of the *Bordetella pertussis* adenylate cyclase catalytic domain (Karimova et al. 2000). Coding regions of the wild-type protein domains were PCR-amplified from genomic DNA, while versions bearing Ala mutations of specific residues in TtCdnLNt or in Tt$\beta_{17–139}$ were synthesized by GeneScript Inc (USA). These were then cloned into the XbaI and BamHI sites of pKT25 (for TtCdnLNt and variants) and pUT18C (for Tt$\beta_{17–139}$ and variants). Given pairs of the pKT25-pUT18 constructs were electroporated into *E. coli* strain BTH101 (adenylate cyclase deficient or *cya*$^-$). Pairs in which only one fusion protein was expressed served as the negative control, while the positive control was the GCN4 leucine zipper. Interaction was assessed qualitatively from the blue colour developed on LB plates containing 40 μg/mL X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside), and quantitatively by measuring β-galactosidase (β-gal) specific activity.

NMR spectra acquisition

NMR samples were prepared at 0.5–1 mM protein concentration in either 0.5 mL or 0.2 mL of H_2O/D_2O (9:1 ratio by volume) or in pure D_2O containing 100 mM NaCl, 50 mM sodium phosphate buffer (pH 7.0), and 0.05 % NaN_3. The pH value was checked with a glass microelectrode and was not corrected for isotope effects. NMR

spectra were acquired on Bruker AV 600 MHz and 800 MHz spectrometers both equipped with z-gradient cryoprobes. A methanol sample was employed to calibrate the NMR probe temperature. ^{1}H chemical shifts were referenced to the internal sodium 2,2-dimethyl-2-silapentane-5-sulphonate (DSS), and the ^{1}H spectrometer frequency assigned to 0 ppm was multiplied by 0.251449530 and 0.101329118, respectively, to indirectly reference the ^{13}C and ^{15}N chemical shifts (Markley et al. 1998). All NMR spectra were processed using TOPSPIN software (Bruker Biospin, Karlsruhe, Germany) and analyzed with Sparky (T. D. Goddard and D. G. Kneller, Sparky 3, University of California, San Francisco, USA). 2D [^{1}H–^{15}N]-HSQC and 3D HNCO, HNCA, CBCANH, CBCAcoNH, HBHANH and HBHAcoNH experiments were acquired in H_2O/D_2O (9:1 vol/vol), while 2D [^{1}H–^{13}C]-HSQC, 3D HCCH-TOCSY and 3D NOESY-[^{1}H–^{13}C]-HSQC spectra (mixing time = 80 ms) were recorded in D_2O using [^{13}C, ^{15}N]-labeled samples. 2D homonuclear [^{1}H,^{1}H]-COSY, [^{1}H,^{1}H]-TOCSY (mixing time = 60 ms) and [^{1}H,^{1}H]-NOESY (mixing time = 150 ms) spectra were recorded in both H_2O/D_2O (9:1 vol/vol) and in D_2O for the unlabeled samples. Heteronuclear ^{15}N-{^{1}H} NOEs for backbone amides were estimated from the peak intensity ratios in [^{1}H–^{15}N]-HSQC data with and without NOE, recorded for a 1 mM [^{13}C, ^{15}N]-labeled TtCdnL sample in H_2O/D_2O 9:1 v/v at 800 MHz.

NMR chemical shift assignment

[^{1}H–^{15}N]-HSQC spectra acquired for [^{13}C,^{15}N]-TtCdnLNt in aqueous solution at 25 °C are characterized by well-dispersed cross-peaks (Fig. 1) that were readily assigned using standard procedures. Cross-peaks for all the amide groups, except for residues S30, S32, R47 and S59, and the AGH N-terminal-tag (that remains after intein cleavage) could be identified in the [^{1}H–^{15}N]-HSQC spectra. Cross-peaks for a 15-residue "N-extein" peptide generated on intein cleavage (IMPACT kit protocols, New England Biolabs) but not fully eliminated despite extensive dialysis were also observed. This peptide, which could be eliminated by gel filtration, has narrow NMR line-widths, chemical shifts very close to reference random coil peptide values, and negative or near-zero ^{15}N-{^{1}H}-NOEs (Supplementary Material (SM) Figure SF1 and Table ST1), suggesting that it is disordered and does not interact with TtCdnLNt. Based on the analyses of a series of 3D NMR spectra (CBCANH, CBCAcoNH, HBHANH, HBHAcoNH, HNCO, HNCA and HNHA; Sattler et al. 1999) recorded with the [^{15}N,^{13}C]-TtCdnLNt in aqueous solution, we achieved near-complete assignment of all backbone (^{15}N, ^{1}HN, $^{1}H_\alpha$, $^{13}C_\alpha$ and ^{13}C') and $^{13}C_\beta$ atoms. The $^{1}H_\alpha$, $^{13}C_\alpha$, ^{13}C' and $^{13}C_\beta$ atoms of the residues whose amide groups

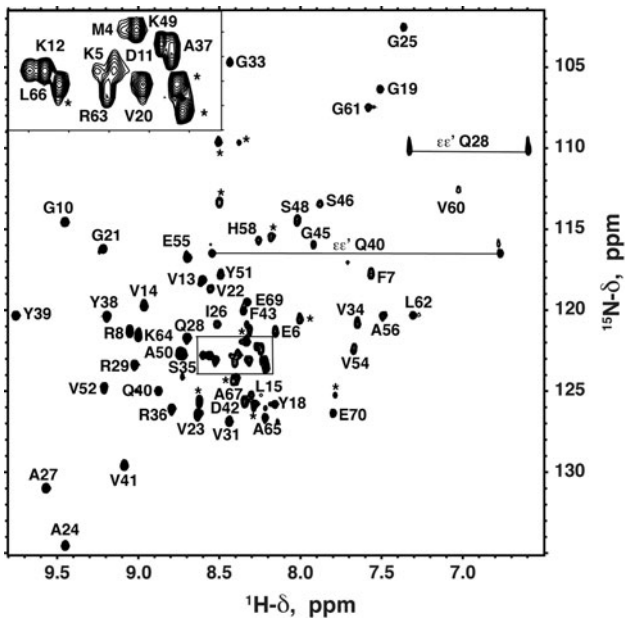


Fig. 1 2D [^{1}H,^{15}N]-HSQC spectra of [^{15}N,^{13}C]-TtCdnLNt. The *boxed* crowded region is shown expanded as an inset. εε' denotes side chain amide protons. Residues are numbered including the three nonnative ones (AGH) that are retained after intramolecular cleavage of the intein-fusion protein. *Asterisks* indicate cross-peaks of a 15-residue peptide generated on intein cleavage that persisted despite extensive dialysis (see text)

were not present in the [^{1}H–^{15}N]-HSQC spectra, as well as the two Q amide side chains, including the ^{13}C carbonyl groups, were also identified from that analysis. The assignment was extended to other side chain atoms, ^{1}H and ^{13}C, by joint analysis of 3D HCCH-TOCSY spectra recorded in D_2O solution, and 2D [^{1}H–^{1}H]-TOCSY and [^{1}H–^{1}H]-COSY spectra acquired for non-labeled TtCdnLNt in both H_2O/D_2O 9:1 v/v and D_2O solutions. All of the ^{1}H and ^{13}C resonances of the seven P residues in TtCdnLNt were assigned starting from the $^{13}C_\alpha$, $^{13}C_\beta$, $^{1}H_\alpha$ and/or $^{1}H_{\beta\beta'}$ identified in the 3D spectra, as we have previously described for another P-rich protein (León et al. 2009). Based on the chemical shift difference between $^{13}C_\beta$ and $^{13}C_\gamma$ carbons (Schubert et al. 2002), the X-P bonds are in the *trans* conformation ($\Delta\delta_{C\beta-C\gamma}$ = 2.4–4.7 ppm) for all P except P17, where it is *cis* ($\Delta\delta_{C\beta-C\gamma}$ = 9.4 ppm). This was confirmed by the sequential NOE observed between the H_α of P16 and P17, and the NOEs between the $H_{\delta\delta'}$ of P9, P16, P44, P53, P57 and P68, and the H_α of the preceding residue to the N-terminus of the given P in the 3D NOESY-[^{1}H–^{13}C]-HSQC and 2D [^{1}H–^{1}H]-NOESY spectra acquired in D_2O. The ring protons of the aromatic residues (F7, Y19, Y39, Y40, F44, Y52) were fully identified by examining [^{1}H–^{1}H]-COSY, [^{1}H–^{1}H]-TOCSY, [^{1}H–^{1}H]-NOESY and [^{1}H–^{13}C]-HSQC 2D-spectra recorded for the unlabeled TtCdnLNt sample in D_2O. In sum, more than 98 % of the total ^{1}H, ^{13}C and ^{15}N resonances of TtCdnLNt

were assigned, and these have been deposited at BioMagResBank (http://www.bmrb.wisc.edu/; accession code BMRB-18193).

NMR structure calculation

Distance constraints for structure calculation were derived from the 80 ms 3D NOESY-[^{1}H-^{13}C]-HSQC spectrum recorded for [^{15}N,^{13}C]-TtCdnLNt in D_2O, and two 150 ms 2D [^{1}H-^{1}H]-NOESY spectra acquired for unlabeled TtCdnLNt in H_2O/D_2O 9:1 v/v and in D_2O, respectively. Dihedral angle restraints for ϕ and ψ angles were obtained from $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$ $^{13}C'$ and ^{15}N chemical shifts using the program TALOS (Cornilescu et al. 1999). Structure calculations were carried out in three stages. First, we used the program CYANA 2.1 to run the standard iterative protocol for automatic NOE assignment consisting of seven cycles of combined automated NOE assignment and structure calculation of 100 conformers per cycle (Güntert 2004). Stereospecific assignments for the $H_{\beta}/H_{\beta'}$ protons of 9 residues (F7, R8, D11, Y18, Y38, Y39, F43, P53 and P57), the $H_{\gamma}/H_{\gamma'}$ protons of 2 residues (R8 and R36), the γ and γ' methyl groups of four V (V23, V41, V52 and V54), and the amide side chain protons of the two Q (Q28 and Q40) were done in the last cycle. The list of distance constraints resulting from the last automatic cycle was checked by inspection of the three NOESY spectra, and ambiguous constraints were removed or relaxed to generate the final list used as input for a standard simulated annealing CYANA 2.1 calculation of 100 conformers. The 20 conformers with the lowest target function values were selected and subjected to 2,000 steps of energy minimization using the generalised Born continuum solvation model with a non-bonded cut-off of 10 Å as implemented in the program AMBER9 (Case DA, Darden TA, Cheatham III TE, University of California, San Francisco, 2006). Table 1 lists the structural statistics for this final ensemble of 20 TtCdnLNt structures (Fig. 2). The quality of these final structures was assessed using PROCHECK/NMR (Laskowski et al. 1996) as implemented at the Protein Structure Validation Suite server (PSVS server: http://psvs-1_4-dev.nesg.org/). All of the residues were either in the most favored or allowed regions of the Ramachandran map (Table 1). The calculated TtCdnLNt structural ensemble has been deposited at the PDB data bank with accession code 2LQK.

The solution structure of TtCdnLNt

TtCdnLNt structural features were examined using MOLMOL (Koradi et al. 1996), PROCHECK/NMR (Laskowski et al. 1996) and PROMOTIF (Hutchinson and Thornton 1996). Besides the N-terminal AGH cloning tag, the calculated structure displays two regions with high pairwise root-mean-square-deviation (rmsd) values (Table 1): the N-terminal segment spanning residues M4, K5, and E6, and a C-terminal segment from A65 to E70. Heteronuclear ^{15}N-{^{1}H} NOEs (SM Fig. SF1) with values less than 0.6 or negative indicate that these two segments are indeed flexible, whereas these NOEs exceed 0.6 for the remaining residues from 7 to 64 that are well-defined with low rmsd (Fig. 2a, b; Table 1). Also, excluding G, A, and P (18 in

Table 1 Structural statistics for the ensemble of the 20 lowest energy structures of TtCdnLNt

Number of distance restraints	
Intraresidue (i–j = 0)	217
Sequential (|i–j| = 1)	207
Medium range (1 < |i–j| < 5)	96
Long-range (|i–j| ≥ 5)	461
Total number	981
Averaged total number per residue	14.2
Number of dihedral angle constraints	
Number of restricted ϕ angles	64
Number of restricted ψ angles	45
Total number	109
Average maximum violations per structure	
Distance (Å)	0.18 ± 0.06
Dihedral angle (°)	7 ± 2
Averaged structure energies	
CYANA target function value	0.53 ± 0.01
AMBER energy (kcal/mol)	−2327.85
van der Waals energy (kcal/mol)	−419.2
Electrostatic energy (kcal/mol)	−3743.2
Deviations from ideal geometry	
Bond length (Å)	0.014
Bond angle (°)	1.8

Pairwise rmsd (Å)	Backbone atoms	All heavy atoms
All residues (4–70)[a]	1.6 ± 0.6	2.4 ± 0.5
Ordered residues (7–64)	0.4 ± 0.1	1.1 ± 0.1
β-strand residues	0.19	
β-strand residues versus TRCF-RID (2EYQ)	1.29	

Ramachandran plot (%)	Ordered residues (7–64)	All residues (4–70)[a]
Most favoured regions	87.8	85.0
Additional allowed regions	12.2	14.9
Generously allowed regions	0.0	0.1
Disallowed regions	0.0	0.0

[a] N-terminal cloning tag, AGH, excluded from these analyses

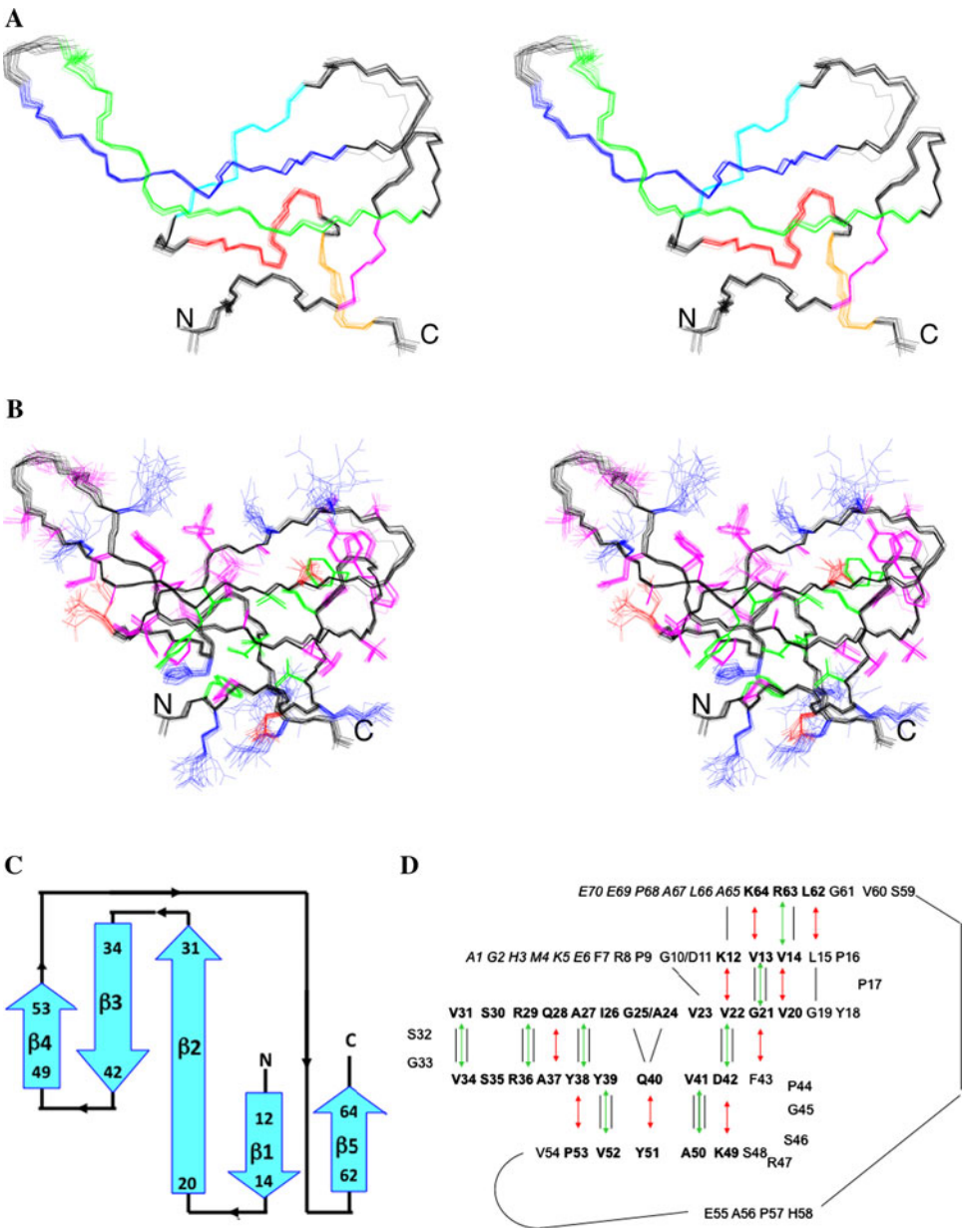


Fig. 2 NMR solution structure of TtCdnLNt. **a** Stereo view of an overlay of the backbone atoms for residues 6–65 in final NMR structures. The β-strands are colored *magenta* (β1), *green* (β2), *blue* (β3), *cyan* (β4), and *orange* (β5), the segment between residues 55 and 61 is shown in *red* and the rest in *black*. **b** Same as in (**a**) with the backbone displayed in *black* and the side chains of residues 7–64 in *blue* if positively charged (K, R, H), *red* if negatively charged (**d**, **e**), *green* for the core residues (those with buried side chains; ASA for side chain atoms $\leq$30 %; F7, V13, L15, A37, Y39, V41, F43, V52, L62), and *magenta* for all others. "N" and "C" indicate the N- and C-termini, respectively. **c** Cartoon representation of the β-sheet topology in TtCdnLNt, with the wide *cyan arrows* indicating β-strands (as labeled) and the delimiting residues of each β-strand numbered. **d** TtCdnLNt sequence showing residue connectivities in the antiparallel β-sheet with the β-strand residues in *bold* face and disordered residues in *italics*. *Vertical lines* indicate H-bonds present in the calculated structures, *red arrows* NOE cross-peaks between H_α of residues facing each other, and *green arrows* NOEs between NH protons of H-bonded residues

all), side chains of 24 residues in the segment from F7 to K64 are well ordered (χ_1 angular rmsd $\leq$ 30°; SM Fig. SF2). These include all of the buried residues (accessible surface area, ASA $\leq$ 20 %) that are nonpolar except for D11, whose carboxylate group may form a salt-bridge with the amino group of K64 that is sufficiently close in the structure (2.8–4.9 Å in 14 of the 20 conformers in the ensemble). Considering the proximity of oppositely charged groups, two other salt bridges may exist in the structure of TtCdnLNt: one between R36 and E55 (3.3–5.0 Å in 15 conformers of the ensemble), and the other between D42 and K49 (2.8–6.0 Å in a quarter of the

conformers). In both, the residues involved are solvent-exposed (ASA $\geq$ 30 %). No buried residue is among the 17 with the more disordered side chains (SM Fig. SF2), although some solvent-exposed residues also have well-defined side chain conformations.

The well-defined region from F7 to K64 contains five antiparallel β-strands spanning residues 12–14, 20–31, 34–42, 49–53 and 62–64, respectively, that coincide reasonably well with secondary structure predictions from the sequence. The five β-strands form a twisted antiparallel β-sheet with a β5-β1-β2-β3-β4 topology (Fig. 2). Most of the cross-strand H-bonds characteristic of anti-parallel β-sheets can be inferred as present in at least half of the 20 calculated structures based on the criteria that the proton-acceptor distance is less than 2.4 Å and the donor–acceptor angle is less than 35° (Fig. 2d). In TtCdnLNt, a regular H-bonded type II β-turn (residues 8–11; RPGD) precedes the N-terminal β1 strand, and forms part of an antiparallel G1 β-bulge in which residues G10 and D11 face residues V23 in the β2 strand. Also, a classic antiparallel β-bulge formed by residues A24 and G25 in strand β2 faces residue Q40 in strand β3 (Fig. 2d). Of the loops connecting the β-strands, that between β1 and β2 contains a *cis* P16-P17 bond and displays two consecutive type IV β-turns (LPPY, PPYG). Strands β2 and β3 are linked by a regular H-bonded I β-turn whose two central residues, S32-G33, exhibit positive values for their φ dihedral angles. In the loop linking strands β3 and β4, residues 43–46 (FPGS) adopt a regular type IV β-turn. The long segment connecting strands β4 and β5 adopts a distorted helix between residues 55–60 (Fig. 2a).

Discussion and conclusions

Comparison with structural homologues

A DALI (Holm and Sander 1993) search for structural homologues of TtCdnLNt yielded numerous hits to proteins with the Tudor, PHD, or SH3-like domain whose antiparallel β-sheet has a β5-β1-β2-β3-β4 topology. The large majority of the significant DALI hits were to eukaryotic proteins or their domains. Nevertheless, the closest structural homologue to TtCdnLNt is the Tudor-like RID of *E. coli* TRCF (PDB accession code: 2EYQ) with a DALI *Z*-score of 6.4/6.2 and rmsd of 3.5 Å for the backbone positions of 66 aligned residues sharing a 26 % identity. DALI hits were also obtained to the Tudor-like domains of two other bacterial proteins known to interact with RNAP: a 54-residue RapA N-terminal domain of RapA (PDB code: 3 DMQ; DALI *Z*-score = 5.5; rmsd = 2.1 Å) and a 50-residue segment of the C-terminal NusG domain (PDB code: 1MLG; DALI *Z*-score = 3.9; rmsd = 2.9 Å). A DALI hit but with a lower *Z*-score (3.7) was to the *T. thermophilus* TRCF-RID (PDB code: 3MLQ), possibly because structural coordinates for this domain are available only for the three central antiparallel β-strands (β2-β3-β4), a 40-residue stretch whose structure overlaps with that of TtCdnLNt (with a 38 % sequence identity) to an rmsd of 1.9 Å (Fig. 3a).

An overlay of the TtCdnLNt structures onto that of *E. coli* TRCF-RID domain (Fig. 3b) revealed that the main differences between the structures lie at the N-terminal segment, and at the loop regions. The number of G in TtCdnLNt is relatively high (10 %), and five of the seven that occur are conserved in the CarD_CdnL_TRCF family (Fig. 3a). The ϕ dihedral angle for three of the conserved G in TtCdnLNt (G10, G19, and G33) is positive but negative for the other two (G21 and G25), as is observed for the equivalent G in *E. coli* TRCF-RID. The less conserved G45 and G61, as well as S32, in TtCdnLNt also have positive ϕ dihedral angles. The β1–β2 linker region in TtCdnLNt contains a P–P pair present as a *cis* rotamer of which one, P16, is conserved within the family (Fig. 3a). P53 in TtCdnLNt, immediately C-terminal to β4, is also highly conserved and may play a structural role in the adequate positioning of the long linker to β5. It is also part of a residue stretch whose equivalent in *T. thermophilus* TRCF-RID has been implicated in contacts with RNAP-β (Fig. 3a; Westblade et al. 2010).

Interactions of TtCdnLNt with RNAP-β

TtCdnLNt and TRCF-RID have highly similar structures, as described above, and both interact with the same sub-domain of the considerably large RNAP β subunit. Also, the RNAP-β interacting side chains of TRCF-RID and the equivalent TtCdnLNt side chains exhibit a similar arrangement, as seen in the superposition of the TtCdnLNt ensemble structure onto the structure of *T. thermophilus* TRCF-RID complexed with an RNAP-β domain shown in Fig. 3c. Hence, we experimentally examined if specific contacts of TRCF-RID with RNAP-β are also conserved in TtCdnLNt. We did this using two-hybrid analysis in *E. coli*, as described in Methods and Results, in which interaction between two test proteins is inferred from the functional complementation of the *Bordetella pertussis* adenylate cyclase catalytic domain, when its T25 and T18 fragments are each fused to a given test protein (Karimova et al. 2000). Figure 3d shows that wild-type TtCdnLNt interacts with Tt$\beta_{17–139}$, as observed previously (Stallings et al. 2009; our unpublished data). We next examined the interaction of TtCdnLNt variants with Y38, Y51, or P53 mutated to A with wild-type Tt$\beta_{17–139}$. These three residues in TtCdnLNt were chosen for mutation to A as their equivalents in the *T. thermophilus* TRCF-RID (Y350,

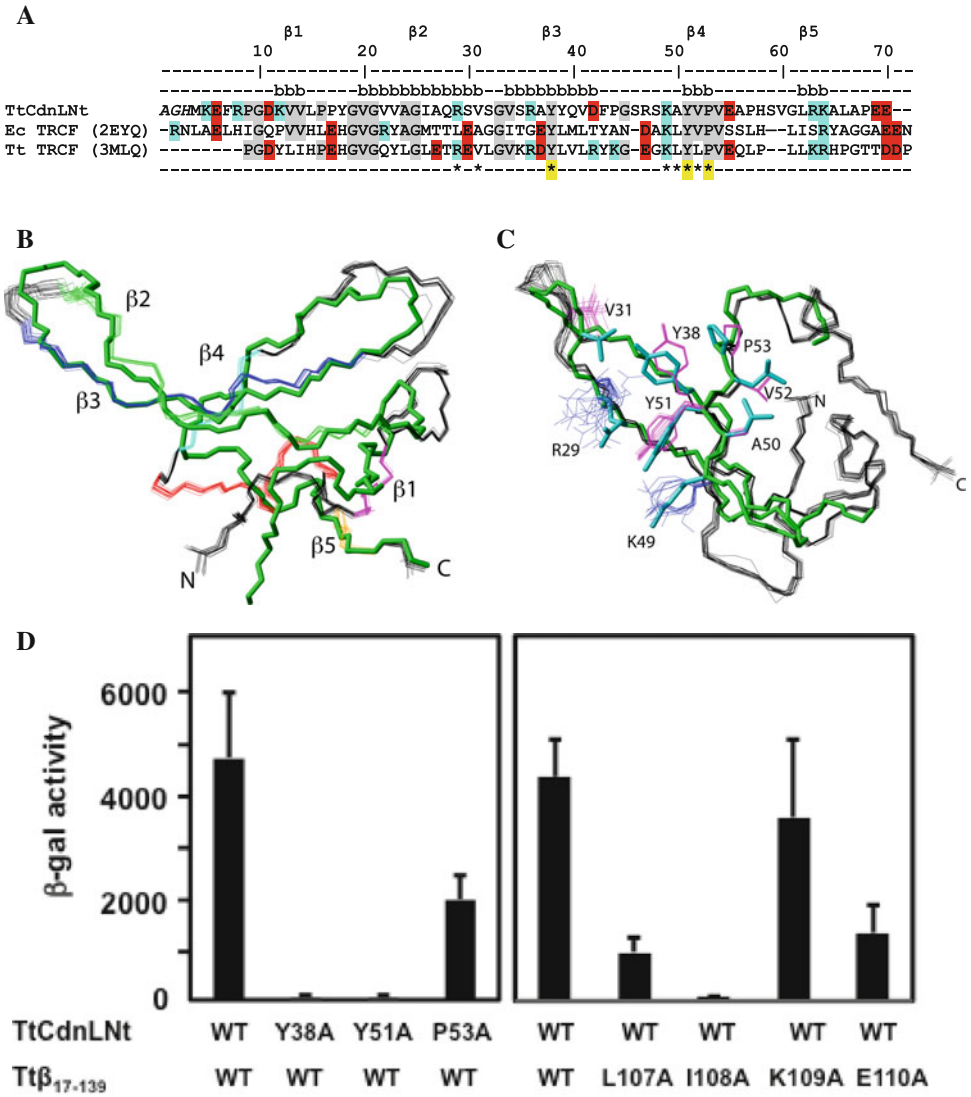


Fig. 3 Comparison of TtCdnLNt with structural homologues. **a** Sequence alignment of TtCdnLNt with *E. coli* and *T. thermophilus* TRCF-RID. Positively and negatively charged residues are indicated by *cyan* and *red backgrounds*, respectively. Residues identical in TtCdnLNt and in the TRCF-RID proteins are highlighted in *grey*. The *asterisks* below indicate residues in *T. thermophilus* TRCF-RID involved in contacts with RNAP-β. Those in TtCdnLNt examined by mutational analysis in this study are on a *yellow background*. **b** Overlay of the TtCdnLNt structural ensemble (in *black*) onto the crystal structure of *E. coli* TRCF-RID (in *green*; PDB code: 2EYQ). **c** The TtCdnLNt structure ensemble (in *black*) superimposed onto the crystal structure of *T. thermophilus* TRCF-RID in the complex with an RNAP-β domain (*green*; PDB code: 3MLQ), with the side chains of TRCF-RID residues interacting with RNAP-β shown in *cyan*. The residues occupying equivalent positions in TtCdnLNt are displayed in blue if positively charged (R29, K49) or in *magenta*, if uncharged (Y38, V31, A50, Y51, V52, P53). **d** Bacterial two-hybrid analysis of the interactions of wild-type (WT) TtCdnLNt and its mutants (as indicated) with Ttβ_{17-139}, and mutants of the latter (as indicated) with wild-type TtCdnLNt (see text for details)

Y362, and P364, respectively) contact RNAP-β in the reported structure of the complex (Westblade et al. 2010). Moreover, the Y38, Y51 and P53 side chains are quite solvent-exposed in native TtCdnLNt (ASA of 42, 45 and 38 % respectively, as expected for likely contact residues; compare with 42, 44 and 54 % for Y350, Y362, and P364, respectively, in TRCF-RID; SM Fig. SF2), and so would not contribute to the native TtCdnLNt core. Hence mutating them to A is unlikely to affect protein stability or folding, and we have experimental evidence suggesting this to be the case (data not shown). As can be seen in Fig. 3d (left panel), interaction with Ttβ_{17-139} was impaired for all three TtCdnLNt mutants examined relative to wild type, being essentially undetected for the Y38A and Y51A variants. Thus, TtCdnLNt employs a surface and residues to contact RNAP-β that are equivalent to those in TRCF-RID. Contacts with TRCF-RID include Ttβ_{17-139} residues L107, I108, K109, and E110, which are highly conserved

among RNAP-β and can be replaced by A to yield stable and properly folded mutants (Deaconescu et al. 2006; Westblade et al. 2010). For these mutants of Ttβ_{17-139}, except K109 to A, interaction with TtCdnLNt was markedly reduced relative to wild type (Fig. 3d, right panel). In the TRCF-RID/RNAP-β complex, K109 is involved in van der Waal′s contacts but not H-bond or charge interactions (Westblade et al. 2010) and these, presumably, are not altered on replacing the K109 sidechain with that of A. Overall, these data suggest that equivalent surfaces and conserved residues in TtCdnL and TRCF-RID interact with the same region of RNAP-β in *T. thermophilus.*

Conclusions

The structure of the N-terminal domain of *T. thermophilus* CdnL determined in this study closely resembles that of the RNAP-interaction domain, RID, of the bacterial transcription repair coupling factor (TRCF), with both adopting the Tudor-like twisted five-stranded antiparallel β-sheet fold with a β5-β1-β2-β3-β4 topology. Domains with this fold are also present in some other RNAP-associated bacterial proteins like the NusG transcription elongation factor and RapA, which mediates RNAP recycling for multi-round transcription, suggesting that it may be a general RNAP-interacting fold in bacteria. Besides sharing the same overall fold, our data indicate that TRCF-RID and TtCdnLNt employ analogous contact surfaces and conserved residues to interact with the same region of RNAP-β. Thus, TtCdnLNt mirrors TRCF-RID in structure as well as in contacts with RNAP. The two could therefore compete with each other for RNAP in vivo with important functional implications. Addressing these can draw upon the structural descriptions of the interactions with RNAP described for TRCF-RID earlier and those for TtCdnL in this study.

Acknowledgments We thank C. López, Dr. D. Pantoja-Uceda and L. de la Vega at Instituto de Química Física Rocasolano, CSIC for technical assistance. In the Universidad de Murcia, we thank José Antonio Madrid for technical assistance, Dr. César Flores–Flores for DNA sequencing, and Alejandro Torrecillas for mass spectrometry analysis. This work was supported by the Ministerio de Ciencia e Innovación, Spain grants: CTQ2008-0080/BQU and CTQ2011-22514 (MAJ), BFU2009-12445-C02-02 (SP), BFU2009-12445-C02-01 (MEA) co-financed by the European Union (FEDER) and Ph.D. fellowships from the Ministerio de Ciencia e Innovación, Spain (AGG) and the Consejo Superior de Investigaciones Científicas, Spain (YM).

References

Cayuela ML, Elías-Arnanz M, Peñalver-Mellado M, Padmanabhan S, Murillo FJ (2003) The *Stigmatella aurantiaca* homolog of *Myxococcus xanthus* HMGA-type transcription factor CarD: insights into the functional modules of CarD and their distribution in bacteria. J Bacteriol 185:3527–3537. doi:10.1128/JB.185.12.3527-3537.2003

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302. doi:10.1023/A:1008392405740

Deaconescu AM, Chambers AL, Smith AJ, Nickels BE, Hochschild A, Savery NJ, Darst SA (2006) Structural basis for bacterial transcription-coupled DNA repair. Cell 124:507–520. doi:10.1016/j.cell.2005.11.045

Elías-Arnanz M, Padmanabhan S, Murillo FJ (2010) The regulatory action of the myxobacterial CarD/CarG complex: a bacterial enhanceosome? FEMS Microbiol Rev 34:764–778. doi:10.1111/j.1574-6976.2010.00235.x

García-Moreno D, Abellón-Ruiz J, García-Heras F, Murillo FJ, Padmanabhan S, Elías-Arnanz M (2010) CdnL, a member of the large CarD-like family of bacterial proteins, is vital for *Myxococcus xanthus* and differs functionally from the global transcriptional regulator CarD. Nucleic Acids Res 38:4586–4598. doi:10.1093/nar/gkq214

Güntert P (2004) Automated NMR protein structure calculation. Prog Nucl Magn Res Spect 43:105–125. doi:10.1016/S0079-6565(03)00021-9

Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. J Mol Biol 233:23–38. doi:10.1006/jmbi.1993.1489

Hutchinson EG, Thornton JM (1996) PROMOTIF-a program to identify and analyze structural motifs in proteins. Protein Sci 5:212–220. doi:10.1002/pro.5560050204

Karimova G, Ullmann A, Ladant D (2000) A bacterial two-hybrid system that exploits a cAMP signaling cascade in *Escherichia coli*. Methods Enzymol 328:59–73

Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:51–58. doi:10.1016/0263-7855(96)00009-4

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486. doi:10.1007/BF00228148

León E, González C, Elías-Arnanz M et al (2009) ^{1}H, ^{13}C and ^{15}N backbone and side chain resonance assignments of a *Myxococcus xanthus* anti-repressor with no known sequence homologues. Biomol NMR Assign 3:37–40. doi:10.1007/s12104-008-9136-2

Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wüthrich K (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids - (IUPAC Recommendations 1998). Pure Appl Chem 70:117–142. doi:10.1351/pac199870010117

Padmanabhan S, Elías-Arnanz M, Carpio E, Aparicio P, Murillo FJ (2001) Domain architecture of a high mobility group A-type bacterial transcriptional factor. J Biol Chem 276:41566–41575. doi:10.1074/jbc.M106352200

Penalver-Mellado M, Garcia-Heras F, Padmanabhan S, Garcia-Moreno D, Murillo FJ, Elias-Arnanz M (2006) Recruitment of a novel zinc-bound transcriptional factor by a bacterial HMGA-type protein is required for regulating multiple processes in *Myxococcus xanthus*. Mol Microbiol 61:910–926. doi:10.1111/j.1365-2958.2006.05289.x

Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog NMR Spectrosc 34:95–158. doi:10.1016/S0079-6565(98)00025-9

Schubert M, Labudde D, Oschkinat H, Schmieder P (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in

proteins based on ^{13}C chemical shift statistics. J Biomol NMR 24:149–154. doi:10.1023/A:1020997118364

Selby CP, Sancar A (1995) Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties. J Biol Chem 270:4882–4889. doi:10.1074/jbc.270.9.4882

Stallings CL, Stephanou NC, Chu L, Hochschild A, Nickels BE, Glickman MS (2009) CarD is an essential regulator of rRNA transcription required for *Mycobacterium tuberculosis* persistence. Cell 138:146–159. doi:10.1016/j.cell.2009.04.041

Westblade LF, Campbell EA, Pukhrambam C et al (2010) Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. Nucleic Acids Res 38:8357–8369. doi:10.1093/nar/gkq692