

DOI: 10.1002/cbic.201300226

# Structural Study of the Partially Disordered Full-Length $\delta$ Subunit of RNA Polymerase from *Bacillus subtilis*

Veronika Papoušková,<sup>[a]</sup> Pavel Kadeřávek,<sup>[a]</sup> Olga Otrusínová,<sup>[a]</sup> Alžběta Rabatinová,<sup>[b, c]</sup>  
Hana Šanderová,<sup>[b]</sup> Jiří Nováček,<sup>[a]</sup> Libor Krásný,<sup>[b]</sup> Vladimír Sklenář,<sup>[a]</sup> and Lukáš Židek<sup>\*[a]</sup>

This article is written in memory of Prof. Ivano Bertini, who was a constant source of energy and inspiration in the field of NMR of proteins in general and of paramagnetic systems in particular; his spectrometers allowed us to test methods crucial for the success of this project before our laboratory was equipped with the optimal instruments.

The partially disordered  $\delta$  subunit of RNA polymerase was studied by various NMR techniques. The structure of the well-folded N-terminal domain was determined based on inter-proton distances in NOESY spectra. The obtained structural model was compared to the previously determined structure of a truncated construct (lacking the C-terminal domain). Only marginal differences were identified, thus indicating that the first structural model was not significantly compromised by the absence of the C-terminal domain. Various <sup>15</sup>N relaxation experiments were employed to describe the flexibility of both domains. The relaxation data revealed that the C-terminal

domain is more flexible, but its flexibility is not uniform. By using paramagnetic labels, transient contacts of the C-terminal tail with the N-terminal domain and with itself were identified. A propensity of the C-terminal domain to form  $\beta$ -type structures was obtained by chemical shift analysis. Comparison with the paramagnetic relaxation enhancement indicated a well-balanced interplay of repulsive and attractive electrostatic interactions governing the conformational behavior of the C-terminal domain. The results showed that the  $\delta$  subunit consists of a well-ordered N-terminal domain and a flexible C-terminal domain that exhibits a complex hierarchy of partial ordering.

## Introduction

Gram-positive Firmicutes bacteria differ from other bacteria in the composition of their RNA polymerase (RNAP). They contain an additional subunit,  $\delta$ , which is encoded by the *rpoE* gene in the model organism *Bacillus subtilis*.<sup>[1]</sup> Previously, the  $\delta$  subunit was reported to increase the specificity of RNAP for a promoter sequence,<sup>[2,3]</sup> and, recently, we reported that the  $\delta$  subunit is involved in regulation of transcription initiation in response to the concentration of the initiating nucleoside triphosphate (iNTP).<sup>[4]</sup> The  $\delta$  subunit destabilizes the open complex, a crucial intermediate of transcription initiation. This makes transcription at certain promoters amenable to regulation by the concentration of the iNTP,<sup>[5]</sup> which plays a critical role in reprogramming gene expression in response to stress.<sup>[6]</sup> Therefore, the  $\delta$  subunit is essential for rapid adaptation to changes in

the environment, as well as for virulence in some bacterial species.<sup>[4,7]</sup>


As an important complement to the in vitro and in vivo functional studies of the  $\delta$  subunit, we investigated its molecular structure. Because the sequence of the  $\delta$  subunit does not exhibit sufficient similarity with any protein of known structure, it was not possible to start with a homology model; rather, it was necessary to employ experimental approaches. The structural characterization of the  $\delta$  subunit was complicated mainly by the physical properties of its C-terminal domain. This region of the sequence is extremely acidic and highly repetitive (Figure 1 A), and it has been demonstrated that this is necessary for regulating RNAP affinity towards nucleic acids.<sup>[8]</sup>

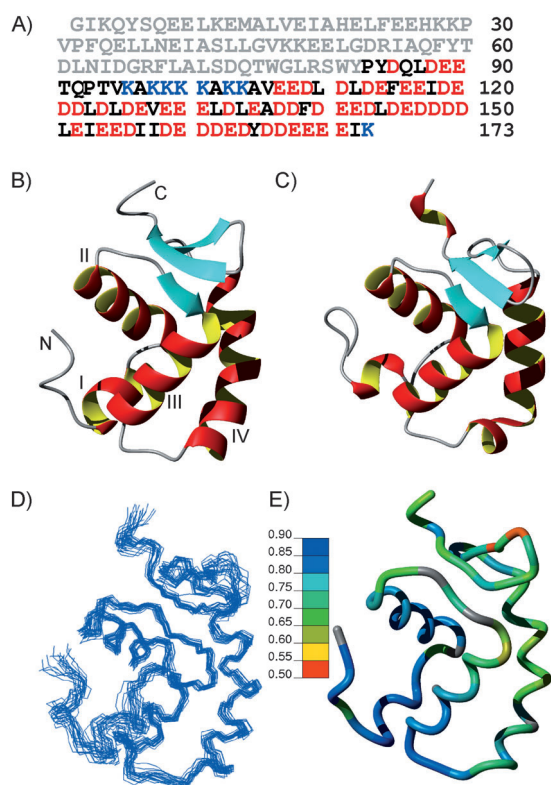
Circular dichroism has shown that the C-terminal domain does not adopt a well-defined structure.<sup>[8]</sup> Consequently, crystallization trials failed, and NMR spectra of the  $\delta$  subunit were difficult to assign because of severe peak overlap. Therefore, we first solved the solution structure of its ordered N-terminal domain, expressed as a truncated His-tagged construct that lacked the disordered C-terminal region.<sup>[9]</sup> The obtained model revealed that the N-terminal domain adopts a "DNA/RNA-binding three-helical bundle fold" (SCOP); this made identification of its structural homologues possible. In order to proceed towards structural characterization of the native, full-length protein, we then developed advanced NMR techniques that allowed us to overcome the technical difficulties mentioned above.<sup>[10,11]</sup>

[a] Dr. V. Papoušková, Dr. P. Kadeřávek, O. Otrusínová, J. Nováček, Prof. V. Sklenář, Dr. L. Židek  
Faculty of Science, NCBR and CEITEC MU, Masaryk University  
Kamenice 5, 625 00 Brno (Czech Republic)  
E-mail: lzidek@chemi.muni.cz

[b] A. Rabatinová, Dr. H. Šanderová, Dr. L. Krásný  
Laboratory of Molecular Genetics of Bacteria, Institute of Microbiology  
Academy of Sciences of the Czech Republic  
Videňská 1083, 142 20 Prague (Czech Republic)

[c] A. Rabatinová  
Faculty of Science, Charles University  
Viničná 5, 128 44 Prague 2 (Czech Republic)

 Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cbic.201300226>.



**Figure 1.** A) Amino acid sequence of the  $\delta$  subunit. The structured N-terminal domain is highlighted in gray; acidic and basic residues of the C-terminal domain are shown in red and blue, respectively. B) Representative structure of the N-terminal domain solved for the full-length  $\delta$  construct (PDB ID: 2M4K). C) Structure determined for the truncated His-tagged N-terminal domain (PDB ID: 2KRC). D) Superimposed backbone traces of 20 calculated 3D structures of the full-length  $\delta$  subunit (lowest energy, without violations). E) Structure of the ordered N-terminal domain of the full-length  $\delta$  protein, color-coded according to paramagnetic relaxation enhancement observed for the K99C mutant: blue ( $I_{\text{para}}/I_{\text{dia}} \geq 0.9$ ) to orange ( $I_{\text{para}}/I_{\text{dia}} \leq 0.5$ ). The disordered C-terminal regions (starting from P83) are omitted for clarity. Figures were prepared with the program MOLMOL.<sup>[12]</sup>

In this paper, we report a thorough structural study of the full-length construct of the  $\delta$  subunit with a histidine tag. Three-dimensional structure of the N-terminal domain, derived from observed nuclear Overhauser effects (NOEs), confirmed that the first structural model was not significantly compromised by the absence of the C-terminal domain. The combination of  $^{15}\text{N}$  relaxation data, paramagnetic labeling, and chemical shift analysis resulted in the first picture of the conformational behavior of the flexible C-terminal domain.

## Results and Discussion

### Solution structure of the N-terminal domain of the full-length $\delta$ subunit

Resonance frequencies of nuclei in the flexible and highly repetitive C-terminal region of the  $\delta$  subunit exhibit very low dispersion. As a consequence, the traditional strategy failed to assign the observed NMR signals. This led to the decision to start the structural studies by determining the 3D structure of

a truncated construct.<sup>[9]</sup> In parallel, development of a novel assignment strategy applicable to the flexible and highly repetitive C-terminal region of the full-length construct was initiated.<sup>[10,11]</sup> The new methodology provided a complete list of resonance frequencies of all observable nuclei, and allowed us to derive structural information from the spectra of the full-length construct of the  $\delta$  subunit.

Standard  $^{15}\text{N}$ -edited and  $^{13}\text{C}$ -edited 3D NOESY spectra<sup>[13]</sup> were recorded, and 2520 distance restraints were obtained. The complete list of frequencies allowed us to predict torsion angles for all residues. The distance restraints and 128 reliably predicted torsion angles were used to calculate the 3D structure of the  $\delta$  subunit. All 525 long-range NOEs (critical for correct determination of the three-dimensional structure) were observed between protons within the N-terminal domain. Therefore, the conformation of the C-terminal domain remained undefined in the obtained structural model, in agreement with the physical behavior of the molecule. Importantly, the core structure of the N-terminal domain (Figure 1B) was well-defined and virtually unchanged relative to the truncated His-tagged protein (Figure 1C), with the exception of the ends of helix IV and residues in loop L71–Q74. A set of 20 structures of the full-length construct with the lowest energies is presented in Figure 1D. The root-mean-square deviations (RMSD) between all and backbone heavy atoms in the mean structures of the truncated and full-length constructs were 2.65 and 2.00 Å, respectively. The RMSD values within each structure and parameters describing the structure calculation and validation are summarized in Table 1.

**Table 1.** Statistics of sets of 20 final structures of the  $\delta$  subunit with the lowest energy and restraints used in the structure calculation.

Parameter Construct PDB ID	Number/value	
	Full length 2M4K	Truncated 2KRC
total NOE restraints	2520	2341
intraresidue, $ i-j ^{[a]}=0$	836	734
sequential, $ i-j =1$	574	499
medium range, $1 <  i-j  < 5$	585	564
long range, $ i-j  \geq 5$	525	544
TALOS dihedral angle restraints	128	132
RDC restraints	0	342
CSA restraints	0	33
total CNS energy <sup>[b]</sup> [kcal mol <sup>-1</sup> ]	-6315.0	-3978.4
distance violations $\geq 5$ Å	0.0	0.0
dihedral angle violations $\geq 5^\circ$	1.1	0.1
RMSD <sup>[c]</sup> backbone heavy atoms/all atoms [Å]	0.97/1.64	0.83/1.49
Ramachandran <sup>[d]</sup> most-favoured region [%]	90.3	87.8
<b>Structure Z-scores (PROCHECK)</b>		
1st-generation packing quality	1.913	1.651
2nd-generation packing quality	5.553	6.566
Ramachandran plot appearance	-2.562	-2.770
$\chi$ -1/ $\chi$ -2 rotamer normality	-2.986	-3.598
backbone conformation	-0.535	-1.038

[a]  $i$  and  $j$  refer to residue numbers. [b] Calculated for the structure with the lowest energy. [c] Average RMSD to the mean structure. [d] See Ramachandran plot in Figure S1.

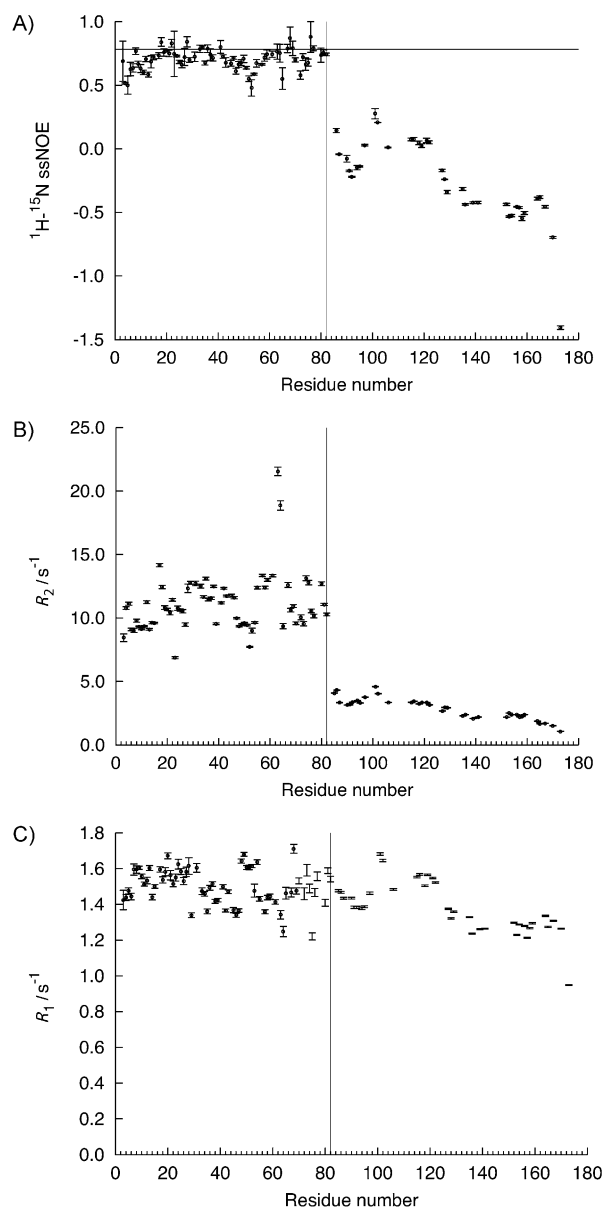
Slight local differences between the structures might reflect different restraints used in the structure calculations. In order to distinguish real structural changes from the bias due to the differences in experimental restraints, the chemical shifts of the N-terminal domains of both constructs were compared. As expected, large chemical shift changes were observed between P83 and Q92. Chemical shifts in the well-ordered region (G2–Y82) differed much less. The largest differences did not exceed 0.03 ppm for  $^1\text{H}$ , 0.3 ppm for  $^{13}\text{C}$ , and 0.4 ppm for  $^{15}\text{N}$ , with a few exceptions ( $\text{C}^\beta$  of S72,  $\text{C}^\alpha$  of L44 and Q57). Chemical shifts of the backbone nuclei (amide  $^{15}\text{N}$  and  $^1\text{H}$ ,  $^1\text{H}^\alpha$ ,  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$ , and carbonyl  $^{13}\text{C}$ ; see Figure S2 in the Supporting Information) differed significantly in helix II and for several residues in helix III (E39, I40) and helix IV (A56, Q57). In summary, comparison of the chemical shifts showed that some of the differences between the calculated structures might reflect conformational variations that are real but small (helix IV and loop region L71–Q74). On the other hand, larger chemical shift differences were found in helix II, which is almost identical in both calculated structures. In any case, differences in the chemical shifts were small, thus confirming that the structures are very similar.

#### Degree of disorder in the C-terminal domain

As the structure calculation did not probe the conformational space of the C-terminal domain properly (because of lack of distance restraints), we used additional analyses to characterize this region of the protein. Longitudinal and transverse  $^{15}\text{N}$  auto-relaxation rates ( $^{15}\text{N}$   $R_1$  and  $^{15}\text{N}$   $R_2$ , respectively) and steady state  $^1\text{H}$ – $^{15}\text{N}$  nuclear Overhauser enhancement ( $^1\text{H}$ – $^{15}\text{N}$  ssNOE) were measured by using standard  $^{15}\text{N}$  relaxation experiments.<sup>[14]</sup> Conventionally, the combination of  $^{15}\text{N}$   $R_1$ ,  $^{15}\text{N}$   $R_2$ , and  $^1\text{H}$ – $^{15}\text{N}$  ssNOE data is used to describe motions of well-ordered proteins in the model-free (MF) approach.<sup>[15]</sup> The standard MF analysis relies on the assumption that fast internal motion and slow global tumbling are dynamically separated, and are described by internal and global correlation times, respectively:  $^{15}\text{N}$   $R_2$  mostly reflects the correlation time of global tumbling, whereas  $^1\text{H}$ – $^{15}\text{N}$  ssNOE is most sensitive to the internal correlation time. Contributions of the two motional modes to the overall dynamics are described by the order parameter  $S^2$ , which ranges from one (completely rigid residues) to zero (completely flexible residues with dynamics dominated by internal motions). Both  $^{15}\text{N}$   $R_2$  and  $^1\text{H}$ – $^{15}\text{N}$  ssNOE decrease with decreasing  $S^2$ , whereas  $^{15}\text{N}$   $R_1$  can increase or decrease, depending on the internal correlation time. In addition,  $^{15}\text{N}$   $R_2$  also includes a contribution from slow conformational or chemical exchange. The MF analysis converts the experimental data into quantitative values of the order parameter and correlation times. However, such an approach was not applicable in our case as the C-terminal domain is disordered, with internal and global motions effectively coupled. Therefore, the results of relaxation measurements were interpreted directly, without transforming them into the order parameters and correlation times.

$^1\text{H}$ – $^{15}\text{N}$  ssNOE was inspected as a good indicator of the flexibility of the polypeptide chain. Although  $^1\text{H}$ – $^{15}\text{N}$  ssNOE values

are positive for rigid residues of macromolecules (close to the limit value of 0.8 at 600 MHz), they decrease significantly if short correlation times dominate the dynamics (with the limit of  $-3.8$ ). The  $^1\text{H}$ – $^{15}\text{N}$  ssNOE values measured for residues of the N-terminal domain exhibited a pattern typical for well-folded proteins (Figure 2). Interestingly, the most significant decrease in  $^1\text{H}$ – $^{15}\text{N}$  ssNOE was detected at the N-terminal of helix IV, where the structural models differed (compare Figure 1B and 1C). This indicates that variations in this region of the calculated structures might arise from local backbone flexibility. The sudden drop in the  $^1\text{H}$ – $^{15}\text{N}$  ssNOE between Q86 and the C-terminus revealed much higher flexibility in the C-terminal



**Figure 2.** A)  $^1\text{H}$ – $^{15}\text{N}$  ssNOE values measured for individual backbone amides of the  $^{15}\text{N}$ -labeled  $\delta$  subunit at 600 MHz. The horizontal line corresponds to a physical limit for the amide group. B)  $^{15}\text{N}$   $R_2$  and C)  $^{15}\text{N}$   $R_1$  values were measured for individual backbone amides of the  $^{15}\text{N}$ -labeled  $\delta$  subunit at 600 MHz. Vertical bar shows the boundary between the N- and C-terminal domains.

region. Nevertheless, the flexibility of the C-terminal domain was not uniform. Several regions of the C-terminal sequence exhibited significantly higher rigidity, most evident in the lysine-rich motif between K96 and K104. A largely negative  $^1\text{H}$ - $^{15}\text{N}$  ssNOE value was observed only for the terminal residue, K173.

The determined  $^{15}\text{N}$   $R_2$  rates exhibit the same trend as the  $^1\text{H}$ - $^{15}\text{N}$  ssNOE values (Figure 2B). The  $^{15}\text{N}$   $R_2$  value ( $\sim 12\text{ s}^{-1}$ ) in the N-terminal domain mostly reflects global tumbling of this structural module. The lower  $^{15}\text{N}$   $R_2$  of the N-terminal amino acids and of residues K47–R54 indicates higher flexibility of the N terminus of helix IV and of the loop between helices III and IV. On the other hand, significantly higher  $^{15}\text{N}$   $R_2$  values reveal slow exchange between two (or more) states, most notable for V17, N63, and I64. The dramatic decrease in  $^{15}\text{N}$   $R_2$  in the C-terminal region correlates well with the  $^1\text{H}$ - $^{15}\text{N}$  ssNOE data, but this is less sensitive to variations in local dynamics, because  $^{15}\text{N}$   $R_2$  is mostly given by the global correlation time while  $^1\text{H}$ - $^{15}\text{N}$  ssNOE predominantly reflects the internal correlation time (as discussed above).

Finally, the  $^{15}\text{N}$   $R_1$  rates complete the picture of the dynamics of the  $\delta$  subunit. The  $^{15}\text{N}$   $R_1$  values agree well with the  $^{15}\text{N}$   $R_2$  and  $^1\text{H}$ - $^{15}\text{N}$  ssNOE data. As expected from the dependence of  $^{15}\text{N}$   $R_1$  on the correlation times<sup>[14]</sup> and from the typical values of the correlation times of disordered and ordered proteins of the given size, N-terminal and C-terminal domains do not differ in average  $^{15}\text{N}$   $R_1$ , but  $^{15}\text{N}$   $R_1$  follows the trends of  $^1\text{H}$ - $^{15}\text{N}$  ssNOE in the disordered region.

### Long-range contacts in the C-terminal domain

As mentioned above, none of the long-range cross-peaks identified in the 3D NOESY spectra could be attributed to a residue in the C-terminal region. This indicates that the amino acids of the C-terminal domain do not make close contacts of sufficient stability to result in the observed inter-proton NOEs. To search for possible interactions of the C-terminal domain, we took advantage of the absence of cysteine residues in the  $\delta$  subunit, and we created a series of single point mutants (selected amino acids in the C-terminal domain were replaced with cysteine). This allowed the attachment of a paramagnetic label at various positions in the C-terminal tail.

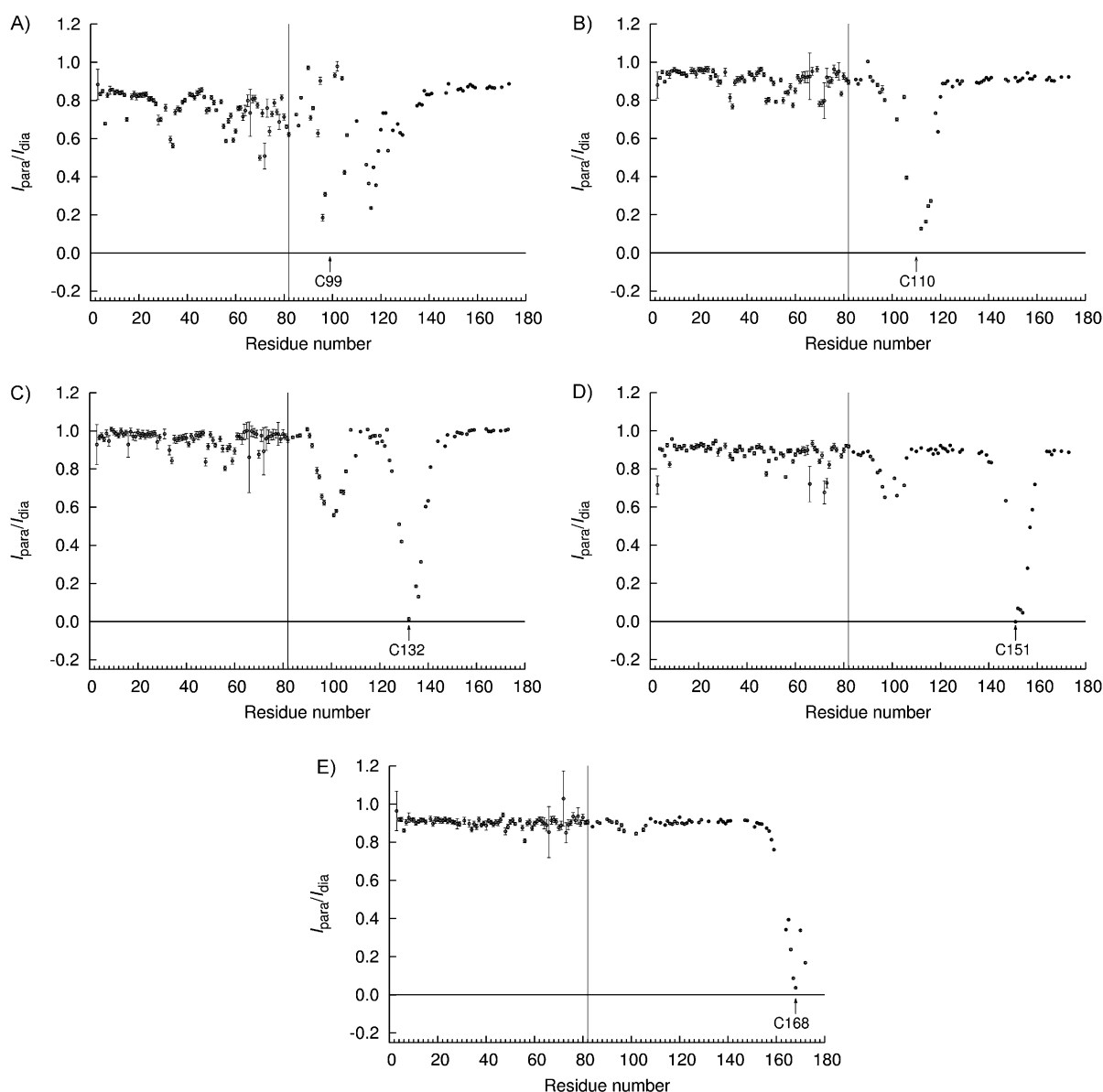
Paramagnetic relaxation enhancement (PRE:  $I_{\text{para}}/I_{\text{dia}}$ , the ratio of peak intensities in spectra of paramagnetic and diamagnetic samples) for mutants K99C, L110C, L132C, L151C, and E168C, is presented in Figure 3. A decrease in peak intensity due to PRE was observed not only in the vicinity of the label, but also for residues in distant regions of the sequence. It shows that the tail forms transient contacts with the well-ordered N-terminal domain and with itself.

PRE values indicating inter-domain interactions were inspected first. Regions of the N-terminal domain affected by the spin label were mostly located in helix IV, loop L71–Q74, and the first strand of the  $\beta$ -sheet (Figure 1E). The peak intensity dropped to approximately 60% in this region when the spin label was placed close to the N-terminal domain (K99C). The residues with the most significant PRE correlate with the differ-

ences between the calculated structures of the full-length and truncated constructs of the  $\delta$  subunit (helix IV and loop L71–Q74, cf. Figure 1B and C). Therefore, it cannot be excluded that the C-terminal tail influences conformation of the mentioned region of the well-folded N-terminal domain. The PRE gradually decreased as the position of the label moved towards the C terminus, thus indicating that the observed contacts are predominantly intramolecular.

After analyzing the contacts between the domains, we turned our attention to interactions within the C-terminal domain. The most significant interaction was between the positively charged lysine-rich sequence K96–K104 and negatively charged residues in the C-terminal tail. This type of interaction was monitored by comparing the PRE profile obtained for the label placed at position 99 (K99C, Figure 3A) with the decrease in peak intensities in the region T94–V106 observed when the label was placed at various positions in the negatively charged C-terminal region: L110C, L132C, L151C, and L168C (Figure 3B–E, respectively). Interpretation of this PRE data was straightforward. In addition to the drop of peak intensities in the vicinity of the spin-labeled residues, all these four mutants exhibited significant PRE in the positively charged region K96–K104. While the paramagnetic probe was too close to the lysine-rich sequence in the L110C mutant (Figure 3B), data for the L132C, L151C, and L168C mutants clearly showed the PRE effect for residues T94–A106: gradually decreasing as the sequential distance of the spin-label from the lysine-rich region increased (Figure 3C–E). These results suggest that the interactions of the C-terminal domain of the  $\delta$  subunit are dominated by an electrostatic attraction between the short positively charged sequence K96–K104 and a long negatively charged region between E114 and D171. The intensity of PRE in the lysine-rich sequence reflects the relative populations of the  $\delta$  conformers forming contacts between K96–K104 and various regions of the acidic C-terminal tail.

Interpretation of the data for mutant K99C (Figure 3A) was somewhat more complicated. As expected, the peak intensity for residues in a close proximity to the paramagnetic label (K96, A97) dropped significantly (the intensity of the labeled K99C could not be evaluated because of peak overlap). Interestingly, low PRE was observed for residues close (in the sequence) to the paramagnetic probe (E90, Q92, T94, V95, K101, A102, and K103). In general, the ratio of peak intensities ( $I_{\text{para}}/I_{\text{dia}}$ ) is given by  $\exp(-R_{\text{para}}t_m)/(1+R_{\text{para}}/R_{\text{dia}})$ ,<sup>[16]</sup> where  $R_{\text{dia}}$  is the effective relaxation rate of the diamagnetic sample,  $R_{\text{para}}$  is the paramagnetic contribution to the relaxation, and  $t_m$  is the total duration of the magnetization transfer periods (11 ms in this case). The exponential term can be neglected for the discussed weak PRE, and  $I_{\text{para}}/I_{\text{dia}}$  can be expressed just as a function of  $R_{\text{para}}/R_{\text{dia}}$  which is approximately  $0.55 < r^{-6} > [4J_e(0) + 3J_e(600\text{ MHz})]/[4J_N(0) + 3J_N(600\text{ MHz})]$ , where  $r$  is the proton–electron distance in nm, and  $J_e(\omega)$  and  $J_N(\omega)$  are spectral density functions of the proton–electron and proton–nitrogen vectors, respectively, given by the order parameters and correlation times.<sup>[16,17]</sup> For such a simplified analysis, the lower sensitivity of the peak intensity ratio ( $I_{\text{para}}/I_{\text{dia}}$ ) to the paramagnetic relaxation in the region E90–K103 can be explained by the long aver-



**Figure 3.** Relative peak intensity ( $I_{\text{para}}/I_{\text{dia}}$ ) of amide cross-peaks in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of A) K99C, B) L110C, C) L132C, D) L151C, and E) E168C mutants with MTSL/MTS attached. Arrows show the spin-label position. Vertical bar shows the boundary between the N- and C-terminal domains.

age distance of the observed amide protons from the nitroxide (longer than expected for a fully flexible peptide chain) and/or by substantial differences in the correlation times and order parameters of the proton–nitrogen and proton–electron vectors. In any case, the data are not consistent with the behavior of a completely disordered protein. This conclusion is in agreement with the  $^{15}\text{N}$  relaxation data: significantly higher  $^{15}\text{N}$   $R_1$ ,  $^{15}\text{N}$   $R_2$ , and  $^1\text{H}$ - $^{15}\text{N}$  ssNOE in this region of the protein, with a maximum at K101. Lower flexibility in the vicinity of the label is also reflected by broadening of the peak of proton H4 of the (1-acetoxy-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate (MTS) label (Figure S3).

Interpretation of the long-range contacts with the C-terminal domain of the K99C mutant was easier. A decrease in  $I_{\text{para}}/I_{\text{dia}}$  indicated strong interaction of the lysine-rich region K96–K104 with residues F115–D119. Weaker PRE effects were observed

for residues closer to the C terminus, and this was marginal for the last 20 amino acids. The  $I_{\text{para}}/I_{\text{dia}}$  ratios of residues close to L110 and L132 in the K99C mutant correlated well with the decrease in the peak intensity of K99 due to the presence of the paramagnetic probe in L110C (Figure 3B) and L132C (Figure 3C), respectively. Decreasing, but still significant, PRE was observed for K99 in mutants with the label at positions L151C and L168C (Figure 3D and E, respectively); the  $I_{\text{para}}/I_{\text{dia}}$  ratio of the L151 and E168 peaks in the K99C mutant was approximately 0.9. In summary, PRE in the sample with the paramagnetic probe placed at K99C confirmed the interactions between the lysine-rich sequence K96–K104 and the long acidic C-terminal region.

It should be noted that attachment of a paramagnetic probe presents a potential risk of perturbation of protein structure and dynamics. No dramatic changes in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spec-



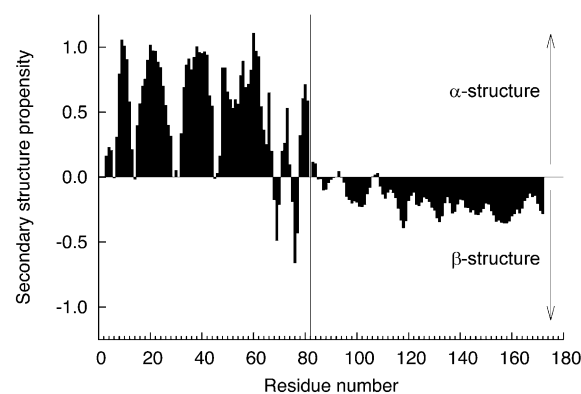
tra of the individual mutants with the diamagnetic form of the label (MTS) were observed (see Figures S4–S33). Nevertheless, the small differences between PRE of the same pairs of residues with the label placed in the positive and negative region of the C-terminal domain suggest that the presence of a label in the lysine stretch might have a disruptive effect. Also, changes in the peak shapes in region E90–A105 observed after attaching either MTS or (1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate (MTSL) to the K99C mutant indicated that the label might influence the conformation of residues in the lysine-rich region. Therefore, data for the K99C mutant served only as an additional piece of information in this study, thus supporting the results of the PRE experiments performed with the spin-label placed in the negatively charged region of the C-terminal domain. More importantly, the observed increase of  $^1\text{H}$ – $^{15}\text{N}$  ssNOE and  $^{15}\text{N}$   $R_1$  (Figure 2) exactly in the regions of the strongest contacts identified by PRE provides an important independent support of the PRE results.

Finally, comparison of the NOE and PRE data allowed us to make a rough estimate of the apparent effective distance  $\langle r^{-6} \rangle^{-1/6}$  between the interacting residues. The detection limits of the methods, 0.5 nm for NOE<sup>[18]</sup> and 2.0 nm for PRE<sup>[19]</sup> (estimated for the conditions used in this study), defined the expected range of the effective average distances. Therefore, the contacts invisible as NOE cross-peaks but clearly manifest as significant PRE (Figure 3) correspond to the effective average distance in the range 0.5–2.0 nm.

### Secondary structure propensities in the C-terminal domain

The mapping of transient inter-residue contacts by PRE was complemented by structural information derived directly from the observed resonance frequencies. Deviations of chemical shifts from their random-coil values provide a very useful description of the effective conformation of an intrinsically disordered polypeptide chain. Analysis of the chemical shifts does not require additional measurements and is not biased by modifying the molecule with paramagnetic labels. The complete list of frequencies provided by the improved assignment methodology<sup>[10,11]</sup> was ideally suited for such a survey. The secondary structure propensity was calculated by using the program SSP<sup>[20]</sup> with neighbor-corrected random chemical shifts<sup>[21]</sup> employed as a reference. The program evaluates deviations of various chemical shifts from reference values and combines them into a single score. The calculated score, weighted by the sensitivity of individual chemical shifts to the  $\alpha$ - and  $\beta$ -structures, corresponds to the content of the conformers present in a secondary structure. A value of +1 indicates that the given residue is present in  $\alpha$ -helix (or in a similar helical structure) in all conformers of the ensemble; a score of –1 corresponds to complete formation of  $\beta$ -sheet (or a similar extended structure). For example, an SSP score of –0.3 predicts that 30% of molecules in the sample have the analyzed residues in a  $\beta$ -structure.

The results of the SSP analysis revealed that the C-terminal domain is disordered with a significant bias towards a  $\beta$ -structure (ca. 30% propensity, Figure 4), presumably due to the



**Figure 4.** The protein secondary structure propensity predicted from  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  chemical shift values by the program SSP.<sup>[20]</sup> Positive values correspond to a helical conformation; negative values refer to a  $\beta$ -sheet or extended structure. Vertical bar shows the boundary between the N- and C-terminal domains.

electrostatic repulsion. Interestingly, this trend was not continuous but was interrupted by two regions with no propensity to form the  $\beta$ -structure: T91–T94 and V106–E108. This pattern might be interpreted as a tendency to form an antiparallel  $\beta$ -sheet, but we do not have any direct evidence supporting such a hypothesis. Comparison with the PRE data indicates a well-balanced interplay of repulsive and attractive forces that keep the C-terminal region relatively extended, but still able to bend and form significant electrostatic contacts within the C-terminal domain.

### Conclusions

NMR structure determination of the  $\delta$  subunit verified the reliability of the previously obtained model of its N-terminal domain. Comparison of the PRE values observed for residues of the N-terminal domain with the differences between chemical shifts and calculated structures of the full-length and truncated construct indicated which regions of the well-ordered domain might be influenced by interaction with the long flexible C-terminal tail. A combination of several approaches was used to characterize transient structures and the dynamics of the C-terminal domain. Chemical shift analysis showed a significant tendency of the C-terminal domain to form  $\beta$ -type structures, in agreement with its polyanionic nature. Results of  $^{15}\text{N}$  relaxation experiments, especially  $^1\text{H}$ – $^{15}\text{N}$  ssNOE, revealed that the regulatory C-terminal domain adopts a flexible (but not random coil) structure. The observed hierarchy of a partial ordering correlated well with the transient contacts identified by paramagnetic labeling. In summary, internal flexibility and a balance between attractive and repulsive electrostatic interactions seem to define the conformational behavior of the C-terminal domain. This characterization will be important for future structure–function studies of the  $\delta$  subunit and its interactions with the core RNAP. Moreover, given the increasing recognition of the importance of non-structured proteins (or their parts) for cell physiology,<sup>[22]</sup> the  $\delta$  subunit is an excellent model molecule for studies of such proteins.

## Experimental Section

### Determination of the 3D structure of the $\delta$ subunit in solution:

The  $\delta$  subunit was overexpressed from vector pFL31<sup>[8]</sup> in *Escherichia coli* BL21 (DE3). The cells were grown in M9 minimal medium<sup>[23]</sup> (2 L) containing  $^{15}\text{NH}_4\text{Cl}$  and  $^{13}\text{C}$ glucose. Expression of the protein was induced by IPTG (0.8 mM) when  $\text{OD}_{600}$  reached 0.6. The protein was induced for 3 h at 25 °C. By taking advantage of the low pI of the protein (3.85 as calculated by the ExpASY Compute pI/ $M_w$  tool),<sup>[24]</sup> ion exchange chromatography and isoelectric precipitation were used to purify the protein, as previously reported.<sup>[8]</sup> The uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled sample of the RNA polymerase  $\delta$  subunit was prepared in an NMR buffer containing phosphate buffer (20 mM, pH 6.6), NaCl (10 mM), and  $\text{NaN}_3$  (0.05%). The sample was finally concentrated to 0.8 mM. All NMR experiments were recorded at 27 °C on an Avance 600 MHz spectrometer (Bruker) equipped with a 5 mm cryogenic  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  Z-gradient TCI probe head. A set of 3D  $^{15}\text{N}$ -edited and  $^{13}\text{C}$ -edited NOESY spectra<sup>[13]</sup> with 1536 real points in the direct dimension, 48 real points in the  $^{15}\text{N}/^{13}\text{C}$  dimension and with 240 points in the indirect  $^1\text{H}$  dimension was recorded to obtain data for structure determination. The  $^{15}\text{N}$  carrier frequency was set to 116.3 ppm; the  $^{13}\text{C}$  carrier was placed to 29.7 and 125.0 ppm in the aliphatic and aromatic NOESY experiments, respectively; the  $^1\text{H}$  carrier was placed at the resonance of a proton in water. The spectral widths were 8389 Hz in the direct  $^1\text{H}$  dimension, 4545 Hz in the indirect  $^{13}\text{C}$  dimensions, 1370 Hz in the indirect  $^{15}\text{N}$  dimension, and 7194 Hz in the indirect  $^1\text{H}$  dimension. The spectral processing and analysis program NMRPipe/NMR-Draw 3.0<sup>[25]</sup> and the graphical NMR assignment and integration software Sparky 3.111 (T. D. Goddard and D. G. Kneller, University of California, San Francisco, CA) were used for spectra processing and analysis, respectively.

The frequency assignment deposited in the BMRB database (<http://www.bmrb.wisc.edu>) under the accession code BMRB-16912<sup>[10]</sup> was used to determine the protein backbone torsion angle restraints by the program TALOS<sup>[26]</sup> and to predict secondary structure propensities by the program SSP.<sup>[20]</sup> The set of 2520 distance restraints among the protons was obtained from the  $^{15}\text{N}$ -edited and  $^{13}\text{C}$ -edited NOESY spectra<sup>[13]</sup> and assigned by the program ARIA 2.1<sup>[27]</sup> The structure calculations were performed by running restrained molecular dynamics in the CNS 1.2 program,<sup>[28]</sup> with standard protocol and input scripts written by the authors of the RECOORD database.<sup>[29]</sup> A total of 300 structures was calculated; 100 were further minimized in the explicit solvent. The quality of the ordered part of the final structures was checked by the program CING (G. W. Vuister. CING Nijmegen: Radboud University Nijmegen, the Netherlands). A set of 20 lowest energy structures was deposited in the Protein Data Bank ([www.pdb.org](http://www.pdb.org)) under the PDB ID: 2M4K.

**Relaxation experiments:** The  $^{15}\text{N}$  relaxation experiments were performed by using a uniformly  $^{15}\text{N}$ -labeled  $\delta$  subunit (0.8 mM), in the same buffer and under the same experimental conditions as used for the structure determination. Spectral width were 8389 Hz in the direct ( $^1\text{H}$ ) dimension and 2000 Hz in the indirect ( $^{15}\text{N}$ ) dimensions; 2048 and 384 real points were collected in the direct and indirect dimensions, respectively. Standard experiments<sup>[14]</sup> were used for the measurements of  $R_1$  (relaxation delays 22.4, 67.2, 134.4\*, 246.4, 380.8, 560\*, 784, 1008, and 1232 ms), and for the measurements of  $R_2$  at the CPMG frequency of 515 Hz (relaxation delays 0, 17.181, 34.362\*, 51.542, 68.723, 85.904\*, 103.085, 137.446, and 171.808 ms). Asterisks denote spectra recorded twice in order to estimate experimental error. The relaxation rates were obtained by fitting peak intensities to a monoexponential decay by using the

program Relax.<sup>[30,31]</sup> The  $^1\text{H}$ - $^{15}\text{N}$  ssNOE values were measured under a steady state condition, achieved by a 5 ms  $^1\text{H}$  irradiation with 226 repeats of 200  $\mu\text{s}$  180° pulses,<sup>[32]</sup> separated by 22.22  $\mu\text{s}$  delays, and with a 20 s interscan relaxation delay (sufficiently long for a quantitative description of a highly flexible polypeptide chain).<sup>[33]</sup> The reference spectra were measured interleaved together with the spectra under the steady state conditions. The experimental error was evaluated based on three independent measurements. The data were deposited in the BMRB database (accession code BMRB-19284).

**Paramagnetic relaxation enhancement:** Site-directed mutagenesis of the *rpoE* gene was performed with the QuikChange Site-Directed Mutagenesis Kit (Stratagene/Life Technologies) with plasmid pFL31<sup>[8]</sup> as the template and primers (see Table S1): 706 and 707 (mutant K99C; strain LK994), 708 and 709 (mutant L110C; LK995), 710 and 711 (mutant L132C; LK996), 712 and 713 (mutant L151C; LK997), and 714 and 715 (mutant E168C; LK998). The results of the site-directed mutagenesis were verified by sequencing. The verified plasmids were transformed into *E. coli* BL21 (DE3) and the proteins were expressed as described for the wild-type protein. The K99C, L110C, L132C, L151C, and E168C mutants were labeled with a paramagnetic compound, (1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate (MTSL), and by its diamagnetic analogue, (1-acetoxy-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate (MTS; both purchased from Toronto Research Chemicals, Inc., North York, ON, Canada) as described by Volkov et al.<sup>[34]</sup> Incorporation of the label was checked by mass spectrometry. The NMR samples contained the following concentrations of the MTS- and MTSL-labeled mutants, respectively: K99C, 0.61 and 0.51 mM; L110C, 0.71 and 0.70 mM; L132C, 0.35 and 0.35 mM; L151C, 0.65 and 0.61 mM; E168C, 0.39 and 0.40 mM. All samples were prepared in the same buffer and studied under the same conditions as described for the structure determination. PRE was monitored by  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments with 2048 real points and spectral width of 8389 Hz in the direct dimension, 512 real points and spectral width of 1370 Hz in the indirect dimension, and with a recycle delay of 5 s, sufficient to start each scan from a thermal equilibrium. The PRE values were calculated as a ratio of peak heights in spectra of paramagnetic and diamagnetic samples ( $I_{\text{para}}/I_{\text{dia}}$ ) scaled to their relative protein concentration, as determined by a Bradford assay.<sup>[35]</sup> As the peaks of MTS-labeled C99 in the K99C mutant and C110 in the L110C mutant could not be distinguished (due to overlap with other signals), the peak of proton H4 of the MTS moiety at approximately 5.6 ppm was monitored in 1D  $^1\text{H}$  NMR spectra for both MTS and MTSL samples. The data showed that the amount of the reduced label was below the detection limit, given by the noise in the spectra of the paramagnetic samples (see Figure S3).

**Secondary structure propensity:** Secondary structure propensities were calculated from  $^1\text{H}^\alpha$ ,  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$ , backbone carbonyl  $^{13}\text{C}$ , and backbone  $^{15}\text{N}$  chemical shifts, deposited previously in the BMRB database under the accession code BMRB-16912,<sup>[10]</sup> by using the program SSP<sup>[20]</sup> with weighted averaging over three residues. Neighbor-corrected random-coil chemical shifts<sup>[21]</sup> were used as the reference values.

## Acknowledgements

The study was supported by grants from Czech Science Foundation No. 204/09/0583 (structure determination of the N-terminal domain and mutagenesis) and No. 13-16842S (structural characterization of the C-terminal region). Partial support by the project

"CEITEC—Central European Institute of Technology" (CZ.1.05/1.1.00/02.0068) from a European Regional Development Fund is also acknowledged. A.R. was in part supported by grant No. 648712 from the Charles University Grant Agency. We thank Marcellus Ubbink and Yoshitaka Hiruma from Leiden University for providing us with a protocol for spin-label preparation and helpful suggestions.

**Keywords:** NMR spectroscopy · partially disordered proteins · protein structures · RNA polymerase ·  $\delta$  subunit

- [1] G. P. Doherty, M. J. Fogg, A. J. Wilkinson, P. J. Lewis, *Microbiology* **2010**, *156*, 3532–3543.
- [2] E. C. Achberger, H. R. Whiteley, *J. Biol. Chem.* **1981**, *256*, 7424–7432.
- [3] X. Xue, J. Tomasch, H. Sztajner, I. Wagner-Döbler, *J. Bacteriol.* **2010**, *192*, 5081–5092.
- [4] A. Rabatinová, H. Šanderová, J. Jiráť Matějčková, J. Korelusová, L. Sojka, I. Barvík, V. Papoušková, V. Sklenář, L. Židek, L. Krásný, *J. Bacteriol.* **2013**, *195*, 2603–2611.
- [5] L. Sojka, T. Kouba, I. Barvík, H. Šanderová, Z. Maděrová, J. Jonák, L. Krásný, *Nucleic Acids Res.* **2011**, *39*, 4598–4611.
- [6] A. Kriel, A. N. Bittner, S. H. Kim, K. Liu, A. K. Tehrani, W. Y. Zou, S. Rendon, R. Chen, B. P. Tu, J. D. Wang, *Mol. Cell* **2012**, *48*, 231–241.
- [7] R. Seepersaud, R. H. V. Needham, C. S. Kim, A. L. Jones, *J. Bacteriol.* **2006**, *188*, 2096–2105.
- [8] F. J. López de Saro, A. Y. M. Woody, J. D. Helmann, *J. Mol. Biol.* **1995**, *252*, 189–202.
- [9] V. Motáčková, H. Šanderová, L. Židek, J. Nováček, P. Padrta, A. Švenková, J. Korelusová, J. Jonák, L. Krásný, V. Sklenář, *Proteins Struct. Funct. Bioinf.* **2010**, *78*, 1807–1810.
- [10] V. Motáčková, J. Nováček, A. Zawadzka-Kazimierzczuk, K. Kazimierzczuk, L. Židek, H. Šanderová, L. Krásný, W. Koźmiński, V. Sklenář, *J. Biomol. NMR* **2010**, *48*, 169–177.
- [11] J. Nováček, A. Zawadzka-Kazimierzczuk, V. Papoušková, L. Židek, H. Šanderová, L. Krásný, W. Koźmiński, V. Sklenář, *J. Biomol. NMR* **2011**, *50*, 1–11.
- [12] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **1996**, *14*, 51–55.
- [13] D. Marion, P. C. Driscoll, L. E. Kay, P. T. Wingfield, A. Bax, A. M. Gronenborn, G. M. Clore, *Biochemistry* **1989**, *28*, 6150–6156.
- [14] D. M. Korzhnev, M. Billeter, A. S. Arseniev, V. Y. Orekhov, *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *38*, 197–266.
- [15] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.
- [16] L. Salmon, G. Nodet, V. Ozenne, G. Yin, M. R. Jensen, M. Zweckstetter, M. Blackledge, *J. Am. Chem. Soc.* **2010**, *132*, 8407–8418.
- [17] J. Iwahara, G. M. Clore, *J. Am. Chem. Soc.* **2010**, *132*, 13346–13356.
- [18] D. Neuhaus, M. P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, 2nd ed., Wiley-VCH, Weinheim, **2000**.
- [19] J. L. Battiste, G. Wagner, *Biochemistry* **2000**, *39*, 5355–5365.
- [20] J. A. Marsh, V. K. Singh, Z. Jia, J. Forman-Kay, *Protein Sci.* **2006**, *15*, 2795–2804.
- [21] K. Tamiola, B. Acar, F. A. A. Mulder, *J. Am. Chem. Soc.* **2010**, *132*, 18000–18003.
- [22] T. Chouard, *Nature* **2011**, *471*, 151–153.
- [23] J. Sambrook, D. W. Russel, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 3rd ed., **2001**.
- [24] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch in *The Proteomics Protocols Handbook* (Ed.: J. M. Walker), Humana, Totowa, **2005**.
- [25] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277–293.
- [26] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **2009**, *44*, 213–223.
- [27] J. P. Linge, M. Habeck, W. Rieping, M. Nilges, *Bioinformatics* **2003**, *19*, 315–316.
- [28] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **1998**, *54*, 905–921.
- [29] A. J. Nederveen, J. F. Doreleijers, W. Vranken, Z. Miller, C. A. E. M. Spronk, S. B. Nabuurs, P. Güntert, M. Livny, J. L. Markley, M. Nilges, E. L. Ulrich, R. Kaptein, A. M. J. J. Bonvin, *Proteins Struct. Funct. Bioinf.* **2005**, *59*, 662–672.
- [30] E. J. d'Auvergne, P. R. Gooley, *J. Biomol. NMR* **2008**, *40*, 107–119.
- [31] E. J. d'Auvergne, P. R. Gooley, *J. Biomol. NMR* **2008**, *40*, 121–133.
- [32] F. Ferrage, D. Cowburn, R. Ghose, *J. Am. Chem. Soc.* **2009**, *131*, 6048–6049.
- [33] C. Renner, M. Schleicher, L. Moroder, T. A. Holak, *J. Biomol. NMR* **2002**, *23*, 23–33.
- [34] A. N. Volkov, J. A. R. Worrall, E. Holtzmann, M. Ubbink, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18945–18950.
- [35] M. M. Bradford, *Anal. Biochem.* **1976**, *72*, 248–254.

Received: April 11, 2013

Revised: June 14, 2013

Published online on July 18, 2013