MASARYKOVA UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
NÁRODNÍ CENTRUM PRO VÝZKUM BIOMOLEKUL

Diplomová práce

BRNO 2017 **HANA ŠTÉGNEROVÁ**

MASARYKOVA UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
NÁRODNÍ CENTRUM PRO VÝZKUM BIOMOLEKUL

Strukturní role elektrostatických interakcí v C-terminální doméně podjednotky delta RNA polymerasy bakterie Bacillus subtilis

Diplomová práce

Hana Štégnerová

Vedoucí práce: prof. Mgr. Lukáš Žídek, Ph.D. **Brno 2017**

MASARYK UNIVERSITY
FACULTY OF SCIENCE
NATIONAL CENTRE FOR BIOMOLECULAR RESEARCH

Structural role of electrostatic interactions in C-terminal domain of delta subunit of RNA polymerase from Bacillus subtilis

Diploma thesis

Hana Štégnerová

Supervisor: prof. Mgr. Lukáš Žídek, Ph.D. **Brno 2017**

Bibliografický záznam

Autor:	Bc. Hana Štégnerová Přírodovědecká fakulta, Masarykova univerzita Národní centrum pro výzkum biomolekul
Název práce:	Strukturní role elektrostatických interakcí v C-terminální doméně podjednotky delta RNA polymerasy bakterie Bacillus subtilis
Studijní program:	Biochemie
Studijní obor:	Biomolekulární chemie
Vedoucí práce:	prof. Mgr. Lukáš Žídek, Ph.D.
Akademický rok:	2016/2017
Počet stran:	$xx + 47$
Klíčová slova:	NMR; delta podjednotka RNA polymerasy; IDP; elektrostatické interakce; RDC; relaxace; PRE

Bibliographic Entry

Author:	Bc. Hana Štégnerová Faculty of Science, Masaryk University National Centre for Biomolecular Research
Title of Thesis:	Structural role of electrostatic interactions in C-terminal domain of delta subunit of RNA polymerase from Bacillus subtilis
Degree Programme:	Biochemistry
Field of Study:	Biomolecular chemistry
Supervisor:	prof. Mgr. Lukáš Žídek, Ph.D.
Academic Year:	2016/2017
Number of Pages:	$xx + 47$
Keywords:	NMR; delta subunit of RNA polymerase; IDP; electrostatic interactions; RDC; relaxations; PRE

Abstrakt

Delta podjednotka je součástí enzymu RNA polymerasy výhradně Gram pozitivních bakterií. Je to postradatelná část RNA polymerasy, nicméně je potřebná pro přežití buněk v kompetitivním prostředí. Delta podjednotka je částečně neuspořádaný protein, který se skládá ze dvou domén. N-terminální doména je strukturovaná, zatímco C-terminální doména je vnitřně neuspořádaná. C-terminální doména obsahuje mnohonásobné repetice negativně nabitých zbytků kyseliny glutamové a asparagové, které na polypeptidovém řetězci generují elektrostatickou repulzi. C-terminální doména také obsahuje konzervovaný úsek s pozitivně nabitými zbytky lyzinu, takzvaný lyzinový trakt. Předchozí studie delta podjednotky odhalily přechodný elektrostatický kontakt mezi pozitivně nabitým lyzinovým traktem a negativně nabitými zbytky v C-terminální doméně.

V této diplomové práci se zabýváme popisem role elektrostatických interakcí v konformačním chování delta podjednotky. Za účelem pozorování vlivu elektrostatickým interakcí byl lyzinový trakt zaměněn za negativně nabité zbytky kyseliny glutamové. Případné změny v konformačním chování delta podjednotky byly pozorovány za použití nukleární magnetické rezonance. Mezi sledované parametry patřily chemické posuny, zbytkové dipolární interakce, zrychlení relaxace v důsledku paramagnetické interakce (paramagnetic relaxation enhancement) a relaxační rychlosti.

Výsledky analýzy potvrdily, že elektrostatické interakce jsou zdrojem částečného uspořádání v C-terminální doméně a řídí její konformační chování. Navíc také zprostředkovávají mezidoménový kontakt mezi negativně nabitým C-terminálním řetězcem a specifickými oblastmi v uspořádané N-terminální doméně. Bylo zjištěno, že elektrostatické interakce v neuspořádané doméně jsou nezbytné pro správné konformační chování celého proteinu delta podjednotky.

Abstract

The δ subunit is a unique part of RNA polymerase of Gram positive bacteria. It is a dispensable subunit of RNA polymerase, however the δ subunit is required for fitness of the cells in the competitive environment. The δ subunit is a partially disordered protein, which comprises two domains. The N-terminal domain is well folded, whereas the C-terminal domain is intrinsically disordered. The C-terminal domain contains multiple repetitions of negatively charged glutamic and aspartic acid residues, generating electrostatic repulsion in the polypeptide chain. Additionally, the conserved lysine stretch with positive charge is present in the C-terminal domain. Previous structural studies of the δ subunit revealed transient electrostatic interactions between the lysine stretch and the negatively charged residues in the C-terminal domain.

In my diploma thesis I focused on description of the role of these electrostatic interaction in the conformational behaviour of the δ subunit. In order to investigate the impact of electrostatic interactions, the positively charged lysine stretch was replaced by negatively charged glutamate residues. The potential changes in the conformational behaviour were monitored by nuclear magnetic resonance. The observed structural parameters were chemical shifts, residual dipolar coupling, paramagnetic relaxation enhancement and relaxation rates.

Our investigation confirmed that electrostatic interactions are the source of the partial ordering in the C-terminal domain and govern its conformational behaviour. Moreover the electrostatic interactions in the flexible C-terminal domain mediate the interdomain contact of the negatively charged C-terminal tail with specific regions in the well folded N-terminal domain. These results revealed that the electrostatic interactions in the disordered domain are crucial for the proper conformational behaviour of the whole δ subunit.

MASARYKOVA UNIVERZITA
Přírodovědecká fakulta

ZADÁNÍ DIPLOMOVÉ PRÁCE

Akademický rok: 2016/2017

Ústav: Národní centrum pro výzkum biomolekul

Studentka: Bc. Hana Štégnerová

Program: Biochemie

Obor: Biomolekulární chemie

Ředitel *Národního centra pro výzkum biomolekul* **PřF MU Vám ve smyslu Studijního a zkušebního řádu MU určuje diplomovou práci s názvem:**

Název práce: Strukturní role elektrostatických interakcí v C-terminální doméně podjednotky delta RNA polymerasy bakterie Bacillus subtilis

Název práce anglicky: Structural role of electrostatic interactions in C-terminal domain of delta subunit of RNA polymerase from Bacillus subtilis

Oficiální zadání:
Cílem práce bude zjistit jakou roli hrají elektrostatické interakce ve strukturním chování podjednotky delta RNA polymerasy bakterie Bacillus subtilis. Úloha elektrostatických interakcí bude zkoumána dvěma přístupy - změnami iontové síly a mutacemi lysinů, o kterých se předpokládá, že jsou zásadní pro interakce s kyselými aminokyselinami neuspořádané části proteinu. Jako základní metoda vyhodnocení strukturních vlastností bude použita nukleární magnetická rezonance. Využity budou zejména pokročilé metody neuniformního vzorkování, které poskytují vysoké rozlišení, nezbytné pro studium daného proteinu.

Literatura:
CAVANAGH, John. *Protein NMR spectroscopy : principles and practice*. 2nd ed. Amsterdam: Elsevier, 2007. xxv, 885. ISBN 9780121644918.

Jazyk závěrečné práce: angličtina

Vedoucí práce: prof. Mgr. Lukáš Žídek, Ph.D.

Datum zadání práce: 23. 10. 2015

V Brně dne: 12. 4. 2017

Souhlasím se zadáním (podpis, datum):

Bc. Hana Štégnerová
studentka

prof. Mgr. Lukáš Žídek, Ph.D.
vedoucí práce

doc. RNDr. Oldřich Janiczek, CSc.
zástupce ředitele Národního centra pro výzkum biomolekul pro pedagogické záležitosti

Acknowledgement

I would like to thank my supervisor Lukáš Žídek in the first place, for his kind and patient guidance throughout my master thesis. Also, I want to express my gratitude to all members of the Protein Structure and Dynamics group for their friendly company during my master studies. Special thank is given to Zuzka Jaseňáková for showing me how to survive in wet lab, and to Milan Zachrdla for all useful advice and his neverending patience.

I want to sincerely thank to my dearest family, my boyfriend Aleš and great friends Karolína and Hany for their support and cheering me up during my studies.

Declaration

I honestly declare that this diploma thesis was elaborated independently, based on literature and other resources stated in bibliography.

Brno 15. května 2017 .....

Hana Štégnerová

Contents

Abbreviations

δKE	mutant of the δ subunit
δWT	wild type of the δ subunit
CTD	C-terminal domain
IDP	intrinsically disordered protein
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NTD	N-terminal domain
RDC	residual dipolar coupling
RNAP	RNA polymerase
SCS	secondary chemical shift
SDFM	spectral density function mapping
ssNOE	steady-state nuclear Overhauser enhancement
SSP	secondary chemical shifts propensity
PRE	paramagnetic relaxation enhancement

Introduction

Intrinsically disordered proteins (IDPs) have come in the forefront of researchers' interests few decades ago, when the structure-function paradigm was called into question. IDPs do not adopt any stable tertiary structure, however their structural plasticity is associated with a wide variety of biological functions, which are inaccessible for well ordered proteins. The abundance of IDPs within proteoms of all organisms documents their importance and remarkable efforts were made in order to develop novel approaches for structural characterization of IDPs. The structural investigation of IDPs might help not only in the understanding of the functionality of proteins in regulatory networks, but also in designing new drugs [1]. Nuclear magnetic resonance is a well-suited experimental technique for the investigation of IDPs with atomic resolution.

The protein of interest in the presented diploma thesis is the δ subunit of the bacterial enzyme RNA polymerase. The δ subunit is a partially unstructured protein, comprising well folded N-terminal domain and intrinsically disordered C-terminal domain. This protein is found exclusively in Gram positive bacteria and although it is a dispensable part of RNA polymerase, the δ subunit is essential for the fitness of cells in competitive environment [2]. The presence of the δ subunit enhances transcription specificity and efficiency of RNA polymerase recycling.

The structure of the δ subunit has been previously studied in our laboratory. The structure of the well folded N-terminal domain was determined by Veronika Papoušková in 2010 [3]. The N-terminal domain comprises four alpha helices and a beta sheet. Afterwards the structure of the C-terminal domain was characterized [4]. The sequence of C-terminal domain contains multiple repetitions of the negatively charged glutamic and aspartic acid residues, which results in electrostatic repulsion and extended structure. Additionally, the lysine stretch with a positive charge is present in the sequence of the C-terminal domain. The electrostatic attraction mediates the intradomain contact between the lysine stretch and the negatively charged C-terminal tail. The flexibility of the polypeptide chain in the C-terminal domain is not uniform and exhibits increased rigidity in the region of the lysine stretch and its vicinity. These results promoted the potential importance of the electrostatic interactions on the conformational behaviour of the C-terminal domain.

Within the framework of my bachelor thesis the influence of the electrostatic interactions was probed by increasing the salt concentration from the 10 mM sodium chloride, used for the structure determination, to the final concentration of 200 mM sodium chloride. The salt ions were supposed to weaken the electrostatic attraction between the lysine stretch and the negatively charged C-terminal tail and potential changes in the structural behaviour were searched for using nuclear magnetic resonance. The effect of the increased salt on the chemical shifts and residual dipolar coupling was not significant. Thus we decided to

investigate the electrostatic interactions in a more controlled manner.

Therefore the mutation was introduced to the lysine stretch and the positively charged lysines were replaced by the glutamic acid residues with negative charge. The potential differences in the conformational behaviour caused by the disruption of electrostatic interactions were monitored by various structural parameters obtained using nuclear magnetic resonance, such as chemical shifts, residual dipolar coupling, paramagnetic relaxation enhancement and relaxation rates.

Chapter 1

Theory

1.1 Delta subunit of RNA polymerase of *Bacillus subtilis*

1.1.1 RNA polymerase

Bacillus subtilis is a well studied model organism, which belongs to the industrially and medically important strain of Firmicutes. It is a Gram positive low G+C bacterium.

DNA-dependent RNA polymerase (RNAP) is a multi-subunit protein complex fundamental for gene expression. This enzyme catalyses the essential process of transcription, where genetic information from the DNA template is converted into RNA. RNAP is controlled by a wide scale of transcription factors. Specificity and precise regulation of the RNAP activity play a crucial role in the control of gene expression. Therefore efficient function of RNAP is necessary for organisms to adapt their environment [5]. Transcription consists of three phases: initiation, elongation and termination.

During initiation, RNAP recognises the DNA promoter sequence, binds to double-stranded DNA (dsDNA) and the closed dsDNA-RNAP complex is created. Subsequently, the open complex is formed by unwinding double-stranded DNA to single-stranded DNA. The open complex is able to synthesize RNA. During the elongation phase, the synthesis of RNA continues. Termination of transcription is catalysed by a specific termination sequence or ρ factor, the RNA transcript is released and RNAP recycles.

The bacterial RNAP core consists of five indispensable subunits - dimer α_2, β, β' and ω. The RNAP core enables elongation and termination. The process of initiation requires forming of RNAP holoenzyme by association of the core with an additional σ subunit. The σ subunit enables recognition of a promoter sequence on DNA and therefore reduces unspecific binding of RNAP. The α and ω subunits maintain assembly and stability of the RNAP core [6, 7]. The β, β' subunits form the active center of the enzyme, carry out the interaction of RNAP with the DNA template, formation of the transcription bubble and NTP binding to RNAP [8].

Composition of the RNAP core is identical in all prokaryotic organisms and the structure of the active site in the core is conserved in both bacterial and eukaryotic organisms [9]. However, RNAP of Gram positive bacteria differs in the presence of two additional subunits: ε and δ subunits.

1.1.2 Delta subunit

The accesory δ subunit is a protein encoded by the *rpoE* gene in Gram-positive bacteria. Sequence of the δ subunit includes 173 amino acids and the protein molecular mass is 20.4 kDa. The δ subunit is a highly acidic protein with the isoelectric point of 3.6 [10].

The δ subunit contains two distinct regions – structured N-terminal domain (NTD) and flexible C-terminal domain (CTD).

N-terminal domain

The NTD provides binding to the RNAP core *in vivo*, however the binding site is unknown [11]. The amino acid sequence of NTD is displayed in grey in Figure 1.1, it contains 82 amino acids and the charge distribution in this region is uniform. Conserved regions are present in NTD, which indicates importance of this region [12].

```
GIKQYSQEELKEMALVEIAH ELFEEHKKPVPFQELLNEIA
SLLGVKKEELGDRIAQFYTD LNIDGRFLALSDQTWGLRSW
YPYDQLDEETQPTVKAKKKK AKKAVEEDLDLDEFEEIDED
DLDLDEVEEELDLEADDFDE EDLDEDDDDLEIEEDIIDED
DEDYDDEEEEIK
```

Figure 1.1: Amino acid sequence of the δ subunit from *Bacillus subtilis*. Grey colour represents the structured N-terminal part. Letters in black, red and blue mark uncharged, negatively charged and positively charged amino acid residues in the C-terminal part, respectively.

The structure of the ordered N-terminal part was solved using both X-ray [13] and NMR [14]. The N-terminal domain consists of four alpha helices (helix I, II, III, IV created by residues Q8–K12, L16-–H27, F33-–L44, G52-–N63, respectively) and antiparallel beta sheet, formed by three beta strands (residues V31-–P32, F68-–A70, T75-–L78) [14, 4]. The structure, shown in Figure 1.2, was published by Papouskova et. al in 2010. The PDB code is 2KRC [14]. The arrangement of NTD is structurally and sequentially similar to the winged helix-turn-helix (wHTH) domain in the ASXL protein, which belongs to the HARE-HTH family (HB1, ASXL, Restriction Endonuclease proteins). Based on the functions of proteins included in this family, HARE-HTH might be a potential DNA-binding domain [15]. Resemblance of δ with HARE-HTH suggests the ability of δ NTD interact not only with RNAP, but also with modified DNA sequences [16].

C-terminal domain

The unstructured C-terminal part contains 92 amino acid residues. The amino acid sequence includes stretches of glutamate and aspartate residues, represented by red colour in Figure 1.1. Glutamic and aspartic acid side chains carry negative charge. Repetitive acidic residues generate an extremely acidic C-terminal tail, which acts as a polyanion. The acidic nature provides CTD the ability to displace nucleic acid bound to RNAP in a binary complex [10].

CTD is highly flexible and does not adopt any 3D structure. However, the chain does not behave as a random coil. The electrostatic repulsion of repetitive negative charges in CTD results in formation of an extended tail with 30% propensity for beta secondary

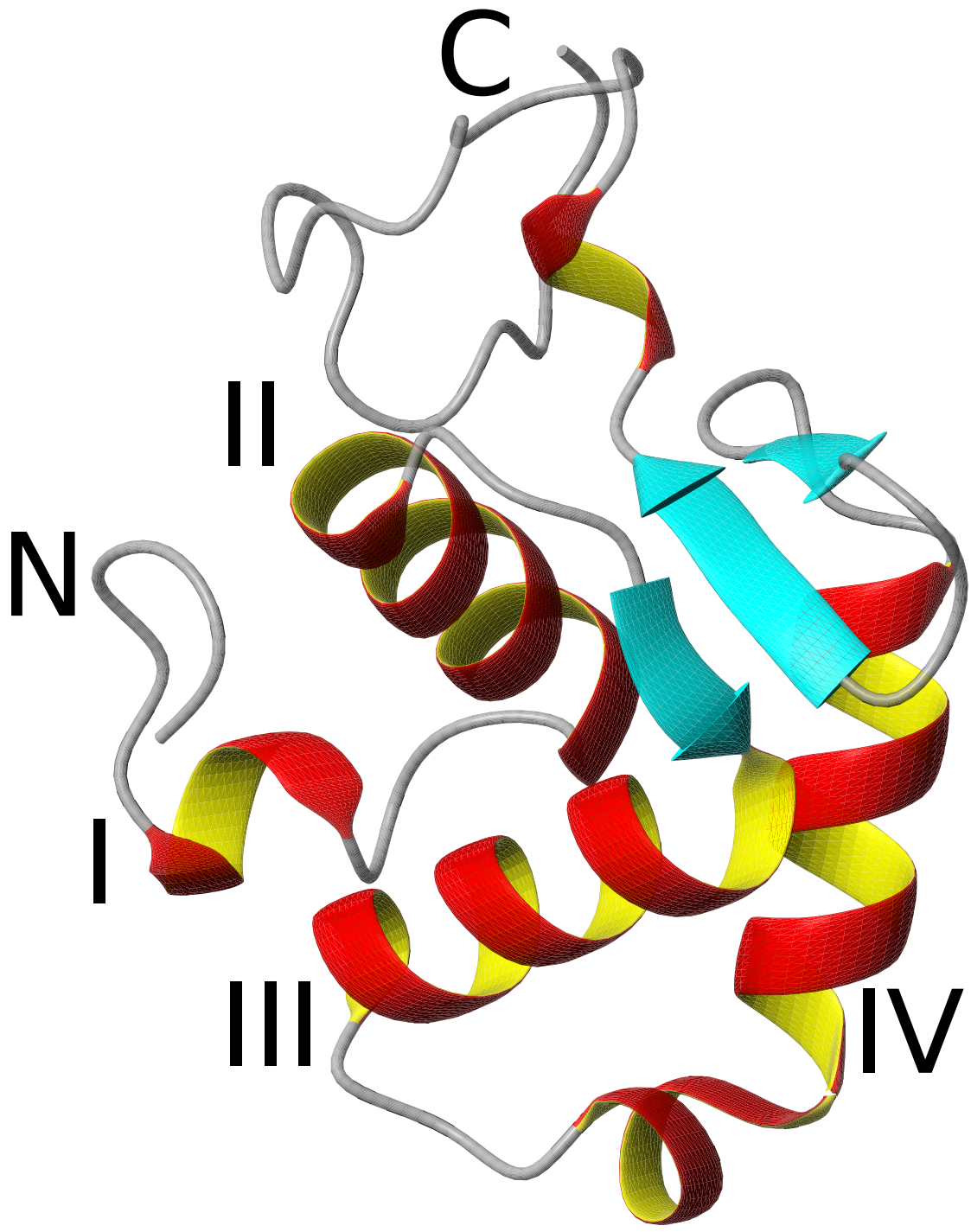

Figure 1.2: The structure of δ N-terminal domain (PDB ID: 2KRC).

structures. Also the flexibility of C-terminal tail is not uniform along the polypeptidic chain. The region with the lysin stretch, shown in blue in Figure 1.1, is significantly more rigid than the rest of the tail [4].

Transient long-range contacts of CTD both with well-ordered NTD and itself were observed. The intradomain interaction is caused by an electrostatic attraction between the positively charged lysin stretch (residues K96–K104) and the negatively charged C-terminal tail (residues E114–D171). Balance between the electrostatic attraction and repulsion governs the conformational behaviour of CTD. The interdomain interaction was identified with helix IV, loop L71–Q74 and the first strand of the beta-sheet [4]. Contacts between the domains might indicate influence of CTD on folding of NTD [16].

Function of δ subunit

The δ subunit is a dispensable part of RNAP and cells are viable after its deletion [17]. However, δ is required for the proper transcriptional function of RNAP, it maintains RNAP specificity and increases efficiency of the transcription machinery.

δ plays an important role in initiation of transcription. It suppresses binding of RNAP at non-promoter DNA sequences, decreases RNAP-DNA open complex stability at weak promoters and thus supports transcription from preferred and strong promoters [18, 19]. Rapid and specific gene regulation by δ provides cells ability to adapt and survive in changing environment [2].

An impact of δ on virulence of Gram positive pathogenic bacteria such as *Streptococcus agalactiae, Streptococcus aureus* was observed. The absence of the δ subunit negatively affects the virulence, decreases the ability to survive in human blood and the resistance against leukocytes [20, 12].

Processes following termination of transcription are also influenced by the δ subunit. NTD bound to RNAP presents its C-terminal tail, which competes with nucleic acid for the RNAP binding. It facilitates efficient RNAP release and recycling after transcription and increases the overall RNAP activity in multi-round transcription [10]. Despite the intensive studies of the δ subunit, the molecular mechanism of its function is still unknown.

The δ subunit, containing the unstructured region in C-terminal domain, belongs to the group of intrinsically disordered proteins. The specific characteristics of this group are discussed in the following section.

1.2 Intrinsically disordered proteins

The structure-function paradigm was called into question few decades ago, after discovering proteins lacking stable tertiary structure, but still performing biological functions. These proteins are termed intrinsically disordered proteins (IDPs) [21]. IDPs exhibit very dynamic nature and high flexibility, characterised by heterogeneous conformational ensembles. The polypeptide chain does not adopt any stable structure, however it forms residual secondary structure motifs or transient long-range contacts [22].

The intrinsic disorder might resemble denatured state of well-structured proteins or random coils, however the disorder is encoded in the amino acid sequence. IDPs contain large amount of disorder-promoting amino acids: Ala, Arg, Gly, Gln, Ser, Glu, Lys and Pro. Occurrence of these residues gives rise to low overall hydrophobicity and hight net

charge of IDPs. The low content of hydrophobic amino acids causes failure of adopting compact structure and the hight net charge leads to repulsion [21, 23].

IDPs are very abundant in living cells. The abundance of IDPs in bacteria and archea is predicted to 16–45% and 26–51%, respectively. Eukaryotic proteins are predicted to contain 25–30% of IDPs and 52–67% of proteins including disordered region [24, 25].

High occurrence of IDPs in organisms indicates their important function in fundamental cellular processes. Pliability of these proteins provides increased interaction surface area and conformational flexibility. IDPs are able to interact with many different binding partners. The interaction is often followed by a disorder-to-order transition and binding to another binding partner with a high specificity and low affinity [22]. The aforementioned properties are crucial for proper function of IDPs in cells. These proteins play a key role in signaling and regulation pathways, cell cycle control or cellular organisation [26, 27, 28].

The improper behaviour of IDPs is associated with diverse neurodegenerative diseases, such as Alzheimer's disease, Down's syndrome, Parkinson's disease or prion diseases [29, 30, 31]. Moreover, the misbehaviour of IDPs is connected to cancer [32, 33].

The investigation of biophysical properties of IDPs is possible using numerous spectroscopic methods such as single molecule Förster resonance energy transfer (smFRET) [34], small angle X-ray scattering (SAXS) [35] or fluorescence correlation spectroscopy (FCS) [36]. The most convenient and universal method for observing structural parameters of IDPs is nuclear magnetic resonance spectroscopy (NMR) [37].

1.3 Structural characterization of IDPs using NMR

Nuclear magnetic resonance spectroscopy is an experimental method which provides information about both structure and dynamics in biomolecules at atomic resolution. The development of high-field NMR technology and using recombinant proteins containing ^{13}C and ^{15}N isotopes enabled highly automated procedures of structural investigation of well-ordered proteins.

However, the well-established methods used to study folded proteins can not be applied in the case of IDPs. The highly flexible nature and conformational plasticity of IDPs lead to extensive signal-overlap and high chemical shift degeneracy of the observed nuclei in spectra. Moreover, exposure of the protein backbone to the solvent enables significant $^{1}H^{N}$ chemical exchange with water. Higher chemical exchange rates cause peak broadening and reducing $^{1}H^{N}$ dispersion, thus using the common ^{1}H-direct detection with high sensitivity results in crowded spectra with low noise-to-signal ratio.

Therefore, the structural characterization of IDPs by NMR requires novel approaches. Very high resolution in spectra is necessary in order to avoid spectral overlaps. Improved resolution is accomplished by increasing spectra dimensions to 4D, 5D or even 7D. Recording this kind of spectra demands using sparse sampling with regards to reduction of the time of measurement. The problems with low ^{1}H dispersion and peak broadening due to the amide proton chemical exchange are solved by employing ^{13}C-direct detection experiments. The development of innovative methodology and instrumentation has promoted NMR the only experimental technique for obtaining high resolution structural information about IDPs.

IDPs do not adopt a unique structure corresponding to the lowest energy conformation, but their "structure" is represented by an ensemble of various conformations. Therefore, determination of a single unique structure for IDPs is not possible. However, the structural constraints characterizing an ensemble of states can be obtained. Residual secondary structure, transient long-range contacts and regions of restricted or enhanced mobility can be detected by various NMR techniques and it can be used for structural characterization of IDPs [38].

1.3.1 Chemical shift

Chemical shifts are highly sensitive probes for detecting residual secondary structure in proteins [39, 40]. This structural parameter can be easily obtained directly from assigned spectra of the protein.

Resonance frequencies of nuclei with non-zero spin numbers in strong magnetic fields are measured in NMR experiments. The resonance frequency of the measured nucleus depends on the applied external magnetic field B_0, magnetogyric ratio of the nucleus γ and chemical shift δ, which describes shielding caused by the local environment of nucleus.

$$\omega = -\gamma(1+\delta)B_0 \tag{1.1}$$

The chemical shift is given by difference of resonance frequencies of the sample and of a reference compound. Widely used reference compounds are 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) in biochemistry or tetremethylsilane (TMS) in chemistry. The chemical shift is expressed in parts per million (ppm), which is convenient for its independence on the applied magnetic field.

$$\delta = \frac{\omega_{\text{sample}} - \omega_{\text{reference}}}{\omega_{\text{reference}}} \tag{1.2}$$

Obtained chemical shifts are very sensitive to the local environment and molecular structure. Therefore the secondary structure propensities can be extracted from deviations of chemical shifts from their random coil values. Random coil shifts are resonance frequencies referring to completely unstructured proteins. The difference between observed and random coil chemical shifts is termed as the secondary chemical shift $\Delta\delta$.

$$\Delta\delta = \delta_{\text{observed}} - \delta_{\text{random coil}} \tag{1.3}$$

It is important to use reliable random coil shifts calibrated with respect to the used pH and temperature [41]. Secondary chemical shifts obtained for various types of nuclei can be combined and used for a more precise calculation of transient secondary structure propensity. Sophisticated methods for determination of secondary structure in IDPs from various secondary chemical shifts with respect to the primary structure were developed [39].

1.3.2 Paramagnetic relaxation enhancement

Measurement of paramagnetic relaxation enhancement provides useful information about transient long-range order in IDPs. Paramagnetic relaxation enhancement (PRE) arises

from the dipole-dipole interaction between the electron spin from the paramagnetic center and the spin of the reporting nucleus. The nuclei in proximity of the electron spin experience rapid relaxation, resulting in line broadening in NMR spectra. The PRE effect is very strong due to the high magnetic moment of unpaired electron, generating great local magnetic field [42]. Therefore the measurement of PRE is often employed in order to map distant transient contacts in range from 1 to 2.5 nm.

The PRE measurement requires paramagnetic labelling, typically thiol-reactive nitroxide compounds are inserted in the protein as paramagnetic spin labels . These labels are attached through disulfidic bonds to the single cysteine residue, which is introduced by site-directed mutation at specified position in the protein sequence. Remaining cysteine residues are removed from the protein [43].

The spin label affects both transverse and longitudinal relaxation rates of the nuclei close in space. The lifetime of excited states in the affected nuclei is decreased and the peak intensities in NMR spectra are lowered. Ten to fifteen residues neighbouring the spin label experience the line broadening, nuclei further in the sequence exhibiting the line broadening indicate the long-range contact. The transient long-range contact is revealed by the comparison of peak intensities between the paramagnetic and reference spectra ($I_{\text{para}}/I_{\text{dia}}$). The distance r_{IS} between the paramagnetic label and the nuclei can be calculated from the increased transverse relaxation rate R_2^{PARA} according to the Solomon-Bloembergen equation:

$$r_{\text{IS}} = \left[\frac{K}{R_2^{\text{PARA}}} \left(4\tau_c + \frac{3\tau_c}{1+\omega_{\text{I}}^2\tau_c^2} \right) \right] \tag{1.4}$$

where τ_c is correlation time, ω_{I} is the Larmor frequency of the nuclei and K is expressed as:

$$K = \frac{1}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \gamma_{\text{I}}^2 g_e^2 \mu_B^2 S(S+1) \tag{1.5}$$

where μ_0 is the permeability of the vacuum, g_e is the g-factor of the electron, μ_B is the Bohr magneton and S is the total electron spin quantum number.

1.3.3 Residual dipolar coupling

Residual dipolar couplings (RDCs) represent another useful probe for detection of transient secondary structure and description of the conformational behaviour of IDPs. RDC originates from interaction between two magnetic dipoles close in space, this effect is called dipole-dipole interaction. The interaction contains information about mutual distances of observed nuclei and orientation of the vector connecting the two spins with respect to the external field. The dipole-dipole interaction in isotropic environment is averaged to zero due to the Brownian motion. Therefore, the sample needs to be transferred into an anisotropic environment, where the symmetry of molecular tumbling is broken and the resonance splitting caused by dipolar interaction appears in addition to the scalar coupling. RDCs are then obtained as the difference between scalar coupling in isotropic and apparent coupling in anisotropic environment. The value of RDC D_{ij} is dependent on the distance r between

nuclei i and j, characterized by gyromagnetic ratios γ_i and γ_j, and on the orientation of the internuclear vector with respect to the static magnetic field as follows:

$$D_{ij} = -\frac{\gamma_i \gamma_j \hbar \mu_0}{8\pi^2 r^3} \left\langle \frac{3cos^2\theta - 1}{2} \right\rangle \tag{1.6}$$

where $\hbar$ is reduced Planck constant, μ_0 is the permeability in vacuum, θ is the angle between the inter-spin vector and the external magnetic field, and the angular parentheses express an average over exchanging conformations [44].

The anisotropic environment can be generated by various orienting media as phospholipid bicells, filamentous phages, purple membrane fragments or stretched polyacrylamide gels [45, 46].

The RDC values of IDPs are usually lower than values of well folded proteins due to weaker alignment of flexible regions [47]. The type of transient secondary structure can be detected from preferred alignment of molecules. NH-bond vectors are typically aligned parallel or orthogonal in helical or extended secondary structure, respectively, which results in a positive or negative values of RDCs [48, 47]. Additionally, combination of different type of couplings along protein backbone, such as RDC of $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ and of carbonyl ^{13}C and ^{15}N in the peptide bond, reveals transient long-range order in IDPs [49, 43]. The direct interpretation of RDCs is not straightforward, therefore computational methods for obtaining long-range and local structure features from RDCs in combination with other experimental data were developed [44].

1.3.4 Conversion of experimental data to structural parameters

In the analysis of the data from δ subunit the software A Selection Tool for Ensemble Representation Of Intrinsically Disordered States (ASTEROIDS) was used [50]. The ASTEROIDS combines the various experimental data in order to create ensemble of the structures, which describes the conformational behaviour of the intrinsically disordered protein. The information about the global structural features is extracted from the Small Angle X-ray Scattering (SAXS) and Paramagnetic Relaxation Enhancement (PRE), whereas the information about the local conformation is encoded in the chemical shifts [51, 49]. The RDCs report on both local and global arrangement [52]. The ASTEROIDS analysis converts the diverse experimental data to structural parameters, which define the average representative ensemble and provide the interpretation of the obtained data.

Initially an ensemble representing all possible states is generated on the basis of the protein sequence using statistical random coil sampling from the *flexible-meccano* approach. In this step pool containing tens thousands of different structures is produced. Then the experimental observables are calculated for each of these structures. Next, a smaller set of hundreds of structures is selected on the basis of the experimental data. The generated structures from the large pool are iteratively exchanged with the smaller ensemble in order to obtain a set of structures with the averaged predicted observables corresponding to the experimental data. This ensemble of hundreds structures represents conformational space populated by the protein. The structural parameters extracted from the representative ensemble provide meaningful description of IDPs behaviour [50].

1.4 Protein dynamics

The IDPs are typically very flexible and their structure can be described by a high number of inter-converting states. The mobility of proteins is not uniform and usually different degrees of flexibility can be distinguished in various regions. Regions with restricted or enhanced mobility are usually connected with important biological function, such as target recognition or binding. Therefore the description of motional properties of IDPs is another desirable feature, which is used for biological function recognition.

NMR spectroscopy is well suited for monitoring protein dynamics at atomic level because molecular motion can be described by the NMR relaxations. The relaxation is a phenomenon when nuclei excited by a radiofrequency pulse return to their thermal equilibrium distribution. The process of relaxations is mediated by random fluctuating magnetic fields arising from internuclear motions and molecular tumbling [53]. The main sources of the fluctuating fields in biomolecules are dipole-dipole interactions and chemical shift anisotropy (CSA). The return of the spin to the equilibrium is a result of longitudinal relaxation T_1. The longitudinal relaxation, also called spin-lattice relaxation, is a process of recovering α and β spin states in the thermal equilibrium through energy transfer to neighbouring nuclei. The transverse relaxation (spin-spin relaxation) is also mediated by random fluctuating fields and represents the loss of coherence of magnetization. The longitudinal and transverse relaxations can be described by longitudinal and transverse relaxation rates R_1 and R_2, respectively. Additionally, the correlation of distinct fluctuations such as dipole-dipole, dipole-CSA or CSA-CSA can be observed. The interference between two different relaxation mechanism is described by cross-correlated longitudinal and transverse relaxation rates Γ_x and Γ_z [54]. The relaxation rates are related to dynamic properties of individual nuclei and report on sampling of different conformations [55]. Another useful parameter characterizing dynamics is heteronuclear steady state NOE enhancement (ssNOE). The exchange of z-magnetisation between nucleus in thermal equilibrium and saturated nucleus coupled by dipolar interaction takes place during the measurement of ssNOE [56]. The ssNOE values are typically used as a good indicators of flexibility of the polypeptide chain.

Relaxation rates report on various molecular motions, occurring on ps–ns time scale for fast motions and on μs–ms time scale for conformational exchanges. The separation of internal motions from the rotational diffusion is not feasible in case of the IDPs, therefore the spectral density function mapping (SDFM) is the only approach how to describe the effect of molecular motions on the relaxation rates [57].

Spectral density function $J(\omega)$ is Fourier transform of the correlation function $G(\tau)$. The correlation function is function of correlation time(s) τ_c and describes the time dependence of random motion. Analogously to the processing of free induction decay, the Fourier transform of the correlation function provides the function dependent on frequency $\mathrm{J}(\omega)$. The spectral density function reports on the frequency distribution of the molecular motion, thus it is used for interpretation of the dynamic behaviour of proteins. The linear combination of the spectral density functions at discrete frequencies determines the relaxation rates. The spectral density values are calculated by the substitution of relaxation rates R_2, R_1, ssNOE expressed as a ratio of peak intensities in the steady-state (I_{ss}) and reference (I_{ref}) spectra, Γ_x and Γ_z, rescaled to the same units as the spectral density values,

in the set of equations:

$$(2R_2 - R_1)/d^2 = 2\xi + 8b\bar{J}(0) + 12J^{\mathrm{dd}}(\omega_{\mathrm{H}}) \tag{1.7}$$

$$R_1/d^2 = 6b\bar{J}(\omega_{\mathrm{N}}) + 2J^{\mathrm{dd}}(\omega_{\mathrm{H}} - \omega_{\mathrm{N}}) + 12J^{\mathrm{dd}}(\omega_{\mathrm{H}} + \omega_{\mathrm{N}}) \tag{1.8}$$

$$\frac{\gamma_{\mathrm{N}}}{\gamma_{\mathrm{H}}}\left(\frac{I_{\mathrm{ss}}}{I_{\mathrm{ref}}} - 1\right)R_1/d^2 = -2J^{\mathrm{dd}}(\omega_{\mathrm{H}} - \omega_{\mathrm{N}}) + 12J^{\mathrm{dd}}(\omega_{\mathrm{H}} + \omega_{\mathrm{N}}) \tag{1.9}$$

$$b\Gamma_x/(pcd) = 2b(4J^{\mathrm{cd}})(0) + 3J^{\mathrm{cd}}(\omega_{\mathrm{N}}))/p \tag{1.10}$$

$$b\Gamma_z/(pcd) = 12bJ^{\mathrm{cd}}(\omega_{\mathrm{N}})/p \tag{1.11}$$

where $b = 1 + c^2/d^2$, $d = -\mu_0 h\gamma_{\mathrm{H}}\gamma_{\mathrm{N}}/(16\pi^2 r_{\mathrm{HN}}^3\sqrt{5})$, $c = \gamma_{\mathrm{N}}B_0\Delta_{\mathrm{N}}/(3\sqrt{5})$, γ_{H} and γ_{N} are the magnetogyric ratios of ^{1}H and ^{15}N respectively, r_{HN} is the ^{1}H - ^{15}N internuclear distance, μ_0 is the permeability of vacuum, h is the Planck's constant, Δ_{N} is the anisotropy of the ^{15}N chemical shift tensor, $p = (3\cos^2\theta_{c,d} - 1)/2$, $\theta_{c,d}$ is the angle between the ^{1}H-^{15}N bond and the symmetry axis of the ^{15}N chemical shift tensor, and B_0 is the external magnetic field induction. The $J^{\mathrm{dd}}(\omega)$ and $J^{\mathrm{cc}}(\omega)$ functions are the auto-correlation spectral density functions and J^{cd} is the cross-correlation spectral density function, $\xi = R_{\mathrm{ex}}/d^2$ is the contribution of slow exchange and $\bar{J}(\omega) = (J^{\mathrm{dd}}(\omega) + c^2J^{\mathrm{cc}}(\omega)/d^2/b)$. The spectral density functions J^{dd}, J^{cc} and J^{cd} can be substituted by a single function $J(\omega)$ for isotropically moving molecules. With the knowledge of relaxation rates then the spectral density functions values at five distinct frequencies $J(0)$, $J(\omega_{\mathrm{N}})$, $J(\omega_{\mathrm{H}})$, $J(\omega_{\mathrm{N}} + \omega_{\mathrm{H}})$ and $J(\omega_{\mathrm{N}} - \omega_{\mathrm{H}})$ are calculated from the set of five linear combinations (Equations 1.7,1.8, 1.9, 1.10, 1.11) [58]. This approach is not convenient from the experimental point of view, therefore the high-frequency spectral density functions $J^{\mathrm{dd}}(\omega_{\mathrm{H}})$, $J^{\mathrm{dd}}(\omega_{\mathrm{N}} + \omega_{\mathrm{H}})$ and $J^{\mathrm{dd}}(\omega_{\mathrm{N}} - \omega_{\mathrm{H}})$ are replaced by a single spectral density function $J^{\mathrm{dd}}(\varepsilon_{\mathrm{H}})$. This reduction is applicable due to the high value of ω_{H} in comparison to the value of ω_{N}. Afterwards various combination of different relaxation rates can be used for interpretation of dynamic behaviour of proteins using reduced spectral density function mapping [58]:

$$(2R_2 - R_1)/d^2 = 2\xi + 8b\bar{J}(0) + 12J^{\mathrm{dd}}(\varepsilon_{\mathrm{H}}) \tag{1.12}$$

$$R_1/d^2 = 6b\bar{J}(\omega_{\mathrm{N}}) + 14J^{\mathrm{dd}}(\varepsilon_{\mathrm{H}}) \tag{1.13}$$

$$\frac{\gamma_{\mathrm{N}}}{\gamma_{\mathrm{H}}}\left(\frac{I_{\mathrm{ss}}}{I_{\mathrm{ref}}} - 1\right)R_1/d^2 = 10J^{\mathrm{dd}}(\varepsilon_{\mathrm{H}}) \tag{1.14}$$

1.5 Aim of thesis

The main aim of the thesis was to investigate the role of electrostatic interactions in structural behaviour of the δ subunit of RNA polymerase from *Bacillus subtilis*. Structure of the δ subunit was previously studied in our laboratory in detail [4, 13, 14]. Various NMR experiments provided a clear evidence of the intramolecular contacts in the δ subunit and of the partial ordering in the unstructured CTD [4]. The presence of the highly conserved lysine stretch in the negatively charged CTD suggests that the well-balanced interplay between the attractive and repulsive interactions in CTD govern the motional properties and influence the global structure of the δ subunit. On the basis of these assumptions we decided to investigate the impact of the electrostatic interactions.

Initially we investigated the electrostatic interactions by increasing the salt concentration from 10 mM to 200 mM NaCl in the sample. The ions of sodium chloride were supposed to weaken the electrostatic attraction between the lysine stretch and the negatively charged C-terminal tail. However the chemical shift analysis and residual dipolar coupling analysis revealed that the impact of increasing salt concentration on the structure is rather small. Therefore we decided to replace the positively charged lysine stretch by glutamate residues in order to examine the influence of the electrostatic interactions in a more controlled manner.

The mutation of the lysine stretch was introduced to the δ subunit. The substitution of the positively charged lysines for the negatively charged glutamates was expected to disrupt the electrostatic interaction and the potential changes in the structure were monitored using nuclear magnetic resonance. Various observables from the NMR experiments were measured and compared with the previously measured data from the δ wild type in order to reveal the influence of electrostatic interactions on conformational behaviour of the δ subunit. The secondary structure propensity of the δ mutant was probed by a chemical shift analysis. The long-range order was inspected using the residual dipolar coupling and the paramagnetic relaxation enhancement. Also ^{15}N relaxation experiments were employed in order to characterize the motional properties of the δKE mutant.

Chapter 2

Materials and Methods

2.1 Protein sample expression and purification

The expression and purification protocol was adapted from the Laboratory of Microbial Genetics and Gene Expression of Libor Krásný at the Institute of Microbiology at the Czech Academy of Sciences. Nowadays, the procedure is established in the laboratory of Protein Structure and Dynamics at CEITEC, Masaryk University.

The δ subunit was prepared both in a wild type form (δWT) and two mutated forms. The first mutation was introduced to lysine stretch (96 KAKKKKAKK 104), where positively charged lysine residues were substituted by negatively charged glutamic acid residues (96 EAEEEEAEE 104) (δKE). The influence of this mutation was supposed to reveal the impact of electrostatic interactions on conformational behaviour of the protein. The other mutant includes the previously mentioned mutation with an additional one-point mutation in L132 for C132 (δKE-L132C). This site-directed mutation was required for paramagnetic resonance enhancement labelling.

The amino acid sequences of the δKE and δKE-L132C mutants are shown in Figures 2.1 and 2.2, respectively.

```
GIKQYSQEELKEMALVEIAH ELFEEHKKPVPFQELLNEIA
SLLGVKKEELGDRIAQFYTD LNIDGRFLALSDQTWGLRSW
YPYDQLDEETQPTVEAEEEE AEEAVEEDLDLDEFEEIDED
DLDLDEVEEELDLEADDFDE EDLDEDDDDLEIEEDIIDED
DEDYDDEEEEIK
```

Figure 2.1: The sequence of the δKE mutant. The color-coding corresponds to that previously used in Figure 1.1. The region highlighted in green represents the mutation of lysines for glutamates.

GIKQYSQEELKEMALVEIAH ELFEEHKKPVPFQELLNEIA
SLLGVKKEELGDRIAQFYTD LNIDGRFLALSDQTWGLRSW
YPYDQLDEETQPTVEAEEEE AEEAVEEDLDLDEFEEIDED
DLDLDEVEEECDLEADDFDE EDLDEDDDDLEIEEDIIDED
DEDYDDEEEEIK

Figure 2.2: The sequence of δKE-L132C mutant. The color-coding corresponds to that previously used in Figure 2.1. The letter C highlighted in olive represents the mutation of leucine for cystein.

Competent *Escherichia coli* cells strain BL21(DE3) transformed with plasmid pET22b containing the δ *rpoE* gene were spread on a Petri dish with ampicilin and cultivated overnight at 37 °C. One colony was inoculated into 5 ml of the LB medium containing ampicilin. The inoculum was incubated for 4 hours at 37 °C and 250 rmp. After incubation, 800 μl of the inoculum was transferred into 80 ml of minimal M9 medium and the preculture was grown overnight.

Next day, the preculture was transferred to 1920 ml of minimal M9 medium and it was grown at 37 °C and 250 rpm in the incubator. After OD_{600} reached the value 0.8, protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to the final concentration at 0.8 mM. The expression lasted 3 hours at the laboratory temperature.

The expression was terminated by cooling the culture on ice for 20 minutes. The cells were harvested by centrifugation for 15 minutes at 4 °C and 7903 g. The pelet was resuspended in the lysis buffer (100 mM bis-Tris pH 6, 2mM EDTA pH 8, 100 mM NaCl, 0.1 mM DTT, 1 mM β-mercaptoethanol, 1mM PMSF) and disrupted by using ultrasonication. The δ subunit, present in the supernatant, was separated from cellular compartments by centrifugation for 15 minutes at 4 °C and 30000 g.

The protein was purified by anion exchange chromatography in a batch mode using Whatman DE52 cellulose. Supernatant was mixed with buffer A (50 mM bis-Tris, pH 6, 100 mM NaCl, 0.1 mM DTT, 1 mM β-mercaptoethanol, 1 mM PMSF) and cellulose gel and the mixture was stirred for 5 hours at 4 °C. Afterwards the mixture was left in a cold room overnight for sedimentation. The sediment was transferred to the chromatographic column PolyProp (BioRad) and the protein was eluted by increasing the salt concentration at 4 °C (range 0.1 mM–0.6 mM NaCl). The presence of protein was analysed by the Bradford protein assay. Selected fractions were tested by 12% polyacrylamide gel electrophoresis (SDS-PAGE).

Fractions containing the recombinant protein were isoelectrically precipitated (sodium acetate, pH 3.6) and centrifuged for 5 minutes at 4 °C and 17000 g. Afterwards, the precipitate was diluted in the dissolving buffer (50 mM Tris-HCl, pH 8, 0.1 mM NaCl) and its pH was adjusted to the range 6–7. The diluted protein was transferred to a dialysis membrane and it was dialysed overnight in dialyses buffer (20 mM sodium phosphate, pH 6.6, 10 mM NaCl).

The last purification step was the ion exchange chromatography performed on a Resource Q chromatography column (GE Healthcare). The purified protein was concentrated by ultrafiltration on a centrifugal membrane with the cut-off 3 kDa. The final concentration was measured using the Bradford protein assay.

2.2 Protein sample preparation for NMR measurement

2.2.1 Isotropic samples

The δWT and δKE samples were measured in the phosphate buffer (20 mM, pH 6.6), containing NaCl (10 mM), NaN_3 (0.05%) with an addition of 10% D_2O.

2.2.2 Anisotropic sample

Partial alignment of the protein sample was achieved by using a stretched polyacrylamide gel as an orienting environment. The aqueous stock solution contained acrylamide and *N*, *N*′-methylenebisacrylamide mixed in a ratio 49:1. Polymerization was induced by adding ammonium persulfate (APS) and tetramethylethylenediamin (TEMED). The final volume of the 5% stress-oriented polyacrylamide gel was 350 μl. The gel was prepared by mixing 58 μl of stock solution, 3.5 μl of APS, 0.35 μl of TEMED and 287.8 μl of water. The mixture was transferred into a gel chamber for 4 hours to polymerise. Afterwards the gel was displaced from the chamber into 500 ml of deionized water. The gel was left in water overnight in order to remove unreacted chemicals. Subsequently the gel was dried at 38 °C and the shrank gel was again transferred into the gel chamber and 250 μl of protein sample was added. The protein solution was soaked into the gel overnight and then it was pushed into an NMR sample tube. The ability of the medium to align molecules was checked by measuring the HDO line splitting in a 1D spectrum. The splitting of the δKE sample was 1.15 Hz.

2.2.3 Spin-labelled sample

The paramagnetic relaxation enhancement measurement requires a paramagnetic spin label present in the protein molecule. The paramagnetic radical (1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate (MTSL; Toronto Research Chemicals, Inc.) was applied for observing transient contact of CTD. Also, the analogous diamagnetic compound (1-acetoxy-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate (MTS; Toronto Research Chemicals, Inc.) was incorporated into the protein molecule in order to precisely reference the difference between the samples. These spin labels form covalent disulfide linkage with cystein in biomolecules.

The protein sample of δKE-L132C, containing a single Cys mutation, was concentrated and a short fast spinning of sample was applied. The sample volume was divided into two parts, for labelling with paramagnetic MTSL and analogous diamagnetic MTS labels. In order to avoid disulfide bridges formation between the mutated cysteins, dithiothreitol (DTT) was added to the final concentration of 5 mM DTT. The protein sample was incubated with DTT for 1 hour on ice. After reduction the sample was purified from DTT by gel filtration on a desalting PD10 column. Concentration of both samples was checked by measuring absorbance at 280 nm at NanoDrop (Thermo Scientific). The paramagnetic or diamagnetic labels were added separately to two protein samples in a 10-fold excess while stirring drop by drop. Labelled protein samples were incubated in a cold room under stirring for 2 hours. The purification of sample was performed by dialysis in the NMR buffer while stirring overnight. The final concentration of each labelled protein sample

was checked by both Bradford protein assay and bicinchoninic acid protein assay (BCA). The concentrations of diamagnetic and paramagnetic samples were 0.30 mM and 0.27 mM respectively.

2.3 NMR measurement

The experimental set-up and measurements of spectra were performed by Lukáš Žídek, Milan Zachrdla, Pavel Srb and Jiří Nováček at various spectrometers according to the given purposes. My main task was the processing and the analysis of the measured spectra. The reference spectra for δWT were measured and analysed by Veronika Papoušková or Jiří Pol, unless stated otherwise.

2.3.1 Resonance assignment

Assignment of resonance frequencies of δKE and δWT was accomplished using 5D CACONCACO experiment for backbone assignment and 5D HC(CC-TOCSY)CONH experiment for side-chain assignment [3], [59]. The spectra were measured on Bruker Avance 600 MHz spectrometer equipped with a cryogenic probehead (5 mm CPQCI ^{1}H-^{31}P/^{13}C/^{15}N/D). The experimental set-up including real points, carrier frequencies, spectral width and number of scans is listed in Tables 2.1, 2.3, 2.2 and 2.4. Temperature was calibrated to 27°C using neat methanol. Chemical shifts were referenced to DSS.

Table 2.1: Parameters used to record 5D CACONCACO spectrum of δWT

	Direct dimension	Indirect dimensions			
	$^{13}C'$	$^{13}C^{\alpha}$	^{15}N	$^{13}C'$	$^{13}C^{\alpha}$
Real points	1024		1200[a]		
Scans	8				
Spectral width / ppm	40	22.71	28.19	11.36	22.71
Carrier frequency / ppm	175.01	55.01	123.56	175.01	55.01

[a] - number of non-uniformly sampled points in indirect dimensions

Table 2.2: Parameters used to record 5D HC-(CC-TOCSY)CACON spectrum of δWT

	Direct dimension	Indirect dimension			
	$^{13}C'$	^{15}N	$^{13}C^{\alpha}$	$^{13}C^{ali}$	$^{1}H^{ali}$
Real points	1024		1200[a]		
Scans	2				
Spectral width / ppm	40	28.18	22.71	71	7.14
Carrier frequency / ppm	175	123.56	55.01	41.01	4.74

[a] - number of non-uniformly sampled points in indirect dimensions

Table 2.3: Parameters used to record 5D CACONCACO spectrum of δKE

	Direct dimension	Indirect dimensions			
	$^{13}C'$	$^{13}C^{\alpha}$	^{15}N	$^{13}C'$	$^{13}C^{\alpha}$
Real points	1024		1200[a]		
Scans	8				
Spectral width / ppm	40	22.71	28.19	11.36	22.71
Carrier frequency / ppm	175	55	123.55	175	55

[a] - number of non-uniformly sampled points in indirect dimensions

Table 2.4: Parameters used to record 5D HC-(CC-TOCSY)CACON spectrum of δKE

	Direct dimension	Indirect dimension			
	$^{13}C'$	^{15}N	$^{13}C^{\alpha}$	$^{13}C^{ali}$	$^{1}H^{ali}$
Real points	1024		1200[a]		
Scans	2				
Spectral width / ppm	40	35.23	22.71	70.98	7.14
Carrier frequency / ppm	175	123.53	55	40.97	4.71

[a] - number of non-uniformly sampled points in indirect dimensions

2.3.2 Residual dipolar couplings

The measurements of δKE RDCs were performed on the Bruker Avance 600 MHz spectrometer at 27 °C. Three types of residual dipolar couplings along the peptide backbone were measured: $^{1}D_{NH}$, $^{1}D_{C^{\alpha}C'}$ and $^{1}D_{C^{\alpha}H^{\alpha}}$. The $^{1}D_{NH}$ was measured using 3D IPAP HNCO experiment with sparse sampling [60], $^{1}D_{C^{\alpha}C'}$ and $^{1}D_{C^{\alpha}H^{\alpha}}$ using 4D IPAP (HA)CACONH experiment [61]. The experimental set-up is listed in Tables 2.5 and 2.6. Deuterium line splitting was 1.15 Hz.

Table 2.5: Parameters used to record 3D HNCO IPAP spectrum of δKE

	Direct dimension	Indirect dimensions	
	$^{1}H^{N}$	$^{13}C'$	^{15}N
Real points	2048	600	
Scans	24		
Spectral width / ppm	14	32.89	13.25
Carrier frequency / ppm	4.74	174.99	118

2.3.3 Relaxation measurements

The reference relaxation rates of δWT were measured and analysed by Pavel Srb in our laboratory. The spectra of δKE were recorded on the Bruker Avance 600 MHz spectrometer at 27 °C. Five ^{15}N relaxation rates were measured: steady state NOE, longitudal relaxation rate R_1, transverse relaxation rate R_2, cross-correlated transverse relaxation rate Γ_x and cross-correlated longitudal relaxation rate Γ_z. Experiments were measured using a double-labelled sample with non-uniform sampling. The delays 0.3, 0.5, 0.7 s were used

Table 2.6: Parameters used to record 4D (HA)CACONH spectrum of δKE

	Direct dimension	Indirect dimensions		
	$^{1}H^{N}$	$^{13}C^{\alpha}$	$^{13}C'$	^{15}N
Real points	2048		850	
Scans	12			
Spectral width / ppm	14	13.25	26.51	32.89
Carrier frequency / ppm	4.74	58.47	175.02	118.06

for measurement of cross-correlated transverse rate Γ_x, 0.1, 0.15, 0.2, 0.25 s for measurement of cross-correlated longitude rate Γ_z, 0.0448, 0.0672, 0.112, 0.1792, 0.1792, 0.2464, 0.3808, 0.784, 1.232 s for measurement of relaxation rate R_1 and delays 0.0288, 0.0576, 0.0864, 0.1152, 0.1728, 0.2592 s for measurement of relaxation rate R_2. The experimental parameters of R_1, R_2, ssNOE and Γ_x, Γ_z measurements are listed in Tables 2.7 and 2.8, respectively.

Table 2.7: Parameters used to record 3D R_1, R_2, ssNOE HNCO spectra of δKE

	Direct dimension	Indirect dimensions	
	$^{1}H^{N}$	$^{13}C'$	^{15}N
Real points	2048	500	
Scans	4		
Spectral width (ppm)	14.03	26.31	13.25
Carrier (ppm)	4.74	175.03	118.07

Table 2.8: Parameters used to record 3D Γ_x, Γ_z HNCO spectra of δKE

	Direct dimension	Indirect dimensions	
	$^{1}H^{N}$	$^{13}C'$	^{15}N
Real points	2048	500	
Scans	16		
Spectral width (ppm)	14.03	26.31	13.25
Carrier (ppm)	4.74	175.03	118.07

2.3.4 Paramagnetic relaxation enhancement

The spectra of δKE were recorded on Bruker Avance 600 MHz spectrometer at 27 °C. Measurement of paramagnetic relaxation enhancement included recording of two ^{1}H-^{15}N HSQC experiment for paramagnetic and diamagnetic sample. Experimental parameters are listed in Table 2.9. The recycle delay was set to 5 s.

2.4 Used software

The uniformly sampled NMR data were processed and analysed using programs NMRPipe and NMRDraw (version 8.9) [62]. The non-uniformly measured NMR data were processed

Table 2.9: Parameters used to record NH-HSQC spectra of δKE for PRE measurement

	Direct dimension $^{1}H^{N}$	Indirect dimensions ^{15}N
Real points	2048	512
Scans	18	
Spectral width (ppm)	13.95	32.89
Carrier (ppm)	4.742	116.334

by softwares reduced and cleaner3d from the package nuft [63]. Program Sparky 3.115 was used for visualisation, analysis and the assignment of the spectra [64]. The peak heights for the NOE, relaxation rate and PRE analysis were also measured in the program Sparky.

The secondary structure propensities from secondary chemical shifts were calculated by the SSP software [39]. The RDC values were validated using program dconv, developed in our laboratory by Petr Padrta. The processing of relaxation data was performed by program Relax 4.0.3 and relaxation rates were obtained by the fitting peak intensities to exponential decay in software Octave [65] using a set of in-house written unix shell scripts, developed in our laboratory by Pavel Kadeřávek. The error of relaxation rates was estimated by bootstrap [66]. The reduced spectral density function mapping was also performed using unix shell scripts, the values were calculated on the basis of equations 1.10,1.11, 1.12, 1.13, 1.14. The visualisation of biomolecule δ subunit was done using the program MOLMOL.

Chapter 3

Results and Discussion

3.1 Secondary structure propensity

In order to map transient secondary structure propensities in the flexible C-terminal domain, resonance frequency assignment was obtained from NMR spectra. Well-established assignment procedures failed in the case of the highly flexible δ subunit with multiple repetitions. High degeneracy and multiple overlaps of signals in spectra required measurement of sparsely sampled spectra with increased resolution.

Backbone assignment was achieved using the ^{13}C detected 5D CACONCACO experiment. This experiment provides sequential connectivity of nuclei in the backbone. First, peak picking in the 3D CACON spectrum was performed. The list of resonance frequencies from the 3D spectrum contained values for ^{15}N, $^{13}C^{\alpha}$ and $^{13}C'$ from single individual acid residues. These ”3D coordinates” were subsequently used to process the 5D CACONCACO spectrum. In order to spare the disk space, only 2D cross-sections were extracted from the 5D CACONCACO spectrum based on fixed frequencies from the CACON list of peaks. The obtained 2D cross-sections included resonance frequencies of $^{13}C^{\alpha}$ and $^{13}C'$ from the preceding amino acid residue. A combination of the 3D CACON spectrum and cross-sections was utilized to identify sequential connectivity and assign ^{15}N, $^{13}C^{\alpha}$, $^{13}C'$ atoms.

The 5D HC(CC-TOCSY)CACON spectrum was acquired for the side chain assignment. The first step of the assignment was identical with the previous procedure, peak picking was applied to the 3D CACON spectrum and the fixed coordinates were used for extracting the 2D cross-sections. Peaks in the cross-sections referred to $^{13}C^{ali}$ and $^{1}H^{ali}$ atoms from the side chain of the given residue. Resonance frequencies of amide hydrogen atoms were obtained from the 3D HNCO spectrum.

The mentioned strategy enabled the $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, $^{13}C'$, $^{13}C^{\beta}$, aliphatic $^{13}C^{ali}$ and $^{1}H^{ali}$ resonance assignment of 84 from 91 residues in the C-terminal domain of δKE and δWT. The experimental set-up was aimed at characterization of the flexible part and peaks for the structured part did not appear in spectra. The assignment covered 92.30% sequence of the C-terminal domain. The chemical shift analysis of δWT was successfully accomplished previously in our laboratory [4]. However, we decided to repeat analysis in order to achieve a precise reference for comparing δWT and δKE chemical shifts.

Chemical shifts are influenced by the local environment and carry structure infor-

mation. Deviations of the experimental shifts from the random coil values, so called secondary chemical shifts (SCS), detect regions with extended or helical structure. Secondary chemical shifts in intrinsically disordered regions are typically smaller than shifts of folded proteins, therefore an accurate calculation of random coil chemical shifts is necessary. Random coil shifts were calculated for the actual temperature and pH (27°C and 6.6, respectively) and neighbour-correction was applied. The following server was used: https://*spin.niddk.nih.gov/bax/nmrserver/PoulsenrcCS/*. Differences between experimental and random coil shifts for δWT and δKE are plotted in Figure 3.1. C^{α} is the most sensitive nucleus for the secondary structure determination, however it is not recommended to use only C^{α} data in case of IDPs for its unreliability to detect random coils and extended structures [67, 68]. Extended stretches of SCS with the same sign and similar magnitude reliably indicate formation of secondary structures. The C^{α} SCS for both δWT and δKE indicated presence of extended structure of β strand in C-terminal domain. Analysis of the $^{1}H^{N}$, ^{15}N, $^{1}H^{\alpha}$, $^{13}C'$ and $^{13}C^{\beta}$ nuclei also confirmed propensity of the C-terminal domain to form β strand.

Obtained values of secondary chemical shifts for δWT and δKE are plotted in Figure 3.1. Data for the δWT and δKE are color-coded in blue and red, respectively, this color-coding is constant within the thesis. The comparison of δWT and δKE shows significant changes in the region of mutation (E96–E104). Values of SCS for δWT in region (K96–K104) indicate random-coil-like behaviour, whereas SCS of δKE exhibit tendency to form extended secondary structure. It shows that substitution of the positively charged lysine stretch by negatively charged glutamic acid residues results in conformational change of C-terminal domain. This phenomenon confirms the hypothesis that a balanced interplay between attractive and repulsive electrostatic forces is crucial for proper conformational behaviour of the C-terminal domain.

The calculated secondary chemical shifts served as an input for the software SSP (Secondary Structure Prediction) [39]. This program combines information from different nuclei into a residue specific secondary structure propensity score. The SSP score ranges from -1 to 1, where 1 or -1 correspond to completely formed α helix or β strand. According to the SSP documentation, nuclei $^{13}C^{\alpha}$, $^{1}H^{\alpha}$ and $^{13}C^{\beta}$ are the most useful in case of the IDP analysis. The output of SSP analysis is plotted in Figure 3.2 for both δWT and δKE. Resulting SSP scores show 20–30% propensity for the β structure. δWT displays two regions (E96–E104, D126–L132) with the SSP score close to zero, indicating less than $\sim$10% propensity for any secondary structure. In the contrary, δKE shows increased SSP scores ($\sim$20%) in those regions. The comparison of δWT and δKE SSP scores are analogous to the direct SCS comparison and also confirm a significant influence of the electrostatic interactions on conformational behaviour of the C-terminal domain.

Another parameter related to the transient secondary structure in IDPs is the residual dipolar coupling. Aligned molecules with restricted motions provide measurable NH dipolar coupling, which is dependent on the type of the secondary structure. Observed values of $^{1}D_{\mathrm{NH}}$ can report on secondary structure because $^{1}D_{\mathrm{NH}}$ of helices and strand have opposite sign. Flexible polypeptide chain exhibits decreased magnitude of RDCs due to its dynamic, thus disordered regions can be detected by analysing RDCs. The $^{1}D_{\mathrm{NH}}$ values of both δKE and δWT, plotted in Figure 3.3, show dominantly negative values in the C-terminal domain. The δKE mutant exhibit increased propensity for formation of beta sheet in the

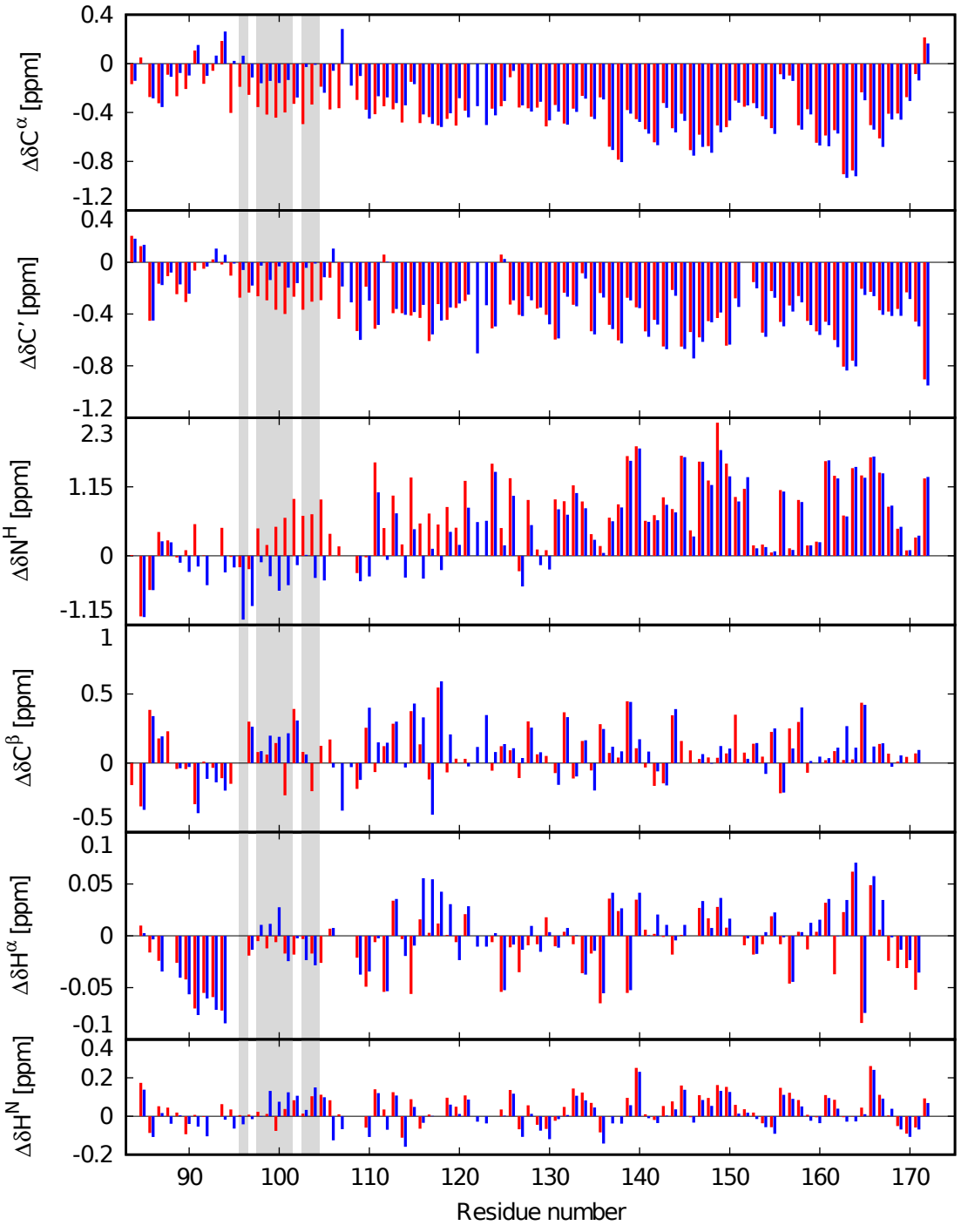

Figure 3.1: Secondary chemical shifts of δWT and δKE in blue and red, respectively. The grey rectangles mark the region of mutation.

region of mutation and in its vicinity in comparison to the δWT. This behaviour can be explained as an electrostatic repulsion between the mutated glutamate residues. Results of the analysis of both chemical shifts and $^1D_{\mathrm{NH}}$ are in agreement with the fact that the C-terminal domain forms extended secondary structure, enhanced in the case of mutated δKE. These results indicate that conformational changes in mutated δKE are induced by disturbing the of electrostatic balance in the C-terminal domain.

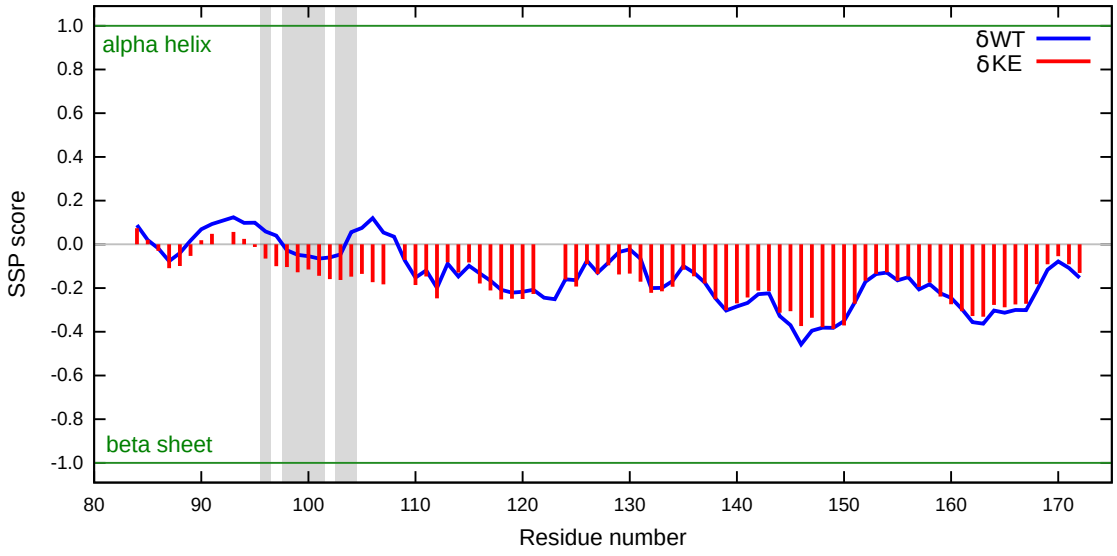

Figure 3.2: Secondary structure propensities for δWT and δKE in blue and red, respectively. The grey rectangles mark the region of the mutation.

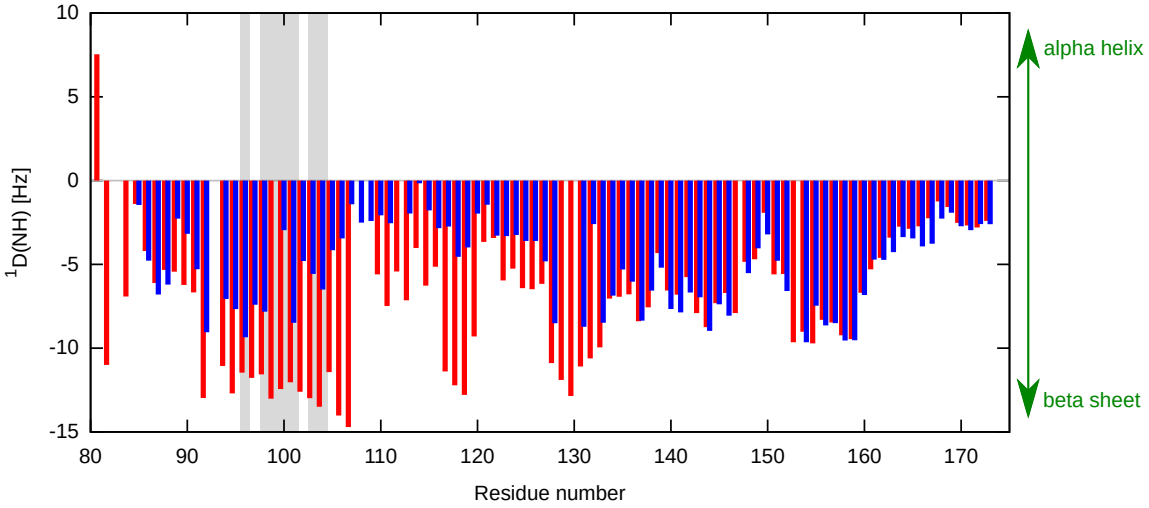

Figure 3.3: The obtained $^1D_{\mathrm{NH}}$ values for the δWT and δKE mutant in blue and red, respectively.

3.2 Long-range order

The previous studies of the long-range order in the δ subunit revealed the presence of the orientational restraints in the CTD and also the intramolecular contacts. The measurements of the residual dipolar couplings and paramagnetic relaxation enhancement were employed in order to characterize the transient tertiary structure of the δ subunit [4]. The long-range order is supposed to be caused by the electrostatic interactions in CTD of the δ subunit. The ASTEROIDS analysis for the δKE mutant, performed by Vojtěch Kubáň in our laboratory, predicted disappearance of the contact between the lysine stretch and negatively charged CTD as expected. However, the results of the ASTEROIDS analysis exhibited also disappearance of interdomain contacts. Therefore we decided to remeasure not only the residual dipolar coupling, but also the paramagnetic relaxation enhancement for the δKE mutant. The impact of the disruption of these electrostatic interactions on the long-range order was monitored by the comparison of the residual dipolar coupling and paramagnetic relaxation enhancement between δWT and the δKE mutant.

3.2.1 Residual dipolar couplings

The RDCs reflect the global conformational behaviour of molecule. Therefore the inspection of RDC values is useful for the obtaining potential changes in the long-range order, caused by the disruption of the electrostatic interaction in the δ subunit. The pattern of RDCs in random polymer chains (e.g. in random-coil polypeptides) can be described by a cosine hyperbolic function of the residue number (Fig 3.4). Deviations from this functional dependence indicate the presence of transient contacts of the polypeptide chain, which restrains the conformational behaviour of the polypeptide chain. The previous RDCs measurements on δWT, performed by Jiří Pol, revealed the orientational restraints in the region K100–E130. These results were interpreted as electrostatic interactions between the negatively and positively charged regions of CTD. Therefore the mutation of the lysine stretch to negatively charged residues was expected to dramatically change the RDCs profile.

In order to align the molecules with respect to the external field, the protein sample of δKE was transferred into the stretched 5% polyacrylamide gel. The HDO splitting reached the value of 1.15 Hz, which is sufficient for obtaining the dipolar coupling in the anisotropic sample. Three types of RDC were measured, namely ${}^1D_{\mathrm{NH}}$, ${}^1D_{\mathrm{C}^\alpha\mathrm{C}'}$ and ${}^1D_{\mathrm{C}^\alpha\mathrm{H}^\alpha}$, using In-Phase Anti-Phase (IPAP) experiments. The IPAP is spin-state-separation approach, where the α and β states of coupled spin are separated in two distinct subspectra. The spectra with in-phase (IP) and anti-phase (AP) components are measured. Subsequently the addition and subtraction of IP and AP spectra provides subspectra containing peaks corresponding to either α or β state. The RDC values D were calculated as the difference between the apparent coupling from anisotropic spectra J' and the scalar coupling from isotropic spectra J:

$$D = J' - J \tag{3.1}$$

Values of 149 ${}^1D_{\mathrm{NH}}$ couplings were extracted from 3D IPAP HNCO spectra, 63 values for the structured part and 86 values for the unstructured part. The ${}^1D_{\mathrm{C}^\alpha\mathrm{C}'}$ and ${}^1D_{\mathrm{C}^\alpha\mathrm{H}^\alpha}$ values were obtained from a 4D IPAP (HA)CACONH experiment. The experiment pro-

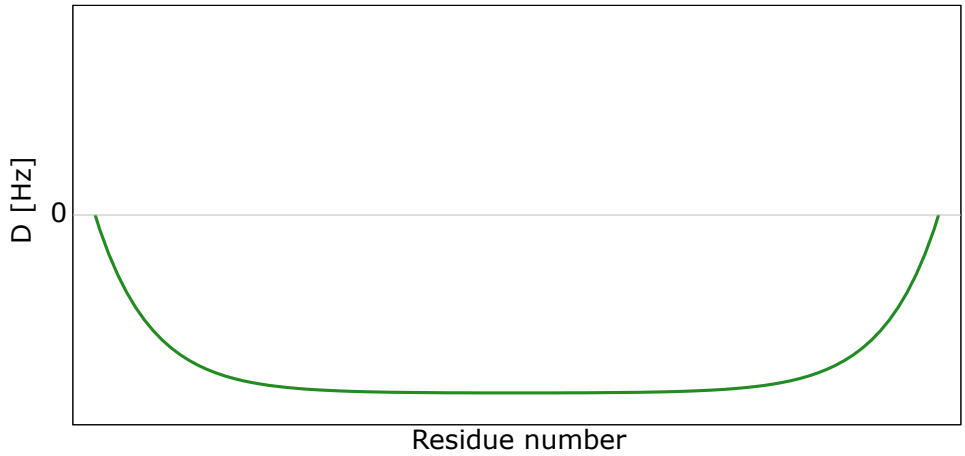

Figure 3.4: The theoretical RDCs trend for a completely disordered polypeptide chain, represented by a hyperbolic cosine function.

vides two 2D cross-sections: the 2D spectrum with ^{15}N-^{1}H correlation and the spectrum with $^{13}C^{\alpha}$-$^{13}C'$ correlation. First the peak picking in the 2D ^{15}N-^{1}H spectrum is applied and assigned resonance frequencies are used for processing 2D $^{13}C^{\alpha}$-$^{13}C'$ spectra. Splitting due to the $J'_{C^{\alpha}C'}$ and $J'_{C^{\alpha}H^{\alpha}}$ is visible in the 2D $^{13}C^{\alpha}$-$^{13}C'$ spectrum as displayed in Figure 3.5. I assigned 164 peaks in the 2D ^{15}N-^{1}H spectrum and 144 peaks in the 2D $^{13}C^{\alpha}$-$^{13}C'$ cross-sections. Due to a lower resolution of certain peaks I was able to extract only 119 $^{1}D_{C^{\alpha}C'}$ and $^{1}D_{C^{\alpha}H^{\alpha}}$ couplings from the spectra, 43 for the structured and 85 for the unstructured part. The obtained RDC values for δKE are plotted in Figure 3.6.

The interpretation of RDCs is not as straightforward as in the case of the chemical shift analysis. Nevertheless, the magnitudes and trends in the RDC coupling values along the polypeptide chain can provide information about the local fragment geometry and transient tertiary structure. The magnitude of the RDC coupling values for NTD is higher than for the flexible domain. The difference in the magnitude between NTD and CTD, most visible in $^{1}D_{NH}$ coupling values, confirms both the folding of NTD and the flexibility of CTD in δKE mutant. The trend of RDCs in the well-folded NTD corresponds to the δWT secondary structures. The RDC values of NTD were validated using the in-house written program dconv by comparing the experimental data with the data recalculated from the δWT structure (PDB ID: 2KRC). The distribution of the data is displayed in Figure 3.7. The experimental $^{1}D_{NH}$ coupling values are in agreement with the calculated data, the experimental data for $^{1}D_{C^{\alpha}C'}$ and $^{1}D_{C^{\alpha}H^{\alpha}}$ perform slight deviations. These deviations are possibly due to a lower resolution in the spectra of the structured part. The resemblance of experimental data from the δKE mutant with the recalculated data from δWT indicates that the structured part of the δ subunit is well-folded in the presence of lysine stretch mutation.

The RDC coupling values of CTD only are plotted in Figure 3.8 in order to observe the trend and to make comparison with δWT. The RDCs form the pattern resembling a cosine hyperbolic function of the residue number, most apparent in the $^{1}D_{C^{\alpha}C^{\alpha}}$ coupling values. Slight deviations of this trend are visible in the region L110–F115, E120–L125 and D150 in the $^{1}D_{NH}$ coupling values. The $^{1}D_{NH}$ coupling is sensitive to the local structural effects

and these effects are most likely the source of the observed deviations. Nevertheless, all three resulting values of RDCs for δKE evince the random polymer behaviour lacking any long-range order in contrast to the data for δWT. The previously observed decrease in the region K100–E130 in δWT is reduced in the δKE mutant. These results indicate that orientational restraints of the CTD clearly originate from the electrostatic interactions between the positively charged lysine stretch and the negatively charged region E114–E171. The disruption of electrostatic interactions in the CTD leads to the disappearance of the long-range order in the CTD and conformational behaviour of the CTD resembles a random polymer chain.

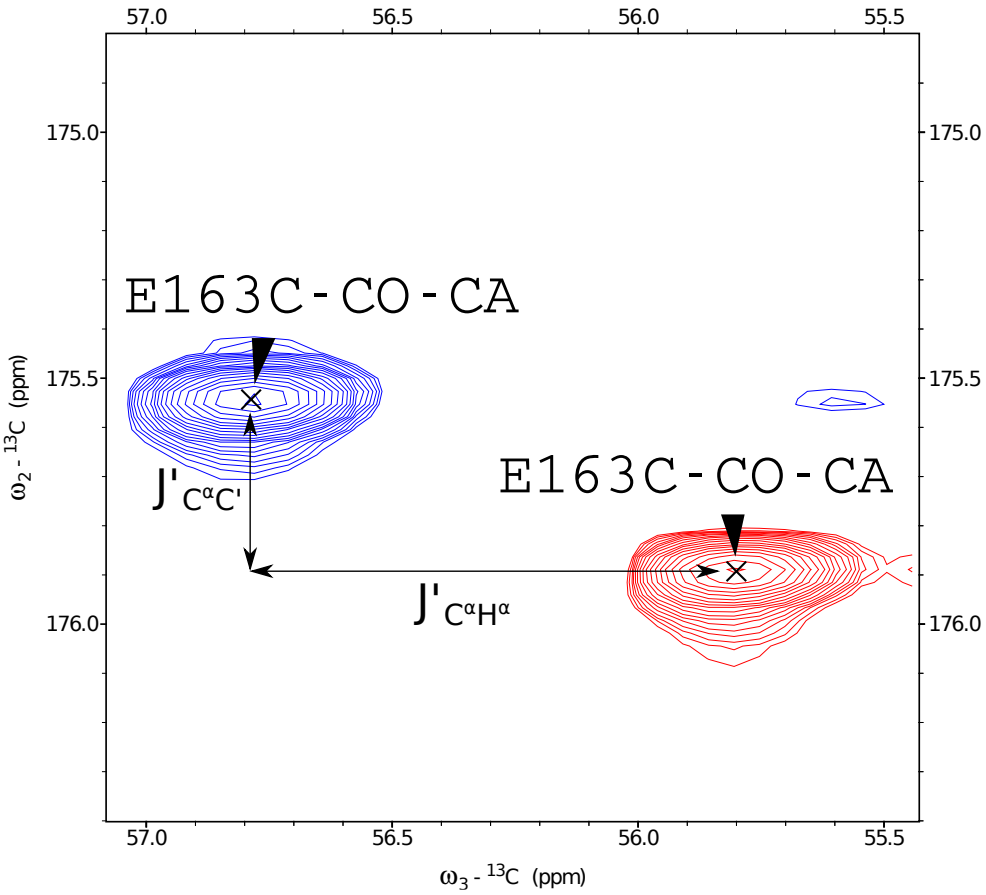

Figure 3.5: Overlaid anisotropic α- and β-state spectra in red and blue, respectively. The difference between the peaks on the horizontal axis refers to the apparent coupling for C^{α} and H^{α} nuclei and the difference on the vertical axis refers to the apparent coupling for C^{α} and C'.

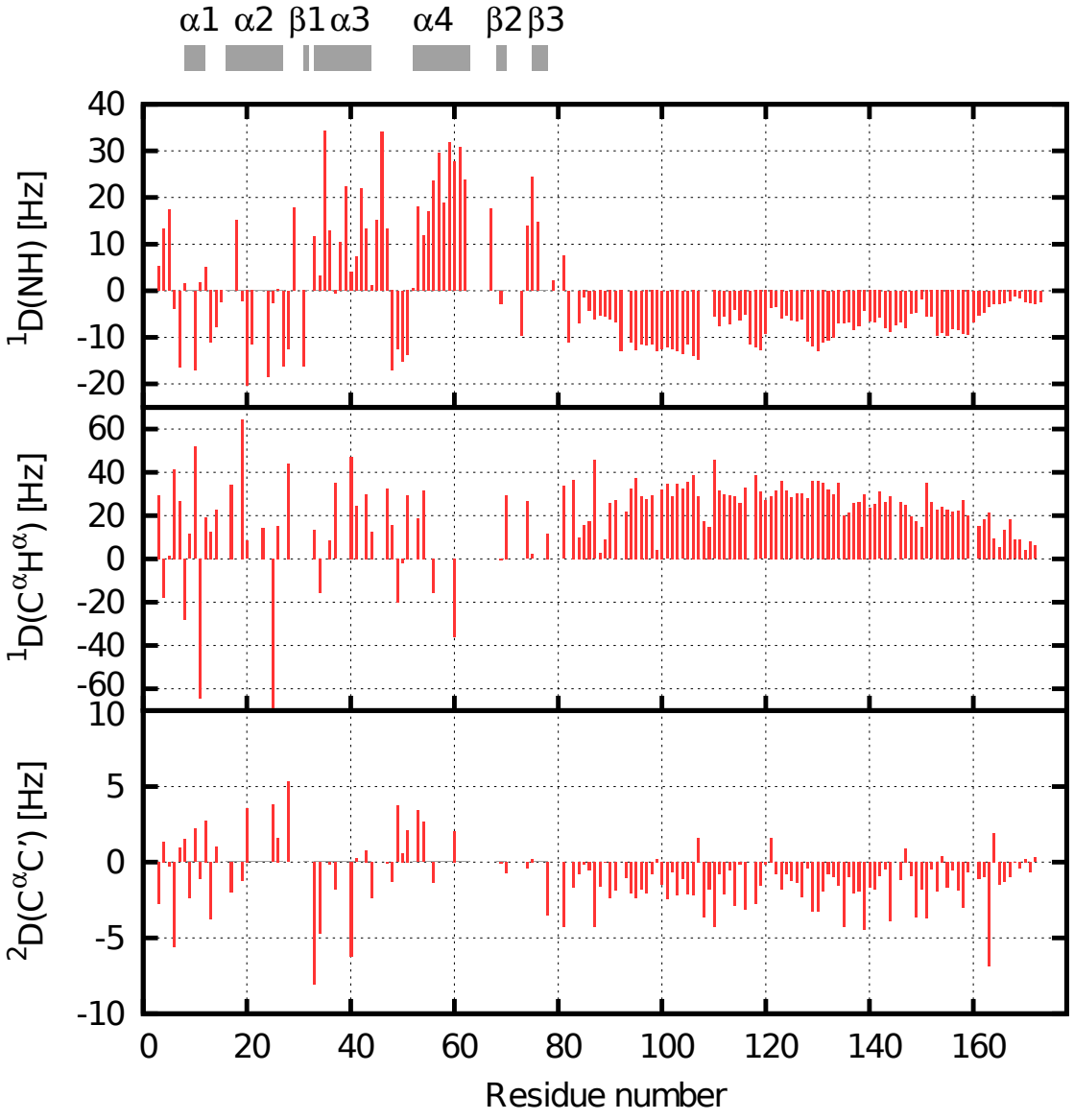

Figure 3.6: The RDC values obtained for the δKE mutant.

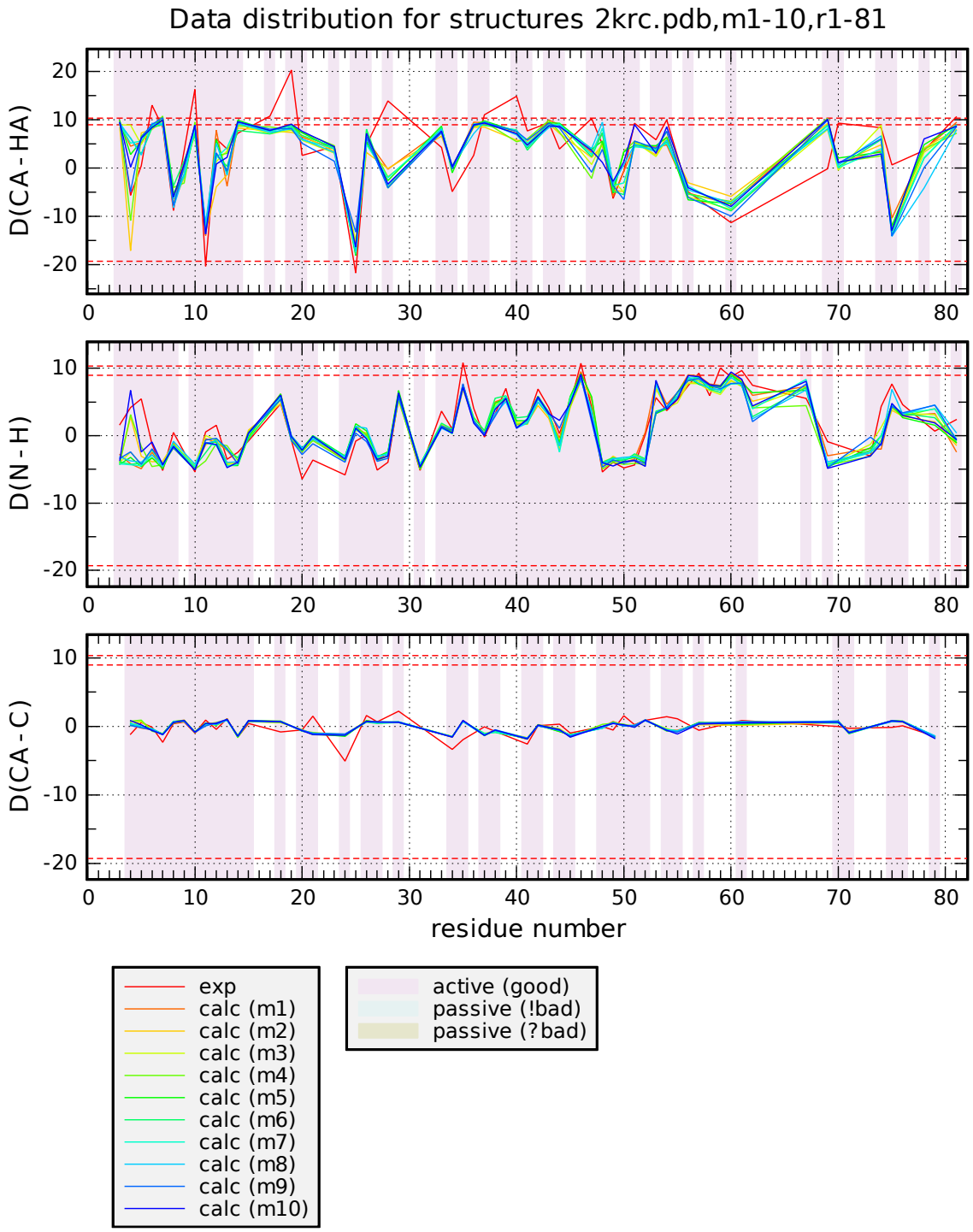

Figure 3.7: The distribution of experimental ${}^1D_{C^\alpha H^\alpha}$, ${}^1D_{NH}$ and ${}^1D_{C^\alpha C'}$ values in comparison to the recalculated data from δWT by the software dconv. The pink rectangles in the background mark the obtained experimental data, whereas the white rectangles represent the missing experimental data.

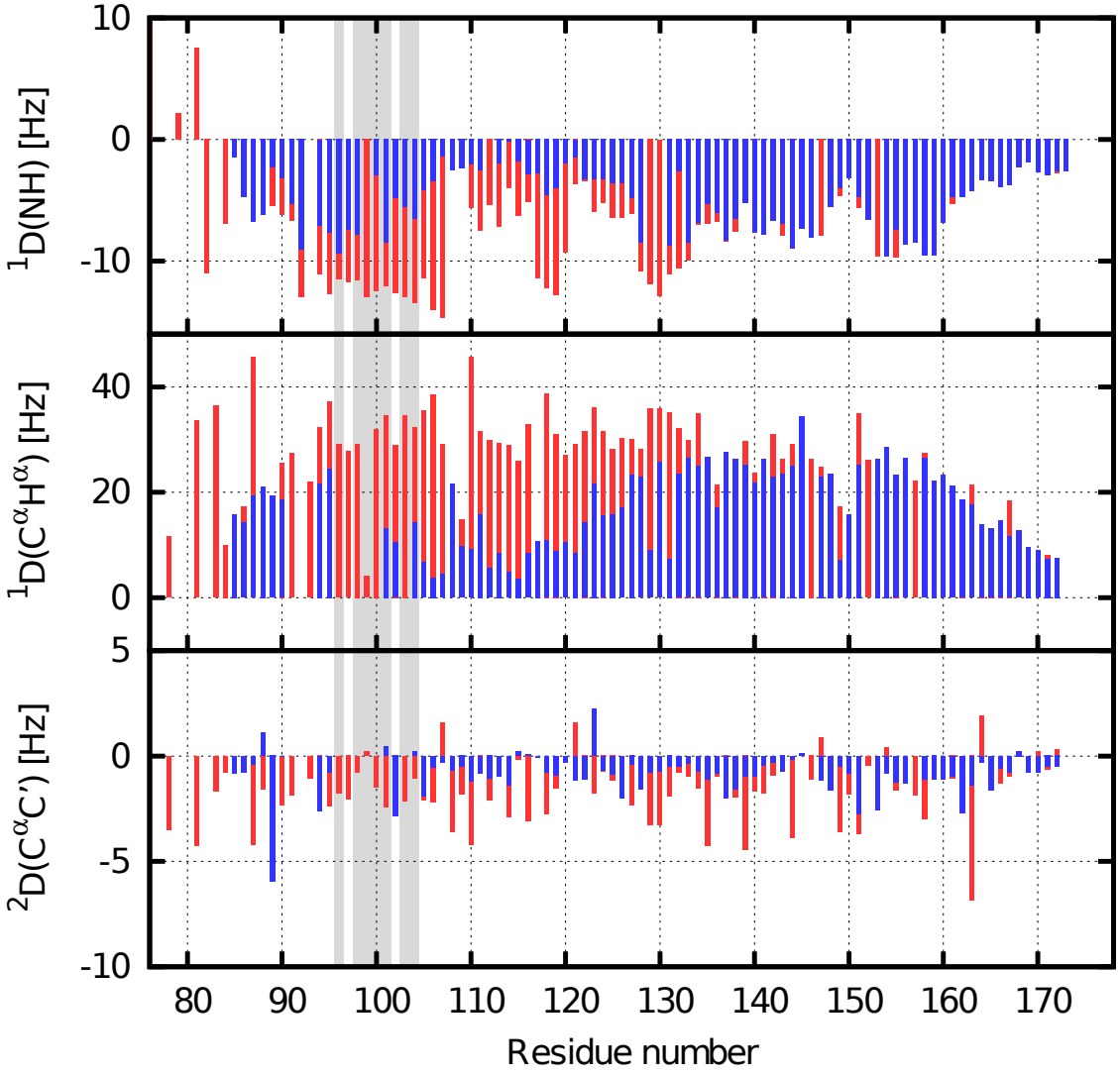

Figure 3.8: The RDC values for CTD of δWT and δKE in blue and red, respectively. Gray rectangles mark the region of mutation.

3.2.2 Paramagnetic relaxation enhancement

The effect of paramagnetic relaxation enhancement is mediated by a direct dipole interaction between an unpaired electron of the spin label and the nucleus spin. The presence of the unpaired electron enhances the transverse relaxation rate R_2, which results in line broadening observed for nuclei in the proximity of the label. The effect is dependent on the distance between the observed nuclei and the electron spin, thus PRE is a highly useful probe for observing transient long-range contacts in macromolecules in the range of 1–2.5 nm [69].

The measurement of PRE in proteins requires attachment of a paramagnetic label. The thiol-reactive methanethiosulfonate spin label (MTSL) proved to be suitable in previous studies of δWT [70, 4]. The label did not influence the structure of the δ subunit and the PRE effect was observed. The previous PRE analysis of CTD revealed both intra and

inter-domain interactions in δWT. The inter-domain contact, detected in helix IV, loop L71–Q74 and the first strand of the beta sheet, indicates a possible influence of the CTD on folding of the well ordered region in the NTD. The other significant interaction was observed between the lysine stretch (T94–V106) and the negatively charged C-terminal tail (E114–D171). This interaction is clearly governed by electrostatic attraction.

The influence of electrostatic interactions on global folding of the δ subunit was examined by monitoring long-range contacts of the δKE mutant. The substitution of the lysine stretch was expected to create highly extended structure without any contacts.

In order to attach the paramagnetic spin label, the site-directed mutation L132C was introduced to the δKE mutant. The label attached to the L132C mutant revealed the most visible contact in the previous measurements [4]. The cystein residue provides a binding site for the paramagnetic (MTSL) and diamagnetic (MTS) label. The label was placed in the negatively charged C-terminal tail, which mediates interaction with the lysine stretch. The dipole-dipole interaction between the label and ^{1}H is most sensitive due to the high magnetogyric constant of proton, thus the ^{1}H-^{15}N HSQC spectrum was recorded for both paramagnetic and diamagnetic sample. The spectrum of δKE-L132C with the diamagnetic label does not contain any information about possible contacts. However, it represents the reference for the spectrum of the paramagnetic sample. Comparison of the the ^{1}H-^{15}N HSQC spectrum of diamagnetic δKE-L132C with ^{1}H-^{15}N HSQC spectrum of δKE confirmed that the structure was not influenced by the attachment of the label. The number of assigned residues, 148, covers 85.5% of the sequence. Some peaks can not be included in the analysis because of a poor resolution, resulting in overlaps and increased peak intensities. An example is presented in Figure 3.9, where the intensity of the peak D166 is strongly influenced due to the overlap with D140. Although the 2D HSQC experiment does not provide sufficient resolution to assign all resonance frequencies of the flexible CTD, it was sufficient for obtaining an overall picture of intramolecular contacts in the δKE mutant.

The PRE effect of the paramagnetic moiety in the molecule was measured by comparing peak intensities (I_{para}/I_{dia}) between the spectra in the presence and in the absence of the paramagnetic spin label. The ratio I_{para}/I_{dia} for δKE is plotted in Figure 3.10. Also the PRE values calculated from ASTEROIDS for the δKE-L132C mutant are plotted in Figure 3.10 (green line), the ASTEROIDS analysis was performed by Vojtěch Kubáň. The residues neighbouring C132 in the sequence are significantly affected by the PRE effect: 10–15 residues up and downstream exhibit line broadening. The deviation of the I_{para}/I_{dia} ratio is decreasing with increasing distance between the given residue and the spin label. The peak heights of residues far enough from the label are apparently unaffected and show uniform values. This trend is typical for a random coil behaviour without any transient contact. The values predicted from ASTEROIDS for the δKE-L132C mutant show the typical pattern of a random coil without any contact. Potential long-range contacts are revealed by a decreased peak intensity ratio of residues distant far in the sequence from the spin label.

The obtained PRE values for the δKE mutant show that either the NTD region or CTD do not exhibit any deviations of peak intensities besides the region of the spin label, except the region of A70. The PRE data for δWT and δKE are plotted in Figure 3.11. Whereas δWT shows a clear pattern of decreased peak intensities in the lysine stretch (K96–K105), δKE peak intensities in the region of mutation (E96–E105) are unaffected by the presence

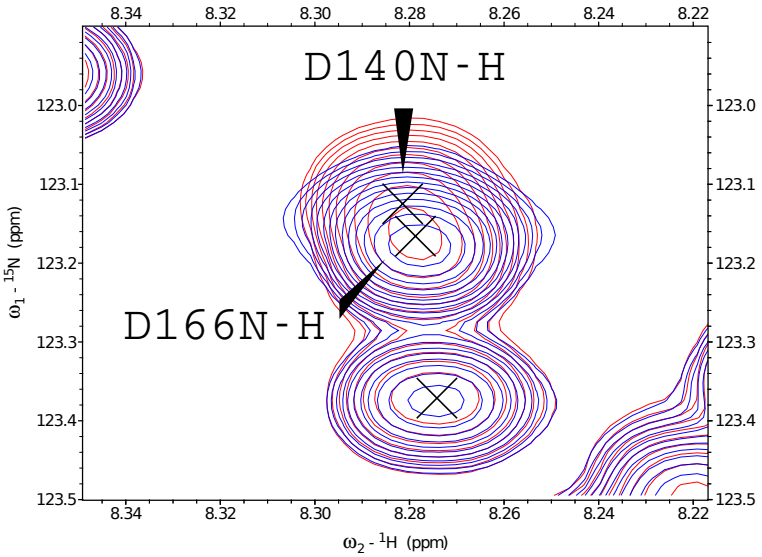

Figure 3.9: Insufficiently resolved peak of D166 in the ^{1}H-^{15}N HSQC spectrum. Two spectra are overlaid in this picture, red color represents the spectrum with the diamagnetic label and blue color represents the spectrum with the paramagnetic spin label. The signal of D140 disappeared under the influence of spin label and it decreases the signal of D166 due to insufficient resolution.

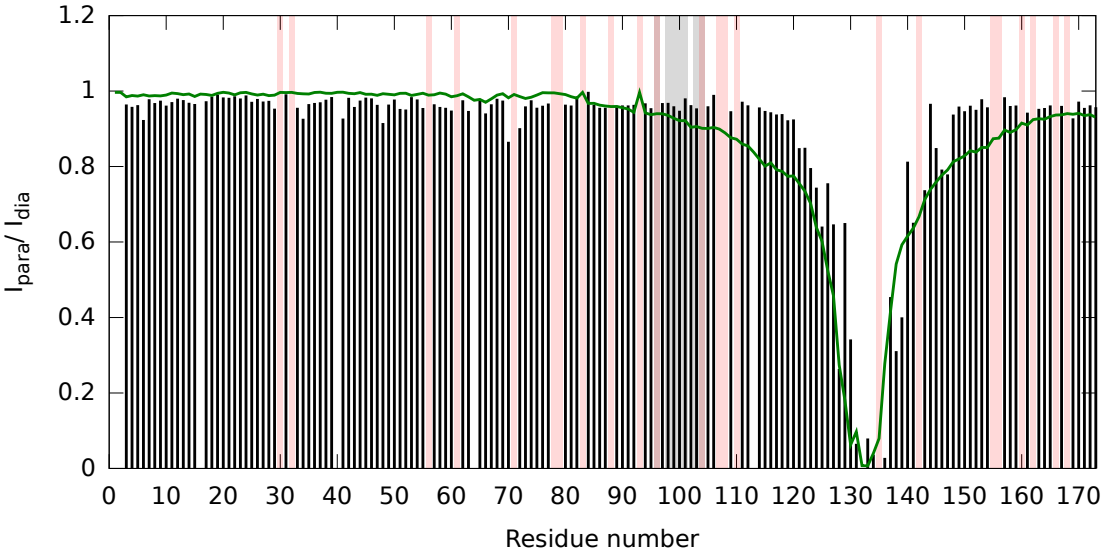

Figure 3.10: Experimental PREs of the δKE-L132C mutant with the MTSL label attached. The green line was calculated using the software ASTEROIDS for the δKE-L132C mutant with the MTSL label attached. The red rectangles marks the missing data, the grey rectangles represents the region of mutation.

of the spin label. Also the interaction of CTD with the structured part is not apparent in δKE, both residues in helix IV (E50–T60) and in the first strand of beta sheet (P30–Q34) show uniform peak intensities. The exception emerges in the region of A70 near the loop in the structured part, the slight decrease in the intensity ratio is present in the δKE mutant.

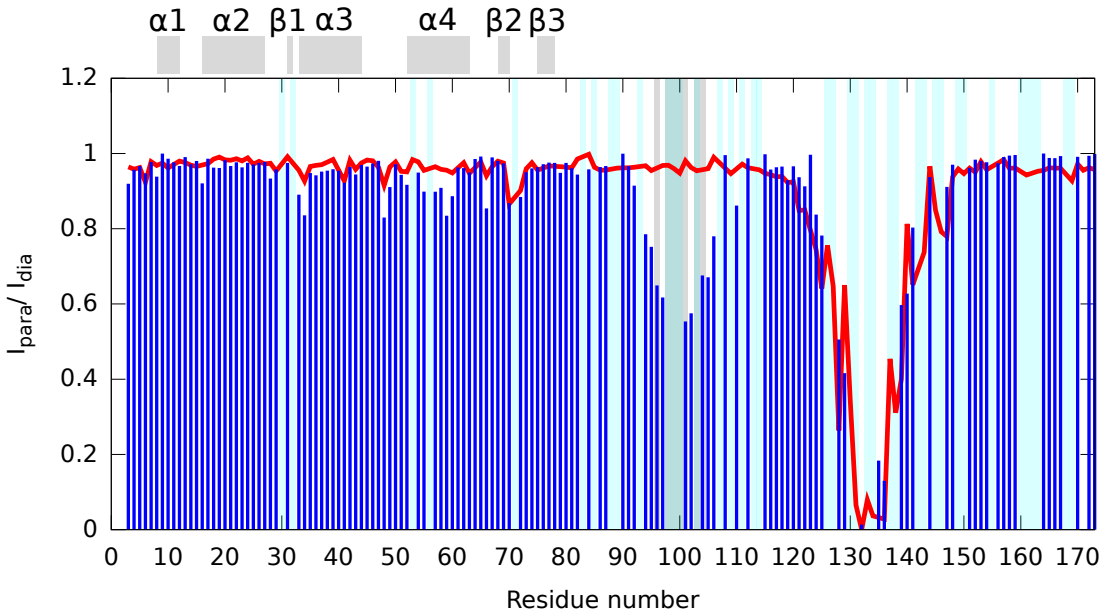

Figure 3.11: Comparison of PREs for δKE and δWT. Blue and red represents PREs of δWT and δKE, respectively. The blue rectangles represent the missing data, the grey rectangles mark the region of mutation in δKE mutant.

The mutation of the lysine stretch affects long-range interactions within CTD. The mutated region (E96–E104) keeps the C-terminal part extended by electrostatic repulsion and the contact of the negative C-terminal tail with the lysine stretch mediated by electrostatic attraction completely disappears. Moreover, the mutation of the lysine stretch does affect not only the electrostatic interaction within CTD, but also the interaction with NTD. The transient contact in helix IV and with the beta sheet is also suppressed by mutation of the lysine stretch, as predicted from the ASTEROIDS analysis. The only interdomain contact in NTD remaining after the disruption of electrostatic interactions was detected in the region of A70.

The changes displayed in NTD distant more than 80 residues from the region of mutation indicate that the electrostatic interactions play a crucial role in global conformational behaviour of the δ subunit. Although CTD is unstructured, the inspection of electrostatic interactions using PRE suggests that the structure of the well-ordered NTD of the δ subunit is influenced by the electrostatic interactions within CTD. The interdomain contact between CTD and the specific regions of NTD, helix IV and beta sheet, is mediated by the electrostatic interaction in CTD. The remaining decrease in the vicinity of A70 might be explained as a contact of a different kind than electrostatic attraction between the lysine stretch and the negatively charged C-terminal tail.

3.3 Dynamics

The studies of dynamic properties are highly useful in order to characterize conformational behaviour of IDPs. The NMR spectroscopy is able to detect regions of increased or decreased mobility with a potential biological function.

The previous investigation of dynamic behaviour of the δ subunit revealed that flexibility of CTD is not uniform along the polypeptide backbone. Regions with decreased flexibility corresponding to the lysine stretch were revealed by interpretation of ^{15}N relaxation experiments, namely steady-state nuclear Overhauser enhancement (ssNOE) and relaxation rates R_1, R_2, Γ_x and Γ_z [4]. The NTD exhibited behaviour of a well-ordered protein with local backbone flexibility in the region of helix IV (E50–T60). In order to investigate the electrostatic interaction in the δ subunit, ^{15}N relaxation experiments including ssNOE, R_1, R_2, Γ_x and Γ_z were measured for the δKE mutant and the resulting values were analysed and compared with previously measured data of δWT [4].

3.3.1 Heteronuclear ^{1}H-^{15}N steady state NOE

Heteronuclear ^{1}H-^{15}N steady state NOE (ssNOE) is used to characterize the N-H bond mobility in the polypeptide backbone. The dipolar relaxation mechanism causes the modulation of peak intensities of coupled nuclei. The ^{1}H-^{15}N steady state NOE measurement includes recording of two spectra: saturated one and reference one. The long selective irradiation of ^{1}H nuclei is applied in order to saturate the spin population difference between α ans β states. During the mixing time, ^{1}H nuclei return to the equilibrium and through longitudinal relaxation mechanism transfer energy to coupled ^{15}N nuclei, perturbing their equilibria [56]. The unsaturated spectrum is used as reference. The values of NOEs are obtained as ratios of saturated and unsaturated peak intensities (I_{ss}/I_{ref}). The typical NOE values for rigid residues in a molecule are positive, reaching 0.5–0.8 at 600 MHz. The fast internal motions result in a decrease in the peak intensity ratio and in negative NOE values.

The 3D ssNOE HNCO spectra with non-uniform sampling were measured on a double labelled ^{13}C, ^{15}N protein sample of δKE mutant. The increased resolution enabled the assignment of 162 residues, the assignment covered 93.6% of the sequence. The careful measurement of peak intensities is crucial for the precise analysis [71]. First, the peak list with exact resonance frequencies of centered peaks was created in the reference spectrum. The peak list of the reference spectrum was loaded into the saturated spectrum and peak heights for each spectrum were measured using fixed frequencies. The ratio of peak heights for saturated and reference spectra is the value of ssNOE. The resulting ssNOEs of δKE are plotted in Figure 3.12. The NTD of the δKE mutant exhibited positive values in the range from 0.5 to 0.8. This trend is uniform along the whole polypeptide backbone of the NTD and it is typical for structured proteins. The slight ssNOE decrease is observed in the region F31–L37. The region of F31–L37 corresponds to helix III (residues F33–L44). The lowered ssNOEs in the structured part indicates increased flexibility in the protein backbone.

The trend is significantly different for the CTD. The sudden drop of ssNOE values to -0.4 to -0.6, visible from the residue Y84, indicates a highly flexible region. The ssNOE values are relatively uniform along the peptide backbone of the CTD, except the region

of mutation (E96–E104). This region exhibits slightly lower values than the rest of the CTD, indicating a higher mobility of the polypeptide chain in this region. Only the residues E170–K173 show more negative values, this behaviour is expected for terminal residues of the polypeptide chain.

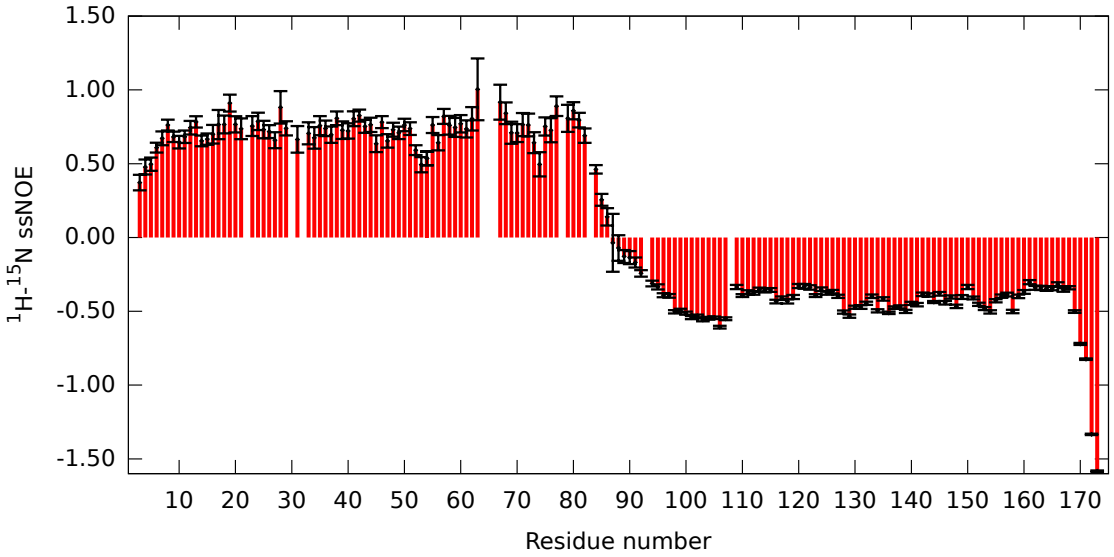

Figure 3.12: The ssNOE values for δKE at 600 MHz.

The comparison of ssNOE values for δWT and δKE is plotted in Figure 3.13. The significant difference between the data reveals a dramatic change in dynamic behaviour of δKE. Although the structured NTD exhibits ssNOE values typical for rigid residues in both δKE and δWT, the residues F31–L37 in the region of helix III show increased flexibility for δKE. Besides this region, the data of the NTD are comparable for δKE and δWT. However, dramatic differences are obtained for the dynamic behaviour of the CTD. Whereas δWT flexibility is not uniform along the CTD backbone and the partial ordering is obvious in the lysine stretch, δKE mutant exhibits uniform flexibility along the whole CTD.

In conclusion, the disruption of electrostatic interactions in CTD strongly influenced the dynamic behaviour of the whole δ subunit. The partial ordering of CTD was completely suppressed and also changes in helix III were observed for δKE mutant. The analysis of ssNOE revealed that the electrostatic interactions in δ subunit governs the conformational flexibility of the CTD and moreover, mildly affects the dynamics of helix III in the structured part.

3.3.2 Transverse and longitudinal auto-relaxation rates

The autorelaxation rates R_1 and R_2 were measured with 9 and 7 relaxation delays, respectively. A distinct spectrum for each delay was recorded (9 spectra for R_1 and 7 spectra for R_2). The peak intensities were measured like in the ssNOE experiment. The peaklist was created in the spectrum with the shortest relaxation delay and then it was used for measurement of peak heights in all spectra. The assignment covered 86% of the protein sequence.

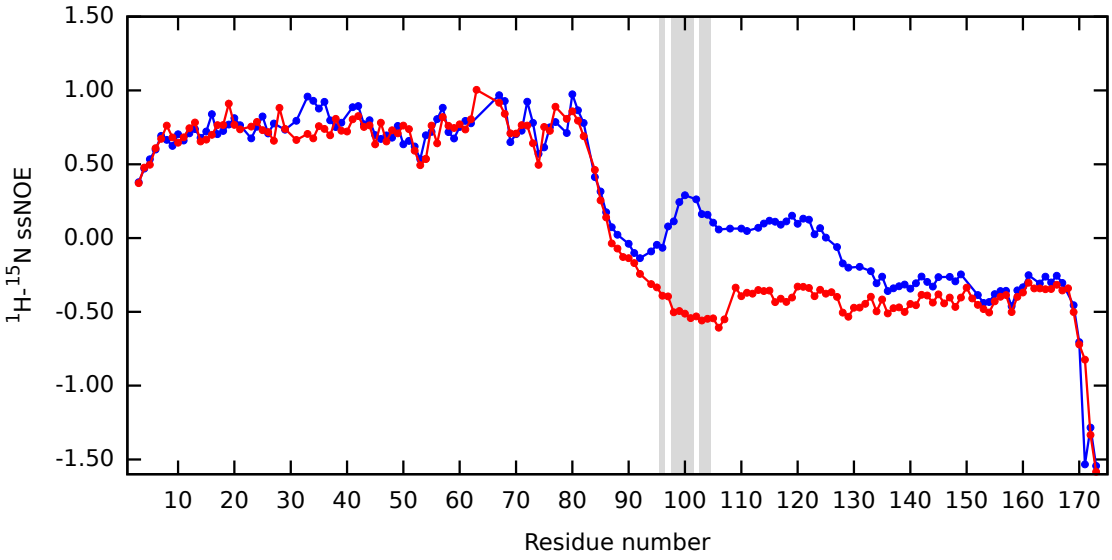

Figure 3.13: The comparison of ssNOE values for δWT and δKE. Blue and red represents PREs of δWT and δKE, respectively. The grey rectangles mark the region of mutation in δKE mutant.

The peak heights for each delay were fitted to a monoexponential decay and relaxation rates R_1 and R_2 were obtained. The resulting autorelaxation rates R_1 and R_2 are plotted in Figures 3.15, 3.14. The errors for structured part are higher due to the experimental set-up targeting the signal of the unstructured domain.

R_2 characterizes the overall tumbling of the structured part and the obtained trend of transverse relaxation rate R_2 resembles the values of ssNOE. The decreased R_2 rates in the regions K47–R54 and A70–Q74 indicate higher flexibility of helix III and the third strand of the beta sheet in the N-terminal part of δKE. The dramatic decrease of R_2 starting at residue D85 reflects dynamic behaviour of the unstructured C-terminal tail of δKE. High values of the R_2, most notable for V17, D61 and W76, indicate presence of the slow exchange.

The longitudinal relaxation rates R_1 are influenced both by internal and global motions of the molecule and depend mostly on the low frequency term ω_N. The R_1 values are not so different between the structured and unstructured parts, because J(ω_N) is in our case similar for different values of the correlation time.The R_1 values also indicate increased flexibility of the region K47–R54 and A70–Q74.

The comparison of relaxation rates of δKE and δWT, plotted in Figures 3.17 and 3.16, shows a significant difference in dynamic behaviour of the CTD for both R_1 an R_2 relaxation rates. The results, indicating the strong influence of electrostatic contacts on motional properties of δ subunit, correlate with the results of the ssNOE analysis.

3.3.3 Transverse and longitudinal cross-correlated relaxation rates

In order to obtain independent data reporting on protein dynamics, transverse cross-correlated relaxation rate Γ_x and longitudinal cross-correlated relaxation rate Γ_z were

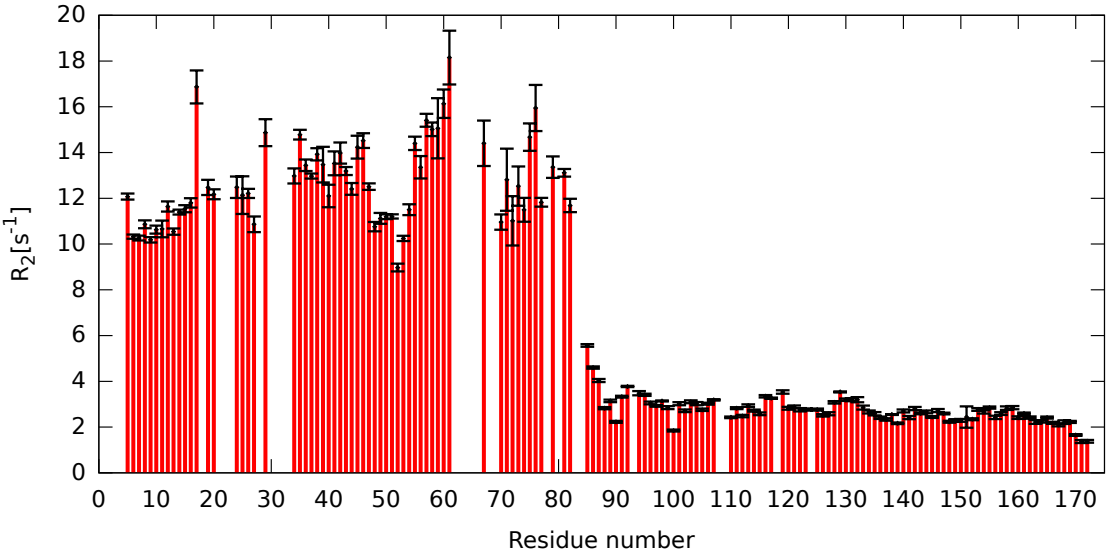

Figure 3.14: The transverse relaxation rates R_2 of δKE mutant.

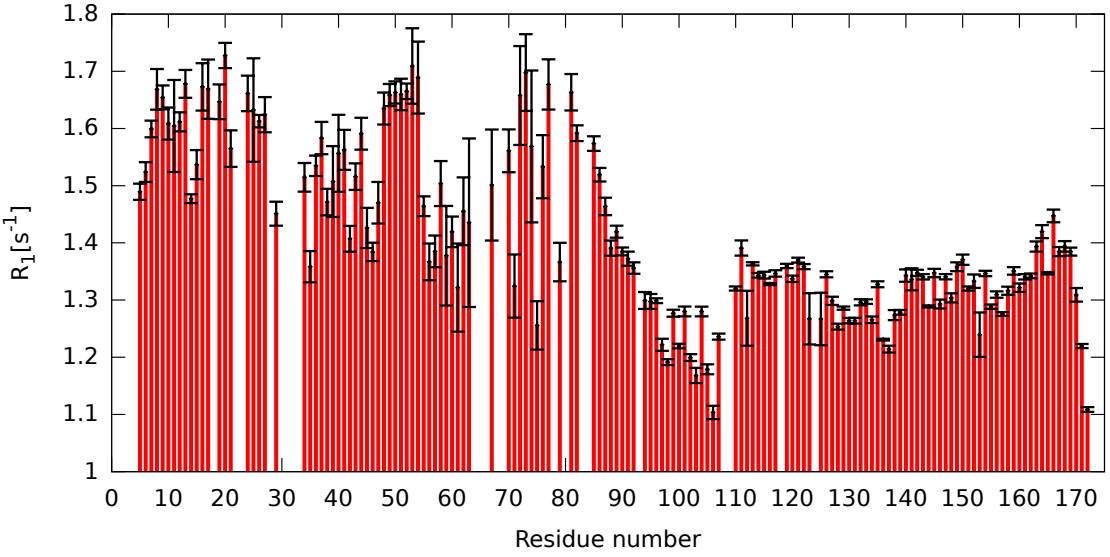

Figure 3.15: The longitudinal relaxation rates R_1 of δKE mutant.

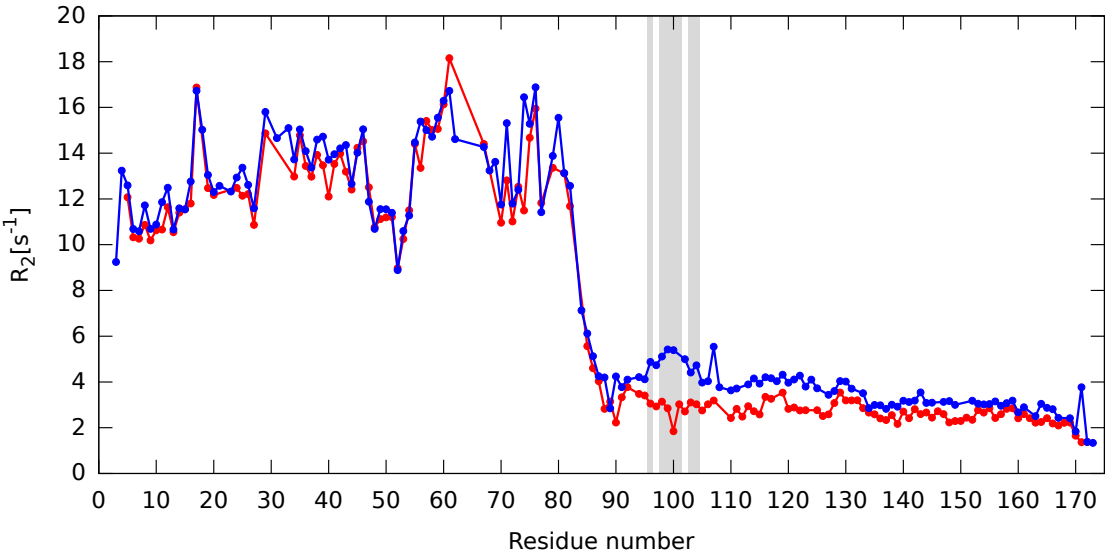

Figure 3.16: The comparison of R_2 rates for δKE and δWT. Blue and red symbols represent the data for δWT and δKE, respectively. The grey rectangles mark the region of mutation in δKE mutant.

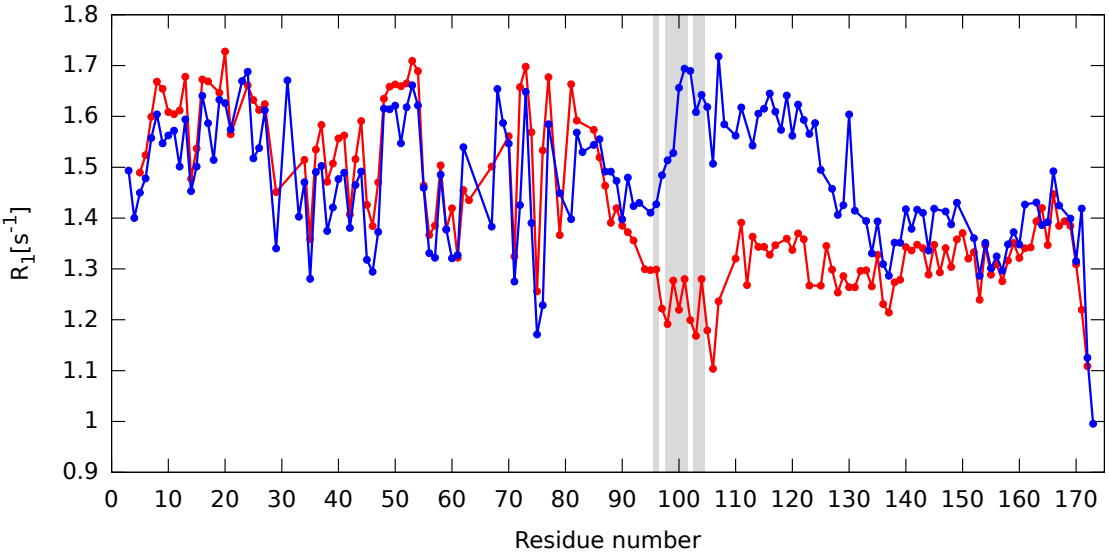

Figure 3.17: The comparison of R_1 rates for δKE and δWT. Blue and red symbols represent the data for δWT and δKE, respectively. The grey rectangles mark the region of mutation in δKE mutant.

measured [72]. The obtained rates Γ_x and Γ_z are plotted in Figures 3.19 and 3.18. The pattern of Γ_x relaxation rates exhibits uniform rigidity in the structured NTD and a sudden drop of rates in the region of Y84. These results correspond to the transverse auto-relaxation rates. Also the trend of the cross-correlated longitudinal relaxation rate Γ_z resembles the longitudinal auto-relaxation rates. The data for the structured part of the cross-correlated longitudinal relaxation rate Γ_z are sparse due to the worsened sensitivity of measurement for the well-folded part of protein. Both cross-correlated relaxation rates show lower values for the flexible part, with a uniform profile along the peptide backbone as observed for the auto-relaxation rates.

In contrast to δWT, the δKE mutant exhibits a uniformly increased flexibility in the C-terminal tail. The comparison of the data for δWT and δKE mutant are plotted in Figures 3.20 and 3.21. The independent data from cross-correlated relaxation rates are in agreement with auto-relaxation rates and also confirm the influence of the electrostatic interactions on the dynamic behaviour of the δ subunit. The regions with increased rigidity between K96 and E130 of the CTD completely disappears with disruption of electrostatic interactions. It confirms that these interactions are crucial in for the proper dynamic behaviour of δ subunit.

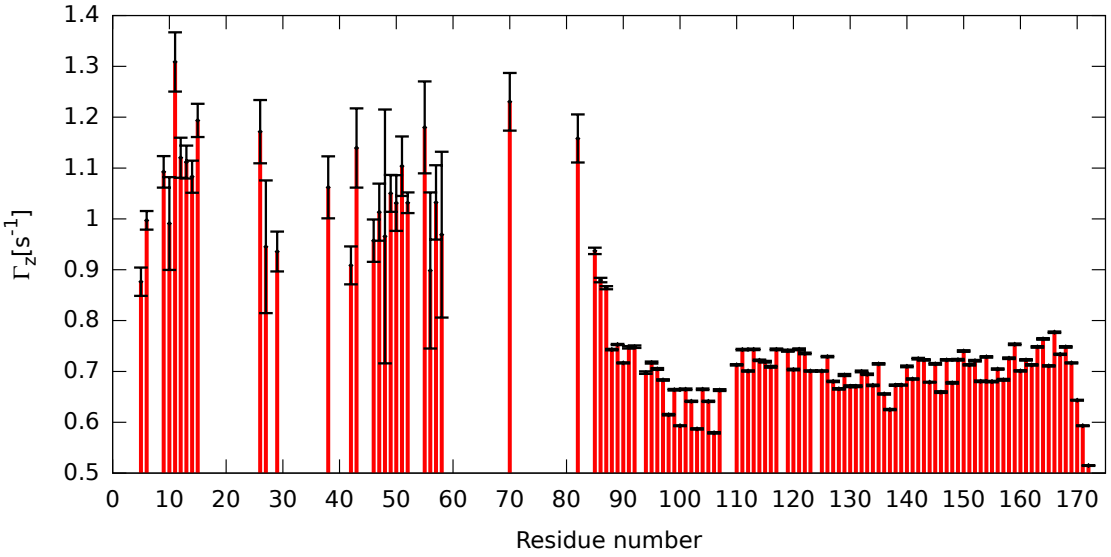

Figure 3.18: The Γ_z cross-relaxation rates for the δKE mutant.

Although the trends of the relaxation rates ssNOE, R_1, R_2, Γ_z and Γ_x report on dynamic properties of the δ subunit and provide useful information about the flexibility of the backbone, the exact description of the contributions of different molecular motions is not apparent from this kind of analysis. Therefore the obtained relaxation rates were analysed by the spectral density mapping. The measurement of cross-correlated rates are also highly useful in order to avoid a bias by the slow exchange in this analysis.

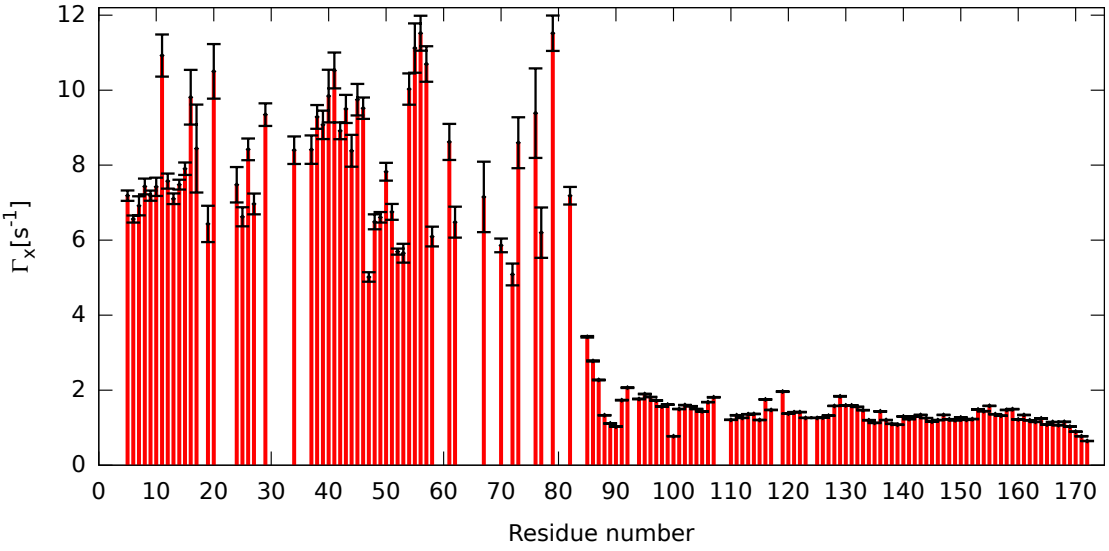

Figure 3.19: The Γ_x cross-relaxation rates for the δKE mutant.

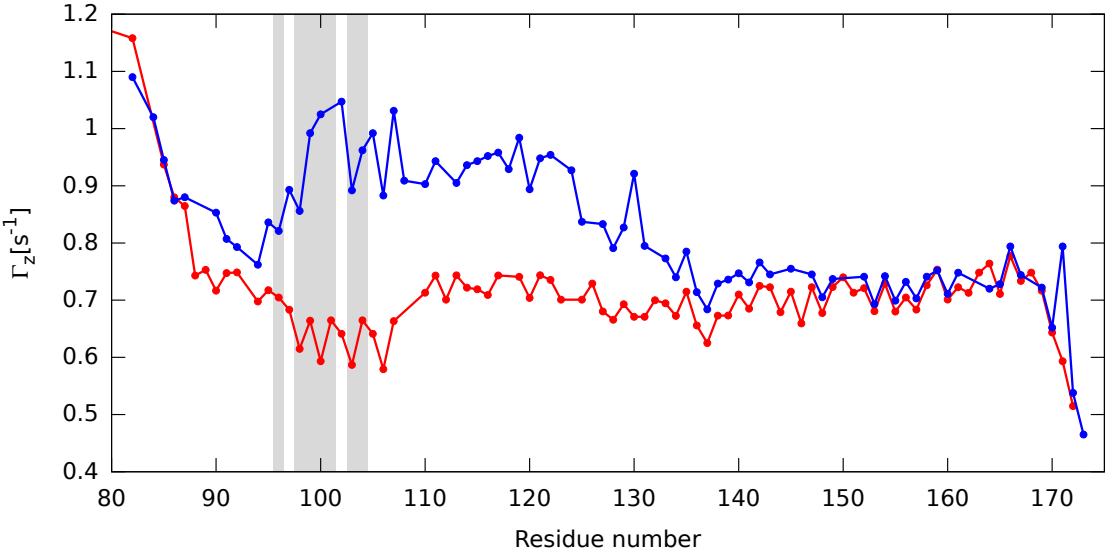

Figure 3.20: The comparison of Γ_z cross-relaxation rates between δWT and δKE mutant in blue and red, respectively.

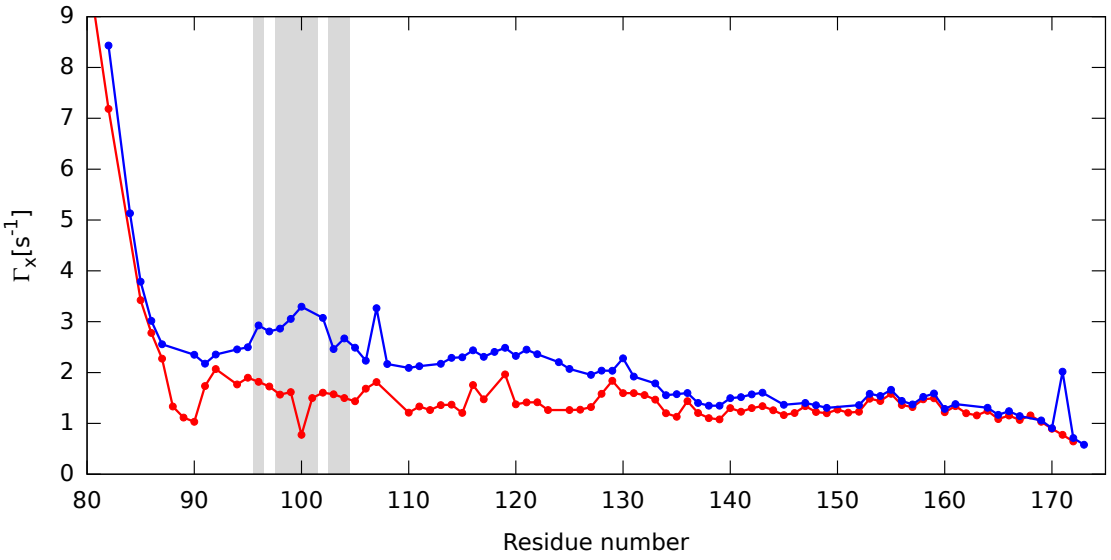

Figure 3.21: The comparison of Γ_x cross-relaxation rates between δWT and δKE mutant in blue and red, respectively.

3.3.4 Spectral density function mapping

The auto-relaxation rates R_1, R_2, ssNOE and cross-correlated relaxation rates Γ_x, Γ_z were analysed by the reduced spectral density function mapping in order to characterize the molecular motions of CTD of the δKE mutant. The analysis of spectral density functions provides information on the time scale of motions governing the dynamics of a given residue. The relaxation rates are linear combinations of five values of the spectral density function $J(\omega)$ at distinct frequencies and the spectral density function values are obtained by solving a linear combination of five equations. The reduced spectral density function mapping approach assumes similarity of the high-frequency terms due to the high value of γ_{H}. Therefore the set of linear combinations is reduced to three, depending on $J(0)$, $J(\omega_N)$ and a single high-frequency term $J(\varepsilon\omega_H)$. This approach requires measurement of only three or less relaxation rates [57, 73].

The values of the spectral density function were obtained from different combinations of the relaxation rates using the protocol developed in our group by Pavel Kadeřávek [58]. The distinct kinds of analysis are described according to the types of used relaxation rates, abbreviated L for longitudinal auto-relaxation rate R_1, T for transversal auto-relaxation rate R_2, N for the heteronuclear ssNOE, X for the transverse cross-correlated relaxation rate Γ_x and Z for the longitudinal cross-correlated relaxation rate Γ_z. The resulting spectral density function values for $J(0)$, $J(\omega_{\mathrm{N}})$ and $J(\omega_{\mathrm{H}})$ calculated from different relaxation rate combinations are plotted in Figures 3.22, 3.23 and 3.24.

The obtained spectral density values are plotted in the Lefèvre's plot as a function of calculated $J(0)$ in Figure 3.25. The Lefèvre's plot is a two dimensional representation of the correlation between distinct spectral density values with the displayed Lorentzian curve, defining the motional limit. This graphical representation facilitates the interpretation the of obtained data.

Four different combinations of relaxation rates were analysed for CTD of the δKE

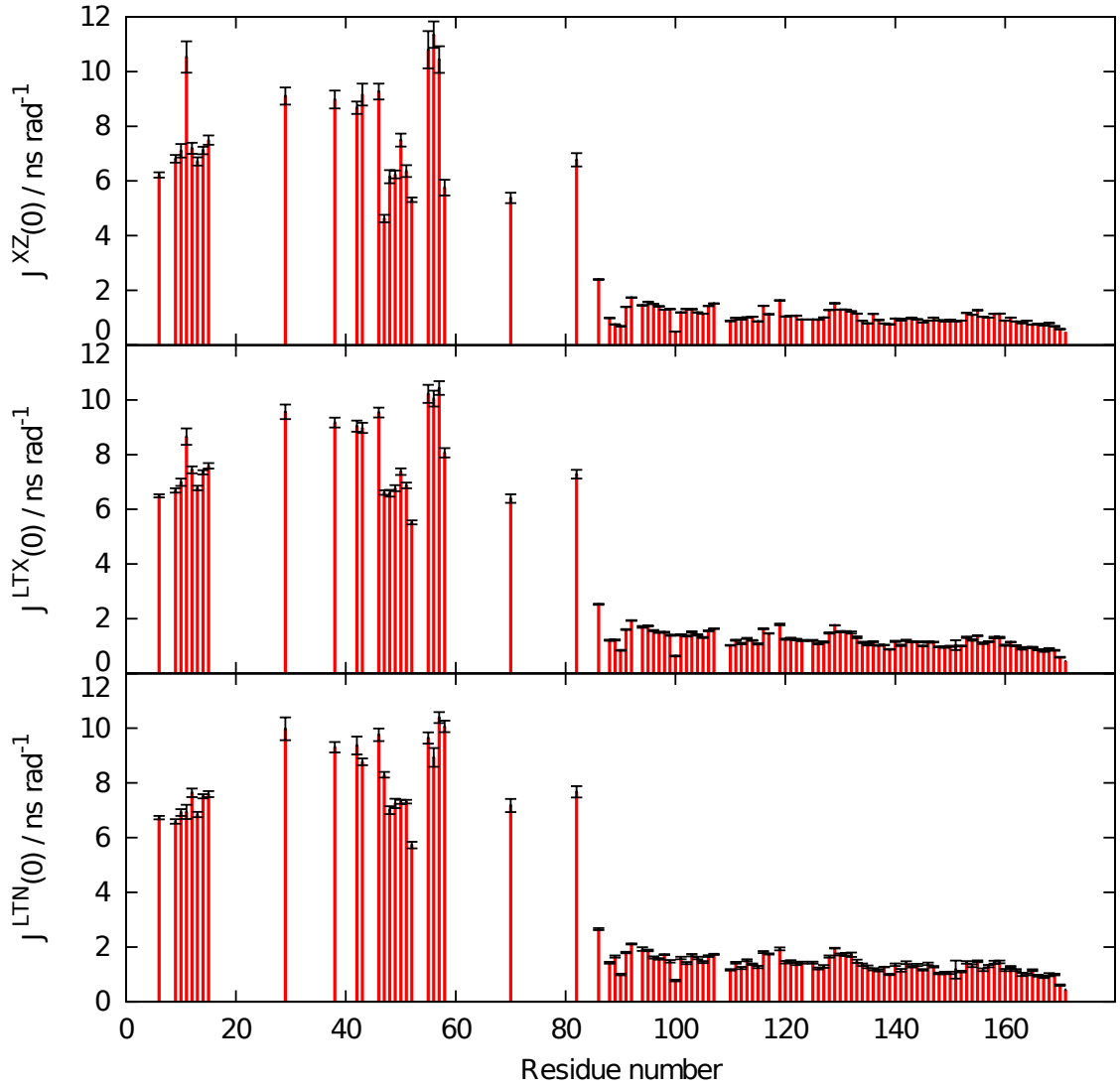

Figure 3.22: The spectral density function values $J(0)$ obtained from relaxation rates Γ_x and Γ_z (XZ), from relaxation rates R_1, R_2 and Γ_x (LTX) and from relaxation rates R_1, R_2, ssNOE (LTN).

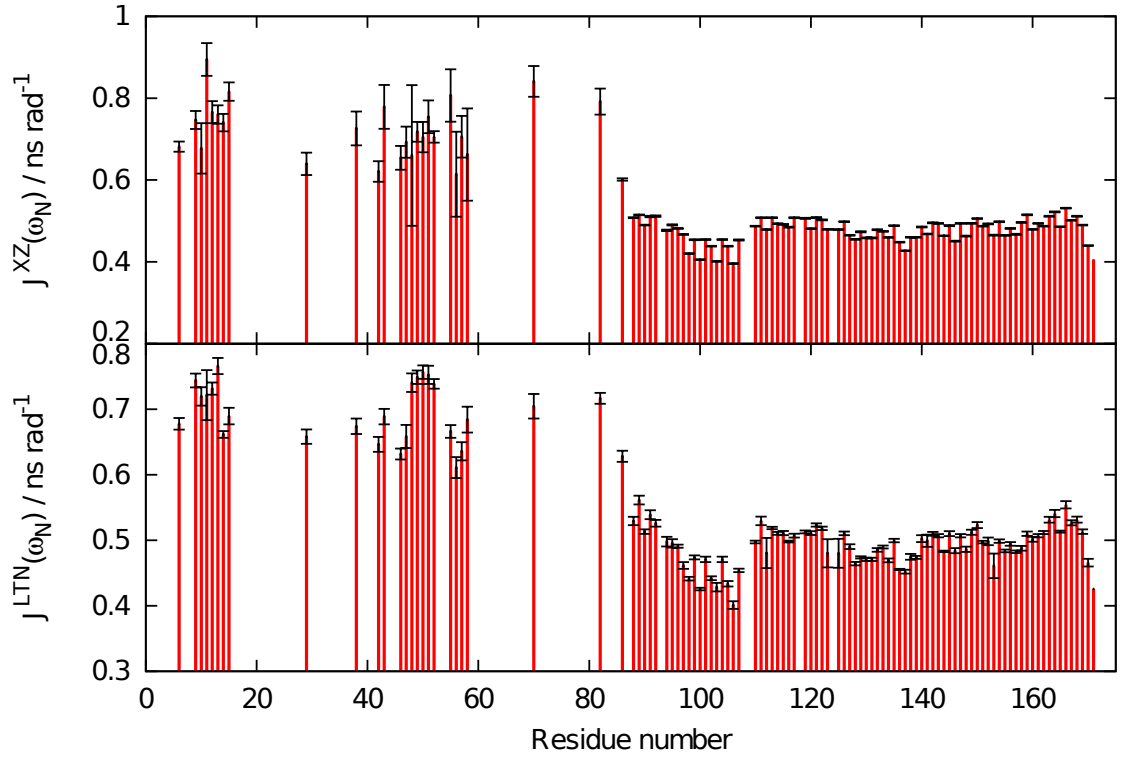

Figure 3.23: The spectral density function values $J(\omega_N)$ obtained from relaxation rates Γ_x and Γ_z (XZ) and from relaxations rates R_1, R_2, ssNOE (LTN).

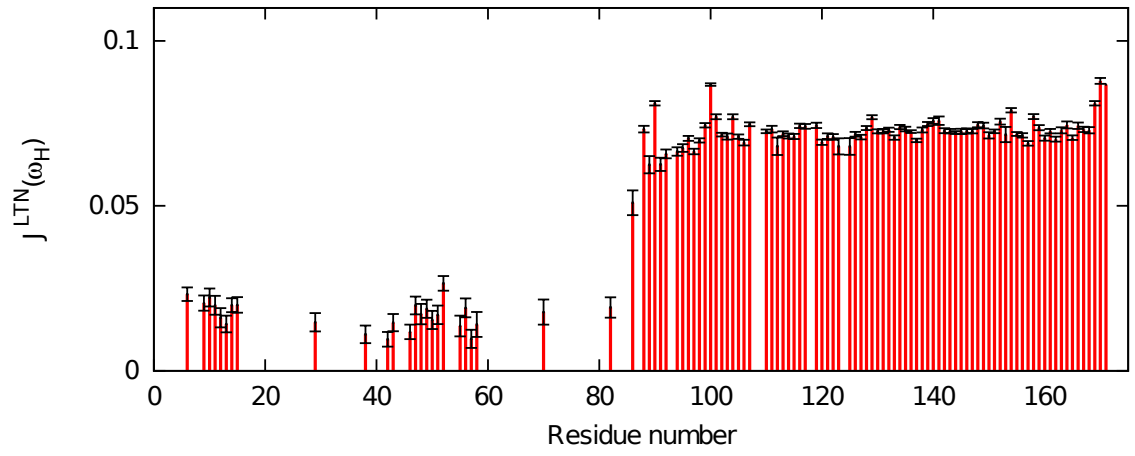

Figure 3.24: The spectral density function values $J(\omega_H)$ obtained from relaxation rates R_1, R_2, ssNOE (LTN).

mutant. All combinations of the obtained spectral density values are plotted in Figure 3.25. The data are color-coded according to the residue number as displayed above the plots in the sequence for clarity. The spectral density values calculated from routinely measured R_1, R_2 and ssNOE values are plotted in Figure 3.25a, d. The $J(0)$ values in Figure 3.25a are low and they are distributed in one cluster close to the Lorentzian curve, signalizing that the residues are mostly influenced by the fast internal motions on the time scale of picoseconds to nanoseconds. The dependence ofthe high-frequency $\mathrm{J}^{LN}(\varepsilon\omega_H)$ on $J^{LTN}(0)$ is plotted in Figure 3.25d and it is useful for the recognition of the slow motion contribution. The LTN analysis is sensitive to the slow exchange due to including the transverse relaxation rate R_2. The R_2 rates include the slow-exchange contribution, which increases the spectral density values and overestimate the contribution of overall rotation diffusion in molecular motion. Consequently, this analysis is biased by the presence of the slow conformational or chemical exchange. Therefore, additional analyses not including the R_2 rates, were preformed.

The exchange free analyses XZ and LNX employing the cross-correlation relaxation rates instead of R_2 were performed in order to avoid the potential exchange bias. Results of the analysis of spectral density values containing the cross-correlated relaxations only are plotted in Figure 3.25c. The analysis of cross-correlated relaxation rates is the most straightforward and it is not influenced by any bias. The pattern of the XZ analysis shows one cluster with low spectral density values, indicating prevalent influence of the fast internal motion on the molecular motions of the protein backbone. Also the results of the LNX analysis combining the auto-relaxation and cross-correlated rates, displayed in Figure 3.25b, reveal a dominant contribution of the internal motion in the residues of CTD. The corresponding results of the independent LNX and XZ spectral density function mapping confirm the interpretation reliability of the data. LNX spectral density values exhibit low $J(0)$ values, indicating fast internal motions on the microseconds to milliseconds and the flexibility along the δKE backbone.

Results of the LNX and XZ analyses are compared with the data from the δWT in Figure 3.26. The data for CTD of δWT are divided into two clusters. C-terminal residues E135–K173 exhibit low values indicating the high flexibility and fast internal motions, whereas the residues D85–E129 are clustered at higher spectral density values and show apparently slower molecular motion. The division into two clusters completely disappears in the δKE mutant, exhibiting uniformly lowered spectral density values. The spectral density function mapping of δKE and its comparison with δWT shows that the electrostatic interactions of the CTD restrict the fast internal motions in the vicinity of the lysine stretch. Disruption of these interactions leads to the higher flexibility of the whole backbone of the CTD.

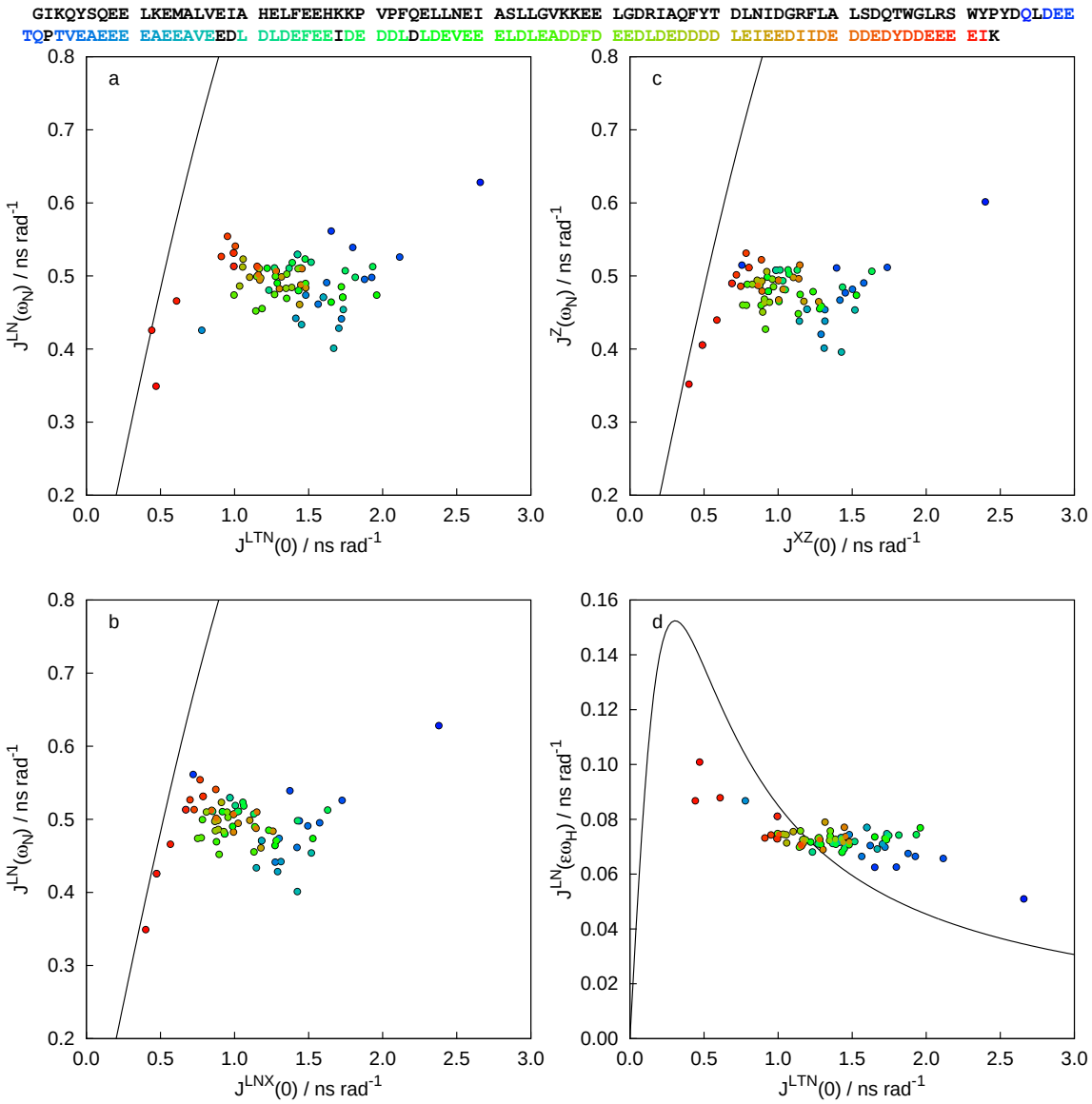

Figure 3.25: Lefèvre's plots for $J^{LTN}(0)$ vs. $J^{LN}(\omega_N)$ (a), $J^{LNX}(0)$ vs. $J^{LN}(\omega_N)$ (b), $J^{XZ}(0)$ vs. $J^{Z}(\omega_N)$ (c) and $J^{LTN}(0)$ vs. $J^{LN}(\varepsilon\omega_H)$ for the unstructured part of the δKE mutant. The plotted values are color-coded according to the position of the given residue in the sequence, displayed above the plots.

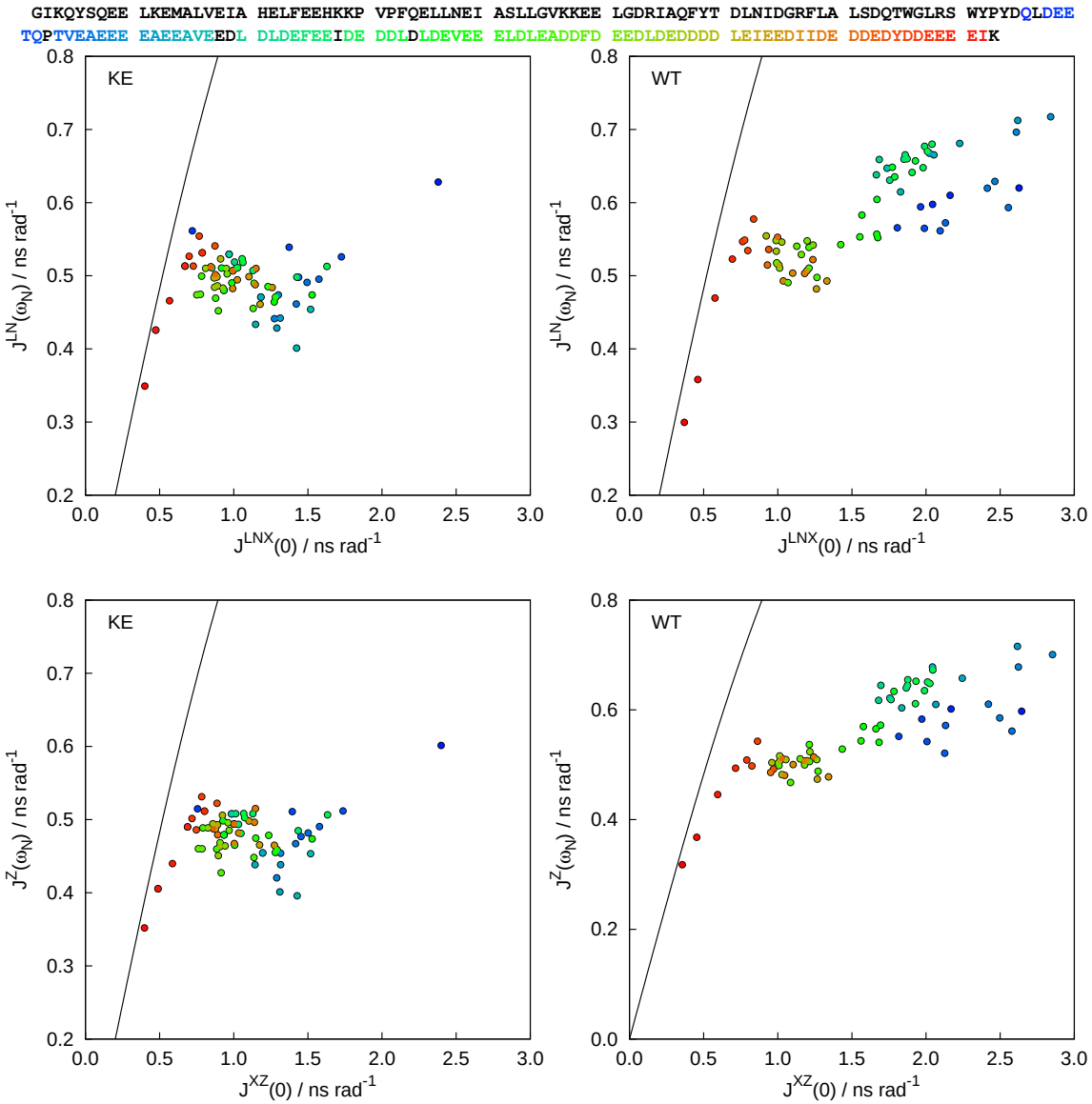

Figure 3.26: The comparison of the Lefèvre's plot $J^{LNX}(0)$ vs. $J^{LN}(\omega_N)$ for the CTD between the δKE and δWT in the upper row and $J^{XZ}(0)$ vs. $J^{Z}(\omega_N)$ in the lower row.

Conclusion

The aim of the thesis was to describe the structural role of the electrostatic interactions in the flexible C-terminal domain of the δ subunit from the *Bacillus subtilis*. These interactions are observed between the positively charged lysine stretch and negatively charged C-terminal tail. In order to obtain the influence of the electrostatic interactions on the global shape of the δ subunit, the electrostatic interaction were disrupted by mutation of the lysine stretch for the negatively charged glutamate residues. The changes in the conformational behaviour of the δKE mutant were monitored by various structural features obtained from NMR experiments.

The comparison of the δKE mutant with the δ wild type revealed significant differences in the structural properties of the δ subunit caused by the disruption of the electrostatic interactions. The increased secondary structure propensity to form the beta sheet was detected for the δKE mutant as a result of electrostatic repulsion of negatively charged glutamate residues in the C-terminal domain. The intradomain contacts in the δ subunit, inspected by the paramagnetic relaxation enhancement, disappeared with the disruption of the electrostatic interactions in the C-terminal domain. The interdomain contacts between CTD and helix IV and beta sheet were also suppressed. The orientational restraints of the polypeptide chain in CTD diminished in case of the δKE mutant. The ^{15}N relaxation measurements of the δKE mutant revealed uniform flexibility along the intrinsically disordered region in the C-terminal domain.

The obtained results show that the balance between the attractive and repulsive electrostatic interactions govern the conformational behaviour of the flexible CTD. Electrostatic interactions restrict the flexibility of the intrinsically disordered CTD and moreover, influence also the well-folded NTD. The interdomain contacts between CTD and specific regions in NTD, namely the beta sheet and helix IV, are mediated by the electrostatic interactions. The dynamics of helix III is restricted by these interactions. The electrostatic interactions in the intrinsically disordered region of the δ subunit are important for the proper structural behaviour of the whole protein.

The importance of the electrostatic interactions is relevant also for the biological function of the δ subunit. The results of the biological experiments, carried out in the collaborating group in the Laboratory of Microbial Genetics and Gene Expression at the Czech Academy of Sciences, indicates that the lysine stretch is vital for proper function of RNA polymerase and for the fitness of the bacteria.

Bibliography

[1] Yugong Cheng, Tanguy LeGall, Christopher J. Oldfield, James P. Mueller, Ya-Yue J. Van, Pedro Romero, Marc S. Cortese, Vladimir N. Uversky, and Keith A. Dunker. Rational drug design via intrinsically disordered protein. *Trends Biotechnol.*, 24(10):435–442, 2006.

[2] Alžbeta Rabatinová, Hana Šanderová, Jitka Jirát Matějčková, Jana Korelusová, Luděk Sojka, Ivan Barvík, Veronika Papoušková, Vladimír Sklenář, Lukáš Žídek, and Libor Krásnỳ. The δ subunit of RNA polymerase is required for rapid changes in gene expression and competitive fitness of the cell. *J. Bacteriol.*, 195(11):2603–2611, 2013.

[3] Veronika Motáčková, Jiří Nováček, Anna Zawadzka-Kazimierczuk, Krzysztof Kazimierczuk, Lukáš Žídek, Hana Šanderová, Libor Krásnỳ, Wiktor Koźmiński, and Vladimír Sklenář. Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. *J. Biomol. NMR*, 48(3):169–177, 2010.

[4] Veronika Papoušková, Pavel Kadeřávek, Olga Otrusinová, Alžbeta Rabatinová, Hana Šanderová, Jiří Nováček, Libor Krásnỳ, Vladimír Sklenář, and Lukáš Žídek. Structural study of the partially disordered full-length δ subunit of RNA polymerase from Bacillus subtilis. *Chembiochem*, 14(14):1772–1779, 2013.

[5] Finn Werner. A nexus for gene expression—molecular mechanisms of spt5 and nusG in the three domains of life. *J. Mol. Biol.*, 417(1):13–27, 2012.

[6] Makoto Kimura and Akira Ishihama. Functional map of the α subunit of Escherichia coli RNA polymerase: Insertion analysis of the amino-terminal assembly domain. *J. Mol. Biol.*, 248(4):756 – 767, 1995.

[7] Leonid Minakhin, Sechal Bhagat, Adrian Brunning, Elizabeth A. Campbell, Seth A. Darst, Richard H. Ebright, and Konstantin Severinov. Bacterial RNA polymerase subunit ω and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc. Natl. Acad. Sci.*, 98(3):892–897, 2001.

[8] Ruth M. Saecker and Thomas M. Record. Mechanism of bacterial transcription initiation: RNA polymerase-promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.*, 412(5):754–771, 2011.

[9] Patrick Cramer. Multisubunit RNA polymerases. *Curr. Opin. Struct. Biol.*, 12(1):89 – 97, 2002.

[10] Francisco J. López de Saro, A-Young Moon Woody, and John D. Helmann. Structural analysis of the Bacillus subtilis δ factor: A protein polyanion which displaces RNA from RNA polymerase. *J. Mol. Biol.*, 252(2):189–202, 1995.

[11] Geoff P. Doherty, Mark J. Fogg, Anthony J. Wilkinson, and Peter J. Lewis. Small subunits of RNA polymerase: localization, levels and implications for core enzyme composition. *Microbiology*, 156(12):3532–3543, 2010.

[12] Andy Weiss, Antonio J. Ibarra, Jessica Paoletti, Ronan K. Carroll, and Lindsey N. Shaw. The δ subunit of RNA polymerase guides promoter selectivity and virulence in Staphylococcus aureus. *Infect. Immun.*

[13] Gabriel Demo, Veronika Papoušková, Jan Komárek, Pavel Kadeřávek, Olga Otrusinová, Pavel Srb, Alžbeta Rabatinová, Libor Krásnỳ, Lukáš Žídek, Vladimír Sklenář, et al. X-ray vs. NMR structure of N-terminal domain of δ-subunit of RNA polymerase. *J. Struct. Biol.*, 187(2):174–186, 2014.

[14] Veronika Motáčková, Hana Šanderová, Lukáš Žídek, Jiří Nováček, Petr Padrta, Alžběta Švenková, Jana Korelusová, Jiří Jonák, Libor Krásnỳ, and Vladimír Sklenář. Solution structure of the N-terminal domain of Bacillus subtilis δ subunit of RNA polymerase and its classification based on structural homologs. *Proteins: Struct., Funct., Bioinf.*, 78(7):1807–1810, 2010.

[15] L. Aravind and Lakshminarayan M. Iyer. The HARE-HTH and associated domains: novel modules in the coordination of epigenetic DNA and protein modifications. *Cell cycle*, 11(1):119–131, 2012.

[16] Andy Weiss and Lindsey N. Shaw. Small things considered: the small accessory subunits of RNA polymerase in Gram-positive bacteria. *FEMS Microbiol. Rev.*, page fuv005, 2015.

[17] Mary Lampe, Craig Binnie, Ruth Schmidt, and Richard Losick. Cloned gene encoding the delta subunit of Bacillus subtilis RNA polymerase. *Gene*, 67(1):13–19, 1988.

[18] Yue-Li Juang and John D. Helmann. The δ subunit of Bacillus subtilis RNA polymerase: an allosteric effector of the initiation and core-recycling phases of transcription. *J. Mol. Biol.*, 239(1):1–14, 1994.

[19] Yue-Li Juang and John D. Helmann. Pathway of promoter melting by Bacillus subtilis RNA polymerase at a stable RNA promoter: Effects of temperature, δ protein, and σ factor mutations. *Biochemistry*, 34(26):8465–8473, 1995.

[20] Amanda L. Jones, Rachel H. Needham, and Craig E. Rubens. The δ subunit of RNA polymerase is required for virulence of Streptococcus agalactiae. *Infect. Immun.*, 71(7):4011–4017, 2003.

[21] Vladimir N. Uversky, Joel R. Gillespie, and Anthony L. Fink. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Struct., Funct., Bioinf.*, 41(3):415–427, 2000.

[22] Jane H. Dyson and Peter E. Wright. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struc. Biol.*, 12(1):54–60, 2002.

[23] Keith A. Dunker, Jeong Oh, Christopher Oldfield, Pedro Romero, Catherine Ratliff, Kerry W. Hipps, Raymond Reeves, Juan Ausio, Charles Kissinger, Robert Bailey, et al. Intrinsically disordered proteins. *J. Mol. Graph. Model.*, 18(4-5):552, 2000.

[24] Keith A. Dunker, Pedro Romero, Zoran Obradovic, Ethan C. Garner, and Celeste J. Brown. Intrinsic protein disorder in complete genomes. *Genome Inform.*, 11:161–171, 2000.

[25] Christopher J. Oldfield, Yugong Cheng, Marc S. Cortese, Celeste J. Brown, Vladimir N. Uversky, and Keith A. Dunker. Comparing and combining predictors of mostly disordered proteins. *Biochemistry*, 44(6):1989–2000, 2005.

[26] Madan M. Babu. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. T.*, 44(5):1185–1200, 2016.

[27] Charles A. Galea, Yuefeng Wang, Sivashankar G. Sivakolundu, and Richard W. Kriwacki. Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits. *Biochemistry*, 47(29):7598–7609, 2008.

[28] Malcolm S. Steinberg, Hisato Shida, George J. Giudice, Mariko Shida, Nipam H. Patel, and Orest W. Blaschuk. On the molecular organization, diversity and functions of desmosomal proteins. *Junctional complexes of epithelial cells*, 1:3, 1987.

[29] Vladimir N. Uversky and Anthony L. Fink. Conformational constraints for amyloid fibrillation: the importance of being unfolded. *BBA-Proteins proteom*, 1698(2):131–153, 2004.

[30] George G. Glenner and Caine W. Wong. Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. *Biochem. Bioph. Res. Co.*, 122(3):1131–1135, 1984.

[31] Stanley B. Pruisner. Shattuck lecture-neurodegenerative diseases and prions. *J. Med.*, 344:1516–1526, 2001.

[32] Lilia M. Iakoucheva, Celeste J. Brown, David J. Lawson, Zoran Obradović, and Keith A. Dunker. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.*, 323(3):573–584, 2002.

[33] Li Ou, Antonio M. Ferreira, Steve Otieno, Limin Xiao, Donald Bashford, and Richard W. Kriwacki. Incomplete folding upon binding mediates Cdk4/cyclin D complex activation by tyrosine phosphorylation of inhibitor p27 protein. *J. Biol. Chem.*, 286(34):30142–30151, 2011.

[34] Benjamin Schuler and William A. Eaton. Protein folding studied by single-molecule FRET. *Current opinion in structural biology*, 18(1):16–26, 2008.

[35] Ian S. Millett, Sebastian Doniach, and Kevin W. Plaxco. Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. *Adv. Protein Chem.*, 62:241–262, 2002.

[36] Huimin Chen and Elizabeth Rhoades. Fluorescence characterization of denatured proteins. *Curr. Opin. Struc. Biol.*, 18(4):516–524, 2008.

[37] Jane H. Dyson and Peter E. Wright. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.*, 104(8):3607–3622, 2004.

[38] David Eliezer. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struc. Biol.*, 19(1):23–30, 2009.

[39] Joseph A. Marsh, Vinay K. Singh, Zongchao Jia, and Julie D. Forman-Kay. Sensitivity of secondary structure propensities to sequence differences between α-and γ-synuclein: Implications for fibrillation. *Protein Sci.*, 15(12):2795–2804, 2006.

[40] Carlo Camilloni, Alfonso De Simone, Wim F. Vranken, and Michele Vendruscolo. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry*, 51(11):2224–2231, 2012.

[41] Magnus Kjaergaard, Søren Brander, and Flemming M. Poulsen. Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J. Biomol. NMR*, 49(2):139–149, 2011.

[42] Marius G. Clore and Junji Iwahara. Theory, practice and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.*, 109(9):4108, 2009.

[43] Simone Kosol, Sara Contreras-Martos, Cesyen Cedeño, and Peter Tompa. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. 18(9):10802–10828, 2013.

[44] Robert Schneider, Jie-rong Huang, Mingxi Yao, Guillaume Communie, Valéry Ozenne, Luca Mollica, Loïc Salmon, Malene Ringkjøbing Jensen, and Martin Blackledge. Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Mol. Biosyst.*, 8(1):58–68, 2012.

[45] Markus Rückert and Gottfried Otting. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.*, 122(32):7793–7797, 2000.

[46] Eike Brunner. Residual dipolar couplings in protein NMR. *Concept. Magnetic Res.*, 13(4):238–259, 2001.

[47] Ronaldo Mohana-Borges, Natalie K. Goto, Gerard J. Kroon, Jane H. Dyson, and Peter E. Wright. Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *J. Mol. Biol.*, 340(5):1131–1142, 2004.

[48] Robert Konrat. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Biomol. NMR*, 241:74–85, 2014.

[49] Loic Salmon, Gabrielle Nodet, Valéry Ozenne, Guowei Yin, Malene Ringkjøbing Jensen, Markus Zweckstetter, and Martin Blackledge. NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.*, 132(24):8407–8418, 2010.

[50] Malene Ringkjøbing Jensen, Rob W.H. Ruigrok, and Martin Blackledge. Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr. Opin. Struc. Biol.*, 23(3):426–435, 2013.

[51] Malene Ringkjøbing Jensen, Loïc Salmon, Gabrielle Nodet, and Martin Blackledge. Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J. Am. Chem. Soc.*, 132(4):1270–1272, 2010.

[52] Pau Bernadó, Carlos W. Bertoncini, Christian Griesinger, Markus Zweckstetter, and Martin Blackledge. Defining long-range order and local disorder in native α-synuclein using residual dipolar couplings. *J. Am. Chem. Soc.*, 127(51):17968–17969, 2005.

[53] Malcolm H. Levitt. *Spin dynamics: basics of nuclear magnetic resonance*. John Wiley & Sons, 2001.

[54] Anil Kumar, Christy Rani Grace, and P. Madhu. Cross-correlations in NMR. *Prog. Nucl. Mag. Res. Sp.*, 37(3):191–319, 2000.

[55] Mark W.F. Fischer, Ananya Majumdar, and Erik R.P. Zuiderweg. Protein NMR relaxation: theory, applications and outlook. *Prog. Nucl. Mag. Res. Sp.*, 33(3):207–272, 1998.

[56] Dmitry Korzhnev, M. Billeter, A. Arseniev, and Yu V. Orekhov. NMR studies of Brownian tumbling and internal motions in proteins. *Prog. Nucl. Mag. Res. Sp.*, 38(3):197–266, 2001.

[57] Neil A. Farrow, Ouwen Zhang, Attila Szabo, Dennis A. Torchia, and Lewis E. Kay. Spectral density function mapping using ^{15}N relaxation data exclusively. *J. Biomol. NMR*, 6(2):153–162, 1995.

[58] Pavel Kadeřávek, Vojtěch Zapletal, Alžbeta Rabatinová, Libor Krásnỳ, Vladimír Sklenář, and Lukáš Žídek. Spectral density mapping protocols for analysis of molecular motions in disordered proteins. *Journal of biomolecular NMR*, 58(3):193–207, 2014.

[59] Jiří Nováček, Anna Zawadzka-Kazimierczuk, Veronika Papoušková, Lukáš Žídek, Hana Šanderová, Libor Krásnỳ, Wiktor Koźmiński, and Vladimír Sklenář. 5D 13C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *J. Biomol. NMR*, 50(1):1–11, 2011.

[60] Perttu Permi and Arto Annila. Transverse relaxation optimised spin-state selective NMR experiments for measurement of residual dipolar couplings. *J. Biomol. NMR*, 16(3):221–227, 2000.

[61] Jiří Nováček, Noam Y. Haba, Jordan H. Chill, Lukáš Žídek, and Vladimír Sklenář. 4D non-uniformly sampled HCBCACON and ^{1}J (NC$^{\alpha}$)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. *J. Biomol. NMR*, 53(2):139–148, 2012.

[62] Frank Delaglio, Stephan Grzesiek, Geerten W. Vuister, Guang Zhu, John Pfeifer, and AD Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, 6(3):277–293, 1995.

[63] Krzysztof Kazimierczuk, Maria Misiak, Jan Stanek, Anna Zawadzka-Kazimierczuk, and Wiktor Koźmiński. Generalized Fourier transform for non-uniform sampled data. In *Novel Sampling Approaches in Higher Dimensional NMR*, pages 79–124. Springer, 2011.

[64] T. Goddard and D. Kneller. SPARKY, 3.

[65] Edward J. d'Auvergne and Paul R. Gooley. Optimisation of NMR dynamic models i. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J. Biomol. NMR*, 40(2):107, 2008.

[66] Bradley Efron and Robert J. Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[67] David S. Wishart and David A. Case. Use of chemical shifts in macromolecular structure determination. *Methods Enzymol.*, 338:3–34, 2000.

[68] Yunjun Wang and Oleg Jardetzky. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.*, 11(4):852–861, 2002.

[69] Joel R. Gillespie and David Shortle. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.*, 268(1):158–169, 1997.

[70] Hayes O. Griffith and Harden M. McConnell. A nitroxide-maleimide spin label. *P. Natl. Acad. Sci.*, 55(1):8–11, 1966.

[71] John H. Viles, Brendan M. Duggan, Eduardo Zaborowski, Stephan Schwarzinger, James A. Huntley, Gerard A. Kroon, Jane H. Dyson, and Peter E. Wright. Potential bias in NMR relaxation data introduced by peak intensity analysis and curve fitting methods. *J. Biomol. NMR*, 21(1):1–9, 2001.

[72] Philippe Pelupessy, Fabien Ferrage, and Geoffrey Bodenhausen. Accurate measurement of longitudinal cross-relaxation rates in nuclear magnetic resonance. *J. Chem. Phys.*, 126(13):134508, 2007.

[73] Rieko Ishima and Kuniaki Nagayama. Quasi-spectral-density function analysis for nitrogen-15 nuclei in proteins. *J. Magn. Reson.*, 108(1):73–76, 1995.