UNIVERSITY OF GOTHENBURG

The structural and dynamical basis of NusA's role in transcription-coupled DNA repair

Damasus C. Okeke
Department of Chemistry and Molecular Biology

Thesis for the degree of Doctor of Philosophy in the Natural Sciences
Gothenburg, 2022

Thesis for the Degree of Doctoral Philosophy in the Natural Sciences

The structural and dynamical basis of NusA's role in transcription-coupled DNA repair

Damasus C. Okeke

Cover: Structural illustration of NusA initial contact with DinB *via* their carboxy-terminal domains and subsequent assistance of other domains to stabilize the interaction based on initial docking performed using the HADDOCK 2.4 webserver (PDB IDs: 7PY3 and 4Q45).

ISBN: 978-91-8069-055-3 (PRINT)

ISBN: 978-91-8069-056-0 (PDF)

Available online at http://hdl.handle.net/2077/73707

Department of Chemistry and Molecular biology

Division of Biochemistry and Structural biology

University of Gothenburg

SE-405 30 Göteborg, Sweden

This thesis is dedicated to my family for their prayers and moral support particularly in the last year of this PhD education.

Abstract

DNA repair processes involves a host of repair machineries that are able to recognize and repair the damaged DNA. Nucleotide excision repair (NER) pathway is one of these processes exploited by cells to remove bulky DNA lesions. In *Escherichia coli*, transcription-repair coupling factors (TRCF) such as Mfd, NusA, and UvrD, mediate coupling of transcription to DNA repair by recruiting necessary repair machineries.

NusA is a flexible 55 kDa protein composed of six domains. Isolated free NusA in solution undergoes autoinhibitory intramolecular interdomain interaction which reduces its nascent RNA binding activity. The role of NusA in transcription regulation as well as DNA repair have been widely reported, however, structural characterization of NusA autoinhibition and interaction complex with repair enzymes has so far not been studied. Furthermore, the structural dynamics that underline the DNA-substrate binding activity of one of the repair enzymes, translesion DNA polymerase IV (DinB), is still missing.

The aim of this thesis is to unravel the structural dynamics behind NusA autoinhibitory phenomenon as well as the role of the DinB-Thumb in translesion DNA repair. I also aim to further characterize the transcription-coupled repair complex of NusA and DinB, alongside the UvrD helicase, using both biochemical and biophysical techniques.

In this thesis, I reported for the first combined backbone and methyl groups resonance assignment of full-length NusA and confirms that the wild-type NusA and "open" NusA mutant are in different state. Relaxation data reports on a low population of the open state in solution and shows a clear trend of released autoinhibition for the open NusA state. For DinB, I also for the first time completed the sequence resonance assignment and elucidate the structural dynamics governing its DNA substrate accommodation in the active site. Relaxation data reveal that the DNA-binding Thumb domain is structurally flexible. I showed for the first time that the stalled RNAP makes contact with DinB. I also observed that the self-autoinhibitory phenomenon of NusA diminishes its affinity to DinB in solution. On the other hand, DinB affinity to NusA further stabilizes in the presence of additional domains. For the NusA:UvrD interaction, I identified an additional binding interface within the central NusA S1-KH1-KH2 region, which is required to stabilize the interaction.

Keywords: Resonance ssignment, DNA repair, transcription-repair coupling factors, transcription-coupled repair, NusA, DinB, UvrD, autoinhibition, DNA substrate, dynamics, interaction.

Sammanfattning på Svenska

DNA-reparationsprocessen involverar en mängd reparationsmaskiner som kan känna igen och reparera det skadade DNA:t. Nukleotid excision reparation (NER) är en process som utnyttjas av celler för att ta bort stora DNA-skador. I *Escherichia coli* förmedlar transkriptions-reparationskopplingsfaktorer (TRCF), såsom Mfd, NusA och UvrD, koppling av transkription till DNA-reparation genom att rekrytera nödvändiga reparationsmaskiner. NusA är ett flexibelt 55 kDa protein som består av sex domäner. Isolerad, fri NusA i lösning genomgår autoinhibitorisk intramolekylär interdomänkomplexbildning som negativt reglerar dess syntetiserade RNA-bindande aktivitet. NusAs roll i transkriptionsreglering såväl som i DNA-reparation har rapporterats i stor utsträckning, däremot har inga tidigare studier fokuserat på strukturell karakterisering av NusA-autoinhibering eller interaktionskomplex med reparationsenzymer. Dessutom saknas information om den strukturella dynamiken som understryker DNA-substratbindande aktiviteter hos ett av reparationsenzymen, translesion DNA-polymeras IV (DinB). Syftet med denna avhandling är att klargöra den strukturella dynamiken bakom NusA autoinhiberande fenomen samt DinB-Thumb-domänens roll i translesions-DNA-reparationsrollen för DinB; och ytterligare karakterisera det transkriptionskopplade reparationskomplexet av NusA och DinB samt NusA och UvrD-helikas, med både biokemiska och biofysiska tekniker.

I denna avhandling slutförde jag för första gången både ryggrads- och metylgruppers resonanstilldelning av NusA i fullängd och bekräftar att vildtypen NusA och den "öppna" NusA-mutanten är i olika tillstånd. Relaxationsdata rapporterar mycket låg population av det öppna tillståndet i lösning och visar en tydlig trend av frisatt autoinhibering för det öppna NusA-tillståndet. På DinB-delen slutförde jag också för första gången sekventiell resonanstilldelningen och belyser den strukturella dynamik som styr dess DNA-substratplacering i det aktiva centret. Relaxationsdata visar att den DNA-bindande ”thumb”-domänen är strukturellt flexibel. Jag visade för första gången att den avstannade RNAP får kontakt med DinB. Vi observerade att det själv-autoinhiberande fenomenet av NusA minskar dess affinitet för DinB i lösning. Däremot stabiliseras DinBs affinitet för NusA i närvaro av ytterligare domäner. I interaktionen mellan NusA och UvrD identifierade vi flertalet bindande interface inom den centrala NusA S1-KH1-KH2-regionen som krävs för att stabilisera interaktionen.

Keywords: Resonance assignment, DNA repair, transcription-repair coupling factors, transcription-coupled repair, NusA, DinB, UvrD, autoinhibition, DNA substrate, dynamics.

List of Publications

This thesis consists of the following research papers:

Paper I: **Okeke D.C.**, Burridge C., Burmann B.M. Structural dynamics underlying NusA autoinhibition. *Manuscript*

Paper II: **Okeke D.C.**, Lidman J., Matečko-Burmann I., Burmann B.M. Thumb-domain dynamics modulate the functional repertoire of DNA-polymerase IV (DinB) *Submitted* (2022)

Paper III: **Okeke D.C.**, Kawale A.A., Burmann B.M. Structural basis of the complex of NusA and the translesion DNA polymerase IV (DinB). *Manuscript*

Paper IV: **Okeke D.C.**, Kawale A.A., Burmann B.M. Structural basis of the interaction between NusA and UvrD helicase during transcription-coupled repair. *Manuscript*

Contribution Report

Paper I. I performed the experiments including molecular cloning, protein expression/ purification, SEC-MALS, BLI and recorded some of the NMR data. The NMR data processing and sequence-specific resonance assignments were done between my supervisor and myself. The data analysis was performed by all co-authors. I, together with my supervisor, jointly wrote the manuscript.

Paper II. I designed and planned the project, together with my supervisor. I performed the experiments including molecular cloning, protein expression/ purification with support from other co-authors, SEC-MALS, BLI and recorded some of the NMR data. I processed the NMR data and did sequence-specific resonance assignments together with my supervisor. The data analysis was performed by all co-authors. My supervisor and I prepared figures and wrote jointly the manuscript.

Paper III. My supervisor and I designed the project. I performed the experiments including molecular cloning, protein expression/ purification, SEC-MALS, BLI and recorded some of the NMR data. The NMR data processing and sequence-specific resonance assignments were done between my supervisor and myself. The structure calculation and the data analysis were done by all co-authors. My supervisor and I prepared figures and wrote the manuscript.

Paper IV. I took part in the project design and planning of the experimental approach, together with my supervisor. I cloned, expressed and purified all the NusA-variants. I also purified UvrD and performed all the experiments as well as data analysis. My supervisor and I prepared figures and jointly wrote the manuscript.

TABLE OF CONTENT

Abbreviations

DNA – Deoxyribonucleic acid

NMR – Nuclear magnetic resonance

RNAP – RNA polymerase

GGR – Global genome repair

TCR – Transcription-coupled repair

TLS – Translesion synthesis

BER – Base excision repair

NER – Nucleotide excision repair

UV – Ultraviolet

Uvr – Ultraviolet-resistant

Nus – N-utilization substance

TRCF – transcription-repair coupling factor

CPD – cyclobutane-pyrimidine dimer

6-4PP – 6-4 photoproduct

BLI – Bio-layer interferometry

SEC-MALS – Size exclusion chromatography-multiangle light scattering

K_D – Dissociation constant

CSP – chemical shift perturbation

HSQC – heteronuclear single quantum correlation

PCR – polymerase chain reaction

IPTG – isopropyl-β-D-1-thiogalactopyranoside

SUMO – Small ubiquitin-like modifier

TROSY - transverse-relaxation optimized spectroscopy

CTD – carboxy-terminal domain

NTD – amino-terminal domain

PAD – polymerase-associated domain

LF – little finger

din – damage-inducible

The Scope of the thesis

The thesis provides in-depth presentation of my characterization of the structural and dynamical studies of the *Escherichia coli* transcription elongation factor, NusA, and the translesion DNA polymerase IV, DinB, as well as structural basis of NusAs role in transcription-coupled DNA repair (TCR) using advanced high-resolution solution-state nuclear magnetic resonance (NMR) spectroscopy and other complementary biophysical and biochemical methods. The thesis is divided into five chapters, as follows:

- **Chapter I: Introduction.** A brief review of the thesis literature background.
- **Chapter II: Methodology.** Description of the methodology basis of the research approach used in the thesis.
- **Chapter III: Results, Discussion.** Highlights of the major findings in the individual projects and their significance in advancing previous knowledge.
- **Chapter IV: Future perspectives.** Outlook for possible future studies to clarify questions raised by my results.

In addition to the above sections, the thesis also includes the following four papers.

- **Paper I (Manuscript)** elucidate the structural dynamics underlying the *Escherichia coli* transcription elongation factor, NusA, it's autoinhibition, and reveal how autoinhibited isolated free NusA is functionally activated.
- **Paper II (Submitted)** Reveals intrinsic dynamics of translesion DNA polymerase IV (DinB) Thumb domain studied under high-resolution solution-state NMR. Also, the underlying binding kinetics that govern DinB-interactions with DNA as well as RNA polymerase core are presented.
- **Paper III (Manuscript)** provides a detailed structural study of the interaction complex of NusA and the translesion DNA polymerase IV (DinB) including binding kinetics, which is a crucial event in the initial stages of the translesion synthesis DNA repair process.
- **Paper IV (Manuscript)** investigates the interaction between NusA and the UvrD helicase during transcription-coupled DNA repair, with structural characterization of the interaction interface together with the apparent binding kinetics.

Chapter I: Introduction

Here, I introduce the different coping mechanisms used by cells to ameliorate DNA damage-induced stress. In addition, I cover in detail the repair machineries involved in the three sub-pathways of transcription-coupled DNA repair (TCR) in *Escherichia coli*.

1.1 DNA Transcription

The deoxyribonucleic acid (DNA) sequence (also called gene) is a fundamental polymer of deoxynucleotides linked together by phosphodiester bonds. Each nucleotide consists a nitrogenous base (either guanine (G), adenine (A), thymine (T) or cytosine (C)), a deoxyribose and a phosphate moiety. The nitrogenous bases are responsible for the double-stranded helix structure of DNA through formation of complementary base pairs linked by hydrogen bonds. Most importantly, the bases encode genetic information required to make ribonucleic acid (RNA) – a transcript – by RNA polymerase (RNAP) in a process known as DNA transcription. During transcription, RNAP moves along the template DNA strand unwinding its double helix and extending growing RNA chain by adding ribonucleotides in the 5' – 3' direction [1]. However, the activity of transcribing RNAP can be hindered by presence of DNA damage or lesions on the template strand [2].

1.2 DNA damage

DNA is a highly reactive macromolecule, and often damaged by both physical and chemical agents that cause deformation of the bases and/or the phosphate backbone. DNA damaging agents induces deleterious effects on the DNA helix, thereby threatening the integrity and stability of the genome. These agents can either be from environmental sources (e.g., Ultraviolet (UV) light, ionizing radiation, and a variety of food-, air- and water-borne chemicals) or metabolic sources (e.g., methylating species and the reactive oxygen species (ROS) that arise during respiration). The most common forms of DNA damage are; (a) hydrolysis of the N-glycosyl bond linking a base to the deoxyribose, (b) hydrolytic deamination and methylation that can directly change one base into another, (d) covalent adduct dimers formed between two adjacent pyrimidines, and (e) oxidative damage to bases and to the phosphodiester backbone.

1.3 DNA repair Mechanisms

Cells are equipped with a variety of efficient tolerance and repair mechanisms to overcome the lethal effects of DNA damage-induced stress. Due to the large diversity of DNA lesions in terms of origin, form, and consequence, cells respond in different ways by either activating repair mechanism, cell cycle checkpoints, or apoptotic cell death pathways [3]. DNA repair processes involves many repair machineries that serve as global sensors or continuously scan the genome and are able to recognize the presence of a mismatched base, an apurinic or apyrimidinic site, or abnormal bases in general [4]. The presence of DNA lesions results in activation of appropriate repair mechanisms, which ultimately leads to the restoration of genetic information. Where repair is not possible, highly mutagenic damage tolerance mechanism involving translesion DNA polymerases is activated [4,5]. Some of the repair and tolerance mechanisms used by bacteria to overcome DNA damage include: photoreactivation, base excision repair (BER), nucleotide excision repair (NER), translesion synthesis, and recombination repair.

1.3.1 Photoreactivation

DNA lesion repair by photoreactivation involves a light-sensitive enzyme, DNA photolyase, which recognizes and reverses UV radiation-induced DNA lesions (such as cyclobutane-pyrimidine dimers (CPDs) and 6-4 photoproducts, 6-4PPs) using the energy from light [6–8]. DNA photolyase is a flavinprotein, which contains two chromophores capable of absorbing blue-light. During the actual repair process, an electron generated by excitation of the flavin cofactor is donated to the bulky dimers, resulting to splitting and reversal of the dimerization [6,8,9]. UV radiation is categorized into three different groups based on their wavelength, namely; UV-A (315 – 400 nm), UV-B (280 – 315 nm), and UV-C (<280 nm). The UV-induced lesions are mainly caused by the UV-B radiation as light of this wavelength is absorbed by native DNA [10,11]. The magnitude of the UV radiation effect on DNA bases depends on flexibility of the DNA strand and the nature of constituting bases. Regions with high pyrimidine base content, poly(dA)-(dT) tracts, are more susceptible to dimeric photoproduct formation [12,13]. These photo-induced dimers exert a lethal effect on cellular metabolism by obstructing the progression of replicative and transcribing polymerases [14,15]. Photolyases are specific to photo-induced dimers, and are classified as either CPD photolyases (removes CPD lesions) or 6-4 photolyases (removes 6-4PP lesions). Photolyases has been found in archaebacteria but not in humans and are considered ancient repair proteins, with

photoreactivation likely one of the oldest repair processes [4,16].

1.3.2 Excision repair

In excision repair, abnormal bases or distorted stretches of DNA are completely removed and new DNA stretches are synthesized in a complex process catalyzed by a host of repair enzymes. Excision repair is further sub-divided into two major pathways; base excision repair (BER) and nucleotide excision repair (NER), as outlined in more detail below.

1.3.2.1 Base excision repair

Damaged, abnormal bases are excised and replaced through the base excision repair (BER) pathway. The BER process is centrally catalyzed by DNA glycosylases, which has damage-specific activity, uracil glycosylase repairs hydrolytic cytosine deamination, while 3-methyladenine glycosylase repairs hypoxanthine, among others. In general, DNA glycosylases remove abnormal bases by cleaving the N-glycosidic bond between the base and the deoxyribose moieties of the nucleotide, thereby creating an apurinic/ apyrimidinic (AP) site. Subsequently, the cleaved site is removed by an AP endonucleases or an AP lyases, which nicks the DNA strand either 5' or 3' to the AP site, respectively. The cleavage of N-glycosidic bond leaves the deoxyribose phosphate part of the nucleotide remaining in the DNA sequence. The deoxyribose phosphate is subsequently excised by the action of a phosphodiesterase. The resulting gap is filled by a repair DNA polymerase and the nick is then sealed by the DNA ligase. Several genes for DNA glycosylases have been identified in *E. coli* (and in higher organisms) [17,18], indicating the importance of this repair pathway in preventing mutations arising from base deformation.

1.3.2.2 Nucleotide excision repair

In contrast to BER, nucleotide excision repair (NER) removes short damage-containing oligonucleotide stretches of about 10 – 12 bases [19], and has been shown to serve as backup system for the repair of single damaged bases such as thymine glycol and 8-oxoguanine [20,21]. In bacteria, NER process depends on activities of four UV-resistant (Uvr) proteins – UvrABCD [22]. The NER proteins can recognize and repair a wide range of DNA distortions including the bulky UV-induced CPDs and 6-4PPs. Even though the proteins are not evolutionary related across all domains of life, they share closely related strategies in their functional activity. The activity of the UvrABCD proteins on DNA lesion consists of three major stages: (*i*) damage recognition, (*ii*) incision and release of excised oligomer, and (*iii*) gap filling and DNA ligation.

Detection of damaged DNA is achieved either by a concerted action of UvrA with UvrB, which continuously scans the genome, or an actively transcribing RNAP, which serves as a global sensor to transcribed strand. Once a lesion is detected, UvrA is released and the ATP-dependent UvrB and UvrC performs dual incisions at specific locations both 5' and 3' to the lesion, leading to excision of 10 – 12 nucleotide-long DNA oligomers in the damage region. These proteins are collectively named after their physiological role – excision nucleases or excinucleases [22]. After excision, the excised short oligomer is displaced by UvrD (also referred to as helicase II) creating a gap on the DNA double strand. Finally, the gap is filled by DNA polymerase I, using the remaining complementary strand as a template, and finally the sealing the newly synthesized strand by DNA ligase [23,24]. The NER process is divided into two sub-pathways: (*i*) transcription-coupled repair (TC-NER), in which presence of DNA lesion is sensed by RNAP and involves preferential repair of transcribed strands, and (*ii*) global genome repair (GG-NER), in which DNA lesion is detected by the repair proteins, UvrAB, which repairs damage in the non-transcribed strand/ genome [25,26].

1.3.2.2.1 Transcription-coupled repair in *Escherichia coli*

RNAP is a global sensor for DNA lesions and during transcription can sense DNA lesions on the transcribed DNA strand. RNAP can bypass DNA lesions on the non-transcribed strand but becomes stalled at lesions on the transcribed strand, leading to preferential repair of this strand. RNAP arrested on the transcribed strand results in the activation of the TCR sub-pathway of NER by the transcription-repair coupling factors,TRCF [27,28] (Figure 1). Stalled RNAP remains tightly bound to the template DNA strand and occludes the damage site [27]. The repair coupling factors (such as Mfd, NusA, and UvrD) have been shown to play a major role in rescuing the stalled RNAP [29,30]. These factors mediate linking of transcription to repair by displacing the RNAP from the lesion site, and then recruiting the UvrABCD repair machineries to the exposed site, in a process known as transcription-coupled DNA repair (TCR) [31,32].

1.3.2.2.1.1 Mfd-dependent transcription-coupled repair

Mutation frequency decline (Mfd) protein was identified about three decades ago as an very important TRCF in *E. coli* [33]. Mfd is a multidomain ATP-dependent translocase that binds DNA upstream of the stalled RNAP, and also interacts with the β subunit of RNAP through its RNAP interaction domain (Mfd-RID). Joo-Seop and Michael reported that Mfd displaces the stalled RNAP by pushing it forward using energy from ATP hydrolysis, as it translocates along

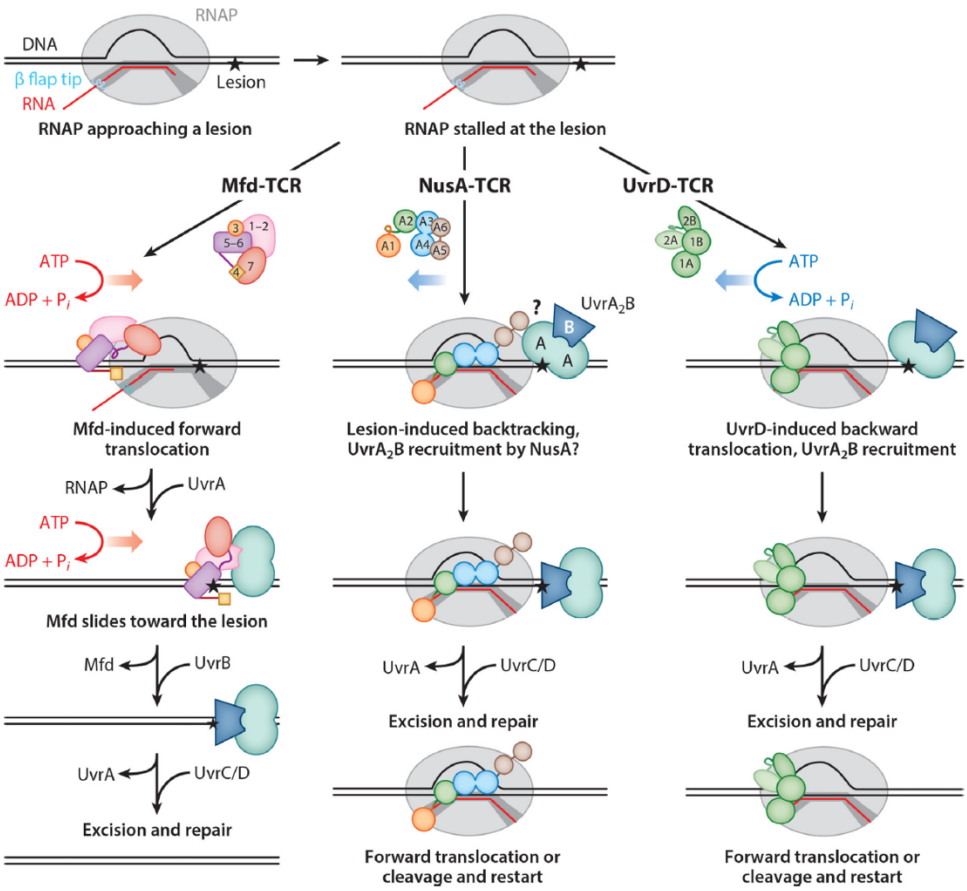

Figure 1. Three pathways of bacteria transcription-coupled repair (TCR). Transcribing elongation complex is stalled upon encounter with DNA lesion (black star) in the template DNA strand. Three transcription-repair coupling factors; Mfd, NusA and UvrD mediate the process. Mfd pushes RNAP forward, releasing it from DNA, and then recruits UvrA. UvrD pushes RNAP backward, exposing the lesion to Uvr proteins. NusA may stabilize a backtracked state induced by RNAP recognition of certain lesions and recruit UvrA. Following recruitment of UvrB, the repair pathway follows the canonical NER mechanism. Structural domains of Mfd, NusA, and UvrD are indicated by numbers. The figure was adopted from Belogurov and Artsimovitch [30].

the DNA strand [34,35]. After pushing the RNAP away from the lesion, Mfd directly interacts with UvrA which in turn recruits UvrB nuclease, thereby initiating Mfd-dependent TC-NER sub-pathway [35,36]. With the UvrAB in place, the repair process proceeds following the NER pathway, as described above. Earlier studies have suggested that Mfd is the only TRCF that mediates TC-NER in *E. coli*. However, a small but surprising effect was observed upon exposure of bacterial cells that lack the *mfd* gene to UV radiation, indicating a possible alternative TCR pathway [37,38]. Also, Mfd-mediated TCR failed for low transcribed genes, which that usually depend on TCR. Studies on the role of Mfd in TCR shows that overexpression of Mfd globally actually inhibits repair, supporting an earlier finding that Mfd is not induced upon

UV stress but is constitutively expressed, and that Mfd is not an abundant protein. Thus, recent findings indicate that Mfd may be recruited by UvrA and that Mfd acts as a DNA "clearing" factor owing to its role in removing the stalled RNAP from lesion site [39,40]. This updated view suggests that the role of Mfd in TCR is to displace the transcription complex that occludes damage site.

1.3.2.2.1.2 NusA-dependent transcription-coupled repair

NusA is a 55 kDa-multidomain protein known as a transcription elongation factor, and is part of TEC, in addition to NusG. NusA also associates with λN and Rho factors, which play a crucial role in transcription antitermination and termination processes, respectively [41]. NusA is composed of six domains (Figure 2A) an amino-terminal domain (NTD), three central S1, KH1, KH2 domains (usually referred to as SKK subdomains) and two carboxy-terminal acidic repeats (AR1 and AR2) domains [42–45]. Through the NTD and AR2 domains, NusA regulates the rates of transcription elongation *via* a NusA-AR2:αCTD-RNAP interaction [43,46] and induces both intrinsic and ρ-dependent termination *via* a NusA-NTD interaction with the flap-tip helix of RNAP in the vicinity of the RNA exit channel [43,47]. The central SKK subdomains bind nascent RNA, and participate in the formation of intramolecular SKK:AR2 interdomain complex which renders the SKK domains unavailable for RNA binding. This process is termed autoinhibition [48–50]. Proteomics studies, as well as far-western blot studies reveal that NusA makes direct contact with UvrA, one of the NER proteins. It has been proposed that NusA operates independently on the TC-NER pathway and recruits UvrA to the damage site. However, it appears to be deficient in RNAP-displacement activity. This observation has led to speculation that NusA may need other accessory factors during NusA-mediated TCR or that RNAP displacement is not required in this sub-pathway [51,52]. There is increasing evidence that NusA-mediated TC-NER requires other factors such as Mfd or UvrD, or both. Mfd has been postulated to act as a DNA "clearing" factor, but details on this role, especially the requirement for NusA, are lacking. On the other hand, UvrD has been shown to cooperate with NusA in Mfd-independent TC-NER. This was supported by recent results from the Nudler lab, which shows that NusA specifically interacts with UvrD helicase during TCR through its amino-terminal domain [52–54]. The recovery rate of cells exposed to damaging agents reveal significant differences in the rate of transcription in cells that lack one or more TRCF compared to wild-type [55]. This implies that different TRCFs cooperate with each other and the TC-NER sub-pathways are simultaneously active.

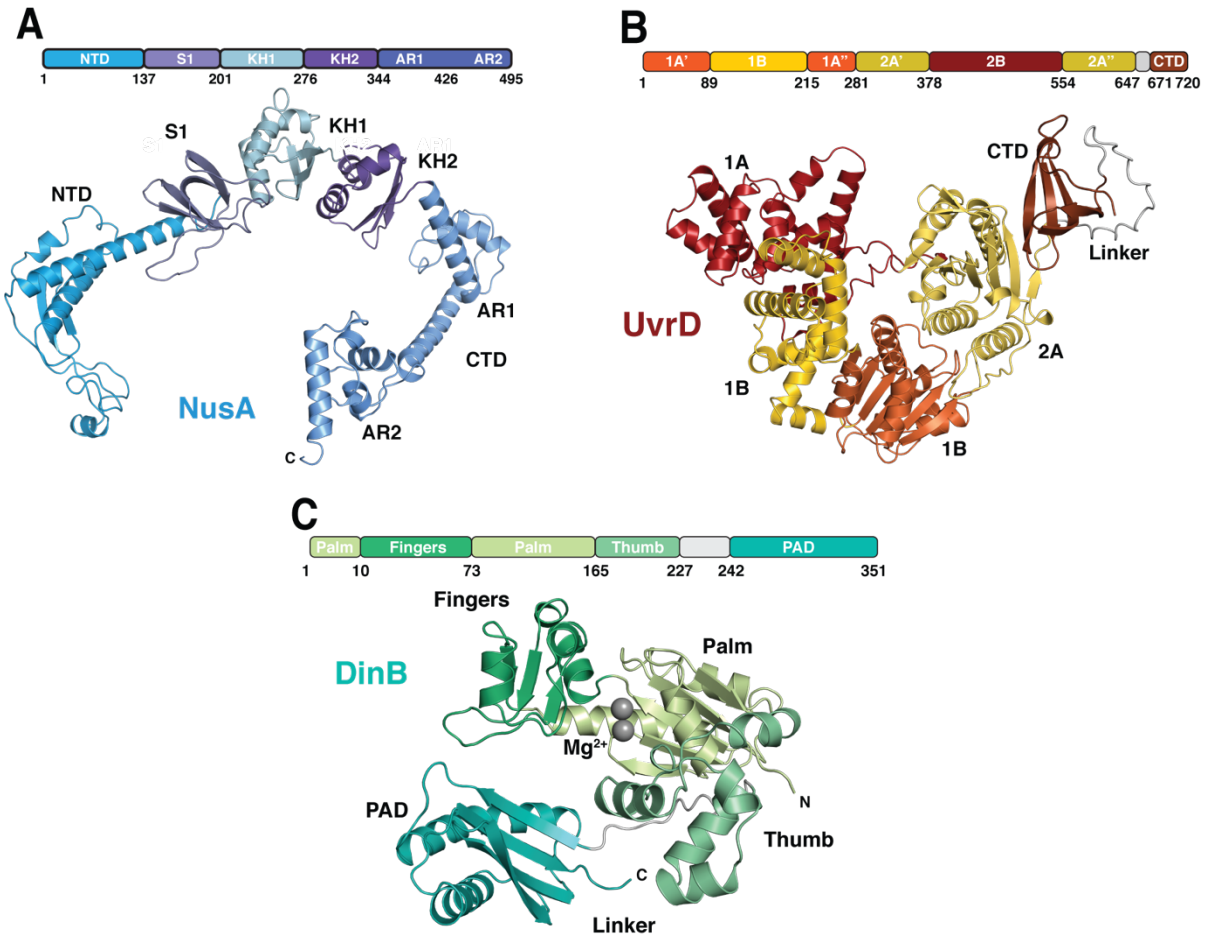

Figure 2: A) Schematic representation of the NusA domain architecture and the AlphaFold2 derived structural model, which is in good agreement with a recent high-resolution EM-structure of NusA bound to RNAP (PDB ID: 7PY3). **B**) Schematic representation of the UvrD domain architecture and the AlphaFold2 derived structural model, which is in good agreement with a high-resolution X-Ray structure of UvrD$^{1–647}$ (PDB: 3LFU) as well as a recent NMR structure of UvrD-CTD (PDB: 6YI2). **C**) Schematic representation of the DinB domain architecture and the X-ray of *E. coli* DinB (PDB ID: 4Q45) determined in complex with damaged dsDNA and incoming deoxyribonucleotides.

to cooperate with NusA in Mfd-independent TC-NER. This was supported by recent result from the Nudler lab, which shows that NusA specifically interacts with UvrD helicase during TCR through its amino-terminal domain [52–54]. The recovery rate of cells exposed to damaging agents reveal significant difference in the rate of transcription in cells that lack one or more TRCF compare to wild-type [55]. That could imply that all the different TRCFs cooperate with each other and the TC-NER sub-pathways are simultaneously active.

1.3.2.2.1.3 UvrD-dependent transcription-coupled repair

Evidence in *E. coli* points to the fact that UvrD is a *bona fide* transcription elongation factor [53]. Transcription elongation factors are known to be associated with transcribing RNAP, forming a transcription elongation complex (TEC). Therefore, UvrD is proposed to be constantly associated with RNAP during transcription. The UvrD helicase is a multidomain protein (Figure 2B) that possess both ATP-dependent DNA translocase as well as helicase activity. UvrD directly interacts with RNAP through its carboxy-terminal region [56–58]. As a DNA translocase in addition to the interaction with RNAP, UvrD slides along the DNA and pushes the stalled elongation complex backwards, utilizing ATP hydrolysis. This backtracking exposes the lesion site to the UvrABC repair machineries, which excise the distorted strands, generating a damaged DNA fragment that is removed through the helicase activity of UvrD [53]. The final stages of gap filling and DNA ligation are accomplished as described earlier. The effect of transcription factors such as GreA, GreB, Rho, and NusA, on the UvrD-mediated TCR has been investigated by cell assays, which show a decline in damage sensitivity of *uvr*D$^+$ cells devoid of *gre*B, *gre*A, and *mfd* [53]. GreA/B is known to assist in the stalled complex release by Mfd [53], suggesting that both Mfd and UvrD may act simultaneously.

1.3.3 Translesion synthesis

DNA lesions in the non-transcribed strand may cause a gap in the transcribed strand, which causes elongating transcription complexes to become arrested at the damage site. NusA, part of these complexes, has been shown to recruit and directly interact with translesion DNA polymerase IV (called DinB in bacteria), a Y-family polymerase, to mediate transcription-coupled translesion synthesis (TC-TLS) [59]. This finding highlights NusA as a key DNA repair factor – playing a role in both TC-NER and TC-TLS pathways. A crystal structure of DinB (40 kDa) in complex with damaged DNA template [60] shows a characteristic catalytic core-flexible linker-extended domain (CC-FL-ED), a structural feature common among Y-family polymerases (Figure 2B). The catalytic core exhibits a right-handed fold consisting of a Palm, Fingers, and Thumb sub-domains. The Little Finger (LF) domain, (sometimes alternatively called polymerase-associated domain (PAD) or C-terminal domain (CTD)) is connected to the extended domain through the flexible linker [61] (Figure 2C). DinB interacts with the template DNA *via* the Thumb domain [62] and with the incoming deoxynucleotide (dNTPs) substrates *via* Fingers domain [63]. Apart from its role in TCR, the presence of DinB at the site of a stalled replicative DNA polymerase was shown to be modulated by the β-clamp subunit of DNA polymerase III as reported by Heltzel and colleagues (2012) [64]. Other protein interaction

partners involved in the DNA damage tolerance and thus the physiological role of DinB include: UmuD, UmuD', RecA, and the molecular chaperone GroEL [65–67].

Unrepaired DNA damage can have deleterious effects including incomplete transcription and/or replication, which can result to pausing of the cell cycle or even cell death. Translesion synthesis is part of the SOS-induced damage tolerance mechanism, explored by cells for survival in the presence of DNA lesions [68]. The polymerase can bypass the lesion and synthesize DNA, thereby filling the gap in the transcribed strand, in a process known as translesion synthesis. Translesion DNA polymerase is a low fidelity and low processivity replicative enzyme, which executes the gap repair at risk of mutagenesis, hence this process is termed mutagenic repair [69].

Chapter II: Methodology

In this section an overview of the methods used in this project are described in details. As far back as 1838, scientists recognized proteins as a unique biomolecules capable of causing physiological changes [70]. From then, investigation on how proteins mediate their biological functions continued to gain interest, which forms the foundation of this project.

2.1 Molecular gene cloning

DNA constructs were produced by standard PCR (polymerase chain reaction) cloning methods or in some cases purchased from GenScript. Primers used for the PCR cloning were either designed manually or using the QuickChange site-directed mutagenesis protocol [71]. Full-length genes (*nus*A, *din*B, and *uvr*D) were purchased from GenScript with an amino(N)-terminal hexa-Histidine (6xHis)-Thrombin tag. *nus*A and *din*B plasmids were used as templates to create short constructs of truncated genes, which have part(s) deleted as well as full-length genes with point mutations. For cloning of SUMO-tagged *nus*A-variants (*nus*A-CTD, *nus*A-NTD, *nus*AΔAR2, *nus*AΔCTD, *nus*AΔNTD, and mutants) and *din*B-PAD, the 6xHis-Thrombin tag of *nus*A and *din*B were first, replaced with 6xHis-SUMO (small ubiquitin-like modifier) tag using the megaprimer PCR technique [72]. Subsequently, all the *nus*A-variants as well as DinB-PAD were cloned as amino-terminally SUMO-fused genes using the newly cloned full-length genes as template. The SUMO tag was used as it can: (*i*) help to enhance solubility of expressed recombinant protein, and (*ii*) efficiently cleaved off by SUMO specific proteases without leaving any additional residues to the target protein [73,74]. All PCR products were treated with DpnI restriction enzyme, which selectively digests methylated template DNA plasmids at 5'-Gm6ATC sites, leaving newly amplified products intact. DpnI does not digest the non-methylated, newly amplified DNA product. DpnI digestion is followed by heat inactivation to stop the DpnI activity. PCR products were transformed into highly efficient, chemically competent XL-10 Gold cells which were then plated on Luria-Bertani (LB) agar. Several colonies were then selected from the LB agar plate for plasmid extraction and subsequently send for DNA sequencing.

2.2 Protein expression

Plasmids with correct insert (gene of interest) were transformed into chemically competent *E. coli* BL21(λDE3) cells to be used as expression system. Bacterial expression systems are usually the first choice for recombinant protein expression due to wide vector adaptability and inexpensive growth conditions. All the genes of interest were cloned into a pET vector containing an inducible T7 RNA polymerase controlled by a set of T7lac regulatory genes, namely: lac promoter, lac repressor (*lac*I), and operator (*lac*O). In the absence of lactose or its non-metabolizable analogue, isopropyl-β-D-1-thiogalactopyranoside (IPTG), repressor proteins are constantly produced and bound to the operator site. As the repressor protein binding site, also called operator, overlaps with the T7 polymerase binding site, the promoter, the repressor protein bound to the operator blocks promoter site. This results in no gene transcription as the T7 polymerase cannot bind to the promoter. Protein expression is initiated by the T7 RNA polymerase-mediated gene transcription, which is induced in the presence of lactose or IPTG. Hydrolyzed lactose (allolactose) or IPTG binds to the repressor protein and inhibits its interaction to the operator. T7 polymerase is then able to bind to the promoter and initiates transcription [75,76].

All protein expressions were induced by the IPTG. Non-labeled protein was expressed in cells cultivated in LB medium. Expression of isotope-labeled proteins was carried out in either H_2O- or 100 % (v/v) D_2O-based 2xM9 minimal medium [77], supplemented with relevant isotopes either $(^{15}NH_4)Cl$ for [*U*-^{15}N]-labeled protein or $(^{15}NH_4)Cl$ and *D*-(^{13}C)-glucose for [*U*-^{13}C,^{15}N]-labeled protein; or deuterated proteins with $(^{15}NH_4)Cl$ for [*U*-^{2}H,^{15}N]-labeled proteins or *D*-(^{2}H, ^{13}C)-glucose for [*U*-^{2}H,^{13}C,^{15}N]-labeled proteins. Deuteration of proteins (molecular weight (Mw)>20 kDa) greatly improves signal-to-noise of NMR signals by decreasing the network of dipolar-coupled protons spin systems (also called spin diffusion suppression). In protonated proteins, there is increased dipole-dipole relaxation, leading to NMR signals broad linewidths. Deuteration reduces this effect, as deuterium has a much smaller gyromagnetic ratio than a proton, leading to reduced dipole-dipole interaction, reduced transverse relaxation, ultimately leading to smaller linewidths and a higher signal-to-noise ratio. In addition, decreased transverse relaxation increases spectral resolution in constant-time experiments, thereby reducing overlap of cross-peaks [78–81]. Specific methyl-group labeling of Methionine, Alanine, Isoleucine, as well as *proS* ^{13}CH group of leucine and valine (MAILV^{proS}) residues was achieved by growing the cells in 100 % (v/v) D_2O-based 2xM9 minimal medium supplemented with 2 % (v/v) Bioexpress rich, $(^{15}NH_4)Cl$, and *D*-(^{2}H,^{12}C)-glucose. Selectively

protonated amino acid precursors: 2-Ketobutyric acid-4-^{13}C,3,3-d_2 sodium salt hydrate (Isoleucine), DLAM-LV^{proS}-kit (2-(^{13}C)-methyl-4-(D_3)-acetolactate (valine/leucine *proS* methyl group only), [*U*-^{2}H, $^{13}CH_3$] methionine, and 2-[^{2}H], 3-[^{13}C] L-alanine, were added to the cell culture 1 hour before induction. Bioexpress helps to suppress any unwanted scrambling to other amino acids during biosynthesis of methionine and alanine. These precursors are easily incorporated into their respective amino acid biosynthesis pathway *in vivo*, as there are similar to the ones normally used by the cells to produce those amino acids. MALVI^{proS}-labeled NMR samples takes advantage of labeling only specific methyl groups to give less crowded and strong signals in [^{13}C-^{1}H]-HMQC spectra. This provides useful information for probing structure and dynamics of large proteins and supra-molecules (up to 1 MDa) [82–86].

2.3 Protein purification

All proteins were purified using fast protein liquid chromatography (FPLC, Äkta) system attached with chromatographic columns such as the immobilized metal ion affinity columns (IMAC, commercially called HisTrap column), ion-exchange chromatography columns, and size exclusion chromatography columns. Our target was to obtain highly pure and active proteins, which are stable for long days experiments at 25 – 37 °C. To achieve this, I applied three purification steps, based on the physico-chemical properties of the protein, to ensure that nearly all contaminants are removed. The first purification step used to isolate target protein from the crude cell lysate is IMAC. This strategy separates hexa-Histidine-tagged proteins from contaminants in the lysate. The imidazole ring of histidine can coordinate with metal ion (either Ni^{2+} or Co^{2+}) immobilized on the column resins [87], resulting to binding of the His-tagged proteins to the column. The bound proteins are eluted by competitive imidazole binding strategy, using high imidazole concentration. Proteins that contain SUMO tag are further treated with SUMO specific proteases, which cleaves off the tag leaving only the target protein without any extra residues [73,74]. Second purification step is by ion-exchange chromatography, which exploits net charge of the target protein. Ion-exchange chromatograph column is of two types; anion (positively charged) and cation (negatively charged) exchangers. The choice of the columns depends on the net charge of the target protein in a selected buffer, as the immobilized charged group binds to oppositely charged proteins. Proteins have no net charge at their isoelectric point (pI). However, they have net positive charge in buffer with pH lower than their pI, and net negative charge in buffer with pH higher than their pI. The pI of our target proteins ranges from 4 to 9, and most of our buffers has pH 7.0 – 8.0. Anion exchanger column

was used for protein with pI<7.0 whilst cation exchanger column was used for protein with pI>8.0. Target proteins bound on the ion-exchanger columns are eluted with high salt buffers. The last purification step is size exclusion chromatography (SEC, also called gel-filtration chromatography). This strategy separates proteins based on their molecular size and shape. Gel-filtration chromatography can separate proteins with the same molecular size but different three-dimensional structural shape due to denaturation [88,89].

2.4 Size exclusion chromatography multiangle light scattering

Size exclusion chromatography multi-angle light scattering (SEC-MALS) is used to characterize the molar mass, overall shape, aggregation, oligomerization, interactions, and purity of proteins and other macromolecules. SEC separates proteins by size based on partial exclusion from the pores of the column's stationary phase. But it is unable to accurately estimate the molecular mass because the retention time is a function of both molecular mass and hydrodynamic radius. The retention time is a function of the interaction between the protein and the stationary phase [90–92]. SEC-MALS is equipped with three detectors connected in series namely: an Agilent multi-wavelength absorbance detector (absorbance at 280 nm and 254 nm), a Wyatt miniDAWN TREOS multiangle light scattering (MALS) detector, and a Wyatt Optilab rEX differential refractive index (dRI) detector (Figure 3). Coupling the SEC column in-line with those detectors provides a more effective strategy for the accurate analysis of molecular mass, oligomeric states, aggregations, binding interactions, and hydrodynamic radii of proteins irrespective of the retention time because of the presence of the light scattering detectors that analyze molecules based on the intensity of scattered light. Furthermore, a combination of analytical SEC with the triple detectors ensures molar mass calculation at any point in the elution chromatogram and characterization of each population in a heterogenous sample [93–95]. In the course of this thesis, SEC-MALS was used to characterize molecular shape of different NusA mutants in comparison with wild-type NusA and binding interactions of NusA-DinB as well as NusA-UvrD.

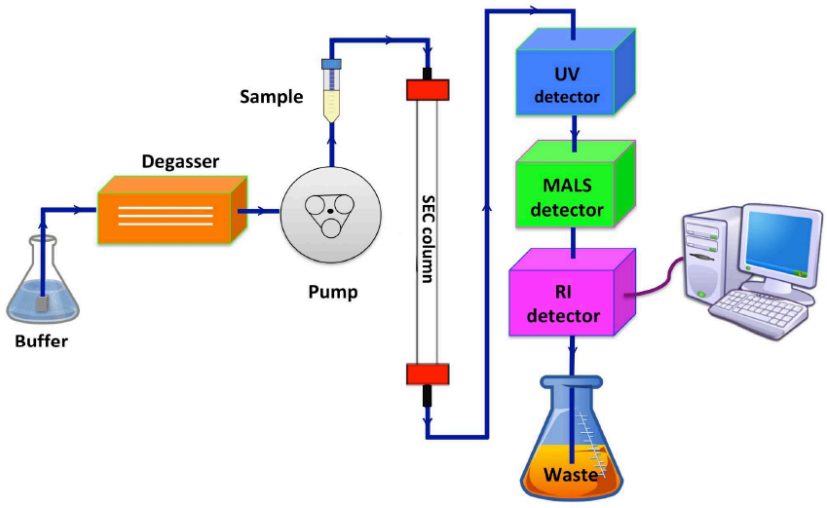

Figure 3. Schematic representation of size-exclusion chromatography multiangle light scattering (SEC-MALS) system. The SEC-MALS equipment consist of a size exclusion chromatography column in-line with three light detectors connected in series. SEC-MALS can provide information on molecular mass, hydrodynamic radius, oligomeric state, aggregations and binding interactions. The figure was adopted from Kwan *et. al.* [96].

2.5 Bio-layer Interferometry

Binding interaction between biomolecules is a natural phenomenon underlying crucial physiological processes such as DNA transcription and translation, cellular transport, signaling, and enzymatic activity. Therefore, studying binding kinetics provides essential information required to understand the mechanism of a particular process. A wide range of analytical techniques are available for accurate measurement of biomolecular binding affinity and kinetics, including: bio-layer interferometry (BLI), isothermal titration calorimetry (ITC), Surface plasmon resonance (SPR), enzyme-linked immunosorbent assay (ELISA), fluorescence resonance energy transfer (FRET), nuclear magnetic resonance (NMR) spectroscopy, etc.

BLI (bio-layer interferometry) uses a label-free approach to provide fast, reliable, and real-time quantification of biomolecular interactions. It uses the Octet® system technology compatible with microplate format and Dip and Read biosensor tips to perform a high-throughput binding study. As an optical instrument, BLI analyzes binding interactions by measuring interference patterns between light waves. The biosensor tip is composed of two optical interfaces: the internal reference layer and the biocompatible layer on the surface of the tip. Target molecules binds to the biocompatible layer and cause increase in thickness of the layer as more molecules bind. This results in a change in the effective distance between the

two optical interfaces, thus shifting the interference pattern of the reflected light (Figure 4). The change in the optical thickness upon binding interaction is observed as spectral shift on the detector and reported as change in wavelength (nm) on sensorgram [97].

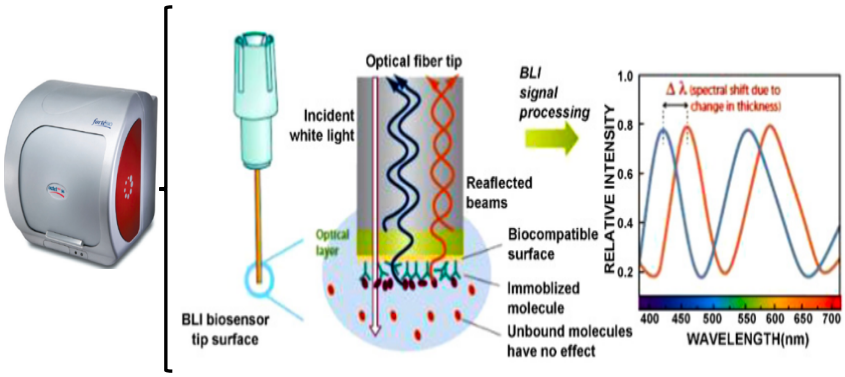

Figure 4. The working principle of the bio-layer interferometry (BLI) technique. Starting from left is the instrument, then the biosensor tip with zoom-in of the tip to show its biocompatible surface on which biotin-tagged ligands are immobilized. The surface is made up of two layers: a layer of immobilized molecules on the biosensor tip, and an internal reference layer. The last panel shows how change in interference pattern defines binding (wavelength shift, $\Delta\lambda$), which is caused by change in the thickness of biocompatible layer. The figure was adopted from Wallner *et al*., [98].

I used BLI to measure kinetics of the interactions identified to be involved in the NusA autoinhibition as well as the NusA-DinB, NusA-UvrD, DinB-RNAP, and DinB-DNA interactions.

2.6 NMR spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is an essential tool for studying molecular motions, interactions, and structure determination. NMR spectroscopy was first discovered in 1946 by Felix Bloch and Edward Purcell [99,100], as a result, they were jointly awarded the Nobel Prize in Physics 1952. Three decades later, two-dimensional (2D) NMR was introduced by Richard Ernst and his colleagues leading to the concept of correlation NMR experiments that involves two spins [101]. Ernst was awarded the Nobel Prize in Chemistry 1991 for this fundamental contribution. Furthermore, Kurt Wüthrich pioneered biomolecular NMR by applying existing methods to achieve first protein structure determination by NMR spectroscopy [102]. Wüthrichs work on NMR-based structure determination of biomolecules was also recognized with the award of the Nobel Prize in Chemistry 2002. Wüthrich pioneered 3D NMR spectroscopy to improve NMR tools for structure determination.

Basically, NMR spectroscopy is a physical phenomenon in which a particle of the nucleus with angular momentum, **I**, (also called nuclear spin) behaves as a magnetic dipole. Each angular momentum has an associated quantum number, *I*, which is specific for each nucleus. The total spin of a nucleus depends on the presence of unpaired protons and/or neutrons and determines the nuclear spin quantum number, *I*. Atoms with even number of both protons and neutrons, have $I = 0$, while atoms with odd number of both protons and neutrons, have $I \neq 0$. Where either the number of protons is even and the number of neutrons is odd or *vice versa*, $I = ½$. Nuclei with $I \neq 0$ and $I = ½$ are said to be NMR active.

The magnetic dipole moment of a nucleus is related to its angular momentum by the relation:

$$\mu = \gamma \mathbf{I}$$

where γ is the gyromagnetic ratio, a constant specific for each nucleus.

Nuclei spins with $I = ½$ have a magnetic moment along two directions, represented as $m \pm ½$, meaning 'spin up' and 'spin down'. These nuclear magnetic spins interact with the external magnetic field, which has its component along the z-axis, usually called B_0. The difference between energy levels with successive values of *m* is given by

$$\Delta E = \gamma h B_0$$

Where *h* is the plank constant. From the relation, it can be observed that the higher the magnetic field, the larger the energy difference between the two energy levels or states. In the absence of an external magnetic field, the spins are equally populated along the state while in the presence of external magnetic field more spins will be in the lower energy state, as described by the Boltzmann distribution. The external field causes the nuclear spin to precess (or resonate) at a frequency called the Larmor frequency. The Larmor frequency which nuclear spins experience is proportional to the strength of the external field and is different for different nuclei, e.g., ^{1}H, ^{13}C or ^{15}N. The difference in the external field experienced by the different nuclei is caused by shielding from electrons, with more shielded nuclei having lower chemical shift. The electronic environment of a given nucleus defines its chemical shift and position of its unique signal on NMR spectrum.

2.6.1 Protein-ligand interaction NMR studies

One of the applications of NMR spectroscopy is to extract information on molecular interaction at the atomic level [103,104]. NMR is able to capture protein-ligand interactions within binding affinities typically in the micro- to millimolar regime [105–107]. Some of the NMR-based

techniques used to study protein-ligand interactions includes: 1D-^{1}H line broadening, chemical shift perturbations among others.

Chemical shift perturbations (CSP) can occur due to the spatial proximity of aromatic rings or due to the non-covalent interactions with ligands and solvent molecules. The perturbed chemical shift of the protein can be assessed when ligand is titrated towards protein [108,109]. Uniformly-labeled ^{15}N or ^{13}C rich protein is need to perform either ^{15}N or ^{13}C, respectively, heteronuclear single quantum correlation (HSQC) titration with unlabeled ligand, which can either be another protein, nucleic acid, or any other molecule. The changes in chemical shift during titration can be monitored by acquiring series of 2D [^{1}H, ^{15}N]-HSQC spectra. Upon binding, specific ^{1}H-^{15}N cross peaks in the 2D spectra are displaced from their original position. Using completed sequence-specific resonance assignment, the displaced cross peaks are identified to belong to residues in the protein located in close proximity to the ligand binding site, and are said to exhibit significant ^{1}H-^{15}N CSP. In addition, CSP can also be observed due to backbone conformational change or allosteric changes in the protein due to ligand binding [110]. Protein-ligand interaction, $P + L \leftrightarrow PL$, is defined by the rate constants for the forward (k_{on}) – complex formation and backward (k_{off}) – complex dissociation reaction. In weak interactions/ fast exchange, k_{off} value is higher than the chemical shift difference between the free and bound protein, resulting to observed change in chemical shift of a single peak which represent average of signals from free protein and complex. Intermediate exchange occurs when k_{off} approximately equal to the chemical shift difference, whilst in slow exchange k_{off} is lower than the chemical shift difference, and show signal from either free or ligand bound protein [111] (Figure 5).

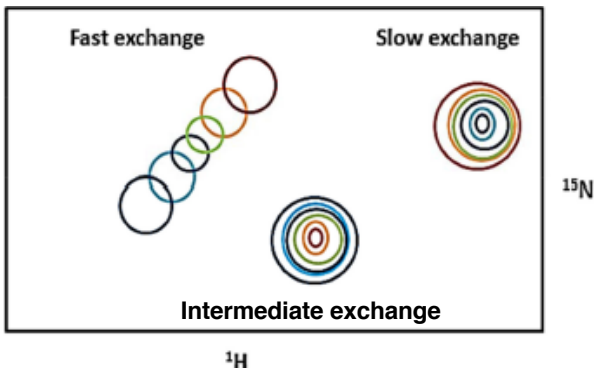

Figure 5. The effect of exchange rate on peak shape in a 2D [^{1}H-^{15}N] HSQC spectrum [111]. The peaks were obtained from a series of titrations from free protein (black) to ligand-bound (brown). Peak undergoing fast exchange show chemical shift change. While peaks undergoing intermediate and slow exchange show intensity loss and intensity gain, respectively.

2.6.2 Sequence-specific resonance assignment

The NMR technique for sequential assignment of proteins was initially developed by Kurt Wüthrich and his co-workers [112]. The starting point of resonance assignment is to acquire 2D [^{1}H-^{15}N]-HSQC using ^{15}N uniformly labeled proteins, which provides chemical shift correlation of the amide proton (H^{N}) and directly connected amide nitrogen (^{15}N) (Figure 6A) [113]. Because each amino acid besides proline residue gives a single peak in the ^{1}H-^{15}N HSQC spectrum, this experiment is a fingerprint as the dispersion of resonances signals correlate with folding properties. Next, triple resonance 3D experiments: HNCO, HNCA, HNCACB, CBCA(CO)NH, and HN(CO)CA [114], are recorded with proteins that are uniformly labeled with ^{13}C and ^{15}N so that, correlation peaks in the ^{1}H-^{15}N HSQC spectrum can be further resolved according to the chemical shift of ^{13}C in the third dimension, thus, the peaks in the triple resonance 3D spectrum gives a chemical shift correlation among ^{1}H, ^{15}N, and ^{13}C. Assignment of proteins in the range 25 – 100 kDa requires extra perdeuterated sample [115]. Perdeuteration is important to reduce the relaxation rates. Another strategy developed for successful 3D experiment acquisition is the transverse-relaxation optimized spectroscopy (TROSY) pulse sequence [116].

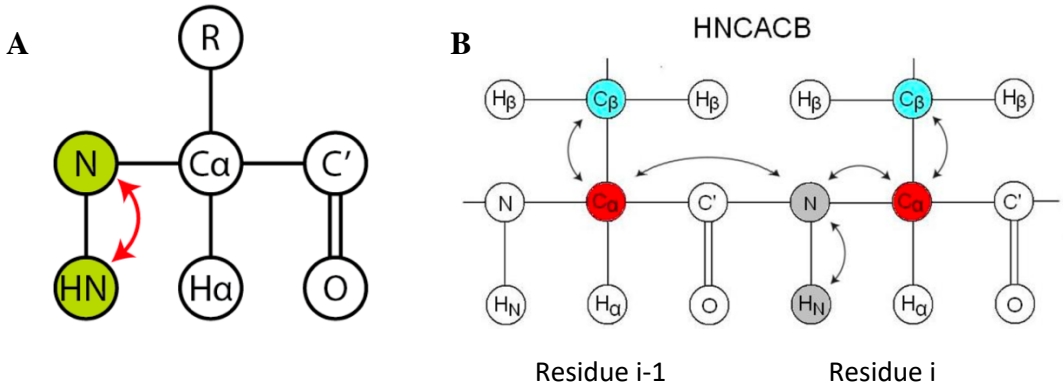

Figure 6. **A**) The correlation of the amide proton and directly connected amide nitrogen. Each amino acid residue contains a ^{H}N-^{15}N bond, and thus gives a single peak in the ^{1}H-^{15}N HSQC spectrum. **B**), correlation of cross peak chemical shifts of $^{1}H_{i}$, $^{15}N_{i}$, and $^{13}C_{i}$ in its own residue and ^{13} C_{i-1} in its preceding residue. The figures were adopted from: en.wikipedia.org/wiki/Triple-resonance_nuclear_magnetic_resonance_spectroscopy

The triple resonance 3D experiments are the general assignment procedure for proteins [114]. Since the nitrogen and carbon atoms (C^{α}, C′ and side chain carbons) are uniformly labeled with ^{15}N and ^{13}C, respectively, protons, carbons, and nitrogens are connected with each other *via* spin-spin couplings. Their chemical shifts are obtained from the 3D experiments including: HNCO HNCA, HNCACB, CBCA(CO)NH, HN(CO)CA [117,118] to assign the backbone C^{α}, C^{β}, C' as well as H^{N} and ^{15}N from ^{1}H-^{15}N HSQC [119]. In the same vain, side chain carbon and proton resonances can be assigned using 2D [^{13}C, ^{1}H]-HSQC spectra with/without constant time (CT) version, as well as 3D (H)CC(CO)NH, H(CC)(CO)NH, and HCCH-TOCSY-experiments [118]. The actual assignment process is performed by correlating cross peak chemical shifts of $^{1}H_{i}$, $^{15}N_{i}$, and $^{13}C^{\alpha}_{i}$ within its own residue and $^{13}C^{\alpha}_{i-1}$ to its preceding residue as well as $^{13}C^{\beta}_{i}$ and $^{13}C^{\beta}_{i-1}$ for example using an HNCACB (Figure 6B). In an HNCA strip at the specific ^{1}H and ^{13}C chemical shift plane, two peaks corresponding to intra-residual ($^{1}H_{i}$, $^{15}N_{i}$, $^{13}C^{\alpha}_{i}$) and inter-residual ($^{1}H_{i}$, $^{15}N_{i}$, $^{13}C^{\alpha}_{i-1}$) cross peaks are observed. In an HN(CO)CA, a single cross peak corresponding to the chemical shift of $^{13}C^{\alpha}_{i-1}$ is observed. In principle, the sequential assignment of the main chain protons, nitrogens, and carbons is done by connecting the chemical shifts of own residues to preceding residue, thereby building a growing sequence. In most cases especially with intermediate and large proteins, the chemical shift dispersion is small in C^{α} chemical shifts so that it is generally difficult to align residues on the sequence solely based on the chemical shifts of C^{α}. As the chemical shift of C^{β} is more disperse, neighboring residues can be uniquely connected using the chemical shifts of both C^{α} and C^{β} [113].

2.6.3 Solution NMR studies of large proteins

Improvements in the NMR methodologies such as development of the TROSY pulse sequence, stable isotopes labeling schemes, and availability of NMR instruments operating at higher magnetic fields have stretched the size limit of protein assemblies possible to be studied by solution-state NMR to as high as 1 MDa [119–122]. Incorporation of TROSY approach in both amides [119] and methyl groups [123] NMR studies has significantly reduced the NMR signals fast decay problem caused by slow tumbling [124]. Introduction of perdeuteration strategy to replace all non-exchangeable protons by deuterons [115] provided additional solution to slow molecular tumbling as described in section 2.2. Furthermore, selective labeling of specific amino acid side chains (aromatic and methyl groups) in a highly deuterated background provides an excellent strategy for obtaining information even in the hydrophobic core of proteins as well as at the interface of biomolecular complexes. Application of the selective labeling has yielded

high sensitivity and resolution, specifically in combination with methyl-TROSY [125,126]. The selective specific methyl group labeling scheme is currently available for all six methyl-bearing amino acids such as Ala, Ile, Leu, Met, Thr, and Val [86,127] as well as aromatic ring labeling [128,129], in a highly efficient manner without scrambling. These amino acids represent about 40 % in protein sequence and distributed randomly throughout the entire chain, thereby serving as a reporter for all parts of the protein. NMR sequence-specific resonance assignment of large proteins is usually achieved by assigning isolated individual domains. In a case where well-dispersed resonances of isolated domains correspond with resonances of the full-length protein, the assignment is directly transferable [130]. Structural characterization without the need of *de novo* determination of large systems has become feasible especially if the structure is available either by crystallographic or cryo-electron microscopy studies or through structure prediction algorithm such as AlphaFold2. A combination of intramolecular NOEs, line broadening and paramagnetic relaxation enhancement (PRE) also provides direct insight into structural and dynamic information [131,132].

2.6.4 NMR-based protein dynamics studies

Proteins are not rigid molecules, rather they have inherent ability to change their conformations rapidly in solution in relation to their biological functions. Protein conformation is usually described as the spatial arrangement of its constituent atoms, which determine the overall shape of the macromolecule [133]. Time-dependent change in the protein conformation is called protein dynamics, which defines the timescale in which structural change occur and the nature of the corresponding change in atomic coordinates. NMR is sensitive to local structure and can capture dynamics over a wide timescale, from picoseconds to hours (Figure 7). Experimental methods used to study protein dynamics and conformations are NMR spectroscopy, cryo-electron microscopy, small-angle X-ray scattering, mass spectrometry, atomic force microscopy and single-molecule fluorescence resonance energy transfer, etc.

NMR spectroscopy is a unique technique as it can capture time-dependent fluctuations in structure at atomic resolution in solution state. Thus, it enables sophisticated protein dynamics studies based on its ability to quantify dynamics (*i*) under equilibrium conditions without external perturbations such as changes in pH, temperature or pressure, (*ii*) using many probes simultaneously with atomic resolution, and (*iii*) over a wide range of time scale, from picoseconds to hours. The longitudinal (R_1) and transverse (R_2) relaxation rates of proton-bound ^{13}C or ^{15}N nuclei are sensitive to protein motions on either the pico- to nanosecond timescale

for R_1 or the micro- to millisecond timescale for R_2, as well as the associated heteronuclear NOEs (hetNOEs) reporting on the reorientation of the bond vector on fast timescale. Several key parameters can be quantified *via* their relationship with the spectral density functions [134], including the rigidity of the bond vector under investigation, which is expressed by the generalized order parameter S^2, the time scales of intra-molecular motions, expressed by the correlation time τ_e, the overall level of molecular tumbling, expressed by the rotational correlation time τ_c, and finally the contributions to chemical exchange, R_{ex}.

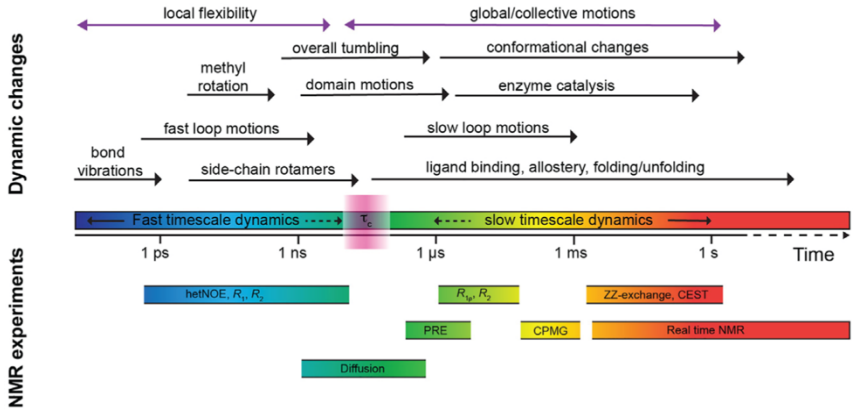

Figure 7. Overview of the protein dynamics, timescale and available NMR methods for quantifying atomic-resolution information for detection of motional events. Also, protein function affected at certain timescale is indicated above. The figure was adopted from Kawale and Burmann [135].

Chapter III: RESULTS AND DISCUSSION

3.1 Probing dynamics of NusA autoinhibition (Paper I)

Isolated NusA exhibits self-regulatory autoinhibition property in solution, in which the carboxy-terminally AR2 domain undergo intramolecular interdomain interaction with the central SKK sub-domain, thereby hindering its activity [136]. Previous studies on this important regulatory mechanism studied interactions between isolated individual domains [48,50]. In **paper I**, I studied the conformational state of full-length NusA in solution and gained insight into its inherent dynamical properties that form the basis of its autoinhibitory properties using mostly advance solution-state NMR spectroscopy. By transferring available assignments of the sub-domains, which were experimental confirmed with uniformly labeled [*U*-^{2}H, ^{13}C, ^{15}N] full-length protein, I was able to obtain for the first time almost a complete sequence-specific resonance assignment of full-length NusA.

To unravel the domain interplay within the NusA autoinhibited state, I performed NMR titrations among the three main NusA sub-domains: the NTD, the SKK, and the CTD. In each titration, unlabeled sub-domain was titrated against [*U*-^{13}C, ^{15}N]-labeled sub-domain. Figure 8 shows [^{1}H,^{13}C]-NMR spectra of the labeled sub-domains with their titration partners. I could observe signal intensity loss when CTD was titrated against [*U*-^{13}C, ^{15}N]-NTD and *vice vera* (**Figure 8A, C**). Surprisingly, titration of SKK against [*U*-^{13}C, ^{15}N]-CTD did not show any significant change in either intensity or chemical shift (**Figure 8D**). In the commonly used [^{1}H,^{15}N]-NMR spectra, only a few chemical shift changes were observed when [*U*-^{13}C, ^{15}N]-CTD was titrated with either SKK or NTD (**Supplementary Figure 1**). However, [^{1}H,^{13}C]-NMR spectra of [*U*-^{13}C, ^{15}N]-SKK in the presence of either CTD or NTD show significant intensity loss (**Figure 8A, C**). In conclusion, the titration data and the plotting of the effects onto the NusA structure revealed that the main interacting domains are the central α-helical part of the NTD, the KH1 domain as well as the AR2 domain (**Figure 8G**). This clearly indicates an interaction between the NTD and the CTD.

Further characterization of the observed interactions by bio-layer interferometry (BLI) yields a K_D value of 1.0 ± 0.01 μM for CTD:SKK (**Figure 8H**) and a notably weaker K_D of 142 ± 94.5 μM for CTD:NTD (**Figure 8I**). These values clearly indicate that the CTD:SKK interaction is likely the driving interaction for the closed state, whereas the CTD:NTD interaction is considered rather transient and might contribute in a stabilizing role to the closed autoinhibited form.

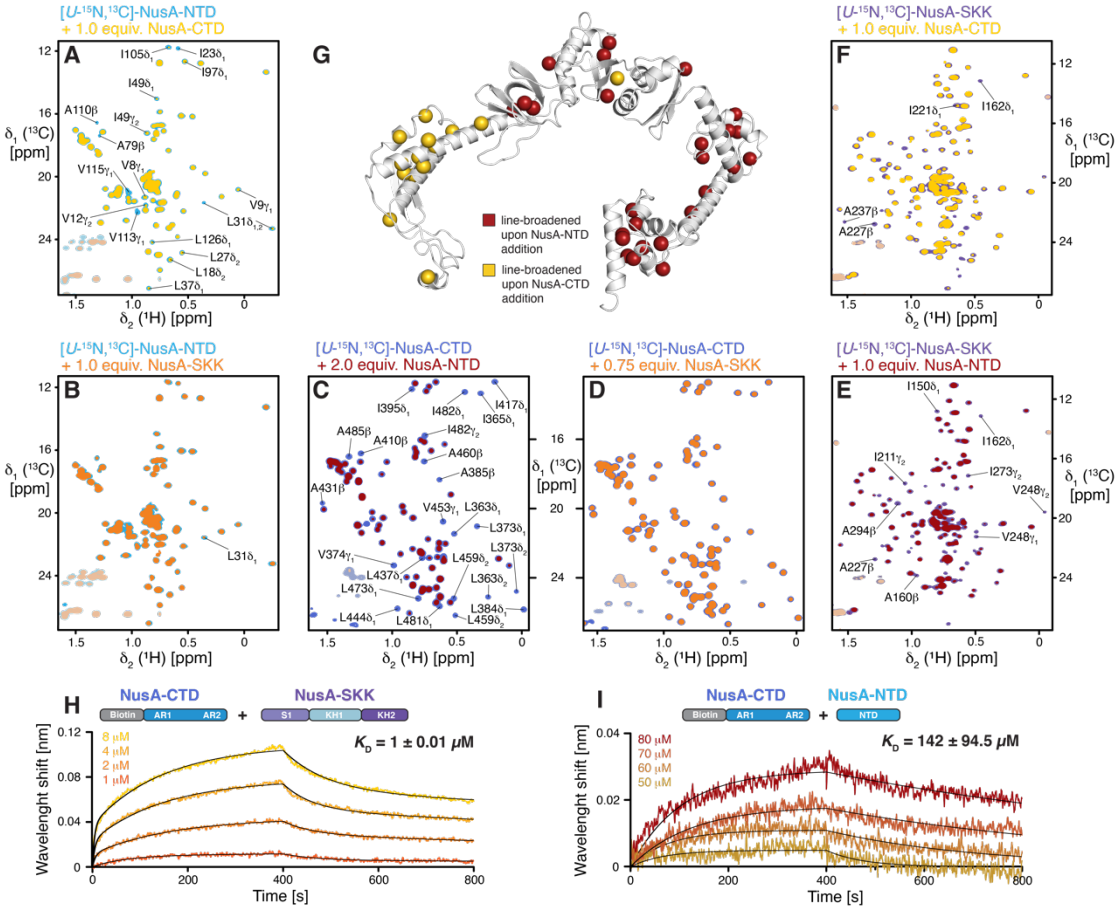

Figure 8. **A–F**) 2D [^{13}C,^{1}H]-NMR spectra of NusA constructs in isolation and in the presence of unlabeled NusA subdomains in NMR-buffer at 310 K. [*U*-^{13}C,^{15}N]-NusA-NTD in isolation and with NusA-CTD (**A**) or NusA-SKK (**B**). [*U*-^{13}C,^{15}N]-NusA-CTD in isolation and with NusA-NTD (**C**) or NusA-SKK (**D**). [*U*-^{13}C,^{15}N]-NusA-SKK in isolation and with NusA-NTD (**E**) or NusA-CTD (**F**). Methyl groups of NusA residues exhibiting line-broadening upon addition of NusA-CTD (yellow) or NusA-NTD (red), indicated as spheres on the NusA structure (PDA-ID: 6FLQ). **H, I**) Biolayer interferometry (BLI) analysis of NusA-SKK (**H**) and NusA-NTD (**I**) binding to biotinylated NusA-CTD. Increasing analyte concentrations are indicated (red to yellow). Non-linear least square fits to the experimental data are shown as black lines with dissociation constants, K_D.

Based on the findings here, which are in agreement with earlier report that the KH1 domain is involved in the autoinhibited closed state of NusA [50]. To disrupt the autoinhibition, we created different combinations of point mutations on the identified residues residing within the KH1 domain. The chromatogram of size-exclusion chromatography multiangle light scattering experiment of all the mutants show that the mutant with four mutations is the most "open" NusA, as it elutes earliest compare to others (See figure in the manuscript).

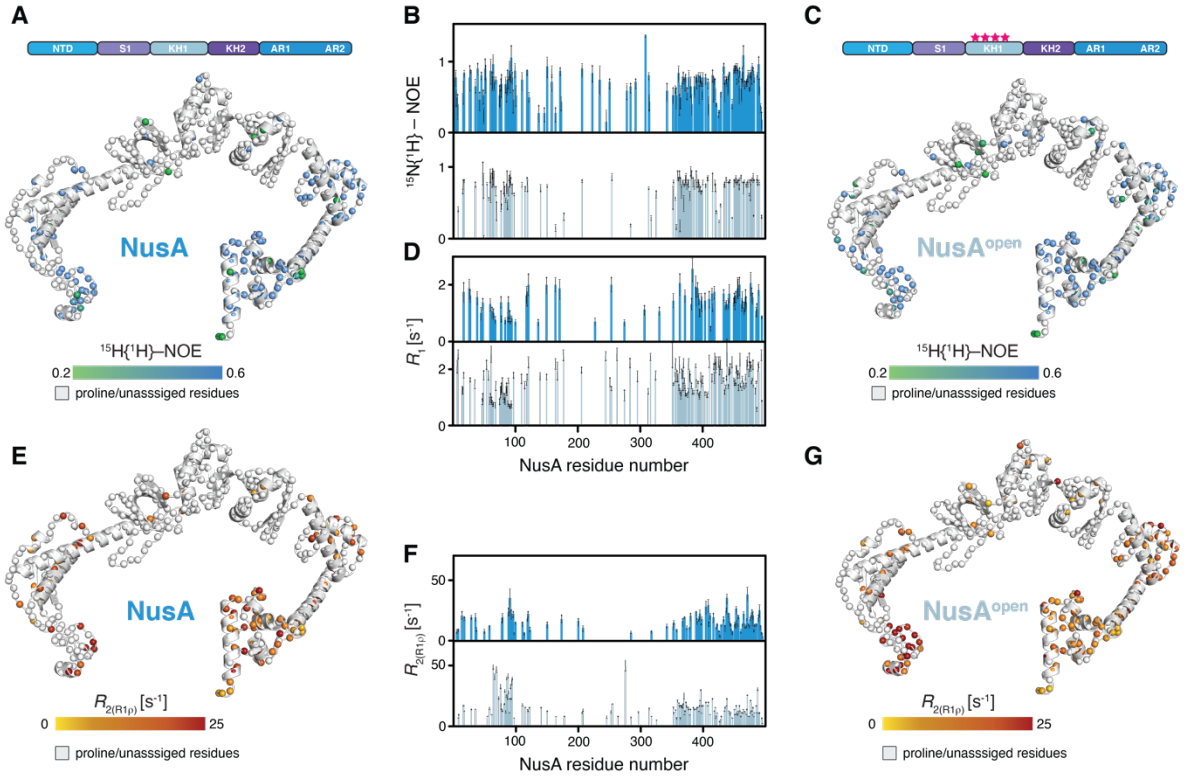

Figure 9. Dynamics of full-length NusA and NusA^{open} probed with backbone amide relaxation measurements. **A**) Pico- to nanosecond dynamics shown on NusA-structure (PDB-ID: 6FLQ), with backbone amide ^{15}N{^{1}H}-NOE values indicated as colored spheres (green to blue gradient). **B**) ^{15}N{^{1}H}-NOE values for NusA (blue) and NusA^{open} (light-blue). **C**) ^{15}N{^{1}H}-NOE values for NusA^{open}, similarly to panel (**A**). **D**) Longitudinal relaxation rates, R_1, for NusA (blue) and NusA^{open} (light-blue). **E**) Micro- to millisecond dynamics shown on NusA-structure (PDB-ID: 7PY3), with $R_{2(R1\rho)}$ values indicated as colored spheres (yellow to red gradient). **F**) Transverse relaxation rates, $R_{2(R1\rho)}$, for NusA (blue) and NusA^{open} (light-blue). **G**) $R_{2(R1\rho)}$ rates for NusA^{open}, similarly to panel (**E**).

With both the NusA and NusA^{open}, we set to analyze their backbone and methyl side-chain (not presented here) dynamics using standard NMR relaxation experiments [137,138]. Our backbone dynamics data show that the ^{15}N{^{1}H}-NOE (hetNOE) and ^{15}N longitudinal R_1 relaxation rates report on fast timescale motions on the pico- to nanosecond timescale. We estimated a stable hetNOE ratio profile, on average of 0.71 for NusA, and observed a sharp decrease towards the carboxy-terminus, consistent with a stable protein with increased flexibility at the carboxy-terminus, which is devoid of stable secondary structure elements (**Figure 9A, B**). The overall picture was highly similar for NusA^{open}, which showed a slightly enhanced average of 0.75 that could possibly indicate a minor modulation of the picosecond movements in the open NusA state (**Figure 9B, C**). The R_1-rates show averages of 1.38 s^{-1} and 1.58 s^{-1}, for NusA and NusA^{open}, respectively (**Figure 9D**), indicating opposite trend on

expected influence of pico- to nanosecond dynamics of the NusA states. We measured the ^{15}N transverse relaxation rates R_2 as derived from the $R_{1\varrho}$ rates ($R_{2(R1\varrho)}$), to explore the motions in the micro- to millisecond timescale. The $R_{2(R1\varrho)}$ rates of NusA and NusA^{open}, report on motions on the lower microsecond timescale (the used spin-lock radio frequency field of 2,000 Hz levels out all exchange contribution (R_{ex}) much slower than 80 μs [137]) show a similar planar behavior as observed for the dynamics on the fast timescale resulting in average values of 16.2 s^{-1} for NusA and of 16.1 s^{-1} for NusA^{open}, respectively (**Figure 9E–G**). Although the average values are highly similar, closer inspection of the obtained $R_{2(R1\varrho)}$ -rates showed a subtle modulation of these μs motions in the β-strand preceding the linker helix between the NTD and the SKK-domain for the open state as well as some reduction of the amplitude of motions within the carboxyterminal AR2 domain. Although these data cannot be interpreted *en detail* at the moment, these subtle modulations of the dynamics of different timescales clearly point to a role of the inherent dynamics of NusA in the autoinhibition process.

3.2 Investigating thumb domain input to the DinB function (Paper II)

Crystallographic studies of *E. coli* DinB reveals three domains: Thumb, Fingers, and Palm, connected to an additional polymerase-associated domain (PAD) through an extended flexible linker [60]. In **paper II,** I aimed to unravel the inherent conformational changes necessitating the translesion DNA repair activity of DinB with closer attention on the DNA-binding Thumb region using advanced solution-state NMR spectroscopy and BLI. Due to the size of DinB, 40 kDa, I applied divide-and-conquer approach to isolate individual Thumb domain and PAD as well as DinBΔPAD. Sequence-specific backbone assignment of [*U*-^{13}C,^{15}N]-PAD yielded 100% and [*U*-^{13}C,^{15}N]-Thumb yielded 84%. For [*U*-^{2}H,^{13}C,^{15}N]-DinBΔPAD, backbone resonance assignment was about 95 % completed. With the data of the different constructs, we were able to transfer and validate the assignments of full-length DinB reaching a final assignment completeness of ~95% for the backbone resonances.

Next, to address the conformational dynamics of DinB *en detail*, we evaluated the backbone dynamics of DinB over a broad range of timescales by NMR relaxation measurements [137,138]. By measuring the steady-state heteronuclear ^{15}N{^{1}H}-NOE (hetNOE) and the ^{15}N longitudinal (R_1) relaxation rates we probed the pico- to nanosecond motions of the H–N bonds. In these type of experiments, high hetNOE values and small R_1 rates indicate rigid and stably folded regions whereas the inverse, low hetNOE values as well as high R_1 rates, points to flexible and unfolded segments. Consistent with the initial structural characterization, the hetNOE data indicated that the folded parts of the Fingers, Palm and PAD are stably folded as evident by high hetNOE values, whereas linker and loop regions are more flexible as indicated by low hetNOE values (**Figure 10A, B**). Slight contrast was the observed behavior of the Thumb domain which showed on average lower hetNOE values of 0.66 compared to the residues comprising sheets and helices of the full-length protein resulting in an average value of 0.73, which in general indicates a stable fold. Thus, the obtained values for the DinB in general are within the theoretical maximum expected at 18.8 T (800 MHz ^{1}H frequency) of 0.86 indicating the presence of global fast motions within the whole domain in solution. These motions are most pronounced within the Thumb domain as this region shows the lowest average values, which might also stem from inherent dynamics as this domain also showed quite distinct chemical shift changes between separated form and within the full-length protein as discussed above. The obtained average R_1 rates for the folded segments of DinB could be determined to be 0.36 s^{-1}, in line with the magnitude of the obtained hetNOE values. We observe only a marginal increase for the Thumb domain to 0.38 s^{-1} and a more notable one for

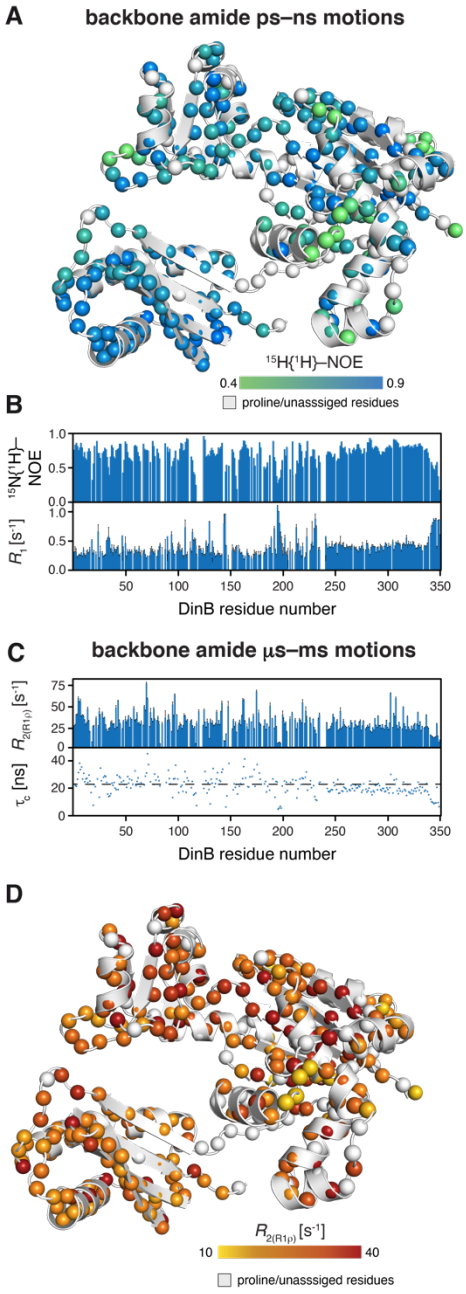

Figure 10. ***A)*** Dynamics on the ps–ns timescale plotted on the DinB structure (PDB-ID: 4Q45). The amide moieties are shown as spheres and the hetNOE values are indicated by the green to blue gradient. **B**) The hetNOE (top) as well as the longitudinal relaxation rate R_1 are plotted against the DinB residue number. **C**) The $R_{2(R1\rho)}$ (top) as well as the rotational correlation time τ_c, obtained from the analysis of the R_1 and $R_{2(R1\rho)}$ relaxation rates, are plotted against the DinB residue number. The broken line indicates the average value of 21.5 ns. **D**) Dynamics on the μs–ms timescale plotted on the DinB structure (PDB-ID: 4Q45). The amide moieties are shown as spheres and the $R_{2(R1\rho)}$ values are indicated by the yellow to red gradient.

the carboxy-terminal residues 342–351 with 0.76, indicating the absence of secondary structure elements in this residue stretch.

In the next step, we analyzed the contributions of motions in the micro- to millisecond regime. To assess these slow timescale motions, we analyzed the ^{15}N transverse relaxation rates. We measured the R_2-rates derived from the $R_{1\rho}$ rates ($R_{2(R1\rho)}$), which report on the motions on the lower micro-second timescale, because under the used spin-lock radio frequency (RF) field of 2,000 Hz all exchange contributions (R_{ex}) much slower than 80 μs would be leveled out. In line with the previous analysis of the fast timescale motions, we observed a largely planar profile for the folded segments with an average value of 29 s^{-1} compared to almost identical 28 s^{-1} for the Thumb domain (**Figure 10C, D**). Based on the R_1 rates as well as the $R_{2(R1\rho)}$ rates we next determined the rotational correlation time τ_C of the DinB with ~22 ns for the structured region (**Figure 10C**), which is in line for a protein of 40 kDa. In summary, a picture emerges where the highly flexible linker region connects a stably folded PAD to the rest of the DinB protein, whereas the Thumb domain exhibits a large scale of fast timescale motions on the pico- to nanosecond timescale, which could be either attributed to inherent flexibility or the priming for interactions.

To obtain a more detailed picture of the underlying dynamics within DinB and especially within its Thumb domain, I exploited the increased sensitivity of methyl groups to get access to the side-chain dynamics of the DNA polymerase IV. I chose specific labelling of isoleucine, alanine, and methionine, as well as stereo-specific labeling of valine and leucine methyl groups as these are well dispersed among the whole DinB, proving specific probes (**Figure 11A**). The exceptional quality of the obtained 2D [^{13}C,^{1}H]-NMR spectrum enabled us to assign ~93% of all methyl groups of the MALVI^{proS}-DinB sample in a sequence specific manner (**Figure 11A**).

Using this methyl-labelled sample we initially determined the product of the side-chain order parameters and the correlation time of the overall molecular tumbling ($S^2_{axis} \bullet \tau_C$), which reports on the extend of the amplitude of motions on the fast NMR timescale [139,140]. The obtained values showed a maximum of ~23 ns, which is in good agreement with the determines τ_c for the protein backbone with 22 ns. The distribution of the different values also indicated some inherent side-chain flexibility within the Fingers and Palm domains. Nevertheless, the quality of the measured data prevented detailed quantitative analysis at the current state providing only information for a sub-set of resonances.

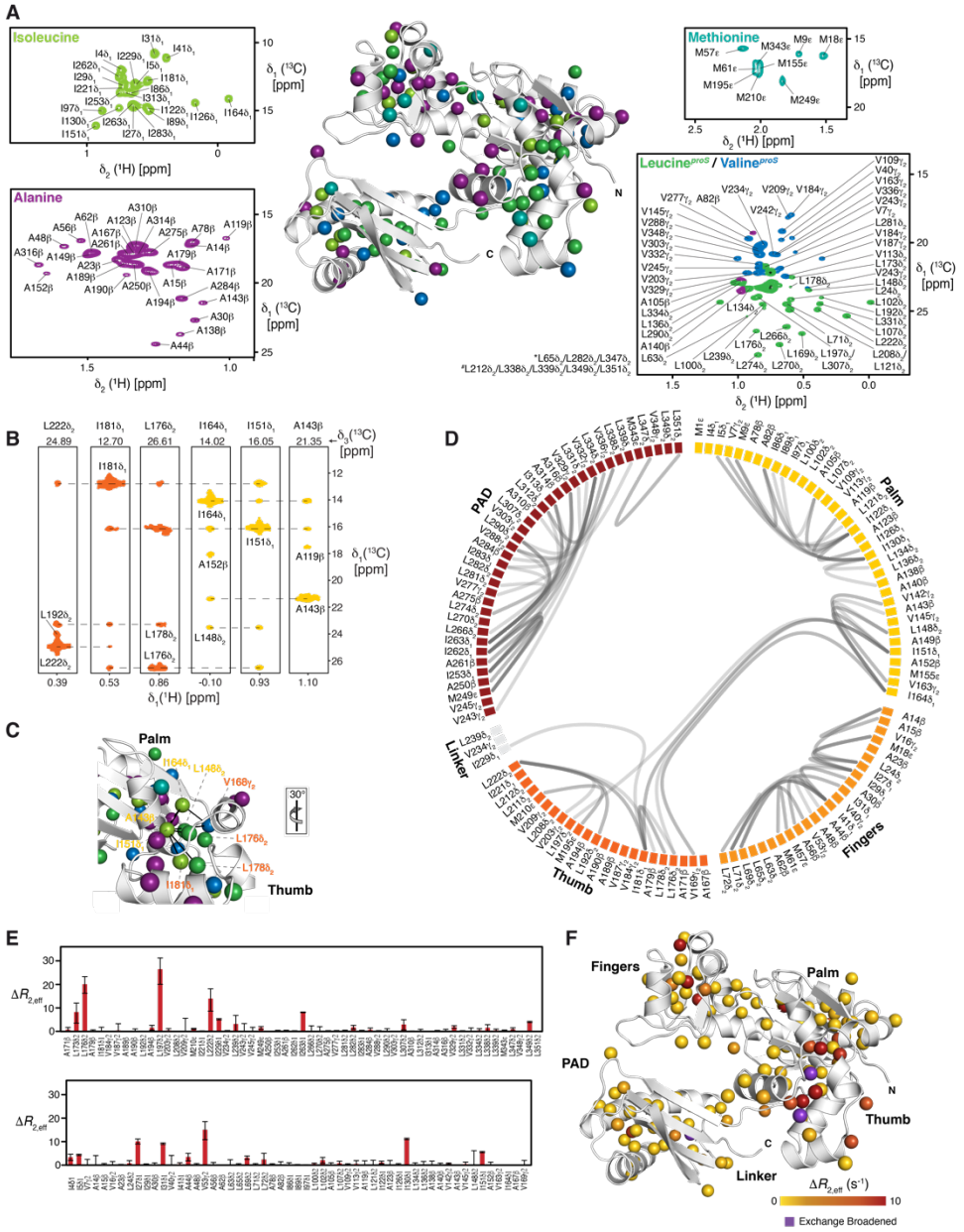

Figure 11. **A**) Distribution and assignment of isoleucine, alanine, methionine, leucine and valine methyl groups using an [*U*-^{2}H, Ile-δ_1-^{13}CH$_3$, Leu-δ_2-^{13}CH$_3$, Val-γ_2-^{13}CH$_3$, Ala-^{13}CH$_3$, Met-^{13}CH$_3$] stereospecific labelled DinB measured in NMR buffer at 298 K. * and # denote leucine residues in the overlapping central region. **B)** Representative NOE strips from a 3D $^{13}C_{methyl}$-$^{13}C_{methyl}$-$^{1}H_{methyl}$ SOFAST NOESY focusing on the interdomain stabilization between Palm (yellow) and Thumb (orange). **C)** NOE network, indicated by the black lines, stabilizing the Thumb on the Palm surface plotted on the DinB structure (PDB-ID: 4Q45). The respective orientation relative to panel **A** is indicated. **D)** Flareplot visualization of the complete methyl-methyl NOE network detected for the MALVI^{proS}-DinB illustrating the connectivity between the individual domains **E, F**) ΔR_{2eff} values for the methyl-groups obtained from the difference of $R_{2,eff}$ at the lowest and highest CPMG frequency υ_{CPMG} (**E**). Structural view of the amplitude of the CPMG relaxation dispersion profiles ΔR_{2eff} at 18.8 T (**F**).

The initial 2D [^{13}C,^{1}H]-NMR spectrum of MALVI^{proS}-DinB had already shown some indications of specific line-broadening, which can possibly be attributed to conformational exchange processes. Therefore, we used to a multiple quantum (MQ) Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiment [141]. We quantified the exchange-induced broadening effects (depicted as $\Delta R_{2,\mathrm{eff}}$) by measuring the difference in the relaxation rates at two different CPMG fields (25 and 750 Hz). The obtained values indicated that the vast of the methyl groups is not involved in chemical exchange processes as the values were close to zero in agreement with a stable protein fold (**Figure 11E, F**). Nevertheless, within the core of the Thumb domain in particular several residues experience large exchange rates or were even exchange broadened, which is further prove of some structural flexibility within this domain, which could possibly be attributed to movements of the different Thumb helices against each other within the core of this domain. In addition, some residues core also showed enhanced relaxation rates also pointing to a possible adaptation mechanism in this region.

3.3 Characterizing the NusA:DinB interaction complex (Paper III)

With the knowledge of inherent dynamical properties of individual NusA and DinB which modulate their respective biological functions. I set to further characterize the previously reported direct interactions between the two proteins during transcription-coupled translesion DNA synthesis (TC-TLS) [59]. The NusA:DinB interaction was previously studied using DinB-immobilized affinity column, far-Western blotting and filter peptide arrays and showed that NusA directly interacts with DinB, and highlighted that the carboxy-terminally 263 amino acids of NusA is mostly involved [59]. Due to limitations of the methods applied, which did not involve full-length protein, it was suggested that the observed interaction might involve other part(s) of NusA.

I had created two sub-constructs of NusA: NusA-CTD and NusAΔCTD as well as two sub-constructs of DinB: DinB-PAD and DinBΔPAD. After a series of experiments, which is covered in detail in the accompanying manuscript, we identified NusA-CTD and DinB-PAD as the minimal interaction interface. To further characterize this interface, initial NMR titration studies reveal that the interaction between the two domains involves fast exchange binding, as demonstrated by change in chemical shifts of both [*U*-^{15}N] NusA-CTD (**Figure 12a**) and [*U*-^{15}N] DinB-PAD (**Figure 12c**) upon contact with unlabeled DinB-PAD and NusA-CTD, respectively. Chemical shift perturbations against residues number reveal most affected or perturbed residues (with $\Delta CSP > 0.14$). These residues are considered to be directly involved in the interaction. In NusA-CTD, the most affected residues include the famous hydrophobic W490 and F491 as well as N487, all located within the AR2 domain (**Figure 12b, e**). These hydrophobic residues have been implicated in most NusA interactions that involve the C-terminally AR2 domain [49,142–144]. Whilst the binding interface in DinB-PAD involves R238, L239, R240, K241, S242, G244, L274, V277, T299, H302, V303, W304, V336, and T337, mostly localized in the concave surface of the domain (**Figure 12d, f**). Interestingly, the same perturbed residues observed in the [*U*-^{15}N] NusA-CTD:DinB-PAD titration spectrum and additional ones were affected upon titrating [*U*-^{15}N] NusA-CTD with unlabeled DinBΔPAD (**Not shown here**). Binding kinetic analysis of NusA-CTD:PAD-DinB complex yields a dissociation constant (K_D) of 1.4 ± 0.04 μM (**Figure 12h**), which is within the μM – mM range typical for weak transient interactions.

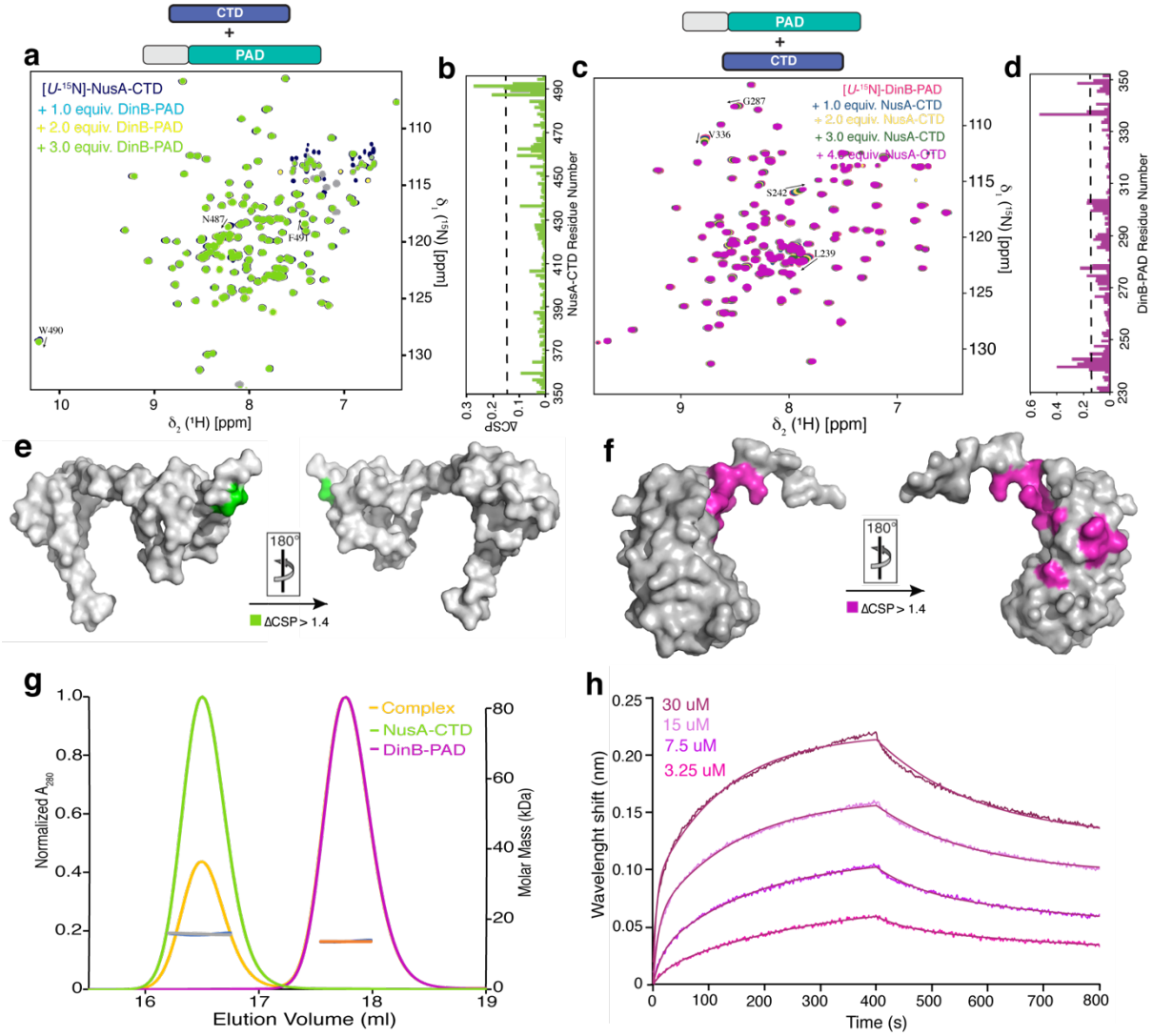

Figure 12: a, NMR titration spectra overlay of [*U*-^{15}N] labeled NusA-CTD (dark blue) at 1.0 (blue), 2.0 (yellow), and 3.0 (green) molar equivalent of unlabeled DinB-PAD unlabeled DinB-PAD. **b,** Overlay of 2D [^{15}N,^{1}H]-NMR spectrum of [*U*-^{15}N] labeled DinB-PAD in the absence (red) and presence of 1.0 (navy blue), 2.0 (yellow), 3.0 (green), and 4.0 (magenta) molar equivalent of unlabeled NusA-CTD. The most perturbed resonances in both spectra are labeled with an arrow indicating direction of the chemical shift. **c–d**, Plot of the NusA-CTD resonances chemical shift changes upon interaction DinB-PAD (3.0 molar equivalent) against the residues number. **d**, Chemical shift changes of the DinB-PAD resonances upon NusA-CTD (4.0 molar equivalent) binding plotted against the residues number. **e–f** Residues undergoing stronger CSPs than twice the standard deviation are mapped in green (for NusA-CTD) or magenta (for DinB-PAD) on the surface representation shown in two different orientations. **g**, Size exclusion chromatography-coupled multiangle light scattering (SEC-MALS) chromatogram of complex co-elution mixed at 1:1 equimolar concentration of the interacting domains. The experiment was performed at room temperature. **h**, Binding kinetic analysis of the CTD:PAD complex by Bio-layer interferometry. Serially diluted solutions of DinB-PAD (30, 15, 7.5, and 3.75 μM) were titrated against biotinylated NusA-CTD at 30 °C.

Structural Models representing the minimal as well as the intertwined complex

Based on the obtained chemical shift changes for the minimal complex formed between the NusA-CTD and the DinB-PAD we used in the next step the HADDOCK 2.4 webserver [145,146] to generate a structural model of the interaction between the two domains. The output of the HADDOCK calculation converged to a single cluster with a target energy function of - 158.7 ± 6.1 resulting in a buried surface are of ~200 Å^2 revealing the packing of the central β-sheet of the DinB-PAD directly against the helix-turn-helix motif of the NusA-AR2 domain (**Figure 13a**).

Encouraged by these promising results we also used the same approach to generate an initial model of the complex formed by the full-length proteins. The output of the HADDOCK calculation converged in this case to two clusters with a target energy function of -49 and -47, respectively, indicating the presence of two distinct states, which are currently not as well defined as the minimal complex as evident from the elevated HADDOCK score. Both clusters yielded extended buried surface areas of ~630 Å^2 and ~580 Å^2, respectively, indicating the contributions of additional domains of both binding partners in the complex formation. Analyzing the differences of the two clusters indicated a subtle re-orientation of the catalytic domain of DinB thus the two clusters might represent the engaging and the stably bound form of the complex (**Figure 13b**). Although, the initial complexes indicates a stable complex formation *via* discrete binding states, additional refinement against experimental data is required to obtain a high-resolution structural model of the NusA:DinB complex.

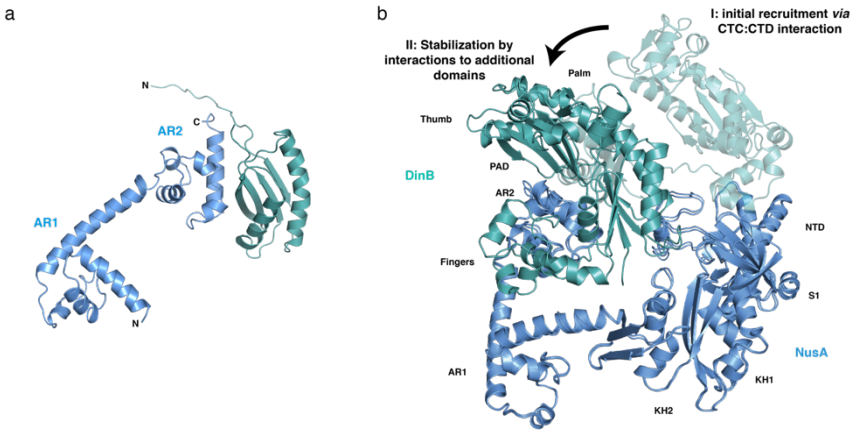

Figure 13: a, Structural model of the NusA-CTD:DinB-PAD complex based on rigid body docking employing chemical shift perturbation data using the HADDOCK 2.4 webserver. **b**, Structural model of the NusA:DinB complex based on rigid body docking as outlined under panel **a**. Two distinct clusters could be obtained, possibly indicating the engagement as well as the stably bound complex.

3.4 Characterizing the NusA:UvrD interaction (Paper IV)

To further explore the role of NusA in TCR, I investigated the recently discovered interaction between NusA and UvrD revealed by cross-linking mass spectrometry [54]. In this paper, I employed NMR spectroscopy, bio-layer interferometry and size-exclusion chromatography to further investigate the proposed NusA-NTD:UvrD interaction. The initial NMR titration result confirms that NusA interacts with UvrD and the interaction is mediated by the NusA-NTD (**Figure 14A**). Analyzing the signals attenuated in a residue specific manner showed that NusA-NTD residues (Thr33–Lys38, S99, V100) with significant backbone amide resonance intensity changes, in the presence of unlabeled UvrD, are mostly located on α-helices (**Figure 14B,C**). Experimental evidence revealed that RNAP and λN binding region on NusA-NTD are located at a groove forming a hydrophobic patch, which the RNAP binding at the concave surface while λN binds directly opposite at convex surface [147]. Both the BLI and SEC-MALS data indicate that the sole NusA-NTD:UvrD interaction is rather transient (estimated $K_D = 52.9 \pm 5$ nM) and might be modulated by the interplay of other binding partners of both NusA and UvrD as well as the access to the central domains of NusA.

The analysis of binding affinity with full-length NusA shows only a slight increase in affinity for UvrD ($K_D = 52 \pm 9.6$ nM; **Table 1**), pointing a possible additional binding surface on NusA which could contribute to stabilizing the interaction. Nevertheless, in this scenario, it has to be taken into account that NusA undergoes an autoinhibitory intramolecular interdomain interaction in solution [136], resulting to a "closed" conformation, thus likely initially preventing direct interaction to the extended binding interface of the central S1, KH1, and KH2 domains.

To test the effect of the NusA autoinhibition phenomenon in NusA:UvrD interaction, we created NusAΔNTD subconstruct (with deleted NTD region) and with a previously generated NusA-mutant, which we could show in a previous study to be constitutively open [148], that harbors four point-mutations in the KH1 domain (**Figure 15E**) and was termed $NusA^{open}$, and subsequently repeated the binding assay. Our result for the NusAΔNTD:UvrD complex provides a K_D value of 4.59 ± 2.55 nM (**Figure 15D**), highlighting that the interaction between NusA and UvrD is enhanced in the absence of NTD. Looking in more detail at the details of the difference underlying this ~10-fold change in K_D showed that the on-rate, k_{on}, was almost unaffected with $2.4 * 10^4$ $M^{-1}s^{-1}$ and $2.9 * 10^4$ $M^{-1}s^{-1}$ for NusA and NusAΔNTD, respectively. In contrast the two off-rates (k_{off}) differ by one order of magnitude from $4.67 * 10^{-4}$ s^{-1} for NusA compared to $8.33 * 10^{-5}$ s^{-1}, respectively, thus the reason for the lowered dissociation constant.

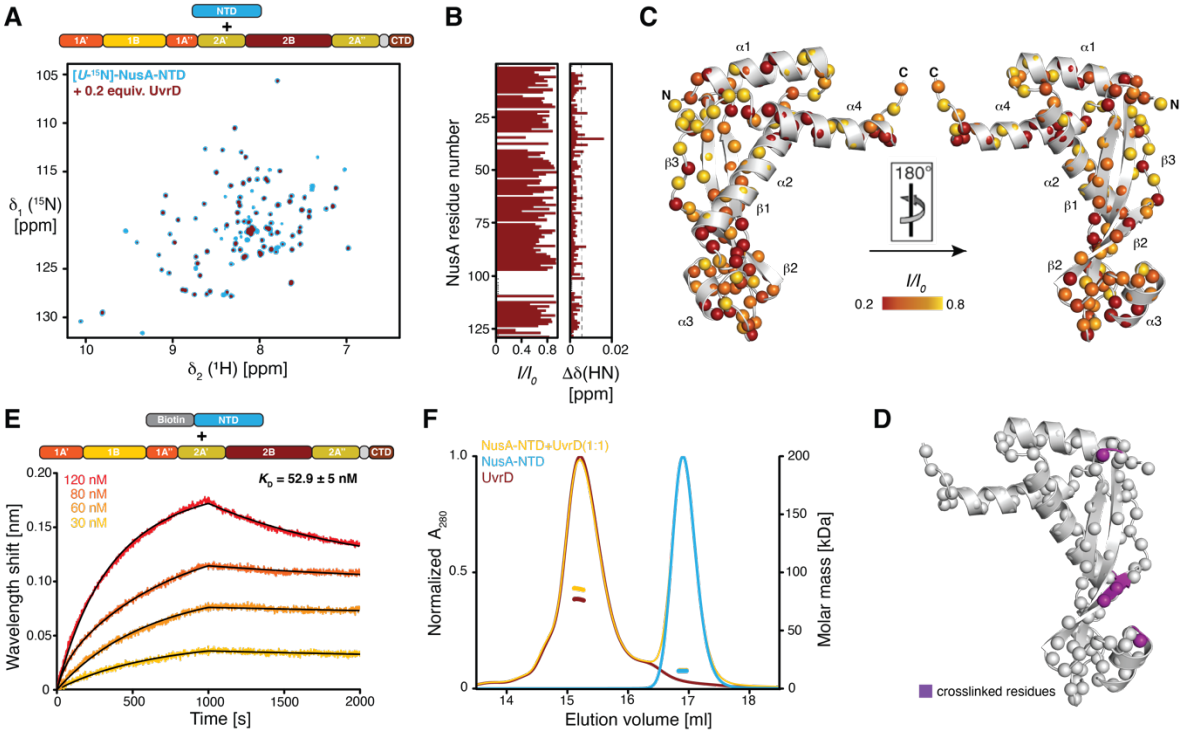

Figure 14: A) Overlay of 2D [^{15}N,^{1}H]-NMR spectra of [*U*-^{15}N] NusA-NTD in the absence (cyan) and after the addition of 0.2 equivalents of UvrD (dark-red) acquired in NMR-buffer at 298 K. **B**) The ratio of the individual peak intensities in the presence of 0.2 equivalents of UvrD to the apo NusA-NTD as well as the observed chemical shift perturbations plotted against the NusA-NTD residue number. **C**) Intensity changes upon UvrD interaction were plotted on the NusA-NTD (PDB: 2KWP) structure by the indicated color gradient. The amide moieties of the NusA-NTD construct are shown as spheres. **D**) Highlighting (purple) the previously determined cross-link positions on NusA side of the NusA:UvrD complex [54]. **E**) Bio-layer interferometry (BLI) data analysis of NusA-NTD binding to UvrD. Analyte concentrations are indicated. Non-linear least square fits to the experimental data are indicated by the black lines. **F**) SEC-MALS elution profiles of NusA-NTD (cyan), UvrD (dark-red), and NusA-NTD:UvrD (yellow). The profiles were recorded on ~1 mg/ml protein in PBS (pH7.4) buffer supplemented with 100 mM NaCl, 2 mM DTT, 1 Protease inhibitor tablet, at room temperature.

In agreement with the observation of a decreased off-rate, we also observed altered elution properties on SEC-MALS chromatogram. Whereas UvrD and NusAΔNTD eluted at separate peaks corresponding to roughly their expected molecular weights of ~77 kDa as well as 56 kDa for NusAΔNTD, which is actually 30% larger than the actual molecular size of 40 kDa. This off-set can likely be attributed to the possibly extended, open form of this NusA variant as both terminal domains, NTD as well as CTD, contribute to the closed, autoinhibited form of NusA [148]. In the NusAΔNTD:UvrD mix we observed a single peak in between the

elution peaks of the individual proteins (**Figure 15C**) indicative of an altered elution behavior of this particular complex on the chromatography column compared to the other NusA-variants tested. The corresponding molecular mass of this peak is ~84 kDa, which is in line with the partial formation of a NusAΔNTD:UvrD complex.

Using the NusA^{open} mutant and UvrD, we performed binding kinetic experiment by BLI, and determine a dissociation constant value of 2.74 ± 0.03 nM (**Figure 15F**). Comparing the kinetic rates showed that the obtained on-rate for this variant with 6.7 * 10^4 $M^{-1}s^{-1}$ is only slightly smaller than for the isolated NusA-NTD, but higher than for the other two variants, showing that the NTD binding determines the association. On the other hand, the obtained dissociation rate of 9.79*10^{-5} s^{-1} is in the same order of magnitude as for the NusAΔNTD-variant with 8.33 * 10^{-5} s^{-1}. Thus, the enhanced complex formation of NusA^{open} to UvrD compared to the wild-type NusA shows that the autoinhibitory feature modulates the interaction between NusA and UvrD.

Table 1: BLI-derived kinetic parameters of the interaction of UvrD with different biotinylated NusA-variants alongside the obtained quality parameters of the non-linear least square fitting.

	NusA-NTD	**NusA**	**NusAΔNTD**	**NusA^{open}**
K_D [nM]	52.9 ± 5	52.0 ± 10	4.59 ± 2.44	2.74 ± 0.3
k_{on} [1/Ms]	8.07 * 10^4	3.4 * 10^4	2.9 * 10^4	6.7 * 10^4
k_{off} [1/s]	1.87 * 10^{-4}	4.67 * 10^{-4}	8.33 * 10^{-5}	9.79 * 10^{-5}
χ^2	0.031	0.050	0.100	0.089
R^2	0.9902	0.9917	0.9941	0.98228

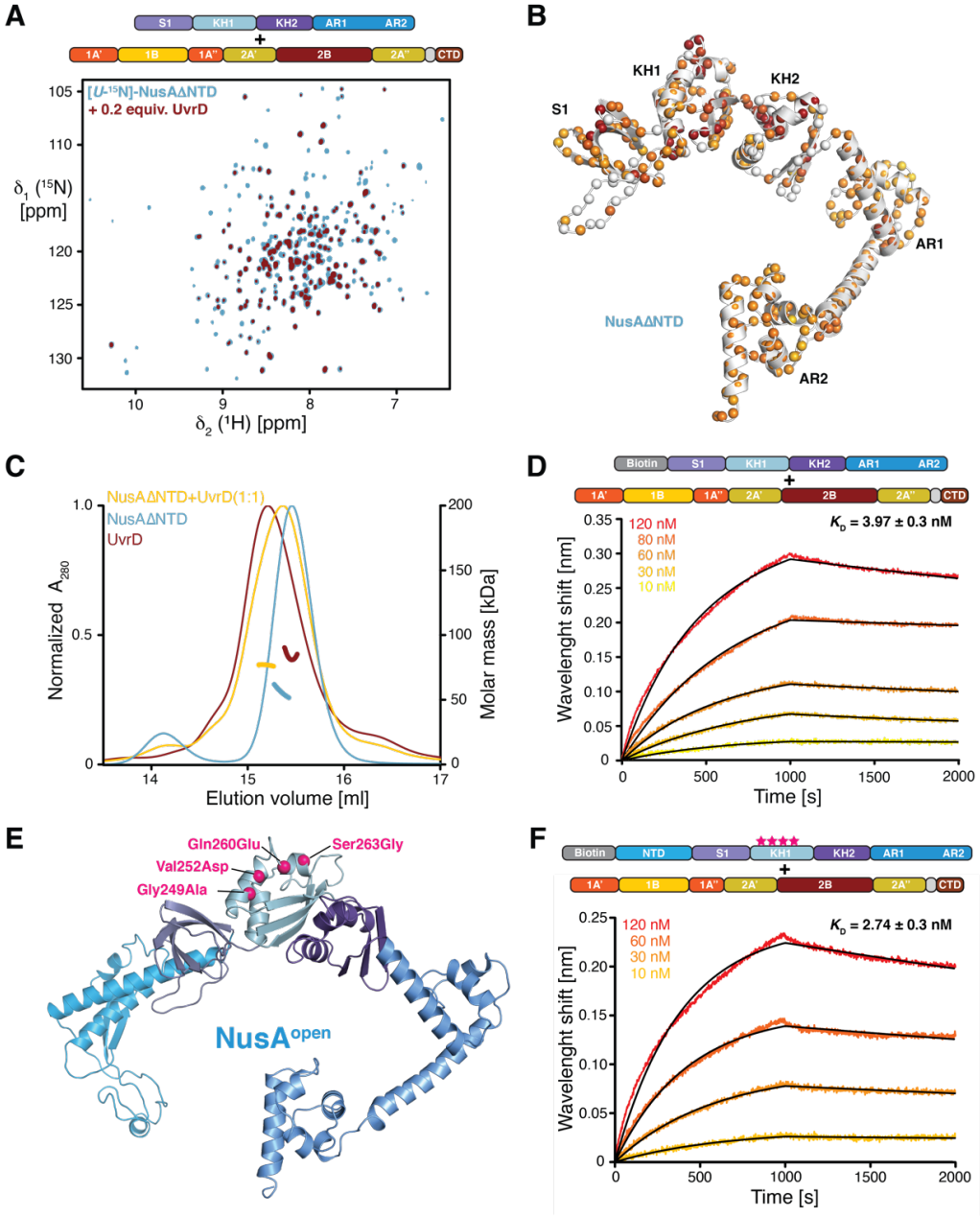

Figure 15: A) Overlay of 2D [^{15}N,^{1}H]-NMR spectra of [*U*-^{2}H,^{15}N] NusAΔNTD in the absence (blue) and after the addition of 0.2 equivalents of UvrD (dark red) acquired in NMR-buffer at 298 K. **B**) Intensity losses upon UvrD interaction were plotted on the NusAΔNTD structural model by the indicated color gradient. The amide moieties of the NusAΔNTD construct are shown as spheres. **C**) SEC-MALS elution profiles of NusAΔNTD (cyan), UvrD (dark-red), and NusAΔNTD:UvrD (yellow). The profiles were recorded on ~1 mg/ml protein in NMR buffer at room temperature. **D**) BLI data analysis of NusAΔNTD binding to UvrD. Analyte concentrations are indicated. Non-linear least square fits to the experimental data are indicated by the black lines. **E**) Position of the point mutations used in a previous study depicted by magenta spheres [148]. The four mutations Gly249Ala, Val252Asp,

Chapter IV: CONCLUDING REMARK AND FUTURE PERSPECTIVES

The year 2015 was named "The year of DNA repair" as it marked the recognition of DNA repair research following the award of the Nobel Prize in Chemistry 2015 to Tomas Lindahl, Paul Modrich, and Aziz Sancar for their pioneering studies in unraveling the molecular mechanism underlying cellular DNA repair. DNA damaging agents of either environmental or endogenous origin constantly threaten the stability and integrity of cellular genome across all domains of life. In this thesis, I studied, in detail, the structural basis of transcription-coupled repair (TCR), a process used by bacteria to overcome the lethal effect of DNA damage.

DNA repair pathways are highly regulated processes, and involve the recruitment of repair proteins to the lesion site within DNA strands. In bacteria, DNA lesions in the template strand stall progression of transcribing RNA polymerase (RNAP), which is modulated by a cascade of transcription elongation factors. One of these modulating factors, NusA, exhibits a self-regulated autoinhibitory phenomenon, in which the carboxy-terminally AR2 domain forms intramolecular interdomain interactions with the central SKK sub-domain, thereby hindering its activity [136]. In **paper I**, I completed both backbone and methyl groups resonance assignments of full-length NusA, which confirmed that the isolated wild-type protein is in the closed state in solution. In order to elucidate the dynamics of this state, I designed and produced an open NusA variant (called $NusA^{open}$), harboring four mutations in the KH1 domain. I used advanced high-resolution solution-state NMR spectroscopy and SEC-MALS to confirms that $NusA^{open}$ is in a different state to the wildtype NusA. I discovered for the first time that the flexible NusA-NTD plays supportive role in the autoinhibition by binding directly to the carboxy-terminal AR domains of NusA. I also investigated NusA dynamics and observed the complete absence of exchange processes on the μs–ms timescale, which points to a low population of the open state in solution. So far, the side-chain order parameter data already clearly shows that in the open state of NusA, the three main parts NTD, SKK, and CTD behaves as beads-on-a-chain with rigid body movement. The newly discovered role of NusA-NTD in NusA autoinhibition provides a clear direction for future structural model of free NusA in solution. The "open" NusA variant used in this study was created with multiple mutations on residues identified to be directly involved in the SKK:AR2 interaction, going forward we recommend a comparative study with an "open" variant formed by interaction complex with any NusA activators.

Active NusA, usually in complex with transcribing RNAP, mediates recruitment of translesion DNA polymerase IV (DinB). In **paper II**, I was able to complete ~95 % sequence-

specific resonance assignment of DinB for the first time, and thus elucidate the structural dynamics governing its DNA substrate accommodation in the active site. Using protein backbone relaxation in addition to CPMG relaxation dispersion experiments of side-chain methyl groups, I showed that the DNA-binding Thumb domain is structurally flexible, a crucial property for accommodation of bulky DNA lesions. Furthermore, binding kinetic data show that the main interaction surface for dsDNA substrate resides within the catalytic core region whilst the extended polymerase-associated domain (PAD) provides additional support to hold the substrate in place. I also showed for the first time that the stalled RNAP makes contact with DinB, and that the contact surface is within the catalytic core region. The catalytic core region consists of three domains, therefore further work is required to map the binding interface on both DinB and RNAP core.

The central aim of this thesis is to characterize the role of NusA in DNA repair in more details but it is necessary to understand the intrinsic solution properties of individual proteins NusA and DinB. With this initial aim achieved, I went further in **paper III**, to produce a detailed characterization of the interaction of these proteins in TC-TLS. This interaction had been proposed based on earlier observations using far-wester blots, but was lacking in structural details [59]. Using NMR titrations together with bio-layer interferometry (BLI) kinetic data, I confirmed NusA:DinB complex formation and map the interaction surface which is located at the carboxy-terminal end of both proteins. We observed that the autoinhibition of NusA reduces its affinity towards DinB. On the other hand, DinB affinity towards NusA further stabilizes in the presence of additional DinB domains. I propose that the NusA-CTD must first dissociate from the stalled RNAP order to recruit DinB. My data also suggests that after this initial contact an intertwined NusA-DinB is formed, which requires additional domains to forms stable complex. The exact role of the large RNAP-NusA-DinB complex in the context of a stalled transcription elongation complex remains to be elucidated in the future.

Lastly in **paper IV**, I investigated another recently reported role of NusA in DNA repair, which involves the UvrD helicase [53,54]. I confirmed that NusA interacts with UvrD, and that the interaction is mediated by NusA-NTD. I identified an additional binding interface within the central NusA S1-KH1-KH2 region, which is required to stabilize the interaction. This region is known to be involved in the NusA autoinhibition [44,49,148]. I observed reduced affinity of full-length NusA towards UvrD. Binding kinetic data showed that point mutations on the KH1 domain, which resulted in creation "open" NusA variant, enhances NusA affinity towards UvrD. Thus, my data provided novel insight into the interaction of NusA and UvrD,

which will be an important foundation for future studies of the interplay of NusA and UvrD in the coming years.

Altogether, this thesis shows that even though the basic principles and mechanisms of DNA repair are now characterized, and the basic mechanism are already textbook knowledge, there are still many open questions regarding the interplay of different proteins and machineries. Therefore, I strongly believe that my work on NusA and its partners in DNA repair has contributed in understanding the mechanism of this crucial processes. The work presented here highlights the future possibilities to apply different structural and biophysical methods to unravel yet uncharacterized biological mechanisms at an unprecedented level.

ACKNOWLEDGEMENTS

It is time to put a full stop to the most interesting and challenging project I ever embarked on. I resigned from my job to seek for something more challenging, but the initial plan was not to do PhD. At this point, I think it was a best decision to go back to university, and most especially, take the advice to do a PhD. The PhD experience which includes the encounter with people and the program activities is filled with memories, which will live with me forever. Therefore, you all have a place in my heart.

I cannot end without saying a word of appreciation to people I encountered during the entire period of program. First, I want to most sincerely thank my supervisor, **Björn**, for the opportunity to undergo PhD education under your supervision and for all support throughout this time. With you, I saw the statement: "Impossible is nothing" in practice.

I would like to thank my examiner immensely, **Gergely**, for all the support and friendliness. It was easy to have a chat with you anytime, anywhere.

To my co-supervisor, **Kristina**, even though you were not directly involved in my project, I felt your presence in other important aspect. Thank you so much for all the kind words and for sharing a lot of useful information.

My sincerely gratitude also goes to all the principal investigators in our division **Richard**, **Gisela**, **Sebastian**, **Julia**, **Johanna**, and **Örjan** for great decisions which have kept the division well organized and progressive. Lundberg laboratory remains a friendly working environment, that is remarkable.

To the Burmann group: **Irena**, I am extremely grateful for all the moral support and for always sacrificing your time to make sure my project is progressing. **Yosh**, thank you all the discussions and for valuable tips. **Emelie**, thank you for your friendliness and for sharing tips and tricks that worked for you in your project. **Ylber & Jens**, a gratitude to one is a gratitude to both, so, many thanks to you guys for sharing both fun time and expertise. **Hannah**, thanks a lot for your kindness. We should clone a Swedish word for "translesion". **Charles**, I am very thankful for your help in my project and for helping me to understand my experiments. I am also grateful for taking your time to proofread this thesis. **Filippo**, thank you for all your friendliness. I will soon be free from this thesis work. **Bozidar**, many thanks for your friendliness in the short time. And have fun with the incubators.

Andrea, it was so good that we are preparing our thesis at the same time. Many thanks for sharing value information with me. **Masoud**, thank you for sharing your PhD and post PhD experience with me. **Szabolcs**, thank you for all the small talks, that meant a lot to me. **Taru**,

thank you for giving me your Äkta time, it was so helpful. **Ann**, thank you for helping with lyophilizing my buffer. I got good result. **Filomena**, thank you for your friendliness. **Katharina**, thanks for the small talks, I appreciate your friendliness. **Vajradhar**, many thanks for all the chats in the kitchen, it was relaxing. **Ulrika**, it was fun when we did the dish washing duty together, thank you for making it easy. **Jessica**, thank you for all the smiles on the corridor, they meant a lot to me. **Laras**, many thanks for all the chats, it was always great. **Torbjörn**, I have only seen you twice, but we greeted each other, thank you for that. **Leona**, many thanks for the small talks, it was really good. **Beatrice**, thanks for the small talks, I appreciate your friendliness. **Owens**, many thanks for all the great chats and laughs, **Doris**, I hope you never recover that picture, thank you for your friendliness, **Jonatan**, many thanks for all the great chats, I appreciate your friendliness, **Gabrielle,** thank you for the nice smiles on the corridor, **Per**, many thanks for all the chats, I appreciate your friendliness**, Greger**, thank you for helping me one time with Prep 1 centrifuge, I still remember. **Giorgia**, thank you for your friendliness and for all the small talks. **Dimitra**, hope you are doing great with your baby, thank you for all the chats. **Analia**, hope you are doing great with your baby as well, thank you all the small talks. **Lucija**, I appreciate your kindness. **Adams**, thank you all the small talks. **Arpitha**, I appreciate your kindness. **Padmini**, I appreciate your kindness. **Bruno & Lars**, many thanks for all the computer-related assistance. It made the work a lot easier, I greatly appreciate it. **Valida**, I would like also extend my gratitude to you for making sure we have all material we need to do our research work.

The Swedish NMR Centre staffs **Göran**, **Zoltan, Ulrika, Cecilia, Anders, Vladislav**, **Arthur**, I most sincerely thank you all for time and support. I want to specially thank **Ashish** for his kindness, time and support through this PhD period. You put a lot of effort in this project in discussing and explaining my data with me. **Weixiao**, thank you for sharing very useful information.

It was not possible to mention everybody I have come across during this PhD period both within and outside the university including my neighbors, I sincerely thank you all.

REFERENCE

1. Chaffey N, Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular biology of the cell. 4th Ed. (2003).

2. Friedberg EC. Relationships between DNA repair and transcription. *Annual review of biochemistry*. **65**,15–42 (1996).

3. Norbury CJ, Hickson ID. Cellular responses to DNA damage. *Annual review of pharmacology and toxicology*. **41**, 367–401 (2001).

4. Carell T, Epple R. Repair of UV light induced DNA lesions: a comparative study with model compounds. *European journal of organic chemistry*. **1998**, 1245–1258 (1998).

5. Penning TM. Chemical Carcinogenesis. *Springer Science & Business Media*. (2011).

6. Thoma F. Light and dark in chromatin repair: repair of UV-induced DNA lesions by photolyase and nucleotide excision repair. *The EMBO journal*. **18**, 6585–6598 (1999).

7. Sancar A. Structure and function of DNA photolyase. *Biochemistry*. **33**, 2–9 (1994).

8. Kim ST, Malhotra K, Smith CA, Taylor JS, Sancar A. Characterization of (6-4) photoproduct DNA photolyase. *Journal of Biological Chemistry*. **269**, 8535–8540 (1994).

9. Sancar A. No" End of History" for photolyases. *Science*. **272**, 48–49 (1996).

10. Herndl GJ. Role of ultraviolet radiation on bacterioplankton activity. *The effects of ozone depletion on aquatic ecosystems*. 143–154 (1997).

11. Sinha R, Singh N, Kumar A, Kumar H, Häder M, Häder DP. Effects of UV irradiation on certain physiological and biochemical processes in cyanobacteria. *Journal of Photochemistry and Photobiology B: Biology*. **32**, 107–113 (1996).

12. Becker MM, Wang Z. Origin of ultraviolet damage in DNA. *Journal of molecular biology*. **210**, 429–438 (1989).

13. Lyamichev V. Unusual conformation of (dA) n·(dT) n-tracts as revealed by cyclobutane thymine–thymine dimer formation. *Nucleic acids research*. **19**, 4491–4496 (1991).

14. Britt AB. Repair of DNA damage induced by ultraviolet radiation. *Plant physiology*. **108**, 891 (1995).

15. Britt AB. DNA damage and repair in plants. *Annual review of plant biology*. **47**, 75–100 (1996).

16. Todo T. Functional diversity of the DNA photolyase/blue light receptor family. *Mutation research DNA repair*. **434**, 89–97 (1999).

17. Seeberg E, Eide L, Bjørås M. The base excision repair pathway. *Trends in biochemical sciences*. **20**, 391–397 (1995).

18. Sakumi K, Sekiguchi M. Structures and functions of DNA glycosylases. *Mutation Research/DNA Repair*. **236**, 161–172 (1990).

19. Truglio JJ, Croteau DL, Van Houten B, Kisker C. Prokaryotic Nucleotide Excision Repair: The UvrABC System. *Chem Rev*. **106**, 233–252 (2006).

20. Cohen SE, Walker GC. New discoveries linking transcription to DNA repair and damage tolerance pathways. *Transcription*. **2**, 37–40 (2011).

21. Kow YW, Wallace SS, Van Houten B. UvrABC nuclease complex repairs thymine glycol, an oxidative DNA base damage. *Mutation Research/DNA Repair*. **235**, 147–156 (1990).

22. Sancar A, Rupp WD. A novel repair enzyme: UVRABC excision nuclease of *Escherichia coli* cuts a DNA strand on both sides of the damaged region. *Cell*. **33**, 249–260 (1983).

23. Caron PR, Kushner SR, Grossman L. Involvement of helicase II (uvrD gene product) and DNA polymerase I in excision mediated by the uvrABC protein complex. *Proceedings of the National Academy of Sciences*. **82**, 4925–4929 (1985).

24. Husain I, Van Houten B, Thomas DC, Abdel-Monem M, Sancar A. Effect of DNA polymerase I and DNA helicase II on the turnover rate of UvrABC excision nuclease. *Proceedings of the National Academy of Sciences*. **82**, 6774–6778 (1985).

25. Lindahl T, Wood RD. Quality control by DNA repair. *Science*. **286**, 1897–1905 (1999).

26. Lehmann AR. Nucleotide excision repair and the link with transcription. *Trends in biochemical sciences*. **20**, 402–405 (1995).

27. Tornaletti S, Hanawalt PC. Effect of DNA lesions on transcription elongation. *Biochimie*. **81**, 139-146 (1999).

28. Witkin EM. Ultraviolet mutagenesis and inducible DNA repair in Escherichia coli. *Bacteriol Rev*. **40**, 869–907 (1976).

29. Ganesan A, Spivak G, Hanawalt PC. Transcription-coupled DNA repair in prokaryotes. *Progress in molecular biology and translational science*. **110**, 25–40 (2012).

30. Belogurov GA, Artsimovitch I. Regulation of transcript elongation. *Annual review of microbiology*. **69**, 49 (2015).

31. Bohr VA, Smith CA, Okumoto DS, Hanawalt PC. DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. *Cell*. **40**, 359–369 (1985).

32. McGlynn P. Linking transcription with DNA repair, damage tolerance, and genome duplication. *Proceedings of the National Academy of Sciences*. **107**, 15314–15315 (2010).

33. Selby CP, Sancar A. Molecular mechanism of transcription-repair coupling. *Science*. **260**, 53–58 (1993).

34. Park JS, Marr MT, Roberts JW. *E. coli* transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell*. **109**, 757–767 (2002).

35. Deaconescu AM, Chambers AL, Smith AJ, et al. Structural basis for bacterial transcription-coupled DNA repair. *Cell*. **124**, 507–520 (2006).

36. Deaconescu AM, Artsimovitch I, Grigorieff N. Interplay of DNA repair with transcription: from structures to mechanisms. *Trends in biochemical sciences*. **37**, 543–552 (2012).

37. Selby CP. Mfd protein and transcription–repair coupling in *Escherichia coli*. *Photochemistry and photobiology*. **93**, 280–295 (2017).

38. Witkin EM. Radiation-Induced Mutations and Their Repair: Bacteria reduce the mutagenic effects of ultraviolet light by repairing DNA damaged by the radiation. *Science*. **152**, 1345–1353 (1966).

39. Martinez B, Bharati BK, Epshtein V, Nudler E. Pervasive Transcription-coupled DNA repair in *E. coli*. *Nature communications*. **13**, 1–10 (2022).

40. Courcelle J, Khodursky A, Peter B, Brown PO, Hanawalt PC. Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics*. **158**, 41–64 (2001).

41. Yang X, Lewis PJ. The interaction between RNA polymerase and the elongation factor NusA. *RNA biology*. **7**, 272–275 (2010).

42. Eisenmann A, Schwarz S, Prasch S, Schweimer K, Rösch P. The *E. coli* NusA carboxy-terminal domains are structurally similar and show specific RNAP- and λN interaction. *Protein Science*. **14**, 2018–2029 (2005).

43. Ha KS, Toulokhonov I, Vassylyev DG, Landick R. The NusA N-Terminal Domain Is Necessary and Sufficient for Enhancement of Transcriptional Pausing via Interaction with the RNA Exit Channel of RNA Polymerase. *Journal of Molecular Biology*. **401**, 708–725 (2010).

44. Prasch S, Jurk M, Washburn RS, Gottesman ME, Wöhrl BM, Rösch P. RNA-binding specificity of *E. coli* NusA. *Nucleic Acids Research*. **37**, 4736–4742 (2009).

45. Worbs M, Bourenkov GP, Bartunik HD, Huber R, Wahl MC. An Extended RNA Binding Surface through Arrayed S1 and KH Domains in Transcription Factor NusA. *Molecular Cell*. **7**, 1177–1189 (2001).

46. Gusarov I, Nudler E. The Mechanism of Intrinsic Transcription Termination. *Molecular Cell*. **3**, 495–504 (1999).

47. Yang X, Molimau S, Doherty GP, et al. The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. *EMBO reports*. **10**, 997–1002 (2009).

48. Burmann BM, Rösch P. The role of *E. coli* Nus-Factors in transcription regulation and transcription. *Transcription*. **2**, 130–134 (2011).

49. Mah TF, Kuznedelov K, Mushegian A, Severinov K, Greenblatt J. The α subunit of *E. coli* RNA polymerase activates RNA binding by NusA. *Genes & development*. **14**, 2664–2675 (2000).

50. Schweimer K, Prasch S, Sujatha PS, Bubunenko M, Gottesman ME, Rösch P. NusA Interaction with the α Subunit of *E. coli* RNA Polymerase Is via the UP-Element Site and Releases Autoinhibition. *Structure*. **19**, 945–954 (2011).

51. Butland G, Peregrín-Alvarez JM, Li J, et al. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*. **433**, 531–537 (2005).

52. Cohen SE, Lewis CA, Mooney RA, et al. Roles for the transcription elongation factor NusA in both DNA repair and damage tolerance pathways in *Escherichia coli*. *Proc Natl Acad Sci USA*. **107**, 15517 (2010).

53. Epshtein V, Kamarthapu V, McGary K, et al. UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature*. **505**, 372–377 (2014).

54. Bharati BK, Gowder M, Zheng F, et al. Crucial role and mechanism of transcription-coupled DNA repair in bacteria. *Nature*. **604**, 152–159 (2022).

55. Schalow BJ, Courcelle CT, Courcelle J. Mfd is required for rapid recovery of transcription following UV-induced DNA damage but not oxidative DNA damage in *Escherichia coli*. *Journal of bacteriology*. **194**, 2637–2645 (2012).

56. Maluf NK, Fischer CJ, Lohman TM. A dimer of *Escherichia coli* UvrD is the active form of the helicase in vitro. *Journal of molecular biology*. **325**, 913–935 (2003).

57. Kawale AA, Burmann BM. UvrD helicase–RNA polymerase interactions are governed by UvrD's carboxy-terminal Tudor domain. *Communications biology*. **3**, 1–13 (2020).

58. Gwynn EJ, Smith AJ, Guy CP, Savery NJ, McGlynn P, Dillingham MS. The conserved C-terminus of the PcrA/UvrD helicase interacts directly with RNA polymerase. *PloS one*. **8**, e78141 (2013).

59. Cohen SE, Godoy VG, Walker GC. Transcriptional Modulator NusA Interacts with Translesion DNA Polymerases in *Escherichia coli*. *J Bacteriol*. **191**, 665 (2009).

60. Kottur J, Sharma A, Gore KR, et al. Unique structural features in DNA polymerase IV enable efficient bypass of the N2 adduct induced by the nitrofurazone antibiotic. *Structure*. **23**, 56–67 (2015).

61. Yang W, Woodgate R. What a difference a decade makes: Insights into translesion DNA synthesis. *Proc Natl Acad Sci USA*. **104**, 15591 (2007).

62. Boudsocq F, Kokoska RJ, Plosky BS, et al. Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication. *Journal of Biological Chemistry*. **279**, 32932–32940 (2004).

63. Yang W. An overview of Y-family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry*. **53**, 2793–2803 (2014).

64. Heltzel Justin M. H., Maul Robert W., Wolff David W., Sutton Mark D. *Escherichia coli* DNA Polymerase IV (Pol IV), but Not Pol II, Dynamically Switches with a Stalled Pol III* Replicase. *Journal of Bacteriology*. **194**, 3589–3600 (2012).

65. Brotcorne-Lannoye A, Maenhaut-Michel G. Role of RecA protein in untargeted UV mutagenesis of bacteriophage lambda: evidence for the requirement for the dinB gene. *Proceedings of the National Academy of Sciences*. **83**, 3904–3908 (1986).

66. Donnelly CE, Walker GC. Coexpression of UmuD'with UmuC suppresses the UV mutagenesis deficiency of groE mutants. *Journal of bacteriology*. **174**, 3133–3139 (1992).

67. Wagner J, Fujii S, Gruz P, Nohmi T, Fuchs RP. The β clamp targets DNA polymerase IV to DNA and strongly increases its processivity. *EMBO reports*. **1**, 484–488 (2000).

68. Walker G. Understanding the complexity of an organism's responses to DNA damage. *Cold Spring Harbor Laboratory Press*. **65**, 1–10 (2000).

69. Goodman MF. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annual review of biochemistry*. **71**, 17–50 (2002).

70. Hartley H. Origin of the word 'protein.' *Nature*. **168**, 244–244 (1951).

71. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology*. **8**, 91 (2008).

72. Bond SR, Naus CC. RF-Cloning. org: an online tool for the design of restriction-free cloning projects. *Nucleic acids research*. **40**, 209–213 (2012).

73. Mikolajczyk J, Drag M, Békés M, Cao JT, Ronai Z, Salvesen GS. Small ubiquitin-related modifier (SUMO)-specific proteases: profiling the specificities and activities of human SENPs. *J Biol Chem*. **282**, 26217–26224 (2007).

74. Kuo D, Nie M, Courey AJ. SUMO as a solubility tag and in vivo cleavage of SUMO fusion proteins with Ulp1. In: *Protein Affinity Tags*. *Springer.* 71–80 (2014).

75. Ramos JL, García-Salamanca A, Molina-Santiago C, Udaondo Z. Operon. In: Maloy S, Hughes K, eds. *Brenner's Encyclopedia of Genetics (Second Edition)*. *Academic Press*. 176–180 (2013).

76. Pelley JW. 16 - RNA Transcription and Control of Gene Expression. *W.B. Saunders.* 137–147 (2012).

77. Azatian SB, Kaur N, Latham MP. Increasing the buffering capacity of minimal media leads to higher protein yield. *Journal of biomolecular NMR*. **73**, 11–17 (2019).

78. Sattler M, Fesik SW. Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure*. **4**, 1245–1249 (1996).

79. Browne D, Kenyon G, Packer E, Sternlicht H, Wilson Dm. Studies of macromolecular structure by carbon-13 nuclear magnetic resonance. ii. specific labeling approach to the study of histidine residues in proteins. *Journal of the American Chemical Society*. **95**, 1316–1323 (1973).

80. Grzesiek S, Anglister J, Ren H, Bax A. ^{13}C line narrowing by ^{2}H decoupling in $^{2}H/^{13}C/^{15}N$-enriched proteins. Application to triple resonance 4D J connectivity of sequential amides. *Journal of the American Chemical Society*. **115**, 4369–4370 (1993).

81. Yamazaki T, Lee W, Revington M, et al. An HNCA pulse scheme for the backbone assignment of ^{15}N, ^{13}C, ^{2}H-labeled proteins: application to a 37-kDa Trp repressor-DNA complex. *Journal of the American Chemical Society*. **116**, 6464–6465 (1994).

82. Tugarinov V, Kay LE. An isotope labeling strategy for methyl TROSY spectroscopy. *Journal of biomolecular NMR*. **28**, 165–172 (2004).

83. Ayala I, Sounier R, Usé N, Gans P, Boisbouvier J. An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *Journal of biomolecular NMR*. **43**, 111–119 (2009).

84. Gelis I, Bonvin AM, Keramisanou D, et al. Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell*. **131**, 756–769 (2007).

85. Sprangers R, Kay LE. Probing Supramolecular Structure from Measurement of Methyl $^{1}H-^{13}C$ Residual Dipolar Couplings. *Journal of the American Chemical Society*. **129**, 12668–12669 (2007).

86. Kerfah R, Plevin MJ, Sounier R, Gans P, Boisbouvier J. Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Current Opinion in Structural Biology*. **32**, 113–122 (2015).

87. Costa S, Almeida A, Castro A, Domingues L. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli* the novel Fh8 system. *Frontiers in microbiology*. **5**, 63 (2014).

88. Scopes RK. Overview of Protein Purification and Characterization. *Current Protocols in Protein Science*. **00**, 1.1.1–1.1.6 (1995).

89. Janson JC. Protein Purification: Principles, High Resolution Methods, and Applications. *John Wiley & Sons*; (2012).

90. Uliyanchenko E. Size-exclusion chromatography—from high-performance to ultra-performance. *Analytical and bioanalytical chemistry*. **406**, 6087–6094 (2014).

91. Sahin E, Roberts CJ. Size-exclusion chromatography with multi-angle light scattering for elucidating protein aggregation mechanisms. *Therapeutic proteins*. 403-423 (2012).

92. Goyon A, Beck A, Colas O, Sandra K, Guillarme D, Fekete S. Evaluation of size exclusion chromatography columns packed with sub-3 μm particles for the analysis of biopharmaceutical proteins. *Journal of Chromatography A*. **1498**, 80–89 (2017).

93. Amartely H, Avraham O, Friedler A, Livnah O, Lebendiker M. Coupling multi angle light scattering to ion exchange chromatography (IEX-MALS) for protein characterization. *Scientific reports*. **8**, 1–9 (2018).

94. Wyatt PJ. Light scattering and the absolute characterization of macromolecules. *Analytica chimica acta*. **272**, 1–40 (1993).

95. Nettleship JE, Brown J, Groves MR, Geerlof A. Methods for protein characterization by mass spectrometry, thermal shift (ThermoFluor) assay, and multiangle or static light scattering. *Structural Proteomics*. 299–318 (2008).

96. Kwan TO, Reis R, Siligardi G, Hussain R, Cheruvara H, Moraes I. Selection of biophysical methods for characterisation of membrane proteins. *International Journal of Molecular Sciences*. **20**, 2605 (2019).

97. Tobias R, Kumaraswamy S. Biomolecular Binding Kinetics Assays on the Octet Platform. *Forte Bio. Appl*. **14**, 1–21(2021).

98. Wallner J, Lhota G, Jeschek D, Mader A, Vorauer-Uhl K. Application of Bio-Layer Interferometry for the analysis of protein/liposome interactions. *Journal of Pharmaceutical and Biomedical Analysis*. **72**, 150–154 (2013).

99. Bloch F. Nuclear induction. *Physical review*. **70**, 460 (1946).

100. Purcell EM, Torrey HC, Pound RV. Resonance absorption by nuclear magnetic moments in a solid. *Physical review*. **69**, 37 (1946).

101. Aue W, Bartholdi E, Ernst RR. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *The Journal of Chemical Physics*. **64**, 2229–2246 (1976).

102. Williamson MP, Havel TF, Wüthrich K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1 H nuclear magnetic resonance and distance geometry. In: *Nmr In Structural Biology: A Collection of Papers by Kurt Wüthrich*. 319–339 (1995).

103. Diercks T, Coles M, Kessler H. Applications of NMR in drug discovery. *Current opinion in chemical biology*. **5**, 285–291 (2001).

104. Aguirre C, Cala O, Krimm I. Overview of Probing Protein-Ligand Interactions Using NMR. *Current protocols in protein science*. **81**, 17-18 (2015).

105. Shapiro M, Wareing J. High resolution NMR for screening ligand/protein binding. *Current Opinion in Drug Discovery & Development*. **2**, 396–400 (1999).

106. Carlomagno T. NMR in natural products: understanding conformation, configuration and receptor interactions. *Natural product reports*. **29**, 536–554 (2012).

107. Harner MJ, Frank AO, Fesik SW. Fragment-based drug discovery using NMR spectroscopy. *Journal of biomolecular NMR*. **56**, 65–75 (2013).

108. Kathir KM, Kumar TKS, Yu C. Understanding the mechanism of the antimitogenic activity of suramin. *Biochemistry*. **45**, 899–906 (2006).

109. Jayaraman G, Krishnaswamy T, Kumar S, Yu C. Binding of nucleotide triphosphates to cardiotoxin analogue II from the Taiwan cobra venom (naja naja atra): elucidation of the structural interactions in the dATP-cardiotoxin analogue II complex. *Journal of Biological Chemistry*. **274**, 17869–17875 (1999).

110. Arunkumar AI, Srisailam S, Kumar TKS, et al. Structure and stability of an acidic fibroblast growth factor from *Notophthalmus viridescens*. *Journal of Biological Chemistry*. **277**, 46424–46432 (2002).

111. Maity S, Gundampati RK, Suresh Kumar TK. NMR methods to characterize protein-ligand interactions. *Natural Product Communications*. **14**, 1934578X19849296 (2019).

112. Wüthrich K. NMR with proteins and nucleic acids. *Europhysics News*. **17**, 11–13 (1986).

113. Inagaki F. Protein NMR Resonance Assignment. *Encyclopedia of Biophysics*. 2033–2037 (2013).

114. Bax A, Grzesiek S. Methodological advances in protein NMR. *Accounts of Chemical Research*. **26**, 131–138 (1993).

115. Kay LE. Nuclear magnetic resonance methods for high molecular weight proteins: a study involving a complex of maltose binding protein and beta-cyclodextrin. *Methods in enzymology*. **339**, 174–203 (2001).

116. Wüthrich K, Wider G. Transverse relaxation–optimized NMR spectroscopy with biomacromolecular structures in solution. *Magnetic Resonance in Chemistry*. **41**, 80–88 (2003).

117. Salzmann M, Pervushin K, Wider G, Senn H, Wüthrich K. TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci USA*. **95**, 13585 (1998).

118. Sattler M, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution. *Progress in nuclear magnetic resonance spectroscopy*. **34**, 93–158 (1999).

119. Pervushin K, Riek R, Wider G, Wüthrich K. Attenuated T_2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA*. **94**, 12366 (1997).

120. Lundström P, Vallurupalli P, Hansen DF, Kay LE. Isotope labeling methods for studies of excited protein states by relaxation dispersion NMR spectroscopy. *Nature protocols*. **4**, 1641–1648 (2009).

121. Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature*. **445**, 618–622 (2007).

122. Wider G. Technical aspects of NMR spectroscopy with biological macromolecules and studies of hydration in solution. *Progress in nuclear magnetic resonance spectroscopy*. **32**, 193–275 (1998).

123. Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. Cross-correlated relaxation enhanced $^{1}H-^{13}C$ NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *Journal of the American Chemical Society*. **125**, 10420–10428 (2003).

124. Jiang Y, Kalodimos CG. NMR studies of large proteins. *Journal of Molecular Biology*. **429**, 2667–2676 (2017).

125. Gardner KH, Kay LE. Production and incorporation of ^{15}N, ^{13}C, ^{2}H (^{1}H-δ_1 methyl) isoleucine into proteins for multidimensional NMR studies. *Journal of the American Chemical Society*. **119**, 7599–7600 (1997).

126. Tugarinov V, Kay LE. Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. *ChemBioChem*. **6**, 1567–1577 (2005).

127. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE. A robust and cost-effective method for the production of Val, Leu, Ile (δ_1) methyl-protonated ^{15}N-, ^{13}C-, ^{2}H-labeled proteins. *Journal of Biomolecular NMR*. **13**, 369–374 (1999).

128. Milbradt AG, Arthanari H, Takeuchi K, Boeszoermenyi A, Hagn F, Wagner G. Increased resolution of aromatic cross peaks using alternate ^{13}C labeling and TROSY. *Journal of biomolecular NMR*. **62**, 291–301 (2015).

129. Teilum K, Brath U, Lundström P, Akke M. Biosynthetic ^{13}C labeling of aromatic side chains in proteins for NMR relaxation measurements. *Journal of the American Chemical Society*. **128**, 2506–2507 (2006).

130. Tzeng SR, Pai MT, Kalodimos CG. NMR studies of large protein systems. *Protein NMR Techniques*. 133–140 (2012).

131. Battiste JL, Wagner G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry*. **39**, 5355–5365 (2000).

132. Tang C, Schwieters CD, Clore GM. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature*. **449**, 1078–1082 (2007).

133. Mittag T, Kay LE, Forman-Kay JD. Protein dynamics and conformational disorder in molecular recognition. *Journal of Molecular Recognition: An Interdisciplinary Journal*. **23**, 105–116 (2010).

134. Abragam A. The Principles of Nuclear Magnetism. *Oxford university press*. (1961).

135. Kawale AA, Burmann BM. Characterization of backbone dynamics using solution NMR spectroscopy to discern the functional plasticity of structurally analogous proteins. *STAR protocols*. **2**, 100919 (2021).

136. Bonin I, Mühlberger R, Bourenkov GP, et al. Structural basis for the interaction of *Escherichia coli* NusA with protein N of phage λ. *Proc Natl Acad Sci USA*. **101**, 13762 (2004).

137. Palmer 3rd A, Kroenke CD, Loria JP. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods in enzymology*. **339**, 204–238 (2001).

138. Lakomek NA, Ying J, Bax A. Measurement of ^{15}N relaxation rates in perdeuterated proteins by TROSY-based methods. *Journal of Biomolecular NMR*. **53**, 209–221 (2012).

139. Sun H, Kay LE, Tugarinov V. An Optimized Relaxation-Based Coherence Transfer NMR Experiment for the Measurement of Side-Chain Order in Methyl-Protonated, Highly Deuterated Proteins. *J Phys Chem B*. **115**, 14878–14884 (2011).

140. Weinhäupl K, Lindau C, Hessel A, et al. Structural Basis of Membrane Protein Chaperoning through the Mitochondrial Intermembrane Space. *Cell*. **175**, 1365–1379.e25 (2018).

141. Korzhnev DM, Kloiber K, Kanelis V, Tugarinov V, Kay LE. Probing Slow Dynamics in High Molecular Weight Proteins by Methyl-TROSY NMR Spectroscopy: Application to a 723-Residue Enzyme. *J Am Chem Soc*. **126**, 3964–3973 (2004).

142. Strauß M, Vitiello C, Schweimer K, Gottesman M, Rösch P, Knauer SH. Transcription is regulated by NusA: NusG interaction. *Nucleic acids research*. **44**, 5971–5982 (2016).

143. Dudenhoeffer BR, Borggraefe J, Schweimer K, Knauer SH. NusA directly interacts with antitermination factor Q from phage λ. *Scientific reports*. **10**, 1–15 (2020).

144. Dudenhoeffer BR, Schneider H, Schweimer K, Knauer SH. SuhB is an integral part of the ribosomal antitermination complex and interacts with NusA. *Nucleic Acids Research*. **47**, 6504–6518.

145. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein− protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*. **125**, 1731–1737 (2003).

146. Van Zundert G, Rodrigues J, Trellet M, et al. The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. *Journal of molecular biology*. **428**, 720–725 (2016).

147. Mishra S, Mohan S, Godavarthi S, Sen R. The interaction surface of a bacterial transcription elongation factor required for complex formation with an antiterminator during transcription antitermination. *Journal of Biological Chemistry*. **288**, 28089–28103 (2013).

148. Okeke DC, Lidman J, Burmann BM. Structural dynamics underlying NusA autoinhibition. *Manuscript*. (2022).