Analysis of NusA's Roles in DNA Repair and Damage Tolerance Pathways in *Escherichia coli*

By

Yimin Zhao

B.S. Chemistry
University of Utah, 2009

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN BIOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

March 2015 [June 2015]

Signature redacted

Signature of Author: .....
Department of Biology
March 31, 2015

Signature redacted

Certified by: .....
Graham C. Walker
Professor of Biology
Thesis Supervisor

Signature redacted

Accepted by: .....
Amy Keating
Professor of Biology
Chairman, Committee for Graduate Students

Analysis of NusA's Roles in DNA Repair and Damage Tolerance Pathways in *Escherichia coli*

By

Yimin Zhao

Submitted to the Department of Biology on March 31, 2015 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Biology

ABSTRACT

NusA has long known to be a *bona fide* transcription factor that functions in both termination and antitermination. However, its role in DNA repair and damage tolerance has not been explored until recently. Cohen *et al.* proposed for the first time that NusA functions as a coupling factor in transcription-coupled translesion synthesis (TC-TLS) and nucleotide excision repair (TCR) processes. A very recent paper by Nudler's lab further supports Cohen's NusA-dependent TCR model. The objectives of the work described here are to gain a deeper understanding of the precise mechanism NusA uses to facilitate DNA repair and damage tolerance.

Thesis Supervisor: Graham C. Walker
Title: Professor of Biology

ACKNOWLEDGEMENTS

I would like to sincerely thank my mentor Graham Walker for his unwavering support and guidance. Graham is not only an extremely talented scientist but also a great person. I will never forget his singing and his encouragement for us to do the same. I will always be grateful for the opportunity to pursue my graduate work in his lab.

I am particularly indebted to Steve Bell who has given me instrumental help and support throughout my graduate studies at MIT. I also want to thank Tania Baker, Mike Laub and Jing-Ke Weng for their input and suggestions as members of my thesis committee.

I would like to thank members of the Walker Lab that I have overlapped with, for their invaluable advice, assistance, kindness, stories, jokes and stimulating conversation. In particular I would like to thank Daniel Ferullo and Jon Penterman, whose advice and encouragement helped to guide through my project. Also to the other members of the Walker Lab team: Sanjay D'Souza, Mohammed Shabab, Markus Arnold, and my recent baymate Charley Gruber, for their thoughtful insights and conversations.

In addition I would like to acknowledge some of my classmates, for being great colleagues and friends. Especially, Theresa Gipson, David Koppstein, and Bena Chen.

I would not be where I am today without the unconditional love and support of my family. I would like to thank my parents for their constant supply of enthusiasm and encouragement.

Analysis of NusA's Roles in DNA Repair and Damage Tolerance Pathways in *Escherichia coli*

ABSTRACT

NusA has long known to be a *bona fide* transcription factor that functions in both termination and antitermination. However, its role in DNA repair and damage tolerance has not been explored until recently. Cohen *et al.* proposed for the first time that NusA functions as a coupling factor in transcription-coupled translesion synthesis (TC-TLS) and nucleotide excision repair (TCR) processes. A very recent paper by Nudler's lab further supports Cohen's NusA-dependent TCR model. The objectives of the work described here are to gain a deeper understanding of the precise mechanism NusA uses to facilitate DNA repair and damage tolerance.

INTRODUCTION

The highly conserved 55 kDa NusA protein has long known to be a core regulator of transcript elongation in bacteria and archaea[1,2]. Bacterial NusA interacts with elongating RNA polymerase (RNAP) and the nascent RNA transcript in ways that stimulate pausing and termination but that can be switched to anti-pausing and anti-termination by other accessory proteins[1-3]. For example, it plays a central role in the architecture of the phage λ N processive antitermination complex, and was therefore named N-utilizing substance[4-6]. Conversely, NusA has also been implicated in facilitating both Rho-dependent and Rho-independent termination processes[7,8]. NusA is composed of an NTD, three RNA-binding motifs S1, KH1, KH2, and two C-terminal acidic repeats (AR1, AR2)[9-12]. The NTD interacts with RNAP[3,13-15]. It is essential

for viability in wild-type *Escherichia coli*[4,5,15]. Here I investigate a novel function of NusA in DNA repair and damage tolerance pathways, specifically its role transcription-coupled translesion synthesis and nucleotide excision repair.

Translesion synthesis (TLS) is an important mechanism of DNA damage tolerance[16]. Complete TLS involves insertion of a nucleotide opposite an adducted base (i.e. lesion bypass) and extension from that position by several subsequent nucleotide additions[16,17]. TLS polymerases are conserved throughout all domains of life, with the majority being members of the Y-family of DNA polymerases[16,17]. DinB (DNA polymerase IV) is the most ubiquitous Y-family DNA polymerase, and one of the two translesion polymerases in *Escherichia coli (E. coli)*[18]. Previous work by Jarosz *et al.* has shown that a Δ*dinB* strain is sensitive to the antibiotic nitrofurazone (NFZ), which generates adducts at the N^2-position of deoxyguanosine (dG)[19,20]. If left unrepaired, this type of lesion blocks transcription. For example, the presence of N^2-furfuryl-dG (N^2-f-dG; a mimic of the major adduct formed by NFZ) on the transcribed strand completely blocks RNAP, stalling transcription at the -4 position relative to the lesion[21].

Later Jarosz *et al.* demonstrated that DinB was highly efficient at catalyzing accurate TLS over N^2-furfuryl-dG[19,20]. It inserts deoxycytidine (dC) opposite N^2-furfuryl-dG with 10-15 fold greater catalytic proficiency than opposite undamaged dG[19,20]. They also showed that DinB was even more proficient (≈ 25-fold relative to an undamaged control) at extending from dC opposite the N^2-furfuryl-dG lesion.[19,20] Given that DinB is present in considerable excess over the replicative DNA polymerase in both SOS-uninduced cells (250 DinB / 10-20 Pol III) and SOS-induced cells (2,500 DinB / 10-20 Pol III)[22], if DinB has such a remarkable ability to carry out TLS over certain N^2dG adducts, could there be classes of N^2-dG lesions (N^2-furfuryl-like) that are essentially invisible to DNA replication but might cause problems for RNA

polymerase during transcription ("stealth lesions")? As mentioned above, Cohen *et al.* showed that the N^2-furfuryl-dG lesion blocks transcription by *E. coli* RNA polymerase when present in the transcribed strand, so does a 14-nucleotide gap[21]. They also isolated RNAP mutants that display an altered sensitivity to NFZ[21]. In search for support for the idea of "stealth lesions", Godoy *et al.* unexpectedly identified NusA, long known to be a *bona fide* transcription factor, as being a potential interacting partner of DinB by DinB affinity column and far-Western blot[23]. Cellulose filter peptide arrays further showed NusA residues that could potentially interact with DinB directly[23].

Further studies on NusA showed that both Δ*nusA* and *nusA11* (G181D, temperature sensitive) were highly sensitive to NFZ at the permissive temperature, more so than Δ*dinB*, suggesting that NusA could play a previously unrecognized role in DNA repair or damage tolerance pathways[21]. Gene expression profiling suggests that this sensitivity is unlikely to be due to an indirect effect on gene expression affecting a known DNA repair or damage tolerance pathway[21]. In addition, the observation that $dinB^+$ is a multicopy suppressor of the temperature sensitivity of *nusA11* strain suggests a biological connection between NusA and TLS[23]. Together, these preliminary results led Cohen *et al.* to propose a model of transcription coupled translesion synthesis (TC-TLS)[23].

In principle, a process of TC-TLS could be helpful when transcription becomes stalled by gaps in the transcribed strand that are opposite lesions in the non-transcribed strand[24]. Such gaps can be caused by lesions that cannot be bypassed by the replicative DNA polymerase[24]. On the lagging strand, gaps are generated by replication resuming at the site of the next Okasaki fragment, while a replication restart can generate similar gaps on the leading strand; gaps formed in this manner are estimated to average about 1,000 nucleotides in length[25,26].

Alternatively, gaps opposite lesions could also be formed by UvrABC-dependent nucleotide excision repair if two lesions are very close together but on opposing strands or by UvrABC-dependent incisions during the repair of an intrastrand cross-link[27]. In the latter two cases, the gaps would be smaller, 12 to 13 nucleotides[27].

Cohen *et al.* hypothesized that if an RNA polymerase encounters one of these gaps in the transcribed strand opposite a lesion, it would stall[23]. In this case, NusA, which is associated with the RNA polymerase throughout the elongation phase of transcription, might then recruit a TLS polymerase to fill in the gap in the template strand[23]. Repairing the gap would permit the transcription of the gene by RNA polymerase[23].

The greater sensitivity of Δ*nusA* and *nusA11* strains to NFZ killing compared to a Δ*dinB* mutant suggests a role for NusA beyond the involvement in TC-TLS[21]. Interactions between NusA and UvrABC nucleotide excision repair complex provided insights into the additional role that NusA might be playing. First, a high-throughput protein interaction screen (tandem affinity purification 'TAP') suggested NusA as a potential interacting partner of UvrA[28], which Cohen *et al.* have confirmed by far-Western blotting[21]. Secondly, Δ*uvrA* strains are sensitive to NFZ, and Uvr-dependent NER is the predominant mechanism for processing NFZ-induced DNA damage in *E. coli*[21,29]. Epistasis analysis with respect to NFZ sensitivity of *nusA11* and Δ*uvrA* alleles revealed that Δ*uvrA* was largely epistatic to *nusA11*, suggesting that NusA plays a role in UvrA-dependent process[21]. Additionally, Δ*uvrB* and Δ*uvrC* alleles are similarly epistatic to *nusA11* with respect to sensitivity to NFZ[21]. Together, these observations led Cohen *et al.* to propose for the first time a transcription-coupled nucleotide excision repair (TCR) model that involves NusA[21]. In this model, NusA recruits nucleotide excision repair (NER) machinery to an RNAP that has been stalled by an NFZ-induced lesion on the transcribed strand[21].

In bacteria, the general NER pathway commences when UvrA and UvrB proteins bind damaged DNA and recruit UvrC to cleave the damaged strand on both sides of the lesion[30-32]. The first incision is made at the fourth or fifth phosphodiester bond 3' to the damage, and the second incision, eight phosphodiester bonds 5' to the damage[30-32]. These cleavages create a 12-13-residue-long single-stranded DNA segment, which is displaced by UvrD helicase[30-32]. DNA polymerase I then fills the gap using the complementary strand as a template[30-32]. NER rates are usually greatest at transcriptionally active genes[30-32]. Moreover, the transcribed DNA strand is preferentially repaired compared to the non-transcribed strand[30-32]. This phenomenon, known as transcription-coupled repair (TCR), is a sub-pathway of global NER[30-32].

There is a well-characterized Mfd-dependent TCR model that is different from the NusA-dependent TCR model proposed by Cohen *et al*[33]. The Mfd model postulates that a DNA lesion blocking the progression of the transcription elongation complex is shielded from NER enzymes by the stalled RNAP[34,35]. The DNA translocase, Mfd, is proposed to bind to the stalled elongation complex through the β subunit of RNAP and to dislodge the complex by 'pushing' it forward[34,36]. Concurrently, Mfd recruits UvrA to the exposed lesion site to expedite NER[34,36].

However, very recently the Mfd model has been challenged by Evgeny Nudler's group[37]. They have proposed an alternative TCR model whose key component is UvrD helicase, which translocates in a 3' to 5' direction using a single-strand, DNA-dependent, ATPase activity[37,38]. In contrast to Mfd, UvrD facilitates NER by pulling RNAP backward from the DNA lesion without causing termination[37]. Their studies also showed evidence that supports the NusA-dependent TCR model: 1. UvrD and NusA map to close proximity on the surface of RNAP; 2. NusA facilitates UvrD-mediated backtracking in vitro; 3. Deletion of *greB* - a transcript cleavage factor which suppresses RNAP backtracking - suppresses sensitivity of

nusA cells (Δ*nusA* and *nusA11*) to NFZ, 4-NQO, mitomycin C, and ultraviolet (UV)[37]. To summarize, in Nudler's model RNAP recruits the NER complex via UvrD/NusA to the damage site[37].

I am interested in dissecting NusA's functions in DNA repair and damage tolerance pathways into more mechanistic details. How does NusA interact with DinB and the NER machinery? Does NusA function differently upon encountering different types of lesions? If so, how? And how does it cooperate with UvrD to initiate backtracking when RNAP is blocked by a lesion?

RESULTS

Physical interactions between NusA and TLS / NER components

My first approach was to demonstrate direct interactions between NusA and TLS / NER components, which could provide strong support to the proposed NusA-dependent models. NusA-DinB was the first pair to test because NusA was found on a DinB affinity column, and *dinB*$^+$ was a multicopy suppressor of the temperature sensitivity of *nusA11* strain[23]. NusA-UvrA was also a possibility because UvrA is the first NER protein complex to assemble on the damage site[32]. Previous work of 'TAP' and far-Western blot have both shown an interaction[21,28]. NusA-UvrD was another interesting one to test because Nudler's group has proposed that UvrD and NusA cooperate in backtracking-mediated NER[37].

A quick way to test potential interactions *in vivo* is via bacterial two-hybrid (BACTH)[39]. The BACTH system I used is based on the reconstitution of the adenylate cyclase activity in *E. coli*[39]. The catalytic domain of this protein is composed of two sub-domains: a 25 kDa fragment (T25) and an 18 kDa fragment (T18)[39]. When these two domains are produced separately, no

cAMP is produced[39]. However these two fragments can be fused separately to interacting polypeptides, X and Y[39]. And hetero-dimerization of these hybrid proteins results in functional complementation between T25 and T18 fragments and, therefore, cyclic AMP (cAMP) synthesis[39]. cAMP further binds to the catabolite activator protein, CAP[39]. The cAMP/CAP complex turns on the expression of the *lac* operon, and can be easily assayed on selective media[39].

Using this BACTH system, so far I have confirmed the NusA-DinB interaction in unstressed conditions, giving support to the TC-TLS model (Figure 1). NusA is associated with all elongating RNAP[3,15], so how could DinB be specifically recruited to RNAP that have been stalled by a gap in the transcribed strand? It is possible that NusA undergoes a conformational change upon the stalling of RNAP, and it is this conformation that DinB can now be recruited. One could also envision a situation where DinB is always associated with elongating RNAP, through an interaction with NusA. In this case, every time the RNAP encounters a gap opposite a lesion, DinB would already be present to fill in this gap. However, there is not likely to be enough DinB to always be associated with all elongating RNA polymerases.

In addition, I have also observed a NusA-UvrD interaction in the same conditions (Figure 1). Given NusA's interaction with RNAP, this interaction suggests that NusA could potentially act as a "coupling factor" that connects UvrD /NER components to the transcription machinery when RNAP encounters a lesion. This NusA-UvrD interaction is intriguing because Nudler has proposed that UvrD backtracks RNAP during blockage[37]. What signals UvrD to start backtracking? Could it be NusA? Is there another protein involved in the signaling process? To answer these questions, further characterization of the NusA-UvrD interaction will be needed.

On the other hand, I have not been able to demonstrate a NusA-UvrA interaction. Unexpectedly I have identified a NusA-NusA self-interaction (Figure 1). This contrasts with published NusA crystal structures in which it has been shown as a monomer[10-12]. It is important to know whether NusA self-interact under certain conditions, e.g. when there is DNA damage. Therefore the discrepancy needs to be resolved via other methods such as size-exclusion or sedimentation through gradients. In addition, it will also be interesting to test all interactions in stressed conditions, e.g. adding NFZ to the plates or irradiate them with UV, to see whether they affect the interactions.

The apparent discrepancy between previous work on NusA-UvrA interaction and my bacterial two-hybrid results could be explained by the fact that UvrA must be a dimer to be functional during the NER process[30-32]. It is possible that UvrA needs to dimerize before binding to NusA. Since UvrA may not be in its native state when fused to BACTH vectors, its ability to interact could be disrupted, thereby accounting for the lack of a positive interaction with NusA. In addition, neither 'TAP' nor BACTH rules out the possibility of an indirect interaction through a 'third protein'[40]. This discrepancy could be resolved using other methods such as co-immunoprecipitation (Co-IP) and FRET. Purified proteins needs to be used in both of these experiments so that a positive signal would suggest a direct interaction. The positive NusA-UvrD interaction observed via BACTH needs to be validated and characterized further. This could be done using the same Co-IP and FRET techniques. To double confirm Co-IP results, one can use an antibody against NusA to attempt to pull down UvrA/UvrD, as well as reciprocally use antibodies against UvrA / UvrD to attempt to pull down NusA. A positive *in vitro* FRET signal using purified proteins strongly suggests a direct interaction because energy transfers between fluorophores are very sensitive to small changes in distance and in practice is

not detectable at >10 nm[40]. For example, NusA can be fused to cyan fluorescent protein (CFP), which can be excited with 440-nm wavelength light, fluoresces and emits light at 480 nm. In addition, UvrA can be fused to yellow fluorescent protein (YFP), which absorbs the 480-nm light and emits light at 535 nm. Both the N-terminus and the C-terminus of NusA and UvrA/D can be fused to fluorophores, and one should test all the above since there is little information about how they might interact at this point.

On the other hand, if no interaction could be observed between NusA and UvrA/D via *in vitro* Co-IP or FRET, there is still the possibility that they may interact only upon damage induction. To test this one can run similar Co-IP assays with NFZ or UV damaged cell lysates. RNAP will be expected to show up since NusA is a *bona fide* transcription factor. Any other protein identified in this assay could potentially be involved in NusA-mediated processes. Further characterizations will be needed.

Isolation and characterization of *nusA* separation-of-function mutants

NusA is essential for viability in wild-type *E. coli*, thus a *nusA* deletion (Δ*nusA*) in normal background (e.g. AB1157) does not survive[4,5,15]. However Δ*nusA* can be introduced into a specialized genetic background strain MDS42, which is deleted for horizontally transferred genes including recombinogenic or mobile DNA and cryptic virulence genes[41,42]. The MDS42 Δ*nusA* strain grows much more slowly than a wild type MDS42 strain in unstressed conditions[42]. On the other hand, *nusA11* (ts) is viable at the permissive temperature, and exhibits good growth profiles in a normal background such as AB1157[43]. As mentioned previously, at the permissive temperature, *nusA11* (ts) in AB1157 is highly sensitive to NFZ; it is also very sensitive to another DNA-damaging compound, 4-Nitroquinoline 1-oxide (4-NQO,)

under the same conditions[21]. These phenotypes suggest that in a *nusA11* (ts) mutant, NusA's function in viability is genetically separable from its function in DNA repair, specifically, in withstanding NFZ/4-NQO induced DNA damage (Figure 2).

Cohen *et al.* initially observed that a Δ*nusA* mutant in MDS42 background was sensitive to NFZ and 4-NQO[21]. They later found that it was additionally sensitive to MMS, and UV, whereas a *nusA11* mutant was not, in either the MDS42 or the normal AB1157 background[21]. NFZ and 4-NQO primarily generate lesions on the N^2-position of guanine[44,45]. During transcription the stalling of RNAP at the -4 position relative to N^2-f-dG places the lesion in the minor groove of the double stranded DNA (dsDNA) ahead of the transcription bubble[21]. This contrasts with other types of transcription blocking lesions such as UV generated pyrimidine dimers, which enter the active site of RNAP[46]. Cohen *et al.* have proposed that the stalling of RNAP at such lesions could aid in their recognition, so that repair proteins could subsequently be recruited via an interaction with NusA[21,47]. Therefore the selective sensitivity of *nusA11* allele to NFZ and 4-NQO, but not to MMS or UV, represents a partial loss-of-function mutation which affects NusA's ability to participate in the repair of N^2-dG type of lesions that is genetically separable from its ability to help recognize other classes of lesions (Figure 2).

To summarize, it is possible to isolate separation-of-function mutants that are viable but sensitive to all DNA damaging agents mentioned above (NFZ, 4-NQO, MMS, UV). Alternatively it is also possible that UV and MMS induced lesions necessitate a different function of NusA to aid in their repair during transcription, hence mutants that confer sensitivity to UV or MMS but resistance to NFZ and 4-NQO (opposite to *nusA11*). To conduct this screen, I mutagenized a $nusA^+$ plasmid with XL1-Red strain, transformed the mutant plasmids into MDS42 Δ*nusA,* and then replica plated onto respective plates with or without

NFZ, 4-NQO, UV, and MMS. Using this method, I screened for mutant clones that confer sensitivity to UV or MMS, and subsequently tested whether they are sensitive or resistant to NFZ/4-NQO. Using a standard genetic background where NusA is essential, I then tested whether the new *nusA* mutants retained the functions required for viability and were recessive (similar to *nusA11*) or dominant.

NusA is a relatively small protein with 1500 base pairs. It would be reasonable to hit every base on average for at least 10 times, hence to screen for 15000 mutants, within 7 to 8 screening procedures. If the afore-mentioned mutants cannot be isolated after that, it could be an indication that such mutants may not exist. This will further suggest that NusA's function in the repair of UV-induced damages cannot be separated from its function in viability, which means that NusA functions differently in the repair of different types of damages, e.g. N^2-furfuryl-dG vs. pyridmine dimers.

So far I have screened for approximately 6,000 mutant colonies. By screening for NFZ and UV sensitivity, I have obtained one mutant that has an I97V mutation. This mutant grows as slowly as Δ*nusA* in unstressed conditions, and seems to have a dominant interference phenotype (Figure 3). Although not a separation-of-function mutant, it is more sensitive to both NFZ and UV than Δ*nusA*. It also appears that I97V impairs both growth and colony morphology even in the absence of any exogenous stress. Concerned about I97V being toxic, and suppressor or reversion arising from selective pressure on cells, I re-cloned this mutant gene into a plasmid with an inducible/repressible promoter in order to turn the toxic gene on at will and control when the cells are exposed to that stress. I also wanted to see whether it is a dominant gene. It will be interesting to test its sensitivity in other DNA damaging agents such as 4-NQO and MMS. In addition, I97V maps to the interface between the N-terminal globular

head domain and the helical body domain[12]. The N terminal domain of NusA interacts with RNAP[10,12,14].

After obtaining separation-of-function mutants as previously described, along with *nusA11*, these mutants can be tested in the bacterial two-hybrid assays to see how they affect NusA's interaction with UvrD/A, as well as its dimerization. One can also test these in both unstressed and stressed conditions. Depending on where the mutants are on NusA's crystal structure, these assays could give mechanistic clues to how NusA interacts with UvrD/A.

The mutants can also be tested in a backtracking assay described in Nudler's Nature paper where they showed that wild type NusA facilitates UvrD-mediated backtracking[37]. This type of assay allows controlled stepwise transcription in solid phase 'walking' along a DNA template, and isolation of defined intermediates in the transcription cycle[48]. The principle behind solid-phase 'walking' is that the initial Elongation Complex (EC) immobilized onto a solid support undergoes rounds of washing to remove the unincorporated nucleotide triphosphates (NTP) substrates, followed by addition of the incomplete set of NTPs (three or less) that allows transcription to proceed to the next DNA position corresponding to the first missing NTP ('chase')[48]. Each component of Elongation Complex, e.g. RNAP, DNA, or RNA, can be tagged for 'walking'[48].

One would expect that the mutant forms of NusA have decreased abilities to participate in UvrD-mediated backtracking in a non-damaged template. Lesions in the DNA template could potentially enhance backtracking. It will be interesting to see whether and how different types of mutants behave differently when encountering different types of damages.

An UvrABC incision assay is also relevant and interesting to answer some of the mechanistic questions of NusA-dependent TCR[34]. It was used to demonstrate that Mfd alone is

capable of stimulating the rate of incision of the transcribed strand[34]. In the original assay, Selby and Sancar labeled a 5.3-kbp DNA fragment at the 3' end of the transcribed strand[34]. The DNA was then irradiated with UV, and a transcription-repair reaction was performed[34]. Without RNAP, the DNA was nicked at many specific sites as a result of UvrABC incision of photoproducts produced at dipyrimidine hot spots[34]. When RNAP was present in the repair reaction, incision was inhibited[34]. However when Mfd was added, the inhibition disappeared and the rate of the incision was stimulated[34].

To provide additional evidence for the involvement of NusA in TCR, one could run the same assay replacing Mfd with wild type or mutant forms of NusA. The wild type NusA and the NusA11 will be expected to have a similar effect on the incision process as Mfd, while the reverse type of separation-or-function mutant (sensitive to UV but resistant to NFZ) will not. Along with the previous two assays, the phenotypes of the mutants and their positions on NusA's crystal structure will give mechanistic insights to NusA's interactions with NER machinery, and how these interactions affect the repair of different types of DNA damages.

Loss of transcription-coupled repair for N^2-furfuryl-dG adduct in Δ*uvrA* or *nusA11* mutants.

In addition to the previous two approaches, another strong support of the proposed NusA-dependent TCR model would be to show that such repair is lost in either a *nusA11* mutant strain or in a Δ*uvrA/B/C/D* mutant strain. Since the NER machinery uses the homologous strand as its template to direct repair, site-specifically modified double stranded DNA (dsDNA) would be required as a proper substrate[32].

John Essigmans's group have long used site-specifically modified single strand-M13 genomes to study the repair of DNA lesions by proteins that work on single-stranded DNA

(ssDNA) by a direct reversal of base damage mechanism, such as Ada, Ogt, and AlkB[49]. However, dsDNA containing a single adduct has not been utilized in parallel DNA repair studies because of the technical issues and potential challenges of discriminating between the lesion-containing and the lesion-free strands[50].

It was not until recently that Robert Fuchs' group developed a method to insert a plasmid containing a single lesion at a precise chromosomal location in *E. coli*[50]. They have used this method to study bypass of a single replication-blocking lesion by damage avoidance in the *E. coli* chromosome[50]. The experimental system, entails two major components: a recipient *E. coli* strain with a single attR site and a non-replicating plasmid construct containing the single lesion of interest and an attL site and an ampicillin resistance gene[50]. The recombination reaction between attL and attR is controlled by ectopic expression of phage lambda int-xis proteins, and leads to the integration of the lesion-containing vector into the chromosome[50]. Integrants are selected on the basis of their resistance to ampicillin[50]. The chromosomal integration region carries the 3'-end of the *lacZ* gene fused to attR, whereas the remaining 5'-end is located on the incoming fragment in fusion with attL, so that precise integration restores a functional β-galactosidase gene (*lacZ*)[50]. The non-damaged opposite strand contains a short sequence heterology that inactivates the *lacZ* gene and serves as a genetic marker that allows strand discrimination[50]. Precise integration of the vector at nucleotide resolution could be shown by the presence of only LacZ+ colonies following the integration of a lesion-free construct[50]. Individual integration events can be analyzed by PCR[50].

We have been communicating with the Essigmann's group about making a similar dsDNA construct containing a single N^2-furfuryl-dG lesion where we can control transcription

with an inducible promoter. Such a construct would allow us to demonstrate loss of transcription-coupled repair for N^2-furfuryl-dG in *nusA11* or *ΔuvrA/B/C/D* mutants.

CONCLUSION:

What role would NusA have in TC-TLS and recruiting TLS polymerases?

If the proposed mechanism of TC-TLS were to exist, there would likely need to be several other proteins in addition to RNAP, NusA and TLS polymerases involved in this process. In order for TLS polymerases to fill in a gap opposite a lesion, it would need to reach the opposite end of the gap relative to RNAP. These gaps could be anywhere from 12-1000 nucleotides in length (discussed in Introduction), and thus there would likely need to be a hand-off of the TLS polymerase from NusA / stalled RNAP to the appropriate primer / termini end.

NusA is classified as a general transcription elongation factor and has been found to be associated with all elongating RNA polymerases. Therefore one can come up with two ideas for the potential role for NusA in the model of TC-TLS. One idea is that since NusA is associated with all elongating RNA polymerases, it is a platform, allowing for recruitment and / or binding of TLS polymerases to the site of the stalled RNAP. This idea would not require NusA's ability to function as a transcriptional terminator / antiterminator. Alternatively, one could hypothesize that NusA, using its abilities to pause or stall elongating RNA polymerases, could sense the upstream gap in the transcribed strand and pause the RNAP before it encounters a situation in which it has run out of template. In this situation, NusA would also function as a "platform" by recruiting TLS DNA polymerases, however it would also utilize its own functions as a transcriptional terminator / antiterminator.

NusA-dependent TCR, an alternative to Mfd-dependent TCR

The proposed model for a new class of transcription-coupled repair suggests that NusA can recruit NER machinery to sites of RNAP stalled by a lesion in the transcribed strand in a fashion that is independent of Mfd. Why the existence of a second class of TCR? One intriguing possible explanation involves a particular class of DNA lesion, such as an N^2-dG lesion, which can be efficiently bypassed by DinB and thus is not a hindrance during replication. However, such lesions would then cause a problem when transcription is attempted. In this case, a novel class of TCR involving NusA could be important for the recognition and removal of these DNA lesions. Since some of these minor groove lesions are not efficiently recognized by normal NER, NusA-TCR could help a cell remove this class of lesions from its DNA.

The C-terminal half of Mfd is comprised of seven helicase domains and a translocase domain, which are important for the translocation of the stalled RNA polymerase. NusA however, does not contain such motifs, and thus it is likely that one or more factors need to be involved in translocating the stalled RNAP in this new class of transcription-coupled repair, for example, UvrD. The alternative TCR model proposed by Nudler states that UvrD helicase facilitates DNA repair by pulling RNAP backwards, and NusA potentiates UvrD-mediated backtracking. On the other hand, excessive backtracking can be detrimental to genomic integrity in cells recovering from genotoxic stress and resuming replication, as frequent co-directional collisions between the replisome and backtracked elongation complexes result in dsDNA breaks (DSBs). By 'pushing' backtracked RNAPs forward, Mfd suppresses DSBs associated with such collisions, and hence diminishes the high frequency of mutations

associated with DSBs repair. Overall, Nudler's studies gave both physical and intellectual support to our NusA-dependent TCR model.

MATERIALS AND METHODS

Bacterial strains. The strains and plasmids used in this study are listed in Table 1 and were constructed using standard genetic techniques.

DNA damage sensitivity assays. *E. coli.* strains were diluted in LB medium overnight. Appropriate dilutions were spread on LB agar plates containing increasing concentrations of NFZ (0-15 μM), 4-NQO (0-15 μM), MMS (0-15 μM). For UV survival assays, cells were plated on LB agar and then irratiated with UV light (0~40 J/m^2) by using a G15T8 UV lamp (GE) at 254 nm, then incubated in the dark. A concentrated stock solution of NFZ or 4-NQO was first made in N,N-dimethylformamide, stored at -20°C, then diluted appropriately for each experiment. Percent survival was determined relative to growth in the absence of DNA-damaging agent.

Bacterial two-hybrid assays. Detection of in vivo interactions between two proteins of interest with the BACTH system requires the co-expression of these proteins as fusions with the T25 and T18 fragments in bacteria that are lacking endogenous adenylate cyclase acitivity (E. coli cya). This is achieved by using two compatible vectors (BACTH kit), one expressing the T25 fusion (pKT25 or pKNT25), the other one expressing the T18 fusion (pUT18 or pUT18C). Genes of interest were cloned into these vector plasmids. The compatible recombinant plasmids were co-transformed into BACTH strain BTH101, and plated on either indicator or selective media to reveal the resulting Cya^+ phenotype. In *E. coli*, expression of the lacZ gene encoding β-galactosidase is positively controlled by cAMP/CAP. Hence, bacteria

expressing interacting hybrid proteins will form blue colonies on rich LB medium in the presence of the chromogenic substrate X-GAL (40 μg/ml), while cells expressing non-interacting proteins will remain white (pale blue). Carbenicillin can be used in replacement of ampicillin to reduce the frequency of incorrect white colonies. And IPTG (0.5 mM) can be included in the medium to increase β-galactosidase expression. By using these rich indicator media functional complementation can be detected within 24~72 hours at 30°C. Overcrowding of indicator plates should be avoided, e.g. maximum 300 ~ 500 colonies per plate, otherwise the detection of positive clones might be difficult.

***nusA* separation-of-function screen.** Three controls were obtained as the following: 1. Negative control, empty vector *pBR322* (Amp^R) transformed into MDS42 Δ*nusA*::Cm^R (made competent cells). 2. Positive control, wild type *pnusA*$^+$, *pNAG2010 (pBR332-nusA*$^+$, Amp^R) transformed into MDS42 Δ*nusA*::Cm^R (made competent cells). 3. *nusA11* control, *pBR332* (Amp^R) transformed into MDS42 *nusA11* Tn10 IQ419 (made competent cells). Mutant *nusA** pools were created by mutagenizing the wild type *pnusA*+ plasmid (*pNAG2010)* in XL1-Red strain. The mutant plasmid pools were extracted and transformed into MDS42 Δ*nusA*::Cm^R (made competent cells), and plated onto Amp^R plates. They were grown at 37°C for approximately 36 hours. Mutant *nusA** colonies were then picked individually by hand, and re-inoculated (LB and glycerol) into 96 well tissue culture plates. These were again grown overnight at 37°C, and can be kept sterile and frozen at -80°C as "frozen stock". Before the plates were put into -80°C, they can be replica plated onto respective LB plates with or without NFZ (5 μM) or UV (20 J/m^2). The plates were incubated at 37°C overnight, and can be screened for mutant clones that confer sensitivity or resistance to NFZ or UV. It is important to grow colonies at 37°C instead of 30°C because Δ*nusA* is cold sensitive. Using a standard

genetic background where NusA is essential (AB1157), the new *nusA* mutants were tested for whether they retained the functions required for viability and were recessive (similar to *nusA11*) or dominant.

FIGURES

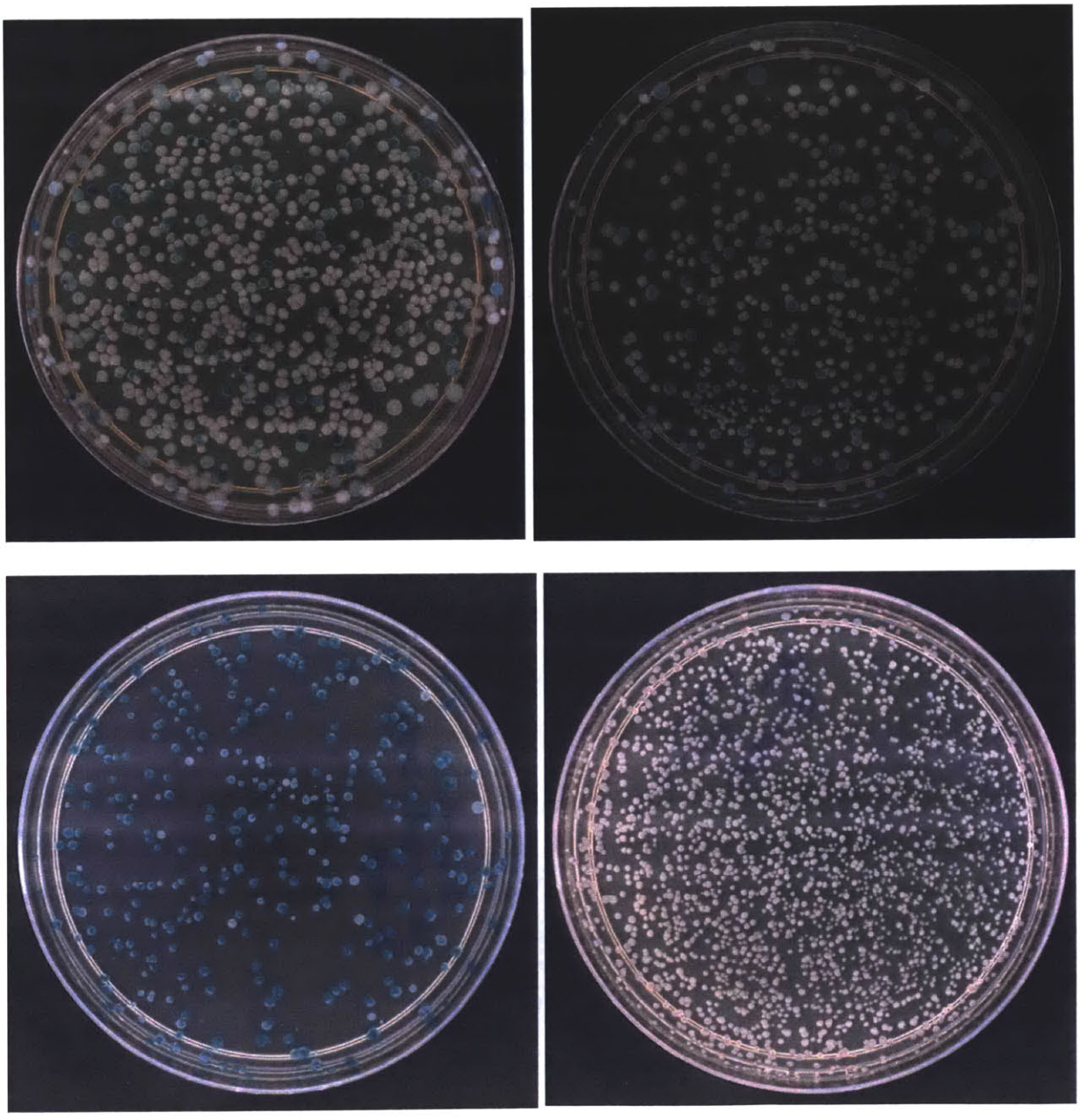

Figure 1. clockwise starting from top left, *pKNT25-nusA + pUT18C-dinB*, *pKNT25-nusA + pUT18C-uvrD*, *pKT25-nusA + pUT18* vector, *pKT25-nusA + pUT18-nusA*. Assay conditions: LB + X-Gal + Amp + Kan, IPTG induced, incubated at 30°C for 72 hours.

Strains	NFZ	4-NQO	MMS	UV
WT MDS42	R	R	R	R
WT AB1157	R	R	R	R
MDS42 Δ*nusA*	S	S	S	S
AB1157 ΔnusA: not viable				
MDS42 *nusA11*	S	S	R	R
AB1157 *nusA11*	S	S	R	R

Figure 2. Phenotypes of different strains upon treatment with different DNA damaging agents.

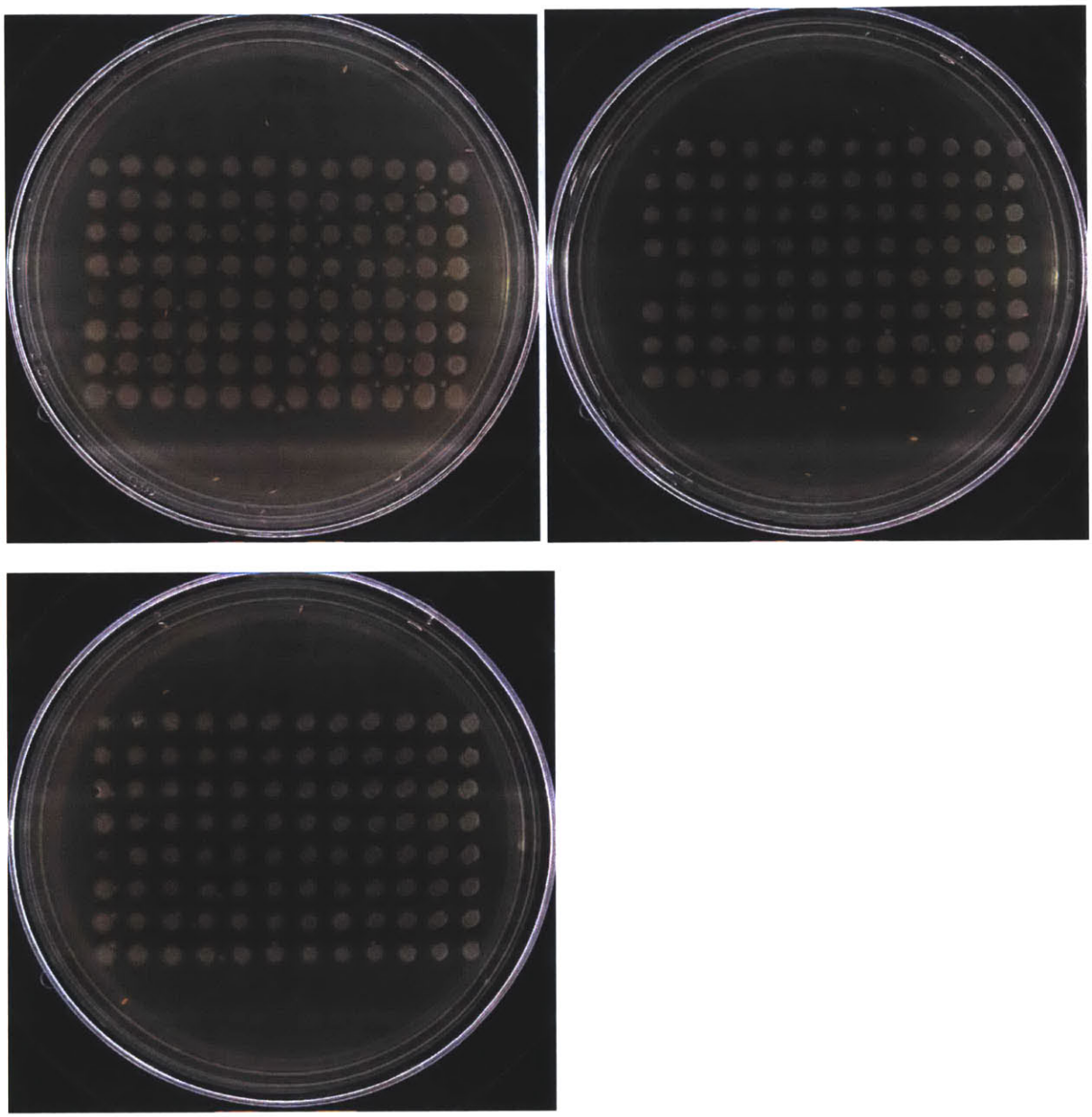

Figure 3. *nusA* separation of function screen. Clockwise from top left: control plate, NFZ treated plate (5 μM), UV treated plate (20 J/m^2). On each plate, the top left three colonies are controls, from left: empty vector transformed into Δ*nusA*, wild type *nusA* transformed into Δ*nusA*, empty vector transformed into *nusA11*. The fifth colony on the first column is I97V.

Table 1. Strains and plasmids used in this study

Strains	Genotype	Reference
MG1655	F^-, λ^-, rph-1	Lab stock
MG1655 pKT25-NusA	F^-, λ^-, rph-1 pKT25 fused to wt nusA	This study
MG1655 pKNT25-NusA	F^-, λ^-, rph-1 pKNT25 fused to wt nusA	This study
MG1655 pUT18-NusA	F^-, λ^-, rph-1 pUT18 fused to wt nusA	This study
MG1655 pUT18C-NusA	F^-, λ^-, rph-1 pUT18C fused to wt nusA	This study
MG1655 pKNT25-DinB	F^-, λ^-, rph-1 pKNT25 fused to wt dinB	This study
MG1655 pUT18-DinB	F^-, λ^-, rph-1 pUT18 fused to wt dinB	This study
MG1655 pUT18C-DinB	F^-, λ^-, rph-1 pUT18C fused to wt dinB	This study
MG1655 pKT25-UvrA	F^-, λ^-, rph-1 pKT25 fused to wt uvrA	This study
MG1655 pKNT25-UvrA	F^-, λ^-, rph-1 pKNT25 fused to wt uvrA	This study
MG1655 pUT18-UvrA	F^-, λ^-, rph-1 pUT18 fused to wt uvrA	This study
MG1655 pUT18C-UvrA	F^-, λ^-, rph-1 pUT18C fused to wt uvrA	This study
MG1655 pUT18C-UvrD	F^-, λ^-, rph-1 pUT18C fused to wt uvrD	This study
MG1655 pKT25	F^-, λ^-, rph-1 pKT25	This study
MG1655 pKNT25	F^-, λ^-, rph-1 pKNT25	This study

MG1655 pUT18	F^-, λ^-, rph-1 pUT18	This study
MG1655 pUT18C	F^-, λ^-, rph-1 pUT18C	This study
MDS42	MG1655 with ~14% of genome deleted	Lab stock
MDS42 ΔnusA	As MDS42 except ΔnusA::Cm	Lab stock
SEC1302	As MDS42 except nusA11 zha0132::Tn10	Lab stock
MDS42 ΔnusA pBR322	pBR322 empty vector transformed into MDS42 ΔnusA	This study
MDS42 ΔnusA pNAG2010	pNAG2010 transformed into MDS42 ΔnusA	This study
Sec 1302 pBR322	pBR322 empty vector transformed into Sec1302	This study
BTH101	F^-, cya-99, araD139, galE15, galK16, rpsL1 (Str^r), hsdR2, mcrA1, mcrB1	Lab stock
XL1-Red	endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac mutD5 mutS mutt Tn10 (Tet^r)[a]	Agilent Technologies
XL1-Blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB $lacI^qZ\Delta M15$ Tn10 (Tet^r)]	Agilent Technologies
Plasmids		
pKT25	Bacterial two-hybrid plasmid, encodes T25, MCS is inserted at the 3' end of T25, low copy number	Lab stock
pKNT25	Bacterial two-hybrid plasmid, encodes T25 fragment that is fused in frame downstream from a MCS, low copy number	Lab stock
pUT18	Bacterial two-hybrid plasmid, encodes T18, MCS is fused to the N-terminal end of T18, high copy number	Lab stock
pUT18C	Bacterial two-hybrid plasmid, encodes T18, MCS is located at the 3' end of T18, high copy number	Lab stock

pBR322	Empty vector, Amp^R	New England Bio labs
pNAG2010	NusA cloned into pBR322 under its own promoter	Lab stock

References:

1 Farnham, P. J., Greenblatt, J. & Platt, T. Effects of NusA protein on transcription termination in the tryptophan operon of Escherichia coli. *Cell* **29**, 945-951, doi:10.1016/0092-8674(82)90457-3.

2 Greenblatt, J. & Li, J. Interaction of the sigma factor and the nusA gene protein of E. coli with RNA polymerase in the initiation-termination cycle of transcription. *Cell* **24**, 421-428, doi:http://dx.doi.org/10.1016/0092-8674(81)90332-9 (1981).

3 Liu, K., Zhang, Y., Severinov, K., Das, A. & Hanna, M. M. Role of E. coli RNA polymerase alpha subunit in modulation of pausing, termination and anti-termination by the transcription elongation factor NusA. *The EMBO journal* **15**, 150-161 (1996).

4 Friedman, D. I. & Baron, L. S. Genetic characterization of a bacterial locus involved in the acitivity of the N function of phage lambda. *Virology* **58**, 141-148 (1974).

5 Greenblatt, J. & Li, J. The nusA gene protein of E. coli. . *J. Mol. Biol.* **147**, 11-23 (1981).

6 Craven, M. G. *et al.* E.coli-Salmonella typhimurium hybrid nusA genes: identification of a short motif required for action of the lambda N transcription antitermination protein. *Journal of bacteriology* **176** (1994).

7 Nudler, E. & Gottesman, M. E. Transcription termination and antitermination in E. coli. *Genes to Cells* **7**, 755-768 (2002).

8 Burns, C. M., Richardson, L. V. & Richardson, J. P. Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in E. coli. *Journal of molecular biology* **278**, 307-316 (1998).

9 Beuth, B., Penneli, S., Arnvig, K. B., Martin, S. R. & Taylor, I. A. Structure of Mycobacterium tuberculosis NusA-RNA complex. *The EMBO journal* **24**, 3578-3587, doi:10.1038/ (2005).

10 Gopal, B. *et al.* Crystal structure of the transcription elongation/antitermination factor NusA from Mycobacterium tuberculosis at 1.7A resolution. *J. Mol. Biol.* **314**, 1087-1095, doi:10.1006/jmbi.2001.5144 (2001).

11 Shibata, R. *et al.* Crystal structure and RNA-binding analysis of the archaeal transcription factor NusA. *Biochemical and biophysical research communications* **355**, 122-128, doi:10.1016/j.bbrc.2007.01.119 (2007).

12 Shin, D. H. *et al.* Crystal structure of NusA from Thermotoga Maritima and functional implication of the N-terminal domain. *Biochemistry* **42**, 13429-13437 (2003).

13 Schweimer, K. *et al.* NusA interaction with the alpha subunit of E. coli RNA polymerase is via the UP element site and releases autoinhibition. *Structure* **19**, 945-954, doi:10.1016/j.str.2011.03.024 (2011).

14 Ha, K. S., Toulokhonov, I., Vassylyev, D. G. & 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. *J. Mol. Biol.* **401**, 708-725, doi:10.1016/j.jmb.2010.06.036 (2010).

15 Mah, T., Li, J., Davidson, A. R. & Greenblatt, J. Functional importance of regions in E. coli elongation factor NusA that interact with RNA polymerase, the bacteriophage lambda N protein and RNA. *Mol. Microbiol.* **34**, 523-537 (1999).
16 Friedberg, E. C., Wagner, R. & Radman, M. Specialized DNA polymerases, cellular survival, and the genesis of mutations. *Science* **296**, 1627-1630 (2002).
17 Goodman, M. F. & Woodgate, R. Translesion DNA polymerases. *Cold Spring Harbor perspectives in biology* **5**, a010363, doi:10.1101/cshperspect.a010363 (2013).
18 Tang, M. *et al.* Roles of E. coli DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature* **404**, 1014-1018 (2000).
19 Jarosz, D. F., Godoy, V. G., Delaney, J. C., Essigmann, J. M. & Walker, G. C. A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates. *Nature* **439**, 225-228, doi:10.1038/nature04318 (2006).
20 Jarosz, D. F., Cohen, S. E., Delaney, J. C., Essigmann, J. M. & Walker, G. C. A DinB variant reveals diverse physiological consequences of incomplete TLS extension by a Y-family DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 21137-21142, doi:10.1073/pnas.0907257106 (2009).
21 Cohen, S. E. *et al.* Roles for the transcription elongation factor NusA in both DNA repair and damge tolerance pathways in E. coli. *Proceedings of the National Academy of Sciences of the United States of America* (2010).
22 S.-R, K., K, M., M, Y., P, G. & T, N. Roles of chromosomal and episomal dinB genes encoding DNA pol IV in targeted and untargeted mutagenesis in Escherichia coli. *Molecular Genetics and Genomics* **266**, 207-215, doi:10.1007/s004380100541 (2001).
23 Cohen, S. E., Godoy, V. G. & Walker, G. C. Transcriptional modulator NusA interacts with translesion DNA polymerases in Escherichia coli. *Journal of bacteriology* **191**, 665-672, doi:10.1128/JB.00941-08 (2009).
24 Kornberg, A. & Baker, T. A. DNA Replication. *Freeman, New York* (1992).
25 Heller, R. C. & Marians, K. J. Replication fork reactivation downstream of a blocked nascent leading strand. *Nature* **439**, 557-562 (2006).
26 Iyer, V. N. & Rupp, W. D. Usefulness of benzoylated naphthoylated DEAE-cellulose to distinguish and fractionate double-stranded DNA bearing different extents of single-stranded regions.. *Biochimica et biophysica acta* **228**, 117-126 (1971).
27 Friedberg, E. C. *et al.* DNA Repair and Mutagenesis. *Am Soc Microbiol, Washington, DC* (2005).
28 Imaizumi-Anraku, H. *et al.* Interaction network containing conserved and essential protein complexes in E coli. *Nature* **433**, 527-531, doi:10.1038/nature03237 (2005).
29 Ona, K. R., Courcelle, C. T. & Courcelle, J. Nucleotide excision repair is a predominant mechanism for processing nitrofurazone-induced DNA damage in Escherichia coli. *Journal of bacteriology* **191**, 4959-4965, doi:10.1128/JB.00495-09 (2009).
30 Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nature reviews. Molecular cell biology* **9**, 958-970, doi:10.1038/nrm2549 (2008).
31 Savery, N. J. The molecular mechanism of transcription-coupled DNA repair. *Trends in microbiology* **15**, 326-333, doi:10.1016/j.tim.2007.05.005 (2007).
32 Truglio, J. J., Croteau, D. L., Houten, B. V. & Kisker, C. Prokaryotic Nucleotide Excision Repair: The UvrABC System. *Chem. Rev.* **106**, 233-252 (2006).
33 Selby, C. P. & Sancar, A. Mechanism of transcription-repair coupling and mutation frequency decline. *Microb. Reviews* **58**, 317-329 (1994).
34 Selby, C. P. & Sancar, A. Molecular mechanism of transcription-repair coupling. *Science* **260**, 53-58 (1993).

35 Selby, C. P. & Sancar, A. Transcription preferentially inhibits nucleotide excision repair of the template DNA strand in vitro. *The Journal of biological chemistry* **265**, 21330-21336 (1990).
36 Selby, C. P., Witkin, E. M. & Sancar, A. E. coli mfd mutant deficient in "mutation frequency decline" lacks strand-specific repair: in vitro complementation with purified coupling factor. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 11574-11578 (1991).
37 Epshtein, V. *et al.* UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* **505**, 372-377, doi:10.1038/nature12928 (2014).
38 Lee, J. Y. & Yang, W. UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell* **127**, 1349-1360, doi:10.1016/j.cell.2006.10.049 (2006).
39 Euromedex. BACTH system kit, bacterial adenylate cyclase two-hybrid system kit.
40 Phizicky, E. M. & Fields, S. Protein-protein interactions: methods for detection and analysis. *Microb. Reviews* **59**, 94-123 (1995).
41 Posfai, G. *et al.* Emergent properties of reduced-genome Escherichia coli. *Science* **312**, 1044-1046, doi:10.1126/science.1126439 (2006).
42 Washburn, R. S. & Gottesman, M. E. Transcription termination maintains chromosome integrity. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 792-797, doi:10.1073/pnas.1009564108 (2011).
43 Nakamura, Y., Mizusawa, S., Court, D. L. & Tsugawa, A. Regulatory defects of a conditionally lethal nusA ts mutant of E. coli, Positive and negative modulator roles of NusA protein in vivo. *J. Mol. Biol.* **189**, 103-111 (1986).
44 Kanojia, D. & Vaidya, M. M. 4-nitroquinoline-1-oxide induced experimental oral carcinogenesis. *Oral oncology* **42**, 655-667, doi:10.1016/j.oraloncology.2005.10.013 (2006).
45 Ryan, A., Kaplan, E., Laurieri, N., Lowe, E. & Sim, E. Activation of nitrofurazone by azoreductases: multiple activities in one enzyme. *Scientific reports* **1**, 63, doi:10.1038/srep00063 (2011).
46 Selby, C. P., Drapkin, R., Reinberg, D. & Sancar, A. RNA polymerase II stalled at a thymine dimer: footprint and effect on excision repair. *Nucleic acids research* **25**, 787-793 (1997).
47 Van Houten, B., Croteau, D. L., DellaVecchia, M. J., Wang, H. & Kisker, C. 'Close-fitting sleeves': DNA damage recognition by the UvrABC nuclease system. *Mutation research* **577**, 92-117, doi:10.1016/j.mrfmmm.2005.03.013 (2005).
48 Kashlev, M. *et al.* Histidin-tagged RNA polymerase: dissection of the transcription cycle using immobilized enzyme. *Gene.* **130**, 9-14 (1993).
49 Delaney, J. C. & Essigmann, J. M. Assays for Determining Lesion Bypass Efficiency and Mutagenicity of Site - Specific DNA Lesions In Vivo. *Methods in Enzymology* **408**, 1-15, doi:10.1016/s0076-6879(06)08001-3 (2006).
50 Pages, V., Mazon, G., Naiman, K., Philippin, G. & Fuchs, R. P. Monitoring bypass of single replication-blocking lesions by damage avoidance in the Escherichia coli chromosome. *Nucleic acids research* **40**, 9036-9043, doi:10.1093/nar/gks675 (2012).