AFRL-OSR-VA-TR-2013-0603

MAINTAINING GENETIC INTEGRITY UNDER EXTREME CONDITIONS - NOVEL DNA DAMAGE REPAIR BIOLOGY IN THE ARCHAEA

STUART MACNEILL

UNIVERSITY OF ST. ANDREWS

11/23/2013
Final Report

AIR FORCE RESEARCH LABORATORY
AF OFFICE OF SCIENTIFIC RESEARCH (AFOSR)/RSL
ARLINGTON, VIRGINIA 22203
AIR FORCE MATERIEL COMMAND

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
13-11-2013	Final Performance Report	15 AUG 2010 - 14 AUG 2013

4. TITLE AND SUBTITLE

Maintaining Genetic Integrity Under Extreme Conditions: Novel DNA Damage Repair Biology in the Archaea

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-10-1-0421

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

MACNEILL, Stuart, A

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of St Andrews, College Gate, St Andrews, Fife KY16 9AJ, UK

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Office of Scientific Research
Suite 325, Room 3112
875 Randolph Street
Arlington, VA 22203-1768

10. SPONSOR/MONITOR'S ACRONYM(S)

AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution A - Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This work tested the hypothesis that the conserved archaeal Nre protein functions as a PCNA-dependent DNA repair factor. By exploiting the genetic tools available in the halophilic euryarchaeal organism Haloferax volcanii, it was shown that [1] Nre does indeed have a role in DNA damage repair (cells lacking Nre are supersensitive to specific DNA damaging agents and show delayed repair of DNA double strand breaks evidenced by DNA pulsed field gel electrophoresis) and [2] that the function of Nre in DNA damage repair in dependent on its ability to interact through its C-terminal PIP motif with PCNA (cells expressing an Nre protein lacking the PIP motif display similar phenotypes to cells lacking Nre altogether). Thus, the original hypothesis was validated. Further genetic analysis indicates that Nre performs its function, at least in part, in coopertation with the UvrABC excinuclease complex. This work contributes significant to our understanding of the diversity of DNA repair functions in archaea.

15. SUBJECT TERMS

Extremophile/DNA damage/DNA repair/archaea/mitomycin C/UvrABC

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr Stuart A. Macneill
U	U	U	UU		**19b. TELEPHONE NUMBER** *(include area code)* +44 1334 46 7268

Maintaining Genetic Integrity Under Extreme Conditions: Novel DNA Damage Repair Biology in the Archaea

Contract/Grant #: FA9550-10-1-0421
Reporting Period: 15 August 2010 to 14 August 2013

Summary of work undertaken

A. Probing the function of the Nre family proteins

- A1 Bioinformatic analysis of the Nre and SplB homologues
- A2 Expression of Nre and SplB proteins in native and non-native hosts
- A3 Genetic analysis of Nre DNA repair function
- A4 Conclusions

B. Widening the net in the search for new DNA-directed enzyme activities

C. New tools for *H. volcanii* research

A. Probing the function of the Nre family proteins

In work leading up to the funding application a previously uncharacterised family of archaeal proteins, termed the Nre family, was identified as being able to bind to the DNA sliding clamp molecule PCNA. The Nre proteins (**Figure 1**) were hypothesised to be novel DNA repair enzymes. The stated aims of the proposal were to use a combination of genetic, biochemical and structural biology methods [1] to investigate the biochemical activities and cellular function of Nre proteins, [2] to probe their relationship with the SplB protein, a homologue of bacterial spore photoproduct (SP) lyase, and [3] to determine the role of the hypothesised Nre-Spl repair system in repairing DNA damage under extreme conditions.

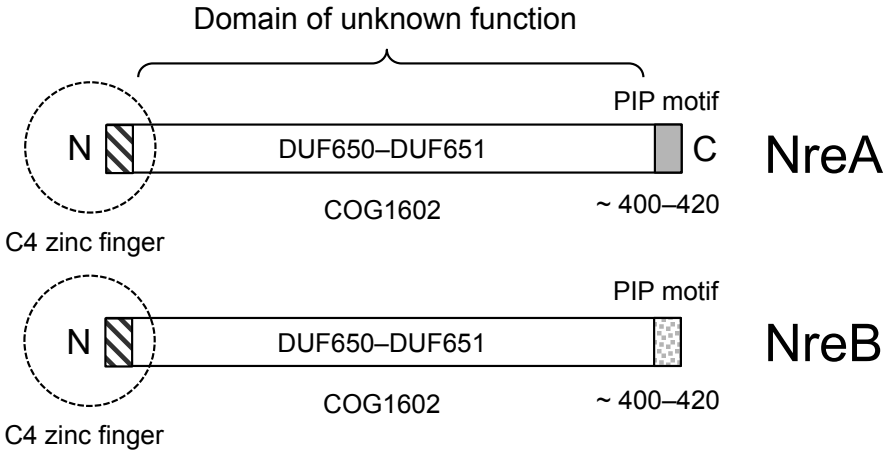


Figure 1 (above): Schematic view of Nre protein structure. Notes: [1] While the majority of archaeal species encode a single Nre protein, some encode two (termed NreA and NreB). [2] The C4 zinc finger is found in all species except halophiles (see below).

A1 Bioinformatic analysis of the Nre and SplB homologues

(Postdoctoral researcher Dr Xavier Giroux, salary supported by the AFOSR grant)

At the start of the project, a detailed bioinformatic analysis of Nre and SplB gene distribution and protein sequence across the archaeal domain was performed. This allowed confirmation of earlier observations on, for example, the co-localisation of Nre and SplB homologues in widely divergent archaeal lineages, the presence or absence of metal binding (putative zinc finger or iron-sulphur cluster) domains and PIP boxes, and the distribution of NreA and NreB proteins in different species. While not providing a clear insight into Nre function, these studies highlighted the near universal nature of Nre in the archaeal domain, and multiple protein sequence alignments of ~ 100 Nre proteins have identified highly conserved amino acids that would make excellent targets for future mutagenesis studies.

A2 Expression of Nre and SplB proteins in native and non-native hosts

(Postdoctoral researcher Dr Xavier Giroux, salary supported by the AFOSR grant)

A2.1 Expression of recombinant NreA, NreB and SplB proteins in *E.coli*

Almost all proteins that interact directly with PCNA are enzymes possessing DNA-directed activities such as nucleases, glycosylases, methylases, polymerases, etc. As the main body of the Nre protein (i.e. the 350 aa region between the N-terminal C4 metal binding domain and the C-terminal PIP motif) is unlike anything in current databases, the function of this putative enzyme cannot be predicted. Thus, a major part of the proposed work aimed to purify the NreA and NreB proteins and assay these for DNA-directed enzymatic activity using a variety of synthetic DNA substrates. In parallel with this, it was intended to express and purify SplB also. In all three cases (NreA, NreB and SplB) protein crystallisation was a secondary goal.

A2.1.1 Recombinant protein expression

The NreA, NreB and SplB ORFs from the thermophilic crenarchaeon *Sulfolobus solfataricus* were amplified by PCR from genomic DNA and cloned into plasmid pET151-D-TOPO (Invitrogen) using the directional TOPO cloning method. ORFs cloned into pET151-D-TOPO are expressed as fusions with a TEV-cleavable N-terminal His6-V5 tag (see **Figure 2** below). All three plasmids were sequenced to ensure the absence of unwanted sequence changes and transformed into the following *E.coli* strains optimised for protein expression: Rosetta 2 (DE3) [pLysS] and Arctic Express (DE3) RP (marketed by EMD Biosciences and Agilent, respectively). Rosetta 2 carries additional genes encoding seven "rare" tRNAs, while Arctic Express carries a subset of the these "rare" tRNA genes and two cold-adapted protein chaperones, Cpn10 and Cpn60 from the psychrophilic bacterium *Oleispira antarctica*, to promote protein folding at low temperatures (10-15°C). Protein expression was then induced by addition of IPTG at various temperatures (12°C, 20°C, 37°C and 44°C). In summary, all three proteins were expressed to high levels after 2 hours at 37°C and 44°C. **Figure 3** shows representative results (SDS-PAGE gels of total protein extracts expressing SsoNreA, SsoNreB and SsoSplB). Reduced expression was seen at 12°C, even 24 hours after induction, and at 20°C. Soluble protein extracts were then prepared by lysing cells (either by sonication or by freeze-thawing following lysozyme treatment) in a variety of extraction buffers (with pH values in the range 5.6 – 9.0, NaCl concentrations from 250 mM to 1M and non-ionic detergent Tween 20 and Triton X100 concentrations from 0.1 – 1%). In all cases, all three expressed proteins SsoNreA, SsoNreB and SsoSplB were largely, if not entirely, insoluble. In subsequent experiments, addition of 5 mM zinc sulfate during protein expression, potentially to aid

folding of the Nre zinc finger domains, did not result in production of soluble SsoNreA or SsoNreB. Similarity, pairwise mixing of protein extracts post-expression (for example, mixing extracts containing SsoNreA with extracts containing SsoNreB) did not result in solubilisation, as might have occurred if NreA, NreB and SplB were natural binding partners.

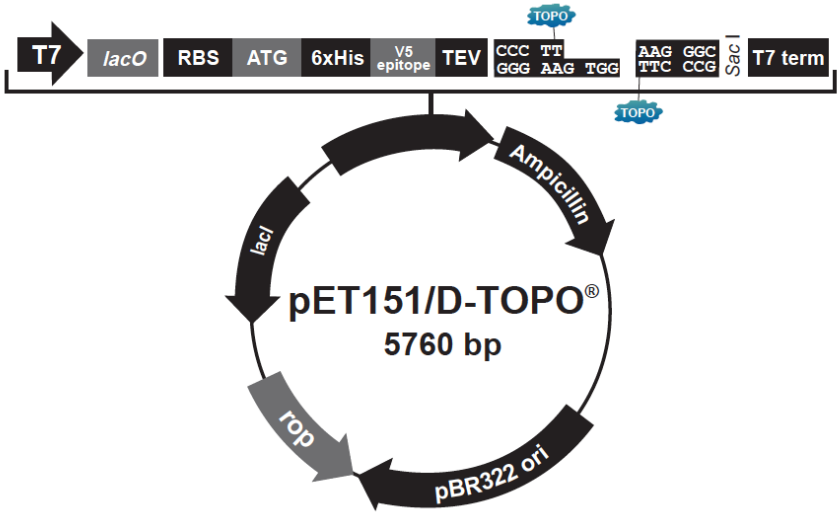


Figure 2 (above): Schematic of pET151/D-TOPO (Invitrogen) showing location of the T7 promoter and lacO operator, ribosome binding site (RBS), start codon (ATG), hexahistidine tag (6XHis), V5 epitope tag, TEV protease cleavage site and transcriptional terminator (T7 term).

A2.1.2 Further attempts at production of soluble protein

In an effort to identify conditions for production of soluble SsoNreA, SsoNreB and SsoSplB proteins from pET151-D-TOPO, a wide range of expression and native cell extract preparation conditions were tested. Tested variables included induction time, induction temperature, salt concentration in lysis buffer, pH of lysis buffer, detergent type and concentration, method of cell lysis, etc. It was not possible to identify conditions where any of the three proteins was present in the soluble fraction, however (data not shown). The use of denaturing conditions was considered but rejected, as the key goal of this work was to purify the proteins in active form and this could not be guaranteed if the protein had been denatured and then re-folded, even assuming that the re-folded protein remained soluble. (Indeed, in earlier work from the lab, it was shown that solubilising recombinant *A. fulgidus* Nre and SplB proteins with urea was possible, but that subsequent slow removal of the urea by dialysis led to protein precipitation.)

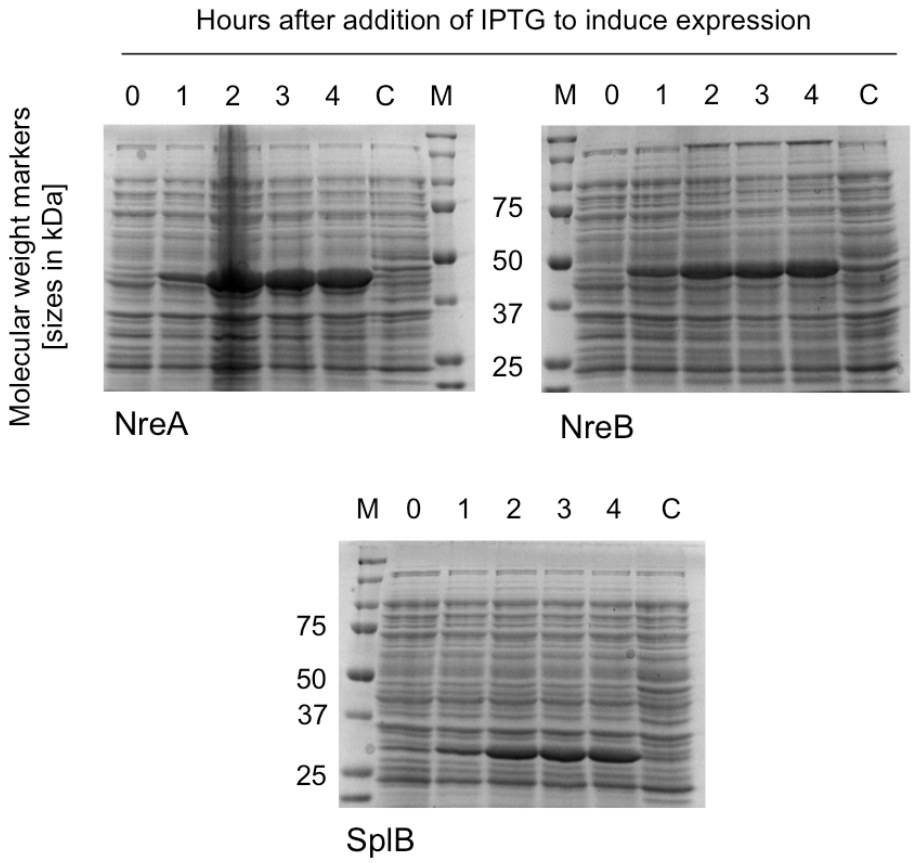


Figure 3 (above): Expression of SsoNreA, SsoNreB and SsoSplB in Rosetta 2 (DE3) [pLysS]. SDS- PAGE gels of total protein samples prepared following IPTG induction at 37°C for 0 – 4 hours. All three proteins are strongly expressed under these conditions but all three are insoluble (data not shown).

A2.1.3 Co-expression of SsoNreA, SsoNreB and SsoSplB using DUET expression system

As noted above, it is possible that the NreA, NreB and SplB proteins interact directly with one another in vivo. Such interactions may be required to promote folding of the individual proteins and facilitate partitioning to the soluble extract by, for example, masking hydrophobic protein surfaces that could cause irreversible protein aggregation. To test whether co-expression of the SsoNreA, SsoNreB and SsoSplB proteins results in their partitioning to the soluble extract, the three ORFs were subcloned into the DUET series expression plasmids (EMD Biosciences). Each of the four DUET plasmids can be used to express two proteins. Thus by combining two plasmids in one cell (this is possible as the plasmids are not incompatible), it was possible to express all three proteins together and test their solubility. However, it was found that the proteins remained resolutely insoluble under all conditions tested (data not shown).

A2.1.4 Expression of untagged proteins, single and in combination

It has been observed in certain cases that adding a sequence tag to a protein sequence (to facilitate protein purification for example, as is the case here) can result in the fusion protein failing to fold correctly and a resulting loss of solubility. To determine if this was the issue with SsoNreA,

SsoNreB and SsoSpIB, the three ORFs were recloned into expression vectors such that the proteins would be expressed without the N-terminal His6, V5 and TEV tags conferred by pET151-D-TOPO. This was done using the DUET vectors mentioned above, potentially permitting expression of individual proteins, or pairwise or three-way combinations.

Figure 4 shows the results of expression of untagged NreA (Sso0872, left panel) and NreB (Sso2743, right panel). Both proteins were insoluble under all conditions tested. **Figure 5** shows the co-expression of untagged NreA and NreB. Again, the proteins remained resolutely insoluble under all conditions tested.

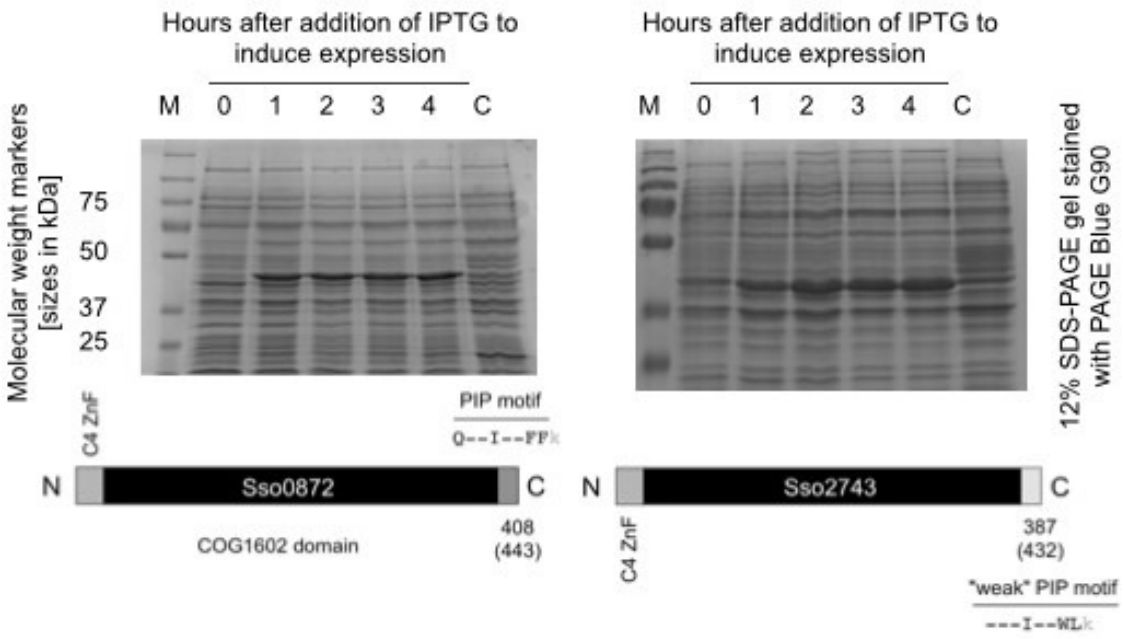
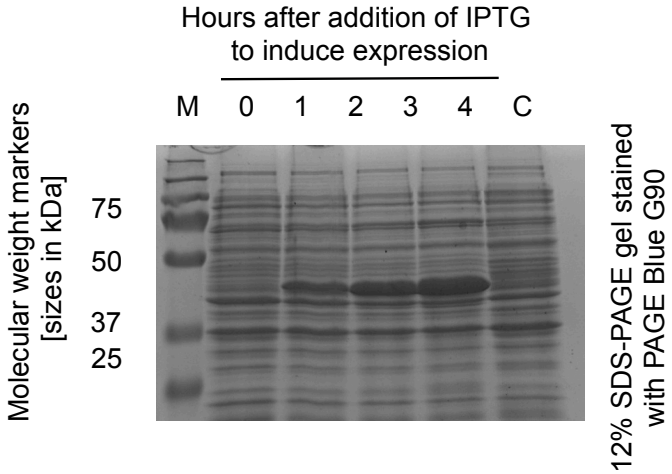


Figure 4 (above): SDS-PAGE analysis of NreA and NreB expression at 37°C for 4 hours in *E.coli* Rosetta 2 [pLysS].

Figure 5 (below): SDS-PAGE analysis of NreA and NreB co-expression at 37°C for 4 hours in E.coli Rosetta 2 [pLysS].



Thus, to summarise the results of sections A2.1, it was not possible to obtain soluble recombinant NreA, NreB or SplB proteins by expression in the non-native host E.coli that could be used for subsequent biochemical analysis.

A2.2 Expression of Sulfolobus solfataricus Nre and SplB proteins in their native host

In order to overcome the problems with *E.coli* expression, expression of the Sso NreA, NreB and SplB in their native host was attempted using the SSV1 virus-based expression system developed in the Albers lab (MPI for Terrestrial Microbiology, Marburg, Germany). To achieve this, the *S.solfataricus nreA*, *nreB* and *splB* genes were first cloned into the "entry vector" plasmid pMZ1, before being transferred into the viral vector pMJ0503. Each protein was expressed from pMJ0503 with C-terminal His10 and StrepII tags, allowing purification on Ni-NTA and StrepTactin resins, respectively, and from an arabinose-inducible promoter. The pMJ0503 plasmids were transformed into *S.solfataricus* and expression of the target protein was then induced by arabinose addition for 4 days at 75°C. Cells were then harvested and protein purification attempted by affinity chromatography. The results of this were as follows:

- No expression of the *S.solfataricus* NreA protein could be detected in this system, possibly indicating that the viral vector is unstable or that expression of NreA is toxic to virally-infected cells.
- Low-level expression of *S.solfataricus* NreB could be detected by immunoblotting induced cell extracts and partially purified fractions with anti-His tag antibodies (see **Figure 6**). However, the yield of protein was so low as to make purification of sufficient material for biochemical analysis infeasible.

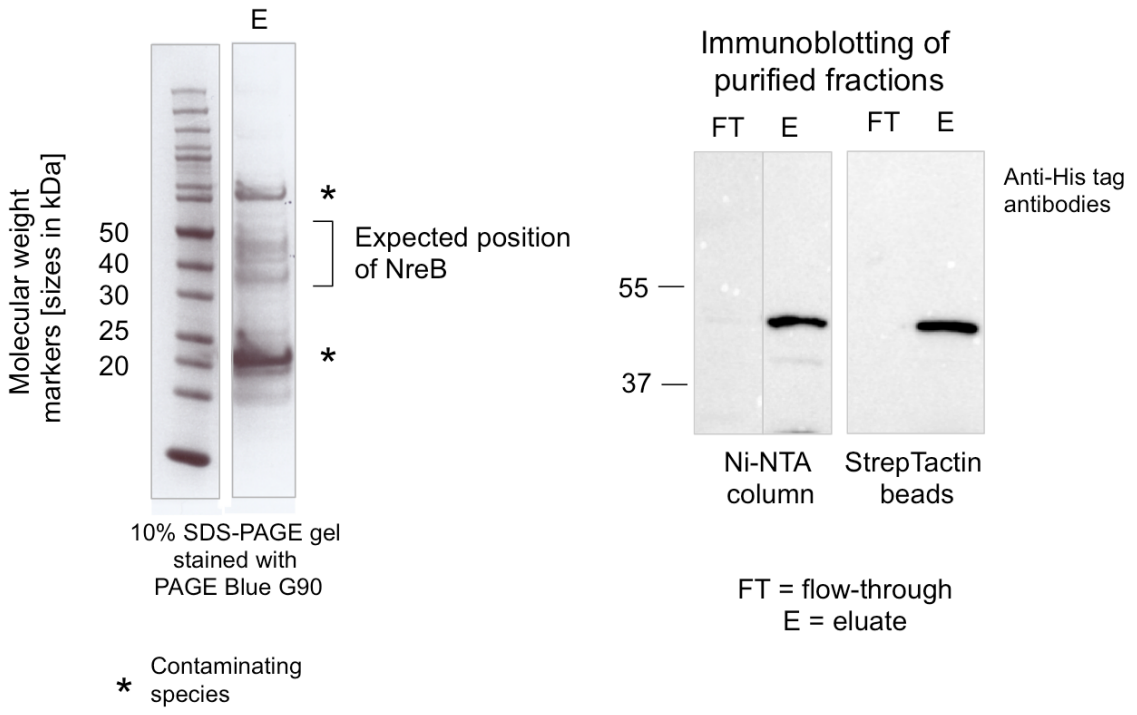


Figure 6 (above): Expression and low yield purification of NreB from *S. solfataricus*. Extracts were prepared from cells expressing NreB and subjected to immobilised metal ion chromatography (IMAC) using Ni-NTA agarose to bind the His-tagged NreB protein. Bound proteins were eluted using imidazole and applied to Strep-Tactin beads to re-bind the StrepII-tagged NreB. Bound proteins were then eluted using desthiobiotin. Left hand panels: PAGE Blue G90 staining of imidazole-eluted fraction from NiNTA column (labelled E). Right hand panel: Immunoblotting of purified fractions using anti-His antibodies detects the NreB protein in the column/bead eluates (E) and not in the column/bead flow-throughs (FT).

- Moderate levels of expression of SplB were detected and this protein could be purified to near homogeneity on Ni-NTA agarose (see **Figure 7**). However, it was not possible to concentrate the purified material - despite repeated attempts under a variety of conditions, the SplB protein precipitated out of solution.

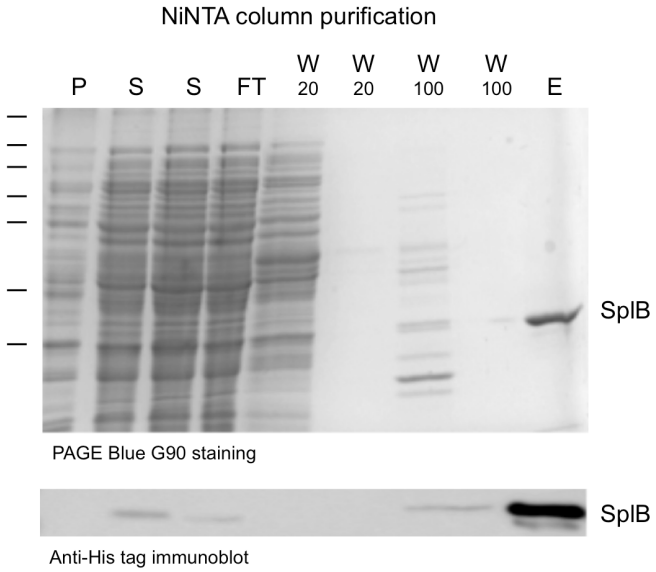


Figure 7 (above): Expression and purification of *S. solfataricus* SplB. Extracts were prepared from cells expressing SplB and subjected to immobilised metal ion chromatography (IMAC) using Ni-NTA agarose to bind the His-tagged SplB protein. The column was washed sequentially with buffers containing 20 mM and 100 mM imidazole (W20,W100 respectively), after which bound protein was eluted with 250 mM imidazole (E). Other fractions shown: insoluble material following lysis (P, pellet), soluble material following cell lysis (S) and Ni-NTA column flow-through (FT). Upper panel: PAGE Blue G90 stained SDS-PAGE gel. Lower panel: immunoblotting of equivalent gel using anti-His tag antibodies.

A2.4 Summary of efforts to obtain Nre and SplB proteins from biochemical analysis

Despite a considerable amount of time and effort being allocated to this part of the project, as befitting the importance of determining the biochemical activity of the NreA, NreB and SplB

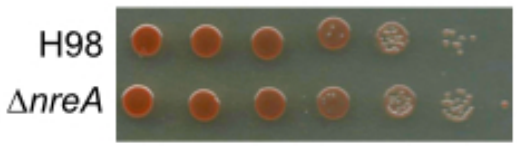
proteins, it was not possible to obtain any of the three in a suitable form for subsequent biochemical analysis, and attention was switched to analysing the *in vivo* function of NreA in *H. volcanii*, as described below.

A3 Genetic analysis of Nre DNA repair function

(Postdoctoral researcher Dr Xavier Giroux, salary supported by the AFOSR grant)

A3.1 Initial work demonstrating a link between NreA function and DNA damage repair

In order to probe the cellular function of the Nre protein, the genetically tractable archaeal organism *Haloferax volcanii* was used as a model system. *H. volcanii* encodes a single Nre protein, designated Nre. Nre possesses a C-terminal PIP motif but in common with other haloarchaeal Nre proteins, no N-terminal zinc finger. Using standard reverse genetic methods, the *nre* gene was deleted previously in the lab (**Figure 8**). Δ*nre* cells were viable but had not been characterised in any detail. To remedy this, the Δ*nre* strain was revived and its sensitivity to various DNA damaging agents tested. With two exceptions, no differences were seen between the Δ*nre* strain and the otherwise isogenic wild-type control. The exceptions were [1] phleomycin and [2] mitomycin C. Phleomycin causes DNA double-strand breaks that can be repaired by several pathways. The sensitivity of Δ*nre* cells to phleomycin implicates Nre in one or more of these pathways. Mitomycin C (hereafter MMC) forms mono-adducts on DNA as well as intra- and inter-strand crosslinks, with the latter leading to double-strand DNA breaks (DSBs) (**Figure 9**). Nre is therefore implicated in the repair of one or more of these MMC-induced DNA lesions. Note that issues involving batch-to-batch variation of commercially sourced phleomycin led to subsequent work largely focusing on cellular responses to MMC.



Serial dilutions spotted onto YPC medium and incubated at 45° for 3 days

Figure 8 (above): Nre is non-essential in *H. volcanii*. The *nre* gene was deleted by standard methods in wild-type *H. volcanii* strain H98 (shown for comparison).

A3.2 Mechanism of DSB repair (DSBR) in Δ*nre* cells

To investigate potential mechanisms of DSB repair in Δ*nre* cells, an assay developed by the Allers lab (University of Nottingham, UK) was used that involves transforming specially modified *H. volcanii* cells with a specific plasmid that has been linearised by restriction enzyme cleavage to create a DSB that can be repaired by the cell. Depending on the nature of the repair reaction (i.e. accurate joining of the broken ends, inaccurate joining of the broken ends, or homologous recombination), three possible outcomes can be distinguished. In wild-type cells, ~ 95% of repair involves accurate end joining, 1.5% involves inaccurate end-joining and ~ 3% homologous recombination (HR). To test the effect of loss of Nre in this assay, an Δ*nre* knock-out was created in

the appropriate genetic background, transformed with the test plasmid, and the results assayed using the Allers lab method. No difference between the wild-type and Δ*nre* strains was found (data not shown), indicating that loss of Nre does alter the balance between repair by end-joining and repair by HR.

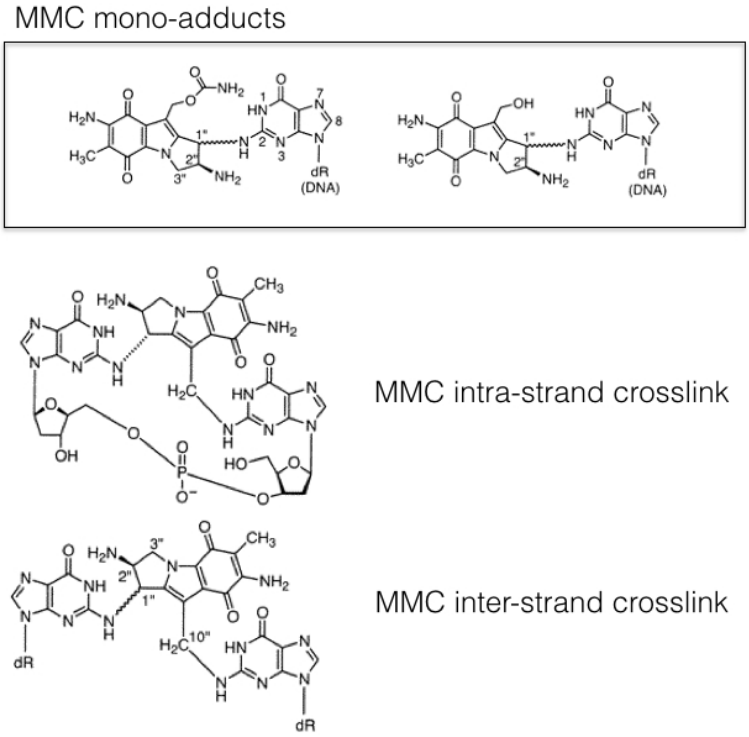


Figure 9 (above): DNA lesions caused by MMC.

A3.3 Genetic epistasis analysis

A3.3.1 Rad50-Mre11 pathway

The Mre11-Rad50 complex is present in all domains of life and is involved in several pathways of DSB repair including HR and NHEJ. Mre11 is a nuclease while Rad50 consists of two globular DNA-binding domains. Together, these two proteins form a complex that binds to and tethers DNA ends, in order to erect a scaffold for the subsequent processing and repair of DSBs. It has previously been shown that, in contrast to the situation with *nre*, deleting the *rad50* and *mre11* genes in *H. volcanii* results in increased resistance to phleomycin. Under these circumstances, there is also significant shift from repair by end-joining to repair by HR (with the latter increasing from 3% to 66% of repair events in the plasmid-based assay). It has been postulated that Rad50-Mre11 binds to the broken DNA ends to prevent them from initiating HR, so that under normal circumstances repair by HR is greatly suppressed (3% of repair events, using the assay described in section A3.2 above).

To examine the effects of loss of *nre* in a Δ*rad50* Δ*mre11* background, an *H. volcanii* Δ*nre* Δ*rad50* Δ*mre11* strain was constructed and its sensitivity to phleomycin- and MMC-induced DNA damage determined. While deleting *nre* led to increased sensitivity to phleomycin in a wild-type background, as above, this effect was not seen in the Δ*rad50* Δ*mre11* background (data not shown). This suggests a functional link between Nre, Rad50 and Mre11. However, the functions of

these proteins are clearly non-equivalent, otherwise Δ*nre* cells would be resistant rather than sensitive to phleomycin. Instead we hypothesise that Nre acts at a late stage in the Rad50-Mre11 pathway and that when Nre is absent it is not possible to divert the broken DNA ends into the HR pathway for repair, as is the case when Rad50-Mre11 is absent. This model is consistent with the data obtained with the plasmid-based repair assay which showed that loss of Nre did not result in an increase in repair by HR. In contrast to the increased resistance to phleomycin seen when Rad50-Mre11 is inactivated, Δ*rad50* Δ*mre11* cells display supersensitivity to MMC (**Figure 10**). Thus the function of Rad50-Mre11 is different when the nature of the DNA damage differs. Significantly, a Δ*nre* Δ*rad50* Δ*mre11* strain was more sensitive to MMC than either Δ*nre* or Δ*rad50* Δ*mre11* (**Figure 10**), indicating that Nre functions in a different pathway to Rad50-Mre11 in the repair of MMC damage.

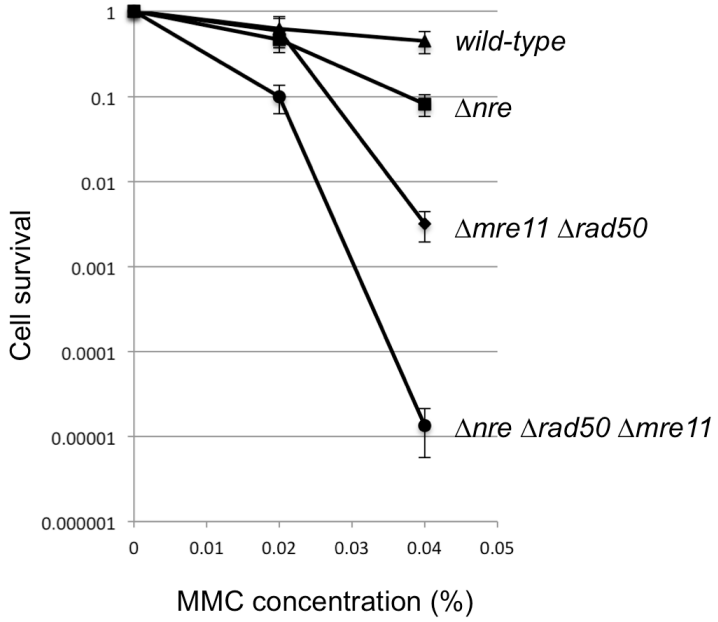
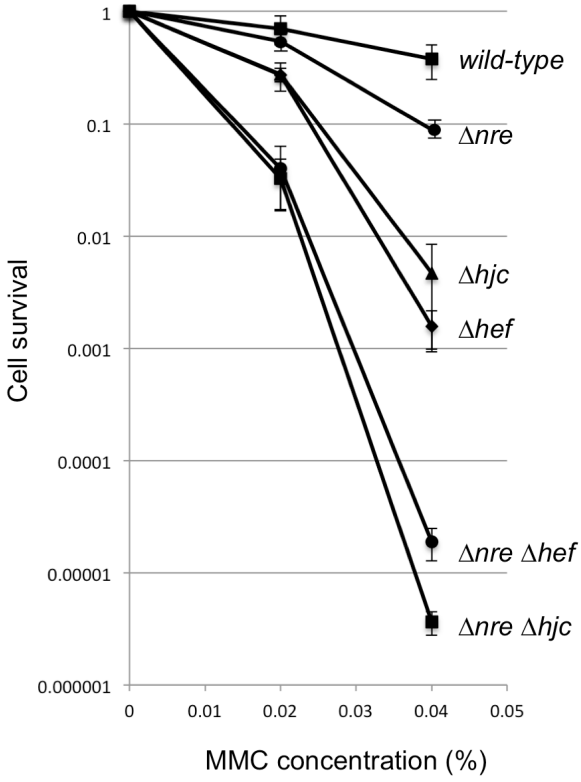


Figure 10 (above): Sensitivity of wild-type, Δ*nre*, Δ*rad50* Δ*mre11* and Δ*nre* Δ*rad50* Δ*mre11* cells to MMC.

A3.3.2 Hef and Hjc

The *H. volcanii* Hef and Hjc proteins are also implicated in the repair of MMC-induced DNA damage. Hef is a member of the XPF/MUS81 nuclease family and is related to the Fanconi anaemia protein FANCM, while Hjc is a Holliday junction resolvase. *H. volcanii* Δ*hef* and Δ*hjc* strains are both supersensitive to MMC. Simultaneous deletion of both *hef* and *hjc* genes is lethal. To test whether Nre functions together with either Hef or Hjc, Δ*hef* Δ*nre* and Δ*hjc* Δ*nre* double mutants were created (both strains were viable) and tested their sensitivity to MMC. Both strains were significantly more sensitive to MMC than the corresponding single mutants Δ*nre*, Δ*hef* or Δ*hjc* (Figure 11). In addition, loss of Nre has a greater impact on survival in Δ*hef* or Δ*hjc* backgrounds compared to a wild-type background, indicating that Nre function is especially important when Hef or Hjc function is compromised. Taken together, these results imply that Nre operates independently of Hef and Hjc.

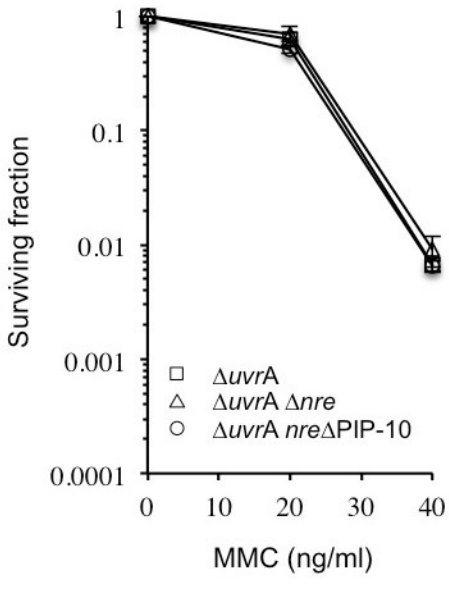
Figure 11 (below): Sensitivity of wild-type, Δ*nre*, Δ*hef,* Δ*hjc,* Δ*nre* Δ*hef and* Δ*nre* Δ*hjc* cells to MMC.



A3.3.3 UvrABC pathway

The primary function of the UvrABC pathway is in the repair of UV-induced damage. *H. volcanii* mutants in *uvrA*, *uvrB* or *uvrC* are acutely sensitive to UV. In some organisms UvrABC also plays a role in the repair of MMC-induced damage, though this has not been tested in *H. volcanii*. To test whether this pathway functions in MMC-induced damage repair in *H. volcanii*, and to probe possible relationships with Nre, various single and double mutant combinations were tested for MMC sensitivity, as above. Results for ΔuvrA and ΔuvrA Δnre are shown in **Figure 12**: similar results were obtained with Δ*uvrB* and Δ*uvrB* single and double (with Δ*nre*) mutants.

Figure 12 (above): Sensitivity of wild-type, Δ*nre*, Δ*uvrA, Δnre Δuvr*A *and Δnre Δhjc* cells to MMC.

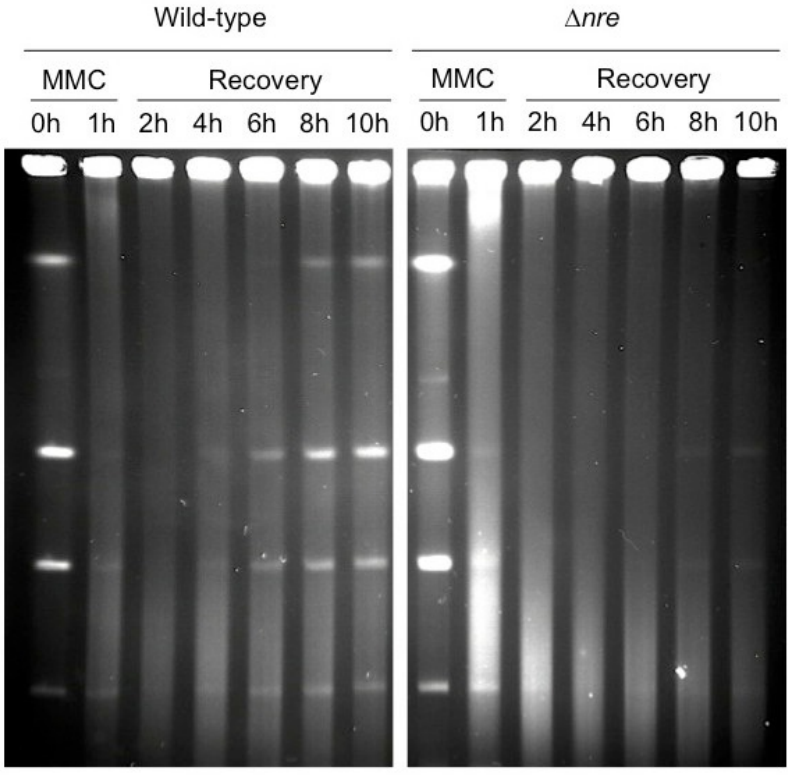


In all three cases, the single Δ*uvr* mutants display sensitivity to MMC similar to that of Δ*nre*, showing for the first time that *H. volcanii* UvrABC has a function beyond that of UV damage repair. Strikingly, the Δ*uvr* Δ*nre* double mutant was no more sensitive than the single mutants (**Figure 12**). These results imply that Nre and UvrABC function in the same pathway of repair of MMC-induced DNA lesions. Interestingly, Δ*nre* mutants are not UV sensitive, implying that Nre and UvrABC do not act in concert to repair all types of DNA damage.

A3.4 Cells lacking NreA are delayed in DNA double-strand break repair (DSBR) following MMC treatment

The sensitivity to DNA damaging agents and the results of the genetic epistasis analysis described above imply an important role for NreA in DNA damage repair. In order to obtain direct physical evidence for the role of NreA in damage repair, pulse field gel electrophoresis (PFGE) was used to follow chromosome structure following MMC treatment. Chromosomal DNA was prepared in agarose gel plugs and digested with the restriction enzymes PmeI to generate appropriately sized fragments for convenient PFGE analysis. Optimised conditions for PFGE of *H. volcanii* chromosomes were supplied by Dr J. DeRuggiero (Johns Hopkins University). Chromosome behaviour was monitored in wild-type and Δ*nreA* cells treated with MMC for 1 hour and during a 10 hour recovery period (**Figure 13**). As can been seen in **Figure 13**, DSBR occurs over a 6 hour period in wild-type cells and is apparently complete by 8 hours. In Δ*nre* cells, little or no repair is seen after 10 hours of recovery. This represents the first direct physical evidence for NreA's role in repair processes.

Figure 13 (below): PFGE analysis of DSB in wild-type and Δ*nre* cells treated with MMC for 1 hour. Repair is significantly delayed in Δ*nre* cells.



A3.5 The C-terminal PCNA-binding PIP motif is essential for NreA function

Taken together, the observed sensitivity to DNA damaging agents, the genetic interactions with known DNA repair pathways and, crucially, the observed delay in DSB repair visualised by PFGE confirm the first part of the hypothesis that formed the basis for the original application, that the Nre family of proteins were novel DNA repair factors. The second part of the hypothesis postulated that NreA function would be dependent on its ability to bind to PCNA via its C-terminal PIP motif.

To test this directly, mutant *H. volcanii* strains were constructed in which sequences encoding the last 10 or last 20 amino acids of NreA were deleted from the chromosome using established methods, resulting in expression of Nre proteins in which the C-terminal PIP motif was absent. These *nreA*-Δ*10* and *nreA*-Δ*20* mutants were then tested [1] for their sensitivity to MMC-induced DNA damage, [2] for their ability to repair DSBs as judged by PFGE and [3] for their genetic interactions with the various DNA repair pathways described above (by construction of double mutants and analysis of the double mutants response to MMC treatment). The results of these experiments were clear: that loss of the PIP motif phenocopied complete loss of NreA function i.e. that the *nreA*-Δ*10* and *nreA*-Δ*20* mutants were indistinguishable from Δ*nreA* in all three functional assays, i.e. *nreA*-Δ*10* and *nreA*-Δ*20* mutants were sensitive to MMC (**Figure 14** shows data for *nreA*-Δ*10* versus wild-type and an Δ*nre* deletion), this increased sensitivity was additive with Δ*rad50* Δ*mre11*, Δ*hjc*, Δ*hef*, Δ*uvrA*, Δ*uvrB*, Δ*uvrC* and Δ*uvrD* (as seen with the Δ*nre* deletion, see above), and *nreA*-Δ*10* and *nreA*-Δ*20* mutants were delayed in DSB repair as evidenced by PFGE (**Figure 15).**

Taken together, these results clearly demonstrate that *H. volcanii* Nre function is dependent on its ability to interact with PCNA via its C-terminal PIP motif.

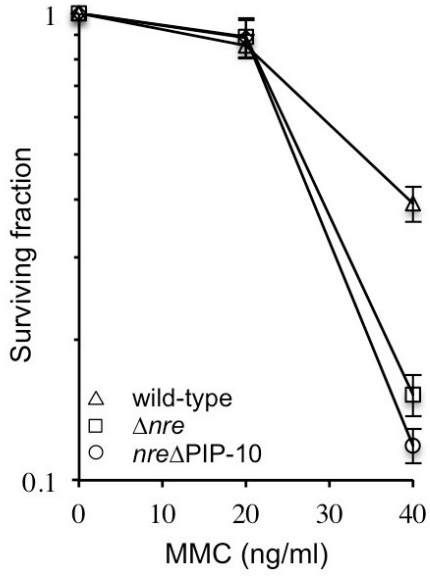
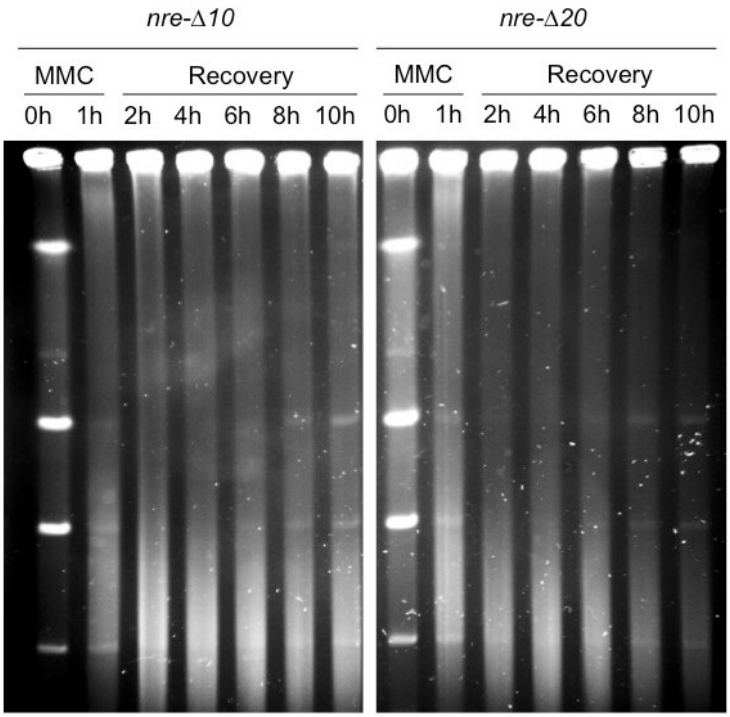


Figure 14 (above): Survival of wild-type, Δ*nre* and Δ*nre*-Δ*10* (here labeled *nre*ΔPIP-10) following MMC exposure.

Figure 15 (below): PFGE analysis of DSB in *nre*-Δ*10* and *nre*-Δ*20* cells treated with MMC for 1 hour. Repair is significantly delayed in both mutants relative to wild-type (cf. **Figure 13**).



A4 Overall conclusions

The original application for AFOSR funding hypothesised that the Nre proteins functioned in a novel PCNA-dependent DNA damage repair pathway, mostly likely as enzymes that act directly on damaged DNA to facilitate its repair. The results obtained during the course of the project allowed the former part of this hypothesis to be proven: NreA is indeed required for the efficient repair of DNA damage (sections A3.2-A3.4 above) and its function, which is related to that of UvrABC, is dependent on its ability to bind to PCNA (section A3.5).

However, the inability to obtain purified NreA or NreA proteins from either native or non-native hosts stymied efforts to determine the catalytic function of the proteins hypothesised in the application. Nevertheless, genes encoding Nre proteins are found in almost all sequenced archaeal genomes and in all cases the PIP motif is present at the C-terminal end of the protein. Thus it is highly likely that the PCNA-dependent DNA damage repair role identified in *H. volcanii* is a widespread feature of DNA damage repair processes in archaea.

B. Widening the net in the search for new DNA-directed enzyme activities

A central tenet of the funded project was the idea that it would be possible to use a candidate-based approach to identify new enzymes involved in the repair of DNA damage in archaea and that this would lead to the discovery of new mechanisms of DNA damage processing and potentially undiscovered types of DNA damage also (“Novel DNA Damage Repair Biology in the Archaea”). The confirmation that Nre plays a role in DNA damage repair (as predicted by the presence of the PCNA-binding motif and genomic co-localisation with *splB*) validates this approach, even though the precise molecular action of Nre remains unknown.

To extend our screen for new DNA-directed activities, a series of projects were initiated focused on additional candidate DNA replication and repair enzymes in *Haloferax volcanii*. For the most part, this work has been undertaken by graduate and undergraduate students (named below) with funding for consumables coming in part from the AFOSR grant. In such cases, the AFOSR funding will be acknowledged in the publications that result.

B1 Single-stranded DNA binding proteins

(Graduate student Agnieszka Skowyra, salary supported by the Scottish Universities Life Science Alliance, SULSA)

Single-stranded DNA binding proteins (SSBs) play vital roles in all aspects of DNA metabolism in all three domains of life and are characterised by the presence of one or more OB fold ssDNA binding domains. Three SSBs were identified in *H. volcanii* by database searching. Two of these (RpaA and RpaB) were shown to be individually non-essential for cell viability but to share an essential function, whereas the gene encoding the third (RpaC) was essential. Loss of RpaC function could be rescued by elevated expression of RpaB, indicative of functional overlap between the two classes of haloarchaeal SSB. Consistent with a role for RpaC in DNA repair, elevated expression of this protein leads to enhanced resistance to DNA damage. Taken together, these results offer important insights into archaeal SSB function and establish the haloarchaea as a valuable model for further studies. This work was published as Skowyra and MacNeill, 2011, *Nucl Acids Res* 40, 1077–1090, with the funding from the AFOSR duly acknowledged. The roles of these proteins in responding to ionising radiation damage are currently being analysed in collaboration with the laboratory of Jocelyn DiRuggiero (Johns Hopkins University).

B2 Genetic and biochemical analysis of family 4 uracil DNA glycosylase (UDG) enzymes

(Undergraduate students Kaori Kashi and Kotryna Temcinaite, research assistant Kacper Sendra, salary supported by the University of St Andrews)

The presence of uracil in DNA (as a result of its erroneous incorporation during DNA synthesis or the spontaneous deamination of cytosine to uracil) and its subsequent excision is a potent source of genetic instability. Uracil DNA glycosylase (UDG) enzymes play a vital role in combating this by removing uracil from DNA as a precursor to damage repair.

Remarkably, *H. volcanii* encodes four putative family 4 UDG proteins. One of these (Udg3) is conserved across all sequenced haloarchaeal species, while the others have a more sporadic distribution. To investigate the functions of these enzymes, all four genes were deleted, singly and in combination, and methods to assay UDG activity in *H. volcanii* cell extracts developed. In addition, the Udg3 enzyme was purified and shown to be an iron-sulphur protein with UDG activity *in vitro* under high-salt conditions. Working with *H. volcanii* cell extracts, Udg3 was shown to be responsible for all UDG activity detectable in extracts. A manuscript describing these findings with Udg3 is in preparation and will acknowledge AFOSR funding.

B3 The RecJ family: novel nucleases with unknown roles

(PI Dr Stuart MacNeill, undergraduate students Agnieszka Janska and Jason Woodier)

In bacteria, RecJ has an important role in DNA damage repair, in particular in the response to replication fork stalling caused by UV-induced DNA damage. RecJ is a 5’-3’ single-stranded DNA exonuclease. Recent work has identified an evolutionary relationship between members of the RecJ family and the Cdc45 protein, a key component of the eukaryotic DNA replication machinery, and has identified putative archaeal Cdc45/RecJ orthologues. During the period covered by this

report, the identity of the *H. volcanii* Cdc45/RecJ orthologue (named RecJ2) was confirmed by virtue of its interactions with another key replication fork component, GINS. Three additional RecJ family members (RecJ1, RecJ3 and RecJ4) were shown not to interact with GINS. Analysis of the function of these proteins is underway in the lab, with a variety of single and multiple deletion strains now constructed. Using *recJ1* deletion and nuclease-deficient *recJ1* strains, a role for one of these (RecJ1) in DNA damage repair has already been demonstrated. Further work is anticipated to reveal new pathways of DNA damage repair.

B4 MutS family enzymes

(Undergraduate student Anna Bota Cherino)

The MutS and MutL proteins, and MutS and MutL homologues in eukaryotes, play vital roles in DNA mismatch repair. *H. volcanii* encodes MutS and MutL homologues, named MutS1a, MutS1b, MutLa and MutLb, in addition to two more distantly related MutS proteins, named MutS5a and MutS5b. During the period covered by this report, strains in which the *H. volcanii mutS1a*, *mutS1b*, *mutS5a* and *mutS5b* genes were deleted were constructed as a precursor to genetic analysis of MutS protein function in mismatch repair. Similar work was carried out by the DiRuggiero lab (Johns Hopkins University, work also funded by the AFOSR). To avoid unnecessary duplication of effort, this project was put on hold. The constructed strains, however, are freely available to other researchers.

C. New tools for *H. volcanii* research

C1 Small molecule inhibitor of DNA ligase function

(Postdoctoral researcher Dr Xavier Giroux, salary supported by the AFOSR grant)

DNA ligases are essential enzymes in DNA replication, repair and recombination processes. Previously, the lab identified and characterised two DNA ligases encoded by the genome of *H. volcanii*, designated LigA and LigN (Zhao et al., 2006, *Mol Micro* 59, 743–752). The former is a typical archaeal ATP-dependent DNA ligase enzyme, while the latter is an NAD-dependent enzyme the gene for which was apparently acquired by lateral gene transfer from bacteria. Biochemical analysis of LigN function (Poidevin and MacNeill, 2006, *BMC Mol Biol* 7, 44) revealed an enzyme whose activity was dependent upon high salt concentrations (> 3M KCl), while genetic analysis revealed that LigA and LigN shared an essential function in *H. volcanii*: either gene could be deleted individually but the two genes could not be simultaneously deleted. Recently, researchers at AstraZeneca described the synthesis of novel NAD-dependent DNA ligase inhibitors (Mills et al., 2011, *Antimicrob Agents Chemother* 55:1088-96). To test whether these were effective in *H. volcanii*, wild-type, Δ*ligN* and Δ*ligA* strains were tested for sensitivity using an inhibitor supplied by AstraZeneca. As predicted, only Δ*ligA* strains (i.e. cells in which LigN alone is present) were sensitive to the drug, opening up the possibility of using this compound to probe diverse aspects of DNA replication and repair in Δ*ligA* cells, in Δ*ligA* cells exposed to sub-lethal levels of DNA damage and in Δ*ligA* cells in which specific DNA damage repair pathways are inactive.

C2 Tandem affinity purification cassette for *H. volcanii*

(PI Dr Stuart MacNeill)

Tandem affinity purification (TAP) is a well-established technique for protein purification under native (non-denaturing) conditions in eukaryotic cells (Puig et al., 2001, Methods 24, 218-229). The target protein is fused to a cassette encoding two protein affinity tags (two IgG binding domains from *Staphylococcus aureus* protein A and a calmodulin binding peptide) separated by a TEV

protease cleavage site. The first step of purification utilises the high affinity binding of protein A to IgG. Purified proteins are then released from the IgG matrix by TEV cleavage and then captured using a calmodulin matrix. Finally, bound proteins are gently released from the calmodulin matrix by calcium chelation (Puig et al., 2001, Methods 24, 218-229).

To apply this method to *H. volcanii*, to facilitate protein purification under native conditions, a novel cassette was constructed containing affinity tags capable of withstanding the extremely high salt conditions required for native purification from *H. volcanii* cell extracts. The cellulose binding domain from the *Clostridium thermocellum* cellulosome protein replaced the protein A tag in the classic TAP construct and a hexahistidine tag replaced the calmodulin binding peptide. The TEV protease site was retained. All three elements were codon-optimised for expression in *H. volcanii* and the cassette placed 3' to a strong constitutively active promoter (P_{fdx}). The resulting expression plasmid (named pH6CT, see **Figure 16**) is currently being used for protein purification in the lab and is freely available to other researchers.

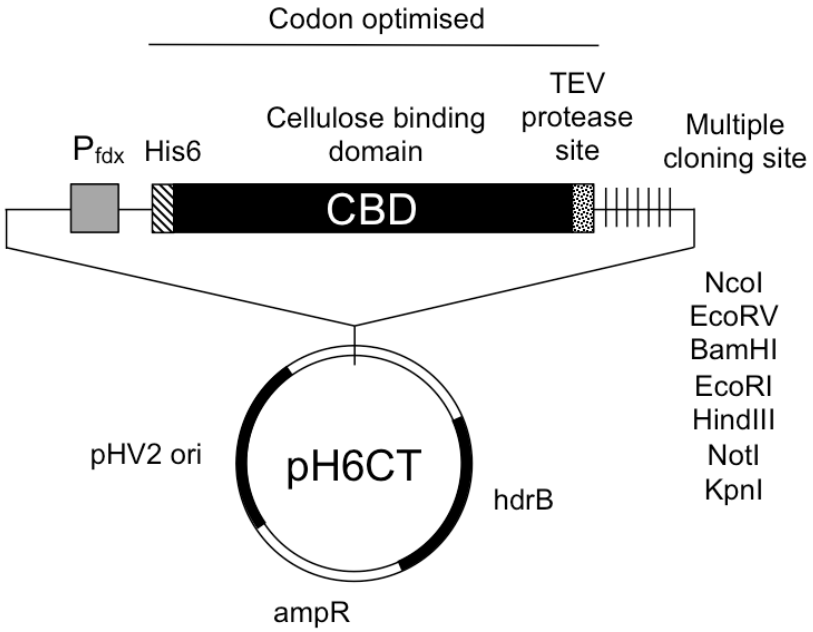


Figure 16 (above): Schematic representation of the expression plasmid pH6CT using for expression of proteins for tandem affinity purification (TAP) under native (high salt) conditions from *H. volcanii* cells extracts.