Defining the Genetic Basis of Ionizing Radiation Resistance

by Rose T. Byrne

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

(Microbiology)

at the

University of Wisconsin-Madison

2013

Examination: 08/12/13
This dissertation is approved by the following members of the Final Oral Committee:

Michael M. Cox, Professor, Biochemistry
Diana Downs, Professor, Bacteriology
James Keck, Associate Professor, Biomolecular Chemistry
Joseph Dillard, Associate Professor, Medical Microbiology & Immunology
Lloyd Smith, Professor, Chemistry

In order to enhance my medical school application, I participated in a summer research program in 2007 where I worked on a DNA repair project in the Foster Lab.

That project turned out to take much longer than I ever expected…

To my past mentors for changing the trajectory of my career and for teaching me how to think and write like a scientist:

Sarah Andres
Ashley (Ash) Williams
Kimberly Mauch Storvik

This is for you.

Acknowledgments

Every year, nestled in the cornfields that surround Bloomington, Indiana, there is a bike race. Some of the finest cyclists in the region race around a velodrome track at lightening fast speeds on single speed bikes in a 500-lap relay known as the Little 500. The Indiana University student body spends the entire week leading up to the race skipping classes and celebrating. Barack Obama stopped by in 2008 and every year various national acts perform during the week leading up to the event. Students at IU really love cycling you would think? Not quite - maybe 10% of the student body even shows up to watch the race. One year, some friends made shirts that read, ***"All of this for a bike race?"*** to encompass the general feeling that the build up to Little 5 is so much bigger than the actual race. I have the same general feeling about my time here at UW-Madison, with my thesis defense being my bike race and the build up to that bike race making up the most pivotal years of my life. During this build up, I have learned so much about myself and so much about what it takes and what it means to be a member of the scientific community. This thesis examination (or bike race) is a simply a celebration of what I have learned in the last four years and I'd like to thank everyone that has made the journey to this bike race so incredible.

I would first like to thank my advisor, Michael Cox. I have always felt like Mike had my back over the years and as a graduate student, that's a great feeling to have. I worked in DNA repair during my undergraduate career and actually came to Wisconsin to work for Mike. I'd like to thank him for believing in me and pushing me to complete this degree in the speedy nature that I did. I would also like to thank Jim Keck. Jim has a very approachable mentoring style and has been an invaluable resource to me. In 2012, Jim and I ran the Madison Marathon with American Cancer Society DetermiNation team. While I couldn't keep up with his crazy 3 hour marathon pace, I thank him for getting me involved with such a wonderful group of people. I would like to thank Diana Downs for being the head of my thesis committee, for writing me letters of recommendation, for being such a valuable genetics resource, and for being such a wonderful example of a strong woman in science. I would also like to thank Joe Dillard and Lloyd Smith for being contributing members to my thesis committee.

Next, I would like to thank all of the Cox Lab members, past and present. First, thank you Liz Wood, for being the cloning queen. If it weren't for Liz, my PhD would have taken 2 years longer. Thank you recent graduates, Asher Page, Khanh Ngo,

Marielle Gruineg, and Audrey Klingele, for always being there to answer questions during my first few years in the lab. Thank you to all the current Cox Lab members. Angie is literally the hardest working individual I have ever encountered. She set a bar of excellence that was always very hard to reach but always provided me that extra bit of motivation to go into lab on the weekend. Erin was such a wonderful addition to the lab and I consider her one of my closest friends. I will never forget turning onto the Washington Ave at mile 25 of my first marathon. It was rainy and I was upset that I hadn't seen any of my friends out on the course and there was Erin wearing something adorable and standing under a designer umbrella cheering me up the hill. Erin, thank you for always escaping lab with me for coffee breaks and girls lunches. You made the lab such a fun place to be. Thank you Stefanie Chen, for being an awesome and fun addition to the lab but more than anything, thank you for taking on YejH. It's a wonderful protein really and I'm sure you'll figure out what the hell it's doing. I would also like to personally thank my past undergraduate mentee, Eileen Moltzberger. Mentoring her and watching her go on to be a research assistant at MIT and then a graduate student at Duke has been the most fulfilling aspect of my graduate school career. Team YejH!

I would like to thank all my Madison friends for making the years of my PhD so incredibly fun. Thank you to my best friend and training partner, Kelly Stecker. Melody Berry and Emily Winkleman - thank you for keeping me laughing over the years. Thank you Trevor Nash for being the kindest and most supportive boyfriend a girl could ask for. I am thrilled he decided to move to Colorado with me for the mountain chapter of our relationship. To the various Berkeley Running Company marathon-training teams and the Endurance House Ironman training team, thank you for keeping me sane over the years and introducing me to the Madison community.

I dedicated my thesis to three very special people that are absolutely the reason I went to graduate school. Sarah Andres was my riding instructor and best friend. I was in high school while she was finishing a degree in Microbiology at Purdue University. She introduced me to this world. Ashley Williams was my first graduate student mentor in the Foster Lab. The first time I had to write a scientific proposal, Ash colored my entire draft in red marker, threw it at me, and told me for every two words I wrote to remove one. He showed me tough love but is the reason I love scientific writing. Ash has read every document I have ever turned in including my preliminary exam. He has an incredible command of the English language and because of this whenever I'm

writing I ask myself, "What would Ash hate about this sentence?" Kim Storvik was my second graduate student mentor in the Foster Lab. Kim and Ash had a good cop/bad cop thing going on. Kim was a very nurturing mentor and has always been there for me over the years. I consider her one of my closest friends.

Last but definitely not least, I would to thank my big crazy family. I would like to thank my mom and dad for bringing me into this world and helping me out when things got tough. These two have helped me move a total of 9 times (yes, I've counted). Thank you mom for fostering my passion for equestrian sports as a child and thank you dad for all the barns and fence building. I learned at a young age that to partake in the type of expensive hobbies that I enjoy (horseback riding and triathlon) I would need to make a decent paycheck and thus I would need to excel in school. I would like to thank my oldest sister, Dawn, for constantly tracking the weather here in Wisconsin and knitting me warm hats every time it dropped below freezing. Dawn has played a very special role in raising me and I could never put into words how grateful I am to her. I would like to thank my brothers, Bobby and Billy, for being awesome big brothers and for always being there when I needed them. My brother Billy just got into running and will be taking on his first marathon in September. I am so incredibly proud of him. I would like to thank Grace, my sister and best friend growing up. Grace is a mother of two beautiful girls and in my opinion, an awesome role model for young working mothers around the world. I would like to thank Mary Kay for pushing me for all of these years. The right thing to do would be to thank her for being such a great role model but if you knew our relationship that wouldn't quite do it justice, so I'd like to thank her for being such a hard to reach target. I'd like to thank her for going to graduate school and for running a 12 hr 49 minute Ironman. I'm lovingly shooting for 12 hrs and 48 minutes at Ironman Wisconsin this year and she knows I will be chasing her for that entire race. Being the youngest of six children makes it hard to be heard at times. While I thank my siblings for the love and support they have given me over the years, I cannot thank them enough for the major contributing factor to my success – an ingrained personal desire to succeed and a fierce competitive nature.

Thank you to anyone I haven't mentioned for making the writing of this thesis possible and for making my time in Madison a time in my life I will always cherish.

"All of this for a bike race?"

Abstract

Exposure to ionizing radiation can cause devastating and even lethal damage to cells. *Deinococcus radiodurans* as well as strains of directly evolved *Escherichia coli* have evolved mechanisms for dealing with this type of stress. The mechanisms responsible for this radiation resistance phenotype have been hotly debated. To address how cells survive exposure to ionizing radiation (IR) we have taken two approaches, a directed evolution approach and a genetic screening approach. By identifying the genetic innovations responsible for the acquisition of radiation resistance and by identifying all of the genes that contribute to survival of IR exposure, we become one step closer to understanding the genetic mechanisms responsible for this largely enigmatic phenotype. From the work discussed in this document we report: 1) Genetic innovations involving pre-existing but clearly malleable DNA repair functions can play a predominant role in the acquisition of an extreme IR resistance phenotype. 2) Recovery from high-level exposure to IR requires the function of at least 47 screen-identified genes, with 14 of these subjected to direct verification. The genes identified affect many different cellular processes. Contributions by genes involved in DNA repair, protein synthesis/stabilization, cellular responses to oxidative damage, cell wall structure and function, and genes of unknown function are important for IR survival. 3) A gene encoding a putative helicase of previously unknown function, *yejH*, strongly contributes to IR survival in *E. coli*. The substantial requirement for radiation resistance is the first

phenotype to be reported for *yejH.* YejH interacts with single strand DNA binding protein, SSB, providing a mechanism for which YejH is localized to damaged DNA and/or the replication fork. We propose the renaming of this gene to *radD* to note its important contribution to radiation resistance.

Table of Contents

List of Tables

Chapter 2: Evolution of Extreme Resistance to Ionizing Radiation via Genetic Adaptation of DNA Repair

Chapter 3: Surviving extreme exposure to ionizing radiation: *Escherichia coli* genes and pathways

List of Figures

Chapter 1: Introduction and Background

Chapter 2: Evolution of Extreme Resistance to Ionizing Radiation via Genetic Adaptation of DNA Repair

Chapter 3: Surviving extreme exposure to ionizing radiation: *Escherichia coli* genes and pathways

Chapter 1: Introduction and Background

1.1 The impact of radiation biology on human health

The damaging effects of IR discussed in this section have been used for decades in sterilizing food and in treating rapidly dividing cancer cells (*1*, *2*). Albeit effective, this cancer treatment damages healthy tissue, has a long list of side effects, and increases the probability of cancers later in life due to its mutagenic nature. While the work discussed in this document is focused on advancing our understanding of basic bacterial physiology, specifically the genetic basis of the extreme radiation resistance phenotype, it also has potential long-term implications in both preventing and treating cancer. Briefly, by advancing our understanding of how bacteria evolve resistance to radiation we can potentially predict evolutionary paths which rapidly dividing cancer cells could utilize to evolve resistance to treatment (Chapter 2). Secondly, by identifying genes involved in radioresistance (Chapter 3) we can potentially identify new human homologs to serve as targets for inhibitors, thus making radiation treatment more effective and lowering the dose of IR used in the clinic.

1.2 What is ionizing radiation?

Ionizing radiation, by definition, is radiation with sufficient energy to ionize biological molecules. IR is produced when radioactive elements decay. There are two types of IR: particulate (α- and β-particles) and electromagnetic (x- and γ-radiation) (*1*). Different forms of ionizing radiation deposit energy at different rates. α- and β-particles

deposit their energy within a short range after entering matter. On the other hand, γ-rays are photons that can pass through dense materials, such as tissue, and ionize molecules over a longer range. Ionized molecules react with other molecules to form radicals on a nanosecond scale. Each energy deposition event can generate 2-5 radical pairs within a radius of a few nanometers (*1*). Because of the devastating effects of γ-rays on biological molecules and the ability of these rays to penetrate cells, we focus our IR research on γ-radiation.

Biological molecules are susceptible to both direct and indirect damage caused by γ-radiation (*2*). Direct damage is caused by absorption of IR by a molecule, which in the case of DNA, can lead to strand breakage and chemical alterations of bases. In contrast, indirect damage occurs when IR is absorbed by water to produce reactive oxygen species (ROS), which interact with biological molecules. In aerobic and aqueous environments, IR generates three major radicals: •OH, H•, and e^- (*3*). The hydroxyl radicals (•OH) are the most immediate threat to DNA upon irradiation. Additionally, superoxide (O_2^-) is a byproduct of radiation exposure that often dismutates to form hydrogen peroxide (H_2O_2) which interacts with iron through Fenton chemistry to increase the concentration of hydroxyl radicals which can diffuse throughout the cell and cause further damage (*3*).

1.3 Cellular responses to oxidative stress

E. coli encodes the following two ROS detoxification enzymes to manage endogenously produced ROS: superoxide dismutase (SodA) and catalase (KatE) (*4*). SodA catalyzes the dismutation of superoxide to hydrogen peroxide, which is then further detoxified by KatE, which catalyzes the decomposition of hydrogen peroxide to water and oxygen. These are basal oxidative defenses in all aerobic organisms.

There are two systems in *E. coli* known to respond to ROS stress: the SoxRS and the OxyR system (reviewed in (*5*)). In the mid-1980s, the Demple and Weiss labs independently demonstrated that the sensor protein SoxR is activated by the oxidation of its [2Fe-2S] cluster (*6-8*). When SoxR is oxidized by redox-cycling drugs such as paraquat, the protein transcriptionally activates SoxS, a transcription factor responsible for activating superoxide dismutase (SodA), oxidation-resistant dehydratase isozymes, proteins involved in iron sulfur cluster repair, drug efflux pumps, and Fur, an iron-uptake regulatory protein (*9*). The SoxR response can be rapidly turned off upon removal of oxidative stress by the reduction of the SoxR iron-sulfur cluster to its inactive form by the *rsxABCDGE* and *rseC* genes (*10*). These genes reset or prime SoxR for another response and maintain the reduced state of SoxR against oxidation during aerobic growth.

The OxyR system responds to hydrogen peroxide stress. The *oxyR* gene encodes a transcription factor and was first discovered in a selection for hyper-H_2O_2 resistant *Salmonella (11)*. The oxidation of two cysteine residues (Cys-199 and Cys-208) by H_2O_2 activates OxyR which then directly interacts with RNA polymerase and positively regulates the expression of catalase (KatE), a manganese importer (MntH), proteins that reduce disulfide bonds, proteins involved in iron-sulfur cluster assembly, aforementioned Fur, and the DNA binding ferritin-like iron-scavenging protein, Dps (*12*) (*13*). When H_2O_2 stress subsides, the OxyR system is turned off by glutaredoxin 1, encoded by *grxA*, which quickly reduces the Cys-199/Cys-208 disulfide bonds (*14*).

1.4 Energy deposition events and their damage to biological molecules

High doses of IR cause an influx of free radicals that often overwhelm the SoxRS and OxyR systems. The effects of high-energy rays and free radicals within a cell are too numerous to discuss thoroughly in this document, but are thought to include numerous types of DNA base damage and breaks, protein ionization, DNA cross-linking, DNA-protein crosslinking, and protein-protein crosslinking. Discussed below are the molecules most frequently reported in the literature as lethal targets of radiation exposure, as well as the pathways known to contribute to repair of these molecules.

1.4.1 IR damages DNA

Radiation causes different types of DNA damage, depending on which part of the molecule the energy deposition event occurs. Hydroxyl radicals, produced *via* the indirect damage path, abstract hydrogen atoms, but more rapidly they add to double bonds, resulting in two major modes of DNA attack 1) hydrogen abstraction from deoxyribose sugars and 2) addition to the double bonds of DNA bases. These two modes of attack result in strand breaks and base damage and/or loss (reviewed in (*15*)).

IR-induced base damage

When the energy of a γ-ray or radical is deposited into a nucleobase, base damage occurs. Damage to bases by ionizing radiation has been studied extensively *in vitro* by the irradiation of free bases in aqueous solution, in aerobic and anaerobic

conditions. From these studies, the chemistry of damaged bases/lesions and their formation have been described (*2*, *16*, *17*). Ionizing radiation induced base damage often results in distortion of the normal helical structure of DNA. These distortions often stall or halt the replication fork and thus cells have evolved mechanisms to repair damaged bases. The base excision repair (BER) pathway repairs base damage and is mediated by a class of DNA repair enzymes called DNA glycosylases (*18*, *19*) . These enzymes cleave the *N*-glycosyl bond that links the damaged base to the sugar-phosphate backbone resulting in a apurinic/abasic (AP) site. The resulting AP site is then repaired by a class of enzymes known as apurinic/apyrimidinic (AP) endonucleases which recognize AP sites and nick the phosphodiester backbone directly 5′ of the AP site. The 5′ terminal deoxyribose-phosphate residue is then removed by an exonuclease to generate a single nucleotide gap in the DNA duplex. The missing nucleotide is then filled in and the backbone sealed in processes called repair synthesis and ligation (reviewed in (*20*)). There are numerous DNA glycosylases encoded by *E. coli* and each respond to different types of DNA lesions.

BER differs from another pathway important for removing damaged bases called nucleotide excision repair, or NER, by the size of the product that is excised. In BER, the base alone is excised whereas in NER, a small oligonucleotide is removed *via* a different cellular pathway. In the process of NER, the damaged base is recognized, a cut is made

on either side of the lesion, the resulting oligonucleotide containing the lesion is excised, the gap is filled in by repair synthesis, and the new DNA is ligated to the old. This function is encoded by *uvrABCD* in *E. coli (21-24).* Briefly and as illustrated in Figure 1, a UvrA dimer binds a UvrB monomer which is then thought to scan the DNA using a DNA helicase activity until a lesion is found. When a damaged base is encountered, UvrA disassociates leaving UvrB bound to the lesion and producing a bent DNA structure. UvrC is recruited to the structure and nicks the DNA on either side of the lesion before the helicase UvrD is recruited to release the oligonucleotide. DNA polymerase I and DNA ligase are required for the subsequent repair synthesis and ligation steps (reviewed in (*25*)).

The UvrABCD system is also implicated in the transcription-coupled repair (TCR) pathway. When the transcription machinery comes in contact with a damaged base, it often stalls, inhibiting transcription and forming a roadblock on the DNA that can halt the replication machinery as well. TCR is mediated by Mfd, a protein that is recruited to stalled RNA polymerases (RNAP) *via* a direct interaction with the stalled complex. Upon binding to RNAP, Mfd removes RNAP and the nascent mRNA from the DNA template and then recruits the UvrABCD complex to preferentially repair the lesion (*26-29*).

One of the most extensively studied lesions to date is 8-hydroxyguanine (8-oxoG), which is the result of oxidation at C8. This oxidation can cause the inversion of the *N*-glycosidic bond to form a stable base analog that can incorrectly base pair with adenine as illustrated in Figure 2A-C. The resulting mispair is not easily detected by the proofreading system, and a high frequency of G ➔ T transversions results (*30*), thus increasing the mutation frequency within the cell. The oxidative damage repair system or 'GO' system, encoded by *mutTMY*, is a specialized system for dealing with 8-oxoG and is diagrammed in Figure 2D. MutT is a oxoGTP hydrolase that hydrolyzes oxo-dGTP to oxo-dGMP thus working upstream of replication and preventing the incorporation of 8-oxoG into the chromosome by cleansing the nucleotide triphosphate pool (*31*). MutM is an 8-oxoG *N*-glycosylase that cleaves the 8-oxoG from the sugar-phosphate backbone resulting in an AP site that is repaired via BER. MutY is an adenine *N*-glycosylase that removes adenine base paired with 8-oxoG leaving the 8-oxoG to pair with a cytosine (*32*).

The NER, BER, TCR, and oxidative damage repair systems are mechanisms that cells have evolved to overcome oxidative base damage, and the importance of some these systems in surviving IR exposure become clear in Chapter 2 and 3.

IR-induced strand breaks

While nucleobases are damaged most frequently, 20% of the time the sugar-phosphate backbone is affected (*33*). When energy is deposited in the DNA backbone, strand breaks occur. Double strand breaks (DSBs) are generated when multiple single strand breaks occur within close proximity, a cluster of hydroxyl radicals introduce strand breaks in both strands at one location, or when cells attempt to enzymatically excise damaged bases present in close proximity on both strands (*1*). DSBs are the most lethal form of DNA damage because they halt DNA replication, cause the collapse of the replication fork, often result in the loss of genomic material, and are difficult to repair (*1, 34*, *35*) .

Double strand break repair (DSBR) and gap repair are highly effective cellular processes that repair strand breaks. These processes rely on homologous recombination, or the exchange of DNA between two different DNA molecules. Homologous recombination is a highly faithful, nonmutagenic mechanism of DNA repair that is catalyzed by the central bacterial recombinase, Recombinase A (RecA) enzyme. This process has been reviewed in depth (*36*) (*37*) (*38*).

The RecBCD pathway mediates DSBR by producing the preferred substrate of RecA and by loading RecA (*39*) (*40*). As illustrated in Figure 3, DSBR is initiated by the binding of RecBCD to the broken dsDNA end. RecBCD, through its combined helicase

and exonuclease functions, unwinds and preferentially degrades the 3′ strand. When the protein complex reaches a chi site (5′- GCTGGTGG-3′) the complex switches to degrading the 5′ strand thus producing a 3′ overhang (*40*). RecA is then loaded onto the 3′ overhang, producing an active nucleoprotein filament which can find and invade an intact homologous double stranded DNA displacing one of the strands to form a D-loop structure. The DNA from the other side of the break is processed similarly to create a second RecA filament, and the original displaced strand is captured by the second RecA filament resulting four-stranded DNA intermediate is called a Holliday Junction (HJ). The 3′ ends serve as primers for continued synthesis of the once broke DNA. The HJs are migrated by RuvAB and RecG and/or resolved by RuvC. Repair of DSB can become complicated by the structure of the break, which is known to be different based on how the break arose. The complex nature of repair may be why mutations in *recN*, *recF*, *recJ*, *radA*, and *uvrD* were shown to disrupt DSBR (*41*). RecN and RadA are particularly interesting because while these proteins are important for DSBR, no function has been assigned to these proteins. The *recN* gene is one of the most highly expressed genes in the SOS response (*42*), the transcriptional response to DNA damage, and shares sequence homology with the eukaryotic chromosome cohesins and condensins that make up the structural maintenance of chromosomes (SMC) family of

proteins. These results suggest that RecN may contribute to effective DSBR by holding together DNA strands (*43*, *44*).

When two single strand breaks occur on the same strand or when a single strand break is processed into a gapped DNA substrate, cells utilize the RecFOR pathway for initiation of homologous recombination at the damage site (*45*). The model proposed for this pathway is as follows: the RecQ helicase and the RecJ exonuclease process the gapped substrate. The single stranded DNA binding protein (SSB) coats the single stranded DNA gap. RecF recognizes the ssDNA-dsDNA junction, is thought to recruit RecOR, and then the RecFOR complex serves to nucleate RecA on the substrate. SSB is displaced and homologous recombination occurs as described for DSBR (*45*).

The importance of proteins that mediate recombination repair are in concordance with a theme that carries through the entirety of this document: at the heart of radiation repair is efficient and effective DNA repair.

1.4.2 IR damages Proteins

Protein Carbonylation

The most well studied irreversible protein oxidation is carbonylation. Protein carbonylation results in covalent modifications of lysine, arginine, proline, and threonine side chains and can result in inactivation of essential proteins and increased susceptibility to aggregation and proteolysis (*46*) (*47*).

As carbonylation is an irreversible chemical modification, there are no reported pathways for repairing carbonylated proteins. However, avoidance of protein carbonylation has been proposed to be highly important for surviving radiation exposure. A higher amount of manganese, especially in relation to iron (high Mn/Fe ratio) is correlated with lower levels of detected protein carbonylation. The observation that many radioresistant organisms have a high Mn/Fe ratio has led to the proposal by M. Daly that remediation of protein oxidation plays a primary role in IR resistance (*48*). Daly and colleagues propose that the abundant Mn(II) in these organisms is used to scavenge ROS and thus protect proteins from inactivation (*49*).

Protein turnover is an important mechanism for avoiding the deleterious effects of protein carbonylation as well. Carbonylated proteins tend to aggregate, and aggregation has been linked to aging and numerous diseases (*50*). Proper degradation of damaged proteins is important for cell viability after IR exposure. This theme is present in both Chapter 2 and 3 of this document.

1.5 The phenomenon of radioresistance

The phenotype of extreme radiation resistance in prokaryotes is particularly puzzling because there are no known natural environments where organisms are subjected to extreme levels of radiation as a natural selection. Many theories have been put forward, some even proposing that radiation resistant organisms were once Martian life forms that made it to earth *via* a galactic meteorite adventure (*51*). A more commonly accepted theory is that the mechanisms required to survive IR exposure evolved to manage other stresses that bacteria face, for instance desiccation, which has been reported to result in an influx in ROS and DSBs (*52*). As mentioned above, surviving IR exposure involves mediating the damage associated with free radicals, which are also natural byproducts of cellular metabolism. It is possible that cells have evolved mechanism for ameliorating the damage of these metabolic byproducts and that extreme radiation resistance is fortuitous consequence of an efficient and effective response to these damaging radicals.

1.5.1 *Deinococcus radiodurans*

D. radiodurans was first isolated in 1956 from canned ground beef that had been γ-irradiated with a dose 250-fold higher than that used to kill *E. coli (53)*. To illustrate just how radiation resistant this organism is, a typical human can survive approximately 3 Gy with the dose of **5 Gy** being absolutely lethal (*54*). *D. radiodurans*, however, can

survive up to **5,000 Gy** without losing any viability (*55*). Previously named *Micrococcus radiodurans, D. radiodurans* is a red pigmented, non-sporeforming, culturable, non-pathogenic, transformable obligate aerobe (*56*). *D. radiodurans* is commonly studied because of its ability to survive tremendous genetic insult by high doses of IR despite the induction of hundreds of DSBs. This unusual ability resulted in the genus name – *Deinococcus-* which is Greek for strange or unusual berry (*57*). The error-free mechanisms of DSB repair in *D. radiodurans* has captivated a field of scientists. While these mechanisms are not yet fully understood, they appear to require both passive and enzymatic processes.

1.5.2 Mechanisms of Radioresistance

Because of the extreme nature of this phenotype and the potential for using *D. radiodurans* for bioremediation purposes, there is demand for understanding the mechanisms underlying the ability of *D. radiodurans* to faithfully repair itself after extreme radiation insult. The field has put forward the following potential mechanisms, both enzymatic and passive.

Passive Mechanisms

1) High cellular manganese concentration

D. radiodurans accumulates high intracellular manganese levels compared to radiation sensitive bacteria (*49*). Daly et al, observed that when starved for Mn(II), *D.*

radiodurans loses radioresistance. Mn(II) scavenges ROSs, an activity reported in *Lactobacillus plantarum* which lacks superoxide dismutase and relies on Mn(II) alone for protection (*58*). Because Mn(II) starvation does not reduce the number of DSBs, Mn(II) likely promotes DNA damage tolerance rather than DNA protection (*49*). There are currently two hypotheses for the role of high cytoplasmic Mn(II) levels in conferring radioresistance. Daly *et al.* proposed that Mn(II) prevents protein damage by ROSs in *D. radiodurans*. Another hypothesis is that the positive charge of Mn(II) may promote condensation of the *D. radiodurans* nucleoid by neutralizing the repulsive forces between the phosphates in the DNA backbone (*59*). The possible benefits of a condensed nucleoid will be discussed below. It is likely that both mechanisms are important.

2) Genome copy number

DSBs are often repaired by an error-free mechanism initiated by homologous recombination (HR); however, such repair requires an undamaged template (*60*). *D. radiodurans* has 8-10 genome equivalents during exponential growth and four during stationary phase (*61*). Because more than one copy is present at all times, the likelihood that an undamaged template is available for HR-dependent repair is increased. Furthermore, inactivation of a gene by IR-induced damage is compensated for by additional copies of the gene (*62*) (*63*).

3) Nucleoid Structure

In 2003, Levin-Zaidman *et al.* showed that the *D. radiodurans* nucleoid assumes a condensed toroidal morphology (*59*). They proposed that the condensed structure might restrict diffusion of DNA double-strand ends, keeping them close together to facilitate repair. Zimmerman and Battista (2005) challenged this hypothesis when they reported that appearance of toroidal nucleoid morphology under different growth conditions does not always confer IR resistance. The authors surveyed nucleoid morphology of radioresistant organisms in the *Deionococcus* and *Rubrobacter* genera and concluded that condensation of the genome, rather than its shape, contributes to radioresistance (*64*).

Enzymatic Mechanisms

The extraordinary faculty of *D. radiodurans* to repair DSBs relies heavily upon extremely efficient, yet not fully understood, enzymatic repair mechanisms. As described by Daly and Minton (1996), *D. radiodurans* repairs its fragmented genome in two distinct phases: a RecA-independent process that repairs one third of the IR-induced DSBs within the first 1.5 hours post-irradiation, and a RecA-dependent process that repairs the remaining breaks (*65*).

1) RecA-independent Genome Reconstitution

The following proposed mechanisms (some outlined in Figure 4) were summarized by Zahradka *et al.* as ways that the cell could rejoin hundreds of genome

fragments in a *recA*-independent fashion (*66*): nonhomologous end joining (NHEJ) (*67, 68*), intra- and interchromosomal single-strand annealing (SSA) (*69*); synthesis- dependent-strand annealing (SDSA) (*69, 70*); break-induced replication (*69, 71*); and copy choice (the switching of DNA replication from fragment to fragment) (*72*).

Most of these mechanisms were dismissed when Radman and colleagues reported that post-irradiated cells contained high levels of newly synthesized DNA. From this work, Extended Synthesis-Dependent Strand Annealing (ESDSA) was proposed as the favored model (*66*). ESDSA (Figure 4) requires DNA polymerase A (DNA pol I). As the model proposes, immediately after IR exposure, chromosomal fragments with overlapping homologies are used as both primers and templates for a strong burst in DNA synthesis by DNA Pol I, which results in complementary single-stranded extensions that can act as 'sticky ends' to anneal and thus join together genome fragments.

2) Homologous Recombination in *D. radiodurans*

Because complete genome reassembly in *D. radiodurans* is RecA-dependent, homologous recombination must play a crucial role in radioresistance. This RecA-dependence was first demonstrated when a radiation sensitive *D. radiodurans* mutant was isolated and the mutation was mapped to the *recA* gene. This mutant, termed *rec30*, and strains of *D. radiodurans* disrupted for *recA* were highly sensitive to IR exposure,

with a 70-fold reduction in the dose required to yield survival of 37% of colony forming units (D_{37}) relative to that of the wild-type (*73*). The *E. coli recA* gene does not complement the *D. radiodurans recA* null mutant (*74*), suggesting that *D. radiodurans* RecA (DrRecA) is intrinsically different from *E. coli* RecA (EcReca) and that this difference contributes to extreme IR resistance.

The DrRecA protein is 57% identical and 73% similar to the EcRecA; however, properties of the two proteins differ in some *in vitro* studies. For example, the DrRecA protein promotes DNA strand exchange with a preferred order of addition for DNA substrates opposite of that seen with EcRecA (*75*); rather than binding single strand DNA first, DrRecA preferentially initiates strand exchange from the duplex DNA substrate (*75*). The properties of DrRecA with respect to filament formation are also distinct. The EcRecA protein nucleates filament formation relatively slowly, and extends the filaments more rapidly, a property that would tend to localize much of the available RecA protein in a single filament. This may reflect a repair system that typically must deal with one or only a few situations requiring recombinational DNA repair in each cell cycle. In contrast, the DrRecA protein nucleates more rapidly and extends the filament more slowly, a property which would tend to create large numbers of shorter filaments (*76*). These properties are consistent with a system that must deal with hundreds of DSBs after desiccation or extreme doses of ionizing radiation. Nucleation and

filamentation properties of DrRecA, EcRecA, and the RecA variants from hyper radiation resistant *E. coli* mutants are explored in Appendix A.

3) Novel Repair Pathways

Novel DNA repair mechanisms may also contribute to the extreme IR resistance of *D. radiodurans*. Tanaka and Early (2004) identified 72 genes upregulated at least 3-fold in response to irradiation and desiccation (*77*). The most highly upregulated genes under both stress conditions had unknown functions and were named *ddrA*, *ddrB*, *ddrC*, *ddrD*, and *pprA* (*77*). Inactivation of these genes caused IR-sensitivity, suggesting that they play novel roles in radioresistance. Genetic analysis by Tanaka and Earl showed that *pprA* and *recA* are epistatic, indicating PprA interacts with RecA genetically in a novel DNA repair pathway (*77*).

Further characterizing the newly found genes, *ddrA* was investigated. Deletion of *ddrA* increases radiation sensitivity slightly; however, when the *ddrA* knockout cells are starved, the cells become over 100-fold more sensitive to radiation than wild type. *In vitro*, DdrA protects the 3' ends of DNA fragments from nucleolytic degradation. The current hypothesis is that by protecting DNA ends, cells can preserve their genomes until their environment is more conducive for growth and repair (*78*).

1.6 Conflicting theories of radiation resistance, the problem with molecule centric ideologies, and the need for a global approach

Ongoing research into the molecular basis of extreme IR resistance has suggested three potential classes of mechanisms (discussed above). These are (A) a condensed nucleoid structure, which could potentially facilitate DNA repair processes (*59, 64*), (B) an enhanced capacity for amelioration of protein damage, which could protect DNA repair systems and make them more readily available following irradiation (*79, 80*), and (C) potential specialized pathways for DNA repair (*66*).

The amelioration of protein oxidation has received extensive experimental support as a mechanism to account for much if not all of the observed IR resistance in *D. radiodurans* (*49, 79-82*). Increases in cytosolic antioxidant capacity, particularly an increase in the cellular Mn/Fe ratio (*49*), appear to make the major contributions. Increasing the Mn/Fe ratio limits the Fenton chemistry that produces ROS (*79, 82*). The hypothesis that the condensed nucleoid of *Deinococcus* facilitates efficient DNA repair has been questioned, since the presence of condensed nucleoids does not correlate reliably with radiation resistance in bacterial species (*64*).

Repair of IR-induced DNA damage is clearly important to survival. However, the constellation of DNA repair functions in *D. radiodurans* is unremarkable by bacterial standards. *D. radiodurans* and other IR resistant species appear to have approximately

the same toolbox of DNA repair pathways as non-resistant species. Thus, the argument has been advanced that specialized DNA repair plays little or no role in IR resistance (*79, 80*). Instead, antioxidants prevent protein oxidation and render the classical DNA repair systems more readily available to correct the effects of IR.

The three proposed theories of resistance have one thing in common, their molecule-centric nature. Although radiation damages all cellular components, current research tends to focus on either DNA damage or protein oxidation as the lethal antagonist of cell viability. The work outlined in this document was performed to provide the field with a more global understanding of how an organism survives radiation exposure. The major results highlighted in this work indicate that surviving radiation exposure requires a global cellular effort consisting of efficient and effective DNA repair, properly regulated cell division, maintenance of cell wall biosynthesis and structure, functional oxidative stress signaling, protein turnover, and potentially new cellular processes performed by several genes of unknown function. These findings were discovered by exploiting two powerful bacterial techniques: directed evolution (Chapter 2) and high-throughput genetic screening (Chapter 3).

1.6.1 Using directed evolution to understand the genetic basis of IR resistance

In 1946, Evelyn Witkins reported that *E. coli* B could acquire mutations that increased radioresistance. Cultures were exposed to doses of radiation that killed 99% of

the population and the surviving 1% was harvested and analyzed for radioresistance. Cells that withstood the first irradiation were significantly more radioresistant than non-irradiated cells (*83*). Similar studies have shown that other organisms can also acquire radioresistance when repeatedly exposed to IR (*84-87*). Davies and Sinsky generated extremely radioresistant *Salmonella enterica* serovar Typhimurium by repeatedly exposing cells to γ-radiation. These mutants were 200-fold more radioresistant than wild type and showed resistance similar as *D. radiodurans* (*85*).

In 2009, Harris *et al.* generated and characterized four extremely radioresistant populations of *E. coli* (*87*). To further elucidate the molecular basis of radioresistance, the specific mutations in these strains were identified using high-throughput sequencing technologies. The idea behind employing this method was that by directly evolving radiation resistance, the cell themselves would reveal how radioresistance evolves (*87*). As the Harris study is the foundation for Chapter 2, I have highlighted the experimental design and major findings in the following sections.

Generation of IR resistant *E. coli* populations. MG1655 was subjected to 20 iterative cycles of irradiation and outgrowth as illustrated in Figure 5. The resulting radioresistant population was designated IR-1-20 and 62 non-identical strains were purified and analyzed. From these strains, a single radioresistant isolate, designated CB1000, was chosen for further analysis. The 20 iterative cycles of irradiation were

performed independently three more times to generate the radioresistant populations IR-2-20, IR-3-20, and IR-4-20. From each of these populations, one radiation resistant single colony isolate was purified and designated CB2000, CB3000, and CB4000, respectively.

Radioresistance of Evolved Strains. Exposure to a dose of 3,000 Gy reduced viability of the Founder strain by four orders of magnitude, while CB1000, CB2000, and CB3000 exhibited 1,500- to 4,500-fold increases in survival relative to the Founder (Figure 6). Based on viability analysis, the evolved strains were nearly as radioresistant as *D. radiodurans,* and after 100 generations without selection, the radioresistance phenotype was maintained. Previous bacterial evolution studies provide evidence for adaptation occurring by sequential base substitutions that appear as successive steps (*88-90*). In agreement with this observation, the radioresistance was acquired in a stepwise manner where each step is most likely fixation of beneficial mutation (Figure 7).

Initial Mutational Analysis of Evolved Strains. In this initial study, microarray-based comparative genome sequencing (Nimblegen Systems Inc.) was used to detect mutations in the Founder, CB1000, CB2000, and CB3000 and Illumina sequencing was used to directly sequence the entire genomes of Founder, CB1000, CB2000, CB3000, and six additional isolates from IR-1-20. Combinations of sequencing methods were used to

increase the likelihood that all mutations were detected. Each evolved strain contained 40 to 70 genomic basepair alterations, and 50 to 60% of these mutations were non-synonymous. Figure 8 is a mutation map for isolates CB1000, CB2000, and CB3000. This figure indicates that there were no mutational hotspots but rather that mutations occurred randomly across the genome and thus there is not a universal mutational profile responsible for the observed radiation resistance of these strains. By analyzing all of the mutations sequenced in this initial study, the authors pointed out that mutations occurred frequently in *recA* (the recombinase), *ruvB* (involved in homologous recombination), *dnaT* (involved in replication restart and double strand break repair), *ftsZ* and *ftsY* (important players in cell division), *clpP* and *clpX* (components of the protein degradosome), *gltS* (a glutamate transporter), *bglH* (a membrane porin for *b*-glucosides), and two genes of unknown function, *yjgL* and *ylbE*. The exact path to resistance was not determined in thus study and is the focus of my work described in Chapter 2. While this study examined the IR-1-20 population, the work discussed in Chapter 2 focuses on the IR-2-20 population which has been more amenable to this type of analysis. IR-1-20 is a complex population that still has multiple subpopulations competing with one another whereas most mutations are fixed in IR-2-20, providing us a more homogenous population to study. Figure 9 highlights the strains sequenced in this

initial Harris 2009 study and also lists the additional isolates we sequenced and analyzed in Chapter 2.

Mutations in *recA*. In the Harris study, it was reported that mutations in *recA* appeared in 17 of the 23 IR resistant strains, suggesting that RecA modifications contribute to IR resistance. Three *recA* alleles, which encode RecA A289S and RecA D276N/A, were studied. Each allele was moved into the Founder strain and their effects on survival of 2,000Gy of IR were significant. All tested *recA* alleles increased the IR resistance of the Founder strain. Recent work (Appendix A) to biochemically characterized the RecA D276A/N variants provide a molecular logic that appears to conform in a reasonable way to the molecular requirements for genome reconstitution after extreme exposure to ionizing radiation. The two studied RecA protein variants form filaments on single strand DNA, hydrolyze ATP, and promote DNA strand exchange, as does the wild type protein. However, the RecA D276A/N proteins (a) exhibit a slower rate of filament extension, leading to a distribution of the protein among a higher number of shorter filaments, (b) promote DNA strand exchange more efficiently in the context of a shorter filament, and (c) are markedly less inhibited by ADP. These adaptations potentially allow the RecA protein to address larger numbers of DSBs in an environment where ADP concentrations are higher due to a compromised cellular metabolism (Appendix A).

Further Characterization of Evolved Strains. The growth rates of strains CB1000 and CB2000 were unaffected in rich and minimal media in comparison to the Founder strain. The growth rate of CB3000 was unaffected in rich medium; however, it is an arginine auxotroph and cannot grow in minimal media. All of the strains were at least ten-fold more resistant to ultraviolet radiation. CB1000 and CB2000 were more resistant to the DNA damaging agent mitomycin C than the Founder strain; in contrast, the mitomycin C resistance of CB3000 was unchanged. No changes were observed in cell or nucleoid morphologies. Also, there were no changes to cytoplasmic Mn(II), DNA, and nucleotide concentrations. Neither passive DNA protection nor obvious protection from protein oxidation was observed.

1.6.2 Using high-throughput genetic screening to define the genetic basis of radiation resistance

In his 2009 review, Thomas Silhavy coined the phrase "high-tech genetics" to describe techniques that combine the power of traditional genetics with large-scale methods (*91*). Due to massive sequencing efforts and these new high-tech genetic methods, we have elucidated a nearly comprehensive list of genes essential for growth in *E. coli* (*92*); however, the functions of nearly one third of the genes encoded by *E. coli* are still unknown. The Y-gene nomenclature assigns temporary systematic names to these genes of unknown function (*93*). Once a Y-gene is functionally characterized *in vivo* and/or *in vitro*, a functionally related gene name is assigned. To facilitate the

elucidation of unknown gene functions, we must find growth conditions or stress conditions where these poorly characterized genes are optimal for growth.

The high-throughput genetic screen described in my work (Chapter 3) utilized the Transposon Directed Insertion Sequencing (TraDIS) method, developed by Langridge and colleagues (*94*). This technique uses transposon mutagenesis, competitive outgrowth (IR survival), and Illumina sequencing from a transposon-specific primer. Due to the selective pressure of IR exposure, individual mutants will decrease or increase in abundance during outgrowth. As radiosensitive cells die, genes that are required for IR resistance will have decreased representation in the IR-treated mutant pools versus the untreated mutant pools, revealing their importance for survival of IR exposure. Using this screen, nearly all non-essential genes can be tested for their contribution to IR resistance in several parallel experiments. From this work we report the first phenotype to date for the gene of previously unknown function, *yejH*. This gene contributes dramatically to the survival of IR exposure and the work described in Chapter 4 describes the preliminary work to determine the function of the *yejH* gene product by biochemical and genetic characterization.

By combining the directed evolution approach described in Chapter 2, the genetic screening approach in Chapter 3 and the further characterization of the new radiation resistance factor, YejH, in Chapter 4, we have begun to define the genetic basis

of radiation resistance and thus are one step closer to a complete understanding of how cells defend themselves against exposure to ionizing radiation.

Figures

Figure 1: The Nucleotide Excision Repair Pathway of *E. coli.*

The UvrA dimer ('A' circles) binds UvrB ('B' square) and scans the genome until it finds a lesion (represented here as a pyrimidine dimer, TT). Upon recognizing the lesion, $UvrA_2$ disassociates leaving UvrB to recruit UvrC ('C' triangle) which nicks the DNA on either side of the lesion. UvrD ('D' circle) then excises the oligonucleotide, DNA polymerase I ('I' circle) fills in the gap, and ligase fills in the gaps. Image adopted from (*95*).

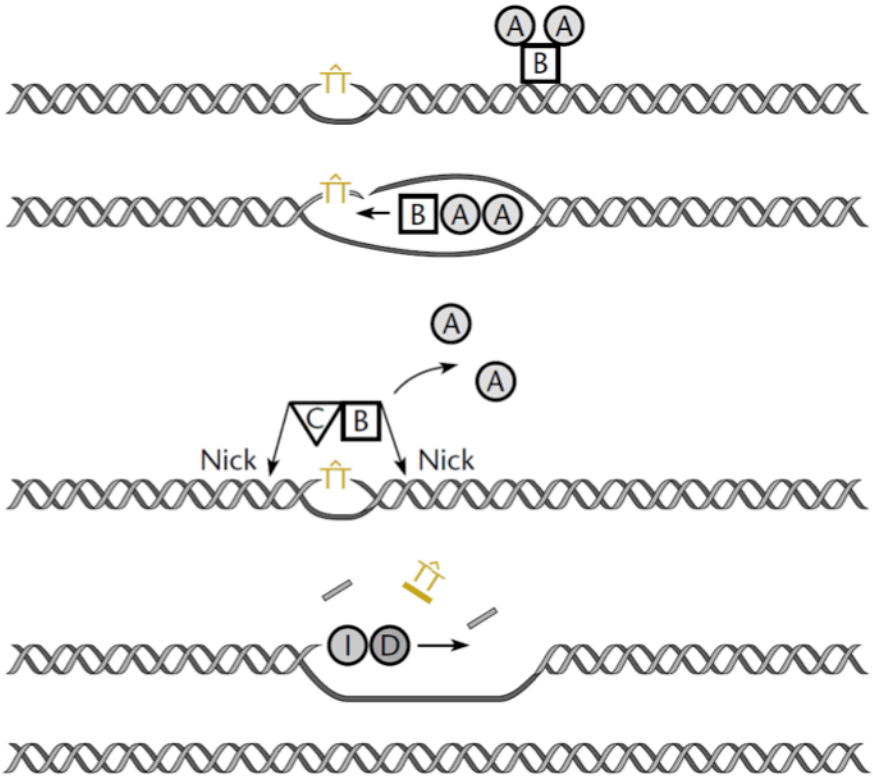

A
A
B
B
A
A
A
A
C
B
Nick
Nick
I
D

Figure 2: Oxidative Damage Repair.

A. 8-oxoG is formed by the oxidation of guanine at C8. **B.** This base analog can form proper base pairing with cytosine in anti formation. Dotted lines indicate hydrogen bonding. **C**. Oxidation of C8 allows for the stable rotation of the N-glycosidic bond to form a base analog that can incorrectly base pair with adenine. Adopted from (*96*). **D.** The "GO system" or Oxidative Damage response pathway. 8-oxoG is represented as a blue 'GO' 1) MutY removes adenines incorrectly incorporated across from 8-oxoG. 2) MutM is a glycosylase that cleaves 8-oxoG from the backbone. The resulting AP site is repaired via BER. 3) MutT cleanses the nucleotide pool of 8-oxodGTP by converting it to 8-oxodGMP. Figure was adopted from (*95*).

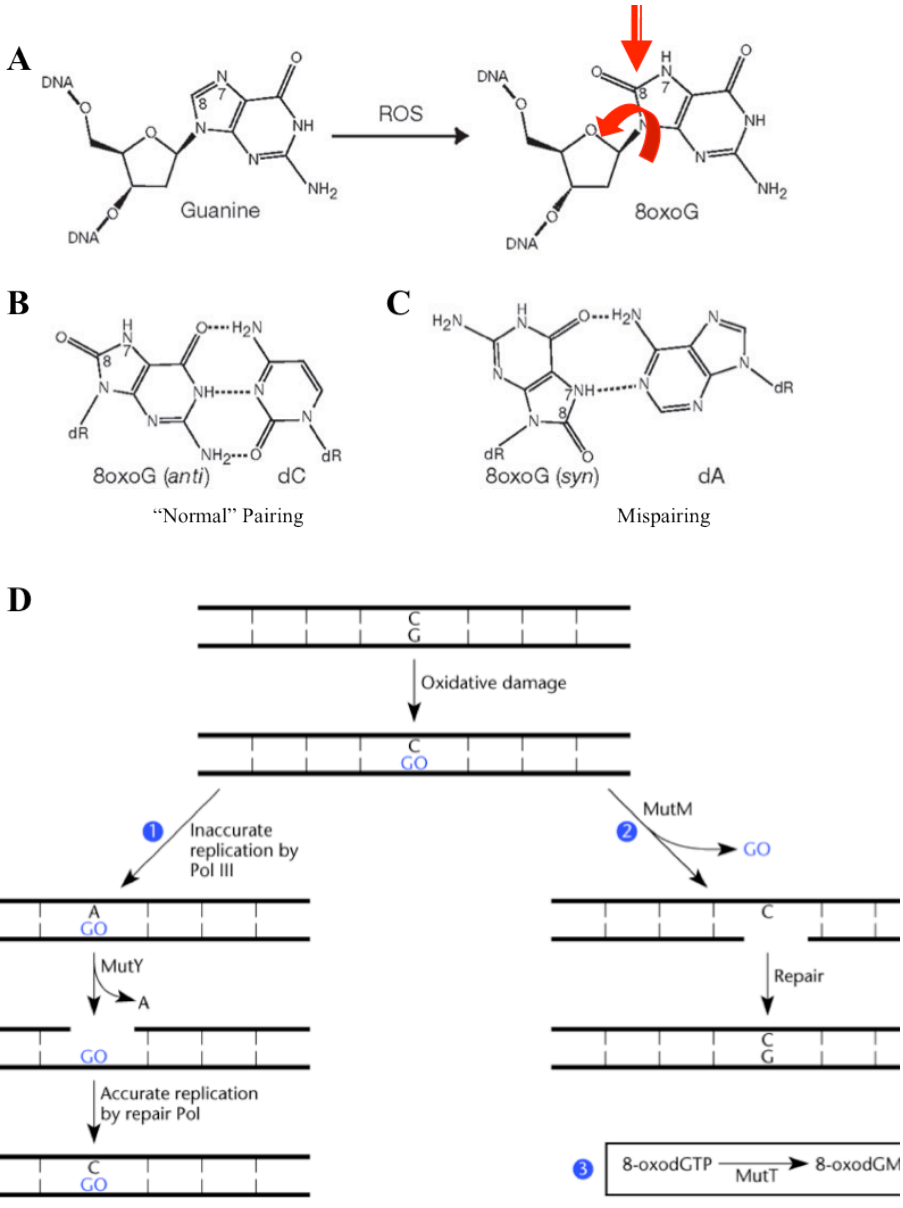

A
DNA
ROS
Guanine
8oxoG
B
C
8oxoG (anti)
dC
8oxoG (syn)
dA
"Normal" Pairing
Mispairing
D
Oxidative damage
Inaccurate replication by Pol III
MutM
GO
MutY
A
Repair
Accurate replication by repair Pol
8-oxodGTP
MutT
8-oxodGM

Figure 3: Double Strand Break Repair mediated by RecBCD

A.The RecBCD complex recognizes and binds the dsDNA end. **B.** The dsDNA is unwound with preferential degradation of the 3′ terminal strand. **C.** Upon recognition of the χ site the 3′ ➔5′ nuclease activity is attenuated and the 5′ ➔3′ nuclease activity is activated forming a 3′ overhang. **D.** RecA (omitted from this image) is loaded onto the 3′ end and catalyzes homologous pairing and DNA-strand exchange. Figure adopted from (*95*).

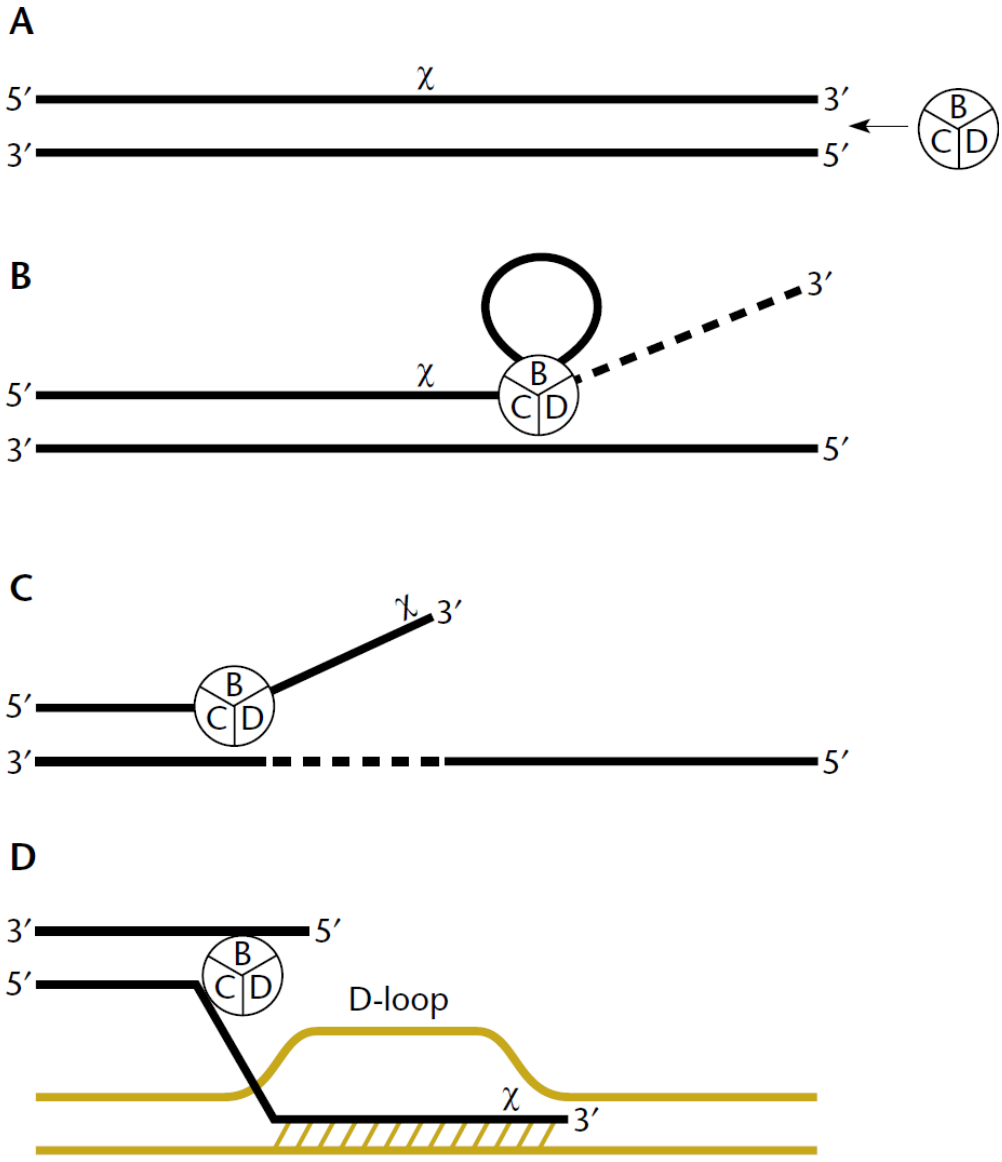

A
χ
5′
3′
3′
5′
B
C
D
B
3′
χ
5′
3′
5′
C
χ
3′
B
C
D
5′
3′
5′
D
3′
5′
B
C
D
5′
D-loop
χ
3′

Figure 4: DNA repair pathways in *D. radiodurans*.

There are multiple documented pathways that *D. radiodurans* can utilize to repair itself following DNA damage. HR, ESDSA, SSA, and NHEJ are discussed below.

A. Homologous recombination is the main pathway involved in DNA double stranded repair in bacteria and yeast. Single stranded DNA overhangs are produced and bound by SSB. RecA is loaded onto this single stranded DNA by mediator proteins like the RecBCD complex or RecFOR, which invades a homologous molecule to form a D loop and primes DNA synthesis.

B. ESDSA occurs early and uses the single stranded tail of a recessed fragment to invade a partially overlapping homologous fragment, thus priming DNA synthesis. Dissociation of this newly synthesized DNA produces long stretches of single stranded DNA that can anneal to overlapping regions elsewhere and thereby facilitate the reconstruction of long double stranded DNA intermediates. These intermediates subsequently recombine to reform the chromosome in a RecA-dependent fashion.

C. The SSA pathway also occurs early in recovery post irradiation and is RecA-independent. In this pathway, the double strand DNA breaks are recessed revealing single stranded DNA tails that can anneal to overlapping regions and prime DNA synthesis.

D. NHEJ is a common repair pathway in eukaryotes and only has recently been described in bacteria. In contrast to the other repair pathways, NHEJ does not require a homologous counterpart, nor single stranded invading tails. In eukaryotes, the double stranded DNA ends are bound by the Ku proteins and a number of end processing proteins are recruited to the site. DNA ligase subsequently joins the two DNA ends. It has been proposed that this pathway exists in *D. radiodurans* and the novel PprA protein acts to bind the double stranded DNA ends and stimulate DNA ligase. NHEJ is inherently error prone thus providing doubt that it is involved in the highly error-free process of DNA repair in *D. radiodurans*.

Figure adopted from (*97*).

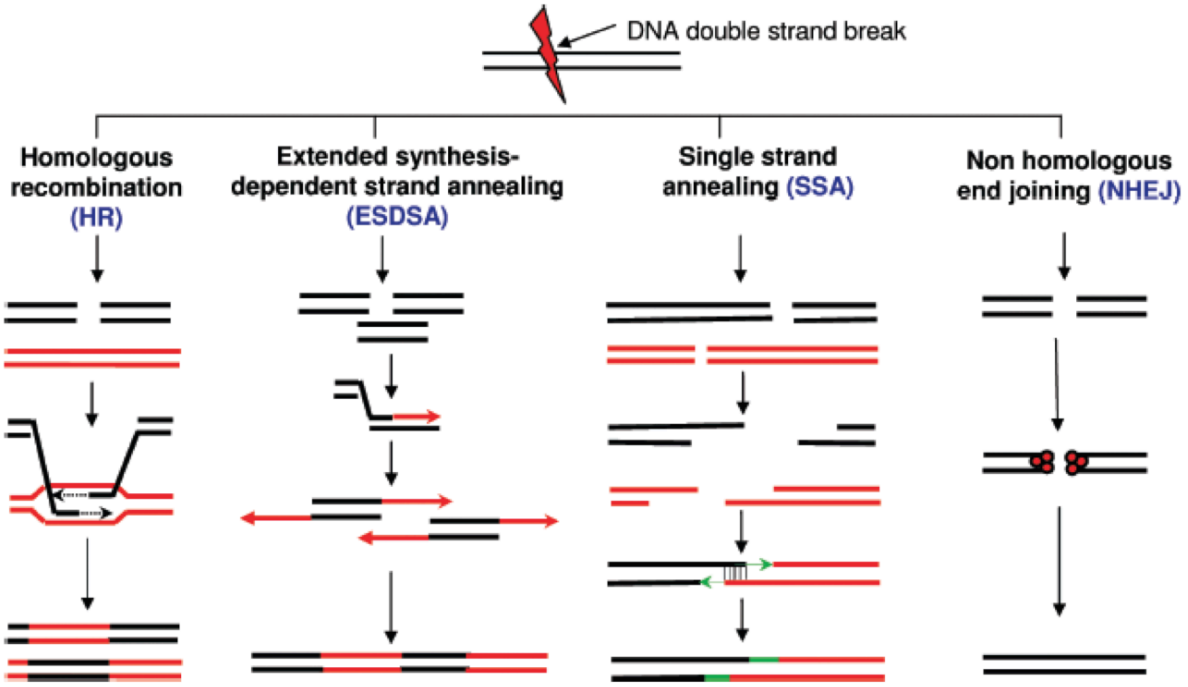

DNA double strand break
Homologous recombination (HR)
Extended synthesis-dependent strand annealing (ESDSA)
Single strand annealing (SSA)
Non homologous end joining (NHEJ)

Figure 5: Experimental design of directed evolution experiment.

A colony of lab strain *E. coli* MG1655, also referred to as Founder in our studies, was grown up and used to inoculate a culture that was irradiated with a dose to kill 99% of the cells. The 1% that survived was grown up and used as the inoculum for the next cycle of irradiation and outgrowth. Twenty cycles of irradiation and outgrowth were performed, four separate times, and isolates termed, CB1000-CB4000 were isolated from each trial and studied further.

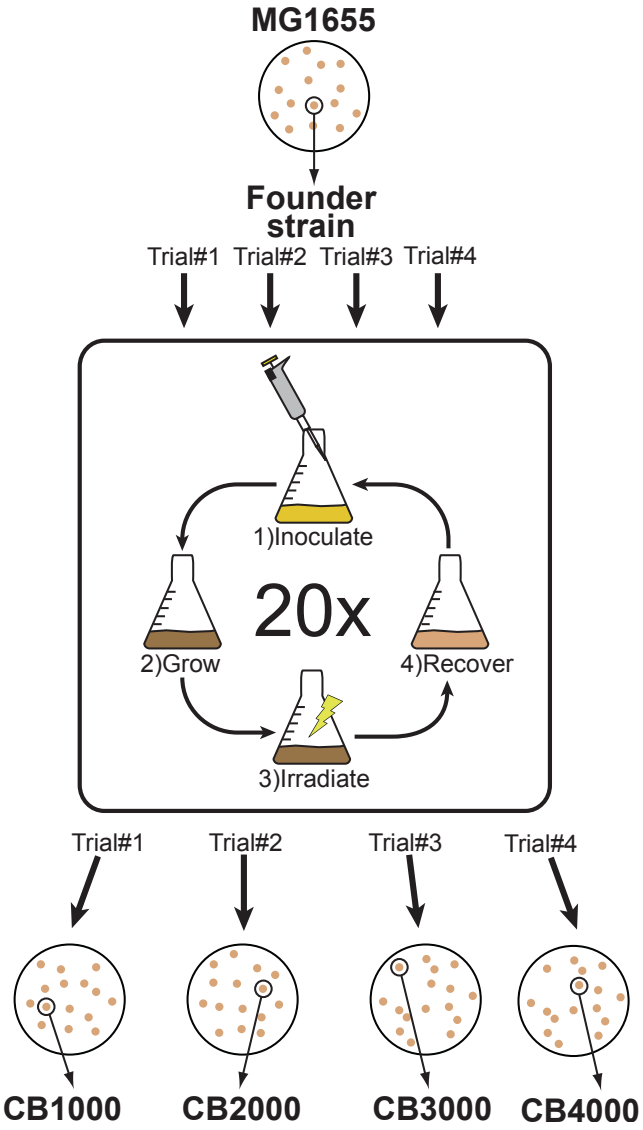

MG1655
Founder strain
Trial#1
Trial#2
Trial#3
Trial#4
1)Inoculate
20x
2)Grow
3)Irradiate
4)Recover
Trial#1
Trial#2
Trial#3
Trial#4
CB1000
CB2000
CB3000
CB4000

Figure 6: Evolved isolates are radiation resistant compared to the Founder strain.

Cultures were grown to mid-logarithmic phase and then subjected to increasing doses of irradiation before being plated for viability. The evolved isolates are 3-4 orders of magnitude more resistant to a dose of 3kGy than the radiosensitive Founder strain (•). Figure from (*87*).

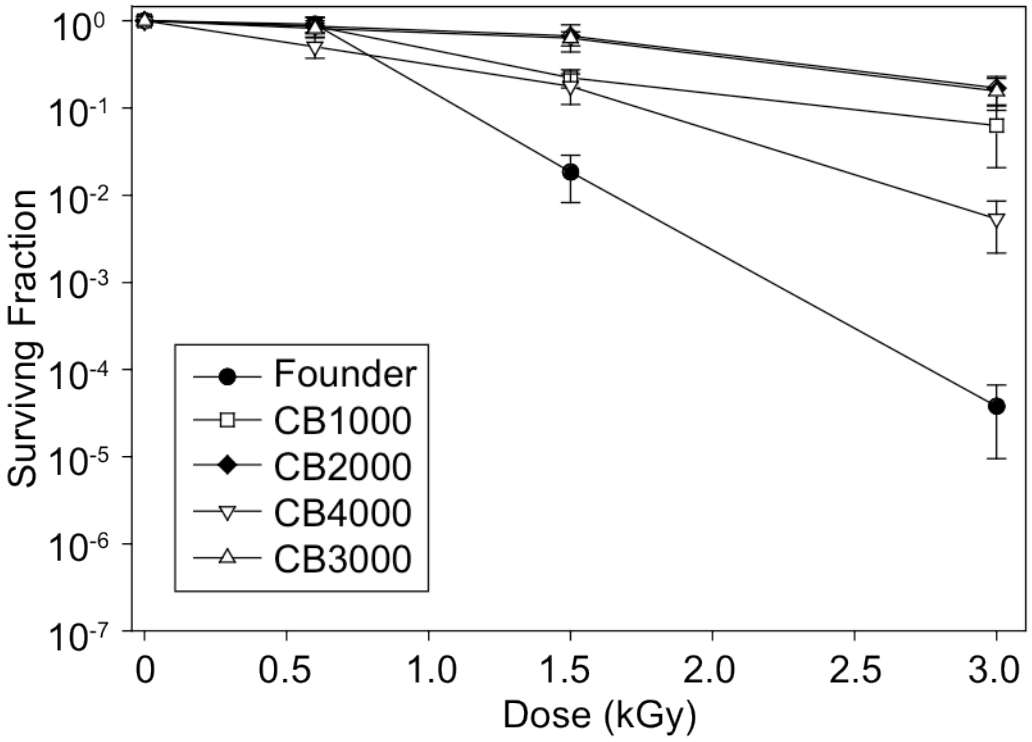

Survivng Fraction
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
0
0.5
1.0
1.5
2.0
2.5
3.0
Dose (kGy)
Founder
CB1000
CB2000
CB4000
CB3000

Figure 7: Radiation resistance was acquired in a step-wise manner.

Increase in radiation resistance during selection. Bars represent relative increase in survival at 3,000 Gy as a function of cycles that lead to isolation of top) CB1000 and bottom) CB2000. Figure from (*87*).

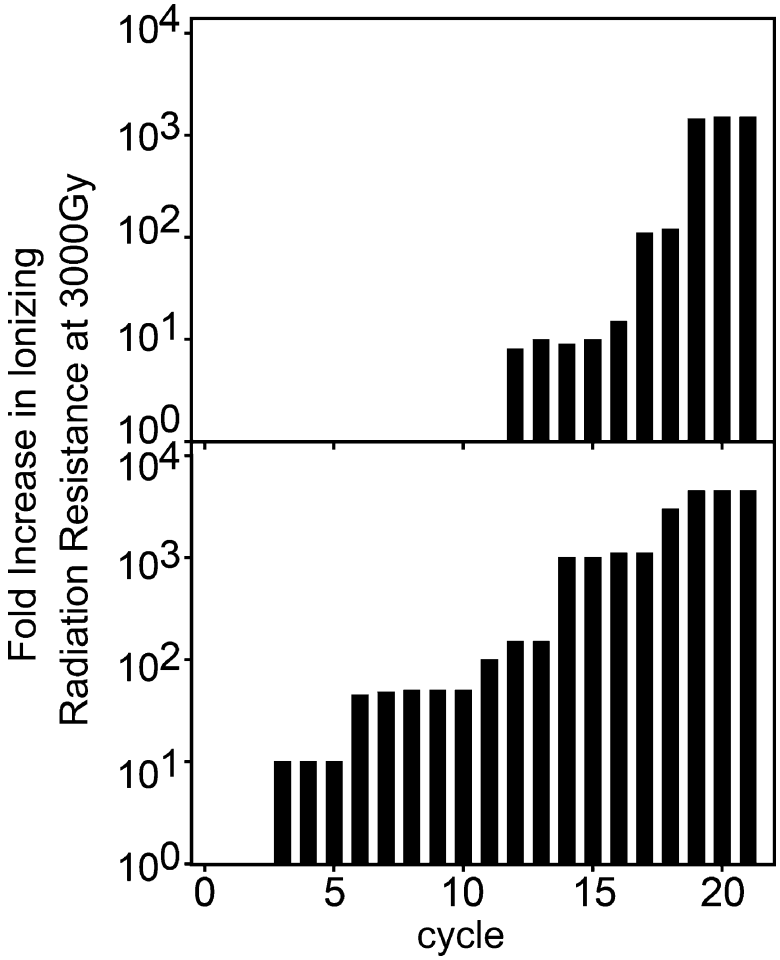
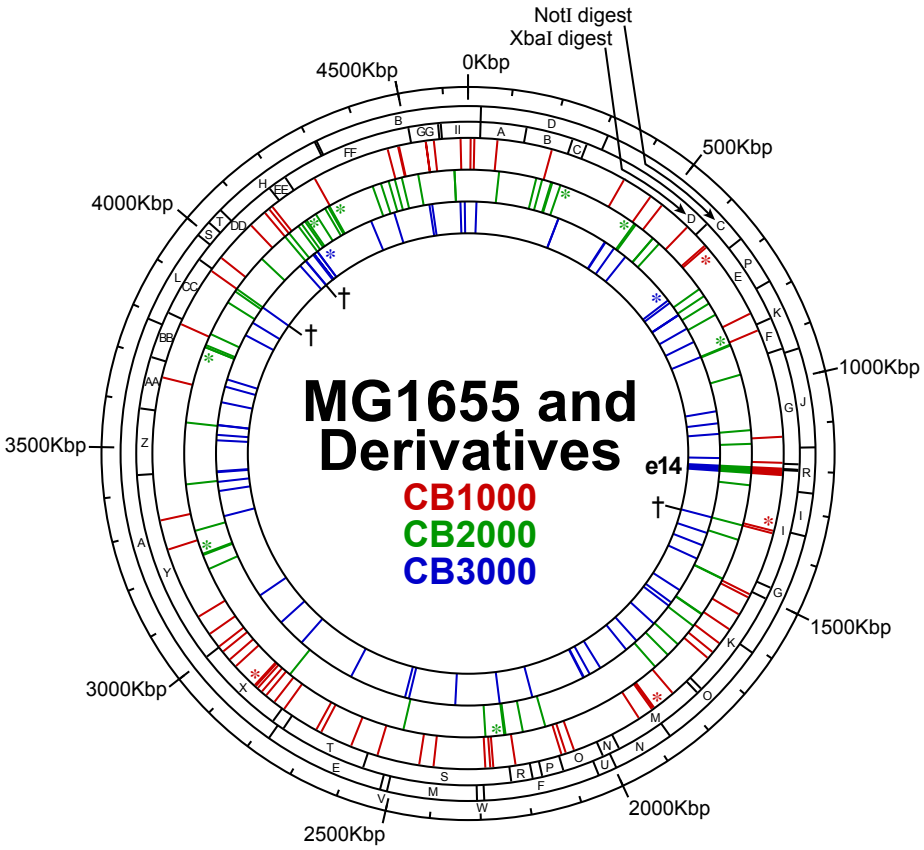

Fold Increase in Ionizing
Radiation Resistance at 3000Gy
10^4
10^3
10^2
10^1
10^0
10^4
10^3
10^2
10^1
10^0
0
5
10
15
20
cycle

Figure 8: Mutational analysis of three radiation resistant isolates.

Mutational pattern for 3 radiation resistant isolates illustrates that mutations occurred randomly across the genome and that there was no obvious mutational pattern shared among these three resistant isolates.



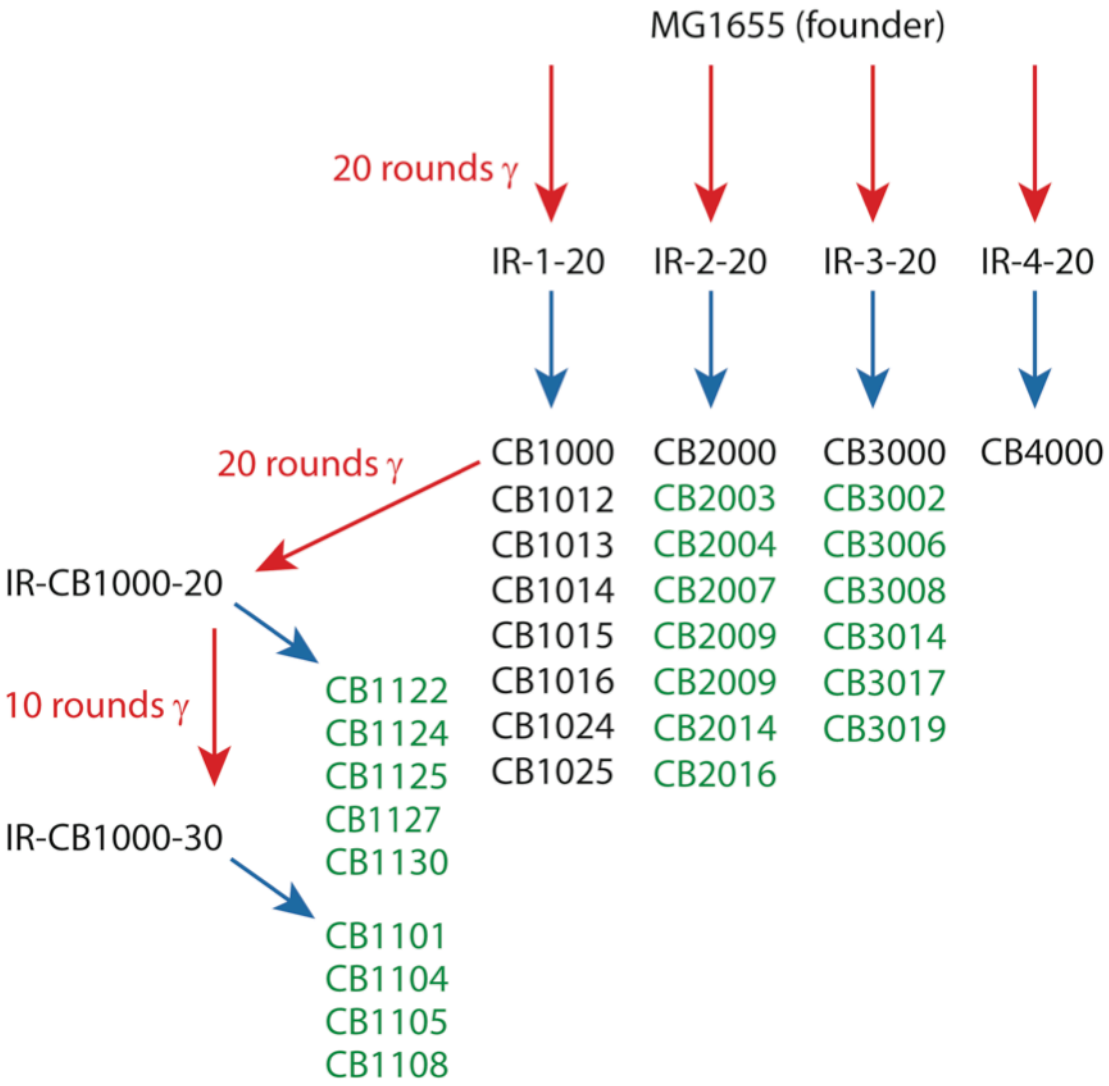
Red represents mutations found in CB1000
Green represents mutations found in CB2000
Blue represents mutations found in CB3000
The e14 prophage is marked as “e14”. The prophage is not present in any of the radiation resistant strains of *E.coli.* 1181794→1207826 (~26Kbp).

* Represents an area where proximity of mutations may make it difficult to visually resolve in this illustration

† The location of the three mutations pre-existing in our Founder strain.
–1349198 (G→A, intergenic)
–3936524 (T→G, RRbsR L91R)
–4122152 (G→A, CytR Q109*)

Figure 9: Radiation resistant isolates sequenced to date.

Schematic representation of evolved project highlighting all of the sequenced isolates to date. From the 20 rounds of irradiation and outgrowth illustrated in Figure 5, the following 4 radiation resistant populations were generated: IR-1-20, IR-2-20, IR-3-20, and IR-4-20. In the Harris 2009 study, the isolates listed in black were sequenced thus providing us a detailed account of mutational innovations that occurred in the IR-1-20 population, while offering a glimpse into how radiation resistance had evolved in the other populations. The work described in Chapter 2, discusses the additional sequences listed in green. This work gives us a more detailed account of how radiation resistance evolved in the IR-2-20 population which for reasons discussed in Chapter 3, is a great population for determining which mutations are responsible for the evolution of radiation resistance.


MG1655 (founder)
20 rounds γ
IR-1-20
IR-2-20
IR-3-20
IR-4-20
20 rounds γ
CB1000
CB1012
CB1013
CB1014
CB1015
CB1016
CB1024
CB1025
CB2000
CB2003
CB2004
CB2007
CB2009
CB2009
CB2014
CB2016
CB3000
CB3002
CB3006
CB3008
CB3014
CB3017
CB3019
CB4000
IR-CB1000-20
10 rounds γ
IR-CB1000-30
CB1122
CB1124
CB1125
CB1127
CB1130
CB1101
CB1104
CB1105
CB1108

References

1. Cox, M. M., and Battista, J. R. (2005) *Deinococcus radiodurans* - The consummate survivor, *Nature Reviews Microbiology 3*, 882-892.
2. von Sonntag, C. (1987) New aspects in the free-radical chemistry of pyrimidine nucleobases, *Free Radical Research Communications 2*, 217-224.
3. Sonntag, C. v. (2005) Free-Radical-Induced DNA Damage and Its Repair: A chemical perspective, *Springer Publishing*.
4. Blattner, F. R., Plunkett, G. r., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado, V. J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome sequence of Escherichia coli K-12, *Science 277*, 1453-1474.
5. Imlay, J. A. (2008) Cellular defenses against superoxide and hydrogen peroxide, *Annu Rev Biochem 77*, 755-776.
6. Greenberg, J. T., Monach, P., Chou, J. H., Josephy, P. D., and Demple, B. (1990) Positive control of a global antioxidant defense regulon activated by superoxide-generating agents in *Escherichia coli*, *Proc Natl Acad Sci U S A 87*, 6181-6185.
7. Tsaneva, I. R., and Weiss, B. (1990) soxR, a locus governing a superoxide response regulon in Escherichia coli K-12, *Journal of Bacteriology 172*, 4197-4205.
8. Hidalgo, E., and Demple, B. (1994) An iron-sulfur center essential for transcriptional activation by the redox-sensing SoxR protein, *EMBO J 13*, 138-146.
9. Gu, M., and Imlay, J. A. (2011) The SoxRS response of *Escherichia coli* is directly activated by redox-cycling drugs rather than by superoxide, *Mol Microbiol 79*, 1136-1150.
10. Koo, M. S., Lee, J. H., Rah, S. Y., Yeo, W. S., Lee, J. W., Lee, K. L., Koh, Y. S., Kang, S. O., and Roe, J. H. (2003) A reducing system of the superoxide sensor SoxR in Escherichia coli, *EMBO J 22*, 2614-2622.
11. Christman, M. F., Morgan, R. W., Jacobson, F. S., and Ames, B. N. (1985) Positive control of a regulon for defenses against oxidative stress and some heat-shock proteins in Salmonella typhimurium, *Cell 41*, 753-762.
12. Tao, K., Fujita, N., and Ishihama, A. (1993) Involvement of the RNA polymerase alpha subunit C-terminal region in co-operative interaction and transcriptional activation with OxyR protein, *Mol Microbiol 7*, 859-864.

13. Slauch, J. M., Russo, F. D., and Silhavy, T. J. (1991) Suppressor mutations in rpoA suggest that OmpR controls transcription by direct interaction with the alpha subunit of RNA polymerase, *Journal of bacteriology 173*, 7501-7510.

14. Zheng, M., Aslund, F., and Storz, G. (1998) Activation of the OxyR transcription factor by reversible disulfide bond formation, *Science 279*, 1718-1721.

15. Friedberg, E. C., Walker, G. C., and Siede, W. (1995) *DNA repair and mutagenesis*, ASM Press, Washington, D.C.

16. Cadet, J., Delatour, T., Douki, T., Gasparutto, D., Pouget, J. P., Ravanat, J. L., and Sauvaigo, S. (1999) Hydroxyl radicals and DNA base damage, *Mutation research 424*, 9-21.

17. Teoule, R. (1987) Radiation-induced DNA damage and its repair, *International journal of radiation biology and related studies in physics, chemistry, and medicine 51*, 573-589.

18. Lindahl, T. (1976) New class of enzymes acting on damaged DNA, *Nature 259*, 64-66.

19. Duncan, J., Hamilton, L., and Friedberg, E. C. (1976) Enzymatic degradation of uracil-containing DNA. II. Evidence for *N*-glycosidase and nuclease activities in unfractionated extracts of Bacillus subtilis, *Journal of virology 19*, 338-345.

20. Wyatt, M. D., Allan, J. M., Lau, A. Y., Ellenberger, T. E., and Samson, L. D. (1999) 3-methyladenine DNA glycosylases: structure, function, and biological importance, *BioEssays : news and reviews in molecular, cellular and developmental biology 21*, 668-676.

21. Howard-Flanders, P., Boyce, R. P., and Theriot, L. (1966) Three loci in Escherichia coli K-12 that control the excision of pyrimidine dimers and certain other mutagen products from DNA, *Genetics 53*, 1119-1136.

22. Howard-Flanders, P., and Boyce, R. P. (1966) DNA repair and genetic recombination: studies on mutants of Escherichia coli defective in these processes, *Radiation research*, Suppl 6:156+.

23. Mattern, I. E., van Winden, M. P., and Rorsch, A. (1965) The range of action of genes controlling radiation sensitivity in Escherichia coli, *Mutation research 2*, 111-131.

24. van de Putte, P., van Sluis, C. A., van Dillewijn, J., and Rorsch, A. (1965) The location of genes controlling radiation sensitivity in Escherichia coli, *Mutation research 2*, 97-110.

25. Van Houten, B., and McCullough, A. (1994) Nucleotide excision repair in E. coli, *Annals of the New York Academy of Sciences 726*, 236-251.

26. Selby, C. P., and Sancar, A. (1995) Structure and function of transcription-repair coupling factor. II. Catalytic properties, *J Biol Chem 270*, 4890-4895.

27. Selby, C. P., and Sancar, A. (1995) Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties, *J Biol Chem 270*, 4882-4889.

28. Selby, C. P., and Sancar, A. (1993) Molecular mechanism of transcription-repair coupling, *Science 260*, 53-58.

29. Savery, N. J. (2007) The molecular mechanism of transcription-coupled DNA repair, *Trends in Microbiology 15*, 326-333.

30. Hogg, M., Wallace, S. S., and Doublie, S. (2005) Bumps in the road: how replicative DNA polymerases see DNA damage, *Curr Opin Struct Biol 15*, 86-93.

31. Maki, H., and Sekiguchi, M. (1992) MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis, *Nature 355*, 273-275.

32. Michaels, M. L., and Miller, J. H. (1992) The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine), *Journal of Bacteriology 174*, 6321-6325.

33. Breen, A. P., and Murphy, J. A. (1995) Reactions of oxyl radicals with DNA, *Free Radical Biology & Medicine 18*, 1033-1077.

34. Sugawara, N., and Haber, J. E. (2006) Repair of DNA double strand breaks: In vivo biochemistry, In *DNA Repair, Pt A*, pp 416-429.

35. Cahill, D., Connor, B., and Carney, J. P. (2006) Mechanisms of eukaryotic DNA double strand break repair, *Frontiers in Bioscience 11*, 1958-1976.

36. Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., and Rehrauer, W. M. (1994) Biochemistry of homologous recombination in *Escherichia coli.* , *Microbiological Reviews 58*, 401-465.

37. Kuzminov, A. (1999) Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda, *Microbiol Mol Biol Rev 63*, 751-813.

38. West, S. C. (1992) Enzymes and molecular mechanisms of genetic recombination, *Annu Rev Biochem 61*, 603-640.

39. Dillingham, M. S., Spies, M., and Kowalczykowski, S. C. (2003) RecBCD enzyme is a bipolar DNA helicase, *Nature 423*, 893-897.

40. Smith, G. R., Kunes, S. M., Schultz, D. W., Taylor, A., and Triman, K. L. (1981) Structure of chi hotspots of generalized recombination, *Cell 24*, 429-436.

41. Sargentini, N. J., and Smith, K. C. (1988) Genetic and phenotypic analyses indicating occurrence of the *recN262* and *radB101* mutations at the same locus in *Escherichia coli*, *Journal of bacteriology 170*, 2392-2394.

42. Finch, P. W., Chambers, P., and Emmerson, P. T. (1985) Identification of the *Escherichia coli recN* gene product as a major SOS protein, *Journal of Bacteriology 164*, 653-658.

43. Jessberger, R. (2002) The many functions of SMC proteins in chromosome dynamics, *Nature reviews. Molecular cell biology 3*, 767-778.

44. Yokomori, K. (2003) SMC protein complexes and the maintenance of chromosome integrity, *Current topics in microbiology and immunology 274*, 79-112.

45. Morimatsu, K., and Kowalczykowski, S. C. (2003) RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair, *Mol Cell 11*, 1337-1347.

46. Moller, I. M., Rogowska-Wrzesinska, A., and Rao, R. S. (2011) Protein carbonylation and metal-catalyzed protein oxidation in a cellular perspective, *Journal of proteomics 74*, 2228-2242.

47. Shacter, E. (2000) Quantification and significance of protein oxidation in biological samples, *Drug metabolism reviews 32*, 307-326.

48. Daly, M. J., Gaidamakova, E. K., Matrosova, V. Y., Vasilenko, A., Zhai, M., Leapman, R. D., Lai, B., Ravel, B., Li, S. M., Kemner, K. M., and Fredrickson, J. K. (2007) Protein Oxidation Implicated as the Primary Determinant of Bacterial Radioresistance, *PLoS biology 5*, e92.

49. Daly, M. J., Gaidamakova, E. K., Matrosova, V. Y., Vasilenko, A., Zhai, M., Venkateswaran, A., Hess, M., Omelchenko, M. V., Kostandarithes, H. M., Makarova, K. S., Wackett, L. P., Fredrickson, J. K., and Ghosal, D. (2004) Accumulation of Mn(II) in, Deinococcus radiodurans facilitates gamma-radiation resistance, *Science 306*, 1025-1028.

50. Dalle-Donne, I., Giustarini, D., Colombo, R., Rossi, R., and Milzani, A. (2003) Protein carbonylation in human diseases, *Trends in molecular medicine 9*, 169-176.

51. Richmond, R. C., Sridhar, R., and Daly, M. J. (1999) Physicochemical survival pattern for the radiophile *Deinococcus radiodurans*: A polyextremophile model for life on Mars, *SPIE 3755*, 210-222.

52. Mattimore, V., and Battista, J. R. (1996) Radioresistance of *Deinococcus radiodurans*: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation, *Journal of Bacteriology 178*, 633-637.

53. Anderson, A. W., Nordon, H. C., Cain, R. F., Parrish, G., and Duggan, D. (1956) Studies on a radioresistant micrococcus. 1. isolation, morphology, cultural characteristics, and resistance to gamma radiation, *Food Technol. 10*, 575-578.

54. Levin, S. G., Young, R. W., and Stohler, R. L. (1992) Estimation of median human lethal radiation dose computed from data on occupants of reinforced concrete structures in Nagasaki, Japan, *Health physics 63*, 522-531.

55. Moseley, B. E., and Mattingly, A. (1971) Repair of irradiation transforming deoxyribonucleic acid in wild type and a radiation-sensitive mutant of *Micrococcus radiodurans*, *Journal of bacteriology 105*, 976-983.

56. Bergey, D. H., Holt, J. G., and Krieg, N. R. (1984) *Bergey's manual of systematic bacteriology*, Williams & Wilkins, Baltimore.

57. Brooks, B. W., and Murray, R. G. E. (1981) Nomenclature for "*Micrococcus radiodurans*" and other radiation-resistant cocci: *Deinococcaceae* fam. nov. and *Deinococcus* gen. nov., including five species, *International Journal of Systematic Bacteriology 31*, 353-360.

58. Archibald, F. S., and Fridovich, I. (1981) Manganese and defenses against oxygen toxicity in *Lactobacillus plantarum*, *Journal of Bacteriology 145*, 442-451.

59. Levin-Zaidman, S., Englander, J., Shimoni, E., Sharma, A. K., Minton, K. W., and Minsky, A. (2003) Ringlike structure of the *Deinococcus radiodurans* genome: A key to radioresistance?, *Science 299*, 254-256.

60. Krasin, F., and Hutchinson, F. (1977) Repair of DNA double-strand breaks in *Escherichia coli,* which requires recA function and the presence of a duplicate genome, *J Mol Biol 116*, 81-98.

61. Makarova, K. S., Aravind, L., Wolf, Y. I., Tatusov, R. L., Minton, K. W., Koonin, E. V., and Daly, M. J. (2001) Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics, *Microbiology & Molecular Biology Reviews 65*, 44-79.

62. Cox, M. M., and Battista, J. R. (2005*) Deinococcus radiodurans* - the consummate survivor, *Nat Rev Microbiol 3*, 882-892.

63. Burrell, A. D., Feldschreiber, P., and Dean, C. J. (1971) DNA-membrane association and the repair of double breaks in x-irradiated *Micrococcus radiodurans*, *Biochimica et biophysica acta 247*, 38-53.

64. Zimmerman, J. M., and Battista, J. R. (2005) A ring-like nucleoid is not necessary for radioresistance in the *Deinococcaceae*, *BMC Microbiology 5*, 17.

65. Daly, M. J., and Minton, K. W. (1996) An alternative pathway of recombination of chromosomal fragments precedes *recA*-dependent recombination in the radioresistant bacterium *Deinococcus radiodurans*, *Journal of bacteriology 178*, 4461-4471.

66. Zahradka, K., Slade, D., Bailone, A., Sommer, S., Averbeck, D., Petranovic, M., Lindner, A. B., and Radman, M. (2006) Reassembly of shattered chromosomes in *Deinococcus radiodurans*, *Nature 443*, 569-573.

67. Wilson, T. E., Topper, L. M., and Palmbos, P. L. (2003) Non-homologous end-joining: bacteria join the chromosome breakdance, *Trends in Biochemical Sciences 28*, 62-66.

68. Lees-Miller, S. P., and Meek, K. (2003) Repair of DNA double strand breaks by non-homologous end joining, *Biochimie 85*, 1161-1173.

69. Paques, F., and Haber, J. E. (1999) Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae, *Microbiology & Molecular Biology Review (Washington, DC) 63*, 349-404.

70. Gloor, G. B., Nassif, N. A., Johnson-Schlitz, D. M., Preston, C. R., and Engels, W. R. (1991) Targeted gene replacement in *Drosophila* via P element-induced gap repair, *Science 253*, 1110-1117.

71. Kraus, E., Leung, W. Y., and Haber, J. E. (2001) Break-induced replication: a review and an example in budding yeast, *Proc Natl Acad Sci U S A 98*, 8255-8262.

72. Lederberg, J. (1955) Recombination mechanisms in bacteria, *Journal of Cellular Physiology. Supplement 45*, 75-107.

73. Daly, M. J., Ouyang, L., Fuchs, P., and Minton, K. W. (1994) In vivo damage and recA-dependent repair of plasmid and chromosomal DNA in the radiation-resistant bacterium *Deinococcus radiodurans*, *Journal of Bacteriology 176*, 3508-3517.

74. Carroll, J. D., Daly, M. J., and Minton, K. W. (1996) Expression of recA in *Deinococcus radiodurans*, *Journal of Bacteriology 178*, 130-135.

75. Kim, J.-I., and Cox, M. M. (2002) The RecA proteins of *Deinococcus radiodurans* and *Escherichia coli* promote DNA strand exchange via inverse pathways, *Proc Natl Acad Sci U S A 99*, 7917-7921.

76. Hsu, H.-F., Ngo, K.V., Chitteni-Pattu, S., Cox, M.M., and Li, H.-W. (2011) Investigating *Deinococcus radiodurans* RecA protein filament formation on double-stranded DNA by a real-time single-molecule approach, *Biochemistry 50*, 8270-8280.

77. Tanaka, M., Earl, A. M., Howell, H. A., Park, M. J., Eisen, J. A., Peterson, S. N., and Battista, J. R. (2004) Analysis of *Deinococcus radiodurans*'s transcriptional response to ionizing radiation and desiccation reveals novel proteins that contribute to extreme radioresistance, *Genetics 168*, 21-33.

78. Harris, D. R., Tanaka, M., Saveliev, S. V., Jolivet, E., Earl, A. M., Cox, M. M., and Battista, J. R. (2004) Preserving Genome Integrity: The DdrA Protein of *Deinococcus radiodurans R1*, *PLoS biology 2*, e304.

79. Daly, M. J. (2012) Death by protein damage in irradiated cells, *DNA repair 11*, 12-21.

80. Daly, M. J. (2009) A new perspective on radiation resistance based on *Deinococcus radiodurans*, *Nat Rev Microbiol 7*, 237-245.

81. Slade, D., and Radman, M. (2011) Oxidative Stress Resistance in *Deinococcus radiodurans*, *Microbiology and Molecular Biology Reviews 75*, 133-+.

82. Krisko, A., and Radman, M. (2010) Protein damage and death by radiation in *Escherichia coli* and *Deinococcus radiodurans*, *Proc Natl Acad Sci U S A 107*, 14373-14377.

83. Witkin, E. M. (1946) Inherited Differences in Sensitivity to Radiation in Escherichia Coli, *Proc Natl Acad Sci U S A 32*, 59-68.

84. Parisi, A., and Antoine, A. D. (1974) Increased radiation resistance of vegetative *Bacillus pumilus*, *Appl Microbiol 28*, 41-46.

85. Davies, R., and Sinskey, A. J. (1973) Radiation-resistant mutants of *Salmonella typhimurium LT2*: development and characterization, *Journal of bacteriology 113*, 133-144.

86. Erdman, I. E., Thatcher, F. S., and Macqueen, K. F. (1961) Studies on the irradiation of microorganisms in relation to food preservation. II. Irradiation resistant mutants, *Can J Microbiol 7*, 207-215.

87. Harris, D. R., Pollock, S. V., Wood, E. A., Goiffon, R. J., Klingele, A. J., Cabot, E. L., Schackwitz, W., Martin, J., Eggington, J., Durfee, T. J., Middle, C. M., Norton, J. E., Popelars, M. C., Li, H., Klugman, S. A., Hamilton, L. L., Bane, L. B., Pennacchio, L. A., Albert, T. J., Perna, N. T., Cox, M. M., and Battista, J. R. (2009) Directed evolution of ionizing radiation resistance in *Escherichia coli*, *Journal of bacteriology 191*, 5240-5252.

88. Elena, S. F., and Lenski, R. E. (2003) Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation, *Nature Reviews Genetics 4*, 457-469.

89. Gerrish, P. J., and Lenski, R. E. (1998) The fate of competing beneficial mutations in an asexual population, *Genetica 103*, 127-144.

90. Lenski, R. E., Rose, M. R., Simpson, S. C., and Tadler, S. C. (1991) Long-Term Experimental Evolution in Escherichia-Coli .1. Adaptation and Divergence During 2,000 Generations, *American Naturalist 138*, 1315-1341.

91. Shuman, H. A., and Silhavy, T. J. (2003) The art and design of genetic screens: *Escherichia coli, Nat Rev Genet 4*, 419-431.

92. Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balazsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I., Gelfand, M. S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M. V., Grechkin, Y., Mseeh, F., Fonstein, M. Y., Overbeek, R., Barabasi, A. L., Oltvai, Z. N., and Osterman, A. L. (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655, *Journal of bacteriology 185*, 5673-5684.

93. Rudd, K. E. (1998) Linkage map of *Escherichia coli* K-12, edition 10: the physical map, *Microbiol Mol Biol Rev 62*, 985-1019.

94. Langridge, G. C., Phan, M. D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E., Dougan, G., Wain, J., Parkhill, J., and Turner, A. K. (2009) Simultaneous assay of every *Salmonella typhi* gene using one million transposon mutants, *Genome Res 19*, 2308-2316.

95. Snyder, L., and Champness, W. (2007) *Molecular genetics of bacteria*, 3rd ed., ASM Press, Washington, D.C.

96. Hsu, G. W., Ober, M., Carell, T., and Beese, L. S. (2004) Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase, *Nature 431*, 217-221.

97. Blasius, M., Hubscher, U., and Sommer, S. (2008) *Deinococcus radiodurans*: What belongs to the survival kit?, *Critical Reviews in Biochemistry and Molecular Biology 43*, 221-238.

Chapter 2: Evolution of Extreme Resistance to Ionizing Radiation via Genetic Adaptation of DNA Repair

Rose T. Byrne[1#], Audrey J. Klingele[1#], Elizabeth A. Wood[1], Eric L. Cabot[2], Wendy Schackwitz[3], Jeff Martin[3], Joel Martin[3], Zhong Wang[3], Christa Pennacchio[3], Len A. Pennacchio[3], Nicole T. Perna[2,4], John R. Battista[5], and Michael M. Cox[1*]

[1]Department of Biochemistry, University of Wisconsin – Madison, Madison, Wisconsin 53706-1544, [2]Genome Center, University of Wisconsin, 425G Henry Mall, Madison, WI 53703, [3]DOE Joint Genome Institute, Walnut Creek, CA 94598, [4]Laboratory of Genetics, University of Wisconsin, 425G Henry Mall, Madison, WI 53706, [5]Department of Biological Sciences, Louisiana State University and A & M College, Baton Rouge, LA 70803

#These authors contributed equally to the work.

*To whom correspondence should be addressed.

Reference:

Byrne RT, Klingele AJ, Wood EA, Cabot EL, Schackwitz W, Martin J, et al. (2013) Evolution of extreme resistance to ionizing radiation via genetic adaptation of DNA repair. **eLife** *submitted*

2.1 Abstract

Life has successfully colonized extreme environments from hot springs in Yellowstone National Park to the ruins of the Chernobyl nuclear reactor. The evolution required for a species of microorganisms to thrive in a challenging environment can occur with surprising rapidity. However, the molecular basis of extremophile adaptation is rarely evident. Here, we show that a complex new phenotype acquired by directed evolution of the common laboratory bacterium *Escherichia coli* – extreme resistance to ionizing radiation (IR) – can be explained by evolutionary adaptation of existing DNA repair systems that are common to most bacterial species. By directed evolution in the laboratory, we have generated strains of *Escherichia coli* that exhibit a 3-4 orders of magnitude increase in IR survival after exposure to 3,000 Gy (*1*), a level of irradiation that is more than 1,000 times the lethal dose for a human. Of 69 mutations identified in CB2000 (one of our most highly adapted isolates), functional experiments now indicate that the IR resistance phenotype is almost entirely accounted for by only three of these nucleotide changes, in the DNA metabolism genes *recA*, *dnaB*, and *yfjK*. Four additional genetic changes make small but measurable contributions. Our results demonstrate that there are multiple molecular contributions to an extreme IR resistance phenotype, and at the same time highlight a particular adaptation mechanism not adequately considered in studies to date. Genetic innovations involving pre-existing but clearly malleable DNA

repair functions can play a predominant role in the acquisition of an extreme IR resistance phenotype.

2.2 Introduction

Ionizing radiation (IR) is encountered medically in the form of X-rays and tumor irradiation, and very rarely in the context of nuclear power plant malfunction. The study of organisms with extreme resistance to IR has received increasing attention as a potential source of mechanistic insights that might permit the modulation of IR resistance in cells. *Deinococcus radiodurans*, which can absorb IR doses in excess of 5 kGy without lethality (over 1,000 times the lethal dose for humans), has become a key model organism for understanding this phenotype (*2, 3*). Mainly through the formation of reactive oxygen species (ROS), IR can lead to the damage of protein, DNA, and all other cellular macromolecules (*4*). Ongoing research into the molecular basis of extreme IR resistance has suggested three potential classes of mechanisms. These are (A) a condensed nucleoid structure, which could potentially facilitate DNA repair processes (*5, 6*), (B) an enhanced capacity for amelioration of protein damage, which could protect DNA repair systems and make them more readily available following irradiation (*7, 8*), and (C) potential specialized pathways for DNA repair (*2, 9*).

The amelioration of protein oxidation has received extensive experimental support as a mechanism to account for much if not all of the observed IR resistance in *D. radiodurans* (*7, 8, 10-12*). Increases in cytosolic antioxidant capacity, particularly an

increase in the cellular Mn/Fe ratio (*10*), appear to make the major contributions. Increasing the Mn/Fe ratio limits the Fenton chemistry that produces ROS (*7, 12*). The hypothesis that the condensed nucleoid of *Deinococcus* facilitates efficient DNA repair has been questioned, since the presence of condensed nucleoids does not correlate reliably with radiation resistance in bacterial species (*5*).

Repair of IR-induced DNA damage is clearly important to survival. However, the constellation of DNA repair functions in *D. radiodurans* is unremarkable by bacterial standards. *D. radiodurans* and other IR resistant species appear to have approximately the same toolbox of DNA repair pathways as non-resistant species. Thus, the argument has been advanced that specialized DNA repair plays little or no role in IR resistance (*7, 8*). Instead, antioxidants prevent protein oxidation and render the classical DNA repair systems more readily available to correct the effects of IR.

Bacteria are being used in long-term laboratory experiments that have elucidated many aspects of evolutionary biology (*13-15*). Directed bacterial evolution can be used as a tool to determine the molecular underpinnings of adaptation to a stress or a new environment. Does IR resistance always arise via enhanced protection of otherwise commonplace DNA repair proteins from oxidative inactivation? Alternatively, can a special facility for DNA repair make a significant contribution to this extremophile phenotype?

In an earlier study, we demonstrated that the naturally IR sensitive bacterium, *E. coli,* could acquire an extreme IR resistance phenotype by directed evolution (*1*). In brief, we carried out 20 rounds of selection, with each round consisting of IR exposure sufficient to kill ~99% of the population, followed by outgrowth. The experiment was carried out four times, generating four separately evolved populations (IR-1-20, IR-2-20, IR-3-20, and IR-4-20). In the evolved populations, survival at 3,000 Gy was typically increased by 3-4 orders of magnitude (*1*). We reported genomic sequences from seven isolates of IR-1-20, one from IR-2-20, and one from IR-3-20. Genetic alterations in the evolved strains were quite numerous, suggesting a complex and somewhat variable pattern of genetic innovation. It remained for us to determine which of those mutations underlay the increase in radiation resistance, and what each of them might contribute to the IR resistance phenotype. This is accomplished in the present study.

2.3 Results

2.3.1 Sequencing of multiple isolates from evolved populations reveals different evolutionary paths to IR resistance.

In the present study, we augmented the earlier data with the complete genomic sequences of 13 new isolates from populations IR-2-20, IR-3-20, and IR-4-20. In addition, we examined the complete genomic sequences of nine isolates derived from either 20 or 30 rounds of further evolution of CB1000, an isolate derived from IR-1-20. We note that IR resistance in this more highly evolved population (IR-CB1000-30) was impossible to

measure accurately. The cells in this population grew more slowly than the Founder, but growth continued during irradiation at 6 Gy/min (John R. Battista, unpublished data). The results from the previous (*1*) and new sequencing efforts are combined in Tables S1 and S2. The results give us a much-enhanced view of mutational patterns that are likely to contribute to the acquired IR resistance.

A summary of mutation types detected in the various sequenced isolates is presented in Table 1. In each case, the genomic sequence was compared to the 4,639,675-bp reference genome. In the isolates derived from the original four evolved populations, there were between 44 and 77 genetic alterations. The numbers of mutations jumped appreciably in the isolates derived from the further evolution of CB1000, with 242 to 267 mutations present in these strains. Transition mutations dominated the mutational spectrum.

One straightforward pattern noted in the earlier study was continued. The only genetic change that is universal to all strains sequenced is deletion of the e14 prophage, a defective lambdoid prophage that is 15.4 kb in length. The deletion of e14 occurs early, within the first three rounds of selection for IR resistance (E. A. Wood and J. R. Battista, unpublished data).

Population IR-1-20 has a particularly complex population structure and is strikingly different from the IR-2-20 and IR-3-20 populations. Outside of the e14 deletion, there are no mutations shared by all seven of the IR-1-20 isolates. There are

three groupings of mutations in subsets of the clonal isolates that reflect clonal interference (*16-18*) in this population (J. R. Battista, unpublished results). In contrast, isolates from the IR-2-20 and IR-3-20 populations each exhibit a number of mutations (23 in IR-2-20 and 30 in IR-3-20) that are present in all seven isolates from their respective populations, and are effectively fixed. The apparent fixation of a mutation in one of these populations may reflect its importance to the phenotype. Alternatively, this may reflect a genetic bottleneck at some stage in the evolution of these populations, and speaks to the close relationship of the isolates.

2.3.2 Identification of mutations contributing to IR resistance in IR-2-20.

To identify the mutations that are most likely to contribute to the IR resistance phenotype, we focused on genetic alterations that exhibited the following criteria: (I) the mutation is present in all or most isolates sequenced in at least one population, (II) the mutated gene was a prominent mutational target in at least one other sequenced population or mutations are found in genes in the same operon or pathway in other populations. Application of the second criterion requires data that was unavailable in our earlier study (*1*), and provides a pattern implicating a particular mutation in the acquisition of an extreme IR resistance phenotype. This criterion also assumes that there is a significant level of phenotypic parallelism (*14*, *19*) between the independently evolved populations. Table 2 summarizes the results of applying these criteria, with 9 altered genes or systems meeting these two criteria. The entire complement of mutations

found in all isolates from the 4 evolved populations, both those sequenced for this study and those analyzed in the earlier study (*1*), are provided in Table S2.

The prominent genetic alterations present in the various evolved isolates are not identical. We thus set out to define the genetic basis of extreme IR resistance in one representative isolate. We focused on the isolated strain CB2000 from the IR-2-20 population for several reasons. First, CB2000 is one of the most radiation resistant of the isolated strains from the original evolved populations (*1*). Second, it is particularly well characterized ((*1*); and this study). Third, it was isolated from a population that lacks any sign of clonal interference, allowing us to focus on patterns present in a relatively simple population. Finally, the patterns of mutations we highlight below overlap patterns that are evident in other populations, making CB2000 a good barometer of mechanisms by which radiation resistance evolved in many of our strains. Of the 9 genes and/or systems reflected in Table 2, seven are represented in CB2000.

The candidate CB2000 genes can be grouped into three functional categories: (I) DNA repair and replication (*recA, dnaB, yfjK*), (II) oxidative damage suppression (*rsxB* and *gsiB*), and (III) cell wall biogenesis (*wcaK* and *nanE*). All of these represent prominent mutational patterns.

DNA metabolism. Every evolved isolate sequenced features a mutation in one or another component of the replication restart primosome (*priA*, *priC*, *dnaB*, *dnaT*), critical for the origin-independent restart of stalled DNA replication forks after recombinational

DNA repair has occurred (*20-22*). A mutation in the *recA* gene appears in 13 of 22 isolates from the original four evolved populations, appearing in every evolved population except IR-3-20. The RecA D276N substitution is fixed in IR-2-20 (including CB2000), and the identical allele also appears in the separately evolved isolate from IR-4-20, CB4000. Two different variants of RecA, D276A and A289S, appear in strains derived from population IR-1-20, helping to define two of the three sub-populations exhibiting clonal interference. Mutations in the cryptic helicase gene *yfjK* are present in three of the four populations, although a mutation appears in only one of three sub-populations of IR-1-20.

Oxidative damage suppression. Over the sequenced isolates, quite a few mutations clustered in components of the *rsxABCDGE* gene cluster and *gsiB*. The *rsx* gene cluster encodes a multi-subunit complex involved in the reduction of SoxR, part of a regulatory circuit for up to 50 target genes. Some of these are directly involved in protection against reactive oxygen species (*23*). The *gsiB* gene encodes an ABC transporter for glutathione. Different mutations in *gsiB* are fixed in IR-2-20 and IR-3-20.

Cell wall biogenesis. Mutations of *wcaK* and *wcaC* appear frequently in populations IR-2-20 and IR-3-20, respectively, and also appear in *wcaA* in the single isolate from IR-4-20. The pattern is reinforced by the appearance of a *wcaL* mutation in 5 of the 9 isolates derived from the continued evolution of CB1000. The products of the *wca* operon are involved in the synthesis of secreted exopolysaccharides that confer

significant resistance to heat and acid stress (*24*). Mutations in *nanE* and *nanT* are fixed in IR-2-20 and IR-3-20, respectively. The pattern is again reinforced by the appearance of a mutation in *nanA* in all nine of the sequenced isolates derived from the continued evolution of CB1000 (Table S3). The *nan* operon is involved in the recycling and/or synthesis of components of the bacterial cell wall, particularly N-acetylmannosamine and N-acetylneuraminate (*25*).

2.3.3 Defining the genetic basis of directly evolved extreme radiation resistance in CB2000.

To assess the importance of these seven CB2000 mutations quantitatively, we took 3 approaches:

(I) Mutations identified in CB2000 were moved into the radiosensitive Founder strain, This was done both individually and collectively with one or more of the other mutations, The e14 deletion (Δe14) was included in the wild type background, since this genetic alteration occurs very early in the evolution of IR resistance in all of our strains, and the effects of this deletion have already been characterized (*1*). If any of the 7 mutations we identify above are contributing to the phenotype, they are doing so in a background that includes the e14 deletion. In addition to the seven mutations identified in Table 2, we selected the *glpD* mutation as a target for analysis. This mutation was fixed in population IR-1-20. Although this gene was not mutated in any other population, a mutation that might affect regulation of *glpD* was found in four isolates of IR-3-20.

(II) The same eight total mutations (including *glpD*) were reverted back to the original Founder sequence in CB2000 both individually and in combination with other mutations. This allows us to determine if the mutations contribute to IR resistance in a genetic background that includes all or most of the other 69 mutations present in CB2000.

(III) We deleted the nonessential genes carrying these mutations in the radiosensitive Founder strain individually. We wished to determine if a definitive knockout of an individual gene could mimic any observed effects of the individual mutations observed in CB2000. If the answer was yes, we reasoned that the mutation we observed in CB2000 was likely to be a loss of function mutation.

All of the strains constructed for this effort were assayed for survival to 3000 Gy to measure the contribution of individual mutations and mutation combinations to the IR resistance phenotype.

2.3.4 Individual mutation contributions to the IR resistance phenotype.

All of the mutations identified above made a measurable contribution to the phenotype, some large and some small. As illustrated in Figure 1A, moving the CB2000 alleles of *gsiB*, *rsxB*, *glpD*, and *wcaK* individually into the Founder Δe14 background resulted in small increases in radioresistance, ranging from 2 (*wcaK*) – 7 (*gsiB*) fold. Each of these mutations thus appears to contribute to the IR resistance phenotype, but in a relatively small way. In contrast, moving the CB2000 alleles encoding the RecA D276N,

YfjK A151D, or DnaB P80H proteins individually into Founder Δe14 resulted in 1-2 log increases in IR resistance, indicating that these mutations contribute in a more substantial way to the observed IR resistance of CB2000. The effects of the individual mutations on IR resistance in the Founder Δe14 background were echoed by the effects of reverting these same mutations to wild type in the CB2000 background (Fig. 1B). Individually reverting the *recA*, *dnaB*, and *yfjK* alleles back to the Founder sequence resulted in 1-2 log decreases in resistance, with the *recA* reversion resulting in the largest loss of resistance with an approximately 50 fold decrease in IR resistance relative to CB2000. Reverting the other mutant alleles (*gsiB, rsxB, glpD, wcaK*, and *nanE*) to wild type in the CB2000 background had significant but relatively small effects.

Mutation combinations. Combining multiple mutations in the Founder Δe14 background allowed us to assess whether mutations were acting additively or affected single pathways. As shown in Figure 1A, the deletion of the e14 prophage along with point mutations in *recA, dnaB*, and *yfjK* accounted for much of the radiation resistance of CB2000. Of the other 5, the *gsiB* mutation appeared to contribute the most (an additional 3 fold) when combined with the DNA repair function mutations. A strain incorporating the *recA, dnaB*, *yfjK, wcaK, and gsiB* mutations accounted for all but 2 fold of the overall increase in radiation resistance in CB2000. Further supporting the importance of the *recA, dnaB*, and *yfjK* mutations, Figure 1B illustrates that the radiation resistant phenotype observed in CB2000 was eliminated in its entirety by converting only these

three genes back to the wild type sequence. The effects of the other mutant alleles (*gsiB*, *rsxB*, *glpD*, *wcaK*, and *nanE*) were small to insignificant in all cases.

Mutations in *recA* and *dnaB* appear to affect the same pathway while the *yfjK* mutation has an additive effect. The *recA* and *dnaB* mutations in CB2000 were individually reverted back to the Founder sequence resulting in a ~50 and ~10 fold decrease in resistance from CB2000, respectively (Figure 1B). Interestingly, a strain carrying both mutations was no more sensitive than the single *recA* mutant, suggesting that the mutations in *recA* and *dnaB* likely affect the same cellular process. Upon reversion of the CB2000 *yfjK* mutation to wildtype, a 15-fold decrease in radiation resistance was observed (Figure 1B). This *yfjK* mutation appears to affect a different pathway than the mutations in *recA* and *dnaB*. When the mutations in *recA*, *dnaB*, and *yfjK* were reverted to wild type together, CB2000 was quantitatively as sensitive to irradiation at 3000Gy as was the initial Founder Δe14 strain, as summarized in Figure 1D.

Several of the key results of Figure 1D were confirmed in direct competition experiments (Fig. 2) (*26*). We incorporated a neutral Ara$^-$ mutation into particular strains (which confers a red color on colonies when grown on tetrazolium arabinose (TA) indicator plates) to permit color based scoring of mixed populations (*26*). When CB2000 (Ara$^-$) was mixed in a 1:1 ratio with CB2000 in which the *recA*, *dnaB*, and *yfjK* mutations had been converted to wild type, and irradiated at 2,000 and 3,000 Gy, CB2000 exhibited

at least a 2 log advantage in survival at the higher IR dose (Fig. 2A). Similarly, Founder Δe14 (Ara⁻) was mixed with the same background incorporating the *recA*, *dnaB*, and *yfjK* mutations. A 2 log advantage for the strain incorporating the three mutations was seen after irradiation at 2,000 Gy (Figs 2B and 2C). When CB2000 in which the *recA*, *dnaB*, and *yfjK* mutations had been converted to wild type was mixed with Founder Δe14 (Ara⁻), survival of the two strains was similar (the CB2000 reversion strain had less than a 0.5 log advantage (Figs 2D and 2E)).

2.3.5 Mutations in *wcaK*, *gsiB*, *yfjK*, *nanE*, and *glpD* likely result in a loss of function.

Nonessential genes mutated in these studies were individually deleted in the Founder Δe14 background and the resulting strains were assayed for survival of 3,000 Gy and shown in Figure 1C. Strains deleted for *wcaK*, *gsiB*, *glpD*, and *nanE* respectively are 2-3 fold more resistant than Founder Δe14. For *wcaK* and *glpD*, this increase in resistance is similar to the resistance observed in Founder Δe14 strains carrying the point mutations in these genes, suggesting that the mutations result in loss of function. The 2-fold increase in resistance conferred by a *gsiB* deletion is less than that of the *gsiB* point mutation, suggesting a more complex situation. The situation with *yfjK* is similar. Founder Δe14 Δ*yfjK* cells are 13 fold more resistant than the isogenic parent strain as illustrated in Figure 1C. However, the Founder Δe14 strain encoding YfjK A151D is an additional 10 fold more resistant than the deletion, and nearly 2 orders of magnitude more resistant than Founder Δe14. It is possible that the mutation in *yfjK* is a partial loss

of function, or that it eliminates one or more activities of the YfjK protein but not all of them.

As detailed in Table 2, two additional mutational patterns are evident in the overall populations that are not manifested in CB2000. Mutations affecting the *clpXP* complex are prominent in populations IR-1-20 and IR-3-20, and new mutations potentially affecting this system also appear in the further evolution of CB1000. *clpXP* encodes a proteolytic system involved in turnover of many key cellular proteins including many involved in DNA metabolism (*27*), and is essential for the transition from growth to stationary phase. It is possible that this complex is essential for the elimination of proteins that have been inactivated by oxidative carbonylation. A mutation in *fnr* is fixed in IR-3-20. While this gene is not targeted in any of the other three original populations, its importance is suggested by the appearance of a mutation that leads to a truncated Fnr protein in all nine isolates derived from the further evolution of CB1000 (Table S3). Fnr is a regulator that oversees the transition between aerobic and anaerobic metabolism (*28*).

2.3.6 Extreme IR resistance does not entail significant changes in the transcriptome, metabolome, or metal ion content.

We examined the baseline metabolic profiles of the wild type Founder strain and several of the IR resistant isolates. The strains were not subjected to irradiation prior to analysis. The results measure the state of the cells when they first encounter irradiation.

We used RNA-Seq (*29*) to directly sequence and map RNAs that are expressed in the Founder, CB1000, and CB2000, with the comparison reported in Table S4. Overall, there were few genes with different expression profiles when comparing the strains with a 1.5-fold cutoff enforced. The only commonality in expression patterns between the evolved strains was the >1.5-fold decrease in the transcript abundance compared to Founder of the following genes: *fruBKA* (fructose metabolism), *sdaC* (serine transport) and *proK* (proline tRNA). Transcription of the entire fimbrial operon is increased in CB1000, possibly due to phase variation of the *fimS* region, but in CB2000, only the *fimC* gene exhibits an increase in transcription of more than 1.5-fold. In both evolved strains *icd* (b4519) transcript levels are increased, likely due to the excision of the e14 prophage, which reconstitutes the *icd* gene with a different 3′ end of the gene containing 2 base substitutions. In general, changes are minimal. There are few genes that display a newly constitutive level of expression that is strikingly higher than in the parent Founder strain, in spite of the existence of mutations in a number of genes encoding global regulators.

Using NMR, we investigated metabolite concentrations in the total soluble fraction of the cytosol from Founder, two evolved radioresistant strains, CB1000 and CB2000, and in CB1013 (an isolate from the IR-1-20 population that has a distinctively different mutational profile relative to CB1000) (Figure 3A). There are no significant changes in metabolites between any of the strains measured, with the possible exception

of small apparent decreases in the levels of acetate and succinate in CB2000 as compared to Founder. Unlike *D. radiodurans*, there is no evidence of accumulation of intermediate metabolites that could act as antioxidants. The measured metabolites included glutathione, a molecule that plays a particularly important role in cellular redox chemistry.

We also used trace metal analysis to measure total metal content of all strains for which we obtained genomic sequences. Mn/Fe ratios are reported in Figure 3B, and complete listings of metal ion measurements are provided in Figure S1. In spite of the demonstrable increases in radiation resistance exhibited by all of our isolates, we did not see a significant change in the Mn/Fe ratio (nor significant increases in the concentration of either metal) in most of our directly evolved highly radioresistant strains of *E. coli*. There is a minor elevation in manganese in the one evolved isolate from population IR-4-20 (CB4000) and in some strains derived from the further evolution of CB1000. There is no universal change in metal content in the evolved strains that mirror the apparent adaptation seen in *D. radiodurans* and other IR-resistant bacteria.

2.4 Discussion

To the three potential mechanisms described in the introduction of this article – amelioration of protein oxidation, novel DNA repair systems, and nucleoid condensation – we now document a fourth. IR resistance can be increased – dramatically – by functional enhancement and/or adaptation of existing DNA repair enzymes.

Although classical DNA repair systems may be shared widely in different IR resistant and IR sensitive bacterial species, those DNA repair systems are biologically malleable. Classical DNA repair pathways can adapt to facilitate more efficient repair when cells are exposed to high levels of IR. In our evolved isolates, adaptations to DNA repair, involving genetic alterations in well-studied enzymes that are present in most bacterial species, represent a substantial and sometimes the dominant adaptation that contributes to the IR resistance phenotype. In one well-characterized isolate, three mutations – all in DNA repair functions – largely account for the increase in IR resistance.

The metabolic profile of our cells indicates that the evolved strains possess little or no unusual gene expression, metabolite concentration, or metal ion adaptations when they first encounter irradiation. We acknowledge that irradiation may lead to an alteration of the profile of genes induced in the IR resistant isolates, and those isolates may thus adapt to the challenge of irradiation more rapidly. In CB2000, changes in the *rsxB* gene might bring about some significant and beneficial changes in the IR response. However, in the case of CB2000, it is clear that changes to DNA repair genes – not regulatory genes – play the major role in the observed IR resistance.

Whereas the results highlight the effects of changes to DNA repair systems, there are other effects clearly evident in the data. Evolution of a complex phenotype, such as extreme resistance to ionizing radiation, is not brought about by changes in one gene.

There are multiple paths, and multiple mutations affecting multiple cellular systems make significant contributions. The mutational patterns suggest that improvements in the amelioration of protein oxidation may also play a role in some of the evolved populations, although that role appears to be relatively minor in CB2000. Given the genes affected, the mechanisms contributing to the amelioration of protein oxidation may be somewhat different than those described to date for *D. radiodurans*. Contributions by additional biochemical processes appear likely, and remain to be explored.

2.5 Materials and Methods

*Bacterial strains used in this study***.** All strains used in this study are *E. coli* K-12 derivatives and are listed in Table S1. Genetic manipulations were performed by site directed mutagenesis (Stratagene) and as previously described (*30*).

Directed evolution of Escherichia coli. Radioresistant populations IR-1-20, IR-2-20, IR-3-20, and IR-4-20 were generated in a directed evolution experiment described previously (*1*). To further evolve the CB1000 isolate, 1 ml of mid logarithmic phase liquid culture grown in LB was placed into two 1.5 ml plastic tubes. One was archived at –80°C, and the other was exposed to IR (^{60}Co source from a Shepherd model 484 irradiator; 19 Gy/min). After irradiation, appropriate dilutions were plated to estimate survival. The balance of the irradiated culture was used to inoculate fresh LB broth. Survivors were allowed to grow to stationary phase (12 to 18 h) and the protocol was

repeated. The dose of radiation administered was adjusted to allow approximately 1% survival after 1 day at 37°C, with the dose increasing as radio-resistance increased. Survivors at the end of 20 and 30 rounds of irradiation constitute a population of cells, designated IR-CB1000-20 and IR-CB1000-30, respectively. Single colony isolates from both populations were isolated and designated with the prefix "CB."

High-throughput sequencing using an Illumina instrument. New evolved isolates described in this study, were sequenced at the Joint Genome Institute, Walnut Creek, CA, as described previously (*1*).

IR survival assay. Cells from a fresh single colony of each strain were cultured in Luria-Bertani (LB) broth (*31*) at 37° C with aeration. After growth overnight, cultures were diluted 1:1000 into 10 mL fresh LB broth in 125 mL flasks and grown at 37° C with shaking until an optical density (OD_{600}) of ~0.2 was reached. Each culture was incubated on ice for 5 minutes before a 1mL sample was transferred to an eppendorf tube and irradiated in a Mark I ^{137}Cs irradiator (from J. L. Shepherd and Associates) for a time corresponding to 3kGy (~7 Gy/min). Irradiated samples as well as the unirradiated control samples for each culture were diluted appropriately, and plated on LB 15% agar medium to determine the total number of colony forming units (CFUs). Percent survival was calculated by dividing the titer of the surviving population by the titer of the unirradiated control sample. For each strain, 3-5 biological replicates were carried out.

IR survival competition assays. Cells from a fresh single colony of each strain were cultured in Luria-Bertani (LB) broth (*31*) at 37° C with aeration. After growth overnight, competition cultures were started by inoculating 10mL fresh LB broth with 35mL of competition Ara+ and Ara$^-$ strains in 125 mL flasks and grown at 37° C with shaking until an optical density (OD_{600}) of ~0.2 was reached. Each competition culture was incubated on ice for 5 minutes before a 1mL sample was transferred to an eppendorf tube and irradiated in a Mark I ^{137}Cs irradiator (from J. L. Shepherd and Associates) for a time corresponding to 2kGy and 3kGy (~7 Gy/min). Irradiated samples were diluted appropriately, and plated on TA plates to determine the total number of surviving red and white colony forming units. A non-irradiated control sample for each competition culture was diluted and plated to determine the titer of each culture and the percent Ara+ vs Ara$^-$ cells before irradiation. For each competition, three biological replicates were carried out. Percent survival was calculated by dividing the titer of the surviving population by the titer of the non-irradiated control sample, for both Ara+ and Ara$^-$ strains. Selection rate, r, also referred to as log(advantage), was calculated as $\log(N_{1(IR)})/(N_{1(No\text{-}IR)}) - \log(N_{2(IR)})/(N_{2(No\text{-}IR)})$, where $N_{1(No\text{-}IR)}$ and $N_{2(No\text{-}IR)}$ represent the initial densities of the two competing strains before IR treatment, and $N_{1(IR)}$ and $N_{2(IR)}$ represent the corresponding densities after IR exposure.

Normally, in a direct competition experiment, plates with fewer than 20 colonies of either competitor are usually excluded to reduce the effect of outliers caused by low

counts (*32*). However, because the differences in fitness after IR treatment is so great between CB2000 Ara⁻ and CB2000 wtRecA wtDnaB wtYfjK and between Founder Δe14 Ara⁻ and Founder Δe14 RecA D276N DnaB P80H YfjK A151D, it was virtually impossible to retrieve at least 20 colonies of the sensitive strains in a range that we could also use to calculate the density of the resistant strains. The selection rates in Figure 2E are approximate, because there were less than 20 colonies counted on plates for the sensitive strains. However, the trend that we show in Figure 1 is strongly conserved. By reverting the three mutations in DNA metabolism genes, CB2000 loses virtually all of its IR resistance. We report in Figure 2E that when treated with 2000Gy, CB2000 Ara- has *at least* a two log fitness advantage over CB2000 wtRecA wtDnaB wtYfjK and inversely, Founder Δe14 Ara- RecA D276N DnaB P80H YfjK P80H has *at least* a two log advantage over Founder Δe14. Because their sensitivities to IR were so similar, we did not have a problem with retrieving more than 20 colonies for each strain in the competition of Founder Δe14 and CB2000 wtRecA wtDnaB wtYfjK . Rather, CB2000 wtRecA wtDnaB wtYfjK had less than a half log advantage (less than 3 fold) over Founder Δe14, again illustrating the importance of mutations in these three genes for extreme radiation resistance.

RNA-Seq. This method is reviewed in (*33*). *Sample preparation.* Samples were prepared as described in (*34*) with modification as described here. *Cell growth.* Overnight cultures from single-cell inoculates grown in LB were used to inoculate 20 mL

of LB in a 125 ml flask with appropriate antibiotic to an initial OD_{600} of 0.02. Cultures were grown at 37 °C with aeration until an OD_{600} of ~0.2 was reached. Aliquots of 15mL of each culture were mixed with 30mL of RNAprotect bacterial reagent (Qiagen), inverted to mix, and incubated at room temperature for 5 minutes. Cells were centrifuged at 4000xg for 20 minutes at 4 degrees and the cell pellets were stored at -80 °C.

Total RNA isolation. Total RNA was isolated using the MasterPure RNA purification kit according to the manufacturer's specifications (Epicentre). Nucleic acid pellets were treated with 0.05 U/μl DNase I for 45 min at 37 °C and then repurified with MasterPure.

mRNA enrichment and cDNA synthesis. Ten μg of total RNA was enriched for mRNA by targeted removal of rRNA using the MICROBExpress™ bacterial mRNA enrichment kit (Ambion). The resulting enriched mRNA was isopropanol precipitated, and the pelleted mRNA resuspended in TE. The enriched mRNA concentration was quantified by A_{260} measurement on a NanoDrop 1000 instrument. Ten μg of purified total RNA was reverse transcribed using the Superscript II double-stranded cDNA kit (Invitrogen) followed by RNase digestion and cDNA purification by phenol chloroform extraction and precipitation.

Illumina Sequencing and Analysis. cDNA samples were submitted to JGI for library preparation and sequencing using the Illumina Genome Analyzer IIx to generate

single-ended 36 bp reads. Libraries were prepared for sequencing according to the manufacturer's instructions.

Analysis. Analysis was performed using the CLC-Bio Genomics Workbench version 3.7. There were two biological replicates for each of the three samples (Founder, CB1000, CB2000). Read ends were trimmed to remove low quality and ambiguous bases and all reads less than 20 nt were discarded. Trimmed reads were mapped to the annotated CDSs of the reference genome (*E. coli* K-12 MG1655 m56 reference genome, RefSeq Accession Number NC_00913.2) with two mismatches allowed and 10 bases of each read were allowed to map beyond ORF boundaries. Expression was calculated independently for each duplicate sample. Any read that could be mapped to more than 4 locations was discarded. Genes encoding rRNA and tRNA transcripts were masked by removing their annotations from the reference genome prior to mapping so that did not affect normalization expression estimates of protein-coding genes. Expression values were reported in RPKM (*35*).

Differential Expression. Differential expression, reported as fold-change, was determined by separately comparing the expression estimates of each pair of duplicate samples to the control samples. CB1000 and CB2000 were also directly compared. Prior to analysis, all expression values (E) were transformed into $\log_2(E+1)$ and then standardized by adjusting each sample to the expression level corresponding to 1 million mapped reads. Fold changes were tested using Baggerly's test (*36*) on the fold

changes estimated both from the original RPKM values and the transformed/standardized values. The resulting p-values were independently corrected using the Bonferroni method and via a determination of the false discovery rate (FDR). Heat maps were generated from the transformed/standardized expression values and dendrograms showing the clustering of genes were computed using complete linkage with Pearson product-moment correlations as the distance metric.

Growth of E. coli for metal analysis. Flasks were soaked in 2 N nitric acid for 12 h and then transferred into 1% (vol/vol) nitric acid for 24 h and rinsed in high-purity water to minimize metal contamination. Cells of CB1000, CB2000, CB3000, and CB4000 were cultured in LB broth at 37° C (*31*) with aeration in polypropylene tubes. After growth overnight, cultures were diluted 1:100 into 25 mL fresh LB broth in 50mL polypropylene tubes and grown at 37° C with shaking until an optical density (OD_{600}) of ~0.2 was reached. This was performed in triplicate for each strain. Cultures were chilled for 10 min on ice before 10ml were pelleted and resuspended in 1mL fresh LB. Aliquots of 0.6 ml of cells were transferred to treated microcentrifuge tubes prepared by the metal analysis facility and pelleted. Cell pellets were stored at -20 °C. To determine the titer of each cell pellet, appropriate dilutions were plated on LB 15% agar medium and incubated overnight at 37 °C.

Metal analysis. Cell pellets were submitted for analysis in acid-cleaned polypropylene vials and treated at 40 °C with ultrapure HNO_3 and subsequently diluted

to volume for analysis with 2% HNO_3. Samples were analyzed in the Trace Element Clean Laboratory at the Wisconsin State Laboratory of Hygiene, Madison, WI, using high-resolution inductively coupled plasma mass spectrometry. Each sample was measured twice. The metal content is reported as mg per pellet of bacteria submitted.

NMR sample preparation. Overnight cultures of Founder, CB1000, CB2000, and CB1013 were diluted into 1L M9 media containing glucose in 6 L flasks to an initial OD_{600} of 0.05 and grown at 37 °C with aeration. When an OD_{600} of ~0.80 was achieved, cultures were chilled on ice for 30 minutes before being centrifuged at 8K rpm for 10 min at 4 °C in a JLA 8.1 rotor. The supernatant was discarded, and cell pellets were washed in 25 mL of 1X M9 salts and transferred to a JA-20 tube before centrifuged 20 min at 5,000 *g* at 4 °C. The supernatant was discarded and 16 mL of boiling water with 250 mM MES (2-(N-morpholino)ethanesulfonic acid) was added to each pellet, vortexed briefly to loosen the pellet, and placed in boiling water for 7.5 min. Tubes were briefly vortexed again, and then centrifuged in a JA-20 rotor at 7,000 *g* for 20 min at 4 °C to clear cell debris. The supernatant was poured off into a clean 50 mL sterile polypropylene tube and frozen.

NMR sample processing. Supernatants were transferred to microfilters (Sartorius Stedim Vivaspin 20, 3000 MWCO). The low MW fraction was frozen and lyophilized. Dried metabolites were dissolved in 800 µL D_2O containing 300 µL DSS and 300 µL NaN_3 and titrated to pH 7.40 (±0.01) with DCl/NaOD as needed. Samples were

transferred to 5 mm NMR tubes (Wilmad). Spectroscopy was performed at the Nuclear Magnetic Resonance Facility at Madison (NMRFAM) in Madison, WI on a 600 MHz Varian Spectrometer with a cryoprobe and VNMRJ software. Two dimensional ^{1}H-^{13}C HSQC spectra were acquired using 4 transits, 32 steady state transits, 0.3 second acquisition time, and 512 increments.

NMR analysis. Analysis was performed using rNMR, an open source software package developed at NMRFAM (*37*). To quantify signals, standard compounds of the observed metabolites were prepared at 2, 5, and 10 mM. The resonances from these compounds were linearly regressed in order to measure concentration as a function of intensity. The samples were normalized to 5 mM MES and their resultant peak intensities used to obtain concentrations of the measured metabolites using the regression slopes.

Figures

Figure 1: Effects of selected mutations on survival of *E. coli.*

Mid-logarithmic phase cultures were irradiated to 3,000 Gy and plated to measure survival as described in Materials and Methods. A, mutations discovered in CB2000 were moved individually and in combination into the Founder Δe14 background. Mutations present in a given strain are indicated by a + symbol. For reasons not understood, it proved impossible to move the *nanE* mutation into this background on its own. CB2000 itself is presented in the final lane. B, the same point mutations (this time including *nanE*) were mutated back to Founder sequence in the CB2000 background. The first lane is CB2000. Mutations converted to the wild type allele in a given strain are indicated by a – symbol. C, non-essential genes assayed in Panel A and B were deleted in the Founder Δe14 background, with the deleted gene indicated by a Δ symbol. D, The effects of mutations in genes *recA*, *dnaB*, and *yfjK* are summarized, with symbols as in panels A-C. All results were obtained with a ^{137}Cs irradiator (~7 Gy/min).

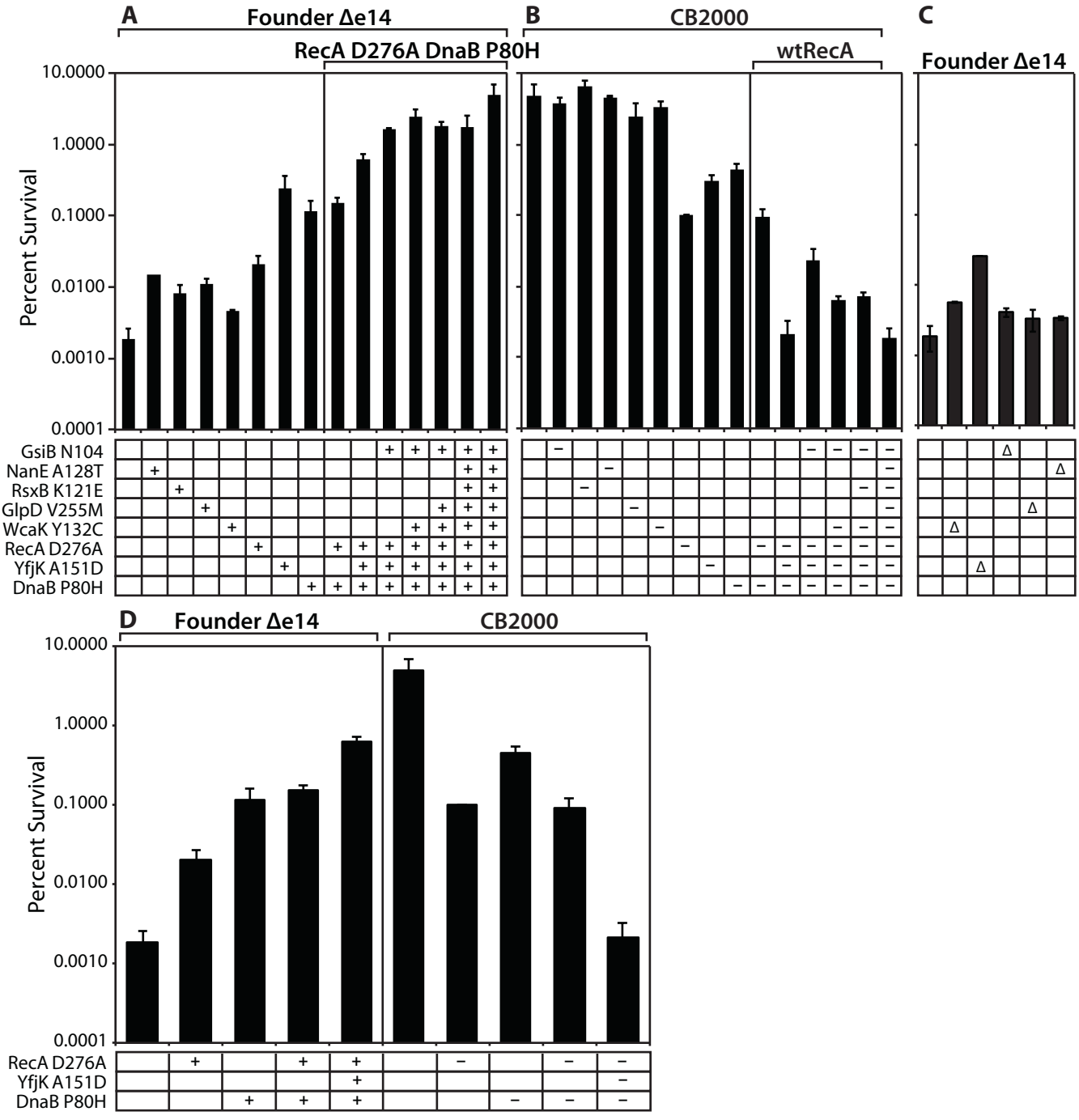

A
Founder Δe14
RecA D276A DnaB P80H
B
CB2000
wtRecA
C
Founder Δe14
Percent Survival
10.0000
1.0000
0.1000
0.0100
0.0010
0.0001
GsiB N104
NanE A128T
RsxB K121E
GlpD V255M
WcaK Y132C
RecA D276A
YfjK A151D
DnaB P80H
D
Founder Δe14
CB2000
Percent Survival
10.0000
1.0000
0.1000
0.0100
0.0010
0.0001
RecA D276A
YfjK A151D
DnaB P80H

Figure 2: IR survival in direct competition assays.

Mid-logarithmic cultures consisting of a 1:1 ratio of the competiting Ara+ and Ara$^-$ strains were irradiated to 2,000 and 3,000 Gy and plated on tetrazolium arabinose indicator plates to distinguish the frequency of survival for both strains. A. CB2000 Ara$^-$ (red, ○), *vs* CB2000 wtRecA wtDnaB wtYfjK, □ B. Founder *Δ*e14 Ara$^-$, (red, △), *vs* Founder *Δ*e14 RecA D276N DnaB P80H YfjK A151D, ◇ C. TA plate of 2,000 Gy survival competition of Founder *Δ*e14 Ara$^-$ *vs* Founder *Δ*e14 RecA D276N DnaB P80H YfjK A151D. The two red Founder *Δ*e14 Ara$^-$ colonies are indicated by red arrows. D. Founder *Δ*e14 Ara$^-$ (red, ◇), *vs* CB2000 wtRecA wtDnaB wtYfjK, □. E. Log advantage in survival to 2000Gy of CB2000 Ara$^-$ over CB2000 wtRecA wtDnaB wt YfjK (in black), Founder Δe14 RecA D276N DnaB P80H YfjK P80H over Founder Δe14 Ara$^-$(in grey), and CB2000 wtRecA wtDnaB wt YfjK over Founder Δe14 Ara$^-$(in white).

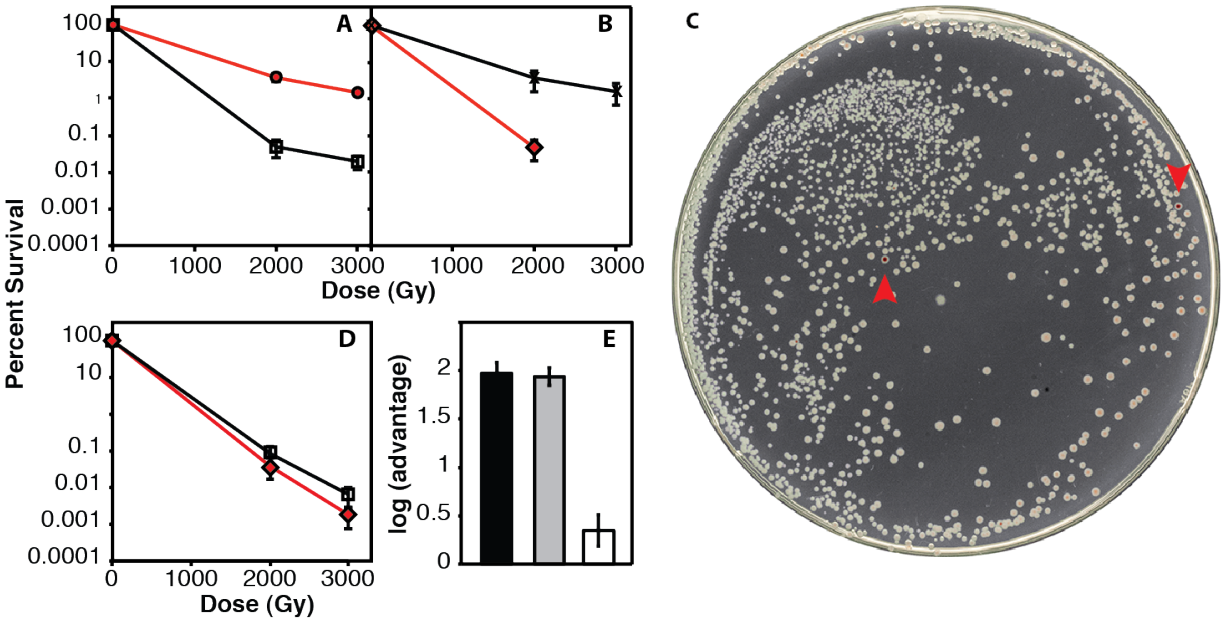

A
B
C
D
E
Percent Survival
Dose (Gy)
log (advantage)
100
10
0.1
0.01
0.001
0.0001
0
1000
2000
3000
2
1.5
1
0.5

Figure 3: Metabolite measurement and Mn/Fe ration analysis of evolved isolates.

A, Measurements of metabolites from three representative evolved *E. coli* strains as compared to Founder. Metabolites from whole cell pellets collected during logarithmic growth in LB were identified using a two-dimensional ^{1}H-^{13}C Heteronuclear Single Quantum Coherence (HSQC) experiment. Each metabolite is expressed as a ratio of the amount measured in the evolved strain (CB1000, CB1013, or CB2000; black, gray, and white bars, respectively) relative to the Founder. B, Ratios of manganese to iron are plotted for all isolates for which genomic sequences were obtained. The average increase in Mn/Fe ratio in strains derived from the further evolution of CB1000 is 1.4-fold.

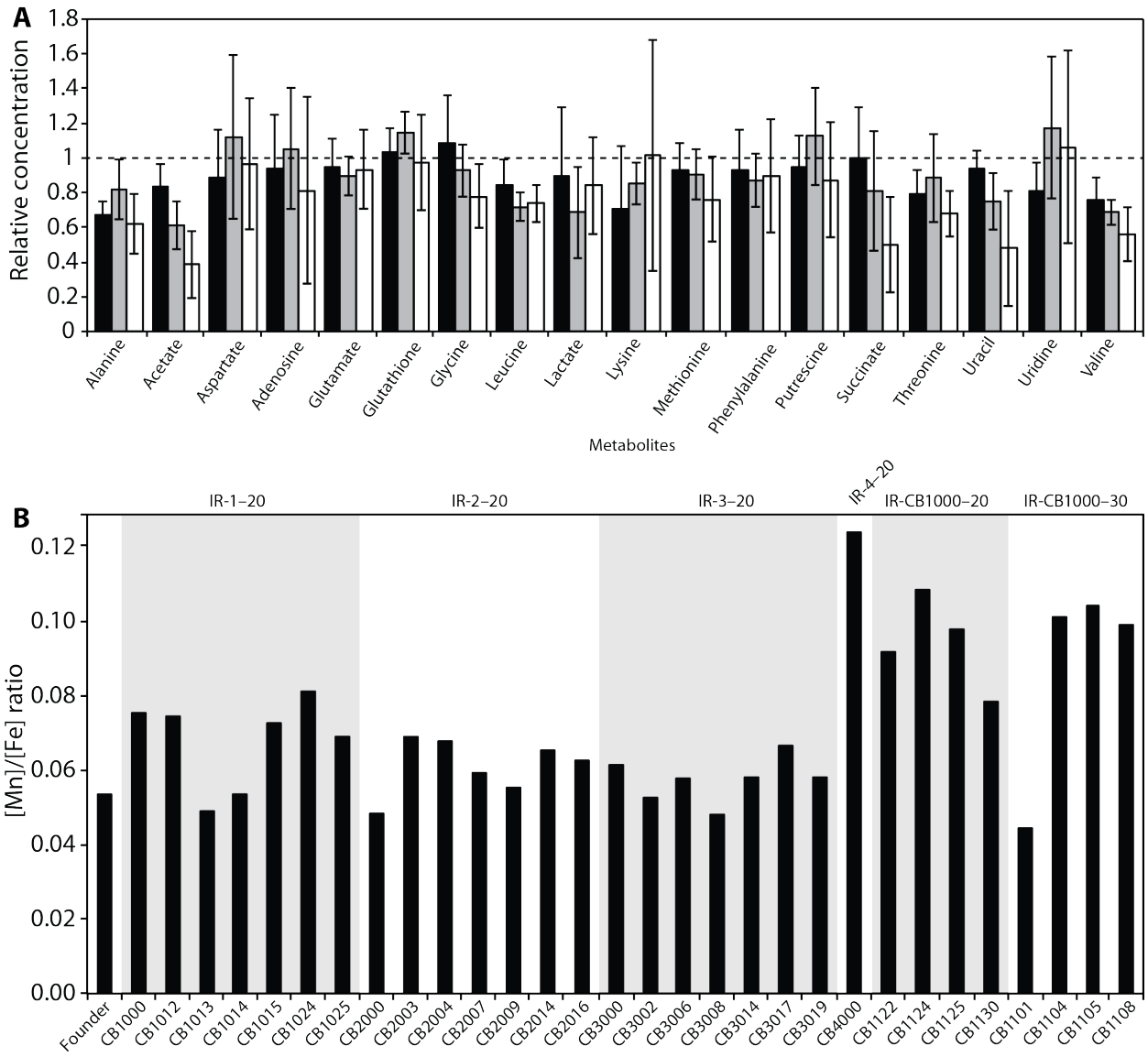

A
Relative concentration
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Alanine
Acetate
Aspartate
Adenosine
Glutamate
Glutathione
Glycine
Leucine
Lactate
Lysine
Methionine
Phenylalanine
Putrescine
Succinate
Threonine
Uracil
Uridine
Valine
Metabolites
B
IR-1–20
IR-2–20
IR-3–20
IR-4–20
IR-CB1000–20
IR-CB1000–30
[Mn]/[Fe] ratio
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Founder
CB1000
CB1012
CB1013
CB1014
CB1015
CB1024
CB1025
CB2000
CB2003
CB2004
CB2007
CB2009
CB2014
CB2016
CB3000
CB3002
CB3006
CB3008
CB3014
CB3017
CB3019
CB4000
CB1122
CB1124
CB1125
CB1130
CB1101
CB1104
CB1105
CB1108

Table 1: Summary of the mutational spectrum observed in strains derived from directed evolution of resistance to ionizing radiation.

The mutational spectrum of the population is inferred based on the genetic alterations observed in the subset of strains sequenced.

	IR-1-20							IR-2-20							IR-3-20							IR-4-20	IR-CB1000-20					IR-CB1000-30			
Transitions	**CB1000**	**CB1012**	**CB1013**	**CB1014**	**CB1015**	**CB1024**	**CB1025**	**CB2000**	**CB2003**	**CB2004**	**CB2007**	**CB2009**	**CB2014**	**CB2016**	**CB3000**	**CB3002**	**CB3006**	**CB3008**	**CB3014**	**CB3017**	**CB3019**	**CB4000**	**CB1122**	**CB1124**	**CB1125**	**CB1127**	**CB1130**	**CB1101**	**CB1104**	**CB1105**	**CB1108**
C → T	19	20	6	10	15	6	9	12	13	11	12	11	9	9	14	14	13	14	13	12	12	14	59	64	69	69	63	60	62	63	62
G → A	18	9	11	16	11	11	12	17	17	16	15	13	13	11	17	11	13	13	12	8	15	10	60	65	78	71	60	57	60	60	60
T → C	7	7	6	4	6	13	4	11	8	16	10	9	9	11	11	4	13	11	7	6	9	7	51	51	40	44	47	47	47	47	47
A → G	15	16	7	8	8	14	11	10	6	6	7	7	6	5	6	14	10	10	9	11	6	8	42	43	42	54	50	46	50	50	50
Transversions																															
G → T	4	4	1	1	1	3	5	3	4	2	3	5	4	2	5	5	4	4	4	6	4	2	10	9	7	8	6	5	6	6	6
C → A	1	1	2	1	1	5	6	5	3	6	3	3	4	6	2	4	2	1	3	4	1	6	3	2	5	5	6	4	5	5	4
A → C	1	1	2					1			1	1	2		1		1	2	2		1	1	7	4	2	2	4	4	3	3	3
T → G			1	1	1	1	1			1				1	1	1	1	2	1	2	2		2	4			1	1	1	1	1
G → C	1	1	1	2	2		1	1			1					1	1			1		1	2	1	1	2	4	3	3	3	3
C → G		1					2	3	1	1	1	1	1						1			2	1	2	1	1	2	2	2	2	2
A → T			1		2			1	1	3	1	2	1	4	3	1	1	2	1	1	1		2	2	1	2	3	3	3	3	3
T → A		1		1		1		1	1	1	1			1	1	1	3	1	1	1	2	1	3	3	2	3	3	3	3	3	2
Insertions																															
767bp IS1	2	1	1	1	1	1	1																								
Deletions	1																														
e14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
93bp															1	1	1	1	1	1	1										
Totals	70	67	44	49	53	59	56	69	55	64	56	53	50	51	64	57	63	61	55	53	54	53	248	256	254	267	256	242	252	252	250

Table 2: Summary of prominent mutational patterns observed in multiple evolved populations.

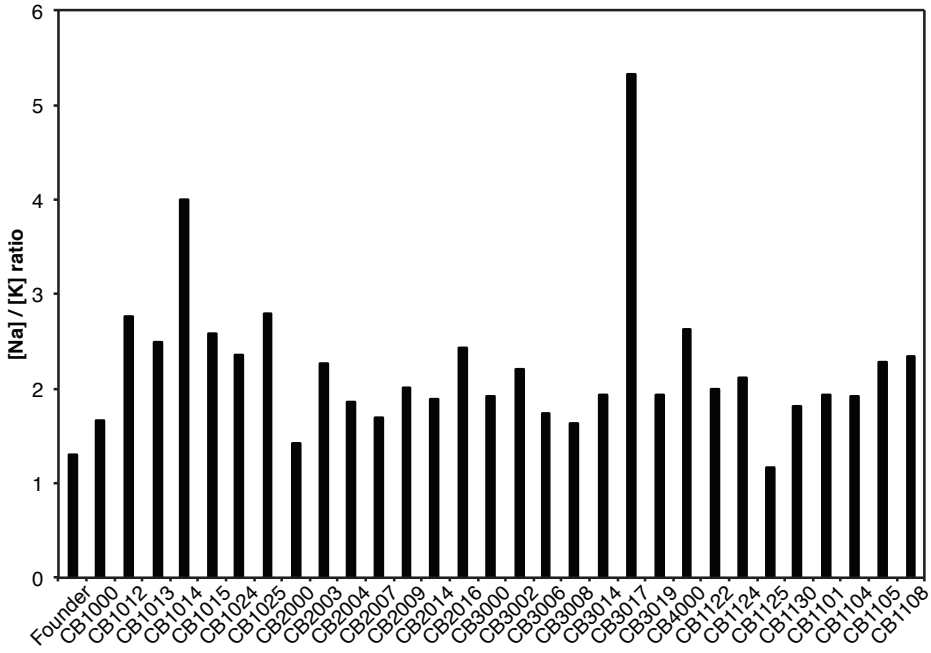
Entries in red denote mutations that are present in CB2000.

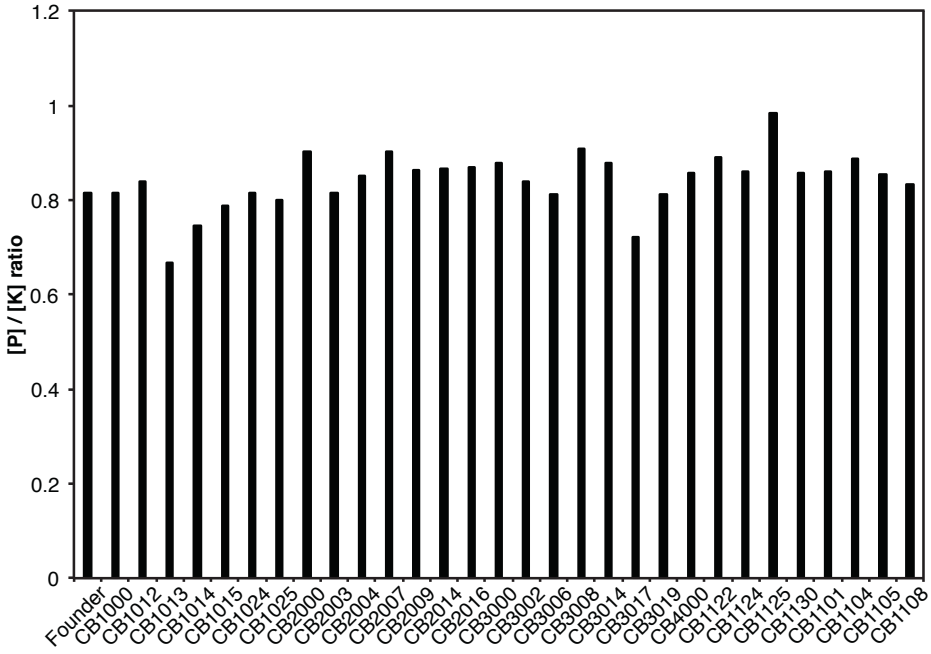
			IR-1-20							IR-2-20								IR-3-20							IR-4-20	IR-CB1000-+		
Gene	Position	Ref. Allele	CB1000	CB1012	CB1013	CB1014	CB1015	CB1024	CB1025	CB2000	CB2003	CB2004	CB2007	CB2009	CB2014	CB2016	CB3000	CB3002	CB3006	CB3008	CB3014	CB3017	CB3019	CB4000			Change	Mutation Type
clpP	456127	A	G	G																							Y75C	N
clpP / clpX	456637	G															A	A	A	A	A	A	A			+	-	I
clpX	457803	A				G																					Y384C	N
gsiB	868947	A								G	G	G	G	G	G	G											N104S	N
gsiB	869499	T															C	C	C	C	C	C	C				L288P	N
gsiB	870075	T																						C			V480A	N
fnr	1396995	A															T	T	T	T	T	T	T			+	F185I	N
rsxB	1704735	A								G	G	G	G	G	G	G										+	K121E	N
rsxD	1707299	T															C	C	C	C	C	C	C				V44A	N
wcaM	2113451	T																						C			N156S	N
wcaK	2116031	T								C	C		C	C	C											+	Y132C	N
wcaC	2129153	A															G	G	G	G	G	G	G				S313S	S
yfjK	2759609	G																						A			H651Y	N
yfjK	2760683	T						C																			K293E	N
yfjK	2760809	G														A											P251S	N
yfjK	2761108	G								T	T	T	T	T	T	T											A151D	N
recA	2820924	C						A																			A289S	N
recA	2820962	T			G	G	G		G																		D276A	N
recA	2820963	C								T	T	T	T	T	T	T								T			D276N	N
nanE	3368674	C								T	T	T	T	T	T	T										+	A128T	N
nanT	3369380	A															G	G	G	G	G	G	G				F415S	N
dnaB	4262560	T						C																			L74S	N
dnaB	4262578	C								A	A	A	A	A	A	A											P80H	N
dnaB	4262935	C																						A		+	P199Q	N
priA	4123174	C															T	T	T	T	T	T	T				V553I	N
priC	489549	A	G	G																							L162P	N
dnaT	4599105	G			A	A	A		A																		R145C	N

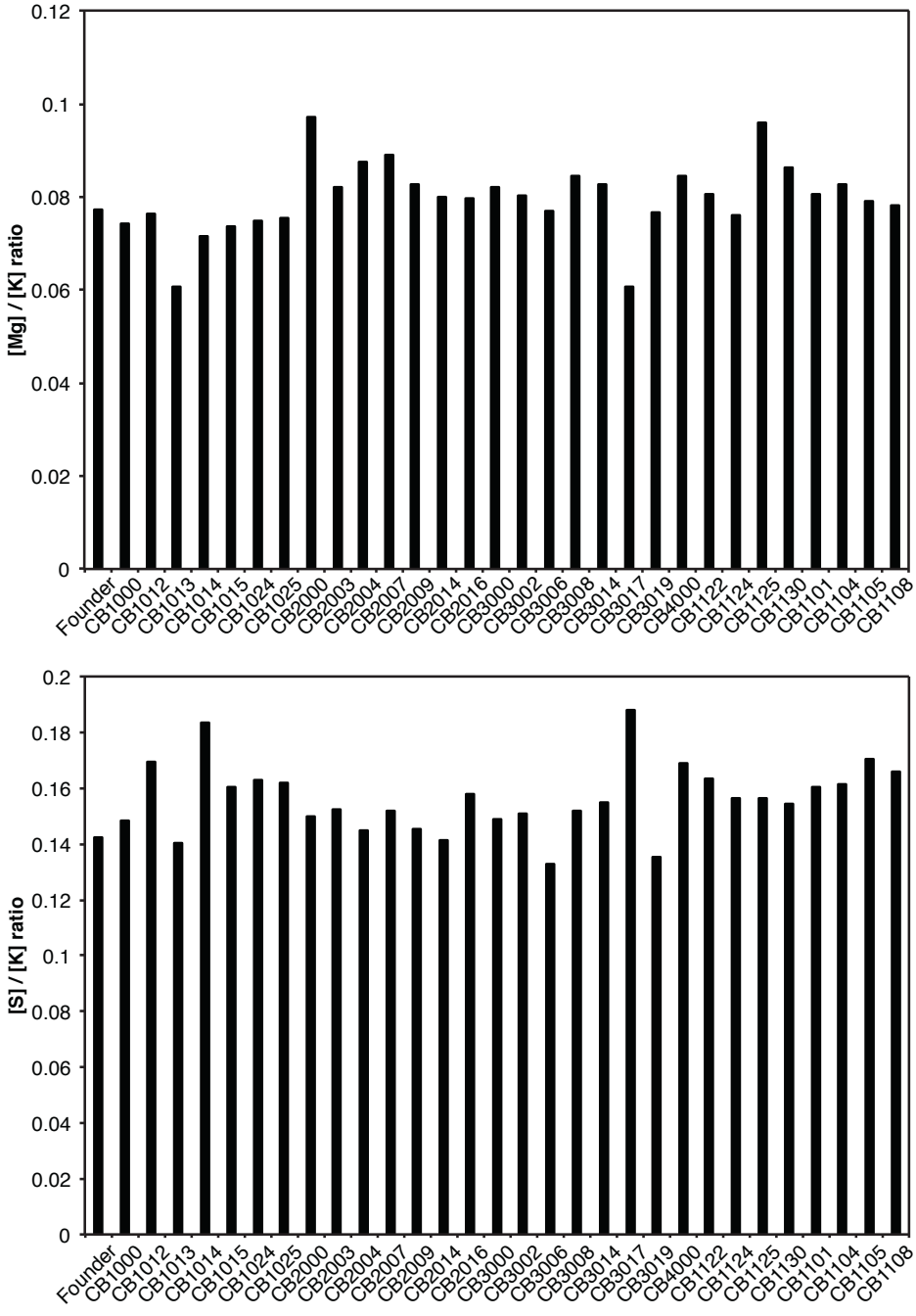
Supplemental Materials

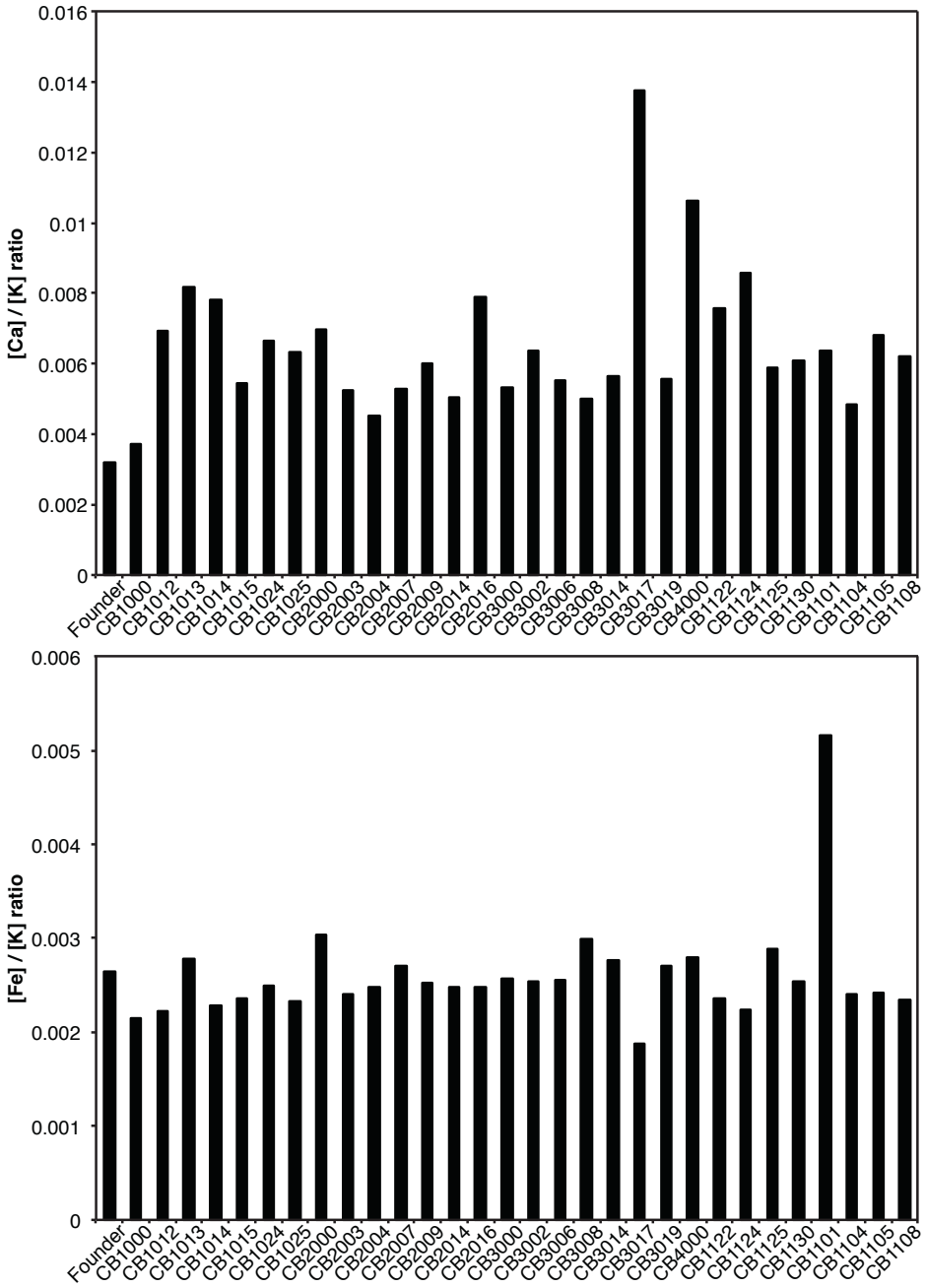
Figure S1: Relative levels of metals are unchanged in evolved *E. coli*.

Cells sampled during mid-logarithmic growth were subjected to trace metal analysis as described in Materials and Methods. The amount of each metal was plotted as a ratio normalized to the amount of potassium present in each sample, K, which was chosen arbitrarily to correct for variation in metal extraction.


[Na] / [K] ratio
0
1
2
3
4
5
6
Founder
CB1000
CB1012
CB1013
CB1014
CB1015
CB1024
CB1025
CB2000
CB2003
CB2004
CB2007
CB2009
CB2014
CB2016
CB3000
CB3002
CB3006
CB3008
CB3014
CB3017
CB3019
CB4000
CB1122
CB1124
CB1125
CB1130
CB1101
CB1104
CB1105
CB1108


[P] / [K] ratio
0
0.2
0.4
0.6
0.8
1
1.2
Founder
CB1000
CB1012
CB1013
CB1014
CB1015
CB1024
CB1025
CB2000
CB2003
CB2004
CB2007
CB2009
CB2014
CB2016
CB3000
CB3002
CB3006
CB3008
CB3014
CB3017
CB3019
CB4000
CB1122
CB1124
CB1125
CB1130
CB1101
CB1104
CB1105
CB1108


[Mg] / [K] ratio
0.12
0.1
0.08
0.06
0.04
0.02
0
[S] / [K] ratio
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Founder
CB1000
CB1012
CB1013
CB1014
CB1015
CB1024
CB1025
CB2000
CB2003
CB2004
CB2007
CB2009
CB2014
CB2016
CB3000
CB3002
CB3006
CB3008
CB3014
CB3017
CB3019
CB4000
CB1122
CB1124
CB1125
CB1130
CB1101
CB1104
CB1105
CB1108


[Ca] / [K] ratio
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
[Fe] / [K] ratio
0.006
0.005
0.004
0.003
0.002
0.001
0
Founder
CB1000
CB1012
CB1013
CB1014
CB1015
CB1024
CB1025
CB2000
CB2003
CB2004
CB2007
CB2009
CB2014
CB2016
CB3000
CB3002
CB3006
CB3008
CB3014
CB3017
CB3019
CB4000
CB1122
CB1124
CB1125
CB1130
CB1101
CB1104
CB1105
CB1108

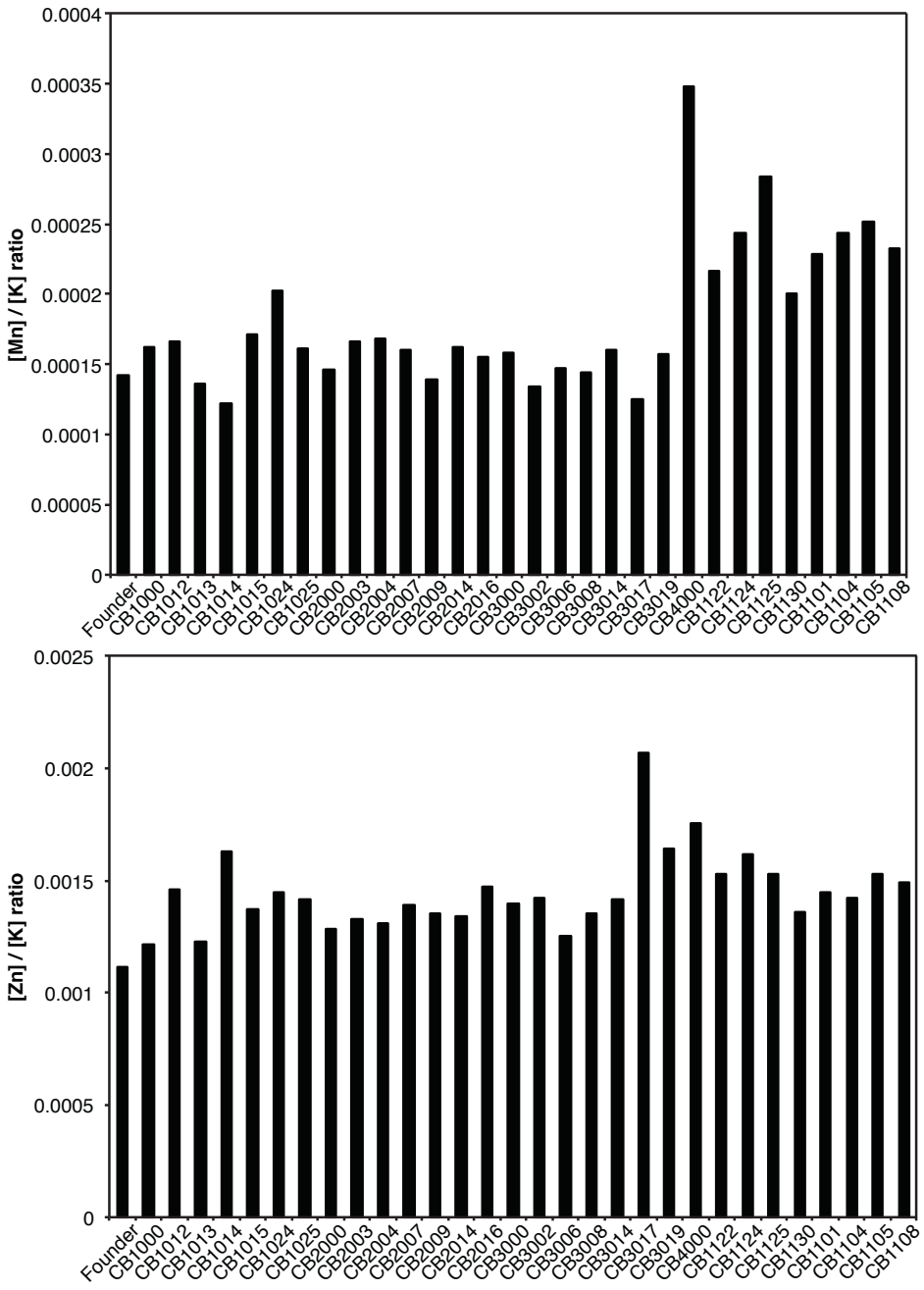

[Mn] / [K] ratio
0.0004
0.00035
0.0003
0.00025
0.0002
0.00015
0.0001
0.00005
0
[Zn] / [K] ratio
0.0025
0.002
0.0015
0.001
0.0005
0
Founder
CB1000
CB1012
CB1013
CB1014
CB1015
CB1024
CB1025
CB2000
CB2003
CB2004
CB2007
CB2009
CB2014
CB2016
CB3000
CB3002
CB3006
CB3008
CB3014
CB3017
CB3019
CB4000
CB1122
CB1124
CB1125
CB1130
CB1101
CB1104
CB1105
CB1108

Table S1: Strains used in this study

Name	Description	Reference
Founder	*E. coli* MG1655 single colony isolate	Harris *et al.*
CB1000	Population IR-1-20 isolate	Harris *et al.*
CB1013	Population IR-1-20 isolate	Harris *et al.*
CB1012	Population IR-1-20 isolate	Harris *et al.*
CB1013	Population IR-1-20 isolate	Harris *et al.*
CB1014	Population IR-1-20 isolate	Harris *et al.*
CB1015	Population IR-1-20 isolate	Harris *et al.*
CB1024	Population IR-1-20 isolate	Harris *et al.*
CB1025	Population IR-1-20 isolate	Harris *et al.*
CB2000	Population IR-2-20 isolate	Harris *et al.*
CB2003	Population IR-2-20 isolate	This study
CB2004	Population IR-2-20 isolate	This study
CB2007	Population IR-2-20 isolate	This study
CB2009	Population IR-2-20 isolate	This study
CB2014	Population IR-2-20 isolate	This study
CB2016	Population IR-2-20 isolate	This study
CB3000	Population IR-3-20 isolate	Harris *et al.*
CB3002	Population IR-3-20 isolate	This study
CB3006	Population IR-3-20 isolate	This study
CB3008	Population IR-3-20 isolate	This study
CB3014	Population IR-3-20 isolate	This study
CB3017	Population IR-3-20 isolate	This study
CB3019	Population IR-3-20 isolate	This study
CB4000	Population IR-4-20 isolate	Harris *et al.*
CB1101	Population IR-CB1000-30 isolate	This study
CB1104	Population IR-CB1000-30 isolate	This study

CB1105	Population IR-CB1000-30 isolate	This study
CB1108	Population IR-CB1000-30 isolate	This study
CB1122	Population IR-CB1000-20 isolate	This study
CB1124	Population IR-CB1000-20 isolate	This study
CB1125	Population IR-CB1000-20 isolate	This study
CB1127	Population IR-CB1000-20 isolate	This study
CB1130	Population IR-CB1000-20 isolate	This study
EAW7704	Founder Δe14	This study
EAW138	Founder Δe14 GsiB N104	This study
EAW256	Founder Δe14 RsxB K121E	This study
EAW19001	Founder Δe14 RecA D276N	This study
EAW221	Founder Δe14 YfjK A151D	This study
EAW212	Founder Δe14 DnaB P80H	This study
EAW257	Founder Δe14 GlpD V255M	This study
EAW261	Founder Δe14 WcaK Y132C	This study
EAW227	Founder Δe14 RecA D276N DnaB P80H	This study
EAW239	Founder Δe14 RecA D276N DnaB P80H YfjK A151D	This study
EAW254	Founder Δe14 RecA D276N DnaB P80H YfjK A151D GsiB N104	This study
EAW265	Founder Δe14 RecA D276N DnaB P80H YfjK A151D GsiB N104 WcaKY132C	This study
EAW269	Founder Δe14 RecA D276N DnaB P80H YfjK A151D GsiB N104 WcaKY132C GlpD V255M	This study
EAW281	Founder Δe14 RecA D276N DnaB P80H YfjK A151D GsiB N104 WcaK Y132C GlpD V255M	This study
EAW248	Founder Δe14 Δ*wcaK*	This study
EAW246	Founder Δe14 Δ*rsxB*	This study
EAW249	Founder Δe14 Δ*yfjK*	This study
EAW246	Founder Δe14 Δ*nanE*	This study
EAW245	Founder Δe14 Δ*glpD*	This study
EAW241	Founder Δe14 Δ*gsiB*	This study

EAW10502	CB2000 wtRecA	This study
EAW220	CB2000 wtYfjK	This study
EAW213	CB2000 wtDnaB	This study
EAW233	CB2000 wtGsiB	This study
EAW234	CB2000 wtGlpD	This study
EAW235	CB2000 wtWcaK	This study
EAW237	CB2000 wtRsxB	This study
EAW238	CB2000 wtNanE	This study
EAW228	CB2000 wtRecA wtDnaB	This study
EAW240	CB2000 wtRecA wtDnaB wtYfjK	This study
EAW255	CB2000 wtRecA wtDnaB wtYfjK wtGsiB	This study
EAW266	CB2000 wtRecA wtDnaB wtYfjK wtGsiB wtWcaK	This study
EAW270	CB2000 wtRecA wtDnaB wtYfjK wtGsiB wtWcaK wtGlpD	This study

Table S2: Mutation table for isolated sequenced from IR-1-20, IR-2-20, IR-3-20, &

Gene	Pos.	Ref. Allele	CB10000	CB10012	CB10013	CB10014	CB10015	CB10024	CB10025	CB20000	CB20003	CB20004	CB20007	CB20009	CB20014	CB20016	CB30000	CB30002	CB30006	CB30008	CB30014	CB30017	CB30019	CB40000	Change	Mutation Type
thrB	2888	G									A			A											A29T	N
yaaA	6114	C	T																						G115S	N
dnaJ	14857	C	T	T																					G229G	S
insL-1	15781	-								T															V113C	N (+1)
ileS	24764	T															C		C	C			C		Y791H	N
djlA / yabP	58448	T										C													-	I
hepA	61526	A						G																	M579T	N
araB	69253	C	T																						V265M	N
thiQ	72922	T													C										L1L	S
sgrR	77213	C							T																L28L	S
leuB	81214	A								G	G		G	G	G										F248L	N
leuA	82538	C											T												R330H	N
ilvI	86320	G																	A						E230K	N
ftsW	98506	A					G																		E34G	N
ftsW	99207	A			G				G																M268V	N
ftsZ	106214	G		A																					D303N	N
lpxC	107033	C				T																			I158I	S
hofC	115115	G			A																				Q203*	N (ter)
hofB	116428	G				A	A																		T223T	S
dksA	160511	T						C																	M31V	N
sfsA	160785	C																				T			L233L	S
hrpB	163292	G												A											L395L	S
hrpB	164386	C																	T						P760L	N
fhuA	168936	C																			T				R484C	N
hemL	173832	C																						A	V350F	N
yadS	177567	A																		G					L19L	S
dgt	179544	C								G	G		G	G	G										T102S	N
ispU	195256	G								A															S117S	S
cdsA	196188	T												C											F170S	N
bamA	198849	A	G	G																					I307V	N
lpxB	203439	A																						G	E30G	N
ldcC	209946	G																						T	D89Y	N
rcsF	219698	T								C	C	C	C	C	C	C									K99E	N

IR-4-20.

metN / gmhB	222810	G								T	T	T	T	T	T	T									-	I
mltD	233097	T							C																T286A	N
yafJ	244974	G								A															Q215Q	S
proA	261858	T				C																			F377L	N
yagF	284326	C							G																G633G	S
yagR	299845	A						G																	I104T	N
yagW	305033	G				C																			Q336E	N
yagX	306927	C				T																			D543N	N
betT	330267	G							T																W526C	N
yahG / yahI	339123	C								A															-	I
yahG / yahI	339247	-								G															-	I (+1)
codA / cynR	356956	T								A															-	I
cynX	359486	C									T			T											L90L	S
cynX	360369	G											C												*384S	N (read-through)
lacZ	364174	C										A				A									P451P	S
mhpF	372177	T																						C	S10S	S
yaiO	379791	T		C																					R91R	S
yaiP	382807	G										A													A117V	N
ampH / sbmA	395819	T										C													-	I
rdgC / mak	409266	T																		G						
acpH	423996	G															A			A			A		D48D	S
tgt	426104	A															G								K247K	S
cyoA	450352	C										T													V160V	S
clpP	456127	A	G	G																					Y75C	N
clpP / clpX	456637	G															A	A	A	A	A	A	A		-	I
clpX	457803	A				G																			Y384C	N
ppiD	461176	T										G				G									V12G	N
ppiD	462852	C								A															L571M	N
cof	467098	C								T	T		T	T											L154F	N
mdlA	468837	T									C														I247T	N
acrB	481645	C																		T					A660A	S
priC	489549	A	G	G																					L162P	N
aes	498328	C															T	T	T	T	T	T	T		T289T	S
ybaS	511709	C																						T	P281L	N
ybaT	512072	G																	C						S90S	S
ybaT	512927	C																			T				A375A	S

ybbJ	513932	A																				C	L50R	N
qmcA	514715	G															A						P94S	N
ybbP	521090	A						G															Q483R	N
rhsD	525902	C													T								L1139L	S
ybbC	527126	A							G														I107V	N
gcl	533699	C															A			A	A		A186D	N
allD	545489	C								T													L32L	S
ylbE'	547836	-	G							G						G							K85E	N (+1) (pseudo)
folD	556376	G				A																	R196C	N
exoD'	565537	C						T															M15I	N (pseudo)
emrE / ybcK	568019	G	A	A																			-	I
ybcK	568141	C											T										A5V	N
tfaD'	581219	T		C																			S112P	N (pseudo)
tafX' / appY	582696	A						G															-	I
nfrA	590143	C										T											W11*	N (ter)
cusA	598463	A			G				G														E175G	N
ybdK	605968	G																				A	Y212Y	S
fepA	609783	C						A															T644T	S
fepE	617525	T										C											Y16H	N
entA	627972	T	C																				L66L	S
ybdL	633092	C	T	T																			T94I	N
ybdN	635197	C															T						G198D	N
dcuC	654488	A																G					V234A	N
ybeF	660420	C																		G			S60T	N
mrdA	665831	G										A											A536V	N
leuS	671496	T							C														K836K	S
ybeL	674532	C					T																Q97*	N (ter)
djlB	677428	G														A	A	A	A	A	A	A	C263Y	N
rihA	683008	T													C								N209D	N
gltL	686276	T														G	G	G	G	G	G	G	E231A	N
asnB	697388	T		A																			K337M	N
nagC	700277	C															T						A180T	N
nagE	703577	C								T	T	T	T	T	T	T							T136T	S
glnS	706854	G							A														A512A	S
ybfM	708603	C					T																F348F	S
ybfF	711554	C																			T		V157M	N

ybfG'	715501	G						A																A106V	N (pseudo)
kdpB	725547	G								A	A	A	A	A	A	A								T237M	N
kdpB	725668	T										C				C								T197A	N
kdpA	726295	T					C																	E553G	N
ybgH / ybgI	741971	T																	C					-	I
ybgP	748601	G								T	T		T	T	T									R109R	S
ybgQ	750279	C																T						G371R	N
mngA	767167	A						G																D653G	N
tolA / tolB	776955	G														T	T	T	T	T	T	T		-	I
ybgF	779278	T											C											V152A	N
nadA	781495	G																A						G62D	N
nadA	782331	G								A	A	A	A	A	A	A								A341T	N
modC	795796	C																			T			S6F	N
pgl	797960	A					G																	Y50C	N
ybhJ	804535	G			A				A															E603K	N
bioC	811323	T																					C	D192D	S
ybhU / uvrB	812708	A											G											-	I
ybhL / ybhM	819857	C	A																					-	I
ybhN	820957	C														T	T	T	T	T	T	T		G254G	S
ybhS	825727	A						G																V249A	N
ybhG	829128	T																	C			C		Y22C	N
rlmF	841636	C				T																		P27S	N
rybA / mntR	852295	A	G																					-	I
ybiT	856409	G						A																W407*	N (ter)
fsaA	863201	G	A																					G112S	N
moeA	865009	A							G															G192G	S
gsiB	868947	A								G	G	G	G	G	G	G								N104S	N
gsiB	869499	T														C	C	C	C	C	C	C		L288P	N
gsiB	870075	T																					C	V480A	N
gsiD	871345	C								T														P77L	N
rimO	876334	G				A	A																	P308S	N
dacC	880774	G													A									E274E	S
ybjG	882133	C																					T	G159D	N
grxA / ybjC	890053	C			T				T															-	I
ybjD	917022	G													A									G442R	N
cspD	921765	A							G															F16L	N

cydC	927416	C				T																		D334N	N
cydC	927829	C															T			T				G196E	N
ftsK	932725	T																					C	G92G	S
pflA	950207	G																					A	R32C	N
pflB	951867	T								C														E303G	N
pflB	952067	C													T									M236I	N
msbA	966949	G			A				A															G368E	N
mukB	978570	C						A																R1007S	N
ycbB	980297	C						A																Q9K	N
asnS	987540	C																		T				L222L	S
asnS / pncB	988243	C																					T	-	I
yccS	1022080	C						T																A342T	N
yccS	1022253	C						T																C284Y	N
yccS	1022496	G										A												A203V	N
yccS	1022833	C					T																	A91T	N
hyaD	1035333	C														T								P113L	N
appC	1037552	T					C																	M196T	N
etk	1041943	T						C																K496K	S
gfcC	1047851	C																					A	A21S	N
torD	1061455	T			C																			V144A	N
rutF	1069446	A																G						D43D	S
rutC	1071011	C														T	T	T	T	T	T	T		A123A	S
phoH	1084873	C															T							A219V	N
phoH	1085097	A		G																				N294D	N
ycdY	1098949	C								T	T	T	T	T	T	T								T28T	S
csgG	1100628	T							C															I93V	N
csgF	1101122	A						G																S76P	N
csgD	1102192	G														T		T	T			T		S75R	N
mdtG / lpxL	1114748	C					T																	-	I
dinI	1120539	C	T	T																				E57K	N
yceM	1126490	A						G																P153P	S
flgG	1134629	G								A	A		A	A	A									V225V	S
mfd	1173068	A																			G			G39G	S
mfd / ycfT	1173277	G														A	A	A	A	A	A	A		-	I
lolC	1175170	C																					T	P173L	N

potD	1181794	C	T	T																				V86I	N
phoQ	1188762	T																					C	K79E	N
mcrA / icdC	1210635	A											G											G	I
ycgE	1212614	C		A																				R222S	N
ymgG	1222139	T																C						N1D	N
ycgI	1223009	A											T											R74W	N
minD	1224060	T															A							K174N	N
emtA	1242766	A							T	T	T	T	T	T	T									N121Y	N
dhaK	1249603	G																					C	T152T	S
narL	1274406	A															G							F215S	N
narG	1282792	G																					A	G1235S	N
rttR,tpr / tyrV	1286465	G			T				T															-	I
purU	1287175	C																					T	D224N	N
adhE	1295192	A					G																	F717S	N
oppD	1303734	A									G													E318E	S
kch	1307113	G			A																			S393L	N
ompW	1312222	T										C												L59P	N
trpE	1320323	A						G																R215R	S
topA	1330073	G									A													M333I	N
yciX / acnA	1333742	T			C																			-	I
yciM	1338957	G																		T				A125S	N
gmr	1343113	G	A																					Q551*	N (ter)
gmr	1344567	C																T						S66N	N
rnb	1345880	T									C													N352D	N
sapA	1354456	G						T																Q226K	N
ycjO	1371867	T									C													F207L	N
ycjP	1372862	C																					A	A249D	N
mpaA	1388275	G							A	A	A	A	A	A	A									N115N	S
ycjZ	1390174	A			C																			T53P	N
mppA	1392424	C						T																Q391*	N (ter)
ynaI	1393257	T																					A	G229G	S
fnr	1396995	A														T	T	T	T	T	T	T		F185I	N
ydaM / ydaN	1405870	G																					T	-	I
ttcA	1409903	G				A																		R23C	N

recE	1413629	G				A																		D593D	S
recE	1413769	T									C													S547G	N
recE	1414611	T									C													D266G	N
ydbK	1435293	G																				A		N1171N	S
ydbK	1435547	G														T	T	T	T	T	T	T		L1087I	N
ydbK	1436124	G							T															V894V	S
ydbK	1436151	G							T															F885L	N
ydbK / ydbJ	1438958	C							A															-	I
ydbH	1442143	G		A																				D356N	N
ydbL	1444103	A				G																		Q66R	N
paaA	1452461	C																T						Q170*	N (ter)
paaK	1460550	T						C																V133V	S
ydbC	1473075	G									A													L276L	S
ydbD	1474165	G												A										R332H	N
ynbB	1476749	C							T															A166V	N
ynbC	1477794	T														C	C	C	C	C	C	C		I210I	S
hrpA	1481250	C																					T	Q55*	N (ter)
hrpA	1483786	A		G																				E900G	N
hrpA	1483817	G											T											P910P	S
ydcJ	1493504	G												A										V64I	N
rimL	1497338	A								C														E125A	N
ydcK	1498255	C								T	T		T	T	T									V72V	S
ydcM	1502187	A	G	G																				V168V	S
ydcS	1510249	T	C	C																				I190T	N
ydcU	1511931	G			C				C															G25A	N
ydcY	1515884	C																		T				T70T	S
rhsE'	1527290	G					A																	W454*	N (ter) (pseudo)
narZ	1538798	T											C											N605S	N
fdnI	1549749	A		G																				T129A	N
adhP / maeA	1551985	C																					A	-	I
maeA	1552317	G																				A		A458A	S
ddpF	1555689	A	G	G																				L124S	N
ddpB	1558739	C		T																				S71N	N
ydeN	1580128	T																			C			T140A	N

ydeP	1583041	G						A																A489A	S
ydeP	1583106	C																					T	A468T	N
ydeP	1583119	G														T								G463G	S
ydeT	1587715	A											G											V129A	N
yneO	1591581	G																A						A1509A	S
lsrK	1597223	G	A																					G336G	S
lsrB	1603853	T							C				C	C	C									V259A	N
lsrF	1604261	G									A				A									P45P	S
lsrG	1605089	C							G															R22G	N
tam	1605806	C																						A145V	N
yneH	1610738	T									A				A									N179Y	N
yneH	1610745	C			T				T															S176S	S
sad	1612428	A									T				T									Y99*	N (ter)
marA / marB	1617987	A																					G	-	I
ydeJ	1623173	C																					T	T125M	N
ydfH	1627062	A														G	G	G	G	G	G	G		*228*	S
ydfP	1637502	A	C																					V16G	N
essQ / cspB	1638996	C														A	A	A	A	A	A	A		-	I
ydfU/rem	1642390	C							T															-	I
ynfF	1659029	G			A				A															L149L	S
clcB	1663526	C	T	T																				A62V	N
clcB	1664042	G					A																	G234E	N
pntA	1675918	C									A				A									G3C	N
ydgH	1677388	A																	G					Y312C	N
ydgI	1678705	G									A													A374A	S
ydgC	1679987	G							A															T22M	N
rstB	1681401	A	G																					K165E	N
ydgA	1689037	C							A															P387T	N
malI	1696691	A					G																	L171L	S
rsxB	1704735	A								G	G	G	G	G	G	G								K121E	N
rsxD	1707299	T														C	C	C	C	C	C	C		V44A	N
tppB	1711320	A													T									K175N	N
mliC	1716204	G				T																		G71G	S
anmK	1717359	C																				T		D89N	N

anmK	1717360	T																				A		R88S	N
sodC	1722190	C																A						G163C	N
nemA / gloA	1725790	G						A																-	I
lhr	1727596	G									A													L161L	S
lhr	1731477	T										C												V1455A	N
mdtK	1742659	G								A	A	A	A	A	A	A	A							T392T	S
ydhT	1747414	T											A											K56I	N
sufS	1757709	T		C																				N279S	N
sufD	1758791	T																	C					E341G	N
ydiI	1763585	C											T											L23L	S
ydiR	1778835	C															T							I136I	S
ydiU	1787894	C								A	A	A	A	A	A	A	A							D458Y	N
pheS	1796818	G									A													P49L	N
rpmI	1797924	T	C																					T33A	N
infC	1798508	T																	C					D51G	N
arpB	1802755	C										T				T								C545C	S
yniA	1806324	A			G				G															I168V	N
yniB	1807084	G										A												I57I	S
ydjN	1809365	T																	C					G135G	S
chbC	1817975	T						C																I421V	N
chbC	1818558	T															C		C	C			C	A226A	S
astB	1825116	G			A				A															L389L	S
ynjC	1835996	T														C								L253S	N
ynjI / topB	1842992	A			T																			-	I
ydjE	1851795	T													C									Y69C	N
ydjG	1853102	T	C																					V297V	S
ydjI	1855594	G	A																					S66L	N
ydjJ	1856269	C																	T					A196T	N
ydjK	1857347	C					T																	V302I	N
ydjL	1858960	T						C																I132V	N
ydjL / yeaC	1859386	A	G	G																				-	I
yeaI	1869338	G				A																		V309V	S
yeaJ	1870557	G				A																		V164I	N
yeaM	1873502	A				G																		S32S	S

sdaA / yoaD	1896416	G	A																					-	I
yobD / yebN	1903606	C		T																				-	I
prc	1911561	C														T	T	T	T	T	T	T		R426H	N
yebT	1917640	A						G																I702V	N
edd	1931575	C																T						G351S	N
yebK	1935367	C																				T		T230I	N
ruvB	1943223	T						C																D52G	N
ruvB	1943323	C			A				A															D19Y	N
aspS	1947323	T					C																	A407A	S
aspS	1947485	A														T								V353V	S
aspS / yecD	1948549	G					A																	-	I
yecD	1949188	T						C																D110D	S
torZ	1953566	T					C																	E488G	N
torY / cutC	1956456	T													C									-	I
argS	1959731	A					G																	N548S	N
cheR	1966843	T														C	C	C	C	C	C	C		Q181Q	S
tar	1969075	A														G	G	G	G	G	G	G		D546D	S
motA	1975011	G																			C			G50G	S
otsA	1978643	G												T										P331T	N
yecR / ftnA	1986630	C							A															-	I
fliC / fliD	2001890	A									T				T									-	I
fliN	2019120	C																					A	D2E	N
yodD	2023202	A																					G	I64V	N
yedQ	2025877	G					A																	S54S	S
yedA	2028240	C			T																			A225A	S
yedV	2035894	T									C													N94D	N
yeeJ	2047156	T													C									S1398P	N
yeeJ	2047545	A																		C				K1527N	N
yeeL	2050572	A																	G					Y260H	N
shiA	2051774	T									C													Y35Y	S
cbl	2058107	G				A																		Q277*	N (ter)
yoeH'	2066390	A	G																					C17C	S (pseudo)
yeeA / sbmC	2078674	C																T						-	I
yeeE	2083022	T																			G			K175N	N

yeeZ	2086775	G	A																					T125T	S
hisC	2091011	C							T															A196V	N
hisC	2091123	C				T																		N233N	S
ugd	2097458	G	A																					H59H	S
wcaM	2113451	T																					C	N156S	N
wcaK	2116031	T								C	C		C	C	C									Y132C	N
wcaC	2129153	A														G	G	G	G	G	G	G		S313S	S
wcaA	2130664	T																A						K252I	N
asmA	2138470	G			A				A															G388G	S
yegE	2143609	G									T													V773F	N
yegL	2150962	C						T																G63D	N
mdtB	2154011	A																C						D241A	N
mdtC	2157538	A							G															M376V	N
yegS	2166910	G													A									A58T	N
gatA	2173035	C																				T		V12I	N
fbaB	2176080	A								G														Y168Y	S
yegW	2180453	A																					G	V116V	S
yohN/yehA	2185344	G				A																		-	I
yehH'	2197475	C						A																P993T	N (pseudo)
yehI	2199704	C							T															A467A	S
yehQ'	2207144	T						C																L7S	N (pseudo)
yehQ'	2207944	G	A																					D274N	N (pseudo)
yehU	2212073	G															T			T	T			R197R	S
yehU	2212593	C									T													S24N	N
yehX	2215324	A				G																		S35P	N
yehY	2215996	C					T																	G194D	N
bglX	2219865	C										T												E48E	S
dld	2221816	C								A	A	A	A	A	A	A								A536E	N
yohF	2225187	C								T	T	T	T	T	T	T								D35N	N
mdtQ'	2226741	G															A							P40S	N (pseudo)
cdd	2230578	C		T																				P237L	N
yeiT	2232132	T			C																			C25C	S
yeiG	2241988	C				T																		H18H	S
fruB	2260443	GTT	---																					Q358-	N (-3)

lpxT	2267554	C	T																					L226L	S
yejB	2272868	A																	G					D222G	N
yejF	2274392	T								C	C	C	C	C	C	C								V23A	N
yejH	2279168	G	A																					G171D	N
ccmA	2295436	G				A																		N76N	S
napA	2299071	A			G																			F568L	N
napA	2299616	A						G																V386A	N
napA	2300747	T																C						K9R	N
eco/mqo	2302576	C								G														-	I
mqo	2303175	C			T																			L533L	S
yojI	2305897	G			A																			I246I	S
rcsD	2313084	G						A																V524V	S
yfaS	2327937	C																					G	L1495L	S
yfaS	2332331	C																					T	G31R	N
yfaT	2332910	G					A																	V23V	S
yfaL / nrdA	2342577	T							C															-	I
yfaU	2356760	G							A															A35A	S
yfaV	2357420	A														C								V264G	N
arnF	2371104	G						A																W63*	N (ter)
menF	2378171	G																				A		H164H	S
nuoL	2392821	C																					G	G82A	N
nuoH	2394562	G				C	C																	G300G	S
nuoG	2396587	C							T															R533H	N
nuoC	2401084	G	T																					H261N	N
nuoC / nuoB	2401962	A							G															-	I
yfbQ	2405997	T										C												S138P	N
yfbT	2409755	T			C																			I118M	N
folX	2419512	A															G							N55D	N
truA	2433126	A												G										V177A	N
pdxB	2434910	C	T																					D321N	N
mnmC	2441067	T				C																		Y427H	N
yfcP	2448580	G		A																				Y10Y	S
yfcR	2449550	G																		A				C18C	S
fadI	2457472	G		A																				I339I	S

fadL	2460009	A													G								N227D	N
fadL	2460103	A														G							D258G	N
fadL / yfdF	2460746	A									G												-	I
yfdF	2461525	G				A	A																G163G	S
gtrS	2467583	C								T													S143L	N
gtrS / tfaS'	2468587	C													T								-	I
yfdL	2469915	C																			T		G56D	N
yfdR	2473119	C															T		T	T			A37A	S
oxc	2489734	G							A														P79L	N
lpxP / yfdZ	2494971	T														C		C	C		C		-	I
fryA	2501830	A																				G	S225S	S
ypdE	2502874	G				A																	G231G	S
glk	2506699	G														A		A	A		A		L249L	S
xapB	2521395	A		G																			F204L	N
xapA	2522710	C																	T				G63D	N
ligA	2527644	G									A				A								N184N	S
pdxK / yfeK	2535315	G	T	T																			-	I
yfeW	2546651	A					G																L175L	S
ypeA / amiA	2550325	A															G						-	I
yffL	2558646	G											A										W122*	N (ter)
eutD	2570543	G				A																	R328C	N
maeB	2575790	T																		C			N203D	N
aegA	2582485	C											T										V354M	N
ypfM / yffB	2589229	A																	G				-	I
hyfB	2600797	G	T	T																			G319C	N
yfgH	2627996	C							T														D60D	S
rlmN	2641504	A										G											S267P	N
yfhM	2645502	G										A											A1602V	N
yfhM	2649164	T																				C	P381P	S
yphB	2672511	C															T						R66H	N
yphC	2673580	G								A			A										N67N	S
yphD	2674644	C																	T				W67*	N (ter)
yphF	2676454	C														T			T		T		T311T	S
yphG	2678056	G	A																				A903A	S

glyA	2683151	G											A										L126L	S
purL	2690904	C	T																				G887S	N
tadA	2695804	G							A														R25W	N
nadB	2709617	G									A												L391L	S
yfiQ	2718258	A																C			C		K94T	N
yfiQ	2720505	A														G			G	G			N843S	N
clpB	2730460	G								A													A578V	N
yfiR	2740344	G						A															A149T	N
rpsP	2744004	A														G			G	G			S67S	S
rpsP / ffh	2744453	C					T																-	I
yfjD	2746811	A																		G			T5A	N
yfjI	2757855	T				C																	F282F	S
yfjI	2758272	T		C																			N421N	S
yfjK	2759609	G																				A	H651Y	N
yfjK	2760683	T						C															K293E	N
yfjK	2760809	G												A									P251S	N
yfjK	2761108	G								T	T	T	T	T	T	T							A151D	N
ypjA	2780582	G	A																				T55I	N
ygaQ'	2785968	C		T																			A516V	N (pseudo)
ygaU	2794524	G																				A	A94A	S
nrdE	2800214	G		T																			R281L	N
proX	2805751	G			A																		E199K	N
ygaH / mprA	2808766	G	A	A																			-	I
serV / csrA	2816671	C					T																-	I
recA	2820924	C						A															A289S	N
recA	2820962	T			G	G	G		G														D276A	N
recA	2820963	C								T	T	T	T	T	T	T						T	D276N	N
mltB	2822744	G												A									G284G	S
hypF	2833488	A	G	G																			W653R	N
ascB	2839926	T									A												F304L	N
hycC	2846936	G	A																				Q109*	N (ter)
fhlA	2852903	C														T							R181W	N
fhlA	2853352	T						C															N330N	S
ygbJ	2859530	C	T	T																			R26C	N

nlpD	2866080	G	A																					V231V	S
truD	2868586	G																	A					R246R	S
ispF	2869428	A	G																					D124D	S
ygcB	2885184	T					C																	K19E	N
cysI	2888089	G										A												V10V	S
cysJ	2888465	A			C																			F485V	N
fucO	2930693	A																		C				F115V	N
fucA / fucP	2932140	C	T	T																				-	I
fucK	2936029	A																			G			G189G	S
fucK	2936460	T								C														M333T	N
fucU	2937272	T														C		C	C			C		V120V	S
ygdL	2943498	A																					G	S122P	N
recB	2953341	T	C																					I228V	N
lysR	2977775	C	T	T																				P244S	N
ygeK'	2992757	A	G																					S119P	N (pseudo)
ygfK	3014204	T																			C			G40G	S
ygfU	3029566	A	G	G																				K59E	N
ygfU	3030225	T				C																		G278G	S
xerD	3037674	C														T	T	T	T	T	T	T		R30H	N
serA	3055925	C																					T	A169T	N
serA	3056348	C		T																				A28T	N
scpA	3060807	G					A																	A645T	N
cmtA	3075666	G																					A	H405Y	N
tktA	3077762	C	T	T																				M631I	N
metK / galP	3086156	G			A				A															-	I
yqgF	3091589	T		C																				I22T	N
mutY	3101729	C																					T	T231I	N
yghD	3108747	C																			T			V133V	S
yghJ	3114599	C															A							E845*	N (ter)
glcD	3125648	C										T												A131A	S
glcD / glcC	3126216	C																					T	-	I
yghT	3132728	G		A																				Q191Q	S
hybC	3139587	G																					A	L474L	S
mqsR	3166480	C		T																				L28L	S

ygiZ / mdaB	3170242	A								G														-	I
ribB / yqiC	3182554	A		G																				-	I
dnaG	3209547	G								A	A		A	A	A									R139H	N
rpoD	3212119	A															G							T350A	N
rpoD	3212771	A								G														N567S	N
ygjR	3235351	A																	G			G		I6V	N
alx	3237555	G									A													Q317Q	S
yqjG	3249966	G							T															G306G	S
yhaH / yhaI	3250857	A	G	G																				-	I
yhaJ	3251758	G											A											S159L	N
tdcE	3260065	A													G									S125P	N
tdcA / tdcR	3265098	G						T																-	I
tdcR	3265424	T						C																I7T	N
yhaV / agaR	3275837	G								A			A											-	I
agaA' / agaS	3279715	T					C																	-	I
yraM	3292561	G						A																A379A	S
diaA	3294171	T														A	A	A	A	A	A	A		L113*	N (ter)
nlpI	3306455	A				G	G																	P163P	S
glmM	3321508	G	T																					R194R	S
ispB	3331922	A						G																N63S	N
yrbC	3335355	T				A																		T186S	N
rpoN	3343785	A		C																				E348D	N
yhcC / gltB	3352609	A																					G	-	I
nanE	3368674	C								T	T	T	T	T	T	T								A128T	N
nanT	3369380	A														G	G	G	G	G	G	G		F415S	N
aaeR	3388466	A																					G	K308E	N
yhdE	3396350	A														G								L16L	S
mreB / csrD	3399224	T								C														-	I
csrD	3400412	G									A													H314Y	N
prmA	3407852	G																A						G253D	N
yhdZ / yrdA	3424869	G														A								-	I
yrdA	3427431	C														T								T129T	S
aroE	3428194	G									A													L222L	S
yhdN	3437459	C		T																				A24T	N

rpsN	3444820	C										T												A28A	S
rplB	3448660	G									A													H242Y	N
gspC	3454125	A		G																				I175V	N
fusA	3471524	T										C												T4A	N
nirB	3493496	C															A			A	A			H487Q	N
cysG	3497167	A				G																		I439V	N
yhfX	3508808	C			A				A															G63V	N
hofP	3518648	C														T	T	T	T	T	T	T		D144N	N
hofO	3519315	G							A															Q52*	N (ter)
yrfF	3524884	A													G									T131A	N
yrfF	3526497	C				T																		A668A	S
ompR	3533999	A										G												V202A	N
ompR / greB	3534616	T														C								-	I
feoB	3538580	A								G														Q50R	N
yhgA / bioH	3542070	T										C												-	I
glpD	3560801	G							A	A	A	A	A	A	A									V255M	N
glpD / unnamed pseudogene	3561617	A															G			G	G			G	I
glgP	3563053	G														T	T	T	T	T	T	T		Q517K	N
glgP	3563638	G														A	A	A	A	A	A	A		H322Y	N
glgB	3569505	T			C																			K673R	N
yhhY / yhhZ	3579865	C							G															-	I
yhhZ	3580097	C																T						P70L	N
yhhN	3604163	G					C																	W129C	N
yhhH	3621491	G				A																		V23I	N
yhiL'	3631270	C														T								V403V	S (pseudo)
pitA	3636547	G		A																				E294K	N
yhiR	3643629	G						A																L73L	S
yhiF/yhiD	3653245	C				A																		-	I
hdeD	3655501	A																G						T161A	N
mdtF	3659044	T						C																V202A	N
mdtF	3660175	C	T																					A579V	N
mdtF	3660805	G																					A	R789H	N
kdgK	3678292	C														A								A356E	N

yhjJ / dctA	3680038	C									T													-	I
bcsC	3686233	C		T																				A321T	N
bcsB	3688983	G		T																				Q549K	N
dppF	3700121	G														A								Y256Y	S
glyS	3720676	G																		T				A581E	N
xylA	3727614	C										G												S391T	N
xylA	3728585	C										A				A								P67P	S
xylG	3731346	C									T													Q374*	N (ter)
xylH	3732701	A																	C					N319T	N
selB	3756684	G						A																P400S	N
selA	3758316	C			T				T															P318P	S
rhsA	3763699	A											C	C										E1164A	N
rhsA	3763700	A											G	G										E1164E	S
yibH	3768842	C		G																				Q186H	N
yibI	3769431	C								T			T											D112N	N
yibL/lldP	3775408	C					T																	-	I
yibN	3783034	T													C									Q1R	N
kbl	3789997	A						G																L192P	N
rfaD	3792548	G	A																					S179N	N
rfaC	3794258	A							G															N85S	N
rfaI	3800614	A								G														V155V	S
rfaS	3802337	A																	T					I267K	N
dfp	3811194	C												T										L146L	S
ligB	3818168	T																				G		K341N	N
gltS	3825922	A			G	G	G		G															V255A	N
yicH	3829137	T						A																W219R	N
yicO	3840679	G														A		A	A			A		N403N	S
ilvB	3849561	A									G													L415P	N
dgoR / yidX	3873322	A										G												-	I
rnpA	3882847	G														A								R110H	N
tnaC / tnaA	3886749	T																C						-	I
purP	3894550	G								A														A27V	N
bglH	3898645	C				T	T																	Q532Q	S
bglH	3899844	C						T																G133S	N

rsmG	3921572	G																		A				R43R	S
mnmG	3922616	G																		A				F346F	S
ravA	3928950	G							A															A55V	N
rbsA	3932767	G																					A	D322N	N
rbsC	3933829	G						T																P172P	S
hdfR	3945354	C																			A			E212*	N (ter)
ilvA	3953940	T																C						V195A	N
yifN	3958073	C												A										R136R	S
rffE	3968890	G											T											V244V	S
rffA	3973721	G									A													A184T	N
wzxE	3974865	C	T																					P188S	N
wzyE	3977007	C																T						V127V	S
tatC	4021383	C																T						P208S	N
yihT	4070533	A																	G					F47L	N
yihU	4071272	T												C										D107G	N
fdhE	4078346	G	A	A																				F301F	S
yiiG	4085221	C					T																	P65L	N
frvR	4086862	A				G																		C338C	S
frvR	4087176	C		T																				G234S	N
rhaA	4093913	A					T																	D30E	N
rhaT	4097654	A																G						Y298H	N
yiiM	4100999	G														A	A	A	A	A	A	A		G61E	N
cpxA	4102033	C							A															A321A	S
pfkA / sbp	4106697	C												T										-	I
priA	4123174	C														T	T	T	T	T	T	T		V553I	N
metL	4129990	C							T	T		T	T											F710F	S
metL	4130134	T						G																H758Q	N
yijF	4135247	T	C																					Y144C	N
gldA	4136015	A							G	G		G	G	G										V347V	S
pflC	4144596	G	C	C																				E105Q	N
ppc	4150978	G														A								G47G	S
fabR	4159346	G				A																		G66D	N
murI / murB	4166244	T														C								-	I
murI / murB	4166503	T														C								-	I

rplK	4176590	G								A	A	A	A	A	A	A								A40T	N
rpoC	4187534	G								A														D1387N	N
thiS	4190847	A																	G					F4S	N
thiC	4192664	A								G														W486R	N
metA	4212676	A						G																K124R	N
aceB	4213910	G																		A				R136H	N
aceA	4216423	A															G			G	G			E430G	N
iclR	4221400	G								C														I83M	N
yjbI'	4249616	G								A														L296L	S (pseudo)
ubiA/plsB	4252034	G						A																-	I
dinF	4256277	C					A																	L170L	S
yjbM'	4258750	T								C														I135T	N (pseudo)
dnaB	4262560	T						C																L74S	N
dnaB	4262578	C								A	A	A	A	A	A	A								P80H	N
dnaB	4262935	C																					A	P199Q	N
alr	4264102	C	T																					H99Y	N
uvrA	4270674	G										A												A406A	S
phnI	4318623	G																					A	R21*	N (ter)
phnH	4319089	G															A							P60L	N
adiY	4335298	A												C										S218A	N
melB / yjdF	4342848	A															G							-	I
dcuR	4347852	G														A	A	A	A	A	A	A		P68L	N
aspA / fxsA	4366510	G					A																	-	I
yjeK	4373038	G											A											R214C	N
yjeP	4386471	T								C	C	C	C	C	C	C								Q307R	N
mutL	4395562	A					T																	D42V	N
yjfJ	4409047	T								C	C	C	C	C	C	C								S157P	N
ulaC	4419905	G															A	A	A	A	A	A	A	V58I	N
ulaD	4420633	A																	G					Q141R	N
cycA	4428586	A										G												T233A	N
ytfE / ytfF	4430033	A								G														-	I
ytfN	4443548	A																						T471P	N
chpB	4446834	T								C	C		C	C										F39F	S
ytfQ	4448890	G	A																					L301L	S

yjgA	4455866	C			T																			W7*	N (ter)
nrdD	4458943	C		T																				V580M	N
treR / mgtA	4465450	T								C														-	I
pyrB	4470036	G					T																	A127D	N
yjgL	4474024	A	G																					N188D	N
yjgL	4475030	A			G				G															D523G	N
argI / rraB	4476359	C	T	T																				-	I
valS	4480618	A											C											S414A	N
lptF	4484936	G															C							A231A	S
yjgB	4493968	T																C						I88V	N
yjgB	4494219	T		C																				K4R	N
yjhV	4508105	G															T							P120P	S
fecE	4509041	G								A														A146V	N
sgcA	4525857	C														T								W48*	N (ter)
yjhR'	4534044	G														A								R335H	N (pseudo)
nanM	4536085	G																				A		T234M	N
fimB / fimE	4539638	C										T												-	I
fimE / fimA	4540704	C				T																		-	I
uxuB	4552285	C				T	T																	A453A	S
kptA	4559160	T	C																					F69L	N
yjiN	4564436	C					T																	M277I	N
mcrB	4576070	T			C																			T430A	N
mrr	4585486	A			G																			E171G	N
yjiY	4588962	C													T									G113D	N
yjiZ	4592346	C					T																	V133M	N
dnaT	4599105	G			A	A	A		A															R145C	N
holD	4605973	A																				G		T49A	N
prfC	4608546	C																					T	H369H	S
deoC / deoA	4616138	A		G																				-	I
deoA	4617315	A														T								D354V	N
lplA	4622080	C	T																					E20K	N
slt	4630017	G														A	A	A	A	A	A	A		W429*	N (ter)
trpR	4630885	C					T																	H34Y	N
ytjC	4632008	G													A									A62A	S

Table S3: Mutation table for isolates sequenced from IR-CB1000-20 and IR-CB1000-30.

Gene	Position	Ref. Allele	CB1000	CB1101	CB1104	CB1105	CB1108	CB1122	CB1124	CB1125	CB1127	CB1130	Change	Mutation Type
yaaX	5431	T		C	C	C	C	C	C	C	C	C	N65N	S
yaaA	6114	C	T	T	T	T	T	T	T	T	T	T	G115S	N
yaaA_yaaJ	6526	T						C	C				-	I
dnaJ	14857	C	T	T	T	T	T	T	T	T	T	T	G229G	S
dapB	28901	G		A	A	A	A	A	A	A	A	A	E175E	S
caiT / fixA	42007	T								C			-	I
fixA	42802	G		-	-	-	-	-	-	-	-	-	V133W	N (-1)
hepA	61345	G									A		R639R	S
araB	69253	C	T	T	T	T	T	T	T	T	T	T	V265M	N
fruR	88978	T								C			G316G	S
mraZ	89791	G								A	A		W52*	N (ter)
murC	101789	A		T	T	T	T	T	T	T	T	T	K341I	N
speD	135052	C								T	T		M176I	N
htrE	152928	A		T	T	T	T	T	T		T	T	S832R	N
bamA	198849	A	G	G	G	G	G	G	G	G	G	G	I307V	N
dnaE	207111	C								T	T		S661S	S
gloB	234092	C						T	T				E230K	N
fadE	242645	T									C		E219G	N
phoE	258722	A									G		D200D	S
insI-1 / eyeA	271456	C								T	T		-	I
afuC	277291	T		C	C	C	C	C	C	C	C	C	T245A	N
ykgR	312393	T						C					Y24C	N
ykgF	322026	C		T	T	T	T					T	V154V	S
ykgF	322888	A									G		I442V	N
yahE	335172	G							A				A7A	S
yahI	339982	G						A	A				P197P	S
yahJ	340743	C		T	T	T	T					T	P131L	N
yahK	342668	A						G	G				G186G	S
yahK	343102	T									C		L331P	N
yahM	344840	G		A	A	A	A					A	M70I	N
prpC / prpD	350406	A						G	G				-	I
prpE	353804	G									A		A624A	S
mhpR / mhpA	367764	T						C	C				-	I
yaiL / frmB	376598	T						C					-	I
yaiL / frmB	376674	T							A				-	I

tauA	384795	A		T	T	T	T					T	I113F	N
brnQ	418956	T		A	A	A	A	A	A	A	A	A	F47I	N
yajG	453141	A		G	G	G	G			G	G	G	L83P	N
clpP	456127	A	G	C	G	G	G	G	G	G	G	G	Y75C	N
clpX	456926	T						A					Y92N	N
clpX lon	458060	C						T	T				-	I
clpX / lon	458070	T		C	C	C	C					C	-	I
ybaV	463287	G			T	T	T					T	A42S	N
mdlB	470032	A							C				N57T	N
acrB	483381	C						T	T				D82N	N
priC	489549	A	G	G	G	G	G	G	G	G	G	G	L162P	N
htpG	495728	A							G				D461G	N
aes	498941	G						A	A				T85M	N
aes / gsk	499269	T		C	C	C	C					C	-	I
fsr	503225	C							T				L231L	S
ybbP	520367	T						C	C				F242S	N
ybbP / rhsD	522316	G						A	A				-	I
glxK	542200	C			T	T	T					T	F362F	S
ylbE'	547836	-	G	G	G	G	G	G	G	G	G	G	K85E	N (+1) (pseudo)
purE / lpxH	552356	G								T			-	I
cysS	554871	G		A	A	A	A	A	A	A	A	A	M345I	N
peaD / insE-3	565952	A		G	G	G	G					G	-	I
emrE / ybcK	568019	G	A	A	A	A	A	A	A	A	A	A	-	I
arrD	577224	G								A	A		G129D	N
appY / ompT	583783	G								A	A		-	I
nfrA	588471	T		C	C	C	C					C	Q568Q	S
nfrA	589125	A						G	G				R350R	S
ybdG	603246	C			T	T	T	T	T	T	T	T	G213D	N
entA	627972	T	C	C	C	C	C	C	C	C	C	C	L66L	S
ybdL	633092	C	T	T	T	T	T	T	T	T	T	T	T94I	N
dsbG	637135	T						C	C				Y220C	N
dcuC / pagP	655683	A		G	G	G	G					G	-	I
ybeQ	675199	G		A	A	A	A	A	A	A	A	A	A190V	N
gltJ / gltI	685905	G						A	A				-	I
miaB	693577	T							C				Q200R	N
nagA	701027	G							A				G315G	S
ybfO	734024	C									T		R222R	S
ybfL	736703	A								C	C		E218A	N
cydB	772443	A			G	G	G	G	G	G	G	G	N59S	N
tolA	775813	G						A	A				M82I	N

modA	794613	G								A	A		R100H	N
modB	795583	A						G					T166A	N
pgl	797875	G								A	A		E22K	N
ybhI	801474	A		G	G	G	G					G	N121S	N
ybhI	801989	G								A			D293N	N
ybhL / ybhM	819857	C	A	A	A	A	A	A	A	A	A	A	-	I
ybhP	823504	G		A	A	A	A			A	A	A	L72F	N
ybhR	824869	T						C	C				N154D	N
ybhS	825931	A						G					F181S	N
ybhG	829017	A						C					V59G	N
rybA / mntR	852295	A	G	G	G	G	G	G	G	G	G	G	-	I
fsaA	863201	G	A	A	A	A	A	A	A	A	A	A	G112S	N
fsaA	863238	C						T	T				A124V	N
moeA / iaaA	865779	T						C	C				-	I
bssR / yliI	877896	T								C	C		-	I
yliI	879024	T							C				Y353H	N
ybjN / potF	892998	G								A	A		-	I
artQ	901422	C								T	T		A17T	N
infA / aat	925698	G								A	A		-	I
rarA	937963	T		C	C	C	C					C	D248D	S
rarA	938325	T		C	C	C	C	C	C	C	C	C	V369A	N
serC	957238	C						T	T				A120A	S
aroA	958459	A								G	G		Y141C	N
ycbJ	971791	C		T	T	T	T	T	T	T	T	T	Q272*	N (ter)
ycbB / ycbK	982177	T		C	C	C	C					C	-	I
ompF	986019	C		T	T	T	T	T	T	T	T	T	A62T	N
ompA	1018621	G							A				S218F	N
appC	1038414	G								A			L483L	S
etp	1043510	A		G	G	G	G					G	H129H	S
gfcD	1046755	C						T	T				L137L	S
gfcA	1048764	C								T	T		V67V	S
gnsA	1051346	G		A	A	A	A					A	A18A	S
yccM	1052145	G								T	T		P146P	S
rutE	1069830	C								T			A116T	N
putA	1075818	T								C	C		N762S	N
pgaC	1085789	T							C				Q426Q	S
pgaA	1090331	G		A	A	A	A					A	S393S	S
ghrA	1097551	G		A	A	A	A	A	A	A	A	A	G147D	N
opgC / opgG	1108332	G							A				-	I
dinI	1120539	C	T	T	T	T	T	T	T	T	T	T	E57K	N
yceM	1126731	A									G		I234V	N

flgA	1129853	T		A	A	A	A			A	A	A	A77A	S
flgJ	1136785	C		G	G	G	G	G	G	G	G	G	S63R	N
flgK	1138428	T									C		N275N	S
rne	1141413	A						G	G				N725N	S
fhuE	1160480	C		A	A	A						A	G98C	N
lolC	1174891	C						A					P80Q	N
lolC	1175091	G						A	A				E147K	N
lolC	1175211	G						A	A				G187S	N
lolC	1175268	G						A	A				D206N	N
lolC	1175427	G						A	A				V259I	N
potD	1181794	C	T	T	T	T	T	T	T	T	T	T	V86I	N
ycgX	1212018	T								C	C		T104A	N
ycgE / ycgF	1213439	C								T	T		-	I
ycgF	1214109	T						C	C				N196S	N
ymgC	1216056	A		G	G	G	G					G	Y28C	N
ycgG	1217221	T									C		I223T	N
ycgI / minE	1223336	G							A				-	I
ycgN / hlyE	1228563	C						T	T				-	I
ycgB	1234936	A		G	G	G	G					G	L509P	N
ycgV / ychF	1255943	A		G	G	G	G					G	-	I
prs	1260916	G						A	A				A60A	S
kdsA	1267611	A		G	G	G	G	G	G	G	G	G	E74G	N
rdlB / ldrC	1269234	T		G	G	G	G					G	-	I
narG	1279575	G								A	A		W162*	N (ter)
purU / ychJ	1287855	A						G	G				-	I
adhE	1296044	C								T	T		G433D	N
cls	1305269	T							G				K466N	N
yciB	1310795	G								A			S39S	S
yciC	1311300	A								G			F129L	N
trpB	1316392	G			A	A	A					A	Y15Y	S
yciO	1322315	T								C	C		F64S	N
yciQ	1323119	G		A	A	A	A	A	A	A	A	A	G116D	N
btuR	1326147	G						A	A				P78S	N
yciK	1326571	A								G	G		V188A	N
acnA	1335842	G						A	A				R662Q	N
gmr	1343113	G	A	A	A	A	A	A	A	A	A	A	Q551*	N (ter)
fabI	1348936	T		C	C	C	C	C	C	C	C	C	K42R	N
ycjG	1387357	A		G	G	G	G					G	N134S	N
mpaA	1388363	C		T	T	T	T	T	T	T	T	T	G86E	N
ycjY	1389842	A								G	G		N11N	S
fnr	1397081	A		-	-	-	-	-	-	-	-	-	M156*	N (-1) (ter)

ydaQ	1411472	G		A	A	A	A					A	A1V	N
ydaE	1415908	C								T	T		C41Y	N
ydaT	1418795	G								A	A		E29K	N
ydbA	1467480	A									G		K1354K	S
ydbD	1474978	T						C	C				L603S	N
ynbB	1476931	A		G	G	G	G					G	T227A	N
hrpA	1482201	C		T	T	T	T					T	R372C	N
hrpA	1484619	G		A	A	A	A					A	A1178T	N
trg	1490862	T		C	C	C	C	C	C	C	C	C	D122D	S
ydcM	1502187	A	G	G	G	G	G	G	G	G	G	G	V168V	S
ydcS	1510249	T	C	C	C	C	C	C	C	C	C	C	I190T	N
yncB	1517552	C		A	A	A	A			A	A	A	Q167K	N
mcbR	1518675	A								G	G		Q129Q	S
yncE	1521646	G						T					G105C	N
ansP	1523804	A		G	G	G	G					G	Y66Y	S
ansP	1523805	T								C	C		Y66C	N
yncH	1525020	T		C	C	C	C					C	I18I	S
narY	1535529	G						T	T				A449D	N
fdnG	1548061	G		A	A	A	A	A	A	A	A	A	A878A	S
maeA	1553133	C			T							T	P186P	S
maeA / sra	1553774	A		G	G	G	G					G	-	I
ddpF	1555689	A	G	G	G	G	G	G	G	G	G	G	L124S	N
dos	1562498	G		A	A	A	A			A	A	A	S419S	S
gadC	1568506	G								A			T2I	N
pqqL	1571072	A		G	G	G	G					G	L718L	S
yddB	1574924	G		A	A	A	A					A	A239A	S
ydeP	1583154	T						C	C				T452A	N
yneO	1591710	A		G	G	G	G			G	G	G	H1466H	S
lsrK	1597223	G	A	A	A	A	A	A	A	A	A	A	G336G	S
lsrR	1599003	C								A	A		R87L	N
lsrC	1601858	G								T	T		L271L	S
tam	1605806	C		T	T	T	T					T	A145V	N
yneK	1614267	C								T	T		P160S	N
eamA	1618518	A		G	G	G	G	G	G	G	G	G	L214S	N
eamA	1618939	A		G	G	G	G	G	G	G	G	G	F74L	N
ydfH	1627048	T									C		F224L	N
ydfI	1628382	G										C	P185A	N
ydfP	1637502	A	C	C	C	C	C	C	C	C	C	C	V16G	N
ydfJ	1637716	C									T		V4382I	N
ydfJ	1640427	G						A	A				P3478L	N
dmsD / clcB	1663308	T		C	C	C	C					C	-	I

clcB	1663526	C	T	T	T	T	T	T	T	T	T	T	A62V	N
clcB	1664519	G		A	A	A	A	A	A	A	A	A	C393Y	N
rstB	1681401	A	G	G	G	G	G	G	G	G	G	G	K165E	N
uidA / uidR	1694125	G		A	A	A	A	A	A	A	A	A	-	I
uidA / uidR	1694384	-						G	G				-	I (+1)
rsxC	1705974	T		C	C	C	C					C	F343F	S
rsxC	1706686	C							G				Q581E	N
pdxH	1715631	G						A					P133L	N
nemA	1725699	G						T	T				R338R	S
ydhR	1744879	G								A			G51G	S
ydhT	1746877	T						C	C				H235R	N
ydhU	1748181	C								T			E63E	S
ydhV	1751021	A						G	G				N277N	S
sufS	1757709	T								N279S			N279S	N
ydiJ	1765046	A									G		V554A	N
ydiK	1767281	C						T	T				R61C	N
rpmI	1797924	T	C	C	C	C	C	C	C	C	C	C	T33A	N
astC	1829386	T									C		Q206Q	S
ydjZ	1833268	A		G	G	G	G	G	G	G	G	G	K145R	N
ynjB	1834303	T		C	C	C	C					C	R68R	S
ydjF	1852539	C								T	T		D113N	N
ydjG	1853102	T	C	C	C	C	C	C	C	C	C	C	V297V	S
ydjI	1855594	G	A	A	A	A	A	A	A	A	A	A	S66L	N
ydjL / yeaC	1859386	A	G	G	G	G	G	G	G	G	G	G	-	I
yeaD	1862148	T						C	C				L91P	N
yeaH	1867072	C		T	T	T	T	T	T	T	T	T	Q31*	N (ter)
yeaN	1874491	G						T	T				L264L	N
yeaU	1880972	C									T		A345V	N
yeaX	1884351	G		A	A	A	A	A	A	A	A	A	Q160Q	S
sdaA / yoaD	1896416	G	A	A	A	A	A	A	A	A	A	A	-	I
rlmA	1904917	T						C	C				Q55Q	S
yebQ	1909237	G		A	A	A	A			A	A	A	W312*	N (ter)
yebT	1916832	T						C	C				F432F	S
zwf yebK	1934483	G							A				-	I
yebK	1935271	G									A		R198H	N
lpxM	1937966	C		T	T	T	T					T	A83A	S
argS	1959762	G								A	A		A558A	S
tar	1969081	T		C	C	C	C	C	C	C	C	C	E544E	S
yecS	1996485	G								A	A		L6L	S
fliC	2000909	T						C					N240S	N
fliD	2002600	G								A	A		E234E	S

fliM	2018391	G								A			R93H	N
yedI	2026476	A		G	G	G	G					G	H304H	S
yedI	2026827	G		A	A	A	A					A	G187G	S
yedZ	2039113	A		G	G	G	G	G	G	G	G	G	K202E	N
yeeJ	2046051	A						C					A1029A	S
yoeH'	2066390	A	G	G	G	G	G	G	G	G	G	G	C17C	S (pseudo)
yoeA'	2066939	G								A			G93D	N
yeeR	2073639	C								T	T		F278F	S
yeeA	2078507	T						C					I36V	N
yeeZ	2086775	G	A	A	A	A	A	A	A	A	A	A	T125T	S
hisD	2089389	T							G				V89G	N
cld	2096199	C								T	T		A42T	N
ugd	2097458	G	A	A	A	A	A	A	A	A	A	A	H59H	S
wbbL'	2101073	C						T	T				S113N	N
wbbK	2102418	G		A	A	A	A	A	A	A	A	A	A38V	N
rfbX	2107420	G		A	A	A	A					A	F62F	S
rfbD	2109667	C		T	T	T	T			T	T	T	G111S	N
wcaL	2114446	C		T	T	T	T					T	G235S	N
cpsB	2121505	T		C	C	C	C	C	C	C	C	C	D346G	N
insF-5 / gatD	2169755	C		G	G	G	G					G	-	I
yegT	2177986	T						C	C				W381R	N
yegT	2178114	T		C	C	C	C					C	V423V	S
yegV	2179532	A								G	G		T138A	N
yehE	2190680	G						A	A				Q46*	N (ter)
yehQ'	2207944	G	A	A	A	A	A	A	A	A	A	A	D274N	N (pseudo)
dld	2221889	C		T	T	T	T			T	T	T	T560T	S
yohK	2229146	G								A	A		V35I	N
fruB	2260443	GTT	---	---	---	---	---	---	---	---	---	---	Q358-	N (-3)
fruB	2261227	G						A					I96I	S
yeiQ	2264628	C						T	T				A120V	N
yeiQ	2265680	C								T	T		Q471*	N (ter)
yeiQ / yeiR	2265742	G								A	A		-	I
lpxT	2267554	C	T	T	T	T	T	T	T	T	T	T	L226L	S
yejB	2272868	A						0					D222G	N
bcr	2276818	A						G					M321T	N
yejH	2279168	G	A	A	A	A	A	A	A	A	A	A	G171D	N
yejK	2281672	C						T	T				G99S	N
yejM	2283731	T							G				V444G	N
insH-8 / narP	2288213	T									C		-	I
mqo	2303175	C		T	T	T	T	T	T	T	T	T	L533L	S

apbE	2309471	G						T	T				T28N	N
rcsD	2311618	A									G		M36V	N
atoA	2322280	C							T				L49L	S
atoE	2323202	G								A	A		C141Y	N
yfaA / gyrA	2334811	A								G			-	I
yfaL	2338904	T						C	C				E1095E	S
yfaL	2341049	A			G	G	G	G	G	G	G	G	G380G	S
yfaE	2346548	A						G	G				T4A	N
arnA	2367707	C		T	T	T	T					T	R548R	S
arnT	2370303	T						G					G457G	S
menB	2374030	T		C	C	C	C	C	C	C	C	C	N270S	N
yfbB	2375312	C		T	T	T	T					T	A100A	S
nuoN	2388646	G		A	A	A	A	A	A	A	A	A	I293I	S
nuoL	2392233	A							C				V278G	N
nuoE	2399613	G		A	A	A	A					A	T153T	S
nuoC	2401084	G	T	T	T	T	T	T	T	T	T	T	H261N	N
yfbS	2407799	T						C	C				T525A	N
yfbS	2407840	A							G				L511S	N
yfbV / ackA	2411283	G		A	A	A	A					A	-	I
yfcC	2415141	G								A	A		W12*	N (ter)
pdxB	2434910	C	T	T	T	T	T	T	T	T	T	T	D321N	N
yfcS	2450356	C								T	T		M0I	N
yfcU	2450507	C										A	L838F	N
gtrA	2466165	G							A				V96I	N
gtrS	2468380	T							C				F409L	N
emrY	2480196	C		T	T	T	T					T	M0I	N
emrK / evgA	2481572	G								A	A		-	I
yfdE	2487121	C								T	T		G23R	N
frc	2490946	G		A	A	A	A					A	H110Y	N
yfdX / ypdI	2492653	T		C	C	C	C					C	-	I
glk / yfeO	2507512	A		G	G	G	G	G	G	G	G	G	-	I
ptsI	2533632	G								A	A		L514L	S
pdxK / yfeK	2535315	G	T	T	T	T	T	T	T	T	T	T	-	I
eutN	2569963	T									C		D36G	N
tktB	2578198	T		C	C	C	C					C	Y180H	N
aegA	2583473	A						G	G				A24A	S
narQ	2584755	A						G	G				S334G	N
dapA	2597522	G							A				S86S	S
hyfB	2600797	G	T	T	T	T	T	T	T	T	T	T	G319C	N
hyfC	2602614	T								C			F248S	N
guaA	2629730	C		T	T	T	T	T	T	T	T	T	M275I	N

bamB	2636600	G		A	A	A	A					A	S24S	S
rodZ	2640863	T		C	C	C	C					C	N1D	N
sseB	2652794	T		C	C	C	C	C	C	C	C	C	L53L	S
suhB	2661810	T		C	C	C	C					C	V115A	N
hcaE	2667245	A									G		G63G	S
yphG	2678056	G	A	A	A	A	A	A	A	A	A	A	A903A	S
hmp	2683874	C						T	T				T5T	S
qseF	2686129	A		G	G	G	G	G	G	G	G	G	F232L	N
purL	2690904	C	T	T	T	T	T	T	T	T	T	T	G887S	N
yfhH	2697351	G						A					V190I	N
lepB	2702664	C		T	T	T	T					T	R222H	N
nadB	2708578	C		T	T	T	T					T	T45I	N
pssA	2720788	C		T	T	T	T					T	L13F	N
pssA	2720802	C		T	T	T	T	T	T	T	T	T	P17P	S
kgtP	2722866	C		T	T	T	T					T	A300A	S
kgtP / rrfG	2723968	A		G	G	G	G					G	-	I
yfiN	2741406	G								A			V333V	S
rpsP	2744061	G		A	A	A	A					A	G48G	S
grpE	2748243	T						C	C				D162G	N
smpA	2751695	G		A	A	A	A	A	A	A	A	A	L22L	S
alpA / yfjI	2756915	G						A	A				-	I
yfjW	2772845	C								T			G501G	S
yfjW / ypjI	2773051	G		A	A	A	A	A	A	A	A	A	-	I
yfjY / ypjJ	2774897	A		G	G	G	G					G	-	I
ypjA	2776539	T		C	C	C	C			C	C	C	T1403A	N
ypjA	2778431	T						C	C				N772S	N
ypjA	2780582	G	A	A	A	A	A	A	A	A	A	A	T55I	N
ypjA / pinH	2780902	A									G		-	I
ypjC / ileY	2783585	G		A	A	A	A			A	A	A	-	I
ygaH / mprA	2808766	G	A	A	A	A	A	A	A	A	A	A	-	I
gshA	2813043	A							G				N472N	S
argQ / argZ	2816019	A								G	G		-	I
alaS	2818431	C		T	T	T	T					T	G534S	N
hypF	2833488	A	G	G	G	G	G	G	G	G	G	G	W653R	N
hycC	2846936	G	A	A	A	A	A	A	A	A	A	A	Q109*	N (ter)
fhlA	2853157	C								T	T		F265F	S
fhlA	2853617	C		-	-		-					-	R419V	N (-1) (frameshift,ter)
ygbJ	2859530	C	T	T	T	T	T	T	T	T	T	T	R26C	N
nlpD	2866080	G	A	A	A	A	A	A	A	A	A	A	V231V	S
truD	2868925	A									G		V133V	S

ispF	2869428	A	G	G	G	G	G	G	G	G	G	G	D124D	S
ygcW / yqcE	2898473	G		C	C	C	C					C	-	I
ygcE / ygcF	2901785	A		G	G	G	G					G	-	I
ygcG	2903865	A		G	G	G	G					G	T44A	N
ygcG	2904514	A									G		D260G	N
fucA / fucP	2932140	C	T	T	T	T	T	T	T	T	T	T	-	I
recB	2953341	T	C	C	C	C	C	C	C	C	C	C	I228V	N
recC	2957276	C								T	T		D1058N	N
aas	2973521	C		T	T	T	T					T	M171I	N
lysR	2977775	C	T	T	T	T	T	T	T	T	T	T	P244S	N
araE / kduD	2980312	C		T	T	T	T	T	T	T	T	T	-	I
yqeG	2984655	G								A			A262T	N
ygeF / ygeG	2989180	C									T		-	I
ygeK′	2992757	A	G	G	G	G	G	G	G	G	G	G	S119P	N (pseudo)
ygeK′	2993044	C		A	A	A	A					A	G23V	N (pseudo)
yqeA / yqeB	3010598	C		T	T	T	T	T	T	T	T	T	-	I
ygfJ	3013185	T								C	C		S1P	N
ygfK	3015018	G									A		V312I	N
ygfK	3015796	C		T	T	T	T					T	A571V	N
ssnA	3018106	G								A	A		A307A	S
ygfT	3028951	C								T	T		M0I	N
ygfU	3029566	A	G	G	G	G	G	G	G	G	G	G	K59E	N
ygfU	3030225	T											G278G	S
recJ	3034750	G		A	A	A	A					A	S459L	N
scpB	3062715	G								A	A		R233H	N
yggC	3072177	G		A	A	A	A	A	A	A	A	A	Q178*	N (ter)
cmtA	3076022	A						G	G				L286S	N
tktA	3077762	C	T	T	T	T	T	T	T	T	T	T	M631I	N
glcB	3120193	C		T	T	T	T	T	T	T	T	T	A544A	S
glcB	3120741	C								A			D362Y	N
hybB	3141359	C		T	T	T	T	T	T	T	T	T	S275S	S
dkgA	3155134	C		T	T	T	T	T	T	T	T	T	H163Y	N
plsC	3161011	G						A	A				H164Y	N
parC / ygiS	3164072	C							T				-	I
ygiS / ygiT	3165858	C								T	T		-	I
ygiZ / mdaB	3170522	T								C			-	I
yqiC	3182918	T						C	C				P18P	S
yqiH	3188148	T						G	G				R81R	S
glgS / yqiJ	3190104	A								G	G		-	I
rfaE	3193941	A						G	G				S278P	N
yqjH / yqjI	3214574	T		C	C	C	C					C	-	I

ebgC	3224043	T								C			Y99Y	S
ygjK	3228684	C									T		T591I	N
psrN	3236439	G						A	A				-	R
alx	3236695	G							A				G31R	N
yhaH / yhaI	3250857	A	G	G	G	G	G	G	G	G	G	G	-	I
yhaI	3251094	T						A	A				L53L	S
tdcD	3261616	C		T	T	T	T					T	D22N	N
tdcA / tdcR	3265311	C		T		T							-	I
yhaB / yhaC	3266422	A		G	G	G	G					G	-	I
garR	3269995	A							C				V259G	N
agaR	3276186	C						T	T				G167S	N
yraQ	3295489	T		C	C	C	C	C	C	C	C	C	L223L	S
yhbO	3297216	G		A	A	A	A	A	A	A	A	A	G73D	N
yhbU	3299827	T								C	C		Y106Y	S
truB	3309955	C							T				R281H	N
infB	3312159	C						T	T				V625V	S
glmM	3321508	G	T	T	T	T	T	T	T	T	T	T	R194R	S
gltB	3356000	T		A	A	A					A	A	L1115H	N
gltF	3359450	A									G		I84V	N
nanA	3371196	C		T	T	T	T	T	T	T	T	T	V134M	N
yhcB / degQ	3378626	G									A		-	I
mdh / argR	3382323	T						C	C				-	I
yhdP	3390484	T		C	C	C	C					C	Q1265R	N
csrD	3400328	G								A	A		R342C	N
yhdH	3401567	A			G	G	G					G	D20G	N
yhdT	3405446	G						A					G16E	N
prmA dusB	3408003	C									A		-	I
rrsD / yrdA	3427066	C								T	T		-	I
fmt	3432295	C		T	T	T	T					T	L19L	S
rpsK	3439854	C								T			P88P	S
rpsJ	3451256	A								G	G		F12L	N
gspD	3455629	A									G		M410V	N
gspI	3460281	A		G	G	G	G					G	T95A	N
slyX	3475758	G						A					A32T	N
slyX	3475865	A						C					P67P	N
slyD	3476315	C								A	A		G68C	N
ppiA	3490194	C		T	T	T	T	T	T	T	T	T	L41L	S
tsgA	3491539	T								C			L316P	N
mrcA	3521312	G								A	A		K139K	S
yrfF	3526602	C		T	T	T	T					T	T703T	S
ompR	3534169	G		A	A	A	A					A	N145N	S

gntX	3543196	T						C	C				L97P	N
malQ	3547947	A										C	V48G	N
livK	3595302	T		C	C	C	C	C	C	C	C	C	K93K	S
yhhK	3596170	C		T	T	T	T					T	T54I	N
dtpB	3639900	A									G		K338R	N
arsR	3646625	C								T	T		L24L	S
slp	3652184	G							A				T66T	S
mdtF	3660175	C	T	T	T	T	T	T	T	T	T	T	A579V	N
gadA / yhjA	3665712	C								T	T		-	I
yhjE	3672895	T						C	C				I28I	S
yhjR	3694066	G								A			A47V	N
eptB	3708412	A									G		F28F	S
xylB	3726787	T		C	C	C	C	C	C	C	C	C	Y202C	N
lyxK	3746232	G						A					V375M	N
yiaS	3748719	C		T	T	T	T			T	T	T	A203V	N
kbl	3790238	C						T	T				D112N	N
rfaD	3792548	G	A	A	A	A	A	A	A	A	A	A	S179N	N
rfaF	3793239	G		C	C	C	C					C	V95V	S
gmk	3820066	G						A	A				A205T	N
recG	3823243	G		T	T	T	T					T	R3L	N
gltS / xanP	3826945	T						C	C				-	I
uhpT	3844428	C						T	T				V254M	N
uhpB	3848040	C						T	T				V39V	S
yidK	3856878	A								G	G		I420T	N
tnaA	3888159	A		C	C	C	C					C	K468N	N
asnC	3924661	T		C	C	C	C			C	C	C	V121V	S
rbsA	3932995	G		A	A	A	A					A	G398S	N
ilvA	3953687	G						C					A111P	N
ilvA	3953690	A						C					T112P	N
rfe	3967035	C		T	T	T	T					T	T365I	N
wzxE	3974865	C	T	T	T	T	T	T	T	T	T	T	P188S	N
corA / yigF	4000438	C						C	C				-	I
tatD	4022329	T		C	C	C	C					C	N250N	S
trkH	4032209	G									C		G347R	N
yihL	4058730	G			A	A	A					A	R86R	S
fdhE	4078346	G	A	A	A	A	A	A	A	A	A	A	F301F	S
fdoG	4082878	A		G	G	G	G	G	G	G	G	G	F322S	N
rhaT	4097747	T		C	C	C	C	C	C	C	C	C	M267V	N
yiiM	4101165	A		-	-	-	-	-	-	-	-	-	G106G	S (frameshift,ter)
metB	4127094	C						T	T				L133L	S

metL	4129026	G		A	A	A	A	A	A	A	A	A	G389E	N
yijF	4135247	T	C	C	C	C	C	C	C	C	C	C	Y144C	N
pflC	4144596	G	C	C	C	C	C	C	C	C	C	C	E105Q	N
yijP	4147956	T		C	C	C	C					C	E110E	S
sroH	4188387	C							T				-	R
yjaH	4199273	A							G				K228E	N
zraS	4200545	C		T	T	T	T					T	R198R	S
zraS	4200748	C						T	T				A266V	N
metA	4212790	A		C	C	C	C					C	K162T	N
yjbH	4236463	T								C	C		V268V	S
malF	4243054	C								T	T		L14L	S
malE	4243337	A									G		V368A	N
qorA	4261577	C								T			L225L	S
alr	4264102	C	T	T	T	T	T	T	T	T	T	T	H99Y	N
tyrB	4266134	A						G	G				E332G	N
yjcF	4280554	C								T	T		C181Y	N
fdhF	4296946	C			A	A	A					A	L147L	S
mdtN	4301574	C		T	T	T	T	T	T	T	T	T	A186T	N
yjcS	4303465	G		A	A	A	A	A	A	A	A	A	L385L	S
alsA	4308946	C								T			A19T	N
alsB	4309520	G								A			F181F	S
phnK	4316452	C				T	T					T	A111A	S
yjdA	4326099	A								G			A313A	S
proP	4328721	T		C	C	C	C	C	C	C	C	C	V65A	N
basS	4330339	A								G	G		G318G	S
eptA	4332857	G		A	A	A	A			A	A	A	L252F	N
adiY	4335748	A						G	G				F68L	N
adiA	4337152	A		G	G	G	G					G	Y464H	N
dcuS / yjdI	4349686	C									T		-	I
cadA	4355149	T									C		M497V	N
yjeJ	4372019	G							A				A79V	N
yjeJ / yjeK	4372469	A								G	G		-	I
frdA	4379497	T		C	C	C	C	C	C	C	C	C	Y281C	N
poxA / yjeM	4381818	C						T	T				-	I
orn	4389975	T								C	C		F116L	N
mutL	4396035	G									T		A200S	N
ulaF	4421992	C								T			G89G	S
ytfA	4426010	T									C		S98P	N
ytfN	4443548	A						C					T471P	N
ytfQ	4448890	G	A	A	A	A	A	A	A	A	A	A	L301L	S
mpl / yjgA	4455255	G								A			-	I

nrdD	4460076	G		A	A	A	A	A	A	A	A	A	T202M	N
nrdD	4460425	C									T		D86N	N
treB	4464115	A						G	G				I29T	N
yjgL	4474024	A	G	G	G	G	G	G	G	G	G	G	N188D	N
argI / rraB	4476359	C	T	T	T	T	T	T	T	T	T	T	-	I
yjgN	4478628	T		C	C	C	C					C	V291V	S
leuX	4494512	A		G	G	G	G			G	G	G	-	R
intB	4495321	C						T	T				R207R	S
yjgX	4498345	G								A	A		Q152*	N (ter)
yjhB	4502332	G								A			M83I	N
insB-7	4517153	C						T	T				A136V	N
yjhF	4519014	T		C	C	C	C	C	C	C	C	C	I343V	N
yjhF	4519853	A		G	G	G	G					G	I63T	N
sgcA	4525621	A						C					V127G	N
fimB	4539173	G								A	A		C64Y	N
fimC	4542397	T		C	C	C	C					C	M23T	N
kptA	4559160	T	C	C	C	C	C	C	C	C	C	C	F69L	N
mcrC	4575017	C						A	A				G321V	N
mcrC	4575497	T		C	C	C	C					C	Y161C	N
hsdR	4582303	T						C	C				T827A	N
tsr	4591284	G			A	A	A					A	A534A	S
yjjM	4593554	C						T	T				G106G	S
deoC	4616054	G									A		A236T	N
deoB	4617796	C						T	T				T56T	S
lplA	4622080	C	T	T	T	T	T	T	T	T	T	T	E20K	N
nadR	4626139	G						T	T				V267F	N
yjtD	4639601	G									T		A212S	N

Table S4: RNA-SEQ results for CB1000 and CB2000 as compared to Founder.

CB1000 vs Founder

Up in CB1000

Gene	Function	Fold Change	Test statistic	P-value	FDR
icdC	isocitrate dehydrogenase	43.50	10.31	0.00	0.00
fimA	pilin	21.08	49.49	0.00	0.00
fimI	fimbrial protein	19.18	14.80	0.00	0.00
fimF	minor component	17.71	6.49	0.00	0.00
fimC	chaperone	17.00	6.01	0.00	0.00
fimD	porin protein	11.21	8.06	0.00	0.00
fimG	minor component	8.64	6.00	0.00	0.00
fimH	minor component	5.58	5.17	0.00	0.00
ppdC	predicted protein	3.88	3.62	0.00	0.03
malP	maltodextrin phosphorylase	2.88	3.67	0.00	0.03
elaB	conserved protein	2.56	4.37	0.00	0.00
yifN	conserved protein	2.37	3.48	0.00	0.04
ydhZ	predicted protein	2.28	7.17	0.00	0.00
rrlH	23S ribosomal RNA	1.87	29.04	0.00	0.00
evgA	RR dna binding acid resistance	1.85	3.78	0.00	0.02
rpmI	50S ribosomal subunit L35	1.79	3.86	0.00	0.02
ycfP	conserved protein	1.61	4.92	0.00	0.00
mprA	negative regulator multi drug efflux pumps	1.54	4.23	0.00	0.00

Down in CB1000

Gene name	Function	Fold Change	Test statistic	P-value	FDR
yjdQ	pseudogene	-7.63	-3.69	0.00	0.03
fruK	1-phosphofructokinase	-6.21	-7.76	0.00	0.00
fruA	part of fructose permease	-5.26	-13.47	0.00	0.00
fruB	part of fructose permease	-5.22	-6.24	0.00	0.00
proK	tRNA for proline	-3.18	-7.74	0.00	0.00
oxyS	oxidative stress regulator	-2.46	-3.48	0.00	0.04
crcB	protein involved in resistance to camphor-induced chromosome	-2.24	-3.69	0.00	0.03

	decondensation				
rrfB	5S ribosomal RNA	-2.22	-3.84	0.00	0.02
metU	tRNA for met for elongation	-1.93	-4.99	0.00	0.00
ydaL	conserved protein, same motif as MutS	-1.83	-3.54	0.00	0.04
ogrK	regulates bacteriophage P2 late transcription	-1.81	-4.30	0.00	0.00
adhE	alcohol dehydrogenase	-1.71	-5.28	0.00	0.00
epd	erythrose 4-phosphate dehydrogenase	-1.71	-8.69	0.00	0.00
yqhA	conserved inner membrane protein	-1.65	-3.85	0.00	0.02
sdaC	serine STP transporter	-1.54	-3.69	0.00	0.03
fbaA	fructose bisphosphate aldolase	-1.51	-4.48	0.00	0.00

CB2000 vs Founder

Up in CB2000					
Gene name	**Function**	**Fold Change**	**Test statistic**	**P-value**	**FDR**
icdC	isocitrate dehydrogenase	59.85	12.45	0.00	0.00
nrdF	ribonucleoside-diphosphate reductase 2, β subunit dimer	6.27	3.79	0.00	0.01
ytfE	protein involved in repair of stress-damaged iron-sulfur clusters	4.23	5.08	0.00	0.00
zinT	cadmium-induced cadmium binding protein	3.72	3.51	0.00	0.02
proP	ProP osmosensory MFS transporter	3.03	7.13	0.00	0.00
asnV	asparagine tRNA	2.70	3.90	0.00	0.00
gcd	glucose dehydrogenase	2.70	3.70	0.00	0.01
fimC	periplasmic chaperone for type 1 fimbrae	2.65	3.17	0.00	0.04
yqaE	predicted membrane protein	2.36	3.44	0.00	0.02
omrB	small RNA, controls outermembrane protein composition	2.25	4.19	0.00	0.00
fecA	outer membrane receptor, citrate-dependent iron transport	2.18	6.71	0.00	0.00
ydbK	predicted pyruvate:flavodoxin oxidoreductase	2.15	6.53	0.00	0.00
ilvE	val, Leu, isoleu synthesis	2.13	3.50	0.00	0.02
yhjR	conserved protein	1.99	3.14	0.00	0.04

glpD	glycerol 3-phosphate dehydrogenase	1.95	3.60	0.00	0.01
fecC	iron dictrate ABC transporter	1.94	4.68	0.00	0.00
pheA	chorismate mutase / prephenate dehydratase	1.87	4.97	0.00	0.00
fecB	iron dictrate ABC transporter	1.87	6.13	0.00	0.00
fecD	iron dictrate ABC transporter	1.85	3.58	0.00	0.01
inaA	induced by ph stress	1.80	3.17	0.00	0.04
add	deoxyadenosine deaminase	1.78	5.22	0.00	0.00
leuX	leucine tRNA	1.77	6.63	0.00	0.00
rhlE	ATP dependent RNA helicase, ribosome maturation	1.72	4.50	0.00	0.00
cspA	cold shock protein A, ups hsn and gyrA	1.62	13.62	0.00	0.00
pepA	aminopeptidase A, controls transcription of operon in pyrimidine sysn	1.61	4.77	0.00	0.00
sdaA	breaks down serine to pyruvate	1.59	4.38	0.00	0.00
rimN	tRNA modifying enzyme	1.58	4.21	0.00	0.00
folE	biosyn of tetrahydrofolate	1.57	4.32	0.00	0.00
yibL	conserved protein, associates with the ribosome	1.53	4.00	0.00	0.00
gpp	pppGpp to ppGpp, also exopolyphosphatase	1.51	3.35	0.00	0.03

Down in CB2000

Gene name	**Function**	**Fold Change**	**Test statistic**	**P-value**	**FDR**
bcsF	predicted protein	-3.73	-3.30	0.00	0.03
rzpD	DLP12 prophage; predicted murein endopeptidase	-3.54	-5.48	0.00	0.00
dhaK	dihydroxyacetone kinase subunit K	-3.26	-7.89	0.00	0.00
borD	bacteriophage lambda Bor protein homolog	-2.92	-9.19	0.00	0.00
ibsE	toxic peptide	-2.65	-3.19	0.00	0.04
nupG	nucleoside transport	-2.58	-5.20	0.00	0.00
treB	trehalose PTS permease	-2.46	-4.64	0.00	0.00
fruB	fructose PTS permease	-2.37	-5.10	0.00	0.00
yjiY	predicted inner membrane protein	-2.32	-3.33	0.00	0.03
malT	transcription factor	-2.28	-4.76	0.00	0.00

symR	small RNA, regulates SymE an SOS induced toxin protein	-2.27	-3.23	0.00	0.04
yehU	predicted sensory kinase in two-component system with YehT	-2.22	-3.40	0.00	0.02
ogrK	DNA-binding transcriptional regulator, prophage P2 remnant	-2.20	-6.17	0.00	0.00
mlaA	putative lipoprotein	-2.20	-7.31	0.00	0.00
dhaM	dihydroxyacetone kinase subunit M	-2.18	-3.92	0.00	0.00
yrbL	predicted protein	-2.18	-6.82	0.00	0.00
ycdZ	predicted inner membrane protein	-2.13	-6.65	0.00	0.00
fruK	1-phosphofructokinase	-2.11	-4.73	0.00	0.00
glnX	glutamine tRNA	-2.04	-7.79	0.00	0.00
dhaL	dihydroxyacetone kinase subunit L	-2.03	-4.05	0.00	0.00
cspD	DNA replication inhibitor	-2.00	-3.91	0.00	0.00
dctA	dicarboxylate transporter	-1.95	-4.23	0.00	0.00
rbsD	ribose pyranase	-1.92	-10.24	0.00	0.00
fruA	fructose PTS permease	-1.91	-7.03	0.00	0.00
yffS	CPZ-55 prophage; predicted protein	-1.90	-4.53	0.00	0.00
ilvB	acetohydroxybutanoate synthase, valine syn	-1.85	-3.94	0.00	0.00
ivbL	ilvB operon leader peptide	-1.85	-3.81	0.00	0.01
ydiE	conserved protein	-1.85	-4.65	0.00	0.00
proK	proline tRNA	-1.84	-4.77	0.00	0.00
glyU	glycine tRNA	-1.81	-3.80	0.00	0.01
maa	maltose acetyltransferase, broad range	-1.81	-4.37	0.00	0.00
mgtA	magnesium transporter	-1.81	-5.16	0.00	0.00
aspA	aspartate ammonia-lyase	-1.79	-3.62	0.00	0.01
ylaC	predicted inner membrane protein	-1.77	-3.80	0.00	0.01
spr	outer membrane lipoprotein	-1.75	-4.69	0.00	0.00
csrB	small RNA, regulates CsrA	-1.75	-8.63	0.00	0.00
manX	mannose permease	-1.70	-4.94	0.00	0.00
ompT	outer membrane protease	-1.66	-7.39	0.00	0.00
rbsC	ribose transport	-1.65	-4.42	0.00	0.00
yggE	conserved protein involved in oxidative stress	-1.62	-3.28	0.00	0.03
tsx	nucleoside channel	-1.60	-4.95	0.00	0.00
sdaC	serine transport	-1.59	-10.36	0.00	0.00
proW	proline transport	-1.58	-3.16	0.00	0.04
ynfB	predicted protein	-1.58	-3.79	0.00	0.01

nupC	nucleoside transport	-1.58	-4.41	0.00	0.00
rbsA	ribose transport	-1.56	-3.70	0.00	0.01
gatY	tagatose-1,6-bisphosphate aldolase	-1.55	-3.26	0.00	0.03
ttcA	predicted C32 tRNA thiolase	-1.54	-4.27	0.00	0.00
lpxT	protein with low undecaprenyl pyrophosphate phosphatase activity	-1.54	-3.76	0.00	0.01
slyB	outer membrane lipoprotein	-1.52	-7.19	0.00	0.00
ygfZ	folate binding protein, poss regulatory	-1.52	-3.35	0.00	0.03
yifL	predicted lipoprotein	-1.51	-3.76	0.00	0.01

References

1. Harris, D. R., Pollock, S. V., Wood, E. A., Goiffon, R. J., Klingele, A. J., Cabot, E. L., Schackwitz, W., Martin, J., Eggington, J., Durfee, T. J., Middle, C. M., Norton, J. E., Popelars, M., Li, H., Klugman, S. A., Hamilton, L. L., Bane, L. B., Pennacchio, L., Albert, T. J., Perna, N. T., Cox, M. M., and Battista, J. R. (2009) Directed evolution of radiation resistance in Escherichia coli, *Journal of bacteriology 191*, 5240-5252.

2. Cox, M. M., and Battista, J. R. (2005) Deinococcus radiodurans - The consummate survivor, *Nature Reviews Microbiology 3*, 882-892.

3. Blasius, M., Hubscher, U., and Sommer, S. (2008) Deinococcus radiodurans: What belongs to the survival kit?, *Critical Reviews in Biochemistry and Molecular Biology 43*, 221-238.

4. Sonntag, C. v. (2005) Free-Radical-Induced DNA Damage and Its Repair: A chemical perspective, *Springer Publishing*.

5. Zimmerman, J. M., and Battista, J. R. (2005) A ring-like nucleoid is not necessary for radioresistance in the Deinococcaceae, *BMC Microbiology 5*, 17.

6. Levin-Zaidman, S., Englander, J., Shimoni, E., Sharma, A. K., Minton, K. W., and Minsky, A. (2003) Ringlike structure of the Deinococcus radiodurans genome: A key to radioresistance?, *Science 299*, 254-256.

7. Daly, M. J. (2012) Death by protein damage in irradiated cells, *DNA repair 11*, 12-21.

8. Daly, M. J. (2009) A new perspective on radiation resistance based on Deinococcus radiodurans, *Nat Rev Microbiol 7*, 237-245.

9. Zahradka, K., Slade, D., Bailone, A., Sommer, S., Averbeck, D., Petranovic, M., Lindner, A. B., and Radman, M. (2006) Reassembly of shattered chromosomes in Deinococcus radiodurans, *Nature 443*, 569-573.

10. Daly, M. J., Gaidamakova, E. K., Matrosova, V. Y., Vasilenko, A., Zhai, M., Venkateswaran, A., Hess, M., Omelchenko, M. V., Kostandarithes, H. M., Makarova, K. S., Wackett, L. P., Fredrickson, J. K., and Ghosal, D. (2004) Accumulation of Mn(II) in, Deinococcus radiodurans facilitates gamma-radiation resistance, *Science 306*, 1025-1028.

11. Slade, D., and Radman, M. (2011) Oxidative Stress Resistance in Deinococcus radiodurans, *Microbiology and Molecular Biology Reviews 75*, 133-+.

12. Krisko, A., and Radman, M. (2010) Protein damage and death by radiation in Escherichia coli and Deinococcus radiodurans, *Proc Natl Acad Sci U S A 107*, 14373-14377.

13. Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E., and Kim, J. F. (2009) Genome evolution and adaptation in a long-term experiment with Escherichia coli, *Nature 461*, 1243-1247.

14. Hindre, T., Knibbe, C., Beslon, G., and Schneider, D. (2012) New insights into bacterial adaptation through in vivo and in silico experimental evolution, *Nat Rev Microbiol 10*, 352-365.

15. Elena, S. F., and Lenski, R. E. (2003) Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation, *Nature Reviews Genetics 4*, 457-469.

16. Gerrish, P. J., and Lenski, R. E. (1998) The fate of competing beneficial mutations in an asexual population, *Genetica 103*, 127-144.

17. Perron, G. G., Lee, A. E., Wang, Y., Huang, W. E., and Barraclough, T. G. (2012) Bacterial recombination promotes the evolution of multi-drug-resistance in functionally diverse populations, *Proceedings. Biological sciences / The Royal Society 279*, 1477-1484.

18. Rozen, D. E., de Visser, J., and Gerrish, P. J. (2002) Fitness effects of fixed beneficial mutations in microbial populations, *Current Biology 12*, 1040-1045.

19. Futuyma, D. J. (1986) Reflections on reflections: ecology and evolutionary biology, *Journal of the history of biology 19*, 303-312.

20. Cox, M. M., Goodman, M. F., Kreuzer, K. N., Sherratt, D. J., Sandler, S. J., and Marians, K. J. (2000) The importance of repairing stalled replication forks, *Nature 404*, 37-41.

21. Marians, K. J. (2008) Understanding how the replisome works, *Nature Structural & Molecular Biology 15*, 125-127.

22. Rangarajan, S., Woodgate, R., and Goodman, M. F. (2002) Replication restart in UV-irradiated Escherichia coli involving pols II, III, V, PriA, RecA and RecFOR proteins, *Molecular Microbiology 43*, 617-628.

23. Demple, B. (1996) Redox signaling and gene control in the Escherichia coli soxRS oxidative stress regulon--a review, *Gene 179*, 53-57.

24. Mao, Y., Doyle, M. P., and Chen, J. (2001) Insertion mutagenesis of wca reduces acid and heat tolerance of enterohemorrhagic Escherichia coli O157:H7, *Journal of bacteriology 183*, 3811-3815.

25. Plumbridge, J., and Vimr, E. (1999) Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminic acid by Escherichia coli, *Journal of bacteriology 181*, 47-54.

26. Lenski, R. E., Rose, M. R., Simpson, S. C., and Tadler, S. C. (1991) Long-Term Experimental Evolution in Escherichia-Coli .1. Adaptation and Divergence During 2,000 Generations, *American Naturalist 138*, 1315-1341.

27. Baker, T. A., and Sauer, R. T. (2006) ATP-dependent proteases of bacteria: recognition logic and operating principles, *Trends in biochemical sciences 31*, 647-653.

28. Kiley, P. J., and Beinert, H. (1998) Oxygen sensing by the global regulator, FNR: the role of the iron-sulfur cluster, *FEMS microbiology reviews 22*, 341-352.

29. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., and Gilad, Y. (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays, *Genome Res 18*, 1509-1517.

30. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products, *Proc Natl Acad Sci U S A 97*, 6640-6645.

31. Miller, J. H. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

32. Breed, R. S., and Dotterrer, W. D. (1916) The Number of Colonies Allowable on Satisfactory Agar Plates, *Journal of bacteriology 1*, 321-331.

33. Croucher, N. J., and Thomson, N. R. (2010) Studying bacterial transcriptomes using RNA-seq, *Current opinion in microbiology 13*, 619-624.

34. Durfee, T., Hansen, A. M., Zhi, H., Blattner, F. R., and Jin, D. J. (2008) Transcription profiling of the stringent response in Escherichia coli, *Journal of bacteriology 190*, 1084-1096.

35. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq, *Nature methods 5*, 621-628.

36. Baggerly, K. A., Deng, L., Morris, J. S., and Aldaz, C. M. (2003) Differential expression in SAGE: accounting for normal between-library variation, *Bioinformatics 19*, 1477-1483.

37. Lewis, I. A., Schommer, S. C., and Markley, J. L. (2009) rNMR: open source software for identifying and quantifying metabolites in NMR spectra, *Magnetic resonance in chemistry : MRC 47 Suppl 1*, S123-126.

Chapter 3: Surviving extreme exposure to ionizing radiation: *Escherichia coli* genes and pathways

Rose T. Byrne[1], Elizabeth A. Wood[1], Eric L. Cabot[2], John R. Battista[3], and Michael M. Cox[1*]

[1]Department of Biochemistry, University of Wisconsin – Madison, Madison, Wisconsin 53706-1544, [2]Genome Center, University of Wisconsin, 425G Henry Mall, Madison, WI 53703, [3]Department of Biological Sciences, Louisiana State University and A & M College, Baton Rouge, LA 70803

*To whom correspondence should be addressed.

Reference:

Byrne RT, Wood EA, Cabot EL, Battista, JR, Cox, MM. (2013) Surviving extreme exposure to ionizing radiation: *Escherichia coli* genes and pathways. **eLife** *submitted*

:

3.1 Abstract

An improved understanding of the mechanisms used by bacterial cells to survive extreme exposure to ionizing radiation (IR) ultimately requires a complete molecular description of the cellular processes that contribute. To further this understanding, we broadly screened for genes involved in IR resistance using a high-throughput genetic screening technique, transposon-directed insertion sequencing (TraDIS) that allowed us to assay nearly all of the non-essential genes in the *E. coli* genome for their contribution to IR recovery. We report that recovery from high-level exposure to IR requires the function of at least 47 screen-identified genes, with 14 of these subjected to direct verification. There are two major conclusions. First, the genes identified affect many different cellular processes. About 40% cluster in DNA repair pathways. The most prominent are *recF*, *recG*, and *recN*, genes known to be central to the process of recombinational DNA repair. Contributions by genes involved in protein synthesis/stabilization, cellular responses to oxidative damage, cell wall structure and function, and genes of unknown function were also evident. Second, a gene encoding a putative helicase of previously unknown function, *yejH*, strongly contributes to IR survival in *E. coli*. The deletion of *yejH* yields cells 100x more sensitive to IR exposure at a dose of 2 kGy than an isogenic parent. We hypothesize that YejH is a DNA repair enzyme involved in repairing damage resulting from radiation toxicity. YejH interacts directly with SSB, and shares sequence similarity with the eukaryotic NER helicase, XPB.

YejH appears to function in a novel repair pathway with the product of the gene *uup*, previously implicated in maintenance of genome stability. The substantial requirement for radiation resistance is the first phenotype to be reported for *yejH.* We propose that *yejH* be renamed to *radD*.

3.2 Introduction

Organisms have evolved mechanisms to maintain genomic integrity in the face of extreme environmental stresses. One class of extremophiles, typified by the bacterium *Deinococcus radiodurans* (*1, 2*), exhibits extraordinary resistance to the effects of high doses of ionizing radiation (IR). The repair of damaged DNA, stalled replication forks, and other damaged cellular components are critical for cells to survive exposure to IR. The DNA sugar-phosphate backbone is particularly susceptible to both direct and indirect damage caused by IR (*3, 4*). Direct damage is caused by absorption of IR by the DNA molecule, which can lead to strand breakage and chemical alterations of bases. In contrast, indirect damage occurs when reactive oxygen species (ROS), such as hydroxyl radicals, formed when IR is absorbed by water, interact with DNA. Hydroxyl radicals produce single-strand DNA breaks. Double-strand DNA breaks (DSB) can occur when two IR-induced single-strand DNA breaks are in close proximity (*5*). DSBs are the most lethal form of DNA damage because they halt DNA replication, cause the collapse of the replication fork, and are difficult to repair (*1, 6, 7*). Cells repair DSBs and other DNA damage caused by IR, utilizing recombinational DNA repair and nonhomologous end-

joining (*6-10*). Because ROSs are also byproducts of aerobic respiration and general metabolism, it is likely that genes involved in IR survival are also involved in preserving DNA integrity under normal conditions, suggesting an essential role in bacteria.

Our understanding of the molecular basis of cellular survival of extreme exposure to IR has proceeded through stages. Early studies focused almost exclusively on the capacity of cells to repair DNA, particularly double strand breaks (*11-19*). Early hypotheses about the enhanced survival of *Deinococcus* similarly focused on a potentially novel DNA repair process (*15, 17-19*). While acknowledging the inherent importance of DNA repair in recovery from high doses of IR, the Daly group, and more recently the Radman group, have focused attention on the importance of amelioration of oxidative damage to proteins (*20-24*). Based on a variety of evidence, it has been argued that extreme radiation resistance does not reflect an enhanced or novel DNA repair functionality, but instead reflects a cellular capacity to protect a set of common DNA repair functions from oxidative inactivation (*24, 25*). An assessment as to whether this represents a complete description of the mechanisms underlying the extreme IR resistance phenotype is still needed.

We have carried out an exercise in directed evolution in which the *Escherichia coli* K-12 strain MG1655 acquired the phenotype of extreme resistance to IR as described in Chapter 2 and (*26*). Four evolved populations of *E. coli* were obtained, exhibiting levels

of IR resistance approaching that of *D. radiodurans*. Analysis of numerous sequenced isolates from these populations allowed us to identify the genetic alterations accounting for most of the acquired IR resistance phenotype (Chapter 2). In one highly evolved isolate, the phenotype is largely explained by three mutations in DNA repair/replication genes *recA*, *dnaB*, and *yfjK*. The modified genes provide the beginning of a molecular explanation for the adaptations needed to survive extreme radiation resistance.

Efforts to understand this phenotype have focused to a large extent on *Deinococcus radiodurans* and related bacteria. Analysis of transcriptome (*27*) and proteome changes upon IR exposure (*28*), as well as careful analysis of how the genome is reconstructed over time have provided some important insights into this bacterium's response to IR. However, broad genetic screens to identify all contributing processes are very difficult to do in *Deinococcus*, reflecting its multiploid genome. *E. coli* strains with an extreme IR resistance phenotype provide an opportunity to utilize a highly tractable and insight-fertile genetic system to more broadly explore the molecular basis of this phenotype. One step towards a more complete description of the genetic requirements for IR resistance would be the identification of contributing genes that were not modified in the directed evolution trials. That identification requires a different genetic approach.

Many screens have been carried out to identify genes involved in DNA repair in *E. coli* (*29-36*). These have resulted in the discovery of many of the key DNA repair enzymes we continue to study today. Screens to identify genes involved in radiation resistance were part of these efforts. The *recG* and *recN* genes were characterized to an extent as genes involved in radiation resistance and given a rad nomenclature (*radC* and *radB*, respectively) until their functions were further understood (*12, 37*). However, modern screening methods are much more sensitive and offer a more robust path to the discovery of new genes with particular functions. We sought to identify the genes involved in survival after extreme IR exposure for three additional reasons: 1) We do not understand the physiological function of one-third of the genes encoded by *E. coli* despite its role as the most extensively studied organisms. 2) Radiation resistance is a complex phenotype whose molecular basis remains the subject of some controversy (*1, 2, 24, 38*). 3) Whereas radiation damages all cellular components, current research tends to focus on either DNA damage or protein oxidation. We thus set out to identify additional genes that contribute to IR resistance and provide a more global assessment of the cellular processes that contribute to IR resistance.

A range of modern screening methods have been described (*39-43*) that utilize transposon mutagenesis in combination with Illumina sequencing. These techniques measure each gene's contribution to fitness on a genome scale through massive sequencing of transposon-genome junctions in highly mutagenized populations. We

employed a relatively new procedure called Transposon Directed Insertion Sequencing (TraDIS) (*39*). In this method, saturating transposon mutagenesis was performed and the resulting insertion mutants were pooled together to make an insertion mutant library that was then subjected to repeated exposures to IR. Genomic DNA from the non-treated population as well as the irradiated populations was isolated and the location of each transposon insertion as well as the frequency of each insertion mutant within the population was determined. The change in frequency of insertion mutant in each gene within the population was calculated, reflecting the effect the insertion has on a strain's ability to survive radiation exposure.

Using TraDIS, we have identified 47 candidate genes that appear to have a significant role in survival after IR exposure. A large number of these were directly validated. This work implicates the enzymes encoded by the uncharacterized *yejH* and minimally characterized *uup* genes in a potentially new pathway with an important role in the recovery of cells from IR damage. A general clustering of IR resistance genes into DNA repair functionalities highlights yet again the importance of efficient DNA repair in this phenotype. However, genes in several other functional classes are also represented. In addition to providing an overview of genes with a role in IR resistance, we initiate the characterization of *yejH* and its protein product, and propose the renaming of this gene to *radD*.

3.3 Results

3.3.1 TraDIS was performed to identify genes involved in IR survival.

The original directed evolution trials were carried out with an aliquot of *E. coli* strain MG1655 obtained from F. R. Blattner (*44*). Deep sequencing revealed 7 mutations in this strain (designated Founder) relative to the type strain database (*26*) A mutant library consisting of 500,000 insertion mutants was generated in Founder, and also in one highly evolved strain, CB2000. Each library was subjected to 5 rounds of irradiation followed by competitive outgrowth as diagramed in Figure 1. A non-treated control was taken through the entire experiment, treated identically except that it sat outside of the irradiator during treatment. To identify genes that contribute to IR survival using TraDIS, genomic DNA from the mutagenized population was isolated after the first and fifth IR treatment passage. TraDIS was carried out using Illumina high-throughput sequencing to identify genes that – when disrupted – caused cells to drop out of the population after IR treatment as illustrated for *uvrB* in Figure 1B. The method can be verified in part by examining the effects of irradiation on the insertion patterns of genes known to have major roles in the repair of DNA damage inflicted by IR. As expected, *uvrB*, which is involved in nucleotide excision repair, exhibited the following two part insertion pattern reflecting an essential role in IR resistance: (1) Numerous insertions were present in the *uvrB* gene in the non-irradiated populations indicating the gene is not essential for normal growth. (2) However, there were reduced transposon directed

sequence reads in this gene after IR treatment passage 1 and no sequence reads in this gene after IR treatment passage 5 (Figure 1). These results suggest that any cells that had insertions in *uvrB* rapidly dropped out of the population upon IR treatment.

3.3.2 General sequencing results.

The protocol described above generated fifteen to thirty million reads per sample. The reads were mapped to the *E. coli* genome, and the number of unique transposon insertion sites and the average base pair distance between inserts for each sample was calculated (Table 1). These results suggest that the mutagenesis was saturating with 1 insertion per 40-50 bases for the mutant pool after 1 passage. For the mutant pool after 5 passages, 1 transposon insertion per 100-200 bases was detected. The decline was expected as it was previously reported that passaging reduces the number of unique mutants in the pool, even in the absence of stress (*39*). This is due to genetic bottlenecks that occur during passaging and competition between strains with different mutations. Insertion densities were calculated for the mutant pools from each passage, as previously described (Table 1) (*45*). The gene length boundaries were calculated to determine the minimum length of a gene (in bps) required to ensure that the absence of sequenced transposon insertions signaled an essential gene function rather than a random chance occurrence (p-value < 0.05). This value differed by sample due to varying insertion densities obtained for each sample.

We note that approximately 670 genes in the *E. coli* genome (about 13% of the total) are required for normal growth in an unstressed environment under our growth conditions as indicated by the absence of insertions in these genes in our No IR control. No transposon insertions appear in these genes in either non-irradiated or irradiated samples. We are thus effectively screening the approximately 3,555 genes denoted non-essential in our non-irradiated sample. Our goal was to identify those genes that are not necessary during normal growth, but which become important when cells are heavily dosed with ionizing radiation.

Essential genes have previously been surveyed in *E. coli*. Of 620 genes denoted as essential in one survey that covered 87% of the genomic open reading frames in *E. coli* (*46*, *47*), only 55% overlap with the essential genes found in our study. A second survey carried out under different conditions produced a list of 300 essential genes (*48*), of which 94% appeared essential in our study. The differences observed between these 3 studies are likely due to different growth conditions, the presence or absence of competitive outgrowth, the approach for distinguishing essential versus non-essential genes, and the depth in which the mutant libraries were assayed. Because of the requirement for outgrowth in our protocol, any gene inactivation that produces a sufficient decline in growth rate under our conditions will lead to that gene's inclusion on the list of essential genes.

3.3.3 Identification of genes involved in IR survival.

After removal of the transposon tag sequence, each read was mapped to the *E. coli* genome. The first genome-derived bp of each read defined the genomic location of each transposon insertion within the mutant pool. The number of transposon insertion locations for each gene was used to calculate the relative contribution of non-essential genes to IR survival. Contribution values were only calculated for genes with at least three independent insertion sites to reduce variability that can result in misleading fitness calculations (*49*). We identified genes that – when disrupted – resulted in reduced IR survival fitness after passage 1 and after passage 5. A total of 47 genes were thus identified in Founder (Table 2). We also note that over 90% of the non-essential genes in the *E. coli* genome exhibited little or no difference in the observed insertion patterns with or without irradiation. The genes of interest in this study are those exhibiting transposon insertion patterns similar to *uvrB* in Figure 1 and are listed in Table 2 Panel A.

Deletion or alteration of some genes involved in DNA repair are known to result in slow growth phenotypes in rich media (*50-58*). The otherwise non-essential *recA* protein, clearly important for IR resistance, is not present in our list because strains with alterations resulting in *recA* gene inactivation grow somewhat slower and are unable to compete with the broader population during outgrowth. A total of 19 of the 47 genes in table 2 exhibited patterns that reflected somewhat slow growth, although the decline in growth rate was insufficient to remove them from our screen at least in passage 1. Cells

disrupted for these genes had reduced fitness upon irradiation in passage 1.

By passage 5, insertions in these genes disappeared from both the irradiated and the non-irradiated samples (Table 2B), eliminated competitively during outgrowth. Interestingly, two genes, *recR* and *rep,* appeared to be essential for IR survival as early as the first IR exposure. By passage 5, there are no insertions in these genes in the non-irradiated control samples. We hypothesize that these genes are essential for surviving IR, but also make a modest contribution to growth in rich media. These genes have been reported to be important for normal growth (*47*, *48*).

To further investigate the importance to general radiation resistance of the 47 genes identified here in Founder, TraDIS was performed on CB2000, a strain of *E. coli* previously reported to be highly radiation resistant (*26*). The 9 genes with the largest contributions to IR resistance (Table 2A) were also identified as top contributors in CB2000 after passage 1, in spite of the presence of all CB2000 mutations that confer an IR resistance phenotype. This result validates their importance for cell survival during radiation stress. A total of 37 of the reported genes for MG1655 (79%) were identified as important in CB2000 as well. Data was ambiguous for 6 of the 12 genes that were required for MG1655, but not CB2000 – this is likely due to the genetic bottlenecks that occur during passaging or slow growth of these mutants. Four genes, *pepP*, *rsxA*, crr, and *tatC* appear to be important for survival in wild type *E. coli* but not the directly

evolved CB2000. This suggests that the mutations arising in CB2000 render these four genes dispensable for IR survival.

3.3.4 IR resistance gene validations.

To directly verify a subset of genes identified as putative IR resistance genes by TraDIS, we separately deleted 14 of the 47 genes identified from wild type *E. coli* MG1655, and assayed each deletion mutant for survival following exposure to 1,000 and 2,000 Gy (Figure 2). All exhibited deficiencies that were easily documented with survival curves, helping to validate that the overall screen was identifying genes of interest. Based on our tests, these 14 deletion mutants were clustered into three different groups based on the overall decline in cell survival upon their deletion: critical (*recF*, *recG*, *recN*), severe (*uvrA*, *uvrB*, *yejH*, *uup*, *sbcB*), and moderate (*rdgC*, *ftsP*, *radA*, *rsxB*, *topB*, *recX*). Those gene products deemed critical produce a 3-5 log decline in survival at 2,000 Gy when deleted (Figure 2). The deletion mutants with severe deficiencies produced survival declines of 2-3 orders of magnitude, and those with moderate deficiencies produced declines of 1-2 orders of magnitude.

3.3.5 A strain deleted for yejH has phenotypes similar to DNA repair deficient mutants.

Of the genes listed in Table 2, 8 have no known function. One of these, *yejH*, is a putative DNA helicase. We focused some additional attention on this gene, in part to help validate our overall approach and in part to examine a possible DNA metabolism enzyme for which IR resistance is the first significant known phenotype. The YejH

protein shares sequence homology with XPB, the nucleotide excision repair helicase, conserved in eukaryotes and archea. By sequence analysis, the protein possesses the 7 helicase motifs central to superfamily 2 helicases in the N-terminal 350 amino acids. *E. coli* YejH is 25% identical and 40% similar to the 383 amino acid long helicase encoding region of human XPB with the defined helicase motifs being 62% identical and 71% similar (Figure 3). YejH does not encode the DNA damage recognition domain encoded by XPB. Thus, the DNA binding capability of YejH (data not shown) is likely encoded by the unique C-terminus. The cysteine rich C-terminus of YejH encodes two CxxC motifs, which could potentially coordinate a Fe-S cluster and thus work as sensor of oxidative stress or could constitute a zinc finger motif responsible for DNA binding.

The YejH protein clearly plays a role in survival when cells are exposed to high levels of IR. As illustrated in Figure 4, cells deleted for *yejH* rapidly dropped out of the mutant pool upon irradiation. Supporting this observation and lending validity to our screening technique, Figure 4B shows the deletion of *yejH* confers a nearly 3 orders of magnitude decrease in survival of the Founder strain at 2,000Gy. Furthermore, a strain deleted for *yejH* has a slight growth defect compared to a wild type strain in rich media. However, this slow growth phenotype is absent in minimal media (Figure 5A). Slow growth phenotypes in rich media have been associated with deficiencies in a variety of DNA repair functions (*55-58*). The faster metabolic rates in rich media have two

consequences that are less prominent during slower growth in minimal media: (1) there are more ROSs produced during metabolism (*59, 60*), and (2) because DNA metabolism (including DNA replication) is faster, more stalled replication forks occur that need to be repaired.

3.3.6 A yejH deletion strain takes longer to recover from IR treatment than a wild type.

A *yejH* deletion strain was assayed for its ability to recover from IR exposure compared to wild type. The Founder strain lacking a functional copy of *yejH* took *1.5* hours longer than wild type to start growing after exposure to 1,000 Gy, as shown in Figure 5B. In 3 separate trials, the lag time after IR treatment increased by essentially the same margin, from 219 minutes for wild type to 305 minutes for the *yejH* mutant in the example shown. Because IR produces DNA damage and *yejH* is a putative DNA metabolism gene via sequence analysis, these findings suggest that *yejH* is required for efficient DNA repair and cellular recovery. Upon mutating the Walker A lysine to alanine, cells again take longer to recover from IR treatment than wild type as seen for the *yejH* deletion (data not shown), indicating that the ATPase activity of YejH is required for its contribution to radiation resistance.

3.3.7 yejH is not epistatic with the nucleotide excision repair pathway.

Single deletions of *yejH*, *uup*, and *uvrAB* resulted in similar critical IR sensitivities as observed in Figure 2. This observation led us to construct double deletion mutations of *yejH* with *uup*, *uvrA*, and *uvrB*. Since YejH shares sequence similarity with a human

NER protein, we hypothesized that a *yejH uvrA* or *yejH uvrB* double mutant would be no more sensitive to IR than the single mutants. However, as shown in Figure 6A this was not the case. A *uvrA yejH* double mutant is substantially more sensitive than the two single mutants, indicating that we have removed two different mechanisms the cell relies upon for survival. The same result was obtained when the *yejH* and *uvrB* mutations were combined. In this case, the cells were so sensitized by these deletions that we did not recover any cells after irradiation with 2,000Gy.

3.3.8 yejH and uup appear to work together in a potentially new path to IR recovery.

A mutant deleted for both *yejH* and *uup* is just as sensitive to radiation as the single mutants as illustrated in Figure 6B. This result suggests that these two genes are working in the same path to IR recovery. Because neither of these genes have been linked to any major cellular function, and certainly not to DNA repair processes, we hypothesize these genes may define and constitute at least part of a radiation repair and/or recovery pathway not previously described. We have not shown a direct interaction between the genes products of *yejH* and *uup*, thus it is also possible that these two genes are not working together directly but rather that one gene works upstream or downstream from another in two different but required processes for IR recovery. Future work will further probe this observed genetic interaction.

3.3.9 The YejH protein interacts with single strand DNA binding protein, SSB.

To learn more about the role that YejH is playing in the cell, we looked for potential protein partners using a tandem affinity purification (TAP) method. We fused YejH to a TAP tag and expressed the tagged protein in wildtype *E. coli*. We purified the TAP tagged protein as previously described (*61, 62*) and submitted the associated proteins to mass spectroscopy analysis. The results, minus the ribosomal proteins, are reported in Figure 7a. The most prominent protein in the pull-down sample was the single-strand DNA binding protein (SSB). Thirteen unique SSB peptides were identified resulting in sequence coverage of 67%. The only protein with higher abundance was YejH itself. Various DNA repair enzymes such as RuvA, RecQ, TopB and TopA were identified in high abundance and are listed in Figure 7A. Whereas we don't know if YejH is interacting with any of these proteins directly, or indirectly via interactions with DNA and/or SSB, the apparent proximity of YejH to known repair enzymes again suggests that YejH is a new DNA repair enzyme.

To validate the potential YejH-SSB interaction, we exploited the convenient precipitation of SSB at very low concentrations of ammonium sulfate (*63*). As illustrated in the gel in Figure 7B, SSB and SSBΔC8 precipitate at 20% Ammonium sulfate but YejH does not. When YejH is incubated with SSB, SSB pulls YejH into the pellet and both bands are observed on the gel indicating the two proteins interact. When YejH is incubated with SSBΔC8, YejH is not pulled into the pellet and no YejH band is observed,

indicating that the interaction between YejH and SSB is tail dependent. The SSB C-terminal tail has been shown to interact with numerous DNA repair and replication proteins (*64*). Thus, this interaction lends support to our hypothesis that YejH is involved in DNA metabolism, possibly in a DNA repair role, and suggests an SSB-mediated mechanism for the localization of YejH to damaged DNA or the replication fork.

3.3.10 Contributions to IR resistance.

The list of gene functions required for survival of extreme IR exposure (Table 2) generally continues themes that were evident in the accompanying study that examines the genetic adaptations required for extreme resistance to IR (Chapter 2). In the overall list presented in Table 2, 20 of the 47 genes (or 43%) can be clearly defined as DNA repair or DNA metabolism functions. In accompanying report (Chapter 2) a small number of mutations in genes involved in DNA repair provide the major contributions to extreme IR resistance in one isolate (CB2000) derived from directed evolution. Adaptation of existing DNA repair systems provides one molecular explanation of the IR resistance phenotype. Not surprisingly, the pattern seen in Table 2 reinforces the theme that DNA repair functions are critical to survival of IR exposure. Whereas the idea that DNA is the major target of IR that results in lethality has been challenged in recent years (*20-24*) a range of key DNA repair systems must be intact in order for the cell to survive extreme IR exposure.

Many additional contributions are evident. An additional 7 of the genes identified (15%) have not been functionally characterized (we do not include *yejH* in this grouping) and represent a new class of genes to be studied for a role in surviving IR exposure. The remaining 20 genes cluster into 5 major categories as defined in Table 3, with roles in cell wall structure and biosynthesis (8), protein stability and turnover (5), cell division (4), oxidative stress signaling (3), and central metabolism (2). Three genes identified in the screen, but not included in Table 3, *trmH*, *yebC*, and *rlmL*, may contribute to IR resistance. However, they may have been identified as IR resistance genes due to possible polar effects on genes immediately downstream that are known to be involved in radiation survival – *recG*, *ruvC*, and *uup* respectively. We have not directly tested these insertions to confirm the presence of polar effects. However, each of these genes are in the same operon as and coexpressed with the indicated downstream genes. Of interest, five of the genes listed in Table 3, *pgi*, *prc*, *tatC, recF*, and *pepP*, are part of a broader network of 93 genes believed to play a role in promoting the stress-induced mutagenesis (SIM) response of *E. coli* K-12 (*65*).

The requirements for genes involved in protein stability and turnover, as well as those involved in oxidative stress signaling, can likely be understood in the context of current research from the Daly and Radman groups indicating that protein oxidation is a major deleterious effect of IR (*22*, *23*, *38*, *66*, *67*). As is the case for DNA, protein is a target of IR-mediated damage. Among the mutations underlying the acquired extreme

resistance to IR documented in an accompanying report (Chapter 2) mutations in *rsxB*, encoding part of a system that controls the cellular response to reactive oxygen species, and in *gsiB*, encoding a glutathione transporter, were fixed in population IR-2-20. Each makes a small but measurable contribution to the acquired IR resistance of evolved strain CB2000. The current study (Table 3) indicates that multiple cellular systems involved in ameliorating the effects of oxidative damage play a significant role in IR resistance.

The requirements for several genes involved in cell wall structure and biosynthesis as well as cell division, continue a theme seen in the accompanying report (Chapter 2). In CB2000, mutations in the genes *wcaK* and *nanE* again made small but measureable contributions to the acquired IR resistance (Chapter 2). These genes encode enzymes involved in the synthesis and/or recycling of peptidoglycan or surface polysaccharides. Essential genes involved in cell division, *ftsZ*, *ftsW, and ftsK*, were found to be mutational targets in our previous study as well (*26*). In the present study, we identified eight additional genes that contribute to cell wall structure, shape, and biosynthesis and six genes involved in cell division with three genes overlapping these two intrinsically linked cellular processes. After DNA repair, genes involved in maintaining the integrity of the cell structure and its proper division make up most of the remaining genes that we identified in our screen. One gene following into both of these categories, *mrcB*, encodes a transpeptidase that plays a role in peptidoglycan

synthesis and facilitates outermembrane constriction during cell division in concert with the Tol-Pal system. Strengthening this observation, we also observed central component of the Tol-Pal complex to be important for IR survival as well and to be mutational targets in our evolved project.

Numerous screen-identified genes cluster into the late stages of cell division indicating the importance of properly executed cell division in surviving exposure. We hypothesize that cells halt division to give themselves more time to repair their genomes. The presence of *slmA*, a gene that encodes a cell division inhibitor, on our list suggests that cells that survive radiation are aided by regulating their cell division during repair. Potentially, the loss of these cell division genes uncoordinates cell division regulation.

The importance of the bacterial cell wall and proper cell division as a target of IR-mediated damage has not yet been adequately assessed. The results of these two studies suggest that – in addition to DNA damage and protein inactivation via oxidation – the integrity of the bacterial cell wall, or particular substructures within it, represents a key factor in the overall lethality of IR. In each case, the gene functions identified in the new study can be dispensed with under normal growth conditions, but become important upon IR exposure. We can suggest at least three mechanisms that might be at work. First, the cell wall, particularly the peptidoglycan, could have substructures that are effectively weak points, particularly sensitive to damage inflicted by IR. Enzymes

involved in replacing such structures could become essential under stress. Second, there may be alterations to the peptidoglycan that are part of a general cellular response to stress that are critical to IR survival. Peptidoglycan plays an important role in osmotic regulation and cell shape (*68*). Changes to its structure accompany a number of cell stresses, including the nutritional stress that leads to the onset of stationary phase (*68*). Third, a somewhat different cell shape may be more optimal for IR survival in ways that are hard to predict. The *ompA* gene, mutated in the evolved project and found to contribute to survival in this study, has been proposed to have a role in mediating cell shape. Cells deleted for ompA have unstable outermembrane structures and cells tend to be spherical (*69*, *70*). Additionally, alterations in one of the 12 penicillin binding proteins that catalyze synthesis of peptidoglycan was documented in a long-term evolution experiment that increased cellular fitness in a particular medium, and this in turn produced alterations in cell shape (*71*). Additional mechanisms may be considered, and this list is not intended to be exhaustive.

Of the 47 genes identified in this study, 9 (19%) were mutated in one or more of the sequenced isolates characterized as part of our original directed evolution study (Chapter 2). As the present study indicates that loss of these gene functions results in significant radiation sensitivity, it is tempting to speculate that the mutations identified in the earlier study may be either neutral or gain of function. This conclusion must be tempered by the fact that each of the IR resistant strains in which the mutations appear

has dozens of additional mutant loci that could potentially act as functional suppressors. Of the nine, the nonsynonymous mutation in *rsxB* has already been discussed. A mutation upstream of *tolA* was fixed in another of the four separately evolved populations (IR-3-20) (Chapter 2) indicating that a change in the expression of this operon could be beneficial to survival. A nonsynonymous mutation was also fixed in *prc* in IR-3-20. While no mutations were found among the evolved isolates in *ftsN* or *ftsP*, mutations were common in *ftsW* and *ftsZ* among isolates of two sub-populations in IR-1-20, suggesting that alterations of the cell division process might contribute to IR resistance. Other mutations in our previous study were identified in *recG*, *dnaJ*, *ompA*, *tatC*, and *yejH*. The *yejH* and *dnaJ* mutations appeared to be fixed in the isolates taken from the further evolution of strain CB1000, and it is possible that the observed mutations in these two genes provide a useful gain of function in the context of extreme exposure to ionizing radiation. However, the mutations in *tatC, recG,* and *ompA* appeared in only one or a few IR resistant isolates, providing little in the way of a pattern to indicate that the mutations in these genes contributed significantly to IR resistance.

3.4 Discussion:

Combined with the Chapter 2 this work reveals a multi-tiered and nuanced cellular approach to surviving IR irradiation. Chapter 2 documents that enhancements to DNA repair processes can make major contributions to an extreme IR resistance phenotype acquired by directed evolution, even in a genetic background that is otherwise unaltered. At the same time, contributing mutations appear that provide potential enhancements to cellular systems for protein oxidation amelioration, protein folding and stability, cell division, and maintenance of cell wall structure and function. Roles for the same cellular functions are evident in the screen carried out here. A more complete description of the molecular basis of IR survival might thus consist of (a) robust DNA repair processes, (b) an enhanced capacity to prevent or ameliorate the effects of protein oxidation – including protein stabilization/re-folding, (c) an appropriate control of cell division to ensure that DNA repair can be completed, and (d) an enhancement of key processes affecting the structure and function of the cell wall.

The current study provides a general screen of gene functions that are not required for normal growth, but which become necessary when cells are exposed to high levels of ionizing radiation. We highlight two major conclusions. First, our screen indicates there are at least 47 genes required for cells to recover from IR exposure. As befits the need to repair IR damage to DNA, DNA repair functions predominate with 20 identified genes falling into this category. Several of these genes, particularly *recF*, *recN*,

and recG make critical contributions to survival. The results again highlight the importance of general DNA repair when cells are exposed to ionizing radiation. At the same time, the requirements for gene functions involved in protein structure stabilization and turnover, response to oxidative damage, and the maintenance of bacterial cell wall structure and function continue themes that were evident in the Chapter 2. Second we identified *yejH* as a significant contributor to survival of IR damage. This is a gene of previously unknown function not associated with any known phenotypes. Cells lacking a functional copy of *yejH* are nearly 100 times more sensitive to 2,000 Gy than wild type. This gene appears to be part of a potentially novel pathway directed at DNA repair. The gene *uup*, associated with the precise excision of transposons and associated with replication fork stability, may be part of the same pathway. The observation that YejH interacts with SSB provides a mechanism in which YejH is localized to damaged DNA or replication forks.

Our assessment of genes that contribute to IR survival implicated 7 genes of previously unknown function in the recovery of cells. By utilizing high-throughput screening techniques with simple organism such as *E. coli* in various growth conditions we can begin to identify the function of these enigmatic genes by identifying growth conditions or stress conditions where these genes become essential. A follow-up study on the cellular role of these 7 genes will begin to unravel the basis of their contribution and potentially define their cellular function. This in turn may help define the role of

their homologs in archaeans and eukaryotes. Of the seven genes, 3/7 have homologs identified in all 3 domains of life and all 7 have homologs in eukaryotes. In no instance has the function of one of these homologs been described.

Whereas this screen identified many new genes potentially involved in IR resistance, we specifically deleted 14 of the genes to study them further and to validate the TraDIS screening method. The deletion mutants were compared to an isogenic parent strain for survival to 1,000 and 2,000 Gy. All 14 mutants lacking these genes were sensitive to IR damage to some level. The 14 mutant strains were clustered into three sensitivity categories: critical, severe, and moderate.

Critical category. The genes with the most prominent effects on IR survival are known DNA repair enzymes. This category contains *recG*, *recF* and *recN*, genes central to recombinational DNA repair. We would likely see a similar result for a *recA* deletion but as discussed above, *recA* mutants were likely outcompeted in passaging and thus were categorized here as essential for growth.

Severe category. Deletion of the nucleotide excision repair pathway as well as exonuclease I (*sbcB*) had a major effect on survival as well. Genes of unknown function, *yejH* and *uup*, had a similar effect on the survival of IR exposure.

Moderate category. *The rdgC*, *radA*, *recX*, *topB*, and *ftsP* genes had mild but significant effects on survival. The *radA* gene has been previously linked to IR survival (*12, 72, 73*). The protein encoded by the *recX* gene is a negative regulator of the central recombinase,

RecA (*74-78*). DNA topoisomerase III is the product of *topB*, a topoisomerase that may play a role in reducing supercoiling for proper DNA repair upon damage (*79*). The product of the *ftsP* gene protects the cell division septum during stress (*80-83*). The RdgC protein binds to duplex DNA and exhibits some inhibitory effects on RecA function *in vitro* (*84, 85*), but its detailed molecular function is not understood.

Among the genes with no previously described phenotype, we were particularly interested in *yejH* because of the dramatic TraDIS profile, the IR sensitivity of cells lacking a functional copy, and its evident homology to the human gene encoding XPB. This gene clusters into the severe category with numerous genes of known function that have been studied for their roles in DNA repair for decades. Based on the gene sequence, our initial characterization (in particular the binding of YejH to SSB), and the new phenotype of the *yejH* gene described here, we hypothesize that YejH is involved in cellular repair of DNA after IR exposure. We propose assigning the name radiation resistance factor D, or *radD* to this gene.

3.5 Materials and Methods

*Bacterial strains and primers used in this study***.** All strains used in this study are *E. coli* K-12 derivatives and are listed in Table S1. Genetic manipulations were performed as previously described (*86*).

Transposome preparation. Transposon mutagenesis was performed using the EPICENTRE EZ-Tn5 transposition system, consisting of a transposase dimer conjugated to transposon DNA (*87*). The transposon is EZ-Tn5 <KAN-2>Tnp and was amplified using phusion polymerase (Stratagene). One hundred nanograms of this DNA were incubated with Tnp EK54/MA56/LP372, a hyperactive transposase with reduced target specificity (*88*), at room temperature for 3 hours. Transposome complexes were dialyzed against TE to remove all salt from the reaction before electroporation.

Preparation of electrocompetent cells for mutagenesis. Cells were cultured in Luria-Bertani (LB) broth (*89*) at 37° C with aeration to an OD600 of 0.4-0.6, chilled at 4° C for 30 minutes with stirring, harvested by centrifugation, and washed three times with ice cold 10% glycerol. In the final wash, cells were resuspended in 1/500 vol ice cold GYT media and stored at -80° C. One hundred microliters of cell suspension were mixed with 10uL of transposomes and electroporated in a 2-mm electrode gap cuvette with a GenePulser II (Bio-Rad). Cells were recovered in 1mL of SOC medium (*90*) and incubated at 37° C for 1 hour and then spread on plates containing 40mg/mL kanamycin and incubated overnight. The total number of colonies was estimated by counting several plates. The

colonies on each plate were pooled together sterilely in LB + 20% glycerol and stored at -80° C. Approximately 5 or more electroporations were performed per strain to generate an insertion mutant pool. The number of mutants per electroporation ranged from 20,000 to 175,000. By estimating the total number of mutants per batch, volumes containing similar number of mutants from each batch were pooled to create mutant libraries to contain 5×10^5 mutants.

IR treatment. An IR dose of 1000 Gy was applied iteratively to the mutant libraries with a ^{137}Cs source. Mutant pools were inoculated into 100mL of LB at an initial OD_{600} of 0.02 and were grown to an OD_{600} of 0.2. Cells were spun down and resuspended in 0.5 mL LB, IR treated, and then allowed to grow for approximately 7 generations to stationary phase before being used to inoculate the next cycle. This was repeated five times. Non-irradiated mock cultures were taken through all five passages in parallel but sat outside of the irradiator during treatment.

Fragment library sample preparation, sequencing, and data analysis. Genomic DNA was purified after each cycle of irradiation and from the non-irradiated cultures. DNA from the first and fifth cycle was sheared to an average size of 300 base pairs using hydroshear sonication. Preparation of the fragment library for sequencing was performed as described by Illumina, except the PCR amplification step was modified so that only sequences flanking transposons were amplified. Amplified fragment libraries were separated on an E-gel Size Select 2% agarose gel (Promega) and 270 bp fragments

were purified. The amplified DNA fragment libraries were sequenced on single end Illumina flow cells for 75 cycles using an Illumina Genome Analyzer IIx. The sequencing primer was modified to only sequence transposon containing DNA fragments. Sequence reads from the Illumina FASTQ FILES were separated into reads with tag and reads without. Reads containing the transposon tag sequence were retained for analysis. The 10bp tag was removed and then reads were trimmed to 50bp and mapped to the *E. coli* genome using BOWTIE (*91*), omitting insertion locations with less than 10 reads and allowing for 1 mismatch. All insertion locations in the first 1% and last 10% of gene regions were removed from further analysis. Reads per gene were normalized by total millions of reads collect for the sample to normalize for variation of total reads in different sequencing runs. Contribution values were calculated as the log ratio of reads in the irradiated sample $n_{g,B}$ to reads in non-irradated sample, $n_{g,A}$, for each gene, g $[\log(n_{g,B}/n_{g,A})]$. Genes were analyzed for the decrease in insertion locations per gene in a parallel analysis. Genes in Table 2 had a contribution value of -.5 to -2, indicating a loss of 1 to 2 logs of reads in the irradiated vs non-irradiated condition. When genes were found to have a low Contribution Value yet a large decrease in insertion locations, the genes were check for single insertion locations with over 1,000 reads which can be artifacts of library amplification. Genes listed in Table 2 that were discovered by loss of insertion locations rather than decreased read counts were in the top 99.999 percentile of genes for insertion location losses.

IR survival assays for gene validation. Of the 47 genes discovered, 14 genes were verified by deleting the gene from the wild type strain using the Wanner Method (*86*). Deletions strains were tested for their ability to survive increasing doses of ionizing radiation in comparison to an isogeneic wild type strain. All strains were tested in biological triplicate. Cells from a fresh single colony of each strain were cultured in Luria-Bertani (LB) broth (*89*) at 37° C with aeration. After growth overnight, cultures were diluted 1:1000 into 10 mL fresh LB broth in 125 mL flasks and grown at 37° C with shaking until an optical density (OD_{600}) of ~0.4 was reached. For each sample, 15mL of culture was spun down and resuspended in 0.8mL of fresh LB. 100uL were set on ice as the non-irradiated control and the other 700uL were irradiated in a Mark I ^{137}Cs irradiator (from J. L. Shepherd and Associates) for times corresponding to 1 and 2Gy (~7 Gy/min). Irradiated samples as well as the non-irradiated control samples for each culture were diluted appropriately, and plated on LB 15% agar medium to determine the total number of colony forming units (CFUs). Percent survival was calculated by dividing the titer of the surviving population by the titer of the non-irradiated control sample. For each strain, 3-5 biological replicates were preformed.

Growth curves. Cells of each strain were cultured in Luria-Bertani (LB) broth (*89*) at 37° with aeration. After growth overnight, cultures were diluted 1:250 into 0.5 mL fresh LB broth or M9 minimal media with glucose in 24 well plates and grown at 37° with shaking in a plate reader (Bio-Tek). The optical density (OD_{600}) was measured every

20 minutes. In a separate experiment, samples were grown to mid-log phase, treated with 1,000Gy and then inoculated as described above. Each strain tested was analyzed in biological triplicate and the error was reported as SEM.

Tandem Affinity Purification. The open reading frame of *E. coli yejH* was amplified by PCR and subcloned in-frame into cloning vector pCN70[3] to produce an expression vector encoding YejH with an N-terminal dual affinity tag (includes protein A and calmodulin peptide binding domains separated by a tobacco etch virus protease cleavage site). *E. coli* K12 strain MG1655 (DE3) transformed with pTAP-YejH was grown at 37 °C in 4 liters of Luria-Bertani medium supplemented with 50 μg/ml ampicillin to midlog phase ($A_{600\ nm}$ of ~0.5), induced by the addition of 1 μM isopropyl 1-thio-β-D-galactopyranoside, and grown for an additional 3 h. Cells were harvested by centrifugation, suspended in 50 ml of Nonidet P-40 buffer (6 mM dibasic sodium phosphate, 4 mM monobasic sodium phosphate, 150 mM NaCl, 2 mM EDTA, 50 mM NaF, 4 mg/liter leupeptin, 0.1 mM sodium vanadate, 19.5 mg/liter benzamidine, 8.7 mg/liter phenylmethylsulfonyl fluoride (PMSF), 1% Nonidet P-40 substitute), and lysed by sonication. Tandem affinity purification was performed as described previously (*92*) except the TCA precipitation pellet was resuspended in buffer, digested with trypsin, and subjected to MALDI-TOF mass spectrometry for identification of peptides (University of Wisconsin Mass Spectrometry facility).

Ammonium sulfate co-precipitation. Co-precipitation experiments were performed as described previously (*63*)and pellet fractions were suspended in 30 µl of loading buffer prior to SDS-PAGE on 4-15% polyacrylamide gradient gels (Bio-Rad).

Figures

Figure 1: TraDIS was used to identify genes that contribute to IR resistance.

A. IR treatment and passaging of mutant pool. The mutant library consisting of 500,000 insertion mutants was grown to mid-log phase. A non-treated sample was taken and the rest of the library was irradiated with a dose 1000 Gy before both samples were diluted into fresh media for competitive outgrowth. After outgrowth, genomic DNA was extracted from both the irradiated and non-irradiated samples (passage 1 results). Outgrown cultures were then used as the inoculum for the next passage of IR treatment. This was repeated five times. After each treatment passage, genomic DNA was extracted for TraDIS analysis. **B.** TraDIS profile for *uvrB*. Detailed plots generated used MochiView show the frequency and distribution of transposon directed insertion-site sequences across a chromosomal region containing *uvrB* for a pool of 500,00 transposon mutants. No IR indicates the mutant pool was subjected to the permissive non-irradiated condition and +IR indicates the pool was subjected to 1000Gy. Data is presented after 1 and 5 passages of irradiation and outgrowth. The y-axis shows the number of mapped sequence reads and the position of annotated genes relative to the plotted sequence reads is indicated below the distribution plot.

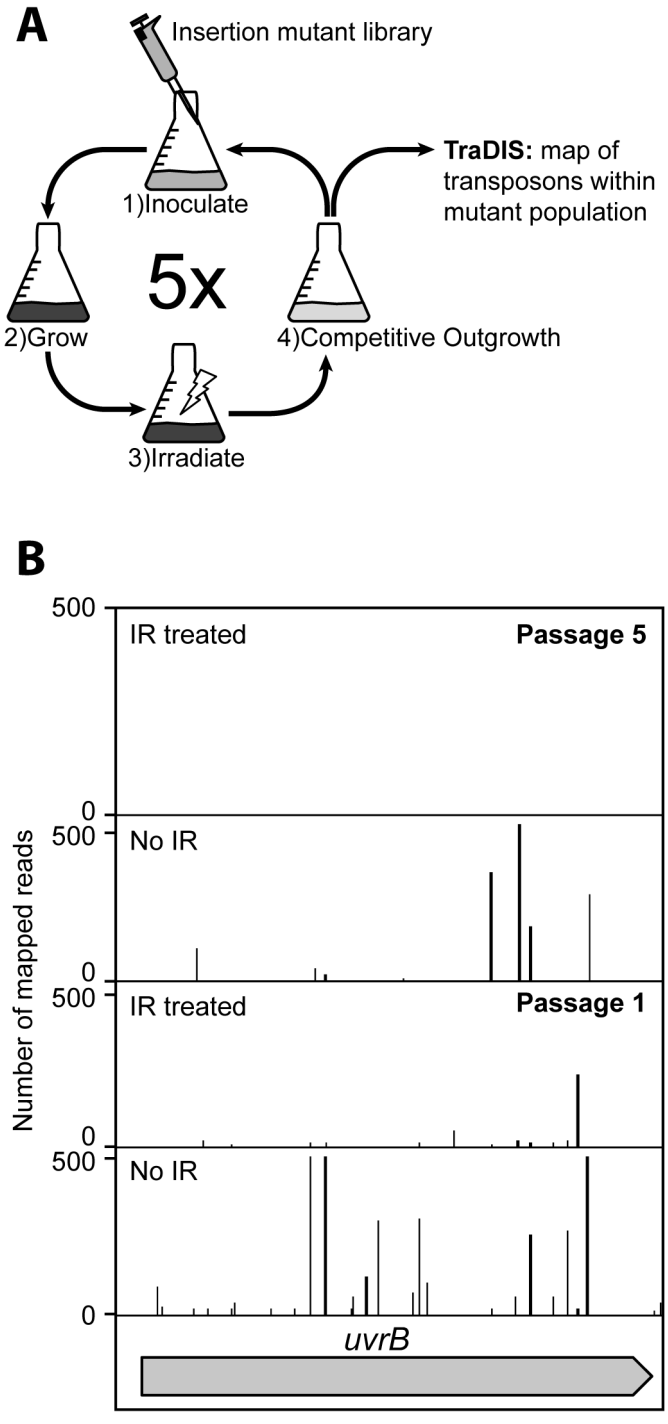

A
Insertion mutant library
1)Inoculate
TraDIS: map of transposons within mutant population
5x
2)Grow
4)Competitive Outgrowth
3)Irradiate
B
Number of mapped reads
500
IR treated
Passage 5
0
500
No IR
0
500
IR treated
Passage 1
0
500
No IR
0
uvrB

Figure 2: Effects of gene deletion on survival of *E. coli* to increasing doses of radiation.

Fourteen genes identified using TraDIS were individually deleted in MG1655 and assayed for survival to 1,000 and 2,000 Gy. Gene deletions strains appear to cluster into three sensitivity groups, Moderate (M), Severe (S), and Critical (C).

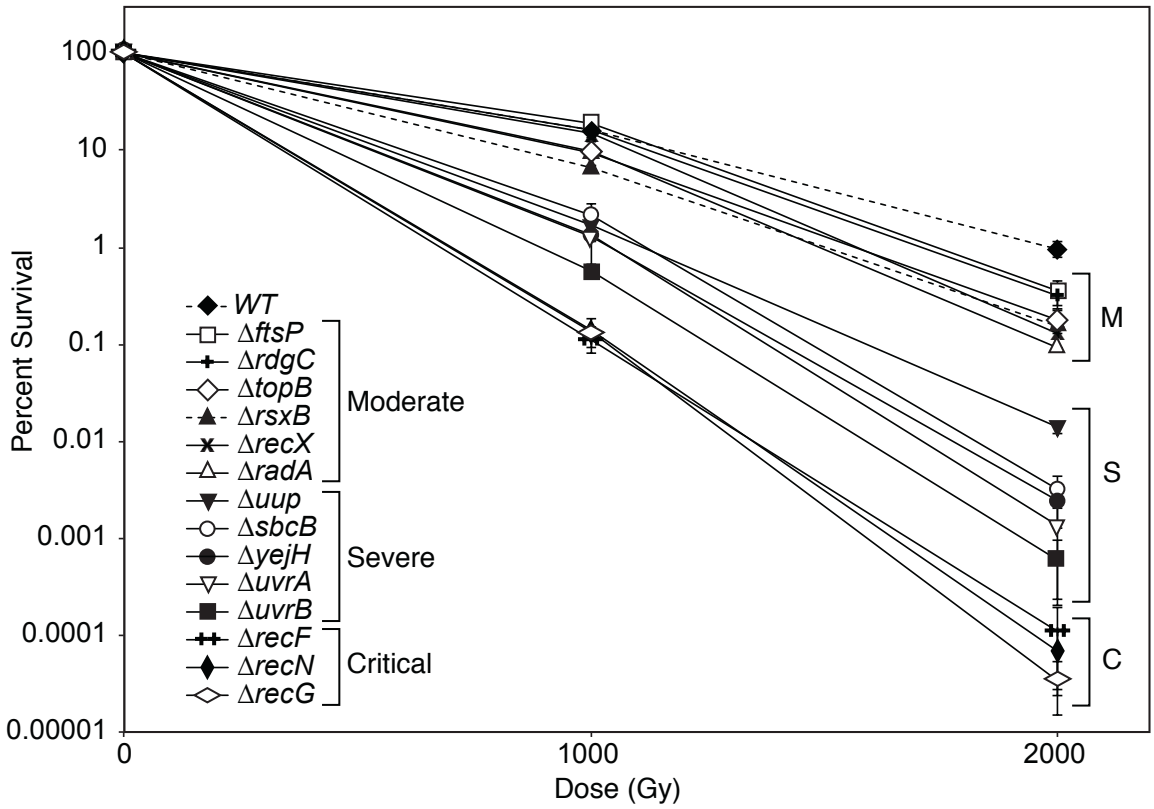

100
10
1
0.1
0.01
0.001
0.0001
0.00001
Percent Survival
0
1000
2000
Dose (Gy)
WT
ΔftsP
ΔrdgC
ΔtopB
ΔrsxB
ΔrecX
ΔradA
Moderate
Δuup
ΔsbcB
ΔyejH
ΔuvrA
ΔuvrB
Severe
ΔrecF
ΔrecN
ΔrecG
Critical
M
S
C

Figure 3: Sequence alignment of YejH homologs from *Archaeoglobus fulgidus* and *Homo sapiens*, the species from which XPB structures have been solved (*93*, *94*).

A. Schematic alignment between *Ec*YejH, *Af*XPB, and *Hs*XPB. Shown are the XPB damage recognition domain (open blue rectangles) the 7 conserved helicase motifs, I-VI, (yellow), the ThM domain (closed blue rectangles), and the 2 CxxC motifs in the YejH C-terminus (olive). **B.** Sequence alignment between the helicase containing regions of *Ec*YejH, *Hs*XPB, and *Af*XPB. Alignments were carried out by Clustal Omega (*95*). Invariant (*) and conserved (: or .) residues are marked. The seven conserved helicase motifs are boxed in yellow.


EcYejH
AfXPB
HsXPB
I
Ia
II
III
IV
V
VI
CxxC
CxxC
1
586
1
452
1
243
686
782
XPB:DRD
Conserved helicase motifs
YejH: Unique C-term
EcYejH 4 LRPYQQEAVDATLNHFRRHKTPAVIVLPTGAGKSLVIAELARLARGRVLVLAHVKELVAQ 64
AfXPB 74 LRDYQEKALERWLVD-----KRGCIVLPTGSGKTHVAMAAINELSTPTLIVVPTLALAEQ 128
HsXPB 316 LRPYQEKSLRKMFGNG--RARSGVIVLPCGAGKSLVGVTAACTVRKRCLVLGNSAVSVEQ 373
EcYejH NHAKYQALGLEADI----FAAGLKRKESHGK----------VVFGSVQSVARNLDAFQ 108
AfXPB WKERLGIFGEEYVG---EFSGRIKELKPLTV----ST----------YDSAYVNAEKLG 170
HsXPB WKAQFKMWSTIDDSQICRFTSDAK-DKPIGCSVAISTYSMLGHTTKRSWEAERVMEWLKT 432
EcYejH GEFSLLIVDECHRIGDDEESQYQQILTHLTKVNPHLRLLGLTATPFRLGKGWIYQFHYHG 168
AfXPB NRFMLLIFDEVHHLPAESYVQIAQMSIA-------PFRLGLTATFEREDGRHE------- 216
HsXPB QEWGLMILDEVHTIPAKMFRRVLTIVQA-------HCKLGLTATLVREDDKIV------- 478
EcYejH MVRGDEKALFRDCIYELPLRYMIKHGYLTP----------PERLDMPVVQYDFSRL-Q 215
AfXPB ----ILKEVVGGKVFELFPDSLA-GKHLAKYTIKRIFVPLAEDERVEYEKREKVYKQFLR 271
HsXPB ----DLNFLIGPKLYEANWMELQNNGYIAKVQCAEVWCPMSPEFYREYVAIK-------- 526
EcYejH AQSNGLFSEADLN----------------RELKKQQRIT-----PHII---SQIMEFAAT 261
AfXPB ARGITLRRAEDFNKIVMASGYDERAYEALRAWEEARRIAFN--SKNKIRKLR--EILERH 327
HsXPB ----------TKKRILLYTMNPNKFRACQFLIKFHER 553
EcYejH -RKGVMIFAATVEHAKEIVGLLPAEDAALITGDTPGAERDVLIENFKAQ-RFRYLVNVAV 309
AfXPB RKDKIIIFTRHNELVYRISKV---FLIPAITHRTSREEREEILEGFRTG-RFRAIVSSQV 383
HsXPB RNDKIIVFADNVFALKEYAIR---LNKPYIYGPTSQGERMQILQNFKHNPKINTIFISKV 610
EcYejH LTTGFDAPHVDLIAILRP-TESVSLYQQIVGRGLRLAPGKTDC----LILDYAGNPHDLY 364
AfXPB LDEGIDVPDANVGVIMSG-SGSAREYIQRLGRILRPSKGKK----EAVLYELISRG---- 434
HsXPB GDTSFDLPEANVLIQISSHGGSRRQEAQRLGRVLRAKKGMVAEEYNAFFYSLVSQD---- 666
EcYejH APEVGTPKGKS 375
AfXPB TGEVNTARRRK 449
HsXPB TQEMAYSTKRQ 677

Figure 4: *yejH* contributes to IR resistance.

A. TraDIS profile for *yejH*. Detailed plot generated used MochiView show the frequency and distribution of transposon directed insertion-site sequences across a chromosomal region containing *yejH* for a pool of 500,00 transposon mutants. No IR indicates the mutant pool was subjected to the permissive non-irradiated condition and +IR indicates the pool was subjected to 1000Gy. The y-axis shows the number of mapped sequence reads and the position of annotated genes relative to the plotted sequence reads is indicated below the distribution plot. **B.** Effects of *yejH* deletion on survival of radiation exposure. The *yejH* gene was individually deleted in MG1655 and assayed for survival to 1,000 and 2,000 Gy and compared to an isogenic parent strain.

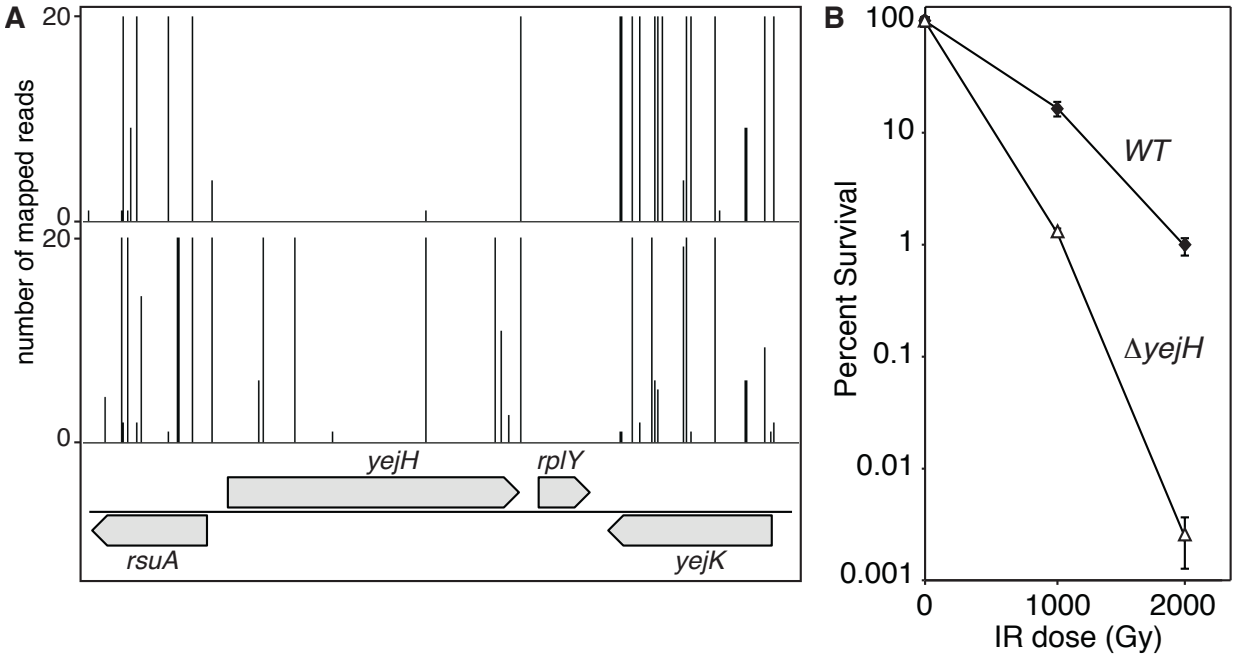

A
20
0
20
0
number of mapped reads
yejH
rplY
rsuA
yejK
B
100
10
1
0.1
0.01
0.001
Percent Survival
WT
ΔyejH
0
1000
2000
IR dose (Gy)

Figure 5. A *yejH* deletion strain has a growth defect in rich media that is absent in minimal media and takes 2 hrs longer than wild type to start growing after treatment with 1,000 Gy.

A. Growth was monitored over time as optical density at 600nm. Strains were grown at 37° in the indicated media. Strains shown are MG1655 (closed squares) and MG1655 ΔyejH (open circle). Experiments were performed in biological triplicate. Error bars represent SEM. **B.** Cells were irradiated with a dose of 1,000 Gy before growth was monitored in rich media as described for A.

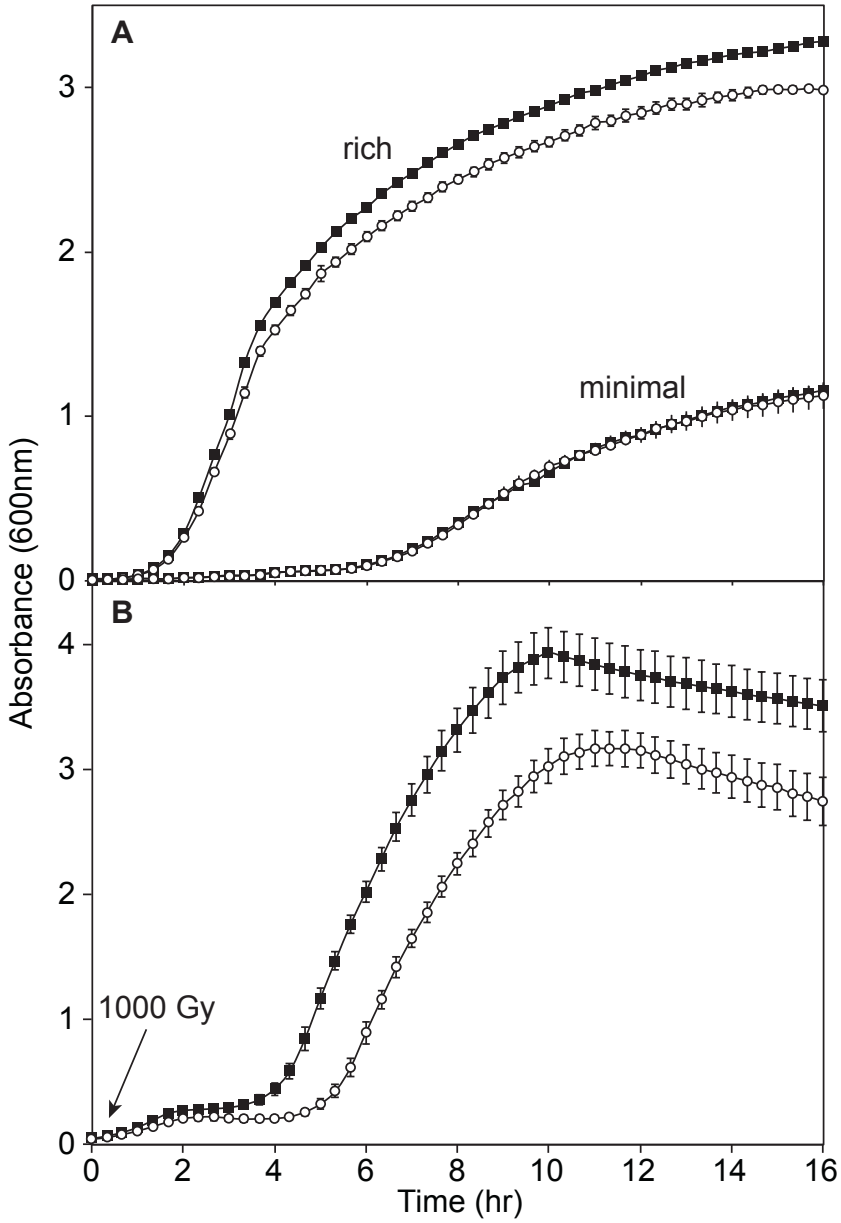

A
B
rich
minimal
1000 Gy
Absorbance (600nm)
Time (hr)
0
1
2
3
4
0
2
4
6
8
10
12
14
16

Figure 6. Epistasis studies.

Effects of selected gene deletion and double gene deletions on survival of *E. coli* to increasing doses of radiation. Mid-log phase cultures were irradiated with increasing doses of radiation and plated to measure survival as described in Materials and Methods. **A.** Epistasis studies with *uvrABC*. **B.** Epistasis studies with *uup*.

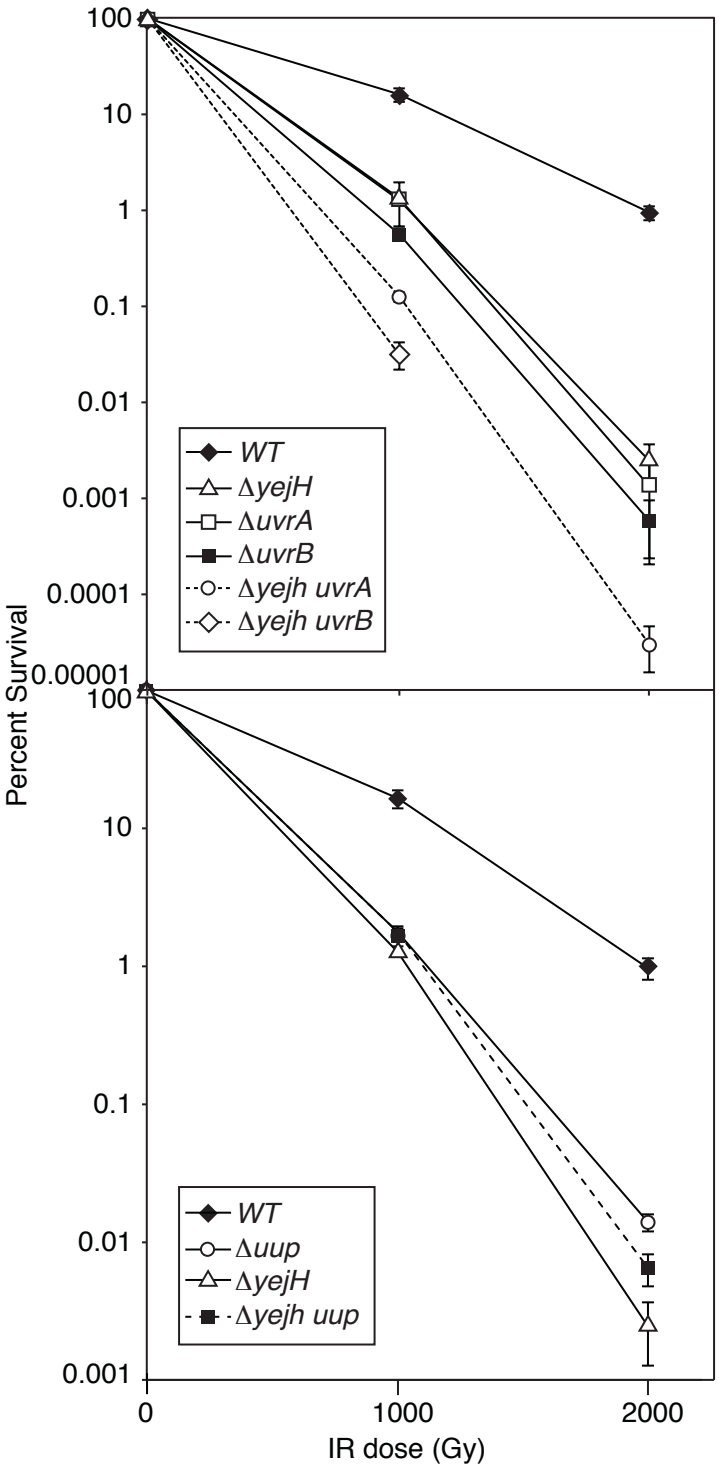

100
10
1
0.1
0.01
0.001
0.0001
0.00001
WT
ΔyejH
ΔuvrA
ΔuvrB
Δyejh uvrA
Δyejh uvrB
100
10
1
0.1
0.01
0.001
WT
Δuup
ΔyejH
Δyejh uup
Percent Survival
0
1000
2000
IR dose (Gy)

Figure 7: YejH interaction studies.

A. TAP purification of YejH-containing protein complexes from *E. coli.* Results are top hits from MALDI-TOF mass spectrometry analysis of protein pellet from final elution of TAP-YejH and co-purifying proteins. Ribosomal proteins were removed from list. **B.** Ammonium sulfate co-precipitation of YejH with SSB and SSBΔC8. 150 g/liter ammonium sulfate was incubated with YejH, SSB, and SSB ΔC8 in reactions as indicated by + symbols and the pellets were washed and analyzed by SDS-PAGE.

A

Gene	Description	MW	Unique peptides	Total spectra	Sequence coverage
SSB	single-stranded DNA-binding protein	18,974	13	1508	61%
YejH	predicted ATP-dependent DNA or RNA helicase	66,414	44	1120	67%
TnaA	tryptophanase/L-cysteine desulfhydrase	52,775	16	136	46%
RuvA	component of RuvABC resolvasome, regulatory subunit	22,086	6	73	30%
TopA	DNA topoisomerase I, omega subunit	97,351	20	48	38%
TopB	DNA topoisomerase III	73,218	16	43	33%
Dps	Fe-binding and storage protein	18,696	9	37	61%
RpoB	RNA polymerase, beta subunit	150,636	22	29	28%
RpoC	RNA polymerase, beta prime subunit	155,164	23	28	30%
SeqA	regulatory protein for replication initiation	20,316	6	26	54%
H-NS	global DNA-binding transcriptional dual regulator H-NS	15,540	4	25	38%
AhpC	alkyl hydroperoxide reductase, C22 subunit	20,762	7	25	54%
RecQ	ATP-dependent DNA helicase	68,365	12	25	29%

B

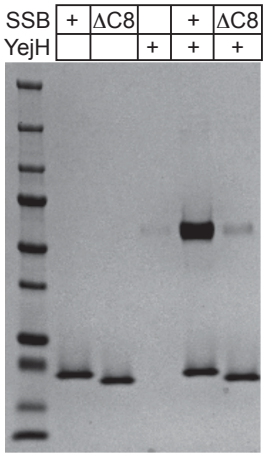


Table 1: General sequencing results.

	IR Treatment	Total number of reads	Number of unique insert sites	Average Distance between inserts (bp)	Gene length cutoff
Wild-type (Founder)					
Passage 1	–	22,778,780	90,920	51.0	122.9
Passage 5	–	33,338,903	25,590	181.3	436.6
Passage 1	+	30,935,721	125,644	36.9	88.9
Passage 5	+	19,557,186	39,185	118.4	285.1

Table 2: Genes found to contribute to survival of radiation exposure.

A. Genes with the largest contribution to radiation resistance after 5 passages of irradiation and outgrowth. Contribution was calculated as described in material and methods. Genes listed here had the largest decrease in reads upon irradiation when compared to the nonirradiated control, with contribution factors between -1 and -2 logs. The * indicates that the given gene was discovered by analyzing loss of unique insertion locations. Validation in this study indicates the gene was assayed for contribution to IR resistance as shown in Figure 2. **B.** Genes identified to be important in passage 1 with no data for passage 5 in either the non-irradiated control or the irradiated sample. Cells lacking these genes are likely outcompeted due to growth defects associated with the deletion of these genes.

A Effect on IR survival

	Wild type		CB2000		
Gene	Passage 1	Passage 5	Passage 1	Passage 5	Validation
recN					This Study✓
trmH					
uvrB					This Study✓
uvrA					This Study✓
mrcB				OC	
recD					(Thomas, 1999)
recF					This Study✓
uvrC	*	*			
yebC				OC	
radA					This Study✓
slmA					
sbcB					This Study✓
endA					
phr					
yafC				OC	
uup					This Study✓
speA					
gph					
rdgC					This Study✓
ybjN				OC	
yejH					This Study✓
yqiA					
yhgF					
topB					This Study✓
rlmL					
yabI					
pgi					
recX				OC	This Study✓
ybgI					

B Effect on IR survival

	Wild type	CB2000	
Gene	Passage 1	Passage 1	Validation
recG			This Study ✓
recO	*		
recR			
uvrD	**		
dnaJ	*		
rsxB			This Study ✓
rsxA			
ompA			
crr			
rfaC			
tolA			
tatC			
pstS			
pepP			
prc			
rep		OC	
ftsN	*		
ftsP			This Study✓

Highest contribution to IR survival

No contribution to IR survival

Table 3: Clustering of identified genes by cellular function.

Cellular Function	No. of genes clustered	% of identified genes	Genes
DNA metabolism	20	43%	*recN , uvrABCD , recD , recF*, recO , sbcB , endA , phr, uup , gph , rdgC , yejH , topB , recX , recR , rep , radA*
Cell wall structure and biosynthesis	8	17%	*rfaC, prc*, ompA, tolA, pstS, tatC*, crr, mrcB**
Unknown	7	15%	*yafC , ybjN , yqiA , yhgF , yabI , ybgI , yebC*
Cell division	4	9%	*mrcB *, slmA , ftsN , ftsP*
Oxidative stress signaling	3	6%	*pgi*, rsxAB*
Protein stability and turnover	3	6%	*pepP*, prc*, dnaJ*
Central metabolism	2	4%	*pgi*, speA*
SIM response	5	11%	*pgi*, prc*, tatC*, recF*, pepP**

* indicated gene is listed more than once

References

1. Cox, M. M., and Battista, J. R. (2005) Deinococcus radiodurans - The consummate survivor, *Nature Reviews Microbiology 3*, 882-892.

2. Blasius, M., Hubscher, U., and Sommer, S. (2008) Deinococcus radiodurans: What belongs to the survival kit?, *Critical Reviews in Biochemistry and Molecular Biology 43*, 221-238.

3. Collins, C., Zhou, X., Wang, R., Barth, M. C., Jiang, T., Coderre, J. A., and Dedon, P. C. (2005) Differential oxidation of deoxyribose in DNA by gamma and alpha-particle radiation, *Radiation research 163*, 654-662.

4. Mikkelsen, R. B., and Wardman, P. (2003) Biological chemistry of reactive oxygen and nitrogen and radiation-induced signal transduction mechanisms, *Oncogene 22*, 5734-5754.

5. Bresler, S. E., Noskin, L. A., Kuzovleva, N. A., and Noskina, I. G. (1979) The nature of the damage to Escherichia coli DNA induced by gamma-irradiation, *International journal of radiation biology and related studies in physics, chemistry, and medicine 36*, 289-300.

6. Cahill, D., Connor, B., and Carney, J. P. (2006) Mechanisms of eukaryotic DNA double strand break repair, *Frontiers in Bioscience 11*, 1958-1976.

7. Sugawara, N., and Haber, J. E. (2006) Repair of DNA double strand breaks: In vivo biochemistry, In *DNA Repair, Pt A*, pp 416-429.

8. Eckardt-Schupp, F., and Klaus, C. (1999) Radiation inducible DNA repair processes in eukaryotes, *Biochimie 81*, 161-171.

9. Hicks, W. M., Yamaguchi, M., and Haber, J. E. (2011) Real-time analysis of double-strand DNA break repair by homologous recombination, *Proc Natl Acad Sci U S A 108*, 3108-3115.

10. Krogh, B. O., and Symington, L. S. (2004) Recombination proteins in yeast, *Annual Review of Genetics 38*, 233-271.

11. Sargentini, N. J., and Smith, K. C. (1985) Growth-medium-dependent repair of DNA single-strand and double-strand breaks in X-irradiated Escherichia coli, *Radiation research 104*, 109-115.

12. Sargentini, N. J., and Smith, K. C. (1986) Quantitation of the involvement of the recA, recB, recC, recF, recJ, recN, lexA, radA, radB, uvrD, and umuC genes in the repair of X-ray-induced DNA double-strand breaks in Escherichia coli, *Radiation research 107*, 58-72.

13. Sargentini, N. J., and Smith, K. C. (1986) Characterization and quantitation of DNA strand breaks requiring recA-dependent repair in X-irradiated Escherichia coli, *Radiation research 105*, 180-186.

14. Sargentini, N. J., and Smith, K. C. (1989) Role of ruvAB genes in UV- and gamma-radiation and chemical mutagenesis in Escherichia coli, *Mutation research 215*, 115-129.

15. Daly, M. J., Ouyang, L., Fuchs, P., and Minton, K. W. (1994) In vivo damage and recA-dependent repair of plasmid and chromosomal DNA in the radiation-resistant bacterium Deinococcus radiodurans, *Journal of bacteriology 176*, 3508-3517.

16. Krisch, R. E., Krasin, F., and Sauri, C. J. (1978) DNA breakage, repair, and lethality accompanying 125I decay in microorganisms, *Current Topics in Radiation Research Quarterly. 12*, 355-368.

17. Minton, K. W., and Daly, M. J. (1995) A model for repair of radiation-induced DNA double-strand breaks in the extreme radiophile Deinococcus radiodurans, *BioEssays : news and reviews in molecular, cellular and developmental biology 17*, 457-464.

18. Repar, J., Cvjetan, S., Slade, D., Radman, M., Zahradka, D., and Zahradka, K. (2010) RecA protein assures fidelity of DNA repair and genome stability in Deinococcus radiodurans, *DNA repair 9*, 1151-1161.

19. Slade, D., Lindner, A. B., Paul, G., and Radman, M. (2009) Recombination and Replication in DNA Repair of Heavily Irradiated Deinococcus radiodurans, *Cell 136*, 1044-1055.

20. Daly, M. J., Gaidamakova, E. K., Matrosova, V. Y., Vasilenko, A., Zhai, M., Leapman, R. D., Lai, B., Ravel, B., Li, S. M., Kemner, K. M., and Fredrickson, J. K. (2007) Protein Oxidation Implicated as the Primary Determinant of Bacterial Radioresistance, *PLoS biology 5*, e92.

21. Fredrickson, J. K., Li, S. M., Gaidamakova, E. K., Matrosova, V. Y., Zhai, M., Sulloway, H. M., Scholten, J. C., Brown, M. G., Balkwill, D. L., and Daly, M. J. (2008) Protein oxidation: key to bacterial desiccation resistance?, *Isme Journal 2*, 393-403.

22. Slade, D., and Radman, M. (2011) Oxidative Stress Resistance in Deinococcus radiodurans, *Microbiology and Molecular Biology Reviews 75*, 133-+.

23. Daly, M. J. (2009) A new perspective on radiation resistance based on Deinococcus radiodurans, *Nat Rev Microbiol 7*, 237-245.

24. Daly, M. J. (2012) Death by protein damage in irradiated cells, *DNA repair 11*, 12-21.

25. Daly, M. J., Gaidamakova, E. K., Matrosova, V. Y., Kiang, J. G., Fukumoto, R., Lee, D. Y., Wehr, N. B., Viteri, G. A., Berlett, B. S., and Levine, R. L. (2010) Small-molecule antioxidant proteome-shields in Deinococcus radiodurans, *PloS one 5*, e12570.

26. Harris, D. R., Pollock, S. V., Wood, E. A., Goiffon, R. J., Klingele, A. J., Cabot, E. L., Schackwitz, W., Martin, J., Eggington, J., Durfee, T. J., Middle, C. M., Norton, J. E., Popelars, M., Li, H., Klugman, S. A., Hamilton, L. L., Bane, L. B., Pennacchio, L., Albert, T. J., Perna, N. T., Cox, M. M., and Battista, J. R. (2009) Directed evolution of radiation resistance in Escherichia coli, *Journal of bacteriology 191*, 5240-5252.

27. Liu, Y. Q., Zhou, J. Z., Omelchenko, M. V., Beliaev, A. S., Venkateswaran, A., Stair, J., Wu, L. Y., Thompson, D. K., Xu, D., Rogozin, I. B., Gaidamakova, E. K., Zhai, M., Makarova, K. S., Koonin, E. V., and Daly, M. J. (2003) Transcriptome dynamics of Deinococcus radiodurans recovering from ionizing radiation, *Proc Natl Acad Sci U S A 100*, 4191-4196.

28. Zhang, C., Wei, J., Zheng, Z., Ying, N., Sheng, D., and Hua, Y. (2005) Proteomic analysis of Deinococcus radiodurans recovering from gamma-irradiation, *Proteomics 5*, 138-143.

29. Konrad, E. B. (1977) Method for the isolation of Escherichia coli mutants with enhanced recombination between chromosomal duplications, *Journal of bacteriology 130*, 167-172.

30. Mahdi, A. A., and Lloyd, R. G. (1989) Identification of the recR locus of Escherichia coli K-12 and analysis of its role in recombination and DNA repair, *Molecular & General Genetics 216*, 503-510.

31. Volkert, M. R., and Nguyen, D. C. (1984) Induction of specific Escherichia coli genes by sublethal treatments with alkylating agents, *Proc Natl Acad Sci U S A 81*, 4110-4114.

32. Kolodner, R., Fishel, R. A., and Howard, M. (1985) Genetic recombination of bacterial plasmid DNA: effect of RecF pathway mutations on plasmid recombination in Escherichia coli, *Journal of bacteriology 163*, 1060-1066.

33. Ohta, T., Sutton, M. D., Guzzo, A., Cole, S., Ferentz, A. E., and Walker, G. C. (1999) Mutations affecting the ability of the Escherichia coli UmuD ' protein to participate in SOS mutagenesis, *Journal of bacteriology 181*, 177-185.

34. Clark, A. J., and Margulies, A. D. (1965) Isolation and characterization of recombination-deficient mutants of *Escherichia coli* K12, *Proc. Natl. Acad. Sci. U.S.A. 53*, 451-459.

35. Howard-Flanders, P. (1968) DNA repair. , *Annual Review of Biochemistry 37*, 175-200.

36. Modrich, P. (1987) DNA mismatch correction, *Annu Rev Biochem 56*, 435-466.

37. Lombardo, M. J., and Rosenberg, S. M. (2000) radC102 of Escherichia coli is an allele of recG, *Journal of bacteriology 182*, 6287-6291.

38. Krisko, A., and Radman, M. (2010) Protein damage and death by radiation in Escherichia coli and Deinococcus radiodurans, *Proc Natl Acad Sci U S A 107*, 14373-14377.

39. Langridge, G. C., Phan, M. D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E., Dougan, G., Wain, J., Parkhill, J., and Turner, A. K. (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants, *Genome Res 19*, 2308-2316.

40. Gallagher, L. A., Shendure, J., and Manoil, C. (2011) Genome-scale identification of resistance functions in Pseudomonas aeruginosa using Tn-seq, *mBio 2*, e00315-00310.

41. Gawronski, J. D., Wong, S. M., Giannoukos, G., Ward, D. V., and Akerley, B. J. (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung, *Proc Natl Acad Sci U S A 106*, 16422-16427.

42. Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., Knight, R., and Gordon, J. I. (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat, *Cell host & microbe 6*, 279-289.

43. van Opijnen, T., Bodi, K. L., and Camilli, A. (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms, *Nature methods 6*, 767-772.

44. Blattner, F. R., Plunkett, G. r., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado, V. J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome sequence of Escherichia coli K-12, *Science 277*, 1453-1474.

45. Balazsi, G. (2008) Statistical evaluation of genetic footprinting data, *Methods Mol Biol 416*, 355-359.

46. Gerdes, S., Edwards, R., Kubal, M., Fonstein, M., Stevens, R., and Osterman, A. (2006) Essential genes on metabolic maps, *Current opinion in biotechnology 17*, 448-456.

47. Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balazsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I., Gelfand, M. S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M. V., Grechkin, Y., Mseeh, F., Fonstein, M. Y., Overbeek, R., Barabasi, A. L., Oltvai, Z. N., and Osterman, A. L. (2003) Experimental determination and system level analysis of essential genes in Escherichia coli MG1655, *Journal of bacteriology 185*, 5673-5684.

48. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection, *Mol Syst Biol 2*, 2006 0008.

49. Brutinel, E. D., and Gralnick, J. A. (2012) Anomalies of the anaerobic tricarboxylic acid cycle in Shewanella oneidensis revealed by Tn-seq, *Mol Microbiol 86*, 273-283.

50. Linn, S., and Imlay, J. A. (1987) Toxicity, mutagenesis and stress responses induced in Escherichia coli by hydrogen peroxide., *J Cell Sci Suppl 6*, 289-301.

51. Miguel, A. G., and Tyrrell, R. M. (1986) Repair of near-ultraviolet (365 nm)-induced strand breaks in Escherichia coli DNA. The role of the polA and recA gene products., *Biophysical journal 49*, 485-491.

52. Morse, L. S., and Pauling, C. (1975) Induction of error-prone repair as a consequence of DNA ligase deficiency in Escherichia coli., *Proc Natl Acad Sci U S A 72*, 4645-4649.

53. Wang, T. C., and Smith, K. C. (1986) Inviability of dam recA and dam recB cells of Escherichia coli is correlated with their inability to repair DNA double-strand breaks produced by mismatch repair, *Journal of bacteriology 165*, 1023-1025.

54. Witkin, E. M., and Roegner, M. V. (1992) Overproduction of DnaE protein (alpha subunit of DNA polymerase III) restores viability in a conditionally inviable Escherichia coli strain deficient in DNA polymerase I., *Journal of bacteriology 174*, 4166-4168.

55. Cox, J. M., Li, H., Wood, E. A., Chitteni-Pattu, S., Inman, R. B., and Cox, M. M. (2008) Defective dissociation of a "Slow" RecA mutant protein imparts an Escherichia coli growth defect, *Journal of Biological Chemistry 283*, 24909-24921.

56. Masai, H., Asai, T., Kubota, Y., Arai, K., and Kogoma, T. (1994) Escherichia coli PriA protein is essential for inducible and constitutive stable DNA replication, *Embo Journal 13*, 5338-5345.

57. Sandler, S. J. (1996) Overlapping functions for *recF* and *priA* in cell viability and UV-inducible SOS expression are distinguished by *dnaC809* in *Escherichia coli* K-12, *Molec. Microbiol. 19*, 871-880.

58. Shibata, T., Hishida, T., Kubota, Y., Han, Y. W., Iwasaki, H., and Shinagawa, H. (2005) Functional overlap between RecA and MgsA (RarA) in the rescue of stalled replication forks in Escherichia coli, *Genes to Cells 10*, 181-191.

59. Nagano, T., Tanaka, T., Mizuki, H., and Hirobe, M. (1994) Toxicity of singlet oxygen generated thermolytically in Escherichia coli, *Chemical & pharmaceutical bulletin 42*, 883-887.

60. Abdul-Tehrani, H., Hudson, A. J., Chang, Y. S., Timms, A. R., Hawkins, C., Williams, J. M., Harrison, P. M., Guest, J. R., and Andrews, S. C. (1999) Ferritin mutants of Escherichia coli are iron deficient and growth impaired, and fur mutants are iron deficient, *Journal of bacteriology 181*, 1415-1428.

61. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification, *Methods 24*, 218-229.

62. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration, *Nat Biotechnol 17*, 1030-1032.

63. Marceau, A. H. (2012) Ammonium sulfate co-precipitation of SSB and interacting proteins, *Methods Mol Biol 922*, 151-153.

64. Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M., and Keck, J. L. (2008) SSB as an Organizer/Mobilizer of Genome Maintenance Complexes, *Critical Reviews in Biochemistry and Molecular Biology 43*, 289-318.

65. Al Mamun, A. A., Lombardo, M. J., Shee, C., Lisewski, A. M., Gonzalez, C., Lin, D., Nehring, R. B., Saint-Ruf, C., Gibson, J. L., Frisch, R. L., Lichtarge, O., Hastings, P. J., and Rosenberg, S. M. (2012) Identity and function of a large gene network underlying mutagenic repair of DNA breaks, *Science 338*, 1344-1348.

66. Ghosh, M., Manna, P., and Sil, P. C. (2011) Protective role of a coumarin-derived schiff base scaffold against tertiary butyl hydroperoxide (TBHP)-induced oxidative impairment and cell death via MAPKs, NF-kappaB and mitochondria-dependent pathways, *Free radical research 45*, 620-637.

67. Granger, A. C., Gaidamakova, E. K., Matrosova, V. Y., Daly, M. J., and Setlow, P. (2011) Effects of Mn and Fe levels on Bacillus subtilis spore resistance and effects of Mn2+, other divalent cations, orthophosphate, and dipicolinic acid on protein resistance to ionizing radiation, *Appl Environ Microbiol 77*, 32-40.

68. Huisman, G. W., D. A. Siegele, M. M. Zambrano, and R. Kolter. (1996) Morphological and physiological changes during stationary phase, In *Escherichia coli and Salmonella: cellular and molecular biology* (al., I. F. C. N. e., Ed.), pp 1672–

1682, ASM Press, Washington, D.C.

69. Sonntag, I., Schwarz, H., Hirota, Y., and Henning, U. (1978) Cell envelope and shape of Escherichia coli: multiple mutants missing the outer membrane lipoprotein and other major outer membrane proteins, *Journal of bacteriology 136*, 280-285.

70. Nikaido, H. (1996) Outer membrane., In *Escherichia coli and Salmonella: Cellular and Molecular Biology* (F.C, N., Ed.), pp 29–47, ASM Press, Washington D.C.

71. Philippe, N., Pelosi, L., Lenski, R. E., and Schneider, D. (2009) Evolution of penicillin-binding protein 2 concentration and cell shape during a long-term experiment with Escherichia coli, *Journal of bacteriology 191*, 909-921.

72. Diver, W. P., Sargentini, N. J., and Smith, K. C. (1982) A mutation (radA100) in Escherichia coli that selectively sensitizes cells grown in rich medium to x- or u.v.-radiation, or methyl methanesulphonate, *International Journal of Radiation Biology & Related Studies in Physics, Chemistry & Medicine 42*, 339-346.

73. Song, Y., and Sargentini, N. J. (1996) Escherichia coli DNA repair genes radA and sms are the same gene, *Journal of bacteriology 178*, 5045-5048.

74. Drees, J. C., Lusetti, S. L., Chitteni-Pattu, S., Inman, R. B., and Cox, M. M. (2004) A RecA filament capping mechanism for RecX protein, *Mol Cell 15*, 789-798.

75. Gruenig, M. C., Stohl, E. A., Chitteni-Pattu, S., Seifert, H. S., and Cox, M. M. (2010) Less is more--Neisseria gonorrhoeae RecX protein stimulates recombination by inhibiting RecA, *Journal of Biological Chemistry in press*.

76. Pages, V., Koffel-Schwartz, N., and Fuchs, R. P. (2003) recX, a new SOS gene that is co-transcribed with the recA gene in Escherichia coli, *DNA Repair. 2*, 273-284.

77. Stohl, E. A., Brockman, J. P., Burkle, K. L., Morimatsu, K., Kowalczykowski, S. C., and Siefert, H. S. (2003) Escherichia coli RecX inhibits RecA recombinase and coprotease activities in vitro and in vivo, *Journal of Biological Chemistry 278*, 2278-2285.

78. Venkatesh, R., Ganesh, N., Guhan, N., Reddy, M. S., Chandrasekhar, T., and Muniyappa, K. (2002) RecX protein abrogates ATP hydrolysis and strand exchange promoted by RecA: Insights into negative regulation of homolgous recombination, *Proc Natl Acad Sci U S A 99*, 12091-12096.

79. DiGate, R. J., and Marians, K. J. (1989) Molecular cloning and DNA sequence analysis of Escherichia coli topB, the gene encoding topoisomerase III, *Journal of Biological Chemistry 264*, 17924-17930.

80. Bageshwar, U. K., Whitaker, N., Liang, F. C., and Musser, S. M. (2009) Interconvertibility of lipid- and translocon-bound forms of the bacterial Tat precursor pre-SufI, *Mol Microbiol 74*, 209-226.

81. Reddy, M. (2007) Role of FtsEX in cell division of Escherichia coli: viability of ftsEX mutants is dependent on functional SufI or high osmotic strength, *Journal of bacteriology 189*, 98-108.

82. Samaluru, H., SaiSree, L., and Reddy, M. (2007) Role of SufI (FtsP) in cell division of Escherichia coli: evidence for its involvement in stabilizing the assembly of the divisome, *Journal of bacteriology 189*, 8044-8052.

83. Tarry, M., Arends, S. J., Roversi, P., Piette, E., Sargent, F., Berks, B. C., Weiss, D. S., and Lea, S. M. (2009) The Escherichia coli cell division protein and model Tat substrate SufI (FtsP) localizes to the septal ring and has a multicopper oxidase-like structure, *J Mol Biol 386*, 504-519.

84. Drees, J. C., Chitteni-Pattu, S., McCaslin, D. R., Inman, R. B., and Cox, M. M. (2006) Inhibition of RecA protein function by the RdgC protein from Escherichia coli, *Journal of Biological Chemistry 281*, 4708-4717.

85. Ryder, L., Sharples, G. J., and Lloyd, R. G. (1996) Recombination-dependent growth in exonuclease-depleted recBC sbcBC strains of Escherichia coli K-12, *Genetics 143*, 1101-1114.

86. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products, *Proc Natl Acad Sci U S A 97*, 6640-6645.

87. Goryshin, I. Y., Jendrisak, J., Hoffman, L. M., Meis, R., and Reznikoff, W. S. (2000) Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes, *Nat Biotechnol 18*, 97-100.

88. Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., Pelletier, J. J., Paulus, H., and Xu, M. Q. (1997) Single-column purification of free recombinant proteins

using a self-cleavable affinity tag derived from a protein splicing element, *Gene 192*, 271-281.

89. Miller, J. H. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

90. Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids, *J Mol Biol 166*, 557-580.

91. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol 10*, R25.

92. Page, A. N., George, N. P., Marceau, A., Cox, M. M., and Keck, J. L. (2011) Structure and properties of the MgsA (RarA) protein of Escherichia coli, *Journal of Biological Chemistry submitted*.

93. Fan, L., Arvai, A. S., Cooper, P. K., Iwai, S., Hanaoka, F., and Tainer, J. A. (2006) Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair, *Mol Cell 22*, 27-37.

94. Hilario, E., Li, Y., Nobumori, Y., Liu, X., and Fan, L. (2013) Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE, *Acta crystallographica. Section D, Biological crystallography 69*, 237-246.

95. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol Syst Biol 7*, 539.

Chapter 4: Preliminary characterization of novel radiation resistance factor, YejH/RadD

4.1 Introduction

The *yejH* gene is a gene of unknown function that was discovered using a high throughput screening technique for genes involved in IR resistance and subsequently renamed *radD* as described in Chapter 3. Within the transposon mutant pool, cells lacking a functional copy of *yejH* (now exclusively referred to as *radD*) dropped out of the population after irradiation with 1000 Gy. In verification experiments, *radD* deletion strains show a decrease in IR survival of nearly 100 fold compared to wild type. At this point, no one has linked YejH/RadD to any cellular functions in the literature, nor reported purifying or characterizing the protein. Sequence analysis of *E. coli* K-12 (*1*) predicted the *E. coli* genome to contain approximately 25 enzymes with nucleic acid unwinding activity; 12 have been biochemically characterized (*2*), two unwind RNA-RNA hybrids (*3-5*), and 11 have never been purified and their helicase activities have yet to be demonstrated (*2*). RadD is one of the eleven un-characterized proteins with predicted helicase function.

From sequence analysis, RadD appears to contain two domains: a conserved N-terminal helicase resembling domain and a unique C-terminal domain. As illustrated in Figure 3 in Chapter 3, the helicase resembling domain is approximately 350 amino acids long and contains the seven helicase motifs found in the Superfamily 2 (SF2) of RNA and DNA helicases. SF2 is the largest helicase superfamily and its members are involved in several diverse cellular processes (*6*, *7*). The second domain contains two CxxC motifs

that we predict coordinate an Fe-S cluster or Zn^{2+} ion. The functions of these domains in IR survival will be discussed in the Future Directions section of Chapter 4.

RadD shares high sequence similarity with the human *Xeroderma pigmentosum* group B helicase, XPB. XPB unwinds dsDNA with a 3′ to 5′ polarity in the presence of ATP (*8*) and is involved in transcription (*9*), nucleotide excision repair (NER), and TFIIH functional assembly (*10*). Loss of *XPB* function causes NER defects in *Xeroderma pigmentosum* group B patients (*11*). These patients are hyperphotosensitive and have a dramatically increased risk of skin cancer (*12*, *13*). A sequence alignment of *E. coli* RadD (*Ec*RadD), *Homo sapien* XPB (*hs*XPB), and the *Archaeoglobus fugidus* XPB homolog (*Af*XPB) is shown in Chapter 3 Figure 3. The helicase containing domain of *Ec*RadD shares 25% sequence identity and 40% similarity with the 383 amino acid long helicase domain of *Hs*XPB. The seven helicase motifs of *Ec*RadD and *Hs*XPB align with 62% identity and 71% similarity. The structure of the two XPB homologs aligned to Ec*RadD* were solved, providing insight into the role of XPB in DNA repair (*14*) and providing structural insight into a potential domain architecture of RadD.

A thorough biochemical characterization of RadD is necessary to begin to unravel the protein's cellular function and role and in radiation damage repair. In this Chapter, I introduce the preliminary biochemical characterization of this enzyme, including the first reported native purification of RadD, characterization of the enzyme's ATP hydrolysis activity and DNA binding properties, as well as the change in domain

architecture upon nucleotide binding. Further in-depth characterizations, both genetic and biochemical, are being performed in the lab and are the topic of discussion in Chapter 5.

4.2 Results:

4.2.1 RadD was purified natively.

RadD was cloned into an expression vector and expression was induced in a Rosetta (DE3) expression strain. The progress of the purification is displayed in Figure 1. Briefly, RadD was precipitated from the supernatant of a polyethylenimine precipitation step in a 30-50% ammonium sulfate cut. RadD was then purified by successive chromatographic steps using a butyl-Sepharose column, a DEAE column, SP-sepharose column, a ceramic hydroxyapatite column, and a Sephacryl S-300 gel filtration column. A sample from each step of the purification was analyzed by gel electrophoresis.

4.2.2 RadD is a DNA-independent ATPase.

Helicases are molecular motors that couple ATP hydrolysis to DNA or RNA unwinding (*15*). For all known helicases, NTP binding and hydrolysis are absolutely required for unwinding. ATP hydrolysis by RadD was monitored by an enzyme-coupled, spectrophotometric assay (*16, 17*) in the absence of DNA and upon addition of DNA substrates presented in Figure 2. Our initial hypothesis was that if RadD is a bona fide DNA helicase, we would observe DNA-dependent ATPase activity. However, we observed DNA-independent ATPase activity (Figure 2). We hypothesize that RadD

could be locked into an active form by the potential redox state of a proposed Fe-S cluster. Anaerobic purification and protein characterization experiments proposed in Chapter 5 will further investigate the potential *in vitro* helicase activity of RadD. It is important to note that the current biochemical characterization of RadD ATPase activity has not been exhaustive. There are numerous DNA substrates that can be tested for inhibitory or stimulatory effects on RadD ATPase activity. The reaction conditions could also potentially be optimized for this enzyme. At the time of doing these experiments, I worried I could spend years trying to optimize conditions to find an activity I did not know existed so I turned to the cell and focused on characterizing RadD via genetic approaches. Work discussed in Chapter 5 performed by our new post-doctoral fellow will come back to these biochemical experiments.

4.2.3 RadD binds ssDNA in the presence of nucleotide.

While a thorough characterization of the DNA binding affinity of RadD has not been performed, we observed binding of RadD to ssDNA in the presence of ATP and the non-hydrolyzable analog, ATPγS, but not in the presence of ADP (Figure 3). This could be due to the results reported below that RadD undergoes a large conformational change upon binding ATPγS. It is possible that RadD must be bound to ATP before it can bind ssDNA. Fluorescent polarization experiments were performed as well but saturation was never reached and thus, the data is not presented here.

4.2.4 RadD undergoes large conformational change upon nucleotide binding.

To determine if RadD undergoes a conformational change upon nucleotide binding we utilized a limited proteolysis approach. Briefly, we incubated chymotyrpsin or subtilisin with RadD in the presence and absence of ATP and ATPγS and ran the proteolysis reactions on SDS-PAGE gels and submitted the samples for mass spec analysis to identify the stable protein fragments. Figure 4 illustrates the results of this experiment.

In the absence of nucleotide, a stable protein fragment consisting of the C-terminal 226 amino acids of RadD is a prominent band following proteolysis with chymotrypsin. In the presence of ATPγS, the stable fragment consists of the N-terminal 463 amino acids upon subtilisin digestion. As illustrated in the Figure 4, the regions of RadD where these two fragments overlap include the two CxxC motifs indicating a potential hinge in this area.

4.2.5 RadD is within close proximity to known DNA repair enzymes and nucleotide associating proteins.

To learn more about the role that RadD is playing in the cell, we looked for potential protein partners using a tandem affinity purification (TAP) method. We fused RadD to a TAP tag and expressed the tagged protein in wildtype *E. coli*. We purified the TAP tagged protein as previously described (*18*, *19*) and submitted the associated proteins to mass spectroscopy analysis. The results, minus the ribosomal proteins, are

reported in Chapter 3 Figure 7a. The most prominent protein in the pull-down sample was the single-strand DNA binding protein (SSB). Thirteen unique SSB peptides were identified resulting in sequence coverage of 67%. The only protein with higher abundance was RadD itself. Various DNA repair enzymes such as RuvA, RecQ, TopB and TopA were identified in high abundance and are listed in Figure 7A. Although we don't know if RadD is interacting with any of these proteins directly, or indirectly via interactions with DNA and/or SSB, the apparent proximity of RadD to known repair enzymes again suggests that RadD is involved in DNA metabolism.

In the list of potential interaction partners obtained from the TAP experiment are the following three nucleoid-associated proteins (NAPs): SeqA, H-NS, and Dps. It has been known for some time that the *E. coli* genome is subjected to the compaction within the nucleoid and that this impaction is imposed by a combination of factors one of which is the interaction of the DNA with proteins such as nucleoid-associated proteins (NAPs) (*20*). Most NAPs bind DNA and have the ability to bend, wrap, or bridge DNA molecules. It has been reported that there is a good correlation between transcriptional activity and the number and stability of looped DNA domains (*21-23*). Many NAPs were actually studied initially as transcriptional regulators and have been reported to have both positive and negative regulation roles. Could RadD remove NAPs and thus upregulate or downregulate the expression of genes beneficial or detrimental for radiation exposure survival? As discussed in the Future Directions section of this

document, experiments will be performed to validate this pulldown result and look at transcription activity in strains expressing functional and mutant RadD.

4.2.6 The RadD-SSB interaction

To validate the potential RadD-SSB interaction, we exploited the convenient precipitation of SSB at very low concentrations of ammonium sulfate (*24*). As illustrated in the gel in Chapter 3 Figure 7B, SSB and SSBΔC8 precipitate at 20% Ammonium sulfate but RadD does not. When RadD is incubated with SSB, SSB pulls RadD into the pellet and both bands are observed on the gel indicating the two proteins interact. When RadD is incubated with SSBΔC8, RadD is not pulled into the pellet and no RadD band is observed, indicating that the interaction between RadD and SSB is tail dependent. The SSB C-terminal tail has been shown to interact with numerous DNA repair and replication proteins (*25*). Thus, this interaction lends support to our hypothesis that RadD is involved in DNA metabolism, possibly in a DNA repair role, and suggests an SSB-mediated mechanism for the localization of RadD to damaged DNA or the replication fork.

4.3 Discussion

An initial characterization of RadD demonstrates that the enzyme (I) hydrolyzes ATP in a DNA-independent fashion, (II) binds ssDNA and undergoes a large comformation change in the presence of ATP, (III) interacts directly the SSB C-terminal interaction domain, and (IV) is found in close proximity of numerous DNA metabolism proteins. As the RadD protein is critical for IR survival, follow up experiments discussed in Chapter 5 will be performed to define the cellular function of this protein.

In no way were the conditions or reaction perameters exhausted in the experiments described in this section. When characterizing a new enzymes, one can spend years trying to optimize conditions to find an activity. To avoid this fate I turned to the cell and focused on characterizing RadD via genetic approaches described in Chapter 3. The experiments described in this section will be repeated and further optimized in the future by a postdoctoral research fellow who has taken over the project.

4.4 Materials and Methods

RadD Purification. The open reading frame of E. coli radD was amplified by PCR and subcloned in-frame into NdeI/BamHI-digested pET21a resulting in plasmid pEAW724. The Rosetta (DE3) strain was transformed with the pEAW724 and cells were grown in Luria-Bertania medium containing 100 μg/mL ampicillin at 37°C to an OD_{600} of ~0.5. RadD expression was induced by the addition of 0.8mM isopropyl 1-thio-β-D-

galactopyranoside and cells were grown for an additional 4 hours at 37°C.

Cells were pelleted and resuspended in sucrose solution (25% (w/v) sucrose, 250mM Tris-Cl (80% cation), pH 7.7, 7mM EDTA, 1μM pepstatin, 1μM leupeptin, 1μM E-64) and lysed by sonication and the addition of 1.6g/liter of lysozyme. Insoluble material was removed by centrifugation. RadD was precipitated by the addition of solid $NH_4(SO_4)_2$. The $NH_4(SO_4)_2$ pellet was resusped in 20mM Tris-Cl (80% cation), pH 7.7, 1mM EDTA, 10% glycerol (R-buffer) containing 800mM $NH_4(SO_4)_2$ and applied to a butyl-sepharose column. RadD was eluted using a liner gradient from 1M to 0mM $NH_4(SO_4)_2$ in R-buffer. RadD containing fractions were pooled, dialyzed into P buffer (20mM phosphate, pH 7.0, 1mM EDTA, and 10% glycerol) and loaded onto a ceramic hydroxyapatite column. RadD came off late in the wash. Fractions containing RadD were dialyzed into R-buffer + 200mM NaCl and loaded onto a source-15Q column. RadD flowed through but a peak containing nuclease came off in gradient before being run over a size exclusion column. The resulting RadD was dialyzed into R-buffer +200mM NaCl, aliquoted, and flash-frozen in liquid nitrogen before being stored at -80°C. The purified protein was >95% pure by gel and free of any detectable nuclease activity.

*ATPase Assay.*A coupled spectrophotometric enzyme assay (63, 64) was used to measure the ATPase activities of RadD. The assays were carried out in a Varian Cary 300 dual beam spectrophotometer equipped with a temperature controller and a 12-position cell changer. The reactions were carried out at 37 °C in 25 mM Tris acetate (80% cation),

1 mM DTT, 3 mM potassium glutamate, 10 mM $Mg(OAc)_2$, 5% (w/v) glycerol, an ATP regeneration system (10 units/ml pyruvate kinase, 2.2 mM phosphoenolpyruvate), a coupling system (3 mM NADH and 10 units/ml lactate dehydrogenase), purified RadD, and the indicated DNA substrate.

Electrophoretic Mobility Shift Assays. A single-stranded carboxyfluorescein-labeled polydT 100-mer oligonucleotide was purchased from Integrated DNA Technologies. DNA binding reactions (20μL) contained the indicated concentration of RadD and 50nM polydT100. Reactions were carried out in the following reaction buffer: 275mM KGlu. mM Tris acetate (80% cation), 75mM potassium glutamate, 1 mM DTT, 3 mM potassium glutamate, 10 mM $Mg(OAc)_2$, and 5% (w/v) glycerol. Samples were incubated for 10minutes at 37°C, and then 5 μL of 6x loading buffer (18% (w/v) Ficoll, 20 mM Tris-OAc 80% cation) was added to each reaction before the reactions were loaded on an 7.5% acrylamide gel and subjected to electrophoresis in TBE buffer (90 mM Tris borate and 10 mM EDTA).

Limited Proteolysis. Escherichia coli RadD (10 μM) was incubated with 0.2 μM protease (a-chymotrypsin or subtilisin) in 20 mM Tris, pH 8.0, 200 mM NaCl, 1 mM DTT, 1 mM EDTA, plus or minus 0.1 μM nucleotide (indicated in Figure 4) at room temperature. Aliquots (10 μL) from the reaction were quenched after 5, 10, 20, 30, 60, 90, 120 and 240 minutes by mixing with 1 mM PMSF and flash-freezing in liquid nitrogen. Samples were then separated by SDS-PAGE on a purchased 4-20% gradient gel (BioRad)

and stained with Coomassie Brilliant Blue. Individual bands (boxed in Figure 4) were submitted to University of Wisconsin Mass Spectrometry facility for peptide identification.

Tandem Affinity Purification. The open reading frame of *E. coli radD* was amplified by PCR and subcloned in-frame into cloning vector pCN70[3] to produce an expression vector encoding RadD with an N-terminal dual affinity tag (includes protein A and calmodulin peptide binding domains separated by a tobacco etch virus protease cleavage site). *E. coli* K12 strain MG1655 (DE3) transformed with pTAP-RadD was grown at 37 °C in 4 liters of Luria-Bertani medium supplemented with 50 μg/ml ampicillin to midlog phase ($A_{600\ nm}$ of ~0.5), induced by the addition of 1 μM isopropyl 1-thio-β-D-galactopyranoside, and grown for an additional 3 h. Cells were harvested by centrifugation, suspended in 50 ml of Nonidet P-40 buffer (6 mM dibasic sodium phosphate, 4 mM monobasic sodium phosphate, 150 mM NaCl, 2 mM EDTA, 50 mM NaF, 4 mg/liter leupeptin, 0.1 mM sodium vanadate, 19.5 mg/liter benzamidine, 8.7 mg/liter phenylmethylsulfonyl fluoride (PMSF), 1% Nonidet P-40 substitute), and lysed by sonication. Tandem affinity purification was performed as described previously (*26*) except the TCA precipitation pellet was resuspended in buffer, digested with trypsin, and subjected to MALDI-TOF mass spectrometry for identification of peptides (University of Wisconsin Mass Spectrometry facility).

Ammonium sulfate co-precipitation. Co-precipitation experiments were performed as described previously (*24*) and pellet fractions were suspended in 30 μl of loading buffer prior to SDS-PAGE on 4-15% polyacrylamide gradient gels (Bio-Rad).

Figures

Figure 1: Native purification of YejH/RadD.

Purification was performed as described in the Materials and Methods. YejH/RadD was purified to >95% purity with no observable nuclease contamination.

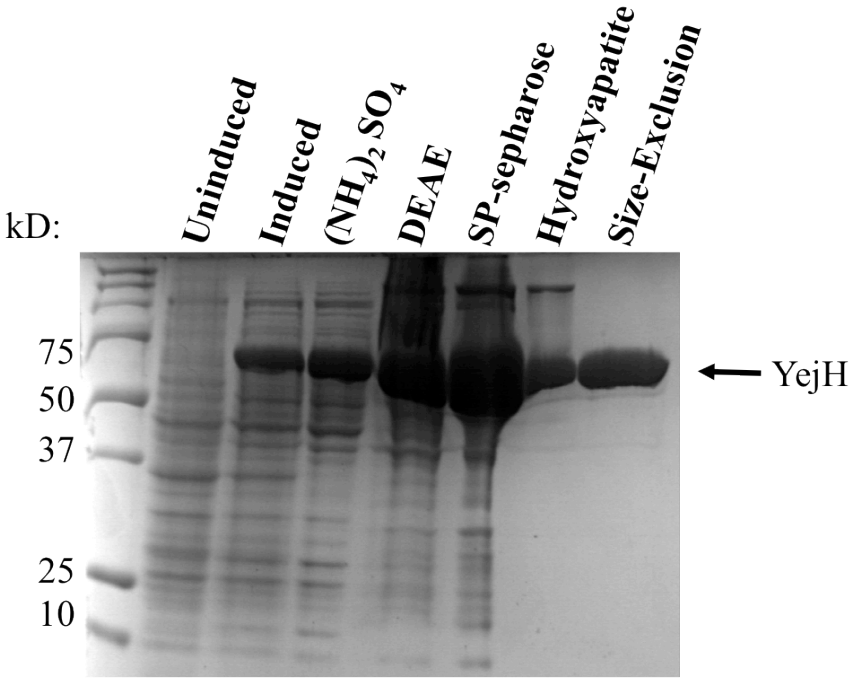

kD:
Uninduced
Induced
$(NH_4)_2 SO_4$
DEAE
SP-sepharose
Hydroxyapatite
Size-Exclusion
75
50
37
25
10
YejH

Figure 2: YejH is a DNA-independent ATPase.

ATP hydrolysis was monitored as described in Materials and Methods with the various DNA substrates listed above. RadD ATPase activity does not appear to be stimulated or inhibited by the presence of these DNA substrates. Rather RadD has DNA-independent ATPase activity.

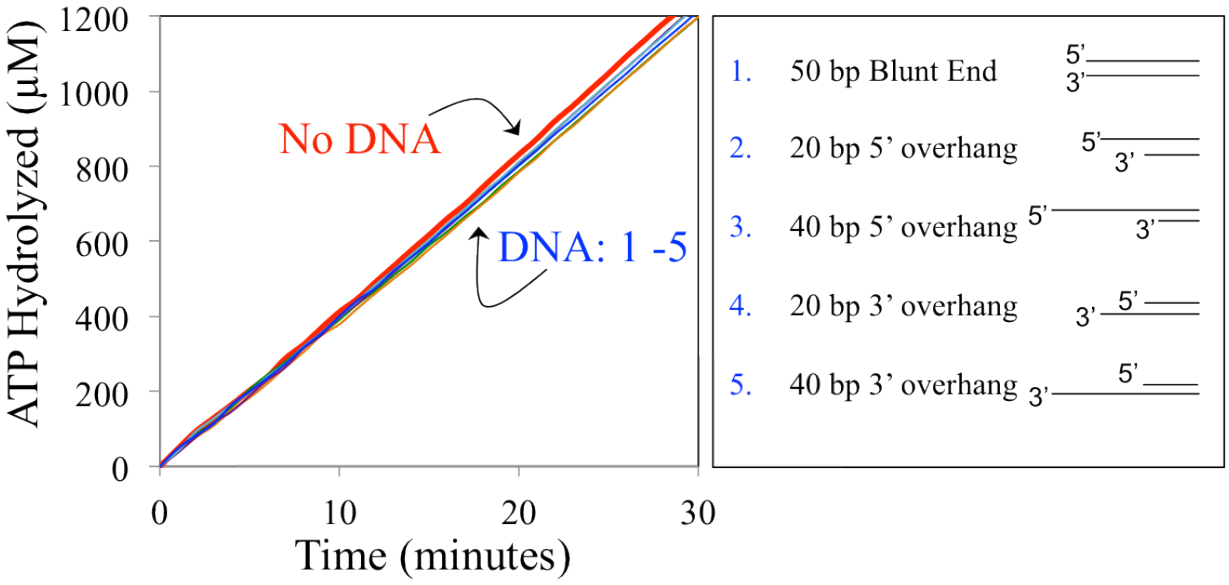

ATP Hydrolyzed (µM)
1200
1000
800
600
400
200
0
0
10
20
30
Time (minutes)
No DNA
DNA: 1 -5
1. 50 bp Blunt End
2. 20 bp 5' overhang
3. 40 bp 5' overhang
4. 20 bp 3' overhang
5. 40 bp 3' overhang
5'
3'

Figure 3: YejH/RadD binds single stranded DNA in the presence of ATP and ATPγS.

Electromobility shift assays were performed as described in Materials and Methods. RadD appears to bind single stranded DNA substrates lacking secondary structure in the presence of ATP and ATPγS, but not ADP.

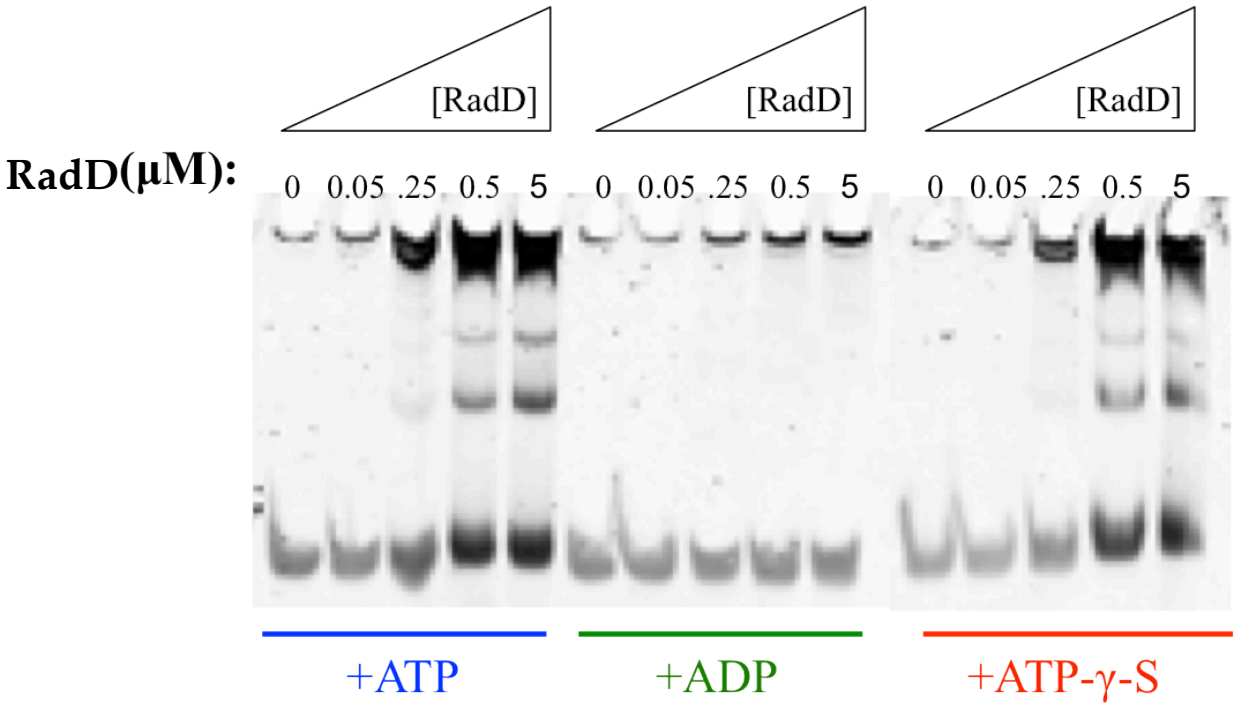
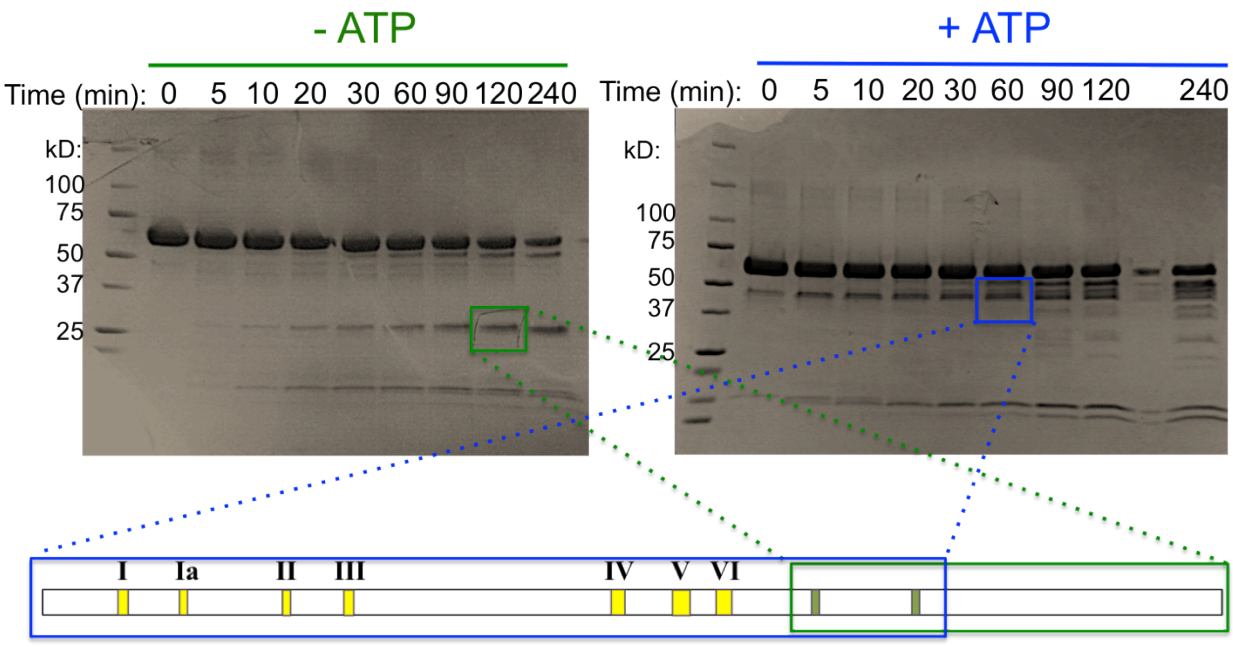

RadD(μM):
[RadD]
[RadD]
[RadD]
0 0.05 .25 0.5 5
0 0.05 .25 0.5 5
0 0.05 .25 0.5 5
+ATP
+ADP
+ATP-γ-S

Figure 4: YejH/RadD undergoes conformational rearrangement upon nucleotide binding.

Limited proteolysis and fragment identification was performed as described in Materials and Methods. RadD was incubated with protease for increasing amounts of time before being subjected to SDS-PAGE in the presence or absence of nucleotide. Green and Blue boxes indicate the stable fragments analyzed my mass spectroscopy. Their identity is indicated on the schematic of RadD, where the 7 conserved helicase domains are highlighted in yellow and the 2 CxxC domains are green.


- ATP
+ ATP
Time (min): 0 5 10 20 30 60 90 120 240
Time (min): 0 5 10 20 30 60 90 120 240
kD:
100
75
50
37
25
kD:
100
75
50
37
25
I
Ia
II
III
IV
V
VI

References

1. Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome sequence of Escherichia coli K-12, *Science 277*, 1453-1462.
2. Voloshin, O. N., Vanevski, F., Khil, P. P., and Camerini-Otero, R. D. (2003) Characterization of the DNA damage-inducible helicase DinG from Escherichia coli, *J Biol Chem 278*, 28284-28293.
3. Henn, A., Medalia, O., Shi, S. P., Steinberg, M., Franceschi, F., and Sagi, I. (2001) Visualization of unwinding activity of duplex RNA by DbpA, a DEAD box helicase, at single-molecule resolution by atomic force microscopy, *Proc Natl Acad Sci U S A 98*, 5007-5012.
4. Diges, C. M., and Uhlenbeck, O. C. (2001) Escherichia coli DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA, *Embo J 20*, 5503-5512.
5. Liou, G. G., Chang, H. Y., Lin, C. S., and Lin-Chao, S. (2002) DEAD box RhlB RNA helicase physically associates with exoribonuclease PNPase to degrade double-stranded RNA independent of the degradosome-assembling region of RNase E, *J Biol Chem 277*, 41157-41162.
6. Singleton, M. R., Dillingham, M. S., and Wigley, D. B. (2007) Structure and mechanism of helicases and nucleic acid translocases, *Annu Rev Biochem 76*, 23-50.
7. Jankowsky, E., and Bowers, H. (2006) Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases, *Nucleic acids research 34*, 4181-4188.
8. Hwang, J. R., Moncollin, V., Vermeulen, W., Seroz, T., van Vuuren, H., Hoeijmakers, J. H., and Egly, J. M. (1996) A 3' --> 5' XPB helicase defect in repair/transcription factor TFIIH of xeroderma pigmentosum group B affects both DNA repair and transcription, *J Biol Chem 271*, 15898-15904.
9. Park, E., Guzder, S. N., Koken, M. H., Jaspers-Dekker, I., Weeda, G., Hoeijmakers, J. H., Prakash, S., and Prakash, L. (1992) RAD25 (SSL2), the yeast homolog of the human xeroderma pigmentosum group B DNA repair gene, is essential for viability, *Proc Natl Acad Sci U S A 89*, 11416-11420.
10. Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J. H., Chambon, P., and Egly, J. M. (1993) DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor, *Science 260*, 58-63.

11. Weeda, G., van Ham, R. C., Vermeulen, W., Bootsma, D., van der Eb, A. J., and Hoeijmakers, J. H. (1990) A presumed DNA helicase encoded by ERCC-3 is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome, *Cell 62*, 777-791.

12. Berneburg, M., and Lehmann, A. R. (2001) Xeroderma pigmentosum and related disorders: defects in DNA repair and transcription, *Adv Genet 43*, 71-102.

13. de Boer, J., and Hoeijmakers, J. H. (2000) Nucleotide excision repair and human syndromes, *Carcinogenesis 21*, 453-460.

14. Fan, L., Arvai, A. S., Cooper, P. K., Iwai, S., Hanaoka, F., and Tainer, J. A. (2006) Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair, *Mol Cell 22*, 27-37.

15. Abdel-Monem, M., Durwald, H., and Hoffmann-Berling, H. (1976) Enzymic unwinding of DNA. 2. Chain separation by an ATP-dependent DNA unwinding enzyme, *Eur J Biochem 65*, 441-449.

16. Morrical, S. W., Lee, J., and Cox, M. M. (1986) Continuous association of Escherichia coli single-stranded DNA binding protein with stable complexes of recA protein and single-stranded DNA, *Biochemistry 25*, 1482-1494.

17. Lindsley, J. E., and Cox, M. M. (1990) Assembly and disassembly of RecA protein filaments occur at opposite filament ends. Relationship to DNA strand exchange, *J Biol Chem 265*, 9043-9054.

18. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification, *Methods 24*, 218-229.

19. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration, *Nat Biotechnol 17*, 1030-1032.

20. Dillon, S. C., and Dorman, C. J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression, *Nat Rev Microbiol 8*, 185-195.

21. Deng, S., Stein, R. A., and Higgins, N. P. (2005) Organization of supercoil domains and their reorganization by transcription, *Mol Microbiol 57*, 1511-1521.

22. Postow, L., Hardy, C. D., Arsuaga, J., and Cozzarelli, N. R. (2004) Topological domain structure of the Escherichia coli chromosome, *Genes Dev 18*, 1766-1779.

23. Stein, R. A., Deng, S., and Higgins, N. P. (2005) Measuring chromosome dynamics on different time scales using resolvases with varying half-lives, *Mol Microbiol 56*, 1049-1061.

24. Marceau, A. H. (2012) Ammonium sulfate co-precipitation of SSB and interacting proteins, *Methods Mol Biol 922*, 151-153.

25. Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M., and Keck, J. L. (2008) SSB as an Organizer/Mobilizer of Genome Maintenance Complexes, *Critical Reviews in Biochemistry and Molecular Biology 43*, 289-318.

26. Page, A. N., George, N. P., Marceau, A., Cox, M. M., and Keck, J. L. (2011) Structure and properties of the MgsA (RarA) protein of Escherichia coli, *Journal of Biological Chemistry submitted*.

Chapter 5: Major Findings and Future Directions

5.1 Major Findings

Ionizing radiation is high-energy radiation that causes devastating biological effects. The energy deposition events that occur during IR exposure lead to a plethora of biological consequences including: (I) DNA strand breaks, (II) DNA base damage, and (III) protein inactivation/carbonylation. Numerous mechanisms and cellular pathways have been described for the repair of these types of damage and were outlined in Chapter 1. *Deinococcus radiodurans* is an extremely radiation resistant organism. The mechanisms responsible for this organisms extraordinary ability to repair a genome shattered into hundreds of pieces has been the focus of numerous research groups and the subject of a heated debate. In order to understand how organisms evolve resistance to IR, we directly evolved normally radiation-sensitive *E. coli* to be nearly as radiation resistant as *D. radiodurans.* In the work described in Chapter 2, we focused our attention on the radiation resistant isolate, CB2000. By testing the effect of identified mutations in this strain we discovered that the IR resistance phenotype is almost entirely accounted for by only three nucleotide changes, in the DNA metabolism genes *recA*, *dnaB*, and *yfjK.* The major finding from this study is that the adaptation of current DNA repair functions is one way in which cells can evolve resistance to IR.

Not all genes important for radiation resistance were mutated in the evolution project so we took a genetic screening approach (Chapter 3) to identify which non-essential genes contribute to IR resistance in *E. coli.* This work was performed to

contribute towards the production of a complete molecular description of the cellular processes that contribute to IR survival. This worked produced 47 screen identified genes that appear to contribute to IR survival. We validated 14 of the 47 genes identified, and the strains deleted for these genes had IR sensitive that ranged from moderate to critical. Results from this study indicate that surviving radiation exposure is a global initiative. While DNA repair mechanisms made up 40% of the genes we identified in the screen, other genes clustered into cell division, cell wall structure and biogenesis, and protein turnover. We also identified 7 genes of unknown function that appear to have a role in radiation recovery. One of these genes of unknown function, *yejH*, was shown to be very important for surviving IR exposure. YejH was further characterized and renamed RadD to annotate the newly described role of the protein in radiation resistance. In Chapter 4, RadD was preliminarily characterized and shown to interact with single stranded DNA binding protein, SSB, providing a molecular mechanism for localization to damaged DNA and/or the replication fork.

Work described in this document begins to describe the genetic basis of radiation resistance. Our results highlight the importance of functional DNA repair mechanisms in IR survival but also suggest that other cellular processes play a role. These results suggest that we should widen our understanding from the field's previously proposed molecular-centric ideologies and illustrate that in surviving IR exposure cells must make a global repair initiative.

5.2 Future Directions

5.2.1 Chapter 2 Future Directions

In Chapter 2, the radiation resistant population IR-2-20 was the focus of our work because the population was stable and numerous mutations were observed in DNA repair genes. We have also identified mutations that appear to be important for the IR-3-20 population as well. The genes commonly mutated in this population cluster into protein turnover and oxidative stress signaling, likely supporting the amelioration of protein oxidation hypothesis proposed by M. Daly. By performing functional experiments similar to those described in Chapter 2, we will be able to identify which of numerous mutations in CB3000 are responsible for the strain's evolved resistance. We propose that there are numerous evolutionary paths to IR resistance, and I hypothesize that analysis of IR-3-20 will support this idea.

5.2.2 Chapter 3 Future Directions

We identified 47 potential IR resistance genes and validated 14 of these genes in the study described in Chapter 3. First, more validation experiments to confirm the remaining genes will be performed. It is important to test the contribution of genes that do not cluster into a DNA repair function to support our current hypothesis that cells rely on numerous cellular processes to survive IR exposure. Second, we identified 7 genes of unknown function in this project, thus potentially providing the first reported phenotype for 7 genes of unknown function. Follow-up experiments to confirm this

phenotype should be performed. Further experiments into the role that these genes may play in radiation biology will be incredibly interesting and begin to provide information about conserved yet enigmatic genes. Lastly, continued dissection of the *yejH*/*radD* pathway will continue to shed light on the cellular role and function of our new radiation resistance gene and potentially describe a new DNA repair pathway never before documented. In work not described in this document, a potentially genetic redundancy with *recG* was discovered. These preliminary findings may suggest of role for YejH/RadD in fork regression and will be discussed further below.

5.2.3 Chapter 4 Future Direction

A preliminary biochemical characterization of YejH/RadD was described in Chapter 4. In the future, the following experiments will be performed to learn more about the role of this radiation resistance protein.

1) Cellular localization of RadD. To determine the cellular localization of RadD, a GFP fusion protein has been created, and cells containing the fusion protein will be observed by fluorescent microscopy to determine the localization of the protein under normal and stressed conditions.

2) Suppressor analysis. We very recently discovered that a Δ*radD* Δ*recG* strain grows very slowly, while both of the single mutants grow normally. This double mutant quickly develops suppressors that regain normal growth. In future studies we will isolate suppressors of this slow growth defect. The genetic changes that result in

suppression should provide an excellent clue as to the process that is limiting growth, and by extrapolation, potentially the function of RadD. Genetic changes will be identified by whole genome sequencing of both the parent double deletion strain and the suppressor strain.

3) Anaerobic purification of RadD. As discussed in Chapter 2 and 3, RadD has two CxxC domains and we hypothesis these cysteine residues may coordinate an Fe-S cluster that could be important for *in vivo* and *in vitro* RadD activity. RadD will be purified anaerobically in the laboratory of Patricia Kiley. A simply spectrophotometric analysis will be performed to determine if RadD coordinates an Fe-S cluster.

4) Helicase/translocation activities. As RadD is predicted to be a super family 2 helicase, the protein will be tested for its ability to modify DNA structure, including unwinding, annealing, translocation, and replication fork regression. These activities can be tested by using relatively simple oligonucleotide-based substrates and by performing gel electrophoresis to observe the products. These experiments will be performed with proteins purified anaerobically if RadD is shown to coordinate an Fe-S cluster.

5) Structural studies. While sequence homology predicts that RadD is a helicase, the three-dimensional structure is unknown. To determine the oligomeric state of RadD in solution, we will use quantitative gel filtration and analytical ultracentrifugation approaches. To gain additional structural information, crystallographic studies will be conducted in collaboration with James Keck.

6) Protein interaction verifications. A tandem affinity purification of RadD produced the list of potential protein interaction partners in Chapter 3 Figure 7. In this list are three Nucleoid Associated Proteins (NAPs). I hypothesize that RadD could play a role in reorganizing the nucleoid. In the future, experiments will be performed to validate the pull-down of these NAPS. Currently the genes that encode these NAPs have been cloned into a yeast-2-hybrid vector and will be tested for direct interactions with RadD. If this result is verified, transcriptomic experiments will be performed in a strain deleted for *radD* to determine if RadD affects transcription by reorganizing the nucleoid.

Appendix A: Directed Evolution of Extreme Resistance to Ionizing Radiation: How RecA Protein Adaptation Contributes

Joseph R. Piechura[1], Tzu-Ling Tseng[2], Hsin-Fang Hsu[2], Rose T. Byrne[1], John R. Battista[3], Hung-Wen Li[2], and Michael M. Cox[1]*

[1]Department of Biochemistry, University of Wisconsin – Madison, Madison, Wisconsin 53706-1544, [2]Department of Chemistry, National Taiwan University, Taiwan, [3]Department of Biological Sciences, Louisiana State University and A & M College, Baton Rouge, LA 70803

*To whom correspondence should be addressed.

B.1 Abstract

Strains of *Escherichia coli* that can survive extreme exposure to ionizing radiation have been generated by directed evolution. Among the evolved isolates, mutations in the *recA* gene are prominent and contribute substantially to the acquired phenotype. Changes at amino acid residue 276, D276A and D276N, occur repeatedly and in separate evolved populations. We find that RecA D276A and RecA D176N exhibit unique adaptations to an environment that can require the repair of hundreds of double strand breaks. These two RecA protein variants form filaments on single strand DNA, hydrolyze ATP, and promote DNA strand exchange, as does the wild type protein. However, the RecA D276A/N proteins (a) exhibit a slower rate of filament extension, leading to a distribution of the protein among a higher number of shorter filaments, (b) promote DNA strand exchange more efficiently in the context of a shorter filament, and (c) are markedly less inhibited by ADP. These adaptations potentially allow RecA protein to address larger numbers of double strand DNA breaks in an environment where ADP concentrations are higher due to a compromised cellular metabolism. The results contribute to a logical framework for thinking about the reconstitution of highly degraded genomes.

B.2 Introduction

Ionizing radiation (IR) is present in the environment in the form of low levels of X-rays and radionuclides. Higher-level doses may be encountered in the form of X-ray devices and radioactive compounds. However, few organisms are normally exposed to substantial IR doses in the environment. Nevertheless, a number of organisms can survive extraordinary exposure to IR. The molecular basis of this resistance is of interest due to the potential for human genomic and cellular damage due to IR exposure related to human activities.

The most studied model system for extreme resistance to IR is the bacterium *Deinococcus radiodurans* (*1-4*). Whereas a 2-5 Gy dose of IR is lethal for a human, *Deinococcus radiodurans* can survive doses in excess of 5,000 Gy with no lethality (*1-5*). The extraordinary resistance of *Deinococcus* to ionizing radiation is likely to be a byproduct of an evolved capacity to withstand long-term desiccation (*6*, *7*).

On the surface, *Deinococcus* appears to possess an unremarkable complement of DNA repair systems. These include recombinational DNA repair (RecA, RecF, RecO, RecR, RecX, RecN, RadA, RuvA, RuvB, RuvC, RecG, RecJ, RecQ, SbcCD), nucleotideexcision repair (UvrA, UvrB, UvrC, UvrD), base excision repair (Xth, AP endonuclease, two AlkA glycosylases, nine DNA glycosylases), and mismatch repair (MutS (2 homologs), MutL, UvrD, no NutH homolog) (*4*, *8*). A handful of novel proteins are present, induced to high levels when *Deinococcus* is exposed to IR (DdrA, DdrB, DdrC, DdrD, PprA (*2*, *9*), and play a demonstrated role in genome reconstitution. However, the activities of at least some of these proteins suggest a role in genome preservation (perhaps in the context of long term desiccation) rather than a role in a novel DNA repair system (*10-13*). The apparent absence of novel DNA repair functions in *Deinococcus* has contributed to arguments that the molecular basis of extreme IR resistance lies in the capacity of the cells to protect its proteins from oxidative damage, rather than any special facility for DNA repair (*3*, *14*, *15*).

Deinococcus does appear to utilize its DNA repair functions in a somewhat unusual two-stage process for genome reconstitution, dubbed extended synthesis-dependent single strand annealing, or ESDSA (*16-18*). The RecA protein of *Deinococcus radiodurans* (DrRecA) also possesses some novel attributes that might be linked to its function in the context of the reconstitution of an IR-fractured genome. First, under at least some conditions, the DrRecA protein promotes DNA strand exchange with a preferred DNA substrate order of addition that is opposite of that seen with the *E. coli* RecA protein (EcRecA) (*19*). Rather than binding single strand DNA first, DrRecA

preferentially initiates strand exchange from the duplex DNA substrate (*19*).

The properties of DrRecA with respect to filament formation are also distinct. The EcRecA protein nucleates filament formation relatively slowly, and extends the filaments more rapidly, properties that would tend to localize much of the available RecA protein in a single filament. This may reflect a repair system that typically must deal with one or only a few situations requiring recombinational DNA repair in each cell cycle. In contrast, the DrRecA protein nucleates more rapidly and extends the filament more slowly, properties which would tend to create large numbers of shorter filaments (*20*). These properties are consistent with a system that must deal with hundreds of double strand breaks after desiccation or extreme doses of ionizing radiation.

To better understand the genetic innovations that produce extreme resistance to IR, we have subjected *Escherichia coli* to directed evolution. We have generated strains of *E. coli* that are nearly as resistant to IR as is *Deinococcus* (*21*), exhibiting increases in survival of 3-4 orders of magnitude when exposed to an IR dose of 3,000 Gy. For one of these strains, an isolate called CB2000, we have attributed nearly the entire phenotype to mutations in the genes *recA*, *dnaB*, and *yfjK* (Chapter 2). All three mutations reflect broader patterns in the four evolved populations. The overall result demonstrates that adaptation of existing DNA repair systems, rather than evolution of novel repair processes, can make major contributions to an extreme IR resistance phenotype (Chapter 2).

In the evolution trials, mutations in the *recA* gene appear prominently in three of four separately evolved populations, and alterations at residue 276 (D276A and D276N) predominate. The D276A RecA mutation makes a substantial contribution to the overall IR resistance phenotype in CB2000 (Chapter 2). The RecA alterations, along with other mutations, demonstrate that new DNA repair systems may not be necessary for extreme IR resistance; existing systems can adapt to the new cellular requirements. In spite of over three decades of investigation of the structure and function of RecA protein, the

two RecA variants at residue 276 have never been isolated or characterized.

The obvious question becomes: what alterations in the molecular repertoire of the RecA protein might contribute to resistance to ionizing radiation? In this report, we investigate the changes in wild type RecA protein function that result from the D276A and D276N mutations, revealing acquired properties that fit the context of a severely degraded genome.

B.3 Results

B.3.1 RecA D276A and RecA D276N hydrolyze ATP moderately faster than Wild type RecA when bound to cssDNA and more readily displace SSB from cssDNA.

We began our investigation of the RecA D276A and RecA D276N proteins with an analysis of ATPase activity on circular single-standed DNA (cssDNA). Excess RecA protein or RecA protein variant was added to cssDNA to allow the protein to bind and filament on the DNA, and ATP was added to initiate ATP hydrolysis. SSB was added with ATP to remove secondary structure from the cssDNA to allow for formation of contiguous RecA filaments. RecA was present in excess of ssDNA, and the 3 mM ATP added was approximately 150 fold greater than the measured Km for wild type RecA protein bound to ssDNA (*22*). ATPase activity is thus reported here as apparent k_{cat}, calculated by dividing the observed rate of ATP hydrolysis by the number of RecA binding sites in solution (the ssDNA concentration in total nucleotides divided by 3). Wild type RecA hydrolyzed ATP with an apparent k_{cat} of 29.3 ± 0.8 min^{-1}, which agrees well with the literature value of 30 min^{-1} (*23-28*). As shown in Figure 1, RecA D276A and RecA D276N displayed a small but significant increase in the rate of ATP hydrolysis compared to wild type RecA, with apparent k_{cat} values of 32.8 ± 0.6 min^{-1} and 33.1 ± 0.9 min^{-1}, respectively.

The capacity of the RecA variants to displace SSB protein from DNA was examined next (Figure 1). In these experiments, ATP and SSB were added to cssDNA first to allow SSB to bind and coat the cssDNA, and then RecA protein was added to

initiate ATP hydrolysis. A long lag was observed before a steady state rate of ATP hydrolysis was reached for all proteins, consistent with the fact that RecA protein has a quite limited capacity to displace SSB from ssDNA (*29-32*). As conditions for ATP hydrolysis were not optimal, k_{cat} was not calculated for these experiments. An apparent lag time was calculated by extrapolating the most linear portion of each curve to the x axis to estimate the time required for the maximum rate observed over a 90m min timecourse to be reached. Wild type RecA hydrolyzed ATP with a maximum observed rate of 22 ± 1 μM/min (compared with the approximately 50 μM/min that would be produced if the ssDNA were saturated by RecA protein). The lag time needed to reach this rate was 36 ± 3 min. Interestingly, RecA D276A hydrolyzed ATP at a maximum rate of 31 ± 2 μM/min with a lag time of 23 ± 3 min and RecA D276N hydrolyzed ATP at a maximum rate of 34 ± 2 μM/min with a lag time of 21.7 ± 0.3 min. This data suggests that RecA D276A and RecA D276N have a somewhat enhanced capacity to displace SSB from DNA, although they do not approach the capacity of classic RecA mutants such as RecA E38K (RecA730) or RecAΔC17 (*30, 31*).

B.3.2 RecA D276A and RecA D276N hydrolyze ATP faster while catalyzing strand exchange but resolve strand exchange intermediates into products somewhat slower than wild type RecA.

Next, the strand exchange activity of the RecA variants was characterized. Contiguous RecA filaments were formed on cssDNA as described above and were allowed to hydrolyze ATP for 20 minutes. Then, homologous linear double-stranded DNA (ldsDNA) was added to initiate strand exchange. RecA filaments on cssDNA pair the DNA with homologous DNA in the ldsDNA and then exchange the cssDNA with the identical strand in the ldsDNA, creating a newly paired nicked circular DNA duplex (Fig. 2A). After addition of ldsDNA, time points were removed, deproteinized, and resolved on an agarose gel to follow the progression of DNA strand exchange. ATP hydrolysis was also measured and activity is expressed as apparent k_{cat} as described

above.

When bound to cssDNA, wild type RecA hydrolyzed ATP with an apparent k_{cat} of 27.9 ± 0.8 min^{-1} in this experiment. This rate is slightly slower than the one reported above, likely due to the potentially inhibitory higher concentration of PEP included in these reactions to extend the lifetime of the ATPase coupling system. Once strand exchange was initiated, the rate of hydrolysis of Wild type RecA decreased by nearly 1/3, to an apparent k_{cat} of 18.7 ± 0.5 min^{-1}. The decline reflects a change in state of the RecA filament that is dependent on the length of homology in the duplex DNA substrate (*33, 34*). Consistent with the above results, RecA D276A and RecA D276N hydrolyzed ATP slightly faster than Wild type RecA when bound to cssDNA, with apparent k_{cat} values of 30.7 ± 0.7 min^{-1} and 31.1 ± 0.4 min^{-1}, respectively. Interestingly, the decline in rates of ATP hydrolysis observed upon addition of homologous duplex DNA was much attenuated with RecA D276A and RecA D276N. The RecA variants hydrolyzed ATP much faster than wild type RecA while catalyzing strand exchange, with apparent k_{cat} values of 26.6 ± 0.6 min^{-1} and 27.7 ± 0.4 min^{-1}, respectively, observed after dsDNA addition. Thus, RecA D276A and RecA D276N hydrolyze ATP faster than wild type RecA when bound to cssDNA and when catalyzing strand exchange, but this difference in activity is much more pronounced during DNA strand exchange.

The agarose gel pictures in figure 2C illustrate the progression of the strand exchange reactions catalyzed by the RecA variants. DNA substrates, reaction intermediates (Int), and the nicked circular duples product (NCP) are readily distinguishable and are labeled on the gels. Figure 2D presents a quantification of the reactions, carried out as described in *Experimental Procedures*. Wild type RecA, RecA D276A, and RecA D276N all produced similar amounts of reaction intermediates 5 minutes after the addition of ldsDNA, suggesting that when RecA is present in excess of cssDNA, these proteins display a similar capacity for pairing homologous DNAs. However, product DNA was visible in the reaction promoted by wild type RecA

reaction about 5 minutes before it was visible in the RecA D276A and RecA D276N reactions. Furthermore, more product DNA was present at significantly higher levels in the wild type RecA reaction at 10, 15, and 20 minutes after the addition of ldsDNA. At later time points, all three RecA variants produced similar amounts of product DNA. This data suggests that RecA D276A and RecA D276N show wild type activity in pairing homologous DNA when RecA protein is present in excess of cssDNA, but the resolution of reaction intermediates into product is slightly impeded in these mutants. Together, the reduced decline in ATP hydrolysis triggered by dsDNA addition, and the slower resolution of intermediates to products, may signal a reduction in the coupling of ATP hydrolysis to DNA strand exchange in these variant RecA proteins. In all of the activities examined in Figures 1 and 2, the effects are relatively small.

B.3.3 RecA protein filament extension on dsDNA is slowed by the D276 mutations.

The relatively slow nucleation of wild type *E. coli* RecA protein on DNA, coupled with more rapid filament extension (*20, 35-38*), would tend to create small numbers of long RecA protein filaments in the cell. This would be appropriate in an unstressed *E. coli* cell dealing with no more than one or two DNA lesions requiring recombinational DNA repair in a cell cycle (*39, 40*). We have recently made use of single-molecule tethered particle motion (TPM) experiments to study the assembly dynamics of *E. coli* and *D. radiodurans* RecA proteins on individual duplex DNA molecules by observing changes in dsDNA tether length and stiffness resulting from RecA binding (*20*). Relative to the *E. coli* RecA, DrRecA protein filaments nucleate more rapidly and extend more slowly (*20*), parameters that would tend to create more numerous and shorter filaments. We applied the TPM technology to examine the kinetic parameters of RecA filament formation with the RecA D276 variants. The DrRecA protein was included in this set of experiments to provide a comparison with a RecA that is adapted to IR resistance.

We immobilized one end of a 382 bp dsDNA segment on a surface and attached

the other end to a bead. The approach takes advantage of the properties of RecA, in which binding to dsDNA leads to a 1.5 fold increase in length and an increase in filament stiffness (*41, 42*). This in turn leads to a measurable change in the bead's Brownian motion (BM). The method of tethered particle motion (TPM) has been widely used in single-molecule studies, addressing problems such as the size of a loop formed by a repressor (*43*), the folding and unfolding state of G-quadruplex(*44*), and translocation on DNA by polymerases (*45*) and RecBCD helicase/nuclease (*46*). When applied to RecA filament formation, changes in the length and rigidity of the DNA tether that occur upon RecA binding result in an increase in the spatial extent of bead BM. This change of bead motion can be measured to nanometer precision using digital image processing techniques that determine the standard deviation of the bead centroid position in light microscope recordings (*44, 46-48*).

Typical time-courses of RecA protein assembly on a 382 bp duplex DNA are shown in the left column of Figure 3. After a 1,000-frames (33 s) recording of the BM of dsDNA tethers, a mixture of RecA protein and ATP was introduced in the reaction chamber at the time indicated with a grey bar, with a deadtime of about 10 seconds due to focus re-stabilization. The same DNA tether was recorded to monitor the RecA assembly process. In all cases, the DNA tether length stayed constant until reaching a point where an obvious, continuous increase in BM was observed. The DNA tether length then stayed constant at the higher BM amplitude, consistent with the BM value we measured separately for fully coated RecA-dsDNA filaments (data not shown).

We define the dwell time between the RecA introduction and the time where apparent BM change occurred as the nucleation time, the continuous BM increase region as the DNA extension caused by RecA assembly (the rate calculated from the slope), and the high BM value where DNA tether length stayed constant as the maximum BM achieved (*20*). Experiments and analysis were first done with wild type RecA. We then applied this method to study the RecA protein variants. The *Deinococcus* RecA protein

(DrRecA) was included in the experiments to provide an interesting point of reference. The nucleation of all of these RecA proteins onto dsDNA is highly pH-dependent (*49, 50*). To permit reasonably convenient observation in real time, these reactions were carried out at pHs below neutrality.

The results are summarized in Figure 3. The more rapid nucleation of the DrRecA protein, relative to wild type EcRecA, is seen in the first column (B) of binned results. The average nucleation time (t) is 51 seconds for DrRecA relative to 449s for EcRecA. A reduction in nucleation time to 146 seconds is also seen for the RecA D276A mutant protein. However, this is not a general pattern, as seen by the increase in t seen for D276N RecA protein to 841 seconds. A consistent patter is observed when the rate of filament extension is considered (C, the second column of binned results). The average rate of filament extension for wild type EcRecA protein is indicated by the vertical dashed line. For DrRecA, the average extension rate is slower, as shown by the leftward shift in the profile of Figure 3. A very similar leftward shift to slower filament extension rates is seen in Figure 3C for both of the RecA D276 variants. We note that the maximum Brownian motion seen with those same D276 variants is identical to that seen with wild type EcRecA protein, as shown in column D. Thus, both RecA D276 variants exhibit slower filament extension rates that could lead to the formation of shorter RecA protein filaments on DNA, at least under these experimental conditions.

In Figure 3E, we examine the pH-rate profiles for the nucleation of filament formation for all of the RecA proteins examined here. Although the data is limited to pHs between 6 and 6.5, a clear dependence of nucleation rates on pH is seen that is consistent with the dependence observed in earlier studies (*50*). We note that the slope of the best fit lines, obtained for this limited data set, is somewhat steeper for the wild type RecA protein than is seen for the D276 variants. Although more work is required, this may suggest that ionization of D276 plays a direct role in the nucleation process in the wild type protein.

B.3.4 RecA D276A and RecA D276N more efficiently catalyze strand exchange when present at sub-saturating levels relative to the cssDNA substrate.

To better characterize the strand exchange activity of RecA D276A and RecA D276N, a series of strand exchange reactions were carried out to examine reaction efficiency as a function of RecA protein concentration. As a RecA monomer binds 3 nucleotides (*41, 51-53*), 100%, 75%, 50%, 33%, 20%, and 10% saturation of 10 μM M13mp18 cssDNA were defined as 3.33 μM, 2.50 μM, 1.67 μM, 1.10 μM, 0.67 μM, and 0.33 μM RecA. RecA protein was allowed to incubate with cssDNA for 10 minutes, 3 mM ATP and 1 μM SSB were added and incubated with RecA for 10 minutes, and then 20 μM M13mp18 ldsDNA were added to initiate the reaction. The reactions were allowed to proceed for 90 minutes before being deproteinized and resolved on a gel for quantification.

At 100% and 75% saturation, all three RecA proteins produced similar amounts of strand exchange products (Figure 4A). Interestingly, at 10%, 20%, 33%, and 50% saturation of cssDNA, At lower RecA protein concentrations, RecA D276A and RecA D276N consistently produced more strand exchange products than wild type RecA in three independent experiments (Figure 4A). The difference is most striking at 10% saturation, where RecA D276A and RecA D276N produce approximately 3 fold more product than wild type RecA. As saturation increases, the fold difference between wild type RecA and the mutants becomes smaller, but remains significant until 75% saturation. RecA D276A and RecA D276N exhibited essentially identical levels of activity in this assay, generating the same amount of product within error at all levels of saturation. Most important, the RecA D276 variants generated substantially higher levels of product than wild type RecA protein at subsaturating RecA protein concentrations. The results suggest that the mutant proteins function better when present as short filaments.

To explore the mechanism of this difference in catalytic activity, a time course

strand exchange reaction was carried out with RecA present at 1.1 μM, or 33% saturation of the available cssDNA. The results from three separate experiments are quantified in Figure 4B. Consistent with the previous experiments, RecA D276A and RecA D276N produced approximately 1.65 fold more product than wild type RecA after the reactions had proceeded for 90 minutes. Even though the final yield of products was higher, RecA D276A and RecA D276N catalyzed the resolution of strand exchange intermediates into product DNA occurred somewhat more slowly than wild type RecA, as seen in Figure 2. The yield advantage was nevertheless evident at all stages of the reaction. At 20 minutes after the addition of ldsDNA, RecA D276A and RecA D276N produced 1.85 fold more intermediates than wild type RecA. Furthermore, these intermediates are not present at high levels at times outside of the times when intermediates are at maximum levels in the normal wild type reaction, as might be expected if RecA D276A and RecA D276N were increasing yield by promoting multiple rounds of strand exchange. Thus, the data presented above supports the hypothesis that RecA D276A and RecA D276N more efficiently catalyze the pairing of homologous DNA strands when RecA is present at sub-saturating levels of cssDNA.

B.3.5 The DNA strand exchange activity of RecA D276A and RecA D276N is much less inhibited by ADP.

An underappreciated property of the wild type RecA protein is its sensitivity to inhibition by ADP. If no ATP regeneration system is included in the reaction, and ADP levels are allowed to rise, the RecA filaments will disassemble entirely as the ADP/ATP ratio approaches 1.0 (*54-56*). Effective DNA strand exchange will halt well before that point. Under normal logarithmic growth conditions, ADP is rarely present at levels sufficient to affect RecA function (*57*). However, heavy irradiation of cells has the potential to compromise metabolic processes, leading to unpredictable alterations in the available nucleotide pools. ADP inhibition of RecA protein has the potential to be a significant barrier to recombinational DNA repair under these circumstances.

We examined the effects of ADP on our D276 variant RecA proteins. Strand exchange reactions were carried out without an ATP regeneration system, and were initiated in the presence of varying levels of ADP. First, 2 μM RecA was incubated with 5 μM M13mp18 cssDNA for 10 minutes. Next, 5 mM ATP and 0.5 μM SSB were added and incubated for 10 minutes to allow formation of contiguous RecA filaments. Then, 10 μM M13mp18 ldsDNA and 0, 1, 2, 3, 4, or 5 mM ADP were added to initiate strand exchange. Strand exchange reactions were allowed to proceed for 60 minutes before being deproteinized and resolved on a gel for quantification.

If no ADP was added with ldsDNA, wild type RecA converted 48% of the DNA substrates to nicked circular product (Figure 5). As the concentration of ADP added increased, the amount of product produced by wild type RecA decreased rapidly (Figure 5). The strand exchange activity of wild type RecA was completely inhibited by the addition of 4 mM ADP to the reaction, with trace amounts of activity seen when 3 mM ADP was added (Figure 5). Strikingly, RecA D276A and RecA D276N produced considerably more product than wildtype RecA at all ADP concentrations (Figure 5). This difference in activity was most striking when 2 mM ADP was added to the reactions, as RecA D276A and RecA D276N produced approximately 6 times more product than wild type RecA. The strand exchange activity of RecA D276A and RecA D276N was completely inhibited by the addition of 5 mM ADP, with trace amounts of activity seen when 4 mM ADP was added to the reactions (Figure 5). This data provides clear evidence that the DNA strand exchange activity of RecA D276A and RecA D276N is much less inhibited by ADP than the activity of wild type RecA.

B.3.6 The ATPase activity of RecA D276A and RecA D276N during strand exchange is less inhibited by ADP than that of Wild type RecA.

Next, the effect of ADP on the ATP hydrolysis activity of RecA D276A and RecA D276N during strand exchange was examined. RecA protein was allowed to hydrolyze ATP while bound to cssDNA for five minutes, and then homologous ldsDNA and

varying concentrations of ADP were added. ATP hydrolysis was monitored in real time using a method that does not rely on ATP regeneration (see Experimental Procedures). DNA and protein concentrations were reduced 3-5 fold from the strand exchange reactions described above to allow for measurable steady state rates of ATP hydrolysis, but the same ratios of RecA:ATP:ADP were employed in these assays. ADP was again added at the same time as the linear duplex DNA was added to initiate the reaction.

When ldsDNA but no ADP was added to filaments of Wild type RecA, a steady state rate of ATP hydrolysis was observed (Figure 6A). However, as the concentration of ADP added was increased, an increasingly abrupt drop in the rate of ATP hydrolysis was observed (Figure 6A). As RecA must be bound to DNA to hydrolyze ATP, this observation suggests that the addition of ADP leads to disassembly of RecA filaments. The apparent extent of filament disassembly caused by the addition of ADP was quantified by determining the instantaneous rate of ATP hydrolysis at the point 7 minutes after the addition of ADP and ldsDNA, as described in *Experimental Procedures*. These rates were then expressed as the percentage of the rate of the reaction in which no ADP was added (Figure 6B). The decline in the rate of observed rate of ATP hydrolysis at time 7 minutes was quite rapid for the wild type protein (Figure 6B). Only 10 % of the uninhibited rate of hydrolysis remained if 0.9 mM ADP was added, and the level of ATP hydrolysis was essentially negligible at time 7 if 1.5 mM ADP was added (Figure 6B).

The effect of ADP on the ATP hydrolytic reaction promoted by RecA D276A or RecA D276N was much less severe (Figure 6B). Significant ATPase activity remained even when 1.2 mM ADP was included in the reaction..

RecA filaments on single-stranded DNA exist in different states before and after strand exchange is initiated (*33, 34*). We thus again examined the effect of ADP on RecA D276A and RecA D276N filaments using the protocol of Figure 6, but in the absence of added linear duplex DNA. As shown in Figures 7A and Figure 7B. the differences in the

effects of ADP between the wild type and variant proteins was much reduced when the proteins were not promoting DNA strand exchange. This was due primarily to a reduction in the inhibitory effects of ADP for the wild type protein in the absence of ongoing strand exchange (compare Figures 6B and 7B). The inhibitory effects of ADP on the RecA D276 mutants was similar in both sets of experiments. The results not only signal a reduction in ADP hydrolytic effects for the RecA variants. Combined with the reduced decline in ATP hydrolysis when strand exchange is initiated (Figure 2), the results suggest that the D276 mutants undergo a less pronounced change in state when duplex DNA is added to initiate DNA strand exchange than is evident for the wild type protein.

B.4 Discussion

Upon irradiation to 3,000 Gy, a bacterial cell must deal with hundreds of DNA double strand breaks as well as substantial damage to other cellular components. In principle, the damage could include alterations in metabolic processes that could affect the pool of metabolites, including nucleotides, available for genome reconstitution. Directed evolution produces strains of *E. coli* that can deal with this situation (described in Chapter 2), and RecA mutations affecting the Asp residue at position 276 contribute substantially to IR resistance in many of those strains. Our results show that the D276 alterations adapt RecA function to IR exposure in at least three important ways.

First, the mutations slow the growth of RecA filaments on dsDNA after nucleation, under the conditions utilized here for single-molecule measurements. If RecA filament growth is slowed under all conditions, this would lead to the formation of shorter and more numerous RecA filaments in the cell. The result would allow formation of RecA filaments at more locations to effect repair of numerous double strand breaks.

Second, the RecA D276A/N mutants promote DNA strand exchange more

efficiently than the wild type proteins when they are present in sub-saturating amounts relative to the DNA substrates. As the improvement in strand exchange product yield is preceded by a large improvement in the generation of strand exchange intermediates, the advantage of the mutant proteins appears to lie at the initiation of strand exchange. With the standard strand exchange reaction we use, productive strand exchange must be initiated via pairing at the end of the linear duplex DNA substrate, requiring the presence of part of a RecA filament at the complementary sequences on the ssDNA. At a particular sub-saturating RecA protein concentration and at a particular moment, the RecA mutants should be no more likely to occupy a particular stretch of DNA than the wild type. The enhanced capacity of the mutant RecA proteins to initiate strand exchange may thus reflect more dynamic filaments that can utilize end-dependent assembly and disassembly processes to more rapidly re-position the shorter filaments on the DNA. Alternatively, the mutant proteins may simply promote the initial steps of DNA pairing more efficiently than the wild type protein, making initiation more likely when the filament is properly positioned. A much broader analysis of the properties of these mutants is needed to distinguish between these possibilities.

Third and finally, the mutant proteins are substantially less inhibited by ADP during strand exchange than is the wild type protein. It is difficult to assess all of the effects of ionizing radiation on cellular functions. In addition, it is likely that those effects would vary from cell to cell depending on which critical macromolecules were randomly inactivated. In principle, the capacity of the cell to generate ATP could be compromised in a number of different ways. Any lesions that led to an increase in cellular ADP concentrations would hamper recombinational DNA repair in a wild type *E. coli* cell, since RecA function would be affected. At a minimum, this particular property change in the D276 mutants expands the environmental space in which RecA can operate effectively.

In sum, the observed alterations in RecA function brought about the D276 mutations exhibit a molecular logic that appears to conform in a reasonable way to the molecular requirements for genome reconstitution after extreme exposure to ionizing radiation. There may be additional effects of these mutations that will be elucidated by more detailed studies.

In the course of this work, we considered several molecular explanations for the effects of the D276 alterations. D276 resides in the C-terminal domain of RecA, a region of the protein that is not involved in DNA or ATP-binding but is observed to make large conformational shifts which could represent a switch from an active and inactive conformation (*41, 42, 58*). The carboxyl group of D276 is approximately 20 angstroms from the 2′ hydroxyl of the ribose moeity of the ADP-AlF_4-Mg complex in the X-ray crystal structure of RecA bound to ssDNA (*41*). Thus, any effect of mutation in this residue on nucleotide binding must be indirect. D276 resides within the helix closest to the connection between the C-terminal domain and the core domain of RecA, and this helix is connected to a turn that makes direct contacts with the adenine and ribose moieties of the ATP in the active site (*41*). Furthermore, D276 forms a salt bridge with K302, which resides within a more distal helix from the core. Thus, It is possible that this salt bridge is responsible for transmitting conformational shifts from the core nucleotide-binding domain to the C-terminal domain of RecA. Mutating an aspartate to either an alanine or an asparagine would eliminate this salt bridge at neutral pH, and could thus render the RecA D276A and RecA D276N variants less able to shift from a strand exchange proficient conformation with a high affinity for ATP to a strand-exchange inactive conformation with a high affinity for ADP.

If the K302-D276 salt bridge were important for transmission of conformational changes related to ADP inhibition, one might expect that mutating K302 to an inert residue would cause a similar defect in nucleotide inhibition. An experiment similar to those in Figure 5 revealed that a K302A RecA variant is inhibited by ADP during strand

exchange to an extent similar to that of wild type RecA (Joseph R. Piechura, unpublished data). This data thus does not support, but does not exclude this structural hypothesis. It is possible that the K302A mutation is more readily compensated for by slight shifts in folding and formation of new salt bridges or hydrogen bonds than the D276 mutations. It is also worth noting that the RecA protein from the extremely radiation resistance bacterium *Deinococcus radiodurans* (DrRecA) contains the inert residue glycine at the same structural position as *E. coli* RecA D276, and thus the DrRecA protein lacks this aspartate to lysine salt bridge in its C-terminal domain. However, the strand exchange activity of DrRecA is inhibited by ADP to a similar extent as the *E. coli* wild type RecA (Joseph R. Piechura, unpublished data).

Another possible, but not mutually exclusive mechanism for the efficient strand exchange phenotype of RecA D276A and RecA D276N is that these mutations affect DNA pairing. The salt bridge with residue K302 allows for less restricted movement of this residue. K302 is part of a large cluster of nine positively charged surface residues found between residues 270 and 328 (*59*). Within this cluster, K302 is one of two Lysines that are particularly well conserved among bacterial RecA proteins. The positively charged cluster has been implicated in having a role in capturing homologous dsDNA for formation of joint molecules (*60*). Without this salt bridge, the K302 residue in the RecA D276A and RecA D276N variants may have more freedom to move in solution and thus aid these proteins in more readily initiating DNA pairing.

In a broader sense, our results speak to the malleability of RecA protein function. While not unique to RecA, this is a feature of the protein that is increasingly evident in comparisons of RecA proteins from different bacteria. The variations are sometimes subtle and sometimes not (*19, 61-64*). The same theme is highlighted in recent work on the *E. coli* RecA protein (*65*). Most important, the variations do not reflect an evolutionary approach to some common and most efficient platform for DNA pairing and strand exchange. Homologous genetic recombination is both a benefit and a

potential hazard to every cell. A maximally efficient recombinase is not the ideal for all cells, and possibly not the ideal for any cell. The RecA protein structure, shared by RecA homologs in virtually all organisms, is a highly adaptable scaffold readily tailored by evolution to the requirements for genome maintenance in each particular organism.

B.5 Materials and Methods

DNA Substrates – Circular single-stranded DNA and supercoiled double-stranded DNA were obtained from bacteriophage M13mp18 using CsCl gradients as described previously (*33*, *66*, *67*). Linear double-stranded DNA for strand exchange reactions was generated by digestion of supercoiled DNA with the restriction enzyme PstI according to the manufacturers instructions. Concentration of DNA substrates was determined using absorbance at 260 nm and the conversion factors 108 μM A_{260}^{-1} for single-stranded DNA and 151 μM A_{260}^{-1} for double-stranded DNA. DNA concentrations in this work are given in total nucleotides.

Protein Purification – Native wild type *E. coli* RecA and single-stranded DNA binding protein were purified as described previously (*68*). Protein concentrations were determined using absorbance at 280 nm and the extinction coefficients 2.23×10^4 M^{-1} cm^{-1} for RecA (*69*) and 2.83×10^4 M^{-1} cm^{-1} for SSB (*70*).

E. coli K12 strain STL2669 (Δ(*recA-srlR*)306:*Tn10xonA2*(*sbcB*$^-$)), a nuclease-deficient strain which was graciously provided by Susan T. Lovett (Brandeis University), was transformed with pT7pol26 and either pEAW416, encoding RecA D276A, or pEAW 446, encoding RecA D276N (*71*). Four liters of Luria-Bertini media supplemented with 100 μg/mL ampicillin and 40 μg/mL kanamycin were inoculated with the appropriate strain and incubated at 37° C with shaking. When the cultures were at an $A_{600\ nm}$ of approximately 0.65, protein expression was induced by the addition of isopropyl 1-thio-

β-D-galactopyranoside to a final concentration of 0.4 mM. The cells were allowed to grow for 3 more hours at 37° C with shaking and then were collected by centrifugation. Cell pellets were flash-frozen in liquid N_2 and then adjusted to a concentration of 20 % (w/v) in a solution of 250 mM Tris-HCl (80% cation, pH 7.7) and 25 % (w/v) sucrose and allowed to thaw at 4° C overnight. All subsequent purification steps were carried out at 4° C. Lysis of the cells containing RecA D276N was initiated by the addition of lysozyme to a final concentration of 2.5 mg/mL. EDTA was added to both cell solutions to a final concentration of 7 mM, and cells were lysed by sonication. Cell debris was removed by centrifugation, and RecA protein was precipitated by the addition of 0.111 mL of 5 % (w/v) polyethyleneimine per mL of lysate. The precipitate was pelleted by centrifugation, washed with R buffer (20 mM Tris-HCl (80% cation, pH 7.7), 10% glycerol, 0.1 mM EDTA, 1 mM DTT) for the RecA D276N pellet or R + 100 mM ammonium sulfate for the RecA D276A pellet. Then, protein was extracted from the pellet with R buffer + 300 mM ammonium sulfate. RecA was then precipitated by the addition of ammonium sulfate to a final concentration of 0.28 g/mL The precipitate was pelleted by centrifugation and washed with R buffer + 0.28 g/mL ammonium sulfate. RecA was then resuspended and dialyzed into R buffer + 100 mM KCl. RecA D276A was further purified using successive Q-sepharose, SP-sepharose, and Sephacryl S-300 gel filtration chromatography steps. RecA D276N was further purified using successive DEAE-sepharose, ceramic hydroxyapatite, Source 15-Q, and Sephacryl S-300 gel filtration chromatography steps. The proteins were determined to be >95% pure as determined by SDS-PAGE, dialyzed into R buffer, flash-frozen in liquid N_2, and stored at -80 ° C. RecA proteins were determined to be free of detectable nuclease activity. The concentrations of RecA D276A and RecA D276N were determined using the wild type RecA protein extinction coefficient.

ATP Hydrolysis (ATPase) Assay – A coupled spectrophotometric enzyme assay (*72, 73*) was employed to measure the rate of ATP hydrolysis of RecA. ADP formed by

ATP hydrolysis was converted back to ATP by an ATP regenerationg system of 10 units/mL of pyruvate kinase and 3.5 mM phosphoenolpyruvate. Pyruvate produced by the regeneration reaction was then utilized by a coupling system of 10 units/mL of lactate dehydrogenase and 1.5 mM NADH to oxidize NADH to NAD+. The conversion of NADH to NAD+ was monitored using absorbance at 380 nm, and the extinction coefficient of $\varepsilon_{380} = 1.21 mM^{-1} cm^{-1}$ was used to convert absorbance to amount of ATP hydrolyzed, assuming a stoichiometry of one NADH molecule oxidized per ATP molecule hydrolyzed. Reactions were carried out at 37° C in buffer containing 25 mM Tris-OAc (80% cation), 10 mM magnesium acetate, 3 mM potassium glutamate, 1 mM DTT, and 5 % (w/v) glycerol. These assays were carried out in a Perkin Elmer Lambda 650 UV/Vis spectrometer equipped with a temperature controller and a 9-position cell changer. Cell path length was 1.0 cm and bandpass was 2 nm. All ATPase assays but the strand exchange reactions not involving the addition of ADP were carried out in this spectrophotometer with the same cuvettes. 5 μM M13mp18 cssDNA was incubated for 10 minutes with either 3 μM RecA or 0.5 μM SSB and 3 mM ATP, and then ATP hydrolysis was initiated by the addition of 0.5 μM SSB and 3 mM ATP or 3 μM RecA to the reactions, respectively.

DNA Strand Exchange Reactions – DNA strand exchange reactions were carried out at 37° C in the same buffer used for ATPase assays. The basic reaction scheme involved incubation of RecA with M13mp18 cssDNA for 10 minutes. Next, ATP and SSB were added to the reaction and incubated for 10 minutes. DNA strand exchange was then initiated with the addition of homologous M13mp18 ldsDNA. Exact concentrations are delineated in figure legends. Ten μL time points were removed and deproteinized by addition of 2 μL of 10% (w/v) SDS and 3 uL of 15% (w/v) Ficoll and 20 mM Tris-OAc (80% cation, pH 7.8). Time points were then resolved on a 0.8 % (w/v) agarose gel and visualized with ethidium bromide staining and exposure to UV light. Gel images were captured using a Fotodyne FOTO/Analyst® CCD camera, PC Image acquisition

software, and a FOTO/Convertible dual transilluminator. Images were inverted using PC Image software and then DNA band intensity was quantified using TotalLab version TL100 software from Nonlinear Dynamics. Product and intermediate amounts are expressed as a percentage of the total intensity of ldsDNA, intermediates, and nicked circular product DNA in the respective lane. If ADP was not added to reactions, an ATP regeneration system of 3 mM phosphoenolpyruvate and 10 units/mL of pyruvate kinase was included. Strand exchange reactions in which ATP hydrolysis was measured and ADP was not added contained an ATP regeneration system of 4 mM phosphoenolpyruvate and 10 units/mL of pyruvate kinase and a coupling system of 4.5 mM NADH and 10 units/mL of lactate dehydrogenase. These reactions were carried out in a Varian Cary 300 dual beam spectrophotometer equipped with a temperature controller and a 12-position cell changer. Cell path length was 0.5 cm and bandpass was 2 nm. ATP hydrolysis was followed using absorbance at 380 nm as described above. Strand exchange reactions involving the addition of ADP were normalized by converting the percentage of nicked circular product DNA produced in a given reaction to the percentage of the amount of product formed in the reaction in which no ADP was added. These values were then fit to a dose response (inhibition) curve with variable slope in Prism software to determine IC50 values.

ATPase Assays involving the addition of ADP – As the NADH-based coupling system for ATPase assays described above relies on the regeneration of ADP to produce a signal, the Enzcheck® Phosphate Assay Kit from Molecular Probes was employed to measure the ATPase activity of RecA when ADP was added to reactions. As inorganic phosphate is released from RecA after an ATP hydrolysis event, the coupling enzyme purine nucleoside phosphorylase utilizes the phosphate ion to convert 2-amino-6-mercapto-7-methyl purine riboside (MESG) to ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine, which absorbs strongly at 360 nm. Assays were conducted at 37° C in the same buffer used for the NADH-based ATPase assays and contained 0.2

mM MESG and 1 unit/μL of purine nucleoside phosphorylase. First, 1 μM M13mp18 cssDNA was incubated with 0.6 μM RecA protein for 10 minutes. Then, ATP hydrolysis was initiated by the addition of 1 mM ATP and 0.1 μM SSB. ADP was then added five minutes after the addition of ATP. If strand exchange reactions were carried out, 2 μM M13mp18 ldsDNA was added at the same time as ADP. Absorbance was measured at 360 nm during the reaction in a Perkin Elmer Lambda 650 UV/Vis spectrometer as was used for the NADH-based assays. A standard phosphate curve was generated in order to convert absorbance at 360 nm to the concentration of phosphate generated by ATP hydrolysis. Briefly, varying amounts of phosphate were added to reaction buffer containing the coupling system and incubated at 37° C until aborbance readings at 360 nm were constant. A linear fit line was applied to the linear portion of the graph of phosphate concentration vs. A_{360} . This analysis generated a conversion factor of 0.0132 A_{360} Pi^{-1} (μM), and revealed that the coupling system produced a linear signal up to a concentration of approximately 100 μM Pi.

Data was further analyzed in the program CurveExpert. If no ADP was added to reactions, the rate of ATP hydrolysis was determined by applying a linear fit to the data. If ADP was added, a polynomial fit (n=6) was applied to the data and differentiated at time 7 minutes after the addition of ADP. Data from three independent experiments was averaged and these values were converted to the percentage of the uninhibited rate of ATP hydrolysis. Standard deviation values were determined using standard statistical methods.

Single-molecule experiments For real-time RecA nucleoprotein assembly observations at the single-molecule level, experimental details are the same as previously described (*20*). Duplex DNA with a length of 382 bp was prepared by polymerase chain reaction (PCR) using a 5′-digoxigenin-labeled primer (5′-dig-ACTACGATACGGGAGGGC), a 5′-biotin-labeled primer (5′-biotin-TGAGTGATAACACTGCGGC) and pBR322 templates. PCR products were gel purified

(Qiagen). Individual dsDNA molecules were tethered on the antidigoxigenin-coated slide, and the distal end of the DNA molecules were attached to streptavidin-labeled beads for visualization under optical microscope. Individual tethered DNA molecules were screened and verified as single DNA tethers by both the BM amplitude of the 382 bp DNA and the symmetric BM trajectory (x/y ratio ~ 1). To observe the RecA nucleoprotein assembly process, a 40 μL mixture of RecA (2 μM) with ATP (2 mM, Aldrich), 10 units/mL pyruvate kinase (Roche) and 3 mM phosphoenolpyruvate (Aldrich) in the single-molecule buffer (25 mM MES, pH 6.20, 10 mM magnesium acetate, 3 mM potassium glutamate, 1 mM dithiothreitol, 5 % glycerol, and 1 mg/mL BSA) was flowed into the reaction chamber. All reactions were conducted in 22^{o}C. Flow deadtime is 10 s.

Figures

Figure 1. RecA D276A and RecA D276N hydrolyze ATP faster when bound to cssDNA and more readily displace the SSB protein from cssDNA than Wild type RecA.

The DNA-dependent RecA-catalyzed hydrolysis of ATP was monitored. Reactions were carried out as described in *Experimental Procedures*. Solid lines represent reactions in which RecA protein was added first to a solution containing cssDNA, and dashed lines represent reactions in which SSB and ATP were added first to cssDNA. The 0 time point corresponds to either the time of addition of ATP and SSB (solid lines) or the addition of RecA protein (dashed lines).

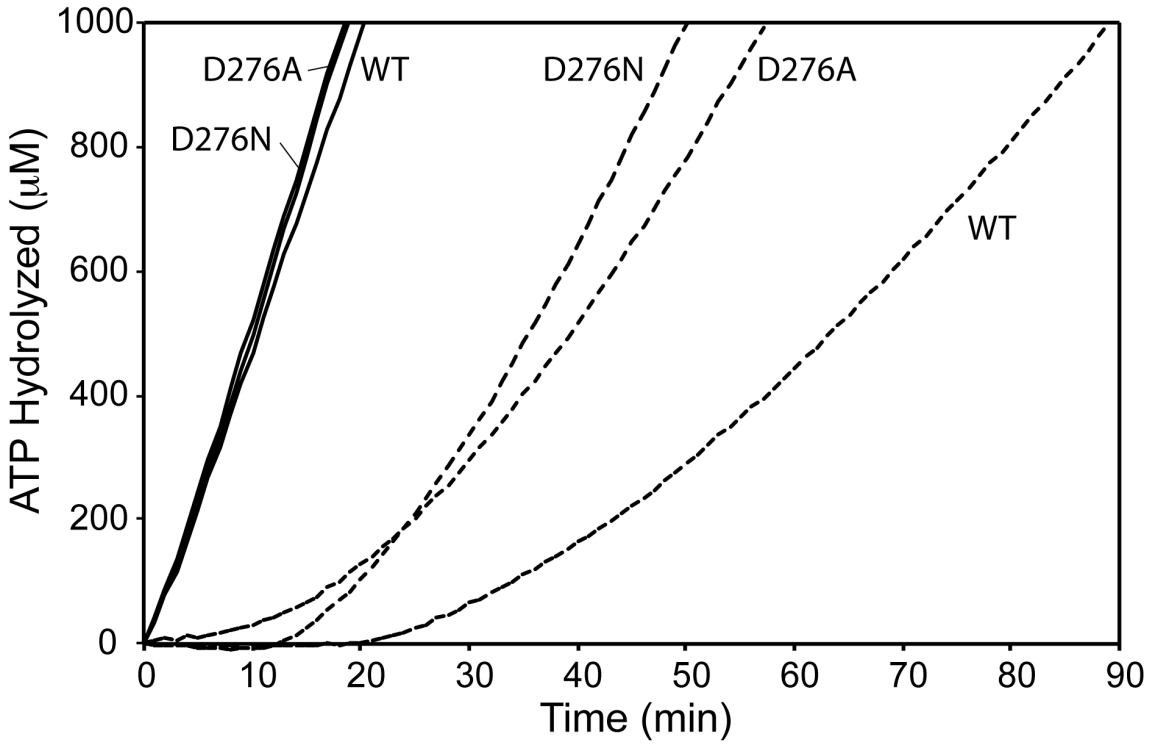

1000
800
600
400
200
0
ATP Hydrolyzed (μM)
0
10
20
30
40
50
60
70
80
90
Time (min)
D276A
WT
D276N
D276N
D276A
WT

Figure 2. RecA D276A and RecA D276N hydrolyze ATP faster while catalyzing strand exchange but resolve strand exchange intermediates into products slower than Wild type RecA.

A, ATP hydrolysis was monitored. Reactions were carried out as described in *Experimental Procedures*. First, 3 μM RecA protein was added to 5 μM M13mp18 cssDNA and incubated for 10 minutes. Then, ATP hydrolysis was initiated by the addition of 3 mM ATP and 0.5 μM SSB. RecA protein was allowed to hydrolyze ATP for twenty minutes. Strand exchange reactions were then initiated by the addition of 10 μM homologous M13mp18 ldsDNA (time marked with an arrow). Time 0 corresponds to the time of addition of ATP and SSB. *B*, Agarose gel pictures displaying the progression of the strand exchange reactions. DNA bands are labeled as seen in the reaction schematic. Time 0 corresponds to the addition of ldsDNA to the reactions. *C*, The amount of nicked circular product DNA was quantified at each time point as the percentage of the total intenstity of ldsDNA, intermediates, and product DNA in the corresponding lane. Wild type RecA reaction points are shown as squares, RecA D276A reaction points are shown as diamonds, and RecA D276N reaction points are shown as triangles. Values represent the averages of three independent experiments, with vertical error bars displaying the standard deviation of these values.

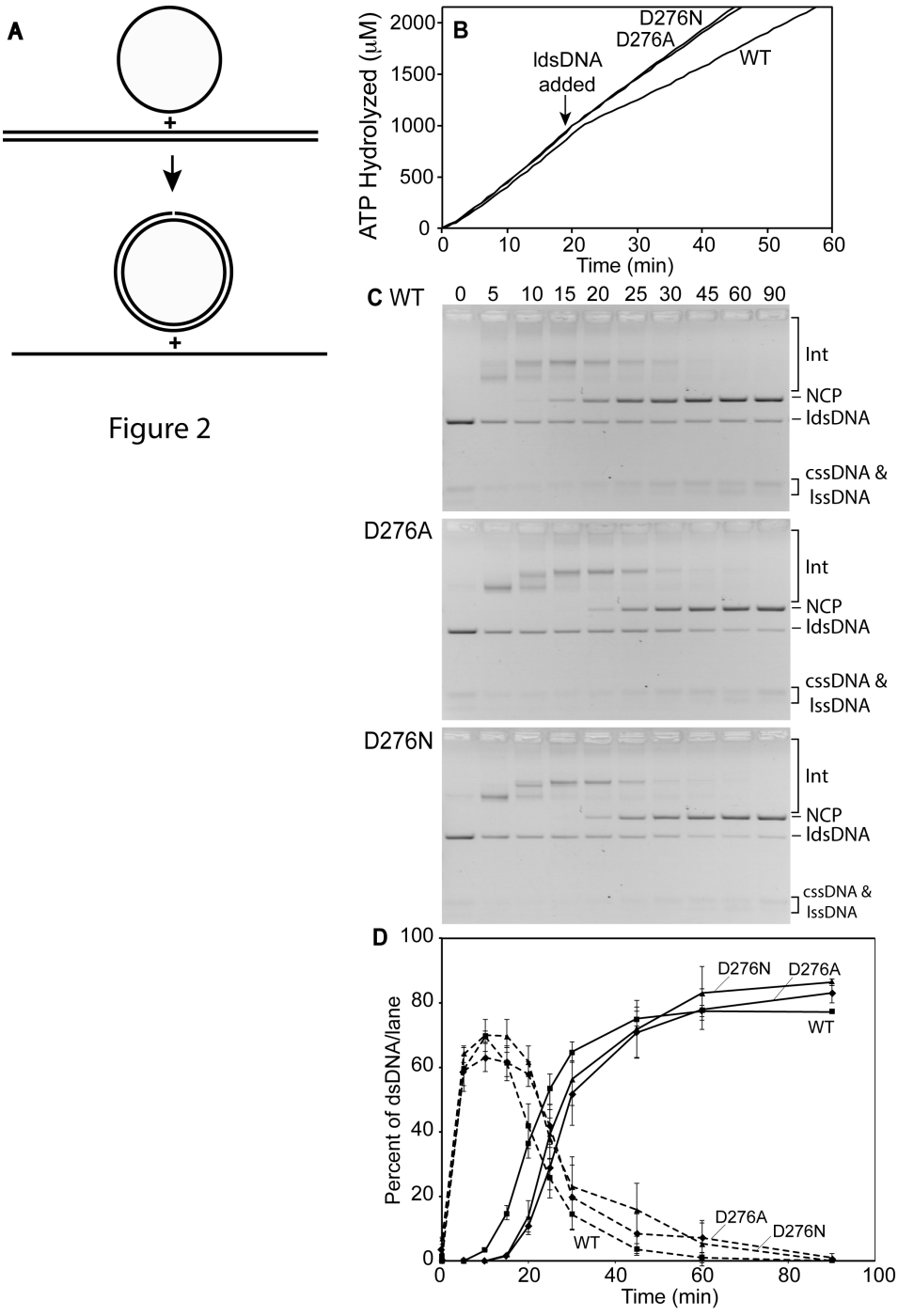

A
+
+
B
ATP Hydrolyzed (μM)
2000
1500
1000
500
0
0
10
20
30
40
50
60
Time (min)
D276N
D276A
WT
ldsDNA
added
C WT
0 5 10 15 20 25 30 45 60 90
Int
NCP
ldsDNA
cssDNA &
lssDNA
D276A
Int
NCP
ldsDNA
cssDNA &
lssDNA
D276N
Int
NCP
ldsDNA
cssDNA &
lssDNA
D
Percent of dsDNA/lane
100
80
60
40
20
0
0
20
40
60
80
100
Time (min)
D276N
D276A
WT
D276A
D276N
WT

Figure 2

Figure 3. Single-molecule TPM observation of nucleoprotein filament assembly process of EcRecA, D276N mutant, D276A mutant and DrRecA.

(A). Exemplary time-courses of the filament assembly process on 382 bp dsDNA at pH 6.20. The RecA mixture was flowed in during the gray bar. Before RecA forms a stable nucleus, the BM amplitude stays constant, the same as that before the gray bar. A stable nucleus is followed by a cooperative extension process, which is represented by the continuous BM increase. After the filament assembly is finished, the BM amplitude reaches a maximum plateau value. (B). Accumulated histograms of observed nucleation times (the point in each trial in which BM begins to increase). (C). Histograms of observed extension rates. (D). Histograms of observed maximum BM amplitudes. The histograms include 116, 146, 293, and 103 molecules of EcRecA, D276N, D276A and DrRecA, respectively.

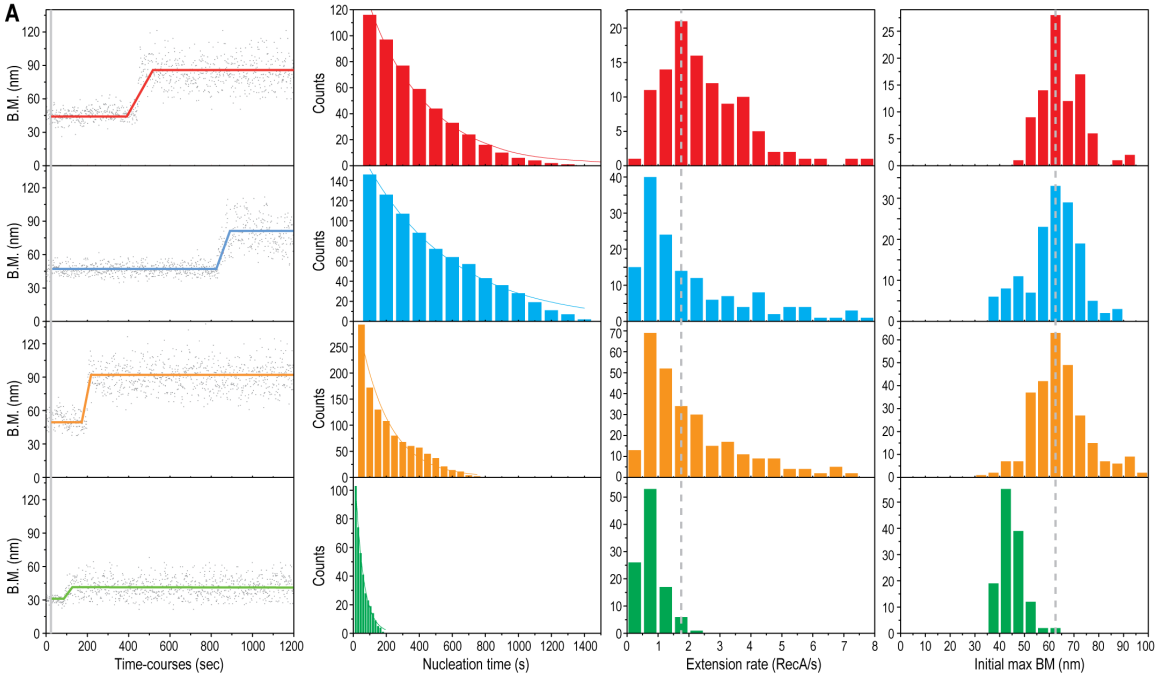

A
B.M. (nm)
Time-courses (sec)
Counts
Nucleation time (s)
Extension rate (RecA/s)
Initial max BM (nm)
B
Log k
pH
EcRecA
D276N
D276A
DrRecA

Figure 4. RecA D276A and RecA D276N more efficiently catalyze strand exchange reactions than Wild type RecA when present at sub-saturating levels of cssDNA.

A, Strand exchange reactions were carried out as described in *Experimental Procedures* with RecA protein present at 10%, 20%, 33%, 50%, 75%, and 100% saturation of cssDNA. Reactions were deproteinized 90 minutes after the addition of ldsDNA, resolved on an agarose gel, and the amount of NCP DNA was quantified as described in Figure 2. Wild type RecA points are shown as squares, RecA D276A points are shown as diamonds, and RecA D276N points are shown as triangles. Values of percent nicked circular product represent the averages of three independent experiments, with vertical error bars displaying the standard deviation of these values. Reactions contained varying amounts of RecA protein, 10 μM M13mp18 cssDNA, 20 μM M13mp18 ldsDNA, 3 mM ATP, and 1 μM SSB. *B*, The amount of intermediates or product DNA at each time point from a strand exchange time course in which each RecA protein is present at 33% saturation of the circular ssDNA, quantified as described in Figure 2. Wild type RecA points are shown as squares, RecA D276A points are shown as diamonds, and RecA D276N points are shown as triangles. Points connected with a dashed line display the amount of reaction intermediates present at each time point, while points connected with a solid line display the amount of product DNA present at each time point. Values displayed represent the averages of three independent experiments, with vertical error bars displaying the standard deviation of these values.

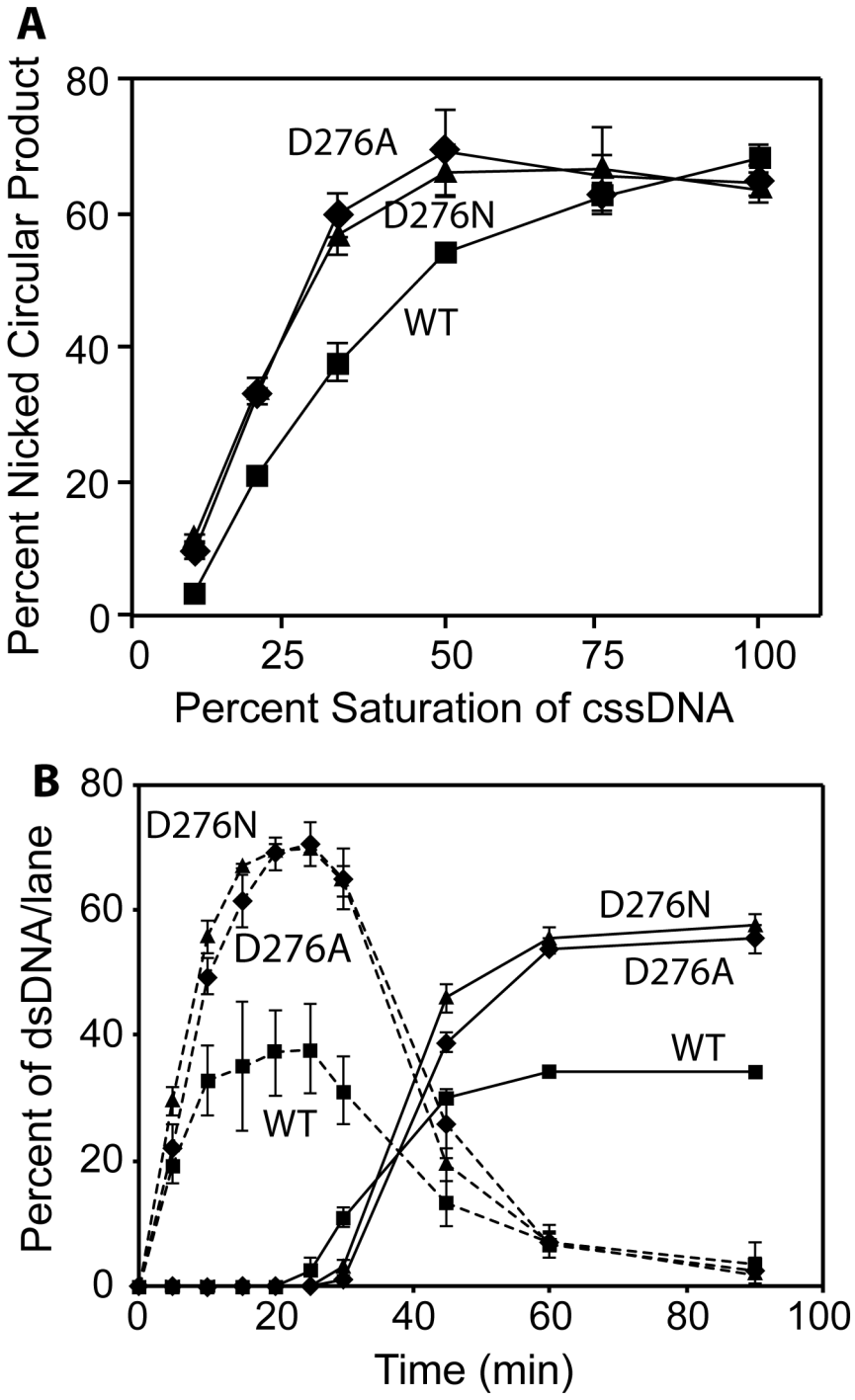

A
Percent Nicked Circular Product
Percent Saturation of cssDNA
D276A
D276N
WT
0
20
40
60
80
0
25
50
75
100
B
Percent of dsDNA/lane
Time (min)
D276N
D276A
WT
D276N
D276A
WT
0
20
40
60
80
100

Figure 5. The strand exchange activitity of RecA D276A and RecA D276N is less inhibited by ADP than that of Wild type RecA.

Strand exchange reactions were carried out as described in *Experimental Procedures*, with varying concentrations of ADP. Reactions contained 5 μM M13mp18 cssDNA, 2 μM RecA protein, 5 mM ATP, and 0.5 μM SSB. To initiate strand exchange, 10 μM M13mp18 ldsDNA was added to the reactions with 0, 1, 2, 3, 4, or 5 mM ADP. Reactions were deproteinized 60 minutes after the addition of ldsDNA, resolved on an agarose gel, and the amount of product DNA was quantified as described in Figure 2. Values displayed on the graph are averages of data from three independent experiments, and vertical error bars represent the standard deviation of these values. Wild type RecA points are shown as squares, RecA D276A points are shown as diamonds, and RecA D276N points are shown as triangles.

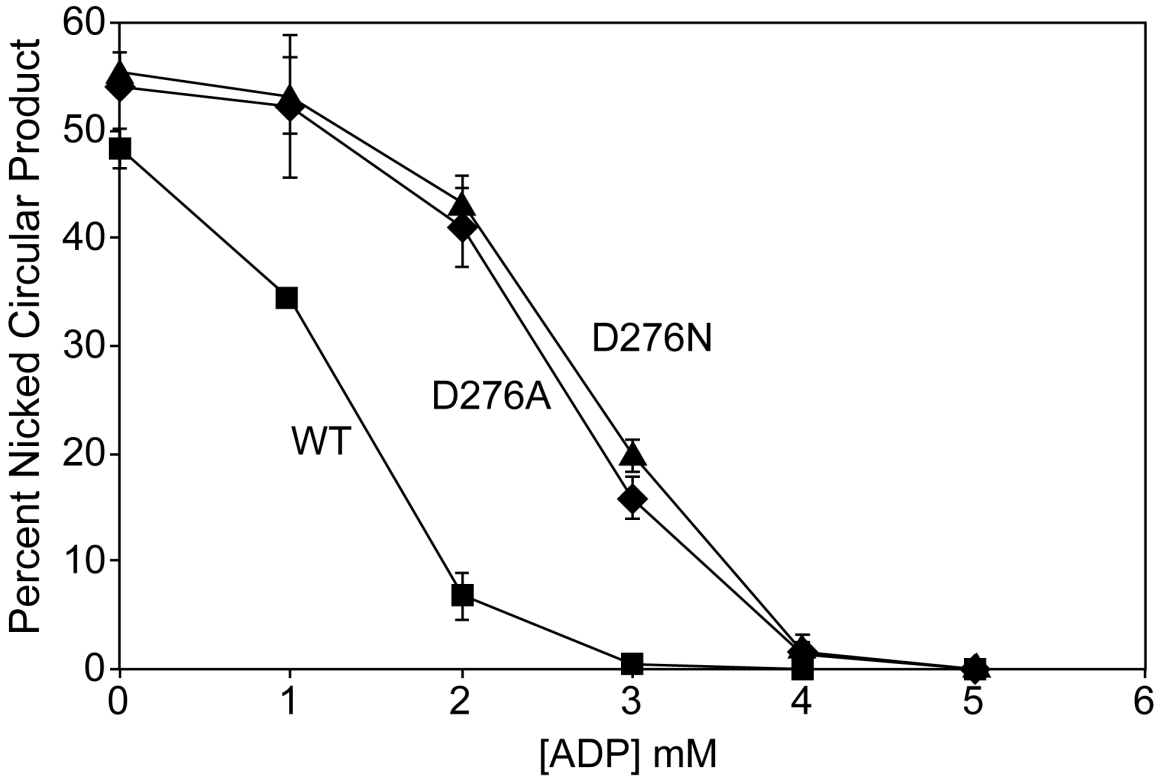

Percent Nicked Circular Product
60
50
40
30
20
10
0
0
1
2
3
4
5
6
[ADP] mM
D276N
D276A
WT

Figure 6. The ATPase activity of RecA D276A and RecA D276N during strand exchange is less inhibited by ADP than that of Wild type RecA.

A, Reactions were carried out as described in *Experimental Procedures*, and ATP hydrolysis was monitored. First, 0.6 μM RecA protein was allowed to hydrolyze ATP in the presence of 1 μM M13mp18 cssDNA, 1.5 mM ATP, and 0.1 μM SSB for five minutes, and then 2 μM homologous M13mp18 ldsDNA and 0, 0.3, 0.6, 0.9, 1.2, or 1.5 mM ADP were added. Graphs are labeled with the variant of RecA protein used in the reactions, and reaction lines are labeled with the concentration of ADP added. Time 0 corresponds to the addition of ldsDNA and ADP. *B*, ATP hydrolysis data was fit with a polynomial curve and this fit was differentiated at time 7 minutes after the addition of ldsDNA and ADP to determine the rate of hydrolysis of ATP at that time. Values from three independent experiments were averaged, and then expressed as the percentage of the rate at 7 minutes in the reaction in which no ADP was added. Vertical error bars display propogated standard deviation. Wild type RecA points are shown as squares, RecA D276A points are shown as diamonds, and RecA D276N points are shown as triangles.

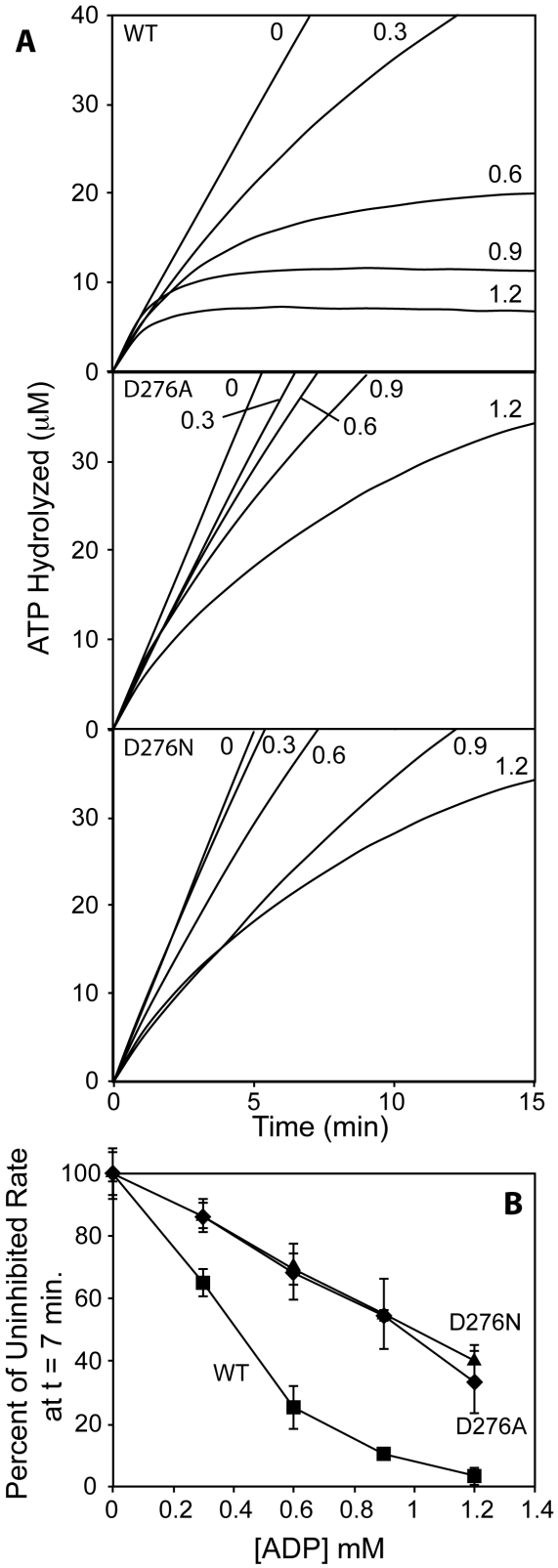
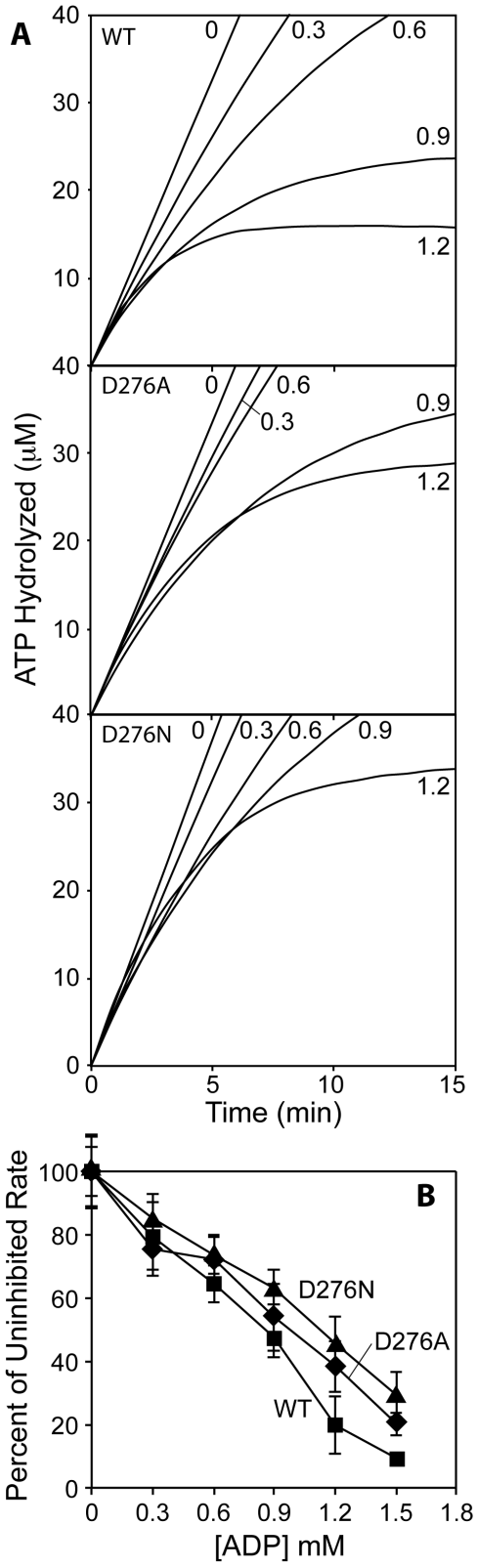

A
WT
0
0.3
0.6
0.9
1.2
D276A
0
0.3
0.6
0.9
1.2
D276N
0
0.3
0.6
0.9
1.2
ATP Hydrolyzed (μM)
Time (min)
B
Percent of Uninhibited Rate at t = 7 min.
D276N
WT
D276A
[ADP] mM

Figure 7. The ATPase activity of RecA D276A and RecA D276N on cssDNA is only slightly less inhibited by ADP than that of Wild type RecA.

A, reactions were carried out as in Figure 6 but without homologous ldsDNA, and ATP hydrolysis was monitored. Graphs are labeled with the variant of RecA protein used in the reactions, and reaction lines are labeled with the concentration of ADP added. Time 0 corresponds to the addition of ADP. *B*, The rate of ATP hydrolysis at time 7 minutes after the addition of ADP was quantified and converted to percentage of the rate of the reaction to which no ADP was added as described in Figure 6. Wild type RecA points are shown as squares, RecA D276A points are shown as diamonds, and RecA D276N points are shown as triangles.


A
WT
0
0.3
0.6
0.9
1.2
D276A
0
0.3
0.6
0.9
1.2
D276N
0
0.3
0.6
0.9
1.2
ATP Hydrolyzed (μM)
Time (min)
B
D276N
D276A
WT
Percent of Uninhibited Rate
[ADP] mM

References

1. Blasius, M., Hubscher, U., and Sommer, S. (2008) Deinococcus radiodurans: What belongs to the survival kit?, *Critical Reviews in Biochemistry and Molecular Biology 43*, 221-238.

2. Cox, M. M., and Battista, J. R. (2005) Deinococcus radiodurans - The consummate survivor, *Nature Reviews Microbiology 3*, 882-892.

3. Daly, M. J. (2012) Death by protein damage in irradiated cells, *DNA repair 11*, 12-21.

4. Slade, D., and Radman, M. (2011) Oxidative Stress Resistance in Deinococcus radiodurans, *Microbiology and Molecular Biology Reviews 75*, 133-+.

5. Battista, J. R. (1997) Against all odds - the survival strategies of deinococcus radiodurans, *Annual Review of Microbiology 51*, 203-224.

6. Mattimore, V., and Battista, J. R. (1996) Radioresistance of *Deinococcus radiodurans*: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation, *Journal of bacteriology 178*, 633-637.

7. Rainey, F. A., Ray, K., Ferreira, M., Gatz, B. Z., Nobre, M. F., Bagaley, D., Rash, B. A., Park, M. J., Earl, A. M., Shank, N. C., Small, A. M., Henk, M. C., Battista, J. R., Kampfer, P., and da Costa, M. S. (2005) Extensive diversity of ionizing-radiation-resistant bacteria recovered from Sonoran Desert soil and description of nine new species of the genus Deinococcus obtained from a single soil sample, *Applied & Environmental Microbiology 71*, 5225-5235.

8. White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D., Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L., Moffat, K. S., Qin, H. Y., Jiang, L. X., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J. J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K. S., Aravind, L., Daly, M. J., Minton, K. W., Fraser, C. M., and et al. (1999) Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1, *Science 286*, 1571-1577.

9. Tanaka, M., Earl, A. M., Howell, H. A., Park, M. J., Eisen, J. A., Peterson, S. N., and Battista, J. R. (2004) Analysis of Deinococcus radiodurans's transcriptional response to ionizing radiation and desiccation reveals novel proteins that contribute to extreme radioresistance, *Genetics 168*, 21-33.

10. Gutsche, I., Vujičić-Žagar, A., Siebert, X., Servant, P., Vannier, F., Castaing, B., B., G., Heulin, T., de Groot, A., Sommer, S., and Serre, L. (2008) Complex oligomeric structure of a truncated form of DdrA: A protein required for the extreme

radiotolerance of Deinococcus, *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics 1784*, 1050-1058.

11. Harris, D. R., Tanaka, M., Saveliev, S. V., Jolivet, E., Earl, A. M., Cox, M. M., and Battista, J. R. (2004) Preserving Genome Integrity: The DdrA Protein of Deinococcus radiodurans R1, *PLoS biology 2*, e304.

12. Harris, D. R., Ngo, K. V., and Cox, M. M. (2008) The stable, functional core of DdrA from Deinococcus radiodurans R1 does not restore radioresistance in vivo, *Journal of bacteriology 190*, 6475-6482.

13. Norais, C. A., Chitteni-Pattu, S., Wood, E. A., Inman, R. B., and Cox, M. M. (2009) DdrB protein, an alternative Deinococcus radiodurans SSB induced by ionizing radiation, *J Biol Chem 284*, 21402-21411.

14. Daly, M. J., Gaidamakova, E. K., Matrosova, V. Y., Vasilenko, A., Zhai, M., Leapman, R. D., Lai, B., Ravel, B., Li, S. M., Kemner, K. M., and Fredrickson, J. K. (2007) Protein Oxidation Implicated as the Primary Determinant of Bacterial Radioresistance, *PLoS biology 5*, e92.

15. Makarova, K. S., Omelchenko, M. V., Gaidamakova, E. K., Matrosova, V. Y., Vasilenko, A., Zhai, M., Lapidus, A., Copeland, A., Kim, E., Land, M., Mavrommatis, K., Pitluck, S., Richardson, P. M., Detter, C., Brettin, T., Saunders, E., Lai, B., Ravel, B., Kemner, K. M., Wolf, Y. I., Sorokin, A., Gerasimova, A. V., Gelfand, M. S., Fredrickson, J. K., Koonin, E. V., and Daly, M. J. (2007) Deinococcus geothermalis: the pool of extreme radiation resistance genes shrinks, *PloS one 2*, e955.

16. Repar, J., Cvjetan, S., Slade, D., Radman, M., Zahradka, D., and Zahradka, K. (2010) RecA protein assures fidelity of DNA repair and genome stability in Deinococcus radiodurans, *DNA repair 9*, 1151-1161.

17. Slade, D., Lindner, A. B., Paul, G., and Radman, M. (2009) Recombination and Replication in DNA Repair of Heavily Irradiated Deinococcus radiodurans, *Cell 136*, 1044-1055.

18. Zahradka, K., Slade, D., Bailone, A., Sommer, S., Averbeck, D., Petranovic, M., Lindner, A. B., and Radman, M. (2006) Reassembly of shattered chromosomes in Deinococcus radiodurans, *Nature 443*, 569-573.

19. Kim, J.-I., and Cox, M. M. (2002) The RecA proteins of *Deinococcus radiodurans* and *Escherichia coli* promote DNA strand exchange via inverse pathways, *Proc Natl Acad Sci U S A 99*, 7917-7921.

20. Hsu, H.-F., Ngo, K.V., Chitteni-Pattu, S., Cox, M.M., and Li, H.-W. (2011) Investigating Deinococcus radiodurans RecA protein filament formation on

double-stranded DNA by a real-time single-molecule approach, *Biochemistry 50*, 8270-8280.

21. Harris, D. R., Pollock, S. V., Wood, E. A., Goiffon, R. J., Klingele, A. J., Cabot, E. L., Schackwitz, W., Martin, J., Eggington, J., Durfee, T. J., Middle, C. M., Norton, J. E., Popelars, M., Li, H., Klugman, S. A., Hamilton, L. L., Bane, L. B., Pennacchio, L., Albert, T. J., Perna, N. T., Cox, M. M., and Battista, J. R. (2009) Directed evolution of radiation resistance in Escherichia coli, *Journal of bacteriology 191*, 5240-5252.

22. Weinstock, G. M., McEntee, K., and Lehman, I. R. (1981) Hydrolysis of nucleoside triphosphates catalyzed by the recA protein of Escherichia coli. Steady state kinetic analysis of ATP hydrolysis, *Journal of Biological Chemistry 256*, 8845-8849.

23. Brenner, S. L., Mitchell, R. S., Morrical, S. W., Neuendorf, S. K., Schutte, B. C., and Cox, M. M. (1987) RecA protein-promoted ATP hydrolysis occurs throughout RecA nucleoprotein filaments, *J. Biol. Chem. 262*, 4011-4016.

24. Cox, M. M. (1994) Why Does RecA Protein Hydrolyze ATP, *Trends in biochemical sciences 19*, 217-222.

25. Cox, J. M., Tsodikov, O. V., and Cox, M. M. (2005) Organized unidirectional waves of ATP hydrolysis within a RecA filament, *PLoS biology 3*, 231-243.

26. Cox, J. M., Abbott, S. N., Chitteni-Pattu, S., Inman, R. B., and Cox, M. M. (2006) Complementation of one RecA protein point mutation by another - Evidence for trans catalysis of ATP hydrolysis, *Journal of Biological Chemistry 281*, 12968-12975.

27. Cox, M. M. (2007) Motoring along with the bacterial RecA protein, *Nature Reviews Molecular Cell Biology 8*, 127-138.

28. Cox, J. M., Li, H., Wood, E. A., Chitteni-Pattu, S., Inman, R. B., and Cox, M. M. (2008) Defective dissociation of a "Slow" RecA mutant protein imparts an Escherichia coli growth defect, *Journal of Biological Chemistry 283*, 24909-24921.

29. Cazaux, C., Larminat, F., Villani, G., Johnson, N. P., Schnarr, M., and Defais, M. (1994) Purification and biochemical characterization of Escherichia coli RecA proteins mutated in the putative DNA binding site, *Journal of Biological Chemistry 269*, 8246-8254.

30. Eggler, A. L., Lusetti, S. L., and Cox, M. M. (2003) The C terminus of the Escherichia coli RecA protein modulates the DNA binding competition with single-stranded DNA-binding protein, *Journal of Biological Chemistry 278*, 16389-16396.

31. Lavery, P. E., and Kowalczykowski, S. C. (1992) Biochemical basis of the constitutive repressor cleavage activity of recA730 protein. A comparison to recA441 and recA803 proteins, *Journal of Biological Chemistry 267*, 20648-20658.

32. Mirshad, J. K., and Kowalczykowski, S. C. (2003) Biochemical basis of the constitutive coprotease activity of RecA P67W protein, *Biochemistry 42*, 5937-5944.

33. Haruta, N., Yu, X. N., Yang, S. X., Egelman, E. H., and Cox, M. M. (2003) A DNA pairing-enhanced conformation of bacterial RecA proteins, *Journal of Biological Chemistry 278*, 52710-52723.

34. Schutte, B. C., and Cox, M. M. (1987) Homology-dependent changes in adenosine 5'-triphosphate hydrolysis during RecA protein promoted DNA strand exchange: evidence for long paranemic complexes, *Biochemistry 26*, 5616-5625.

35. Joo, C., McKinney, S. A., Nakamura, M., Rasnik, I., Myong, S., and Ha, T. (2006) Real-time observation of RecA filament dynamics with single monomer resolution, *Cell 126*, 515-527.

36. Lusetti, S. L., Hobbs, M. D., Stohl, E. A., Chitteni-Pattu, S., Inman, R. B., Seifert, H. S., and Cox, M. M. (2006) The RecF protein antagonizes RecX function via direct interaction, *Mol Cell 21*, 41-50.

37. Bell, J. C., Plank, J. L., Dombrowski, C. C., and Kowalczykowski, S. C. (2012) Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA, *Nature 491*, 274-278.

38. Galletto, R., Amitani, I., Baskin, R. J., and Kowalczykowski, S. C. (2006) Direct observation of individual RecA filaments assembling on single DNA molecules, *Nature 443*, 875-878.

39. Cox, M. M., Goodman, M. F., Kreuzer, K. N., Sherratt, D. J., Sandler, S. J., and Marians, K. J. (2000) The importance of repairing stalled replication forks, *Nature 404*, 37-41.

40. Cox, M. M. (2001) Historical overview: Searching for replication help in all of the rec places, *Proc Natl Acad Sci U S A 98*, 8173-8180.

41. Chen, Z. C., Yang, H. J., and Pavletich, N. P. (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures, *Nature 453*, 489-494.

42. Story, R. M., Weber, I. T., and Steitz, T. A. (1992) The structure of the *E. coli* RecA protein monomer and polymer, *Nature 355*, 318-325.

43. Wong, O. K., Guthold, M., Erie, D. A., and Gelles, J. (2008) Interconvertible lac repressor-DNA loops revealed by single-molecule experiments, *PLoS biology 6*, e232.

44. Chu, J. F., Chang, T. C., and Li, H. W. (2010) Single-molecule TPM studies on the conversion of human telomeric DNA, *Biophysical journal 98*, 1608-1616.

45. Schafer, D. A., Gelles, J., Sheetz, M. P., and Landick, R. (1991) Transcription by single molecules of RNA polymerase observed by light microscopy, *Nature 352*, 444-448.

46. Fan, H. F., and Li, H. W. (2009) Studying RecBCD helicase translocation along Chi-DNA using tethered particle motion with a stretching force, *Biophysical journal 96*, 1875-1883.

47. Dohoney, K. M., and Gelles, J. (2001) Chi-sequence recognition and DNA translocation by single RecBCD helicase/nuclease molecules, *Nature 409*, 370-374.

48. Yin, H., Landick, R., and Gelles, J. (1994) Tethered particle motion method for studying transcript elongation by a single RNA polymerase molecule, *Biophysical journal 67*, 2468-2478.

49. Pugh, B. F., and Cox, M. M. (1987) Stable binding of RecA protein to duplex DNA. Unraveling a paradox, *J. Biol. Chem. 262*, 1326-1336.

50. Pugh, B. F., and Cox, M. M. (1988) General mechanism for RecA protein binding to duplex DNA, *J. Mol. Biol. 203*, 479-493.

51. Cox, M. M. (2004) The RecA Protein, In *The Bacterial Chromosome* (Higgins, N. P., Ed.), pp 369-388, American Society of Microbiology, Washington, D. C.

52. Cox, M. M. (2007) The bacterial RecA protein: structure, function, and regulation, In *Topics in Current Genetics: Molecular Genetics of Recombination* (Rothstein, R., and Aguilera, A., Eds.), pp 53-94, Springer-Verlag, Heidelberg.

53. Lusetti, S. L., and Cox, M. M. (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks, *Annual Review of Biochemistry 71*, 71-100.

54. Cox, M. M., Soltis, D. A., Lehman, I. R., DeBrosse, C., and Benkovic, S. J. (1983) ADP-mediated dissociation of stable complexes of RecA protein and single-stranded DNA, *J. Biol. Chem. 258*, 2586-2592.

55. Lee, J. W., and Cox, M. M. (1990) Inhibition of RecA protein-promoted ATP hydrolysis. I. ATPγS and ADP are antagonistic inhibitors, *Biochemistry 29*, 7666-7676.

56. Lee, J. W., and Cox, M. M. (1990) Inhibition of RecA protein-promoted ATP hydrolysis. II. Longitudinal assembly and disassembly of RecA protein filaments mediated by ATP and ADP, *Biochemistry 29*, 7677-7683.

57. Buckstein, M. H., He, J., and Rubin, H. (2008) Characterization of nucleotide pools as a function of physiological state in Escherichia coli, *Journal of bacteriology 190*, 718-726.

58. Story, R. M., and Steitz, T. A. (1992) Structure of the RecA Protein-ADP complex, *Nature 355*, 374-376.

59. Roca, A. I., and Cox, M. M. (1990) The RecA protein: structure and function, *CRC Critical Reviews in Biochemistry and Molecular Biology 25*, 415-456.

60. Kurumizaka, H., Aihara, H., Ikawa, S., Kashima, T., Bazemore, L. R., Kawasaki, K., Sarai, A., Radding, C. M., and Shibata, T. (1996) A possible role of the C-terminal domain of the RecA protein. A gateway model for double-stranded DNA binding, *Journal of Biological Chemistry 271*, 33515-33524.

61. Stohl, E. A., Gruenig, M. C., Cox, M. M., and Seifert, H. S. (2011) Purification and characterization of the RecA protein from Neisseria gonorrhoeae, *PloS one 6*, e17101.

62. Angov, E., and Camerini, O. R. (1994) The recA gene from the thermophile Thermus aquaticus YT-1: cloning, expression, and characterization, *Journal of bacteriology 176*, 1405-1412.

63. Pierré, A., and Paoletti, C. (1983) Purification and characterization of a RecA protein from *Salmonella typhimurium*, *J. Biol. Chem. 258*, 2870-2874.

64. Ganesh, N., and Muniyappa, K. (2003) Characterization of DNA strand transfer promoted by Mycobacterium smegmatis RecA reveals functional diversity with Mycobacterium tuberculosis RecA, *Biochemistry 42*, 7216-7225.

65. Bakhlanova, I. V., Dudkina, A. V., Baitin, D. M., Knight, K. L., Cox, M. M., and Lanzov, V. A. (2010) Modulating cellular recombination potential through alterations in RecA structure and regulation, *Molecular Microbiology in press*.

66. Messing, J. (1983) New M13 vectors for cloning, *Methods Enzymol. 101*, 20-78.

67. Neuendorf, S. K., and Cox, M. M. (1986) Exchange of RecA protein between adjacent RecA protein-single-stranded DNA complexes, *J. Biol. Chem. 261*, 8276-8282.

68. Petrova, V., Chitteni-Pattu, S., Drees, J. C., Inman, R. B., and Cox, M. M. (2009) An SOS Inhibitor that Binds to Free RecA Protein: The PsiB Protein, *Mol Cell 36*, 121-130.

69. Craig, N. L., and Roberts, J. W. (1981) Function of nucleoside triphosphate and polynucleotide in Escherichia coli recA protein-directed cleavage of phage lambda repressor, *Journal of Biological Chemistry 256*, 8039-8044.

70. Lohman, T. M., and Overman, L. B. (1985) Two binding modes in *Escherichia coli* single strand binding protein-single stranded DNA complexes. Modulation by NaCl concentration, *J. Biol. Chem. 260*, 3594-3603.

71. Mertens, N., Remaut, E., and Fiers, W. (1995) Tight transcriptional control mechanism ensures stable high-level expression from T7 promoter-based expression plasmids., *Biotechnology 13*, 175-179.

72. Lindsley, J. E., and Cox, M. M. (1990) Assembly and disassembly of RecA protein filaments occurs at opposite filament ends: relationship to DNA strand exchange, *J. Biol. Chem. 265*, 9043-9054.

73. Morrical, S. W., Lee, J., and Cox, M. M. (1986) Continuous association of *Escherichia coli* single-stranded DNA binding protein with stable complexes of RecA protein and single-stranded DNA, *Biochemistry 25*, 1482-1494.