5-2023

Computational Modeling of the Fanconi Anemia Gene Network and its Connection to Cancer

Alyssa Warren-Belford

Computational Modeling of the Fanconi Anemia Gene Network and its Connection to Cancer

An Honors Thesis submitted in partial fulfillment of the requirements for Honors Studies in Biology

By

Alyssa Warren-Belford

Spring 2023

Biology

Fulbright College of Arts and Sciences

The University of Arkansas

Abstract

Fanconi anemia (FA) is a rare genetic condition in which the cell's DNA repair machinery is dysregulated, significantly increasing the chances of tumorigenesis. Further research is being done in order to improve patient outcomes and incidences of cancer. Our group created a computational model of the FA DNA repair gene network, which removes interstrand crosslinks found in damaged DNA and repairs it so DNA synthesis can continue. Computer simulations show the number of DNA damage indicators decreased as the pathway continued. This was expected as the FA pathway repairs DNA damage. The goal of this project was to provide further understanding of the protein interactions so that other biologically significant pathways may be added in the future.

Table of Contents

Introduction

Fanconi anemia (FA) is a rare and autosomal recessive genetic disease that affects the body's ability to repair damaged deoxyribonucleic acid (DNA). Symptoms of FA can vary widely among individuals but typically include bone marrow failure, which can lead to anemia, low platelet counts, and a weakened immune system. Individuals with FA have a notably higher risk

of developing certain types of cancer, such as leukemia and solid tumors, compared to the general population. Other physical abnormalities can present, which include abnormal facial features, small head size, and skeletal abnormalities (Green & Kupfer, 2009). Diagnosis of FA is made through genetic testing, which can identify mutations in the genes associated with the condition. Treatment for the disease typically involves supportive care to manage symptoms and complications, such as blood transfusions to treat anemia and antibiotics to prevent infections. Bone marrow transplants may also be used to treat the condition by replacing damaged bone marrow with healthy stem cells (Green & Kupfer, 2009).

FA is caused by mutations in a subset of at least 22 known genes involved in the repair of DNA damage (Auerbach, 1995). These genes are collectively referred to as the FA pathway. When DNA is damaged, these genes are intended to work together to repair the DNA and prevent further damage from occurring. However, FA individuals have a pathway that is dysregulated, leading to an accumulation of DNA damage and a higher risk of cancer (Niraj et al., 2019). Early diagnosis and comprehensive symptom management are critical to improving outcomes for patients with FA.

There is currently no cure for FA. Ongoing research is focused on developing new therapies and improving understanding of the disease. This project focuses on developing a general computational model for how the FA genes interact and lead to the development of cancer. By using computational modeling, we created a mathematical representation of the DNA damage repair pathway that imitates the FA pathway in humans. This model can be manipulated to replicate the accumulation of DNA damage that occurs in FA patients. In the future, the goal is to use this model, integrated with computational models of other processes known to be important in FA (e.g.,

secretion of factors from the oral microbiome), to develop improved treatments and preventative measures for FA patients.

Background: The Fanconi anemia DNA repair pathway

A key characteristic of FA is hypersensitivity to crosslinking agents (Green & Kupfer, 2009). Crosslinking agents are substances introduced into the DNA that cause nucleotides to bind together; the resulting binding is referred to as a DNA interstrand crosslink (ICL). ICLs prevent DNA synthesis and repair in cells. DNA repair is a normal part of cell maintenance from exposure to sources of DNA damage that naturally occur due to interactions with the environment and chemical agents. It is the cell's responsibility to repair DNA damage to prevent accumulation of mutations, which can lead to cancer. The FA pathway detects ICLs and initiates repair. Numerous proteins function in this pathway and are divided into groups based on their function: ICL detection, cleavage, and repair (Green & Kupfer, 2009).

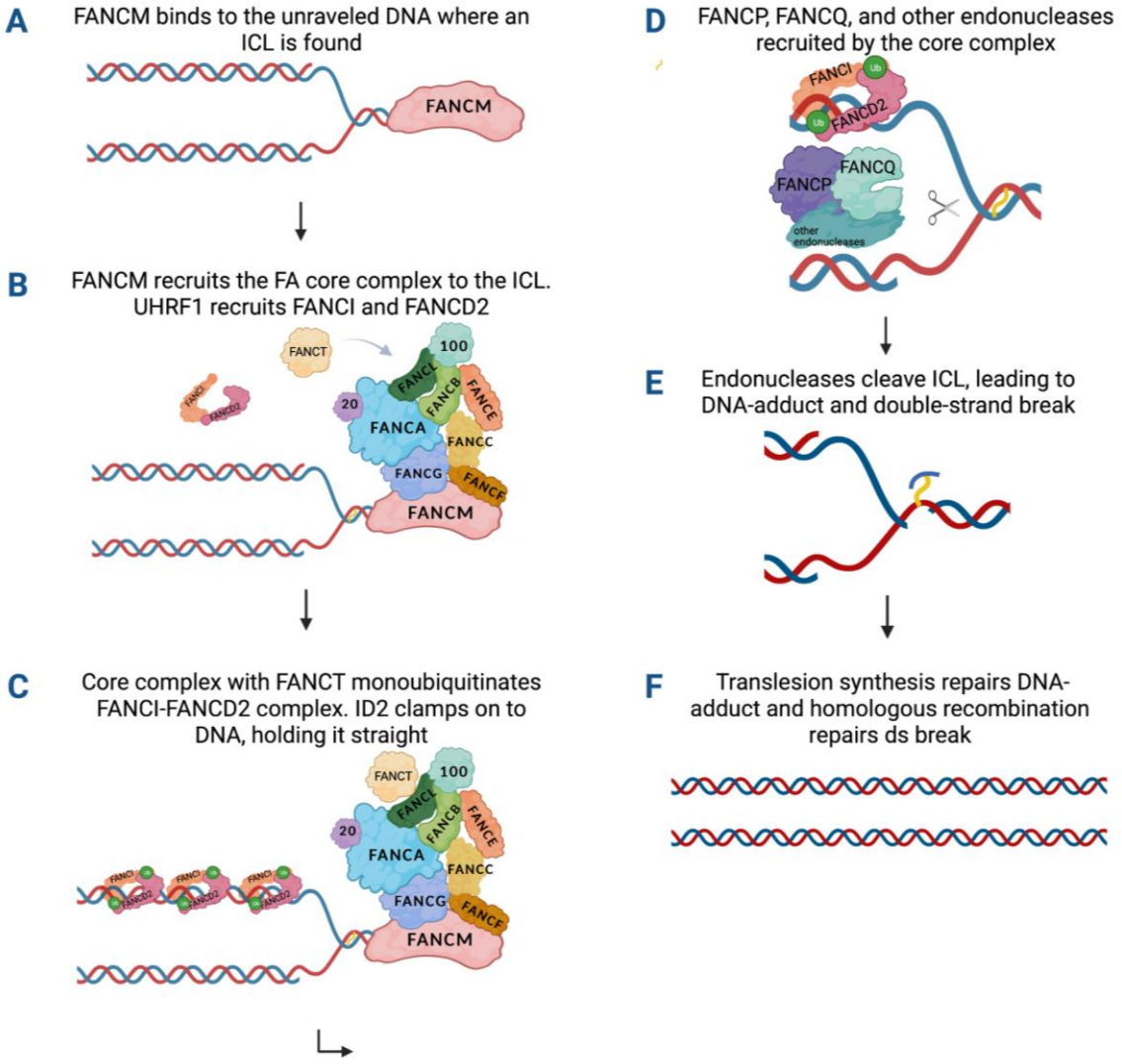

Figure 1. Schematic representation of the Fanconi anemia DNA repair pathway. See text for details. Created with BioRender.com.

Detection of ICLs involves 10 FA proteins: FANCA, FANCB, FANCC, FANCE, FANCF, FRANCG, FANCL, FANCM, FAAP100, and FAAP20. The abbreviation "FANC" before the protein name indicates that the protein is in the FA pathway, while the abbreviation "FAAP" indicates that the protein is an FA-associated protein. FANCM detects an ICL in the DNA (Figure 1A) and acts as a recruiter and landing site for the FA "core complex" (Figure 1B), which is composed of nine proteins, divided into three subcomplexes: FANCB-FANCL-FAAP100

(BL100), FANCA-FANCG-FAAP20 (AG20), and FANC-FANCE-FANCF (CEF). The BL100 subcomplex guides the assembly of the core complex, AG20 is required for localization to the cell nucleus (where DNA repair takes place), and CEF forms a bridge for the interaction between the FA core complex and its target, the FANCI-FANCD2 (ID2) complex (Rodríguez & D'Andrea, 2017).

The ICL cleavage phase begins with the recruitment of an additional FA pathway protein, FANCT, to the core complex (Figure 1B), which is required for the subsequent monoubiquitylation of the ID2 complex, mentioned above. Monoubiquitylation refers to the bonding of a ubiquitin molecule to a protein. This bonding is required before the pathway can continue. After the ID2 complex is ubiquitinated (Figure 1C), it binds and stabilizes the DNA strand, which is essential for recruiting two additional FA proteins: FANCP and FANCQ. FANCP, a scaffolding protein, is especially important because of its responsibility in recruiting additional endonucleases (Figure 1D), as well as downstream proteins hypothesized to aid in the process of homologous recombination (Moldovan & D'Andrea, 2009). The recruited endonucleases cleave the ICL, leaving behind a DNA lesion (or adduct) and a double-strand break (DSB; Figure 1E).

There are several pathways associated with repair of DNA lesions and DSBs. I focused specifically on the FA pathway's connection to the nucleotide excision repair (NER) pathway

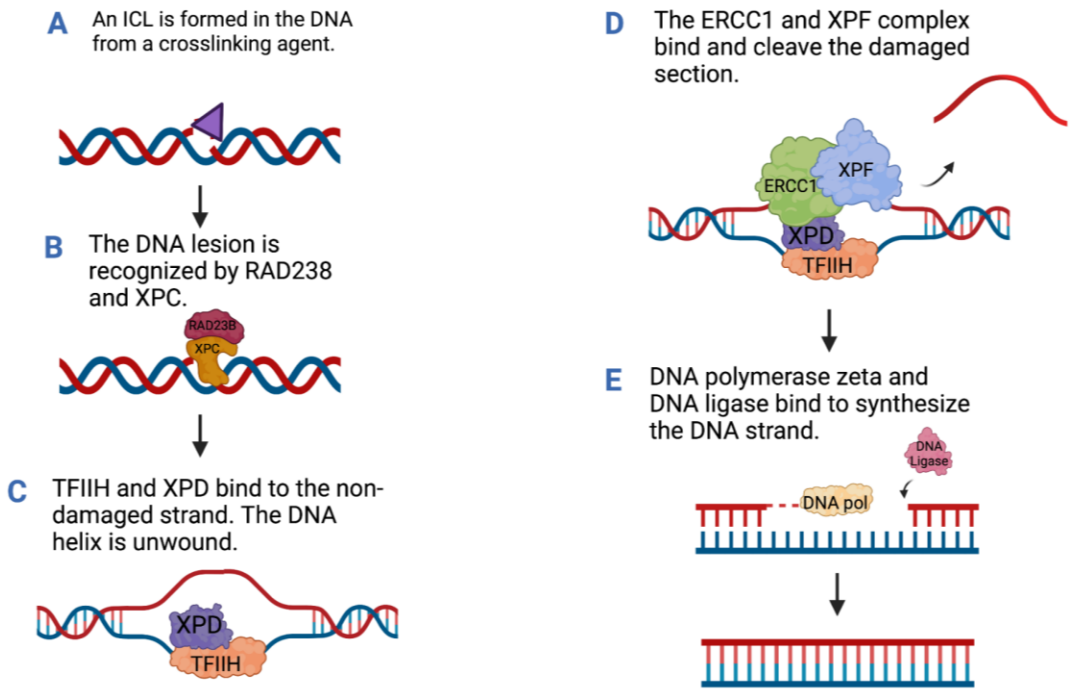

Figure 2. Schematic of the NER pathway and its role in repairing DNA lesions generated by the FA pathway. See text for details. Created with BioRender.com.

(Figure 2). The NER pathway is a mechanism used by eukaryotes to remove DNA lesions, which are a type of molecular lesion that affects the structure of the DNA molecule (Schärer, 2013). After an ICL is formed (Figure 2A), it is recognized by the nucleotide excision repair proteins XPC and RAD23B (Figure 2B). This is followed by binding of the TFIIH complex and then the helicase XPD, which unwinds the DNA (Figure 2C). Endonucleases ERCC1 and XPF, recruited previously by FANCP, then remove damaged DNA strands to protect the rest of the strand by cleaving the 5' and 3' ends (Figure 2D) (Schärer, 2013). Finally, DNA polymerase and DNA ligase bind to synthesize the strand (Figure 2E), allowing for replication and DNA synthesis to continue normally.

Methods

We constructed an integrated computational model of the FA pathway (Figure 1), the NER pathway (Figure 2), and the homologous recombination pathway for repairing DSBs (not shown), using PySB, a Python-based modeling and simulation platform (Lopez et al., 2013). PySB utilizes a "rule-based" modeling (RBM) approach (Chylek et al., 2014, 2015), which is designed to simplify the construction of complex intracellular pathways, such as that studied here. The basic building block of a PySB model is the `Monomer`, which represents a protein. Importantly, PySB monomers are structured objects, meaning they are composed of binding sites and states (e.g.,

```
# Monomers
Monomer("RAD23B", ["xpc"])
Monomer("XPC", ["rad23b", "lesion"])
Monomer("TFIIH", ["lesion", "xpd"])
Monomer("XPD", ["tfiih"])
Monomer("ERCC1", ["xpf", "lesion"])
Monomer("XPF", ["ercc1"])

# Rules
Rule("XPC_binds_Lesion",
     XPC(lesion=None, rad23b=None) + Lesion(b=None) | XPC(lesion=1, rad23b=None) % Lesion(b=1), kf_XPC_lesion, kr_XPC_lesion)

Rule("RAD23B_binds_XPC_lesion", RAD23B(xpc=None) + XPC(lesion=ANY, rad23b=None) | RAD23B(xpc=1) % XPC(lesion=ANY, rad23b=1),
     kf_RAD23B_XPC_lesion, kr_RAD23B_XPC_lesion)

Rule("TFIIH_binds_lesion", TFIIH(lesion=None, xpd=None) + Lesion(b=2) % RAD23B(xpc=1) % XPC(lesion=2, rad23b=1) >>
     TFIIH(lesion=3, xpd=None) % Lesion(b=3) + RAD23B(xpc=None) + XPC(lesion=None, rad23b=None), k_TFIIH_lesion)

Rule("XPD_binds_TFIIH_lesion", XPD(tfiih=None) + TFIIH(lesion=ANY, xpd=None) | XPD(tfiih=1) % TFIIH(lesion=ANY, xpd=1),
     kf_XPD_TFIIH_lesion, kr_XPD_TFIIH_lesion)

Rule("ERCC1_binds_XPF", ERCC1(xpf=None, lesion=None) + XPF(ercc1=None) | ERCC1(xpf=1, lesion=None) % XPF(ercc1=1),
     kf_ERCC1_XPF, kr_ERCC1_XPF)

Rule("ERCC1_XPF_binds_lesion", ERCC1(xpf=1, lesion=None) % XPF(ercc1=1) + Lesion(b=3) % XPD(tfiih=2) % TFIIH(lesion=3, xpd=2) >>
     ERCC1(xpf=1, lesion=3) % XPF(ercc1=1) % Lesion(b=3) + XPD(tfiih=2) % TFIIH(lesion=None, xpd=2), k_ERCC1_XPF_lesion)

Rule("Pol_Zeta_binds_lesion", Pol_Zeta(dna=None) + ERCC1(xpf=1, lesion=2) % XPF(ercc1=1) % Lesion(b=2) >>
     Pol_Zeta(dna=3) % Lesion(b=3) + ERCC1(xpf=None, lesion=None) + XPF(ercc1=None), k_Pol_Zeta_lesion)

Rule("Ligase_binds_lesion", Ligase(dna=None) + Pol_Zeta(dna=1) % Lesion(b=1) >> Ligase(dna=1) % Lesion(b=1) + Pol_Zeta(dna=None),
     k_Ligase_lesion)

Rule("Ligase_Repairs_Lesion", Ligase(dna=1) % Lesion(b=1) >> Ligase(dna=None), k_ligase_repairs_lesion)
```

Figure 3. Screenshot of the monomers and rules from the PySB model I constructed of the translesion synthesis pathway. This submodel contains six monomers, shown at the top, and nine rules. Some of the rules are unidirectional (indicated by the '>>' operator), while others are bidirectional (indicated by the '|' operator).

phosphorylated/unphosphorylated, ubiquitinated/deubiquitinated). A biological process is represented in PySB as a `Rule`, which defines the proteins involved in an interaction or transformation and the products that result. For example, binding of protein A to protein B can be represented in PySB as follows:

```
A(b=None) + B(a=None) >> A(b=1) % B(a=1), k_AB_bind
```

Here, the monomer '`A`' contains one binding site, '`b`', and the monomer '`B`' contains one binding site, '`a`'. The Python keyword '`None`' indicates that the binding site is unbound, which is an obvious necessary condition for binding to occur. The '`>>`' operator indicates that the reactants on the left are transformed into the products on the right. The label '`1`' on the right-hand side of the rule represents the bond that is formed between the '`b`' binding site of the '`A`' protein and the '`a`' binding site of the '`B`' protein. Every PySB rule is also associated with a `Parameter` object, called '`k_AB_bind`' above, that represents how fast the reaction proceeds. In all, our integrated FA DNA repair pathway model includes 35 monomers, 76 rate parameters, and 61 rules. Python code showing the monomers and rules for the "translesion synthesis" submodel, which represents the processes shown in Figure 2, are included in Figure 3.

Results

After constructing our computational model of the FA DNA repair pathway, we ran simulations to test the feasibility of the model. In Figure 4A, we track the amounts of the three subcomplexes ("`AG20_total`", "`BL100_total`", and "`CEF_total`") that form the FA core complex, the unbound core complex ("`FAcpx_free`"), ubiquitinated ID2 bound to the core complex ("`FAcpx_ID2ub`"), and the ubiquitinated ID2 complex bound to FANCP and FANCQ ("`FANCQ_FANCP_ID2Ub`"). As described above, this last complex cleaves the ICL, leaving

behind a DNA lesion and a DSB. We see that the amounts of the AG20, BL100, and CEF complexes initially decrease as they bind to form the FA core complex. The amount of unbound core complex rises quickly but then equilibrates to a steady level as it binds to FANCM, followed by recruitment of FANCT and ID2 (Figure 1B,C). We also see small increases in the amounts of ubiquitinated ID2 bound to the FA complex and the FANCP-FANCQ-ID2 complex. These increases are short-lived, however, because cleavage of the ICL proceeds quickly, followed by dissociation of these complexes, at which point the whole process begins again with a new ICL.

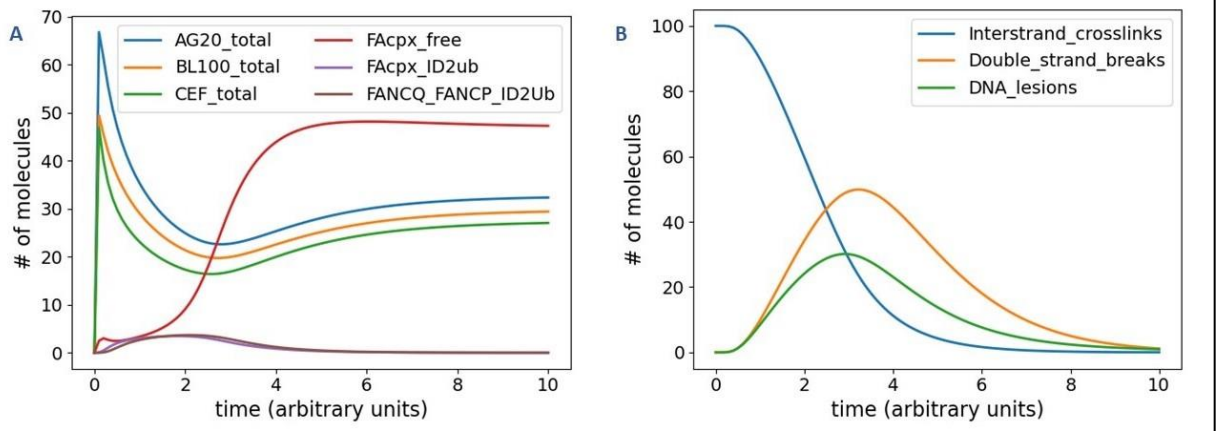

Figure 4. Time courses for numerous species in the FA DNA repair pathway model. Note that time is in arbitrary units because the rate parameters were chosen somewhat arbitrarily for a "proof of principle" demonstration of the model.

In Figure 4B, we plot time courses of ICLs, DNA lesions, and DSBs, demonstrating how the model captures the DNA repair process. We begin the simulation with 100 ICLs. As described above, these ICLs are gradually cleaved and converted into DNA lesions and DSBs. We see that the number of ICLs in the simulation declines over time, while the number of DNA lesions and DSBs increases. The DNA lesions are then repaired by the NER pathway (Figure 2) and the DSBs are repaired by homologous recombination. Eventually, all the DNA mutations are repaired, and the simulation ends. Note that this simulation was performed with a preliminary set of rate parameters and initial protein amounts as a "proof of principle" demonstration of the capabilities

of the model. In the future, more biologically realistic values will be used based on experimental and clinical data obtained from the literature and experimental collaborators.

Conclusion

Providing further understanding of FA is the goal of this project. The FA pathway is a collection of proteins that function to repair damaged DNA. Individuals with FA have a dysregulated DNA repair pathway, and this leads to increased risk of cancer and other problems. We built a computational model of the interactions between the proteins in the FA pathway to the best of our knowledge. The model can be manipulated to include initial amounts of proteins, rate constants, and alternative hypothesized interactions. This allows researchers to see how the cell is affected. The limitations to this type of modeling are based on the limitations in knowledge. Many parts are unclear in their function and structure. Nevertheless, we were able to capture the process by which ICLs are repaired by the Fanconi anemia DNA repair pathway using integrated computational modeling. Research is still being conducted to fully understand the FA pathway and DNA repair systems. FA research is important for improving our understanding of this rare genetic disease, improving patient outcomes, and advancing our understanding of cancer biology.

References

Auerbach, A. D. (1995). Fanconi Anemia. *Dermatologic Clinics*, *13*(1), 41–49. https://doi.org/10.1016/S0733-8635(18)30105-0

Chylek, L. A., Harris, L. A., Faeder, J. R., & Hlavacek, W. S. (2015). Modeling for (physical) biologists: an introduction to the rule-based approach. *Physical Biology*, *12*(4), 045007. https://doi.org/10.1088/1478-3975/12/4/045007

Chylek, L. A., Harris, L. A., Tung, C., Faeder, J. R., Lopez, C. F., & Hlavacek, W. S. (2014). Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. *WIREs Systems Biology and Medicine*, *6*(1), 13–36. https://doi.org/10.1002/wsbm.1245

Green, A. M., & Kupfer, G. M. (2009). Fanconi Anemia. *Hematology/Oncology Clinics of North America*, *23*(2), 193–214. https://doi.org/10.1016/j.hoc.2009.01.008

Lopez, C. F., Muhlich, J. L., Bachman, J. A., & Sorger, P. K. (2013). Programming biological models in Python using PySB. *Molecular Systems Biology*, *9*(1), 646. https://doi.org/10.1038/msb.2013.1

Moldovan, G. L., & D'Andrea, A. D. (2009). How the Fanconi Anemia Pathway Guards the Genome. *Http://Dx.Doi.Org/10.1146/Annurev-Genet-102108-134222*, *43*, 223–249. https://doi.org/10.1146/ANNUREV-GENET-102108-134222

Niraj, J., Färkkilä, A., & D'Andrea, A. D. (2019). The Fanconi anemia pathway in cancer. *Annual Review of Cancer Biology*, *3*(1), 457–478. https://doi.org/10.1146/ANNUREV-CANCERBIO-030617-050422

Rodríguez, A., & D'Andrea, A. (2017). Fanconi anemia pathway. *Current Biology*, *27*(18), R986–R988. https://doi.org/10.1016/J.CUB.2017.07.043

Schärer, O. D. (2013). Nucleotide excision repair in eukaryotes. *Cold Spring Harbor Perspectives in Biology*, *5*(10), a012609. https://doi.org/10.1101/cshperspect.a012609