5-2023

Development of a Computational Model to Investigate Pathways and the Effects of Treatment in Fanconi Anemia

Sabrina Kellett

Development of a Computational Model to Investigate Pathways and the Effects of Treatment in Fanconi Anemia

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

by

Sabrina J. Kellett

Spring 2023

Biological Sciences

J. William Fulbright College of Arts and Sciences

The University of Arkansas

Acknowledgements

I could not have completed this project if it were not for the guidance of my honors thesis advisor, Dr. Leonard Harris. I could not thank him enough for his patience and dedication to helping complete this project. This experience has given me exposure to a new field of scientific research, which would not have been possible if he had not welcomed me into his lab. My lab partner for this project, Alyssa Warren-Belford, has also provided me constant support throughout this process, and it was a joy to work with her. I would also like to thank Dr. Shilpa Iyer, Dr. Kate Walker, and Dr. Lindsey Aloia for graciously volunteering their time to serve on my thesis committee. I am grateful for this opportunity and for the people who have helped me get to this point in my educational career.

Table of Contents

Abstract

Fanconi Anemia (FA) is a rare type of anemia that is not easily studied and can have very detrimental effects. This disease compromises the bone marrow, resulting in decreased hemopoiesis. Symptoms of FA also include abnormalities in the brain and spinal cord, incorrect formation of the kidneys, abnormal formation of the heart and lungs, and a dramatically increased risk of developing cancer. FA can be caused by various mutations in any of the 22 genes that encode for proteins involved in what is called the FA DNA repair pathway. In healthy individuals, this pathway specifically repairs interstrand cross-links (ICLs) recognized in the S phase of the cell cycle. A series of steps leads from the ICL to the formation of a DNA lesion and double-strand break. The DNA lesion is then repaired by translesion synthesis, and the double-strand break by homologous recombination. The goal of this project is to build a computational model, using the PySB software package, that simulates the biochemical reactions that take place in the FA pathway between proteins and DNA. Once the model was constructed, it produced results that showed DNA damage decreasing over time as it would in the FA pathway, and that protein concentrations changed relative to what events in the pathway were taking place. This model, integrated with other computational models simulating biological events related to the FA pathway, could be used to further the search for druggable targets and improve treatments for individuals with FA.

Introduction

More than one-fourth of people on Earth are affected by anemia, making it the most common blood disorder in the world. Anemia is characterized by a lack of healthy red blood cells, resulting in inadequate transport of oxygen throughout the body. Deficient oxygen can cause fatigue, weakness, dizziness, shortness of breath, and numerous other complications (*Anaemia*, n.d.). Fanconi anemia (FA) is a specific type of anemia that, though not as well studied, can have very detrimental effects. This rare condition, which only affects 1 in 160,000 individuals worldwide, mainly affects the bone marrow, resulting in a decrease in all blood cells, not just erythrocytes. Symptoms of FA include developmental and functional problems of the bones, abnormalities in the brain and spinal cord, incorrect formation of the kidneys, and abnormal formation of the heart and lungs (*Fanconi Anemia: MedlinePlus Genetics*, n.d.).

Individuals with FA are also more susceptible to infection due to bone marrow failure and decreased hemopoiesis (Furquim et al., 2017). The oral microbiome is full of bacteria and other microorganisms. Thus, if there is an injury or malformation in the mouth of someone with FA, these microorganisms can invade and cause infection or other types of oral disease (Goswami et al., 2016). This includes cancer, specifically oral squamous cell carcinomas (Niraj et al., 2019). Furthermore, researchers have found that some of the genes in the FA pathway, when mutated, are associated with susceptibility to ovarian and breast cancer. Additional studies have shown that at least one gene of the FA pathway is altered in ~40% of all cancers (Niraj et al., 2019).

FA can be caused by various mutations in any of the 22 genes that encode for proteins involved in what is called the FA DNA repair pathway (Rodríguez & D'Andrea, 2017). This pathway is activated when DNA damage, particularly interstrand cross-links (ICLs), is sensed during DNA replication in the S-phase of the cell cycle and the replication fork is stalled. The FA

pathway executes the removal and repair of ICLs mainly through homologous recombination (HR) and translesion synthesis (Niraj et al., 2019). When there are mutations in genes of the FA pathway, the ICLs are not properly repaired and continue to accumulate, resulting in uncontrolled cell growth or abnormal cell death (Rageul & Kim, 2020). ICLs attach the two strands of a DNA double-helix through a covalent bond, as opposed to the normal hydrogen bonds connecting the two. ICLs can be induced in a multitude of ways, such as exogenous (environmental) and endogenous metabolites, as well as chemotherapeutics. Though all four types of nucleotides can participate in cross-link formation, purine nucleotides (bases adenine and guanine) are typically more reactive. These cross-links obstruct DNA replication and, therefore, must be repaired for cells to properly replicate. In individuals with FA, accumulation of ICLs can lead to bone marrow failure, thus putting them at greater risk of developing cancer (Clauson et al., 2013; Niraj et al., 2019). The goal of this Honors thesis project is to construct a computational model of the biochemical events involved in the FA DNA repair pathway to better understand how dysregulation of the pathway leads to the accumulation of DNA damage and, hence, increased risk of tumorigenesis.

The Fanconi Anemia Pathway

The FA pathway is initiated during DNA replication when an ICL is recognized and the replication fork stalls. The stressed fork is then recognized by the FANCM protein complex (Figure 1A), which then recruits the FA "core complex" to the site of damage (Semlow & Walter, 2021) (Figure 1B). The core complex is made up of three subcomplexes: AG20, BL100, and CEF. The AG20 complex is composed of the proteins FANCA, FANCG, and FA-associated protein 20 (FAAP20). This subcomplex helps guide the formed core complex to the nucleus. The BL100 complex is composed of the FANCB, FANCL, and FAAP100 proteins, and guides the assembly

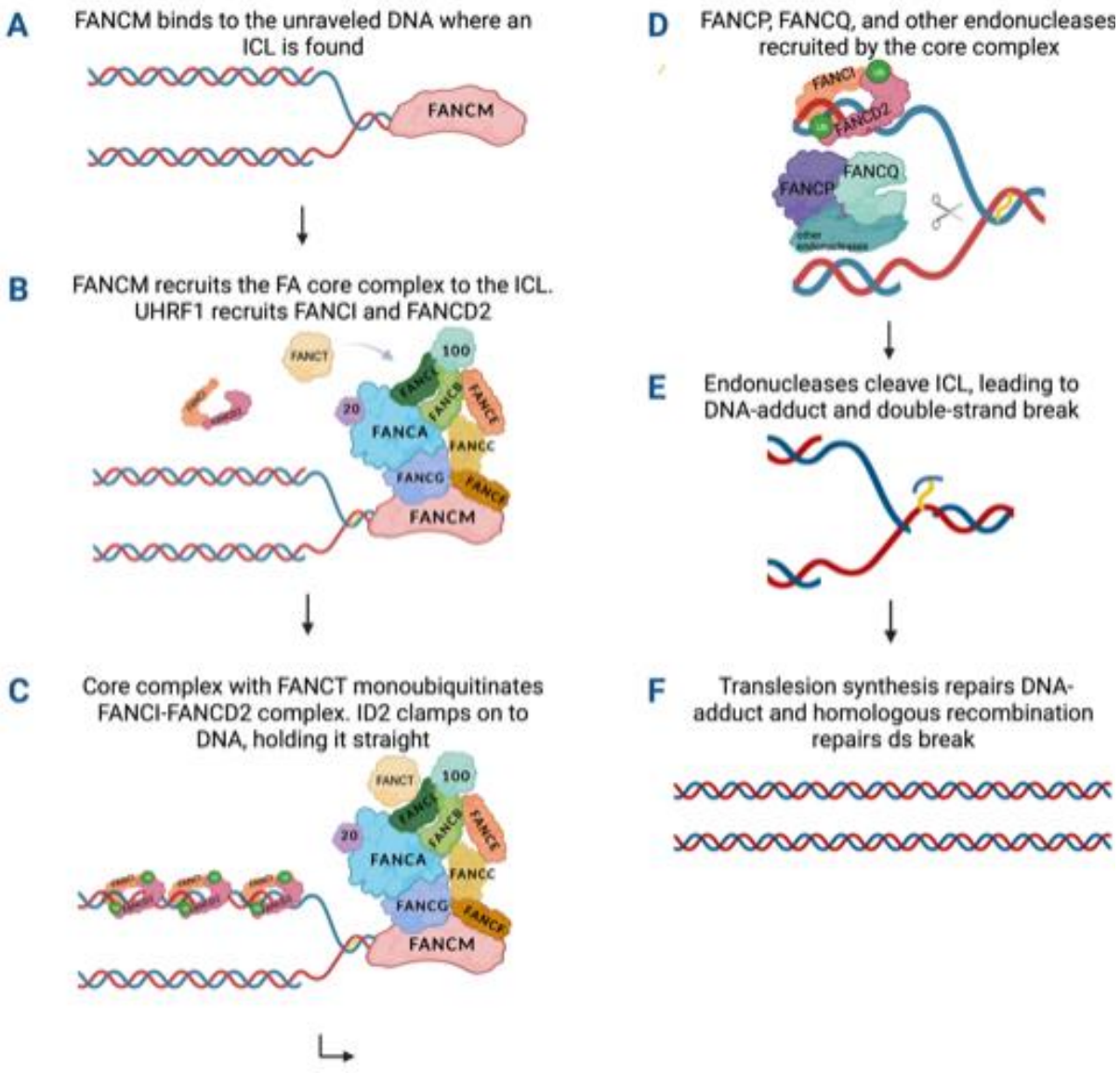

Figure 1. A Biochemical Representation of the Fanconi Anemia DNA Repair Pathway. See text for details. Created with BioRender.com

of the core complex as a whole. The CEF complex is composed of the FANCC, FANCE, and FANCF proteins. The CEF subcomplex allows the core complex to associate with FANCM and the FANCI-FANCD2 (ID2) complex (Huang et al., 2014; Rodríguez & D'Andrea, 2017). At the same time FANCM is recruiting the core complex, UHRF1 protein also recognizes the ICL and recruits FANCI and FANCD2, which form the ID2 complex (Figure 1B). The FA core complex works in conjugation with the protein FANCT to monoubiquitinate FANCI and FANCD2 (Figure

1C). FANCL works as the E3 ubiquitin ligase, and FANCT serves as the E2 ubiquitin-conjugating enzyme.

Monoubiquitylation of ID2 allows it to clamp onto the DNA near the damage, stabilizing it and providing space for the "unhooking machinery" (Sharp et al., 2021). The ID2 complex recruits FANCP, FANCQ, and other structure-specific nucleases (Figure 1D) to cleave the cross-link from one strand of the DNA helix, a process known as unhooking (Niraj et al., 2019). After the endonucleases cleave one segment of the DNA cross-link, replication continues, bypassing the lesion left hanging on one strand. The result is two helices of DNA – one with a double-strand break and one with a DNA adduct (lesion) still attached (Figure 1E) (Moldovan & D'Andrea, 2009).

Homologous Recombination and the Fanconi Anemia Pathway

After the unhooking process, the two strands of DNA are no longer linked together, but still need to be repaired. The strand with the remaining adduct is repaired through translesion synthesis (TLS). In this thesis, I focus on the methods of repair of the DNA strand with the double-strand break by a process known as homologous recombination with a sister chromatid (Hashimoto et al., 2016) (Figure 2). First, the double-strand break is detected by the MRN (Mre11-Rad50-Nbs1) complex, which binds to the 5' ends of the break, and resections the damaged DNA, leaving a 3' overhang on either side of the break (Qiu & Huang, 2021) (Figure 2A). MRN unbinds and replication protein A (RPA) attaches to the DNA strands to prevent them from recoiling and stabilize them until Rad51 arrives (Figure 2B). Rad51 proteins replace the RPAs, forming a filament with the DNA (Figure 2C). The damaged DNA then must associate with a similar template DNA for repair, preferably a sister chromatid. Rad51 molecules help guide one damaged strand of DNA to the appropriate segment of DNA template (Figure 2D). This association with the

sister chromatid DNA is called invasion (Krejci et al., 2012). Once the appropriate template is found, DNA polymerase binds and extends the damaged DNA segment (Figure 2E). Another round of DNA synthesis follows for the other damaged strand in the same way (Figure 2F). Once both strands have been properly extended, DNA ligase seals the nicks on each newly synthesized strand, and the DNA is restored to its normal state (Çağlayan, 2019) (Figure 2G).

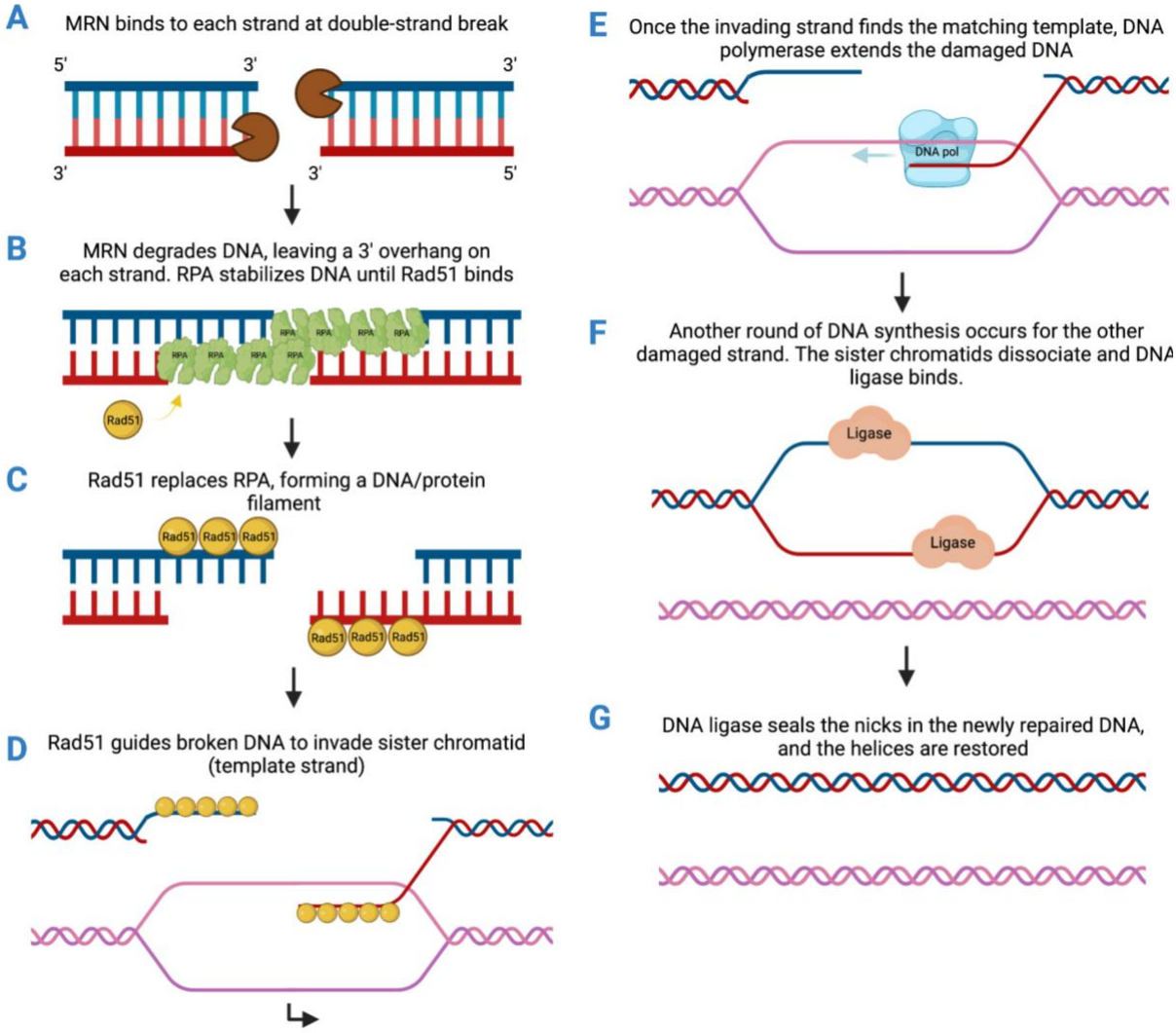

Figure 2. A Biochemical Representation of Homologous Recombination in Regard to Interstrand Cross-link Repair in the Fanconi Anemia Pathway. This model elaborates on the homologous recombination that takes place in Figure 1F. See text for a detailed description of the mechanism. Created with BioRender.com.

Methods

The main computational model of the FA DNA repair pathway was constructed using PySB, a computer modeling and simulation platform based on Python (Lopez et al., 2013). Two smaller models were also created with PySB and integrated into the main model. These described the TLS pathway (not shown) and homologous recombination (Figures 2). This platform utilizes a "rule-based" modeling approach, which is designed to simplify the translation of complex intracellular pathways into computer code (Chylek et al., 2014, 2015). The most basic component of a PySB model is called the `Monomer`. These are the proteins. One important feature of PySB monomers is that they are structured objects, meaning they have binding sites and can be present in different states, such as ubiquitinated and deubiquitinated in this model. An actual biological process or event is represented as a `Rule` in PySB. Rules define the proteins involved in the process or interaction as well as the resulting products. 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
```

In this interaction, the monomer '`A`' has one binding site, '`b`', and the monomer '`B`' has one binding site, '`a`'. The keyword '`None`' indicates that nothing is currently occupying the binding site. The '`>>`' operator indicates that the reactants (left) are being transformed into the products (right). The '`%`' between monomers indicates that they are in the same complex. The label of the binding sites changes from '`None`' to '`1`' in this simple reaction. The '`1`' represents the bond that is formed between the '`b`' binding site of the '`A`' protein and the '`a`' binding site of the '`B`' protein. Each PySB rule must have a `Parameter` object, named '`k_AB_bind`' in the example above, that represents the rate at which the reaction proceeds. The overall integrated FA DNA repair pathway model for this project includes 35 monomers, 61 rules, and 76 rate parameters. Python code

showing the monomers and rules for the homologous recombination submodel, which represents the processes shown in Figure 2, is included in Figure 3.

```
# Monomers
Monomer("MRN", ["dsb"])
Monomer("RPA", ["dsb"])
Monomer("Rad51", ["dsb"])

# Rules
Rule('MRN_binds_DSB', MRN(dsb=None) + DSB(b=None) | MRN(dsb=1) % DSB(b=1), kf_MRN_DSB, kr_MRN_DSB)

Rule('RPA_binds_DSB', RPA(dsb=None) + MRN(dsb=1) % DSB(b=1) >> RPA(dsb=1) % DSB(b=1) + MRN(dsb=None),
     k_RPA_DSB)

Rule('Rad51_binds_DSB',Rad51(dsb=None) + RPA(dsb=1) % DSB(b=1) >> Rad51(dsb=1) % DSB(b=1) +
     RPA(dsb=None), k_Rad51_BRCA2_DSB)

Rule('Pol_Zeta_binds_DSB',Pol_Zeta(dna=None) + Rad51(dsb=1) % DSB(b=1) >> Pol_Zeta(dna=1) % DSB(b=1) +
     Rad51_BRCA2(dsb=None), k_Pol_Zeta_DSB)

Rule('Ligase_binds_DSB', Ligase(dna=None) + Pol_Zeta(dna=1) % DSB(b=1) >> Ligase(dna=1) % DSB(b=1) +
     Pol_Zeta(dna=None), k_Ligase_DSB)

Rule('Ligase_repairs_DNA', Ligase(dna=1) % DSB(b=1) >> Ligase(dna=None), k_Ligase_repairs_DNA)
```

Figure 3. Monomers and Rules for the PySB Model of Homologous Recombination in ICL Repair. This image taken from the PySB homologous recombination submodel features the monomers and rules used to construct the working simulation. Parameters and initial amounts of reactants are not included.

Results

Once the main FA pathway model and two submodel were constructed and integrated, simulations were run and plots produced of the amounts of different protein complexes over time. Figure 4 tracks the amounts of the three subcomplexes of the FA core complex (AG20, BL100, and CEF), the unbound FA core complex, the FA core complex bound to ubiquitinated ID2, and the ubiquitinated ID2 complex bound to FANCP and FANCQ. The figure shows that the amount of free FA core complex quickly increases as the three subcomplexes associate with each other. The amounts of AG20, BL100, and CEF decrease as the core complex is formed, but then increase slightly to a steady equilibrium. The amount of the core complex bound to ID2Ub and the FANCP

and FANCQ endonucleases bound to ID2Ub slightly increase for a time, then level off at zero due to the rapid binding, ICL cleavage, and unbinding of these complexes.

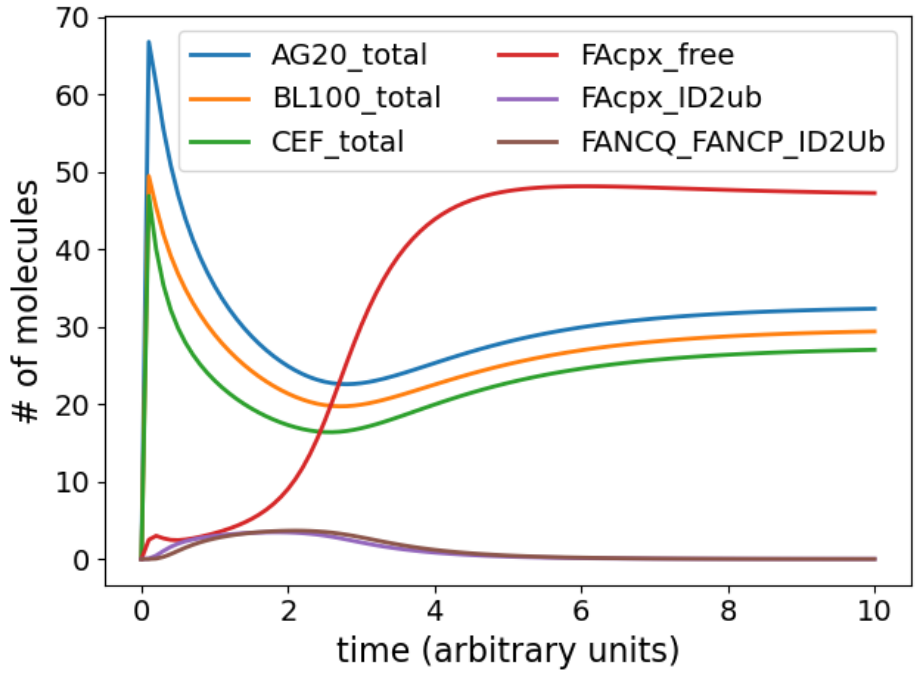

Figure 4. The Quantities of Fanconi Anemia Pathway Protein Complexes Over Time. The initial values of AG20, BL100, and CEF proteins were input at different values so that the change over time of each complex would be visible on the plot. The term "FAcpx" refers to the FA core complex. FAcpx_free is the value of core complex that is not bound to anything else. FAcpx_ID2ub is the value of the core complex that is bound to FANCM, FANCT, and the ID2 complex as it momentarily ubiquitinates ID2. FANCQ_FANCP_ID2ub refers to the moment ID2Ub recruits the FANCQ and FANCP endonucleases to the ICL, which is then cleaved.

In Figure 5, the time courses of ICLs, DNA lesions, and double-strand breaks are plotted, demonstrating how the model simulates the DNA repair process. The simulation begins with 100 ICLs, which are gradually cleaved and converted into DNA lesions and double-strand breaks. The plot shows that the number of ICLs in the simulation declines over time, while the number of DNA lesions and double-strand breaks increases for a period of time. The numbers then decrease as homologous recombination (Figure 2) and TLS take place. All DNA damage is eventually repaired and the simulation ends. Note that this simulation was performed with a preliminary set of initial protein values and rate parameters to demonstrate the functional capabilities of the model. The

numbers of molecules are all relative to each other, and not set to biological reality. Values used in future simulations will be based on data obtained from literature and experimental and clinical collaborators.

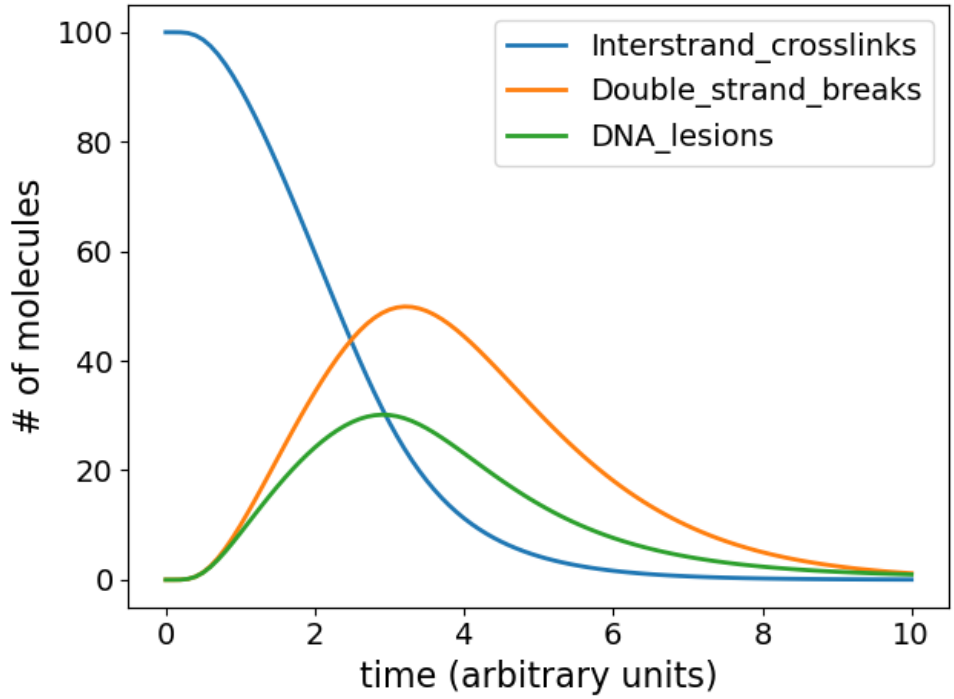

Figure 5. The Prevalence of DNA Damage as the Fanconi Anemia Pathway Progresses. Interstrand cross-links are automatically present in this model. Each ICL is then converted to a double strand break and a DNA lesion by endonuclease activity (Figure1). Then, the lesions are repaired through translesion synthesis and the double strand breaks are repaired through homologous recombination (Figure 2). The rate constant for homologous recombination was set slower than the rate constant fore translesion synthesis so that all curves would be visible on the plot. All DNA damages are eventually repaired in this model.

Discussion

The plots in Figures 4 and 5 demonstrate that the computational model was properly constructed and is running successfully. The levels of DNA damage change, as expected, in relation to the levels of the protein complexes, eventually reaching zero. In the model, no ICLs were added after the simulation began. In a living cell, new ICLs may form as old ones are repaired. This simulation did not account for this, but if it did one would expect the values of DNA damage

to level off and reach equilibrium at a value greater than zero. If the simulation was modeled based on a normally functioning FA pathway, the damage should level off at a relatively low number. However, if the simulation was modeled based on an abnormally functioning FA pathway, the DNA damage would level off at a much higher value, which would be associated with a higher incidence of cancer.

The purpose of this project was to construct a computational model of the FA pathway in humans using computational modeling and simulation software that is freely available. Since FA is such a rare condition, there have been limited studies conducted and data collected about how mutations in the FA pathway affect other parts of the body on a biochemical level. Computational models are able to simulate complex processes, such as the FA repair pathways, that can be difficult to study experimentally. Such models allow researchers to collect data in a time-efficient and cost-effective manner. Additionally, computational modeling can help identify gaps in current knowledge, generating new insights and hypotheses for topics not yet completely understood. This can further lead to new discoveries and innovations in a multitude of fields, particularly medicine (Brodland, 2015).

The ultimate goal is for this model to be used for further research on FA, and to be used as a tool in identifying druggable targets in an effort to find a cure for this disease. Another computational model of the oral microbiome is also under construction for a different project. In the future, these models could be combined to further study the connection between the FA DNA repair pathway and the increased susceptibility of FA patients to oral health issues. Various computational models of biochemical pathways in the human body could one day be integrated together to form a simulation very similar to what occurs inside of a living person. These models could eventually be made patient-specific, meaning the models could predict how a certain

patient's body would react to various drugs or treatments before they physically take them. This could be very beneficial for patients with FA, who are more sensitive to many chemotherapeutic treatments. For this large, integrated model to come to fruition, smaller, more specific models must be constructed. Thus, this project is just one step in the direction of developing effective treatments and cures for FA.

Conclusion

Fanconi anemia is still not fully understood. It is a very rare condition that predominantly affects children, making it difficult to study the biochemistry of the disease. The purpose of this project was to develop a tool that could be used to study and experiment with the FA biochemical pathway. The results show that this model accurately mimics, to a degree, the events that take place inside the cells of someone with a properly functioning FA repair pathway, given that the value of ICLs, lesions, and double strand breaks all go to zero. Using this model, one would be able to study the outcomes of various mutations to genes in the FA pathway. For instance, whilst constructing the model, if one protein or action was incorrectly coded, the values of DNA damage would not go to zero, indicating that they were not properly repaired.

This model could be used to further the search for a druggable target and treatment for those with Fanconi anemia. Improving our understanding of the molecular pathways underlying FA and the factors that lead to cancer has the potential to greatly improve patient therapies. However, there is much research to be done in order to completely understand the precise reactions of molecules and proteins that take place in the FA pathway. The computational model was created by compiling the little that is known about this pathway. Hopefully this and future models can serve as a gap between what is known and what is to be discovered about the FA DNA repair pathway.

References

Anaemia. (n.d.). Retrieved 10 April 2023, from https://www.who.int/health-topics/anaemia#tab=tab_2

Brodland, G. W. (2015). How computational models can help unlock biological systems. *Seminars in Cell & Developmental Biology*, *47–48*, 62–73. https://doi.org/10.1016/J.SEMCDB.2015.07.001

Çağlayan, M. (2019). Interplay between DNA polymerases and DNA ligases: Influence on substrate channeling and the fidelity of DNA ligation. *Journal of Molecular Biology*, *431*(11), 2068. https://doi.org/10.1016/J.JMB.2019.04.028

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). https://doi.org/10.1088/1478-3975/12/4/045007

Chylek, L. A., Harris, L. A., Tung, C. S., 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. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*, *6*(1), 13–36. https://doi.org/10.1002/WSBM.1245

Clauson, C., Schärer, O. D., & Niedernhofer, L. (2013). Advances in Understanding the Complex Mechanisms of DNA Interstrand Cross-Link Repair. *Cold Spring Harbor Perspectives in Biology*, *5*(10). https://doi.org/10.1101/CSHPERSPECT.A012732

Fanconi anemia: MedlinePlus Genetics. (n.d.). Retrieved 10 April 2023, from https://medlineplus.gov/genetics/condition/fanconi-anemia/

Furquim, C. P., Soares, G. M. S., Ribeiro, L. L., Azcarate-Peril, M. A., Butz, N., Roach, J., Moss, K., Bonfim, C., Torres-Pereira, C. C., & Teles, F. R. F. (2017). The Salivary Microbiome and Oral Cancer Risk: a Pilot Study in Fanconi Anemia. *Journal of Dental Research*, *96*(3), 292–299. https://doi.org/10.1177/0022034516678169

Goswami, M., Bhushan, U., & Goswami, M. (2016). Dental Perspective of Rare Disease of Fanconi Anemia: Case Report with Review. *Clinical Medicine Insights. Case Reports*, *9*, 25. https://doi.org/10.4137/CCREP.S37931

Hashimoto, S., Anai, H., & Hanada, K. (2016). Mechanisms of interstrand DNA crosslink repair and human disorders. *Genes and Environment*, *38*(1), 1–8. https://doi.org/10.1186/S41021-016-0037-9/FIGURES/2

Huang, Y., Leung, J. W. C., Lowery, M., Matsushita, N., Wang, Y., Shen, X., Huong, D., Takata, M., Chen, J., & Li, L. (2014). Modularized Functions of the Fanconi Anemia Core Complex. *Cell Reports*, *7*(6), 1849. https://doi.org/10.1016/J.CELREP.2014.04.029

Krejci, L., Altmannova, V., Spirek, M., & Zhao, X. (2012). Homologous recombination and its regulation. *Nucleic Acids Research*, *40*(13), 5795. https://doi.org/10.1093/NAR/GKS270

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). 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

Qiu, S., & Huang, J. (2021). MRN complex is an essential effector of DNA damage repair. *Journal of Zhejiang University. Science. B*, *22*(1), 31. https://doi.org/10.1631/JZUS.B2000289

Rageul, J., & Kim, H. (2020). Fanconi anemia and the underlying causes of genomic instability. *Environmental and Molecular Mutagenesis*, *61*(7), 693–708. https://doi.org/10.1002/EM.22358

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

Semlow, D. R., & Walter, J. C. (2021). Mechanisms of Vertebrate DNA Interstrand Cross-Link Repair. *Annual Review of Biochemistry*, *90*, 107–135. https://doi.org/10.1146/ANNUREV-BIOCHEM-080320-112510

Sharp, M. F., Bythell-Douglas, R., Deans, A. J., & Crismani, W. (2021). The Fanconi anemia ubiquitin E3 ligase complex as an anti-cancer target. *Molecular Cell*, *81*(11), 2278–2289. https://doi.org/10.1016/J.MOLCEL.2021.04.023