The diversity of germinal center antibodies using a genetic algorithm.

Ruchina Shakya
Dept. of Computer Science
University of New Mexico
rshakya@unm.edu

Nicholas Bacon
Dept. of Computer Science
University of New Mexico
Nbacon@unm.edu

Yingfan Wang
Dept. of Computer Science
University of New Mexico
yingfan3@unm.edu

***Abstract*—This paper will introduce how B cells and antibodies at germinal centers work when the body gets infected, how they evolve or how they are selected. To simulate the whole process, we used a genetic algorithm based on Germinal Center dynamics. We initialized the population of antibodies and epitopes to see what will happen after iterations. As follows are our main findings: a.Fig.2 tells us , the initial population of antibodies is 8000 and we assume every epitope has 8 amino acids, after 25 days of duplications and mutations, about 1100 antibodies have survived and they only focused on 2 amino acids. b.Comparing Fig.1 and Fig.3 ,we can see as the affinity becomes higher, the odds of antibodies mutation gets lower.**

I. Introduction

In this paper, we will focus on the mutation of B cells and their antibodies in Germinal centers using a model called GCGA (Genetic Algorithm based on Germinal Center dynamics). Genetic algorithm shows the process to find the fittest individual through natural selection .Germinal centers are the places where antibodies diversified and affinity maturation occurs. We will use this model to simulate how the fittest B cells and antibodies are selected to protect people from infection.

We are going to represent a single germinal center in our GCGA. Germinal centers are transiently formed structures within the B cell zone (follicles) in secondary lymphoid organs – lymph nodes, ileal Peyer's patches, and the spleen [12]. This GCGA is an analog model of an immune response. The working definition of an immune response is the way that body defends itself against substances it sees as harmful or foreign. In our GCGA we are going to represent Antibody and Epitope. Antibody is a large, Y-shaped protein used by the immune system to identify and neutralize foreign objects such as pathogenic bacteria and viruses and it's produced by B cells [13]. Epitope is the specific piece of the antigen to which an antibody binds [14]. B cells are a type of white blood cell and a lymphocyte sub type that are responsible for the fighting of an infection [15].

When making our analog of a germinal center in our GCGA. Once the body is infected by antigens ,the immune system will recognise them as enemies and start working. B cells will produce a protein called antibody and mutation might happen during replication. Antigen has a structure called epitopes which contains 6-8 amino acids ,what antibodies do is to bind to those epitopes. Once antibodies bind to epitopes successfully, B cells that produce those antibodies survive and continue somatic hyper mutation . We assume hyper mutation of our B cells will happen after the third day when all the naive b cells are made. Then for the next 3 weeks every 8 hours mutate, which means B cells can directly replicate up to 63 times. We would like to figure out given our model, what the diversity of the ending antibodies are in blood lines (how many of the original naive b cells have children) and target antigens.

II. Methods and Results

The exploration versus exploitation trade-off is a well-known problem. The goal of exploration is to make small changes that seem meaningless but can pay off with a great reward. The goal of exploitation is to repeat actions that give good results in the past. We designed a simulation to represent Germinal Center dynamics using a genetic algorithms model. This genetic algorithms model uses the exploration versus exploitation trade-off to get closer to the goal.

A. Methods

In the genetic algorithm, we are representing 8 static epitopes as the goals and a large dynamic population up to 8000 of antibodies [1]. These epitopes and antibodies are both made up of nucleic acid. We used a randomly generated 48-bit string that can be used to represent the epitopes and antibodies. We came up with 48 by picking a non-arbitrary number of 8 for the amino acids. Each amino acid is made out of three nucleic acids and there are 4 nucleic acids(ACGT). Therefore a 2bits/nucleic acids* 3nucleic acids/amino acids * 8 amino acids is 48 bits.

This genetic algorithm uses a modified euclidean-overlapping metric (EOM) for the standard fitness function. There are many metrics we could use like bit wise hamming distance, pairwise hamming distance, but we chose to use minimum amino acids distance. For minimum amino acids

distance we first pair up 6 bits (1 amino acid) of the antibodies and an epitope then to filter out all pairs that are parts of the same amino acid group. With the resulting list we find the sum of the minimum amount of pairwise edits to change the antibodies amino acid into the one of the epitopes amino acids group. We repeated this process 7 more times and took the minimum number of the pairwise edits. We find this to be more accurate than a hamming distance metric because with hamming distance we only care about the bits that are different. But because there are 64 acid pairs and their 20 groups we could use any pair of three and say this is how many mutations it will take to make it in a random group.

The population of this genetic algorithm is controlled by 3 piece wise functions. All of these functions are based on the works of Philippe Robert and his team of researches for their paper "How to simulate a germinal center" [1], we did not need any of the advanced knowledge of Germinal Center or over modeling so we kept our model very basic by hard coding the function.

$$Pop(x) = \begin{cases} 20^x & x < 3.0 \\ -266.667(x^5) + 6333.33(x^4) & 3 \leq x \leq 6 \\ \quad -59666.7(x^3) + 278917(x^2) & \\ \quad -647317x + 602500 & \\ 10250 - (7775 * x)/9 & 6 \leq x \leq 41 \\ \quad +(250 * (x^2))/9 & \\ \quad -(25 * (x^3))/81 & \\ 0 & else \end{cases}$$

This population function states the maximum amount of antibodies at any given point. This function is an analog model to the human body infection because [2] states that in many cases the antibody responses with multiple rapid peaks and decay in our case 2.

$$fitness(x) = \begin{cases} 0 & x < 3.0 \\ -0.733333 + 0.337037x & \\ \quad -0.0285494(x^2) & \\ \quad +0.00113169(x^3) & 3 \leq x \leq 16 \\ \quad -0.0000171468(x^4) & \\ fitness(16) & else \end{cases}$$

The next function is affinity cut off rate. This function does not have a lot of peer evidence but makes sense as time goes on the acceptance rate and affinity of an antibody should be higher.

$$probsofMut(x) = \begin{cases} 0 & x < 3.0 \\ 1.07533 - 0.295852x & 3 \leq x \leq 16 \\ \quad +0.033784(x^2) & \\ \quad -0.00163374(x^3) & \\ \quad +0.0000281207(x^4) & \\ 0.1 & else \end{cases}$$

The last function is the odds of an antibody mutating. Another function with very low peer evidence but anecdotally makes a lot of sense as time at the beginning there should be a lot of hyper mutation to explore the space and get a lucky mutation that raises the fitness by a large percent. And near the end mutations should be rare because the left over antibodies are very close to their goal epitopes.

The mutating function of the population for the genetic algorithm is simple. The algorithm picks a random number between 0 and 24 then. Saying the random number is N we use the change the bits of the 2*N and 2*N+1 potions to a random bit so it does have a 1/ 4 chance of staying the same. We choose to model this way because according to peer evidence the offspring have point mutations or single-base substitutions and insertions and deletions are rare [3].

Puting all of this into a genetic algorithm. We get this hill climbing algorithm:**Algorithm 1**

B. Results

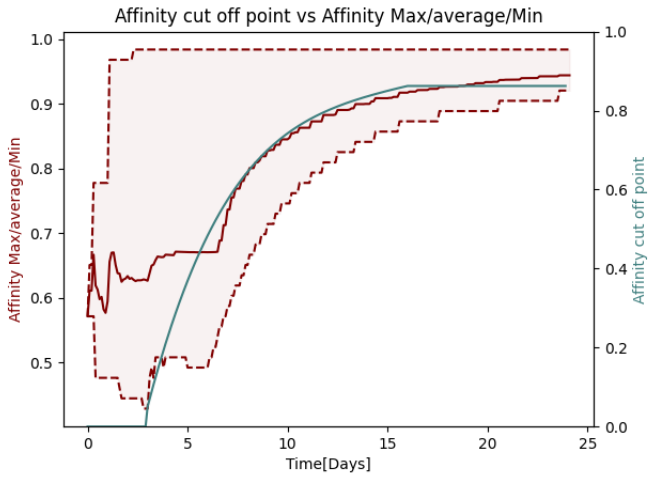

Fig. 1. Represents the binding affinity and the affinity cut-off point of an antibody as the day of infection increases.

The figure [fig. 1] shows an increase in affinity as the day of infection increases. The increase in affinity indicates

Algorithm 1 Genetic Algorithm based on Germinal Center dynamics

```
Antibody ← list
epitopes ← list
day ← 0
fill epitopes with epitopes
while day < end do
    while Antibody.length < Pop(day) do
        fill Antibody with random Antibody
    end while
    while Antibody.length > Pop(day) do
        remove Antibody with weakest Antibody
    end while
    if day > 3 then
        Antibody ←filter Antibody with
                λx=fitness(day) > x.fitness
        while Antibody.length < Pop(day) do
            fill Antibody with random from  Antibody
        end while
        Antibody ←map  Antibody with
                λx.while (random > mutatingOdds(day)) then
                        mutating(x)
                end while
```

naive B cells are proliferating more rapidly leading to cell survival i.e producing more antibodies(causes lower death rates).The two peaks in the solid red line indicate that the antibodies with the increased affinity were produced right after the infection. This shows that the B cell encountered the pathogen that matches its membrane bound antibody. The antibody quickly divided and differentiated into memory B cell or an effector B cell to produce more antibodies [6]. The binding of antigen-antibody was high.The affinity cut off starts increasing after 3 days. There is a steady increase in the affinity cut off point. It indicates that antibody levels decay over time.The cut off point stabilizes after a certain time(i.e after 15 days).

The figure [fig. 2] demonstrates the population of naive B cell antibodies produced after an infection and the antigen population. The red line in the graph represents the uniqueness.Some of the antibodies produced to neutralize the pathogen do not bind with the antigen. Therefore, it is counted under unique bloodlines.The antigen production rises up from the first day of infection.The epitope has 8 amino acids. There is a jump in the antibody population and unique blood lines after 3 days.The antibody production is close to 8000 in the beginning of the infection. This is because the epitope on the antigen is high (8 amino acids high) i.e the amount of antibody produced by the body depends on the number of antigen that invades the body.The antibody population peak starts decreasing and two points(population of antibody and unique bloodline) diverge from each other. The unique blood lines decrease as the pathogens start mutating faster than

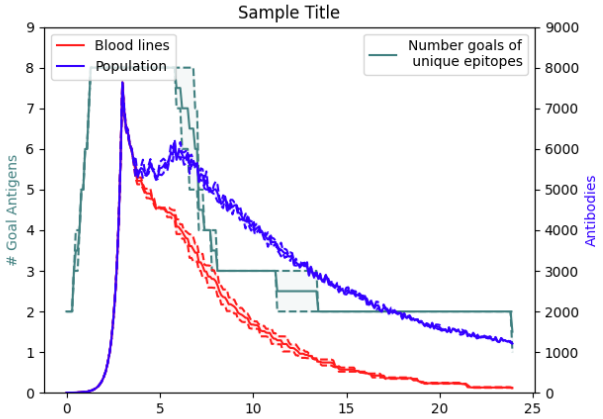

Fig. 2. Comparing the goal antigens(green), bloodline(red) and antibody population(blue) in the B cell as the day of infection increases.

the unique antibody B cells.The antibody production of the same parent B cell decreases. At 5500 antibody population we can assume that the antibody production decreased due to previous antigen-antibody binding action. The antibody neutralized the antigen.

Fig. 3. Comparing the mutations and odds of mutation in the antibody production as the day of infection increases. The odds of mutation increases fitness.

The figure [fig. 3] shows the mutations and the odds of mutation.B lymphocytes recognize pathogens through the binding of specific B-cell receptors. B-cell receptors undergo a higher rate of mutation during proliferation than the normal rate mutation [6].In the figure,the mutation takes a high peak after the third day of infection. The somatic hypermutation in B-cells does not distinguish between favorable and unfavorable mutations. [3]Therefore the antibodies produced might have higher,lower or no change in affinity for the antigen. The two points diverge from each other after the third

day of infection.The second spike appears in the mutations per day after diverging. The second spike observed is the high mutation caused due to the replication of B cells in every 8 hours. The fluctuation observed in the odds of mutation implies that there is a most likely chance of mutation in every 8 hours of replication.

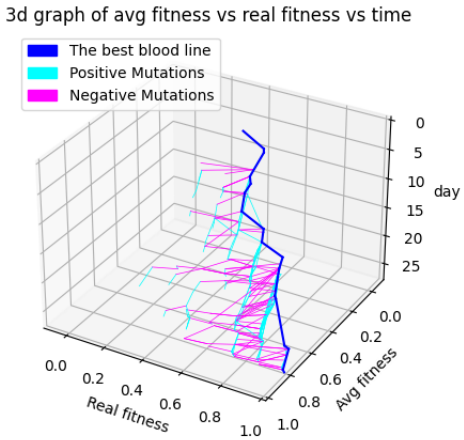

Fig. 4. A 3D representation of real and average fitness as the day of infection increases.

The 3D figure [fig. 4] demonstrates survival of B cells based on the fitness(ability to bind to epitopes that happen to be in that germinal center.Each new mutation in an individual can increase its fitness, decrease its fitness, or have no effect on its fitness. In the figure,the pink scribble represents a negative mutation. Some of the positive mutations are found switching to negative mutations. Therefore, disappearing in the end. The average fitness is always increasing in positive mutation with the increase in days of infection.The negative mutation peak is noticed on day 15 of infection.We can observe a positive correlation between real fitness and average fitness.

The silent mutation would lead to non-functional antibody production [7]. Due to this reason, vaccines are necessary to boost up the functional antibody production.

Hundreds of simulations were run to visualize the antibody production after the pathogen invaded the body with the increase in day of infection. After observing all the graphs, we can clearly see that there is diversity in the antibody population. The responsiveness of the B cell was quite impressive. Antibody production reaction was seen after 3 days.The recognition of antigen took quite some time. We can assume that the antigen was new to the B cell. The memory cell could not detect the antigen.But once the B cell recognized the antigen(toxic), the antibody started binding with the antigen. We can clearly observe that the number of antigen starts decreasing as the antibody population increases with the high affinity. The mutation on the antibody is high. Therefore, we can see some of the antibodies with high fitness. The mutation later on decreases but the binding affinity gets higher. The antibody becomes better at binding. The antibody production with the vaccine injected would be a great analysis to compare and contrast with normal B cell antibody production. Currently, we are only focusing on the normal B cell infection and its responsiveness.

III. Discussion Conclusions

From the simulation we know that as the infection day increases, the amount of affinity also goes up, which means antibodies with higher affinity are kept; the ones that are not kept, no longer reproduce and mutate and eventually die off . These antibodies have stronger ability to bind to epitopes. By forcing the population of antibodies,affinity cut off points to always be increasing, and the odds of antibodies being mutated in their respective ways, means that B cells in a germinal center may be laser focused on one or two epitopes at the end of the infection. This laser focus is good because infection will be fought with one or two epitopes that will make the infection die out faster. The body will only use the remaining antibodies against antigens that match so if the virus mutates someone might be able to be reinfected.

It's easy to know that the concentration of antibodies increases very fast at the beginning and reaches the peak value around the third day, then goes back to 5000 around the tenth day and 1000 on the twenty-fifth day. Likewise, if someone gets the first covid shot and after a few months, the antibodies will also be limited. After that time, all antibodies will dwindle down to the strongest one but have not been tested so might be ejected from the population as a hole. So B cells have a lower chance to mutate but viruses have billions of hosts that could mutate them. Thus , we highly recommend people to get booster shots with a different variant to get new antibodies and keep the one we made in the first shot .

As we talked in class, B cells have the initial DNA called Immature B Cell DNA, then DNA combines with different gene segments to form antibodies. At Lymph Nodes, they have the chance to meet and match antigen, B cells are selected and some copies of them remain as memory cells. If the body gets infected again, they will play a role (produce large amounts of antibodies in a short time). Those B cells and antibodies that have higher affinity will be kept ,this is the evolution of B cells. From our findings, antibodies can only last for a short time in a human's body. More precisely, antibodies to covid can last several months. [9]Therefore, getting shot regularly might be a good way against covid. This can help get a better herd immunity against covid 19 and the new mutations.

Since our findings are based on genetic algorithms and our values came from a germinal center simulation, some of them

may not be able to be applied in the real world. For example, we assume the frequency of B cells replication is every 8 hours, but actually, it's not a constant time. In fact, location of cells or body's age can both make a difference to the speed of replication. [9] We think that the analog model we have made is the best under our knowledge and circumstances.

ACKNOWLEDGMENT

We would like to start out by thank Philippe Robert and his team of researches for their paper "How to simulate a germinal center", the we used there paper to make formals to our genetic algorithm. Next we would like thank Sal Khan from Khan Academy and Hank Green form crash course for teaching us biology. We want to thank our super-group in no particular order: Eliza Gilber, Wesley Swedenburg, Viacheslav Zhuravlev, Liangkun Yu and Chaeeun Park.

CONTRIBUTIONS

Nicholas wrote the code to generate all the four figures. He provided the method section. Yingfan did the Abstract, Introduction, Discussion and Conclusions parts.Ruchina analyzed the graphs with the help of Nicholas,explained the results and formatted the references. Nicholas formatted the paper. Everyone took part to edit the final paper.

REFERENCES

[1] Robert, Philippe A., et al. "How to simulate a germinal center." Germinal Centers (2017): 303-334.
[2] Hammarlund, Erika, et al. "Plasma cell survival in the absence of B cell memory." Nature communications 8.1 (2017): 1-11.
[3] Wikipedia contributors. "Somatic hypermutation." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 3 Mar. 2022. Web. 28 Mar. 2022.
[4] Post, Nathan, et al. "Antibody response to SARS-CoV-2 infection in humans: a systematic review." PloS one 15.12 (2020): e0244126.
[5] Pilzecker, Bas, and Heinz Jacobs. "Mutating for good: DNA damage responses during somatic hypermutation." Frontiers in immunology 10 (2019): 438..
[6] Kaji, Tomohiro, et al. "Both mutated and unmutated memory B cells accumulate mutations in the course of the secondary response and develop a new antibody repertoire optimally adapted to the secondary stimulus." International immunology 25.12 (2013): 683-695..
[7] Yaari, Gur, et al. "The mutation patterns in B-cell immunoglobulin receptors reflect the influence of selection acting at multiple time-scales." Philosophical Transactions of the Royal Society B: Biological Sciences 370.1676 (2015): 20140242.
[8] Laidlaw, Brian J., and Ali H. Ellebedy. "The germinal centre B cell response to SARS-CoV-2." Nature Reviews Immunology 22.1 (2022): 7-18.
[9] Turner, Jackson S., et al. "SARS-CoV-2 infection induces long-lived bone marrow plasma cells in humans." Nature 595.7867 (2021): 421-425.
[10] Mitchell, Melanie. Complexity: A guided tour. Oxford University Press, 2009.
[11] Flake, Gary William. The computational beauty of nature: Computer explorations of fractals, chaos, complex systems, and adaptation. MIT press, 2000.
[12] Wikipedia contributors. "Germinal center." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 11 Feb. 2022. Web. 29 Mar. 2022.
[13] Wikipedia contributors. "Antibody." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 18 Mar. 2022. Web. 29 Mar. 2022.
[14] Wikipedia contributors. "Epitope." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 23 Feb. 2022. Web. 29 Mar. 2022.
[15] Wikipedia contributors. "B cell." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 25 Jan. 2022. Web. 29 Mar. 2022.