LUMEN ET VIRTUS

IN SILICO ANALYSIS OF INTERLEUKIN 10 (IL 10) VARIANTS THAT ARE CANDIDATES FOR AUTOIMMUNE DISEASES IN HUMANS

https://doi.org/10.56238/levv15n41-002

Vitória Giovanna de Souza Cassiano[1], Nara Suzy Aguiar de Freitas[2], Maria Helena Queiroz de Araújo Mariano[3], Eliézer Rushansky[4] and Maria de Mascena Diniz Maia[5]

ABSTRACT

IL 10 is a cytokine with immunoregulatory functions, acting both in the suppression and stimulation of the immune system. Polymorphisms in the promoter region of the IL 10 gene can influence its expression and the production of IL 10, which varies between individuals due to genetic factors. Using bioinformatics tools, the study analyzed three variants of the gene and found associations of these with autoimmune diseases, such as rheumatoid arthritis, systemic lupus erythematosus, and inflammatory bowel disease. The study also investigated the phylogeny of the IL 10 gene in mammals, including Homo sapiens, Mus musculus (mouse), Equus caballus (horse), Rattus norvegicus (rat), and Mesocricetus auratus (Syrian hamster). The phylogenetic analysis demonstrated that the IL 10 gene is highly conserved among species, suggesting that it evolved from a common ancestor and has been preserved as a result of its adaptive value, being a gene conserved even after divergent evolution among the species studied. Analysis of the dN/dS ratio revealed a negative selection, indicating that changes in the amino acid sequence were avoided by natural selection. A phylogenetic tree was constructed using the "Maximum Likelihood" method, revealing a close relationship between the genes of R. norvegicus, M. musculus and M. auratus, while the genes of H. sapiens and E. caballus were grouped with high reliability. The research reinforces the importance of understanding the evolution of genes for comparative medicine and the classification of organisms, helping to identify adaptive changes between homologous genes.

Keywords: Phylogeny. IL10 Gene. Polymorphism. Autoimmune Diseases.

[1] Federal Rural University of Pernambuco, Recife, Brazil
[2] Doctor in Genetics from the Federal University of Pernambuco
Federal Rural University of Pernambuco
E-mail: nara.safreitas@ufrpe.br
[3] Doctor in Biological Sciences from the Federal University of Pernambuco
University of Pernambuco
Rua Arnóbio Marques, Santo Amaro, Recife – PE, Brazil
E-mail: irpe.diretoria@gmail.com
[4] Post-graduation in Internal Medicine from the University of Pernambuco
University of Pernambuco
Rua Arnóbio Marques, Santo Amaro, Recife – PE, Brazil
E-mail: eliezer.rushansky@upe.pe.gov.br
[5] Doctor in Biological Sciences from the Federal University of Pernambuco
Federal Rural University of Pernambuco
Rua Dom Manuel de Medeiros, Dois Irmãos, Recife – PE, Brazil
E-mail: maria.dmaia@ufrpe.br

INTRODUCTION

Interleukin 10 (IL-10) is an anti-inflammatory cytokine produced by various types of immune cells, such as macrophages, T and B lymphocytes, dendritic cells, and mast cells. The IL-10 gene, in humans, is located on chromosome 1q31-32 and is composed of five exons and four introns. This interleukin plays an important role in regulating the immune response, inhibiting the production of pro-inflammatory cytokines and stimulating the differentiation of regulatory T cells (Silva, 2022).

The expression and function of cytokines, such as IL-10, can be affected by genetic variants capable of altering the structure or regulation of these molecules. These variants can be classified as synonyms, when nucleotide substitution does not change the encoded amino acid or non-synonymous or missense when a nucleotide substitution leads to an amino acid substitution that may or may not result in a pathogenic variant depending on the effect of the amino acid substitution on protein function and structure (Borges, 2021). Some Missense variants may be neutral, not significantly affecting the protein, while others may be harmful, causing loss or gain of function, instability, aggregation, or abnormal interaction with other molecules (Costa, 2022).

In silico analysis is a computational approach that uses bioinformatics tools to predict the impact of genetic variants on the structure and function of proteins. These tools can use different methods, such as evolutionary conservation analysis, molecular modeling, dynamics simulation, interaction network analysis, or machine learning, which can be useful to identify Missense variants that are associated with autoimmune diseases, which are conditions characterized by an abnormal immune response against the body's own antigens. These diseases can affect various organs and systems, such as the skin, joints, blood, nervous system, gastrointestinal tract, and endocrine system.

Similarly, phylogenetic analysis is a crucial technique in biology when seeking to understand the evolutionary relationships between different organisms. Using morphological, behavioral and, mainly, molecular data, this analysis allows the reconstruction of the evolutionary history of species (Caldart et al., 2016). The importance of phylogenetic analysis lies in its ability to elucidate how species are related and how they have evolved over time. This is essential for the classification of living beings and for the understanding of biodiversity (Santos; Klassa, 2012). In addition, phylogeny has practical applications in areas such as virology, immunology, and molecular epidemiology, where it is used to track the evolution of pathogens and understand the spread of diseases (Caldart et al., 2016).

Considering the above, the present article aimed to investigate the mutations of the IL10 gene at the nucleotide and amino acid level to verify which mutations are missenses using the Ensembl platform (http://www.ensembl.org/index.html) and also to analyze the degree of pathogenicity and classify as pathogenic or benign the missense variants provided by the databases and visualize the IL 10 protein in 3D, highlighting the position of the amino acids where mutation occurred. Finally, we used the alignment of the gene sequences of the placental mammal species Homo sapiens (human), Mus musculus (mouse or house rat), Equus caballus (horse), Rattus norvegicus (Twister) and a marsupial mammal Mesocricetus auratus (Syrian Hamster) in order to analyze the possible similarities and the type of pressure prevailing throughout evolutionary history, exposing the results in a phylogenetic tree.

MATERIAL AND METHODS

TOOLS USED

Ensemble: is a vertebrate genome navigator that supports research in comparative genomics, evolution, sequence variations, and transcription regulation. This platform allows you to check gene annotations, generate multiple alignments, predict regulatory functions, and gather disease-related data. Among the tools available in Ensembl are BLAST, BLAT, BioMart, and the variant effect predictor (VEP), covering all the species it supports.

MEGA: Molecular Evolutionary Genetics Analysis is a bioinformatics software widely used for statistical analysis of molecular sequence data for the construction of phylogenetic trees. It offers a variety of tools to align sequences, calculate evolutionary distances, infer phylogenies using various methods, such as Maximum Likelihood, Maximum Parsimony, and Neighbor-Joining Method, and perform statistical tests to evaluate hypotheses about molecular evolution.

NCBI: National Center for Biotechnology Information provides access to a wide variety of biological databases, data analysis tools, and research resources, including GenBank, one of the largest collections of genomic sequences in the world. In addition, NCBI develops and maintains software tools such as BLAST for sequence comparison, and offers platforms such as PubMed, which allows access to a vast library of scientific articles in the biomedical field.

Mutation Assessor: predicts the functional impact of amino acid substitutions in proteins, such as mutations identified in cancer or missense polymorphisms. The evaluation of the functional impact is made based on the evolutionary conservation of the altered

amino acid in homologous proteins. This method has been validated in an extensive set of polymorphic and disease-associated variants (OMIM), covering about 60 thousand variants.

SIFT: analyzes whether an amino acid substitution affects the function of a protein based on sequence homology and the physical properties of amino acids. SIFT can be applied to both non-synonymous polymorphisms, which occur naturally, and laboratory-induced missense mutations.

Polyphen 2: Tool that predicts the possible impact of an amino acid substitution on the structure and function of a human protein, using physical and direct comparative considerations.

EXECUTION OF THE PHYLOGENETIC ANALYSIS

Data on the IL-10 gene of the respective mammals: Homo sapiens (human), Mus musculus (mouse or house rat), Equus caballus (horse), Rattus norvegicus (Twister) and an example of marsupial mammal Mesocricetus auratus (Syrian Hamster) were extracted from the Ensembl database (https://www.ensembl.org/index.html) and the National Center for Biological Information (NCBI) database (http://www.ncbi.nlm.nih.gov/snp). They were searched one by one, selecting the cDNA and opting for the coding region of the gene in order to facilitate the identification of the IL-10 protein in the genetic code. The files were saved to be aligned and for this the MEGA 11 (Molecular Evolutionary Genetics) software was used, available at: (https://www.megasoftware.net/). The previously collected nucleotide sequences were loaded in "multiple" into a single file that contained only the "Cds" regions or coding regions in the MEGA 11 program. The alignment was performed by the software, and it was possible to observe all the similarities and divergences between the sequences of each species. A similarity matrix was generated by MEGA 11 (Table 1) and the Maximum Likelihood Estimate (VME) that estimates the matrix parameters, which usually include the transition and transversion rates between the bases. These parameters represent the probability that one database will be replaced by another over time (ScienceDirect Topics, 2024).

In EMV (Table 1), each entry shows the probability of substitution (r) from one base (row) to another base (column) (Tamura, 2004). To put it simply, the sum of the values of r is equal to 100. The rates of different transitional substitutions are shown in bold, and those of transverse substitutions are shown in italics. All ambiguous positions were removed for each pair of sequences (pair exclusion option). A total of 750 positions were verified in the final dataset. Starting from the same alignment, we obtained a new table (Table 2) where the difference between the non-synonymous and synonymous distances per location

between sequences is shown. The analyses were performed using the Kumar model (Kumar, 2000). This analysis involved 5 nucleotide sequences. All ambiguous positions were removed for each pair of sequences (pair exclusion option). A total of 178 positions were observed in the final dataset (Table 2). The method used to calculate it was the Kumar, Kimura 2 model comparing the substitution rates between the orthologous sequences. Finally, the sequences were submitted to phylogeny testing using the "Bootstrap method" and the creation of the evolutionary tree using the following configurations: selecting the nucleotide substitutions using the "Tamura-nei" model with the number of replications in 1000.

A phylogenetic tree was generated using the MEGA software, which consists of a diagram capable of showing the evolutionary relationships between the organisms studied, built based on the hypotheses presented by the algorithms of the mentioned software, used for the alignment. The tree was generated from the "Maximum likelihood tree" method according to the Tamura-nei model that uses the likelihood patterns already mentioned above as well as the maximum parsimony patterns, that is, the main molecular markers observed were used to gather the groups based on the main similarities between the IL-10 gene, in the species studied.

PERFORMING VARIANT ANALYSIS

To analyze the different types of variants and their consequences for the function of IL-10, we searched the Ensemble software for the cDNA sequence of the human IL-10 gene and selected the IL 10-206 transcript (ENST00000659642.2) which is composed of 6 exons, 13 domains and approximately 2911 alleles from which the variants were selected: rs750010814, rs1674874871 and rs568879359 to be submitted to tests in the Mutation Assessor software, Polyphen 2 and SIFT, whose results were listed in tables for a better understanding of these. The Mutation assessor, Polyphen 2, and SIFT are bioinformatics tools that predict the functional impact of amino acid substitutions on proteins, assessed based on the evolutionary conservation of the affected amino acid in protein homologues, using physical and direct comparative considerations. The use of different tools helps us to analyze and prove or not the probability of a deleterious potential among the variants studied.

RESULTS AND DISCUSSION

After the alignment and correction of the sequences of the five species studied, it was possible to observe the existence of 537 regions among them, of which 387 were

100% conserved, that is, when we observe a conserved region, it means that it was present in the common ancestral species and was preserved in the contemporary species that were submitted to the analysis. The full conservation of a region indicates that it has been maintained by natural selection. However, an identical nucleotide at a given position may have been preserved due to selection versus alteration in sequence, but not all conserved characters may be functional (Mayr, E. 1982). Similarly, it was possible to observe the existence of 150 variable regions that consist of specific segments of the DNA molecule, which exhibit variations between individuals. These variations may involve mutations, insertions, or deletions of bases that may be associated with specific genes, regulators, or other biological functions (Cruzito, 2024).

Regions with variations are fundamental during the evolutionary process because they allow species to adapt to adverse occasions, in addition to the fact that variations perceived as beneficial to the organism become common in populations, since they can mean advantages over others. It was also possible to observe the occurrence of regions called parsimony informational or "parsim-info" regions, which consist of regions where nucleotides replaced in common are present in at least two species, which may indicate characteristics acquired in common and maintained throughout evolutionary history because they are important for the function of the protein and thus are important to group them into a phylogenetic tree. About 84 singleton regions or SNPs have been identified and are characterized by manifesting the substitution of a single nucleotide in an individual or in a small population. In coding regions, SNPs are subclassified into synonyms when the polymorphism does not result in the alteration of the encoded amino acid, while non-synonymous ones cause this alteration.

The function of a likelihood is constructed based on the observed data (DNA sequences) and measures the probability of observing the data under the model with the estimated parameters. The estimated values of the matrix parameters indicate the relative rates of substitution between the bases and calculates an estimated average percentage of the occurrences of substitutions between them considering the alignment between the five species. For example, a high value for the transition rate from A to G suggests that this substitution occurs frequently. The results shown in Table 1 indicate the nucleotide frequencies where A = 28.27%, T/U = 21.34%, C = 24.95% and G = 25.44%.

Table 1. Estimation of maximum likelihood of the substitution matrix and total nucleotide frequencies (A), (T/U), (C) and (G).

	The	T	C	G	TOTAL
The	-	4,01	4,69	20,05	28,27%

T	5,31	-	10,83	4,78	21,34%
C	5,31	9,26	-	4,78	24,95%
G	22,28	4,01	4,69	-	25,44%

Source: MEGA 11 / survey data.

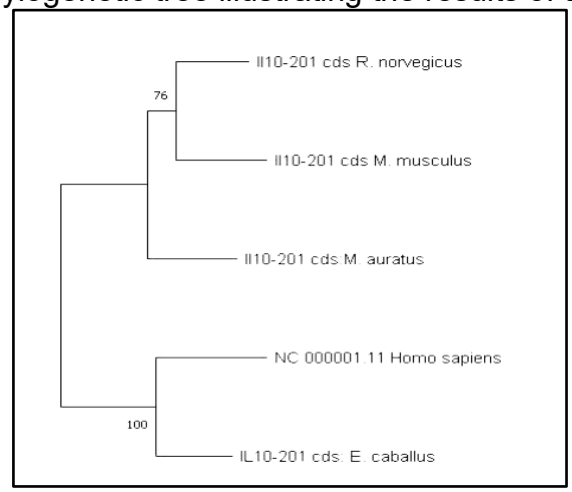
Figure 1 shows the values of the selective signatures where ω = dN/dS in a graph that represents the ratio between the nonsynonymous mutations and the synonymous mutations between the genes submitted to the alignment. The values found for each species were: H.sapiens/M.musculus (-0.2265), H. sapiens/R.norvegicus (-0.1938), H.sapiens/M.auratus (-0.1814), H.sapiens/E.caballus (-0.0978), R.norvegicus/M.musculus (-0.1027), R.norvegicus/M.auratus (-0.1215), R.norvegicus/E.caballus (-0.1907), M.auratus/M.musculus (-0.1191), M.auratus/E.caballus (-0.2546), E.caballus/M.musculus (-0.2431). These data suggest that the IL-10 gene is under pressure from negative or purifying selection, since all the values found were less than one (<1) and negative selection will act in the opposite direction, reducing the frequency or even eliminating deleterious mutations from populations over time (Hartl; Clark, 1997).

Fig 1. Calculation of the dN/dS ratio between the sequences analyzed.

	H. sapiens	M. musculus	R. norvegicus	M. auratus	E. caballus
H. sapiens					
M. musculus	-0,2265				
R. norvegicus	-0,1938	-0,1027			
M. auratus	-0,1814	-0,1191	-0,1215		
E. caballus	-0,0978	-0,2431	-0,1907	-0,2546	

For the construction of the phylogenetic tree (Figure 2), the number of 1000 "Bootstrap" replications was used as a parameter that gathered as a "sister group" with 76% reliability, suggesting a greater diversity in the sequence of the gene studied among the group of rodents R. norvegicus, M. musculus and M. auratus. While the genes of H.sapiens and E.caballus continued to be highly conserved ontogenetically, even after the evolutionary divergence between both groups, since more similarities were found in the genes since the software gathered with 100% reliability in the replications performed.

Figure 2 - Phylogenetic tree illustrating the results of the alignment.



In the Ensemble platform, it was possible to identify the existence of 139 SNP variants in the IL 10-206 (ENST00000659642.2) transcript of the interleukin 10 gene. For this study, three non-synonymous variants (Table 2) were selected to be analyzed in the software to verify the damage that possible mutations could cause to the structure and function of the protein. The variants were submitted to bioinformatics tools to determine whether they were neutral or deleterious according to the score indicated by the software indicated in Table 2. According to Nykamp et al. (2017), a variant is pathogenic if it disrupts a gene product in a way that leads to human disease and is benign if it has an effect that does not lead to disease in even though a single base can alter an amino acid and cause a pleiotropic effect. In the Mutation assessor, Polyphen 2 and SIFT software, the variants rs750010814, rs1674874871 and rs568879359 presented scores that indicate a significant possibility of risk during the formation of proteins that can result in changes in the structure and function of the protein and consequently result in diseases (Table 2).

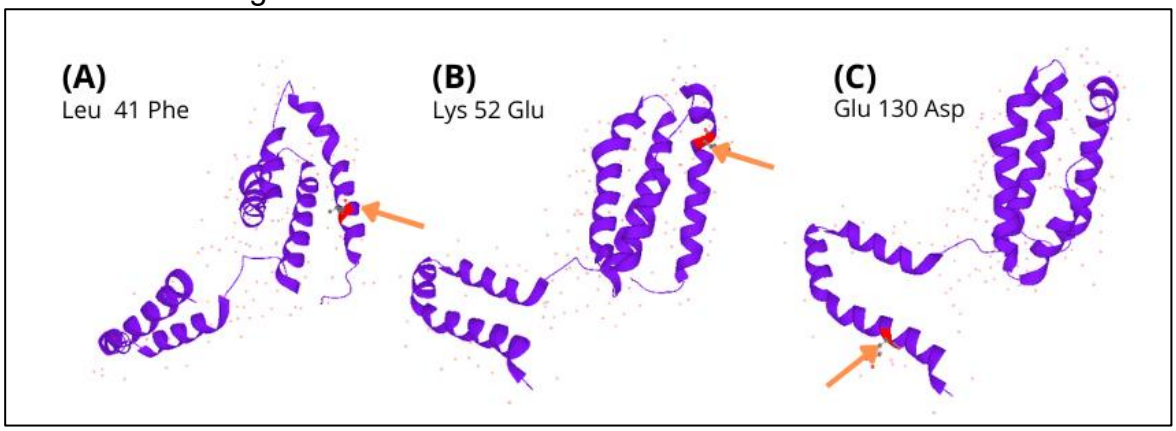
Table 2 - Variants of the human IL 10 gene for comparison and in silicon analysis for pathogenicity.

rs ID	Exodus	Subs. Nt	Sub.Aa	Allele	MAF	MutationAssessor	Polyphen2	SCORE	SIFT	SCORE
rs750010814	2	G>A	L41E	G	<0.01	deleterious	0.919	0.998	deleterious	0.01
rs1674874871	2	T>C	K52E	T	<0.01	deleterious	0.919	0.995	deleterious	0
rs568879359	6	A>C	E130D	T/G	<0.01	deleterious	0.919	0.989	deleterious	0.03

Source: survey data.

We can also observe a 3D model (Figure 3) with 85% of the protein structure that indicates exactly where the amino acid substitution occurred in the respective variants. Studies from recent years show that the C allele of rs568879359 was present in an analysis that associated it with patients with primary immunodeficiency disorders (Chi et al., 2018). Other studies have associated IL 10 with its use in its integral form as gene therapy in the treatment of rheumatoid arthritis. (Keystone et al., 1998).

Figure 3 - 3D model indicating the location where the amino acid exchange occurred in each variant according to Table 2.



FINAL CONSIDERATIONS

The in silicon analysis in this study suggested that the variants rs750010814, rs1674874871 and rs568879359 of the IL10 gene, point to an association with a deleterious potential due to structural and functional changes generated in the protein and can be used as risk markers for autoimmune diseases. Regarding the phylogenetic analysis for the IL 10 gene in the species studied, two clades in the phylogenetic tree showed a relationship between the species, with a reliability test of 76% for R.norvegicus, M.musculus and M.auratus and 100% for H.sapiens and E.caballus, this relationship being highly conserved between the species, suggesting that it evolved from a common ancestor and has been preserved as a result of its adaptive value, being a gene conserved even after divergent evolution among the species studied Therefore, we can see that molecular phylogeny as a branch of phylogeny is a crucial tool to obtain information about the evolutionary relationships between organisms. Therefore, it is possible that the gene for 'IL 10' has the same selective value for the immune systems of both horses and humans, allowing both to defend against pathogens (viruses, fungi, parasites, bacteria, etc.) in a similar way, both as an innate immune response and an adaptive immune response, according to their evolutionary patterns.

REFERENCES

1. Borges, P. (2021). *Comparação de ferramentas in silico para avaliação de patogenicidade de variantes missense* (Trabalho de Conclusão de Curso, Graduação em Ciências Biológicas). Universidade Federal do Rio Grande do Sul, Porto Alegre. Disponível em: https://lume.ufrgs.br/handle/10183/250399.

2. Caldart, E. T., Mata, H., Canal, C. W., & Ravazzolo, A. P. (2016). Análise filogenética: conceitos básicos e suas utilizações como ferramenta para virologia e epidemiologia molecular. *Acta Scientiae Veterinariae*, 44, 1392.

3. Chi, Z. H., et al. (2018). Técnica de sequenciamento de alto rendimento direcionada para o diagnóstico molecular de distúrbios de imunodeficiência primária. *Medicina*, 97(40), e12695. https://doi.org/10.1097/md.0000000000012695.

4. Costa, M. F. (2022). *Análise dos métodos in silico frente à estudo in vivo: uma revisão integrativa* (Trabalho de Conclusão de Curso, Graduação em Ciências Biológicas). Universidade Federal do Rio de Janeiro, Macaé. Disponível em: https://pantheon.ufrj.br/bitstream/11422/20649/1/MFCosta.pdf.

5. Cruzito. Regiões não codificantes de DNA: sequências e explicação. Disponível em: https://pt.estudyando.com/regioes-nao-codificantes-de-dna-sequencias-e-explicacao/.

6. De Waal Malefyt, R., et al. (1991). Interleukin 10 (IL-10) inhibits cytokine synthesis by human monocytes: an autoregulatory role of IL-10 produced by monocytes. *Journal of Experimental Medicine*, 174(5), 1209–1220.

7. Hartl, D., & Clark, A. G. (1997). *Principles of Population Genetics* (3rd ed.). Sinauer Associates.

8. Keystone, E., Wherry, J., & Grint, P. (1998). IL-10 as a therapeutic strategy in the treatment of rheumatoid arthritis. *Rheum Dis Clin North Am*, 24(3), 629-639. https://doi.org/10.1016/s0889-857x(05)70030-2.

9. Kosiol, C., & Anisimova, M. (2019). Selection Acting on Genomes. In *Methods in Molecular Biology* (pp. 373–397).

10. *Maximum Likelihood Method - an overview* | ScienceDirect Topics. Disponível em: https://www.sciencedirect.com/topics/medicine-and-dentistry/maximum-likelihood-method. Acesso em: 18 de Março de 2024.

11. Mayr, E. (1982). Speciation and macroevolution. *Evolution*, 36(6), 1119–1132.

12. Nei, M., & Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press.

13. Nykamp, K., et al. (2017). Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria. *Genetics in Medicine*, 19(10), 1105–1117.

14. Santos, C. M. D., & Klassa, B. (2012). Sistemática filogenética hennigiana: revolução ou mudança no interior de um paradigma? *Scientiae Studia*, 10, 593–612.

15. Saraiva, M., Vieira, P., & O'Garra, A. (2020). Biology and therapeutic potential of interleukin-10. *Journal of Experimental Medicine*, 217(1), e20190418. https://doi.org/10.1084/jem.20190418.

16. Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msab120.

17. Tamura, K., Nei, M., & Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences (USA)*, 101, 11030-11035.

18. Yao, Y., Simard, A. R., Shi, F. D., & Hao, J. (2013). IL-10-producing lymphocytes in inflammatory disease. *International Review of Immunology*, 32(3), 324-336. https://doi.org/10.3109/08830185.2012.762361.

19. Yin, Z., Bahtiyar, G., Zhang, N., Liu, L., Zhu, P., Robert, M. E., McNiff, J., Madaio, M. P., & Craft, J. (2002). IL-10 regulates murine lupus. *Journal of Immunology*, 169(4), 2148-2155. https://doi.org/10.4049/jimmunol.169.4.2148.