genes

MDPI

Article

In Silico Functional and Structural Analysis of *STAT4* Variants of Uncertain Significance

Karla Mayela Bravo-Villagra [1,2], Eric Jonathan Maciel-Cruz [3], Rosa Michel Martínez-Contreras [2,3], Itzae Adonai Gutiérrez-Hurtado [2,4], Alexis Missael Vizcaíno-Quirarte [5,6], José Francisco Muñoz-Valle [5,*] and Andres López-Quintero [1,2,*]

1 Centro Universitario de Ciencias de la Salud, Instituto de Nutrigenética y Nutrigenómica Traslacional, Universidad de Guadalajara, Guadalajara 44340, Jalisco, Mexico; karla.bravo2318@alumnos.udg.mx
2 Programa de Doctorado en Genética Humana, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara 44340, Jalisco, Mexico
3 Centro de Investigación Biomédica de Occidente, Instituto Mexicano del Seguro Social, Guadalajara 44340, Jalisco, Mexico
4 Departamento de Biología Molecular y Genómica, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara 44340, Jalisco, Mexico
5 Centro Universitario de Ciencias de la Salud, Instituto de Investigación en Ciencias Biomédicas, Universidad de Guadalajara, Guadalajara 44340, Jalisco, Mexico
6 Programa de Doctorado en Psicología de la Salud, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara 44340, Jalisco, Mexico
* Correspondence: drjosefranciscomv@cucs.udg.mx (J.F.M.-V.); andres.lopezq@academicos.udg.mx (A.L.-Q.); Tel.: +52-33-1058-5200 (ext. 34236) (A.L.-Q.)

Abstract

Background: The *STAT4* gene plays a key role in immune regulation and is associated with susceptibility to autoimmune diseases such as rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE). Objectives: The objective of this study is to analyze variants of uncertain significance (VUSs) in *STAT4* using bioinformatics tools to predict their functional and structural impact. Methods: A total of 48,295 variants of the *STAT4* gene (ENSG00000138378) were retrieved from the Ensembl database. A tiered filtering approach was used to assess VUS pathogenicity, integrating in silico prediction tools such as SIFT, PolyPhen, MutPred2, and Align-GVGD, as well as structural modeling platforms including Chimera, ModRefiner, Missense3D, HOPE, and DynaMut2. Results: Eighty missense VUSs were identified; of these, 13 were prioritized based on concordant signals across multiple computational predictors. These variants showed significant alterations in the physicochemical properties of the protein, including changes in hydrophobicity and disruption of hydrogen bonding. Notably, the rs140675301 (Glu128Val) variant lies within a conserved loop, and in silico analyses suggest that this mutation may alter kinase specificity regarding the phosphorylation of serine 130. Conclusions: The integrative use of the bioinformatic tools employed represents a valuable preliminary step prior to undertaking more complex and resource-intensive functional studies. This complementary strategy strengthens the interpretative framework for VUS, guiding subsequent experimental validation and supporting a structured assessment of variant relevance, particularly in the context of immune-related genes such as *STAT4*.

Keywords: *STAT4* gene; variants of uncertain significance (VUS); in silico analysis; bioinformatics tools; pathogenicity prediction

Check for updates

Academic Editor: Hongyan Xu

Received: 16 December 2025
Revised: 31 December 2025
Accepted: 5 January 2026
Published: 7 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Autoimmune diseases are complex and multifactorial disorders characterized by an inappropriate immune response against the body's own components. Genomic studies have identified multiple genetic susceptibility loci, among which genes from the major histocompatibility complex (*HLA*) stand out due to their central role in antigen presentation [1,2]. However, several non-*HLA* genes involved in immune regulation have also been identified. Among these, *STAT4* has been widely associated with increased susceptibility to various autoimmune diseases, highlighting its relevance as a genetic factor of interest in the study of these conditions [3–7].

The *STAT4* gene, located on chromosome 2 at cytogenetic band *2q32.2*, encodes the signal transducer and activator of transcription 4 (*STAT4*) protein [8]. This gene comprises 24 exons that encode various functional domains of the protein, including the DNA-binding and transactivation domains [9,10].

As a member of the STAT family, STAT4 plays a pivotal role in immune system regulation and has been implicated in a range of autoimmune and inflammatory diseases [11,12]. STAT4 activation occurs via the JAK-STAT signaling pathway; once phosphorylated, the protein translocates to the nucleus, where it regulates the expression of genes involved in immune responses [8,11,12].

Several single-nucleotide variants (SNVs) in *STAT4* have been associated with increased susceptibility to diseases such as RAand SLE. These variants have also been linked to higher disease activity in RA and elevated levels of anti-cyclic citrullinated peptide (anti-CCP) antibodies [5,13]. These findings underscore the importance of *STAT4* as a shared genetic risk factor across multiple immune-mediated diseases [3]. However, the interpretation of VUS in the *STAT4* gene remains a significant challenge in medical genetics, as such variants may influence disease risk and therapeutic responses [14]. Therefore, this study aimed to apply bioinformatics tools to predict the pathogenic potential of these variants using a systematic filtering and analysis approach.

2. Materials and Methods

2.1. In Silico Prediction of STAT4 Variants

The in silico prioritization of *STAT4* VUS was conducted using a tiered filtering approach, integrating several in silico platforms categorized by similar predictive properties. A variant was retained for further analysis only if it was flagged as potentially impactful by more than one tool.

2.2. Data Collection and Variant Filtering

Information on the *STAT4* gene was obtained from Ensembl (ENSG00000138378) (https://www.ensembl.org/info/index.html) accessed on 26 September 2025 [15,16]. The data were processed in Microsoft Excel, and variants were selected based on the criteria: "SNP," "missense," "uncertain significance," and "dbSNP source." A total of 80 VUSs were identified for computational analysis. The STAT4 canonical protein sequence (Q14765) was retrieved in FASTA format from UniProt (https://www.uniprot.org) accessed on 26 September 2025 [17].

2.3. Functional and Structural Predictions of Variants

In the initial screening phase, all 80 variants were analyzed to assess their potential functional impact, particularly with respect to structural and regulatory relevance. Predictive tools such as SIFT (https://sift.bii.a-star.edu.sg/www/SIFT_dbSNP.html) accessed on 13 October 2025 were employed to evaluate whether amino acid substitutions might impair protein function. SIFT assigns scores between 0 and 1; values below 0.05 indicate a likely

deleterious effect [18]. PolyPhen (http://genetics.bwh.harvard.edu/pph2/) accessed on 13 October 2025 [19] was used to assess physicochemical differences between wild-type and substituted residues, with scores above 0.5 suggesting probable functional disruption.

2.4. Evaluation of Protein Stability

Fifty-four variants underwent further evaluation using Align GVGD and MutPred2. Align GVGD (http://agvgd.hci.utah.edu/index.php) accessed on 14 October 2025, integrates evolutionary conservation and biochemical properties to classify variants on a scale from C0 to C65, where C15–C65 are suggestive of potential functional impact [20,21]. MutPred2 (http://mutpred2.mutdb.org/index.html) accessed on 15 October 2025 evaluates structural and functional impacts, providing scores from 0 to 1. Scores above 0.51 are indicative of potential structural or functional relevance [22].

2.5. Structural Protein Modeling

Structural models of the 54 filtered variants were generated using Chimera, ModRefiner, ERRAT, and PROCHECK. Chimera (https://www.cgl.ucsf.edu/chimera/) accessed on 16 October 2025, was used to visualize three-dimensional molecular changes [23]. ModRefiner (https://zhanggroup.org/ModRefiner/) accessed on 16 October 2025 [24] refined structural geometry. The quality of the models was further assessed by ERRAT and PROCHECK (https://saves.mbi.ucla.edu/) accessed on 16 October 2025, which evaluate stereochemical accuracy based on empirical data [25].

2.6. Structural Impact of Variants

Missense3D (http://missense3d.bc.ic.ac.uk/missense3d/) accessed on 17 October 2025 was used to predict the 3D structural consequences of missense variants [26]. In parallel, HOPE (https://www3.cmbi.umcn.nl/hope/) accessed on 18 October 2025 provided detailed information about physicochemical changes (e.g., size, charge, polarity) and their potential effects on protein structure and function [27]. A total of 54 variants were assessed with these tools.

2.7. Molecular Dynamics and Structural Stability

Based on predictions from HOPE and Missense3D, 13 variants were selected for molecular dynamics analysis using DynaMut (https://biosig.lab.uq.edu.au/dynamut/) accessed on 19 October 2025 [28]. Structural stability was evaluated using the ΔΔG scale: values < 0 indicate destabilization, values > 0 indicate stabilization, and values = 0 suggest no significant effect.

2.8. Statistical Analysis

Statistical analyses were performed using R and RStudio (version 4.4.1). The Shapiro–Wilk test was applied to assess the normality of the ΔΔG data. Data dispersion was visualized using boxplots, and correlations among variables were depicted using heatmaps. A p-value of <0.05 was considered statistically significant.

2.9. Workflow Diagram

A comprehensive workflow summarizing the bioinformatics tools and sequential steps used in the variant prediction pipeline is presented in Figure 1, detailing the platforms applied at each stage to assess the potential structural and functional impact of the selected *STAT4* variants.

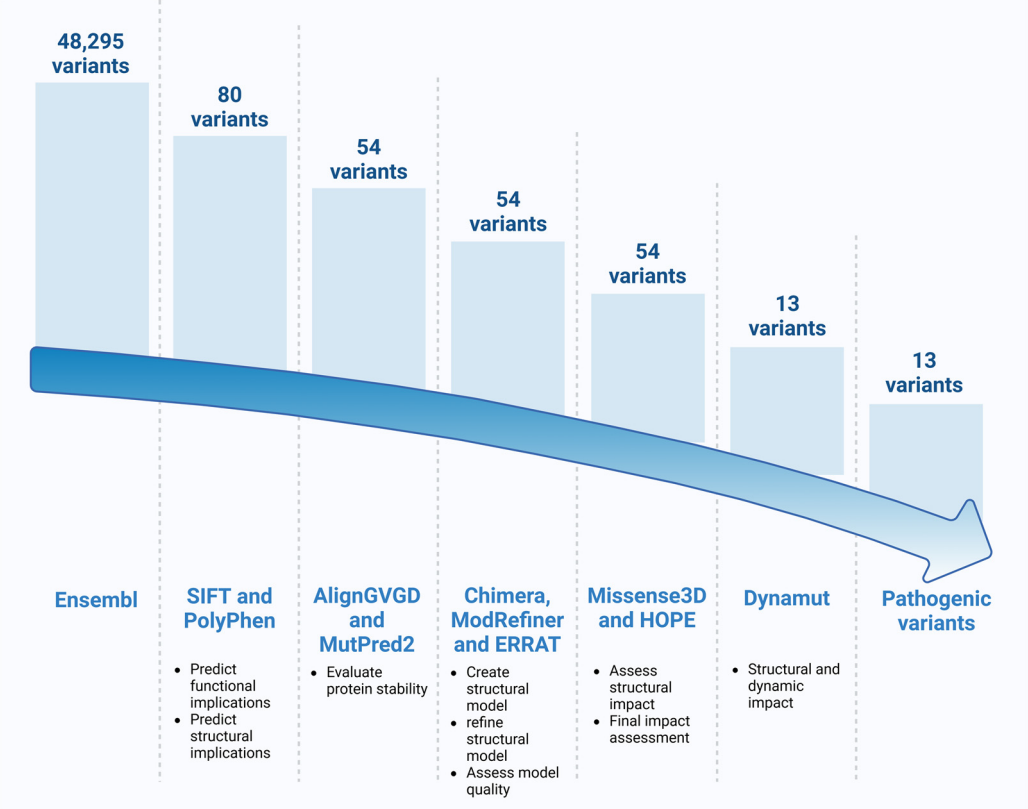


Figure 1. Diagram of Bioinformatics Tool Workflow for Variant Prediction. Eighty missense variants of uncertain significance were identified from the Ensembl database. Predictive tools included SIFT, PolyPhen2, MutPred2, and Align GVGD. Structural and molecular analyses were performed using Chimera, ModRefiner, Missense3D, HOPE and DynaMut2. Image created with BioRender.

3. Results

A total of 48,295 variants of the *STAT4* gene (ENSG00000138378) were retrieved from the Ensembl database. From these, 80 missense VUSs were selected for downstream analysis based on the criteria described. These variants were processed using a structured bioinformatics workflow, applying multiple in silico prediction tools grouped according to their methodological scope.

3.1. Functional and Structural Predictions of Variants

Of the 80 variants analyzed, SIFT predicted 34 variants (42.5%) as tolerated and 46 variants (57.5%) as deleterious, while PolyPhen2 predicted 42 variants (52.5%) as benign and 38 (47.5%) as possibly damaging. Concordance between both platforms identified 26 variants with consistent benign predictions, whereas 34 variants showed concordant damaging predictions. Notably, 20 variants exhibited discordant classifications, highlighting variability among functional predictors and supporting their use for variant prioritization rather than definitive inference.

3.2. Evaluation of Protein Stability

Following the initial screening, 54 variants were subjected to further analysis using Align GVGD and MutPred2. Align GVGD categorized 3 variants (5.55%) as benign and 51 (94.44%) showed non-benign grades, indicating potential functional relevance. MutPred2 predicted 23 variants (42.59%) as low-risk and 32 variants (59.26%) as high-risk based on its scoring framework. Integration of predictions across platforms revealed that 2 variants were flagged by a single tool, 14 variants by two tools, 12 variants by three tools, and 26 variants

were consistently flagged across all four platforms, supporting their prioritization for downstream structural analysis.

3.3. Evaluation of the Structural Impact of Variants

HOPE and Missense3D were applied to the complete set of 54 missense variants to evaluate their potential structural impact on the STAT4 protein. Based on a harmonized interpretation of the structural outputs, 41 variants (75.9%) showed no consistent evidence of structural damage, whereas 13 variants (24.1%) were prioritized as candidates due to the presence of physicochemical changes and/or structural signals reported by at least one of the predictors used.

Complementary analysis using HOPE provided detailed descriptions of the physicochemical changes associated with each variant, while Missense3D and DynaMut2 contributed structural damage indicators and protein stability estimates, respectively. Importantly, these results are intended for variant prioritization and in silico hypothesis generation, and not for inferring biological pathogenicity. A summary of these findings is presented in Table 1 and visualized in Figure 2.

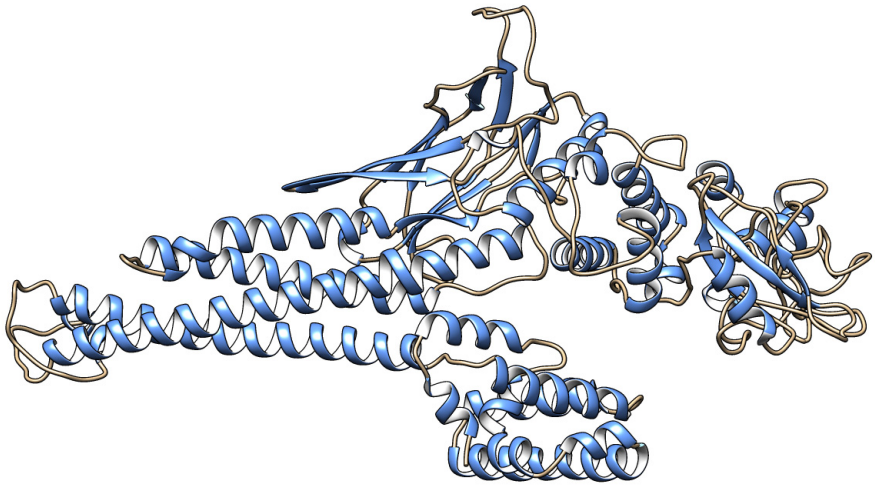


Figure 2. Three-dimensional representation of the STAT4 protein structure. The molecular structure of the STAT4 protein was modeled and visualized using Chimera. Structural alterations resulting from selected missense variants were assessed in relation to their spatial localization and potential impact on conserved domains.

3.4. Molecular Dynamics and Structural Stability Analysis

To further explore the potential structural consequences of amino acid substitutions, a molecular dynamics-based approach was applied using DynaMut2. Stability predictions were derived from changes in Gibbs free energy (ΔΔG), providing estimates of the relative impact of each substitution on protein stability.

Thirteen variants prioritized through the integrated in silico workflow were analyzed (Table 1). Most variants showed negative ΔΔG values, suggesting a tendency toward reduced structural stability (Figure 3A). Additionally, correlation analysis incorporating structural and physicochemical variables derived from HOPE and Missense3D revealed a statistically significant association between changes in amino acid size and ΔΔG values (Figure 3B), indicating that charge alterations may influence predicted stability effects.

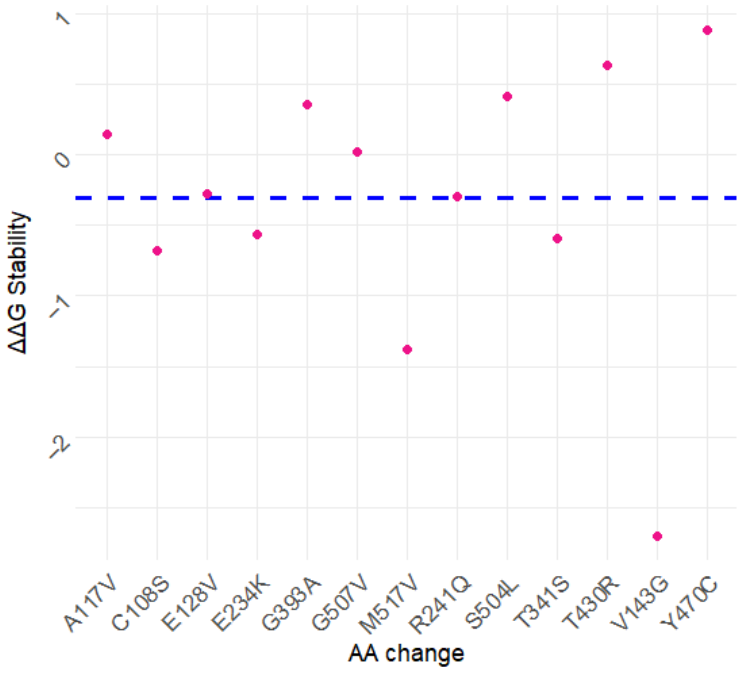
Table 1. Structural Prediction Analysis of 13 Missense Variants in *STAT4* using HOPE, Missense3D, and DynaMut2.

Variant	Allele Frequencies (gnomAD)	Amino Acid Change	ACMG	VarSome	Score MetaRNN	HOPE Main Prediction	Missense 3D Main Prediction	Dynamut2 Predicted Stability Change (ΔΔG)
rs755317297	T: 1.000 A: 6.841×10^{-7} C: 2.736×10^{-6}	M517V	VUS	VUS B	0.843	Smaller, mutant residue located near a highly conserved region. The mutant residue does not prefer α-helices as secondary structure	No structural damage detected	−1.379
rs1192576162	C: 0.999992036 A: 7.96400×10^{-6}	G507V	VUS	P VUS	0.77	Bigger, more hydrophobic, residue is located near a conserved region	Glycine in Bend Buried Glycine Replaced	0.021
rs1235014939	G: 0.999992036 A: 7.96400×10^{-6}	S504L	VUS	VUS B	0.796	Bigger, more hydrophobic, and residue is located a highly conserved region	Cavity altered	0.417
rs1380306157	T: 0.999992036 C: 7.96400×10^{-6}	Y470C	VUS	VUS	0.941	Smaller and more hydrophobic that will affect hydrogen bond formation	Cavity altered	0.879
rs758109437	G: 0.999952217 A: 3.98190×10^{-5} C: 7.96400×10^{-6}	T430R	VUS	VUS B	0.933	Bigger, charge positive, less hydrophobic and residue is located near a conserved region	Buried/exposed switch	0.637
rs199633613	C: 0.99993629 G: 6.37100×10^{-5}	G393A	VUS	VUS	0.548	Bigger, more hydrophobic and mutant residue reduced flexibility	Buried/exposed switch Buried Glycine replaced	0.358
rs746642521	T: 1.000 A: 8.894×10^{-6} C: 1.368×10^{-6}	T341S	VUS	VUS	0.917	Smaller and residue is located near a conserved region	No structural damage detected	−0.598
rs764990697	G: 0.999 T: 0.001	R241Q	VUS	B	0.889	Smaller, uncharged residue; potential disruption of a salt bridge	Buried charge replaced Buried salt bridge breakage	−0.301
rs758217844	C: 1.000 T: 4.723×10^{-5}	E234K	VUS	VUS	0.812	Bigger, positively charged; potential disruption of a salt bridge	Cavity altered Buried/exposed switch	−0.563

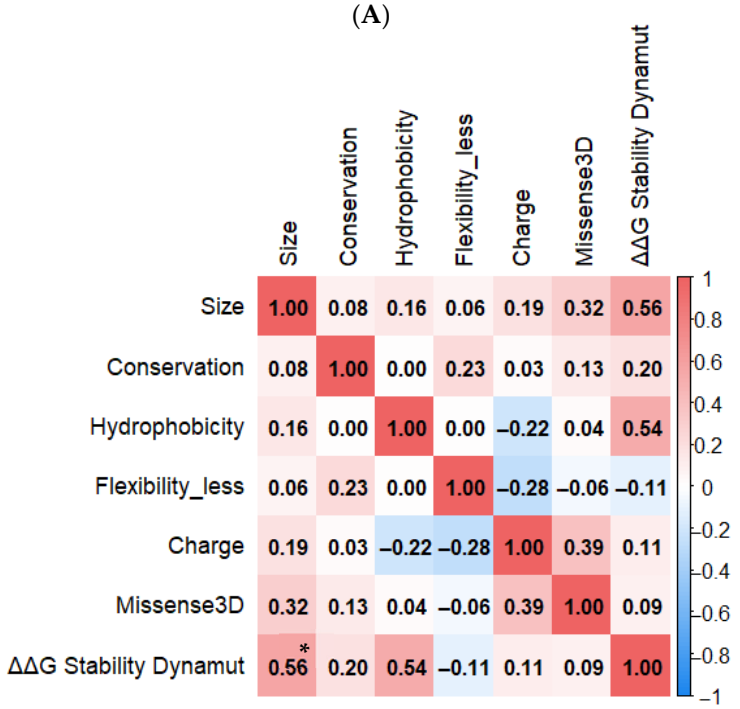
Table 1. *Cont.*

Variant	Allele Frequencies (gnomAD)	Amino Acid Change	ACMG	VarSome	Score MetaRNN	HOPE Main Prediction	Missense 3D Main Prediction	Dynamut2 Predicted Stability Change (ΔΔG)
rs866566754	A: 1.000 C: 4.105×10^{-6}	V143G	VUS	VUS	0.615	Smaller, less hydrophobic and the mutation will cause loss of hydrophobic interactions in the core of the protein	Cavity altered Buried/exposed switch	−2.698
rs140675301	T: 0.999 A: 0.001	E128V	B	VUS	0.031	Smaller, more hydrophobic, charge is neutral can cause loss of interactions with other molecules	No structural damage detected	−0.274
rs2125268711	Not reported	A117V	VUS	VUS B	0.453	Bigger and mutant residue located at the protein surface can disturb interactions with other molecules	No structural damage detected	0.142
rs1697316431	C: 1.000 G: 2.053×10^{-6}	C108S	VUS	VUS	0.887	More hydrophobic and residue is located near a highly conserved region	No structural damage detected	−0.676

ACMG (American College of Medical Genetics); P = Pathogenic; VUS = Variants of Uncertain Significance; B = Benign Abbreviations; ΔΔG = Gibbs free energy change; negative values indicate reduced structural stability. MetaRNN score (0–1), reflecting pathogenicity probability. Complementary to HOPE structural analysis.



(**A**)



(**B**)

Figure 3. Structural Impact and Correlation Analysis of STAT4 Missense Variants. (**A**) Scatter plot showing the predicted impact of amino acid substitutions on protein stability (ΔΔG values) as determined by DynaMut2. Each point represents a missense variant, with the dashed line indicating the mean ΔΔG value, serving as a reference for assessing relative destabilization effects. (**B**). Heatmap illustrating the correlation matrix between seven structural and physicochemical variables. Numerical values within each cell represent Spearman's correlation coefficients. Color intensity denotes the strength and direction of the correlation, with red indicating positive and blue negative associations. * A statistically significant positive correlation was observed between the amino acid size change and ΔΔG stability ($\rho = 0.56$, $p < 0.05$).

3.5. Validation of In Silico Predictive Tools

To strengthen the methodological framework of this study, a comparative analysis was performed using seven additional variants of the *STAT4* gene previously reported and classified in clinical databases as benign (B), likely benign (PB), possibly pathogenic (PP), or pathogenic (P). This validation allowed us to assess the concordance between the in silico predictions obtained through SIFT, PolyPhen-2, Align-GVGD, MutPred2, Missense3D, HOPE, and DynaMut, and the clinical classifications based on ACMG criteria (see Supplementary Materials Tables S1 and S2).

Among the seven variants analyzed, only rs35279173 maintained full concordance with its benign classification across all predictors. In the remaining six variants, discrepancies were observed between computational predictions and clinical classifications, highlighting the limitations of these tools in reproducing the true functional impact. For instance, rs2470644108 and rs2470644386, which are clinically classified as pathogenic, were estimated by DynaMut and Missense3D as stabilizing variants, whereas HOPE identified significant alterations in residue charge and in the surrounding domain environment. Similarly, variants classified as PB and PP (rs61756200, rs3024839, and rs2470644258) displayed negative $\Delta\Delta G$ values, suggesting a possible destabilizing effect on protein structure and showing relevant physicochemical changes in conserved regions.

4. Discussion

VUS represents a major challenge in genetic counseling and clinical decision-making due to the difficulty in interpreting their functional impact. In this context, in silico analysis of VUSs that lead to amino acid substitutions can offer valuable insights into potential structural and functional consequences, supporting efforts toward variant prioritization and hypothesis generation, rather than definitive reclassification [29].

Several studies have underscored the limitations in VUS interpretation. While in vitro experiments provide highly accurate functional characterization, they are resource-intensive and capable of evaluating only a limited number of variants [30,31]. In silico methodologies have therefore emerged as a complementary approach, offering predictive assessments often based on evolutionary conservation and biophysical modeling [29,30,32]. Recent advances in artificial intelligence have further accelerated variant interpretation, with structure-based models showing promising performance [31,33]. However, these predictors should be used with caution. Several studies have documented frequent discrepancies between the results obtained through computational predictions and the data obtained through in vitro validations, highlighting the importance of integrating multiple lines of evidence [31].

Tools such as SIFT and PolyPhen-2 employ distinct algorithms: SIFT uses nonsynonymous single-nucleotide variants to calculate the probability of each substitution (SIFT score) [34], while PolyPhen-2 relies on eight sequence-based and three structural features as inputs for a Bayesian classifier. These methodological differences may limit their clinical applicability in classifying VUS as pathogenic [18,35]. On the other hand, Align-GVGD uses logistic regression divided into four categories and is based exclusively on physicochemical changes resulting from amino acid substitutions relative to observed variability, providing a certain robustness to the evidence it generates [36]. MutPred2 is a machine learning model trained on unlabeled data that estimates probabilities, allowing the modeling of the impact of variants on protein structure and function while assigning potential molecular effects. It has proven useful for predicting de novo mutations that are more frequently pathogenic in cases than in controls [22].

Similarly, the HOPE, Missense3D, and DynaMut tools are valuable for predicting the structural and functional impact of variants. HOPE (web-based server, accessed on 18

October 2025) employs BLAST-formatted sequences, followed by automated homology modeling, and finally analyzes the characteristics of wild-type and mutant amino acids through web services [27]. Missense3D focuses on identifying three-dimensional structural changes in the protein, performing predictions through its web server [26], whereas DynaMut2, through a programming-based framework with defined input and output variables, estimates stability changes using molecular dynamics simulations to model the atomic behavior of the protein [28]. Given the variability of their underlying mechanisms—ranging from evolutionary conservation and physicochemical properties to structural modeling and dynamic simulation—each tool contributes a unique layer of evidence. Therefore, relying on a single predictive model is insufficient for robust variant classification. Instead, integrating multiple computational tools allows for a more comprehensive assessment of possible functional consequences.

In this study, 80 missense VUSs in the *STAT4* gene were identified and systematically evaluated. Thirteen of these (16.25%) were prioritized based on concordant signals across multiple predictors [37], reflecting physicochemical changes such as alterations in amino acid size, hydrophobicity, or charge, and their localization within or near conserved regions. Accumulating evidence suggests that surface hydrophobicity and residue charge play important roles in protein folding and structural integrity [38]. Disruptions in these properties may influence hydrogen bonding networks, local conformational stability, and protein dynamics [39]. Nevertheless, these observations should be interpreted as structural hypotheses rather than direct evidence of functional impairment.

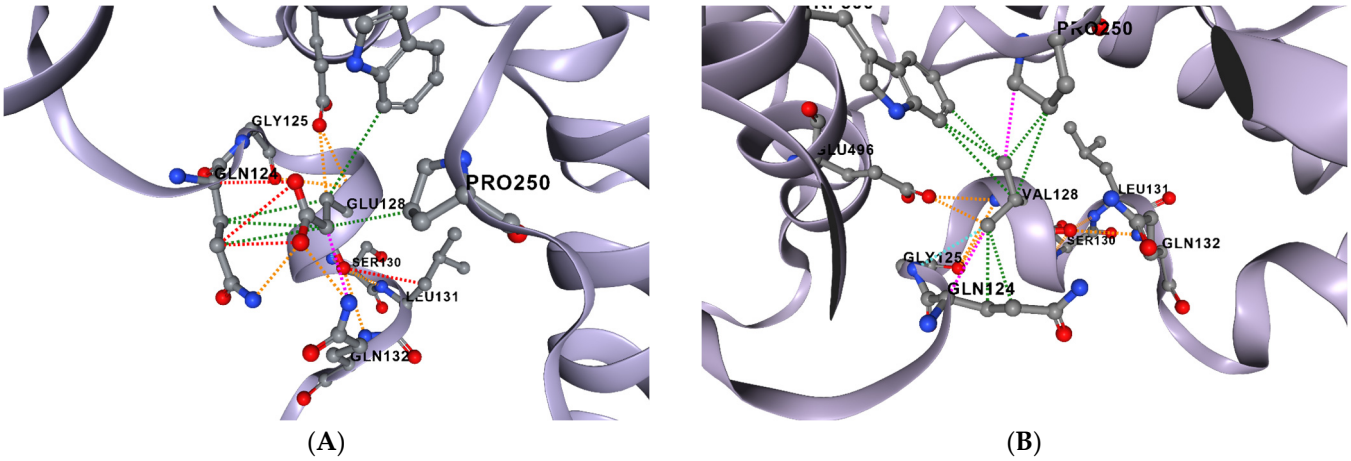
Despite these in silico predictions, the ACMG classification framework [40] categorized all 13 variants as either benign or of uncertain significance. This divergence may be attributed to the lack of supporting functional or association data, often due to the extremely low allele frequencies of these variants (global minor allele frequency < 0.01), which limits the statistical power of existing datasets to confirm pathogenicity. Population frequency data were considered at a global level, and no population-stratified analyses were performed; therefore, population-specific inferences were beyond the scope of this study.

Among the 13 analyzed variants, rs140675301 (Glu128Val) illustrates how structural modeling can inform hypothesis generation. This substitution involves the replacement of a negatively charged glutamic acid with a hydrophobic valine, altering the charge, hydrophobicity, and size of the residue. Structural modeling revealed that the variant is located within a surface-exposed, conserved loop, in proximity to a putative phosphorylation site at Ser130. In silico predictions suggest that this substitution may influence kinase recognition or local phosphorylation dynamics, highlighting a potential regulatory mechanism that warrants further functional investigation, rather than confirming a pathogenic role [41,42]. These predicted features are illustrated in Figure 4.

Analysis of molecular dynamics outputs from DynaMut2, integrated with structural descriptors derived from HOPE and Missense3D, further emphasized the contribution of amino acid size and hydrophobicity changes to predicted stability effects [43]. However, these predictions remain probabilistic and context-dependent, underscoring the limitations of computational tools in capturing complex three-dimensional and interaction-specific effects.

To strengthen the methodology, we validated seven missense variants classified as B, LB, LP, and P. Only the variant rs35279173 maintained full concordance with its benign classification across all predictors. The impact of amino acid substitutions on protein function is multifactorial and depends on the specific location of the residue within the protein structure, its involvement in molecular interactions, and its role in regulatory mechanisms [41]. Modifications at critical positions may result in either loss or gain of function and could interfere with interactions involving other molecules, ultimately influencing regulation and stability of the protein [8,14]. These findings emphasize that

structural and functional predictors do not always capture three dimensional or interaction effects, which may result in falsely neutral predictions. Integrating multiple computational tools and balancing evolutionary and structural information is recommended for more accurate variant classification.



(**A**) (**B**)

Figure 4. Structural impact of the Glu128Val mutation in the STAT4 protein. (**A**) Wild-type structure highlighting the glutamic acid (Glu) at position 128. (**B**) Mutant structure displaying the valine (Val) substitution at the same position. The mutation results in changes in charge, size, and hydrophobicity of the residue, potentially affecting local structural integrity and phosphorylation dynamics. Protein modeling and visualization were performed using DynaMut2.

A major limitation of this study is the absence of experimental validation or association analyses to directly assess biological impact. While computational approaches offer valuable insights, their findings must be validated through in vitro or in vivo experiments to confirm their biological significance. Nevertheless, the present work offers a reproducible methodological framework for variant prioritization, identifying candidates for downstream functional studies and contributing to a more structured interpretation of VUS in *STAT4*.

5. Conclusions

In this study, an integrative in silico variant prioritization framework was applied to missense variants of the *STAT4* gene, allowing the identification of thirteen variants prioritized based on predicted structural and physicochemical features. The potential impact of these variants is influenced by their location within specific protein domains and by the nature of the amino acid substitutions involved. Variants such as p.E128V, located near a known phosphorylation site, illustrate how structural modeling can generate testable hypotheses regarding possible effects on protein regulation. However, these observations should not be interpreted as evidence of biological pathogenicity. Overall, this work highlights the value of combining multiple computational tools to support the interpretation and prioritization of VUS. While functional validation remains essential, the proposed framework provides a reproducible methodological approach that can be extended to other genes and datasets in future studies. In this context, the proposed workflow may support genetic counseling and translational research by providing a structured approach to prioritize VUS for targeted functional testing and follow-up, without replacing clinical or experimental evaluation.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes17010072/s1, Figure S1: Comparison of ionic interactions between wild-type and mutant residues for the 13 analyzed variants. Figure S2: Comparison of ionic interactions between wild-type and mutant residues for the 7 analyzed variants of validation. Table S1: Comparative analysis of clinical and functional prediction tools for the seven validated STAT4 variants. Table S2: Comparative structural and stability assessment of the seven STAT4 variants used for validation.

Author Contributions: Conceptualization, E.J.M.-C. and K.M.B.-V.; formal analysis, K.M.B.-V. and R.M.M.-C.; data curation, E.J.M.-C. and K.M.B.-V.; methodology, K.M.B.-V., E.J.M.-C. and R.M.M.-C.; validation, A.L.-Q., I.A.G.-H. and J.F.M.-V.; investigation, K.M.B.-V. and R.M.M.-C.; resources, A.L.-Q., I.A.G.-H. and A.M.V.-Q.; writing—original draft preparation, K.M.B.-V.; writing—review and editing, A.M.V.-Q., A.L.-Q. and I.A.G.-H.; visualization, K.M.B.-V., R.M.M.-C. and A.M.V.-Q.; supervision, J.F.M.-V. and A.L.-Q.; project administration, E.J.M.-C. and A.L.-Q.; funding acquisition, A.L.-Q. and J.F.M.-V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article and Supplementary Materials. Further inquiries can be directed towards the corresponding authors.

Acknowledgments: This project was supported by the National Council of Humanities, Science and Technology (Sistema Nacional de Posgrados, Consejo Nacional de Humanidades, Ciencias y Tecnologías) through doctoral scholarships awarded to K.M.B.-V. [CVU 1098186] and R.M.M-C. [CVU 1097097].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Balsa, A.; Cabezón, A.; Orozco, G.; Cobo, T.; Miranda-Carus, E.; López-Nevot, M.Á.; Vicario, J.L.; Martín-Mola, E.; Martín, J.; Pascual-Salcedo, D. Influence of HLA DRB1 Alleles in the Susceptibility of Rheumatoid Arthritis and the Regulation of Antibodies against Citrullinated Proteins and Rheumatoid Factor. *Arthritis Res. Ther.* **2010**, *12*, R62. [CrossRef]
2. Das, S.; Baruah, C.; Saikia, A.K.; Bose, S. Associative Role of HLA-DRB1 SNP Genotypes as Risk Factors for Susceptibility and Severity of Rheumatoid Arthritis: A North-East Indian Population-Based Study. *Int. J. Immunogenet.* **2018**, *45*, 1–7. [CrossRef]
3. Lamana, A.; López-Santalla, M.; Castillo-González, R.; Ortiz, A.M.; Martín, J.; García-Vicuña, R.; González-Álvaro, I. The Minor Allele of Rs7574865 in the STAT4 Gene Is Associated with Increased mRNA and Protein Expression. *PLoS ONE* **2015**, *10*, e0142683. [CrossRef]
4. Hagberg, N.; Joelsson, M.; Leonard, D.; Reid, S.; Eloranta, M.L.; Mo, J.; Nilsson, M.K.; Syvänen, A.C.; Bryceson, Y.T.; Rönnblom, L. The STAT4 SLE Risk Allele Rs7574865[T] Is Associated with Increased IL-12-Induced IFN-γ Production in t Cells from Patients with SLE. *Ann. Rheum. Dis.* **2018**, *77*, 1070–1077. [CrossRef]
5. de Durán-Avelar, M.J.; Vibanco-Pérez, N.; Hernández-Pacheco, R.R.; Castro-Zambrano, A.d.C.; Ortiz-Martínez, L.; Zambrano-Zaragoza, J.F. STAT4 Rs7574865 G/T Polymorphism Is Associated with Rheumatoid Arthritis and Disease Activity, but Not with Anti-CCP Antibody Levels in a Mexican Population. *Clin. Rheumatol.* **2016**, *35*, 2909–2914. [CrossRef]
6. Bravo-Villagra, K.M.; Muñoz-Valle, J.F.; Baños-Hernández, C.J.; Cerpa-Cruz, S.; Navarro-Zarza, J.E.; Parra-Rojas, I.; Aguilar-Velázquez, J.A.; García-Arellano, S.; López-Quintero, A. STAT4 Gene Variant Rs7574865 Is Associated with Rheumatoid Arthritis Activity and Anti-CCP Levels in the Western but Not in the Southern Population of Mexico. *Genes* **2024**, *15*, 241. [CrossRef] [PubMed]
7. Esparza Guerrero, Y.; Vazquez Villegas, M.L.; Nava Valdivia, C.A.; Ponce Guarneros, J.M.; Perez Guerrero, E.E.; Gomez Ramirez, E.E.; Ramirez Villafaña, M.; Contreras Haro, B.; Martinez Hernandez, A.; Cardona Muñoz, E.G.; et al. Association of the STAT4 Gene Rs7574865 Polymorphism with IFN-γ Levels in Patients with Systemic Lupus Erythematosus. *Genes* **2023**, *14*, 537. [CrossRef] [PubMed]
8. Yang, C.; Mai, H.; Peng, J.; Zhou, B.; Hou, J.; Jiang, D. STAT4: An Immunoregulator Contributing to Diverse Human Diseases. *Int. J. Biol. Sci.* **2020**, *16*, 1575–1585. [CrossRef]

9. Berg, A.; Gräber, M.; Schmutzler, S.; Hoffmann, R.; Berg, T. A High-Throughput Fluorescence Polarization-Based Assay for the SH2 Domain of STAT4. *Methods Protoc.* **2022**, *5*, 93. [CrossRef] [PubMed]
10. Hu, X.; Li, J.; Fu, M.; Zhao, X.; Wang, W. The JAK/STAT Signaling Pathway: From Bench to Clinic. *Signal Transduct. Target. Ther.* **2021**, *6*, 402. [CrossRef]
11. Malemud, C.J. The Role of the JAK/STAT Signal Pathway in Rheumatoid Arthritis. *Ther. Adv. Musculoskelet. Dis.* **2018**, *10*, 117–127. [CrossRef]
12. Salaffi, F.; Giacobazzi, G.; Carlo, M.D. Pain and the JAK-STAT Pathwsy. *Pain Res. Manag.* **2018**, *2018*, 15. [CrossRef]
13. Beltrán Ramírez, O.; Mendoza Rincón, J.F.; Barbosa Cobos, R.E.; Alemán Ávila, I.; Ramírez Bello, J. STAT4 Confers Risk for Rheumatoid Arthritis and Systemic Lupus Erythematosus in Mexican Patients. *Immunol. Lett.* **2016**, *175*, 40–43. [CrossRef] [PubMed]
14. Shree, M.; Vaishnav, J.; Gurudayal; Ampapathi, R.S. In-Silico Assessment of Novel Peptidomimetics Inhibitor Targeting STAT3 and STAT4 N-Terminal Domain Dimerization: A Comprehensive Study Using Molecular Docking, Molecular Dynamics Simulation, and Binding Free Energy Analysis. *Biochem. Biophys. Res. Commun.* **2024**, *733*, 150584. [CrossRef]
15. Hunt, S.E.; Moore, B.; Amode, R.M.; Armean, I.M.; Lemos, D.; Mushtaq, A.; Parton, A.; Schuilenburg, H.; Szpak, M.; Thormann, A.; et al. Annotating and Prioritizing Genomic Variants Using the Ensembl Variant Effect Predictor-A Tutorial. *Hum. Mutat.* **2022**, *43*, 986–997. [CrossRef]
16. Hunt, S.E.; McLaren, W.; Gil, L.; Thormann, A.; Schuilenburg, H.; Sheppard, D.; Parton, A.; Armean, I.M.; Trevanion, S.J.; Flicek, P.; et al. Ensembl Variation Resources. *Database* **2018**, *2018*, bay119. [CrossRef] [PubMed]
17. The UniProt Consortium. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef]
18. Chen, J.; Bataillon, T.; Glémin, S.; Lascoux, M. Hunting for Beneficial Mutations: Conditioning on SIFT Scores When Estimating the Distribution of Fitness Effect of New Mutations. *Genome Biol. Evol.* **2022**, *14*, evab151. [CrossRef] [PubMed]
19. Hicks, S.; Wheeler, D.A.; Plon, S.E.; Kimmel, M. Prediction of Missense Mutation Functionality Depends on Both the Algorithm and Sequence Alignment Employed. *Hum. Mutat.* **2011**, *32*, 661–668. [CrossRef]
20. Mathe, E.; Olivier, M.; Kato, S.; Ishioka, C.; Hainaut, P.; Tavtigian, S.V. Computational Approaches for Predicting the Biological Effect of P53 Missense Mutations: A Comparison of Three Sequence Analysis Based Methods. *Nucleic Acids Res.* **2006**, *34*, 1317–1325. [CrossRef]
21. Tavtigian, S.V.; Deffenbaugh, A.M.; Yin, L.; Judkins, T.; Scholl, T.; Samollow, P.B.; de Silva, D.; Zharkikh, A.; Thomas, A. Comprehensive Statistical Study of 452 BRCA1 Missense Substitutions with Classification of Eight Recurrent Substitutions as Neutral. *J. Med. Genet.* **2006**, *43*, 295–305. [CrossRef] [PubMed]
22. Pejaver, V.; Urresti, J.; Lugo-Martinez, J.; Pagel, K.A.; Lin, G.N.; Nam, H.-J.; Mort, M.; Cooper, D.N.; Sebat, J.; Iakoucheva, L.M.; et al. Inferring the Molecular and Phenotypic Impact of Amino Acid Variants with MutPred2. *Nat. Commun.* **2020**, *11*, 5918. [CrossRef]
23. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef] [PubMed]
24. Xu, D.; Zhang, Y. Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-Step Atomic-Level Energy Minimization. *Biophys. J.* **2011**, *101*, 2525–2534. [CrossRef]
25. Altunkulah, E.; Ensari, Y. Protein structure prediction: An in-depth comparison of approaches and tools. *Eskişehir Tek. Üniversitesi Bilim Ve Teknol. Derg.-C Yaşam Bilim. Ve Biyoteknoloji* **2024**, *13*, 31–51. [CrossRef]
26. Ittisoponpisan, S.; Islam, S.A.; Khanna, T.; Alhuzimi, E.; David, A.; Sternberg, M.J.E. Can Predicted Protein 3D Structures Provide Reliable Insights into Whether Missense Variants Are Disease Associated? *J. Mol. Biol.* **2019**, *431*, 2197–2212. [CrossRef]
27. Venselaar, H.; te Beek, T.A.; Kuipers, R.K.; Hekkelman, M.L.; Vriend, G. Protein Structure Analysis of Mutations Causing Inheritable Diseases. An e-Science Approach with Life Scientist Friendly Interfaces. *BMC Bioinform.* **2010**, *11*, 548. [CrossRef]
28. Rodrigues, C.H.M.; Pires, D.E.V.; Ascher, D.B. DynaMut2: Assessing Changes in Stability and Flexibility upon Single and Multiple Point Missense Mutations. *Protein Sci.* **2021**, *30*, 60–69. [CrossRef]
29. Caballero-Avendaño, A.; Gutiérrez-Angulo, M.; Ayala-Madrigal, M.d.l.L.; Moreno-Ortiz, J.M.; González-Mercado, A.; Peregrina-Sandoval, J. In Silico Analysis of the Missense Variants of Uncertain Significance of CTNNB1 Gene Reported in GnomAD Database. *Genes* **2024**, *15*, 972. [CrossRef]
30. Katsonis, P.; Wilhelm, K.; Williams, A.; Lichtarge, O. Genome Interpretation Using in Silico Predictors of Variant Impact. *Hum. Genet.* **2022**, *141*, 1549–1577. [CrossRef]
31. Li, C.; Luo, Y.; Xie, Y.; Zhang, Z.; Liu, Y.; Zou, L.; Xiao, F. Structural and Functional Prediction, Evaluation, and Validation in the Post-Sequencing Era. *Comput. Struct. Biotechnol. J.* **2024**, *23*, 446–451. [CrossRef]
32. Cannon, S.; Williams, M.; Gunning, A.C.; Wright, C.F. Evaluation of in Silico Pathogenicity Prediction Tools for the Classification of Small In-Frame Indels. *BMC Med. Genom.* **2023**, *16*, 36. [CrossRef]
33. Kamal, E.; Kaddam, L.A.; Ahmed, M.; Alabdulkarim, A. Integrating Artificial Intelligence and Bioinformatics Methods to Identify Disruptive STAT1 Variants Impacting Protein Stability and Function. *Genes* **2025**, *16*, 303. [CrossRef]

34. Ernst, C.; Hahnen, E.; Engel, C.; Nothnagel, M.; Weber, J.; Schmutzler, R.K.; Hauke, J. Performance of in Silico Prediction Tools for the Classification of Rare BRCA1/2 Missense Variants in Clinical Diagnostics. *BMC Med. Genom.* **2018**, *11*, 35. [CrossRef]
35. Poon, K.-S. In Silico Analysis of BRCA1 and BRCA2 Missense Variants and the Relevance in Molecular Genetic Testing. *Sci. Rep.* **2021**, *11*, 11114. [CrossRef]
36. Young, C.C.; Feng, B.-J.; Mackenzie, C.B.; Girard, E.; Hu, D.; Iwasaki, Y.; Momozawa, Y.; Lesueur, F.; Ziv, E.; Neuhausen, S.L.; et al. Evaluation of ACMG Rules for In Silico Evidence Strength Using An Independent Computational Tool Absent of Circularities on ATM and CHEK2 Breast Cancer Cases and Controls. *bioRxiv* **2019**. [CrossRef]
37. MacielCruz, E.; FigueraVillanueva, L.; GómezFloresRamos, L.; HernándezPeña, R.; GallegosArreola, M. In-Silico Method for Predicting Pathogenic Missense Variants Using Online Tools: AURKA Gene A as a Model. *Iran. J. Biotechnol.* **2024**, *22*, e3787. [CrossRef]
38. Koop, J.; Merz, J.; Schembecker, G. Hydrophobicity, Amphilicity, and Flexibility: Relation between Molecular Protein Properties and the Macroscopic Effects of Surface Activity. *J. Biotechnol.* **2021**, *334*, 11–25. [CrossRef]
39. Kemp, M.T.; Lewandowski, E.M.; Chen, Y. Low Barrier Hydrogen Bonds in Protein Structure and Function. *Biochim. Biophys. Acta BBA Proteins Proteom.* **2021**, *1869*, 140557. [CrossRef]
40. Ghosh, R.; Oak, N.; Plon, S.E. Evaluation of in Silico Algorithms for Use with ACMG/AMP Clinical Variant Interpretation Guidelines. *Genome Biol.* **2017**, *18*, 225. [CrossRef] [PubMed]
41. Latini, A.; Borgiani, P.; De Benedittis, G.; Ciccacci, C.; Novelli, L.; Pepe, G.; Helmer-Citterich, M.; Baldini, I.; Perricone, C.; Ceccarelli, F.; et al. Large-Scale DNA Sequencing Identifies Rare Variants Associated with Systemic Lupus Erythematosus Susceptibility in Known Risk Genes. *Gene* **2024**, *907*, 148279. [CrossRef] [PubMed]
42. Saevarsdottir, S.; Stefansdottir, L.; Sulem, P.; Thorleifsson, G.; Ferkingstad, E.; Rutsdottir, G.; Glintborg, B.; Westerlind, H.; Grondal, G.; Loft, I.C.; et al. Multiomics Analysis of Rheumatoid Arthritis Yields Sequence Variants That Have Large Effects on Risk of the Seropositive Subset. *Ann. Rheum. Dis.* **2022**, *81*, 1085–1095. [CrossRef] [PubMed]
43. Peng, Y.; Alexov, E.; Basu, S. Structural Perspective on Revealing and Altering Molecular Functions of Genetic Variants Linked with Diseases. *Int. J. Mol. Sci.* **2019**, *20*, 548. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.