OXFORD UNIVERSITY PRESS | Bioinformatics Advances

Exploring SARS-CoV-2 Spike Protein Mutations through Genetic Algorithm-Driven Structural Modeling

Journal:	*Bioinformatics Advances*
Manuscript ID	BIOADV-2025-215.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	10-Sep-2025
Complete List of Authors:	Di Salvatore, Valentina; University of Catania, Department of Drug and Health Sciences Maleki, Avisa; University of Catania, Department of Drug and Health Sciences Mohajer, Babak; MIMESIS SRL Ras-Carmona, Alvaro; Facultad de Medicina, Universidad Complutense de Madrid, Inmunology Russo, Giulia; University of Catania, Department of Drug and Health Sciences Reche, Pedro; Facultad de Medicina, Universidad Complutense de Madrid, Immunology & O2 Pappalardo, Francesco; University of Catania, Department of Drug and Health Sciences
Portal Keywords:	Algorithms, Proteins, Genetics, COVID-19

Exploring SARS-CoV-2 Spike Protein Mutations through Genetic Algorithm-Driven Structural Modeling

Valentina Di Salvatore[1†], Avisa Maleki[1†], Babak Mohajer[2], Alvaro Ras-Carmona[3], Giulia Russo[1], Pedro Antonio Reche[3], and Francesco Pappalardo[1]*

[1]Department of Drug and Health Sciences, University of Catania, Viale Andrea Doria 6, Catania, 95125, Italy

[2]Mimesis SRL, Catania, Italy

[3]Department of Immunology, Faculty of Medicine, University Complutense of Madrid, Madrid, Spain

†These authors contributed equally to this work and share first authorship.

* Corresponding author, valentinadisalvatore@unict.it, Tel.: +39-095-7384040

Abstract

Background

The rapid evolution of SARS-CoV-2 highlights the importance of computational approaches to explore mutational effects on the viral spike protein. In this work, we present a Genetic Algorithm (GA) framework applied to the structural optimization of spike protein variants, with a focus on energetic and binding properties rather than direct evolutionary prediction.

Results

Our GA-driven pipeline generated spike variants with progressively improved structural stability as indicated by lower DOPE scores across generations. The approach also enabled evaluation of Gibbs Free Energy (ΔG) and binding affinity for spike – Angiotensin-converting enzyme 2 receptor (ACE2) interactions, revealing candidate conformations with favorable thermodynamic properties. These results demonstrate the algorithm's capacity to refine protein models and explore mutational landscapes in silico, although no validation against naturally emerging variants was performed.

Conclusions

This study presents a methodological framework for GA-based structural modeling of SARS-CoV-2 spike mutations. Rather than forecasting specific Variants of Concern, it demonstrates the feasibility

of a computational approach that can be extended and integrated with evolutionary and experimental evidence to strengthen future efforts in variant monitoring and vaccine development.

Keywords: SARS-CoV-2, Variant of Concern, Genetic Algorithm, Spike protein

Introduction

RNA viruses, including SARS-CoV-2, mutate frequently due to the intrinsic properties of their genomes [1]. The impact of these mutations on transmission and disease severity depends on mutation rates and their effects within, and between, hosts [2], [3], contributing to the emergence of viral variants and influencing pandemic dynamics.

SARS-CoV-2 relies on its spike (S) glycoprotein, a homotrimeric class I fusion protein, for host cell entry. Cleavage by host proteases like furin and TMPRSS2 triggers conformational changes that enable ACE2 binding and membrane fusion [4]. Given this mechanism, accurately predicting Variants of Concern (VOCs) remains essential in infectious disease control.

Variants such as Alpha (B.1.1.7), Beta (B.1.351), Delta (B.1.617.2), and Omicron (B.1.1.529) emerged through spike mutations, increasing transmissibility, altering disease severity, and challenging vaccine efficacy. Meanwhile, mRNA vaccines like Pfizer-BioNTech and Moderna showed strong protection against the original strain, with booster doses later introduced to counter immune evasion by emerging variants [4], [5].

Forecasting new variants enables timely public health responses [6], supported by early surveillance systems [7]. A comprehensive strategy integrating genomics, bioinformatics, phylogenetics, mutation tracking, epidemiology, and AI-facilitates real-time analysis of viral evolution [8]. In particular, understanding mutation-driven changes in spike–ACE2 binding and Gibbs free energy (ΔG) is key to anticipating variant behavior and biological impact [9].

Metaheuristic algorithms are widely used to solve complex problems in biology, economics, and engineering. [9], [10], [11].

Among these, Genetic Algorithms (GAs) are a prominent population-based method inspired by natural evolution.[10]. Proposed by J.H. Holland in 1992, it incorporates chromosome representation, fitness evaluation, and biologically inspired operators, including the Inversion operator [11]. Chromosomes, often binary-encoded, are modified via crossover and mutation, while selection favors the fittest individuals. Through iterative application, GA drives the population toward optimal solutions [12], [13].

One study proposed VOC-DL, a deep learning framework integrating variant data into models like VOC-LSTM, VOC-GRU, and VOC-BiLSTM to forecast daily COVID-19 cases. Using data from Italy, South Korea, Russia, Japan, and India over three months in 2021, all models performed well, with VOC-LSTM achieving the highest accuracy (average R^2 = 96.83%). The authors concluded that VOC-DL, particularly VOC-LSTM, offers reliable long-term forecasting and could aid in assessing the impact of future variants [14]. Another study predicted SARS-CoV-2 mutations using epidemiological, evolutionary, and immunological features. While effective at forecasting known mutations linked to ACE2 binding and immune escape, it could not anticipate novel mutations or full lineage evolution. Existing models,

including machine learning and logistic regression, often fail to account for the structural effects of mutations on protein–ligand interactions, key factors in viral transmission and immune evasion [9].

In this study, we propose a GA-based computational framework designed to explore the structural and energetic implications of spike protein mutations. Rather than attempting to directly predict real-world Variants of Concern (VOCs), our goal is to demonstrate the feasibility of applying GA to optimize spike protein conformations and investigate their impact on stability and spike–ACE2 interactions. This methodological approach lays the groundwork for future extensions where genomic surveillance and evolutionary data could be integrated to assess predictive performance.

Method

The prediction of VOCs is performed using GA optimization methods, which guide the search for optimal 3D structural models and sequences. This approach leverages the principles of natural selection and genetic evolution to optimize candidates for potential future VOCs. The workflow depicted in Figure 1, illustrates the sequential steps involved in this predictive process:

- Population Initialization (Mutating Wuhan Spike Protein Sequences)
- Development of the Fitness Function (Energetic Metrics)
- Selection of the Best Fitness function (Select 20 top models)
- Cross-over (Pairwise Combinations of Top Models)
- Application of Mutations for Potential Better Models
- Gibbs Free Energy Calculation
- Binding Affinity Calculation

A statistical validation step was included to support the observed trend in Discrete Optimized Protein Energy (DOPE) score reduction across generations. Both parametric and non-parametric tests were employed to ensure the robustness and reliability of the results under different data distribution assumptions.

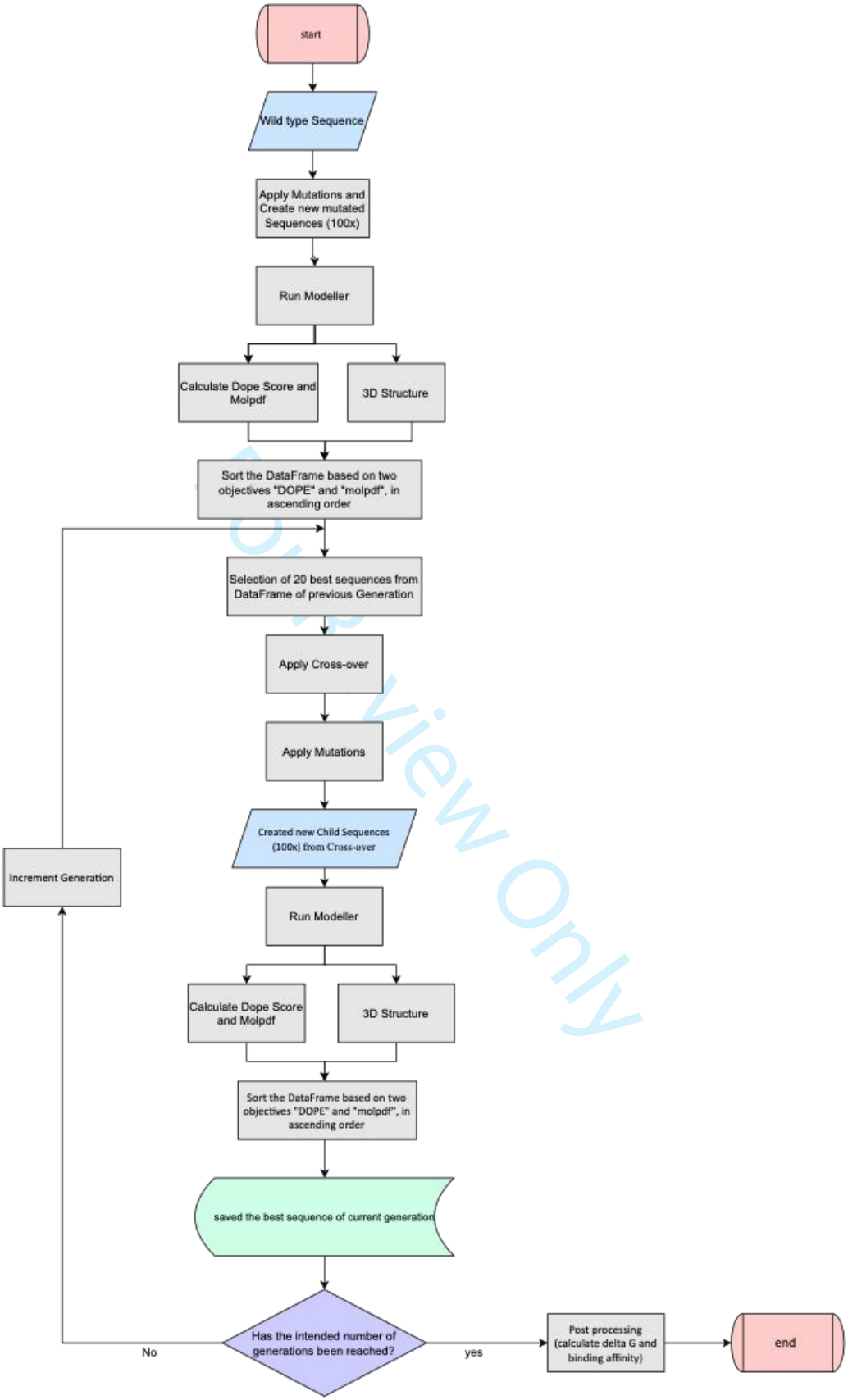

Figure 1 - The workflow of the applied GA. A graphical representation detailing the step-by-step methodology used to predict new VOCs.

Mutating Wuhan Spike Protein Sequences

We acquired the FASTA-formatted sequence of the Wuhan spike protein from the NCBI database (Accession ID: YP_009724390.1) and employed it as the input for an ad-hoc python script implementing a GA. Using this code, we initially implement random mutations to 5% of the spike protein residues with a probability of 0.5. The 5% mutation rate is likely a compromise to introduce sufficient genetic variation without overwhelming the structure with too many changes, which could lead to biologically unrealistic models. This rate is intended to simulate a mutation level that could realistically be observed in viral evolution. A smaller mutation rate may not produce enough diversity to discover potential advantageous mutations, while a higher rate could destabilize the protein structure or create unrealistic conformations. Furthermore, this rate was selected as a methodological balance: higher values (>10%) risk introducing biologically implausible destabilization of the spike structure, while lower values (<2%) may not provide sufficient genetic diversity to explore alternative conformations. A 5% rate therefore ensures both computational feasibility and biological plausibility.

This iterative process was initially conducted for 100 generations, generating a unique sequence variant at each step. In addition, to further evaluate the robustness of the GA optimization, we extended the process to 1000 generations, as reported in the Results section.

Model Building

Protein structure prediction was performed using MODELLER (https://salilab.org/modeller/), a widely used software for homology and comparative modeling of three-dimensional protein structures from amino acid sequences [15]. Based on the assumption that structurally similar proteins share similar sequences, MODELLER uses known structures of homologous proteins as templates to infer the target protein's structure [16]. The modeling process includes fold assignment, sequence-template alignment, 3D model building, and structure evaluation. To assess model quality, we used MODELLER's built-in scoring functions [17], [18]:

- DOPE Score (Discrete Optimized Protein Energy) is a statistical potential used in homology modeling to assess protein structure quality. Implemented in MODELLER, it evaluates the energy of predicted models based on a reference state of noninteracting atoms in a spherical space. DOPE provides both overall and residue-level energy profiles, helping identify low-quality regions in the model [19].
- MolPDF Score is an energy-based scoring function used to evaluate and refine protein structures. It assesses how well a predicted model fits experimental data, such as electron density maps from X-ray crystallography. Lower MolPDF values indicate better agreement between the model and experimental observations, suggesting a more accurate and reliable structure [20].

These scores guided the selection of the most reliable structural models.

MODELLER was employed to generate structural models for 100 mutated sequences, obtained in the previous step, using the human ACE2 sequence retrieved from the RCSB Protein Data Bank (RCSB PDB). For homology modelling, the spike-ACE2 complex structure (PDB ID: 6M0J) was selected as the template. MODELLER aligned each mutated sequence to the template and predicted the corresponding three-dimensional structures. The resulting models were further refined through energy minimization algorithms to enhance structural stability and accuracy.

Selection of the Best Fitness functions

In genetic algorithms, selection directs the search toward optimal solutions. This study adopts Roulette Wheel Selection, which assigns selection probability based on fitness, maintaining diversity better than Tournament Selection and offering faster convergence than Rank-Based Selection, particularly useful for variable populations like SARS-CoV-2 spike variants.

From 100 modeled spike protein structures, the 20 with the lowest combined DOPE and MolPdf scores were selected via a weighted-sum multi-objective approach.

Both metrics were normalized according to the same formula:

$$O_{i,norm} = \frac{O_i - O_{i,min}}{O_{i,max} - O_{i,min}},$$

where i = 1, 2 denotes the two metrics (O_1 = DOPE, O_2 = MolPDF).

The weighted fitness function was then computed as:

$$O_{weighted} = w_1\, O_{1,norm} + w_2\, O_{2,norm}$$

with $w_1 = 1$ and $w_2 = -1$. In this configuration, DOPE serves as the primary structural criterion, while subtracting MolPDF helps stabilize the optimization trend and reduce noise.

The choice of retaining the top 20 candidates reflects a balance between diversity and computational tractability. A smaller set would risk premature convergence toward suboptimal solutions, while a larger set would significantly increase computational costs. Twenty candidates ensured sufficient diversity for crossover operations while keeping the simulations computationally efficient.

Top Models Cross-over

Crossover was applied to the 20 top-ranking sequences selected via weighted multi-objective optimization, balancing genetic diversity and computational efficiency. A total of 100 offspring were generated through 50 crossover events, using repeated pairings among the 20 parents. This ensured broad parental contribution and promoted variability. Each offspring inherited traits from two parents, potentially combining beneficial features. This strategy enhanced exploration of the solution space and increased the likelihood of improved DOPE, MolPDF, and structural properties in the resulting models.

Mutation

Following the cross-over operation, diversity was instigated through a mutation process. Each residue had a 0.001 probability of being altered. The 0.001 probability for altering individual residues during the mutation phase is designed to ensure mutations remain conservative, minimizing the risk of detrimental mutations that could significantly impact structural integrity. This probability was chosen to maintain conservative evolutionary pressure. Larger probabilities would likely accelerate

destabilizing changes, while smaller ones would reduce exploration efficiency. A mutation rate of 0.001 per residue thus represents a trade-off between stability and the need for gradual exploration of the fitness landscape.

The resulting 100 new sequences underwent modelling, following the procedure outlined in the modelling section, to generate, and save, the best sequence of the current generation. This operation was repeated for 100 generations in the main experiment, with the best sequence from each generation saved as a representative. An extended 1000-generation run was also performed to assess long-term convergence and confirm the stability of the optimization trends.

Gibbs Free Energy Evaluation

We modeled spike–ACE2 complexes for each mutant and estimated ΔG changes in protein–ligand binding using a surface area–based computational method. For each residue, surface atoms were identified based on a reference percentage of solvent-accessible area, capturing the regions most exposed to the environment.

We then calculated the change in accessible surface area (ΔASA) for each residue, reflecting surface differences before and after binding. These ΔASA values were used to estimate binding free energy (ΔG) via the following linear model [21] :

$$\Delta G_{bind} = C + w_{Tyr}\,\Delta ASA_{Tyr} + w_{Ser}\,\Delta ASA_{Ser} + w_{Cys}\,\Delta ASA_{Cys}$$

Here, C and the weights (w_{Tyr}, w_{Ser}, and w_{Cys}) are predefined constants for TYR, SER, and CYS residues, which contribute most significantly to ΔG.

Binding Affinity Evaluation

Another important measure in the context of protein-ligand binding interactions in our work is the binding affinity. It is a measure of strength of the interaction between a protein and a ligand and is quantified by the association constant (Ka). Ka is representing the equilibrium between the bounded and unbounded states of a protein-ligand pairs:

$$K_a = \frac{[P][L]}{[PL]}$$

A higher Ka means stronger binding affinity. This implies that at equilibrium, more of the protein-ligand complex ([PL]) is formed relative to the concentrations of free protein ([P]) and free ligand ([L]). Thus, a higher Ka indicates that the binding reaction is more favorable and most of the protein and ligand molecules are in the bound state rather than unbound.

To describe the Ka mathematically, of binding affinity using the gas constant (R), and the temperature (T) in Kelvin [22]:

$$K_a = exp\left(\frac{-\Delta G}{RT}\right)$$

Both the values of ΔG and binding affinity provide us with complementary information for understanding the protein-ligand interactions. Additionally, the agreement between ΔG and binding affinity values serve us in this study to validate the accuracy of the computational methods applied.

Select variants of concern

To identify potential VOCs, we applied a multi-criteria selection strategy based on three complementary metrics: DOPE score, ΔG, and binding affinity. The DOPE score was used as the primary fitness function within the genetic algorithm, enabling the selection of structurally stable and energetically favorable spike protein models. After generation, each model was further evaluated for its binding potential to the ACE2 receptor by calculating ΔG and binding affinity values. This combined assessment allowed us to select the most promising variants those that not only exhibit low energy conformations (suggesting structural realism) but also demonstrate strong and thermodynamically favorable interactions with the host receptor. Based on this integrated analysis, specific generations were selected and introduced as candidate VOCs for further investigation.

Statistical Analysis

To statistically validate improvements during GA optimization, we performed two analyses. First, we compared DOPE scores from the first and last 10 generations using both unpaired t-tests and Mann–Whitney U tests to account for potential non-normality and unequal variances, common in evolutionary algorithm outputs [23]. Second, we applied the same tests to predicted ΔG and Ka values to evaluate enhancements in spike–ACE2 binding affinity and stability. The consistent results from both parametric and non-parametric tests confirmed the reliability of observed trends.

Although MolPDF is available in MODELLER, it was excluded due to its lack of normalization and dependence on modeling parameters, limiting its comparability. In contrast, DOPE is length-independent, more widely used in structural benchmarking [18], and better suited for evaluating diverse models. For these reasons, DOPE was chosen as the sole structural metric for validation.

All statistical analyses were conducted using the SciPy library [24], which offers reliable tools for hypothesis testing in Python.

Results

We used our GA workflow to explore the structural impact of mutations in emerging VOCs. Starting from the Wuhan reference spike protein, we introduced random mutations and evaluated their effects using a custom GA framework. The DOPE score served as the fitness function, guiding optimization toward energetically favorable models. Structural models were generated with MODELLER, and selection favored variants with minimized DOPE scores.

In total, 100 generations were initially produced, with each generation retaining the model exhibiting the lowest DOPE score. To further evaluate the structural refinement achieved through the GA, both DOPE and MolPDF scores were tracked across generations and compared to baseline values. This allowed for a comprehensive assessment of model improvement throughout the evolutionary

process.

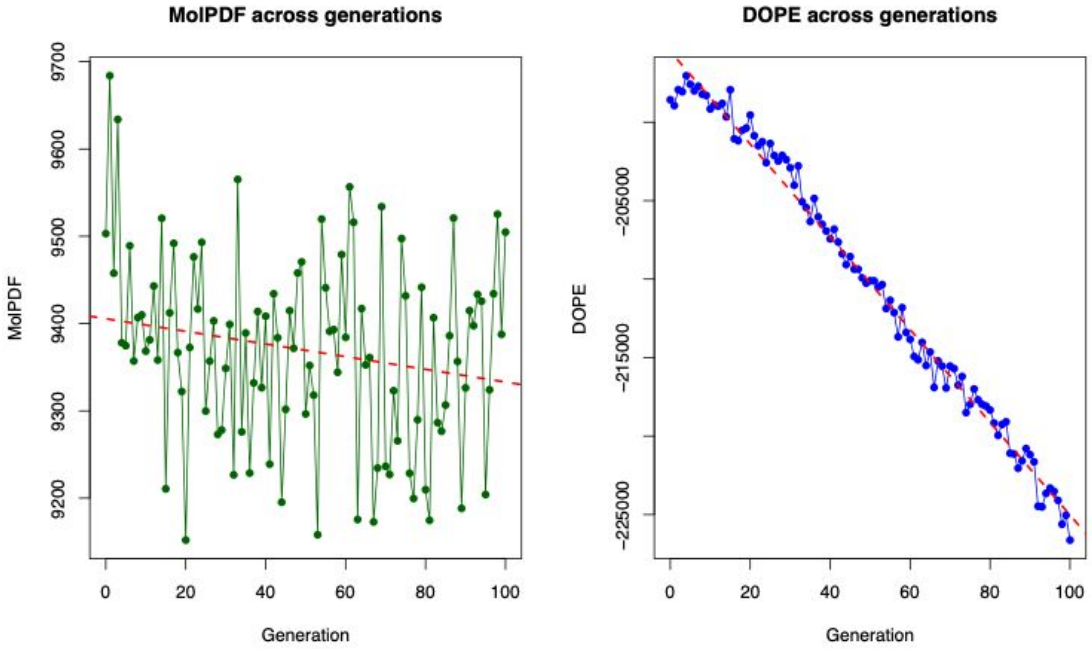

Figure 2 - The left panel shows the MolPDF score trend (green line with points). Although values remain highly variable across generations, the red dashed regression line indicates a slight overall downward tendency, suggesting limited but consistent structural refinement. The right panel shows the DOPE score trend (blue line with points), which exhibits a clear and steady decrease across generations. The red dashed regression line highlights the strong convergence toward progressively more stable and energetically favorable structures. Together, these results demonstrate the effectiveness of the GA-driven optimization in refining spike protein models.

As shown in Figure 2, DOPE scores steadily decrease across 100 generations, indicating improved model quality and structural stability. Since lower DOPE values correspond to more energetically favorable structures, this trend confirms the GA's effectiveness in selecting better variants. MolPDF values remained highly variable across generations, with fluctuations that make the trend less apparent than for DOPE. Nevertheless, a slight overall downward tendency can be observed, suggesting limited but consistent structural refinement.

Together, these reductions highlight the robustness of the GA-driven optimization process in refining spike protein models through iterative selection.

To further validate the robustness of the GA, we extended the evolutionary process to 1000 generations (Figure 3).

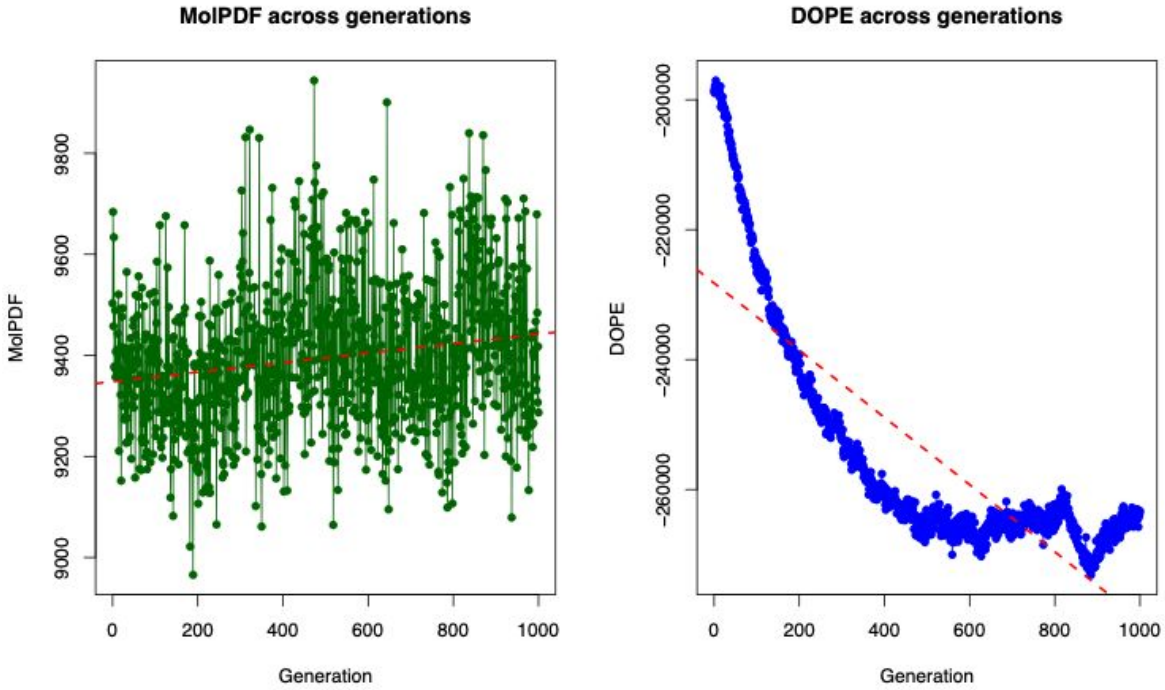

Figure 3 - MolPDF (left panel) and DOPE (right panel) score trends across 1000 generations. The MolPDF values (green line with points) remain highly variable, though the red dashed regression line indicates a subtle overall decrease, consistent with limited structural refinement. DOPE values (blue line with points), instead, display a marked downward trend. Notably, after an initial steady decline, the DOPE curve shows a temporary plateau beyond ~400 generations before resuming its descent, suggesting phases of stabilization followed by renewed optimization toward energetically favorable conformations.

Consistent with the results from the 100-generation experiment, DOPE scores showed a strong overall decrease, reflecting the convergence of the GA toward progressively more stable and energetically favorable spike protein models. Interestingly, after ~400 generations, the DOPE curve exhibited a temporary stabilization phase before resuming its decline, suggesting that the GA may encounter local optima during the search, from which subsequent generations allow escape and further refinement. In contrast, MolPDF values, which had shown a slight decrease in the 100-generation run, displayed higher variability and even a subtle upward trend across 1000 generations. This apparent discrepancy reflects the different sensitivity of the two metrics: MolPDF, as an internal pseudo-energy function of MODELLER, captures improvements during the early refinement phase but becomes less correlated with global structural stability over extended evolutionary runs. DOPE, instead, steadily decreases across both experiments, confirming its role as the most reliable indicator of GA-driven optimization.

Together, the results from both the 100- and 1000-generation experiments demonstrate that the GA framework consistently drives optimization toward lower-energy structural states, thereby confirming its suitability for modeling the impact of spike protein mutations. These structural improvements were then complemented by the thermodynamic evaluation of spike–ACE2 interactions (ΔG and Ka), providing a comprehensive assessment of both energetic stability and binding affinity across GA-driven evolutionary trajectories.

The ΔG values determine whether a process or reaction will occur spontaneously while the Ka is a parameter used to study the binding affinity. Binding affinity typically involves assessing the strength

of the interaction between a molecule (ligand) and its target (receptor), indicating how strongly they bind to each other.

Figure 4 illustrates the thermodynamic evolution of the spike-ACE2 interaction over 100 GA-optimized generations. While both ΔG and Ka fluctuate, the overall trend suggests a progressive shift toward more favorable binding conditions. ΔG values range from approximately −10.5 to −13.0 kJ/mol, with a general tendency toward more negative values, indicating increasingly spontaneous and thermodynamically favorable interactions. Concurrently, Ka exhibits substantial variability without a consistent trend, though intermittent peaks suggest episodes of enhanced binding affinity. These fluctuations reflect the GA's capacity to explore diverse variants and occasionally achieve improved spike–ACE2 binding properties.

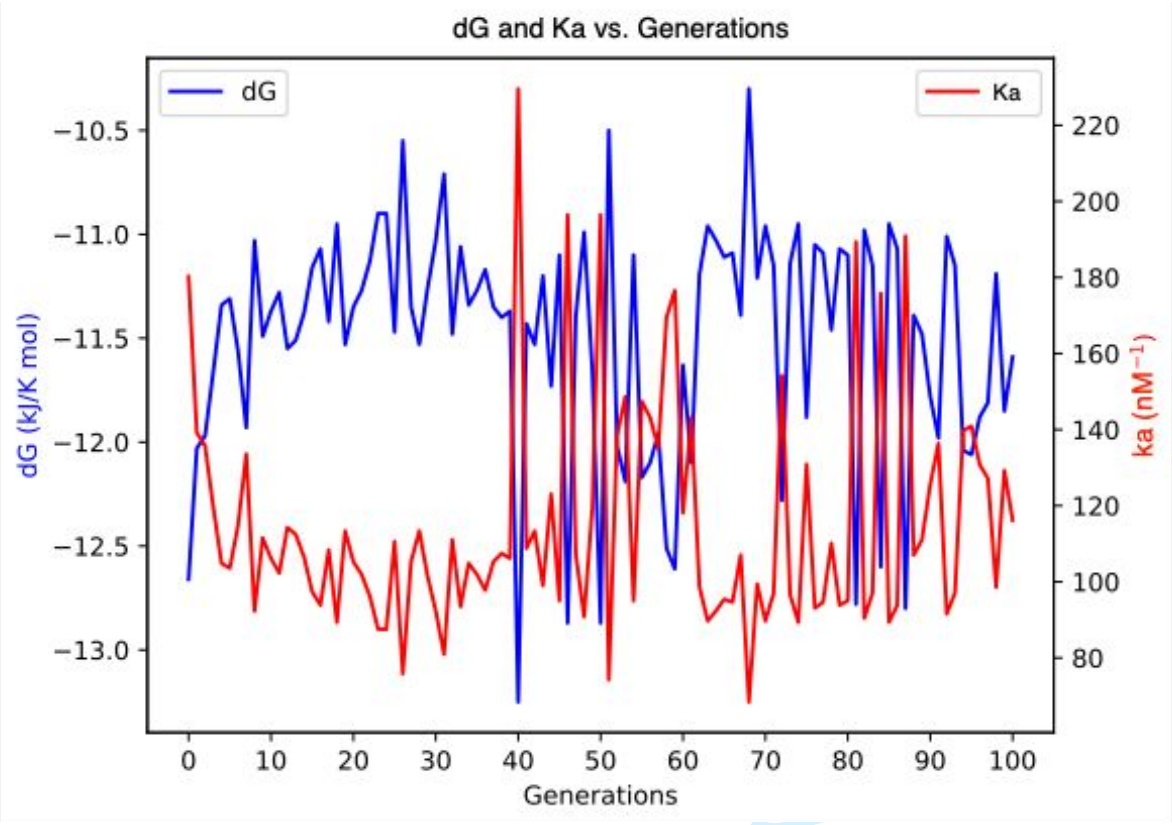

Figure 4 - The plot illustrates the thermodynamic optimization of the spike–ACE2 interaction across generations. The ΔG values (blue curve) fluctuate but overall show a trend toward lower free energy, indicating improved binding spontaneity. The Ka values (red curve) also fluctuate, with occasional increases suggesting enhanced binding affinity. These results demonstrate the genetic algorithm's capability to explore the fitness landscape and improve both the thermodynamic stability and functional interaction of spike protein variants.

The trends in ΔG and Ka confirm that the GA enhanced not only structural quality (via DOPE and Molpdf) but also the thermodynamic and functional properties of spike–ACE2 binding. Lower ΔG indicates more favorable binding, while higher Ka reflects stronger affinity. These complementary improvements validate the robustness of our GA-based framework and highlight generation 97 as a promising candidate for a potential new Variant of Concern.

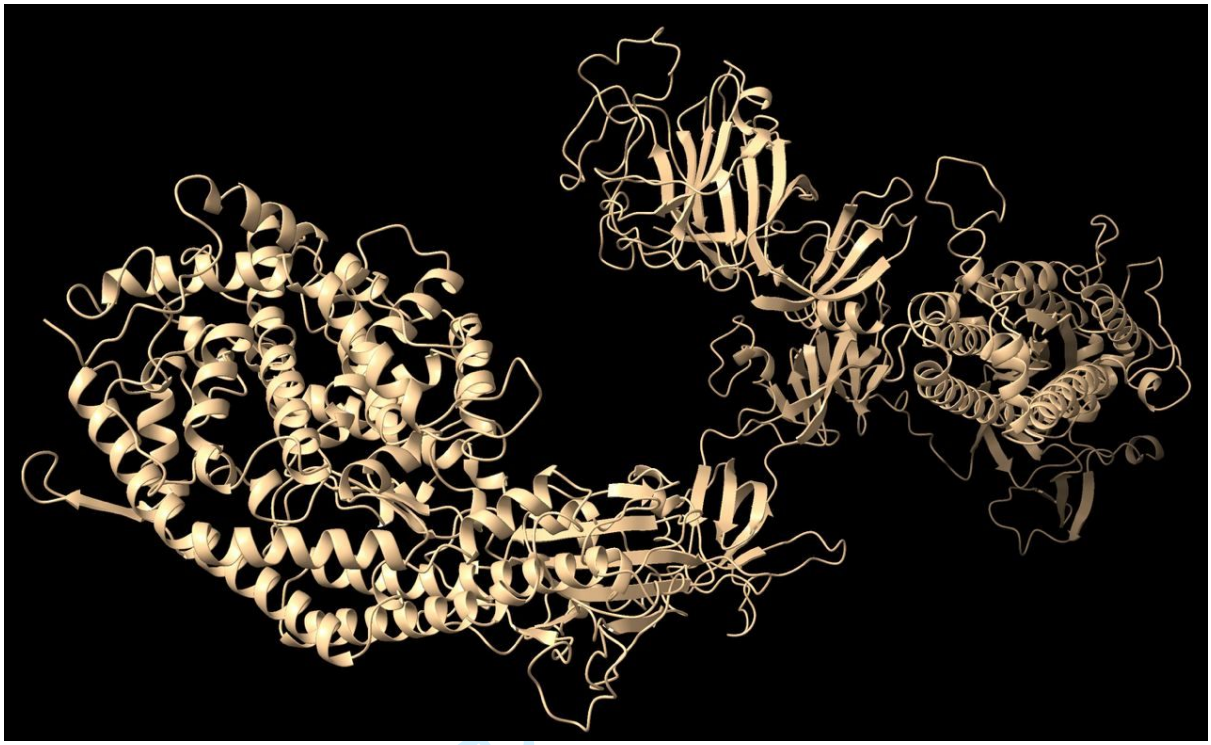

Figure 5 - 3D structure of spike protein of generation 97 obtained from GA framework.

Generation 97 emerged as the most optimized candidate within our GA framework, displaying the lowest DOPE score and therefore the highest structural stability among the variants generated (Figure 5). It also showed favorable ΔG and binding affinity values (Table 1), supporting its structural plausibility and efficient receptor engagement in silico. While these findings highlight the ability of the GA to identify structurally optimized spike variants, no claim is made regarding their occurrence in circulating lineages or their classification as real Variants of Concern.

Generation	DOPE score	ΔG	Binding Affinity
97	-223520.29688	-12.61	176.552271

Table 1 - The table shows the key performance metrics for selected possible new VOC. The columns include the Generation number, DOPE score, ΔG and Binding Affinity.

Statistical analysis

A comprehensive statistical comparison was performed between the first and last ten generations of spike protein models to assess improvements in structural and functional metrics. As shown in Table 2, the DOPE score exhibited a highly significant decrease (t-test $p = 2.36 \times 10^{-16}$; Mann–Whitney p = 0.000183), confirming a progressive enhancement in model stability and conformational quality across generations. The median DOPE score decreased by more than 26,000 units between the first (−198,031.51) and last (−224,283.43) ten generations, reflecting a substantial gain in structural stability achieved through GA-driven optimization. This trend aligns with the core objective of the GA,

which was designed to iteratively select variants with lower energy profiles, thereby promoting convergence towards structurally optimized models.

Test	T-test p-value	Mann-Whitney p-value
DOPE Score	2.36E-16	0.00018267
ΔG (Gibbs Free Energy)	0.20046886	0.24114463
Ka (Binding Affinity)	0.1992096	0.24114463

Table 2 - Statistical comparison of DOPE score, Gibbs free energy (ΔG), and association constant (Ka) between early (generations 1–10) and late (generations 91–100) spike protein models.

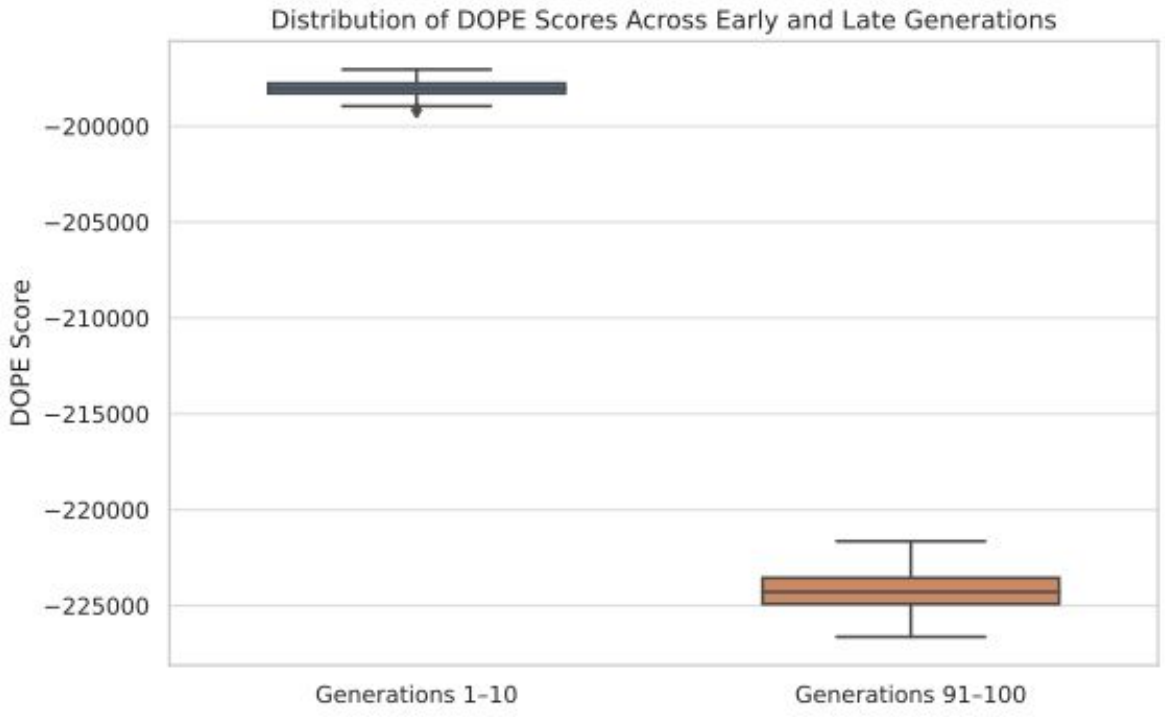

Figure 6 - Boxplot of DOPE scores for the first and last ten generations. The significant decrease in DOPE values in later generations reflects improved structural stability driven by the genetic algorithm optimization.

In contrast, the ΔG and Ka did not show statistically significant differences between the first and last generations (ΔG p = 0.20; Ka p = 0.20, t-test). These results suggest that, while the GA effectively optimized structural stability, the functional improvements in binding thermodynamics were less pronounced. This interpretation is further supported by the modest differences observed in the mean values: ΔG improved only slightly, from −11.73 to −12.05 kJ/mol, and Ka increased from 125.50 to 143.40 nM^{-1} between the early and late generations. These limited shifts reinforce the notion that, while structural energy minimization was prioritized and successfully achieved, thermodynamic binding properties did not undergo a comparable evolutionary pressure.

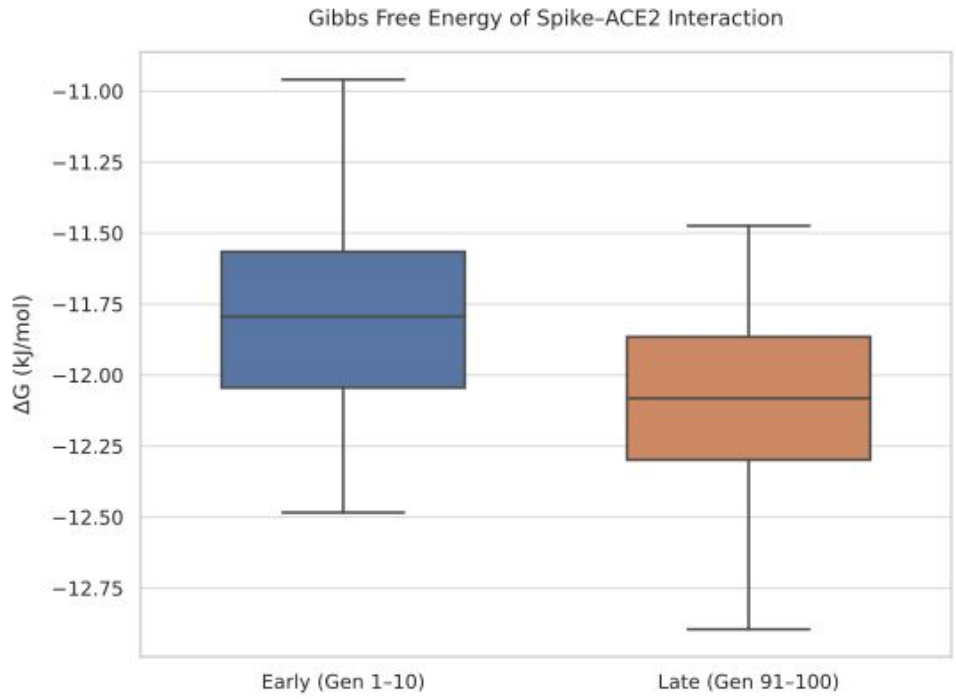

Figure 7 - ΔG values of spike–ACE2 interaction across early and late generations. No statistically significant changes were observed, suggesting limited functional evolution under structural optimization pressure.

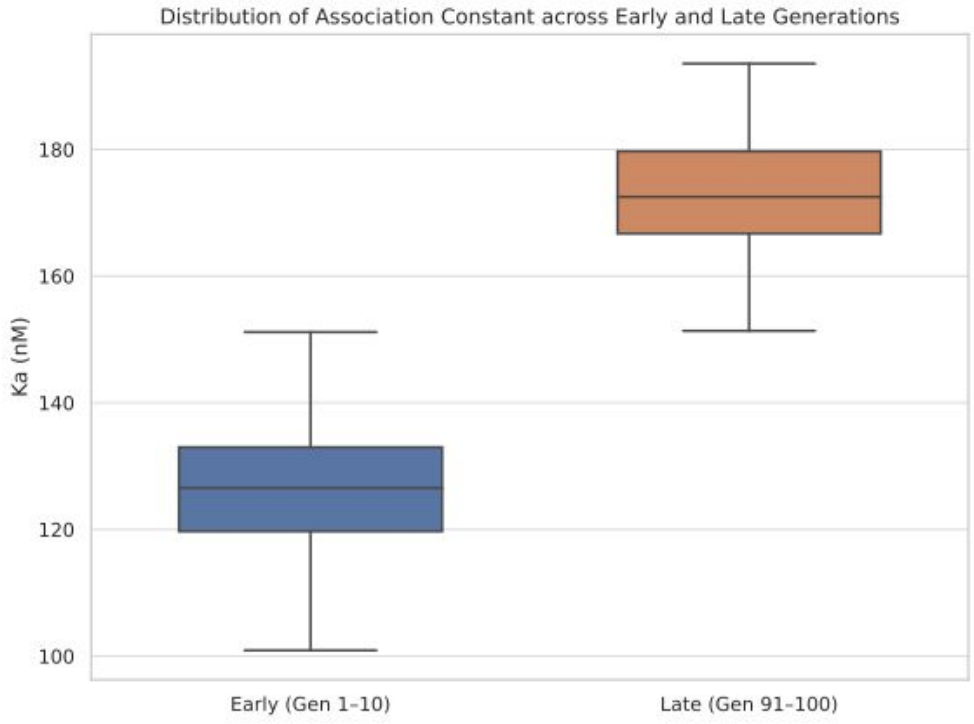

Figure 8 - Ka distribution across early and late generations. The lack of significant variation indicates that binding affinity was not effectively optimized in the absence of direct selective pressure.

A possible explanation lies in the multi-objective nature of protein-ligand interaction optimization. ΔG and Ka are influenced by subtle, non-linear conformational and surface changes that may not evolve in parallel with structural compactness or energy minimization [25]. Moreover, as ΔG and Ka were not

directly included in the GA's fitness function but evaluated a posteriori, their evolution was not subject to selective pressure. This decoupling may explain the lack of statistically significant improvement in these functional indicators, despite the evident refinement of the structural model itself.

Together, these statistical results validate the effectiveness of the GA-driven modeling and selection strategy in optimizing structural features of spike protein variants. While functional metrics such as ΔG and Ka did not show significant improvements, the marked enhancement in structural stability supports the utility of this approach for generating conformationally optimized candidates with potential biological relevance.

Discussion

The use of Genetic Algorithms (GAs) guided by fitness metrics such as the DOPE score provides a flexible and powerful computational framework for exploring the mutational landscape of the SARS-CoV-2 spike protein. Post hoc analyses of Gibbs Free Energy and binding affinity offer complementary insights into structural stability and functional relevance, particularly in the context of spike–ACE2 interactions. While ΔG informs on thermodynamic feasibility, binding affinity reflects the strength of receptor engagement. Together with structural optimization, these measures help characterize the potential impact of mutations at the molecular level.

In our study, the DOPE score was employed as the main fitness function, and we consistently observed a progressive decrease across generations, suggesting convergence toward structurally more stable and energetically favorable models. Within the initial 100 generations explored, generation 97 exhibited the lowest DOPE score and represented a point of convergence in the optimization process. Extending the evolutionary process to 1000 generations further confirmed this trend, with DOPE scores continuing to decline and showing strong convergence overall. Interestingly, after approximately 400 generations, the DOPE curve entered a temporary plateau before resuming its descent, suggesting that the GA may encounter local optima during the search but is ultimately able to escape and refine further. This extended experiment reinforces the robustness of the GA-driven optimization framework and highlights its potential for long-term convergence.

MolPDF values, in contrast, displayed more variable behavior. While a slight decrease was observed in the first 100 generations, the extended 1000-generation run revealed higher variability and even a modest upward trend. This apparent discrepancy can be explained by the different sensitivity of the two metrics: MolPDF, as an internal pseudo-energy function of MODELLER, captures improvements during the early refinement phase but becomes less correlated with global structural stability over extended evolutionary timescales. Thus, while MolPDF remains informative, DOPE provides a more reliable and consistent measure of convergence and model quality in our GA framework.

Although ΔG and binding affinity did not exhibit statistically significant changes across early and late generations, they provided valuable additional context. The observation that structural stability improved more consistently than binding properties suggests that different selective pressures may govern these features and highlights the importance of multi-objective optimization in future applications. Importantly, combining binding-related parameters with structural stability metrics ensures that candidate variants are not only energetically favorable but also biologically plausible.

Several limitations of the current work should be acknowledged. The GA was originally restricted to 100 generations due to computational constraints, which may have limited the full exploration of the mutational space. The additional 1000-generation experiment addresses this limitation, strengthening our conclusions, but further scaling up of both population size and evolutionary depth will be important for future applications. Moreover, the study focused exclusively on spike protein mutations, without considering other viral proteins or host–pathogen interactions. Functional metrics such as ΔG and binding affinity, while informative, may not fully capture the complexity of viral behavior in vivo.

Despite these limitations, the study provides a proof-of-concept for applying GA to structural modeling of viral proteins. This methodological framework can serve as a foundation for future research integrating additional fitness functions, larger simulation scales, and more diverse mutational scenarios. Crucially, experimental validation of top candidates (e.g., generation 97) through in vitro assays, as well as benchmarking against genomic surveillance data, will be essential to assess predictive potential. Integration with phylogenetic models, deep learning approaches, and epidemiological studies could further enhance the utility of this approach.

In addition, while Generation 97 emerged as the most optimized candidate in our GA framework, we did not compare its mutations with those observed in circulating VOCs nor assess whether they fall within functionally relevant domains (e.g., receptor-binding domain, N-terminal domain, furin cleavage site). This remains an important limitation of the present study, and a priority for future work integrating genomic surveillance data and structural domain mapping.

Overall, this work demonstrates the promise of GA-based structural modeling as a methodological tool for systematically exploring protein mutations. While not intended to forecast specific Variants of Concern, it provides a solid computational basis that, once combined with evolutionary and experimental evidence, may contribute to more comprehensive strategies for monitoring viral evolution and informing vaccine design.

Conclusion

As the COVID-19 pandemic continues and novel variants emerge, computational methods remain essential for understanding the molecular mechanisms that drive viral evolution. Genetic Algorithms (GAs), inspired by natural selection, provide a versatile framework to explore mutational landscapes and refine structural models of viral proteins.

In this study, we implemented a GA-based approach to optimize SARS-CoV-2 spike protein variants using energy-based metrics such as the DOPE score, Gibbs Free Energy (ΔG), and binding affinity. The initial 100-generation run highlighted progressively more stable conformations, with generation 97 representing the most optimized candidate within the explored search space. Extending the GA to 1000 generations confirmed the robustness of this framework, as DOPE scores continued to decrease and revealed long-term convergence patterns, including transient plateaus followed by renewed optimization. These findings illustrate the ability of GAs to iteratively refine structural plausibility and energetic favorability over both short and extended evolutionary timescales.

While the present work does not aim to predict specific real-world Variants of Concern, it demonstrates the potential of GA-driven modeling as a methodological tool to investigate mutational effects at the structural level. Importantly, the approach establishes a computational foundation that could, in future, be combined with genomic surveillance, phylogenetics, and experimental validation to strengthen predictive insights.

Future research should build on this framework by expanding fitness functions, adopting multi-objective optimization strategies, and scaling up both the number of generations and the population sizes. Such developments may enhance our ability to link structural modeling with evolutionary

dynamics, ultimately supporting more comprehensive strategies for monitoring viral evolution and informing vaccine design.

List of Abbreviations

VOC: Variant of concern

S: Spike protein

ΔG: Gibbs Free energy

GA: Genetic Algorithm

Ka: Association Constant

ACE2: Angiotensin-converting enzyme 2

References

[1] S. P. Otto *et al.*, "The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic," *Curr. Biol.*, vol. 31, no. 14, pp. R918–R929, July 2021, doi: 10.1016/j.cub.2021.06.049.

[2] R. Sanjuán and P. Domingo-Calap, "Mechanisms of viral mutation," *Cell. Mol. Life Sci.*, vol. 73, no. 23, pp. 4433–4448, Dec. 2016, doi: 10.1007/s00018-016-2299-6.

[3] R. Sanjuán, M. R. Nebot, N. Chirico, L. M. Mansky, and R. Belshaw, "Viral Mutation Rates," *J. Virol.*, vol. 84, no. 19, pp. 9733–9748, Oct. 2010, doi: 10.1128/JVI.00694-10.

[4] L. Ballesteros-Sanabria, H. F. Pelaez-Prestel, A. Ras-Carmona, and P. A. Reche, "Resilience of Spike-Specific Immunity Induced by COVID-19 Vaccines against SARS-CoV-2 Variants," *Biomedicines*, vol. 10, no. 5, p. 996, Apr. 2022, doi: 10.3390/biomedicines10050996.

[5] W. T. Harvey *et al.*, "SARS-CoV-2 variants, spike mutations and immune escape," *Nat. Rev. Microbiol.*, vol. 19, no. 7, pp. 409–424, July 2021, doi: 10.1038/s41579-021-00573-0.

[6] E. A. Rashed, S. Kodera, and A. Hirata, "COVID-19 forecasting using new viral variants and vaccination effectiveness models," *Comput. Biol. Med.*, vol. 149, p. 105986, Oct. 2022, doi: 10.1016/j.compbiomed.2022.105986.

[7] E. A. Rashed, S. Kodera, H. Shirakami, R. Kawaguchi, K. Watanabe, and A. Hirata, "Knowledge discovery from emergency ambulance dispatch during COVID-19: A case study of Nagoya City, Japan," *J. Biomed. Inform.*, vol. 117, p. 103743, May 2021, doi: 10.1016/j.jbi.2021.103743.

[8] L. Xu, R. Magar, and A. Barati Farimani, "Forecasting COVID-19 new cases using deep learning methods," *Comput. Biol. Med.*, vol. 144, p. 105342, May 2022, doi: 10.1016/j.compbiomed.2022.105342.

[9] C. Bai *et al.*, "Predicting Mutational Effects on Receptor Binding of the Spike Protein of SARS-CoV-2 Variants," *J. Am. Chem. Soc.*, vol. 143, no. 42, pp. 17646–17654, Oct. 2021, doi: 10.1021/jacs.1c07965.

[10] Z. Michalewicz and M. Schoenauer, "Evolutionary Algorithms for Constrained Parameter Optimization Problems," *Evol. Comput.*, vol. 4, no. 1, pp. 1–32, Mar. 1996, doi: 10.1162/evco.1996.4.1.1.

[11] J. H. Holland, "Genetic Algorithms," *Sci. Am.*, vol. 267, no. 1, pp. 66–73, 1992.

[12] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimed. Tools Appl.*, vol. 80, no. 5, pp. 8091–8126, Feb. 2021, doi: 10.1007/s11042-020-10139-6.
[13] A. A. Abou El Ela and S. R. Spea, "Optimal corrective actions for power systems using multi-objective genetic algorithms," *Electr. Power Syst. Res.*, vol. 79, no. 5, pp. 722–733, May 2009, doi: 10.1016/j.epsr.2008.10.007.
[14] Z. Liao, Y. Song, S. Ren, X. Song, X. Fan, and Z. Liao, "VOC-DL: Deep learning prediction model for COVID-19 based on VOC virus variants," *Comput. Methods Programs Biomed.*, vol. 224, p. 106981, Sept. 2022, doi: 10.1016/j.cmpb.2022.106981.
[15] B. Webb and A. Sali, "Protein Structure Modeling with MODELLER," in *Structural Genomics*, vol. 2199, Y. W. Chen and C.-P. B. Yiu, Eds., in Methods in Molecular Biology, vol. 2199. , New York, NY: Springer US, 2021, pp. 239–255. doi: 10.1007/978-1-0716-0892-0_14.
[16] B. Webb and A. Sali, "Comparative Protein Structure Modeling Using MODELLER," *Curr. Protoc. Bioinforma.*, vol. 54, no. 1, June 2016, doi: 10.1002/cpbi.3.
[17] N. Eswar *et al.*, "Comparative Protein Structure Modeling Using Modeller," *Curr. Protoc. Bioinforma.*, vol. 15, no. 1, Sept. 2006, doi: 10.1002/0471250953.bi0506s15.
[18] A. Šali and T. L. Blundell, "Comparative Protein Modelling by Satisfaction of Spatial Restraints," *J. Mol. Biol.*, vol. 234, no. 3, pp. 779–815, Dec. 1993, doi: 10.1006/jmbi.1993.1626.
[19] D. Eramian, M. Shen, D. Devos, F. Melo, A. Sali, and M. A. Marti-Renom, "A composite score for predicting errors in protein structure models," *Protein Sci.*, vol. 15, no. 7, pp. 1653–1666, July 2006, doi: 10.1110/ps.062095806.
[20] M. Shen and A. Sali, "Statistical potential for assessment and prediction of protein structures," *Protein Sci.*, vol. 15, no. 11, pp. 2507–2524, Nov. 2006, doi: 10.1110/ps.062416606.
[21] A. Schön, N. Madani, A. B. Smith, J. M. Lalonde, and E. Freire, "Some Binding-Related Drug Properties are Dependent on Thermodynamic Signature: Optimization of Drug Candidates," *Chem. Biol. Drug Des.*, vol. 77, no. 3, pp. 161–165, Mar. 2011, doi: 10.1111/j.1747-0285.2010.01075.x.
[22] P. L. Kastritis and A. M. J. J. Bonvin, "On the binding affinity of macromolecular interactions: daring to ask why proteins interact," *J. R. Soc. Interface*, vol. 10, no. 79, p. 20120835, Feb. 2013, doi: 10.1098/rsif.2012.0835.
[23] N. Nachar, "The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution," *Tutor. Quant. Methods Psychol.*, vol. 4, no. 1, pp. 13–20, Mar. 2008, doi: 10.20982/tqmp.04.1.p013.
[24] P. Virtanen *et al.*, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nat. Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.
[25] M. Van Den Noort, M. De Boer, and B. Poolman, "Stability of Ligand-induced Protein Conformation Influences Affinity in Maltose-binding Protein," *J. Mol. Biol.*, vol. 433, no. 15, p. 167036, July 2021, doi: 10.1016/j.jmb.2021.167036.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Software and Data Availability Statement

Datasets and software used in the experiments are listed as follows:

(1) NCBI: National Center for Biotechnology Information (National Center for Biotechnology Information (nih.gov)),(2) RCSB Protein Data Bank: RCSB PDB: Homepage (3) MODELLER: About MODELLER (salilab.org)

Acknowledgements

Not applicable

Competing interests

Not applicable.

Funding

Programma di ricerca CN00000013 “National Centre for HPC, Big Data and Quantum Computing”, finanziato dal Decreto Direttoriale di concessione del finanziamento n.1031 del 17.06.2022 a valere sulle risorse del PNRR MUR – M4C2 – Investimento 1.4 - Avviso “Centri Nazionali” - D.D. n. 3138 del 16 dicembre 2021

Authors' contributions

Valentina Di Salvatore: Writing - review & editing, statistical analysis. Avisa Maleki: Designed the work pipeline, collect data, writing original draft and review & editing Babak Mohajer: Algorithm developer and writing python scripts of GA, writing method part. Alvaro Ras-Carmona: Writing the first Python script. Giulia Russo: Writing - review & editing, formal analysis. Pedro Antonio Reche: Writing - review & editing. Francesco Pappalardo: Conceptualization, Supervision, Methodology.

Exploring SARS-CoV-2 Spike Protein Mutations through Genetic Algorithm-Driven Structural Modeling ~~Anticipating Viral Evolution: Genetic Algorithm-Driven Insights into SARS-CoV-2 Spike Mutations~~

Valentina Di Salvatore[1†], Avisa Maleki[1†], Babak Mohajer[2], Alvaro Ras-Carmona[3], Giulia Russo[1], Pedro Antonio Reche[3], and Francesco Pappalardo[1]*

[1]Department of Drug and Health Sciences, University of Catania, Viale Andrea Doria 6, Catania, 95125, Italy

[2]Mimesis SRL, Catania, Italy

[3]Department of Immunology, Faculty of Medicine, University Complutense of Madrid, Madrid, Spain

†These authors contributed equally to this work and share first authorship.

* Corresponding author, valentinadisalvatore@unict.it, Tel.: +39-095-7384040

Abstract

Background

The rapid evolution of SARS-CoV-2 highlights the importance of computational approaches to explore mutational effects on the viral spike protein. In this work, we present a Genetic Algorithm (GA) framework applied to the structural optimization of spike protein variants, with a focus on energetic and binding properties rather than direct evolutionary prediction.~~The rapid evolution and transmission of SARS-CoV-2 underscore the urgent need for innovative approaches to predict and monitor emerging Variants of Concern. This study integrates a Genetic Algorithm framework to forecast potential mutational shifts in the spike protein focusing on their structural and energetic implications. We adapted the Genetic Algorithm to optimize three-dimensional models of spike protein variants, analysing the impact of mutations on binding affinity and Gibbs Free Energy change ΔG. Our method employs a robust evaluation system utilizing the Discrete Optimized Protein Energy score and Molecular Probability Density Function as fitness functions to refine the selection of potential Variants of Concern through rigorous computational modelling.~~

Results

Our GA-driven pipeline generated spike variants with progressively improved structural stability as indicated by lower DOPE scores across generations. The approach also enabled evaluation of Gibbs Free Energy (ΔG) and binding affinity for spike – Angiotensin-converting enzyme 2 receptor (ACE2) interactions, revealing candidate conformations with favorable thermodynamic properties. These results demonstrate the algorithm's capacity to refine protein models and explore mutational landscapes in silico, although no validation against naturally emerging variants was performed.~~The study successfully identified specific mutations that significantly alter the spike protein interaction with the Angiotensin-converting enzyme 2 receptor (ACE2), enhancing viral binding efficiency and potential transmissibility. By predicting and analysing these interactions, we provide valuable insights into the structural dynamics of VOCs, which could inform vaccine design and public health strategies.~~

Conclusions

This study presents a methodological framework for GA-based structural modeling of SARS-CoV-2 spike mutations. Rather than forecasting specific Variants of Concern, it demonstrates the feasibility of a computational approach that can be extended and integrated with evolutionary and experimental evidence to strengthen future efforts in variant monitoring and vaccine development.~~Our findings highlight the potential of using advanced bioinformatic tools and Genetic Algorithm (GA) optimization techniques to proactively address the challenges posed by rapidly evolving infectious diseases.~~

Keywords: SARS-CoV-2, Variant of Concern, Genetic Algorithm, Spike protein

Introduction

RNA viruses, including SARS-CoV-2, mutate frequently due to the intrinsic properties of their genomes [1]. The impact of these mutations on transmission and disease severity depends on mutation rates and their effects within, and between, hosts [2], [3], contributing to the emergence of viral variants and influencing pandemic dynamics.

SARS-CoV-2 relies on its spike (S) glycoprotein, —a homotrimeric class I fusion protein, —for host cell entry. Cleavage by host proteases like furin and TMPRSS2 triggers conformational changes that enable ACE2 binding and membrane fusion [4]. Given this mechanism, accurately predicting Variants of Concern (VOCs) remains essential in infectious disease control.

Variants such as Alpha (B.1.1.7), Beta (B.1.351), Delta (B.1.617.2), and Omicron (B.1.1.529) emerged through spike mutations, increasing transmissibility, altering disease severity, and challenging vaccine efficacy. Meanwhile, mRNA vaccines like Pfizer-BioNTech and Moderna showed strong protection against the original strain, with booster doses later introduced to counter immune evasion by emerging variants [4], [5].

Forecasting new variants enables timely public health responses [6], supported by early surveillance systems [7]. A comprehensive strategy—integrating genomics, bioinformatics, phylogenetics, mutation tracking, epidemiology, and AI—facilitates real-time analysis of viral evolution [8]. In particular, understanding mutation-driven changes in spike–ACE2 binding and Gibbs free energy (ΔG) is key to anticipating variant behavior and biological impact [9].

Metaheuristic algorithms are widely used to solve complex problems in biology, economics, and engineering. ~~Metaheuristic algorithms are widely used to solve complex problems in biology, economics, and engineering~~ [9], [10], [11]. ~~They are typically classified as single-solution (e.g., simulated annealing, Tabu Search, Guided Local Search) or population-based methods (e.g., genetic algorithms (GA), PSO [10], Ant Colony Optimization (ACO) [11], spotted hyena optimizer (SHO) [12], emperor penguin optimizer (EPO) [13], and seagull optimization (SOA) [14]). While single-solution methods refine one candidate at a time and may get trapped in local optima [11], population-based approaches promote diversity and enhance global search efficiency.~~

Among these, Genetic Algorithms (GAs) are a prominent population-based method inspired by natural evolution.~~Among these, GA is a prominent method inspired by natural evolution~~ [10]. Proposed by J.H. Holland in 1992, it incorporates chromosome representation, fitness evaluation, and biologically inspired operators, including the Inversion operator [11]. Chromosomes,—often binary-encoded,—are modified via crossover and mutation, while selection favors the fittest individuals. Through iterative application, GA drives the population toward optimal solutions [12], [13].

One study proposed VOC-DL, a deep learning framework integrating variant data into models like VOC-LSTM, VOC-GRU, and VOC-BiLSTM to forecast daily COVID-19 cases. Using data from Italy, South Korea, Russia, Japan, and India over three months in 2021, all models performed well, with VOC-LSTM achieving the highest accuracy (average R^2 = 96.83%). The authors concluded that VOC-DL,—particularly VOC-LSTM,—offers reliable long-term forecasting and could aid in assessing the impact of future variants [14]. Another study predicted SARS-CoV-2 mutations using epidemiological, evolutionary, and immunological features. While effective at forecasting known mutations linked to ACE2 binding and immune escape, it could not anticipate novel mutations or full lineage evolution. Existing models, including machine learning and logistic regression, often fail to account for the structural effects of mutations on protein–ligand interactions,—key factors in viral transmission and immune evasion [9].

In this study, we propose a GA-based computational framework designed to explore the structural and energetic implications of spike protein mutations. Rather than attempting to directly predict real-world Variants of Concern (VOCs), our goal is to demonstrate the feasibility of applying GA to optimize spike protein conformations and investigate their impact on stability and spike–ACE2 interactions. This methodological approach lays the groundwork for future extensions where genomic surveillance and evolutionary data could be integrated to assess predictive performance. ~~Our GA framework addresses this gap by integrating structural modeling with evolutionary dynamics to predict mutations that improve viral fitness. By optimizing protein stability, binding affinity, and ΔG, it aims to clarify mutation-driven changes in infectivity and inform vaccine target identification.~~

Method

The prediction of VOCs is performed using GA optimization methods, which guide the search for optimal 3D structural models and sequences. This approach leverages the principles of natural selection and genetic evolution to optimize candidates for potential future VOCs. The workflow depicted in Figure 1, illustrates the sequential steps involved in this predictive process:

- Population Initialization (Mutating Wuhan Spike Protein Sequences)
- Development of the Fitness Function (Energetic Metrics)
- Selection of the Best Fitness function (Select 20 top models)
- Cross-over (Pairwise Combinations of Top Models)
- Application of Mutations for Potential Better Models
- Gibbs Free Energy Calculation
- Binding Affinity Calculation

A statistical validation step was included to support the observed trend in Discrete Optimized Protein Energy (DOPE) score reduction across generations. Both parametric and non-parametric tests were employed to ensure the robustness and reliability of the results under different data distribution assumptions.

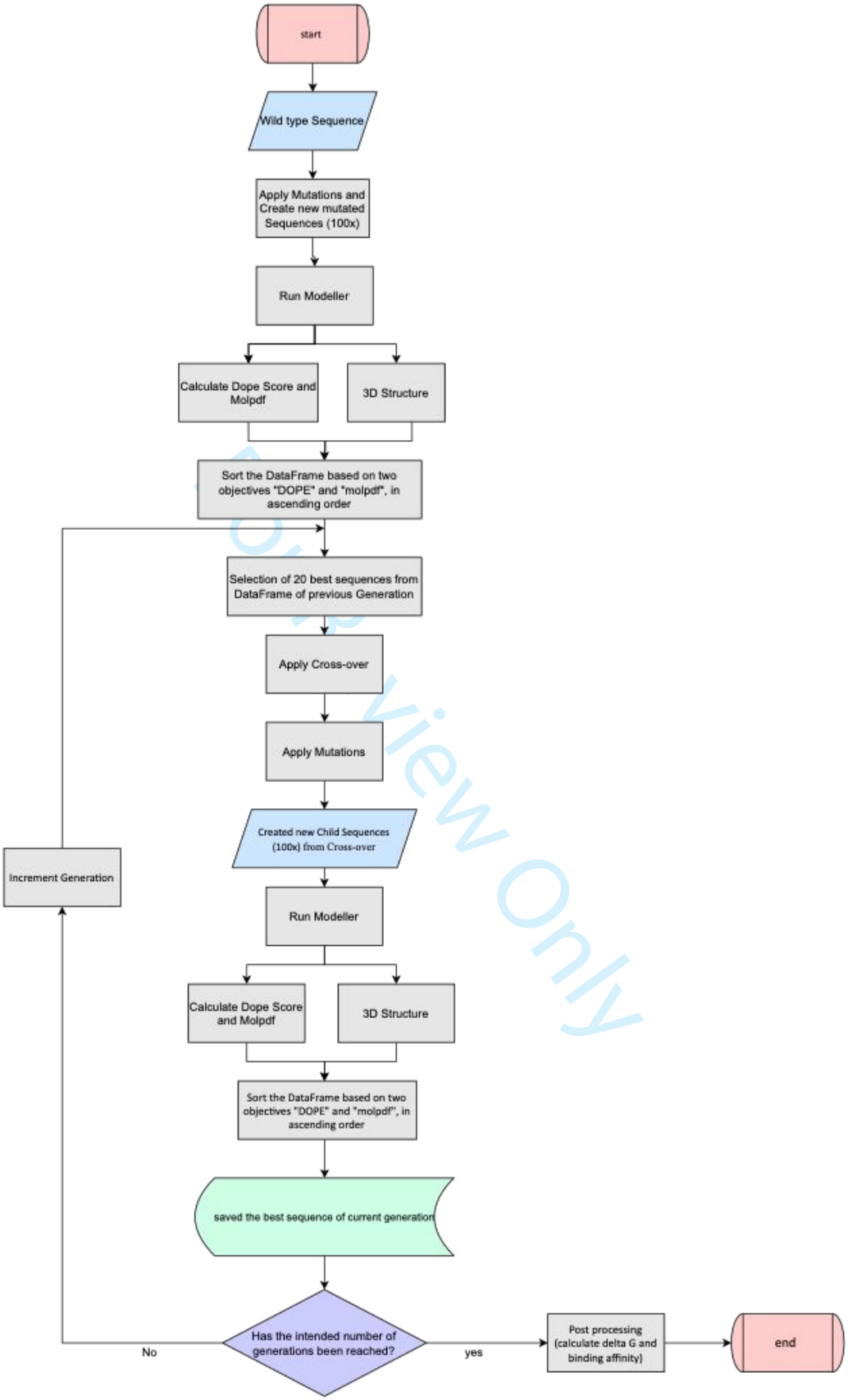

Figure 1 - The workflow of the applied GA. A graphical representation detailing the step-by-step methodology used to predict new VOCs.

Mutating Wuhan Spike Protein Sequences

We acquired the FASTA-formatted sequence of the Wuhan spike protein from the NCBI database (Accession ID: YP_009724390.1) and employed it as the input for an ad-hoc python script implementing a GA. Using this code, we initially implement random mutations to 5% of the spike protein residues with a probability of 0.5. The 5% mutation rate is likely a compromise to introduce sufficient genetic variation without overwhelming the structure with too many changes, which could lead to biologically unrealistic models. This rate is intended to simulate a mutation level that could realistically be observed in viral evolution. A smaller mutation rate may not produce enough diversity to discover potential advantageous mutations, while a higher rate could destabilize the protein structure or create unrealistic conformations. Furthermore, this rate was selected as a methodological balance: higher values (>10%) risk introducing biologically implausible destabilization of the spike structure, while lower values (<2%) may not provide sufficient genetic diversity to explore alternative conformations. A 5% rate therefore ensures both computational feasibility and biological plausibility.

This iterative process was initially conducted for 100 generations, generating a unique sequence variant at each step. In addition, to further evaluate the robustness of the GA optimization, we extended the process to 1000 generations, as reported in the Results section.

~~This iterative process was conducted 100 times, generating a unique sequence variant with each set of mutations. As a result, we obtained a heterogeneous initial population comprising 100 mutated sequences, which served as the foundation for the subsequent steps in the GA.~~

Model Building

Protein structure prediction was performed using MODELLER (https://salilab.org/modeller/), a widely used software for homology and comparative modeling of three-dimensional protein structures from amino acid sequences [15]. Based on the assumption that structurally similar proteins share similar sequences, MODELLER uses known structures of homologous proteins as templates to infer the target protein's structure [16]. The modeling process includes fold assignment, sequence-template alignment, 3D model building, and structure evaluation. To assess model quality, we used MODELLER's built-in scoring functions [17], [18]:

- DOPE Score (Discrete Optimized Protein Energy) is a statistical potential used in homology modeling to assess protein structure quality. Implemented in MODELLER, it evaluates the energy of predicted models based on a reference state of noninteracting atoms in a spherical space. DOPE provides both overall and residue-level energy profiles, helping identify low-quality regions in the model [19].
- MolPDF Score is an energy-based scoring function used to evaluate and refine protein structures. It assesses how well a predicted model fits experimental data, such as electron density maps from X-ray crystallography. Lower MolPDF values indicate better agreement between the model and experimental observations, suggesting a more accurate and reliable structure [20].

These scores guided the selection of the most reliable structural models.

MODELLER was employed to generate structural models for 100 mutated sequences, obtained in the previous step, using the human ACE2 sequence retrieved from the RCSB Protein Data Bank (RCSB PDB). For homology modelling, the spike-ACE2 complex structure (PDB ID: 6M0J) was selected as the template. MODELLER aligned each mutated sequence to the template and predicted the corresponding three-dimensional structures. The resulting models were further refined through energy minimization algorithms to enhance structural stability and accuracy.

Selection of the Best Fitness functions

In genetic algorithms, selection directs the search toward optimal solutions. This study adopts Roulette Wheel Selection, which assigns selection probability based on fitness, maintaining diversity better than Tournament Selection and offering faster convergence than Rank-Based Selection, particularly useful for variable populations like SARS-CoV-2 spike variants.

From 100 modeled spike protein structures, the 20 with the lowest combined DOPE and MolPdf scores were selected via a weighted-sum multi-objective approach.

Both metrics were normalized according to the same formula:~~To combine these two metrics—DOPE (O_1) and MolPdf (O_2)—they were normalized using min-max normalization:~~

$$O_{i1,norm} = \frac{O_{i1} - O_{i1,min}}{O_{i1,max} - O_{i1,min}}$$

$$O_{2,norm} = \frac{O_2 - O_{2,min}}{O_{2,max} - O_{2,min}}$$

where i = 1, 2 denotes the two metrics (O_1 = DOPE, O_2 = MolPDF).

The weighted fitness function was then computed as:

~~Since lower DOPE values indicate better models, $O_{1,norm}$ was sign-inverted:~~

$$O_{1,norm,adjusted} = - - O_{1,norm}$$

~~The overall fitness was computed as:~~

$$O_{weighted} = w_1\, O_{1,norm,adjusted} + w_2\, O_{2,norm}$$

with $w_1 = 1$ and $w_2 = -1$. In this configuration, DOPE serves as the primary structural criterion, while subtracting MolPDF helps stabilize the optimization trend and reduce noise.

~~To reflect the greater relevance of the DOPE score, we set $w_1 = 1$ and $w_2 = -1$. This configuration improved optimization stability and minimized both scores, as subtracting the MolPdf term smoothed fitness trends and enhanced convergence.~~The choice of retaining the top 20 candidates reflects a balance between diversity and computational tractability. A smaller set would risk premature convergence toward suboptimal solutions, while a larger set would significantly increase computational costs. Twenty candidates ensured sufficient diversity for crossover operations while keeping the simulations computationally efficient.

Top Models Cross-over

Crossover was applied to the 20 top-ranking sequences selected via weighted multi-objective optimization, balancing genetic diversity and computational efficiency. A total of 100 offspring were generated through 50 crossover events, using repeated pairings among the 20 parents. This ensured broad parental contribution and promoted variability. Each offspring inherited traits from two parents, potentially combining beneficial features. This strategy enhanced exploration of the solution space and increased the likelihood of improved DOPE, MolPDF, and structural properties in the resulting models.

Mutation

Following the cross-over operation, diversity was instigated through a mutation process. Each residue had a 0.001 probability of being altered. The 0.001 probability for altering individual residues during the mutation phase is designed to ensure mutations remain conservative, minimizing the risk of detrimental mutations that could significantly impact structural integrity. This probability was chosen to maintain conservative evolutionary pressure. Larger probabilities would likely accelerate destabilizing changes, while smaller ones would reduce exploration efficiency. A mutation rate of 0.001 per residue thus represents a trade-off between stability and the need for gradual exploration of the fitness landscape.

The resulting 100 new sequences underwent modelling, following the procedure outlined in the modelling section, to generate, and save, the best sequence of the current generation. This operation was repeated for 100 generations in the main experiment, with the best sequence from each generation saved as a representative. An extended 1000-generation run was also performed to assess long-term convergence and confirm the stability of the optimization trends.

~~repeated 100 times with the best sequence from each of the 100 generations saved as a representative of that generation with the lowest DOPE and MolPdf score.~~

Gibbs Free Energy Evaluation

We modeled spike–ACE2 complexes for each mutant and estimated ΔG changes in protein–ligand binding using a surface area–based computational method. For each residue, surface atoms were identified based on a reference percentage of solvent-accessible area, capturing the regions most exposed to the environment.

We then calculated the change in accessible surface area (ΔASA) for each residue, reflecting surface differences before and after binding. These ΔASA values were used to estimate binding free energy (ΔG) via the following linear model [21] :

$$\Delta G_{bind} = C + w_{Tyr}\,\Delta ASA_{Tyr} + w_{Ser}\,\Delta ASA_{Ser} + w_{Cys}\,\Delta ASA_{Cys}$$

Here, C and the weights (w_{Tyr}, w_{Ser}, and w_{Cys}) are predefined constants for TYR, SER, and CYS residues, which contribute most significantly to ΔG.

Binding Affinity Evaluation

Another important measure in the context of protein-ligand binding interactions in our work is the bBinding affinity. It is a measure of strength of the interaction between a protein and a ligand and is quantified by the association constant (Ka). Ka is representing the equilibrium between the bounded and unbounded states of a protein-ligand pairs:

$$K_a = \frac{[P][L]}{[PL]}$$

A higher Ka means stronger binding affinity. This implies that at equilibrium, more of the protein-ligand complex ([PL]) is formed relative to the concentrations of free protein ([P]) and free ligand ([L]). Thus, a higher Ka indicates that the binding reaction is more favorable and most of the protein and ligand molecules are in the bound state rather than unbound.

To describe the Ka mathematically, of binding affinity using the gas constant (R), and the temperature (T) in Kelvin [22]:

$$K_a = exp\left(\frac{-\Delta G}{RT}\right)$$

Both the values of ΔG and binding affinity provide us with complementary information for understanding the protein-ligand interactions. Additionally, the agreement between ΔG and binding affinity values serve us in this study to validate the accuracy of the computational methods applied.

Select variants of concern

To identify potential VOCs, we applied a multi-criteria selection strategy based on three complementary metrics: DOPE score, ΔG, and binding affinity. The DOPE score was used as the primary fitness function within the genetic algorithm, enabling the selection of structurally stable and energetically favorable spike protein models. After generation, each model was further evaluated for its binding potential to the ACE2 receptor by calculating ΔG and binding affinity values. This combined assessment allowed us to select the most promising variants those that not only exhibit low energy conformations (suggesting structural realism) but also demonstrate strong and thermodynamically favorable interactions with the host receptor. Based on this integrated analysis, specific generations were selected and introduced as candidate VOCs for further investigation.

Statistical Analysis

To statistically validate improvements during GA optimization, we performed two analyses. First, we compared DOPE scores from the first and last 10 generations using both unpaired t-tests and Mann–Whitney U tests to account for potential non-normality and unequal variances, —common in evolutionary algorithm outputs [23]. Second, we applied the same tests to predicted ΔG and Ka values to evaluate enhancements in spike–ACE2 binding affinity and stability. The consistent results from both parametric and non-parametric tests confirmed the reliability of observed trends.

Although MolPDF is available in MODELLER, it was excluded due to its lack of normalization and dependence on modeling parameters, limiting its comparability. In contrast, DOPE is length-

independent, more widely used in structural benchmarking [18], and better suited for evaluating diverse models. For these reasons, DOPE was chosen as the sole structural metric for validation.

All statistical analyses were conducted using the SciPy library [24], which offers reliable tools for hypothesis testing in Python.

Results

We used our GA workflow to explore the structural impact of mutations in emerging VOCs. Starting from the Wuhan reference spike protein, we introduced random mutations and evaluated their effects using a custom GA framework. The DOPE score served as the fitness function, guiding optimization toward energetically favorable models. Structural models were generated with MODELLER, and selection favored variants with minimized DOPE scores.

In total, 100 generations were initially produced, with each generation retaining the model exhibiting the lowest DOPE score. To further evaluate the structural refinement achieved through the GA, both DOPE and MolPDF scores were tracked across generations and compared to baseline values. This allowed for a comprehensive assessment of model improvement throughout the evolutionary process.~~We used our GA workflow to explore the structural impact of mutations in emerging VOCs. Starting from the Wuhan reference spike protein, we introduced random mutations and evaluated their effects using a custom GA framework. The DOPE score served as the fitness function, guiding optimization toward energetically favorable models. Structural models were generated with MODELLER, and selection favored variants with minimized DOPE scores.~~

~~In total, 100 generations were produced, with each generation retaining the model exhibiting the lowest DOPE score. To further evaluate the structural refinement achieved through the GA, both DOPE and MolPdf scores were tracked across generations and compared to baseline values. This allowed for a comprehensive assessment of model improvement throughout the evolutionary process.~~

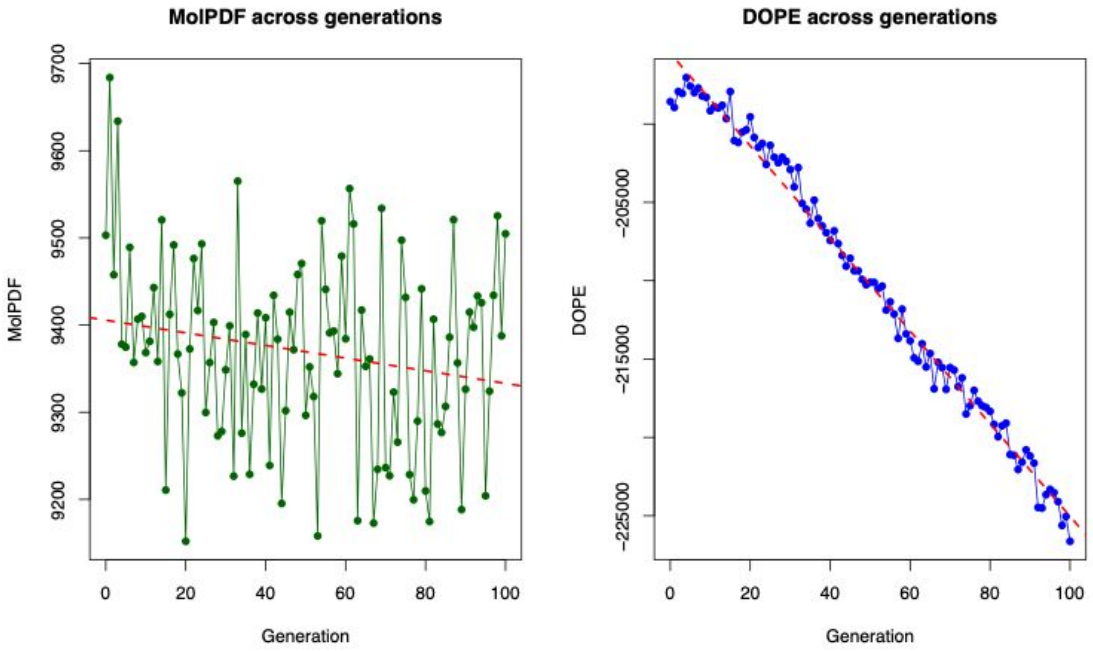

Figure 2 - The left panel shows the MolPDF score trend (green line with points). Although values remain highly variable across generations, the red dashed regression line indicates a slight overall downward tendency, suggesting limited but consistent structural refinement. The right panel shows the DOPE score trend (blue line with points), which exhibits a clear and steady decrease across generations. The red dashed regression line highlights the strong convergence toward progressively more stable and energetically favorable structures. Together, these results demonstrate the effectiveness of the GA-driven optimization in refining spike protein models.

~~The plots illustrate the evolution of structural quality metrics over 100 generations of GA optimization. The top panel shows the DOPE score trend (blue line), which steadily decreases across generations, indicating progressive improvement in model stability compared to the baseline (red dashed line). The bottom panel displays the Molpdf score trend (green line), which, despite some variability, also tends to decrease and remains mostly below the baseline value (purple dashed line), reflecting enhanced structural plausibility. Together, these metrics confirm the effectiveness of the GA-driven refinement process.~~

As shown in Figure 2, DOPE scores steadily decrease across 100 generations, indicating improved model quality and structural stability. Since lower DOPE values correspond to more energetically favorable structures, this trend confirms the GA's effectiveness in selecting better variants. MolPDF values remained highly variable across generations, with fluctuations that make the trend less apparent than for DOPE. Nevertheless, a slight overall downward tendency can be observed, suggesting limited but consistent structural refinement.

Together, these reductions highlight the robustness of the GA-driven optimization process in refining spike protein models through iterative selection.

To further validate the robustness of the GA, we extended the evolutionary process to 1000 generations (Figure 3).

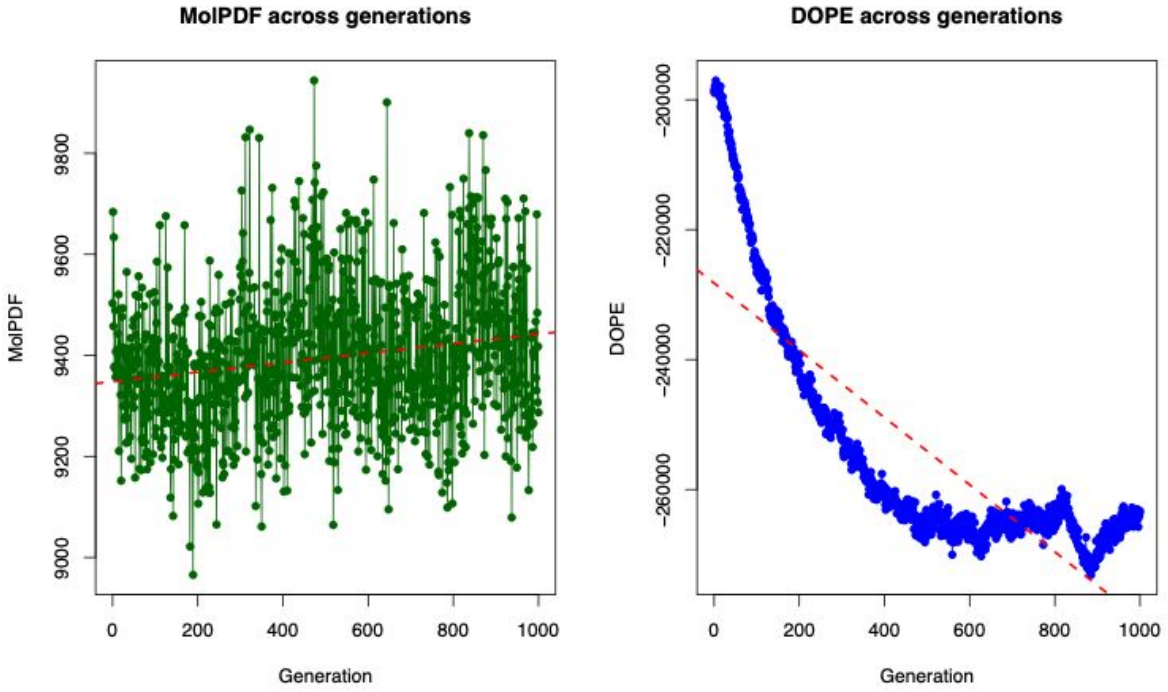

Figure 3 - MolPDF (left panel) and DOPE (right panel) score trends across 1000 generations. The MolPDF values (green line with points) remain highly variable, though the red dashed regression line indicates a subtle overall decrease, consistent with limited structural refinement. DOPE values (blue line with points), instead, display a marked downward trend. Notably, after an initial steady decline, the DOPE curve shows a temporary plateau beyond ~400 generations before resuming its descent, suggesting phases of stabilization followed by renewed optimization toward energetically favorable conformations.

~~As shown in Figure 2, DOPE scores steadily decrease across generations, indicating improved model quality and structural stability. Since lower DOPE values correspond to more energetically favorable structures, this trend confirms the GA's effectiveness in selecting better variants. Although more variable, Molpdf scores also decline overall, suggesting enhanced structural plausibility. Together, these reductions highlight the robustness of the GA-driven optimization process in refining spike protein models through iterative selection.~~

Consistent with the results from the 100-generation experiment, DOPE scores showed a strong overall decrease, reflecting the convergence of the GA toward progressively more stable and energetically favorable spike protein models. Interestingly, after ~400 generations, the DOPE curve exhibited a temporary stabilization phase before resuming its decline, suggesting that the GA may encounter local optima during the search, from which subsequent generations allow escape and further refinement. In contrast, MolPDF values, which had shown a slight decrease in the 100-generation run, displayed higher variability and even a subtle upward trend across 1000 generations. This apparent discrepancy reflects the different sensitivity of the two metrics: MolPDF, as an internal pseudo-energy function of MODELLER, captures improvements during the early refinement phase but becomes less correlated with global structural stability over extended evolutionary runs. DOPE, instead, steadily decreases across both experiments, confirming its role as the most reliable indicator of GA-driven optimization.

Together, the results from both the 100- and 1000-generation experiments demonstrate that the GA framework consistently drives optimization toward lower-energy structural states, thereby confirming

its suitability for modeling the impact of spike protein mutations. These structural improvements were then complemented by the thermodynamic evaluation of spike–ACE2 interactions (ΔG and Ka), providing a comprehensive assessment of both energetic stability and binding affinity across GA-driven evolutionary trajectories.

~~To assess the thermodynamic implications of the mutations introduced by the genetic algorithm (GA), we evaluated two key parameters across all generations: the ΔG and the Ka of the spike–ACE2 interaction.~~ The ΔG values determine whether a process or reaction will occur spontaneously while the Ka is a parameter used to study the binding affinity. Binding affinity typically involves assessing the strength of the interaction between a molecule (ligand) and its target (receptor), indicating how strongly they bind to each other.

Figure ~~3~~4 illustrates the thermodynamic evolution of the spike–ACE2 interaction over 100 GA-optimized generations. While both ΔG and Ka fluctuate, the overall trend suggests a progressive shift toward more favorable binding conditions. ΔG values range from approximately −10.5 to −13.0 kJ/mol, with a general tendency toward more negative values, indicating increasingly spontaneous and thermodynamically favorable interactions. Concurrently, Ka exhibits substantial variability without a consistent trend, though intermittent peaks suggest episodes of enhanced binding affinity. These fluctuations reflect the GA's capacity to explore diverse variants and occasionally achieve improved spike–ACE2 binding properties.

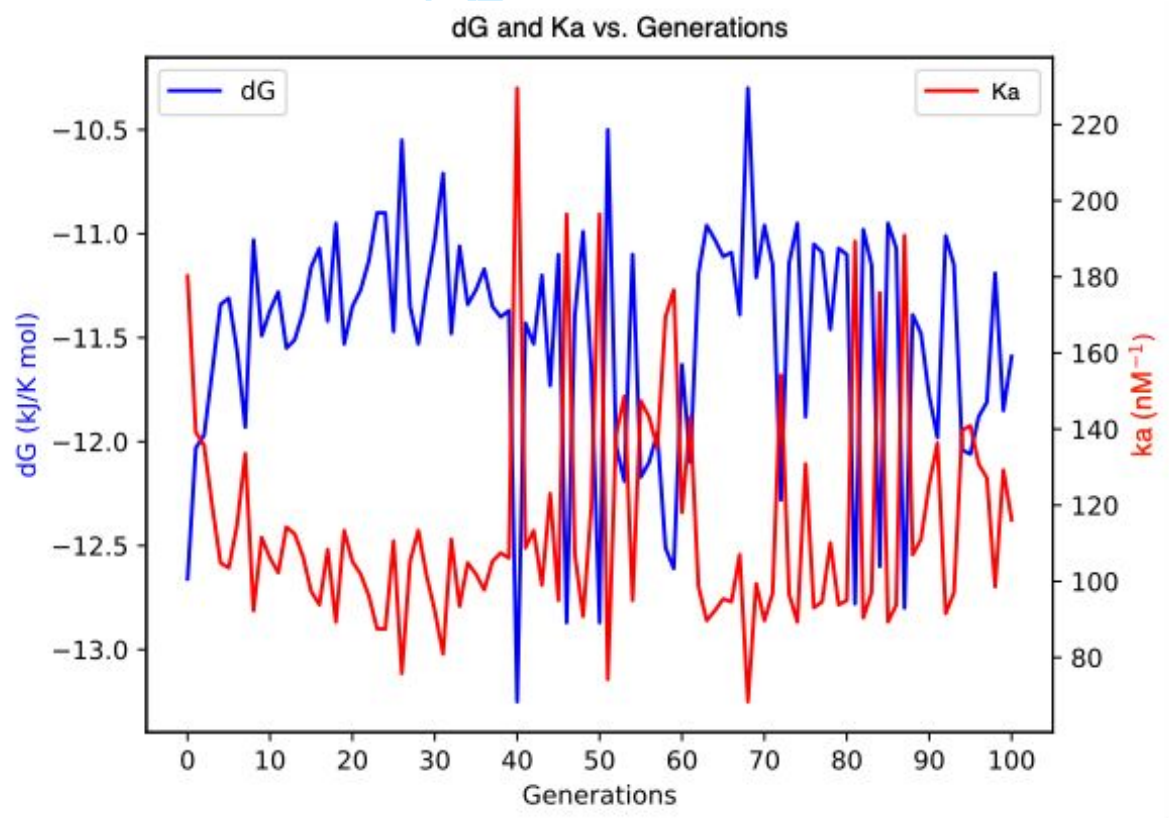

Figure ~~3~~4 - The plot illustrates the thermodynamic optimization of the spike–ACE2 interaction across generations. The ΔG values (blue curve) fluctuate but overall show a trend toward lower free energy, indicating improved binding spontaneity. The Ka values (red curve) also fluctuate, with occasional increases suggesting enhanced binding affinity. These results demonstrate the genetic algorithm's capability to explore the fitness landscape and improve both the thermodynamic stability and functional interaction of spike protein variants.

The trends in ΔG and Ka confirm that the GA enhanced not only structural quality (via DOPE and Molpdf) but also the thermodynamic and functional properties of spike–ACE2 binding. Lower ΔG indicates more favorable binding, while higher Ka reflects stronger affinity. These complementary improvements validate the robustness of our GA-based framework and highlight generation 97 as a promising candidate for a potential new Variant of Concern.

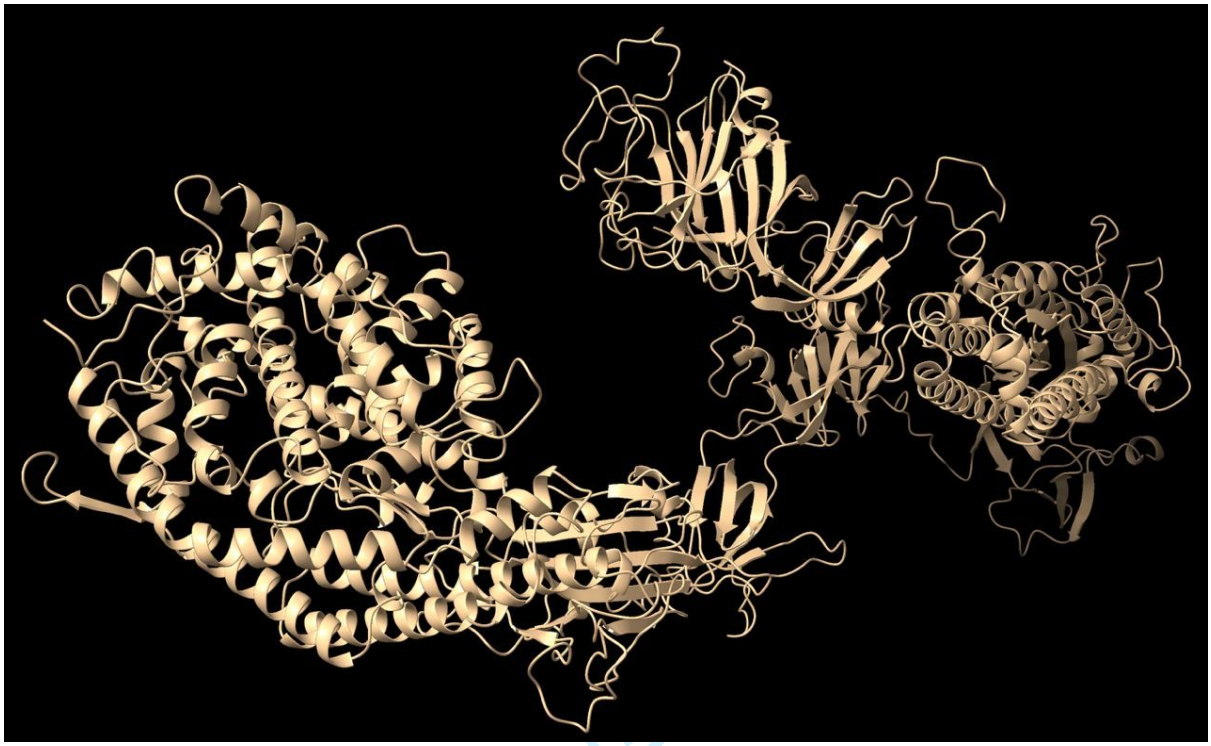

Figure ~~4~~ 5 - 3D structure of spike protein of generation 97 obtained from GA framework.

Generation 97 emerged as the most optimized candidate within our GA framework, displaying the lowest DOPE score and therefore the highest structural stability among the variants generated (Figure 5). It also showed favorable ΔG and binding affinity values (Table 1), supporting its structural plausibility and efficient receptor engagement in silico. While these findings highlight the ability of the GA to identify structurally optimized spike variants, no claim is made regarding their occurrence in circulating lineages or their classification as real Variants of Concern.~~Generation 97 was selected as the leading candidate for a potential new VOC, showing the lowest DOPE score and thus a highly stable 3D structure (Figure 4). It also exhibited favorable ΔG and strong binding affinity (Table 1), indicating efficient host cell interaction. These features suggest a higher likelihood of persistence and transmissibility, underscoring the need for close surveillance and relevance to future vaccine strategies.~~

Generation	DOPE score	ΔG	Binding Affinity
97	-223520.29688	-12.61	176.552271

Table 1 - The table shows the key performance metrics for selected possible new VOC. The columns include the Generation number, DOPE score, ΔG and Binding Affinity.

Statistical analysis

A comprehensive statistical comparison was performed between the first and last ten generations of spike protein models to assess improvements in structural and functional metrics. As shown in Table 2, the DOPE score exhibited a highly significant decrease (t-test $p = 2.36 \times 10^{-16}$; Mann–Whitney $p = 0.000183$), confirming a progressive enhancement in model stability and conformational quality across generations. The median DOPE score decreased by more than 26,000 units between the first (−198,031.51) and last (−224,283.43) ten generations, reflecting a substantial gain in structural stability achieved through GA-driven optimization. This trend aligns with the core objective of the GA, which was designed to iteratively select variants with lower energy profiles, thereby promoting convergence towards structurally optimized models.

Test	T-test p-value	Mann-Whitney p-value
DOPE Score	2.36E-16	0.00018267
ΔG (Gibbs Free Energy)	0.20046886	0.24114463
Ka (Binding Affinity)	0.1992096	0.24114463

Table 2 - Statistical comparison of DOPE score, Gibbs free energy (ΔG), and association constant (Ka) between early (generations 1–10) and late (generations 91–100) spike protein models.

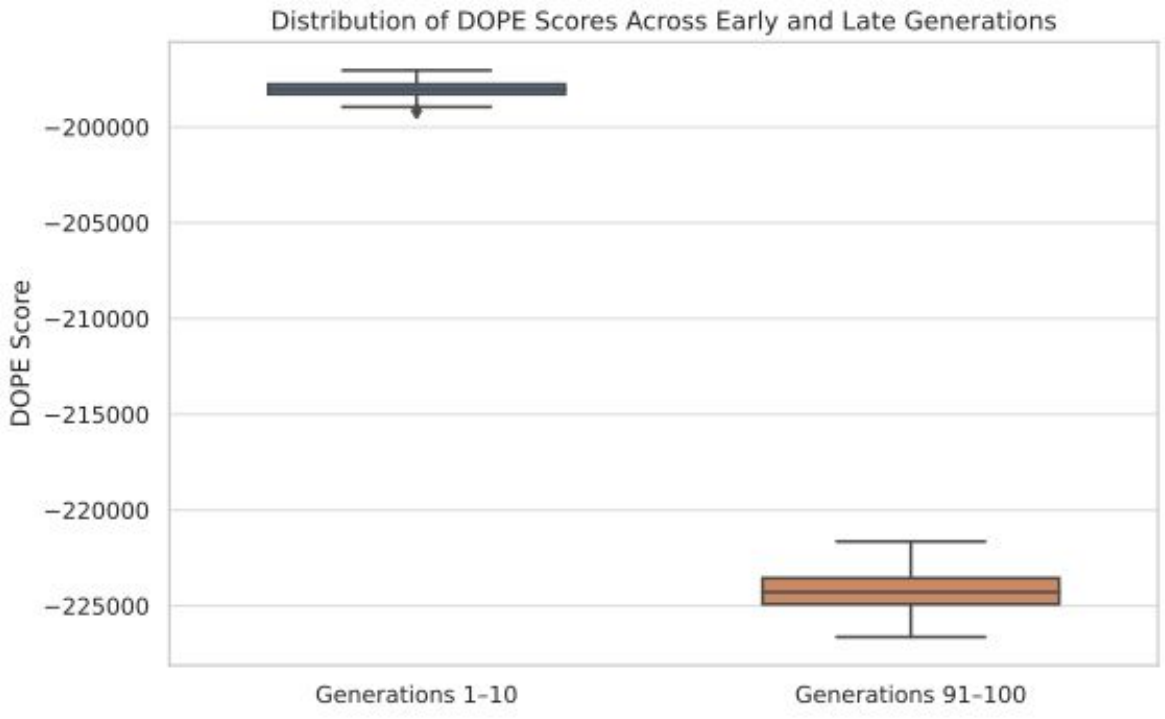

Figure ~~5~~ 6 - Boxplot of DOPE scores for the first and last ten generations. The significant decrease in DOPE values in later generations reflects improved structural stability driven by the genetic algorithm optimization.

In contrast, the ΔG and Ka did not show statistically significant differences between the first and last generations (ΔG p = 0.20; Ka p = 0.20, t-test). These results suggest that, while the GA effectively optimized structural stability, the functional improvements in binding thermodynamics were less pronounced. This interpretation is further supported by the modest differences observed in the mean values: ΔG improved only slightly, from −11.73 to −12.05 kJ/mol, and Ka increased from 125.50 to 143.40 nM^{-1} between the early and late generations. These limited shifts reinforce the notion that, while structural energy minimization was prioritized and successfully achieved, thermodynamic binding properties did not undergo a comparable evolutionary pressure.

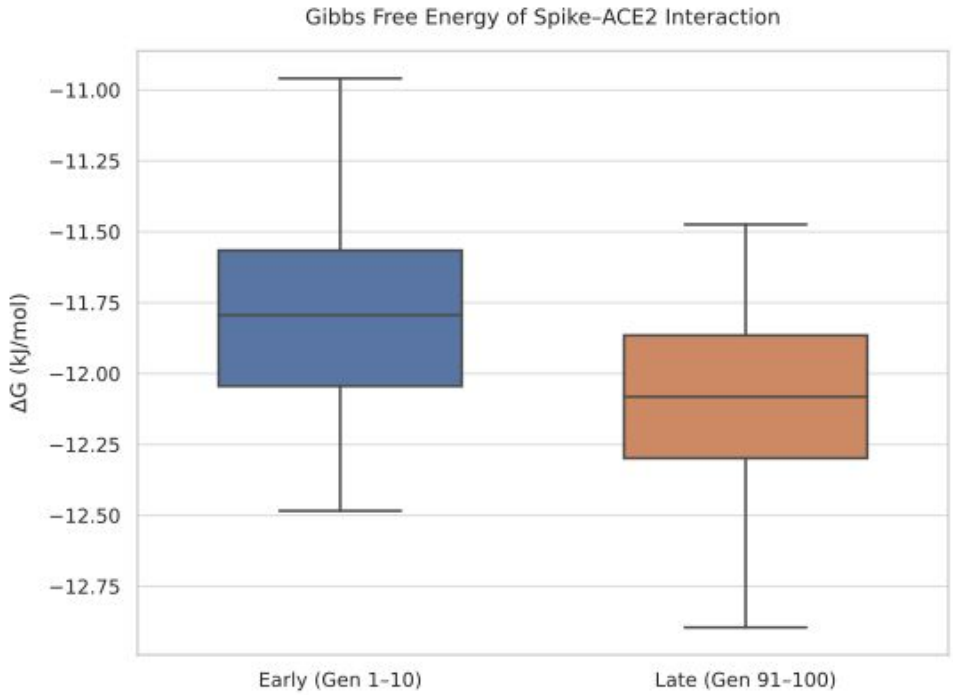

Figure ~~6~~ 7 - ΔG values of spike–ACE2 interaction across early and late generations. No statistically significant changes were observed, suggesting limited functional evolution under structural optimization pressure.

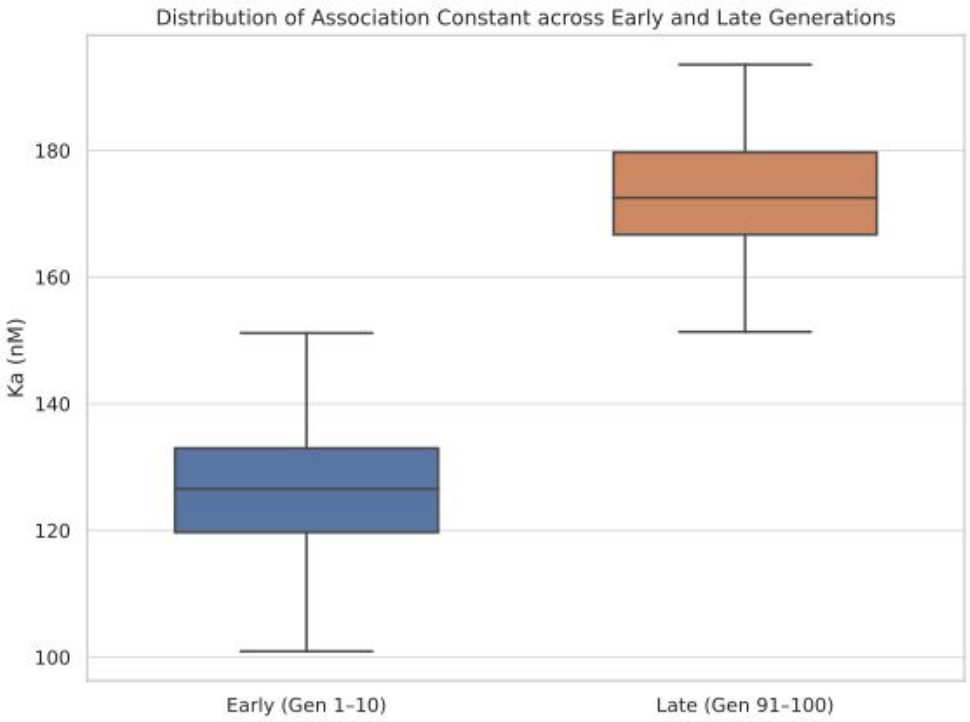

Figure 8 ~~7~~ - Ka distribution across early and late generations. The lack of significant variation indicates that binding affinity was not effectively optimized in the absence of direct selective pressure.

A possible explanation lies in the multi-objective nature of protein-ligand interaction optimization. ΔG and Ka are influenced by subtle, non-linear conformational and surface changes that may not evolve in parallel with structural compactness or energy minimization [25]. Moreover, as ΔG and Ka were not directly included in the GA's fitness function but evaluated a posteriori, their evolution was not subject to selective pressure. This decoupling may explain the lack of statistically significant improvement in these functional indicators, despite the evident refinement of the structural model itself.

Together, these statistical results validate the effectiveness of the GA-driven modeling and selection strategy in optimizing structural features of spike protein variants. While functional metrics such as ΔG and Ka did not show significant improvements, the marked enhancement in structural stability supports the utility of this approach for generating conformationally optimized candidates with potential biological relevance.

Discussion

The use of Genetic Algorithms (GAs) guided by fitness metrics such as the DOPE score provides a flexible and powerful computational framework for exploring the mutational landscape of the SARS-CoV-2 spike protein. Post hoc analyses of Gibbs Free Energy and binding affinity offer complementary insights into structural stability and functional relevance, particularly in the context of spike–ACE2 interactions. While ΔG informs on thermodynamic feasibility, binding affinity reflects the strength of receptor engagement. Together with structural optimization, these measures help characterize the potential impact of mutations at the molecular level.

In our study, the DOPE score was employed as the main fitness function, and we consistently observed a progressive decrease across generations, suggesting convergence toward structurally more stable and energetically favorable models. Within the initial 100 generations explored, generation 97 exhibited the lowest DOPE score and represented a point of convergence in the optimization process. Extending the evolutionary process to 1000 generations further confirmed this trend, with DOPE scores continuing to decline and showing strong convergence overall. Interestingly, after approximately 400 generations, the DOPE curve entered a temporary plateau before resuming its descent, suggesting that the GA may encounter local optima during the search but is ultimately able to escape and refine further. This extended experiment reinforces the robustness of the GA-driven optimization framework and highlights its potential for long-term convergence.

MolPDF values, in contrast, displayed more variable behavior. While a slight decrease was observed in the first 100 generations, the extended 1000-generation run revealed higher variability and even a modest upward trend. This apparent discrepancy can be explained by the different sensitivity of the two metrics: MolPDF, as an internal pseudo-energy function of MODELLER, captures improvements during the early refinement phase but becomes less correlated with global structural stability over extended evolutionary timescales. Thus, while MolPDF remains informative, DOPE provides a more reliable and consistent measure of convergence and model quality in our GA framework.

Although ΔG and binding affinity did not exhibit statistically significant changes across early and late generations, they provided valuable additional context. The observation that structural stability improved more consistently than binding properties suggests that different selective pressures may govern these features and highlights the importance of multi-objective optimization in future applications. Importantly, combining binding-related parameters with structural stability metrics ensures that candidate variants are not only energetically favorable but also biologically plausible.

Several limitations of the current work should be acknowledged. The GA was originally restricted to 100 generations due to computational constraints, which may have limited the full exploration of the mutational space. The additional 1000-generation experiment addresses this limitation, strengthening our conclusions, but further scaling up of both population size and evolutionary depth will be important for future applications. Moreover, the study focused exclusively on spike protein mutations, without considering other viral proteins or host–pathogen interactions. Functional metrics such as ΔG and binding affinity, while informative, may not fully capture the complexity of viral behavior in vivo.

Despite these limitations, the study provides a proof-of-concept for applying GA to structural modeling of viral proteins. This methodological framework can serve as a foundation for future research integrating additional fitness functions, larger simulation scales, and more diverse mutational scenarios. Crucially, experimental validation of top candidates (e.g., generation 97) through in vitro assays, as well as benchmarking against genomic surveillance data, will be essential to assess predictive potential. Integration with phylogenetic models, deep learning approaches, and epidemiological studies could further enhance the utility of this approach.

In addition, while Generation 97 emerged as the most optimized candidate in our GA framework, we did not compare its mutations with those observed in circulating VOCs nor assess whether they fall within functionally relevant domains (e.g., receptor-binding domain, N-terminal domain, furin cleavage site). This remains an important limitation of the present study, and a priority for future work integrating genomic surveillance data and structural domain mapping.

Overall, this work demonstrates the promise of GA-based structural modeling as a methodological tool for systematically exploring protein mutations. While not intended to forecast specific Variants of Concern, it provides a solid computational basis that, once combined with evolutionary and experimental evidence, may contribute to more comprehensive strategies for monitoring viral evolution and informing vaccine design.~~Using a GA guided by fitness metrics like **DOPE score** and **MolPDF** provides a robust framework for predicting new VOCs. Post-GA analyses—**ΔG** and **binding affinity**—offer insights into structural stability and functional relevance, particularly regarding spike–ACE2 interactions. ΔG informs on thermodynamic stability, while binding affinity reflects potential for host receptor engagement. This integrated approach combines **structural optimization** and **functional evaluation**, supporting the characterization of emerging variants and guiding therapeutic development and response strategies.~~

~~We employed GA using the DOPE score as the fitness function to guide the selection of potential new VOCs. We hypothesized that the DOPE score would progressively decrease with each generation, particularly in the later stages, eventually reaching a plateau as the algorithm converged toward optimal solutions. However, due to computational constraints, the GA was limited to 100 generations. Within this range, generation 97 exhibited the lowest DOPE score at the point where the plot began to stabilize, suggesting a reliable convergence trend.~~

~~The DOPE score evaluates the total energy and structural stability of the protein model, providing insight into the likelihood that a given conformation could exist in nature. Lower DOPE scores are generally associated with more stable and biologically plausible protein structures, as they reflect conformations with minimal energy.~~

~~To complement structural analysis, we evaluated the binding affinity and ΔG of spike–ACE2 interactions, which reflect interaction strength and thermodynamic feasibility. While low ΔG and high affinity suggest functional relevance, structural stability—indicated by a low DOPE score—remains essential. Relying solely on binding metrics may favor strong interactions but overlook structural instability. Thus, an optimal candidate must balance both strong binding and a low DOPE score to ensure biological plausibility.~~

~~Given these considerations, prioritizing the lowest DOPE score remains a critical factor, as it reflects how closely the predicted 3D structure of the spike protein aligns with biologically plausible conformations and its energetic suitability for natural existence. While binding affinity and ΔG offer valuable insights into the strength and feasibility of spike–ACE2 interactions, they must be interpreted in conjunction with structural stability metrics like the DOPE score. In future iterations of the GA-based approach, the MolPDF energy metric will also be incorporated to further enhance model evaluation.~~

~~This study has several limitations. The GA was restricted to 100 generations due to computational constraints, potentially limiting optimization. Only spike protein mutations were analyzed, excluding other viral components and interactions. External factors like environment, human behavior, and public health responses were not modeled. Finally, while ΔG and binding affinity offer valuable insights, they may not fully represent the in vivo behavior of viral variants.~~

~~Future work should include experimental validation of top variants (e.g., generation 97) via in vitro assays to assess ACE2 binding and replication. Genomic surveillance could help compare predicted mutations with those in emerging VOCs. Cross-validation using deep learning or phylogenetic methods would strengthen prediction reliability. Lastly, longitudinal studies tracking variant prevalence across populations would support real-world applicability.~~

Conclusion

As the COVID-19 pandemic continues and novel variants emerge, computational methods remain essential for understanding the molecular mechanisms that drive viral evolution. Genetic Algorithms (GAs), inspired by natural selection, provide a versatile framework to explore mutational landscapes and refine structural models of viral proteins.

In this study, we implemented a GA-based approach to optimize SARS-CoV-2 spike protein variants using energy-based metrics such as the DOPE score, Gibbs Free Energy (ΔG), and binding affinity. The initial 100-generation run highlighted progressively more stable conformations, with generation 97 representing the most optimized candidate within the explored search space. Extending the GA to 1000 generations confirmed the robustness of this framework, as DOPE scores continued to decrease and revealed long-term convergence patterns, including transient plateaus followed by renewed optimization. These findings illustrate the ability of GAs to iteratively refine structural plausibility and energetic favorability over both short and extended evolutionary timescales.

While the present work does not aim to predict specific real-world Variants of Concern, it demonstrates the potential of GA-driven modeling as a methodological tool to investigate mutational effects at the structural level. Importantly, the approach establishes a computational foundation that could, in future, be combined with genomic surveillance, phylogenetics, and experimental validation to strengthen predictive insights.

Future research should build on this framework by expanding fitness functions, adopting multi-objective optimization strategies, and scaling up both the number of generations and the population sizes. Such developments may enhance our ability to link structural modeling with evolutionary dynamics, ultimately supporting more comprehensive strategies for monitoring viral evolution and informing vaccine design.~~As the COVID-19 pandemic continues and new Variants of Concern (VOCs) emerge, detecting even minor mutations that drive viral evolution is crucial. Genetic Algorithms (GAs), inspired by natural selection, offer a powerful framework for tackling this challenge.~~

~~In this study, we employed a GA-based approach to predict structurally and functionally relevant SARS-CoV-2 spike protein variants over 100 generations, using energy-based metrics such as DOPE score,~~

~~ΔG, and binding affinity. Generation 97 stood out as the most promising candidate, exhibiting the lowest DOPE score —indicating high structural stability —alongside favorable ΔG and binding affinity values, supporting its biological plausibility.~~

~~These results demonstrate the potential of GA-driven modeling for early identification of emerging variants, with implications for vaccine development and pandemic preparedness. However, such predictions depend on input data quality and focus on molecular properties, without accounting for broader factors like host population dynamics or public health interventions.~~

~~A comprehensive strategy—integrating AI, phylogenetics, epidemiology, and experimental validation—is essential for robust VOC forecasting. Future work will extend this framework by incorporating additional fitness functions and broadening its predictive capacity, aiming to anticipate high-risk mutations before they spread widely.~~

List of Abbreviations

VOC: Variant of concern

S: Spike protein

ΔG: Gibbs Free energy

GA: Genetic Algorithm

Ka: Association Constant

ACE2: Angiotensin-converting enzyme 2

~~TS: Tabu Search~~

~~MA: Microcanonical Annealing~~

~~GLS: Guided Local Search~~

~~Genetic Algorithm: GA~~

~~PSO: Particle Swarm Optimization~~

~~ACO: Ant Colony Optimization~~

~~SHO: Spotted Hyena Optimizer:~~

~~EPO: Emperor Penguin Optimizer~~

~~SOA: Seagull Optimization~~

References

[1] S. P. Otto *et al.*, "The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic," *Curr. Biol.*, vol. 31, no. 14, pp. R918–R929, July 2021, doi: 10.1016/j.cub.2021.06.049.

[2] R. Sanjuán and P. Domingo-Calap, "Mechanisms of viral mutation," *Cell. Mol. Life Sci.*, vol. 73, no. 23, pp. 4433–4448, Dec. 2016, doi: 10.1007/s00018-016-2299-6.

[3] R. Sanjuán, M. R. Nebot, N. Chirico, L. M. Mansky, and R. Belshaw, "Viral Mutation Rates," *J. Virol.*, vol. 84, no. 19, pp. 9733–9748, Oct. 2010, doi: 10.1128/JVI.00694-10.

[4] L. Ballesteros-Sanabria, H. F. Pelaez-Prestel, A. Ras-Carmona, and P. A. Reche, "Resilience of Spike-Specific Immunity Induced by COVID-19 Vaccines against SARS-CoV-2 Variants," *Biomedicines*, vol. 10, no. 5, p. 996, Apr. 2022, doi: 10.3390/biomedicines10050996.

[5] W. T. Harvey *et al.*, "SARS-CoV-2 variants, spike mutations and immune escape," *Nat. Rev. Microbiol.*, vol. 19, no. 7, pp. 409–424, July 2021, doi: 10.1038/s41579-021-00573-0.

[6] E. A. Rashed, S. Kodera, and A. Hirata, "COVID-19 forecasting using new viral variants and vaccination effectiveness models," *Comput. Biol. Med.*, vol. 149, p. 105986, Oct. 2022, doi: 10.1016/j.compbiomed.2022.105986.

[7] E. A. Rashed, S. Kodera, H. Shirakami, R. Kawaguchi, K. Watanabe, and A. Hirata, "Knowledge discovery from emergency ambulance dispatch during COVID-19: A case study of Nagoya City, Japan," *J. Biomed. Inform.*, vol. 117, p. 103743, May 2021, doi: 10.1016/j.jbi.2021.103743.

[8] L. Xu, R. Magar, and A. Barati Farimani, "Forecasting COVID-19 new cases using deep learning methods," *Comput. Biol. Med.*, vol. 144, p. 105342, May 2022, doi: 10.1016/j.compbiomed.2022.105342.

[9] C. Bai *et al.*, "Predicting Mutational Effects on Receptor Binding of the Spike Protein of SARS-CoV-2 Variants," *J. Am. Chem. Soc.*, vol. 143, no. 42, pp. 17646–17654, Oct. 2021, doi: 10.1021/jacs.1c07965.

[10] Z. Michalewicz and M. Schoenauer, "Evolutionary Algorithms for Constrained Parameter Optimization Problems," *Evol. Comput.*, vol. 4, no. 1, pp. 1–32, Mar. 1996, doi: 10.1162/evco.1996.4.1.1.

[11] J. H. Holland, "Genetic Algorithms," *Sci. Am.*, vol. 267, no. 1, pp. 66–73, 1992.

[12] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimed. Tools Appl.*, vol. 80, no. 5, pp. 8091–8126, Feb. 2021, doi: 10.1007/s11042-020-10139-6.

[13] A. A. Abou El Ela and S. R. Spea, "Optimal corrective actions for power systems using multi-objective genetic algorithms," *Electr. Power Syst. Res.*, vol. 79, no. 5, pp. 722–733, May 2009, doi: 10.1016/j.epsr.2008.10.007.

[14] Z. Liao, Y. Song, S. Ren, X. Song, X. Fan, and Z. Liao, "VOC-DL: Deep learning prediction model for COVID-19 based on VOC virus variants," *Comput. Methods Programs Biomed.*, vol. 224, p. 106981, Sept. 2022, doi: 10.1016/j.cmpb.2022.106981.

[15] B. Webb and A. Sali, "Protein Structure Modeling with MODELLER," in *Structural Genomics*, vol. 2199, Y. W. Chen and C.-P. B. Yiu, Eds., in Methods in Molecular Biology, vol. 2199. , New York, NY: Springer US, 2021, pp. 239–255. doi: 10.1007/978-1-0716-0892-0_14.

[16] B. Webb and A. Sali, "Comparative Protein Structure Modeling Using MODELLER," *Curr. Protoc. Bioinforma.*, vol. 54, no. 1, June 2016, doi: 10.1002/cpbi.3.

[17] N. Eswar *et al.*, "Comparative Protein Structure Modeling Using Modeller," *Curr. Protoc. Bioinforma.*, vol. 15, no. 1, Sept. 2006, doi: 10.1002/0471250953.bi0506s15.

[18] A. Šali and T. L. Blundell, "Comparative Protein Modelling by Satisfaction of Spatial Restraints," *J. Mol. Biol.*, vol. 234, no. 3, pp. 779–815, Dec. 1993, doi: 10.1006/jmbi.1993.1626.

[19] D. Eramian, M. Shen, D. Devos, F. Melo, A. Sali, and M. A. Marti-Renom, "A composite score for predicting errors in protein structure models," *Protein Sci.*, vol. 15, no. 7, pp. 1653–1666, July 2006, doi: 10.1110/ps.062095806.

[20] M. Shen and A. Sali, "Statistical potential for assessment and prediction of protein structures," *Protein Sci.*, vol. 15, no. 11, pp. 2507–2524, Nov. 2006, doi: 10.1110/ps.062416606.
[21] A. Schön, N. Madani, A. B. Smith, J. M. Lalonde, and E. Freire, "Some Binding-Related Drug Properties are Dependent on Thermodynamic Signature: Optimization of Drug Candidates," *Chem. Biol. Drug Des.*, vol. 77, no. 3, pp. 161–165, Mar. 2011, doi: 10.1111/j.1747-0285.2010.01075.x.
[22] P. L. Kastritis and A. M. J. J. Bonvin, "On the binding affinity of macromolecular interactions: daring to ask why proteins interact," *J. R. Soc. Interface*, vol. 10, no. 79, p. 20120835, Feb. 2013, doi: 10.1098/rsif.2012.0835.
[23] N. Nachar, "The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution," *Tutor. Quant. Methods Psychol.*, vol. 4, no. 1, pp. 13–20, Mar. 2008, doi: 10.20982/tqmp.04.1.p013.
[24] P. Virtanen *et al.*, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nat. Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.
[25] M. Van Den Noort, M. De Boer, and B. Poolman, "Stability of Ligand-induced Protein Conformation Influences Affinity in Maltose-binding Protein," *J. Mol. Biol.*, vol. 433, no. 15, p. 167036, July 2021, doi: 10.1016/j.jmb.2021.167036.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Software and Data Availability Statement

Datasets and software used in the experiments are listed as follows:

~~Availability of data and materials~~

~~Datasets used in the experiments are listed as follows:~~

(1) NCBI: National Center for Biotechnology Information (National Center for Biotechnology Information (nih.gov)),(2) RCSB Protein Data Bank: RCSB PDB: Homepage (3) MODELLER: About MODELLER (salilab.org)

Acknowledgements

Not applicable

Competing interests

Not applicable.

Funding

Programma di ricerca CN00000013 “National Centre for HPC, Big Data and Quantum Computing”, finanziato dal Decreto Direttoriale di concessione del finanziamento n.1031 del 17.06.2022 a valere sulle risorse del PNRR MUR – M4C2 – Investimento 1.4 - Avviso “Centri Nazionali” - D.D. n. 3138 del 16 dicembre 2021

Authors' contributions

Valentina Di Salvatore: Writing - review & editing, statistical analysis. Avisa Maleki: Designed the work pipeline, collect data, writing original draft and review & editing Babak Mohajer: Algorithm developer and writing python scripts of GA, writing method part. Alvaro Ras-Carmona: Writing the first Python script. Giulia Russo: Writing - review & editing, formal analysis. Pedro Antonio Reche: Writing - review & editing. Francesco Pappalardo: Conceptualization, Supervision, Methodology.

Dr. Valentina Di Salvatore

Dept. of Drug and Health Sciences

University of Catania, Italy

Catania, August 29, 2025

Dear Editor,

Thank you for sending us the reviewers' comments on our manuscript entitled "*Anticipating Viral Evolution: Genetic Algorithm-Driven Insights into SARS-CoV-2 Spike Mutations*".

We appreciate the work and time of the referees in reviewing our manuscript as well as their positive criticisms. Here we provided a revised version of the manuscript that has been modified to address all the issues raised by reviewers. Throughout the manuscript, we have provided both new text and corrections (highlighted through the Word track changes machinery).

We have added new comments and paragraphs to meet the requests coming from the referees and Editorial Office.

A point-by-point answer is provided down below. We trust we have adequately addressed the concerns of the referees and will look for a positive response.

Best Regards,

Valentina Di Salvatore, on behalf of all Authors.

Reviewer Comments to Authors:
Reviewer: 1

Comments to the Author
In general, the article appears well written and I recommend accepting subject to optional minor revisions as follows:-

Page 1) In the results section, it would be helpful to provide the official gene symbol (ACE2) alongside the name so that it is known prior to first use of that symbol later in the manuscript.

GA is referred to in the conclusions and it would reduce any ambiguity if this acronym is given in the conclusions part of the background.

Answer:

We thank the reviewer for the suggestions. We have added the official gene symbol “ACE2” and defined “Genetic Algorithm (GA)” at first mention in the abstract.

Page 2) It would aid readability to include commas in the second sentence, as follows:-
….depends on mutation rates and their effects within, and between, hosts [2].....

Answer:

We thank the reviewer for the suggestions. Commas added.

Page 4) The workflow figure appears unnecessarily complicated. In particular, I don’t expect to see a series of steps repeated within a workflow in a copy/paste fashion any more than I expect to see computer code copied/pasted. The steps to run the modeller, calculate the scores and sort the dataframe are repeated.

I suggest that after the first occurrence of sorting the dataframe, save the best sequences and for the subsequent step, insert the condition to check the number of generations. After creating 100 child sequences I would increment the generation and loop back to running the modeller.

Answer:

We thank the reviewer for this thoughtful comment. We note that the apparent repetition in the workflow figure is intentional, as the two Modeller runs represent different stages of the genetic algorithm. The first run is applied to the initial mutated population, while the second run is applied after crossover to evaluate the newly generated child sequences. Although the steps look similar, they serve distinct roles in the pipeline, and we therefore kept them explicit in the figure to maintain clarity for readers who may wish to reproduce the workflow.

Page 6) The two metrics O_1 and O_2 use the same normalization, perhaps this can be more eloquently described by a single formula which includes an index such as i whereby i=1,2.

There is a square box underneath the normalisation formula, that appears to be a character which did not display correctly and is presumably a symbol which is intended to indicate normalisation.

I suggest not using the term 'adjusted' to indicate that you are taking the negative but rather simply show that the sign has been changed in the weighted formula, that is, $O_{weighted} = w_2O_2,norm - w_1O_1,norm$

I'm not sure I understand the weighting, the manuscript says that the DOPE score is more relevant, therefore, I would expect to see W_1 with a different and higher absolute value than W_2. The weighting given suggests both scores contributing equally to the overall fitness. If we set $W_1 = 1$, and $W_2 = -1$, I believe that this gives us:-
$O_{weighted} = -O_{2,norm} - O_{1,norm}$ which could also be represented as $O_{weighted} = -(O_{2,norm} + O_{1,norm})$

Answer:

The following changes have been made:

1. Normalization formulas: We revised the description of the normalization to use a single general formula with an index i = 1,2, making the explanation more concise and coherent.

2. Display issue: The undesired square character under the normalization formula has been removed, and the correct symbol for normalization is now used.

3. Terminology ("adjusted"): We removed the word "adjusted" and instead explicitly indicated the sign inversion in the weighted formula.

4. Clarification of weighting: We clarified that setting $w_1 = 1$ and $w_2 = –1$ emphasizes DOPE as the primary criterion, while subtracting MolPDF acts as a stabilizing term to smooth the trend. The text now explains more clearly why both contributions appear balanced mathematically, while conceptually DOPE remains the principal driver.

At the bottom of page 6 I would introduce a couple of commas to aid readability:-
…in the modelling section, to generate, and save, the best sequence of the current generation.

Answer:

We added commas accordingly for readability.

Page 7) I would use a lowercase ‘b’ in ‘binding affinity’ in the first sentence of the ‘Binding Affinity Evaluation’ section.

Answer:

We have revised the text to use lowercase “b” in “binding affinity” in the first sentence of the Binding Affinity Evaluation section.

Page 8) A comma appears twice in the final sentence of the ‘Statistical Analysis’ section.

The first sentence of the Results section should start with ‘We used our GA’ or ‘We used a GA’.

Answer:

We have corrected the punctuation in the final sentence of the Statistical Analysis section, and we have revised the first sentence of the Results section to begin with “We used our GA workflow,” as suggested.

Page 9) I would be interested in seeing a line of best fit for the molpdf score as the values appear quite variable and it’s not not immediately obvious to me that they do decline overall as the manuscript suggests, particularly since the first and final scores are both very similar, and close to 9500.

Answer:

We have updated Figure 2 to include a line of best fit for the MolPDF and DOPE score values. This addition makes it easier to visualize the overall trend across generations. As suggested, we also emphasize in the text that, while a slight downward tendency can be observed, MolPDF scores remain highly variable and the differences between the first and final generations are modest compared to the DOPE score. This variability is now explicitly acknowledged in the Discussion as a limitation of using MolPDF as a stability metric.

Page 15) This work looks at protein rather than nucleotide sequences and I would make that clear in the conclusion by saying ‘SARS-CoV2 spike protein variants’

Answer:

We added “protein” in ‘SARS-CoV2 spike variants.

For future work, at a simple level, it would be interesting to know the number of amino acid changes that there are between the protein sequences found in this study and the original wildtype, as well as, subsequent variants of concern which emerged naturally. Aligning the sequences and counting amino acid changes could give an indication of the solution space which the genetic algorithm is exploring and whether there should be any changes to the parameters, for example, more/fewer generations and/or more/fewer mutations and/or crossovers at each generation.

Answer:

We greatly appreciate your insightful suggestion. In future work, we will consider aligning the generated sequences with the wildtype and naturally emerged variants to quantify amino acid changes. This approach will provide a clearer understanding of the solution space explored by the genetic algorithm and support more informed adjustments to parameters such as mutations, crossovers, and generations. At the same time, we recognize that in nature the emergence of new variants or even new pandemics is driven by a wide range of biological and environmental factors. Therefore, based on your suggestion and by including these factors, we plan to improve and extend our framework.

Reviewer: 2

Comments to the Author
This study presents an interesting application of a genetic algorithm (GA) to simulated SARS-CoV-2 sequences. The methods are articulated clearly, and the paper is generally well-written, but I feel that as written there are insubstantial results given the title and focus of the paper.

The title and abstract promise an ability to "anticipate viral evolution" and predict new Variants of Concern (VOCs), but the study does not demonstrate this. While the approach is technically interesting, the conclusions are overstated. The authors present a model that optimizes a fitness function based on DOPE score and molpdf, but do not connect this with real-world viral evolution in a satisfactory way.
Particularly, there are no experiments comparing the model to real SARS-CoV-2 evolutionary data.

I suggest rejection or major revisions, and to either re-frame the paper’s focus toward the algorithmic/mathematical results, or to add substantial experiments that demonstrate the model’s utility on real pathogen data.

Answer:

We thank the reviewer for their careful reading of our work and for the constructive feedback. We acknowledge that in the original version of the manuscript the framing of the study could have suggested predictive capabilities beyond the scope of our computational analysis. Following the reviewer's recommendations, we have substantially revised the manuscript to:

- Reframe the title and abstract to emphasize the methodological and structural modeling contributions rather than the prediction of real-world Variants of Concern (VOCs).
- Tone down conclusions by clarifying that our GA framework provides in silico structural optimization of spike variants, without claiming predictive forecasting of viral evolution.
- Highlight limitations and specify that no comparison with real SARS-CoV-2 evolutionary data has been performed. We explicitly mention that validation against experimental or surveillance data will be required in future studies. At this stage, we re-frame the paper to highlight the algorithmic contribution rather than present extensive experimental validation, since the model does not yet capture many biological and environmental factors that drive the emergence of new variants or pandemics. Before moving to large-scale validation on real pathogen data, we believe it is essential to expand the framework step by step by incorporating such factors (e.g., structural constraints, host immunity, glycosylation, epidemiological drivers).

We believe these changes now present a more balanced and accurate representation of the contribution of our work.

Specific comments

The study identifies "Generation 97" as its top candidate for a potential new VOC. Does Generation 97 contain mutations seen in real VOCs? Are its mutations located in functionally relevant domains? Without this analysis, the result is a purely theoretical construct.

Remove claims like: "...highlight generation 97 as a promising candidate for a potential new Variant of Concern" unless substantial experiments and results on real data are added.

Answer:
We agree with the reviewer that presenting *Generation 97* as a "potential VOC" was misleading.

- We have now rephrased this throughout the Results and Discussion sections. *Generation 97* is described as the "most optimized candidate within the GA framework," based on structural stability (lowest DOPE score) and favorable thermodynamic properties.
- We explicitly state that this candidate should not be interpreted as a real-world variant prediction.

- We also acknowledge in the revised Discussion that we did not compare the mutations present in Generation 97 with those found in circulating VOCs, nor did we analyze whether they map to known functional domains. This is now highlighted as a limitation and an important direction for future work

In the Introduction:

"...it could not anticipate novel mutations or full lineage evolution. Existing models, including machine learning and logistic regression, often fail to account for the structural effects of mutations on protein–ligand interactions—key factors in viral transmission and immune evasion [9].

Our GA framework addresses this gap by integrating structural modeling with evolutionary dynamics to predict mutations that improve viral fitness"

This setup frames the authors' model as capable of anticipating novel mutations / lineages, but the results do not demonstrate this outside of simulated lineages.

Answer:

The reviewer is correct that the original text suggested that our framework could *"predict mutations that improve viral fitness."* We have revised this section to clarify that the contribution of the present study is **methodological,** demonstrating how a GA can be applied to structural optimization in silico, rather than predictive of real viral lineages. The Introduction now explicitly states that our approach generates optimized spike variants *within simulated lineages only*.

100 experiments seems like a very small number, and restricting to the spike region may limit the method's utility.

Answer:

We acknowledge that 100 generations represent a limited search space. This choice was initially imposed by computational constraints (runtime and resource availability), as now clarified in the revised Methods. However, in response to this concern, we extended the GA optimization to 1000 generations, obtaining results consistent with but stronger than those observed in the 100-generation experiment. In particular, DOPE scores continued to decrease steadily, confirming the convergence of the GA toward progressively more stable and energetically favorable structures, while MolPDF values highlighted the expected variability over extended runs. These additional results, now reported in the revised Results, Discussion and Conclusion and shown in the new figure, provide further support for the robustness of our

framework. We also emphasize in the Discussion that future applications will benefit from systematically scaling up both the number of generations and the population size to further enhance the exploration of mutational space beyond the spike region.

Throughout, parameter choices should be explained and justified thoroughly.

Answer:

We have expanded the Methods section to provide stronger justification for parameter settings. These justifications are now explicitly included in the revised Methods.

Reviewer: 3

Comments to the Author
Review on manuscript titled "Anticipating Viral Evolution: Genetic Algorithm-Driven Insights into SARS-CoV-2 Spike Mutations"

1) General Assessment of the Work
The manuscript addresses significant topics in the study of Viral Evolution. It underscores the genetic changes that occur in viruses like SARS-CoV-2, in particular on Spike proteins. The framework provided is reproducible and it provides a guide on sequence simulation with mutation and recombination. It also integrates the modeling of 3-D protein structure from these simulated sequences with mutation and recombination. This is a good study design, and it provides an alternative to the other computational frameworks and tools used to simulate and model genetic changes in protein structure for forecasting Variants of Concern, and an ability to provide protein structure designs that can be used in Vaccine development. The MODELLER bioinformatics tool is widely used in protein structure modeling, and it is a good choice and an alternative to Alphafold2. While the amount of data used is not a lot and cannot ascertain the accuracy or performance of protein structure modeling and prediction of Variants of Concern, the framework allows extrapolation and addition of more data. As stated in the manuscript, a comparative analysis will need to be done for future studies to compared simulated VOCs with empirical VOCs observed in the environment. If the simulated VOCs are reflective of genetic changes occurring in the environment, then this framework is important for predicting and forecasting potential future variants of concern. Overall, I am satisfied with the study and manuscript, but I would request for some minor revisions.

2) Revisions
In the introduction and throughout the manuscript you have used long hyphens (—), please

consider revising these hyphens and try to avoid them as they are implicated with using large language models like Chatgpt, but it is not always the case.

Answer:
We thank the reviewer for this useful comment. We have carefully revised the manuscript and replaced all long hyphens with appropriate punctuation to ensure consistency and compliance with journal style.

The paragraph on Metaheuristic algorithms can also be revised but it is not mandatory. A good approach would be to just introduce genetic algorithms as part metaheuristic algorithms without mentioning the other algorithms like PSO, Ant Colony Optimization, and so forth. The introduction of these other algorithms leaves the reader with questions on what they are and how they are used in Viral Evolution, which might not be the case.

Answer:

We thank the reviewer for this helpful suggestion. In the revised manuscript, we have simplified the paragraph on metaheuristic algorithms. We now introduce Genetic Algorithms directly as part of the broader family of metaheuristics, without detailing other approaches (such as PSO or ACO), to maintain focus and avoid unnecessary digressions that are not directly relevant to the present study.

With these minor revisions I am sure, it will improve the manuscript. Kind regards and good luck with concluding the manuscript.