

## Systems Biology

# Exploring SARS-CoV-2 spike protein mutations through genetic algorithm-driven structural modeling

Valentina Di Salvatore<sup>1,\*</sup>, Avisa Maleki<sup>1,‡</sup>, Babak Mohajer<sup>2</sup>, Alvaro Ras-Carmona<sup>3</sup>,  
Giulia Russo<sup>1</sup>, Pedro Antonio Reche<sup>3</sup>, Francesco Pappalardo<sup>1</sup>

<sup>1</sup>Department of Drug and Health Sciences, University of Catania, Catania, 95125, Italy

<sup>2</sup>Mimesis SRL, Catania, 95125, Italy

<sup>3</sup>Department of Immunology, Faculty of Medicine, University Complutense of Madrid, Madrid, 28040, Spain

\*Corresponding author. Department of Drug and Health Sciences, University of Catania, Viale Andrea Doria 6, Catania, 95125, Italy. E-mail: valentinadisalvatore@unicat.it

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

Associate Editor: Nicola Mulder

## Abstract

**Motivation:** 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 discrete optimized protein energy scores across generations. The approach also enabled evaluation of Gibbs free energy and binding affinity for spike—Angiotensin-converting enzyme 2 receptor 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. 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.

**Availability and implementation:** All the Python and R scripts are available upon request to the authors.

## 1 Introduction

RNA viruses, including SARS-CoV-2, mutate frequently due to the intrinsic properties of their genomes (Otto *et al.* 2021). The impact of these mutations on transmission and disease severity depends on mutation rates and their effects within, and between, hosts (Sanjuán and Domingo-Calap 2016, Sanjuán *et al.* 2010), 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 Angiotensin-converting enzyme 2 receptor (ACE2) binding and membrane fusion (Ballesteros-Sanabria *et al.* 2022). 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 (Ballesteros-Sanabria *et al.* 2022) (Harvey *et al.* 2021).

Forecasting new variants enables timely public health responses (Rashed *et al.* 2022), supported by early surveillance systems (Rashed *et al.* 2021). A comprehensive strategy integrating genomics, bioinformatics, phylogenetics, mutation tracking, epidemiology, and AI facilitates real-time analysis of viral evolution (Xu *et al.* 2022). In particular, understanding mutation-driven changes in spike-ACE2 binding and Gibbs free energy ( $\Delta G$ ) is key to anticipating variant behavior and biological impact (Bai *et al.* 2021).

Metaheuristic algorithms are widely used to solve complex problems in biology, economics, and engineering (Michalewicz and Schoenauer 1996, Holland 1992).

Among these, genetic algorithms (GAs) are a prominent population-based method inspired by natural evolution (Michalewicz and Schoenauer 1996). Proposed by J.H. Holland in 1992, it incorporates chromosome representation, fitness evaluation, and biologically inspired operators, including the inversion operator (Holland 1992). 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 (Katoch *et al.* 2021, Abou El Ela and Spea 2009).

One study proposed VOC-DL, a deep learning framework integrating variant data into models like VOC-LSTM, VOC-

Received: May 30, 2025; Revised: September 10, 2025; Accepted: November 3, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

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 (Liao *et al.* 2022). 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 (Bai *et al.* 2021).

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.

## 2 Methods

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 Fig. 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.

### 2.1 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.

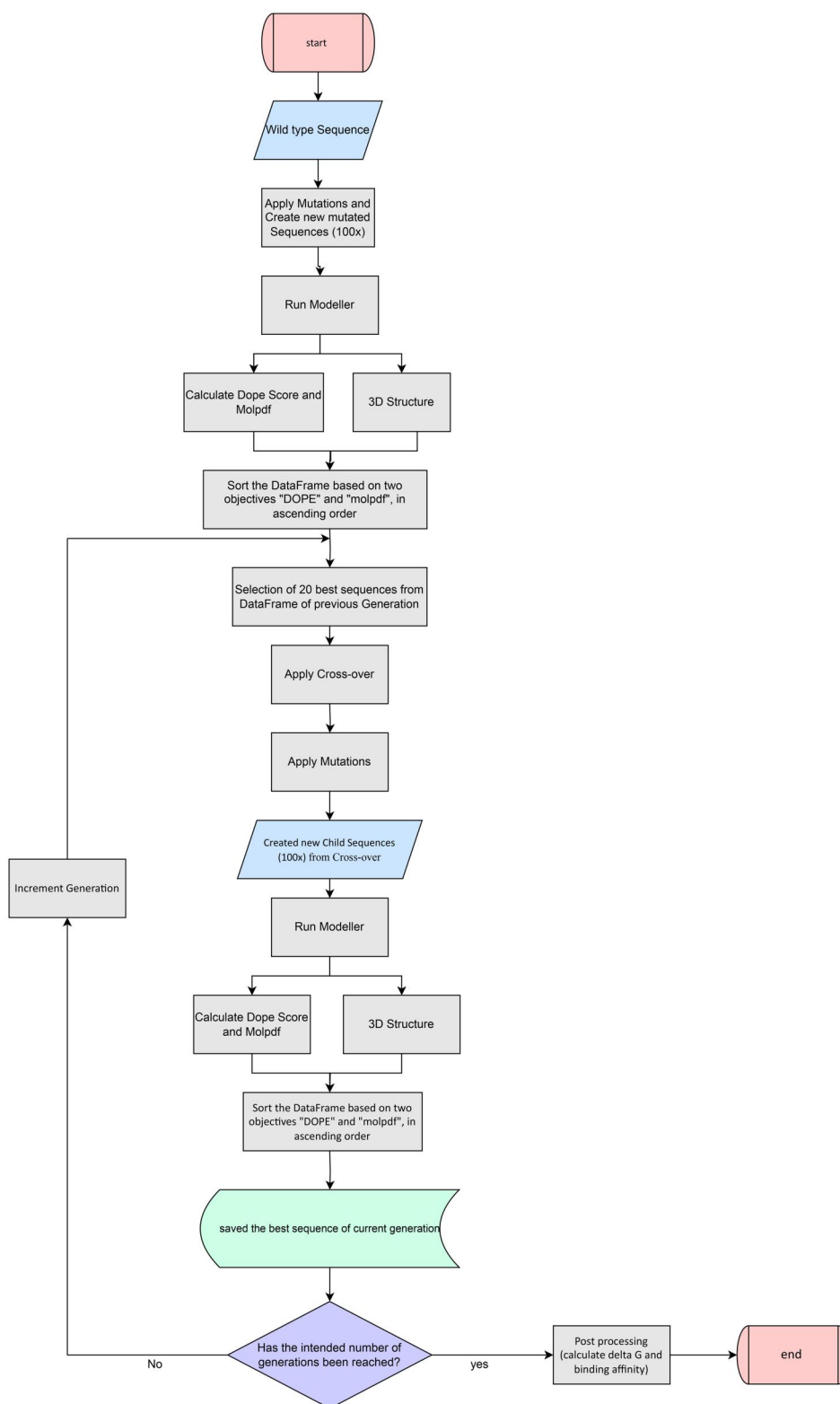
### 2.2 Model building

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 (Webb and Sali 2021). 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 (Webb and Sali 2016). 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 (Webb and Sali 2016, Sali and Blundell 1993):

- DOPE Score (discrete optimized protein energy) is a statistical potential used in homology modelling to assess protein structure quality. Implemented in MODELLER, it evaluates the energy of predicted models based on a reference state of non-interacting atoms in a spherical space. DOPE provides both overall and residue-level energy profiles, helping identify low-quality regions in the model (Eramian *et al.* 2006).
- 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 (Shen and Sali 2006).

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.



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

### 2.3 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:

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

where  $i=1, 2$  denotes the two metrics ( $O_1 = \text{DOPE}$ ,  $O_2 = \text{MolPDF}$ ).

The weighted fitness function was then computed as:

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

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 sub-optimal solutions, while a larger set would significantly increase computational costs. Twenty candidates ensured sufficient diversity for crossover operations while keeping the simulations computationally efficient.

## 2.4 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.

## 2.5 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 the 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.

## 2.6 Gibbs free energy evaluation

We modeled spike–ACE2 complexes for each mutant and estimated  $\Delta G$  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 ( $\Delta \text{ASA}$ ) for each residue, reflecting surface differences before and after binding. These  $\Delta \text{ASA}$  values were used to estimate binding free energy ( $\Delta G$ ) via the following linear model (Schön *et al.* 2011):

$$\Delta G_{\text{bind}} = C + w_{\text{Tyr}} \Delta \text{ASA}_{\text{Tyr}} + w_{\text{Ser}} \Delta \text{ASA}_{\text{Ser}} + w_{\text{Cys}} \Delta \text{ASA}_{\text{Cys}}$$

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

## 2.7 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 the strength of the interaction between a protein and a ligand and is quantified by the association constant ( $K_a$ ).  $K_a$  represents the equilibrium between the bounded and unbounded states of a protein–ligand pairs:

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

A higher  $K_a$  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  $K_a$  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  $K_a$  mathematically, of binding affinity using the gas constant ( $R$ ), and the temperature ( $T$ ) in Kelvin (Kastritis and Bonvin 2013):

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

Both the values of  $\Delta G$  and binding affinity provide us with complementary information for understanding the protein–ligand interactions. Additionally, the agreement between  $\Delta G$  and binding affinity values serves us in this study to validate the accuracy of the computational methods applied.

## 2.8 Select variants of concern

To identify potential VOCs, we applied a multi-criteria selection strategy based on three complementary metrics: DOPE score,  $\Delta G$ , 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  $\Delta G$  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.

## 2.9 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 (Nachar 2008). Second, we applied the same tests to predicted  $\Delta G$  and  $K_a$  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 (Sali and Blundell 1993), 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 (SciPy 1.0 Contributors *et al.* 2020), which offers reliable tools for hypothesis testing in Python.

## 3 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.

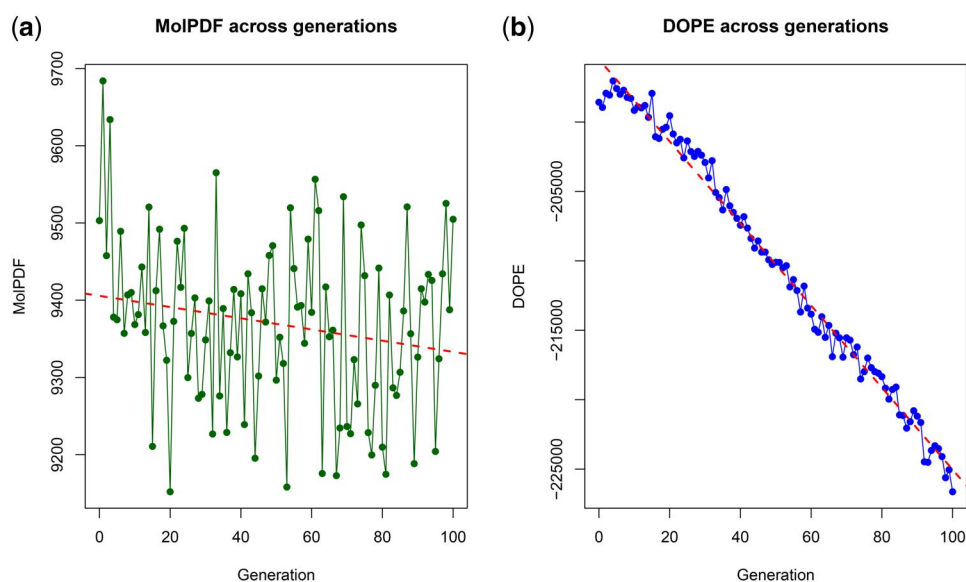
As shown in Fig. 2, DOPE scores steadily decrease across 100 generations, indicating improved model quality and structural stability. Since lower DOPE values correspond to more energetically favorable structures, this trend confirms the GA's effectiveness in selecting better variants. MolPDF values remained 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.

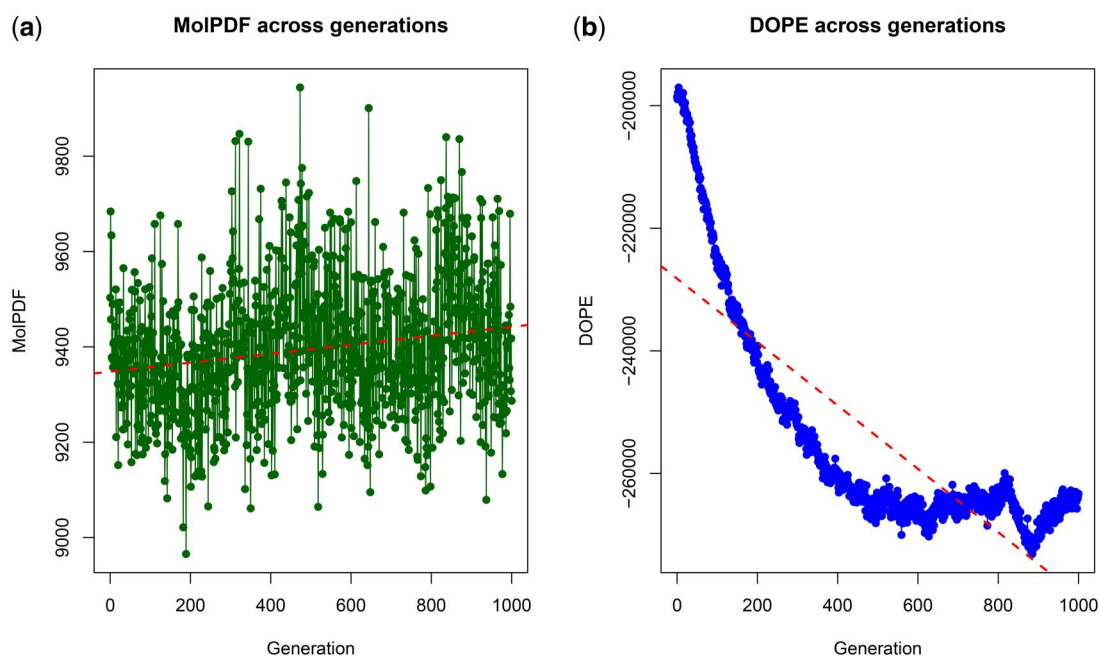
To further validate the robustness of the GA, we extended the evolutionary process to 1000 generations (Fig. 3).

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  $\sim 400$  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.

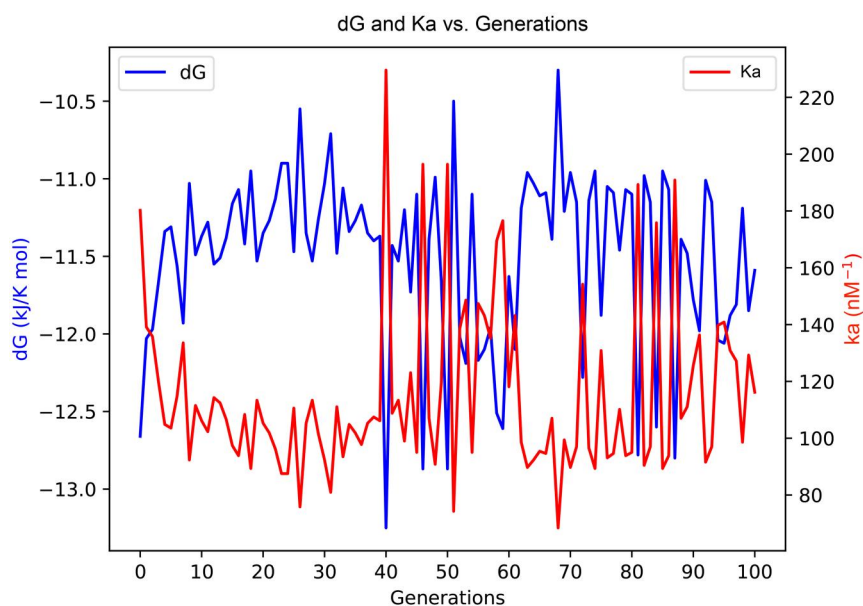
Together, the results from both the 100- and 1000-generation experiments demonstrate that the GA framework consistently drives optimization toward lower-energy structural states, thereby confirming its suitability for modeling the



**Figure 2.** (a) 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. (b) 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.



**Figure 3.** MolPDF (a) and DOPE (b) 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.

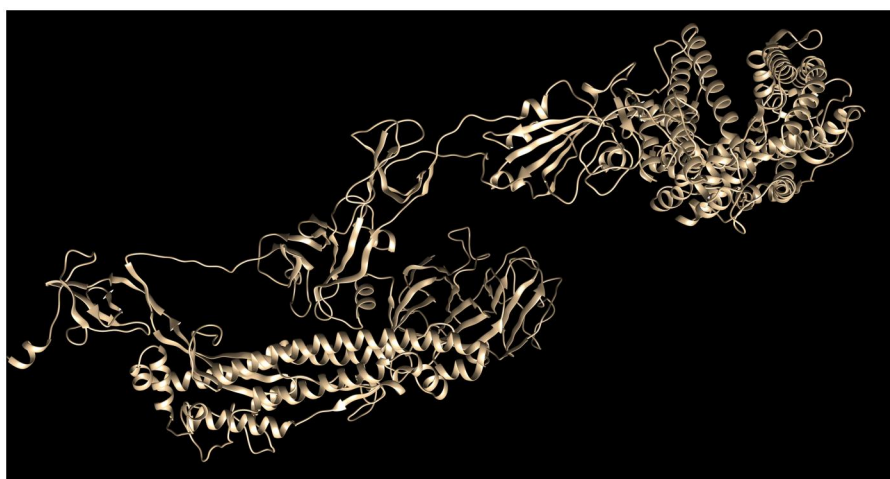


**Figure 4.** The plot illustrates the thermodynamic optimization of the spike–ACE2 interaction across generations. The  $\Delta G$  values (blue curve) fluctuate but overall show a trend toward lower free energy, indicating improved binding spontaneity. The  $K_a$  values (red curve) also fluctuate, 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.

impact of spike protein mutations. These structural improvements were then complemented by the thermodynamic evaluation of spike–ACE2 interactions ( $\Delta G$  and  $K_a$ ), providing a comprehensive assessment of both energetic stability and binding affinity across GA-driven evolutionary trajectories.

The  $\Delta G$  values determine whether a process or reaction will occur spontaneously, while the  $K_a$  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  $\Delta G$  and  $K_a$  fluctuate, the overall trend suggests a progressive shift toward more favorable binding conditions.  $\Delta G$  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,  $K_a$  exhibits substantial variability without a consistent trend, though intermittent peaks suggest episodes of enhanced binding affinity. These fluctuations reflect



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

the GA's capacity to explore diverse variants and occasionally achieve improved spike–ACE2 binding properties.

The trends in  $\Delta G$  and  $K_a$  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  $\Delta G$  indicates more favorable binding, while higher  $K_a$  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.

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 (Fig. 5). It also showed favorable  $\Delta G$  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.

### 3.1 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  $P = .000183$ ), 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. Figure 6 illustrates this difference, showing the marked reduction in DOPE scores between early and late generations. 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.

In contrast, the  $\Delta G$  and  $K_a$  did not show statistically significant differences between the first and last generations ( $\Delta G$   $P = .20$ ;  $K_a$   $P = .20$ ,  $t$ -test), as illustrated in Figure 7 and Figure 8. These results suggest that, while the GA effectively optimized structural stability, the functional improvements in binding thermodynamics were less pronounced. This

**Table 1.** Performance metrics of the optimized model at Generation 97.

Generation	DOPE score	$\Delta G$	Binding Affinity
97	$-223520.296$	$-12.61$	176.552

The table shows the key performance metrics for selected possible new VOCs. The columns include the generation number, DOPE score,  $\Delta G$ , and binding affinity.

**Table 2.** Statistical comparison of spike protein models.

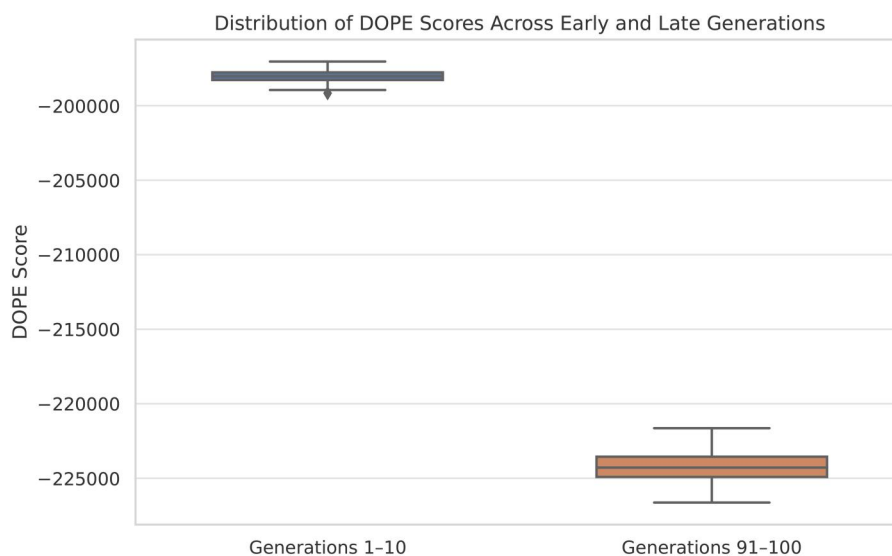
Test	$t$ -test $P$ -value	Mann–Whitney $P$ -value
DOPE Score	$2.36E-16$	.00018267
$\Delta G$ (Gibbs Free Energy)	.20046886	.24114463
$K_a$ (Binding Affinity)	.1992096	.24114463

The table shows a statistical comparison of DOPE score, Gibbs free energy ( $\Delta G$ ), and association constant ( $K_a$ ) between early (generations 1–10) and late (generations 91–100) spike protein models.

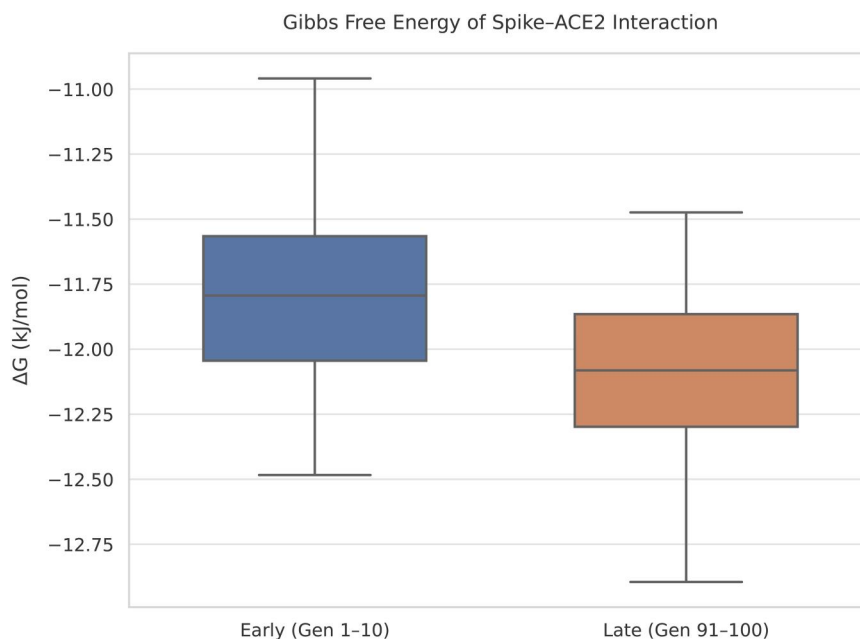
interpretation is further supported by the modest differences observed in the mean values:  $\Delta G$  improved only slightly, from  $-11.73$  to  $-12.05$  kJ/mol, and  $K_a$  increased from  $125.50$  to  $143.40$  nM $^{-1}$  between the early and late generations. These limited shifts reinforce the notion that, while structural energy minimization was prioritized and successfully achieved, thermodynamic binding properties did not undergo a comparable evolutionary pressure.

A possible explanation lies in the multi-objective nature of protein–ligand interaction optimization.  $\Delta G$  and  $K_a$  are influenced by subtle, non-linear conformational and surface changes that may not evolve in parallel with structural compactness or energy minimization (Van Den Noort *et al.* 2021). Moreover, as  $\Delta G$  and  $K_a$  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  $\Delta G$  and  $K_a$  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.



**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.



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

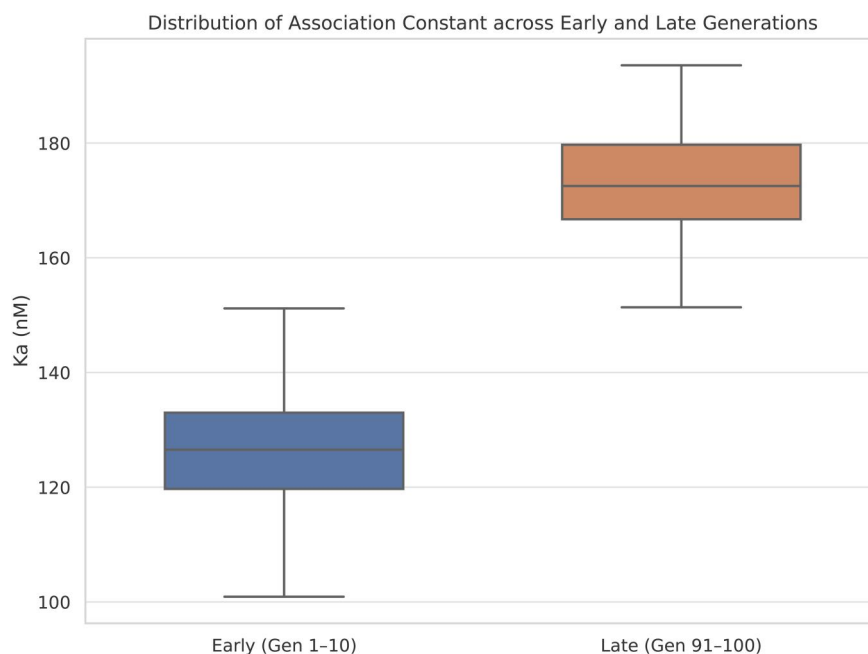
## 4 Discussion

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 insight into structural stability and functional relevance, particularly in the context of spike-ACE2 interactions. While  $\Delta G$  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



**Figure 8.**  $K_a$  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.

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 time-scales. Thus, while MolPDF remains informative, DOPE provides a more reliable and consistent measure of convergence and model quality in our GA framework.

Although  $\Delta G$  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  $\Delta G$  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.

## 5 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 ( $\Delta G$ ), 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 the 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.

## Author contributions

Valentina Di Salvatore (Writing—review & editing, statistical analysis), Avisia 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), and Francesco Pappalardo (Conceptualization, Supervision, Methodology)

## Conflict of interest

None declared.

## Funding

This work was supported by 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.

## Data availability

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](http://nih.gov))), (2) RCSB Protein Data Bank: RCSB PDB: Homepage, (3) MODELLER: About MODELLER ([salilab.org](http://salilab.org)). All the Python and R scripts used for the analysis are available upon request to the authors.

## References

- Abou El Ela AA, Spea SR. Optimal corrective actions for power systems using multi-objective genetic algorithms. *Electr Power Syst Res* 2009;79:722–33. <https://doi.org/10.1016/j.epsr.2008.10.007>
- Bai C, Wang J, Chen G *et al.* Predicting mutational effects on receptor binding of the spike protein of SARS-CoV-2 variants. *J Am Chem Soc* 2021;143:17646–54. <https://doi.org/10.1021/jacs.1c07965>
- Ballesteros-Sanabria L, Pelaez-Prestel HF, Ras-Carmona A *et al.* Resilience of spike-specific immunity induced by COVID-19 vaccines against SARS-CoV-2 variants. *Biomedicines* 2022;10:996. <https://doi.org/10.3390/biomedicines10050996>
- Eramian D, Shen M, Devos D *et al.* A composite score for predicting errors in protein structure models. *Protein Sci* 2006;15:1653–66. <https://doi.org/10.1110/ps.062095806>
- Harvey WT, Carabelli AM, Jackson B, COVID-19 Genomics UK (COG-UK) Consortium *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 2021;19:409–24. <https://doi.org/10.1038/s41579-021-00573-0>
- Holland JH. Genetic algorithms. *Sci Am* 1992;267:66–72.
- Kastritis PL, Bonvin AMJJ. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *J R Soc Interface* 2013;10:20120835. <https://doi.org/10.1098/rsif.2012.0835>
- Katoch S, Chauhan SS, Kumar V. A review on genetic algorithm: past, present, and future. *Multimed Tools Appl* 2021;80:8091–126. <https://doi.org/10.1007/s11042-020-10139-6>
- Liao Z, Song Y, Ren S *et al.* VOC-DL: deep learning prediction model for COVID-19 based on VOC virus variants. *Comput Methods Programs Biomed* 2022;224:106981. <https://doi.org/10.1016/j.cmpb.2022.106981>
- Michalewicz Z, Schoenauer M. Evolutionary algorithms for constrained parameter optimization problems. *Evol. Comput* 1996;4:1–32. <https://doi.org/10.1162/evco.1996.4.1.1>
- Nachar N. The Mann-Whitney U: a test for assessing whether two independent samples come from the same distribution. *TQMP* 2008;4:13–20. <https://doi.org/10.20982/tqmp.04.1.p013>
- Otto SP, Day T, Arino J *et al.* The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. *Curr Biol* 2021;31:R918–R929. <https://doi.org/10.1016/j.cub.2021.06.049>
- Rashed EA, Kodera S, Hirata A. COVID-19 forecasting using new viral variants and vaccination effectiveness models. *Comput Biol Med* 2022;149:105986. <https://doi.org/10.1016/j.combiomed.2022.105986>
- Rashed EA, Kodera S, Shirakami H *et al.* Knowledge discovery from emergency ambulance dispatch during COVID-19: a case study of nagoya city, Japan. *J Biomed Inform* 2021;117:103743. <https://doi.org/10.1016/j.jbi.2021.103743>
- Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779–815. <https://doi.org/10.1006/jmbi.1993.1626>
- Sanjuán R, Domingo-Calap P. Mechanisms of viral mutation. *Cell Mol Life Sci* 2016;73:4433–48. <https://doi.org/10.1007/s00018-016-2299-6>
- Sanjuán R, Nebot MR, Chirico N *et al.* Viral mutation rates. *J Virol* 2010;84:9733–48. <https://doi.org/10.1128/JVI.00694-10>
- Schön A, Madani N, Smith AB *et al.* Some binding-related drug properties are dependent on thermodynamic signature: optimization of drug candidates. *Chem Biol Drug Des* 2011;77:161–5. <https://doi.org/10.1111/j.1747-0285.2010.01075.x>
- Shen M, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* 2006;15:2507–24. <https://doi.org/10.1110/ps.062416606>
- Van Den Noort M, De Boer M, Poolman B. Stability of ligand-induced protein conformation influences affinity in maltose-binding protein. *J Mol Biol* 2021;433:167036. <https://doi.org/10.1016/j.jmb.2021.167036>
- Virtanen P, Gommers R, Oliphant TE, SciPy 1.0 Contributors *et al.* Author correction: sciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods* 2020;17:352–<https://doi.org/10.1038/s41592-020-0772-5>

- Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics* 2016;54:5.6.1–5.6.37. <https://doi.org/10.1002/cpbi.3>
- Webb B, Sali A. Protein structure modeling with MODELLER. In: Chen YW, Yiu C-PB (eds.), *Structural Genomics, Methods in Molecular Biology*. Springer US, New York, NY, 2021, 239–255. [https://doi.org/10.1007/978-1-0716-0892-0\\_14](https://doi.org/10.1007/978-1-0716-0892-0_14)
- Xu L, Magar R, Barati Farimani A. Forecasting COVID-19 new cases using deep learning methods. *Comput Biol Med* 2022;144:105342. <https://doi.org/10.1016/j.combiomed.2022.105342>