Article

A Computational Workflow to Predict Biological Target Mutations: The Spike Glycoprotein Case Study

Pietro Cozzini [1,*] **, Federica Agosta** [1] **, Greta Dolcetti** [2] **and Alessandro Dal Palù** [2]

1   Molecular Modeling Lab, Food and Drug Department, University of Parma, Parco Area delle Scienze 17/A, 43121 Parma, Italy; federica.agosta@unipr.it
2   Department of Mathematical, Physical and Computer Sciences, University of Parma, 43121 Parma, Italy; greta.dolcetti@studenti.unipr.it (G.D.); alessandro.dalpalu@unipr.it (A.D.P.)
*   Correspondence: pietro.cozzini@unipr.it

Abstract: The biological target identification process, a pivotal phase in the drug discovery workflow, becomes particularly challenging when mutations affect proteins' mechanisms of action. COVID-19 Spike glycoprotein mutations are known to modify the affinity toward the human angiotensin-converting enzyme ACE2 and several antibodies, compromising their neutralizing effect. Predicting new possible mutations would be an efficient way to develop specific and efficacious drugs, vaccines, and antibodies. In this work, we developed and applied a computational procedure, combining constrained logic programming and careful structural analysis based on the Structural Activity Relationship (SAR) approach, to predict and determine the structure and behavior of new future mutants. "Mutations rules" that would track statistical and functional types of substitutions for each residue or combination of residues were extracted from the GISAID database and used to define constraints for our software, having control of the process step by step. A careful molecular dynamics analysis of the predicted mutated structures was carried out after an energy evaluation of the intermolecular and intramolecular interactions using the HINT (Hydrophatic INTeraction) force field. Our approach successfully predicted, among others, known Spike mutants.

Keywords: COVID-19; in silico mutation prediction; molecular modeling; HINT

1. Introduction

Identifying a biological target and analyzing its structural characteristics and mechanism of action is the first essential step in the drug discovery process. Thanks to the availability of omics data and three-dimensional protein structures, different computational approaches have been employed to predict pharmacological targets, assuming that a drug may interact with proteins characterized by the same mechanism of action and/or similar binding pocket characteristics [1]. Moreover, the increased development and application of artificial intelligence techniques led to the implementation machine learning methods in the drug discovery process to predict new targets and screen for active compounds [2].

The biological target identification process becomes more challenging when mutations alter the protein mechanism of action, modifying or compromising drug interactions, as in the case of the COVID-19 pandemic [3], whose rapid spread has prompted researchers worldwide to analyze its structural and epidemiological characteristics to find efficacious treatments in a short time [4–8].

Protein mutations are casual and occur with a very low frequency [9,10]. Spike glycoprotein mutations have proven to affect viral replication, antibody neutralization susceptibility, and drug efficacy [11,12]. Predicting new possible stable mutations is crucial for the development of drugs, monoclonal antibodies (mAbs), and vaccines.

The COVID-19 fighting is like the cop and robber's game. To win, the cop must be ahead of the robber, predicting his moves. We can distinguish three actors in the COVID-19 game: the Spike glycoprotein (the robber), the human ACE2 (the victim), and the

antibodies (the cop). To survive, the mutated Spike glycoprotein must maximize its affinity toward ACE2 and reduce, through amino acid substitutions or deletions, interactions with antibodies or drugs.

This work aims to develop a procedure that, starting from chemical structure and genomic sequence information, can predict a comprehensive set of possible new Spike mutations evaluated through an intermolecular and intramolecular energy approach.

To reach this goal, we combine a type of artificial intelligence stemming from computational logic, constraint programming with molecular modeling, and molecular dynamics.

Since the analysis of this game evolution must rely on the quality of a large sequence dataset, we focused on the repository GISAID (https://www.gisaid.org/), which, on 10 January 2020, released the first complete genome sequence of SARS-CoV-2 and, since then, has become the main genomic sequence database in the world [13,14]. To date, it has collected more than 13 M virus sequences. GISAID contained more than 6.9 M sequences at the beginning of this research work.

Since data uploaded to the repository is collected with different standards and variable quality, we validated the reliability of such experimental data. In our view, the main limitation of the GISAID initiative for our purposes is the possibility of loading data from anywhere without an actual preliminary quality check.

In a recent opinion paper in MedChem Letters [15], we argued about the need to have a reliable COVID-19 sequences database, and we proposed to adopt big data and blockchain technologies in a scenario where every National Health Authority is the local collector, owner, and data-quality responsible.

We identified a remarkable number of unusable strings collected in GISAID after spending several months deciphering and cleaning up the data. We built a database of 171 K COVID-19 validated strings, starting from the original database of 6.9 M sequences. The database was designed as a standard relational SQL database (Supplemental Material S1).

Given the large number of sequences and the potential combinations of mutations to be considered, we resort to specific techniques from artificial intelligence that can learn rules from data faster than manual work and use the same rules to generate new hypothetical and plausible sequences.

Different studies have been recently carried out employing artificial intelligence techniques to predict new possible Spike mutations [16,17]. In this work, we focus on declarative programming [18–21] and constraint programming, which, in contrast to machine learning and popular neural networks, allows us to build a model that can be directly understood and modified. We designed a workflow that mimics human deductive capabilities in problem-solving: we aimed to automatically deduce logic rules by observing the data we gathered, to explain and validate them thanks to the Chemistry knowledge base, and finally to use refined rules to model and generate new possible variants of SARS-CoV-2.

This method was implemented with statistical and chemical analysis due to severe limitations caused by data quality, even on the validated dataset.

Biological computational analysis was focused on the evaluation of the effect of extracted mutations, considering protein stability and the impact on interactions with ACE2 and different classes of antibodies.

2. Results and Discussion

Single nucleotide polymorphisms (SNPs) are the most common genetic alterations that lead to substitutions, deletions, or insertions in the protein aminoacidic encoded sequence [22]. Mutations can affect protein structure, stability, and function and sometimes have been related to pathological conditions and/or drug resistance [23,24].

Among mutated proteins, the COVID-19 Spike glycoprotein is sadly known for its multiple mutations, sometimes responsible for the virus's increased diffusion and pathogenicity [25–27], mainly when these mutations affect aminoacidic residues of the receptor binding domain (RBD), the main antigenic region involved in the interaction with the human ACE2 [28,29]. The prediction of new possible RBD stable mutations could help conceive

efficient weapons for pandemic containment, such as developing specific and effective drugs, vaccines, and/or monoclonal antibodies.

The first step of our research consists of the analysis of the known Spike protein's aminoacidic sequences, with a focus on the RBD domain. We retrieved data from GISAID and collected 6,908,513 sequences updated on 12 January 2022, in FASTA format. We focused on the aminoacidic sequence and metadata, e.g., the virus collection date and the world region. After filtering out low-quality non-human regions with no "original" passage history and redundant sequences, we retained 171,862 sequences and designed a database to allow some data analysis and extract the rules to create our predictive model.

Statistical analysis revealed the frequency of single nucleotide polymorphisms for specific amino acids. It allowed us to derive a first set of probabilistic rules describing the probability of observing a specific mutation at a specific position. The problem is computationally demanding due to the number of possible combinations and requires a set of rules that allow and/or avoid specific sub-combinations to reduce the set of non-plausible combinations. Due to the discussed reliability of the Spike sequence data [15] and the consequent difficulty in their analysis and evaluation, the structure-activity relationship plays a critical preliminary role in constructing our predictive method. The evaluation of the effects of known mutations on the stability of the protein and the interaction with the target and different classes of antibodies has allowed the identification of potentially mutable sites and to define the type of substitution to which amino acids of different positions could go against.

Each aminoacidic sequence generation randomly selects mutated amino acids independently. The approach resembles a Montecarlo-like generation, which, however, cannot control the interrelationships between co-occurrent mutations.

The structural analysis of the effect of known mutations was focused on three fundamental points: the stability of the mutated proteins, the interaction with the human target, and the susceptibility to antibody neutralization.

Intramolecular analysis of mutated receptor binding domains showed that all mutated proteins are as stable as or more stable than wild-type proteins [30]. This could explain the virus's ability to live and replicate despite the considerable number of mutations characterizing some variants of concern [31–33]. Analyzing the mutations distribution, it is possible to observe how some residues are highly conserved, especially cysteines (C336, C361, C379, C432, C391, C525, C480, and C488) involved in sulfide bridges and residues located into the five ß-sheets of the core region (except for antigenic residues 375 to 379) that seem to be crucial for the protein stability, compensating for the higher flexibility of the large random coil regions. These residues were considered non-mutable amino acids and excluded from the prediction algorithm.

To detect all RBD antigenic residues, we studied the interaction of different antibodies with a focus on monoclonal ones, divided into four different classes based on their epitope distribution [34], thanks to the availability of several experimental data sets to validate our computational predictions.

Modeling the different mutations of the variants of concern, we observed that mutations affect antigenic residues, and some non-conservative substitutions at residues K417, E484, Q493, G446, Q498, N501, L452, S477, T478, and S371 directly influenced the interactions with one or more different classes without compromising the interaction with ACE2 (Supplemental Materials, S2).

The Spike RBD and the human ACE2 interactions are well-known and characterized [35–37]. This involves some hydrogen bonds (N487-Y83, T500-D355, Q493-E35, Y449-D38, Q498-K353, T500-Y41, Y505-E37), two main salt bridges (K417-D30 and E484-K31), and some hydrophobic interactions (F486 and Y489 of the RBD with F28, L79 and M82 of ACE2, as well as L455 and F456 with T27).

Mutated RBD-ACE2 complexes are at least as stable as the wild-type [30]. For this last point, it is necessary to underline how all the VOCs, except for the alpha variant, which was characterized by a single mutation, present different substitutions at the RBD residues at

the same time. Even if some of these could compromise local interactions with ACE2, such as E484A, that preclude the salt-bridge instauration with residue K31 of the human target, these negative substitutions occur with others able to optimize the interaction with the target, such as polar or basic ones, generating additional hydrogen bonds or electrostatic interactions. Among these, the well-known N501Y guarantees a new π-π interaction with the residue Y41 of ACE2. In contrast, the introduction of basic residues (Q493R, Q498R, and Y505H) in Omicron variants is responsible for electrostatic interactions with the target that present a prevalent negative polarization and/or the establishment of new hydrogen bonds [35,38], as described in Figure 1.
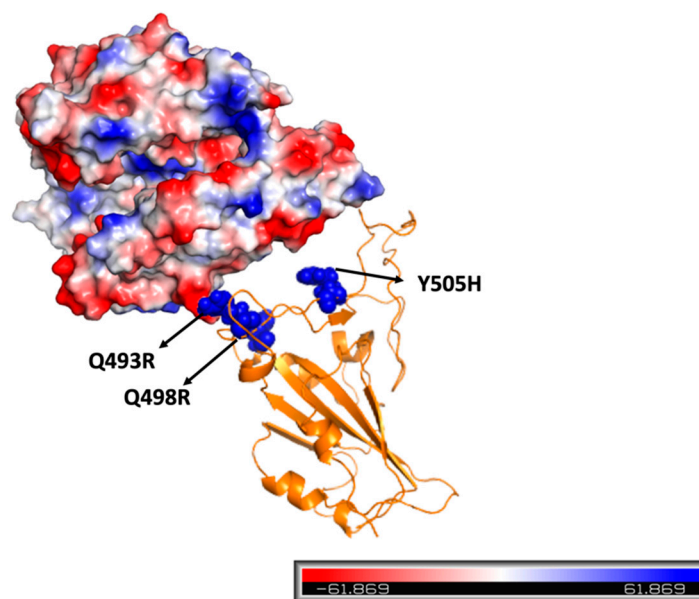

Figure 1. Omicron RBD-ACE2 complex stability. Even if the omicron variant is characterized by a huge number of RBD mutations, the introduction of basic amino acids (blue regions) such as arginine and histidine generates favorable interactions with ACE2 that present a prevalent negative polarization (red regions).

All these aspects agree that for a mutation to be established, the RBD must remain stable, maintain interactions with the target, and reduce affinity toward antibodies.

This structural analysis laid the foundation for the elaboration of our predictive algorithm. In particular, it allowed us to detect 55 possible antigenic residues that could be mutated (Figure 2) and that, to simplify the prediction algorithm, were divided into three different clusters, called A, B, and C, such that the residues of each were 10 Å away from those of the others (Supplemental Materials S3). Residues belonging to cluster A are located in the main random coil regions and are mainly responsible for the interactions with ACE2.

Since Omicron variants are the most common, we hypothesized that the evolution of these could generate a new variant. Therefore, we developed a reference model characterized by common mutations shared between BA.1 and BA.2: G339D, S373P, S375F, K417N, S477N, T478K, E484A, Q493R, Q498R, N501Y, and Y505H. The number of mutations produced was added to the already-imposed Omicron mutations.

Then, we produced the corresponding three-dimensional structures according to the number of additional mutations reported in Table 1.
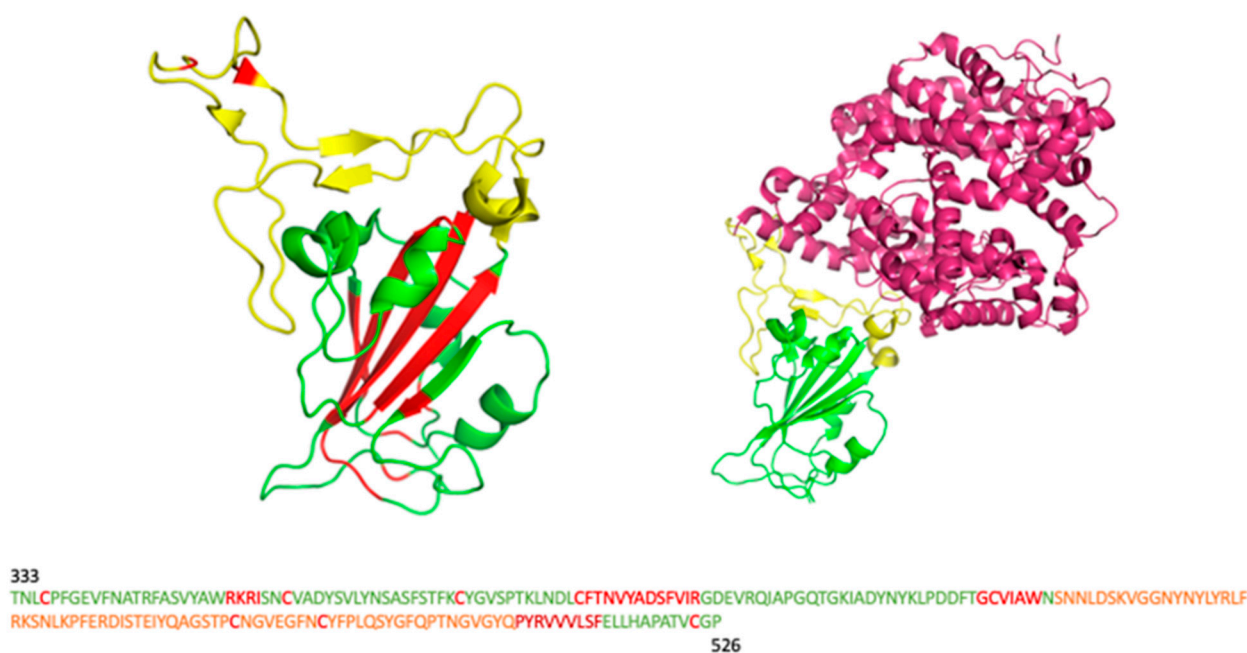
333
TNL**C**PFGEVFNATRFASVYAW**RKRI**SN**C**VADYSVLYNSASFSTFK**C**YGVSPTKLNDL**CFTNVYADSFVIR**GDEVRQIAPGQTGKIADYNYKLPDDFT**GCVIAW**NSNNLDSKVGGNYNYLYRLF
RKSNLKPFERDISTEIYQAGSTP**C**NGVEGFN**C**YFPLQSYGFQPTNGVGYQ**PYRVVVLS**FELLHAPATV**C**GP
526

Figure 2. Receptor Binding Domain Main Features: The receptor binding domain (residues 333–526) is a structural domain responsible for the interaction with human ACE2 and different antibodies. Cysteines and core region amino acids (red residues) essential for protein stability were excluded from prediction algorithm development. The receptor binding motif, including residues 438 to 506 (yellow residues), is the region involved in the interaction with ACE2, as shown on the right side. Antigenic residues (some green and yellow residues) were considered possible mutable residues.

Table 1. Number of generated models for each cluster. Each cluster presents different structures with different mutations (one to four) added to the reference system.

|	1 Add. Mutation	2 Add. Mutations	3 Add. Mutations	4 Add. Mutations
Group A	130 structures	8178 structures	=	=
Group B	10 structures	33 structures	36 structures	=
Group C	33 structures	461 structures	3531 structures	16,002 structures

These generated models were analyzed in HINT (Hydrophatic INTeractions), a force field that uses the experimental amino acid residues LogP$_{o/w}$ values (partition coefficient for solute transfer in 1-octanol/water) to calculate intermolecular and intramolecular interactions [30,39,40]. Each model was compared with the reference system (Supplemental Materials S4). The HINT score is automatically calculated as a summation of hydropathic interactions (hydrogen bond, acid-base, hydrophobic, and Coulombic).

In this screening phase, the intramolecular HINT scoring function was used as a sensitive, reliable, and fast method for evaluating stable mutants. Among the predicted, there are known additional mutations that characterized the omicron sub-variants that had been excluded for the construction of the reference structure (BA.1 and BA.2 mutations) or that characterized the most recent variants, such as BA.4, BA.5, and XE, which have spread after the beginning of our work. These mutants are at least as stable as the reference model, as shown in Table 2, underscoring the robustness, sensitivity, and reliability of our predictive and energy evaluation methods.

Table 2. Validation phase. Some of the predicted additional mutations characterized shared by most recent variants, such as omicron subvariants, were excluded from algorithm prediction construction and unknown at the beginning of this work. They all characterized structures that are at least stable as the reference model.

Additional Mutations	HINTscore	ΔHINTscore	Variant
R408S	10,972.15	223.76	BA.2, BA.4, BA.5
L452R	11,552.56	804.17	BA.4, BA.5, XE
D405N	10,978.56	229.72	BA.4, BA.5, XE
G476S	11,384.25	635.86	BA.1
D405N, L452R	10,860.45	112.06	BA.4, BA.5, XE
D405N, R408S	11,047.57	299.18	BA.2, BA.4, BA.5, XE
R408S, L452R	10,818.82	70.43	BA.4, BA.5, XE
S371F	10,926.9	178.51	BA.2, BA.4, BA.5, XE
T376A	11,025.84	277.45	BA.2, BA.4, BA.5, XE
S371F, T376A	11,216.69	468.3	BA.2, BA.4, BA.5, XE

Since additional mutations belonging to cluster A are related to residues of the receptor binding motif, which include all residues taking part in the interaction with the target, their effect on the affinity and stability toward hACE2 was analyzed in HINT, considering both the intermolecular connections (target affinity) and the intramolecular energy (complex stability). This preliminary analysis allowed the selection of 95 different RBD-ACE2 complexes with higher affinity and stability than the reference complex. A molecular dynamics simulation was performed to evaluate the stability of each complex over time. Since all the analyzed VOCs achieved their stability within 125 ns (Figure 3A), the simulation time was set to 250 ns. The RMSD analysis confirmed the HINT prediction, showing the stability of all the analyzed complexes. The RMSD of some representative models is reported in Figure 3B, while all the other data are reported in Supplemental Materials S5.

(**A**)
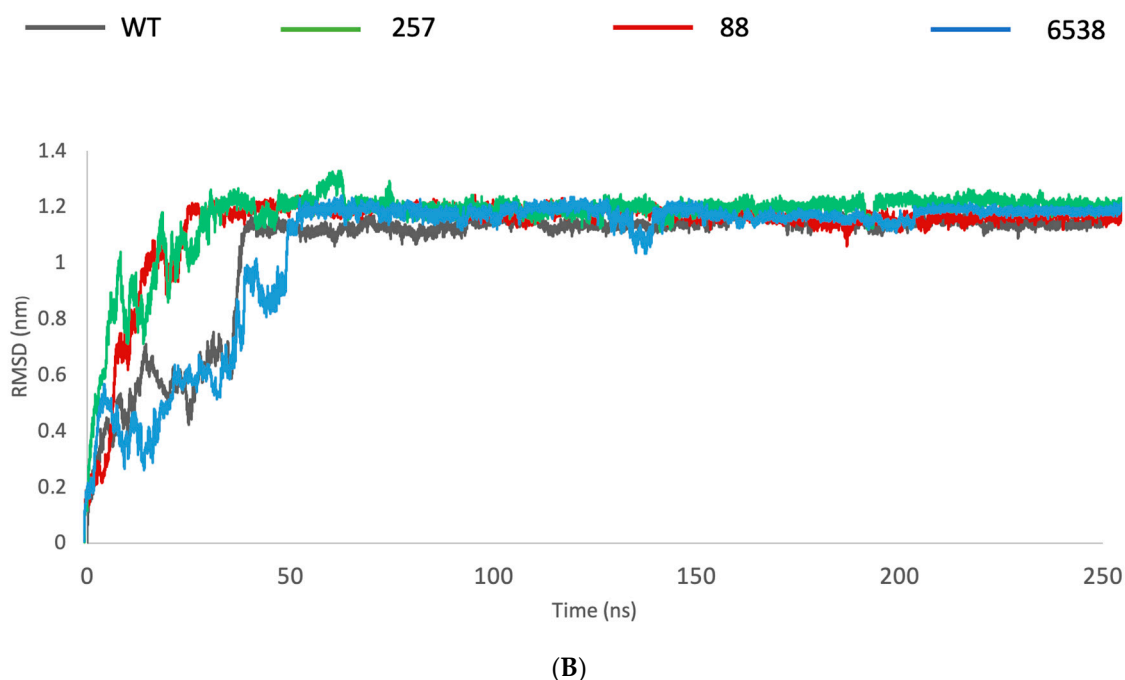
Figure 3. *Cont.*

(B)
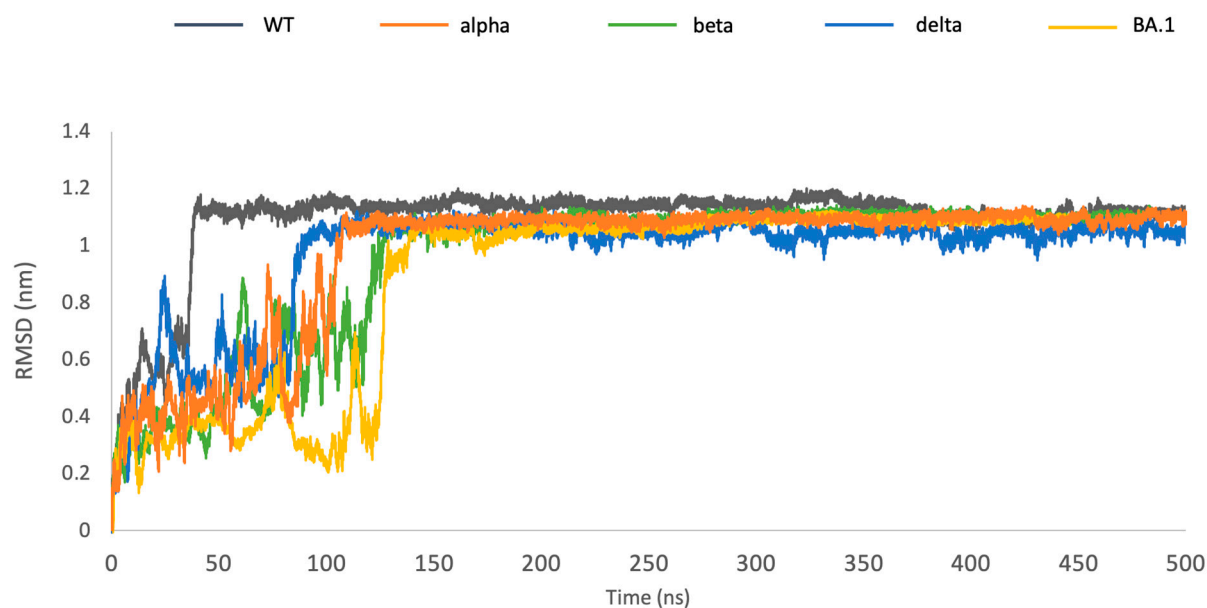
Figure 3. (**A**) RMSD of VOCs. A molecular dynamics simulation analyzed the wild-type, alpha, beta, delta, and BA.1 variants. The RMSD shows that all these complexes reach stability at 125 ns. (**B**) RMSD of some representative models compared with the wild-type system (grey curve). The RMSD analysis shows the higher stability of the generated models during all simulation times.

3. Materials and Methods

This research work aims at developing and applying a computational procedure able to predict a comprehensive set of possible new RBD Spike mutations, starting from chemical structure and genomic sequence information and combining constraint programming with molecular modeling. Generated mutants were evaluated through an intermolecular and intramolecular HINT energy approach (Figure 4).
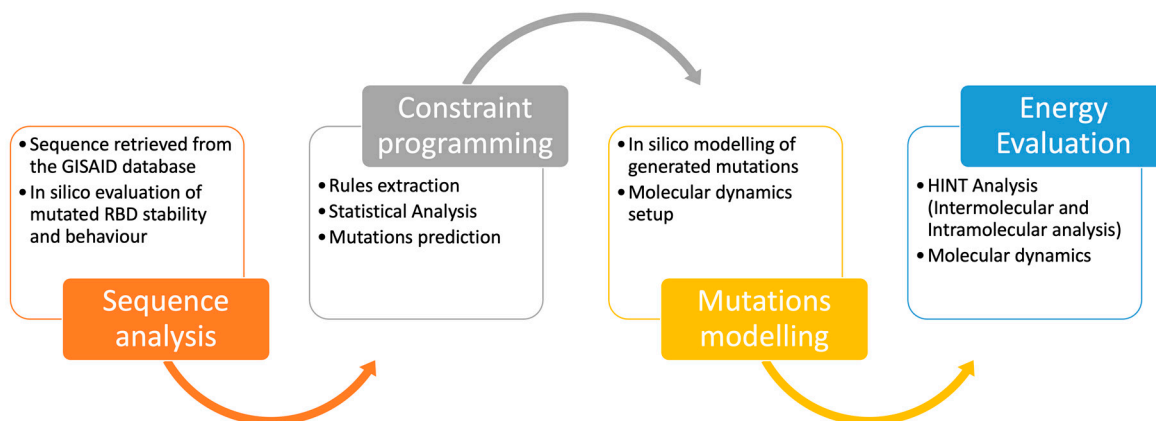
Figure 4. Computational workflow. COVID-19 sequences were retrieved from the GISAID repository and organized as a SQL database. Mutation rules were derived from statistical and chemical analysis and used to develop a procedure based on the constraint programming language. Produced mutations were analyzed through molecular dynamics and HINT-based analysis.

3.1. Constraint-Based Modeling

We used Minizinc [41], a programming language that models constraint satisfaction problems. It allows the developer to express constraints in a fast, declarative way (at a high level).

We modeled how mutations occur in SARS-CoV-2: each amino acid was represented by its position in sequence; we enforced positions where no mutation would be generated (the ones in which we observed few to no occurrences of mutations); for the other ones, we were able to constrain the potential mutation set to the one we observed inside GISAID data.

We scored mutations according to the logarithm of the single nucleotide polymorphism (SNP) frequency to occur, and we maximized the sum of each location's contribution. Optimal sequences had a weak correspondence with expected relevant mutations, leading to the next-generation model.

We replaced the scoring function with a chemistry-based one, considering the positive or negative interaction with the antibodies and ACE2 protein. We changed the set constraints to impose Omicron's mutations occurring inside the Receptor Binding Domain (RBD), impose conserved positions that could not mutate, and impose mutable ones as the result of a possible single base change in the DNA triplet that encodes the amino acid.

3.2. Protein Structures Selection

Several structures of Spike protein in complex with the human Angiotensin Converting Enzyme 2 (ACE2, EC number 3.4.17.23) or different antibodies are available in the Protein Data Bank (https://www.rcsb.org/).

For our preliminary analysis to evaluate the effect of different mutations on Spike stability and behavior, we chose the most representative complex of monoclonal antibodies belonging to different classes and currently approved for the treatment of COVID-19. We selected X-ray or Cryo-EM structures characterized by a resolution better than 3 Å, a B value (isotropic) from Wilson Plot lower than 35 Å, and without mutations in the RBD.

Among the structures of Spike-ACE2 complexes, we selected 6M0J, an X-ray diffraction structure with a resolution of 2.45 Å [37].

3.3. RBD VOC Mutations

Mutations characterizing the receptor binding domain of different variants of concern (VOC) were retrieved from GISAID: Alpha (B.1.1.7): N501Y; Beta (B.1.351): K417N, E484K, N501Y; Gamma (P.1): K417T, E484K; N501Y; Delta (B.1.351): L452R, T478K; Omicron 1 (BA.1): G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, Q498R, N501Y, Y505H; Omicron 2 (BA.2): G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y, Y505H.

Since omicron variants are widespread, we generated a starting model characterized by common mutations shared between BA1 and BA2: G339D, S373P, S375F, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y, and Y505H.

3.4. Protein Preparation

Mutations were introduced in the RBD structure using the Pymol wizard mutagenesis module (https://pymol.org/2/). Proteins were prepared and minimized in Gromacs v.2021.4 (https://www.gromacs.org), choosing the Amber force field.

Proteins were solvated in a triclinic box of 15 Å radius with a TIP3P water model. Each system was neutralized and salted to 0.15 M NaCl. Energy minimization was performed using the steepest descendent minimization algorithm and stopped when the maximum force was less than 100 KJ/(mol nm).

3.5. HINT Calculation

HINT (Hydrophatic Interactions) is a LogP-based force field [39,40] chosen for the evaluation of protein stability (intramolecular scoring function) [30,42] and the calculation of hydrophobic and polar interactions between proteins (intermolecular scoring function). In this work, it was used to validate the generated three-dimensional models, define the effect of mutations on RBD stability and its interactions with the human target, and use different antibody classes to predict new mutations and evaluate their effect.

Protein partitioning was performed in HINT under neutral pH conditions, using the "Dictionary" option and setting a semi-essential hydrogen treatment that explicitly polarized, unsaturated, and alpha-to heteroatom hydrogens.

The HINT score was calculated, starting from the minimized structure, as a summation of hydropathic interactions between all atom pairs, considering the hydrophobic atom constant (a), the solvent accessible surface area (Sasa), and the functional distance behavior for the interaction (R).

$$B = \sum \sum b_{ij}$$

$$b_{ij} = S_i \ a_i \ S_j \ a_j \ R_{ij}$$

For intermolecular calculations, *i* and *j* are atoms of two interacting proteins, while for intramolecular calculations, *i* and *j* are two atoms of the same macromolecule where *i* is not equal to, covalently bonded to, or involved in a 1–3 interaction with *j*.

The output analysis reveals specific information about atom-atom interactions: positive values represent favorable binding situations, while a negative value represents unfavorable interactions such as acid-acid, base-base, hydrophobic-polar (desolvation), and steric clashes. Interactions between residue sidechains and their environment are analyzed as a three-dimensional map that is a backbone angle-dependent library of interaction profiles that shows interaction types, strength, and optimal 3D position [43,44].

The intramolecular HINT force field has already been employed for the evaluation of mutated RBD stability and behavior with respect to available experimental data. As HINT has proven to be a fast and reliable tool to estimate small energy changes related to stability induced by mutations, it is used for an accurate evaluation of generated mutants' stability with a focus on RBD stability (intramolecular analysis) and RBD behavior (RBD-ACE2 intermolecular affinity).

3.6. Molecular Dynamics Simulations

Molecular dynamics simulations were carried out in Gromacs, choosing the Amber force field. Proteins were prepared as previously described in Section 3.4.

Three steps of 0.1 ns of NVT equilibration were carried out using the Langevin thermostat, followed by 0.5 ns of NPT equilibration to heat the system to 300 K gradually and set the pressure at 1 bar.

The total simulation time was 250 ns. To analyze the stability of each system during the simulation time, the root mean square deviation (RMSD), the root mean square fluctuation (RMSF), and the number of hydrogen bonds between the two interacting chains were calculated using the gmx rmsdist, gmx rmsf, and gmx hbond, respectively (Supplemental Materials S5 and S6).

4. Conclusions

In this research, we applied a combined technique of special programming and energy function estimation to discover new possible COVID-19 mutants produced following rules from the computational biological analysis of the known mutations extracted from the GISAID repository.

Because the data collected in the GISAID repository is our milestone in defining mutation rule prediction, we spent much of our efforts cleaning up this huge database to extract a reliable data set.

The stability of the generated mutants was considered the first crucial aspect to be analyzed in determining survival. The HINT force field proved to be a sensitive and reliable tool that allowed us to quickly select only mutants containing additional mutations compared with those in the stable reference model. Among these, mutations that characterized the variants of recent diffusion have been predicted, confirming our method's reliability and predictive power.

A deeper analysis of the stability and affinity toward ACE2 was conducted, combining an energy evaluation of minimized models with molecular dynamics simulation to evaluate complexes' stability over time.

Because it is very hard for many labs to produce new Spike mutants, we validated our approach against wild-type and known variants.

The final goal is to be ready to develop new drugs or new antibodies against new COVID-19 mutants.

References

1. Dai, Y.-F.; Zhao, X.-M. A Survey on the Computational Approaches to Identify Drug Targets in the Postgenomic Era. *BioMed Res. Int.* **2015**, *2015*, 239654. [CrossRef] [PubMed]
2. Mayr, A.; Klambauer, G.; Unterthiner, T.; Steijaert, M.; Wegner, J.K.; Ceulemans, H.; Clevert, D.-A.; Hochreiter, S. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chem. Sci.* **2018**, *9*, 5441–5451. [CrossRef]
3. Lamers, M.M.; Haagmans, B.L. SARS-CoV-2 pathogenesis. *Nat. Rev. Microbiol.* **2022**, *20*, 270–284. [CrossRef] [PubMed]
4. Kumari, R.; Kumar, V.; Dhankhar, P.; Dalal, V. Promising antivirals for PLpro of SARS-CoV-2 using virtual screening, molecular docking, dynamics, and MMPBSA. *J. Biomol. Struct. Dyn.* **2023**, *41*, 4650–4666. [CrossRef] [PubMed]
5. Dhankhar, P.; Dalal, V.; Kumar, V. Screening of Severe Acute Respiratory Syndrome Coronavirus 2 RNA-Dependent RNA Polymerase Inhibitors Using Computational Approach. *J. Comput. Biol.* **2021**, *28*, 1228–1247. [CrossRef] [PubMed]
6. Kumar, K.A.; Sharma, M.; Dalal, V.; Singh, V.; Tomar, S.; Kumar, P. Multifunctional inhibitors of SARS-CoV-2 by MM/PBSA, essential dynamics, and molecular dynamic investigations. *J. Mol. Graph. Model.* **2021**, *107*, 107969. [CrossRef]
7. Dhankhar, P.; Dalal, V.; Singh, V.; Tomar, S.; Kumar, P. Computational guided identification of novel potent inhibitors of N-terminal domain of nucleocapsid protein of severe acute respiratory syndrome coronavirus 2. *J. Biomol. Struct. Dyn.* **2022**, *40*, 4084–4099. [CrossRef]
8. Eweas, A.F.; Osman, H.-E.H.; Naguib, I.A.; Abourehab, M.A.S.; Abdel-Moneim, A.S. Virtual Screening of Repurposed Drugs as Potential Spike Protein Inhibitors of Different SARS-CoV-2 Variants: Molecular Docking Study. *Curr. Issues Mol. Biol.* **2022**, *44*, 3018–3029. [CrossRef]
9. Lynch, M.; Ackerman, M.S.; Gout, J.F.; Long, H.; Sung, W.; Thomas, W.K.; Foster, P.L. Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* **2016**, *17*, 704–714. [CrossRef]
10. Sanjuán, R.; Domingo-Calap, P. Mechanisms of viral mutation. *Cell. Mol. Life Sci.* **2016**, *73*, 4433–4448. [CrossRef]

11. Saifi, S.; Ravi, V.; Sharma, S.; Swaminathan, A.; Chauhan, N.S.; Pandey, R. SARS-CoV-2 VOCs, Mutational diversity and clinical outcome: Are they modulating drug efficacy by altered binding strength? *Genomics* **2022**, *114*, 110466. [CrossRef]

12. Akkiz, H. Implications of the Novel Mutations in the SARS-CoV-2 Genome for Transmission, Disease Severity, and the Vaccine Development. *Front. Med.* **2021**, *8*, 636532. [CrossRef]

13. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* **2017**, *1*, 33–46. [CrossRef] [PubMed]

14. Khare, S.; Gurry, C.; Freitas, L.; Schultz, M.B.; Bach, G.; Diallo, A.; Akite, N.; Ho, J.; Lee, R.T.C.; Yeo, W.; et al. GISAID's Role in Pandemic Response. *China CDC Wkly.* **2021**, *3*, 1049–1051. [CrossRef]

15. Cozzini, P.; Agosta, F.; Dolcetti, G.; Righi, G. How a Blockchain Approach Can Improve Data Reliability in the COVID-19 Pandemic. *ACS Med. Chem. Lett.* **2022**, *13*, 517–519. [CrossRef]

16. Han, W.; Chen, N.; Xu, X.; Sahil, A.; Zhou, J.; Li, Z.; Zhong, H.; Gao, E.; Zhang, R.; Wang, Y.; et al. Predicting the antigenic evolution of SARS-COV-2 with deep learning. *Nat. Commun.* **2023**, *14*, 3478. [CrossRef]

17. Saldivar-Espinoza, B.; Garcia-Segura, P.; Novau-Ferré, N.; Macip, G.; Martínez, R.; Puigbò, P.; Cereto-Massagué, A.; Pujadas, G.; Garcia-Vallve, S. The Mutational Landscape of SARS-CoV-2. *Int. J. Mol. Sci.* **2023**, *24*, 9072. [CrossRef] [PubMed]

18. Palu, A.D.; Torroni, P. 25 years of applications of logic programming in Italy. In *A 25-Year Perspective on Logic Programming*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 300–328.

19. Falkner, A.; Friedrich, G.; Schekotihin, K.; Taupe, R.; Teppan, E.C. Industrial Applications of Answer Set Programming. *Künstliche Intell.* **2018**, *32*, 165–176. [CrossRef]

20. Erdem, E.; Gelfond, M.; Leone, N. Applications of Answer Set Programming. *AI Mag.* **2016**, *37*, 53–68. [CrossRef]

21. Apt, K.R.; Marek, V.W.; Truszczynski, M.; Warren, D.S. *The Logic Programming Paradigm: A 25-Year Perspective*; Springer: Berlin/Heidelberg, Germany, 2012.

22. Edwards, D.; Forster, J.W.; Cogan, N.O.I.; Batley, J.; Chagné, D. Single Nucleotide Polymorphism Discovery. In *Association Mapping in Plants*; Springer: New York, NY, USA, 2007; pp. 53–76. [CrossRef]

23. Petukh, M.; Kucukkal, T.G.; Alexov, E. On Human Disease-Causing Amino Acid Variants: Statistical Study of Sequence and Structural Patterns. *Hum. Mutat.* **2015**, *36*, 524–534. [CrossRef]

24. Kucukkal, T.G.; Petukh, M.; Li, L.; Alexov, E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 18–24. [CrossRef]

25. Harvey, W.T.; Carabelli, A.M.; Jackson, B.; Gupta, R.K.; Thomson, E.C.; Harrison, E.M.; Ludden, C.; Reeve, R.; Rambaut, A.; COVID-19 Genomics UK (COG-UK) Consortium; et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409–424. [CrossRef]

26. Rees-Spear, C.; Muir, L.; Griffith, S.A.; Heaney, J.; Aldon, Y.; Snitselaar, J.L.; Thomas, P.; Graham, C.; Seow, J.; Lee, N.; et al. The effect of spike mutations on SARS-CoV-2 neutralization. *Cell Rep.* **2021**, *34*, 108890. [CrossRef]

27. Gómez, S.A.; Rojas-Valencia, N.; Gómez, S.; Cappelli, C.; Restrepo, A. The Role of Spike Protein Mutations in the Infectious Power of SARS-COV-2 Variants: A Molecular Interaction Perspective. *ChemBioChem* **2022**, *23*, e202100393. [CrossRef]

28. Lin, H.-X.; Feng, Y.; Wong, G.; Wang, L.; Li, B.; Zhao, X.; Li, Y.; Smaill, F.; Zhang, C. Identification of residues in the receptor-binding domain (RBD) of the spike protein of human coronavirus NL63 that are critical for the RBD–ACE2 receptor interaction. *J. Gen. Virol.* **2008**, *89*, 1015–1024. [CrossRef]

29. Li, F.; Li, W.; Farzan, M.; Harrison, S.C. Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science* **2005**, *309*, 1864–1868. [CrossRef]

30. Agosta, F.; Kellogg, G.E.; Cozzini, P. From oncoproteins to spike proteins: The evaluation of intramolecular stability using hydropathic force field. *J. Comput. Aided Mol. Des.* **2022**, *36*, 797–804. [CrossRef] [PubMed]

31. Teng, S.; Sobitan, A.; Rhoades, R.; Liu, D.; Tang, Q. Systemic effects of missense mutations on SARS-CoV-2 spike glycoprotein stability and receptor-binding affinity. *Brief. Bioinform.* **2021**, *22*, 1239–1253. [CrossRef] [PubMed]

32. Laha, S.; Chakraborty, J.; Das, S.; Manna, S.K.; Biswas, S.; Chatterjee, R. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infect. Genet. Evol.* **2020**, *85*, 104445. [CrossRef]

33. Salleh, M.Z.; Derrick, J.P.; Deris, Z.Z. Structural Evaluation of the Spike Glycoprotein Variants on SARS-CoV-2 Transmission and Immune Evasion. *Int. J. Mol. Sci.* **2021**, *22*, 7425. [CrossRef] [PubMed]

34. Barnes, C.O.; Jette, C.A.; Abernathy, M.E.; Dam, K.-M.A.; Esswein, S.R.; Gristick, H.B.; Malyutin, A.G.; Sharaf, N.G.; Huey-Tubman, K.E.; Lee, Y.E.; et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **2020**, *588*, 682–687. [CrossRef] [PubMed]

35. Brown, E.E.F.; Rezaei, R.; Jamieson, T.R.; Dave, J.; Martin, N.T.; Singaravelu, R.; Crupi, M.J.F.; Boulton, S.; Tucker, S.; Duong, J.; et al. Characterization of Critical Determinants of ACE2–SARS CoV-2 RBD Interaction. *Int. J. Mol. Sci.* **2021**, *22*, 2268. [CrossRef]

36. Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020**, *581*, 221–224. [CrossRef] [PubMed]

37. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, *581*, 215–220. [CrossRef] [PubMed]

38. Giordano, D.; de Masi, L.; Argenio, M.A.; Facchiano, A. Structural Dissection of Viral Spike-Protein Binding of SARS-CoV-2 and SARS-CoV-1 to the Human Angiotensin-Converting Enzyme 2 (ACE2) as Cellular Receptor. *Biomedicines* **2021**, *9*, 1038. [CrossRef] [PubMed]

39. Kellogg, G.E.; Semus, S.F.; Abraham, D.J. HINT: A new method of empirical hydrophobic field calculation for CoMFA. *J. Comput. Aided Mol. Des.* **1991**, *5*, 545–552. [CrossRef]

40. Kellogg, G.E.; Abraham, D.J. Hydrophobicity: Is LogPo/w more than the sum of its parts? *Eur. J. Med. Chem.* **2000**, *35*, 651–661. [CrossRef]

41. Nethercote, N.; Stuckey, P.J.; Becket, R.; Brand, S.; Duck, G.J.; Tack, G. MiniZinc: Towards a standard CP modelling language. In Proceedings of the International Conference on Principles and Practice of Constraint Programming, Providence, RI, USA, 23–27 September 2007; Volume 4741 of LNCS.

42. Agosta, F.; Cozzini, P. Hint approach on Transthyretin folding/unfolding mechanism comprehension. *Comput. Biol. Med.* **2023**, *155*, 106667. [CrossRef] [PubMed]

43. Ahmed, M.H.; Catalano, C.; Portillo, S.C.; Safo, M.K.; Scarsdale, N.; Kellogg, G.E. 3D interaction homology: The hydropathic interaction environments of even alanine are diverse and provide novel structural insight. *J. Struct. Biol.* **2019**, *207*, 183–198. [CrossRef]

44. Mughram, M.H.A.L.; Catalano, C.; Bowry, J.P.; Safo, M.K.; Scarsdale, J.N.; Kellogg, G.E. 3D Interaction Homology: Hydropathic Analyses of the 'π–Cation' and 'π–π' Interaction Motifs in Phenylalanine, Tyrosine, and Tryptophan Residues. *J. Chem. Inf. Model.* **2021**, *61*, 2937–2956. [CrossRef]