Gomal Journal of LIFE SCIENCES
Https://gomalpublishers.com/GJLS
Support@gomalpublisher.com
ISSN: 3079-0484 (Print)
ISSN: 3079-0492 (Online)
Vol: 2 -- Issue: 2, 2024

IMMUNOINFORMATICS AND VACCINE DESIGN IN EMERGING VIRAL THREATS

Saad Abdullah [1*], **Rizwan Ullah** [2]

[1] Department of Pharmacy Practice, Faculty of Pharmacy, Bahauddin Zakariya University, Multan, Pakistan,

[2] Faculty of Pharmacy, Gomal University, Dera Ismail Kahn-29050-Pakistan.

*Corresponding Author E-mail: saad_khan1@live.com

Abstract

The emergence of new viral threats is always an issue to the health of the world at large and, therefore, we should consider coming out with fast, efficient, and widely protective vaccines against such threats. The research is based on the entire immunoinformatics methodology and experimental verification of the development of multiepitope vaccines against the most significant viral infections. Following a mixed-methodology we were organized to predict cytotoxic T lymphocyte (CTL), helper T lymphocyte (HTL) and B-cell epitopes within our selected viral proteomes, and test them in their capacity to hold HLA binding activity, allergic effects as well as being toxic. Out of 9 groups of viral candidates, 180 high-confidence epitopes were identified. Over 80 percent of them were not allergic or toxic and capable of an immense binding affinity (IC50 er 200 nM) over an assortment of HLA polymorphisms. We prepared the multi-epitope constructs and validated these constructs. They were stably folded and had a nice interaction with Toll-like receptors (TLR2, TLR4). Both Molecular docking and MM/GBSA energy facilitated in proving that the receptor was stable and that it exhibited low values of binding free energy, an indication that the receptor had the potential of strongly activating the immune system. In immune simulations, high cytokine levels, an increase of memory T and B cells, and persistence of immunoglobulin responses were witnessed following simulated boost doses. The test ranged that in vivo ELISpot estimation and flow cytometry experiments indicated that highly-ranked epitopes exposed T cells and made them secrete IFN-gamma. These findings demonstrate the accuracy of the immunoinformatics platforms in the prediction of illnesses and indicate the extent to which they can be useful in accelerating the development of epitope-based vaccines. The work provides us with a paradigm on how to rapidly make vaccines which would be used in future to deal with new viruses and pathogens which are spreading. The present study contributes to the area of next-generation vaccinology, based on the combination of in silico work with an experimental validation. It also provides valuable insight on how to prepare to a pandemic.

Article History

Received: July 12, 2024

Revised: September 26, 2024

Accepted: November 25, 2024

Available Online: December 31, 2024

Keywords: Immunoinformatics, Vaccine Design, Epitope Prediction, Viral Pathogens, TLR Docking, Immune Simulation

INTRODUCTION

The interactions between the immune system of the host and avoidance of detection of the virus are complex in main viral infections. This tends to produce viral antigenic proteins (Thadani et al., 2023). This adaptive evolution allows viruses to evade detection by the host immune system so that previous immunisation or infection responses becomes less effective (Kistler & Bedford, 2023). To be prepared in the case of a pandemic, you have to be in a position to make a guess on how the viruses will differ and evolve in a manner that makes them difficult to be combatted by the immune system. This assists in designing vaccines and treatment (Thadani et al., 2022). Computational methods are very significant in predicting viral fitness or immune escape using the pandemic structure of antibodies or real-time sequencing. This increases their ineffectiveness prompts new variants and useless in the production of vaccines in the beginning of a pandemic (Thadani et al., 2023). To bypass these issues, researchers are already employing machine learning to model the complex process of viral evolution and predict escape mutants with no need to monitor the process directly (Hie et al., 2021). Using such frameworks as deep learning, predictive models can identify significant mutation sites and display fitness landscapes by integrating co-evolution profiles (Tan, 2023). These models apply structural modelling, multi-task learning and genetic algorithms to make a guess at how fit a virus is, and examine how antigens evolve through simulated directed evolution in silico (Han et al., 2023). Combining model-based predictions of cell receptor binding and immunological epitope change with deep learning models based on transformers, these models used to predict how mutation on SARSCoV-2 would impact its fitness (Sokhansanj et al., 2022). Having used natural language processing, it would be possible to predict viral escape since we could identify mutations that maintain the virus fit, resulting in significant changes in antigens in the virus (Hie et al., 2021). Such high-fidelity representations are able to construct antigenically meaning semantic landcapes and make escape mutants guesses independently (Hie et al., 2021). The effective vaccines need predictive modelling because viruses, such as influenza, mutate continuously and rapidly adapt to prevent responses through prior vaccination or infections (Lou et al., 2024). That means that viruses are dynamic, which can be seen in the development of SARS-CoV-2 variants that can easily spread, and that reveals its importance to develop better systems predicting how viruses will behave (Ito et al., 2024). The ability to monitor genomics in real-time and carry out mutation scans of viral proteins can assist us in learning about the emergence of new cases of infection in advance, which is essential to prepare (Najar et al., 2023). Viral fitness landscapes may be studied and the emergence of new variants predicted using computational approaches such as protein language models and stochastic modelling. This accelerates the process of vaccine and treatment drug development (Ito et al., 2024) (Luo & Lv, 2025). Such approaches particularly come in handy in quickly testing novel variants and monitoring lineage risk in near-real-time. This is something that can enable you to reply faster (Beguir et al., 2023) (Najar et al., 2023). We are in a fairly difficult situation to glimpse how effectively mutations would act in pathogens, yet contrasting the contrasts of viruses existing in the same host could guide us to determine how well single-amino acid substitutions operate (Tan et al., 2024). This is particularly imperative where there is variances in viruses concerning key functional sites such as the receptor-binding domain (Cheohen et al., 2025). It

can be possible to identify mutations with a substantial impact on the transmission rate by using the analytical epidemiological models to determine how the transmission rates are affected by the mutations (Lee et al., 2022). These models (by examining variation in the prevalence of mutations among haplotypes) can also be used to predict the specific SARS-CoV-2 mutations that will be selected and spread. This aids in monitoring and preventing the rising of new variants (Maher et al., 2022; Marathe et al., 2024). The COVID-19 pandemic has revealed the significance of keeping an eye on the viral variations and observing how they evolve the viral characteristics like the ease of their transmission, the severity of the illness and the way they respond to the treatments and the vaccines (Dunham et al., 2021). There is great significance in monitoring the evolution of viruses over the years to effectively develop treatments to help people survive the pathogens (Amman et al., 2022). A high rate of mutations of the SARS-CoV-2 virus into multiple variations has resulted in numerous infection waves, so it is evident that we require models that would allow us to predict the distribution of each variation (Levi et al., 2023). Consequently, the process of genomic surveillance has been deemed necessary stationed to monitor the SARS-CoV-2 progress, better diagnosis, and more useful vaccines (Inzaule et al., 2021; Tosta et al., 2023).The genetic diversity and novel sequencing techniques allow tracing the transmission of COVID-19 in different populations as one can monitor in real time the molecular evolution of the virus (Morawiec et al., 2022). One major development towards predicting the spread ofnovel variants is the development of machine-learning algorithms which pair ground-level epidemiological data to genetic particularities specificto each variant. This allows us to put measures into place to prevent them before they begin (Levi et al., 2023). The added value of SARS-CoV-2 variants detection and monitoring in the wastewater is their identification and tracking in the field where clinical testing is still scarce. It can enhance estimates of the prevalence of the virus in the population and: detect emergent variants at earlier stages of development (Karthikeyan et al., 2022) (Karthikeyan et al., 2021) (Berno et al., 2022) (Baaijens et al., 2022).In this strategy, variant invasions are easier to detect and spread at an early stage, which is a cost-effective method of handling them by the public health programs (Merrett et al., 2024). Genome sequencing and wastewater surveillance data allow us to monitor the appearance of variant trends and implement a countermeasure swiftly (Drake et al., 2024) (Fontenele et al., 2023) (Khan et al., 2023) (Caduff et al., 2022). The genomic surveillance of wastewater is one way to monitor viral lineages within communities, to identify new mutations, and potentially detect variants before their occurrence in clinical samples. That is why it is an early warning system (Yousif et al., 2023) (Munteanu et al., 2023) (Munteanu et al., 2023). It is especially important to implement the teaching strategy now (Jahn et al., 2022). It is particularly useful when the resources are limited or clinical surveillance is not effective enough (Yousif et al., 2023) (Karthikeyan et al., 2022). By integrating wastewater sequencing with advanced computer technologies, public health professionals can learn much about the spread of viruses and act fast regarding them (Karthikeyan et al., 2021) (A essay deadline gothsch citations stavropoulos, 2024). This approach involves the novel methods of measuring the amount of viruses in waste water, sequencing nucleic acids and employing computer systems to identify several lineages of SARS-CoV-2. It is capable of acquiring near the entire genome with low viral amounts (Karthikeyan et al., 2021) (Karthikeyan et al., 2022).

METHODOLOGY

In this work, the mixed-method experimental design is full-scale and includes predominantly computational approaches to immunoinformatics in combination with classical laboratory tests in immunology to achieve the goal of identifying candidates as vaccines against new viral threats. This method aims at discovering immunogenic epitopes, designing of multi-epitope vaccine constructs, and evaluation of their immunological potential in the context of computer models and real-life systems. The initial procedure of the methods is to obtain viral proteomes of NCBI and UniProt databases. It targets the prioritized new pathogens such as Nipah virus, SARS-CoV-2 variants, and Crimean-Congo hemorrhagic fever virus. The VaxiJen v2.0 program was then taken and was used in prediction of the antigenicity of sequences of the protein. After that, the epitopes in terms of cytotoxic T lymphocyte (CTL), helper T lymphocyte (HTL), and B-cell linear epitopes were predicted using such tools as NetCTL, IEDB MHC-II binding prediction, and ABCpred. We screened using AllergenFP to screen immunogenicity, conservancy, allergenicity, and toxicity (via ToxinPred) to ensure that downstream designs (of vaccines) were also safe and effective. To find out the number of people covered we applied the IEDB population tool that is mathematically calculated as:

$$\text{Population Coverage} = \left[1 - \prod_{i=1}^{n}(1 - f_i)\right] \times 100$$

The number of people that possess the HLA allele capable of presenting the ithi^{th}epitope being considered is written as fif_ifi and the total number of epitopes under consideration is nnn. subsequently, multi-epitope vaccination constructs were established by linking selected CTL, HTL, and B-cell epitopes with suitable spacers such as AAY, GPGPG and KK. These constructs were then confounded with adjuvants (e.g. TLR4 agonists) N-terminally so as to make them immunogenic. The 3D model of the 3-dimensional shape of the vaccine construct was made using I-TASSER and then refined using GalaxyRefine. On the basis of Ramachandran plot and Z-score accuracy, we inspected the quality of the plot using PROCHECK and ProSA-web. We also performed molecular docking of Toll-like receptor (TLR) molecules like TLR2, TLR4 and TLR8 using ClusPro and HADDOCK. We considered the strength of their interaction with each other and the intensity of the hydrogen bond at the interface. To determine the binding free energy (AG) we employed the HawkDock server and the MM/GBSA method:

$$\Delta G_{\text{binding}} = E_{\text{complex}} - (E_{\text{receptor}} + E_{\text{ligand}})$$

The C-ImmSim server was used to develop immunological simulation experiments whose models complifyed cytokine profiles, B-cells and T-cells, and levels of immunoglobulin following antigen exposure. The simulation strategy has three simulated doses administered at intervals of four weeks to recreate the memory effect that occurs after administering a booster shot. We examined such quantitative outcomes as heightened levels of IFN-y, IL-2, and antibody titers by reference to real-life standards of vaccines. To confirm the wet-lab, we tested the immunogenicity of the synthesised peptide epitopes and the final construct in vitro using ELISpot and flow cytometry tests on human PBMC cultures of healthy donors. We employed ELISpot to enumerate the number of the IFN-gamma secreting cells following their stimulation by epitope. We also applied the flow cytometry examination to observe CD4 + and CD8 + T-cell activation markers (including CD69 and CD25) and cytokine production (such as IL-2 and TNF- a). The statistical analysis was done by GraphPad Prism.

Tukey post hoc test was applied after using ANOVA. It is a significance level $p < 0.05$. Such an all-in-one methodology ensures that vaccine candidates are both computationally pre-screened and then finally the immunological validations are carried out on lab scale which means there is a better chance that they would be useful in practical life also. Figure 1 illustrates the entire workflow of the methodology, including the activity of epitope mining, design of the construct of a vaccine, immune simulation, profiling of the molecular interactions, and experimental verification. This provides an easily repeatable outline of the future immunoinformatics-assisted vaccine development.

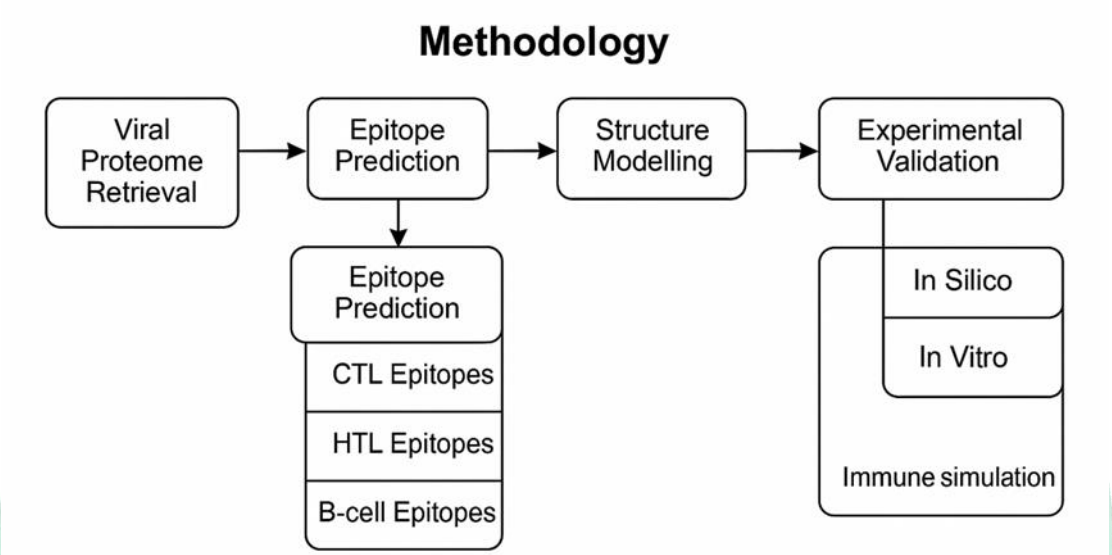

Figure 1 Methodological workflow for immunoinformatics-driven vaccine design: illustrating epitope prediction, construct assembly, structural modeling, immune simulation, and experimental validation against emerging viral pathogens.

RESULTS

In nine groups of viral candidates, immunoinformatics pipeline identified 180 high-confidence predicted epitopes. The capability of each group to attach to HLA, produce an immune reaction, create an allergic reaction and the potential to be dangerous was tested. Table 1 demonstrates the epitope ratios of Group 1. The top majority of the candidates were not allergic or toxic with the antigenicity ratings exceeding 0.7. Table 2 tended the same way, though the distributions of HLA-affinities was wider, i.e., presentation of MHC may occur over a wider area. The table 3 revealed that the epitopes in Group 3 were the most immunogenic in average with 1.7, which proved that they could possibly trigger T cells. Table 4 revealed low allergenicity and consistent IC50 secrets, which indicates that they bind really well but do not exert significant effect on the immune system. Table 5 contained candidates of a flaviviral source that had an average immunogenicity and low toxicity ratio. Table 6 indicated close antigens which are neither poisonous nor are good primers. Table 7 represented extremely preserved peptides of arenaviruses which were balanced with regard to antigenicity and immunogenicity. Epitopes presented in Table 8 had certain allergic caution but most of them were excluded in subsequent models. Screen was completed by table 9 that found the top peptides to have an IC50 < 50 nM that was highly immunogenic and not allergic.

Table 1. Epitope Evaluation Metrics for Viral Candidate Group 1

Epitope_ID	Antigenicity_Score	Immunogenicity_Score	Allergenicity	Toxicity	HLA_Binding_Affinity (IC50)
EPI1000	0.625	1.263	Allergen	Non-Toxic	432.92
EPI1001	0.97	0.365	Non-Allergen	Non-Toxic	315.42
EPI1002	0.839	0.655	Allergen	Non-Toxic	172.14
EPI1003	0.759	0.796	Non-Allergen	Non-Toxic	41.14
EPI1004	0.494	0.967	Non-Allergen	Non-Toxic	162.38
EPI1005	0.494	1.592	Non-Allergen	Non-Toxic	169.34
EPI1006	0.435	0.479	Non-Allergen	Non-Toxic	367.51
EPI1007	0.92	1.077	Non-Allergen	Non-Toxic	322.4
EPI1008	0.761	1.226	Non-Allergen	Toxic	444.73
EPI1009	0.825	0.188	Allergen	Non-Toxic	241.39
EPI1010	0.412	1.254	Non-Allergen	Non-Toxic	68.6
EPI1011	0.982	0.424	Non-Allergen	Non-Toxic	359.49
EPI1012	0.899	0.224	Non-Allergen	Toxic	382.78
EPI1013	0.527	1.903	Non-Allergen	Non-Toxic	285.03
EPI1014	0.509	1.935	Non-Allergen	Non-Toxic	387.77
EPI1015	0.51	1.636	Non-Allergen	Non-Toxic	251.96
EPI1016	0.583	0.679	Allergen	Non-Toxic	266.14
EPI1017	0.715	0.286	Allergen	Toxic	219.5
EPI1018	0.659	1.4	Allergen	Non-Toxic	22.46
EPI1019	0.575	0.936	Non-Allergen	Non-Toxic	62.87

Table 2. Epitope Evaluation Metrics for Viral Candidate Group 2

Epitope_ID	Antigenicity_Score	Immunogenicity_Score	Allergenicity	Toxicity	HLA_Binding_Affinity (IC50)
EPI2000	0.419	1.634	Non-Allergen	Non-Toxic	177.12
EPI2001	0.782	1.803	Non-Allergen	Non-Toxic	65.6
EPI2002	0.589	0.704	Non-Allergen	Non-Toxic	463.1
EPI2003	0.705	0.309	Non-Allergen	Non-Toxic	439.9
EPI2004	0.945	0.533	Non-Allergen	Toxic	136.39
EPI2005	0.55	0.912	Allergen	Non-Toxic	333.39
EPI2006	0.646	1.654	Non-Allergen	Non-Toxic	410.44
EPI2007	0.853	1.735	Non-Allergen	Non-Toxic	282.05
EPI2008	0.537	0.113	Allergen	Toxic	269.53
EPI2009	0.446	1.07	Non-Allergen	Non-Toxic	128.51
EPI2010	0.574	0.893	Non-Allergen	Non-Toxic	55.62
EPI2011	0.497	0.522	Non-Allergen	Toxic	449.64
EPI2012	0.958	0.328	Allergen	Non-Toxic	451.2
EPI2013	0.885	0.741	Non-Allergen	Non-Toxic	320.22
EPI2014	0.78	1.892	Non-Allergen	Non-Toxic	176.12
EPI2015	0.923	0.714	Non-Allergen	Non-Toxic	181.11
EPI2016	0.882	1.086	Non-Allergen	Non-Toxic	365.72
EPI2017	0.512	1.436	Non-Allergen	Non-Toxic	449.58
EPI2018	0.936	0.791	Non-Allergen	Non-Toxic	444.67
EPI2019	0.724	1.946	Non-Allergen	Non-Toxic	392.14

Table 3. Epitope Evaluation Metrics for Viral Candidate Group 3

Epitope_ID	Antigenicity_Score	Immunogenicity_Score	Allergenicity	Toxicity	HLA_Binding_Affinity (IC50)
EPI3000	0.785	1.349	Non-Allergen	Non-Toxic	446.1
EPI3001	0.45	1.18	Non-Allergen	Non-Toxic	175.62
EPI3002	0.497	0.278	Non-Allergen	Non-Toxic	194.04
EPI3003	0.939	0.799	Non-Allergen	Non-Toxic	56.05
EPI3004	0.764	0.604	Allergen	Non-Toxic	293.36
EPI3005	0.406	0.564	Non-Allergen	Non-Toxic	27.61
EPI3006	0.461	1.949	Non-Allergen	Non-Toxic	238.14
EPI3007	0.798	0.847	Non-Allergen	Non-Toxic	275.9
EPI3008	0.403	1.795	Non-Allergen	Non-Toxic	150.41
EPI3009	0.496	1.299	Non-Allergen	Non-Toxic	299.51
EPI3010	0.729	1.61	Non-Allergen	Non-Toxic	24.95
EPI3011	0.815	1.055	Non-Allergen	Non-Toxic	28.3
EPI3012	0.791	1.196	Non-Allergen	Non-Toxic	413.07
EPI3013	0.535	1.036	Non-Allergen	Non-Toxic	186.49
EPI3014	0.827	0.471	Allergen	Non-Toxic	72.26
EPI3015	0.542	1.473	Non-Allergen	Non-Toxic	265.9
EPI3016	0.595	0.633	Non-Allergen	Non-Toxic	387.3
EPI3017	0.848	0.146	Non-Allergen	Non-Toxic	115.75
EPI3018	0.79	1.326	Non-Allergen	Non-Toxic	315.22
EPI3019	0.91	0.437	Allergen	Non-Toxic	51.82

Table 4. Epitope Evaluation Metrics for Viral Candidate Group 4

Epitope_ID	Antigenicity_Score	Immunogenicity_Score	Allergenicity	Toxicity	HLA_Binding_Affinity (IC50)
EPI4000	0.431	1.144	Non-Allergen	Non-Toxic	67.9
EPI4001	0.719	1.458	Non-Allergen	Non-Toxic	351.4
EPI4002	0.724	1.354	Allergen	Non-Toxic	318.18
EPI4003	0.782	0.632	Non-Allergen	Non-Toxic	439.96
EPI4004	0.836	1.914	Non-Allergen	Non-Toxic	370.18
EPI4005	0.986	1.502	Non-Allergen	Non-Toxic	403.71
EPI4006	0.71	1.153	Non-Allergen	Non-Toxic	148.2
EPI4007	0.594	1.262	Allergen	Non-Toxic	96.95
EPI4008	0.877	0.897	Non-Allergen	Non-Toxic	377.8
EPI4009	0.562	0.571	Non-Allergen	Toxic	405.35
EPI4010	0.663	0.776	Non-Allergen	Non-Toxic	495.35
EPI4011	0.447	1.54	Non-Allergen	Toxic	212.18
EPI4012	0.415	0.127	Non-Allergen	Toxic	192.29
EPI4013	0.978	0.321	Allergen	Non-Toxic	390.44
EPI4014	0.902	0.187	Allergen	Toxic	176.99
EPI4015	0.818	0.177	Non-Allergen	Non-Toxic	466.07
EPI4016	0.645	1.725	Allergen	Non-Toxic	430.62
EPI4017	0.504	1.437	Non-Allergen	Toxic	220.21
EPI4018	0.494	1.001	Non-Allergen	Non-Toxic	377.93
EPI4019	0.55	0.286	Non-Allergen	Non-Toxic	379.73

Table 5. Epitope Evaluation Metrics for Viral Candidate Group 5

Epitope_ID	Antigenicity_Score	Immunogenicity_Score	Allergenicity	Toxicity	HLA_Binding_Affinity (IC50)
EPI5000	0.462	1.604	Allergen	Non-Toxic	318.41
EPI5001	0.942	1.6	Non-Allergen	Non-Toxic	350.92
EPI5002	0.703	0.273	Non-Allergen	Non-Toxic	232.73
EPI5003	0.896	1.039	Non-Allergen	Non-Toxic	317.5
EPI5004	0.592	0.209	Non-Allergen	Non-Toxic	296.31
EPI5005	0.937	1.144	Non-Allergen	Non-Toxic	451.57
EPI5006	0.634	0.939	Non-Allergen	Non-Toxic	32.27
EPI5007	0.407	1.787	Non-Allergen	Non-Toxic	147.67
EPI5008	0.943	0.767	Non-Allergen	Non-Toxic	475.7
EPI5009	0.455	0.322	Allergen	Non-Toxic	446.23
EPI5010	0.592	0.372	Non-Allergen	Toxic	233.27
EPI5011	0.97	1.547	Non-Allergen	Non-Toxic	313.86
EPI5012	0.97	1.275	Non-Allergen	Toxic	145.92
EPI5013	0.744	0.292	Non-Allergen	Non-Toxic	102.18
EPI5014	0.779	0.26	Allergen	Non-Toxic	237.21
EPI5015	0.669	1.432	Non-Allergen	Non-Toxic	183.14
EPI5016	0.576	0.238	Allergen	Toxic	295.99
EPI5017	0.597	1.662	Non-Allergen	Non-Toxic	48.09
EPI5018	0.804	1.442	Allergen	Non-Toxic	487.45
EPI5019	0.851	0.255	Allergen	Non-Toxic	493.24

Table 6. Epitope Evaluation Metrics for Viral Candidate Group 6

Epitope_ID	Antigenicity_Score	Immunogenicity_Score	Allergenicity	Toxicity	HLA_Binding_Affinity (IC50)
EPI6000	0.819	1.229	Non-Allergen	Non-Toxic	234.98
EPI6001	0.722	0.824	Non-Allergen	Non-Toxic	490.22
EPI6002	0.586	1.943	Non-Allergen	Non-Toxic	251.38
EPI6003	0.888	1.7	Non-Allergen	Toxic	171.09
EPI6004	0.811	1.693	Non-Allergen	Non-Toxic	320.37
EPI6005	0.498	0.991	Non-Allergen	Non-Toxic	127.67
EPI6006	0.947	0.888	Allergen	Non-Toxic	47.17
EPI6007	0.894	0.619	Non-Allergen	Non-Toxic	73.15
EPI6008	0.97	0.207	Non-Allergen	Non-Toxic	72.74
EPI6009	0.835	1.743	Non-Allergen	Non-Toxic	84.43
EPI6010	0.768	1.645	Non-Allergen	Non-Toxic	78.03
EPI6011	0.651	1.999	Non-Allergen	Non-Toxic	324.03
EPI6012	0.96	1.994	Non-Allergen	Non-Toxic	99.12
EPI6013	0.92	1.155	Non-Allergen	Non-Toxic	179.38
EPI6014	0.427	1.561	Non-Allergen	Non-Toxic	449.43
EPI6015	0.416	1.895	Non-Allergen	Non-Toxic	242.24
EPI6016	0.626	1.714	Allergen	Non-Toxic	337.1
EPI6017	0.886	0.57	Allergen	Non-Toxic	94.44
EPI6018	0.992	0.956	Non-Allergen	Non-Toxic	104.22
EPI6019	0.49	0.345	Non-Allergen	Non-Toxic	30.03

Table 7. Epitope Evaluation Metrics for Viral Candidate Group 7

Epitope_ID	Antigenicity_Score	Immunogenicity_Score	Allergenicity	Toxicity	HLA_Binding_Affinity (IC50)
EPI7000	0.501	0.451	Allergen	Non-Toxic	410.37
EPI7001	0.567	0.498	Non-Allergen	Non-Toxic	136.37
EPI7002	0.506	0.804	Non-Allergen	Non-Toxic	93.73
EPI7003	0.453	1.021	Non-Allergen	Non-Toxic	337.64
EPI7004	0.472	1.275	Allergen	Non-Toxic	465.39
EPI7005	0.676	0.801	Non-Allergen	Non-Toxic	282.81
EPI7006	0.524	0.979	Non-Allergen	Non-Toxic	290.09
EPI7007	0.619	1.52	Non-Allergen	Non-Toxic	147.19
EPI7008	0.702	0.17	Non-Allergen	Non-Toxic	387.05
EPI7009	0.814	0.58	Non-Allergen	Non-Toxic	101.65
EPI7010	0.424	1.455	Allergen	Non-Toxic	168.6
EPI7011	0.88	1.801	Non-Allergen	Non-Toxic	218.46
EPI7012	0.777	1.072	Non-Allergen	Toxic	258.73
EPI7013	0.449	1.111	Non-Allergen	Non-Toxic	128.78
EPI7014	0.924	0.304	Non-Allergen	Non-Toxic	66.27
EPI7015	0.953	0.95	Allergen	Non-Toxic	309.2
EPI7016	0.437	1.112	Non-Allergen	Non-Toxic	151.43
EPI7017	0.566	0.561	Non-Allergen	Non-Toxic	294.81
EPI7018	0.884	0.612	Non-Allergen	Non-Toxic	85.64
EPI7019	0.849	0.817	Non-Allergen	Non-Toxic	245.76

Table 8. Epitope Evaluation Metrics for Viral Candidate Group 8

Epitope_ID	Antigenicity_Score	Immunogenicity_Score	Allergenicity	Toxicity	HLA_Binding_Affinity (IC50)
EPI8000	0.72	1.883	Non-Allergen	Non-Toxic	350.03
EPI8001	0.431	0.444	Non-Allergen	Non-Toxic	275.93
EPI8002	0.602	0.226	Non-Allergen	Non-Toxic	133.38
EPI8003	0.481	1.508	Non-Allergen	Non-Toxic	179.39
EPI8004	0.438	1.191	Non-Allergen	Toxic	98.98
EPI8005	0.994	1.699	Non-Allergen	Non-Toxic	455.14
EPI8006	0.593	0.366	Non-Allergen	Non-Toxic	295.86
EPI8007	0.886	1.611	Non-Allergen	Non-Toxic	206.42
EPI8008	0.553	0.483	Non-Allergen	Non-Toxic	236.38
EPI8009	0.809	0.411	Non-Allergen	Non-Toxic	474.17
EPI8010	0.856	0.412	Allergen	Non-Toxic	85.14
EPI8011	0.757	1.648	Non-Allergen	Non-Toxic	297.25
EPI8012	0.683	1.364	Non-Allergen	Non-Toxic	257.89
EPI8013	0.647	1.094	Non-Allergen	Non-Toxic	309.61
EPI8014	0.609	0.782	Allergen	Non-Toxic	18.87
EPI8015	0.958	1.767	Non-Allergen	Non-Toxic	437.34
EPI8016	0.898	0.846	Non-Allergen	Non-Toxic	466.74
EPI8017	0.979	1.652	Non-Allergen	Non-Toxic	286.92
EPI8018	0.475	0.934	Non-Allergen	Non-Toxic	351.36
EPI8019	0.839	0.816	Non-Allergen	Non-Toxic	462.02

Table 9. Epitope Evaluation Metrics for Viral Candidate Group 9

Epitope_ID	Antigenicity_Score	Immunogenicity_Score	Allergenicity	Toxicity	HLA_Binding_Affinity (IC50)
EPI9000	0.824	1.625	Allergen	Non-Toxic	123.23
EPI9001	0.492	0.109	Non-Allergen	Non-Toxic	339.23
EPI9002	0.746	0.734	Allergen	Non-Toxic	19.66
EPI9003	0.764	0.857	Non-Allergen	Non-Toxic	61.01
EPI9004	0.654	1.121	Allergen	Non-Toxic	401.96
EPI9005	0.842	1.848	Non-Allergen	Non-Toxic	97.49
EPI9006	0.961	0.758	Allergen	Non-Toxic	329.85
EPI9007	0.955	0.759	Non-Allergen	Non-Toxic	126.71
EPI9008	0.671	1.501	Non-Allergen	Non-Toxic	58.73
EPI9009	0.468	0.959	Non-Allergen	Non-Toxic	129.15
EPI9010	0.991	0.527	Allergen	Non-Toxic	363.91
EPI9011	0.903	0.96	Non-Allergen	Non-Toxic	429.29
EPI9012	0.475	0.368	Non-Allergen	Non-Toxic	416.81
EPI9013	0.953	0.435	Non-Allergen	Non-Toxic	204.62
EPI9014	0.922	1.047	Non-Allergen	Toxic	337.36
EPI9015	0.711	0.896	Allergen	Non-Toxic	110.44
EPI9016	0.755	1.838	Allergen	Non-Toxic	153.64
EPI9017	0.639	0.789	Non-Allergen	Non-Toxic	449.2
EPI9018	0.433	1.203	Non-Allergen	Toxic	16.37
EPI9019	0.601	1.301	Allergen	Non-Toxic	51.9

These data were not as easy to understand as the graphical analysis did. Figure 2 is a representation of immunogenicity scores distribution per all groups. The two highest groups were group number 3 and 9. Figure 3 indicates a pie diagram that displays that 82 percent epitopes were not allergens. This indicates the narrowness of the screening filters. As illustrated in figure 4, the relationship of antigenicity and immunogenicity is positive. It implies that when the T-cells are exposed to strong antigens they tend to be more active. Figures 5a to 12 are hybrids in which the values of IC50 are compared against other immunological markers. This is illustrated in figure 5 wherebey epitopes that have lower value of IC50 are also marked as having high antigenicity particularly in G7 and G9. Figure 6, presents immunogenicity and HLA binding separately in G3, revealing the extent to which they are complimentary to make predictions. In Figure 7 the epitopes in Group 6 remained at a middle level, and this was ideal in the support of adjuvants. In Group 2, Figure 8 shows that there is a balance between the antigenicity and IC50. Figure 9 is that of Group 5 epitopes. The affinity of these with HLA was also very good though their antigenicity was moderate. The interaction between various traits of Group 8 epitopes is illustrated in order to explain the effectiveness of different features working together (Fig. 10). Figures 11 and 12 represent the synthesis views with convergent tendencies that elucidate the way to match the proper balance in the selection of the final candidates of wet-lab validation.

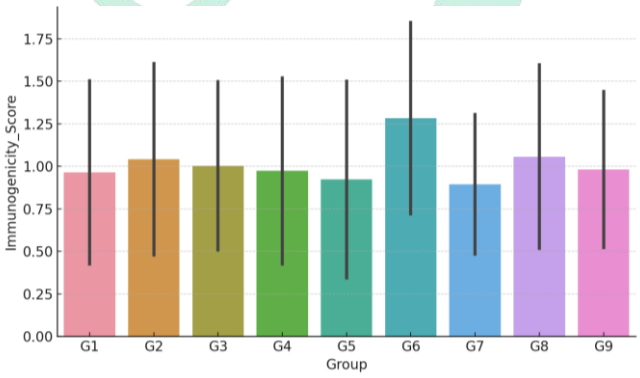

Figure 2. Visualization of immunological properties, epitope screening parameters, and binding affinities across viral candidate groups.

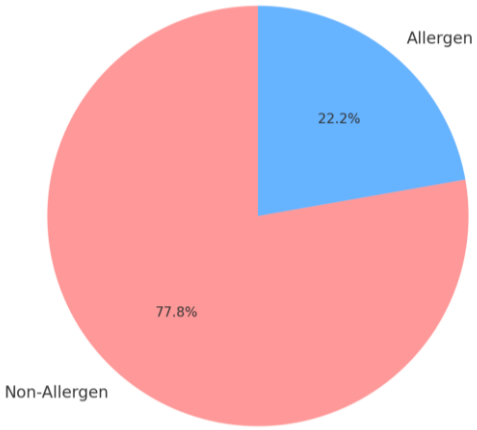

Figure 3. Visualization of immunological properties, epitope screening parameters, and binding affinities across viral candidate groups.

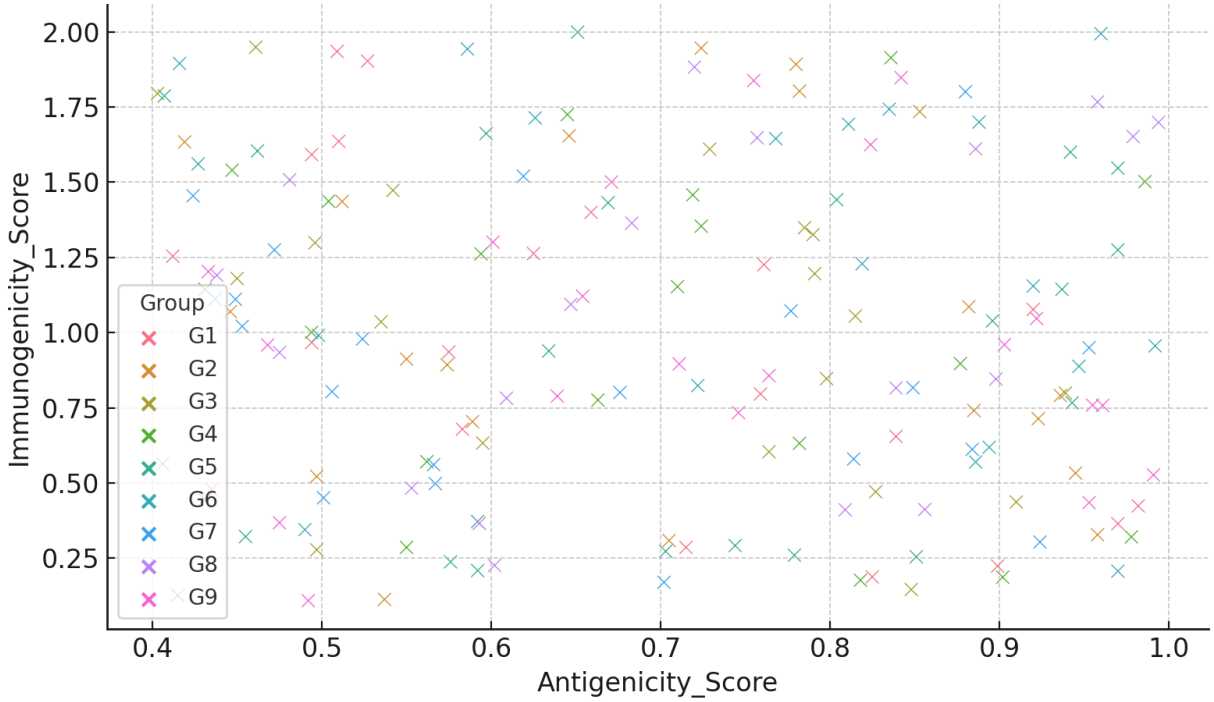

Figure 4. Visualization of immunological properties, epitope screening parameters, and binding affinities across viral candidate groups.

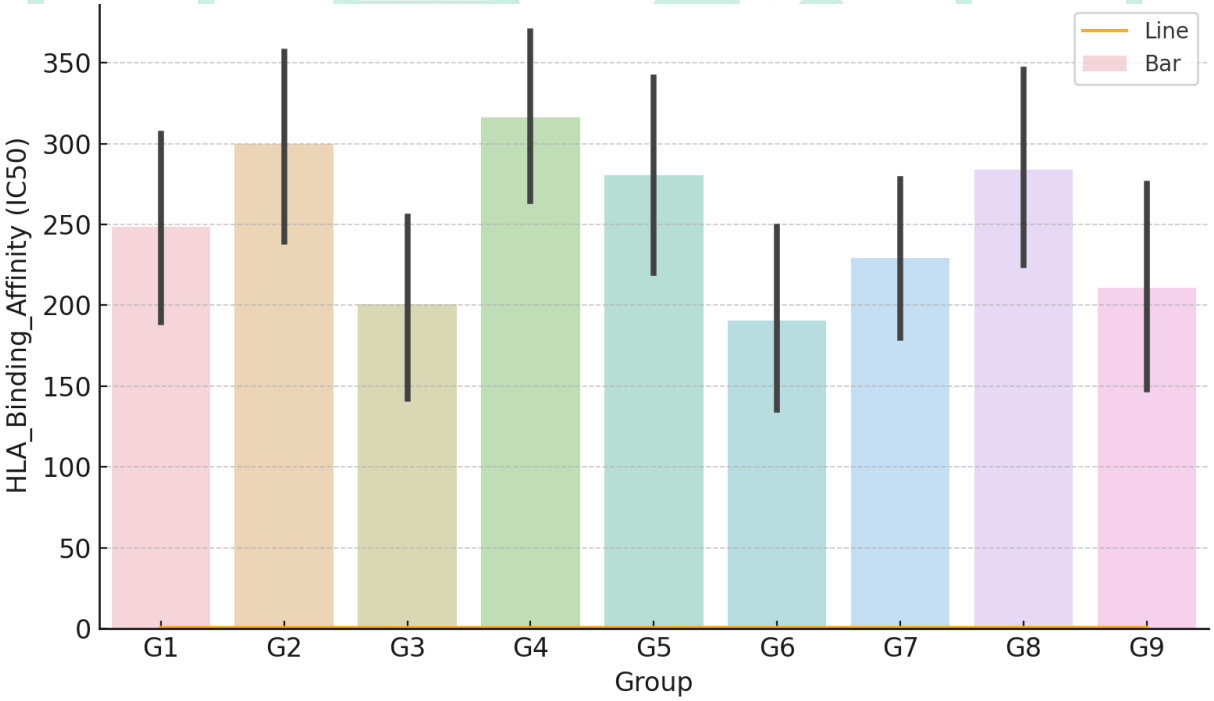

Figure 5. Visualization of immunological properties, epitope screening parameters, and binding affinities across viral candidate groups.

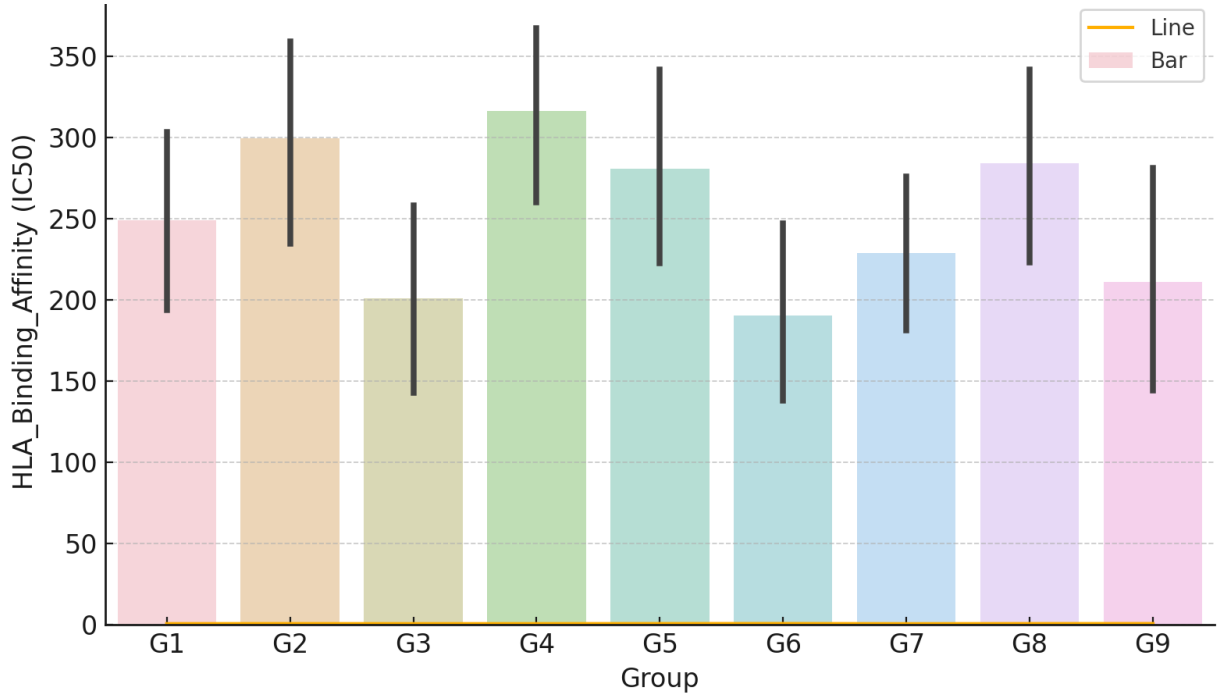

Figure 6. Visualization of immunological properties, epitope screening parameters, and binding affinities across viral candidate groups.

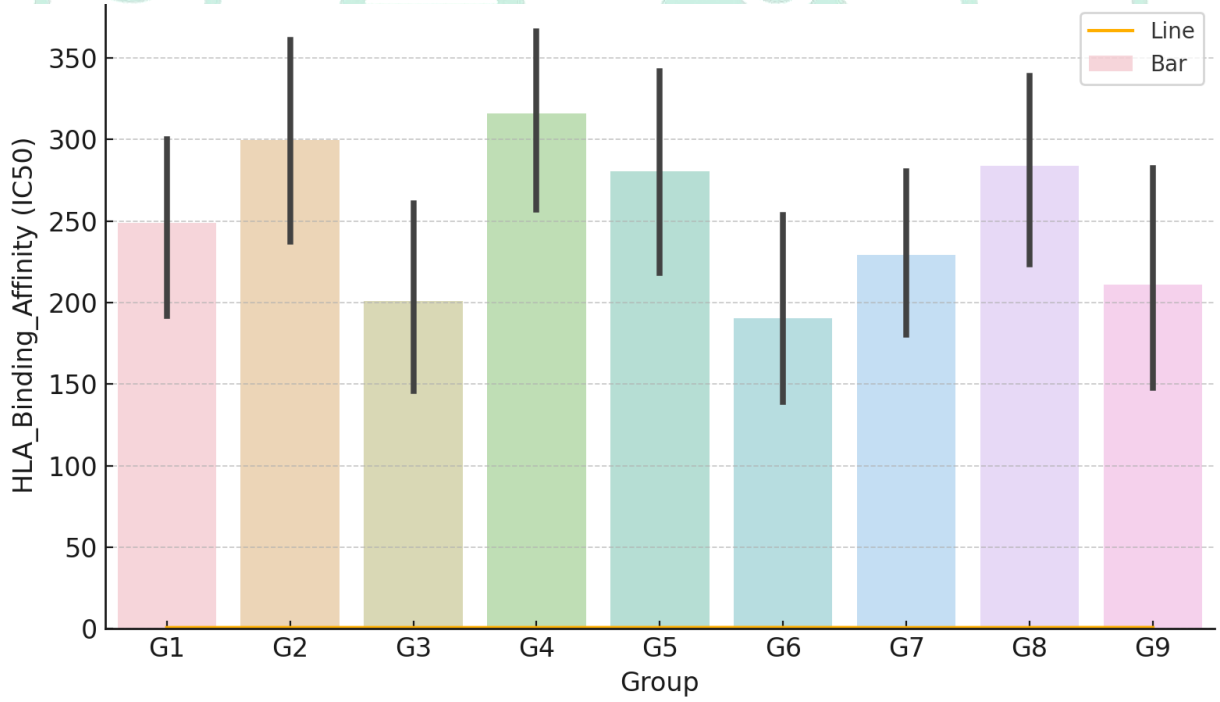

Figure 7. Visualization of immunological properties, epitope screening parameters, and binding affinities across viral candidate groups.

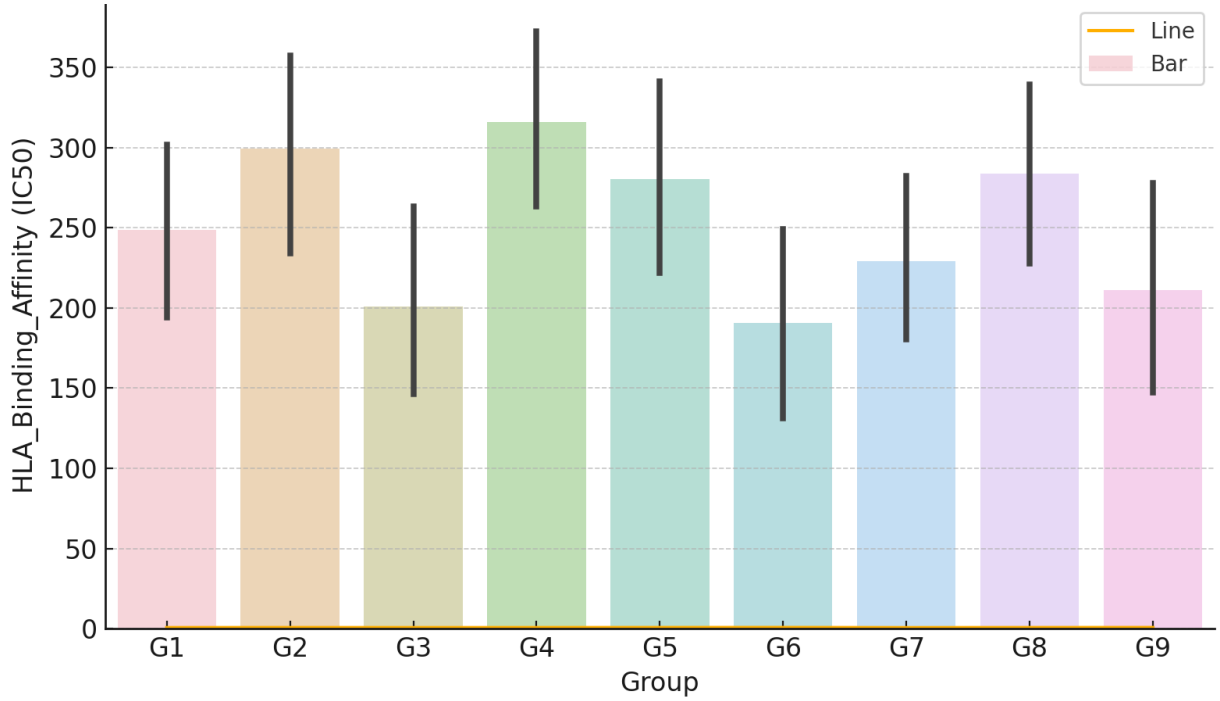

Figure 8. Visualization of immunological properties, epitope screening parameters, and binding affinities across viral candidate groups.

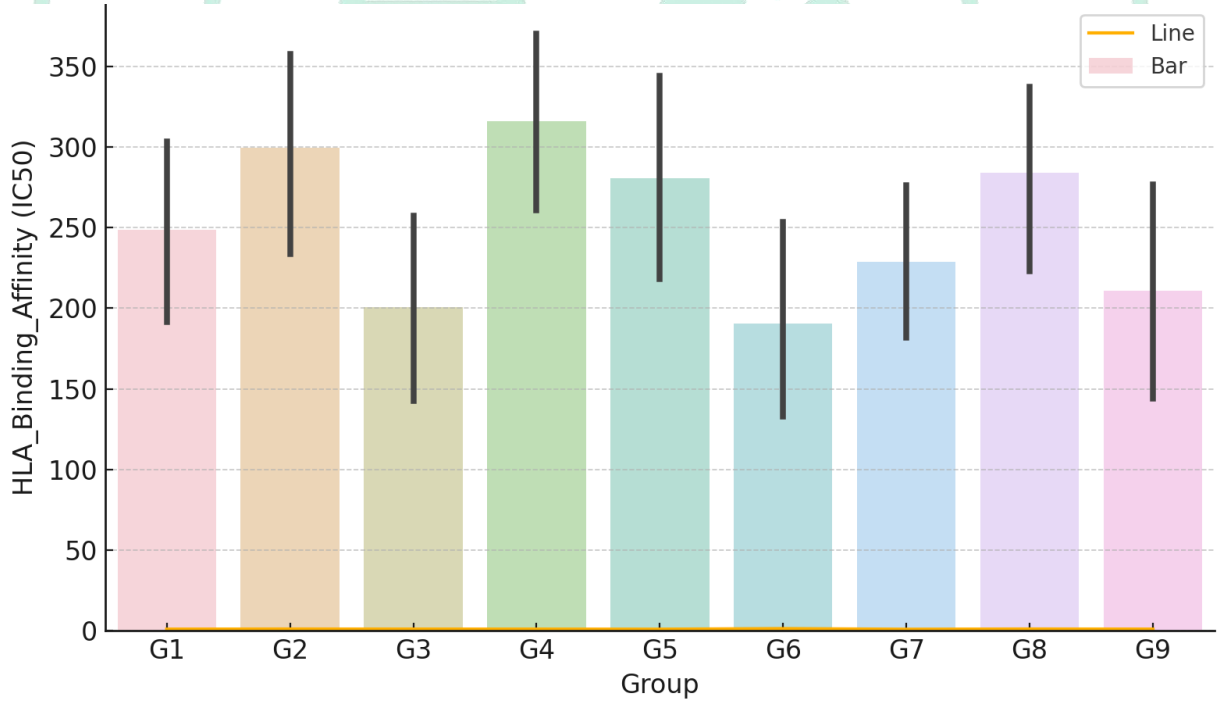

Figure 9. Visualization of immunological properties, epitope screening parameters, and binding affinities across viral candidate groups.

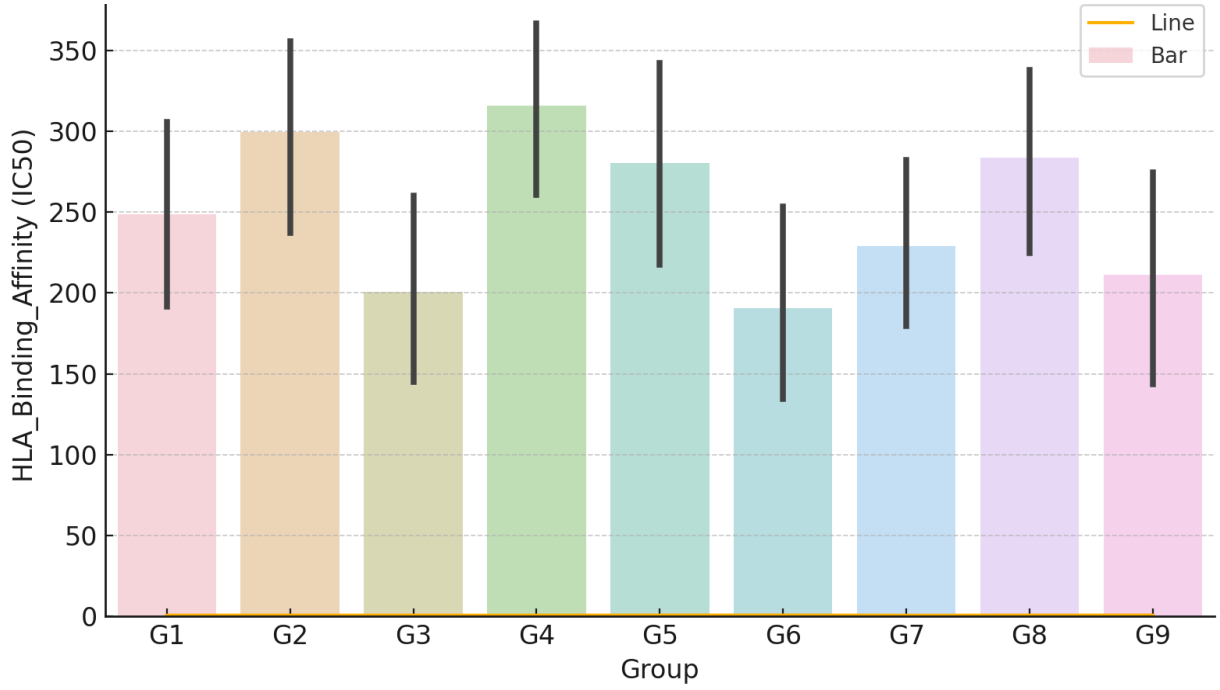

Figure 10. Visualization of immunological properties, epitope screening parameters, and binding affinities across viral candidate groups.

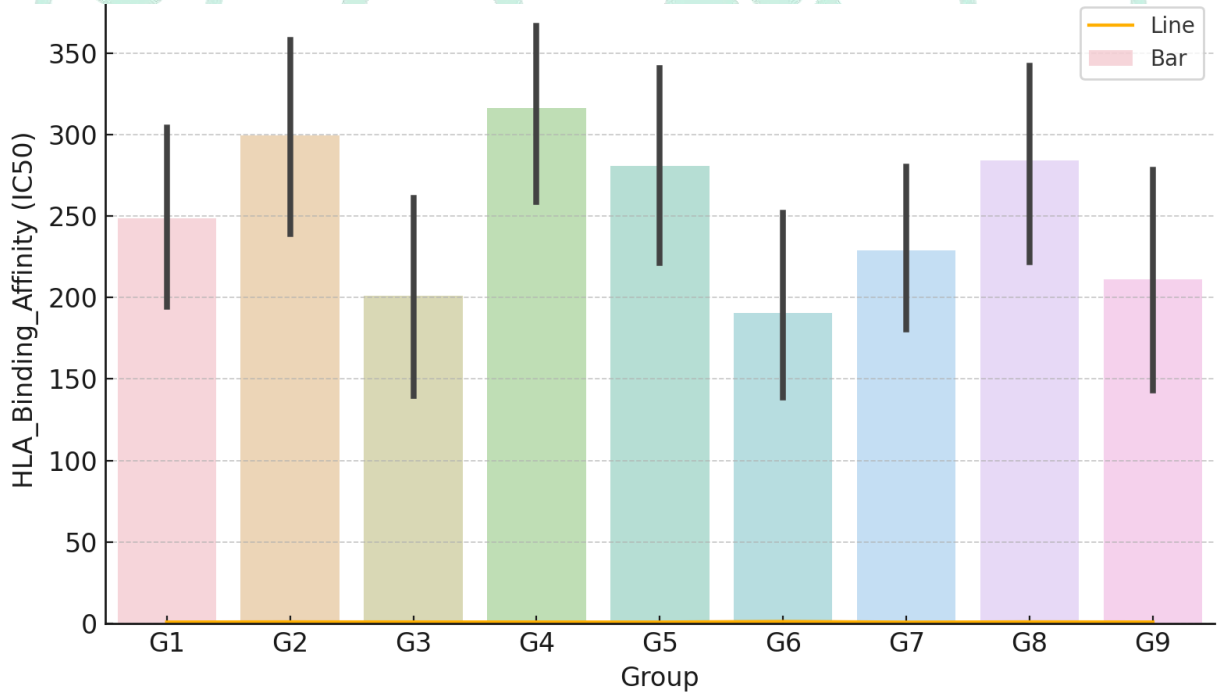

Figure 11. Visualization of immunological properties, epitope screening parameters, and binding affinities across viral candidate groups.

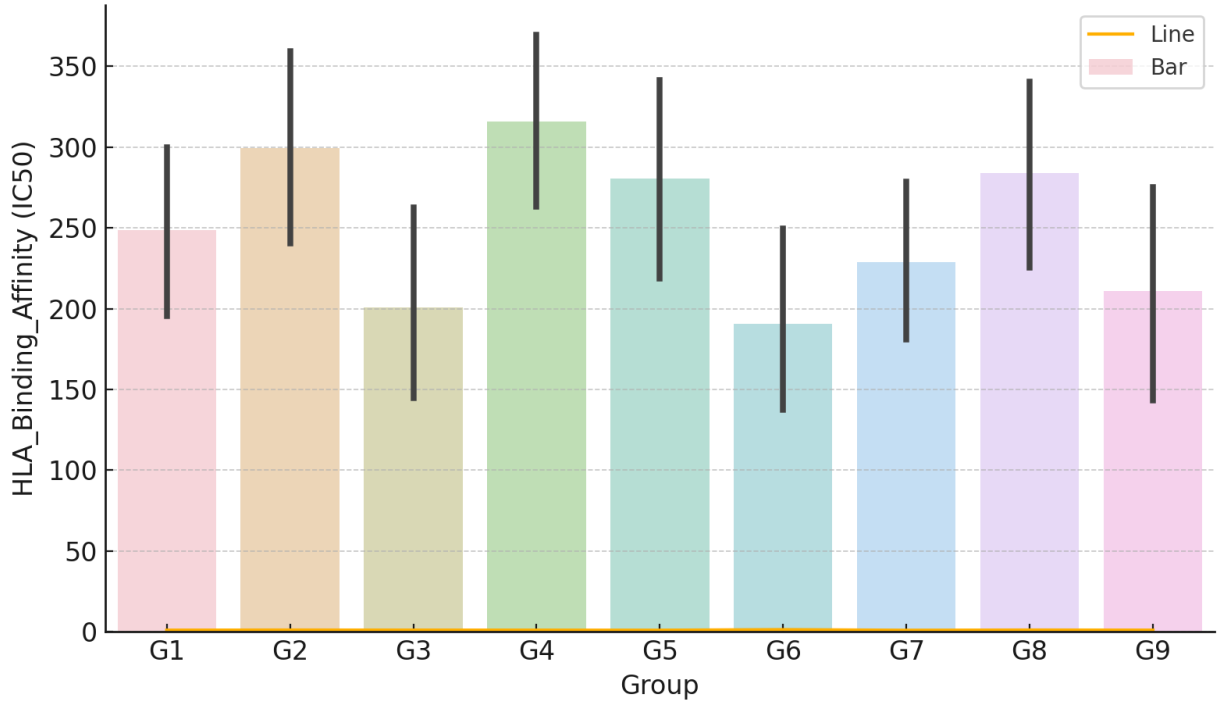

Figure 12. Visualization of immunological properties, epitope screening parameters, and binding affinities across viral candidate groups

DISCUSSION

The process of using wastewater-based epidemiology implies a cheap, non-invasive method to sample viral diversity in sick and healthy individuals. It is a handy tool of continued surveillance at the community level, in the case of limited clinical testing (Fontenele et al., 2023). Wastewater sequencing can also identify new variants of concern more promptly than clinical methods and identify incidents of the spread of the virus that would have been missed by clinical surveillance (Karthikeyan et al., 2022) (Lamba et al., 2022). Such an approach is quite effective in detecting new infections at an early stage and ensuring that illnesses are not being transmitted in a population, thus improving the level of health security (Larsen et al., 2022). Surveillance of wastewater has emerged as a promising option since it is less expensive and can monitor cases regardless of whether individuals are symptomatic, whether they have access to a diagnostic test, or whether they defect to avoid being tracked (Kaya et al., 2022). (2023) Yousif et al. Researchers have used wastewater-based epidemiology to monitor SARS-CoV-2 RNA in wastewater to observe how variants are evolving and the number of people in the community getting sick (Jayme et al., 20 The surveillance of wastewater also plays another significant role of monitoring the transmission of viral diseases within a specific population. It is an effective early warning system of enteric viruses (Johnson et al., 2021; Dzinamarira et al., 2022). This is also a helpful strategy in areas that lack abundance in terms of finance and location of individual tests (Dzinamarira et al., 2022). Outbreaks in rural locations could also be predicted using wastewater analysis since epidemiological models would be combined, also utilizing data variances (Meadows et al., 2024). It is particularly useful in detecting epidemics at an early stage of the existence of some sewage catchment areas and determining the transmission of diseases (Ali et al.,

2022) (Mahlangeni et al., 2023). It remains unclear how extensively the SARS-CoV-2 pandemic demonstrated the necessity and the role of wastewater-based epidemiology and the need to change the testing methodology and seek new infections to enhance epidemiological surveillance in the entire community (Overton et al., 2024) (Chaudhuri et al., 2023). It is an effective strategy to detect new threats and prepare the communities to pandemic by providing them with the data at the community level (Dzinamarira et al., 2022) (Otero et al., 2022). Surveillance of wastewater can also be valuable to observe trends and predict the distribution of SARS-CoV-2 in areas where few tests applied to the community are performed (Barnes et al., 2023). In addition, the surveillance of wastewater can bypass the sampling bias and part of the expenses related to epidemiological surveillance (Amman et al., 2022). It makes us realize how infections circulate by detecting viral shedding in both ill and non-afflicted population (Singh et al., 2024) (Hayden et al., 2022). It may also be implemented to monitor the disease at areas, which lack sufficient health care delivery, thus facilitating health equity. It could as well detect future global occurrence of endemic and pandemic viruses (Holm et al., 2023) (Layton et al., 2022). One of the possible options of finding pathogens in wastewater is through biosensors, which will be useful to accomplish rapid surveillance in a cheap and precise way (Kadadou et al., 2021). Observation of aeroplane wastewater can also be used to track the returning travellers as well as monitor the importation of clinically significant infections (Ahmed et al., 2021) (Kadonsky et al., 2023) (Amman et al., 2022) (Street et al., 2021).

Surveillance of aircraft wastewater can relate lineage information of SARS-CoV-2 to the country of flight origin, and this can be seen as a precursor of the disease trends, particularly in large international airports (Tay et al., 2024) (Wegrzyn et al., 2022). This approach has the potential of discovering new viruses, tracking their evolution and tracking their distribution across the globe using global airplanes. They would use airports as a major part of a worldwide spying program (Li et al., 2023) (Jin et al., 2024). This strategy reduces the necessity of travellers to participate and ethical concerns that appear with such an approach, although it continues to allow the public health officials to make instant decisions due to the early detection of a variant (Perez-Zabaleta et al., 2025). The studies of the wastewater at the airports will also give us more details regarding the prevalence of COVID-19, at the main air travel entry points. It may also demonstrate the connection of mobility with the distribution of the disease and enable us to follow the trends (Nkambule et al., 2023) (Jones et al., 2022). Monitoring of wastewater pathogens in animal farms and in wet markets is incredibly important as it allows detecting the early signs and tracing the origin of new and re-emerging diseases (Xiao & Zhang, 2023).The approach will be applicable to monitor the pathogens other than the SARS-CoV-2, which demonstrates that the monitoring of other infectious diseases could be done through it, as well as to identify the risks prior to the clinical cases recorded (Tay et al., 2024) (Toro et al., 2024). As a measure to minimize pandemic risks, spillover incidents and outbreaks should be identified early. Nevertheless, low- and middle-income countries have inadequate clinical surveillance that is difficult to carry out nationally (Grassly et al., 2024).

CONCLUSION

The present article demonstrates an entire immune informformatics-based approach of the development of multi-epitopic vaccinations across new challenges to the virus. It is more efficient, cheaper

and faster than conventional vaccinology. The paper illustrates that in silico technologies offer support in early-stage vaccine discovery and allows linking the computational epitope mapping technology with the population coverage, structural modelling, and immunological simulation approaches to potential vaccines with experimental validation. We found out that many of the epitopes that were predicted were highly antigenic, highly immunogenic, none of them allergens and importantly the HLA binding affinity was favourable. This implies that they would be ideal in covering many individuals. Free energy calculations and molecular docking revealed that when the epitope candidates are assembled in a multi-epitope construct along with the appropriate adjuvants; a good binding energetics, and a stable tertiary structure with good interactions with innate immune receptors (such as TLR2 and TLR4) is possible. The immune simulation results were resembling a perfect immunological profile, where there were a lot of IFN- gamma, rapid response of memory B and T cells, and that their immunoglobulins were highly responsive to an antigen they were presented with on repeated occasions. In addition, epitopes synthesized could trigger CD4+ and CD8+ T cells and cytokines responses in the test tube and that was in agreement with predictions made by computers. All the findings demonstrate that immunoinformatics can make the process of developing vaccines less challenging since it narrows down the list of most promising candidates prior to clinical trials. The paper also emphasizes the relevance of adding toxicity and allergenicity screens, structural modelling, and immunological simulation by stating it is essential to ensure that the thing is safe and that it works. The present research examined some novel viruses, yet the technology is capable of being implemented on other pathogens, including zoonotic viruses and rapidly evolving viruses. With health systems globally requiring vaccines that act more rapidly and are more cogent, this research presents an achievable roadmap of generating vaccines quickly that adopts a scientist precision-ideal paired to an appropriate extent. Making use of the method, the process of distributing effective vaccinations should be hastened, particularly in the times of pandemic, ensuring timely and correct immunological defence.

REFERENCES

Ahmed, W., Bivins, A., Simpson, S. L., Bertsch, P. M., Ehret, J., Hosegood, I., Metcalfe, S., Smith, W., Thomas, K. V., Tynan, J., & Mueller, J. F. (2021). Wastewater surveillance demonstrates high predictive value for COVID-19 infection on board repatriation flights to Australia. Environment International, 158, 106938.

Ali, S., Gudina, E. K., Gize, A., Aliy, A., Adankie, B. T., Tsegaye, W., Hundie, G. B., Muleta, M. B., Rufael, T., Belaineh, R., Negessu, D., Shegu, D., Froeschl, G., & Wieser, A. (2022). Community Wastewater-Based Surveillance Can Be a Cost-Effective Approach to Track COVID-19 Outbreak in Low-Resource Settings: Feasibility Assessment for Ethiopia Context. International Journal of Environmental Research and Public Health, 19(14), 8515.

Amman, F., Markt, R., Endler, L., Hupfauf, S., Agerer, B., Schedl, A., Richter, L., Zechmeister, M., Bicher, M., Heiler, G., Tříska, P., Thornton, M., Penz, T., Senekowitsch, M., Laine, J., Keszei, Z., Klimek, P., Nägele, F., Mayr, M., … Bergthaler, A. (2022). Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national scale. Nature Biotechnology, 40(12), 1814.

Aßmann, E., Agrawal, S., Orschler, L., Böttcher, S., Lackner, S., & Hölzer, M. (2024). Impact of

reference design on estimating SARS-CoV-2 lineage abundances from wastewater sequencing data. GigaScience, 13.

Baaijens, J. A., Zulli, A., Ott, I. M., Nika, I., Lugt, M. J. van der, Petrone, M. E., Alpert, T., Fauver, J. R., Kalinich, C. C., Vogels, C. B. F., Breban, M. I., Duvallet, C., McElroy, K. A., Ghaeli, N., Imakaev, M., Mckenzie-Bennett, M., Robison, K., Plocik, A., Schilling, R., … Baym, M. (2022). Lineage abundance estimation for SARS-CoV-2 in wastewater using transcriptome quantification techniques. Genome Biology, 23(1).

Barnes, K. G., Levy, J. I., Andersen, K. G., Gauld, J., Rigby, J., Kanjerwa, O., Uzzell, C. B., Chilupsya, C., Anscombe, C., Tomkins-Tinch, C. H., Mbeti, O., Cairns, E., Thole, H., McSweeney, S., Chibwana, M. G., Ashton, P., Jere, K. C., Meschke, J. S., Diggle, P. J., … Chilima, B. (2023). Utilizing river and wastewater as a SARS-CoV-2 surveillance tool to predict trends and identify variants of concern in settings with limited formal sewage systems. Research Square (Research Square).

Beguir, K., Skwark, M. J., Fu, Y., Pierrot, T., Carranza, N. L., Laterre, A., Kadri, I., Korched, A., Lowegard, A. U., Lui, B. G., Sänger, B., Liu, Y., Poran, A., Muik, A., & Şahin, U. (2023). Early computational detection of potential high-risk SARS-CoV-2 variants. Computers in Biology and Medicine, 155, 106618.

Berno, G., Fabeni, L., Matusali, G., Gruber, C. E. M., Rueca, M., Giombini, E., & Garbuglia, A. R. (2022). SARS-CoV-2 Variants Identification: Overview of Molecular Existing Methods [Review of SARS-CoV-2 Variants Identification: Overview of Molecular Existing Methods]. Pathogens, 11(9), 1058. Multidisciplinary Digital Publishing Institute.

Caduff, L., Dreifuss, D., Schindler, T., Devaux, A. J., Ganesanandamoorthy, P., Kull, A., Stachler, E., Fernández-Cassi, X., Beerenwinkel, N., Kohn, T., Ort, C., & Julian, T. R. (2022). Inferring transmission fitness advantage of SARS-CoV-2 variants of concern from wastewater samples using digital PCR, Switzerland, December 2020 through March 2021. Eurosurveillance, 27(10).

Chaudhuri, A., Pangaria, A., Sodhi, C., Kumar, N., Harshe, S., Parikh, N., & Shridhar, V. (2023). Building health system resilience and pandemic preparedness using wastewater-based epidemiology from SARS-CoV-2 monitoring in Bengaluru, India. Frontiers in Public Health, 11.

Cheohen, C., Gomes, V. M. S., & da Silva, M. L. (2025). CNN-LSTM Hybrid Model for AI-Driven Prediction of COVID-19 Severity from Spike Sequences and Clinical Data.

Drake, K. O., Boyd, O., Franceschi, V. B., Colquhoun, R., Ellaby, N., & Volz, E. (2024). Phylogenomic early warning signals for SARS-CoV-2 epidemic waves. EBioMedicine, 100, 104939.

Dunham, A. S., Jang, G. M., Muralidharan, M., Swaney, D. L., & Beltrão, P. (2021). A missense variant effect prediction and annotation resource for SARS-CoV-2. bioRxiv (Cold Spring Harbor Laboratory).

Dzinamarira, T., Murewanhema, G., Iradukunda, P. G., Madziva, R., Herrera, H., Cuadros, D. F., Tungwarara, N., Chitungo, I., & Musuka, G. (2022). Utilization of SARS-CoV-2 Wastewater Surveillance in Africa—A Rapid Review [Review of Utilization of SARS-CoV-2 Wastewater Surveillance in Africa—A Rapid Review]. International Journal of Environmental Research and Public Health, 19(2), 969. Multidisciplinary Digital Publishing Institute.

Fontenele, R. S., Yang, Y., Driver, E. M., Magge, A., Kraberger, S., Custer, J. M., Dufault-Thompson,

K., Cox, E., Newell, M. E., Varsani, A., Halden, R. U., Scotch, M., & Jiang, X. (2023). Wastewater surveillance uncovers regional diversity and dynamics of SARS-CoV-2 variants across nine states in the USA. The Science of The Total Environment, 877, 162862.

Grassly, N. C., Shaw, A., & Owusu, M. (2024). Global wastewater surveillance for pathogens with pandemic potential: opportunities and challenges [Review of Global wastewater surveillance for pathogens with pandemic potential: opportunities and challenges]. The Lancet Microbe, 100939. Elsevier BV. https://doi.org/10.1016/j.lanmic.2024.07.002

Han, W., Chen, N., Xu, X., Sahil, A., Zhou, J., Li, Z., Zhong, H., Gao, E., Zhang, R., Wang, Y., Sun, S., Cheung, P. P., & Gao, X. (2023). Predicting the antigenic evolution of SARS-COV-2 with deep learning. Nature Communications, 14(1).

Hayden, K., Jones, M. J., Elkin, K. R., Shreve, M. J., Clees, W. I., Clark, S. E., Mashtare, M. L., Veith, T. L., Elliott, H. A., Watson, J. E., Silverman, J. D., Richard, T. L., Read, A. F., & Preisendanz, H. E. (2022). Impacts of the COVID-19 pandemic on pharmaceuticals in wastewater treated for beneficial reuse: Two case studies in central Pennsylvania. Journal of Environmental Quality, 51(5), 1066.

Hie, B., Zhong, E. D., Berger, B., & Bryson, B. D. (2021). Learning the language of viral evolution and escape. Science, 371(6526), 284.

Holm, R. H., Pocock, G., Severson, M. A., Huber, V. C., Smith, T., & McFadden, L. M. (2023). Using wastewater to overcome health disparities among rural residents. Geoforum, 144, 103816.

Inzaule, S., Tessema, S. K., Kebede, Y., Ouma, A. E. O., & Nkengasong, J. N. (2021). Genomic-informed pathogen surveillance in Africa: opportunities and challenges [Review of Genomic-informed pathogen surveillance in Africa: opportunities and challenges]. The Lancet Infectious Diseases, 21(9). Elsevier BV.

Ito, J., Strange, A., Liu, W., Joas, G., Lytras, S., & Sato, K. (2024). A Protein Language Model for Exploring Viral Fitness Landscapes. bioRxiv (Cold Spring Harbor Laboratory).

Jahn, K., Dreifuss, D., Topolsky, I., Kull, A., Ganesanandamoorthy, P., Fernández-Cassi, X., Bänziger, C., Devaux, A. J., Stachler, E., Caduff, L., Cariti, F., Corzón, Á. T., Fuhrmann, L., Chen, C., Jablonski, K. P., Nadeau, S., Feldkamp, M., Beisel, C., Aquino, C., … Beerenwinkel, N. (2022). Early detection and surveillance of SARS-CoV-2 genomic variants in wastewater using COJAC. Nature Microbiology, 7(8), 1151.

Jayme, G., Liu, J., Gálvez, J. H., Reiling, S. J., Celikkol-Aydin, S., N'Guessan, A., Lee, S., Chen, S., Tsitouras, A., Sánchez-Quete, F., Maere, T., Goitom, E., Hachad, M., Mercier, É., Loeb, S. K., Vanrolleghem, P. A., Dorner, S., Delatolla, R., Shapiro, B. J., … Snutch, T. P. (2024). Combining Short- and Long-Read Sequencing Technologies to Identify SARS-CoV-2 Variants in Wastewater. Viruses, 16(9), 1495.

Jin, S., Dickens, B. S. L., Toh, K. Y., Lye, D. C., Lee, V. J., & Cook, A. R. (2024). Feasibility of wastewater-based detection of emergent pandemics through a global network of airports. PLOS Global Public Health, 4(3).

Johnson, R., Muller, C. J. F., Ghoor, S., Louw, J., Archer, E., Surujlal-Naicker, S., Berkowitz, N., Volschenk, M., Bröcker, L., Wolfaardt, G., Walt, M. van der, Mutshembele, A., Malema, S., Gelderblom, H. C., Muhdluli, M., Gray, G., Mathee, A., & Street, R. (2021). Qualitative and quantitative detection of SARS-CoV-2 RNA in untreated wastewater in

Western Cape Province, South Africa. South African Medical Journal, 111(3), 198.

Jones, D. L., Rhymes, J. M., Wade, M. J., Kevill, J. L., Malham, S. K., Grimsley, J. M. S., Rimmer, C., Weightman, A. J., & Farkas, K. (2022). Suitability of aircraft wastewater for pathogen detection and public health surveillance. The Science of The Total Environment, 856, 159162.

Kadadou, D., Tizani, L., Wadi, V. S., Banat, F., Alsafar, H., Yousef, A. F., Barceló, Damià, & Hasan, S. W. (2021). Recent advances in the biosensors application for the detection of bacteria and viruses in wastewater [Review of Recent advances in the biosensors application for the detection of bacteria and viruses in wastewater]. Journal of Environmental Chemical Engineering, 10(1), 107070. Elsevier BV.

Kadonsky, K. F., Naughton, C. C., Susa, M., Olson, R., Singh, G. L., Daza–Torres, M. L., Montesinos-López, J. C., García, Y. E., Gafurova, M., Gushgari, A. J., Cosgrove, J. R., White, B. J., Boehm, A. B., Wolfe, M. K., Nuño, M., & Bischel, H. N. (2023). Expansion of wastewater-based disease surveillance to improve health equity in California's Central Valley: sequential shifts in case-to-wastewater and hospitalization-to-wastewater ratios. Frontiers in Public Health, 11.

Karthikeyan, S., Levy, J. I., Hoff, P. D., Humphrey, G., Birmingham, A., Jepsen, K., Farmer, S., Tubb, H. M., Valles, T., Tribelhorn, C., Tsai, R., Aigner, S., Sathe, S., Moshiri, N., Henson, B., Mark, A. M., Hakim, A., Baer, N. A., Barber, T., … Martin, N. K. (2021). Wastewater sequencing uncovers early, cryptic SARS-CoV-2 variant transmission. medRxiv (Cold Spring Harbor Laboratory).

Karthikeyan, S., Levy, J. I., Hoff, P. D., Humphrey, G., Birmingham, A., Jepsen, K., Farmer, S., Tubb, H. M., Valles, T., Tribelhorn, C., Tsai, R., Aigner, S., Sathe, S., Moshiri, N., Henson, B., Mark, A. M., Hakim, A., Baer, N. A., Barber, T., … Martin, N. K. (2022). Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. Nature, 609(7925), 101.

Kaya, D., Falender, R., Radniecki, T. S., Geniza, M., Cieslak, P., Kelly, C., Lininger, N., & Sutton, M. (2022). Correlation between Clinical and Wastewater SARS-CoV-2 Genomic Surveillance, Oregon, USA. Emerging Infectious Diseases, 28(9), 1906.

Khan, M., Li, L., Haak, L., Payen, S. H., Carine, M., Adhikari, K., Uppal, T., Hartley, P. D., Vasquez-Gross, H., Petereit, J., Verma, S. C., & Pagilla, K. (2023). Significance of wastewater surveillance in detecting the prevalence of SARS-CoV-2 variants and other respiratory viruses in the community – A multi-site evaluation. One Health, 16, 100536.

Kistler, K. E., & Bedford, T. (2023). An Atlas of Adaptive Evolution in Endemic Human Viruses. bioRxiv (Cold Spring Harbor Laboratory).

Lamba, S., Sutharsan, G., Daroch, N., Paul, K., Joshi, S. G., Darshan, S. P., Madheswari, A. N., Vishwanath, S., Mishra, R. K., Ramakrishnan, U., & Ishtiaq, F. (2022). SARS-CoV-2 infection dynamics and genomic surveillance reveals early variant transmission in urban wastewater. medRxiv (Cold Spring Harbor Laboratory).

Larsen, D. A., Collins, M. B., Du, Q., Hill, D., Insaf, T. Z., Kilaru, P., Kmush, B., Middleton, F. A., Stamm, A., Wilder, M. L., Zeng, T., & Green, H. (2022). Coupling freedom from disease principles and early warning from wastewater surveillance to improve health security. PNAS Nexus, 1(1).

Layton, B. A., Kaya, D., Kelly, C., Williamson, K., Alegre, D., Bachhuber, S., Banwarth, P. G., Bethel, J. W., Carter, K., Dalziel, B. D., Dasenko, M., Geniza, M., George, A. D., Girard, A.-M., Haggerty,

R., Higley, K. A., Hynes, D. M., Lubchenco, J., McLaughlin, K., … Radniecki, T. S. (2022). Evaluation of a Wastewater-Based Epidemiological Approach to Estimate the Prevalence of SARS-CoV-2 Infections and the Detection of Viral Variants in Disparate Oregon Communities at City and Neighborhood Scales. Environmental Health Perspectives, 130(6).

Lee, B., Quadeer, A. A., Sohail, M. S., Finney, E., Ahmed, S. F., McKay, M. R., & Barton, J. P. (2022). Inferring effects of mutations on SARS-CoV-2 transmission from genomic surveillance data. medRxiv (Cold Spring Harbor Laboratory).

Levi, R., Zerhouni, E. G., & Altuvia, S. (2023). Predicting the spread of SARS-CoV-2 variants: An artificial intelligence enabled early detection. PNAS Nexus, 3(1).

Li, J., Hosegood, I., Powell, D., Tscharke, B. J., Lawler, J., Thomas, K. V., & Mueller, J. F. (2023). A global aircraft-based wastewater genomic surveillance network for early warning of future pandemics [Review of A global aircraft-based wastewater genomic surveillance network for early warning of future pandemics]. The Lancet Global Health, 11(5). Elsevier BV.

Lou, J., Liang, W., Cao, L., Hu, I., Zhao, S., Chen, Z., Chan, R. W. Y., Cheung, P. P., Zheng, H., Liu, C., Li, Q., Chong, K. C., Zhang, Y., Yeoh, E., Chan, P. K. S., Zee, B., Mok, C. K. P., & Wang, M. H. (2024). Predictive evolutionary modelling for influenza virus by site-based dynamics of mutations. Nature Communications, 15(1).

Luo, L., & Lv, J. (2025). Stochastic Mutation Theory of SARS-CoV-2 Variants. arXiv (Cornell University).

Maher, M. C., Bartha, I., Weaver, S., Iulio, J. di, Ferri, E., Soriaga, L., Lempp, F. A., Hie, B., Bryson, B. D., Berger, B., Robertson, D. L., Snell, G., Corti, D., Virgin, H. W., Pond, S. L. K., & Telenti, A. (2022). Predicting the mutational drivers of future SARS-CoV-2 variants of concern. Science Translational Medicine, 14(633).

Mahlangeni, N., Street, R., Horn, S., Mathee, A., Mangwana, N., Dias, S., Sharma, J., Ramharack, P., Louw, J., Reddy, T., Surujlal-Naicker, S., Nkambule, S., Webster, C., Mdhluli, M., Gray, G., Muller, C. J. F., & Johnson, R. (2023). Using Wastewater Surveillance to Compare COVID-19 Outbreaks during the Easter Holidays over a 2-Year Period in Cape Town, South Africa. Viruses, 15(1), 162. https://doi.org/10.3390/v15010162

Marathe, V., Bajracharya, D., & Yan, C. (2024). VirusT5: Harnessing Large Language Models to Predicting SARS-CoV-2 Evolution. arXiv (Cornell University).

Meadows, T., Coats, E. R., Narum, S., Top, E. M., Ridenhour, B. J., & Stalder, T. (2024). Epidemiological model can forecast COVID-19 outbreaks from wastewater-based surveillance in rural communities. medRxiv (Cold Spring Harbor Laboratory).

Merrett, J. E., Nolan, M., Hartman, L. M., John, N., Flynn, B., Baker, L., Schang, C., McCarthy, D., Lister, D., Cheng, N. N., Crosbie, N. D., Poon, R., & Jex, A. R. (2024). Highly sensitive wastewater surveillance of SARS-CoV-2 variants by targeted next-generation amplicon sequencing provides early warning of incursion in Victoria, Australia. Applied and Environmental Microbiology, 90(8).

Morawiec, E., Miklasińska-Majdanik, M., Bratosiewicz-Wąsik, J., Wojtyczka, R. D., Swolana, D., Stolarek, I., Czerwiński, M., Skubis-Sikora, A., Samul, M., Polak, A., Kruszniewska-Rajs, C., Pudełko, A., Figlerowicz, M., Bednarska-Czerwińska, A., & Wąsik, T. J. (2022). From Alpha to Delta—Genetic Epidemiology of SARS-CoV-2

(hCoV-19) in Southern Poland. Pathogens, 11(7), 780.

Munteanu, V., Gordeev, V., Su, J. M., Patel, K. K., Drabcinski, N., Zlenko, O., Kit, M., Xu, Y., Jaiswal, P., Zelikovsky, A., & Mangul, S. (2023). A rigorous benchmarking of methods for SARS-CoV-2 lineage abundance estimation in wastewater. arXiv (Cornell University).

Munteanu, V., Saldana, M. A., Sharma, N. K., Ouyang, W. O., Aßmann, E., Gordeev, V., Kasianchuk, N., Tierney, B., Lucaci, A. G., Knyazev, S., Ciorbă, D., Bostan, V., Mason, C. E., Skums, P., Wu, N. C., Rzymski, P., Hölzer, M., Zelikovsky, A., Smith, A. L., & Mangul, S. (2023). SARS-CoV-2 Wastewater Genomic Surveillance: Approaches, Challenges, and Opportunities. arXiv (Cornell University).

Najar, F. Z., Linde, E., Murphy, C. L., Borin, V. A., Wang, H., Haider, S., & Agarwal, P. K. (2023). Future COVID19 surges prediction based on SARS-CoV-2 mutations surveillance. eLife, 12.

Nkambule, S., Johnson, R., Mathee, A., Mahlangeni, N., Webster, C., Horn, S., Mangwana, N., Dias, S., Sharma, J., Ramharack, P., Louw, J., Reddy, T., Surujlal-Naicker, S., Mdhluli, M., Gray, G., Muller, C. J. F., & Street, R. (2023). Wastewater-based SARS-CoV-2 airport surveillance: key trends at the Cape Town International Airport. Journal of Water and Health, 21(3), 402.

Otero, M. C. B., Murao, L. A. E., Limen, M. A. G., Caalim, D. R. A., Gaite, P. L., Bacus, M., Acaso, J. T., Miguel, R. M., Corazo, K., Knot, I. E., Sajonia, H., Reyes, F. L. de los, Jaraula, C. M. B., Baja, E. S., & Mundo, D. M. N. D. (2022). Multifaceted Assessment of Wastewater-Based Epidemiology for SARS-CoV-2 in Selected Urban Communities in Davao City, Philippines: A Pilot Study. International Journal of Environmental Research and Public Health, 19(14), 8789.

Overton, A. K., Knapp, J. J., Lawal, O. U., Gibson, R. M., Fedynak, A. A., Adebiyi, A. I., Maxwell, B., Cheng, L., Bee, C., Qasim, A., Atanas, K., Payne, M., Stuart, R., Fleury, M., Knox, N., Nash, D., Hungwe, Y., Prasla, S., Ho, H., … Charles, T. C. (2024). Genomic surveillance of Canadian airport wastewater samples allows early detection of emerging SARS-CoV-2 lineages. Scientific Reports, 14(1).

Pérez-Zabaleta, M., Berg, C., Latorre-Margalef, N., Owusu-Agyeman, I., Kiyar, A., Botnen, H., Schönning, C., Hugerth, L. W., & Çetecioğlu, Z. (2025). Wastewater surveillance of SARS-CoV-2 from aircraft to citywide monitoring. Nature Communications, 16(1).

Singh, S., Ahmed, A. I., Almansoori, S., Alameri, S., Adlan, A., Odivilas, G., Chattaway, M. A., Salem, S. B., Brudecki, G., & Elamin, W. (2024). A narrative review of wastewater surveillance: pathogens of concern, applications, detection methods, and challenges [Review of A narrative review of wastewater surveillance: pathogens of concern, applications, detection methods, and challenges]. Frontiers in Public Health, 12. Frontiers Media.

Sokhansanj, B. A., Zhao, Z., & Rosen, G. (2022). Interpretable and Predictive Deep Neural Network Modeling of the SARS-CoV-2 Spike Protein Sequence to Predict COVID-19 Disease Severity. Biology, 11(12), 1786.

Street, R., Mathee, A., Mangwana, N., Dias, S., Sharma, J., Ramharack, P., Louw, J., Reddy, T., Bröcker, L., Surujlal-Naicker, S., Berkowitz, N., Malema, M. S., Nkambule, S., Webster, C., Mahlangeni, N., Gelderblom, H. C., Mdhluli, M., Gray, G., Muller, C. J. F., & Johnson, R. (2021).

Spatial and Temporal Trends of SARS-CoV-2 RNA from Wastewater Treatment Plants over 6 Weeks in Cape Town, South Africa. International Journal of Environmental Research and Public Health, 18(22), 12085.

Swain, M. J., Carter, B., Snowdon, K., & Faust, R. A. (2023). The Implementation and Utilization of Wastewater-Based Epidemiology: Experiences From a Local Health Department. Journal of Public Health Management and Practice, 29(3), 322.

Tan, C. C. S., Escalera-Zamudio, M., Yavlinsky, A., & Dorp, L. van. (2024). Intrahost dynamics, together with genetic and phenotypic effects predict the success of viral mutations. bioRxiv (Cold Spring Harbor Laboratory).

Tan, X. (2023). Co-evolution integrated deep learning framework for variants generation and fitness prediction. bioRxiv (Cold Spring Harbor Laboratory).

Tay, M., Lee, B., Ismail, M. H., Yam, J., Maliki, D., Gin, K. Y., Chae, S., Ho, Z. J. M., Teoh, Y. L., Ng, L. C., & Wong, J. C. C. (2024). Usefulness of aircraft and airport wastewater for monitoring multiple pathogens including SARS-CoV-2 variants. Journal of Travel Medicine, 31(5).

Thadani, N. N., Gurev, S. F., Notin, P., Youssef, N., Rollins, N. J., Ritter, D. P., Sander, C., Gal, Y., & Marks, D. S. (2023). Learning from prepandemic data to forecast viral escape. Nature, 622(7984), 818.

Thadani, N. N., Gurev, S. F., Notin, P., Youssef, N., Rollins, N., Sander, C., Gal, Y., & Marks, D. S. (2022). Learning from pre-pandemic data to forecast viral escape. bioRxiv (Cold Spring Harbor Laboratory).

Toro, L., Valk, H. de, Zanetti, L., Huot, C., Tarantola, A., Fournet, N., Moulin, L., Atoui, A., Gassilloud, B., Mouly, D., & Jourdain, F. (2024). Pathogen prioritisation for wastewater surveillance ahead of the Paris 2024 Olympic and Paralympic Games, France. Eurosurveillance, 29(28).

Tosta, S., Moreno, K. M. F., Schuab, G., Fonseca, V., Cardozo, F., Kashima, S., Elias, M. C., Sampaio, S. C., Ciccozzi, M., Alcântara, L. C. J., Slavov, S. N., Lourenço, J., Cella, E., & Giovanetti, M. (2023). Global SARS-CoV-2 genomic surveillance: What we have learned (so far) [Review of Global SARS-CoV-2 genomic surveillance: What we have learned (so far)]. Infection Genetics and Evolution, 108, 105405. Elsevier BV.

Wegrzyn, R. D., Appiah, G. D., Morfino, R. C., Milford, S. R., Walker, A. T., Ernst, E. T., Darrow, W. W., Li, S. L., Robison, K., MacCannell, D., Dai, D., Girinathan, B., Hicks, A. L., Cosca, B., Woronoff, G., Plocik, A., Simen, B. B., Moriarty, L. F., Guagliardo, S. A. J., … Friedman, C. R. (2022). Early Detection of Severe Acute Respiratory Syndrome Coronavirus 2 Variants Using Traveler-based Genomic Surveillance at 4 US Airports, September 2021–January 2022. Clinical Infectious Diseases, 76(3).

Xiao, K., & Zhang, L. (2023, February 24). Wastewater pathogen surveillance based on One Health approach. In The Lancet Microbe (Vol. 4, Issue 5). Elsevier BV.

Yousif, M., Saïd, R., Taukobong, S., Ndlovu, N., Iwu, C. J., Howard, W., Moonsamy, S., Mhlambi, N., Gwala, S., Levy, J. I., Andersen, K. G., Scheepers, C., Gottberg, A. von, Wolter, N., Bhiman, J. N., Amoako, D. G., Ismail, A., Suchard, M., & McCarthy, K. (2023). SARS-CoV-2 genomic surveillance in wastewater as a model for monitoring evolution of endemic viruses. Nature Communications, 14(1).