DOI: 10.1002/jmv.27017

COMMENTARY

JOURNAL OF MEDICAL VIROLOGY WILEY

Bioinformatics and immunoinformatics to support COVID-19 vaccine development

Stephanie Ishack[1] | Shari R. Lipner[2]

[1]New York University School of Medicine, New York, New York, USA

[2]Department of Dermatology, Weill Cornell Medicine, New York, New York, USA

Correspondence

Shari R. Lipner, 1305 York Avenue, New York, NY 10021, USA.
Email: shl9032@med.cornell.edu

Abstract

Severe acute respiratory syndrome coronavirus 2 has infected over 109 000 000 people with 2 423 443 deaths as of February 17, 2021. Currently, there are no approved or consistently effective treatments, and conventional vaccines may take several years for development and testing. In silico methods of bioinformatics, vaccinogenomics, immunoinformatics, structural biology, and molecular simulations can be used for more rapid and precise vaccine design. This paper highlights two major immunoinformatics strategies that are used in designing novel and effective vaccines and therapeutics: reverse vaccinology and structural vaccinology.

KEYWORDS

bioinformatics, computational biology, COVID-19, immunoinformatics, reverse vaccinology, SARS-CoV-2, structural vaccinology, vaccine, vaccinogenomics

1 | INTRODUCTION

The novel coronavirus 2019 (COVID-19) infection, also known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has infected over 109 000 000 people with 2 423 443 deaths as of February 17, 2021. The mainstay of COVID-19 treatment is supportive care, but there is a high mortality rate, especially in older individuals and those with co-morbidities. Therefore, there is an urgent need to develop safe and effective vaccines. Conventional vaccinology (CV) methods are inadequate for this pandemic because of: (1) time-consuming antigen identification; (2) lack of antigenic diversity; (3) extensive pathogen cultivation in wet labs; and (4) high costs.[1–2] In silico methods of bioinformatics, vaccinogenomics, immunoinformatics, structural biology, and molecular simulations can be effectively applied to advanced vaccine design, with faster processing time than CV.[1–3] In this paper, we highlight two major immunoinformatic strategies, reverse vaccinology (RV) and structural vaccinology (SV), and their application to potential COVID-19 vaccines.

2 | BIOINFORMATICS AND IMMUNOINFORMATICS

Bioinformatics is an interdisciplinary field using computational simulation methods to analyze biological data and make predictions on gene regulation networks.[1–2] It has been successfully utilized for vaccine research, including preclinical, clinical, and postvaccine

Abbreviations: CoV E protein, envelope protein; COVID-19, coronavirus 2019; CV, conventional vaccinology; 3D, three-dimensional; IEDB, Immune Epitope Database; MD, molecular dynamics; MHC, major histocompatibility complex; ORFs, open reading frames; RV, reverse vaccinology; SARS-CoV, SARS-associated coronavirus; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SV, structural vaccinology; Vaccine Design, Vaxign; VC, vaccine candidate; ViPR, virus pathogen resource.

This work has not been previously presented.

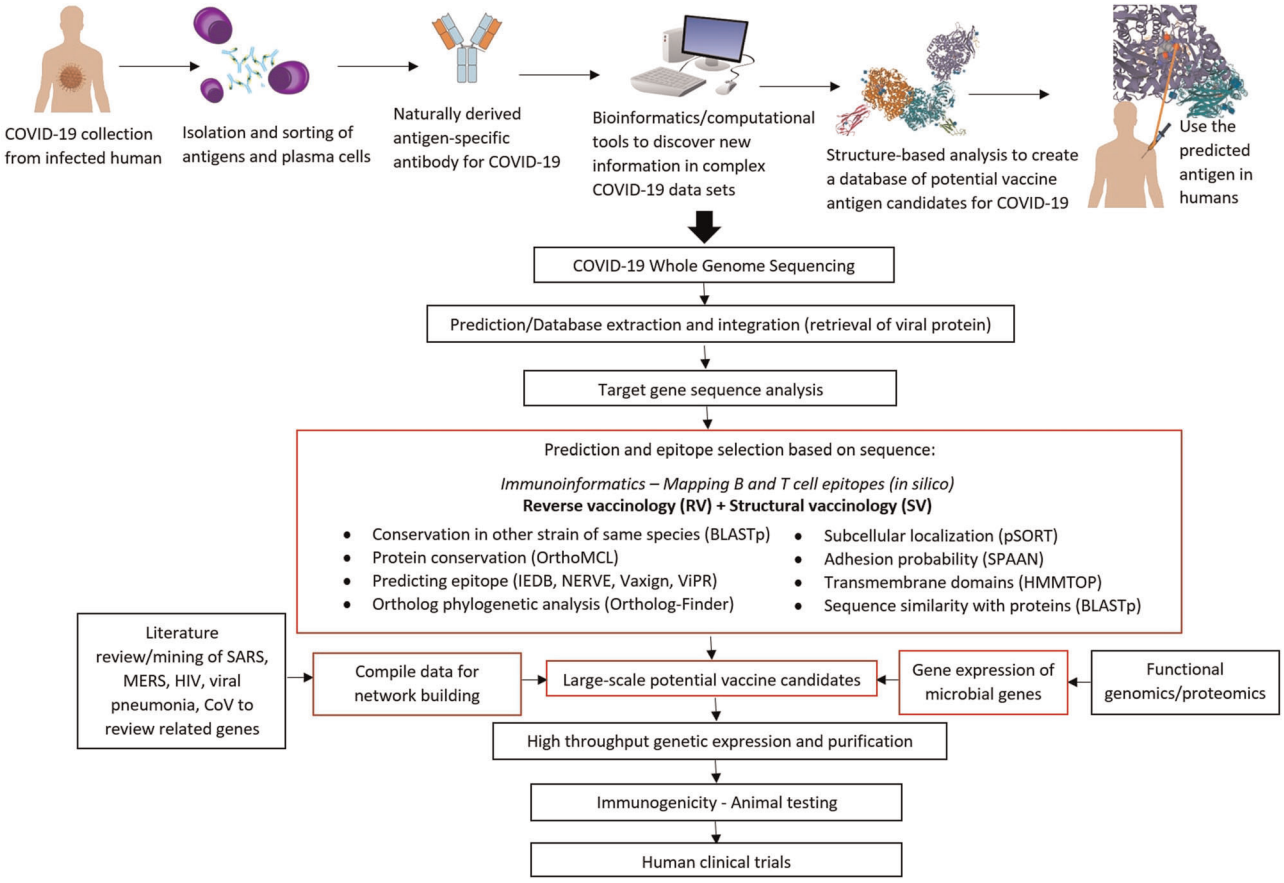

FIGURE 1 Displays the workflow of a vaccine design. The complete COVID-19 genomic sequence is the starting point in developing drugs, diagnostic tools, and a vaccine. Potential vaccine candidates (VCs) are identified from the COVID-19 gene sequence and can then be expressed as recombinants to assess immunogenicity. Reverse vaccinology (RV) and structural vaccinology (SV) can be employed in the COVID-19 vaccine development process. RV relies on genomic information to determine relevant antigens and to design B- and T-cell epitope mapping algorithms for diagnostic or vaccine purposes. SV involves analysis of the 3D structure and vaccine testing of individual domains. COVID-19, coronavirus 2019

phases.[1–3] Immunoinformatics is a branch of bioinformatics utilizing mathematical and computational approaches to process and develop immunological data and make predictions on immunity and disease pathogenesis.[2–3] Epitope and multiepitope vaccines are composed of amino acid peptides that are immunogenic targets. Using the already sequenced COVID-19 genome, computational tools, and searchable databases can aid in predicting potential B and T cell epitopes for vaccine design, immunity protein analysis, and immunization modeling.[4–7]

3 | IN SILICO—RV AND SV

RV identifies specific epitopes from viruses, bacteria, cancer cells, or allergens that activate specific immune responses.[1–3] COVID-19 is an excellent candidate for this strategy because the entire genomic sequence of the pathogen has been available since December 2019.[4] Computational programs (i.e., ORFfinder, GetOrf) are used to identify all open reading frames (ORFs) in the sequences. Then, the identified antigens can be mapped and screened for expression by the pathogen (i.e., COVID-19) and for immunogenicity during infection. For COVID-19, a comprehensive analysis of T-cell and B-cell epitope prediction is preferred for vaccine candidate (VC) selection. The immune system is classified as humoral or cellular. For a humoral response, bioinformatics databases (i.e., Immune Epitope Database [IEDB]) are used to identify B-cell epitopes. For cellular responses, databases are used to search for antigens that can be identified by major histocompatibility complex (MHC) molecules found in T cells.[1–2] Thereafter, in vitro and in vivo assays are performed to confirm VCs. Thus, immunoinformatics can be used to better understand infectious disease pathogenesis, diagnosis, immune system response, and vaccine development.

SV can be used to develop effective peptide-based vaccines by (1) assessing 3D conformational structure (i.e., X-ray crystallography, electron microscopy) of the epitope or antigen–antibody complex; (2) using molecular dynamics simulations to predict and model the epitope; (3) incorporating reengineered antigen into immunoinformatics platforms (i.e., Epitome); and (4) testing VCs for efficacy and safety in vivo.[2] Structural biology strategies provide insights into the structure of an entire virus, viral envelopes, and

antigen–antibody complexes that can be used to develop vaccines targeting novel viruses, such as COVID-19. Moreover, the three-dimensional (3D) modeling platforms provide critical information on the tertiary and quaternary protein architecture and position of the viral epitopes.

RV and SV technologies involve multiple steps, which are delineated in Figure 1. RV and SV serve different purposes but are equally important for vaccine development. CV platforms may take years to unravel information on antigens and disease pathogenesis, whereas, in silico vaccinology typically takes less than a year to discover the same or even more relevant information.[2–3]

4 | RECENT COVID-19 BIOINFORMATICS RESEARCH

There is some preliminary research using bioinformatics for COVID-19 vaccine development.[3–7] Grifoni et al.[4] used the IEDB and virus pathogen resource to show that SARS-associated coronavirus (SARS-CoV) and SARS-CoV-2 have high gene sequence similarity and comparable B- and T-cell epitopes. In another study, immunoinformatics and comparative genomic methods were used to assess a potential T-cell epitope peptide-vaccine by targeting the COVID-19 envelope protein (CoV E). Using comparative sequencing, 10 MHC Class I and MHC Class II peptides were found that are promising VCs for COVID-19.[5] Furthermore, Enayatkhani et al.[6] used RV to analyze three COVID-19 antigenic proteins (nucleocapsid, ORF3a, and membrane protein, [NOM]) and developed a potential multiepitope COVID-19 vaccine that can stimulate both CD4+ and CD8+ T-cell immune responses. Ong et al.[7] utilized the Vaxign platform and Vaxign-ML machine-learning tool to successfully predict a COVID-19 VC called, "Sp/Nsp cocktail." There was sequence conservation of protein nsp3 among SARS-CoV-2, SARS-CoV, and MERS-CoV and that nsp3-domain contained MHC-I T-cell, MHC-II T-cell, and B-cell epitopes.[7] The next steps include clinically testing and validating these proposed COVID-19 VCs to ensure efficacy and safety.

5 | LIMITATIONS

RV and SV are subject to several limitations. There is evidence that antigen residues may become epitopes under specific conditions thereby complicating prediction methods.[3] Thus, RV requires advanced prediction algorithms to filter and analyze nonepitopic antigen surface residues. On rare occasions, sequence data from the high-throughput analysis may contain computing errors thereby affecting the quality of immunoinformatics predictions. The main disadvantage of SV is that newly predicted epitopes may be buried deep within the protein, making it difficult to detect by potential COVID-19 antibodies, and therefore, making the vaccine less effective.[2]

6 | CONCLUSION

The rapid development of an effective and safe vaccine is necessary to reduce global mortality from COVID-19. RV and SV are promising techniques that can aid in the development of multiepitope vaccines against COVID-19 by inducing comprehensive B- and T-cell immunity. Since these immunoinformatics tools are readily available, continued leverage of these technologies will result in a shortened time to vaccine development.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

ORCID

Stephanie Ishack https://orcid.org/0000-0002-2627-7791
Shari R. Lipner https://orcid.org/0000-0001-5913-9304

REFERENCES

1. Soria-Guerra RE, Nieto-Gomez R, Govea-Alonso DO, Rosales-Mendoza S. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. *J Biomed Inform*. 2015;53:405-414. https://doi.org/10.1016/j.jbi.2014.11.003
2. Ribas-Aparicio RM, Castelán-Vega JA, Jiménez- Alberto A. The impact of bioinformatics on vaccine design and development. *Vaccines*. 2017;7:123-147. https://doi.org/10.5772/intechopen.69273
3. Dhama K, Sharun K, Tiwari R. COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics. *Hum Vaccin Immunother*. 2020;0(0):1-7. https://doi.org/10.1080/21645515.2020.1735227
4. Grifoni A, Sidney J, Zhang Y. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe*. 2020;24(4):671-680. https://doi.org/10.1016/j.chom.2020.03.002
5. Abdelmageed MI, Abdelmoneim AH, Mustafa MI, et al. Design of multi epitope-based peptide vaccine against E protein of human COVID-19: an immunoinformatics approach. *Biomed Res Int*. 2020;2020:2683286. https://doi.org/10.13140/RG.2.2.17111.42405
6. Enayatkhani M, Hasaniazad M, Faezi S, et al. Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. *Biomol Struct Dyn*. 2020:1-19. https://doi.org/10.1080/07391102.2020.1756411
7. Ong E, Wong MU, Huffman A, et al. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *bioRxiv*. 2020. https://doi.org/10.1101/2020.03.20.000141

How to cite this article: Ishack S, Lipner SR. Bioinformatics and immunoinformatics to support COVID-19 vaccine development. *J Med Virol*. 2021;93:5209-5211. https://doi.org/10.1002/jmv.27017