


The Relevance of Bioinformatics Applications in the Discovery of Vaccine Candidates and Potential Drugs for COVID-19 Treatment

Bioinformatics and Biology Insights
Volume 15: 1–8
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11779322211002168



Onyeka S Chukwudozie¹ , Vincent C Duru²,
Charlotte C Ndiribe¹, Abdullahi T Aborode³, Victor O Oyeibanji⁴
and Benjamin O Emikpe^{4,5}

¹Department of Cell Biology and Genetics, University of Lagos, Lagos, Nigeria. ²Molecular Genetics Unit, Institute of Child Health, College of Medicine, University of Ibadan, Ibadan, Nigeria. ³Department of Chemistry, University of Ilorin, Ilorin, Nigeria. ⁴Department of Veterinary Pathology, University of Ibadan, Ibadan, Nigeria. ⁵School of Veterinary Medicine, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.

ABSTRACT: The application of bioinformatics to vaccine research and drug discovery has never been so essential in the fight against infectious diseases. The greatest combat of the 21st century against a debilitating disease agent SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) virus discovered in Wuhan, China, December 2019, has piqued an unprecedented usage of bioinformatics tools in deciphering the molecular characterizations of infectious pathogens. With the viral genome data of SARS-COV-2 been made available barely weeks after the reported outbreak, bioinformatics platforms have become an all-time critical tool to gain time in the fight against the disease pandemic. Before the outbreak, different platforms have been developed to explore antigenic epitopes, predict peptide-protein docking and antibody structures, and simulate antigen-antibody reactions and lots more. However, the advent of the pandemic witnessed an upsurge in the application of these pipelines with the development of newer ones such as the Coronavirus Explorer in the development of efficacious vaccines, drug repurposing, and/or discovery. In this review, we have explored the various pipelines available for use, their relevance, and limitations in the timely development of useful therapeutic candidates from genomic data knowledge to clinical therapy.

KEYWORDS: COVID-19, SARS-CoV-2, vaccine, drug discovery, bioinformatics

RECEIVED: December 5, 2020. **ACCEPTED:** February 14, 2021.

TYPE: Review

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Onyeka S Chukwudozie, Department of Cell Biology and Genetics, University of Lagos, Lagos, Nigeria. Email: solomononyeka84@gmail.com

Introduction

Applications of bioinformatics pipelines are indispensable in enabling predictions of potential biomolecules that aid the mechanism of action, treatment, and prevention of infectious diseases. The use of bioinformatics tools and techniques to analyze biological data generated from genomics, transcriptomics, proteomics, and structural omics is gaining tremendous momentum and providing solutions to urgent biological problem.¹

With the increasing number of databases and various curated online repositories, bioinformatics has become a veritable platform for data obtained from epidemiological studies to be analyzed. Through this means, pathogen identification, molecular pathogenesis, and emerging diagnostic methods are highly feasible. In recent times, bioinformatics is used in concert with next generation sequencing techniques in the diagnosis of infectious bacterial and viral diseases.^{2–4}

These omics tools have revolutionized the methods in vaccinology and drug repurposing. Proteomics and transcriptomics can be used as complementary approaches to genomics, for instance, to identify surface proteins during host-pathogen interactions.⁵

In the past, traditional screening for lead compounds in vaccine and drug production took years to attain. Several computational pipelines were used to predict vaccine candidates

before further validation for clinical trials. To improve on the process, new approaches were developed to reduce the duration and cost involved in drug production, particularly with the aid of sophisticated computational pipelines. For vaccines, predicting antigenic peptide components, selecting immunogenic carriers amid arrays of immuno-adjuvants that speed up immunogenic response could be achieved using these computational pipelines.⁶

In December 2019, an unprecedented pneumonia disease outbreak was reported in Wuhan, China.⁷ There was a global rapid spread of the disease from the epicenter to the rest of the world, with thousands of infected persons and several reported death cases, elevating the disease to a pandemic.⁷ By December 31, 2019, the disease outbreak was traced to a novel strain of coronavirus,⁷ which was later termed SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) and the disease COVID-19 by the World Health Organization (WHO).⁷

SARS-CoV-2 has since been identified as a new strain from group 2B Coronaviruses, with approximately 70% genetic homology to SARS-CoV from the 2003 outbreak.⁸ Because the virus has a 96% similarity to a bat coronavirus, it is widely suspected to originate from bats.^{9,10} The pandemic resulted in unprecedented nationwide and international travel restrictions for several countries, alongside grievous economic upheavals.⁷



In the search for solutions, developing a universally available therapeutic vaccine and drug is critical.

Coronaviruses are responsible for respiratory tract infections that range from mild cases of the common cold and low fever, to more severe cases of SARS (severe acute respiratory syndrome), MERS (Middle East respiratory syndrome), and COVID-19.⁸ In the face of the current COVID-19 pandemic, the comprehensive application of bioinformatics techniques on relevant genome sequences deposited in online repositories could tremendously aid the search for a potent vaccine and drug that will combat the spread of the virus. The objective of this study therefore is to assess the different computational tools employed, alongside their suitability in the development of vaccine candidates and therapeutic targets required to effectively combat the menace of the current pandemic.

In this review, we will categorically discuss the applications of computational pipelines and algorithmic intricacies that have tremendously facilitated in the quest for the development of vaccines and drugs for COVID-19 infection. We will also summarize and make reference to few works that contributed in providing much insights of the applications of bioinformatics in providing solutions to COVID-19 biological questions.

Applications of Bioinformatics in Vaccine Design for COVID-19 Treatment

Since the advent of the COVID-19 pandemic, research has been accelerated to develop a potent vaccine candidate against the viral disease. This has produced unprecedented levels of synergy between government, academic, and private organizations that have come to work together to rapidly develop vaccines and antibody countermeasures targeted against the structural and non-structural proteins of the viral genome.¹¹ Such partnerships leverage on available technologies, including bioinformatics to accelerate response to the rampaging disease. This has resulted in the fastest vaccine development initiatives known in history, with some groups developing vaccine candidates and attaining clinical trials within few months. Presently, there are several bioinformatics approaches to design and develop safe, stable, and effective vaccines through reverse vaccinology, immune-informatics, and structural vaccinology.

Reverse vaccinology (RV) is a process of vaccine development that involves the identification of novel antigens through analysis of the genomic information of an organism.¹² A methodology relies on bioinformatics tools to identify target allergens using information on the genetic makeup of the pathogen. RV is able to reveal the genes that encode proteins that could lead to good epitopes.¹²⁻¹⁵ It employs relevant software systems and programs to identify the open reading frames (ORFs) of the organism's genome and determine various antigenic and physicochemical properties associated with the antigenic epitopes using tools such as the VaxiJen server.¹⁶

RV is cost-effective and reduces the time spent in the traditional drug design approach. It refines the number of proteins

to be studied and enables the identification of antigens present in small amounts or that are expressed only at certain stages in the organism's life cycle. It speeds up allergen selection process and allows for the study of pathogens that cannot be cultivated *in vitro*.¹⁴⁻¹⁷

The feasibility of RV is dependent on the availability of genomic information for the pathogen. Where available, it is possible in theory to identify all the antigens seen by conventional means and novel antigens that work on a very different concept. RV has been used in COVID-19 vaccine design by few researchers,^{18,19} leveraging on the available SARS-COV-2 genome sequence information. In addition, it has been employed in the development of multi-epitope chimeric vaccines against the SARS-COV-2.²⁰⁻²⁴

Immuno-informatics, ie, bioinformatics approach to immunology, involves the analysis of the complete information on an organism's immunomics (ie, all genes and proteins of cells that take part in mounting immune responses), and using the generated data to make predictions of immune responses against specific molecules.¹⁷ Immuno-informatics methods have been employed in the determination of the humoral and cellular immune cells attachment sites (ie, epitopes) on COVID-19 virus. Various immuno-informatics tools have also been used to predict whether a region of the SARS-COV-2 genome, usually a protein, can generate an immune response by itself. This is referred to as antigenicity test and includes TEpredict, CTLPred, NetMHC, and Epitopemap. Some research has employed deep learning and machine learning algorithms to predict potential immunogenic subunits from the viral genome sequences.

These tools aid the understanding of the genetic polymorphism of major histocompatibility complex (MHC) classes I and II in target human populations,²⁵ and in predicting the epitopes for cytotoxic and helper T lymphocytes. In addition, these tools reduce the time required to identify immunogenic targets and enhance the development of potentially safe vaccine candidates, as pertains to the current COVID-19 vaccine.²⁶

Several studies²⁷⁻³² have gone further to join epitopes and immunogenic domains of the SARS-COV-2 proteins to develop novel vaccine structures, multi-epitope vaccines or chimeric vaccine structures, either with or without predicted adjuvants to improve the immunogenicity of their candidate products. Similarly, the structural vaccinology approach focuses on the conformational features of the viral epitope that makes them good candidate antigens.¹⁷ This involves the analysis of the vaccine candidate to ascertain the conformational structures and a property that may elicit the best immunological response from monoclonal antibodies since reports of van Regenmortel³³ suggest monoclonal antibodies recognize conformational rather than linear epitopes.

Here, structural attributes such as the structural stability of the peptides, the solvent exposure, the hydrophobicity, and the

codon optimization are used to map antigenic epitopes to detect conformational features that could affect immunogenicity. This is achieved via molecular docking, dynamics simulations, and homology modeling to deduce antigen and antibody structure. Structural vaccinology techniques have been applied in predicting vaccine candidates for the SARS-COV-2^{34,35} and these have aided the identification of structurally stable, safe, and effective peptides as vaccine candidates.

Steps Involved in Immuno-Informatics

Chukwudozie et al^{6,34} have extensively provided a comprehensive and simplified step in designing peptide vaccines and also provided various approaches in validating them using *in silico* checkpoints. The steps involved in designing an epitopic peptide vaccine adopting immuno-informatics method include proteomic data retrieval of the protein of interest and adoption of computational immunology analysis of the data, which includes test for antigenicity, allergenicity, peptide toxicity, and epitopic conservancy.

Process

Curated nucleotide or proteomic data set is retrieved from biorepository and properly edited using bio-editing tools for better precisions. In most cases, deletion sites when discovered are manually expunge from the data set. These manually or automated curated sequences are then tested for their physico-chemical parameters, eg, amino acid composition, half-life index, solubility, carbon, hydrogen, and oxygen contents. The B- and T-cell (CD4 and CD8) epitopes are predicted using several machine learning tools that are readily available online, or standalone resources. These predicted peptide promiscuities are also predicted based on their putative restrictions to MHC I and II alleles. Other quality checks such as their respective antigenicity, allergenicity, peptide toxicity, and peptide conservancy, which characterize their immunogenicity profile, are considered. A well-summarized and applied process of these steps can be seen in a study conducted by Chukwudozie et al.^{6,34}

The New SARS-CoV-2 Viral Strain and Impact on Vaccine Development

In December 2020, a new variant of COVID-19 was discovered in the United Kingdom and was made public in the UK parliament by the country's health secretary.³⁶ This new variant was presumed to be more deadly than the wide type. The variant was named VUI-202012/01 and characterized with a total of 17 mutations.³⁶ The most important of them was the N501Y residual mutation embedded in the spike protein, which is pivotal for the binding fusion with the targeted human ACE2 receptor.³⁷ This intimates that the mutation could make the virus more infectious and spread rapidly. In this scenario, bioinformatics pipelines or next generation DNA sequencing can be adopted to structurally examine and detect possible mutations. Mutation calling algorithms can be fully developed and

optimized for the simultaneous analysis of multiple samples of the virus.

Limitations of Immuno-Informatics for Predicting Vaccine Candidates

Although numerous bioinformatics tools and algorithms have been employed in a bid to develop a suitable vaccine candidate against the COVID-19, there are several limitations that pose drawbacks to the full integration of this predictive approach in conventional vaccine design and development. A major shortfall of bioinformatics in the quest for an effective vaccine against the COVID-19 is its inability to address the persistent issue of selecting a suitable animal model for testing vaccine candidates. In the case of SARS-COV, some of the vaccine candidates developed was unable to provide complete protective immunity in ferrets and monkeys used for the trials. Some stimulated antibodies against the SARS-COV spike protein, but were unable to provide complete protection, whereas some vaccine candidates were associated with lung inflammation when immunized mice were later infected with the virus.^{38,39}

Despite the prospects offered by reverse vaccinology (RV) as an effective improvement over the traditional approach of vaccine development, its major drawback, however, is that it can only target proteins. This is unlike the traditional vaccine design approach, which can find other biomolecular targets, such as polysaccharides which may improve immune response.¹⁴ Also, the bioinformatics-based approach of vaccine design may identify linear and discontinuous epitopes, yet several predicted epitopes may be available buried within the viral protein and may not be easily detected by antibodies *in vivo* as against *in silico* predictions and analysis. This may lead to the failure in the candidate vaccine.⁴⁰

In addition, a critical limitation of immuno-informatics prediction of epitopes for antibodies binding is the pleomorphism of MHC class I molecules. There are more than 20 000 MHC class I alleles that are known to translate into 11 000 different MHC class A, B, and C molecules.^{41,42} With the current urgency for the development of a COVID-19 vaccine, this would prevent the generation of large data sets for training the prediction tools to improve their precision, and thus may mislead the vaccine design.⁴⁰ In addition, there will be incomplete representation of genetic diversities prevalent in low- and middle-income countries of the world, especially in Africa and Asia,^{43,44} which may result in vaccine failure among such populations.

Finally, with the urgency to develop a safe and effective measure to eradicate the COVID-19 pandemic with vaccine design strategies useful to confront possible future outbreaks of SARS-CoV-2 and related coronaviruses, it has become expedient to accelerate the design process with bioinformatics tools available. They remain a potent, cost-effective, and time-saving solution for the rapid development of vaccines. However, it is necessary to address these limitations to enhance the prospects

in the design of safe and effective vaccine candidates against infectious microbial diseases, such as SARS-COV-2 and other related viruses.

The Role of Bioinformatics in Drug Design for COVID-19 Treatment

Before the identification of effective drug candidates, intensive laboratory research efforts and various clinical trials are conducted. Several predictive measures with good precision are used in drug design, leveraging on the vast information on genome biology and disciplines such as bioinformatics. As regard the current COVID-19 pandemic, following the identification of the crystal structure of SARS-CoV-2 main protease (M^{Pro}) alongside its associated structures, there are several reported studies on coronaviruses in attempts to repurpose some existing drugs with varying degrees of success using computational methods.⁴⁵⁻⁴⁸ Repurposing drugs against the COVID-19 disease with known preclinical, pharmacokinetic, pharmacodynamic, and toxicity profiles can hasten treatment regimens for use, directly in clinical settings. For instance, a novel network-based drug repurposing platform to identify potential drugs for the treatment of COVID-19 has been reported by Lin et al.⁴⁹ They analyzed genome sequences of SARS-CoV-2 and identified SARS as the closest disease, based on the genome similarity between both causal viruses, MERS and other human coronavirus diseases.

Given the incomplete trials of vaccine and the urgent need, coupled with the novel drug discovery that is known to take several years, drug repurposing is apparently the best strategy to rapidly yield effective therapies against COVID-19. Drug repurposing can produce new therapies at a faster rate than novel drug discovery if the safety profiles of the repurposed drugs are evaluated in the context of drug development for another disease. This can occur at an even faster rate when the drugs approved for the treatment of other diseases and post-marketing safety surveillance data are available.^{50,51} By relying on existing preclinical, pharmacokinetic, pharmacodynamic, and toxicity profiles of the drugs being repurposed, it is possible to effectively increase the rapidity of the response against a disease condition with unmet clinical needs, especially in a pandemic, where drug proven safe can be readily tested in trials or administered to patients as compassionate treatment.

In silico methods offer a way to methodically and rapidly yield additional repurposing candidates.⁵² For instance, when drug targets associated with a disease of interest are known, and when the protein structures or that of close homologs are available, it is possible to use structural bioinformatics to virtually screen (eg, using molecular docking) a library of existing drugs against these known targets.⁵³

Similarly, a study published on February 27, 2020, relied on this approach, using the predicted structure of all SARS-CoV-2 proteins based on their homology with other known

coronavirus protein structures to identify several compounds with potential antiviral activity.⁵⁴ Although the standard reverse transcription polymerase chain reaction (RT-PCR) and DNA sequencing protocols have been developed for diagnosis, no drugs are fully established to cure the COVID-19 disease as at the time of writing this review.

Even with limited successes reported with repurposed anti-malaria drugs such as chloroquine phosphate and other antiviral drugs, structural bioinformatics pipelines were still applied to validate these claims.⁵⁵ In the space of online molecular docking methods, it was found that all tested repurposed drugs were associated accordingly with the SARS-CoV-2 protease enzyme that plays a role in viral replication.⁵⁶ This implies that certain repurposed drugs could be proposed as drug candidates for the treatment of COVID-19, after clinical trials or rigorous laboratory testing.

It is most encouraging to realize that with the bioinformatics approach, a consortium of Chinese university researchers was able to specifically use structural bioinformatics to elucidate the 3-dimensional (3D) structure of the SARS-CoV-2 protease. Knowledge of the protein structure of SARS-CoV-2 provided a template for further applications of bioinformatics tools in the COVID-19 drug development. The 3D structural conformation of the viral protease could only then be applied to drug repurposing and tested with bioinformatics tools, such as molecular docking and dynamics simulations in the discovery of a drug candidate.⁵⁴

Bioinformatics applications in drug design uses the computer-aided drug design (CADD) method that integrates the methods of lead compound Quantitative Structure-Activity Relationship (QSAR) optimization, sequence, structural homology, stereo-chemical validation, molecular docking, and 2-dimensional (2D) molecular interaction examination.⁵⁷⁻⁵⁹

The sequence and structural alignment of the protease enzyme of interest are used to establish the basis of the drug design.⁶⁰ The CADD method was developed by leveraging on the protein sequence analysis pipeline.⁶¹ The first step to predicting the feasibility of lead compounds of a drug is to determine the QSAR annotation for the presence or absence of protease and/or peptidase inhibition activity. All known compounds have protease or peptidase inhibition activity prediction. Importantly, this could shed novel light on the inhibition activities of the SARS-CoV-2 protease enzyme⁶¹ and is used as an early indicator of efficacy in drug design.

As repurposed drugs have already undergone clinical trials for the treatment of other diseases, with the toxicological and pharmacological indicators already elucidated, the ADME-TOX computational prediction may not be necessary since ADME-TOX prediction can be useful where the drugs' functional groups are been modified. Moreover, it is important to note that based on actual clinical experience, some drugs have

already been listed in the COVID-19 prevention and treatment handbook published by Zhejiang University Medical School.⁶²

Based on structural bioinformatics research, chloroquine phosphate, lopinavir, remdesivir, and other related antiviral drugs have the potential to be leveraged upon as lead compounds for a COVID-19 drug.⁶³

Using the docking method, the top candidates that emerged as a potential solution are lopinavir and remdesivir. With the sophisticated repurposing of these drugs, they could be revalidated by clinical trials for the production of more potent variants and a COVID-19 vaccine.

In bioinformatics research, offline programs, such as Autodock, could be used to provide richer customization options.⁶³ Another approach to drug repurposing is the construction of so-called “disease-related molecular networks,” ie, interactions between gene products (this is sometimes with cellular metabolites) involved in the etiology and symptoms of that disease.⁶⁴

There are several ways to identify such disease-related genes, which include using genomic data (eg, Genome-Wide Association Studies), gene expression data (eg, RNAseq differential expression analysis), or data collected directly from the scientific literature (eg, text mining or expert curation, either analyzed in-house or via recognized structured databases). Compared with virtual screening, where the candidate targets are known from the start, network biology methods can identify additional, unanticipated targets, which are part of the same molecular pathway than previously known targets for the disease of interest.⁶⁵

Network Bioinformatics Approach in Drug Repurposing Predictions

To contextualize and better understand, at the systems level, the molecular and physiological role of the discovered COVID-19-related genes, newly developed algorithm used in building a molecular protein network was adopted. A study conducted by Lin et al⁴⁹ showed how novel network-based drug repurposing was used to identify potential drugs for the treatment of COVID-19. The work explored the dynamics of phylogenomics in the analyses of genomic sequences of SARS-CoV-2 among other closely related viral family with similar proteomic structure and functions. Following the molecular network and drug prioritization algorithms, they were able to identify 30 potential repurposable drugs with activity against COVID-19, some of which includes pseudoephedrine, andrographolide, chloroquine, abacavir, and thalidomide.

One of the vital tools that has been adopted and massively applied in the SARS-CoV-2 drug prediction is the CoVex: Coronavirus Explorer.⁶⁶ The CoVex is a network and systems-medicine web platform that integrates experimental human-virus protein interactions for SARS-CoV,⁶⁷ SARS-CoV-2,⁶⁸ human protein-protein interactions,⁶⁹ and drug-protein

interactions⁷⁰⁻⁷² into a large-scale interactome. It is used by biomedical and clinical researchers to predict novel drug targets and drug repurposing candidates using various state-of-the-art graph analytical methods.⁷³ With CoVex, expert knowledge on virus replication, immune-related biological processes, or drug mechanisms can be applied in the compilation of a host or set of viral proteins (referred to as seeds).⁷⁴

Alternatively, users can upload a proteins list (eg, differentially expressed genes), a list related to a molecular mechanism of interest, or proteins targeted by drugs of interest,⁷⁵ which could be a set of known effective drugs⁷⁶ that can be used as seeds to guide the analysis. Based on the selected seeds, CoVex offers 3 main actions:

1. Searches the human interactome for viable drug targets,
2. Identifies repurposable drug candidates, and
3. A combination of actions that starts from a selection of virus or virus-interacting proteins. Users can explore the interactome for suitable drug targets for which, subsequently, suitable drugs are identified. In summary, CoVex allows researchers to systematically identify already approved drugs that could be repurposed to treat SARS-CoV-2, which is faster than developing new drugs from raw processes.

Drug Target for the Mpro of SARS-CoV-2

Currently, there are still no validated or approved lead drug targeting the Mpro of the SARS-CoV-2. The Mpro of SARS-CoV-2 as the key enzyme plays a crucial role in aiding viral replication and transcription.⁷⁷ This annotated function makes the main protease a likable drug target. Many potential inhibitors have been identified through several methods, including the in silico-based approach. A study by Yang and colleagues⁷⁷ identified a mechanism-based inhibitor adopting CADD and also determined the crystal structure of Mpro of the virus in a combined complex with the identified inhibitor. Yang et al were able to screen about 10 000 arrays of compounds, including approved drugs, drugs undergoing clinical trials, and pharmacological active compounds. This study was able to identify 6 potential compounds that inhibited the enzymatic activities of Mpro. One of the identified compounds was ebsele, which, according to them, exhibited promising antiviral activity in cell-based assays.

Machine Learning and Its Applications for Vaccine and Drug Design

Machine learning incorporates the application of artificial intelligence (AI) that provides the computer or systems technology with the automated ability to learn and improve experiences without been explicitly programmed. Tools whose mode of operation are centered on AI and deep learning were fully explored during the surge of the pandemic in predicting and designing both vaccine and drug candidates. For vaccine

Table 1. Computational tools used in the prediction of antigenic peptides for vaccine design.

TOOLS	DESCRIPTION	LINKS
Sprint	Sequence-based prediction of protein-peptide binding sites using support vector machine	https://sparks-lab.org/
modIAMP	Python 3 package designed for working with peptides, proteins, or any sequence of natural amino acids	https://modlamp.org/index.html
pepATTRACT	Peptide-protein docking	https://bioserv.rpbs.univ-paris-diderot.fr/services/pepATTRACT/
PIGSPRO	Prediction of Ab structures	https://openebench.bsc.es/tool/pigspro
ACCLUSTER	Predicts peptide-binding site	http://zougrouptoolkit.missouri.edu/accluster/
PEPstrMOD	Predicts the tertiary structure of small peptides varying between 7 and 25 residues	http://osddlinux.osdd.net/raghava/pepstrmod/
TepiTool	Pipeline for computational prediction of T-cell epitope candidates	http://tools.iedb.org/tepitool/
PEP-FOLD3	Faster de novo structure prediction for linear peptides in solution	https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/
IEDB	Epitope prediction and analysis tools that make predictions based on Parker hydrophilicity, beta-turn prediction, and surface accessibility	http://tools.iedb.org/main/
ABCpred	Predicts linear B-cell epitopes using amino acid anchoring pair composition	http://crdd.osdd.net/raghava/abcpred/
BEpro	Discontinuous B-cell epitope prediction	http://pepito.proteomics.ics.uci.edu/
DiscoTope	Predicts discontinuous B-cell epitopes from protein 3-dimensional structures	http://www.cbs.dtu.dk/services/DiscoTope-2.0/
PEASE	Predicts epitopes using antibody sequence	http://www.ofranlab.org/PEASE
Expitope	Server for epitope expression. The tool permits users to find all known proteins containing their peptide of interest. It can exclude any cross-reactivity in early stages of T-cell receptor selection for use in design of adoptive T-cell immunotherapy	http://webclu.bio.wzw.tum.de/expitope/
AllerTop	Server for in silico prediction of allergens	http://www.pharmfac.net/allertop/
ToxinPred	Tool for predicting: (1) toxicity or non-toxicity of peptides, (2) minimum mutations in peptides for increase or decrease of toxicity, and (3) toxic regions in proteins	http://crdd.osdd.net/raghava/toxinpred/
IgBLAST	Immunoglobulin variable domain sequence analysis tool	https://www.ncbi.nlm.nih.gov/igblast/
ProtParam	ExpASY proteomics server for computational analysis of the physical and chemical parameters of protein sequences	https://web.expasy.org/protparam/
SVMTriP	Predicts antigenic epitopes. It is for the realistic prediction of protein surface regions that are preferentially recognized by antibodies (antigenic epitopes). It aids the design of vaccine components and immuno-diagnostic reagents	http://sysbio.unl.edu/SVMTriP/
BepiPred	Prediction of B-cell epitopes	http://www.cbs.dtu.dk/services/BepiPred/
ElliProt	Prediction of continuous and discontinuous B-cell epitopes	http://tools.immuneepitope.org/tools/ElliPro/iedb_input
MHCPred	Predicts cytotoxic T-cell epitopes	http://www.ddg-pharmfac.net/mhcpred/MHCPred/
MHC2Pred	Predicts helper T cells	http://www.imtech.res.in/raghava/mhc2pred/

designs, the following tools have been extensively used for predicting antigenic peptides (Table 1).

Table 2 summarizes few machine learning tools that have been used for designing drug candidates for COVID-19 infection.

Conclusions

Bioinformaticians around the world have reacted quickly to the COVID-19 pandemic by providing specific tools to advance research on SARS-CoV-2 for the rapid detection and treatment of the disease. This has necessitated the development of

Table 2. Computational tools used in drug design.

TOOLS	DESCRIPTION	LINKS
CHARMM	It stands for Chemistry at Harvard Macromolecular Mechanics. It is a well-known simulation program.	https://www.charmm.org/charmm/
Amber	Amber is a suite of biomolecular simulation programs.	http://ambermd.org/
DOCK	Open-source docking program for academic purposes.	http://dock.compbio.ucsf.edu/
Zinc Pharmer	Open-source pharmacophore-based screening program.	http://zincpharmer.csb.pitt.edu/
Patchsearch	It is an R package for target prediction.	https://github.com/MTiPatchSearch/patchsearch
GANDI	This is a program for structure-based fragment-based ab initio ligand design.	http://www.biochem-caflisch.uzh.ch/download/
Hyde	Commercial program for binding affinity prediction.	https://www.biosolveit.de/Hyde/
LUDI	Automated program for structure-based ligand design. It comes incorporated in the Discovery Studio suite.	http://www.3dsbiovia.com/products/collaborative-science/biovia-discovery-studio/
CATALYST	Pharmacophore design and analysis program. It is a part of Discovery Studio suite.	http://www.3dsbiovia.com/products/collaborative-science/biovia-discoverystudio/pharmacophore-and-ligand-baseddesign.html
NAMD	It is suitable for parallel molecular dynamics simulations for larger biomolecular systems.	https://www.ks.uiuc.edu/Research/namd/

therapy and vaccination strategies against COVID-19 from various laboratories around the world. However, it is increasingly paramount that workers understand the molecular mechanisms underlying the disease pathogenesis for rapid identification of effective drug and vaccine candidates for clinical trials. Some bioinformatics applications hold promising therapeutic prospects in drug and vaccine design and development. We advocate drug repurposing to facilitate the identification of potential drugs via the screening of existing drugs and protein interactions, which is comparatively cheaper and more time-effective than traditional approaches of drug design.

Author Contributions

Onyeka S. Chukwudozie conceptualize the idea and made extensive literature search before drafting the manuscript. Vincent C Duru, Aborede Tunde and Victor Oyebanji contributed with the drafting of the manuscript. Charlotte Ndiribe assisted with the editorial works, while Benjamin Emikpe provided professional internal review work before submitting the manuscript.

ORCID iD

Onyeka S Chukwudozie  <https://orcid.org/0000-0002-0814-6798>

REFERENCES

- Gill S, Christopher A, Gupta V, Bansal P. Emerging role of bioinformatics tools and software in evolution of clinical research. *Perspect Clin Res.* 2016;7:115-122. doi:10.4103/2229-3485.184782.
- Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. *J Gen Virol.* 2012;93:1853-1868.
- Loman JC, Constantinidou JZ, Chan M, et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol.* 2012;10:599-606.
- Weinstock GW. Genomic approaches to studying the human microbiota. *Nature.* 2012;489:250-256.
- Anderson ST, Kaforou M, Brent AJ, et al. Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Engl J Med.* 2018;370:1712-1723.
- Chukwudozie OS, Chukwuanukwu RC, Iroanya OO, et al. Attenuated subcomponent vaccine design targeting the SARS-CoV-2 nucleocapsid phosphoprotein RNA binding domain: in silico analysis. *J Immunol Res.* 2020;2020:2837670. doi:10.1155/2020/2837670.
- World Health Organization (WHO). Statement regarding cluster of pneumonia cases in Wuhan, China. www.who.int. Published 2020.
- Hui DS, Azhar EI, Madani TA, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis.* 2020;91:264-266. doi:10.1016/j.ijid.2020.01.009.
- Cohen J. Wuhan seafood market may not be source of novel virus spreading globally. *Science.* 2020;10:1-9.
- Eschner K. We are still not sure where the COVID-19 really came from. *Popul Sci.* 2020;599:600.
- Sempowski GD, Saunders KO, Acharya P, Wiehe KJ, Haynes BF. Pandemic preparedness: developing vaccines and therapeutic antibodies for COVID-19. *Cell.* 2020;181:1458-1463. doi:10.1016/j.cell.2020.05.041.
- Ullah M, Sarkar B, Islam SS. Exploiting the reverse vaccinology approach to design novel subunit vaccines against Ebola virus. *Immunobiology.* 2020;225:151949. doi:10.1101/2020.01.02.20016311.
- Seib KL, Zhao X, Rappuoli R. Developing vaccines in the era of genomics: a decade of reverse vaccinology. *Clin Microbiol Infect.* 2012;18:109-116. doi:10.1111/1/j.1469-0691.2012.03939.
- Rappuoli R. Reverse vaccinology. *Curr Opin Microbiol.* 2000;3:445-450. doi:10.1016/S1369-5274(00)00119-3.
- Adu-Bobie J, Capecci B, Serruto D, Rappuoli R, Pizza M. Two years into reverse vaccinology. *Vaccine.* 2003;21:605-610. doi:10.1016/S0264-410X(02)00566-2.
- Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumor antigens and subunit vaccines. *BMC Bioinformatics.* 2007;8:4. doi:10.1186/1471-2105-8-4.
- María RR, Arturo CJ, Alicia JA, Paulina MG, Gerardo AO. The impact of bioinformatics on vaccine design and development. *Vaccines.* 2017;2:3-6. doi:10.5772/intechopen.69273.
- Enayatkhani M, Hasaniyazad M, Faezi S, et al. Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in-silico study [published online ahead of print May 2, 2020]. *J Biomol Struct Dyn.* doi:10.1080/07391102.2020.1756411.
- Ong E, Wong MU, Huffman A, He Y. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *Front Immunol.* 2020;11:1581. doi:10.3389/fimmu.2020.01581.
- Abraham Peele K, Srihansa T, Krupanidhi S, Vijaya Sai A, Venkateswarulu TC. Design of multi-epitope vaccine candidate against SARS-CoV-2: a in-silico study [published online ahead of print June 1, 2020]. *J Biomol Struct Dyn.* doi:10.1080/07391102.2020.1770127.
- Kar T, Narsaria U, Basak S, et al. A candidate multi-epitope vaccine against SARS-CoV-2. *Sci Rep.* 2020;10:10895. doi:10.1038/s41598-020-67749-1.
- Lee CH, Koohy H. In silico identification of vaccine targets for 2019-nCoV. *F1000Res.* 2020;9:145. doi:10.12688/f1000research.22507.2.

23. Norizadehtazehk M, Hajipour O. COVID-19 and bioinformatics. *Acta Sci Microbiol.* 2019;2:4-7.
24. Saha R, Prasad BVLS. In silico approach for designing of a multi-epitope-based vaccine against novel coronavirus (SARS-CoV-2). *bioRxiv.* doi:10.1101/2020.03.31.017459.
25. Ramaiah A, Arumugaswami V. Insights into cross-species evolution of novel human coronavirus 2019-nCoV and defining immune determinants for vaccine development. *bioRxiv.* doi:10.1101/2020.01.29.925867.
26. Delany I, Rappuoli R, Seib KL. Vaccines, reverse vaccinology, and bacterial pathogenesis. *Cold Spring Harb Perspect Med.* 2013;3:a012476. doi:10.1101/cshperspect.a012476.
27. Abdelmageed MI, Abdelmoneim AH, Mustafa MI, et al. Design of a multi-epitope-based peptide vaccine against the E protein of human COVID-19: an immunoinformatics approach. *Biomed Res Int.* 2020;2020:2683286. doi:10.1155/2020/2683286.
28. Ahmad S, Navid A, Farid R, et al. Design of a novel multi epitope-based vaccine for pandemic coronavirus disease (COVID-19) by vaccinomics and probable prevention strategy against avenging zoonotics. *Eur J Pharm Sci.* 2020;151:105387. doi:10.1016/j.ejps.2020.105387.
29. Bhattacharya M, Sharma AR, Patra P, et al. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-CoV-2): immuno-informatics approach. *J Med Virol.* 2020;92:618-631. doi:10.1002/jmv.25736.
30. Joshi A, Joshi BC, Mannan MA, Kaushik V. Epitope based vaccine prediction for SARS-CoV-2 by deploying immuno-informatics approach. *Inform Med Unlocked.* 2020;19:100338. doi:10.1016/j.imu.2020.100338.
31. Pourseif MM, Parvizpour S, Jafari B, Dehghanian J, Naghili B, Omid Y. Prophylactic domain-based vaccine against SARS-CoV-2, causative agent of COVID-19 pandemic [published online ahead of print December 10, 2020]. *J Virol.* doi:10.21203/rs.3.rs-32539/v1.
32. Yang Z, Bogdan P. An in-silico deep learning approach to multi-epitope vaccine design: a SARS-CoV-2 case study [published online ahead of print July 21, 2020]. *J Virol.* doi:10.21203/rs.3.rs-36528/v1.
33. van Regenmortel MHV. Structure-based reverse vaccinology failed in the case of HIV because it disregarded accepted immunological theory. *Int J Mol Sci.* 2016;17:3-7. doi:10.3390/ijms17091591.
34. Chukwudozie OS, Gray CM, Fagbayi TA, et al. Immuno-informatics design of a multimeric epitope peptide based vaccine targeting SARS-CoV-2 spike glycoprotein. *bioRxiv.* 2020. doi:10.1101/2020.07.30.228221.
35. Kames J, Holcomb DD, Kimchi O, et al. Sequence analysis of SARS-CoV-2 genome reveals features important for vaccine design. *bioRxiv.* doi:10.1101/2020.03.30.016832.
36. Pai K. What is the new coronavirus strain in UK and will vaccines work on it? *ThePrint.* 2020;1:1-2.
37. Ranganathan V, Madham P, Somayaji A. COVID-19: a novel strain challenging the human race. *J Microbiol Exp.* 2020;8:98-107. doi:10.15406/jmen.2020.08.00292.
38. Liu W, Fontanet A, Zhang PH, et al. Two-year prospective study of the humoral immune response of patients with severe acute respiratory syndrome. *J Infect Dis.* 2006;193:792-795. doi:10.1086/500469.
39. Tseng CT, Sbrana E, Iwata-Yoshikawa N, et al. Immunization with SARS coronavirus vaccines leads to pulmonary immunopathology on challenge with the SARS virus. *PLoS ONE.* 2012;7:e35421. doi:10.1371/journal.pone.0035421.
40. Silva-Arrieta S, Goulder PJR, Brander C. In silico veritas? Potential limitations for SARS-CoV-2 vaccine development based on T-cell epitope prediction. *PLoS Pathog.* 2020;16:e1008607. doi:10.1371/journal.ppat.1008607.
41. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. *BMC Immunol.* 2008;9:1. doi:10.1186/1471-2172-9-1.
42. Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics.* 2011;63:325-335. doi:10.1007/s00251-011-0513-0.
43. Frahm N, Yusim K, Suscovich TJ, et al. Extensive HLA class I allele promiscuity among viral CTL epitopes. *Eur J Immunol.* 2007;37:2419-2433. doi:10.1002/eji.200737365.
44. Sanchez-Mazas A. African diversity from the HLA point of view: influence of genetic drift, geography, linguistics, and natural selection. *Hum Immunol.* 2001;62:937-948. doi:10.1016/S0198-8859(01)00293-2.
45. Wang Y, Zhang S, Li F. Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.* 2020;48:D1031-D1041.
46. Mothay D, Ramesh KV. Binding site analysis of potential protease inhibitors of COVID-19 using AutoDock. *VirusDisease.* 2020;31:194-199. doi:10.1007/s13337-020-00585-z.
47. Eleftheriou P, Amanatidou D, Petrou A, Geronikaki A. In silico evaluation of the effectivity of approved protease inhibitors against the main protease of the novel SARS-CoV-2 virus. *Molecules.* 2020;25:2529. doi:10.3390/molecules25112529.
48. Ghosh AK, Takayama J, Aubin Y, et al. Structure-based design, synthesis, and biological evaluation of a series of novel and reversible inhibitors for the severe acute respiratory syndrome-coronavirus papain-like protease. *J Med Chem.* 2009;52:5228-5240. doi:10.1021/jm900611t.
49. Lin D, Liu L, Zhang M. Co-infections of SARS-CoV2 with multiple common respiratory pathogens in infected patients. *Sci China Life Sci.* 2020;63:606-609.
50. Li W, Moore MJ, Vasilieva MN. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature.* 2003;426:450-454.
51. Hamming I, Timens M, Bulthuis J. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol.* 2004;203:631-637.
52. Klenk HD, Garten W. Host cell proteases controlling virus pathogenicity. *Trends Microbiol.* 1994;2:39-43.
53. Steinhauer DA. Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology.* 1999;258:1-20.
54. Walls AC, Park YJ, Tortorici MA. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell.* 2020;181:281-292.
55. Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov.* 2020;18:41-58.
56. Parikesit AA, Nurdiansyah R. The challenge of protein domain annotation with supervised learning approach: a systematic review. *J Mat Sains.* 2014;24:1-9. <https://jms.fmipa.itb.ac.id/index.php/jms/article/view/775>.
57. Valeska MD, Adisurja GP, Bernard S, Wijaya R, Aldino M, Parikesit AA. The role of bioinformatics in personalized medicine: your future medical treatment. *Cermin Dunia Kedokt.* 2019;46:785-788. <http://www.kalbedmed.com/Desktop-Modules/EasyDNNNews/DocumentDownload.ashx?portalid=0&moduleid=471&articleid=9018&documentid=1012>.
58. Roy K, Kar S, Das RN, Roy K, Kar S, Das RN. Background of QSAR and historical developments. In: Roy K, Kar S, Narayan Das R, eds. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment.* Amsterdam, The Netherlands: Elsevier; 2018:1-46. <http://linkinghub.elsevier.com/retrieve/pii/B9780128015056000016>.
59. Moroy G, Martiny VY, Vayer P, Villoutreix BO, Miteva MA. Toward in silico structure-based ADMET prediction in drug discovery. *Drug Discov Today.* 2019;17:44-55. <http://www.ncbi.nlm.nih.gov/pubmed/22056716>.
60. Oprea TI. Virtual screening in lead discovery: a viewpoint. *Molecules.* 2006;7:51-62. <http://www.mdpi.com/1420-3049/7/1/51>.
61. Parikesit AA, Nurdiansyah R. Supplementary material for the project: drug repurposing option for COVID-2019 with structural bioinformatics of chemical interactions approach. *Mendeley Data.* 2020;13:205. doi:10.17632/sw43d9g2r6.2.
62. Liang T. Handbook of COVID19 prevention and treatment. *Zhenjiang.* <https://covid-19.alibabacloud.com>. Published 2020.
63. Arli A, Parikesit R, Nurdiansyah H. Drug repurposing option for COVID-19 with structural bioinformatics of chemical interactions approach. *Cermin Dunia Kedokteran.* 2020;47:222-226.
64. Corman VM, Landt O, Kaiser M. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance.* 2020;25:3-7.
65. Moore SC, Penrice-Randal R, Alruwaili M. Amplicon based MinION sequencing of SARS-CoV-2 and metagenomic characterization of nasopharyngeal swabs from patients with COVID-19. *MedRxiv.* <https://www.medrxiv.org/content/10.1101/2020.03.05.20032011v1>. Published 2020.
66. Sadeh S, Matschinske J, Blumenthal DB. Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. *arXiv.* <https://arxiv.org/ftp/arxiv/papers/2004/2004.12420.pdf>. Published 2020.
67. Pfeifferle S, Schöpf J, Kögl M, et al. The SARS Coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. *PLoS Pathog.* 2011;7:e1002331.
68. Gordon DE, Jang GM, Bouhaddou M. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature.* 2020;583:459-468.
69. Kotlyar M, Pastrello C, Malik Z, et al. IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res.* 2018;47:D581-D589.
70. Mendez D, Gaulton A, Bento AP, et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* 2019;47:D930-D940.
71. Wishart DS, Feunang YD, Guo AC. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2017;46:D1074-D1082.
72. Wang J. Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study. *J Chem Inf Model.* 2020;60:3277-3286. doi:10.1021/acs.jcim.0c00179.
73. Sajid A, Singh Y, Shukla P. Computational tools for modern vaccine development. *Hum Vaccin Immunother.* 2019;16:723-735. doi:10.1080/21645515.2019.1670035.
74. Sanami S, Zandi M, Pourhossein B, et al. Design of a multi-epitope vaccine against SARS-CoV-2 using immune-informatics approach. *Int J Biol Macromol.* 2020;164:871-883. doi:10.1016/j.ijbiomac.2020.07.117.
75. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses.* 2020;12:254. doi:10.3390/v12030254.
76. Tunde AA. A systematic review on the potency and how safe chloroquine is for the treatment of COVID-19. *AJIR Preprint.* 2020;2:3-6.
77. Jin Z, Du X, Xu Y, et al. Structure of M^{pro} from SARS-CoV-2 and discovery of its inhibitors. *Nature.* 2020;582:289-293. doi:10.1038/s41586-020-2223-y.