OXFORD

Infectious disease mRNA vaccines and a review on epitope prediction for vaccine design

Xinhui Cai, Jiao Jiao Li, Tao Liu, Oliver Brian and Jinyan Li

Corresponding author: Jinyan Li. Data Science Institute, Faculty of Engineering & IT, University of Technology Sydney, 15 Broadway, Ultimo, 2007, New South Wales, Australia. Tel.: +61 295149264; Fax: +61 295149264; E-mail: jinyan.li@uts.edu.au

Abstract

Messenger RNA (mRNA) vaccines have recently emerged as a new type of vaccine technology, showing strong potential to combat the COVID-19 pandemic. In addition to SARS-CoV-2 which caused the pandemic, mRNA vaccines have been developed and tested to prevent infectious diseases caused by other viruses such as Zika virus, the dengue virus, the respiratory syncytial virus, influenza H7N9 and Flavivirus. Interestingly, mRNA vaccines may also be useful for preventing non-infectious diseases such as diabetes and cancer. This review summarises the current progresses of mRNA vaccines designed for a range of diseases including COVID-19. As epitope study is a primary component in the *in silico* design of mRNA vaccines, we also survey on advanced bioinformatics and machine learning algorithms which have been used for epitope prediction, and review on user-friendly software tools available for this purpose. Finally, we discuss some of the unanswered concerns about mRNA vaccines, such as unknown long-term side effects, and present with our perspectives on future developments in this exciting area.

Key words: mRNA vaccines; machine learning; epitope prediction; COVID-19; other infectious diseases.

Introduction

Since corona virus disease-2019 (COVID-19) emerged in December 2019, it has grown into a global pandemic with more than a hundred million people infected and over two million deaths worldwide [86]. Although preventative measures such as physical distancing and improved hygiene practices have been effective in many countries, the number of positive cases is still rising rapidly in numerous regions around the world. Some countries have shown consistent high positive test rates, including Mexico (36.87%, 7/100 000 tested), Argentina (63.77%, 83/100 000 tested), Ecuador (40.56%, 34/100 000 tested) and India (21.64% with daily confirmed case numbers above 350 000) [2]. Even in developed countries such as the USA, Germany and Italy, the observed case fatality ratio remains above 1%, whereas in Mexico the figure stands at an alarming 9.1% [81].As a result, stopping the spread of COVID-19 continues to be the top priority for global communities and healthcare systems.

Xinhui Cai is a higher degree research (HDR) student at Data Science Institute in University of Technology Sydney (UTS). His research interests include machine learning and COVID-19 vaccine design.

Dr Jiao Jiao Li is a lecturer in Biomedical Engineering at UTS, a National Health and Medical Research Council (NHMRC) Early Career Fellow, a chief investigator on the Australian Research Council Training Centre for Innovative BioEngineering, and a Science & Technology Australia 2021–22 Superstars of STEM. Her research in regenerative medicine focuses on using stem cells and other approaches to treat chronic musculoskeletal conditions such as osteoarthritis.

Dr Tao Liu is an associate professor at Children's Cancer Institute Australia, leading the Gene Dysregulation Group. He is also a conjoint associate professor in the UNSW Faculty of Medicine. His research covers tumor suppressor gene expression, anticancer efficacy of various inhibitors and identification of the mechanisms important for the stability and/or degradation of Myc oncoproteins.

Prof. Brian Oliver has strong interests to identify and develop new ways of treating respiratory diseases. He currently leads the Respiratory Cellular and Molecular Biology Group with laboratory facilities located at both UTS and the Woolcock Institute. The work from his group is recognised to be amongst the best in the world, evidenced by selection for presentation at symposia at both national and large international conferences as well as resulting in prestigious publications.

Prof. Jinyan Li is a Professor at Data Science Institute in University of Technology Sydney (UTS). He has a broad range of interests to invent data mining algorithms to solve challenging problems in genomics and computational biology.

Vaccination is the most effective method for preventing the spread of infectious diseases, including COVID-19. Since the first vaccine was developed for smallpox more than 200 years ago [46], vaccines have greatly improved population health for communities around the world by significantly reducing the disease burden of common infections such as measles, mumps, rubella, diphtheria, typhoid and tetanus. Vaccines induce immune responses by simulating the infection process inside the human body and enabling it to develop lasting immunity. When the body is re-exposed to the target antigens, such as those carried by real pathogens, it is able to produce antibodies which neutralise the antigens and therefore prevents infection.

Messenger RNA (mRNA) vaccines (first used as a new class of therapeutic drugs in 1989 [83]) encode the desired antigens from an mRNA sequence. In practice, the injected mRNAs provides cells information to produce desired proteins in the cytoplasm, which are subsequently presented on the cell surface to trigger immune responses involving antigen-presenting cells (APCs) or antibodies/immunoglobulin, resulting in immunity to specific diseases. This review will focus on this type of vaccines.

Due to the speed and efficiency with which mRNA vaccines can be developed and produced, they have recently been considered as a strong vaccine candidate in the fight against COVID-19. For example, Moderna [15] had only just completed the first clinical batch of its mRNA vaccines, mRNA-1273, in February 2020. Only a year later, mRNA-1273 has already been approved for vaccination in many countries. Compared to other types of vaccines, the characteristics of mRNA vaccines are highly suited for targeting diseases with high infectivity and genetic instability.

Epitopes are a primary yet often overlooked aspect for boosting the effectiveness of mRNA vaccines. These can be classified into T-cell and B-cell epitopes depending on which part of the immune system is being triggered. B-cell epitopes are either conformational epitopes or linear epitopes. A conformational B-cell epitope is a subset of residues that are closely compacted in 3D space but not continuous in primary amino acid sequence, whereas a linear B-cell epitope contains a continuous stretch of amino acids in the primary sequence [7]. Epitope-based vaccine design has already been used to develop peptide-based vaccines, including those for SARS-CoV-2 such as UB-612 [24] and NVX-CoV2373 [80]. However, mRNA vaccines can similarly benefit from epitope-based design approaches, where both B-cell and T-cell epitopes can be used for vaccine design. The epitope properties determine whether the mRNA vaccine can trigger an immune response, and which types of responses will be triggered. Epitope prediction allows researchers to find effective epitopes that can offer both immunogenicity and cross-reactivity for a target pathogen [49]. For viruses, the predicted epitope usually overlap with the receptor binding domain (RBD), a key structure on the viral spike (S) protein. The S protein is the main antigenic component that binds to cell receptors and facilitates entry of the viral genome into the host cell [78]. The epitopes of many viruses can be found in online databases such as the Immune Epitope Database (IEDB), which also offer tools such as prediction models to analyse and apply the information [1].

In this review, we summarise the general characteristics and mechanisms of mRNA vaccines, and compare them with other vaccines types. We also present an overview on the latest progress in mRNA vaccine development for SARS-CoV-2, as well as a range of other diseases both infectious [Zika virus (ZIKV), Dengue virus (DENV), respiratory syncytial virus (RSV), influenza virus H7N9 and flavivirus] and non-infectious (diabetes and cancer). Offering a fresh perspective into mRNA vaccine design, we review *in silico* methods based on computational epitope prediction by machine learning, as well as user-friendly software tools and databases available for this purpose. The main topics covered in this review are schematically summarised in Figure 1. The focus of our review is on mRNA vaccines and computational tools used to optimise mRNA vaccine design, which is a unique direction compared to other recent reviews focusing on COVID-19 mRNA vaccines [36], which have detailed the progress of COVID-19 vaccines in preclinical and clinical studies, and specifically evaluated their safety and efficacy among other aspects.

Characteristics and mechanisms of mRNA vaccines and the comparison with other types of vaccines

mRNA vaccines properties

Two major types of mRNA have been studied as vaccines: non-replicating mRNA and virally derived, self-amplifying mRNA [83]. A key advantage of mRNA as a source of antigen is its efficiency in evoking MHC-I presentation and eliciting cytotoxic T-lymphocyte responses. This allows a high degree of versatility in the type and number of antigenic determinants, allowing fast vaccine development. The basic structure of *in vitro* transcribed (IVT) mRNA consists of a protein-encoding open reading frame (ORF) flanked by 5' and 3' untranslated regions, ending with a 7-methyl guanosine 5' cap structure and a 3' poly(A) tail. Upon entering the host cell, the mRNA is translated by the cell to produce proteins. This makes mRNA vaccines suitable for delivering cytosolic or transmembrane proteins to the correct cellular compartments [57]. Figure 2 is an illustration of the mechanism by which mRNA vaccines mediate immune responses. Other advantages of mRNA vaccines include their generic production procedure, ease of design and relative safety. These are discussed further in the following section.

The common administration routes for mRNA vaccines are needle-syringe and intra-muscular injection [92]. Other methods such as needle-free and intra-dermal injection are also possible, and have been shown to have a similar level of effectiveness as traditional administration routes in experimental studies [92]. Lipid nanoparticles (LNPs), as advanced delivery systems, are a viable approach to increase the efficacy and stability of mRNA vaccines [83, 92]. Compared to naked mRNA, encapsulation within nanoparticles offers much better protection against degradation, and also allows flexibility in controlling biodistribution, cellular targeting and cellular uptake mechanisms. Charge-altering releasable transporters is another type of delivery system that has proven effective in some cases [25], but is not commonly used in prophylactic vaccines.

Despite of promising results in animal experiments and some human trials, there are still concerns about mRNA vaccines efficiency due to disadvantages such as short exposure time to the immune system, low stability of mRNA, the need for adjuvants, and immunodominant effects. Thess et al. [79] engineered mRNA sequences to enhance protein expression and suppress cytokine secretion, and the unmodified engineered mRNA was found to outperform its pseudouridine-modified counterpart. In contrast, Liang et al. [39] proposed that modified mRNA vaccines can offer precision in antigen design, as well as good tolerability and broad immune responses. The common belief in the research community is that unmodified sequences
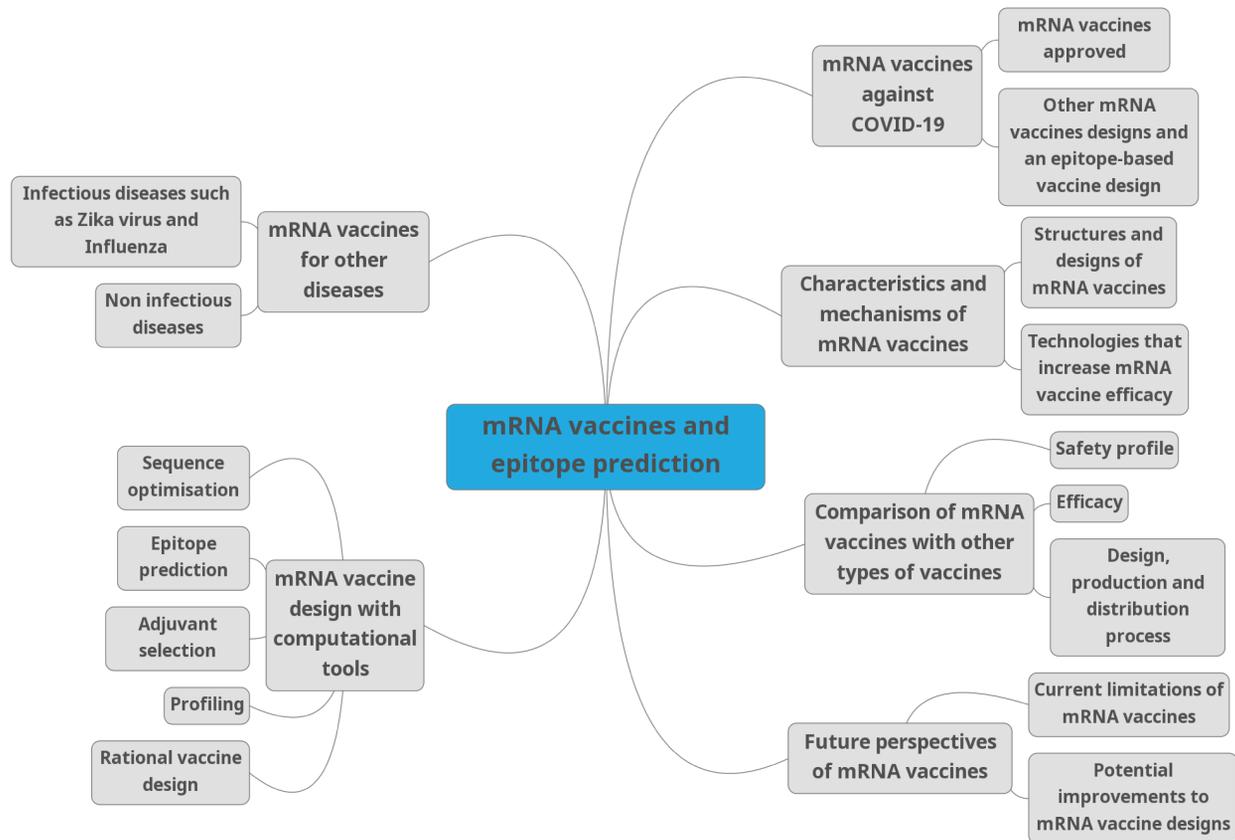
Figure 1. Schematic summarising the main topics in this review.

do not provide attractive features, whereas modification offers much greater potential in improving immunogenicity and pharmacokinetics [34]. Another common measure to increase the immune response for mRNA vaccines is to use adjuvants, but this topic is beyond the scope of the current review.

Comparison of mRNA vaccines with other types of vaccines

To better understand the advantages of mRNA vaccines, this section compares their characteristics with other existing vaccine types such as live attenuated vaccines and subunit vaccines.

Safety profile

The safety profile is arguably the most crucial property of a vaccine. mRNA vaccines typically possess a good safety profile as they do not contain live viruses. As recent researches [37, 47, 55, 83] pointed out, mRNA vaccines also do not pose any risks of genetic integration. Through experiments with SARS-CoV-2 vaccines, Krammer [36] noted that the adverse effects for mRNA vaccines were dose-dependent. The two tested mRNA vaccines in this study showed a superior safety profile, especially when compared to viral vector ChAdOxnCoV-19 or Ad5-based vaccines. However, when Liu [44] compared mRNA vaccines with plasmid DNA vaccines, the author noted potential toxicity issues with mRNA vaccines that were not predicted by preclinical safety testing due to species differences between humans and the animal models used.

Efficacy

Efficacy is the primary measure for the success of a vaccine. Since their conception and development, mRNA vaccines have been criticised for their instability and low translation efficiency [83]. However, many technologies such as LNPs and adjuvants have largely resolved these concerns. For SARS-CoV-2 vaccine candidates, Krammer [36] stated that adjuvanted protein-based vaccines have the best performance in terms of immunogenicity, followed by mRNA vaccines, the viral vector ChAdOx1 and finally the AdV5-based vaccines. Pandey et al. [55] suggested that recombinant vector vaccines could better trigger immune responses than mRNA vaccines due to their longer exposure time *in vivo*. Liu [44] noted that DNA vaccines could persist for a longer period of time than mRNA, but their translation efficiency might be reduced due to the need to enter the nucleus. Nevertheless, in some animal experiments, mRNA vaccines were shown to be more effective than DNA vaccines [37, 56]. Immunisation strategies such as escalated antigen delivery and cross-reactive antigen design proposed by Burton and Walker [9] could help to increase the performance of mRNA vaccines.

Design, production and distribution

The flexible and straightforward design process of mRNA vaccines has enabled their timely development against rapidly evolving infections such as COVID-19. Traditional vaccines such as live attenuated or inactivated vaccines require thorough researching to develop adequate processes for culturing the target virus, and killing or attenuating the virus before
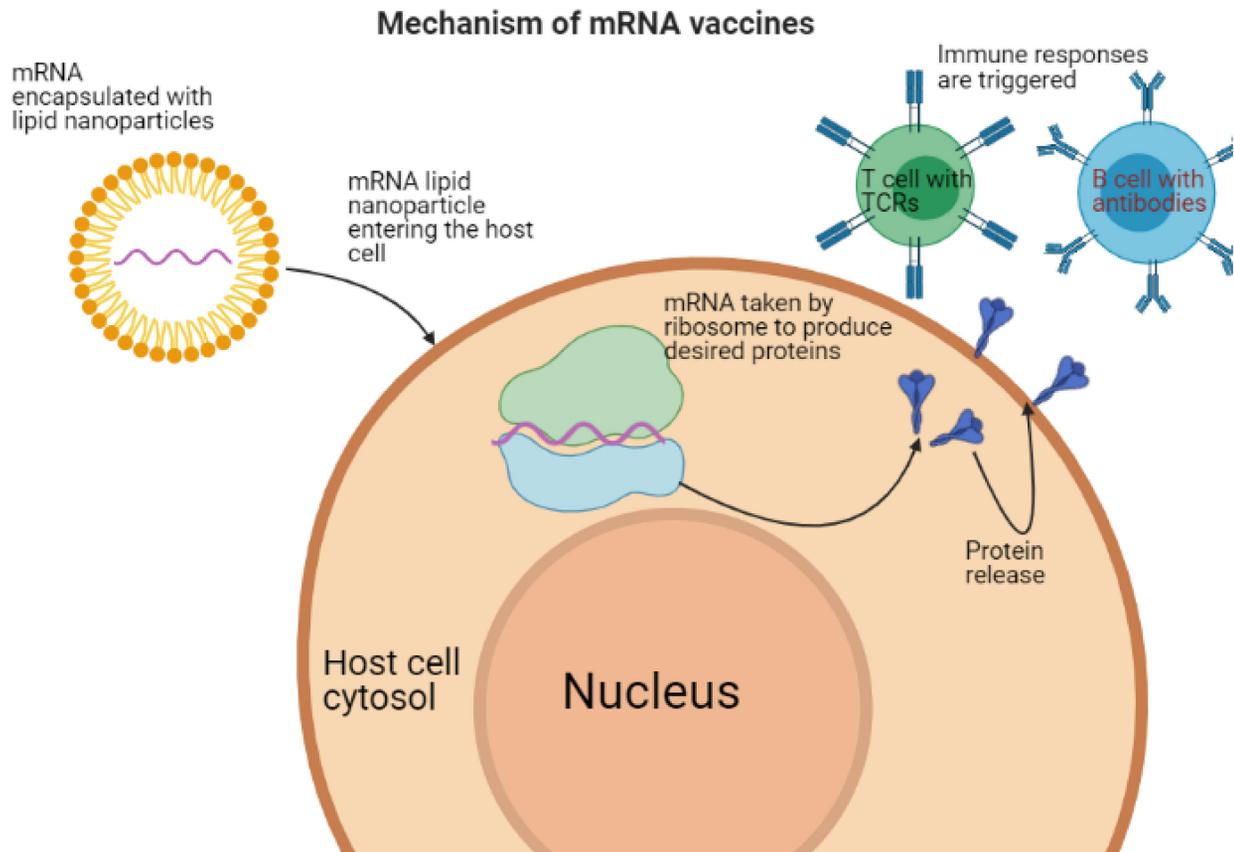
Figure 2. Mechanism of mRNA vaccines.

production [22]. In comparison, mRNA vaccines only require the viral sequence and structural information for vaccine design [14], using a universal platform with the only difference being in the ORF for different viruses or strains. As an example, the two leading mRNA vaccines for SARS-CoV-2 use LNP encapsulation, together with a generic mRNA vaccine IVT process [15, 84]. This greatly increases the speed of mRNA vaccine design and development, as seen in Table 1 where mRNA-1273 was the first vaccine for SARS-CoV-2 to enter clinical trials.

The requirements for production and distribution are often overlooked in designing vaccines, but are important considerations for their widespread applications. mRNA vaccines require frozen delivery chains and storage, creating a significant challenge for large-scale immunisation particularly in less developed regions. These practical restrictions are a serious shortcoming of mRNA vaccines compared to other vaccine types [36]. Nevertheless, challenges in the distribution of mRNA vaccines are somewhat offset by their simple and straightforward production [47, 55]. Faced with the current massive demand for SARS-CoV-2 vaccines, the ability of mRNA vaccines to offer efficient production and up-scaling is a distinct advantage. However, most mRNA vaccines have strict storage and transportation requirements, but there have been some successful attempts to lower such requirements and make mRNA vaccines more viable and accessible Zhang et al. [94].

The ability of mRNA vaccines to provide long-term global-scale immunisation is difficult to determine since most are yet to be tested in large-scale human trials. Although mRNA vaccines may not be suitable for use against all diseases, they do present clear advantages in combining a good safety profile, acceptable immunogenicity, and quick design and development. In light of the urgent need to develop a vaccine for a fast-evolving and unstable RNA virus that has caused a global pandemic, mRNA vaccines could provide a viable solution to be used in the fight against SARS-CoV-2.

mRNA vaccines currently designed for COVID-19

Several mRNA vaccines have been designed for immunisation against COVID-19, some of which have been approved for public use after demonstrating superior safety and efficacy in clinical trials. Of the vaccine candidates reported in April 2020 (Table 1), those with the most advanced progress are the BNT162b1 from Mulligan et al. [51], BNT162b2 from Walsh et al. [84], and mRNA-1273 from Corbett et al. [15], all of which had quickly and safely entered phase 3 clinical trials. Among these, the Moderna vaccine Corbett et al. [15] was the first to commence clinical trials and be implemented for large-scale immunisation, including in the USA, Canada, European Union, Israel and Switzerland (list expanding). BNT162b2 has been approved for use in the UK, USA, Canada, Singapore, Saudi Arabia, Argentina, Switzerland, the European Union and some other Latin American countries (list expanding).

On 9 November 2020, Pfizer and BioNTech announced that BNT162b2 was found to be more than 90% effective during

Table 1. Progress of SARS-CoV-2 vaccines in April 2020 and May 2021 (NA means no data)

Institutions	Approaches	April 2020 progress	May 2021 progress
Moderna (mRNA-1273)	mRNA	Phase 1 clinical trial	Completed trials and immunised massively
Curevac	mRNA	Preclinical phase	Phase 2b/3 trial
ExpresS2ion	Recombinant-protein	Preclinical phase	Clinical trial
iBio	Recombinant-protein	Preclinical phase	Completed toxicity studies
Novavax	Recombinant-protein	Preclinical phase	Completed clinical trial and in production
Baylor College of Medicine	Recombinant-protein	Preclinical phase	Clinical trial
University of Queensland	Recombinant-protein	Preclinical phase	Stopped at phase 1 trial
Sichuan Clover Biopharmaceuticals	Recombinant-protein	Preclinical phase	Phase 2/3 trial
Vaxart	Viral-vector	Preclinical phase	Completed phase 1 trial
Geovax	Viral-vector	Preclinical phase	Animal testing
University of Oxford	Viral-vector	Preclinical phase	Completed trials and immunised massively, stopped in some countries
Cansino Biologics	Viral-vector	Preclinical phase	Phase 3 trial
Inovio	DNA	Preclinical phase	Phase 2/3 trial
Applied DNA Sciences	DNA	Preclinical phase	Completed phase 1 trial
Codagenix	Live-attenuated vaccine	Preclinical phase	Phase 1 trial and approved in 40 countries
Pfizer	mRNA	Phase 1 clinical trial	Completed trials and immunised massively
Anhui Zhifei Longcom	Viral-vector	Preclinical phase	Completed clinical trial and approved in 2 countries
Bharat Biotech	Inactivated	Preclinical phase	Completed clinical trial and approved in 9 countries
Chumakov Center	Inactivated	NA	Approved in one country
FBRI	Protein subunit	Preclinical phase	Approved in one country
Gamaleya	Viral-vector	NA	Phase 3 trial
Janssen	Viral-vector	Preclinical trial	Completed clinical trial and immunised massively, stopped in some countries
RIBSP Kazakhstan	Inactivated	NA	Phase 3 clinical trial approved in 1 country
Beijing/Wuhan Institute of Biological Products	Inactivated	Phase 1 trial	Completed trials and immunised massively
Sinovac	Inactivated	Phase 1 trial	Completed trials and immunised massively

phase 3 clinical trials. However, some researchers have raised doubts about the actual efficacy of BNT162b2 since a number of suspected cases of COVID-19 in the study population were not included in the efficacy analysis, and Pfizer had been prompted to provide further clarification on this issue. Nevertheless, in Israel, mass vaccination on a nationwide level with over 1.5 million participants using two doses of BNT162b2 showed 92% effectiveness for documented infection [17]. On 16 November 2020, Moderna announced a 94.5% effective result in the phase 3 trial of mRNA-1273. Both BNT162b2 and mRNA-1273 mRNA vaccines adopted the same design principle [87], using the LNP approach with similar encoding regions, which comprise the spike protein with two proline mutations to lock in the prefusion conformation.

Other mRNA and protein-based vaccine designs against SARS-CoV-2

The two leading mRNA vaccines described in the previous section require strict cold-chain delivery and frozen storage to maintain their effectiveness. To circumvent distribution requirements, Zhang et al. [94] designed an mRNA-LNP vaccine, ARCoV, that could be stored at room temperature for 1 week. By encoding SARS-CoV-2 RBD into the vaccine and transfecting the target antigen in multiple cell lines, a high expression of recombinant RBD was obtained in culture supernatants. Kinetics analysis demonstrated a high affinity for recombinant human ACE2 from RBD protein brought by the mRNA. In mice challenged with

SARS-CoV-2, the control mice showed high levels of viral RNA in the trachea and lungs, while all mice receiving mRNA-LNP were fully protected.

As an example of how new technologies could be implemented in the development of COVID-19 vaccines, Ahammad and Lira [3] proposed an immunoinformatic approach for vaccine design. A model workflow was demonstrated using epitope prediction technology, involving retrieval of the SARS-CoV-2 spike glycoprotein sequence, followed by computational prediction of epitopes for cytotoxic T-lymphocytes, helper T-lymphocytes, and linear B-lymphocytes as well as epitope antigenicity. Although this work is yet to reach experimentation on animals or humans, it provides a prime example of utilising computational tools for mRNA vaccine design.

Ong et al. [54] utilised machine learning as a specific form of computational tools, to help design a COVID-19 vaccine using reverse vaccinology. The process began by running Vaxign-ML to predict conserved regions of SARS-CoV-2, and calculating the protective antigenicity score. An optimised supervised machine learning model was used, with manually annotated training data consisting of bacterial and viral protective antigens. The most promising vaccine candidate was predicted to be the S protein, followed by the nsp3 protein which had not been tested in other coronavirus vaccine studies. Nsp3 was predicted to contain promiscuous MHC-I and MHC-II T-cell epitopes (28 and 42, respectively), which covered the majority of the world's population, as well as linear B-cell epitopes found to be localised on the surface of the protein.
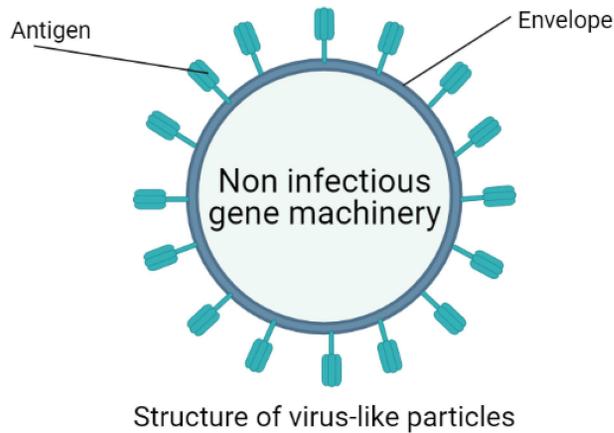
Figure 3. Virus-like particles (VLP) are viral protein complexes which mimic the native virus genetic material but are non-pathogenic and non-replicative. The surface of VLP could carry proteins of interest and since it resembles native virus, the immune system could recognize it quickly. Some SARS-CoV-2 vaccines are based on VLP, such as the Medicago's VLP vaccine [85], but it is also possible to encode VLP in mRNA vaccines.

In addition to the popular choice of encoding proteins and genes, a different technique using virus-like particles (VLPs) can be applied to SARS-CoV-2 mRNA vaccine design. Lu et al. [45] proposed three possible formulations of mRNA vaccines: encoding the RBD of the S protein, encoding the full-length S protein or VLPs with an mRNA cocktail encoding the S, M (membrane) and E (envelope) structural proteins. When the immunogenicity of these three formulations was tested in mice by serum evaluation, the VLP formulation produced superior immune responses (NAbs and T-cell reponses). No adverse effects were observed for any of the three formulations at the site of injection. The graphical representation of a VLP is shown in Figure 3, and a representation of mRNA vaccines encoding the S protein is shown in Figure 4. An additional option for mRNA vaccines is to encode epitopes [49], as shown in Figure 5.

Only a few mRNA vaccines to date have entered clinical trials. Nevertheless, their high flexibility allows mRNA vaccines to adopt a range of different designs. The epitope-focused design approach is gaining increasing popularity among researchers developing SARS-CoV-2 mRNA vaccines. This reinforces the need to develop computational models that allow epitope information to be easily applied in vaccine design.

mRNA vaccines designed for preventing other diseases

Though not yet common, mRNA vaccines have already been developed and tested to prevent several viral infections such as ZIKV, DENV, RSV, influenza virus (specifically H10N8 and H7N9) and flavivirus. Other targets of mRNA vaccines include non-infectious diseases such as diabetes and cancer. It is worth noting that several mRNA vaccines developed to treat cancer used an *ex vivo* dendritic cell (DC) loading approach instead of nanoparticles or LNPs, a technique that is rarely used in other types of vaccine designs for infectious diseases. A list of the mRNA vaccines discussed in this section is presented in Table 2.

Epitope studies proven to be helpful in the understanding of SARS-CoV-2 [70] could similarly inform vaccine design for other diseases. In this section, we discuss examples of the tools and delivery systems, as well as testing methods and outputs used in the development of mRNA vaccines against infectious and non-infectious diseases (other than COVID-19).

Zika virus

ZIKV is an mRNA virus that can be spread by mosquito bites, for which there is no vaccine or specific treatment other than supportive therapy. Richner et al. [66] designed an mRNA vaccine that targets ZIKV infection with optimised LNPs encapsulating modified mRNA to induce high levels of protein expression *in vivo*. It included full-length prM and E genes (encoding prM and E proteins; both are crucial to ZIKV) from an Asian ZIKV strain. Due
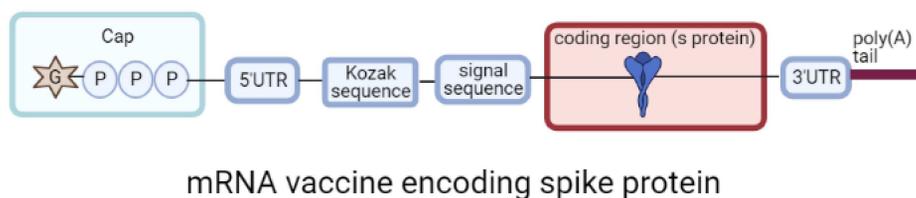
Figure 4. mRNA vaccines encoding the spike protein is the design of most mRNA vaccines for SARS-CoV-2, including the two leading ones, mRNA-1273 and BNT162B2. The spike protein is located in the ORF, and the entire structure is encapsulated by LNP.
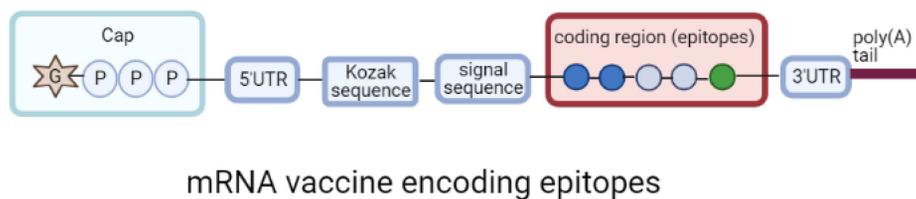
Figure 5. mRNA vaccines encoding epitopes is not a common design of mRNA vaccines. Instead of carrying genes or proteins, the ORF carries a sequence of epitopes. This is a novel method of presenting antigens, which demonstrates the flexibility of mRNA vaccines.

Table 2. mRNA vaccines for infectious and non-infectious diseases (other than COVID-19)

Author	Target	Design of encodings	Experiment subjects
Richner et al. [66]	ZIKV	Full length prM and E gene	Mouse
Roth et al. [67]	DENV	DENV1-NS poly-epitope	Mouse
Espeseth et al. [19]	RSV	mF RNA	Rat and mouse
Feldman et al. [20]	H10N8 & H7N9	HA glycoproein	Human
Firdessa-Fite and Creusot [21]	Diabetes	Epitope sequence	Mouse
VanBlargan et al. [82]	Flavivirus	prM and E genes	Mouse
Cafri et al. [10]	Cancer	Neoantigens	Human
Chahal et al. [12]	ZIKV	prM and E proteins	Mouse

to the similarities between ZIKV and DENV, there is a theoretical concern that vaccine-induced cross-reactive antibodies could cause more severe DENV infection through antibody-dependent enhancement. As a result, the researchers hence modified the ZIKV mRNA vaccine on four mutations in or near the E-DII-FL epitope (an epitope on the DENV E protein). This modification was shown to be protective against ZIKV while diminishing the production of antibodies enhancing DENV infection in cells or mice. Richner et al. [66] assessed the immunogenicity and protective activity of their mRNA vaccine by challenging AG129 mice with ZIKV virus 42 days after vaccination. With the exception of one animal, all the mice that received mRNA vaccines, with or without a boost, survived.

Chahal et al. [12] encoded prM and E proteins of ZIKV into an RNA replicon vector as ORF, and used IVT to make the vaccine. The researchers generated an overlapping 15-mer peptide library spanning amino acids 105 to 713 of the ZIKV polyprotein, along with seven peptides to induce T-cell response. They evaluated H-2Db- and H-2Kb-binding epitopes from all seven individual 'hits' with ANN-based models from IEDB. The two best-ranked peptides were selected for solid-phase peptide synthesis (SPPS). Chahal et al. [12] used mice to test the vaccine efficacy, where C57BL/6 mice were immunised by intra-muscular injection of the test vaccine, and the control group was immunised with a similar RNA replicon vaccine encoding the Ebola virus glycoprotein. Only two mice in the control group exhibited seropositivity above the detection limit of the assay, compared to all animals in the vaccine group.

Although both mRNA vaccines chose to encode the prM-E proteins as the immunogen, they are structurally different. Although the design from Richner et al. [66] used LNP as encapsulation and required two doses, and Chahal et al. [12] designed a modified dendrimer nanoparticle encapsulated mRNA vaccine that only required one dose. The LNP encapsulation has been the preferred choice due to its maturity and effectiveness, but the modified dendrimer nanoparticle encapsulation is also worth exploiting. Additionally, the work of Richner et al. [66] measured the efficacy of serum neutralization titers, whereas Chahal et al. [12] targeted T-cell responses. In addition to immunity, Richner et al. [66] took reducing the cross-reactivity issue with DENV infection into consideration and tackled it by modifying the epitopes.

Dengue virus

DENV is another RNA virus that has become a global public health threat in recent decades. Roth et al. [67] proposed that $CD4^+$ and $CD8^+$ T-cells contribute to protection against DENV, with each targeting different structural/non-structural proteins.

A prototype consensus sequence based on epidemic strains of DENV1 was selected, which was the DENV1-NS T-cell poly-epitope based on the analysis of $CD8^+$ T-cell epitopes from human donors. Using this, an LNP-encapsulated modified mRNA vaccine encoding DENV1-NS was prepared. The vaccine and boost were given at a 4-week interval, with the mice challenged 4 weeks after the boost to test the vaccine efficacy. The assessment included quantification of viral loads, interferon (IFN) splenocytes, and neutralisation assay. Two injections of the mRNA vaccine were found to provide significant protection against the virus.

Respiratory syncytial virus

RSV is a common cause of respiratory infection in children. Espeseth et al. [19] designed a modified mRNA LNP-based vaccine against RSV infection. Due to its conserved nature among serotypes, neutralising antibodies induced by natural RSV infection predominantly target the RSV F protein, providing an attractive vaccine target. Since RSV F protein is difficult to express and purify, mRNA vaccines offer the ability to efficiently explore multiple antigen designs. Espeseth et al. [19] injected cotton rats challenged with the RSV-A2 virus with their designed mRNA vaccines, expressing either prefusion stabilised or native forms of RSV F protein. The mRNA vaccines showed better immunogenicity compared to a protein subunit vaccine. Since the F protein is conserved across strains, the mRNA vaccines designed for the RSV-A strain also provided cross-protection for RSV-B. The rats injected with mRNA vaccines showed no signs of vaccine-elicited respiratory disease, which was a potential safety concern.

Influenza H10N8 and H7N9

Influenza is highly contagious and the most common infectious disease worldwide. Feldman et al. [20] evaluated the immunogenicity and safety of the first mRNA vaccines against H10N8 and H7N9 influenza viruses. These vaccines consisted of chemically modified mRNAs encoding the full-length, membrane-bound form of the hemagglutinin (HA) glycoprotein from the H10N8 or H7N9 influenza strains. Both used an LNP delivery system. In phase 1 randomised, placebo-controlled, double-blinded clinical trials, participants were tested for immunogenicity by hemagglutination inhibition and micro-neutralisation assays, as well as peripheral blood mononuclear cell persistence. After 6 months of immunisation, 22 of 23 participants receiving the H10N8 vaccine remained seropositive, whereas for the H7N9 vaccine the protection rate was 52%.

Flavivirus

Powassan virus (POWV) is a flavivirus transmitted by ticks, which may cause life-threatening encephalitis. VanBlargan et al. [82] designed an LNP-encapsulated modified mRNA vaccine, consisting of base-modified mRNA encoding the prM and E genes of POWV Spooner, a POWV strain. This mRNA was preceded by the prM signal sequence (also known as leader peptide or transit peptide) of POWV, or signal sequence of the Japanese encephalitis virus (JEV). C57BL/6 mice were immunised with POWV, JEV or placebo LNPs followed by a second boost. When challenged with POWV Spooner, 100% of the mice in the placebo-controlled group died, whereas all POWV and JEV immunised mice were survived.

Diabetes

Firdessa-Fite and Creusot [21] compared LNPs with DCs as two major types of delivery vehicles for mRNA vaccines. The mRNA construct encoded multiple epitopes from different antigens. A potential application for this platform is antigen-specific immunotherapy for type I diabetes. The two delivery routes could be used to target different lymphoid tissues. After intravenous injection, the mRNA delivered by DCs and LNPs was localised mainly in the lungs and spleen, respectively. After local intradermal administration, both delivery routes resulted in mRNA expression at the injection site, as well as robust T-cell responses in draining lymph nodes.

Cancer

Based on the knowledge that T-cells recognising neoantigens are present in most cancers, Cafri et al. [10] developed an mRNA vaccine for cancer treatment. The vaccine named mRNA-4650 was composed of an mRNA backbone that encoded up to 20 different neoantigens expressed by the autologous cancer. Mutations generated from *in silico* prediction were included in the mRNA construct. The vaccine sequences were electronically submitted to Moderna Therapeutics for manufacturing, and used to treat four patients with metastatic gastrointestinal cancer. The mRNA vaccine for each patient encoded different predicted neoantigens based on sequencing data. The T-cell analysis from patients showed increased mutation-specific responses against predicted neoantigens that were not detected before the vaccination. The vaccine was shown to be safe, demonstrating potential for future personalised treatment of patients with common epithelial cancers.

Epitope predictions for mRNA vaccine design

Epitopes are antigenic determinants that can be recognised by the immune system. T-cell epitopes are on the surface of an APC and bound to major histocompatibility, whereas B-cell epitopes are bound by antibodies or immunoglobulin. The previous section introduces several examples of mRNA vaccines that encoded epitopes [21, 49, 52, 67], as well as others that used the properties of epitopes in vaccine design [12, 29, 66]. Nevertheless, epitope-based mRNA vaccine design is currently not a popular approach, and still holds much unexplored potential for increasing vaccine safety, immunogenicity and cross-reactivity. As Correia et al. [16] showed, epitope-focused vaccine design can be used to tackle highly antigenically variable viruses such as HIV. Although mRNA vaccines have the capability to encode any gene of interest, common practice in even the most recent designs is to encode sequences of original genes from the natural virus. Epitope prediction can play an important role in assisting mRNA vaccine design, specifically by guiding sequence design and vaccine structure. Several models currently exist for predicting vaccine efficacy, but some of these do not work for mRNA vaccines, and none is optimised for mRNA vaccines [6, 33]. This section describes the models that can be used for epitope prediction for mRNA or DNA vaccine design, together with discussions on the parameters and targets of the predictions.

As proposed by Kanekiyo et al. [31], a primary challenge in antigen design is to elicit broadly protective responses against antigenically variable viruses such as influenza. It is important to overcome the presence of less conserved, distracting epitopes. Epitope-focused vaccine design, an extreme variation of the subdomain approach, can be used to solve this problem. The researchers proposed that computational protein design can be used to incorporate epitopes into unrelated scaffold proteins with high structural fidelity.

A few of the current mRNA vaccine designs have utilised epitope prediction, with the majority focusing on T-cell epitopes. Epitope prediction allows researchers to find the best combination of sequences to put into ORFs, thereby enabling guided and rational vaccine design. Typically, publicly available trained computational models are used for T-cell epitope prediction, since this is more efficient than developing new models. However, for B-cell epitopes, the available models cannot achieve satisfactory accuracy particularly for predicting conformational B-cell epitopes. A model that can solve this problem would allow more researchers to apply B-cell epitope prediction in mRNA vaccine design to induce better humoral responses.

Epitope prediction methods

Epitope prediction methods can be classified into sequence-based and structure-based methods [59]. For sequence-based methods, an outdated but still most commonly used idea is motif search, where the neural network provides a convenient process for finding relationships and describing non-linear data. The support vector machine is another widely used model type in epitope prediction, exemplified by famous models such as COBEPRO (linear B-cell epitope prediction model [77] and Pcleavage (cleavage sites prediction model [5]). Other methods include hidden markov chain and quantitative matrices. For structure-based models, common computational methods include docking of peptides, knowledge-based threading algorithms, and binding energy and molecular dynamics [59].

Epitope prediction is needed in T-cell mediated clinical trials to enable the selection of peptides for cancer vaccination therapy or immune-monitoring of tumour-specific T-cells. Hu et al.[28] predicted potential epitopes binding to HLA class I molecules in Caucasian and East Asian populations with the NetMHCpan4.0 algorithm, using an 8-11 mer epitope length. In epitope prediction of tumour neoantigens, the average epitope numbers of missense mutations were different between the two populations. From these findings, the researchers proposed that they could reduce costs and improve efficiency in clinical immunotherapy.

The current models for B-cell epitope prediction are far from perfect. The first issue is the prevailing binary classification paradigm, which mandates the problematic dichotomisation of continuous outcome variables [11]. Second, these models cannot explicitly simulate the biological consequences of immunisation that are highly relevant to vaccine safety and efficacy. A much broader and deeper systematic view of immunology is required for a thorough understanding of B-cell epitope prediction. This

Table 3. Epitope prediction methods for vaccine design (NA = data not given)

Author/name	Algorithms	AUC	Type
Haste Andersen et al. [26]/Discotope	Linear combination of hydrophilicity scale and epitope log-odds ratios	0.567	B-cell
Sweredoski and Baldi [76]/Bepro(PEPITO)	Linear combination of epitopic residue propensity and half sphere exposure values	0.570	B-cell
Liang et al. [42]/ElliPro	Residue protrusion index	0.585	B-cell
Sun et al. [74]/SEPPA	Linear combination of epitopic residue propensity and contactness of neighbouring residues	0.576	B-cell
Rubinstein et al. [68]/EPITOPIA	Naive Bayesian classifier	0.579	B-cell
Liang et al. [40]/EPCES	Linear method with voting mechanism	0.586	B-cell
Liang et al. [41]/EPSVR	SVR	0.597	B-cell
Yao et al. [90]/Bpredictor	Random forest classifier	0.598	B-cell
Liang et al. [41]/EPMeta	Combination of EPSVR,ref40,ref68,ref74, PEPITO and Discotope	0.638	B-cell
Sela-Culang et al. [69]/PEASE	Random forest	NA	B-cell
Solihah et al. [72]/CluSMOTE	Support Vector Machine and Decision Tree	0.766	B-cell
Dalkas and Rooman [18]/SEPIa	Random forest and Gaussian Naive Bayes	0.65	B-cell
Jespersen et al. [30]/BepiPred-2.0	Random forest	0.62	B-cell
Reynisson et al. [65]/NetMHCpan-4.1	Machine learning	0.994	T-cell
O'Donnell et al. [53]/MHCflurry-2.0	Neural network	0.992	T-cell
Moutaftsi et al. [50]/IEDB Consensus	Scoring-matrix	0.988	T-cell
Kim et al. [32]/SMMPMBEC	SMM with amino acid similarity matrix	0.978	T-cell
Peters and Sette [60]/SMM	Stabilised matrix	0.977	T-cell
Bui et al. [8]/ARB	Average relative binding coefficient matrix	0.962	T-cell
Reche et al. [64]/Rankpep	Position specific scoring matrix ranking	0.903	T-cell
Parker et al. [58]/BIMAS	Table of coefficients	0.942	T-cell
Stojanovic [73]/ MHCLovac	Physicochemical properties modeling	0.628	T-cell
Rammensee et al. [63] / STYPEITHI	Allele-specific peptide motifs	0.983	T-cell
Altuvia et al. [4]/PREDEP	Peptide ranking with template peptide	0.844	T-cell
Singh and Raghava [71]/ProPred1	Matrix-based prediction	0.869	T-cell
Liu et al. [43]PAComplex	Template-based scoring	0.902	T-cell

could enable better comprehension of a range of crucial aspects in vaccine design that are currently neglected [11].

Yao et al. [90] provided an overview of current algorithms and models for conformational B-cell epitope prediction, and a comparison of their performance with common binding site prediction methods. These predictors use conservation scores, structural features, geometric characteristics and amino acid features, integrated by a linear combination of machine learning algorithms. The performance of these methods is measured by area under the receiver operating characteristic curve (AUC-ROC). The current level of accuracy for all of these predictors is not yet sufficient, with the highest accuracy being 25.6% by EPMeta.

Yao et al. [90] also mentioned that researchers frequently apply protein binding site prediction methods in conformational epitope prediction, since epitopic patches may be considered as protein binding sites. The major difference between the two approaches lies in the training data. The protein binding site prediction method uses all known protein–protein binding complexes for training, whereas epitope prediction uses only antibody-antigen complexes. However, general protein binding site prediction methods achieve significantly lower performance than all conformational epitope prediction methods mentioned above. This is because protein binding site prediction methods are designed based on the conservation and hydrophobicity of binding patches. B-cell epitopic patches are neither conserved nor more hydrophobic compared with other protein–protein surfaces.

Table 3 provides a list of epitope prediction models. B-cell epitope prediction models still display low accuracy compared to T-cell models, due to the complicated spatial folding of B-cell epitopes [7]. As more data become available, many models such as NetMHCpan have turned to a machine learning approach for epitope prediction, which can exploit a large amount of data in a sophisticated manner to achieve improved performance.

Examples of epitope prediction in vaccine design

Lucas Michel-Todó et al. [49] designed an epitope-based vaccine for *Trypanosoma cruzi*, a protozoan parasite that causes Chagas disease. T-cell epitopes were predicted on the $H > 0.5$ *T. cruzi*-masked proteome, using IEDB MHC-I binding prediction algorithms such as artificial neural network. These were narrowed down to 18 peptides that had identities < 70% to any human or human microbiome proteins, which were analysed by calculating their projected protection coverage (PPC). Individually, all of the 18 peptides had a PPC > 10%, while together they provided a PPC of 88.3%. For B-cell epitopes, both structure-based and sequence-based approaches for prediction were applied. A potential B-cell epitope in the KMP11 protein of the parasite was identified, and residues with relative solvent accessibilty > 50% were considered good candidates as part of a potential epitope. The sequence-based approach identified 10 potential B-cell epitopes with <70% identity to human proteins. In the final design, 30 epitopes were included in the vaccine ensemble, comprising 18 *CD8*+ T-cell epitopes, two selected *CD4*+ T-cell epitopes and 10 B-cell epitopes.

An epitope-based DNA vaccine for a bacterial pathogen was designed by Gregory et al. [23], which involved prediction and validation of the HLA class II-restricted epitopes. They adapted the methods [48] using EpiMatrix and Epivax to score the peptides screened from two sets of bacterial proteins, one set being putative secreted proteins and the other set being sourced from the literature. A final selection of 14 epitopes were incorporated into the vaccine, among which six were 100% conserved and four were partially conserved. After vaccination, mice that were challenged with a lethal dose of the pathogen showed an increased rate of survival.

mRNA vaccine design using software and computational tools

A range of computational tools that could assist in vaccine design include epitope prediction models, immunogenicity/antigenicity prediction models, protein/gene databases, epitope identification models, and others. This section outlines the tools that have been used or could be helpful in mRNA vaccine design.

Overview of computational tools in vaccine development

Sunita et al. [75] have provided an introduction to modern vaccine development that could benefit from computational tools. In systems biology and structural antigen design, research has demonstrated successful biomarker prediction using systematic simulation-based meta-analytical frameworks. Databases such as Vaxjo can help researchers find potential adjuvants for a vaccine. In rational vaccine design, an online software known as VaxiJen has been designed for antigen prediction, which is widely recognised and has been referenced in various studies. Other bioinformatics methods applied in vaccine development include structural approaches, molecular dynamics simulations and docking. However, computational tools have not been widely adopted in mRNA vaccine design due to unsatisfactory accuracy of many prediction models. The use of computational tools may also require a steep learning curve for researchers, since many tools require a certain level of understanding in programming or algorithms to be used efficiently. Nevertheless, as more data becomes publicly available, new technologies such as machine learning are beginning to improve the use of computational tools for vaccine design. During the design process, computational tools can offer substantial assistance that is otherwise not available to researchers, such as in predicting epitopes, optimising sequence design, and profiling the target population. By reducing uncertainty during the design phase, computational tools can help increase the efficacy of vaccines. A summary of current computational tools that can be used in vaccine design are provided in Table 4.

Sequence optimisation

The main disadvantages of mRNA vaccines, such as their low stability and protein translation efficiency, could be addressed by improving the sequence design of the vaccine. To achieve better efficiency for an mRNA vaccine, Zhang et al. [93] proposed an algorithm known as LinearDesign to optimise the mRNA sequence design. The basis of the algorithm is the intersection between a Stochastic Context Free Grammar (SCFG) and a Deterministic Finite Automaton (DFA). The SCFG represents the folding free energy model and the DFA represents the set

Table 4. Computational tools for vaccine design

Author	Application	Software/ algorithm
Zhang et al. [93]	Sequence optimisation	LinearDesign
Multiple works	Epitope prediction	Table 3
Chaudhury et al. [13]	Adjuvant selection	Machine learning
Lee et al. [38]	Immunogenicity prediction	DAMIP
Xu et al. [88]	Profiling	VirScan
Rahman et al. [62]	Rational vaccine design	UGENE

of all possible synonymous mRNA sequences that code a given protein. The mRNA design problem was formulated as follows:

Given a protein sequence $p = p_1 \ldots p_m$ where each p_i is an amino acid, search the best mRNA sequence $r * (p)$ among all possible mRNA sequences r that translate into protein p, defined as the sequence that has the structure with minimum folding free energy charge argmin of $MFE(r)$. This is defined as the minimum free energy charge of the structure for an RNA sequence r according to an energy model (see formulas (1),(2) below).

$$r^*(p) = \mathrm{argmin}_{r \in \mathrm{RNA}(p)} MFE(r) \qquad (1)$$

$$MFE(r) = \min_{S \in \mathrm{structures}(r)} \Delta G^\circ(r, s) \qquad (2)$$

The implementation of the algorithm uses dynamic programming on the Turner nearest neighbour free energy model. It achieves the complexity of $O(b^{2}*n)$ where b is the number of results kept at each step. The algorithm also reduces the redundancies and secondary structure at the 5' end leader region. When tested on the coding mRNA for the S protein of SARS-CoV-2, this algorithm was shown to work as intended. Although shown to be effective in optimising sequence design for mRNA vaccines, the algorithm needs to be tested in real applications of vaccine design. Nevertheless, this work provided insightful information on methods that could be used to manipulate mRNA codons into a more optimal sequence.

Epitope prediction and SARS-CoV-2

Immunogenicity maps can be used to inform multiple modalities of vaccine development. Yarmarkovich et al. [91] presented such a map with peptide sequences that were expected to be safe as well as immunogenic for T-cell-based vaccination. The conserved regions of SARS-CoV-2 were identified through comparison with closely related alpha and beta coronaviruses. The antigens were chosen based on (i) higher predicted safety for autoimmunity, and (ii) higher immunogenicity of dissimilar peptides.

B-cell epitopes were also assessed, including conformational B-cell epitopes in the S protein of SARS-CoV-2. The tools used were BepiPred and DiscoTope2.0, both of which required structural data. A 33-mer peptide was derived from the S protein, based on linear and conformational B-cell epitope scoring. The accuracy of prediction was estimated by comparing the generated 33-mer peptides with epitopes derived from IEDB. A highest ranked peptide sequence was generated, containing five acquired residues which increased S protein binding to ACE2.

The researchers proposed the use of multivalent constructs in the mRNA vaccine composed of the SARS-CoV-2 minigenes encoding subsets of the B- and/or T-cell epitopes for expression in APCs.

Adjuvant selection

An adjuvant is a chemical compound used to increase vaccine efficacy for vaccine platforms such as mRNA vaccines and subunit vaccines. mRNA vaccines usually require additional measures such as adjuvants to achieve optimal efficacy, since the mRNA sequence alone cannot successfully induce immune responses under most circumstances. Chaudhury et al. [13] used machine learning to identify adjuvant-specific immune response characteristics that could guide rational adjuvant selection. They profiled human immune responses induced by vaccines with two similar, clinically relevant adjuvants, AS01B and AS02A. The researchers established adjuvant-mediated immune signatures by generating a broad immunoprofile. They then integrated the immunoprofiling data and identified the combination of immune features that distinguished vaccine-induced responses by the adjuvants using machine learning. The combination of immune features identified by computational analysis could distinguish subjects by adjuvant with 71% accuracy.

Immunogenicity prediction

Immunogenicity prediction is another aspect of vaccine design that involves heavy use of computational methods. Lee et al. [38] proposed a novel machine learning framework that predicted vaccine and antibody responses by uncovering gene signatures. This framework integrated a combinatorial feature-selection algorithm and an optimisation-based classification model known as discriminant analysis through mixed-integer programming (DAMIP). The training data on gene expression related to immunological responses, cell motility and biopolymer metabolism were collected using high-throughput technologies from an experiment. The prediction accuracy was found to be at least 80% among eight DAMIP rules, with some reaching blind-prediction rates of 90%. Similar tests were performed on influenza vaccines and resulted in 85% accuracy.

Profiling

Vaccine design could benefit greatly from information on the viruses or viral strains to which certain populations have been exposed. Important information on viral exposure could be provided by serological profiling. Xu et al. [88] designed a computational method, VirScan, to identify the set of viruses to which an individual has been exposed. This method identified viruses by setting threshold numbers empirically and tallying the number of enriched peptides from each virus. VirScan could achieve very high sensitivities and specificities of 95% or higher in measurements using serum samples from patients infected with HIV and hepatitis C. It could also be used on blood samples to detect viruses that do not cause viremia. Such tools could be extremely useful in vaccine design, particularly for specific populations.

Rational vaccine design

Rational vaccine design is a design strategy that seeks out and improves suboptimal approaches in current vaccine design, which is applicable to mRNA vaccines Pollard et al. [61]. mRNA vaccines have long suffered from degradation by RNases within a short amount of time, as well as eliciting only moderate DC activation. A two-component mRNA vaccine was recently proposed, consisting of protamine complexed mRNA mixed with naked mRNA, which when injected intradermally induced both prophylactic and therapeutic antitumour responses [27]. Current mRNA vaccines contain structural modifications such as anti-reverse cap analogs and elongated poly(A) tails. Replacement of uridine and cytidine with 2-thiouridine and 5-methyl-cytidine could increase transfection efficiency in human and murine epithelial cells [35]. Type-I IFNs are important molecules in antiviral host defense mechanisms. However, they may interfere with the expression of mRNA vaccines and should be considered in rational vaccine design.

An example of rational vaccine design with epitope prediction is about selecting predicted conserved epitopes. Rahman et al.[62] used UGENE (an application that provides integrated bioinformatics tools) to identify the conserved region of ZIKV. They used three computational models to predict conserved regions, and common peptides in all three methods were chosen as candidate epitopes. One epitope was found in the E protein, and another two in the NS5 protein region. All three candidate epitopes were found to be 100% conserved amid all 305 sequences from the ZIKV.

Many computational tools, such as epitope prediction and sequence optimisation, still only manage mediocre performance levels when assisting mRNA vaccine design. This is mainly due to limited technological advances, but in recent years vast online data have become available and new models have been designed to use demanding hardware resources such as Tensorflow. Machine learning algorithms are capable of epitope prediction. Many current epitope prediction methods have integrated certain machine learning algorithms, but their performance still lacks accuracy. The same is true for immunogenicity prediction. The current pandemic caused by SARS-CoV-2 continues to call for more advanced computational methods to assist in vaccine design.

Summary and future perspectives of mRNA vaccine design

Compared to other vaccine types, mRNA vaccines offer a distinct advantage of having a quick and easy design and production process. This is particularly advantageous for unstable pathogens such as RNA viruses. The design process of mRNA vaccines offers significant versatility, since the same platform can be used for different antigens, and cost-effectiveness due to the reduction in both vaccine development time and the need to commit significant financial resources.

Thermostability has been a persisting issue for mRNA vaccines, including for the two most advanced mRNA vaccine candidates for SARS-CoV-2, both of which require storage and delivery temperatures in the cold-chain environment at -70°C. This greatly limits the availability of mRNA vaccines in rural areas and imposes additional costs. The current implementation of thermostable mRNA vaccines is to apply a freeze-dry protocol, which can maintain a part of the vaccine efficacy and requires recovery before use. The optimisation of vaccine formulation may help to improve the thermostability of mRNA vaccines [94].

In terms of their safety profile, some argue that mRNA vaccines have potential risks to humans that do not show in animal experiments. However, the same concern can be voiced for other types vaccines. Although not yet widely implemented, mRNA

vaccines so far show a relatively good safety profile compared to traditional vaccines.

For the SARS-CoV-2 mRNA vaccines, the two current top competitors both used the same delivery system and immunogen design. Although this design elicits an immune response similar to a natural viral infection, and has shown high efficacy in clinical trials, it still has room for optimisation to present epitopes in a more effective and stable manner. First, the pre-fusion conformation might require additional conditions to be maintained after entering the body, since the macromolecules could be changed by environmental factors. Second, another immunogen such as the N protein could be encoded to provide an alternative target for immune responses. The addition of a different immunogen could help increase the vaccine efficacy and reduce the possibility of mutation escape. Third, although using the original conformation of the S protein in the vaccine could mimic a natural viral infection, the T-cells and neutralising antibodies only bind to certain peptides rather than the entire protein. Encoding only the necessary epitopes could free more space and increase the stability of the mRNA vaccine. More-over, if the SARS-CoV-2 underwent a dramatic mutation in its S protein that could evade the immunity provided by current vaccines, it would be easier to re-adjust the epitopes to adapt to the new mutation rather than designing and encoding a new S protein conformation. Epitope prediction becomes a useful tool in this approach, since it provides a preliminary view of the epitopes that would be generated by the designed sequence *in vivo*, and therefore, the overall vaccine performance. In recent years, methods for the computational prediction of epitopes have evolved dramatically, from the primitive motif-search technique to powerful models such as neural networks. An important advantage of using computational models in vaccine design is their flexibility. For instance in the case of machine learning models, adapting to new purposes or data by adjusting feature space or weights only requires minor coding modifications.

The majority of epitope prediction methods discussed in this review are for T-cell epitopes, which cannot be applied to evaluating humoral responses. For B-cell epitopes, the limited computational prediction methods available only employ linear B-cell epitopes, since methods for conformational B-cell epitopes still struggle to generate satisfactory results in real world applications. Training an accurate conformational B-cell epitope prediction model could be the next step in improving epitope-based vaccine design.

Overall in this review, we have discussed the characteristics of mRNA vaccines and compared these to other types of vaccines. mRNA vaccines generally outperform the majority of other vaccine types in their safety profile, as well as simplicity and efficiency of the design and production process. Prior to their application in the battle against COVID-19, the use of mRNA vaccines had been explored in the prevention of a range of other infectious and non-infectious diseases, including influenza and cancer. Recent developments in bioinformatics have enabled a range of computational tools to be developed or adapted for mRNA vaccine design. Among these, epitope prediction models based on machine learning hold great potential for improving mRNA vaccine design. Applying epitope-based design processes as a part of rational vaccine design for mRNA vaccines could greatly increase vaccine immunogenicity and efficacy for practical applications. The emergence of COVID-19 is perhaps one of the most important moments in the history of mRNA vaccines, which are now being administered massively around the world and have produced excellent results with high levels of immunity. As the world becomes more familiar with the efficacy,

design and characteristics of mRNA vaccines, their future in human disease prevention looks bright.

Key Points

- mRNA vaccines have been studied for decades and applied to prevent infectious diseases and treat cancer.
- COVID-19 mRNA vaccines are globally used at this moment to combat the spread of SARS-CoV-2.
- The current design of mRNA vaccines can be optimised through computational methods.
- Epitope-based vaccine design is a promising method to boost the efficacy of mRNA vaccines.
- Compared to other types of vaccines, mRNA vaccines have good overall performance with balanced efficacy, safety, flexibility and speed of development.

Authors' contributions statement

X.C. conducted the survey and wrote the text. J.L. provided supervision and guidance on the content and organisation of the work. JJ.L. revised the text and structure of the work. B.O. and T.L. advised on the content of the work .

Acknowledgments

Competing interests

The authors declare no competing interests.

References

1. Iedb.org: Free epitope database and prediction resource. URL https://www.iedb.org/.
2. How does testing in the u.s. compare to other countries? URL https://coronavirus.jhu.edu/testing/international-comparison.
3. Ahammad I, Lira SS. Designing a novel mrna vaccine against sars-cov-2: An immunoinformatics approach. *International Journal of Biological Macromolecules* 2020. ISSN 0141-8130; **162**:820–37. doi: 10.1016/j.ijbiomac.2020.06.213. http://www.sciencedirect.com/science/article/pii/S0141813020336655.
4. Altuvia Y, Schueler O, Margalit H. Ranking potential binding peptides to mhc molecules by a computational threading approach. *Journal of Molecular Biology* 1995; **249**(2): 244–50. doi: 10.1006/jmbi.1995.0293. https://www.sciencedirect.com/science/article/pii/S0022283685702938.
5. Bhasin M, Raghava GPS. Pcleavage: an svm based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. *Nucleic Acids Research* 2005; **33**:W202–7. doi: 10.1093/nar/gki587.
6. Bonomo ME, Deem MW. Predicting Influenza H3N2 Vaccine Efficacy From Evolution of the Dominant Epitope. *Clinical Infectious Diseases* 2018 ISSN 1058-4838; **67**(7): 1129–31. doi: 10.1093/cid/ciy323.
7. Hans Rudolf Bosshard. 11 - epitope mapping with peptides. In Bernd Gutte, editor, *Peptides*, pages 419–54. Academic

Press, San Diego, 1995. ISBN 978-0-12-310920-0. doi: 10.1016/B978-012310920-0/50012-7.

8. Bui H-H, Sidney J, Peters B, *et al*. Automated generation and evaluation of specific mhc binding predictive tools: Arb matrix applications. *Immunogenetics* 2005; **57**:304–14. doi: 10.1007/s00251-005-0798-y.

9. Burton DR. Laura M. Walker. Rational vaccine design in the time of covid-19. *Cell Host and Microbe* may 2020; **27**(5). doi: 10.1016/j.chom.2020.04.022.

10. Cafri G, Gartner JJ, Zaks T, *et al*. mrna vaccine-induced neoantigen-specific t cell immunity in patients with gastrointestinal cancer. *The Journal of Clinical Investigation* 2020; **130**. doi: 10.1172/JCI134915.

11. Caoili SEC. B-cell epitope prediction for peptide-based vaccine design: Towards a paradigm of biological outcomes. *Association for Computing Machinery, page* 2010;602–10. doi: 10.1145/1854776.1854889.

12. Chahal JS, Fang T, Woodham AW, *et al*. An rna nanoparticle vaccine against zika virus elicits antibody and cd8+ t cell responses in a mouse model. *Scientific Reports* 2017; **7**(03). doi: 10.1038/s41598-017-00193-w.

13. Chaudhury S, Duncan EH, Atre T, *et al*. Combining immunoprofiling with machine learning to assess the effects of adjuvant formulation on human vaccine-induced immunity. *Human Vaccines and Immunotherapeutics* 2019; **16**:400–11. doi: 10.1080/21645515.2019.1654807.

14. Corbett KS, Edwards DK, Leist SR, *et al*. Sars-cov-2 mrna vaccine design enabled by prototype pathogen preparedness. *Nature* 2020a. doi: 10.1038/s41586-020-2622-0.

15. Corbett KS, Flynn B, Foulds KE, *et al*. Evaluation of the mrna-1273 vaccine against sars-cov-2 in nonhuman primates. *New England Journal of Medicine* 2020b. doi: 10.1056/nejmoa2024671.

16. Correia BE, Bates JT, Loomis RJ, *et al*. Proof of principle for epitope-focused vaccine design. *Nature* 2014; **507**:201–6. doi: 10.1038/nature12966. https://www.ncbi.nlm.nih.gov/pubmed/24499818.

17. Dagan N, Barda N, Kepten E, *et al*. Bnt162b2 mrna covid-19 vaccine in a nationwide mass vaccination setting. *The New England Journal of Medicine* 2021. doi: 10.1056/NEJMoa2101765. https://pubmed.ncbi.nlm.nih.gov/33626250/.

18. Dalkas GA, Rooman M. Sepia, a knowledge-driven algorithm for predicting conformational b-cell epitopes from the amino acid sequence. *BMC Bioinformatics* 2017; **18**(02). doi: 10.1186/s12859-017-1528-9.

19. Espeseth AS, Cejas PJ, Citron MP, *et al*. Modified mrna/lipid nanoparticle-based vaccines expressing respiratory syncytial virus f protein variants are immunogenic and protective in rodent models of rsv infection. *npj Vaccines* 2020; **5**(02). doi: 10.1038/s41541-020-0163-z.

20. Feldman RA, Fuhr R, Smolenov I, *et al*. mrna vaccines against h10n8 and h7n9 influenza viruses of pandemic potential are immunogenic and well tolerated in healthy adults in phase 1 randomized clinical trials. *Vaccine* 2019; **37**(25): 3326–34. doi: 10.1016/j.vaccine.2019.04.074. http://www.sciencedirect.com/science/article/pii/S0264410X19305626.

21. Firdessa-Fite R, Creusot RJ. Nanoparticles versus dendritic cells as vehicles to deliver mrna encoding multiple epitopes for immunotherapy. *Molecular Therapy - Methods & Clinical Development* 2020. ISSN 2329-0501; **16**:50–62. doi: 10.1016/j.omtm.2019.10.015. http://www.sciencedirect.com/science/article/pii/S2329050119301238.

22. Gao Q, Bao L, Mao H, *et al*. Development of an inactivated vaccine candidate for sars-cov-2. *Science* 2020. doi: 10.1126/science.abc1932.

23. Gregory SH, Mott S, Phung J, *et al*. Epitope-based vaccination against pneumonic tularemia. *Vaccine* 2009. ISSN 0264-410X; **27**(39): 5299–306. doi: 10.1016/j.vaccine.2009.06.101. http://www.sciencedirect.com/science/article/pii/S0264410X09009839.

24. Guirakhoo F, Kuo L, Peng J, *et al*. A novel sars-cov-2 multitope protein/peptide vaccine candidate is highly immunogenic and prevents lung infection in an adeno associated virus human angiotensin-converting enzyme 2 (aav hace2) mouse model. *bioRxiv* 2020. doi: 10.1101/2020.11.30.399154.

25. Haabeth OAW, Blake TR, McKinlay CJ, *et al*. mrna vaccination with charge-altering releasable transporters elicits human t cell responses and cures established tumors in mice. *Proceedings of the National Academy of Sciences* 2018. ISSN 0027-8424; **115**(39): E9153–61. doi: 10.1073/pnas.1810002115. https://www.pnas.org/content/115/39/E9153.

26. Andersen PH, Nielsen M, Lund O. Prediction of residues in discontinuous b-cell epitopes using protein 3d structures. *Protein Science: A Publication of the Protein Society* 2006; **15**:2558–67. doi: 10.1110/ps.062405906. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2242418/\#:∼:text-Prediction\%20of\%20residues\%20in\%20discontinuous\%20B-cell\%20epitopes\%20using

27. Hess PR, Boczkowski D, Nair SK, *et al*. Vaccination with mrnas encoding tumor-associated antigens and granulocyte-macrophage colony-stimulating factor efficiently primes ctl responses, but is insufficient to overcome tolerance to a model tumor/self antigen. *Cancer Immunology, Immunotherapy* 2005; **55**:672–83. doi: 10.1007/s00262-005-0064-z.

28. Hu W, He M, Li L. Hla class i restricted epitopes prediction of common tumor antigens in white and east asian populations: Implication on antigen selection for cancer vaccine design. *PLOS ONE* 2020; **15**:e0229327. doi: 10.1371/journal.pone.0229327.

29. Hu Z, Jiao X, Liu X. Antibody immunity induced by h7n9 avian influenza vaccines: Evaluation criteria, affecting factors, and implications for rational vaccine design. *Frontiers in Microbiology* 2017. ISSN 1664-302X; **8**(1898). doi: 10.3389/fmicb.2017.01898. https://www.frontiersin.org/article/10.3389/fmicb.2017.01898.

30. Jespersen MC, Peters B, Nielsen M, Marcatili P. Bepipred-2.0: improving sequence-based b-cell epitope prediction using conformational epitopes. *Nucleic Acids Research* 2017; **45**:W24–9. doi: 10.1093/nar/gkx346.

31. Kanekiyo M, Ellis D, King NP. New vaccine design and delivery technologies. *The Journal of Infectious Diseases* 2019. ISSN 0022-1899; **219**:S88–96. doi: 10.1093/infdis/jiy745. https://academic.oup.com/jid/article-pdf/219/Supplement_1/S88/28263989/jiy745.pdf.

32. Kim Y, Sidney J, Pinilla C, *et al*. Derivation of an amino acid similarity matrix for peptide:mhc binding and its application as a bayesian prior. *BMC Bioinformatics* 2009; **10**:394. doi: 10.1186/1471-2105-10-394.

33. Kohberger RC, Jemiolo D, Noriega F. Prediction of pertussis vaccine efficacy using a correlates of protection model. *Vaccine* 2008. ISSN 0264-410X; **26**(27): 3516–21. doi: 10.1016/j.vaccine.2008.04.016. https://www.sciencedirect.com/science/article/pii/S0264410X08004532.

34. Kormann MSD, Hasenpusch G, Aneja MK, *et al*. Expression of therapeutic proteins after delivery of chemically modified mrna in mice. *Nature Biotechnology* 2011a; **29**:154–7. doi: 10.1038/nbt.1733. https://pubmed.ncbi.nlm.nih.gov/21217696/.

35. Kormann MSD, Hasenpusch G, Aneja MK, *et al*. Expression of therapeutic proteins after delivery of chemically

modified mrna in mice. *Nature Biotechnology* 2011b; **29**:154–7. doi: 10.1038/nbt.1733. https://pubmed.ncbi.nlm.nih.gov/21217696/.

36. Krammer F. Sars-cov-2 vaccines in development. *Nature* 2020. doi: 10.1038/s41586-020-2798-3.

37. Lagunas-Rangel FA, Viveros-Sandoval ME, Reyes-Sandoval A. Current trends in zika vaccine development. *Journal of Virus Eradication* 2017; **3**:124–7. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5518240/.

38. Lee EK, Nakaya HI, Yuan F, *et al*. Machine learning for predicting vaccine immunogenicity. *Interfaces* 2016; **46**:368–90. doi: 10.1287/inte.2016.0862.

39. Frank Liang, Gustaf Lindgren, Ang Lin, *et al*. Efficient targeting and activation of antigen-presenting cells in vivo after modified mrna vaccine administration in rhesus macaques. *Molecular Therapy*, **25**(12): 2635–47, 2017. ISSN 1525–0016. doi: https://doi.org/10.1016/j.ymthe.2017.08.006. URL http://www.sciencedirect.com/science/article/pii/S1525001617303659.

40. Liang S, Zheng D, Zhang C, Zacharias M. Prediction of antigenic epitopes on protein surfaces by consensus scoring. *BMC bioinformatics* 2009; **10**:302. doi: 10.1186/1471-2105-10-302. https://www.ncbi.nlm.nih.gov/pubmed/19772615/.

41. Liang S, Zheng D, Standley DM, *et al*. Epsvr and epmeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics* 2010a; **11**:381. doi: 10.1186/1471-2105-11-381. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2910724/.

42. Liang S, Zheng D, Standley DM, *et al*. Epsvr and epmeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics* 2010b; **11**:381. doi: 10.1186/1471-2105-11-381. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2910724/.

43. Liu I-H, Lo Y-S, Yang J-M. Pacomplex: a web server to infer peptide antigen families and binding models from tcr-pmhc complexes. *Nucleic Acids Research* 2011 ISSN 0305-1048; **39**:W254–60. doi: 10.1093/nar/gkr434.

44. Liu MA. A comparison of plasmid dna and mrna as vaccine technologies. *Vaccines* 2019; **7**:37. doi: 10.3390/vaccines7020037.

45. Lu J, Lu G, Tan S, *et al*. A covid-19 mrna vaccine encoding sars-cov-2 virus-like particles induces a strong antiviral-like immune response in mice. *Cell Research* 2020. doi: 10.1038/s41422-020-00392-7.

46. MARK CATHERINE, RIGAU-PÉREZ JOSÉG. The world's first immunization campaign: The spanish smallpox vaccine expedition, 1803–1813. *Bulletin of the History of Medicine* 2009; **83**:63–94. https://www.jstor.org/stable/44448715?seq=1.

47. Maruggi G, Zhang C, Li J, *et al*. mrna as a transformative technology for vaccine development to control infectious diseases. *Molecular Therapy* 2019. ISSN 1525-0016; **27**(4): 757–72. doi: 10.1016/j.ymthe.2019.01.020. http://www.sciencedirect.com/science/article/pii/S1525001619300413.

48. McMurry JA, Gregory SH, Moise L, *et al*. Diversity of francisella tularensis schu4 antigens recognized by t lymphocytes after natural infections in humans: Identification of candidate epitopes for inclusion in a rationally designed tularemia vaccine. *Vaccine* 2007; **25**:3179–91. doi: 10.1016/j.vaccine.2007.01.039.

49. Michel-Todó L, Reche PA, Bigey P, *et al*. In silico design of an epitope-based vaccine ensemble for chagas disease. *Frontiers in Immunology* 2019; **10**(11). doi: 10.3389/fimmu.2019.02698. https://www.frontiersin.org/articles/10.3389/fimmu.2019.02698/.

50. Moutaftsi M, Peters B, Pasquetto V, *et al*. A consensus epitope prediction approach identifies the breadth of murine t(cd8+)-cell responses to vaccinia virus. *Nature biotechnology* 2006; **24**:817–9. doi: 10.1038/nbt1215. https://www.ncbi.nlm.nih.gov/pubmed/16767078.

51. Mulligan MJ, Lyke KE, Kitchin N, *et al*. Phase 1/2 study of covid-19 rna vaccine bnt162b1 in adults. *Nature* 2020;1–8. doi: 10.1038/s41586-020-2639-4.

52. Nielsen JS, Wick DA, Tran E, *et al*. An in vitro-transcribed-mrna polyepitope construct encoding 32 distinct hla class i-restricted epitopes from cmv, ebv, and influenza for use as a functional control in human immune monitoring studies. *Journal of Immunological Methods* 2010. ISSN 0022-1759; **360**(1): 149–56. doi: 10.1016/j.jim.2010.07.003. http://www.sciencedirect.com/science/article/pii/S0022175910001997.

53. O'Donnell TJ, Rubinsteyn A, Laserson U. Mhcflurry 2.0: Improved pan-allele prediction of mhc class i-presented peptides by incorporating antigen processing. *Cell Systems* 2020. ISSN 2405-4712; **11**(1): 42–48.e7. doi: 10.1016/j.cels.2020.06.010. https://www.sciencedirect.com/science/article/pii/S2405471220302398.

54. Ong E, Wong MU, Huffman A, He Y. Covid-19 coronavirus vaccine design using reverse vaccinology and machine learning. *Frontiers in Immunology* 2020; **11**(07). doi: 10.3389/fimmu.2020.01581.

55. Pandey SC, Pande V, Sati D, *et al*. Vaccination strategies to combat novel corona virus sars-cov-2. *Life Sciences* 2020 ISSN 0024-3205; **256**:117956. doi: 10.1016/j.lfs.2020.117956.

56. Pardi N, Hogan MJ, Pelc RS, *et al*. Zika virus protection by a single low-dose nucleoside-modified mrna vaccination. *Nature* 2017; **543**:248–51. doi: 10.1038/nature21428.

57. Pardi N, Hogan MJ, Porter FW, Weissman D. mrna vaccines-a new era in vaccinology. *Nature Reviews Drug Discovery* 2018; **17**:261–79. doi: 10.1038/nrd.2017.243. https://www.nature.com/articles/nrd.2017.243.

58. Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential hla-a2 binding peptides based on independent binding of individual peptide side-chains. *The Journal of Immunology* 1994. ISSN 0022-1767; **152**(1): 163–75. https://www.jimmunol.org/content/152/1/163.

59. Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. *Open Biology* 2013; **3**:120139. doi: 10.1098/rsob.120139.

60. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 2005; **6**:132. doi: 10.1186/1471-2105-6-132.

61. Pollard C, Koker SD, Saelens X, *et al*. Challenges and advances towards the rational design of mrna vaccines. *Trends in Molecular Medicine* 2013. ISSN 1471-4914; **19**(12): 705–13. doi: 10.1016/j.molmed.2013.09.002. http://www.sciencedirect.com/science/article/pii/S1471491413001664.

62. Rahman A, Hossain M, Tamanna S, Ullah MN. Recognition of a highly conserved dsrcptq epitope in envelope protein of zika virus through in silico approaches. *bioRxiv* 2020. doi: 10.1101/2020.02.11.943530.

63. Rammensee HG, Bachmann J, Emmerich NPN, *et al*. Syfpeithi: database for mhc ligands and peptide motifs. *Immunogenetics* 1999; **50**:213–9. doi: 10.1007/s002510050595.

64. Reche PA, Glutting J-P, Reinherz EL. Prediction of mhc class i binding peptides using profile motifs. *Human Immunology* 2002. ISSN 0198-8859; **63**(9): 701–9. doi: 10.1016/S0198-8859(02)00432-9. https://www.sciencedirect.com/science/article/pii/S0198885902004329.

65. Reynisson B, Alvarez B, Paul S, *et al*. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research* 2020. ISSN 0305-1048; 10.1093/nar/gkaa379 **48**(W1): W449–54. https://doi.org/10.1093/nar/gkaa379.

66. Richner JM, Himansu S, Dowd KA, *et al*. Modified mrna vaccines protect against zika virus infection. *Cell* 2017; **168**:1114–1125.e10. doi: 10.1016/j.cell.2017.02.017. https://www.cell.com/cell/fulltext/S0092-8674(17)30195-2.

67. Roth C, Cantaert T, Colas C, *et al*. A modified mrna vaccine targeting immunodominant ns epitopes protects against dengue virus infection in hla class i transgenic mice. *Frontiers in Immunology* 2019; **10**(06). doi: 10.3389/fimmu.2019.01424. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6598640/.

68. Rubinstein ND, Mayrose I, Pupko T. A machine-learning approach for predicting b-cell epitopes. *Molecular Immunology* 2009; **46**:840–7. doi: 10.1016/j.molimm.2008.09.009. https://www.ncbi.nlm.nih.gov/pubmed/18947876/.

69. Sela-Culang I, Ashkenazi S, Peters B, Ofran Y. PEASE: predicting B-cell epitopes utilizing antibody sequence. *Bioinformatics* 2014 ISSN 1367-4803; **31**(8): 1313–5. doi: 10.1093/bioinformatics/btu790.

70. Ellen Shrock, Eric Fujimura„ *et al*. Viral epitope profiling of covid-19 patients reveals cross-reactivity and correlates of severity. *Science*, 2020. ISSN 0036-8075. doi: 10.1126/science.abd4250. URL https://science.sciencemag.org/content/early/2020/09/28/science.abd4250.

71. Singh H, Raghava GPS. ProPred1: prediction of promiscuous MHC Class-I binding sites. *Bioinformatics* 2003 ISSN 1367-4803; **19**(8): 1009–14. doi: 10.1093/bioinformatics/btg108.

72. Solihah B, Azhari A, Musdholifah A. Enhancement of conformational b-cell epitope prediction using clusmote. *PeerJ Computer Science* June 2020 ISSN 2376-5992; **6**:e275. doi: 10.7717/peerj-cs.275.

73. Stefan Stojanovic. *stefs304/mhclovac*, 2021. URL https://github.com/stefs304/mhclovac.

74. Sun J, Di Wu TX, Wang X, *et al*. Seppa: a computational server for spatial epitope prediction of protein antigens. *Nucleic Acids Research* 2009; **37**:W612–6. doi: 10.1093/nar/gkp417. https://www.ncbi.nlm.nih.gov/pubmed/19465377/.

75. Sunita AS, Singh Y, Shukla P. Computational tools for modern vaccine development. *Human Vaccines & Immunotherapeutics* 2020; **16**(3): 723–35. doi: 10.1080/21645515.2019.1670035 PMID: 31545127.

76. Sweredoski MJ, Baldi P. Pepito: improved discontinuous b-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics (Oxford, England)* 2008; **24**:1459–60. doi: 10.1093/bioinformatics/btn199. https://pubmed.ncbi.nlm.nih.gov/18443018/.

77. Sweredoski MJ, Baldi P. Cobepro: a novel system for predicting continuous b-cell epitopes. *Protein Engineering, Design and Selection* 2009; **22**:113–20. doi: 10.1093/protein/gzn075. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2644406/.

78. Tai W, Zhang X, Drelich A, *et al*. A novel receptor-binding domain (rbd)-based mrna vaccine against sars-cov-2. *Cell Research* 2020. doi: 10.1038/s41422-020-0387-5. https://www.nature.com/articles/s41422-020-0387-5∖#citeas.

79. Thess A, Grund S, Mui BL, *et al*. Sequence-engineered mrna without chemical nucleoside modifications enables an effective protein therapy in large animals. *Molecular Therapy* 2015; **23**:1456–64. doi: 10.1038/mt.2015.103.

80. Jing-Hui Tian, Nita Patel„ *et al*. Sars-cov-2 spike glycoprotein vaccine candidate nvx-cov2373 immunogenicity in baboons and protection in mice. *Nature Communications*, **12**: 372, 2021. doi: 10.1038/s41467-020-20653-8. URL https://www.nature.com/articles/s41467-020-20653-8.

81. John Hopkins University. *Mortality analyses*, 2020. URL https://coronavirus.jhu.edu/data/mortality.

82. VanBlargan LA, Himansu S, Foreman BM, *et al*. An mrna vaccine protects mice against multiple tick-transmitted flavivirus infections. *Cell Reports* 2018; **25**(12): 3382–3392.e3. doi: 10.1016/j.celrep.2018.11.082. http://www.sciencedirect.com/science/article/pii/S2211124718318734.

83. Verbeke R, Lentacker I, Stefaan C, *et al*. Three decades of messenger rna vaccine development. *Nano Today* 2019; **28**:100766. doi: 10.1016/j.nantod.2019.100766.

84. Walsh EE, Frenck RW, Falsey AR, *et al*. Safety and immunogenicity of two rna-based covid-19 vaccine candidates. *New England Journal of Medicine* 2020. doi: 10.1056/nejmoa2027906.

85. Brian WJ, Philipe G, Annie S, *et al*. Phase 1 trial of a candidate recombinant virus-like particle vaccine for covid-19 disease produced in plants. *medRxiv* 2020. 10.1101/2020.11.04.20226282.

86. Worldometer. *Coronavirus toll update*: Cases & deaths by country of wuhan, china virus - worldometer, 2021. URLhttps://www.worldometers.info/coronavirus/.

87. Wrapp D, Wang N, Corbett KS, *et al*. Cryo-em structure of the 2019-ncov spike in the prefusion conformation. *Science* 2020. ISSN 0036-8075; **367**(6483): 1260–3. doi: 10.1126/science.abb2507. https://science.sciencemag.org/content/367/6483/1260.

88. George JX, Kula T, Qikai X, *et al*. Comprehensive serological profiling of human populations using a synthetic human virome. *Science* 2015. ISSN 0036-8075; **348**(6239). doi: 10.1126/science.aaa0698. https://science.sciencemag.org/content/348/6239/aaa0698.

89. Yao B, Zheng D, Liang S, Zhang C. Conformational b-cell epitope prediction on antigen protein structures: A review of current algorithms and comparison with common binding site prediction methods. *PLoS ONE* 2013a; **8**(04). doi: 10.1371/journal.pone.0062249. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3631208/\#pone.0062249-Zhang1.

90. Yao B, Zheng D, Liang S, Zhang C. Conformational b-cell epitope prediction on antigen protein structures: A review of current algorithms and comparison with common binding site prediction methods. *PLOS ONE* 2013b; **8**(4): 1–4. doi: 10.1371/journal.pone.0062249.

91. Yarmarkovich M, Warrington JM, Farrel A, Maris JM. Identification of sars-cov-2 vaccine epitopes predicted to induce long-term population-scale immunity. *Cell Reports Medicine* 2020. ISSN 2666-3791; **1**(3): 100036. doi: 10.1016/j.xcrm.2020.100036. http://www.sciencedirect.com/science/article/pii/S2666379120300483.

92. Zhang C, Maruggi G, Shan H, Li J. Advances in mrna vaccines for infectious diseases. *Frontiers in Immunology* 2019; **10**(03). doi: 10.3389/fimmu.2019.00594.

93. He Zhang, Liang Zhang, Ziyu Li, Kaibo Liu, Boxiang Liu, David H. Mathews, and Liang Huang. Lineardesign: Efficient algorithms for optimized mrna sequence design. *arXiv:2004.10177 [q-bio]*, 2020a. URL https://arxiv.org/abs/2004.10177v4.

94. Zhang N-N, Li X-F, Deng Y-Q, *et al*. A thermostable mrna vaccine against covid-19. *Cell* 2020b; **182**:1271–1283.e16. doi: 10.1016/j.cell.2020.07.024. https://www.sciencedirect.com/science/article/pii/S0092867420309326.