



# Computational Immune Proteomics Approach to Target COVID-19

Bruno Tilocca, Domenico Britti, Andrea Urbani, and Paola Roncada\*



Cite This: *J. Proteome Res.* 2020, 19, 4233–4241



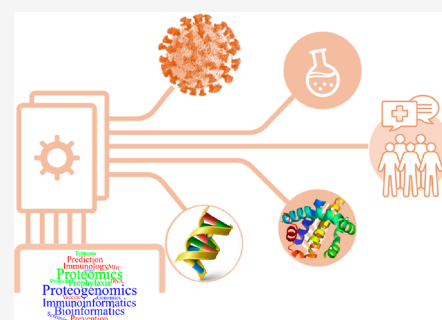
Read Online

ACCESS |

Metrics & More

Article Recommendations

**ABSTRACT:** Progress of the omics platforms widens their application to diverse fields, including immunology. This enables a deeper level of knowledge and the provision of a huge amount of data for which management and fruitful integration with the past evidence requires a steadily growing computational effort. In light of this, immunoinformatics emerges as a new discipline placed in between the traditional lab-based investigations and the computational analysis of the biological data. Immunoinformatics make use of tailored bioinformatics tools and data repositories to facilitate the analysis of data from a plurality of disciplines and help drive novel research hypotheses and in silico screening investigations in a fast, reliable, and cost-effective manner. Such computational immunoproteomics studies may as well prepare and guide lab-based investigations, representing valuable technology for the investigation of novel pathogens, to tentatively evaluate specificity of diagnostic products, to forecast on potential adverse effects of vaccines and to reduce the use of animal models. The present manuscript provides an overview of the COVID-19 pandemic and reviews the state of the art of the omics technologies employed in fighting SARS-CoV-2 infections. A comprehensive description of the immunoinformatics approaches and its potential role in contrasting COVID-19 pandemics is provided.



**KEYWORDS:** SARS-CoV-2, COVID-19, computational immune proteomics, bioinformatics, personalized medicine, diagnostics, vaccines

## ■ COVID-19 PANDEMIC: AN OVERVIEW

The current COVID-19 pandemic is an infectious disease manifested as a heterogeneous and complex clinical picture, most commonly reported as severe respiratory failure, renal problems, gastroenteritis, and death. The high contagiousness of the infectious disease and its rapid spread over boundaries warned health organizations, leading to the adoption of restrictive containment measures that, in turn, amplify the socio-economic impact the pathology is having worldwide.<sup>1,2</sup>

The causal agent of COVID-19 is a single-stranded positive-sense RNA virus belonging to the Coronavirus family (Coronaviridae).<sup>3</sup> Coronaviruses have been known as important animal pathogens for a long time. The first description of coronavirus dates back to the 1930s with the identification of the chicken Infectious Bronchitis Virus (IBV).<sup>4</sup> Since then, the veterinary community experienced the isolation of a wide variety of coronaviruses with both tropism for other animals (e.g., porcine CoV; bovine CoV; canine CoV, etc.) and different serotypes and genotypes from already known viral strains.<sup>5,6</sup> These are likely due to the relatively high frequency of mutation and recombination events of the RNA viruses and the large genome size of the coronaviruses (approximately 30Kbs) along with the poor fidelity of the viral RNA polymerase.<sup>7</sup> The rapid evolution of the viral strains affects the adoption of effective control measures (e.g., identification, diagnostic tests, vaccine formulations) besides hampering the viral classification. Nowa-

days, the classification of the coronaviruses relies mainly on genetic analysis, enabling the distinction of four different genera of Coronavirus, namely, alphacoronavirus, betacoronavirus, deltacoronavirus, and gammacoronavirus. Of these, the “sole” alpha- and beta- coronavirus genera are infective for humans.<sup>8,9</sup> The causal agent of the ongoing pandemic is a betacoronavirus, taxonomically related to the causal agent of the SARS epidemic registered in 2002 and the MERS epidemic of 2012, enabling the currently adopted nomenclature of SARS-CoV-2.<sup>3</sup> Since the wide distribution of Coronaviridae and the close sequence homology in tentative antigenic regions, there is an urgent need for a systematic evaluation of the immunological and structural relationships between viral sequences and the individual phenotypic immunological framework. Such knowledge will provide a rationale for the development of personalized prophylactic strategies which may well take into account the One-Health approach.

**Special Issue:** Proteomics in Pandemic Disease

**Received:** July 20, 2020

**Published:** September 11, 2020



## ■ STATE OF THE ART OF OMICS RESEARCH TO CONTRAST COVID-19

Since its early identification, numerous studies have been performed targeting the viral genome to track back SARS-CoV-2 origin and plausible diffusion routes, in an attempt to provide instruction while defining suitable containment measures, including the development of robust diagnosis tools and efficient prophylaxis programs.<sup>10,11</sup> Genome sequence alignment of the SARS-CoV-2 revealed a plausible origin of the novel viral strain from animal reservoirs, suggesting the bat coronavirus as the closest relative to the COVID-19 causal agent.<sup>12,13</sup> Nevertheless, discarding results have been registered, as the pangolin and the mink, among other wild animals, have also been indicated as the potential source of the SARS-CoV-2.<sup>14–18</sup>

Real-time reverse-transcription PCR (real-time RT-PCR) is, to date, the gold standard test for virus detection, but it does not provide useful information concerning the disease progression and/or the pathogenetic mechanism of the novel virus.<sup>19,20</sup> In this context, the almost exclusive study of the viral genome has soon disclosed intrinsic limits, raising the need to extend the focus of studies toward the viral proteins for comprehensive elucidation of the SARS-CoV-2 biology and delineate its transmission route(s).<sup>20</sup> The viral proteins are, indeed, the major structural components of the viral particle and the key effectors driving the pathogenetic mechanisms the virus activates during the diverse phases of the infective process.<sup>3</sup>

Nowadays, the power of mass spectrometry (MS), while providing rapid, precise, and versatile information, gives rise to a variety of MS-based proteomics aimed at complementing genomic information and targeting diverse aspects of this, yet unknown, pathogen. A very recent study by Cardozo et al.<sup>19</sup> proposes a high-throughput targeted proteomics assay as an alternative, or an integration, to the use of real-time RT-PCR for the detection of SARS-CoV-2 specific proteins. The study targets three peptides from the nucleocapsid (NC) protein of SARS-CoV-2 enabling the analysis of over 500 patients' respiratory tract samples per day in a fully automated manner. Similarly, by setting up a serum and plasma proteomic platform, Messner and colleagues<sup>21</sup> identified 27 putative biomarkers predictive of the diverse levels of severity of the COVID-19, according to the WHO disease classification. Another biomarker discovery-oriented study compensates for the lack of reference immunoproteomics studies by employing a SARS-CoV-2 proteome microarray for the robust investigation of the pathogen moieties involved in antibody interactions.<sup>22</sup> Authors preliminarily reported a dynamic range of 2 orders of magnitude and a limit of detection of 94 pg/mL, but further investigations on a larger clinical cohort to support these achievements are highly needed. Proteomics is also employed for the assessment of suitable therapeutic targets and/or to provide guidance to physicians in the choice of the most adequate therapeutic intervention. A study of Shen and colleagues<sup>23</sup> integrates the proteomics and metabolomics approaches to train a machine-learning model for the identification of COVID-19 patients and categorize them according to the severity of the disease, assessed in dependence of the proteomic and metabolomic profiling of patients' sera. Similarly, the proteomics survey of human Caco-2 cells experimentally infected with SARS-CoV-2 revealed the alteration of the central metabolism including carbon

metabolism, nucleic acids metabolism, splicing, and translation, suggesting novel drug targets. In addition, the study proves that administration of translation inhibitors (e.g., ribavirin) resulted in a reduced viral replication, shedding light on a new potential target to be further investigated for the design of efficient therapeutic intervention.<sup>24</sup>

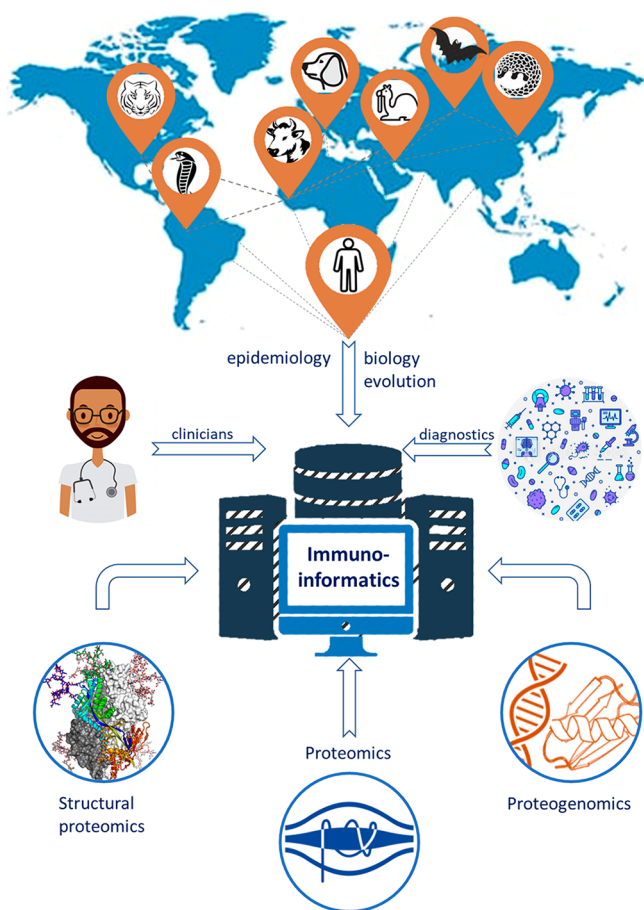
Regardless of the final goal of the above research, the common point to all reports is the need for further investigations on a wider representative sample population to either confirm or reject the diverse suggestions and hypotheses postulated based on their achievements, before providing the health systems with novel and efficient solutions. This limitation is further exacerbated by the adoption of stringent containment measures that, almost on a global scale, restricted access to laboratories to a very limited number of researchers, besides the impossibility, for obvious reasons, of working directly with the pathogen isolates and the virus manipulation that would speed up the research activity of many research groups.

The difficulties above, along with the steadily growing confidence in the *in silico* approaches have given rise to the use of computationally based strategies for the characterization of the novel pathogen, among these being immunoinformatics.

## ■ COMPUTATIONAL IMMUNE PROTEOMICS

During evolution, both vertebrate and invertebrate animals optimized their own immune system where a variety of diverse biological processes and biomolecules are finely orchestrated for the efficient protection of the host organism from diseases that result from either pathogen infection(s) or the altered behavior of the self-components (e.g., cancerous cell).<sup>25,26</sup> To acknowledge its importance, the immune system has been widely investigated. The number of studies performed, along with the heterogeneous approaches employed, has resulted in cumulative knowledge whose complementary integration has helped in a clear depiction of the multifacets of the immune system of each living system.<sup>27,28</sup>

Typically, immunologists use high-throughput methods of analysis and integrate the experimental data with clinical and epidemiological evidence.<sup>29–31</sup> The current progress of the omics platforms enables a deeper level of knowledge resulting in a huge amount of data, for which the management and fruitful integration with past findings requires a steadily growing computational effort.<sup>28,32</sup> In light of this, immunoinformatics has emerged as a new discipline placed in between the traditional wet-lab based investigations and the computational analysis of the biological data. Immunoinformatics, or more specifically for this review, computational immune proteomics, make use of tailored bioinformatics tools and data repositories to facilitate the analysis of the huge load of immunologic data such as immunomics, immunogenomics, and immunoproteomics and drive novel experimental research finalized at the design of innovative control measures against infectious disease and/or postulating and validating hypotheses<sup>25,28</sup> (Figure 1). Implementation of computational immune proteomics is, indeed, of crucial importance while validating serological laboratory tests of novel and/or unknown etiological agents, enabling the early diagnosis of infections. Immunoinformatics screening might also be involved in the prediction of the immunogenic potential of selected proteins of microbial origin and forecasting of the potential adverse effects of novel vaccinal formulates. Importantly, computational immune proteomics may be used to route research directions



**Figure 1.** Potential of computational immune proteomics in the control of the COVID-19 pandemic. The figure depicts a common workflow of the computational immune proteomics approach. Epidemiologic, genetic, and biologic data from a variety of hosts are acquired and stored in dedicated databases. Information is complemented with other data arising from either clinical evidence (clinical outcomes and diagnostics) or research activities such as structural proteomics studies, proteogenomics studies, and proteomics surveys. Altogether, data feed immunoinformatics and bioinformatics tools and software that, on the basis of a variety of algorithms, enable the thorough integration of all information and help draw conclusions and drive future/novel research directions.

and prevent researchers from inconclusive investigation; thus, among others, a significant reduction of animal testing benefits the ethical concern of the research trials.

The pivotal step in immunology is the identification of known epitope sequences and the prediction of novel ones. This information allows the study of the potential binding sites of the pathogen molecules and, in turn, draws the routes activated while eliciting the immune response. This is of fundamental importance for the development of novel and/or improved diagnostic strategies as well as the design of safe and effective prophylactic measures.<sup>33,34</sup>

## ■ COMPUTATIONAL IMMUNE PROTEOMICS IN EPITOPES PREDICTION

Epitopes are the portion of the antigen, commonly of proteinaceous nature, recognized by the receptor of the B-cell and T-cell (i.e., the major effectors of the adaptive immunity). B-cell epitopes are either linear or discontinuous sequences, typically solvent-exposed, of variable length (5–30

amino acids). B-cell binding stimulates the clonal expansion of the target B-cell and the massive release of antibodies that, binding the epitopes exposed by the pathogen cell, suppress the pathogen infectivity.<sup>35,36</sup>

Similarly to B-epitopes, T-epitopes are primarily proteins portions of 9–16 residues and can be further distinguished into two subcategories depending on their capability to be recognized by major histocompatibility complex (MHC) class I and class II molecules. T-epitope recognition is, therefore, a two-step process where epitopes are first processed by antigen presenting cells (APC) and the processed epitope is mounted onto either class I or class II MHC to be presented to CD8 T-cells and CD4 T-cells, respectively. In turn, CD8 T-cells and CD4 T-cells trigger a specific response aimed at killing and eliminating the pathogen.<sup>25,28,37</sup>

A thorough understanding of these molecular dynamics is crucial for the clear depiction of the immune response triggered against specific pathogens. To this purpose, the steadily increasing use of genomic sequencing, proteomics, and clinical and epidemiologic data result in accumulating immunological data whose management and integration are efficiently accomplished by the immunoinformatics enabling data mining to provide epitope prediction in a safe, rapid, and reproducible manner.<sup>38,39</sup> By acknowledging the importance of the computational immune proteomics and, broadly, the immunoinformatics, several software, tools, and databases are being optimized. Principal tools employed in the epitope prediction are listed in Table 1, whereas a comprehensive review of the most suitable bioinformatics to be employed in these disciplines is provided by Soria-Guerra and colleagues.<sup>40</sup> In addition, parallel to the Human Proteome Project (HPP), the Human Immuno-Peptidome Project (HIPP, <https://www.hupo.org/Human-Immuno-Peptidome-Project>) aims at studying and mapping the inventory of human peptides presented by HLA molecules through state-of-the-art methodologies and instruments and make this information accessible to immunologists, clinical investigators, and other research groups.

### Prediction of B-epitopes

Prediction and mapping of linear and/or conformational B-cell epitopes have been recently made available through mass spectrometry-based protocols. For linear epitopes, the construction of random peptide libraries to be surveyed with either monoclonal antibodies or infected sera is among the most common approach.<sup>41–43</sup> This approach has been effective in the investigation of the B-cell epitopes of diverse pathogens such as the Dengue virus, leading to the identification of two epitopes as suitable candidates to be employed in the development of novel diagnostic tools and/or vaccinal strategies.<sup>44</sup>

Most of the B-cell epitopes are conformational or discontinuous; here, distant portions of the epitope sequence are brought together by the specific protein folding, resulting in a tertiary structure with specific complementarity versus its antibody or receptor.<sup>28,45</sup> Typical study of the conformational epitopes expects the isolation of the epitope–receptor complex and its downstream characterization by means of a multitude of techniques such as mass spectrometry and/or crystallography, depending on the aim of the survey. Nevertheless, such an approach is laborious and costly, since it requires screening of a huge number of peptides for each pathogen.<sup>46,47</sup> A promising solution in the screening for B-epitopes relies on the

**Table 1. Major Bioinformatic Tools in Epitope Prediction<sup>a</sup>**

tool	description	link
EpiJen	T-epitope prediction Prediction of TAP epitopes	<a href="http://www.ddg-pharmfac.net/epijen/EpiJen/EpiJen.htm">http://www.ddg-pharmfac.net/epijen/EpiJen/EpiJen.htm</a>
SYFPEITHI	T-epitope prediction	<a href="http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm">http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm</a>
ANNPRED	T-epitope prediction	<a href="http://www.imtech.res.in/raghava/nhlapred/neural.html">http://www.imtech.res.in/raghava/nhlapred/neural.html</a>
MHCPred	T-epitope prediction	<a href="http://www.ddg-pharmfac.net/mhcpred/MHCPred/">http://www.ddg-pharmfac.net/mhcpred/MHCPred/</a>
NetMHC	T-epitope prediction	<a href="http://www.cbs.dtu.dk/services/NetMHC/">http://www.cbs.dtu.dk/services/NetMHC/</a>
PREDEP	T-epitope prediction	<a href="http://margalit.huji.ac.il/Teppred/mhc-bind/index.html">http://margalit.huji.ac.il/Teppred/mhc-bind/index.html</a>
RANKPEP	T-epitope prediction	<a href="http://imed.med.ucm.es/Tools/rankpep.html">http://imed.med.ucm.es/Tools/rankpep.html</a>
SVMHC	T-epitope prediction	<a href="http://abi.inf.uni-tuebingen.de/Services/SVMHC">http://abi.inf.uni-tuebingen.de/Services/SVMHC</a>
IEDB	T-epitope prediction	<a href="http://tools.immuneepitope.org/analyze/html/mhc_processing.html">http://tools.immuneepitope.org/analyze/html/mhc_processing.html</a>
EpiVax	T-epitope prediction	<a href="http://www.epivax.com/">http://www.epivax.com/</a>
NetCTL	T-epitope prediction Prediction of TAP epitopes	<a href="http://www.cbs.dtu.dk/services/NetCTL">http://www.cbs.dtu.dk/services/NetCTL</a>
SVRMHC	T-epitope prediction	<a href="http://c1.accurascience.com/SVRMHCdb/">http://c1.accurascience.com/SVRMHCdb/</a>
WAPP	Prediction of TAP epitopes	<a href="http://abi.inf.uni-tuebingen.de/Services/WAPP/information">http://abi.inf.uni-tuebingen.de/Services/WAPP/information</a>
BepiPred	Linear B-epitopes	<a href="http://www.cbs.dtu.dk/services/BepiPred/">http://www.cbs.dtu.dk/services/BepiPred/</a>
BEST	Linear B-epitopes	<a href="http://biomine.cs.vcu.edu/">http://biomine.cs.vcu.edu/</a>
EPCES	Conformational B-epitopes	<a href="http://sysbio.unl.edu/services/EPCES/">http://sysbio.unl.edu/services/EPCES/</a>
DiscoTope	Conformational B-epitopes	<a href="http://www.cbs.dtu.dk/services/DiscoTope/">http://www.cbs.dtu.dk/services/DiscoTope/</a>
BEPro	Conformational B-epitopes	<a href="http://pepito.proteomics.ics.uci.edu/">http://pepito.proteomics.ics.uci.edu/</a>
EpiSearch	Prediction of discontinuous B-cell epitopes	<a href="http://curie.utmb.edu/episearch.html">http://curie.utmb.edu/episearch.html</a>
Epitopia	Linear and Conformational B-epitopes	<a href="http://epitopia.tau.ac.il/">http://epitopia.tau.ac.il/</a>
PepSurf	Linear and Conformational B-epitopes	<a href="http://pepito.tau.ac.il">http://pepito.tau.ac.il</a>
ElliPro	Linear and Conformational B-epitopes	<a href="http://tools.immuneepitope.org/tools/ElliPro/iedb_input">http://tools.immuneepitope.org/tools/ElliPro/iedb_input</a>

<sup>a</sup>Adapted from Soria-Guerra et al.<sup>40</sup>

use of the novel immunoinformatic approaches. Here, tailored databases and algorithms are available for the rapid and reproducible screening of both linear and conformational epitopes.<sup>48</sup> A crucial step of the B-epitope prediction is represented by matching of a potential antigenic protein against specific databases of known B-epitopes. Several data repositories are nowadays available to run a fast and reliable prediction of linear B-type epitopes based on the sequence similarity. Examples of this databases are Bcipep, a comprehensive data repository providing the peptide sequences of known B-epitopes, enabling the classification of the queried sequence(s) on the basis of the immunogenicity.<sup>49</sup> Nevertheless, these databases do not contain information on conformational B-epitopes and are being overcome by other, more versatile, repositories that complement the sequence information with structural and conformational data including X-ray diffraction, nuclear magnetic resonance, and a collection of previous empirically derived data that allow for robust predictions of both linear and conformational epitopes.<sup>50</sup> With

the progress of the immunoinformatics, several algorithm and web-based tools are available for the customizable prediction of both linear and conformational epitopes through the use of diverse criteria, mainly classified as (i) sequence-based methods, (ii) amino acid propensity scale, and (iii) machine learning.<sup>51–53</sup> Current sequence-based approaches are the evolution of the “sole” sequence matching and comparison taking into account other factors such as the surface accessibility of the putative epitope sequence and the physicochemical features of the B-cell receptor that binds the epitope. Nevertheless, this approach is suitable for the prediction of linear peptides.<sup>48</sup> The amino acid propensity scale provides a more comprehensive survey of the queried sequences, as it evaluates the antigenic propensity of the amino acids as means of a score computed by considering the position of the amino acid residues and other parameters such as the surface exposure, hydrophobicity, molecular flexibility, polarity turns, and accessibility of the amino acids to the receptor.<sup>28,54</sup> Machine-learning is the last frontier in the field of immunoinformatics (and bioinformatics in general), enabling the prediction of epitopes at high accuracy (60–66%) acknowledging the continuous improvement of both the informatics approaches and the accumulating knowledge in epitope prediction and immunology.<sup>55</sup> To date, a combination of the amino acid propensity scale and machine learning results in a powerful prediction of linear epitopes.<sup>28</sup> However, conformational epitope prediction is also efficiently accomplished by machine learning approaches, as it takes into account data from the tridimensional characterization of the queried polypeptides such as X-ray crystallography and NMR.<sup>56</sup> Moreover, conformational epitope prediction may also be performed for the proteins whose tridimensional structure has not yet been elucidated. Here, a structure prediction step is required for computing the most likely 3D structure of the polypeptide,<sup>57</sup> thus computing the candidate conformational epitopes. Nevertheless, the accuracy level of the prediction is far lower than the predictions based on high-affinity structural models.

A study performed by Adhikari and colleagues employed immunoinformatics for the design of an epitope-based peptide vaccine against Oropouche virus (OROV), an emerging pathogen causing Oropouche fever and meningitis in humans.<sup>58</sup> The study employed the BepiPred algorithm which is trained on epitopes identified by antigen–antibody complex structures and LBtope, a method for linear epitopes which is based on experimentally validated data. These identified five linear and eight conformational B-epitopes as suitable candidates for the design of a vaccinal strategy. Tools based on the amino acid propensity scale such as the BCPred algorithm, instead, were efficiently applied for the identification of B-cell epitopes expressed by the Dengue virus.<sup>52</sup> A very recent study of Liu et al. employed a machine learning approach to predict N-linked glycosylation sites and B-cell epitopes of the influenza A viruses.<sup>59</sup> The methods focus on RNA sequence conservation, conserved B-antigenic regions and N-glycosylation, three peculiar characteristics of the influenza A virus and thoroughly processed virus subtype representatives, proving the robustness of the methods in the assessment of the pre-vaccine and the vaccine invalidation assays.

**Table 2. Selected Computational Immune Proteomics-Based Research Targeting the Major Structural Proteins of SARS-CoV-2**

target protein	aim	method	refs
Spike	Epitope sharing through One-Health approach	Protein sequences alignments	70
	Prophylactic-oriented study	B-epitope prediction	72
	Diagnostic deliverables	T-epitope prediction	
	Pathogenetic mechanism elucidation	Multiple protein sequence alignment and comparison	74
	Prophylactic-oriented study	T-epitopes prediction from convalescent patients	75
		Epitopes cross reaction test	
	Prophylactic-oriented study	B-epitopes prediction	76
	Diagnostic deliverables	T-epitopes prediction	
		Prediction of epitopes conservation and population coverage	
		Conformational epitopes prediction	
Nucleocapsid	Prophylactic-oriented study	Conformational epitopes mapping	77
	Prophylactic-oriented study	B-epitopes prediction	
	Prophylactic-oriented study	T-epitopes prediction	
	Epitope sharing through One-Health approach	B-epitopes prediction	73,78
	Prophylactic-oriented study	T-epitopes prediction	
	Diagnostic deliverables	Conformational epitopes mapping	
	Prophylactic-oriented study	B-epitope prediction	72
	Diagnostic deliverables	T-epitope prediction	
	Prophylactic-oriented study	T-epitopes prediction from convalescent patients	75
		Epitopes cross reaction test	
Envelope	Prophylactic-oriented study	B-epitopes prediction	76
	Diagnostic deliverables	T-epitopes prediction	
	Prophylactic-oriented study	Prediction of epitopes conservation and population coverage	
	Diagnostic deliverables	Conformational epitopes prediction	
	Prophylactic-oriented study	Conformational epitopes mapping	
	Epitope sharing through One-Health approach	B-epitopes prediction	79
	Prophylactic-oriented study	T-epitopes prediction	
	Diagnostic deliverables	3D-folding prediction Conformational epitopes mapping	
	Prophylactic-oriented study	B-epitope prediction	80
		T-epitope prediction	
Membrane	Prophylactic-oriented study	B-epitope prediction	72
	Diagnostic deliverables	T-epitope prediction	
	Prophylactic-oriented study	Multiple sequence alignment	71
	Prophylactic-oriented study	Protein molecular docking	
	Diagnostic deliverables	Knowledgebase discovery	
	Prophylactic-oriented study	Therapeutic deliverables	
	Prophylactic-oriented study	Prophylactic-oriented study	
	Diagnostic deliverables	Prophylactic-oriented study	
	Prophylactic-oriented study	Prophylactic-oriented study	
	Diagnostic deliverables	Prophylactic-oriented study	
Membrane	Prophylactic-oriented study	B-epitope prediction	72
	Diagnostic deliverables	T-epitope prediction	
	Prophylactic-oriented study	T-epitopes prediction from convalescent patients	75
		Epitopes cross reaction test	
	Prophylactic-oriented study	B-epitopes prediction	76
	Diagnostic deliverables	T-epitopes prediction	
		Prediction of epitopes conservation and population coverage	
		Conformational epitopes prediction	
		Conformational epitopes mapping	

### Prediction of T-epitopes

Prediction of protein sequences capable of eliciting a T-cell response is a fundamental step for a diversity of applications, ranging from the development of vaccinal strategies to the optimization of diagnostics for the timely and specific diagnosis of a plethora of pathogens.<sup>60</sup> Despite B-cell epitopes, the binding of T-cell epitopes is mediated by the interaction with a specific MHC molecule that, in turn, enable the recognition of the epitope by the T-cell receptor. In this context, elucidating the interaction between the epitope and the MHC molecule is of paramount importance for the accurate epitope prediction. Thus, the available T-epitope prediction tool targets the interaction between the peptide and the MHC molecule.<sup>28,37,61</sup> Nevertheless, MHC molecules are highly polymorphic molecules whose variability relies on diverse factors, such as

the polygeny of the MHC genes, the codominance of the parental alleles, the polymorphism of the genetic variants, and the antigenic peptide splicing. This complicates the prediction of the epitopes that is currently performed by assessing the binding affinity of the queried sequences against a range of MHC molecules.<sup>62,63</sup> The choice of the target MHC is empirical when possible; otherwise, the most common MHC alleles within a given population are to be chosen.

Major methods for epitope prediction are the hidden Markov model (HMM), artificial neural networks (ANNs), support vector machine (SVM), and quantitative matrices (QM). These statistics use the data as mathematical models and integrate multiple levels of knowledge to perform the probabilistic prediction of epitopes on the basis of diverse criteria.<sup>48,64</sup> Accordingly, several databases and tools to serve

T-epitope prediction have been developed and are continually updated as reviewed by Bahrami et al.<sup>28</sup> Nielsen et al. employed ANN for the efficient prediction of T-epitopes from hepatitis C virus<sup>65</sup> and to prove its ability to outperform other commonly employed prediction criteria. In contrast, Zhao and colleagues suggested SVM as the best performing strategy in the prediction of T-epitopes,<sup>66</sup> and a variety of viral biology studies make use of the SVM as reviewed here.<sup>67</sup> Nevertheless, the latest machine learning methods are univocally recognized as the most reliable and efficient strategy so far available in the field of immunoinformatics.<sup>36,68</sup> A very recent study performed by Paul and colleagues<sup>69</sup> developed a machine learning approach that covers all the major T-epitope prediction tools to both make a pairwise comparison of the epitope prediction tools and perform a comprehensive survey of the vaccinia virus epitopes capable of eliciting a T-mediated response. Similarly, Adhikari et al.<sup>58</sup> implemented multiple tools for assessing the potential epitopes of an emerging virus of human concern and provide potential candidates for the design of effective control measures. The study screened at first the most immunogenic protein, and its amino acid sequence was queried for epitope survey. The NetCTL repository was used to predict CD8+ T-cell epitopes for a plurality of MHC-I haplotypes, taking into account the proteasomal C terminal cleavage, MHC class I binding, and TAP transport efficiency which, in turn, are predicted using artificial neural networks and the weight matrix, respectively.

#### ■ CONTRIBUTION OF COMPUTATIONAL IMMUNE PROTEOMICS IN FACING COVID-19 EMERGENCE

Most immunogenic proteins of SARS CoV-2 are four major structural proteins involved in both virion assembly and the infective process. The spike (S) protein is a homotrimeric molecule that constitutes the spikes on the viral surface for the recognition and attachment of the viral particle to the host cell.<sup>70</sup> The envelope protein (E) is mainly involved in viral assembly and budding, besides taking part in the pathogenic processes.<sup>71</sup> The membrane protein (M) is the major constituent of the viral membrane; it is constituted of three transmembrane domains responsible for the viral shape and the binding with the nucleocapsid (NC) protein.<sup>72</sup> NC is the fourth structural protein at the interface between the outer surface and the inner viral particle. This protein consists of two domains that can bind the viral genome and are involved in a variety of processes, including the modulation of the viral replication and the pathogenetic mechanism.<sup>73</sup> The structural proteins of SARS CoV-2 show interesting immunogenic properties and are the objective of diverse research lines aimed at developing efficient vaccinal strategies and/or optimizing sensitive diagnostic tools based on the antigen–antibody complex formation, including the immunoinformatic approaches. A summary of the most recent research activities targeting the viral structural proteins along with the computational immune proteomics method employed is provided in Table 2.

Advantages of the immunoinformatics enabled its steadily increasing usage by diverse research groups aimed at providing a first immunological featuring of the novel pathogen in a fast and reproducible manner.<sup>81</sup> In addition, immunoinformatics broadened the plethora of scientists working on SARS-CoV-2, allowing screening studies without the need of equipped laboratories and regardless of access to the virus isolates. Immunoinformatics-based study of the SARS-CoV-2 spans

from complementing the genomics-based studies to providing guidance for the development of effective diagnostics and prophylactic oriented studies.<sup>82</sup> By following the One-Health approach, our research group employed immunoinformatics, as a means of computational immune proteomics, and investigated three major structural proteins of the SARS-CoV-2 (i.e., Spike, S; Membrane, M; Envelope, E) and compared them with the protein sequences of other coronaviruses with tropism for other synanthropic and domestic animals such as dog, bovine, camel, and dromedary.<sup>70,71,73,79</sup> All studies reported a high level of similarity between the structural protein of SARS-CoV-2 and the counterparts of the bat and pangolin coronaviruses, confirming the results of molecular studies aimed at tracking the origin of the novel pathogen.<sup>11,74,83</sup>

Epitope prediction of the SARS-CoV-2 proteins resulted in the provision of peptide sequences with the potential of eliciting both humoral antibody production and the stimulation of the cell-mediated immune response. Interestingly, a handful of epitopes of the SARS-CoV-2 structural proteins share very high similarity with the protein portion, and structures, of the other coronaviruses with tropism for the animal humans are frequently interacting with.<sup>70,79,73</sup> This evidence is suggestive of a partial protective immunity elicited by the previous exposure of humans to the animal coronaviruses; in turn, this would explain, at least in part, the varying degree of severity of the COVID-19. Such information is of valuable importance for the clear definition of the SARS-CoV-2 immunogenicity and might also be implemented in the design of diagnostic tools capable of differential diagnosis and/or the sketch of a safe and efficient vaccinal formulation. Similarly, a previous study of Tetro<sup>84</sup> takes into account the epitope sharing among the novel pathogen and other coronaviruses and hypothesize the antibody dependent enhancement (ADE) as a plausible explanation of the worsening of the clinical picture of some infected patients due to their putative previous exposure to epitopes shared by other viral strains. In the context of the prophylactic oriented studies, a preliminary survey<sup>78</sup> employed immunoinformatics for predicting SARS-CoV-2 B and T epitopes and compared them against the epitope prediction of the causal agent of the SARS pandemic in 2003. Acknowledging the genetic similarity and epitopes shared by authors leveraged the immunological data already available for SARS-CoV in an attempt to deliver efficient vaccinal solutions in a time-effective manner. The SARS-CoV-2 epitope assay was also performed by Grifoni et al.<sup>72</sup> through the use of HLA class I and II predicted peptide “megapools” to be used for the test of serum samples from both healthy and SARS-CoV-2 infected patients. CD8+ and CD4+ T cell response was registered in the infected patients as stimulated by the major structural proteins and a few nonstructural proteins. In addition, CD4+ T cell response was identified in 40–60% of unexposed individuals, suggesting the occurrence of cross-reaction between SARS-CoV-2 and the taxonomically related viruses.<sup>75</sup> By following another immunoinformatics approach, Mukherjee and colleagues predicted both B and T epitopes of SARS-CoV-2 with the potential for development of a vaccinal formulation.<sup>76</sup> The study provides a list of candidate immunodominant epitopes capable of eliciting both humoral and cell-mediated immunity in a rapid and cost-effective manner, whose efficiency is to be tested experimentally. Another study specifically targeting the surface glycoprotein provides a list of five cytotoxic T-cell epitopes, three linear, and

five more structural B-cell epitopes.<sup>77</sup> Further featuring of the T-epitopes highlighted the capability of binding the class I MHC molecules by taking multiple contacts in the groove of the molecules through hydrogen bonds and salt bridge anchors, which support the immunostimulant potential of these sequences, deserving of experimental validation.

## CONCLUSION

The current pandemic condition requires a huge effort of the worldwide scientific community in providing fast and efficient measures for the control of the SARS-CoV-2 diffusion and/or recurrence. Despite the enormous advances of the state of the art technologies, the lack of knowledge of the novel pathogen strain poses heavy limitations in all aspects of scientific research. For instance, only a very limited number of research consortia have access to the SARS-CoV-2 research, depending on their capability of obtaining viral isolates from clinical samples and their safe manipulation in properly equipped laboratories. On the other hand, the steady diffusion of the immunoinformatics along with the great skillfulness of the research groups enabled the inclusion of a much wider plethora of scientists that, by employing an ethical approach, tested diverse arrays of hypothesis remotely and *in silico*. This sped up the overall research progress by suggesting novel research directions and assessing statistical solidity to the “rare” experimental data and clinical observations, especially in the immunogenicity of SARS-CoV-2 and the potential host immune response tailored against the novel pathogen. Accumulating pieces of evidence support immunoinformatics as a reliable and reproducible discipline, although empirical confirmations are required to confirm the computer-based predictions especially concerning their ability for early/improved diagnosis and the safety of the vaccinal formulations. Empirical confirmations may also serve as further support of the robustness of the immunoinformatics, and we are confident that the steadily growing employment of this discipline will contribute to its self-improvement and its active inclusion while planning experimental researches aimed at studying both known and yet uncharacterized pathogens: COVID-19 pandemic is just the latest example.

## AUTHOR INFORMATION

### Corresponding Author

**Paola Roncada** – Department of Health Sciences, University “Magna Graecia” of Catanzaro, Catanzaro 88100, Italy;  
orcid.org/0000-0002-0114-5872; Phone: (+39)0961-3694284; Email: roncada@unicz.it

### Authors

**Bruno Tilocca** – Department of Health Sciences, University “Magna Graecia” of Catanzaro, Catanzaro 88100, Italy  
**Domenico Britti** – Department of Health Sciences, University “Magna Graecia” of Catanzaro, Catanzaro 88100, Italy  
**Andrea Urbani** – Department of Basic Biotechnological Sciences, Intensivological and Perioperative Clinics, Università Cattolica del Sacro Cuore, Roma 00168, Italy; Dipartimento di Scienze di laboratorio e infettivologiche, Fondazione Policlinico Universitario Agostino Gemelli, Roma 00168, Italy

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acs.jproteome.0c00553>

## Notes

The authors declare no competing financial interest.

## REFERENCES

- (1) Pedersen, S. F.; Ho, Y. C. SARS-CoV-2: A Storm Is Raging. *J. Clin. Invest.* **2020**, *130*, 2202–2205.
- (2) van Doremalen, N.; Bushmaker, T.; Morris, D. H.; Holbrook, M. G.; Gamble, A.; Williamson, B. N.; Tamin, A.; Harcourt, J. L.; Thornburg, N. J.; Gerber, S. I.; Lloyd-Smith, J. O.; de Wit, E.; Munster, V. J. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *N. Engl. J. Med.* **2020**, *382*, 1564–1567.
- (3) Chen, Y.; Liu, Q.; Guo, D. Emerging Coronaviruses: Genome Structure, Replication, and Pathogenesis. *J. Med. Virol.* **2020**, *92*, 418–423.
- (4) Ramakrishnan, S.; Kappala, D. Avian Infectious Bronchitis Virus. *In Recent Advances in Animal Virology* **2019**, *6*, 301–319.
- (5) Buonavoglia, C.; Decaro, N.; Martella, V.; Elia, G.; Campolo, M.; Desario, C.; Castagnaro, M.; Tempesta, M. Canine Coronavirus Highly Pathogenic for Dogs. *Emerging Infect. Dis.* **2006**, *12*, 492–494.
- (6) Decaro, N.; Lorusso, A. Novel Human Coronavirus (SARS-CoV-2): A Lesson from Animal Coronaviruses. *Vet. Microbiol.* **2020**, *244*, 108693.
- (7) Pachetti, M.; Marini, B.; Benedetti, F.; Giudici, F.; Mauro, E.; Storici, P.; Masciovecchio, C.; Angeletti, S.; Ciccozzi, M.; Gallo, R. C.; Zella, D.; Ippodrino, R. Emerging SARS-CoV-2 Mutation Hot Spots Include a Novel RNA-Dependent-RNA Polymerase Variant. *J. Transl. Med.* **2020**, *18*, 179.
- (8) Wu, D.; Wu, T.; Liu, Q.; Yang, Z. The SARS-CoV-2 Outbreak: What We Know. *Int. J. Infect. Dis.* **2020**, *94*, 44–48.
- (9) Masters, P. S. The Molecular Biology of Coronaviruses. *Adv. Virus Res.* **2006**, *66*, 193–292.
- (10) Andersen, K. G.; Rambaut, A.; Lipkin, W. I.; Holmes, E. C.; Garry, R. F. The Proximal Origin of SARS-CoV-2. *Nat. Med.* **2020**, *26*, 450–452.
- (11) Zhou, P.; Yang, X.; Lou, W.; Wang, X. G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H. R.; Zhu, Y.; Li, B.; Huang, C. L.; Chen, H. D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R.; Di, L.; Liu, M. Q.; Chen, Y.; Shen, X. R.; Wang, X.; Zheng, X. S.; Zhao, K.; Chen, Q. J.; Deng, F.; Liu, L. L.; Yan, B.; Zhan, F. X.; Wang, Y. Y.; Xiao, G. F.; Shi, Z. L. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature* **2020**, *579*, 270–273.
- (12) Tang, X.; Wu, C.; Li, X.; Song, Y.; Yao, X.; Wu, X.; Duan, Y.; Zhang, H.; Wang, Y.; Qian, Z.; Cui, J.; Lu, J. On the Origin and Continuing Evolution of SARS-CoV-2. *Natl. Sci. Rev.* **2020**, *7*, 1012–1023.
- (13) Forster, P.; Forster, L.; Renfrew, C.; Forster, M. Phylogenetic Network Analysis of SARS-CoV-2 Genomes. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 9241–9243.
- (14) Zhang, T.; Wu, Q.; Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr. Biol.* **2020**, *30*, 1578.
- (15) Yang, Y.; Peng, F.; Wang, R.; Guan, K.; Jiang, T.; Xu, G.; Sun, J.; Chang, C. The Deadly Coronaviruses: The 2003 SARS Pandemic and the 2020 Novel Coronavirus Epidemic in China. *J. Autoimmun.* **2020**, *109*, 102434.
- (16) Lam, T. T. Y.; Jia, N.; Zhang, Y. W.; Shum, M. H. H.; Jiang, J. F.; Zhu, H. C.; Tong, Y. G.; Shi, Y. X.; Ni, X. B.; Liao, Y. S.; Li, W. J.; Jiang, B. G.; Wei, W.; Yuan, T. T.; Zheng, K.; Cui, X. M.; Li, J.; Pei, G. Q.; Qiang, X.; Cheung, W. Y. M.; Li, L. F.; Sun, F. F.; Qin, S.; Huang, J. C.; Leung, G. M.; Holmes, E. C.; Hu, Y. L.; Guan, Y.; Cao, W. C. Identifying SARS-CoV-2-Related Coronaviruses in Malayan Pangolins. *Nature* **2020**, *583*, 282–285.
- (17) Guo, Q.; Li, M.; Wang, C.; Fang, Z.; Wang, P.; Tan, J.; Wu, S.; Xiao, Y.; Zhu, H. Host and Infectivity Prediction of Wuhan 2019 Novel Coronavirus Using Deep Learning Algorithm. *bioRxiv* **2020**, *1* DOI: [10.1101/2020.01.21.914044](https://doi.org/10.1101/2020.01.21.914044).
- (18) Oreshkova, N.; Molenaar, R.-J.; Vreman, S.; Harders, F.; Munnink, B. B. O.; Honing, R. W. H.; Gerhards, N.; Tolsma, P.;

Bouwstra, R.; Sikkema, R.; Tacken, M.; Rooij, M. M. T. de; Weesendorp, E.; Engelsma, M.; Bruschke, C.; Smit, L. A. M.; Koopmans, M.; Poel, W. H. M. van der; Stegeman, J. SARS-CoV2 Infection in Farmed Mink, Netherlands, April 2020. *bioRxiv* **2020**, 25, 25.

(19) Cardozo, K. H. M.; Lebkuchen, A.; Okai, G. G.; Schuch, R. A.; Viana, L. G.; Olive, A. N.; Lazari, C. S.; Fraga, A. M.; Granato, C. F. H.; Carvalho, V. M. Fast and Low-Cost Detection of SARS-CoV-2 Peptides by Tandem Mass Spectrometry in Clinical Samples. *Research Square* **2020**, 1.

(20) Corman, V. M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D. K. W.; Bleicker, T.; Brünink, S.; Schneider, J.; Schmidt, M. L.; Mulders, D. G. J. C.; Haagmans, B. L.; Van Der Veer, B.; Van Den Brink, S.; Wijsman, L.; Goderski, G.; Romette, J. L.; Ellis, J.; Zambon, M.; Peiris, M.; Goossens, H.; Reusken, C.; Koopmans, M. P. G.; Drosten, C. Detection of 2019 Novel Coronavirus (2019-nCoV) by Real-Time RT-PCR. *Eurosurveillance* **2020**, 25, 2000045.

(21) Messner, C. B.; Demichev, V.; Wendisch, D.; Michalick, L.; White, M.; Freiwald, A.; Textoris-Taube, K.; Vernardis, S. I.; Egger, A.-S.; Kreidl, M.; Ludwig, D.; Kilian, C.; Agostini, F.; Zeleznik, A.; Thibeault, C.; Pfeiffer, M.; Hippenstiel, S.; Hocke, A.; von Kalle, C.; Campbell, A.; Hayward, C.; Porteous, D. J.; Marionni, R. E.; Langenberg, C.; Lilley, K. S.; Kuebler, W. M.; Müllleder, M.; Drosten, C.; Witznath, M.; Kurth, F.; Sander, L. E.; Ralser, M. Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Syst.* **2020**, 11, 11–24.

(22) Wang, H.; Hou, X.; Wu, X.; Liang, T.; Zhang, X.; Wang, D.; Teng, F.; Dai, J.; Duan, H.; Guo, S.; Li, Y.; Yu, X. SARS-CoV-2 Proteome Microarray for Mapping COVID-19 Antibody Interactions at Amino Acid Resolution. *bioRxiv* **2020**, 1 DOI: 10.1101/2020.03.26.994756.

(23) Shen, B.; Yi, X.; Sun, Y.; Bi, X.; Du, J.; Zhang, C.; Quan, S.; Zhang, F.; Sun, R.; Qian, L.; Ge, W.; Liu, W.; Liang, S.; Chen, H.; Zhang, Y.; Li, J.; Xu, J.; He, Z.; Chen, B.; Wang, J.; Yan, H.; Zheng, Y.; Wang, D.; Zhu, J.; Kong, Z.; Kang, Z.; Liang, X.; Ding, X.; Ruan, G.; Xiang, N.; Cai, X.; Gao, H.; Li, L.; Li, S.; Xiao, Q.; Lu, T.; Zhu, Y.; Liu, H.; Chen, H.; Guo, T. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* **2020**, 182, 59–72.

(24) Bojkova, D.; Klann, K.; Koch, B.; Widera, M.; Krause, D.; Ciesek, S.; Cinatl, J.; Münch, C. Proteomics of SARS-CoV-2-Infected Host Cells Reveals Therapy Targets. *Nature* **2020**, 583, 469–472.

(25) Tomar, N.; De, R. K. Immunoinformatics: An Integrated Scenario. *Immunology* **2010**, 131, 153–168.

(26) Kimbrell, D. A.; Beutler, B. The Evolution and Genetics of Innate Immunity. *Nat. Rev. Genet.* **2001**, 2, 256–267.

(27) Bansal, A. K. Bioinformatics in Microbial Biotechnology - A Mini Review. *Microb. Cell Fact.* **2005**, 4, 19.

(28) Bahrami, A. A.; Payandeh, Z.; Khalili, S.; Zakeri, A.; Bandehpour, M. Immunoinformatics: In Silico Approaches and Computational Design of a Multi-Epitope, Immunogenic Protein. *Int. Rev. Immunol.* **2019**, 38, 307–322.

(29) Ma, H.; Shieh, K.-J.; Lee, S. Study of ELISA Technique. *Nature and Science* **2006**, 4, 36–37.

(30) Yates, A.; Chan, C. C.; Callard, R. E.; George, A. J.; Stark, J. An Approach to Modelling in Immunology. *Briefings Bioinf.* **2001**, 2, 245–257.

(31) Davey, H. M. Flow Cytometric Techniques for the Detection of Microorganisms. *Methods Cell Sci.* **2002**, 24, 91–97.

(32) Gardy, J. L.; Lynn, D. J.; Brinkman, F. S. L.; Hancock, R. E. W. Enabling a Systems Biology Approach to Immunology: Focus on Innate Immunity. *Trends Immunol.* **2009**, 30, 249–262.

(33) Martins, L. M. S.; Andrade, H. M.; De; Vainstein, M. H.; Wanke, B.; Schrank, A.; Balaguez, C. B.; Santos, P. R.; Dos; Santi, L.; Pires, S. D. F.; Da Silva, A. S.; De Castro, J. A. F.; Brandão, R. M. S. D. S.; Do Monte, S. J. H. Immunoproteomics and Immunoinformatics Analysis of *Cryptococcus Gattii*: Novel Candidate Antigens for Diagnosis. *Future Microbiol.* **2013**, 8, 549–563.

(34) Blythe, M. J.; Flower, D. R. Benchmarking B Cell Epitope Prediction: Underperformance of Existing Methods. *Protein Sci.* **2005**, 14, 246–248.

(35) Van Regenmortel, M. H. V. What Is a B-Cell Epitope? *Methods Mol. Biol.* **2009**, 524, 3–20.

(36) Sanchez-Trincado, J. L.; Gomez-Perosanz, M.; Reche, P. A. Fundamentals and Methods for T- and B-Cell Epitope Prediction. *J. Immunol. Res.* **2017**, 2017, 1–14.

(37) Moon, J. J.; Chu, H. H.; Hataye, J.; Pagán, A. J.; Pepper, M.; McLachlan, J. B.; Zell, T.; Jenkins, M. K. Tracking Epitope-Specific T Cells. *Nat. Protoc.* **2009**, 4, 565–581.

(38) Korber, B.; LaButte, M.; Yusim, K. Immunoinformatics Comes of Age. *PLoS Comput. Biol.* **2006**, 2, e71.

(39) Davies, M. N.; Flower, D. R. Harnessing Bioinformatics to Discover New Vaccines. *Drug Discovery Today* **2007**, 12, 389–395.

(40) Soria-Guerra, R. E.; Nieto-Gomez, R.; Govea-Alonso, D. O.; Rosales-Mendoza, S. An Overview of Bioinformatics Tools for Epitope Prediction: Implications on Vaccine Development. *J. Biomed. Inf.* **2015**, 53, 405–414.

(41) Wu, C. H.; Liu, I. J.; Lu, R. M.; Wu, H. C. Advancement and Applications of Peptide Phage Display Technology in Biomedical Science. *J. Biomed. Sci.* **2016**, 23, 8.

(42) Molek, P.; Strukelj, B.; Bratkovic, T. Peptide Phage Display as a Tool for Drug Discovery: Targeting Membrane Receptors. *Molecules* **2011**, 16, 857–887.

(43) Devlin, J. J.; Panganiban, L. C.; Devlin, P. E. Random Peptide Libraries: A Source of Specific Protein Binding Molecules. *Science* **1990**, 249, 404–406.

(44) Amin, N.; Aguilar, A.; Chamacho, F.; Vázquez, Y.; Pupo, M.; Ramirez, J. C.; Izquierdo, L.; Dafnis, F.; Stott, D. I.; Perez, E. M.; Acosta, A. Identification of Dengue-Specific B-Cell Epitopes by Phage-Display Random Peptide Library. *Malaysian. J. Med. Sci.* **2009**, 16, 4–14.

(45) Ramanathan, B.; Poh, C. L.; Kirk, K.; McBride, W. J. H.; Aaskov, J.; Grollo, L. Synthetic B-Cell Epitopes Eliciting Cross-Neutralizing Antibodies: Strategies for Future Dengue Vaccine. *PLoS One* **2016**, 11, e0155900.

(46) Tong, J. C.; Ren, E. C. Immunoinformatics: Current Trends and Future Directions. *Drug Discovery Today* **2009**, 14, 684–689.

(47) Parker, J. M. R.; Guo, D.; Hodges, R. S. New Hydrophilicity Scale Derived from High-Performance Liquid Chromatography Peptide Retention Data: Correlation of Predicted Surface Residues with Antigenicity and X-Ray-Derived Accessible Sites. *Biochemistry* **1986**, 25, 5425–5432.

(48) Backert, L.; Kohlbacher, O. Immunoinformatics and Epitope Prediction in the Age of Genomic Medicine. *Genome Med.* **2015**, 7, 119.

(49) Saha, S.; Bhasin, M.; Raghava, G. P. S. Bcipep: A Database of B-Cell Epitopes. *BMC Genomics* **2005**, 6, 79.

(50) Huang, J.; Honda, W. CED: A Conformational Epitope Database. *BMC Immunol.* **2006**, 7, 7.

(51) Saha, S.; Raghava, G. P. S. Prediction of Continuous B-Cell Epitopes in an Antigen Using Recurrent Neural Network. *Proteins: Struct., Funct., Genet.* **2006**, 65, 40–48.

(52) Chen, J.; Liu, H.; Yang, J.; Chou, K. C. Prediction of Linear B-Cell Epitopes Using Amino Acid Pair Antigenicity Scale. *Amino Acids* **2007**, 33, 423–428.

(53) Manavalan, B.; Govindaraj, R. G.; Shin, T. H.; Kim, M. O.; Lee, G. IBCE-EL: A New Ensemble Learning Framework for Improved Linear B-Cell Epitope Prediction. *Front. Immunol.* **2018**, 9, 1695.

(54) Rubinstein, N. D.; Mayrose, I.; Pupko, T. A Machine-Learning Approach for Predicting B-Cell Epitopes. *Mol. Immunol.* **2009**, 46, 840–847.

(55) Söllner, J.; Mayer, B. Machine Learning Approaches for Prediction of Linear B-Cell Epitopes on Proteins. *J. Mol. Recognit.* **2006**, 19, 200–208.

(56) Bhasin, M.; Raghava, G. P. S. Prediction of CTL Epitopes Using QM, SVM and ANN Techniques. *Vaccine* **2004**, 22, 3195–31204.



- (57) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Humana Press, 2005; pp 571–607.
- (58) Adhikari, U. K.; Tayebi, M.; Mizanur Rahman, M. Immunoinformatics Approach for Epitope-Based Peptide Vaccine Design and Active Site Prediction against Polyprotein of Emerging Oropouche Virus. *J. Immunol. Res.* **2018**, *2018*, 1–22.
- (59) Liu, J. H.; Chang, C. C.; Chen, C. W.; Wong, L. T.; Chu, Y. W. Conservation Region Finding for Influenza A Viruses by Machine Learning Methods of N-Linked Glycosylation Sites and B-Cell Epitopes. *Math. Biosci.* **2019**, *315*, 108217.
- (60) Rosa, D. S.; Ribeiro, S. P.; Cunha-Neto, E. CD4+ T Cell Epitope Discovery and Rational Vaccine Design. *Arch. Immunol. Ther. Exp.* **2010**, *58*, 121–130.
- (61) Tomar, N.; De, R. K. Immunoinformatics: A Brief Review. *Methods Mol. Biol.* **2014**, *1184*, 23–55.
- (62) Gyles, C. L.; Prescott, J. F.; Songer, J. G.; Thoen, C. O. *Pathogenesis of Bacterial Infections in Animals*, 4th ed.; Wiley-Blackwell, 2010. DOI: 10.1002/9780470958209.
- (63) Vigneron, N.; Stroobant, V.; Chapiro, J.; Ooms, A.; Degiovanni, G.; Morel, S.; Van Der Bruggen, P.; Boon, T.; Van Den Eynde, B. J. An Antigenic Peptide Produced by Peptide Splicing in the Proteasome. *Science* **2004**, *304*, 587–590.
- (64) Trolle, T.; Metushi, I. G.; Greenbaum, J. A.; Kim, Y.; Sidney, J.; Lund, O.; Sette, A.; Peters, B.; Nielsen, M. Automated Benchmarking of Peptide-MHC Class I Binding Predictions. *Bioinformatics* **2015**, *31*, 2174–2181.
- (65) Nielsen, M.; Lundegaard, C.; Warming, P.; Lauemøller, S. L.; Lamberth, K.; Buus, S.; Brunak, S.; Lund, O. Reliable Prediction of T-Cell Epitopes Using Neural Networks with Novel Sequence Representations. *Protein Sci.* **2003**, *12*, 1007–1017.
- (66) Zhao, Y.; Pinilla, C.; Valmori, D.; Martin, R.; Simon, R. Application of Support Vector Machines for T-Cell Epitopes Prediction. *Bioinformatics* **2003**, *19*, 1978–1984.
- (67) Shapshak, P.; Somboonwit, C.; Sinnott, J. T.; Menezes, L. J.; Kanguane, P.; Balaji, S.; Chiappelli, F. *Global Virology III: Virology in the 21st Century*; Springer: Cham, 2019. DOI: 10.1007/978-3-030-29022-1.
- (68) Desai, D. V.; Kulkarni-Kale, U. T-Cell Epitope Prediction Methods: An Overview. *Methods Mol. Biol.* **2014**, *1184*, 333–364.
- (69) Paul, S.; Croft, N. P.; Purcell, A. W.; Tschärke, D. C.; Sette, A.; Nielsen, M.; Peters, B. Benchmarking Predictions of MHC Class I Restricted T Cell Epitopes. *PLoS Comput. Biol.* **2020**, *16*, 694539.
- (70) Tilocca, B.; Soggiu, A.; Musella, V.; Britti, D.; Sanguinetti, M.; Urbani, A.; Roncada, P. Molecular Basis of COVID-19 Relationships in Different Species: A One Health Perspective. *Microbes Infect.* **2020**, *22*, 218–220.
- (71) Maio, F.; De; Cascio, E.; Lo; Babini, G.; Sali, M.; Longa, S.; Della; Tilocca, B.; Roncada, P.; Arcovito, A.; Sanguinetti, M.; Scambia, G.; Urbani, A. Enhanced Binding of SARS-CoV-2 Envelope Protein to Tight Junction-Associated PALS1 Could Play a Key Role in COVID-19 Pathogenesis. *Research Square* **2020**, DOI: 10.21203/rs.3.rs-30903/v1.
- (72) Grifoni, A.; Sidney, J.; Zhang, Y.; Scheuermann, R. H.; Peters, B.; Sette, A. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* **2020**, *27*, 671–680.
- (73) Tilocca, B.; Soggiu, A.; Sanguinetti, M.; Musella, V.; Britti, D.; Bonizzi, L.; Urbani, A.; Roncada, P. Comparative Computational Analysis of SARS-CoV-2 Nucleocapsid Protein Epitopes in Taxonomically Related Coronaviruses. *Microbes Infect.* **2020**, *22*, 188–194.
- (74) Zhang, T.; Wu, Q.; Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr. Biol.* **2020**, *30*, 1346–1351.
- (75) Grifoni, A.; Weiskopf, D.; Ramirez, S. I.; Mateus, J.; Dan, J. M.; Moderbacher, C. R.; Rawlings, S. A.; Sutherland, A.; Premkumar, L.; Jazi, R. S.; Marrama, D.; de Silva, A. M.; Frazier, A.; Carlin, A. F.; Greenbaum, J. A.; Peters, B.; Krammer, F.; Smith, D. M.; Crotty, S.; Sette, A. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell* **2020**, *181*, 1489–1501.
- (76) Mukherjee, S.; Tworowski, D.; Detroja, R.; Mukherjee, S. B.; Frenkel-Morgenstern, M. Immunoinformatics and Structural Analysis for Identification of Immunodominant Epitopes in SARS-CoV-2 as Potential Vaccine Targets. *Vaccines* **2020**, *8*, 290.
- (77) Baruah, V.; Bose, S. Immunoinformatics-Aided Identification of T Cell and B Cell Epitopes in the Surface Glycoprotein of 2019-NCoV. *J. Med. Virol.* **2020**, *92*, 495–500.
- (78) Ahmed, S. F.; Quadeer, A. A.; McKay, M. R. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses* **2020**, *12*, 254.
- (79) Tilocca, B.; Soggiu, A.; Sanguinetti, M.; Babini, G.; De Maio, F.; Britti, D.; Zecconi, A.; Bonizzi, L.; Urbani, A.; Roncada, P. Immunoinformatic Analysis of the SARS-CoV-2 Envelope Protein as a Strategy to Assess Cross-Protection against COVID-19. *Microbes Infect.* **2020**, *22*, 182–187.
- (80) Abdelmageed, M. I.; Abdelmoneim, A. H.; Mustafa, M. I.; Elfadl, N. M.; Murshed, N. S.; Shantier, S. W.; Makhawi, A. M. Design of Multi Epitope-Based Peptide Vaccine against E Protein of Human 2019-NCoV: An Immunoinformatics Approach. *BioMed Res. Int.* **2020**, *2020*, 1.
- (81) Schönbach, C.; Ranganathan, S.; Brusica, V. *Immunoinformatics*; Springer: New York, NY, 2008. DOI: 10.1007/978-0-387-72968-8.
- (82) Bianchi, M.; Benvenuto, D.; Giovanetti, M.; Angeletti, S.; Pascarella, S. Sars-CoV-2 Envelope and Membrane Proteins: Differences from Closely Related Proteins Linked to Cross-Species Transmission? *Preprints* **2020**, DOI: 10.20944/preprints202004.0089.v1.
- (83) Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.; Bi, Y.; Ma, X.; Zhan, F.; Wang, L.; Hu, T.; Zhou, H.; Hu, Z.; Zhou, W.; Zhao, L.; Chen, J.; Meng, Y.; Wang, J.; Lin, Y.; Yuan, J.; Xie, Z.; Ma, J.; Liu, W. J.; Wang, D.; Xu, W.; Holmes, E. C.; Gao, G. F.; Wu, G.; Chen, W.; Shi, W.; Tan, W. Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *Lancet* **2020**, *395*, S65–S74.
- (84) Tetro, J. A. Is COVID-19 Receiving ADE From Other Coronaviruses? *Microbes Infect.* **2020**, *22*, 72–73.