Indonesian Journal of Life Sciences Vol. 02 | Number 01 | March (2020)
http://journal.i3l.ac.id/ojs/index.php/IJLS/

RESEARCH ARTICLE

What do T cells see in SARS-CoV2? Immunoinformatics analysis to identify T cell epitopes from SARS-CoV2 ORF1ab polyprotein

Marsia Gustiananda

Department of Biomedicine, Indonesia International Institute for Life Sciences, Jakarta, Indonesia

Email: marsia.gustiananda@i3l.ac.id

ABSTRACT

The current epidemic caused by a novel coronavirus SARS-CoV2 as well as two previously documented pandemic caused by SARS-CoV and MERS-CoV imposes that a spillover of an animal coronavirus to humans is a continuous threat. The zoonotic nature of the infection contributes to the unpredictability of the pandemic. In such situations, the availability of the 'off the shelf' vaccines that target the conserved region of the coronavirus might help in preventing the spread of the diseases. Therefore, efforts to generate such vaccines should be considered as a priority. The whole genome of SARS-CoV2 is readily available in the public database one month after the first case was identified. The platform technology known as the "genome to vaccine" approach would provide useful start to identify parts of the virus proteome which can be the candidate for vaccine components. This study used an immunoinformatic approach to identify T cell epitopes from SARS-CoV2 ORF1ab polyprotein in an attempt to design a genome-derived epitope-based universal coronavirus vaccine.

Keywords: immunoinformatics; ORF1ab; SARS-CoV2; T cell epitopes; epitope-based vaccine

INTRODUCTION

The outbreak of novel coronavirus SARS-CoV2 which started in Wuhan-China in December 2019 has spread to 85 countries, including Indonesia, and causes in total 95333 confirmed cases and 3282 deaths (WHO situation reports 45). Individuals infected by SARS-CoV2 developed symptoms of a severe acute respiratory syndrome similar to SARS-CoV and MERS-CoV, which include viral pneumonia, fever, difficulty breathing, and in most severe cases the infiltration of the bilateral lung (Huang et al., 2020).

SARS-CoV2, MERS-CoV, SARS-CoV and several other human coronaviruses, which cause the common cold, belong to the same virus family, known as Coronavirideae. During SARS-CoV epidemy in 2002, 8422 people were reported to be infected and resulted in 916 dead. MERS-CoV epidemy in 2012 caused 1401 infections and 543 death (WHO, 2018; Koh et al., 2010). The periodic emergence of coronaviruses as a spillover from the animal causing global epidemy showed that the virus is a continuous threat to human health and well beings. The zoonotic nature of the viral infection contributes to the unpredictability of the occurrence of coronavirus infections which can only be detected when outbreaks occur. Therefore, efforts to generate vaccines

for a pandemic should be considered as a priority.

The platform technology known as the "genome to vaccine" approach would provide useful start to identify a part of the virus proteome which can be the candidate for vaccine components (Moise et al., 2011; Backert et al., 2015). Using the immunoinformatics approach, one can identify the region on the virus that will be recognized by receptors of the immune cells.

The complete genome sequence of SARS-CoV2 was released on 10 January 2020, which is about one month after the first case was identified and 2 days after the virus was isolated from a Wuhan pneumonia patient on 8 January 2020. SARS-CoV2 genome is a positive sense, single-stranded, and non-segmented RNA with a size of about 30 kB (Lu et al., 2020). SARS-CoV2 is an enveloped virus with nucleocapsid that adopts a helical symmetry. The genome is translated into several proteins namely: ORF1ab polyprotein, surface glycoprotein (S), ORF3a, envelope protein, membrane protein, and several accessory proteins (ORF6, ORF7a, ORF7b, ORF8), and nucleocapsid protein (N).

The SARS-CoV2 ORF1ab occupies two-third of the virus genome, similar to that in other coronaviruses, with nucleotide identities of ⩾83.6% (Chan et al., 2020). The ORF1ab polyprotein was translated from the partial overlapping sequence of the ORF1ab gene. The ~7200 amino acid polyprotein processed by a protease to generate 16 putative nonstructural proteins (nsps). These nsps are proteases such as nsp3 (papain-like protease) and nsp5 (chymotrypsin-like protease), and some other enzymes involved in the transcription and replication of the RNA genomes, such as nsp12 (RNA-dependent RNA polymerase-RdRp) and nsp13 (helicase) (Chan et al., 2020). There is seven replicase domain within ORF1ab polyprotein, which, due to the high conservancy among coronavirus species, were used for CoV species classification. It is reported that these replicase domains of SARS-CoV2 are 94.6% identical with the domains in SARS-CoV (Zhou et al., 2020). Another study implied that ORF1ab especially the RdRp region is highly conserved among human coronaviruses and could potentially become a good candidate for universal vaccine components against coronavirus infection (Sharmin and Islam, 2014).

The vaccine against infectious pathogens ideally should induce both arms of adaptive immunity namely humoral (antibody-mediated) and cellular (cell-mediated) immunity. Neutralizing antibodies can bind to the viral surface protein and prevent viral entry into the host cells. Cell-mediated immunity, which is our second line of adaptive defenses, will be able to limit the infection and might reduce the symptoms to subclinical levels. Cellular immunity is mediated by CD8+ T cells and CD4+ T cells. CD8+ T cell is a cytotoxic T cell that will kill the virus-infected cells. CD4+ T cell will help B cells to differentiate into an antibody-secreting plasma cell. Recent studies have indicated the importance of T cells in viral clearance during primary infection of SARS-CoV. SARS-specific memory T cell responses against the structural proteins, S, M, E, and NP are detected in the blood of SARS survivors even after several years recovered from infection (Fan et al. 2009; Li et al., 2008; Liu et al. 2010).

Not only that the SARS-specific memory T cells recognize epitopes from the structural proteins mentioned above, the T cells recognizing epitopes from the non-structural protein such as replicase has also been identified. In this study, the overlapping peptides spanning the entire SARS-CoV genome were synthesized and tested against the PBMCs from 1-year post-infected patients. The result showed that eight out of the fourteen SARS-CoV proteins (replicase, S,

N, E, M, 3a, 3b, and 9b) contain peptides that induce the production of IFNγ by T cells (Oh et al., 2012). This study showed that replicase which is part of the ORF1ab polyprotein can also become the target for T cell responses. The replicase which represents two-thirds of the SARS-CoV genome were reported to be poorly immunogenic since the study reported that only 7 out of 1386 overlapping peptides originated from replicase ORF1ab are recognized by T cells from SARS-CoV recovered patients (Li et al., 2008.). The detail sequences of these 7 peptides, however, were not reported in the paper. However, the IFNγ-ELISPOT result showed that the response generated from T cells recognizing the peptides originated from replicase is the highest compare to other proteins (Li et al., 2008). IEDB search also revealed that there are around 1533 peptides originated from ORF1ab polyprotein that were potential to become T cell epitopes as they are shown to bind HLA molecules with high affinity in the biochemical peptide binding assay (www.iedb.org. access on 5 February 2020) (Vita et al., 2018).

However, detailed analysis and possible epitope definition of T cell responses against ORF1ab polyprotein of SARS-CoV2 are currently lacking. Here, we analyzed the presence of T cell epitopes from SARS-CoV2 ORF1ab polyprotein using an immunoinformatics approach to identify vaccine components. ORF1ab is a quite stable protein, and there are significant levels of conservancy between SARS CoV and SARS-CoV2, making it an attractive target as sources of T cells epitopes for vaccine components. Vaccines that target the conserved T cell epitopes will solve problems related to viral mutation and evolution and could become an "off the shelf" vaccine for emerging diseases caused by coronaviruses. Since T cells recognize peptide antigen in the context of a complex with HLA molecule, which is the most polymorphic gene in the human genome, the T-cell epitope-based vaccine design should take into account the HLA alleles that are present in the population.

At the time the manuscript was written, there had been 27 confirmed cases of SARS-CoV2 in Indonesia. Knowing what T cells see in the SARS-CoV2 will guide the identification of immune epitopes useful for vaccine formulation. The candidate epitopes, in the forms of peptides with 9 amino acid residues, will be presented by HLA alleles present in the Indonesian population. The HLA allele types and frequencies of the Indonesian population, especially the peopling of Java and Sunda-Java have been well characterized (Yuliwulandari et al., 2009). The information about such T cell epitopes will be useful for both vaccine design and immune diagnostics

MATERIAL AND METHODS

Virus and proteins sequences

The genome sequence of the coronavirus isolate Wuhan-Hu-1 (Genbank MN908947.3/ Ref Seq NC_045512) is the first isolate of Wuhan seafood market pneumonia virus made publicly available and housed in the Genbank. The sequences contain 12 coding sequences. The ORF1ab polyprotein (NCBI Reference Sequence: YP_009724389.1) was used in the current analysis as a source of nonamer peptides that could become T cell epitopes.

Immunoinformatic tools for T cell epitope prediction

Prediction server netCTLpan (http://www.cbs.dtu.dk/services/NetCTLpan/) (Stranzl et al., 2010) and netMHCIIpan (http://www.cbs.dtu.dk/services/NetMHCIIpan/) (Jensen et al., 2018) was used to analyzed The SARS-CoV ORF1ab polyprotein (YP_009724389.1) for the cytotoxic T lymphocyte and T helper cell epitopes, respectively. Prediction of both CTL and T

helper epitopes were conducted for the 9-mer peptides that are presented by HLA Class I and Class II alleles found in the Javanese and Sundanese-Javanese population. The data about HLA alleles and the frequency in this population were taken from the curated HLA database of HLA allele frequency (http://www.allelefrequencies.net/) (Gonzalez-Galarza et al., 2015) and reported in the published paper by Yuliwulandari et al. (2009). Since many of the HLA alleles of the Indonesian population is not well studied, the peptide binding to those HLA molecules is largely unknown. Therefore, in this analysis, the pan-based prediction servers were chosen for its ability to predict peptide binding based on the sequence similarity of the binding groove to the known HLA (Stranzl et al., 2010). The netCTLpan method was able to predict the probability of each step in the MHC Class I pathway which consists of protein proteasomal cleavage to generate the peptides, efficiency of peptides transport to the ER and peptide binding to HLA molecules (Larsen et al., 2005). The netMHCIIpan can predict if the 9-mer peptides will become the core peptide for HLA Class II. In this analysis, the selection for the 9-mer peptides predicted by netCTLpan and netMHCIIpan was based on the highest prediction score (< 1%).

SARS-CoV2 ORF1ab peptide sequence cross-conservation with human self-peptides.

The peptides having sequences that are cross-conserved with the sequence of the self-peptides might induce either autoimmune response if it is used for vaccine or reduce the immunogenicity of the vaccine, due to tolerogenic T cell responses. Therefore blastp analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was done to check sequence similarity between the nonamer peptides derived from SARS-CoV2 and the 9-mer peptides derived from the non-redundant human protein sequences (taxid: 9606). The blast algorithm parameter was automatically adjusted for short input sequences, word size was 2; expect threshold was 30000, matrix used was PAM30, gap cost was set to existence=9 and extension = 1, the compositional parameter was set to no adjustment and low complexity filter was disabled. Using Microsoft excell, we screened out the peptides which shared at least contiguously 7 identical amino acid residues with the human 9-mer peptides with no gap and no mismatches residue.

SARS-CoV2 ORF1ab peptide sequence cross-conservation with SARS-CoV peptides curated in IEDB.

The cross-conservation of SARS-CoV2 and SARS-CoV peptides was checked using epitope search tools in the Immune Epitope Database (www.IEDB.org) (Vita et al., 2018). IEDB contains curated epitopes that had been confirmed experimentally either by T-cell assay or HLA binding assay and the majority of the coronavirus T cell epitopes curated in IEDB are originated from SARS-CoV. The nonamer sequences of SARS-CoV2 that are cross-conserved with peptides reported in IEDB indicated that the peptides are part of the protein which is essential for function and not easily mutate since the sequence has been there since SARS-CoV outbreak in 2004.

SARS-CoV2 ORF1ab peptide sequence cross-conservation across 52 SARS-CoV2 isolates and other coronaviruses

The complete sequences of SARS-CoV2 isolates were downloaded from Genbank (5 March 2020) and ORF1ab polyprotein sequences were curated. The conservancy of the predicted epitopes across 52 isolates of SARS-CoV2 was done using the IEDB antigen conservancy tools (http://tools.immuneepitope.org/tools/conservancy/iedb_input) (Bui et al., 2007). The duplicate sequences were eliminated from the analysis

to avoid bias. The cross-conservancy analysis was also conducted against 202 sequences of ORF1ab from other coronaviruses including SARS-CoV, MERS-CoV, Hu-CoV causing common cold and animal coronaviruses. The sequences were retrieved from Genbank on 15 February 2020.

Population coverage of the SARS-CoV2 ORF1ab peptides.

The potential of the 9-mer peptides to be recognized by the Indonesian population immune system were calculated using the population coverage analysis tool (http://tools.immuneepitope.org/tools/population/ iedb_input) (Bui et al., 2006). Two input files for this analysis are the file containing the list of peptides and the HLA that bind to them, and another file containing the HLA allele frequencies of the population of interest.

RESULTS AND DISCUSSIONS

The outbreak of recent coronavirus which is caused by SARS-CoV2 and the previous two outbreaks caused by coronaviruses such as SARS-CoV and MERS-CoV emphasize the importance of the creation of the pandemic coronavirus vaccine. The Wuhan-Hu-1 (MN908947.3/ Ref Seq NC_045512) is the first isolate of SARS-CoV2 that was made publicly available in Genbank. The availability of the genomic information of this virus has allowed scientific community to analyze the genomes for both drug design (Gao et al., 2020; Xu et al., 2020; Zhavoronkov et al., 2020) as well as vaccine components (Kumar, 2020; Ahmed et al., 2020; Sarkar et al., 2020).

As an attempt to prepare for a pandemic, the availability of the “off-the-shelf” vaccine should become the priority. The components of the universal vaccine should be generated from the conserved immunogenic region that the virus will not mutate. ORF1ab, which occupies two-thirds of the coronavirus genome and encodes nonstructural proteins and enzymes involved in viral replications (Ziebuhr 2005; Chan et al, 2020), could be the target for T cell epitopes. IFNγ ELISPOT assay which tested overlapping peptides from SARS-CoV showed that patients recovered from SARS-CoV infection has T cells recognizing peptides originated from the ORF1ab polyprotein (Oh et al., 2012) and that the amount of T cell response to these epitopes are quite high, although compared to other coronavirus proteins, the ORF1ab is categorized as only intermediately immunogenic (Li et al., 2008). Another study reported the conserve peptides (WDYPKCDRA) within the RdRp protein, which is part of the ORF1ab polyprotein that considers a good target as B cell epitopes (Sharmin et al., 2014). Here we set out to identify if the other region within the ORF1ab polyprotein, which encodes the important enzymes for virus replication also contains conserved peptides epitope for both CD8+ and CD4+ T cells.

All overlapping 9-mer peptides spanning the entire ORF1ab polyprotein were generated in silico and were screened for potential T-cell epitopes using two prediction servers netCTLpan (peptide presented by HLA Class I) and netMHCIIpan (peptide presented by HLA Class II). The prediction servers netCTLpan and netMHCIIpan housed in CBS website are based on pan method which allows prediction for peptide binding to all HLA alleles with a known protein sequence, regardless of the absence of experimental data. Pan method predicts peptide binding to HLA alleles based on sequence similarity of the peptide-binding groove between HLA alleles of the same supertype.

HLA Class I alleles present peptides to CD8+ T cells, which are cytotoxic T cells that will kill the virus-infected cells. HLA A*24:07

(allele frequency of 22%) is one of the HLA Class I alleles that is important for the Javanese and Sundanese-Javanese population. The peptide binding specificity of this allele has not been well characterized and due to differences in the residue number 70, the A*24:07 is not classified into the A*24 supertype (Sidney et al., 2008) which is represented by HLA-A*24:02. Residue number 70 is part of the B pocket of the peptide-binding groove on the HLA molecules. In HLA-A*24:07 residue number 70 is glutamine while in HLA-A*24:02 is histidine. The difference could have some impact on binding specificity that makes HLA-A*24:07 differs from the rest of the members of the A*24 supertype (Gustiananda et al., 2013). The limitation of this study is, therefore, the unavailability of the tool to predict peptide binding to HLA Alleles that are not part of the supertype. However, that limitation was indirectly alleviated by choosing only the peptides that bind to more than one HLA allele (the promiscuous peptides).

ORF1ab protein of Wuhan-Hu-1 (MN908947.3) is about 7096 amino acid length, which means that being the largest polyprotein within coronavirus proteome, the possible number of nonamer peptides generated from this protein is 7088. The peptides were screened using NetCTLpan and netMHCIIpan and those having the top highest prediction score (rank <1%) were selected to minimize the false positives. The highest score peptides are more likely to become T cell epitopes in vivo because netCTLpan not only predict peptide binding to HLA Class I but also the peptide processing inside the cells (Larsen et al., 2005).

Our analysis revealed that ORF1ab polyprotein was capable of contributing 9-mer peptides which will be presented by HLA Class I and Class II of the Indonesian population (Figure 1). This is in agreement with published papers which predicted that 370 peptides from SARS-CoV2 ORF1ab will bind to HLA-A, 516 to HLA-B and 775 to HLA-DRB1 (Fast and Chen, 2020). Our result is also in line with the study of cell-mediated immune response against the whole proteome of SARS-CoV, which revealed that PBMCs from 1-year post-infected patients contained T cells responding to eight out of the fourteen SARS-CoV proteins (S, N, E, M, 3a, 3b, 9b, and replicase) (Oh et al., 2012). This study showed experimentally that replicase which is part of the ORF1ab polyprotein can also become the target for T cell responses. Even though the SARS replicase which represent two-thirds of the SARS-CoV genome was reported to be poorly immunogenic, in which that only 7 out of 1386 overlapping peptides originated from replicase ORF1ab recognized by T cells from SARS-CoV recovered patients (Li et al., 2008.) although the detail sequences of the epitopes were not reported in the paper. However, the IFNγ-ELISPOT result showed that the response generated from T cells recognizing the ORF1ab peptide is highest compare to other proteins (Li et al., 2008). IEDB search revealed that there are around 1533 peptides originated from ORF1ab polyprotein that has the potential to become T cell epitopes as they were shown to bind to HLA molecules with strong affinity in the biochemical peptide binding assay (www.iedb.org access 5 February 2020) (Vita et al., 2018).

The netCTLpan and netMHCIIpan results show that there are 945 nonamer peptides predicted to be presented by HLA Class I alleles and Class II alleles (Figure 1). These peptides are promiscuous binders, meaning that they bind to more than one HLA allele. To develop the vaccine, the chosen peptides must cover as many HLA alleles as possible to maximize the population coverage. At this stage, only the promiscuous peptides were selected for the next step in the analysis. In the final list, the peptide capable of binding

the greatest number of alleles was VMYMGTLSY which could bind 27 HLA class I alleles and 6 HLA Class II alleles.

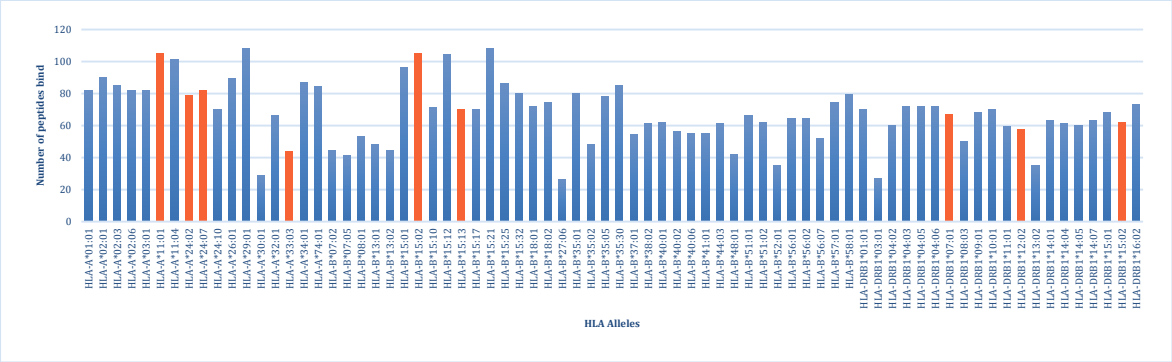

Figure 1. The number of predicted peptid
es across HLA Class I and Class II alleles of the Javanese and Sundanese-Javanese population. Bar in orange color is the HLA with alleles frequency higher than 10% in the population. The total number of predicted epitopes was 945.

Immunoinformatic analysis of ORF1ab revealed that HLA-A*29:01 and HLA-B*15:21 were predicted to bind the highest number of peptides, both bind 108 out of 945 predicted peptides. However, the HLA-A*29:01 allele was present at low frequency in the population (0,8 %) and it might not be relevant to include it in the vaccine construct. HLA-B*15:21, however, present at 6,2%. HLA A*24:07, with 21.52% allele frequency is the most frequent HLA A allele among the Javanese and Sundanese-Javanese population. This allele bind to 82 predicted peptides. The other most frequent HLA A alleles are A*11:01 (16.03%) bound 105 peptides, A*33:03 (15.61%) bound 44 peptides and A*24:02 (14.35%) bound 79 peptides. The two most common HLA B alleles are HLA B*15:02 (11.6%) bound 105 peptides and HLA B*15:13 (11.18%) bound 70 peptides (Figure 1).

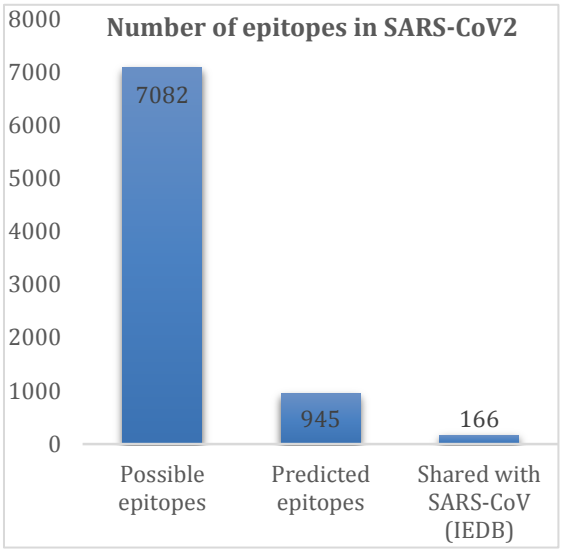

Figure 2. Predicted T cell epitopes from SARS-CoV2 ORF1ab polyprotein. 166 epitopes of SARS-CoV2 ORF1ab were found to have a 100% sequence identity with SARS-CoV which was curated in the Immune Epitope Database.

The predicted SARS-CoV2 ORF1ab-derived epitopes were further checked for the level of cross-conservation with the SARS-CoV-derived epitopes that have been curated in the Immune Epitope Database. We found that out of 945 predicted nonamer peptides, 166 cross-conserve with the SARS-CoV nonamer peptides that are reported as HLA binders in the Immune Epitope Database (Figure 2). The 166 nonamer peptides of SARS-CoV2 have the same amino acid sequences (100% sequence identity) with SARS-CoV epitopes. The exact sequence

similarity indicates that these epitopes are conserved and might be part of the important protein/enzyme of the coronavirus so that the mutation is highly unlikely. Moreover, the conserved epitopes are good candidates for the universal coronavirus vaccine component. Our result is in line with the recently reported analysis that calculated the percentage identity of the SARS-CoV2 predicted epitope with the SARS-CoV T cell epitope curated in IEDB. The study showed that the Orf1ab epitopes, were moderately conserved, in which 3 out of 7 SARS-CoV epitopes shared ≥ 85% sequence identity match with SARS-CoV2 epitopes (Grifoni et al., 2020).

The epitope conservancy was also observed between all 52 complete genomes of SARS-CoV2 housed in the Genbank (access on 5 March 2020), which indicates that the ORF1ab genome sequence is quite stable and make it a good candidate as vaccine target (Figure 3 left panel). Moreover, some of the SARS-CoV2 ORF1ab predicted epitopes were also cross-conserved with other ORF1ab coronaviruses (Figure 3 right panel). The epitopes that are conserved across all coronaviruses could become the best target for universal coronavirus vaccine for the pandemic. The results are in agreement with the published paper which stated that the ORF1ab gene of coronavirus encodes proteins and enzymes that are important for viral replications such as RdRp and helicase. The region ORF1b especially is considered as the most conserved region among all coronaviruses (St Jean et al., 2004).

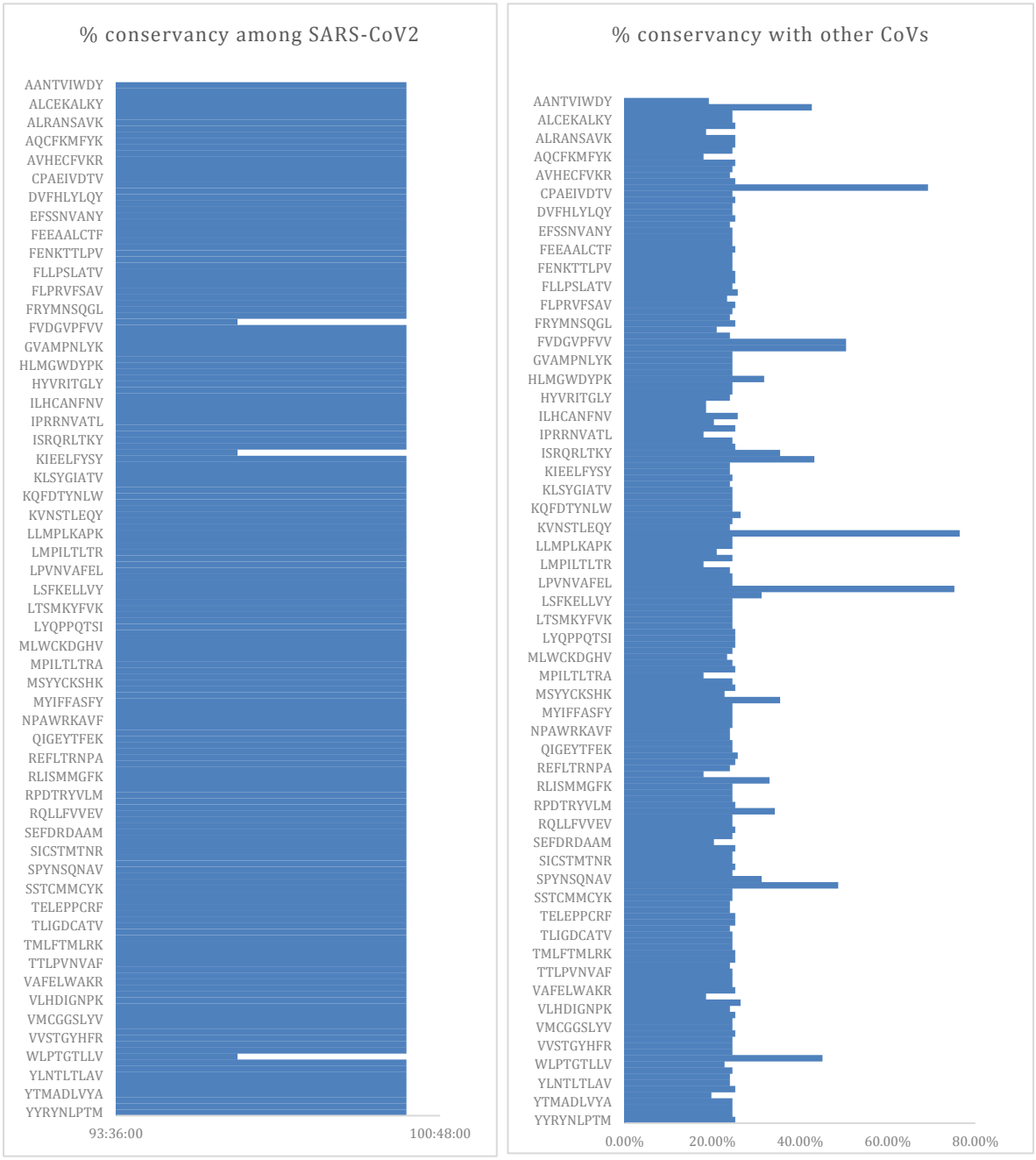

Figure 3. Epitopes cross-conservation among SARS-CoV2 (left) and all other coronaviruses (right). Almost all 166 peptides are conserved within SARS-CoV2 isolates (left). Five of the 166 peptides are cross-conserved with other coronaviruses (right): KYTQLCQYL, LPYPDPSRI, AYANSVFNI, FVSLAIDAY, and FVDGVPFVV.

The 166 predicted peptides were blasted against the human proteome to screen out the ones that mimic the human self-peptides. Using the Microsoft Excel program, the Blast results were further analyzed to identify peptides that have 100% identity with the peptides from the human proteome, with no gap or mismatches residues. The blast results containing predicted sequences that match the human peptides with fewer than six contiguous identical residues (≤ 6/9 similarity) were disregarded. The probability of peptides matching five or fewer residues was high and non-significant (Tan et al. 2010). However, the blast results containing predicted 9-mer peptides that share 8 contiguous amino acids with human self-peptides are significant. The 8-mer peptide is the shortest peptide that can

occupy the binding groove on the HLA molecules (Biddison et al., 2001), therefore, the peptides shared 8 contiguous amino acids with the human peptides were likely to induce an autoimmune or tolerogenic response if used as a vaccine.

In the analysis we found 18 peptides having sequence similarity with the human peptides (Table 1). Two peptides that shared 8 contiguous amino acid residues with the human peptides: FLLPSLATV sequences were found in the human protein hCG2040879 (ID: EAX04876.1) and NAAISDYDY were found in human protein immunoglobulin heavy chain junction region (ID: MON69829.1). Another 16 peptides were shown to have sequences that are similar to human peptides which shared 7 contiguous amino acids residue. The sequence which is similar to human peptide might induce the response of tolerogenic T cells which characterized as having the capacity to produce IL-10 (Bettini et al., 2009; Arce-Sillas et al., 2016). It has been demonstrated that SARS-CoV2 patients have very high levels of serum IL-10 following the SARS-CoV2 infection, while also displaying high levels of the PD-1 and Tim-3 exhaustion markers on their T cells, suggesting that IL-10 might be mechanistically responsible (Diao et al., 2020).

Table 1. Nonamer peptides from ORF1ab that cross-conserve with human self-peptides.

Position	ORF1ab SARS-CoV2 Peptides	Human proteins in which 7-8 contiguous amino acid sequences of SARS-CoV2 peptide sequences were found
476	I**LASFSAS**T	NP_001352622.1; AAI20873.1; NP_004659.2; AAC39568.2; AAC39568.2; AAL83560.1; EAW51985.1; 3TON_A; EAW78729.1
1146	AP**LLSAGIF**	NP_787117.3; EAW56016.1
3639	F**LLPSLATV**	EAX04876.1
3641	L**PSLATVA**Y	EAX04876.1; AAA53375.1; NP_004422.2; NP_001316019.1; 2X10_A; 2X11_A; ACF47642.1; 3MX0_A; 3MBW_A
3913	**SLLSVLL**SM	2X10_A; BAC06004.1; AAM43506.1; AAQ13618.1; NP_054901.1; BAD96900.1
4506	I**SRQRLTK**Y	Q96IX9.1
4634	**MPILTLTRA**	NP_001017403.1; BAB39854.1; NP_067649.2; AAQ88486.1; NP_001017404.1
4713	FP**PTSFGPL**	EAW55845.1
4839	N**AAISDYDY**	MON69829.1
4841	**AISDYDY**YR	MON69829.1
5073	**SSGDAT**TAY	AAL65133.2; ARD71238.1; ARD71237.1; NP_078966.2; EAW83999.1; EAW84001.1; EAW83998.1; EAW84002.1; EAW84003.1
5267	QE**YADVFHL**	MOP95148.1; MOQ08493.1
5268	E**YADVFHL**Y	MOP95148.1; MOQ08493.1
5455	K**LFAAETL**K	AAB28072.2; NP_005537.3; 3QGW_A; 4HCT_A; 3T9T_A; 3MIY_A; 3V5J_A; 4KIO_A; 1SM2_A
5772	**IVDTVSA**LV	AAI31729.1; NP_065796.2; AAH33169.1; AAH13593.1; 6NOW_A; 6NLY_A; AAI28176.1; 6NLQ_A
5863	S**EYDYVIF**T	AAB29125.1
6749	LL**LDDFVEI**	AAC64943.1; NP_001139325.1; BAG35744.1; AAC28380.1
6814	KM**QRMLLEK**	AAF78783.1; NP_061198.2; AAF28912.1

The predicted peptides were further analyzed for the population coverage using the IEDB epitope analysis tool (Bui et al., 2006). The analysis was done for all the 166 peptides and calculated for the coverage for the Javanese and Sundanese-Javanese population. The results showed that the 166 peptides set had good population coverage for both HLA Class I (100 %)and Class II (100 %). The peptide which has the lowest

population coverage of 0,98% is NMLRIMASL (presented by HLA-B*08:01 and HLA-B*48:01). The peptide with the largest population coverage of 96.62% is VMYMGTLSY which binds to most HLA class I and II alleles.

The predicted epitopes were rationally chosen to be incorporated into a multi-epitope vaccine. Out of 166 predicted peptides, we selected 5 peptides to be included in the design of the coronavirus draft vaccine construct (Table 2) based on the following criteria: fully conserve in all SARS-CoV2 isolates, cross-conserve with other coronaviruses (>60 %), bound to HLA class I as well as class II, and no similarity with human peptides. Based on the HLA data, this set of peptides, when used as the vaccine will be able to cover 94,13% of the Indonesian population. The vaccine developed need to be an intranasal vaccine, following the result of one challenge study in mice which showed that the protection against coronavirus infection is mediated by airway memory T cells after the mice were intranasally vaccinated (Zhao et al., 2016).

Table 2. List of the predicted epitopes from SARS-CoV2 ORF1ab.

Position	Epitopes	HLA alleles	Indonesian population coverage
6844	KYTQLCQYL	HLA-A*24:02; HLA-A*24:07; HLA-A*24:10	60,92%
5221	LPYPDPSRI	HLA-B*07:05; HLA-B*35:02; HLA-B*51:01; HLA-B*51:02; HLA-B*52:01; HLA-B*56:01; HLA-B*56:07	21,04%
5080	AYANSVFNI	HLA-A*24:02; HLA-A*24:07; HLA-A*24:10	60,92%
5251	FVSLAIDAY	HLA-A*01:01; HLA-A*26:01; HLA-A*29:01; HLA-A*34:01; HLA-B*15:02; HLA-B*15:21; HLA-B*35:01; HLA-B*35:05; HLA-B*35:30; HLA-DRB1*04:05	61,00%
4726	FVDGVPFVV	HLA-A*02:01; HLA-A*02:06	19,35%
		Epitope set	**94,13%**

CONCLUSION

This immunoinformatics study showed that ORF1ab polyprotein, which is the largest portion of coronavirus proteome and carries important function for virus replication, could be the source of conserve T cell epitopes. The 166 nonamer peptides from SARS-CoV2 ORF1ab fully cross-conserve with SARS-CoV and partially cross-conserve with other coronaviruses. The fact that T cell plays a crucial role in the clearance of SARS-CoV infection emphasizes the need to do a detailed analysis of T cell response to the current SARS-CoV2. The identification of what T cells see in SARS-CoV2, especially the T cell epitopes presented by HLA alleles of the Indonesian population is an important step in designing the SARS-CoV2 vaccine. The epitopes predicted in this study could become a starting point for such a study. The predicted epitopes need to be tested experimentally involving among other assays, the MHC peptide-binding assay and IFNγ ELISPOT assay using PBMCs from the SARS-CoV2 recovered patients from Indonesia..

REFERENCES

Arce-Sillas A, Álvarez-Luquín DD, Tamaya-Domínguez B, Gomez-Fuentes S, Trejo-García A, Melo-Salas M, Cárdenas G,

Rodríguez-Ramírez J, Adalid-Peralta L. (2016) Regulatory T Cells: Molecular Actions on Effector Cells in Immune Regulation, J Immunol Res. 2016:1720827.

Backert, L., & Kohlbacher, O. (2015). Immunoinformatics and epitope prediction in the age of genomic medicine. Genome medicine 7: 119.

Bettini, M., & Vignali, D. A. (2009). Regulatory T cells and inhibitory cytokines in autoimmunity. Current opinion in immunology 21(6): 612–618.

Biddison WE, Martin R (2001) Peptide binding motifs for MHC class I and II molecules. Curr Protoc Immunol, Appendix 1: Appendix 1I.

Bishajit Sarkar, Md. Asad Ullah, Fatema Tuz Johora, Masuma Afrin Taniya, Yusha Araf (2020) The Essential Facts of Wuhan Novel Coronavirus Outbreak in China and Epitope-based Vaccine Designing against 2019-nCoV, bioRxiv 2020.02.05.935072; doi: https://doi.org/10.1101/2020.02.05.935072.

Bo Diao, Chenhui Wang, Yingjun Tan, Xiewan Chen, Ying Liu, Lifeng Ning, Li Chen, Min Li, Yueping Liu, Gang Wang, Zilin Yuan, Zeqing Feng, Yuzhang Wu, Yongwen Chen, (2020) Reduction and Functional Exhaustion of T Cells in Patients with Coronavirus Disease 2019 (COVID-19), medRxiv, doi: https://doi.org/10.1101/2020.02.18.20024364.

Bui H. H, Sidney J, Dinh K, Southwood S, Newman M. J, Sette A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics 17:153.

Bui H. H,Sidney J, Li W, Fusseder N, Sette A., (2007) Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics 8(1):361.

Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Kesmir C, Peters B. (2013) Properties of MHC class I presented peptides that enhance immunogenicity. PloS Comp. Biol. 8(1):361.

Ethan Fast, Binbin Chen (2020) Potential T-cell and B-cell Epitopes of 2019-nCoV, bioRxiv 2020.02.19.955484; doi: https://doi.org/10.1101/2020.02.19.955484.

Fan YY, Huang ZT, Li L, Wu MH, Yu T, Koup RA, Bailer RT, Wu CY. (2009) Characterization of SARS-CoV-specific memory T cells from recovered individuals 4 years after infection. Arch Virol: 154(7):1093-9.

Gonzalez-Galarza FF, Takeshita LY, Santos EJ, Kempson F, Maia MH, Silva AL, Silva AL, Ghattaoraya GS, Alfirevic A, Jones AR and Middleton D (2015) Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. Nucleic Acid Research 28, D784-8.

Grifoni, Alba and Sidney, John and Zhang, Yun and Scheuermann, Richard H. and Peters, Bjoern and Sette, Alessandro, Candidate Targets for Immune Responses to 2019-Novel Coronavirus (nCoV): Sequence Homology- and Bioinformatic-Based Predictions. CELL-HOST-MICROBE-D-20-00119. SSRN: https://ssrn.com/abstract=3541361 or http://dx.doi.org/10.2139/ssrn.3541361.

Gustiananda M, Cox D, McMurtrey C, Jackson KW, Mojsilovic D, Bardet W, Schafer F,

Cate S, Yaciuk J, Kaabinejadian S and Hildebrand WH (2013). Motifs of the naturally processed peptides presented by HLA-A*24:07. Front. Immunol. Conference Abstract: 15th International Congress of Immunology (ICI). doi: 10.3389/conf.fimmu.2013.02.01033.

Hsueh-Ling Janice Oh, Samuel Ken-En Gan, Antonio Bertoletti & Yee-Joo Tan (2012) Understanding the T cell immune response in SARS coronavirus infection, Emerging Microbes & Infections 1:1, 1-6.

Huang C, Wang Y, Li X, et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 395(10223):497-506.

Janice Oh HL, Ken-En Gan S, Bertoletti A, Tan YJ. (2012) Understanding the T cell immune response in SARS coronavirus infection. Emerg Microbes Infect. 1(9):e23.

Jasper Fuk-Woo Chan, Kin-Hang Kok, Zheng Zhu, Hin Chu, Kelvin Kai-Wang To, Shuofeng Yuan & Kwok-Yung Yuen (2020) Genomic characterization of the 2019 novel human pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan, Emerging Microbes & Infections, 9:1, 221-236.

Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, Sette A, Peters B, Nielsen M. (2018) Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology. 2018 Jul;154(3):394-406.

Kaifu Gao, Duc Duy Nguyen, Rui Wang, Guo-Wei Wei (2020) Machine intelligence design of 2019-nCoV drugs, bioRxiv 2020.01.30.927889; doi: https://doi.org/10.1101/2020.01.30.927889.

Koh D, Sng J. (2010) Lessons from the past: perspectives on severe acute respiratory syndrome. Asia-Pacific journal of public health. 22(3 Suppl):132s-136s.

Kumar, S. (2020) Drug and Vaccine Design against Novel Coronavirus (2019-nCoV) Spike Protein through Computational Approach. Preprints 2020, 2020020071 (doi: 10.20944/preprints202002.0071.v1).

Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, Nielsen M. (2005) An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. Eur J Immunol. 35(8):2295-303.

Li, C. K., Wu, H., Yan, H., Ma, S., Wang, L., Zhang, M., Tang, X., Temperton, N. J., Weiss, R. A., Brenchley, J. M., Douek, D. C., Mongkolsapaya, J., Tran, B. H., Lin, C. L., Screaton, G. R., Hou, J. L., McMichael, A. J., & Xu, X. N. (2008). T cell responses to whole SARS coronavirus in humans. Journal of Immunology 181(8): 5490–5500.

Liu J, Sun Y, Qi J, Chu F, Wu H, Gao F, Li T, Yan J, Gao GF. (2010) The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. J Infect Dis. 15;202(8):1171-80.

Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, et al., (2020) Genomic characterisation and

epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding.Lancet, 6736, 1–10.

Moise, L., Cousens, L., Fueyo, J., & De Groot, A. S. (2011). Harnessing the power of genomics and immunoinformatics to produce improved vaccines. Expert opinion on drug discovery, 6(1), 9–15.

Nathan P. Croft, Stewart A. Smith, Jana Pickering, John Sidney, Bjoern Peters, Pouya Faridi, Matthew J. Witney, Prince Sebastian, Inge E. A. Flesch, Sally L. Heading, Alessandro Sette, Nicole L. La Gruta, Anthony W. Purcell, David C. Tscharke (2019) Most viral peptides displayed by class I MHC on infected cells are immunogenic, Proceedings of the National Academy of Sciences 116 (8) 3112-3117.

Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, Hui-Dong Chen, Jing Chen, Yun Luo, Hua Guo, Ren-Di Jiang, Mei-Qin Liu, Ying Chen, Xu-Rui Shen, Xi Wang, Xiao-Shuang Zheng, Kai Zhao, Quan-Jiao Chen, Fei Deng, Lin-Lin Liu, Bing Yan, Fa-Xian Zhan, Yan-Yi Wang, Gengfu Xiao, Zheng-Li Shi. (2020) Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin, bioRxiv 2020, 01.22.914952; doi: https://doi.org/10.1101/2020.01.22.914952.

Schietinger, A., & Greenberg, P. D. (2014). Tolerance and exhaustion: defining mechanisms of T cell dysfunction. Trends in immunology, 35(2), 51–60.

Sharmin, R., Islam, A.B.M.M.K. (2014) A highly conserved WDYPKCDRA epitope in the RNA directed RNA polymerase of human coronaviruses can be used as epitope-based universal vaccine design. BMC Bioinformatics 15: 161.

St-Jean, J. R., Jacomy, H., Desforges, M., Vabret, A., Freymuth, F., & Talbot, P. J. (2004). Human respiratory coronavirus OC43: genetic stability and neuroinvasion. Journal of virology, 78(16), 8824–8834.

Stranzl T, Larsen MV, Lundegaard C, Nielsen M (2010) NetCTLpan: panspecific MHC class I pathway epitope predictions. Immunogenetics 62:357-368.

Syed Faraz Ahmed, Ahmed A. Quadeer, Matthew R. McKay (2020) Preliminary identification of potential vaccine targets for 2019-nCoV based on SARS-CoV immunological studies, bioRxiv 2020.02.03.933226; doi: https://doi.org/10.1101/2020.02.03.933226.

Sylvester-Hvid C, Nielsen M, Lamberth K, Røder G, Justesen S, Lundegaard C, Worning P, Thomadsen H, Lund O, Brunak S, Buus S. (2004) SARS CTL vaccine candidates; HLA supertype-, genome-wide scanning and biochemical validation. Tissue Antigens. 63(5):395-400.

Tan PT, Heiny AT, Miotto O, Salmon J, Marques ET, Lemonnier F, August JT (2010) Conservation and diversity of influenza A H1N1 HLA-restricted T cell epitope candidates for epitope-based vaccines. PLoS One 5:e8754.

Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. (2018) The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 47(D1): D339-D343.

WHO COVID-19 situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

WHO MERS Global Summary and Assessment of Risk https://www.who.int/csr/disease/coronavirus_infections/risk-assessment-august-2018.pdf?ua=1

Xu, X., & Dang, Z. (2020). Promising Inhibitor for 2019-nCoV in Drug Development. https://doi.org/10.31219/osf.io/3hcm6.

Yuliwulandari R, Kashiwase K, Nakajima H, Uddin J, Susmiarsih TP, Sofro AS, Tokunaga K (2009) Polymorphisms of HLA genes in Western Javanese (Indonesia): close affinities to Southeast Asian populations. Tissue Antigens 73:46-53.

Zhao J, Zhao J, Mangalam AK, Channappanavar R, Fett C, Meyerholz DK, Agnihothram S, Baric RS, David CS, Perlman S. (2016) Airway Memory CD4(+) T Cells Mediate Protective Immunity against Emerging Respiratory Coronaviruses. Immunity 44(6):1379-91.

Zhavoronkov, Alex; Aladinskiy, Vladimir; Zhebrak, Alexander; Zagribelnyy, Bogdan; Terentiev, Victor; Bezrukov, Dmitry S.; et al. (2020): Potential 2019-nCoV 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.11829102.v1.

Ziebuhr J. (2005) The coronavirus replicase. Curr Top Microbiol Immunol. 287:57-94.