Cross-Reactivity Between SARS-CoV-2 Proteins and Proteins in Pneumococcal Vaccines May Protect Against Symptomatic SARS-CoV-2 Disease and Death

Robert Root-Bernstein, Ph. D., Professor of Physiology, Michigan State University, East Lansing, MI 48824 USA; rootbern@msu.edu

Abstract

A significant inverse correlation exists between rates of pneumococcal vaccination, at both national and local levels, and symptomatic cases of SARS-CoV-2 infection and death. No correlations exist to BCG, Hib, diphtheria-tetanus-pertussis, measles-mumps-rubella, or poliovirus vaccinations. This paper explored the possibility that pneumococcal vaccines contain antigens that might be cross-reactive with SARS-CoV-2 antigens and that such cross-reactive antigens are absent from other vaccines. Comparison of the glycosylation structures of SARS-CoV-2 with the polysaccharide structures of pneumococcal vaccines yielded no obvious similarities. However, while pneumococcal vaccines are primarily composed of capsular polysaccharides, they also contain about three percent protein contaminants, including the pneumococcal surface proteins PsaA, PspA and probably PspC. These proteins have very high degrees of similarity, using very stringent criteria, with several SARS-CoV-2 proteins including the spike protein, membrane protein and replicase 1a. Equivalently similarities were found at statistically significantly lower rates, or were completely absent, among the proteins in diphtheria, tetanus, pertussis, measles, mumps, rubella, and poliovirus vaccines. Appropriate data were not available for testing Hib and BCG similarities. Notably, PspA and PspC are highly antigenic and new pneumococcal vaccines based on them are currently in human clinical trials so that their effectiveness against SARS-CoV-2 disease is easily testable.

Keywords: COVID-19; SARS-CoV-2; pneumococcal; Streptococcus pneumoniae; vaccine; vaccination; cross-reactivity; similarity; protection

Introduction

In attempting to explain why people in some nations and locales are apparently much more susceptible to serious disease, a very significant inverse correlation has been found between rates of pneumococcal vaccination at both national and local population levels and rates of SARS-CoV-2 infections and death (Root-Bernstein, 2020). No correlation was found to the tuberculosis vaccine BCG (Bacillus Calmette Guerin), Haemophilus influenzae type B (Hib), diphtheria-tetanus-pertussis, measles-mumps-rubella, or poliovirus vaccinations. The purpose of this paper is to provide a possible mechanism for how pneumococcal vaccines might protect against SARS-CoV-2 while the other vaccines do not.

The hypothesis tested is that antigens in pneumococcal vaccines induce antibodies protective against SARS-CoV-2 by means of cross-reactivity with similar SARS-CoV-2 antigens. There are two types of antigens that might play such a role, one being the capsular polysaccharide antigens in current pneumococcal vaccines and the other the proteins that they contain. An extensive search for polysaccharide structures shared by SARS-CoV-2 glycosylated proteins (Watanabe, et al., 2020) and S. pneumoniae serotypes (Shajahan, et al., 2020) failed to identify any, so the search then shifted to possible protein similarities.
While current pneumococcal vaccines are composed primarily of capsular polysaccharides, they also contain one or both of two types of proteins. The polysaccharide component is never pure, generally containing around three percent of the cell surface proteins to which the polysaccharides are attached (WHO, 2010; Morais, et al., 2018; Lee, et al., 2020). Proteins identified in pneumococcal vaccines include pneumococcal surface protein A (PspA) and pneumococcal surface adhesin A (PsaA) (Yu, et al., 1999; Yu, et al., 2003). Because the presence of PsaA was identified only by immunological methods and PsaA cross-reacts strongly with an additional pneumococcal surface protein, PspC (also known as CbpA and SpsA) (Brooks, et al., 1999; Oggunniyi, et al., 2001), it is likely that PspC is also present in capsular polysaccharide-based pneumococcal vaccines. Additionally, pneumococcal conjugate vaccines covalently attach the polysaccharides to a modified diphtheria toxin protein called Cross-Reactive Material 197 (CRM197) which is also present in Hib and meningitis vaccines (Möginger, et al., 2016).

This study reports that SARS-CoV-2 proteins contain significant regions that mimic sequences within pneumococcal surface proteins but not CRM197 or proteins present in other vaccines.

METHODS

In order to ascertain whether PspA, PsaA, PspC and CRM197 have regions of significant similarity to SARS-CoV-2 proteins, LALIGN (at [www.expasy.org](https://www.expasy.org)) was employed to perform pair-wise protein comparisons. The parameters chosen were 20 best alignments to show; BLOSSUM80 (in order to maximize small, local similarities); E = 10; gap penalty of -10.0 (to maximize continuous sequence similarities). SARS-CoV-2 sequences were retrieved from [https://viralzone.expasy.org/8996](https://viralzone.expasy.org/8996) as HTML files or using the accession numbers from the UniProtKB database (UniProtKB accession numbers PODTC1-PODTC9). *Streptococcus pneumoniae* PspA, PsaA and PspC sequences were retrieved as accession numbers (provided in the Tables below) from the UniProtKB database. Because different streptococcal serotypes have slightly different versions of these proteins, several were randomly selected for each search and the sequences similarities displayed in FIGURE 1 are representative of several serotype results.

The LALIGN results were culled by applying the criterion that any sequence similarity reported must have a region containing at least six out of ten identities, where a pair of acceptable substitutions (as determined by the BLOSSUM80 program) counted as one identity. This criterion is based on a number of experimental studies involving the average length of peptide recognized by major histocompatibility receptors and T cell receptors (Rudensky, et al., 1991; Hemmer, et al., 2000; Ekeruche-Makinde, et al., 2013) and the degree of similarity between two antigens that is likely to induce cross-reactive immune responses (Cunningham, et al., 1989; Hemmer, et al., 2000; Root-Bernstein, 2009; Root-Bernstein and Podufaly, 2012; Root-Bernstein, 2014).

As controls for the LALIGN results, all SARS-CoV-2 proteins were used to search for similarities to bacterial proteins used in diphtheria, pertussis, and tetanus vaccines (TABLE 1) and viral proteins incorporated into the measles, mumps, rubella and polio vaccines. The only identified proteins in Hib and meningitis vaccines are CRM197 or meningococcal outer membrane complex protein, so these were also examined for similarities to SARS-CoV-2 proteins (TABLE 1). The same criteria used above to screen the results for sequences having at least six identities in a span of ten amino acids. Bacillus Calmette Guerin (BCG) vaccine (a version of *Mycobacterium bovis*) was not available in the UniProtKB complete proteome dataset and was therefore not subject to similarity searching.
Only results that met the six-in-ten antigenic-cross-reactivity criterion just stated as well as an E value equal to or less than 0.1 and a Watermann-Eggert (W-E) score of 60 or above are reported since these are statistically unlikely to have appeared by chance.

Two statistical tests were used to evaluate the results. A student’s T-test (https://www.usablestats.com/calcs/2samplet) was used to test the significance of the difference between the number of pneumococcal protein similarity matches satisfying the criteria listed above compared with the number of matches from other vaccine proteins satisfying the criteria. A chi squared test (https://www.graphpad.com/quickcalcs/chisquared2/) was used to determine the significance of the difference in the percent of protein pairs that had at least one significant similarity as compared with the number that had no similarities (36 possibilities for 9 SARS-CoV-2 proteins versus 4 streptococcal proteins; 288 possibilities for 9 SARS-CoV-2 proteins versus the 32 bacterial and viral proteins listed in TABLE 1).

RESULTS

Results of the similarity searches that satisfy the W-E score of 60 or greater, E values of 0.1 or less, and in which at least six identical amino acids are present in a sequence of ten, are presented in FIGURES 1 and 2. FIGURE 1 displays the similarities between nine SARS-CoV-2 proteins (PODTC1-9) and streptococcal proteins PsaA, PspA, PspC. Twenty significant similarities were observed, ten of which are indicated in the figure in bold type as sequences that repeat within pair of proteins. Note that a significant sequence similarity was also found between SARS-CoV-2 proteins and the S. pneumoniae GRAM positive anchor protein (Q8DRK2), which serves as an anchor site for capsular polysaccharides. It is not known at this time whether this protein is among those contaminating capsular polysaccharide preparations but because of its association with polysaccharide anchoring, it is likely to be such a contaminant of the polysaccharide material used in pneumococcal vaccines. Each of the four streptococcal proteins was tested against each of the SARS-CoV-2 proteins yielding 36 pairwise tests. Six of these combinations yielded one or more matches that satisfied all similarity criteria employed here.

No significant similarities were found between CRM197 and any SARS-CoV-2 protein or between meningococcal outer membrane protein complex and any SARS-CoV-2 protein. FIGURE 2 displays the results for the pairwise tests of the nine SARS-CoV-2 proteins with the additional bacterial and viral proteins listed in TABLE 1 that are present in measles, mumps, rubella, polio, diphtheria, pertussis, and tetanus vaccines, for a total of 32 microbial proteins. Of these, six yielded one or more significant similarities for a total of nine matches out of 288 possible pairwise combinations.

Statistical tests demonstrate that the results reported above are highly significant. All four of the pneumococcal proteins tested had significant similarities to at least one of the nine SARS-CoV-2 proteins. Altogether, six of the 36 possible permutations of pneumococcal protein pairs yielded significant similarities, or 16.9 percent. In contrast, only six of the 32 other viral and bacterial vaccine proteins tested had significant matches to any of the nine SARS-CoV-2 proteins (2.2% of the 288 pairwise comparisons). The four pneumococcal proteins yielded 21 significant matches, for an average of 5.25 per protein, while the 32 other vaccine proteins yielded only nine significant matches, for an average of 0.28 per protein. The t-test comparing the number of matches from FIGURES 1 and 2 yielded a T Test Statistic t= 3.8081 corresponding to a P value of 0.0002. The chi squared test comparing the 16.7% of protein comparisons that yielded at least one significant similarity from FIGURE 1 (6 categories
out of 36 pairwise comparisons) with the 2.2% (6 categories out of 288 pairwise comparisons) from FIGURE 2 yielded a chi squared value of 114.796 corresponding to a P value of 0.0001.

DISCUSSION

The Results of this study indicate that proteins known to contaminate pneumococcal vaccines significantly mimic SARS-CoV-2 proteins (FIGURE 1) while the CRM197 used to conjugate various polysaccharide vaccines and the proteins contained in other vaccines are statistically significantly less likely to do so (FIGURE 2). In particular, the Results point to potential cross-reactivity between SARS-CoV-2 proteins and the pneumococcal proteins PspA and PsaA, which are known to contaminate polysaccharide-based pneumococcal vaccines (WHO, 2010; Morais, et al., 2018; Lee, et al., 2020) as well as PspC, which it is reasonable to assume is another such contaminant since it derives from the same outer membrane protein complex and is highly cross-reactive with the antibodies against PspA used to demonstrate the presence of PspA in vaccines (Brooks, et al., 1999; Ogunniyi, et al., 2001).

The concentration of protein contaminants in pneumococcal vaccines is sufficient to induce immunity. In Prevnar-13, for example, there are 30.4 µg of capsular polysaccharides and 34.0 µg of CRM197 for a total of 64.4 micrograms of antigen per dose (FDA, 2017). Protein contaminants may make up an additional 3%, or 1.92 µg, of antigenic material according to WHO guidelines and confirmed by laboratory analysis (WHO, 2010; Morais, et al., 2018; Lee, et al., 2020). This 1.92 µg of protein is virtually identical to the 2.2 µg of each of twelve of the capsular polysaccharides present (plus 4.4 µg of serotype 6) or the 2.3 micrograms of CRM197 conjugated to each polysaccharide type (FDA, 2017) and is therefore quite sufficient to induce an immune response, especially since PspA and PspC are strongly cross-reactive. Pneumovax-23, in contrast, has 25 µg of each capsular polysaccharide, adding up to a total of 575 µg of antigen. The three percent protein contamination allowed by WHO (WHO, 2010; Morais, et al., 2018; Lee, et al., 2020) could result in 17.25 µg of PsaA, PspA and PspC per dose, which is certainly sufficient to induce immunity. For comparison, each 0.5-mL dose of Adacel®, a diphtheria-tetanus-pertussis vaccine (Sanofi Pasteur) contains 2.5 µg detoxified pertussis toxoid, 5 µg FHA, 3 µg pertactin and 5 µg FIM acellular pertussis antigens (CDC, 2020).

In addition to being present in concentrations that could induce protective immunity, the pneumococcal-SARS-CoV-2 similarities reported here satisfy multiple criteria involving sequence identities and statistical measures for predicting potential antigenic cross-reactivity so that it is possible that pneumococcal vaccination can protect individuals against SARS-CoV-2 disease. Evidence of protection against SARS-CoV-2 by T cells reactive to unidentified, cross-reactive microbes has been reported by Grifoni, et al. (2020). The study reports that 40- to 60% of people unexposed to SARS-CoV-2 had SARS-CoV-2-reactive CD4+ T cells. The assumption made by the authors is that the cross-reactivity is to coronaviruses that cause colds. However, the study also reports that this cross-reactive immunity is greatest in young people and least in older people, which is not consistent with cold virus exposures. Such waning immunity is, however, consistent with waning childhood vaccination immunity. In light of the data presented here, it is therefore possible that at least some proportion of individuals with cross-reactive immunity developed it through exposure to pneumococcal vaccinations. Such cross-reactivity would also explain the epidemiological observation that pneumococcal vaccination rates, but not vaccination rates with any other commonly used vaccine (DTP, MMR, polio, meningitis), correlate inversely with rates of serious SARS-CoV-2 disease and death (Root-Bernstein, 2020).
The observation that viral and bacterial proteins exhibit antigens similar enough to be cross-reactive may be surprising but it is not novel. Härkönen, et al. (2000) found that rabbit antibodies to HSP65 of *Mycobacterium bovis* (from which BCG is derived) recognized capsid protein VP1 of coxsackievirus A9, VP1, and/or VP2 of coxsackievirus B4. Misko, et al. (1999) demonstrated that Epstein-Barr virus mimicked a *Staphylococcus aureus* replication initiation protein and induced antibodies cross-reactive with it. Trama, et al., (2014) and Williams, et al. (2015) have documented antibodies against the gp41 protein of human immunodeficiency virus that cross-react with commensal bacteria in the human gut. Ross, et al. (1990) reported that sera from chickens inoculated with infectious bursal disease viruses or infectious bursal disease vaccines cross-reacted with *Mycoplasma gallisepticum* and *Mycoplasma synoviae*. And Bordenave (1973) found that antibodies against *Salmonella abortusequii* also recognized tobacco mosaic virus. In short, while the phenomenon may be rare – and, indeed, the data reported here suggests that such similarities may occur at a rate of about 1/200 pairwise combinations -- bacterial antigens are known to occasionally induce antibodies that cross-react with viral antigens or vice versa.

The rarity of antigenic mimicry between bacteria and viruses is emphasized by the finding that there are no significant similarities between SARS-CoV-2 proteins and the vaccine conjugate proteins, CRM197 and meningococcal outer membrane protein. This lack of antigenic mimicry suggests that neither protein is likely to contribute to the possible SARS-CoV-2 protection associated with conjugated pneumococcal, Hib or meningococcal vaccines. This negative result is consistent with the absence of any epidemiological association between Hib and meningococcal vaccines and SARS-CoV-2 rates of disease or death (Root-Bernstein, 2020).

The almost completely negative results reported here for antigenic mimicry between SARS-CoV-2 proteins and proteins from poliovirus, measles, mumps, diptheria, pertussis and tetanus are also consistent with the lack of association between these vaccines and SARS-CoV-2 rates of disease or death (Root-Bernstein, 2020). Franklin, et al., (2020) report “significant” similarities between both rubella and measles proteins and SARS-CoV-2, the key results or which were independently reproduced here in FIGURE 2. However, there are significantly fewer similarities between measles and rubella proteins and those of pneumococcal proteins and epidemiological evidence does not support measles containing vaccines (which often include rubella) as protective against SARS-CoV-2 (Root-Bernstein, 2020). The suggestion that polio vaccine be tested as a SARS-CoV-2 (Chumakov and Gallo, 2020) is likewise not supported by the data presented here (FIGURE 2) or by epidemiological data (Root-Bernstein, 2020).

Tuberculosis (BCG) vaccination has also been proposed to be epidemiologically associated with protection against SARS-CoV-2, but unfortunately, because the Bacillus Calmette Guerin (BCG) proteome is not in the UniProtKB database, it was not possible to perform the type of LALIGN analysis of BCG vaccine that was performed for the other vaccines. BCG vaccination was found to be associated with SARS-CoV-2 protection in one epidemiological study (Miller, et al., 2020) but not in another (Root-Bernstein, 2020) so that the current literature is divided on the issue and further research will be needed.

The observation of an inverse association of pneumococcal vaccinations with rates of SARS-CoV-2 rates of disease and death makes sense in terms of the particular protein contaminants of pneumococcal vaccines that are identified in this study as being potentially protective. These are PspA, PsAa and PspC, the specific regions identified in this study being known to be highly antigenic (van de
Garde, et al., 2019). Moreover, these proteins are under active investigation as more effective and broadly protective pneumococcal vaccine components to replace the polysaccharide-based vaccines (Briles, et al, 2000; Ferreira, et al., 2009; Schachern, et al., 2014; Lagousi, et al., 2019). Some of these vaccine candidates are already in human trials (Lagousi, et al., 2019; Masomiam, et al., 2020). Thus, it should be possible rapidly to determine whether such pneumococcal protein-based vaccines are effective deterrents to SARS-CoV-2 disease and these vaccines may provide needed protection until a SARS-CoV-2 vaccine is produced in sufficient quantities to be effective worldwide.

**Acknowledgements:** This research was not funded.

**Conflicts of Interest:** The author declares no conflicts of interest.

**References**


Kamboj K, King C, Greensapn N, Kirchner L, Schreiber J. Immunization with Haemophilus influenzaeType b-CRM197 Conjugate Vaccine Elicits a Mixed Th1 and Th2 CD4+ T Cell Cytokine Response That Correlates with the Isotype of Antipolysaccharide Antibody. Journal of Infectious Diseases. 184(7):931–935, OCTOBER 1, 2001 DOI: 10.1086/323342


Root-Bernstein RS, Podufaly A. 2012. Autoreactive T-cell receptor (Vbeta/D/Jbeta) sequences in diabetes recognize the insulin receptor, and each other, and are targets of insulin antibodies. Open Autoimmunity J., 4: 10-22, DOI: 10.2174/1876894601204010010


WHO (2009). Recommendations to Assure the Quality, Safety and Efficacy of Pneumococcal Conjugate Vaccines. Geneva Available online at: http://www.who.int/biologicals/areas/vaccines/pneumo/Pneumo_final_23APRIL_2010.pdf?ua=1


TABLE: UniProtKB accession numbers for viral and bacterial proteins used in this study.

**Mumps**
P11235|HN_MUMPM (HN)RecName: Full=Hemagglutinin-neuraminidase
P30929|L_MUMPM (L)RecName: Full=RNA-directed RNA polymerase L
P09458|FUS_MUMPR (F)RecName: Full=Fusion glycoprotein F0
P30928|V_MUMPM (P/V)RecName: Full=Non-structural protein V
P22112|SH_MUMPM (SH)RecName: Full= Small hydrophobic protein

**Measles**
P08362|HEMA_MEASE (H)RecName: Full=Hemagglutinin glycoprotein
Q89933|NCAP_MEASF (N)RecName: Full=Nucleoprotein
P12576|L_MEASE (L)RecName: Full=RNA-directed RNA polymerase L
Q786F3|FUS_MEASC (F)RecName: Full=Fusion glycoprotein F0
P0C774|V_MEASC (P/V)RecName: Full=Non-structural protein V

**Rubella**
P08563|POLS_RUBVM RecName: Full=Structural polyprotein
Q86500|POLN_RUBVM RecName: Full=Non-structural polyprotein p200

**Poliovirus**
P03301|POLG_POL1S RecName: Full=Genome polyprotein; 2209 aa CONTAINS:
RecName: Full=P3;
RecName: Full=Protein 3AB;
RecName: Full=P1;
RecName: Full=Capsid protein VP0;
RecName: Full=Capsid protein VP4;
RecName: Full=Capsid protein VP2;
RecName: Full=Capsid protein VP3;

**Pertussis**
P04977|TOX1_BORPE (ptxA)RecName: Full=Pertussis toxin subunit 1
P04978|TOX2_BORPE (ptxB)RecName: Full=Pertussis toxin subunit 2
P04979|TOX3_BORPE (ptxC)RecName: Full=Pertussis toxin subunit 3
P0A3R5|TOX4_BORPE (ptxD)RecName: Full=Pertussis toxin subunit 4
P04981|TOX5_BORPE (ptxE)RecName: Full=Pertussis toxin subunit 5
P35077|FHAC_BORPE (fhaC)RecName: Full=Filamentous hemagglutinin transporter protein FhaC
P14283|PERT_BORPE (prn)RecName: Full=Pertactin autotransporter
P05788|FM2_BORPE (fim2)RecName: Full=Serotype 2 fimbrial subunit
P17835|FM3_BORPE (fim3)RecName: Full=Serotype 3 fimbrial subunit

**Tetanus**
P04958|TETX_CLOTE (tetX)RecName: Full=Tetanus toxin

**Diphtheria**
Q5PY51|Q5PY51_CORDP SubName: Full=Diphtheria toxin;[Corynebacterium diphtheriae]

**Meningococcus**
Multiple proteins have been identified as potential targets for vaccine development. For example, recombinant major outer membrane protein (0DH58|OMPA_NEIMB) and diphtheria toxin (CRM197|Q6NK15_CORDI) have been studied.

- **SARS-CoV-2**
  - PODTC1 Replicase polyprotein 1a (pp1a)
  - PODTC2 Spike glycoprotein (S)
  - PODTC3 Protein 3a (NS3a)
  - PODTC4 Envelope small membrane protein (E)
  - PODTC5 Membrane protein (M)
  - PODTC6 Non-structural protein 6 (NS6)
  - PODTC7 Protein 7a (NS7a)
  - PODTC8 Non-structural protein 8 (NS8)
  - PODTC9 Nucleoprotein (N)

Pneumococcal proteins PsaA, PspA, PspC and Gram-positive anchor protein have multiple variants; accession numbers of representative variants are provided in FIGURE 1.
FIGURE 1: Similarities between the four known or probable pneumococcal vaccine protein contaminants PsaA, PspA, PspC and Gram-positive anchor protein and SARS-CoV-2 proteins. Multiple variants for each protein were examined and results provided here are representative of results but each serotype presents slightly different matches. 36 pairs of proteins were searched Only similarities satisfying criteria laid out in Methods are shown.

COVID-19 Replicase 1AB 7096 aa vs S. pneumoniae pspA O34097 653 aa
Waterman-Eggert score: 100; 35.6 bits; E(1) < 8.7e-05

SP pspA O34097 EKERKASEKIAEATKEVQQAYLAYLQASNESQRKEADKKIK

COVID Rep1A KSEKQVEQKIAEIPKEEVKPFITESKPSVE-QRKQDDKKIK

COVID-19 Spike Protein 1273 aa vs. S. pneumoniae pspA Q9LAZ1 395 aa
Waterman-Eggert score: 60; 23.2 bits; E(1) < 0.05

SP pspA Q9LAZ1 PLQSKLDTJKAKLSK

COVID SP PLQPELDSKIFEELDK

COVID-19 Spike Protein 1273 aa vs. S. pneumoniae pspA B2IRK1 609 aa
Waterman-Eggert score: 62; 23.9 bits; E(1) < 0.049

SP pspA B2IRK1 QAKIAELENQVHRLEQDLKDINES

COVID SP NASVVNIQEIDRLNEVAKNLNES

COVID-19 Nucleoprotein P59595 422 aa vs. S. pneumoniae pspA Q9LAY4
Waterman-Eggert score: 72; 22.7 bits; E(1) < 0.027

SP pspA Q9LAY4 QKAFLILREAQEQLSKRPNNKTKAQQ

COVID NP QPQGTQTVKKSAEAASKPKRQKRTATKQ

COVID-19 Nucleoprotein 417 aa vs. S. pneumoniae pspA Q9LAZ1
Waterman-Eggert score: 67; 22.1 bits; E(1) < 0.037

SP pspA Q9LAZ1 QLKLKKYLDGRNLSNSSLKKEEAAEKKDEKQ

COVID NP QLESKMSGKSOQQGQVTVKSSAAEASKPKRQKR

COVID-19 Spike Protein 1273 aa vs. S. pneumoniae psaA 309 aa
Waterman-Eggert score: 76; 22.7 bits; E(1) < 0.0025

SP psaA P0A4G2 CASGKKTTSGQKLKVVATNSIA

COVID SP CASYQTQNPPRRARSVASQSIQA

COVID-19 Replicase 1AB 7096 aa vs. S. pneumoniae psaA 310 aa
Waterman-Eggert score: 62; 22.7 bits; E(1) < 0.028

SP psaA P42363 FTKLVRNANKVNKENFVAAASDGVEV

COVID Rep1AB LNKATNMAMQVESDDYIATNGPLKV

12
COVID-19 Replicase 1AB 7096 aa vs \textit{S. pneumoniae} pspC 792 aa
Waterman-Eggert score: 98; 32.6 bits; E(1) < 0.00053

COVID19 Repla EIPKEEVKPFITESKPSVEQRKQDDKK
| || |||| : : || : : |||
SP pspC F2WN4 EKPKEVKPQLEKPKPDNSKPKQADDKK
710 720 730

COVID-19 Replicase 1AB 7096 aa vs \textit{S. pneumoniae} pspC 792 aa
Waterman-Eggert score: 79; 26.9 bits; E(1) < 0.027

COVID19 Repla EIPKEEVKPFITESKPSVEQRKQDDKK
| || |||| : : || |: : |
SP pspC F2WN4 EKPKEVKPQLEKPKPQPEKPK
660 670 680

\textbf{(also 653-671)}

COVID-19 Replicase 1AB 7096 aa vs \textit{S. pneumoniae} pspC 792 aa
Waterman-Eggert score: 77; 26.3 bits; E(1) < 0.039

COVID19 Repla EIPKEEVKPFITESKPSVE
| || |||| : / || |:
SP pspC F2WN4 EKPKEVKPQLEKPKPQPEKPK
670 680

\textbf{(also 565-570; 576-601; 587-612; 598-623; 609-634; 620-645; 631-656; 681-707)}

COVID19 Spike Protein 1273 bp vs. \textit{S. pneumoniae} Gram-positive anchor protein Q8DRK2 1161 aa
Waterman-Eggert score: 72; 24.9 bits; E(1) < 0.1

COVID19 Repla EIPKEEVKPFITESKPSVEQRKQDDKK
| || |||| / || |: \ : |
SP pspC F2WN4 EKPKEVKPQLEKPKPQPEKPK
560 570

\textbf{(also 565-590; 576-601; 587-612; 598-623; 609-634; 620-645; 631-656; 681-707)}

COVID19 SPIKE PROT GKADLTLVATKNDINIGL
| \ |||| : / |: | |||

\textbf{Preprints (www.preprints.org) \mid NOT PEER-REVIEWED \mid Posted: 8 July 2020}
doi:10.20944/preprints202007.0141.v1
FIGURE 2: Similarities between nine SARS-CoV-2 proteins and 32 proteins from measles, mumps, rubella, polio, Hib, meningitis, diphtheria, pertussis and tetanus vaccines (TABLE 1). 288 pairwise combinations were searched. Only similarities satisfying criteria laid out in Methods are shown.

RUBELLA NON-STRUCTURAL PROT Q86500 2116 aa vs. COVID19 REPL1a 4405 aa
Waterman-Eggert score: 83; 29.9 bits; E(1) < 0.009

RUB NSF Q86500 VVVNAANEGLLLAGSGVCGAI
COVID19 REPL1a VVVNAANVYLKGHHGGVAGAL

RUBELLA NON-STRUCTURAL PROT Q86500 2116 aa vs. COVID19 REPL1a 4405 aa
Waterman-Eggert score: 73; 26.8 bits; E(1) < 0.076

RUB NSF Q86500 PLLGAGVGWSAAEASLRAAATR
COVID19 REPL1a PLLSAGIFGAPIHSLRVCDTVR

MEASLES HEMA P08362 617 aa vs. COVID19 REPL1a 4405 aa
Waterman-Eggert score: 70; 25.8 bits; E(1) < 0.045

MEASLE HEM P08362 YVLATYDTSRVEHAVVYVSFS
COVID19 REPL1a YVLPNDDTLRVEAFYHTDPS

MEASLES FUSION GLYCOPROTEIN Q786F3 550 aa vs. COVID19 PROT REPL1A 4405 aa
Waterman-Eggert score: 65; 24.5 bits; E(1) < 0.097

MEASLES FGP Q786F3 IVYILIAVCLGLLI
COVID19 REPL1A IWFLLLSVCGLSLI

PERTUSSIS TOXIN 1 P04977 269 aa vs. COVID19 REPL1a 4405 aa
Waterman-Eggert score: 69; 25.3 bits; E(1) < 0.029

PERT TOX1 P04977 YTSRRSVASIVGTL
COVID19 REPL1a YTSKTTVASLINTL

PERTUSSIS TOXIN 4 P0A3R5 152 aa vs. COVID19 Repl1a 4405 aa
Waterman-Eggert score: 62; 23.2 bits; E(1) < 0.069

PERT TOX4 P0A3R5 FPTRTAPQGGARRSRVRALLLAS
COVID19 REPL1A FVDRQTAQAGTDTTITVNVNLALYAA

TETANUS TOXOID P04958 1315 aa vs. COVID19 PROTEIN 3a 275 aa
Waterman-Eggert score: 64; 23.7 bits; E(1) < 0.026

TETANUS TOX P04958 NPLRYDTEYLY
COVID19 PROT 3a NPLLYDANYFL

140