

# Correlated Mutations of Positions Among Structural Proteins in Delta and Omicron Variants for SARS-CoV-2 Amino Acid Sequences

Yuichi Shimaya and Kouich Hirata

*Kyushu Institute of Technology, Kawazu 680-4, Iizuka 820-8502, Japan*

**Keywords:** Correlated Mutation, Structural Proteins, Spike Protein Substitution, Amino Acid Sequences, SARS-CoV-2, Delta Variant, Omicron Variant.

**Abstract:** In this paper, we find the *correlated mutations* of positions among structural proteins of spike, envelop, membrane and nucleocapsid proteins in amino acid sequences of SARS-CoV-2. Here, we adopt the algorithm designed by Shimada *et al.* (2012) of finding the correlated mutations formulated by joint entropy. In particular, we discuss whether or not the found correlated mutations contains spike protein substitutions in SARS-CoV-2 Delta and Omicron variants.

## 1 INTRODUCTION

Coronavirus disease 19 (COVID-19) is caused by a novel coronavirus designated as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The SARS-CoV-2 is an enveloped virus with a positive-sense, single-stranded RNA genome. Figure 1 illustrates the schematic presentation of the SARS-CoV-2 genome organization, the canonical subgenomic mRNAs and the virion structure (Kim *et al.*, 2020).

Coronaviruses (CoVs) were named because they are characterized by spike protein projections on the surface of the viral particles, and their shape resembles a crown (corona) under electron microscopy, the virion structure in Figure 1.

The CoVs carry the largest genomes among all RNA virus families. The genomic RNA is translated to produce nonstructural protein (nsps) from two open reading frames (ORFs), ORF1a and ORF1b. ORF1a and ORF1b encode 11 and 5 nonstructural proteins; nsp1 to nsp11 and nsp12 to nsp16, respectively. Whereas ORF1a is translated directly from the genomic RNA, the expression of ORF1b requires a -1 ribosomal frameshift near the end of ORF1, resulting in a single ORF1ab polypeptide.

Downstream from the ORF1ab, there exist ORFs encoding a few or more than 10 structural/nonstructural proteins. The common structural proteins of CoVs are spike (S), envelope (E), membrane (M) and nucleocapsid (N) proteins. SARS-CoV-2 is also known to have at least 6 accessory proteins, ORF3a, ORF6, ORF7a, ORF7b, ORF8 and ORF10. However, the ORFs have not yet been exper-

imentally verified for expression. Therefore, it is currently unclear which accessory genes are actually expressed from this compact genome (Kim *et al.*, 2020).

On the other hand, since the *correlated mutations* of positions in amino acid sequences are frequently observed among spatially close residues and valuable for analyzing the structure of proteins, many researchers (*cf.* (Jeong and Kim, 2010; Lee and Kim, 2009; Oliveria *et al.*, 2002)) have developed to find them. In their works, the correlated mutations are formulated by using an entropy (Oliveria *et al.*, 2002), a CM-score (Lee and Kim, 2009) or log-odds scores (Jeong and Kim, 2010).

In this paper, we adopt the correlated mutations of positions based on a *joint entropy ratio* introduced by Shimada *et al.* (Shimada *et al.*, 2012). Here, the joint entropy ratio of positions is the ratio of the minimum entropy in each position in them to the joint entropy (*cf.*, (Cover and Thomas, 2006)) of them. Then, we say that positions are *correlated* (Shimada *et al.*, 2012) if the joint entropy ratio of them is greater than a given *joint entropy threshold*  $\tau$  ( $0 < \tau \leq 1$ ). Furthermore, based on their correlated mutations, they have designed the algorithm FINDCM<sup>1</sup> of finding all of the correlated mutations with the joint entropy threshold  $\tau$  and an *exclusion threshold*  $\rho$  ( $\rho > 0$ ) to exclude the positions in which elements do not change well.

Recently, Yonashiro *et al.* (Yonashiro *et al.*, 2022) have applied the algorithm FINDCM to find the correlated mutations of positions for structural proteins of

<sup>1</sup>The algorithm FINDCM in this paper coincides with an algorithm named by FINDCM2 in (Shimada *et al.*, 2012).

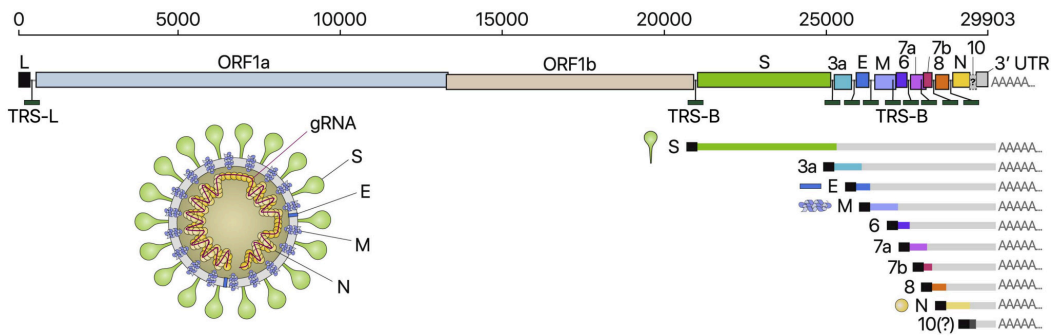


Figure 1: Schematic presentation of the SARS-CoV-2 genome organization, the canonical subgenomic mRNAs and the virion structure (Kim et al., 2020).

S, E, M and N proteins from amino acid sequences of SARS-CoV-2, by focusing on the positions observed as *spike protein substitutions* for SARS-CoV-2 Delta variant reported from CDC<sup>2</sup> and NIIB<sup>3</sup>.

Here, SARS-CoV-2 Delta and Omicron variants are classified to VOCs (Variant of Concerns) from June in 2021 to April in 2022 and from November in 2021 to today, respectively, from CDC. Here, a VOC is a variant for which there is evidence an increase in transmissibility, more severe disease, significant reduction in neutralization by antibodies generated during previous infection or vaccination, reduced effectiveness of treatments or vaccines, or diagnostic detection failures<sup>2</sup>.

Their variants are characterized as the *spike protein substitutions*. Table 1 illustrates the spike protein substitutions for SARS-CoV-2 Delta and Omicron variants. Here,  $a_1na_2$  denotes that an amino acid  $a_1$  at the position  $n$  is substituted to another amino acid  $a_2$ . Also “(\*)” denotes the substitution detected in some sequences but not all. Furthermore, “del211” denotes that the amino acid at the position 211 is deleted, “del69-70” denotes that the amino acids from the position 69 to 70 are deleted and “ins214EPE” denotes that the amino acids EPE are inserted from the position 214 to 215.

Hence, in this paper, after finding the correlated mutations obtained by the algorithm FINDCM under  $\tau$  and  $\rho$ , we discuss whether or not the found correlated mutations among all of the structural proteins contains spike protein substitutions in not only SARS-CoV-2 Delta variant and but also SARS-CoV-2 Omicron variant. Note that the results for the Delta variant are updated from the previous work (Yonashiro et al., 2022) by using newly amino acid sequences.

<sup>2</sup>CDC, Centers for Disease Control and Prevention. <https://www.cdc.gov>.

<sup>3</sup>NIIB, National Institute of Infectious Diseases. <https://www.niid.go.jp/niid/ja>.

Table 1: The spike protein substitutions for SARS-CoV-2 Delta and Omicron variants.

Delta variant				
T19R	(V70F*)	T95I	G142D	del156
del157	R158G	(A222V*)	(W258L*)	(K417N*)
L452R	T478K	D614G	P681R	D950N
Omicron variant				
A67V	del69-70	T95I	del142-144	Y145D
del211	L212I	ins214EPE	G339D	S371L
S373P	S375F	K417N	N440K	G446S
S477N	T478K	E484A	Q493R	G496S
Q498R	N501Y	Y505H	T547K	D614G
H655Y	N679K	P681H	N764K	D796Y
N856K	Q954H	N969K	L981F	

## 2 FINDING CORRELATED MUTATION

In this section, we introduce the algorithm FINDCM designed by (Shimada et al., 2012).

Let  $\Sigma$  be an alphabet and  $\Sigma^n$  the set of all strings on  $\Sigma$  with length  $n$ . Also let  $[n]$  be  $\{1, \dots, n\}$ . For a string  $w \in \Sigma^n$  and a set  $I = \{i_1, \dots, i_k\} \subseteq [n]$  ( $1 \leq i_1 < \dots < i_k \leq n$ ), we denote the string  $w[i_1] \dots w[i_k]$  constructed from concatenating the symbols at from  $i_1$  to  $i_k$  by  $w[I]$ . Furthermore, let  $P(w[I] = v)$  be the probability that  $w[I]$  is  $v \in \Sigma^k$ .

**Definition 1.** Let  $S \subseteq \Sigma^n$  and  $I \subseteq [n]$ . Then, we define the *joint entropy*  $H_S(I)$  of  $S$  at a set  $I$  of positions (cf., (Cover and Thomas, 2006)) as:

$$H_S(I) = - \sum_{v \in \Sigma^{|I|}, w \in S} \left( P(w[I] = v) \times \log P(w[I] = v) \right).$$

An *entropy of  $S$  at a position  $i \in [n]$*  coincides with  $H_S(\{i\})$  (cf., (Cover and Thomas, 2006)), which we denote by  $H_S(i)$ . Then, we define the *joint entropy ratio*  $R_S(I)$  of  $S$  at a set  $I$  of positions as:

$$R_S(I) = \frac{\min\{H_S(i) \mid i \in I\}}{H_S(I)}.$$

For the joint entropy ratio, the following lemma holds.

**Lemma 1.** (Shimada et al., 2012) *Let  $S \subseteq \Sigma^n$  and  $I, J \subseteq [n]$ . If  $I \subseteq J$ , then  $H_S(I) \leq H_S(J)$ .*

Then, we formulate the *correlated mutation* by introducing the *joint entropy threshold* as the main topic in this paper.

**Definition 2.** (Shimada et al., 2012) *Let  $S \subseteq \Sigma^n$  and  $I \subseteq [n]$ . Also let  $0 < \tau \leq 1$  be a *joint entropy threshold*. Then, we say that  $I$  is *correlated* if  $R_S(I) \geq \tau$ . Furthermore, we call the mutations at  $I$  in  $S$  *correlated mutations* of  $S$ .*

When setting  $\tau$  to 1, that is,  $R_S(I) = 1$ , the correlated mutations are *exact*. On the other hand, when setting  $\tau$  less than 1, that is,  $R_S(I) < 1$ , the correlated mutations contains exception. Furthermore, we use another threshold  $\rho \geq 0$ , called an *exclusion threshold*, to exclude the positions in which elements do not change well.

The algorithm FINDCM (Shimada et al., 2012) in Algorithm 1, which is based on the set enumeration algorithm (Rymon, 1992), describes the algorithm of finding all of the correlated mutations with an *entropy pruning* at lines from 5 to 7. Here, the correctness of the entropy pruning follows from the following theorem.

**Theorem 1.** (Shimada et al., 2012) *Let  $S \subseteq \Sigma^n$  and  $I \subseteq [n]$ . If  $R_S(I) \geq \tau$ , then it holds that  $\tau H_S(i) \leq H_S(j) \leq H_S(i)/\tau$  for every  $i, j \in I$ .*

### 3 EXPERIMENTAL RESULTS

In this section, by focusing on the structural proteins, that is, spike (S), envelope (E), membrane (M) and nucleocapsid (N) proteins, we give experimental results of finding correlated mutations among them.

We use amino acid sequences of the four structural proteins for a Delta variant and an Omicron variant provided from NCBI<sup>4</sup>, whose length of amino acid sequences for S, E, M and N proteins is 1,273, 75, 222 and 419, respectively (Nakagawa and Miyazaki, 2020). The amino acid sequences of the Delta variant are stored from April 1 in 2021 to May 31 in 2022 and the number of sequences is 834,870. On the other hand, the amino acid sequences of the Omicron variant are stored from October 1 in 2021 to July 29 in 2022 and the number of sequences is 393,655.

We use amino acids as 20 characters of A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W and Y, so the

<sup>4</sup>NCBI, National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov>.

```

procedure FINDCM( $S, \tau, \rho$ )
  /*  $S \subseteq \Sigma^n, 0 < \tau \leq 1, \rho \geq 0$  */
  1  $C \leftarrow \{1, \dots, n\};$ 
  2 for  $i = 1$  to  $n$  do
  3   if  $H_S(i) \geq \rho$  then
  4      $C \leftarrow C \setminus \{i\}; C_{next} \leftarrow C;$ 
  5     foreach  $j \in C_{next}$  do
  6       if  $H_S(j) \geq \rho$  and  $(H_S(j) < \tau H_S(i)$ 
  7         or  $H_S(i)/\tau < H_S(j))$  then
  8            $C_{next} \leftarrow C_{next} - \{j\};$ 
  9     EXPAND( $S, \{i\}, C_{next}, \tau$ );

procedure EXPAND( $S, I, C, \tau$ )
  /*  $I, C \subseteq \{1, \dots, n\}$  */
  9  $C_{next} \leftarrow C;$ 
  10 foreach  $i \in C$  do
  11    $C_{next} \leftarrow C_{next} \setminus \{i\};$ 
  12   if  $R_S(I \cup \{i\}) \geq \tau$  then
  13     output  $I \cup \{i\};$ 
     EXPAND( $S, I \cup \{i\}, C_{next}, \tau$ );

```

Algorithm 1: FINDCM.

set of the above characters is regarded as  $\Sigma$ . Also we insert a gap symbol ‘-’ at the position where an amino acid is deleted. Furthermore, since a character ‘X’ is inserted if an amino acid is unknown in the amino acid sequences provided from NCBI<sup>4</sup>, we replace ‘X’ at a position as the most frequent amino acid at the position in all the amino acid sequences.

#### 3.1 Running Time

In this subsection, we investigate the running time of the algorithm FINDCM. Here, the computer environment is that OS is WSL2 over Windows 10, CPU is Intel Core i7-7700 3.6GHz and RAM is 16GB.

Table 2 and Table 3 illustrate the running time of the algorithm FINDCM for amino acid sequences of the Delta variant and the Omicron variant, respectively.

Tables 2 and 3 show that, for  $0.80 \leq \tau \leq 0.90$ , the computation time of the Delta variant is smaller than that of the Omicron variant for  $\rho = 0.10$  and  $0.05$ , the computation time of the Delta variant is much larger than that of the Omicron variant for  $\rho = 0.01$ . Note that the number of sequences for the Delta variant is 834,870 and that for the Omicron variant is 393,655, so the former is more than twice the latter. Hence, by comparing the computation time for FINDCM with the number of the sequences, the number of candidates, not pruned by  $\tau$  and  $\rho$  in FINDCM, of the correlated mutations for the Omicron variant is much larger than that for the Delta variant for small  $\rho$ .

Table 2: The running time (sec.) of the algorithm FINDCM for amino acid sequences of the Delta variant.

$\tau$	$\rho$	time (sec.)
0.90	0.10	56
	0.05	74
	0.01	1,366
0.85	0.10	89
	0.05	114
	0.01	2,089
0.80	0.10	144
	0.05	173
	0.01	2,817

Table 3: The running time (sec.) of the algorithm FINDCM for amino acid sequences of the Omicron variant.

$\tau$	$\rho$	time (sec.)	$\tau$	$\rho$	time (sec.)
0.90	0.10	2	0.75	0.10	5
		9	0.70		5
		11,840	0.65		5
			0.60		6
0.85	0.10	3	0.55		6
		11	0.50		7
		80,522	0.45		21
			0.40		30
0.80	0.10	4	0.35		40
		16	0.30		78
		427,746	0.25		127
			0.20		242

### 3.2 Correlated Mutation for Delta Variant

In the following subsections, we discuss the correlated mutations found by the algorithm FINDCM. Here, we denote the correlated mutations by the set of the following form:

$$\langle \text{structural protein} \rangle \text{position.}$$

In this subsection, we investigate the found correlated mutations for the Delta variant. When we fix  $\rho = 0.01$ , Table 4 illustrates the found correlated mutations for the Delta variant obtained by varying  $\tau$  as 0.90, 0.85 and 0.80. Here, “id” means the index “CM $i$ ” of the  $i$ -th correlated mutations. Also, we denote the positions in the spike protein substitutions for SARS-CoV-2 Delta and Omicron variants in Table 1 in Section 1 by bold faces. Furthermore, we denote the newly added positions in a correlated mutation as underlined.

Table 4 shows that, whereas the correlated mutations of CM1 and CM4 do not change even if  $\tau$  is varied from 0.90 to 0.80, the correlated mutations of CM2 and CM3 are added to new positions in correlated mutations. Also, all of the correlated mutations

Table 4: The correlated mutations for the Delta variant obtained by varying  $\tau$  as 0.90, 0.85 and 0.80 under  $\rho = 0.01$ .

$\tau$	id	correlated mutation
0.90	CM1	$\langle S \rangle 570, \langle S \rangle 716, \langle S \rangle 982, \langle S \rangle 1118, \langle N \rangle 3, \langle N \rangle 235$
	CM2	$\langle S \rangle \mathbf{452}, \langle S \rangle \mathbf{478}, \langle M \rangle 82$
	CM3	$\langle S \rangle 190, \langle N \rangle 80$
	CM4	$\langle S \rangle 501, \langle N \rangle 204$
0.85	CM1	$\langle S \rangle 570, \langle S \rangle 716, \langle S \rangle 982, \langle S \rangle 1118, \langle N \rangle 3, \langle N \rangle 235$
	CM2	$\langle S \rangle \mathbf{19}, \langle S \rangle \mathbf{452}, \langle S \rangle \mathbf{478}, \langle M \rangle 82, \langle N \rangle \mathbf{377}$
	CM3	$\langle S \rangle \mathbf{20}, \langle S \rangle 190, \langle S \rangle \mathbf{655}, \langle N \rangle 80$
	CM4	$\langle S \rangle \mathbf{501}, \langle N \rangle 204$
0.80	CM1	$\langle S \rangle 570, \langle S \rangle 716, \langle S \rangle 982, \langle S \rangle 1118, \langle N \rangle 3, \langle N \rangle 235$
	CM2	$\langle S \rangle \mathbf{19}, \langle S \rangle \mathbf{452}, \langle S \rangle \mathbf{478}, \langle S \rangle \mathbf{681}, \langle S \rangle \mathbf{950}, \langle M \rangle 82, \langle N \rangle 63, \langle N \rangle 203, \langle N \rangle 377$
	CM3	$\langle S \rangle 20, \langle S \rangle \mathbf{26}, \langle S \rangle 190, \langle S \rangle 655, \langle S \rangle 1027, \langle S \rangle 1176, \langle N \rangle 80$
	CM4	$\langle S \rangle \mathbf{501}, \langle N \rangle 204$
	CM5	$\langle S \rangle 289, \langle N \rangle 18$

in Table 4 contain the positions of S and the other structural proteins.

In particular, the correlated mutation CM2 at  $\tau = 0.80$  contains five positions for spike protein substitutions that are just positions in S. Note that the five positions are known to mainly characterize SARS-CoV-2 Delta variant (Chen et al., 2022).

### 3.3 Correlated Mutation for Omicron Variant Under $\rho = 0.01$

In this and next subsections, we investigate the correlated mutations for Omicron variant.

First, when we fix  $\rho = 0.01$ , Table 5 illustrates the found correlated mutations for the Omicron variant obtained by varying  $\tau$  as 0.90, 0.85 and 0.80. Here, “+” means that the correlated mutation is obtained by adding the upper correlated mutation to the presented positions.

Table 5: The correlated mutations for the Omicron variant obtained by varying  $\tau$  as 0.90, 0.85 and 0.80 under  $\rho = 0.01$ .

$\tau$	correlated mutation
0.90	$\langle S \rangle 109, \langle S \rangle 110, \langle S \rangle 114, \langle S \rangle 115, \langle S \rangle 116, \langle S \rangle 117, \langle S \rangle 119, \langle S \rangle 121, \langle S \rangle 122, \langle S \rangle 125, \langle S \rangle 126, \langle S \rangle 129, \langle S \rangle 131, \langle S \rangle 136, \langle S \rangle 139, \langle S \rangle 140, \langle S \rangle 149$
0.85	$+ \langle S \rangle 113, \langle S \rangle 120, \langle S \rangle 122, \langle S \rangle 124, \langle S \rangle 127, \langle S \rangle 137, \langle S \rangle \mathbf{144}, \langle S \rangle 150$
0.80	$+ \langle S \rangle 130, \langle S \rangle 134, \langle S \rangle 141, \langle S \rangle 154$

Table 5 shows that, under  $\rho = 0.01$  and  $0.80 \leq \tau \leq 0.90$ , we can find one correlated mutation, which is concerned with just S with positions from 109 to 154. Also the position of 144 is concerned with spike protein substitution in Table 1.

### 3.4 Correlated Mutation for Omicron Variant Under $\rho = 0.10$

Next, when we fix  $\rho = 0.10$ , Table 6 illustrates the found correlated mutations for the Omicron variant obtained by varying  $\tau$  from 0.70 to 0.20 decreasing by 0.05.

Table 6: The correlated mutations for the Omicron variant obtained by varying  $\tau$  from 0.70 to 0.20 decreasing by 0.05 under  $\rho = 0.10$ .

$\tau$	id	correlated mutation
0.70	CM1	$\langle S \rangle 289, \langle N \rangle 18$
0.65–0.55	CM1	$\langle S \rangle 289, \langle N \rangle 18$
	CM2	$\langle S \rangle 222, \langle N \rangle 215$
0.50	CM1	$\langle S \rangle 289, \langle N \rangle 18$
	CM2	$\langle S \rangle 112, \langle S \rangle 222, \langle N \rangle 215$
0.45	CM1	$\langle S \rangle 5, \langle S \rangle 289, \langle S \rangle 809, \langle S \rangle 1104, \langle S \rangle 1264, \langle N \rangle 9, \langle N \rangle 18, \langle N \rangle 63$
	CM2	$\langle S \rangle 112, \langle S \rangle 222, \langle N \rangle 215$
	CM3	$\langle S \rangle 95, \langle S \rangle 142$
0.40–0.35	CM4	$\langle S \rangle 5, \langle S \rangle 112, \langle S \rangle 222, \langle S \rangle 289, \langle S \rangle 809, \langle S \rangle 1104, \langle S \rangle 1264, \langle N \rangle 9, \langle N \rangle 18, \langle N \rangle 63, \langle N \rangle 215$
	CM3	$\langle S \rangle 95, \langle S \rangle 142$
0.30–0.20	CM5	$\langle S \rangle 5, \langle S \rangle 95, \langle S \rangle 112, \langle S \rangle 142, \langle S \rangle 222, \langle S \rangle 289, \langle S \rangle 809, \langle S \rangle 1104, \langle S \rangle 1264, \langle N \rangle 9, \langle N \rangle 18, \langle N \rangle 63, \langle N \rangle 215$

Table 6 shows that all of the correlated mutations contain the positions of S and the other structural proteins. Also, the correlated mutation CM3 contains the positions for spike protein substitutions.

Also Table 6 shows that the correlated mutation CM4 is the combination of CM1 and CM2, and the correlated mutation CM5 is the combination of CM3 and CM4. Hence, we can regard that the correlated mutation CM5 is the convergence of other found correlated mutations.

On the other hand, the number of positions in spike protein substitutions occurring in the correlated mutations is just two, which is small. Then, it is a future work to find the correlated mutations containing more positions in spike protein substitutions.

## 4 CONCLUSION

In this paper, we have found the correlated mutations of positions among structural proteins in amino acid sequences of SARS-CoV-2 Delta and Omicron variants by using the algorithm FINDCM designed by (Shimada et al., 2012). Then, we have obtained the correlated mutations containing the positions among several structural proteins and containing the positions occurring in the spike protein substitutions in SARS-CoV-2 Delta and Omicron variant.

In particular, we have found the correlated mutation CM5 in Table 6 as the convergence of several correlated mutations containing the positions in the spike protein substitutions. On the other hand, it is a future work to investigate the positions except a spike protein of CM2 at  $\tau = 0.80$  in Table 4 and CM5 in Table 6 in the genomic viewpoints.

The algorithm FINDCM is based on the set enumeration algorithm (Rymon, 1992). Then, it is a future work to design the algorithm of finding correlated mutations based on another enumeration algorithm, with introducing another thresholds like  $\tau$  and  $\rho$ .

Whereas we have found the correlated mutations concerned with the positions in the spike protein substitutions, the number of them is small, in particular, for the Delta variant. Also, whereas the algorithm FINDCM finds all of the correlated mutations under given  $\tau$  and  $\rho$ , it is necessary to find the correlated mutations concerned with the positions in the spike protein substitutions directly and efficiently. Hence, it is a future work to design a new algorithm of finding the correlated mutations *containing given several positions* like as the positions in spike protein substitutions, which is possible to be more efficient than FINDCM.

## REFERENCES

- Chen, K.-W. K., Huang, D. T.-N., and Huang, L.-M. (2022). SARS-CoV-2 variants – evolution, spike protein, and vaccines. *Biomed. J.*, 45:573–579.
- Cover, T. M. and Thomas, J. A. (2006). *Elements of information theory (Second edition)*. John Wiley & Sons.
- Jeong, C. and Kim, D. (2010). Linear predictive coding representation of correlated mutation for protein sequence alignment. *BMC Bioinform.*, 11:52.
- Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell*, 181.
- Lee, B.-C. and Kim, D. (2009). A new method for revealing correlated mutations under the structural and functional constraints in proteins. *Bioinform.*, 25:2506–2513.

- Nakagawa, S. and Miyazaki, T. (2020). Genome evolution of SARS-CoV-2 and its virological characteristics. *Inflamm. Regen.*, 40:17.
- Oliveria, L., Paiva, A. C. M., and Vriend, G. (2002). Correlated mutation analyses on very large sequence families. *ChemBioChem*, 3:1010–1017.
- Rymon, R. (1992). Search through systematic set enumeration. In *Proc. KR'92*, pages 539–550.
- Shimada, T., Hazemoto, T., Makino, S., Hirata, K., Yonezawa, K., and Ito, K. (2012). Finding correlated mutations among rna segments in H3N2 influenza viruses. In *Proc. SCIS-ISIS'12*, pages 1696–1705.
- Yonashiro, K., Shimaya, Y., and Hirata, K. (2022). Finding correlated mutations of positions among structural proteins in SARS-CoV-2 amino acid sequences. In *Proc. ESKM'22*, pages 61–64.

