How much energy is stored in SARS-CoV-2 and its structural elements?

Bartu Şimşek1 | Mustafa Özilgen2 | Feride Şermin Utku1

1Department of Biomedical Engineering, Faculty of Engineering, Yeditepe University, Istanbul, Turkey
2Department of Food Engineering, Faculty of Engineering, Yeditepe University, Istanbul, Turkey

Correspondence
Mustafa Özilgen, Yeditepe University, Faculty of Engineering, Department of Food Engineering, Kayisdagi Cad., Istanbul 34755, Turkey.
Email: mozilgen@yeditepe.edu.tr

Abstract
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the virus causing the COVID-19 disease. Data regarding the morphological properties of this virus are collected from the literature and then the energy stored in each structural element is calculated with Domalski and Hearing’s group contribution method. Viruses, including the Corona viruses, derive all of their energy from the host cell and carry out all of their activities with this energy. SARS-CoV-2 construct a vehicle needed for the delivery of its mRNA to other hosts to inflict them with the disease. Upon transfer of the viral RNA to the new host, the remaining parts of the viral structure are discarded. Structural and molecular assessments showed that the chemical formula of SARS-CoV-2 virus is C_{7336852}H_{12384463}N_{1247424}O_{1915357}P_{100231}S_{25084} and its enthalpy of formation is $-8.70 \times 10^{-16}$ kJ. Comparison of SARS-CoV-2 with the other viruses shows that its elemental composition does not like any of the others. The results of this study are expected to improve our knowledge of the thermodynamic properties of this virus.

Keywords
disease spreading vehicle, COVID-19 disease, elemental composition, energy storage, enthalpy of formation, group contribution method, SARS-CoV-2, structural elements

1 INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Figure 1) is the virus, which causes the COVID-19 disease. Viruses are non-living sub-microscopic agents, they replicate obligatorily inside a host cell,1 and carry out their activities and construct a vehicle to deliver their mRNA to other hosts to inflict them with the disease with the energy driven from the hosts (Figure 2). Upon transfer of the viral RNA to a new host, this vehicle is discarded. An mRNA vaccine is made of a coating that makes delivery possible and prevents the host from damaging it during entry. The viral vehicle, which is focused on this study, aims infecting the host, whereas a vaccine aims preventing the infection. Popovic and Minceva2,3 after calculating the Gibbs energy of the proliferation of the SARS-CoV-2 and those of the synthesis of the host cell components, reported that the Gibbs energy of the viral proliferation was more negative, and considered it as the physical explanation for the ability of the viruses to hijack the host cell metabolism. Popovic and Minceva2,3 argued later that this information can open new paths for attenuated vaccine design. When an infection takes control of the metabolism, the disease scavenges the healthy tissues to afford energy and material to invade the body.4,5 Distortion and exploiting of the energy management is observed also in many other health problems, including cancer6 and heart failure.7 Popovic and Minerva3 made considerable contribution to our knowledge of the thermodynamic aspects of the virus...
and the host cell interaction during the development of the infections.

Virus emergence is a three-step process: first, a virus acquires the ability to infect a new host, in the second step it adapts to the new host and finally gains ability to spread in the host population. SARS-CoV-2 virus in pangolins and bats of Southeast Asia evolved very fast due to rapid viral proliferation rate, small viral genome, and very high animal population in the region. Under such conditions, the virus had $10^{-3}$ to $10^{-6}$ mutations per nucleotide in every replication of the genome.\(^8\) Genes from different viruses may combine to generate a virus, which may access into a host cell without triggering its immunity. For example, the host receptor (spike protein) of the chicken bronchitis corona virus was replaced by a spike gene from an unknown source, thus the new virus gained ability to infect turkeys but lost its previous ability of infecting the chickens.\(^9\) The new viral species can survive, if they should have energetically feasible structures\(^10\); therefore, determination of the energy storage in a virion may be of significant importance to understand the stability and emergence of new viruses. A complete virus that consists of an RNA or DNA core with an external envelope in extracellular infectious form of a virus is called a virion. Free energy and entropy are the important thermodynamic properties determining evolution of such a system and its outcomes.\(^11\)

Thermodynamic system and subsystem boundaries of the SARS-COV-2 virus and its structural elements are described schematically in Figure 2. Energy stored within the subsystem boundary (virus) and its structural elements is the focus of this study. Double-stranded DNA of the DNA viruses are tightly confined within the virus envelope and create tens of atmosphere pressure. Viruses store this energy and wait to be unleash it until encountering a host cell. When the

**Figure 1** Schematic description of the structure of the SARS-CoV-2 virus envelope (not in scale). S: spike protein, M: M protein, E: E protein, vRNP: viral ribonucleoprotein complex, RNA: viral RNA) (not in scale)

**Figure 2** Schematic description of the system and the subsystem boundaries employed in thermodynamic analyses (not in scale)
virus gets into the host, cell turns it into a virus factory. With the mRNA viruses, the situation is a bit more complicated. When the single-stranded viral RNA genome, (+) ssRNA, passes through the host cell boundaries, it dissociates itself from the N protein and translated immediately to produce polyproteins and non-structural proteins. The viral replication organelles consist of convoluted membranes, vesicles, and spherules to provide the necessary microenvironment not only for the replication of the viral genomic RNA, but also for the transcription of the subgenomic mRNAs. The structural proteins are translated and translocated into the endoplasmic reticulum (ER) membranes, transit through the ER-to-golgi intermediate compartment (ERGIC), where the genomic RNA is associated with the N protein and buds into the secretory vesicular compartments. Figure 2 shows that the new viruses are synthesized with the building material provided from the host and are secreted out through the host cell system boundary.

The enthalpy of formation, \( \Delta H^0 \), is defined as the energy released or consumed for the formation of a substance from its elements in their most stable reference states at the chosen temperature and pressure. Since it is impossible to find experimental values of the enthalpy of formation of the components of the virus in the literature, theoretical calculations are made in this study by employing the group contribution method (GCM) of reference 14. Moreover, to the best of our knowledge, this study is the first attempt to use the GCM to calculate the enthalpy of formation of a virus. Since the outbreak of the SARS-CoV-2 pandemic in 2020, distinguishing features of the SARS-CoV-2 have been researched to understand its structure and properties, this study was performed with the expectation of learning of the thermodynamic properties of the virus better.

2 | MATERIALS AND METHODS

Schematic description of the method of calculations employed in the present study is provided in Figure 3.

2.1 | Morphological calculations

Like the other (+) ssRNA viruses (single-stranded RNA viruses with positive RNA polarity), SARS-CoV-2 buds and assembles in the ERGIC membrane. The ERGIC membrane may consist of 55% phosphatidylcholine (PC), 25% phosphatidylethanolamine (PE), 10% phosphatidylinositol (PI), 5% phosphatidylserine (PS), and 5% sphingomyelin (SM) and, its Chol (Cholesterol)/PS molar ratio may be approximately 3.5. This membrane is made of lipid bilayers with a monolayer separation of 3.6 nm (thickness \( d = 7.2 \) nm), after assembly of the virus, has a nearly spherical shape with a diameter of approximately 90 nm (radius \( r = 45 \) nm) and referred to as the viral envelope. Within the viral envelope,
approximately 30-kb long (+) ssRNA and viral ribonucleoprotein (vRNP) complexes are located. According to the NCBI database, the reference genome of Wuhan strain of SARS-CoV-2 (NC_045512.2) is made up of 8954 adenine, 5863 guanine, 5492 cytosine, and 9594 uracil bases. In approximate numbers, 38 individual vRNP

| Amino acid | S | E | M | N |
|------------|---|---|---|---|
| Ala        | 79 | 2 | 4 | 2 | 19 | 2 | 3 | 4 | 37 |
| Cys        | 40 | 0 | 3 | 0 | 4  | 1 | 1 | 1 | 0  |
| Asp        | 62 | 0 | 1 | 0 | 6  | 0 | 0 | 0 | 24 |
| Glu        | 48 | 0 | 2 | 0 | 7  | 0 | 0 | 0 | 25 |
| Phe        | 77 | 1 | 5 | 3 | 11 | 3 | 2 | 3 | 13 |
| Gly        | 82 | 2 | 1 | 0 | 14 | 1 | 0 | 2 | 43 |
| His        | 17 | 0 | 0 | 0 | 5  | 0 | 0 | 0 | 4  |
| Ile        | 76 | 5 | 3 | 1 | 20 | 2 | 2 | 3 | 14 |
| Lys        | 61 | 0 | 2 | 0 | 7  | 0 | 0 | 0 | 31 |
| Leu        | 108| 3 | 14| 7 | 35 | 5 | 5 | 4 | 27 |
| Met        | 14 | 2 | 1 | 0 | 4  | 0 | 0 | 2 | 7  |
| Asn        | 88 | 0 | 5 | 1 | 11 | 1 | 0 | 0 | 22 |
| Pro        | 58 | 0 | 2 | 0 | 5  | 0 | 1 | 0 | 28 |
| Gln        | 62 | 0 | 0 | 0 | 4  | 1 | 0 | 0 | 35 |
| Arg        | 42 | 0 | 3 | 0 | 14 | 0 | 1 | 1 | 29 |
| Ser        | 99 | 0 | 8 | 1 | 15 | 0 | 0 | 2 | 37 |
| Thr        | 97 | 1 | 4 | 1 | 13 | 1 | 1 | 0 | 32 |
| Val        | 97 | 2 | 13| 5 | 12 | 1 | 3 | 1 | 8  |
| Trp        | 12 | 2 | 0 | 0 | 7  | 2 | 2 | 1 | 5  |
| Tyr        | 54 | 1 | 4 | 0 | 9  | 1 | 1 | 1 | 11 |

**TABLE 1** List of S, E, M, and N structural proteins full (F) and transmembrane domain (TM) sequences

**FIGURE 4** Cubic representation of transmembrane domain of the structural proteins and average lipid molecule. The phospholipid component of the lipid bilayer has been indicated as two rectangular prisms with an equivalent surface area and volume occupied by the transmembrane domains of the proteins.
complexes surrounds 800 nucleotides of this genome with 12 copies of N (ID: YP_009724397.2) protein of 419 residues.23,24 S (ID: YP_009724390.1), E (ID: YP_009724392.1), M (ID: YP_009724393.1) proteins with 1273, 222, and 75 residues, respectively, are localized in the viral envelope,25-27 also, these proteins assemble into trimers (S3), dimers (M2), and monomers (E), respectively. S protein’s Trp1,214-Leu1,234; M protein’s Trp20-Ala40, Leu51-Try71, and Ile80-Phe100; E protein’s Val14-Leu34 residues are transmembrane (TM) domain.28-30 List of the full and transmembrane domain sequences of the S, E, M, and N structural proteins shown in Table 1. SARS-CoV-2 enters the host cells by the fusion of viral and cellular membranes with the densely glycosylated spike protein. The S protein is a viral fusion protein and an important target for antibody neutralization and vaccine development.31 S proteins are perpendicular, and partially (1.65%) embedded in the viral envelope. There is an average of 26 S3 proteins per a viral structure.32 Unlike the S protein, the M and E proteins have a quarter of their sequence embedded in the envelope. Even though their average amounts are not yet determined for SARS-CoV-2, the M protein is known to be highly enriched in the viral envelope,33 whereas the E protein is deficient and found in negligibly small amounts.34 On the other hand, morphological characteristics of mouse hepatitis virus (MHV), feline coronavirus (FCoV), SARS-CoV, and transmissible gastroenteritis virus (TGEV) are better characterized.35,36 It is estimated that there are approximately 20 copies of disproportionately distributed monomeric E protein structures in the TGEV.34,35,37 According to Neuman et al.,36 there may be approximately 1100 M2 molecules in SARS-CoV, MHV, and FCoV. Due to a lack of the actual amounts of the SARS-CoV-2’s M and E proteins, the average copy numbers of these structural proteins may be derived from other coronaviruses. The total number of lipids, nLipid, of the viral envelope was calculated with the same method as reference 2. In this respect, structural elements in the viral envelope were represented as cubic shape as shown in

### TABLE 2

| Group | ΔH_0^{solid} (kJ/mol) |
|-------|------------------------|
| Carbons |                        |
| C – (H)2(X)a | -46.74 |
| C – (H)2(C)2 | -29.41 |
| C – (H)(C)3 | -5.98 |
| C – (H)2(C)(C)a | -24.35 |
| C – (H)2(C)(C)b | -22.10 |
| C – (H)2(O)(C) | -33.00 |
| C – (H)(O)(C)2, (alcohols, peroxides) | -29.08 |
| C – (H)2(C)(CO) | -27.90 |
| C – (H)2(C)(N) | -34.00 |
| C – (H)(C)2(N) | -13.90 |
| C – (H)2(CO)(N) | -30.95 |
| C – (H)(CO)(N) | -11.65 |
| C – (H)(C)2 | 25.48 |
| C – (C)(C)a2 | 13.90 |
| C – (H)(C)b2 | 6.53 |
| C – (O)(C)a2 | 1.00 |
| C – (C)(N)(C)b | 6.53 |
| C – (H)(N)(C)b | 6.53 |
| Oxygen |                        |
| O – (C)2 | -119.00 |
| O – (H)(C) | -199.66 |
| O – (H)(C)b | -199.25 |
| O – (H)(CO) | -282.15 |
| O – (C)(CO) | -210.60 |
| Carbonyl |                        |
| CO – (C)(N) | -194.60 |
| CO – (C)(O) | -153.60 |
| Nitrogen |                        |
| N – (H)2(C), (first, amino acids) | -6.30 |
| N – (H)2(C), (second, amino acids) | -46.00 |
| N – (H)(C)2 | 47.80 |
| N – (H)(CO), (amino acids) | -59.75 |
| N – (H)(C)(CO), (amino acids) | -5.50 |
| N – (H)(C)(CO), (amines, ureas) | -9.80 |
| N – (C)(C)a2 | 89.30 |
| N – (H)(C)b2 | 45.40 |
| N – (C)(C)a(CO) | 72.00 |
| N – (H)(C)a(CO) | -3.50 |
| N – (C)(C)b(CO) | 55.00 |
| Chlorine |                 |
| Cl | -167.16b |

(Continues)

---

wagman et al.39

methyl group’s X can be used freely. Hence, [C – (H)3(C)] = [C – (H)3(N+)], etc.
Figure 4. The total viral envelope area, \( A_{\text{Envelope}} \), is equal to the sum of the outer, \( A_{\text{out}} \), and inner surface area, \( A_{\text{in}} \):

\[
A_{\text{Envelope}} = A_{\text{out}} + A_{\text{in}} = 4\pi r^2 + 4\pi (r - d)^2
\]  

(1)

This area consists of the surface area of the phospholipids, \( A_{\text{Lipid}} \), and the surface area of the TM domains of the structural proteins, \( A_{\text{TM}} \). A surface coverage of a TM of one molecule of a structural protein may be calculated by dividing molecular weight of its TM domain, \( M_{\text{WTM}} \) of \( X \), into average protein density, \( \rho = 1.36986 \times 10^{-21} \text{ g/nm}^3 \) Avogadro’s number (\( n_{\text{Avogadro}} \)), and its length (same as thickness) \( d \). To calculate coverage of both surfaces, it was multiplied with two and when the result is multiplied by the amount of the structural protein \( (n_X) \) it gives the surface area of a structural protein \( (A_{\text{TM of } X}) \):

\[
A_{\text{TM of } X} = 2 \cdot \frac{1}{d} \frac{n_X [M_{\text{WTM of } X}/n_{\text{Avogadro}}]}{\rho}
\]  

(2)

Sum of the surface areas of all structural proteins is \( A_{\text{TM}} \). Subtracting \( A_{\text{TM}} \) from \( A_{\text{Envelope}} \) gives \( A_{\text{Lipid}} \). Hence, the total number of lipids, \( n_{\text{Lipid}} \), can be calculated by dividing \( A_{\text{Lipid}} \) into the average area of a lipid molecule, \( a = 0.533 \text{ nm}^2 \) \(^{2,2,38} \):

\[
n_{\text{Lipid}} = \frac{A_{\text{Lipid}}}{a}
\]  

(3)

### 2.2 Estimation of enthalpy of formation by group contribution calculations

Calculating the energy stored in SARS-CoV-2, \( \Delta \mu_{H}^{0} \) (SARS-CoV-2), entails calculating the enthalpy of formation of its large proteins, phospholipids, and (+) ssRNA. Since there is no experimental data available for the enthalpy of formation of these components, theoretical calculations were made with the GCM.\(^{14} \)

Domalski and Hearing\(^{14} \) presented estimation of thermodynamic properties of organic compounds in condensed phase at standard state \( (T = 298.15 \text{ K} \text{ and } P = 101325 \text{ Pa}) \) by GCM. Their study was applied to 1512 carbon, hydrogen, oxygen, nitrogen, sulfur, and halogen containing compounds both in the gas, liquid, and solid phases. However, the group values of reference 14 listed in Table 2 are not sufficient to calculate completely some of the amino acids, lipids, and nucleobases. When the contribution value of a group was not readily available in the literature, compounds that include those groups as described in Table 3 were found in the literature and enthalpy of formation of the

| Compound | \( \Delta \mu_{H}^{0} \) \( \text{solid (kJ/mol)} \) | CAS | Reference |
|----------|-------------------------------------------------|-----|-----------|
| 2,3-dimethyl-1 hindele | 4.20 | 91-55-4 | 40 |
| 2-ethylimidazole | \(-21.30\) | 1072-62-4 | 41 |
| Tetramethylammonium chloride (TMACl) | \(-276.40\) | 75-57-0 | 42 |
| Tetraethylammonium chloride (TEACl) | \(-369.40\) | 56-34-8 | 42 |
| Methionine | \(-577.50\) | 63-68-3 | 43 |
| Cysteine | \(-529.20\) | 52-90-4 | 43 |
| Proline | \(-515.2\) | 147-85-3 | 43 |
| d-Arginine | \(-623.60\) | 157-06-2 | 44 |
| Phosphoric acid | \(-1284.38\) | 7664-38-2 | 45 |
| Triethyl phosphate | \(-1186.70\) | 78-40-0 | 46 |
| Adenosine | \(-653.60\) | 58-61-7 | 47 |
| Adenine | 96.90 | 73-24-5 | 48 |
| Guanine | \(-183.90\) | 73-40-5 | 49 |
| Cytosine | \(-221.30\) | 71-30-7 | 50 |
| Uracil | \(-429.56\) | 66-22-8 | 51 |
| 5-(1α,2β,3β,4α)-1,2,3,4-Cyclohexenetetrol (Conduritol E) | \(-836.30\) | 526-87-4 | 52 |
| Chol | \(-674.80\) | 57-88-5 | 53 |
unknown groups were calculated from enthalpy of formation of these compounds by adding or subtracting those of the known groups. Equations that are employed to calculate some of the new group contribution values are presented in Table 4. The equations of enthalpies of formation of structural proteins, viral RNA, and lipids, which are structural elements of the virus, are presented in Table 5, Table 6, and Table 7, respectively. In Table 5, enthalpy of formation of each amino acid in a protein sequence was calculated to determine the enthalpy of formation of long proteins, by distinguishing between a single or terminal amino acid and an amino acid in a sequence, as shown in Figure 5, without the NH₂ ([N – (H)₂(C)], COOH ([O – (H)CO]) and [CO – (C)(O)]) groups and Zwitterion energy. When enthalpy of formation of these amino acid values is multiplied by their number of copies in the protein sequence (Table 1) and summed, and then added with the correction value for long proteins (long protein corr), it gives ΔfH°(X structural protein). The longer the protein sequence, more likely it will contain one of the 20 amino acids. Any protein containing His, Met, Phe, Tyr, or Trp amino acid is aromatic, and Arg, Lys, Asn, or Gln amino acid carries a second amino group ([N – (H)₂(C), (second, amino acids)]). S, E, M, and N structural proteins satisfy these properties; hence, their correction value is equal to the replacement of the amino and hydroxyl groups at the N and C terminus of the sequence with the appropriate values and the addition of the Zwitterion energy (aromatic I). In Table 6, RNA corr, a correction, was applied to hydroxy groups of the appropriate phosphate and ribose sugars at the 5’ and 3’ ends of RNA. Since the experimentally calculated values were available for Chol; therefore, it was not calculated in Table 7.

2.3 Estimation of the enthalpy of formation by Battley’s method

When the chemical formula of the virus is known, its standard enthalpy of formation may also be estimated with the Battley and Patel-Erickson equation⁵⁴,⁵⁵:

\[
\Delta_f H^0(\text{virus}) = n_C \Delta_f H^0(\text{CO}_2) + \frac{1}{2} n_H \Delta_f H^0(\text{H}_2\text{O}) + \frac{1}{4} n_P \Delta_f H^0(\text{P}_4\text{O}_{10}) + n_S \Delta_f H^0(\text{SO}_3) + n_{\text{Mg}} \Delta_f H^0(\text{MgO}) + n_{\text{Ca}} \Delta_f H^0(\text{CaO}) + n_{\text{Cl}} \Delta_f H^0(\text{HCl}) - \Delta_f H^0(\text{virus}) \tag{4}
\]

\[
\Delta_C H^0(\text{virus}) = -111.14 \text{ kJ mol}^{-1} (n_C + n_{\text{H}} - 2n_O - 0n_N + 5n_P + 6n_S) \tag{5}
\]

where, \(\Delta_f H^0(\text{virus})\) is the enthalpy of formation of the virus and \(\Delta_C H^0(\text{virus})\) is the enthalpy of combustion of the virus in (kJ/virus), \(n_C\), \(n_{\text{H}}\), \(n_O\), \(n_N\), \(n_P\), and \(n_S\) are the number of the C, H, O, N, P, and S in the chemical formula of the virus, respectively. Popovic⁵⁶ estimated that 5.36% of uncertainty is involved in Equation (5). Enthalpies of formation of carbon dioxide (g), phosphorous deoxide (cr), sulfur trioxide (g), and water (l) at standard conditions are presented in Table 8.

| Table 4 | Equations that are employed to calculate new group contribution values |
|---------|-------------------------------------------------------------------------|

\[
\begin{align*}
[\text{Pyrrolidine rsc}] & = [\Delta_f H^0(\text{Proline})] - [\text{O} - (\text{H})(\text{CO})] + [\text{CO} - (\text{C})(\text{O})] + [\text{C} - (\text{H})(\text{C})(\text{CO})(\text{N})] + [\text{N} - (\text{H})(\text{C})_2] + [\text{C} - (\text{H})(\text{C})(\text{N})] \\
& + [2 \times [\text{C} - (\text{H})(\text{O})(\text{C})_2]] + [\text{Zwitterion energy; aliphatic}] \\
[\text{C} - (\text{H})(\text{O})(\text{C})(\text{N})] & = [\Delta_f H^0(\text{Adenosine})] - [\Delta_f H^0(\text{Adenine})] - [\text{N} - (\text{H})(\text{C})_2] + [\text{N} - (\text{H})(\text{C})(\text{N})] + [3 \times [\text{O} - (\text{H})(\text{C})]] + [\text{C} - (\text{H})(\text{O})(\text{C})_2] + [\text{C} - (\text{H})(\text{O})(\text{C})(\text{N})] \\
& + [3 \times [\text{C} - (\text{H})(\text{O})(\text{C})(\text{N})] + [\text{Zwitterion energy; aliphatic}]]/2 \\
[\text{C} - (\text{H})(\text{O})(\text{C})(\text{S})] & = [\Delta_f H^0(\text{Conduritol E})] - [\Delta_f H^0(\text{Adenosine})] - [\text{N} - (\text{H})(\text{C})_2] + [\text{N} - (\text{H})(\text{C})(\text{N})] + [3 \times [\text{O} - (\text{H})(\text{C})]] + [\text{C} - (\text{H})(\text{O})(\text{C})_2] \\
& + [\text{C} - (\text{H})(\text{O})(\text{C})(\text{N})] + [\text{C} - (\text{H})(\text{O})(\text{C})(\text{S})] + [\text{Zwitterion energy; aliphatic}] \\
[\text{C} - (\text{H})(\text{O})(\text{C})(\text{PO})] & = [\Delta_f H^0(\text{phosphoric acid})]/3 + [\Delta_f H^0(\text{triethyl phosphate})] - [3 \times [\text{C} - (\text{H})(\text{O})(\text{C})(\text{N})]]/3 \\
& + [\text{C} - (\text{H})(\text{O})(\text{C})(\text{PO})] + [2 \times [\text{O} - (\text{H})(\text{C})]] + [\text{C} - (\text{H})(\text{O})(\text{C})(\text{N})] + [\text{Zwitterion energy; aliphatic}] \\
[\text{S} - (\text{C})(\text{H})(\text{C})(\text{S})(\text{C})] & = [\Delta_f H^0(\text{methionine})] - [\text{C} - (\text{H})(\text{O})(\text{C})(\text{N})] + [\text{C} - (\text{H})(\text{O})(\text{C})(\text{S})] + [\text{Zwitterion energy; aliphatic}] \\
& + [\text{C} - (\text{H})(\text{O})(\text{C})(\text{N})] + [\text{C} - (\text{H})(\text{O})(\text{C})(\text{S})] + [\text{Zwitterion energy; aliphatic}] \\
[\text{N} - (\text{H})(\text{N})] & = [\Delta_f H^0(\text{TMACI})]/4 - [4 \times [\text{C} - (\text{H})(\text{N})]] - [\text{Cl}^-] - [4 \times [\text{CH}_3 \text{corr (quaternary)]}] \\
[\text{C} - (\text{H})(\text{C})(\text{N})(\text{N})] & = [\Delta_f H^0(\text{TEACI})]/4 - [4 \times [\text{C} - (\text{H})(\text{C})(\text{N})]] + [\text{N}^+ - (\text{C})(\text{N})] - [\text{Cl}^-]/4
\end{align*}
\]
TABLE 5 Equations that are employed to calculate the enthalpy of formation of amino acids in protein sequences with their molecular weights and S, E, M, and N structural proteins

| Common part of amino acids in protein (CH3ON)           |
|-------------------------------------------------------|
| $\Delta H_f^\circ$ (CPaa) = [N – (H)(C)(CO), (amino acids)] + [C – (H)(C)(O)(N)] + [CO – (C)(N)]          |
| Alanine (C3H7NO), Mw: 71.09 g/mol                       |
| $\Delta H_f^\circ$ (Ala) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(C)(O)] + [N – (H)(C)(O)] |
| Cysteine (C3H7NO2S), Mw: 103.15 g/mol                   |
| $\Delta H_f^\circ$ (Cys) = $[\Delta H_f^\circ$ (CPaa)] + [S – (C)(H)2] + C – (H)(C)(S) |
| Aspartic Acid (C6H11NO4), Mw: 115.10 g/mol              |
| $\Delta H_f^\circ$ (Asp) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(C)(O)] + [CO – (C)(O)] + [O – (H)(C)] |
| Glutamic acid (C6H11NO4), Mw: 129.13 g/mol              |
| $\Delta H_f^\circ$ (Glu) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(C)(O)] + [CO – (C)(O)] + [O – (H)(C)] |
| Phenylalanine (C9H11NO), Mw: 147.19 g/mol               |
| $\Delta H_f^\circ$ (Phe) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(C)(C)(C)] |
| Tryptophan (C9H35NO), Mw: 113.18 g/mol                  |
| $\Delta H_f^\circ$ (Trp) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(C)(C)(C)] + [C – (H)(C)(C)(S)] + [5 x (C) – (H)(C)] |
| Glycine (C3H7NO), Mw: 57.06 g/mol                       |
| $\Delta H_f^\circ$ (Gly) = [N – (H)(C)(C)(O), (amino acids)] + [C – (H)(C)(C)(O)(N)] + [CO – (C)(N)] |
| Histidine (C6H12N2O2), Mw: 137.16 g/mol                 |
| $\Delta H_f^\circ$ (His) = $[\Delta H_f^\circ$ (CPaa)] + [2 x (C) – (H)(C)(C)(C)] |
| Isoleucine (C4H9NO), Mw: 113.18 g/mol                    |
| $\Delta H_f^\circ$ (Ile) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(C)(C)] |
| Lysine (C6H15N2O2), Mw: 128.2 g/mol                      |
| $\Delta H_f^\circ$ (Lys) = $[\Delta H_f^\circ$ (CPaa)] + [3 x (C – (H)(C)(C)(C))] |
| Leucine (C6H15N2O2), Mw: 113.18 g/mol                     |
| $\Delta H_f^\circ$ (Leu) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(C)(C)] + [C – (H)(C)(C)(C)] + [2 x (C) – (H)(C)(C)] |
| Methionine (C3H7NO2S), Mw: 131.21 g/mol                  |
| $\Delta H_f^\circ$ (Met) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(C)(C)] + [S – (C)(C)2] + [C – (H)(C)(C)(S)] + [C – (H)(C)(C)] |
| Asparagine (C4H11NO2), Mw: 107.12 g/mol                    |
| $\Delta H_f^\circ$ (Asn) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(C)(C)] |
| Proline (C4H9NO2), Mw: 97.13 g/mol                          |
| $\Delta H_f^\circ$ (Pro) = [N – (C)(C)] + [C – (H)(C)(C)(N)] + [2 x (C) – (H)(C)(C)] + [Pyrrrolidine rsc] + [CO – (C)(N)] |
| Glutamine (C6H14N2O3), Mw: 128.15 g/mol                   |
| $\Delta H_f^\circ$ (Gln) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(C)(C)] + [C – (H)(C)(C)(N)] + [CO – (C)(N)] + [N – (H)(C)(C)(O)] |
| Arginine (C5H14N4O3), Mw: 156.22 g/mol                      |
| $\Delta H_f^\circ$ (Arg) = $[\Delta H_f^\circ$ (CPaa)] + [2 x (C) – (H)(C)(C)(C)] |
| Serine (C3H7NO2), Mw: 87.09 g/mol                             |
| $\Delta H_f^\circ$ (Ser) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(O)(C)] + [O – (H)(C)] |
| Threonine (C4H9NO2), Mw: 101.12 g/mol                        |
| $\Delta H_f^\circ$ (Thr) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(O)(C)(C)] + [2 x (C) – (H)(C)(C)] |
| Valine (C5H11NO2), Mw: 99.15 g/mol                               |
| $\Delta H_f^\circ$ (Val) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(C)(C)] |
| Tryptophan (C5H11NO2), Mw: 186.23 g/mol                      |
| $\Delta H_f^\circ$ (Trp) = $[\Delta H_f^\circ$ (CPaa)] + [2 x (C) – (H)(C)(C)] |
| Tyrosine (C5H10NO2), Mw: 163.19 g/mol                         |
| $\Delta H_f^\circ$ (Tyr) = $[\Delta H_f^\circ$ (CPaa)] + [C – (H)(C)(C)(C)] + [C – (H)(C)(C)(C)] + [4 x (C) – (H)(C)] + [O – (H)(C)] |

Correction for long proteins (H2O), Mw: 18.02 g/mol

Long protein corr = $-$ [N – (H)(C)(C)(O), (amino acids)] + [CO – (C)(N)] + [O – (H)(C)(O)] + [CO – (C)(O)] + [N – (H)(C)(C)(O)] (second, amino acids) + [Zwitterion energy; aromatic I]
### TABLE 5 (Continued)

X structural protein (S, E, M, and N)

\[ \Delta H_f^0(\text{Structural Protein}) = [n_{\text{Ala}} \times \Delta H_f^0(\text{Ala})] + [n_{\text{Cys}} \times \Delta H_f^0(\text{Cys})] + [n_{\text{Asp}} \times \Delta H_f^0(\text{Asp})] + [n_{\text{Glu}} \times \Delta H_f^0(\text{Glu})] + [n_{\text{Phe}} \times \Delta H_f^0(\text{Phe})] + [n_{\text{Gly}} \times \Delta H_f^0(\text{Gly})] + [n_{\text{His}} \times \Delta H_f^0(\text{His})] + [n_{\text{Leu}} \times \Delta H_f^0(\text{Leu})] + [n_{\text{Met}} \times \Delta H_f^0(\text{Met})] + [n_{\text{Thr}} \times \Delta H_f^0(\text{Thr})] + [n_{\text{Val}} \times \Delta H_f^0(\text{Val})] + [n_{\text{Tyr}} \times \Delta H_f^0(\text{Tyr})]].

### TABLE 6

Equations that are employed to calculate the enthalpy of formation of the nucleobases in RNA and viral RNA

| Phosphate group (HPO4) | \[ \Delta H_f^0(\text{PO}) = [2 \times (C - (H)_2(O)(C)) + \frac{2}{C_2}] + [2 \times (C - (C)(H)P)]. \] |
|------------------------|--------------------------------------------------|
| Ribose sugar (C6H12O6) | \[ \Delta H_f^0(\text{Ribose}) = [C - (H)_2(O)(C)] + [3 \times (C - (H)(O)(C)_2, (alcohols, peroxides))]. \] |
| Guanine (C5H4N4O2)     | \[ \Delta H_f^0(G) = [\Delta H_f^0(\text{Adenine})] - [N - (H)\text{(Cp)}] + [N - (C)\text{(Cp)}]. \] |
| Cytosine (C4H4N2O2)    | \[ \Delta H_f^0(C) = [\Delta H_f^0(\text{Guanine})] - [N - (H)\text{(Cp)}] + [N - (C)\text{(Cp)}]. \] |
| Uracil (C4H4N2O2)      | \[ \Delta H_f^0(U) = [\Delta H_f^0(\text{Uracil})] - [N - (H)\text{(Cp)}] + [N - (C)\text{(Cp)}]. \] |

**Correction for 5’ and 3’ ends (HPO3)**

- RNA corr = \([\text{PO} - (O)] + [2 \times (\text{O} - (H)(\text{PO})) + [\text{O} - (C)(\text{PO})] + [\text{O} - (H)(\text{C})].\)
- Viral RNA \[ \Delta H_f^0(\text{RNA}) = [n_{\text{nt}} \times (\text{nt})] + [n_{\text{nt}} \times \Delta H_f^0(\text{Ribose})] + [n_{\text{nt}} \times \Delta H_f^0(\text{Adenine})] + [n_{\text{nt}} \times \Delta H_f^0(\text{Cytosine})] + [n_{\text{nt}} \times \Delta H_f^0(\text{Uracil})] + [\text{RNA corr}].\]

### TABLE 7

Equations that are employed to calculate the enthalpy of formation of phospholipids

**Common part of PC, PE, PI, and PS (C16H23O4P)**

\[ \Delta H_f^0(\text{PC}) = [2 \times (C - (H)(O)(C)) + [18 \times (C - (H)(C)(O)) + [4 \times (C - (H)(C)(C)(O)) + [4 \times (C - (H)(O)(C)(O)) + [2 \times (C - (H)(O)(C)(O)) + [2 \times (C - (C)(H)(O)(C)(O)). \] |

**Phosphatidylcholine (PC16:1/7Z/16:1/9Z), (C36H72NO14P)**

\[ \Delta H_f^0(\text{PC}) = [\Delta H_f^0(\text{PCpl})] + [C - (H)(O)(C)] + [C - (H)(C)(N)] + [N + (C) - (C)]. \] |

**Phosphatidylethanolamine (PE16:1/7Z/16:1/9Z), (C37H76NO10P)**

\[ \Delta H_f^0(\text{PE}) = [\Delta H_f^0(\text{PCpl})] + [C - (H)(O)(C)] + [C - (H)(C)(N)] + [N + (C) - (C)]. \] |

**Correlation factors**

- Viral RNA \[ \Delta H_f^0(\text{RNA}) = [n_{\text{nt}} \times (\text{nt})] + [n_{\text{nt}} \times \Delta H_f^0(\text{Ribose})] + [n_{\text{nt}} \times \Delta H_f^0(\text{Adenine})] + [n_{\text{nt}} \times \Delta H_f^0(\text{Cytosine})] + [n_{\text{nt}} \times \Delta H_f^0(\text{Uracil})] + [\text{RNA corr}].\]

### 3 RESULTS AND DISCUSSION

Almost all the experimental and theoretical enthalpy of formation calculations in the literature have been performed at the standard state, without taking the differences in the thermodynamic properties of the biological constituents at room and physiological temperatures. Therefore, in the present study, the standard enthalpy of formation of the virus, \( \Delta H_f^0(\text{SARS-CoV-2}) \), was calculated with the assumption that a droplet of virus was suspended in the air at standard conditions. In order to calculate \( \Delta H_f^0(\text{SARS-CoV-2}) \), the data were collected from the literature. Data, such as thermodynamic properties, the total copy number of compounds, etc., which were not available in the literature, were derived from other coronaviruses or found by calculating within the framework of certain assumptions.

Structural elements of the viruses are not generally scattered in a regularly manner, but the scatter is also not totally random, but found in a narrow range. While Yao et al.\(^{32}\) state the number of the S\(_3\) proteins SARS-CoV-2 are within the range of 26 ± 15, Turňoňová et al.\(^{33}\) state the average number as 40. In the close relative SARS-CoV of SARS-CoV-2, the distribution is between 50 and 100.\(^{57}\) When we consider these values given in the literature, we cannot expect an irrelevant number of
copies. For instance, Neuman et al. noted that axial view of M packing of coronaviruses approximates a rhombus with sides of 4.0 and 4.5 nm and an interior angle of about 75°, and an average spacing of 4-5 nm between M2 proteins, and this would give approximately 1100 M2 molecules per average coronaviruses. According to spacing data of this research, an average SARS-CoV-2 may have around 1300 M2. As a result, the copy numbers of other structural proteins are expected to be within a narrow range as well. On the other hand, the number of lipids in the viral envelope varies depending on the numbers of S, E, and M proteins present. When considered average copy numbers of these proteins, according to the results of Equation (3), the viral envelope contains 70,328 phospholipid molecules. Due to the small size of the Chol, when compared with other phospholipids, its surface area has been neglected on $A_{\text{Envelope}}$.

In addition to the S protein, they are glycosylated with 66 N-linked glycans and approximately 3% of them have post-fusion and the rest has prefusion conformations. However, post-fusion conformation of the S protein and the contribution of the glycans in the virus would be negligible, and their calculations were omitted due to their complex aromatic structure. Also, it assumed that all protein and (+) ssRNA structures were of primary importance, as other structure orders would not have a significant effect on its enthalpy of formation. From the equations given in Table 4, the new group contribution values are calculated and listed in Table 9. The values listed in Table 9 were either not available in the literature or calculated with other methods than that of reference 14; therefore, recalculated to be consistent with the rest of the study. The enthalpy of formation values of adenosine monophosphate (AMP) and choline chloride were available in the literature but calculated once more to test the accuracy of our calculations as presented in Table 10. The similarity of the literature values and the results of our calculations confirm that our results were accurate. Within the framework of these assumptions, the data collected about the structural components are employed to carry out the “morphological calculations.”

### Table 8

| Substance and its state | Formula | $\Delta fH^0_{\text{solid}}$ (kJ/mol) |
|------------------------|---------|-----------------------------------|
| Carbon dioxide (g)     | CO$_2$  | $-393.51$                         |
| Phosphorous decoxide (cr) | P$_4$O$_{10}$ | $-2984.03$                     |
| Sulfur trioxide (g)    | SO$_3$  | $-395.72$                         |
| Water (l)              | H$_2$O  | $-285.83$                         |

Abbreviations: cr, crystal; g, gaseous; l, liquid.

### Table 9

| Group          | $\Delta fH^0_{\text{solid}}$ (kJ/mol) |
|----------------|-----------------------------------|
| Carbons        |                                   |
| C – (H)(O)(C)(C$_4$) | $-29.90$                        |
| C – (H)(O)(C)(N)      | $31.17$                         |
| C – (H)$_3$(C')(N$^+$) | $-27.60$                        |
| Phosphate       |                                   |
| PO – (O)$_3$ + (2 x O – (H)(PO)) + O – (C)(PO) | $-1172.08$                |
| PO – (O)$_3$ + O – (H)(PO) + (2 x O – (C)(PO)) | $-1059.78$                |
| Nitrogen        |                                   |
| N$^+$ – (C)$_4$    | $95.12$                         |
| Sulfur          |                                   |
| S – (C)$_2$ + C – (H)$_2$(C)(S) | $7.41$                           |
| S – (C)(H) + C – (H)$_2$(C)(S) | $-20.40$                        |
| Corrections     |                                   |
| Pyrrolidine rsc  | $32.32$                         |

### Figure 5

A schematic drawing describing four amino acids in the sequence of M protein. The common sequence and the different N and C termini are marked.

### Table 10

| Substance and its state | Formula | $\Delta fH^0_{\text{solid}}$ (kJ/mol) |
|------------------------|---------|-----------------------------------|
| Carbon dioxide (g)     | CO$_2$  | $-393.51$                         |
| Phosphorous decoxide (cr) | P$_4$O$_{10}$ | $-2984.03$                     |
| Sulfur trioxide (g)    | SO$_3$  | $-395.72$                         |
| Water (l)              | H$_2$O  | $-285.83$                         |
employed to calculate the enthalpy of formation of the nucleobases in RNA and viral RNA are presented in Table 2. Equations that are employed to calculate the enthalpy of formation of phospholipids are presented in Table 7. Standard enthalpies of formation of groups, which are calculated in the present study, are presented in Table 3. We employed the copy numbers of the structural elements listed in Table 11 in our calculations and determined $C_7,336,852H_1,247,424O_1,915,375P_{100,231}S_{25,084}$ as the chemical formula of the SARS-CoV-2 virion. Δ$H_f^0$ (SARS-CoV-2) was $-523\,933\,291.60$ kJ/mol. After dividing into Avogadro's constant, Δ$H_f^0$ (SARS-CoV-2) of one molecule of the virus became $-8.70 \times 10^{-16}$ kJ.

When elemental composition of the SARS-CoV-2 is used in Battley's method, according to the Equations (4) and (5), Δ$H_f^0$ (SARS-CoV-2) was $-456\,989\,508.14$ kJ/mol and after dividing this value into $n_{\text{Avogadro}}$, Δ$H_f^0$ (SARS-CoV-2) of one molecule of the was $-7.59 \pm (3.81) \times 10^{-16}$ kJ with 5.36% uncertainty in Δ$H_f^0$.

Estimated values were assigned to viruses with minimum and maximum numbers of structural elements, based on the copy numbers of the structural proteins reported in the literature, and standard enthalpy of formation of 1 molecule of SARS-CoV-2 virion with these features was calculated in Table 12 by using the group contribution and Battley's methods.

The specific heat of the virus, $C_p(X)$, may be used to calculate the enthalpy of formation Δ$H_f^0 (X)$ of the virus at any temperature $T_1$, based on the enthalpy of formation, Δ$H_f^0 (X)$, at the body temperature at $T_0$.

$$\Delta H_f^0 (X) = \Delta H_f^0 (X) + C_p (X) (T_1 - T_0) \quad (6)$$

Another severe epidemic recorded in the history was the 1918 Influenza A (H1N1) pandemic. H1N1 has a morphological structure that is similar to SARS-CoV-2. Both have a distinctive crown structure, and their average diameters are close to each other. However, they belong to different virus families. Morphological
characteristics of H1N1 have been extensively researched. H1N1 buds in the plasma membrane. Therefore, the lipid proportions are not the same with SARS-CoV-2. When a calculation is made using the lipid and plasma membrane percentages from Leventis and Grinstein, the number of (HA, NA, NP, and M1) structural proteins from Mahy et al and Fujiyoshi et al; the genomic and protein sequences in of the Influenza A virus, the \( \Delta f_{H_0}(H1N1) \) was found around \( 1.11 \times 10^{-15} \text{kJ} \) by our group contribution calculations and \( 9.72 \pm (0.51) \times 10^{-16} \text{kJ} \) by Battley’s method. A virus that does not cause an epidemic, poliovirus, which is approximately one-third smaller than SARS-CoV-2 and Influenza A, is composed of a protein capsid that consist of 60 copies of four structural proteins VP1, VP2, VP3, and VP4 and a 7.44 kb long (+) ssRNA. When its capsid’s proteins (+) ssRNA analyzed and its chemical formula was found as \( C_{331,577}H_{490,156}N_{98,028}O_{130,921}S_{2,400} \). On the other hand, Wimmer reported it as \( C_{332,652}H_{492,388}N_{98,245}O_{131,196}P_{7,501}S_{2,340} \). When calculated with the GCM, \( \Delta f_{H_0}(Poliovirus) \) was \( 5.12 \times 10^{-17} \text{kJ} \), and with Battley’s method it was \( 4.50 \pm (1.55) \times 10^{-17} \text{kJ} \). Table 13 shows that SARS-CoV-2 is a substantially bigger virus than the poliovirus, its C, H, N, O, P, and S content is 22, 25, 9.5, 14.6, 13.4, and 10.7 times of those of the poliovirus, respectively; and its enthalpy of formation is 17 times of that of the poliovirus. Jover et al reported number of the C, N, and P atoms in some phages, for example, a virus that infects and replicates within bacteria. The number of these atoms in the phages showed a very wide range. Popovic and Minceva collected the unit carbon formula of 19 viruses, where they divided the number of the H, O, N, P, and S atoms, that is \( n_H, n_O, n_N, n_P, \) and \( n_S \), respectively into the number of the C atoms, \( n_C \), and reported \( 1.493 \leq (n_H/n_C) \leq 1.592, \ 0.3150 \leq (n_O/n_C) \leq 0.4780, \ 0.2563 \leq (n_N/n_C) \leq 0.2985, \ 0.00071 \leq (n_P/n_C) \leq 0.04868 \) and \( 0.00596 \leq (n_S/n_C) \leq 0.292. \) When we consider the unit carbon formula of SARS-CoV-2, we find out that \( (n_H/n_C) = 1.688, \ (n_O/n_C) = 0.261, \ (n_N/n_C) = 0.17, \ (n_P/n_C) = 0.014, \) and \( (n_S/n_C) = 0.0034. \) Among these ratios \( (n_H/n_C), \ (n_O/n_C) \) and \( (n_N/n_C) \) are not within the range; therefore, we may conclude that SARS-CoV-2 shows considerable structural difference than the other viruses.

Interest in energy storage by microorganisms is not limited with their disease-causing potential. Recently Shen et al reviewed the energy storage and conversion potential of bacterium, fungus, and viruses with the purpose of possible employment in high-performance electrodes due to their strong abilities of fast reproduction, biomineralization, gene modification and self-assembly. Although considerable work has so far been achieved,

| Diameter | Minimum | Average | Maximum |
|----------|---------|---------|---------|
| 64.8     | 11      | 15      | 18      |
| 90       | 26      | 30      | 40      |
| 90       | 30      | 1100    | 1910    |
| 90       | 1100    | 4680    | 69      |
| 90       | 4680    | 69      | GCM     |

**TABLE 12**

| Method      | Diameter | Minimum | Average | Maximum |
|-------------|----------|---------|---------|---------|
| Battley’s method | 4.88 ± (0.56) \times 10^{-15} | -4.31 ± (2.68) \times 10^{-16} | -3.59 ± (3.88) \times 10^{-16} | -4.49 ± (10^{-15}) |
substantially more information must be available before commercialization of such processes.

4 | CONCLUSION

Both SARS-CoV-2 and Influenza A viruses share many morphological and genomic similarities with their relatives, as well as similar enthalpy of formation values. In addition to the fact that the ability of the virus to proliferate without causing disease in the host is a major factor in an outbreak, a virus that could cause an epidemic is likely to have similar morphological characteristics and close enthalpy of formation values to these two viruses. In addition, these thermodynamic properties of the virus can be used in vaccine design. Spike protein is the binding site of the virus to the host cell. If we should have sufficient knowledge about the thermodynamic properties of the SARS-CoV-2 virus, we may design a new vaccine or a medication to neutralize the virus via binding to the spike protein. If the Gibbs growth energy of the virus is less negative than that of the host cell, the virus loses its virulence.2 The enthalpy of formation can also be used to make this interpretation, and its metabolic effect on cell destruction can be modeled. In the present study, enthalphy of formation of the SARS-CoV-2 virion is calculated with the GCM. As long as similar thermodynamic information becomes available in the literature, fighting against the viruses will be easier.

ACKNOWLEDGEMENTS

The authors would like to thank Çağla Sezerer for drawing of the figures.

CONFLICT OF INTEREST

The authors do not declare conflict of interest with any parties.

DATA AVAILABILITY STATEMENT

All the data are collected from the literature by citing the sources

ORCID

Bartu Şimşek https://orcid.org/0000-0001-9524-7417

Mustafa Özilgen https://orcid.org/0000-0003-0522-3644

Feride Şermin Utku https://orcid.org/0000-0002-5143-3602

REFERENCES

1. Flint SJ, Racaniello VR, Rall GF, Skalka MR, Enquist LW. Principles of Virology. 4th ed. Washington, D.C.: American Society of Microbiology Press; 2015.

2. Popovic M, Mineeva M. Thermodynamic insight into viral infections 2: empirical formulas, molecular compositions and thermodynamic properties of SARS, MERS and SARS-CoV-2 (COVID-19) viruses. *Heliyon*. 2020a;6(9):e04943. https://doi.org/10.1016/j.heliyon.2020.e04943

3. Popovic M, Mineeva M. A thermodynamic insight into viral infections: do viruses in a lytic cycle hijack cell metabolism due to their low Gibbs energy? *Heliyon*. 2020b;6(5):e03933. https://doi.org/10.1016/j.heliyon.2020.e03933

4. Öngel ME, Yıldız C, Özilgen M, Yılmaz B. Nutrition and disease-related entropy generation in cancer. *Int J Energy*. 2021; 34(4):411-423. https://doi.org/10.1504/IJEX.2021.114091

5. Thaker SK, Ch’ng J, Christofk HR. Viral hijacking of cellular metabolism. *BMC Biol*. 2019;1(59):789. https://doi.org/10.1186/s12915-019-0678-9

6. Garland J. Energy management – a critical role in cancer induction? *Crit Rev Oncol Hematol*. 2013;88(1):198-217. https://doi.org/10.1016/j.critrevonc.2013.04.001

### Table 13

| Virus | Chemical formula | $\Delta f^\circ$ (kJ) | Battley’s method |
|-------|------------------|----------------------|------------------|
|       |                  | GCM                  | Battley’s method |
| SARS-CoV-2 | $C_{7,333,821}H_{12,384,463}O_{1,915,357}N_{1,247,424}S_{1,250,084}$ (present study) | $-8.70 \times 10^{-16}$ (present study) | $-7.59 \pm (3.81) \times 10^{-16}$ (present study) |
| Poliovirus | $C_{332,652}H_{992,368}N_{98,245}O_{151,16}S_{7,50}P_{2,340}$ (Wimmer) | $-5.12 \times 10^{-17}$ (present study) | $-4.50 \pm (1.55) \times 10^{-17}$ (present study) |
| Enterobacteriaceae phage T4 | $C_{7,333,835}N_{2,309,463}P_{337,806}$ (Jover et al) | - | - |
| Enterobacteriaceae phage $\lambda$ | $C_{1,697,088}N_{621,823}P_{97,004}$ (Jover et al) | - | - |
| Enterobacteriaceae phage HK97 | $C_{1,640,059}N_{536,495}P_{79,464}$ (Jover et al) | - | - |
| Enterobacteriaceae phage T7 | $C_{1,550,610}N_{517,444}P_{78,874}$ (Jover et al) | - | - |
| Escherichia spp. phage N4 | $C_{2,765,271}N_{981,989}P_{140,306}$ (Jover et al) | - | - |
| Synechococcus phage Syn5 | $C_{1,664,611}N_{563,058}P_{52,428}$ (Jover et al) | - | - |
| Bacillus spp. phage $\Phi$29 | $C_{1,058,584}N_{236,209}P_{38,564}$ (Jover et al) | - | - |
7. Barclay CJ. Getting energy to where it is needed is the problem in the failing heart. J Physiol. 2008;586(21):5037-5038. https://doi.org/10.1113/jphysiol.2008.163246

8. Wacharapluesadee S, Tan CW, Maneenorn P, et al. Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. Nat Commun. 2021;12:972. https://doi.org/10.1038/s41467-021-21240-1

9. Dennehy JI. Evolutionary ecology of virus emergence. Ann N Y Acad Sci. 2017;1389(1):124-146. https://doi.org/10.1111/nyas.13304

10. Mahmoudabadi G, Milo R, Phillips R. Energetic cost of building a virus. Proc Natl Acad Sci U S A. 2017;114(22):E4333-E4333. https://doi.org/10.1073/pnas.1701670114

11. Desgranges C, Delhomme J. Towards a machine learned thermodynamics: exploration of free energy landscapes in molecular fluids, biological systems and for gas storage and separation in metal–organic framework. Mol Syst Des Eng. 2021;6:52. https://doi.org/10.1039/d0me00134a

12. Bauer D, Evilevitch A. Influence of internal DNA pressure on λ. J Mol Biol. 2021;457(20):3189-3200. https://doi.org/10.1016/j.jmb.2021.03.006

13. V’kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. J Phys Chem Ref Data Monogr. 2016;43(3):155-170. https://doi.org/10.1088/s41579-020-00468-6

14. Domalski ES, Hearing ED. Estimation of the thermodynamic properties of C-H-N-O-S-halogen compounds at 298.15 K. J Phys Chem Ref Data Monogr. 1993;22(4):805-1159. https://doi.org/10.1063/1.555927

15. WHO. Coronavirus disease (Covid-19) dashboard; 2020. Accessed May 30, 2021.

16. Gatti P, Ilamathi HS, Todkar K, Germain M. Mitochondria targeted viral replication and survival strategies – prospective on SARS-CoV-2. Front Pharmacol. 2020;11:1-9. https://doi.org/10.3389/fphar.2020.578599

17. Klein S, Cortese M, Winter SL, et al. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. Nat Commun. 2020;11(1):1. https://doi.org/10.1038/s41467-020-19619-7

18. Kuzmin A, Orekhov P, Astashkin R, Gordeliy V, Gushchin I. Structure and dynamics of the SARS-CoV-2 envelope protein monomer. BioRxiv. 2021;434722. https://doi.org/10.1101/2021.03.10.434722

19. Leventis PA, Grinstein S. The distribution and function of phosphatidylserine in cellular membranes. Annu Rev Biophys. 2010;39(1):407-427. https://doi.org/10.1146/annurev.biophys.093008.131234

20. Vance JE. Phospholipid synthesis and transport in mammalian cells. Traffic. 2015;16(1):1-18. https://doi.org/10.1111/tra.12230

21. Wang S, Trilling M, Sutter K, et al. A crowned killer’s resume: genome, structure, receptors, and origin of SARS-CoV. Virol Sin. 2020;35:673-684. https://doi.org/10.1007/s12025-020-00298-z

22. NCBI. Nucleotide Data Base. Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, co – Nucleotide – NCBI; 2020d. https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2/. Accessed May 2, 2020.

23. Lu S, Ye Q, Singh D, et al. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. Nat Commun. 2021;12(1):1-15. https://doi.org/10.1038/s41467-020-228023

24. NCBI. Nucleocapsid phosphoprotein (severe acute respiratory syndrome coronavirus 2); 2020c. https://www.ncbi.nlm.nih.gov/protein/179817425/. Accessed May 2, 2020.

25. NCBI. Envelope glycoprotein (severe acute respiratory syndrome coronavirus 2); 2020a. https://www.ncbi.nlm.nih.gov/protein/1796318600/. Accessed May 2, 2020.

26. NCBI. Membrane glycoprotein (severe acute respiratory syndrome coronavirus 2); 2020b. https://www.ncbi.nlm.nih.gov/protein/1796318601/. Accessed May 2, 2020.

27. NCBI. Surface glycoprotein (severe acute respiratory syndrome coronavirus 2); 2020e. https://www.ncbi.nlm.nih.gov/protein/1796318598/. Accessed May 2, 2020.

28. UniProt. E – Envelope small membrane protein – severe acute respiratory syndrome coronavirus 2 (2019-nCoV) – E gene & protein; 2021a. https://www.uniprot.org/uniprot/P0DTC4/. Accessed May 21, 2021.

29. UniProt. M – Membrane protein – severe acute respiratory syndrome coronavirus 2 (2019-nCoV) – M gene & protein; 2021b. https://www.uniprot.org/uniprot/P0DTC5/. Accessed May 21, 2021.

30. UniProt. S – Spike glycoprotein precursor – severe acute respiratory syndrome coronavirus 2 (2019-nCoV) – S gene & protein; 2021c. https://www.uniprot.org/uniprot/P0DTC2/. Accessed May 21, 2021.

31. de Queiroz NMGP, Marinho FV, Chagas MA, et al. Vaccines for COVID-19: perspectives from nucleic acid vaccines to BCG as delivery vector system. Microbes Infect. 2020;22:515-524. https://doi.org/10.1016/j.micinf.2020.09.004

32. Yao H, Song Y, Chen Y, et al. Molecular architecture of the SARS-CoV-2 virus. Cell. 2020;183(3):730-738. https://doi.org/10.1016/j.cell.2020.09.018

33. Turoňová B, Sikora M, Schürmann C, et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science. 2020;370(6513):203-208. https://doi.org/10.1126/science.abb5223

34. Singh P, Mishra N, Singh N, et al. Credible protein targets and curing strategies for COVID-19: a review. SN Compr Clin Med. 2020;2:1. https://doi.org/10.1007/s42399-020-00526-3

35. Godet M, L’Haridon R, Vautherot J-F, Laude H. TGEV coronavirus ORF4 encodes a membrane protein that is incorporated into virions. Virology. 1992;188(2):666-675. https://doi.org/10.1016/0042-6822(92)90521-3

36. Neuman BW, Kiss G, Kunding AH, et al. A structural analysis of the SARS-CoV-2 envelope protein as a strategy to assess cross-protection against COVID-19. Microbes Infect. 2020;24:5182-187. https://doi.org/10.1016/j.micinf.2020.05.013

37. Ingólfsson H, Melo MN, van Eerden FJ, et al. Lipid organization of the plasma membrane. J Am Chem Soc. 2014;136(41):14554-14559. https://doi.org/10.1021/ja507832e

38. Wagman DD, Evans WH, Parker VB, et al. The NBS Tables of Chemical Thermodynamic Properties Selected Values for Inorganic and C1 and C2 Organic Substances in SI Units. New York: American Chemical Society; 1990.
Methods. 2019;3(12):1900596. https://doi.org/10.1002/smt.201900596

75. Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S. The spike protein of SARS-CoV – a target for vaccine and therapeutic development. Nat Rev Microbiol. 2009;7(3):226-236. https://doi.org/10.1038/nrmicro2090

How to cite this article: Şimşek B, Özlilgen M, Utku FŞ. How much energy is stored in SARS-CoV-2 and its structural elements? Energy Storage. 2022;4(2):e298. doi:10.1002/est2.298