DOI: 10.14529/jsfi220304

# Calculation of Electrostatic Potential Field of Coronavirus S Proteins for Brownian Dynamics Simulations

*Ekaterina P. Vasyuchenko*[1], *Vladimir A. Fedorov*[1], *Ekaterina G. Kholina*[1], *Sergei S. Khruschev*[1], *Ilya B. Kovalenko*[1,2,3,4], *Marina G. Strakhovskaya*[1,2]



The Brownian dynamics method can give insight into the initial stages of the interaction of antiviral drug molecules with the structural components of bacteria or viruses. RAM of conventional personal computer allows calculation of Brownian dynamics of interaction of antiviral drugs with individual coronavirus S protein. However, scaling up this approach for modeling the interaction of antiviral drugs with the whole virion consisting of thousands of proteins and lipids is difficult due to high requirements for computing resources. In the case of the Brownian dynamics method, the main amount of RAM in the calculations is occupied by an array of values of the virion electrostatic potential field. When the system is increased from one S protein to the whole virion, the volume of data increases significantly. The standard protocol for calculating Brownian dynamics uses a three-dimensional grid with a spatial step of 1Å to calculate the electrostatic potential field. In this work, we consider the possibility of increasing the grid spacing parameter for calculating the electrostatic potential field of individual coronavirus S proteins. In this case, the amount of RAM occupied by the electrostatic potential field is reduced, which makes it possible to use personal computers for calculations. We performed Brownian dynamics simulations of interaction of an antiviral photosensitizer molecule with S proteins of three coronaviruses SARS-CoV, MERS-CoV, and SARS-CoV-2, and demonstrated that reduction of detalization of electrostatic potential field does not influence the results of Brownian dynamics much.

*Keywords: Brownian dynamics, coarse grain, spike protein, SARS-CoV-2, SARS-CoV, MERS-CoV, phthalocyanine, photosensitizer.*

## Introduction

Over the past twenty years, the world has suffered three outbreaks of viral diseases caused by coronavirus, which are expressed mainly in the defeat of the respiratory system and are called acute respiratory syndrome. The causative agents of these epidemics include coronaviruses SARS-CoV, MERS-CoV and SARS-CoV-2. The last pathogen (SARS-CoV-2) was the cause of the COVID-19 pandemic, which affected almost all countries of the world and claimed the lives of 6.33 million people (as of June 2022). Coronaviruses are enveloped viruses consisting of a nucleocapsid and an outer envelope. The outer envelope is represented by a lipid membrane and contains spike (S), membrane (M) and envelope (E) proteins. While E proteins and M proteins are involved in the assembly of virions, the S protein plays a key role in the early stages of eukaryotic cell infection, as it recognizes and binds to the host ACE2 receptor and is responsible for the penetration of the genetic material of the virus into the cell [7]. In this regard, the structures of S proteins are being actively studied, since it is a good target for the search for small molecules that would limit or make it impossible for coronavirus virions to bind to cells. At the moment, the spatial structures of the S proteins of the above viruses are already known. Previ-

[1]Lomonosov Moscow State University, Moscow, Russia
[2]Federal Scientific and Clinical Center of Specialized Types of Medical Care and Medical Technologies, Federal Medical-Biological Agency of Russia, Moscow, Russia
[3]Astrakhan State University, Astrakhan, Russia
[4]Institute of Computer Science and Mathematical Modeling, Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia

ously, a study of the interaction of zinc phthalocyanine and methylene blue molecules with the S protein of coronavirus was carried out using the Brownian dynamics method using the ProKSim software package, the main areas of landing of these small molecules on the protein surface and their interaction with key amino acids were shown [2]. In that study, the full-atomic structure of the S proteins of three coronaviruses was used, and the calculation of the electrostatic potential was carried out in 1Å steps. However, scaling up this approach for modeling the interaction of small molecules with potential antiviral activity with the structure of the whole virion is difficult due to high requirements for computer RAM. In the case of the Brownian dynamics method, the main amount of RAM in the calculations is occupied by an array of values of the electrostatic potential field. When the system is increased from one protein to the whole virion, the volume of the data increases significantly. The standard protocol for calculating Brownian dynamics uses a three-dimensional grid with a step of 1Å to calculate the electrostatic potential field. In this work, we will consider the possibility of increasing the grid spacing parameter for calculating the electrostatic potential field to 2Å using the examples of individual SARS-CoV, MERS-CoV, and SARS-CoV-2 S proteins. Thus, the amount of RAM occupied by the electrostatic potential field is reduced by a factor of 8, which makes it possible to use personal computers for modeling the interaction of antiviral drugs with the whole virion.

The article is organized as follows. Section 1 is devoted to the methods used in this work and the specifics of the design of models of studied molecules. In Section 2, we described and discuss the main results of this study. The Conclusion summarizes the results of the study and indicates directions for further work.

# 1. Methods

## 1.1. *3D* Protein Models

In this work, three-dimensional structures of S proteins of three human coronaviruses (MERS-CoV, SARS-CoV, SARS-CoV-2) were used. The spatial structure of the SARS-CoV-2 S protein was taken from [9]. The structures of the "heads" of the MERS-CoV and SARS-CoV S proteins were taken from the PDB (Protein Data Bank), PDB ID 6NB3 and 5X58. Unresolved remains in the "head" of the spike were completed using the i-TASSER program [10]. Based on the amino acid sequence from the UniProt database (A0A140AYW5 for MERS-CoV and P59594 for SARS-CoV), their secondary structure was predicted using the Jpred4 package [1]. Based on the predicted secondary structure, the unresolved parts of the "core" part of the S protein were completed using the Modeller-9.19 program [8] using the PDB ID 2WPQ as template.

The model of the Zn-PcChol$^{8+}$ molecule was taken from [6]. In that work, to determine possible positions of choline substitutes in the phthalocyanine macrocyclic core, we calculated Fukui functions, reflecting the probability of electrophilic attack of substituents to isoindole rings. Taking into account these results, the steric effects and the electrostatic repulsion of substituents, we determined the most plausible structure of Zn-PcChol$^{8+}$ with specific positions of choline substituents. Then quantum mechanical calculations were carried out in the framework of the density functional theory using PC GAMESS/Firefly software. We performed two-cycle geometry optimization of Zn-PcChol$^{8+}$ molecule using the SBKJC basis set with energy-consistent pseudopotential (ECP) and cc-PVDZ basis set for all atoms but Zn for the final optimization. Atomic partial charges were fit to reproduce the ab initio electrostatic potential using the RESP algorithm. Then we obtained all-atom topology for the predicted Zn-PcChol$^{8+}$ structure in the

GROMOS 54a7 force field using the online service Automated Topology Builder (ATB) and performed all-atomic MD for 50 ns. On the basis of auxiliary all-atom simulation models we created the CG model of Zn-PcChol$^{8+}$ molecule in the Martini force field. The force field parameters of bonded interactions of CG model obtained from the series of short CG simulations (10 ns) were iteratively optimized based on all-atom MD simulation. The distributions of bond and angle terms governing the equilibrium bond/angle and force constant values at each CG iteration step were compared with all-atom MD simulation until 85% coverage between all-atom and CG distributions. The final CG model of Zn-PcChol$^{8+}$ molecule consists of 62 CG beads, 50 bond terms, 18 constraints, 24 angle and 20 dihedrals terms.

To convert an all-atom structure of S protein to CG structure and create the suitable topology in Martini 3 force field, we used martinize2 program (`http://cgmartini.nl`). The general principle of Martini CG modeling of proteins is based on the suggestion that each amino acid residue consisted of one backbone bead and several side chain CG beads. Each CG bead represents group of 2–4 heavy atoms which belong to the certain chemical group. Martini CG force field suggests that the majority of CG bead types are neutrally charged, while few of them have +1 or –1 elementary electric charge.

## 1.2. Brownian Dynamics Simulation

For Brownian dynamics simulation, calculation of the electrostatic field and modeling of long-range electrostatic interactions, the ProKSim software package was used [4]. The values of the charges of coarse-grained models were taken according to the Martini-3.0.b.3.2 force field (`http://cgmartini.nl/index.php/martini3beta`). The values of the electrostatic potential created by the molecules were calculated on a cubic spatial grid using the Poisson-Boltzmann equation [3] (for more details, see [5]). The grid step was a parameter and was taken equal to 1Å or 2Å. Protein molecules and Zn-PcChol$^{8+}$ were described as rigid particles placed in a 30x30x30 nm cubic virtual reaction volume with mirror boundary conditions such that the TM (transmembrane) and CP (cytoplasmic) domains were outside the reaction volume for imitation of the viral membrane. Electric permittivity of proteins $\epsilon$=2, water $\epsilon$=80. The ionic strength of the solution was 100 mM, temperature T=300K. Water was taken into account as an implicit solvent with a water viscosity coefficient corresponding to a temperature of 300 K. The effect of thermal motion of water molecules is taken into account by introducing a random force with normal distribution acting on Brownian particles. Van der Waals interactions are taken into account using the excluded volume effect: particles cannot approach a shorter distance than the sum of their van der Waals radii. For each of the three S proteins, 20000 Brownian dynamics simulations were performed. In each simulation, the Zn-PcChol$^{8+}$ molecule was placed in a random position in the reaction volume and after diffusion and electrostatic interactions with S protein it approached the protein and formed encounter complex. In this process, the Zn-PcChol$^{8+}$ molecule diffused until the energy of electrostatic attraction in such a complex would be not less than specified threshold value of 8kT. The structure of the complex was then saved for further analysis. Visualisation of the complexes was carried out in PyMol-2.5.2 software (`https://pymol.org`).

## 2. Results and Discussion

The main amount of memory in the Brownian dynamics calculation is occupied by the electrostatic potential of the interacting particles and its gradient. In this work, we calculated the electrostatic potential for the S protein molecules of MERS-CoV, SARS-CoV, and SARS-CoV-2 coronaviruses on the spatial grid with cell size of 1Å and 2Å. Table 1 shows the quantitative characteristics of these calculations; in particular, the amount of memory occupied by the arrays for storing the electrostatic potential and its gradient, as well as the total virtual memory occupied by the program.

**Table 1.** The amount of memory occupied for each S protein/Zn-PcChol$^{8+}$ system

| | Grid spacing for ES potential calculation (Å) | Number of cells per edge of the cube of the ES potential field | Array size of ES potential (MB) | Size of the ES potential gradient array (MB) | Process virtual memory (MB) |
|---|---|---|---|---|---|
| MERS-CoV | 1 | 530 | 1136 | 3408 | 4586 |
| SARS-CoV | 1 | 542 | 1215 | 3645 | 4901 |
| SARS-CoV-2 | 1 | 544 | 1228 | 3684 | 4955 |
| MERS-CoV | 2 | 266 | 144 | 431 | 595 |
| SARS-CoV | 2 | 272 | 154 | 460 | 635 |
| SARS-CoV-2 | 2 | 272 | 154 | 460 | 635 |

RAM of conventional personal computer allows calculation of Brownian dynamics of interaction of antiviral drugs with individual coronavirus S protein. For whole SARS-CoV-2 virion envelope the amount of memory occupied would be as much as 150 GB, which significantly exceeds the memory of a typical personal computer. The amount of RAM required to store arrays of the electrostatic potential and its gradient is about 8 times less when using a 2Å grid compared to a 1Å grid. From Tab. 1 it follows that almost all the virtual memory of the program is occupied by the ES potential and its gradient.

We performed Brownian dynamics simulations using two grid spacing values of the electrostatic potential to reveal differences in the interaction of Zn-PcChol$^{8+}$ with coronavirus S proteins. Figure 1 shows individual electrostatic encounter complexes of the photosensitizer molecule with S proteins of MERS-CoV, SARS-CoV and SARS-CoV-2 coronaviruses calculated with two different values of electrostatic potential grid spacing, namely 1Å and 2Å. The position of the Zn-PcChol$^{8+}$ molecule in the encounter complex is represented by the zinc atom visualized with a marine-color sphere.

It can be seen that Zn-PcChol$^{8+}$ molecules mainly bind to the protein in the upper part of the stalk, at the junction of the stalk with the head, which is consistent with the results of previous studies [2]. Zn-PcChol$^{8+}$ molecules are distributed highly heterogeneously on the surfaces of S proteins, forming distinct dense areas. Size and population of each area are characterized by average pairwise RMSD of these molecules from each other, and the number of Zn-PcChol$^{8+}$ molecules forming it (Tab. 2). Positions of these areas on S protein surface are pretty the same in cases of 1Å and 2Å electrostatic potential grid. In general, the size of the areas slightly differs (no more than 0.3Å). With increased 2Å step of the grid of electrostatic potential, small and thinly populated areas located at the head of coronavirus S protein disappear, thus the population of large areas increases and the portion of Zn-PcChol$^{8+}$ bound with the stalk of coronavirus

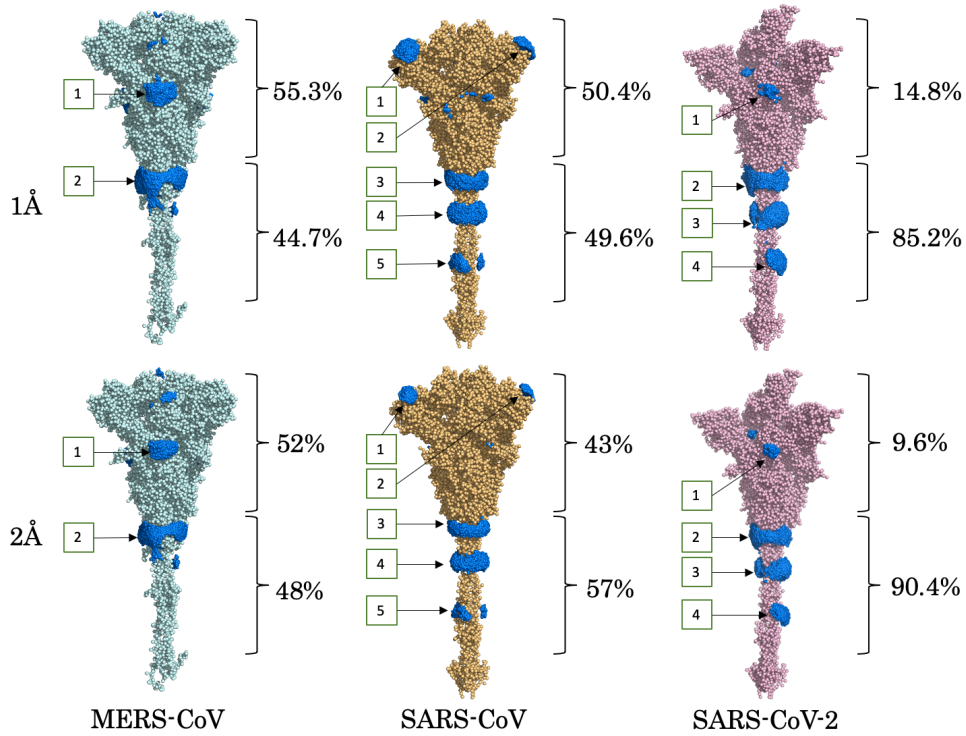


**Figure 1.** Encounter complexes of the Zn-PcChol$^{8+}$ photosensitizer molecule in complexes with S proteins of the studied MERS-CoV, SARS-CoV and SARS-CoV-2 coronaviruses at a grid step for calculating the electrostatic potential of 1Å (upper row) and 2Å (lower row) obtained by the Brownian dynamics method. Final positions of Zn-PcChol$^{8+}$ molecules obtained in individual BD simulations are represented by the zinc atoms (marine-color spheres). S proteins are presented as a coarse-grained model and colored, respectively, MERS-CoV – pale cyan, SARS-CoV – light orange, SARS-CoV-2 – light pink. Isolated Zn-PcChol$^{8+}$ binding areas are numbered. Ratio of Zn-PcChol$^{8+}$ bound with head and stalk of coronavirus is given in percent

**Table 2.** Percentage of structures forming individual binding areas and average pairwise RMSD for this area in Å (value given in parentheses)

| | Grid spacing for ES potential calculation (Å) | 1 binding area | 2 binding area | 3 binding area | 4 binding area | 5 binding area |
|---|---|---|---|---|---|---|
| MERS-CoV | 1 | 15.7 (1.8) | 44.7 (3.0) | | | |
| | 2 | 17.8 (1.8) | 48.0 (2.9) | | | |
| SARS-CoV | 1 | 12.9 (1.5) | 12.7 (1.4) | 22.5 (2.7) | 16.1 (2.4) | 11.0 (2.3) |
| | 2 | 14.9 (1.2) | 12.7 (1.1) | 22.9 (2.5) | 28.8 (2.2) | 5.4 (2.2) |
| SARS-CoV-2 | 1 | 9.7(1.1) | 47.4 (2.8) | 25.1 (2.3) | 12.8 (1.7) | |
| | 2 | 2.0 (1.1) | 49.7 (2.7) | 27.7 (2.2) | 12.9 (1.4) | |

S protein increases. The results show that bigger step of electrostatic potential grid leads to slight changes in distribution of Zn-PcChol$^{8+}$ molecules on coronavirus S proteins. Nevertheless,

the main areas of binding of photosensitizer molecules can be found using rough electrostatic potential field which takes not so much memory resources.

In calculations with different values of the grid spacing of the electrostatic potential, the regions of the protein interacting with Zn-PcChol$^{8+}$ do not change. However, the occupation of minor binding sites differs between simulations with two cell grid steps (1Å and 2Å), since in these positions the specified energy of electrostatic attraction (8kT) cannot be achieved due to the coarse representation of the electrostatic potential field when using 2Å resolution. Nevertheless, this does not change the general distribution of Zn-PcChol$^{8+}$. Thus, the coarsening of the spatial grid for calculation of electrostatic potential makes it possible to perform Brownian dynamics simulations using a smaller amount of RAM while obtaining a reliable result.

## Conclusion

With the advent of a large number of molecular spatial structures of biological objects, there is a need to create models of large systems, such as viral envelopes, microtubules, membrane pigment-protein complexes, lipopolysaccharide membranes, etc. However, the available computing resources remain limited. In this work, we evaluated the possibility of reducing the requirements for computer RAM by reducing the detail of the electrostatic potential of molecules, without losing the quality of the data obtained. In the case of the Brownian dynamics method, the main amount of RAM in the calculations is occupied by an array of values of the electrostatic potential field. We considered the possibility of increasing the grid spacing for calculating the electrostatic potential field to 2Å using the individual S proteins of SARS-CoV, MERS-CoV and SARS-CoV-2 and found that the main binding regions of the photosensitizer molecule can be found using a coarse electrostatic potential field, which requires not so many memory resources. Indeed, with a twofold increase in the grid spacing parameters, the amount of RAM occupied by the electrostatic potential field decreases by a factor of 8, which makes it possible to perform electrostatic potential and Brownian dynamics calculations on a personal computer for large molecular biological systems.



## References

1. Drozdetskiy, A., Cole, C., Procter, J., Barton, G.J.: JPred4: a protein secondary structure prediction server. Nucleic acids research 43(W1), W389–W394 (2015). https://doi.org/10.1093/nar/gkv332

2. Fedorov, V.A., Kholina, E.G., Khruschev, S.S., *et al.*: What binds cationic photosensitizers better: Brownian dynamics reveals key interaction sites on spike proteins of SARS-CoV, MERS-CoV, and SARS-CoV-2. Viruses 13(8), 1615 (2021). https://doi.org/10.3390/v13081615

3. Fogolari, F., Brigo, A., Molinari, H.: The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology. Journal of Molecular Recognition 15(6), 377–392 (2002). https://doi.org/10.1002/jmr.577

4. Khruschev, S.S., Abaturova, A.M., Diakonova, A.N., *et al.*: Brownian-dynamics simulations of protein–protein interactions in the photosynthetic electron transport chain. Biophysics 60(2), 212–231 (2015). https://doi.org/10.1134/S0006350915020086

5. Kovalenko, I.B., Khruschev, S.S., Fedorov, V.A., *et al.*: The role of electrostatic interactions in the process of diffusional encounter and docking of electron transport proteins. Doklady Biochemistry and Biophysics 468(1), 183–186 (2016). https://doi.org/10.1134/S1607672916030066

6. Orekhov, P.S., Kholina, E.G., Bozdaganyan, M.E., *et al.*: Molecular mechanism of uptake of cationic photoantimicrobial phthalocyanine across bacterial membranes revealed by molecular dynamics simulations. The Journal of Physical Chemistry B 122(14), 3711–3722 (2018). https://doi.org/10.1021/acs.jpcb.7b11707

7. Schoeman, D., Fielding, B.C.: Coronavirus envelope protein: current knowledge. Virol J. 16(1), 1–22 (2019). https://doi.org/10.1186/s12985-019-1182-0

8. Webb, B., Sali, A.: Comparative protein structure modeling using modeller. Current protocols in bioinformatics 54(1), 5–6 (2016). https://doi.org/10.1002/cpbi.3

9. Woo, H., Park, S.J., Choi, Y.K., *et al.*: Developing a fully glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane. The Journal of Physical Chemistry B 124(33), 7128–7137 (2020). https://doi.org/10.1021/acs.jpcb.0c04553

10. Zheng, W., Zhang, C., Li, Y., *et al.*: Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. Cell reports methods 1(3), 100014 (2021). https://doi.org/10.1016/j.crmeth.2021.100014