

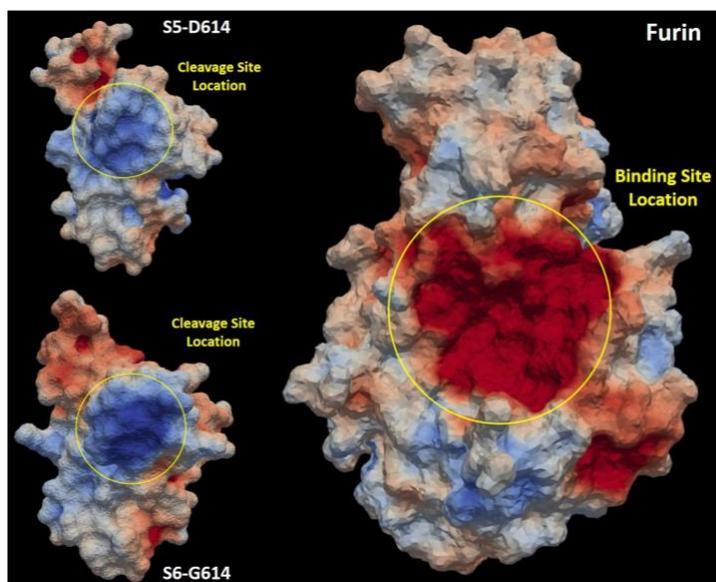
## DRAFT: Structural and Electrostatic Differences of the SARS-CoV-2 D614G Mutation in the S1/S2 Furin Cleavage Domain

CRTC students (Angelos, Spiros, Chris, ...) and Nikos Chrisochoides  
Center for Real-Time Computing, ODU

Kurt Deshayes, John Kenney  
Antibody Solutions, Santa Clara, CA

### Abstract

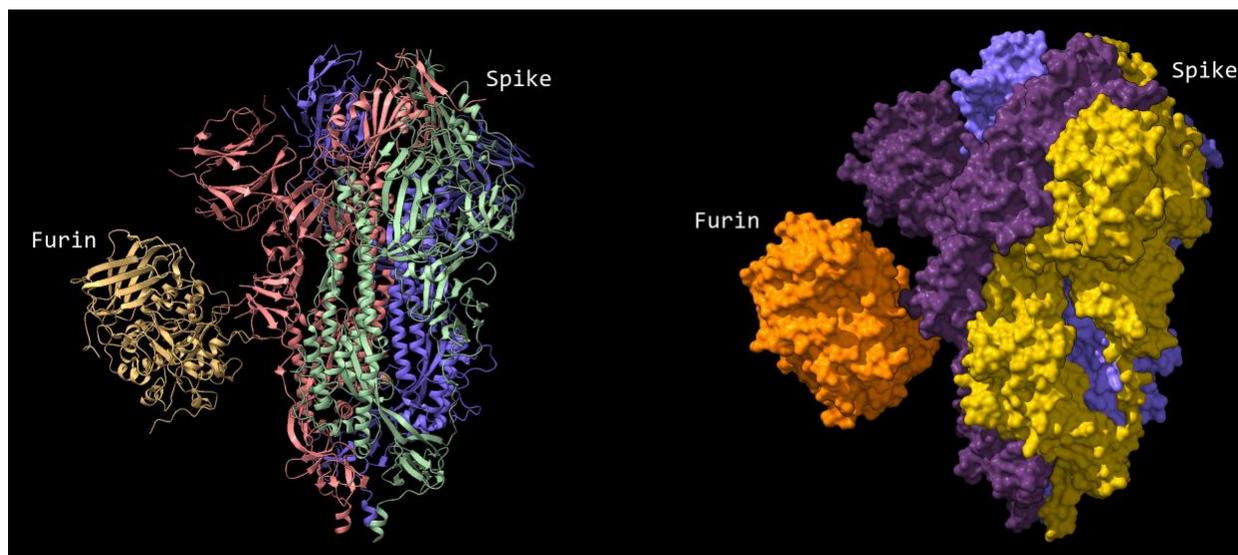
The spike protein mediates the attachment of the virus to the human cell via binding to an extracellular ACE2 protein with the processing of the S1 and S2 subunits leading to membrane fusion and virus entry into the cell. It has been shown that a key component of the spike protein is the cleavage at the *recently evolved* Furin site 682-RRARS-686 with cleavage occurring between R685 and S686. Specifically, it has been observed that the G614 mutant is more infective in vivo and in vitro. The focus of this paper is to computationally investigate how G614 mutation changes the reactivity of the Furin cleavage site, and potentially the rate of virus processing and infection. Namely, we show that the G614 (DG614) mutation leads to large-scale changes in electrostatic potential along with many other structural changes that may increase the attraction of Furin to the cleavage site thereby increasing the rate of virus processing and infection.



### Introduction

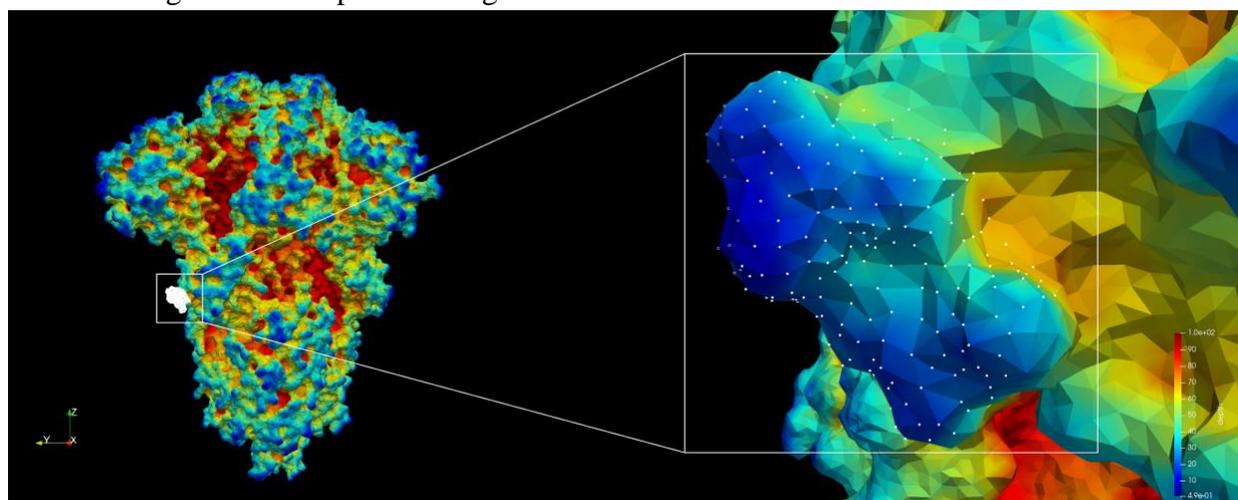
As the SARS-Cov2 virus has passed through human hosts it has evolved from a virus primarily targeting Microchiroptera to one optimized for the infection of Homo sapiens. The mutation of the Aspartic acid at position 614 on the spike protein to Glycine has been observed to correlate with the number of passages. It has also been observed that G614 mutant is more infective in vivo and in vitro, and recently in a clinical setting. Such an increase in infectivity explains why the G614 mutant came to dominate the virus population in the first months of the pandemic. The spike protein mediates the attachment of the virus to the cell via binding to an extracellular ACE2 protein

with the processing of the S1 and S2 subunits leading to membrane fusion and virus entry into the cell. It has been shown that a key component in all these steps is the cleavage of the spike protein at the recently evolved Furin cleavage site 682-RRARS-686 with cleavage occurring between R685 and S686 (see Fig. 1). It has been postulated the addition of this cleavage site allowed the virus to jump species. We and others (Ref) have computationally investigated how the G614 mutation changes the reactivity of the Furin cleavage site, and potentially the rate of virus processing and infection. We show here that the change leads to a number of structural changes in addition to large-scale changes in electrostatic potential in the cleavage site that may increase the attraction of Furin and thereby increase the rate of virus processing and infection.



**Figure 1:** The ribbon (left) and surface (right) visualization of the interaction of the furin and the spike in the cleavage site.

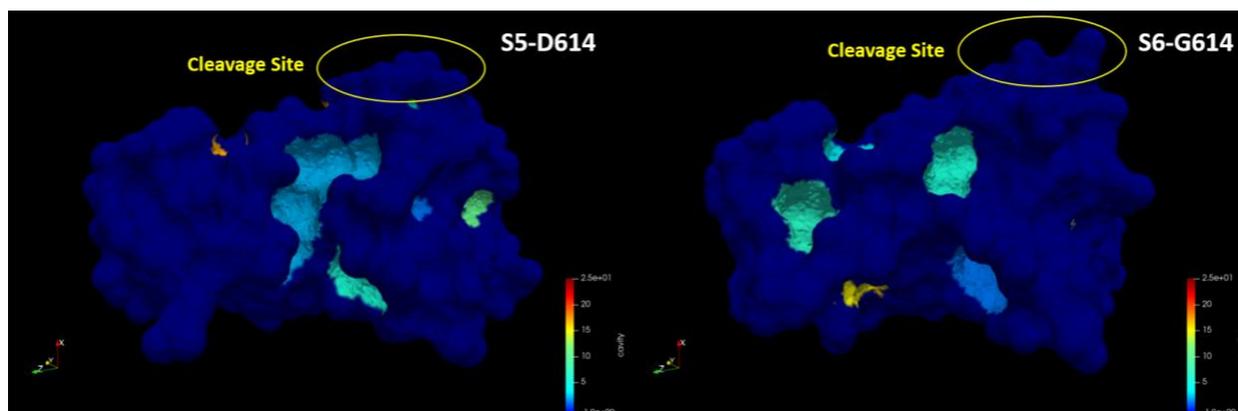
The focus of this paper is to computationally study and report preliminary findings on the impact of G614D mutation in terms of structural and electrostatic changes to SARS-CoV-2 in the S1/S2 Furin cleavage domain depicted in Fig. 2.



**Figure 2:** The depth zoom-out (left) and zoom-in (right) visualization of the furin cleavage site of the spike. Blue colors depict lower surface depth (extrusions), while colors closer to red depict higher depth (cavities).

## Background

Protein pockets or cavities are recessions in a protein's surface, and they can be extracted from PDB data using a Solvent-Excluded Surface (SES). Cavities can take on many different shapes and vary in-depth as it is shown in Figure 3, for Furin (right) S5-D614 and S6-G614 (left) of SARS-CoV-2 in the S1/S2 Furin cleavage domain, for the original and mutated strains, respectively.



**Figure 3:** Drastic differences between the number and type of cavities we detected in the D614 (left) and G614 (right) spike furin cleavage domains. Each cavity has a different color as shown in the legend, based on its number.

Cavities in general are classified as clefts, invaginations, tunnels, channels, or voids [1]. Cavities are important in protein-protein interaction as the active sites of proteins are often located inside some cavities [1]. In drug discovery, one of the first steps is to screen proteins of interest for cavities that can work as binding sites. Then, various ligands are considered for possible interaction with these pockets [2]. Choosing ligands must consider the shape of the target cavities [2]. Emil Fisher as early as 1894 realized that a cavity is like a keyhole, and a ligand is a key. It is important to find the right key that minimizes the energy barrier for the interaction and maximizes its speed and stability [15].

For both a cavity and a ligand, their geometry and various biochemical and biophysical properties are important for analyzing their interaction [1]. In terms of geometry, their shape, depth, and overall size need to be considered. Biochemical and biophysical properties of interest, among others, include pH, hydrophobicity, polarity, and electrostatic potential. All these factors play a role in enabling the interaction between two different molecules [1]. In this paper, we will give special consideration to structural features like cavities created by the D614G mutation and electrostatic potential, which is one of the most important factors in protein-protein interaction [3].

There are several types of algorithms for detecting protein cavities. One of these categories is geometry-based, which is subdivided into several subcategories, the primary being: sphere-, grid-

, surface-, and tessellation-based. In this paper, we use a High-Performance Computing (HPC) tessellation-based method developed at the Center for real-Time Computing (CRTC) over the last twenty-five years for applications in Medical Image Computing [5, 6] and the Aerospace Industry [7].

Most tessellation-based algorithms for protein cavity detection utilize Delaunay triangulation in one way or the other. For most of these, Delaunay triangulation is a first step from which later alpha shapes, beta shapes, or Voronoi diagrams are computed. Alpha shapes are used to get a surface approximation from a protein atom cloud. Beta shapes are like alpha shapes, but are a generalization, which in this case utilize the Van der Waals radii of atoms. Voronoi tessellation is the dual of the Delaunay triangulation and is often used as a graph on which path finding is performed to identify tunnels and channels in proteins. Some notable tessellation-based algorithms make use of Voronoi.

A major shortcoming of currently available algorithms for detecting protein cavities is that they are not HPC i.e., are not parallelized. This means that their runtime performance for very large proteins is poor. In addition to this, not many algorithms work on multi-model PDB files. The lack of this feature effectively bars such algorithms from working on large protein complexes. One way to partially overcome this limitation is for the user to break up large proteins and feed each model individually. However, this can be computational-intensive and will likely cause the algorithms to miss cavities that may be formed on the intersections of models. Consequently, these algorithms are not well suited for large protein complexes. A method that accurately works for all protein sizes, in real-time, needs to be developed.

## **Material and Methods**

In this paper, we use PDB data and a new High-Performance Computing (HPC) algorithm, which is built around the intuitive concept that the surface of pockets and cavities has a greater distance from the protein boundary compared to other parts of the protein. The core of the algorithm consists of (1) creating a tessellated representation of the SES using known and well-understood mesh generation technologies developed for Finite Element/Volume Analysis (2) computing the unsigned distance of the tessellated SES from an exterior triangulated surface that envelops the protein (similar to alpha shapes), and (3) clustering vertices based on unsigned distance value thresholds in order to identify pockets and cavities. The result of the algorithm is a tessellated solvent-excluded surface with unsigned distance labels and identified pockets and cavities.

### **Materials**

The PDB files used for this analysis are I-TASSER-derived protein models using as a template the SARS-CoV-2 spike protein [4] with PDB ID 5X58. This template is the highest resolution PDB structure with the highest homology to SARS-CoV-2 [4]. The AA region of 591 to 710 was analyzed because it contains the site of the DG614 mutation and the Furin cleavage site with a

secondary structure nearby both the sites [4]. The two PDB files are referred to as S5-D614 and S6-G614 [4]. The former is from the original D614 SARS-CoV-2 strain whereas the latter is from the newer G614 strain. The template modeling (TM) score, a metric used to measure the similarity of two protein structures [12], for S5-D614 and S6-G614 is 0.59 and 0.58 respectively [4]. A model with a TM score greater than 0.5 is regarded as topologically correct [4].

## Methods

*Geometric Protein Cavity Detection (GPCD)* is a high-performance software for detecting protein pockets and cavities, which are referred to as geometric features. The algorithm was designed to detect all types of geometric features, including clefts, invaginations, tunnels, and channels. A detailed technical description and analysis will appear elsewhere [8]. The key advantage of GPCD over other pocket detection algorithms is that it can leverage parallel computing (i.e. can use many processing CPU and/or GPU cores). GPCD is designed from the ground up to leverage parallelism across most of its runtime. For even greater performance, it will be extended to utilize the GPU as well as to work in a distributed environment. Parallelism grants GPCD a near-linear speedup to improve end-user productivity. The parallel nature of GPCD makes it capable of processing large proteins and protein complexes in real-time, something those other algorithms cannot handle. The algorithm can work with proteins of all sizes and has an edge on larger ones.

GPCD uses parallel mesh generation and ray casting to detect geometric features. Parallel mesh generation [8] is used to create a tessellated mesh representation of a solvent-excluded surface. A volumetric image is first created from a protein PDB file, and then Euclidean distance transform is used to create a solvent-excluded image [10]. This grid is then used by the Parallel Optimistic Delaunay Mesher (PODM) [9] to create a mesh of the solvent-excluded surface. Ray casting and clustering are employed to detect geometric features. Once the cavities are detected, several properties can be assigned, such as electrostatic potential.

Detecting pockets and cavities is just the first step in the process. It is important to quantitatively analyze detected features to understand their function and how they affect the protein. GPCD can compute and map physicochemical properties to the protein surface. Some advanced properties, like electrostatic potential (ESP), are imported from other software. In the case of GPCD, electrostatic potential is retrieved from the Adaptive Poisson-Boltzmann Solver (APBS) and mapped to the surface [13]. GPCD can compute various statistics for ESP. These statistics can be computed for the entire surface and for each individual cavity. ESP is the only property currently available with GPCD, but other properties such as hydrophobicity and pH will be added in the near future.

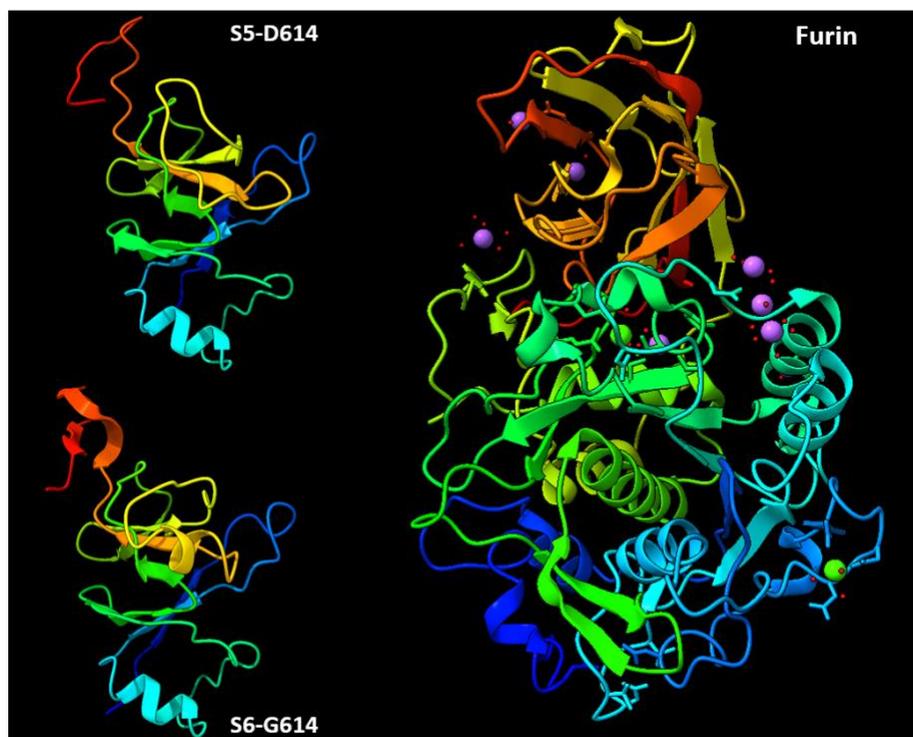
*Validation:* GPCD's results compared with the four widely-used software tools for analyzing structures in biomacromolecules: (i) MOLE 2.0 [Ref], MolAxis [Ref], and CAVER 3.0 [Ref]. A detailed analysis will appear elsewhere [GPCD-Ref]. In this section, we summarize the results pertinent to SARS-CoV-2 D614G Mutation in the S1/S2 Furin Cleavage Domain.

## Results

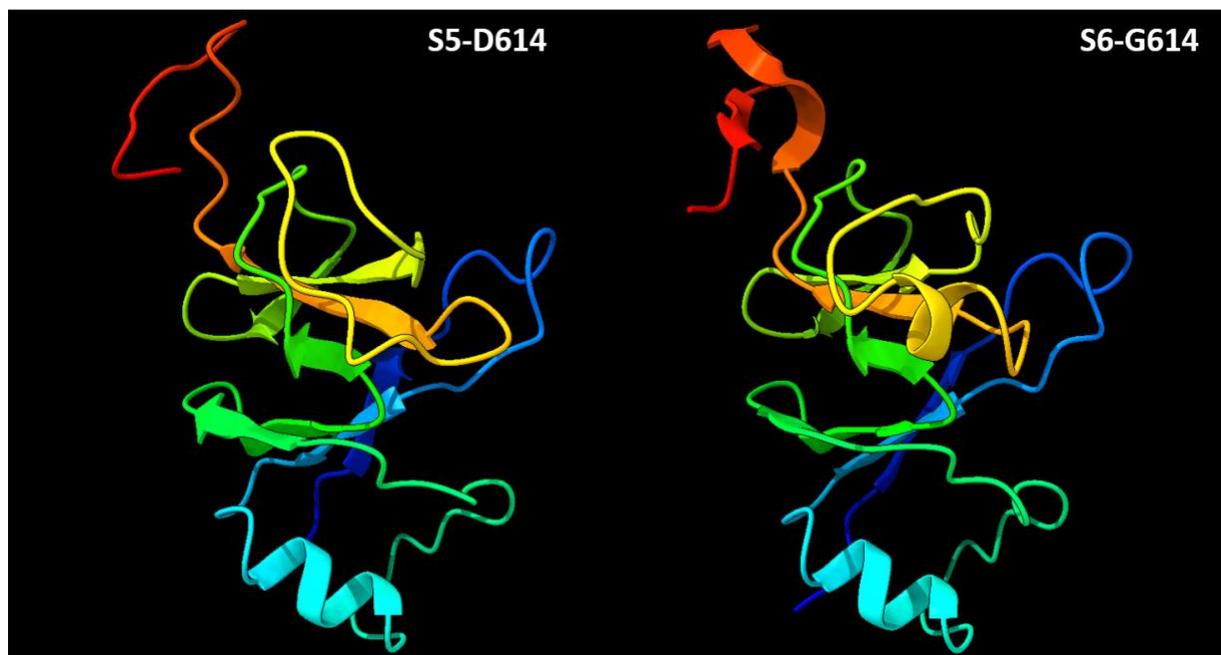
The GPCD algorithm/software is used to create a protein surface mesh using PODM and to identify pockets and other cavities. Quantitative data for D614 and G614 spike furin cleavage domains are summarized in Table 1.

**Table 1:** Lists the number of each type of cavity in the D614 and G614 spike furin cleavage domains.

Cavity Type	S5-D614	S6-G614
Cleft	11	13
Invagination	2	1
Tunnel	1	2
Void	3	2
Channel	0	0



**Figure 4:** Shows the secondary structure of the S5-D614 and S6-G614 cleavage domains (left) and furin (PDB 5JXG).



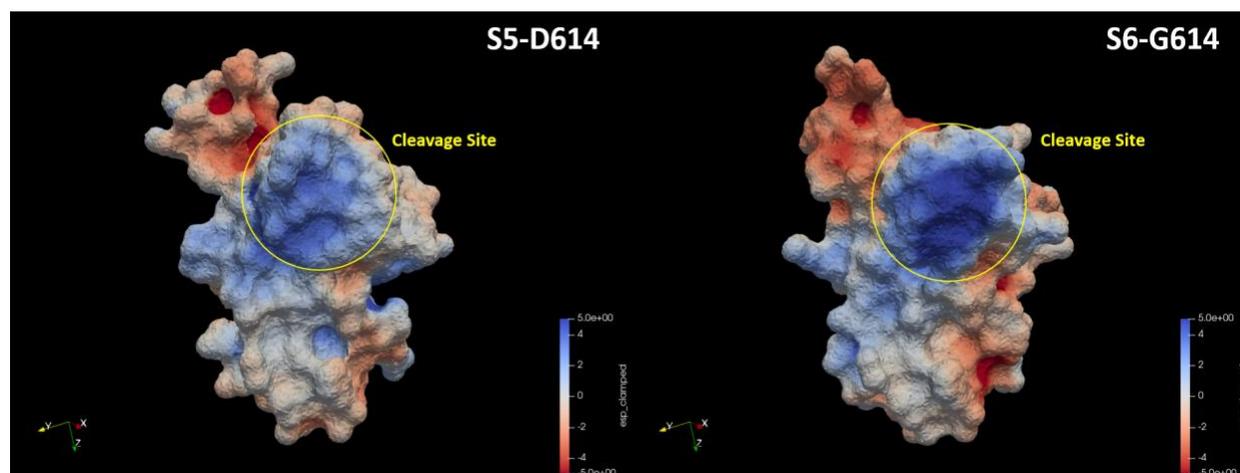
**Figure 5:** Secondary structure of the SARS-CoV-2 D614 (left) and G614 (right) furin cleavage domain.

It has been shown in a clinical study that patients infected with the G614 strain have a higher viral load than those infected with the D614 strain [14]. There are several hypotheses as to why the new strain appears to be more infectious [4]. One theory is that an increased rate of virus processing and cell entry leads to an increased rate of infection and more efficient reproduction of the virus. An increase in the rate of Furin cleavage of the glycoprotein spike before binding with ACE2 would lead to such an outcome, and therefore examining the effects of the D614G mutation on the properties of the Furin cleavage site could yield insight into the validity of this mechanism. Here, we explore how changes in electrostatic potential interaction between the cleavage domain and Furin may indeed lead to more efficient spike protein processing, which increases viral infectivity. The new strain has a more electrostatically positive cleavage site than the original strain, which may lead to faster and more energetically efficient interaction with the electronegative binding region of Furin.

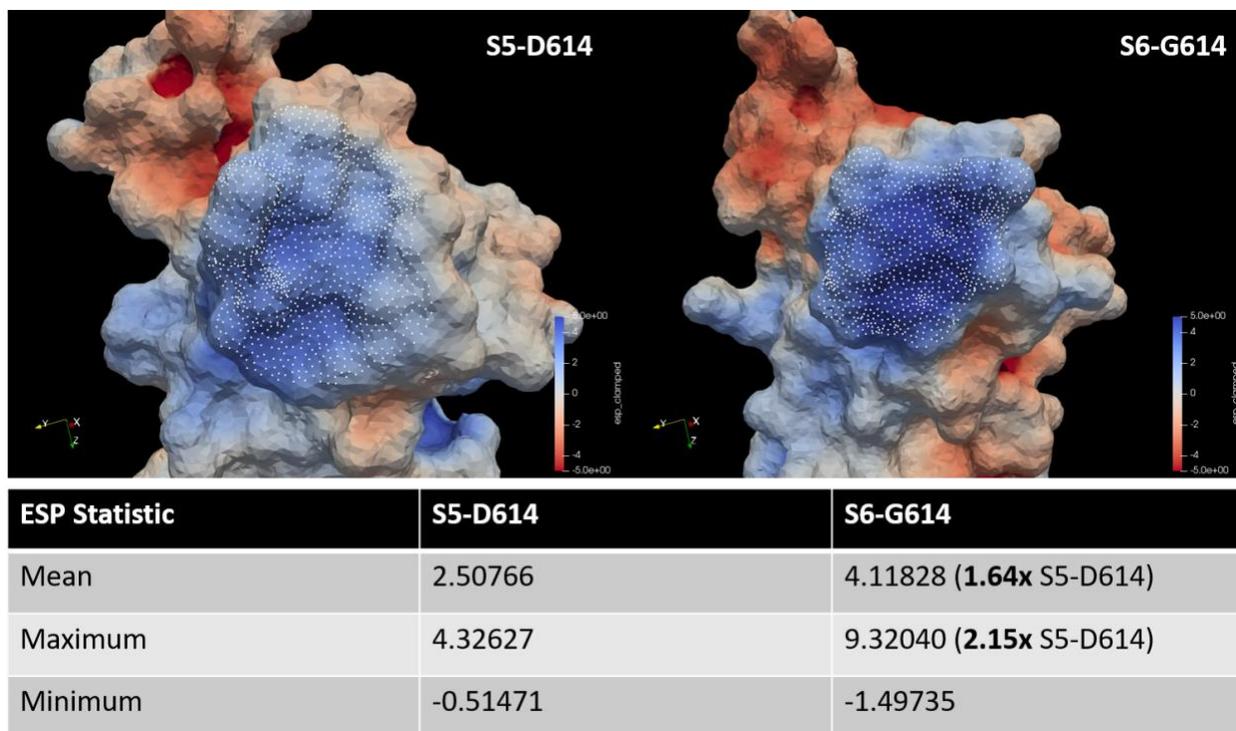
The electrostatic potential is important for protein-protein interaction, protein folding, and enzyme catalysis [6]. It is one of the factors that enables two proteins to be attracted to each other. For this to happen, one must be electropositive, and the other electronegative. Due to this, electrostatic potential is the best candidate for explaining why enzyme catalysis occurs, and in fact, it happens almost exclusively because of electrostatics [11]. Electrostatics plays a major role in the long-range attraction of molecules [6]. There is a large variety of proteins present both inside and outside a cell [6]. As such, a way for proteins to quickly recognize each other is necessary. This recognition process is enhanced by long-range electrostatics, and electrostatic forces bring two proteins close to each other [6]. This guidance of two proteins to each other is called electrostatic steering [6].

Greater electro positivity indicates a greater area of effect for electrostatic interaction. With Furin having an electrostatically negative binding region, this greater positive region could benefit the interaction of the two molecules, with more efficient attraction and electrostatic steering leading to faster cleavage. Furthermore, as can be seen in Figure 6, the cleavage site in the G614 strain is more electrostatically positive than in the older strain, which might also indicate stronger attraction. Figure 7 shows that the mean electrostatic potential in the G614 strain is 1.64 times greater than in the D614, whereas the maximum value is 2.15 times greater.

APBS was used to compute an electrostatic potential grid which was then mapped by GPCD to its protein mesh output using trilinear interpolation. The results were computed on an 8-core/16 GB RAM desktop computer with an approximate time of 4 seconds per protein, using a very high-resolution solvent-excluded surface. Using lower resolution equivalents, the process took about 1.3 seconds. APBS computation time is not included in these timings.

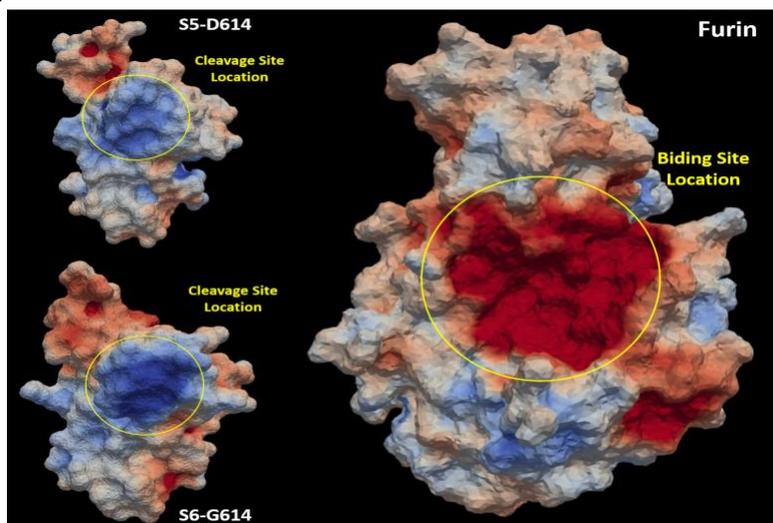


**Figure 6:** Electrostatic potential surface for the SARS-CoV-2 D614 (left) and G614 (right) furin cleavage domain. The cleavage site is marked with a yellow circle. As seen, the cleavage site in S6-G614 is more positive.

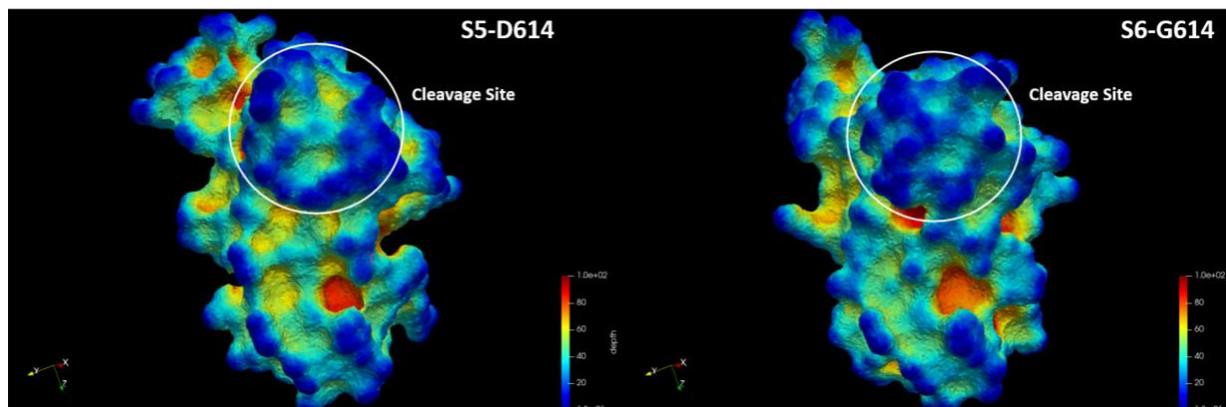


**Figure 7:** Electrostatic potential statistics for the SARS-CoV-2 D614 (left) and G614 (right) furin cleavage site. The statistics are computed for the points marked in white.

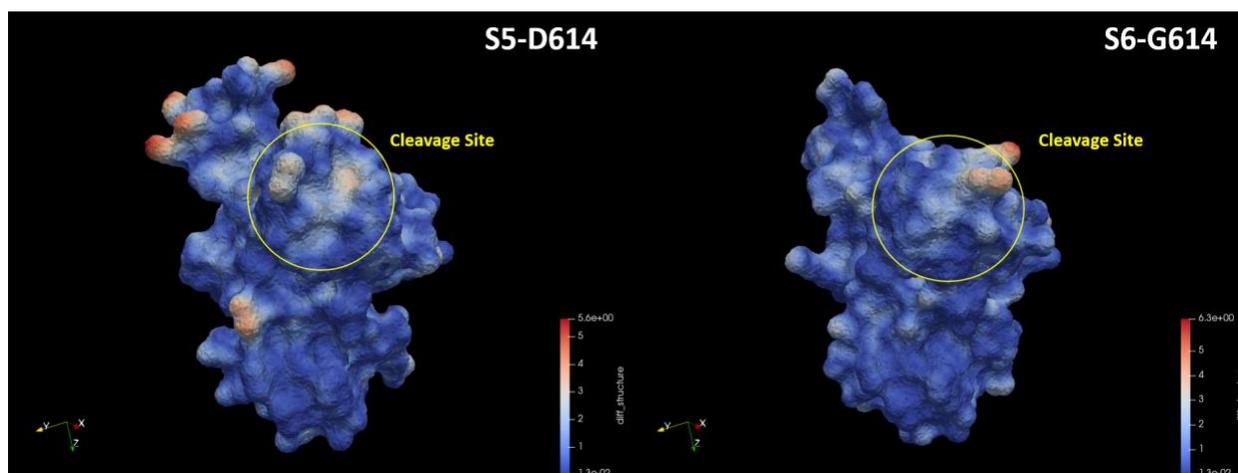
In the G614 strain, the region of the Furin cleavage site and some nearby regions are more positive, whereas side regions are more negative. The hypothesis is that this change can benefit Furin cleavage. Furin may be able to more easily locate and position itself for cleavage because of a stronger positive attraction in the cleavage site and more negative surrounding areas, which promotes repulsion from those regions. This setup can allow Furin to more effectively be guided to the correct location via the means of electrostatic steering, which can lead to faster cleavage and overall greater infectivity. Figure 8 shows the cleavage domains of the two strains together with Furin side-by-side.



**Figure 8:** Shows the S5-D614 and S6-G614 cleavage domains (left) and furin (PDB 5JXG) with the approximate binding site location highlighted (right).



**Figure 9:** Surface depth visualization of the SARS-CoV-2 D614 (left) and G614 (right) furin cleavage domain. The cleavage site is marked with a white circle. Blue colors indicate low surface depth, while yellow and red indicate greater surface depth.



**Figure 10:** Surface geometry differences of S5-D614 from S6-G614 (left), S6-G614 from S5-D614 (right). The colorization reflects the Euclidean Distance of the displayed protein surface from each other. Greater distances have a colorization closer to red.

Apart from electrostatic differences, it is also possible that differences in the geometry near the cleavage site play a role in more efficient cleavage. Figure 9 shows a depth visualization of the surface of the cleavage Furin cleavage domains for the D614 and G614 strains of SARS-CoV-2. In the D614 strain, there are overall more recessions near the cleavage site, whereas in the G614 strain these areas are flatter. Figure 10 visualizes the surface differences in the Furin cleavage domain of the two strains. In the new strain, there are some extruding atoms on the top right of the cleavage site. It is possible that the positions of these atoms can allow for a stronger binding between Furin and the cleavage site, promoting more stability. However, this cannot easily be quantified.

## Conclusion

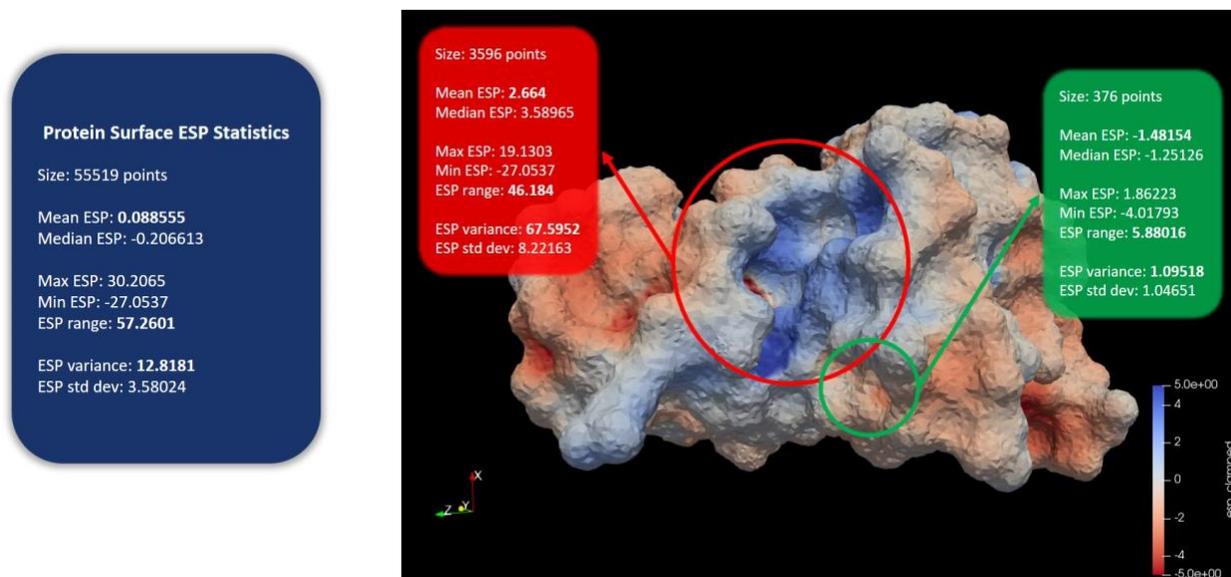
Our findings and analysis support the hypothesis and observations that the SARS-CoV-2 G614 strain is more infectious [14] than the original D614 Wuhan strain. The changes in electrostatic potential in the G614 strain's Furin cleavage domain imply easier interaction between Furin and the spike protein's cleavage site. The high electronegativity of the region in and around the Furin catalytic site leads to a lesser energy barrier when interacting with the more electropositive G614 Furin cleavage site.

## Discussion

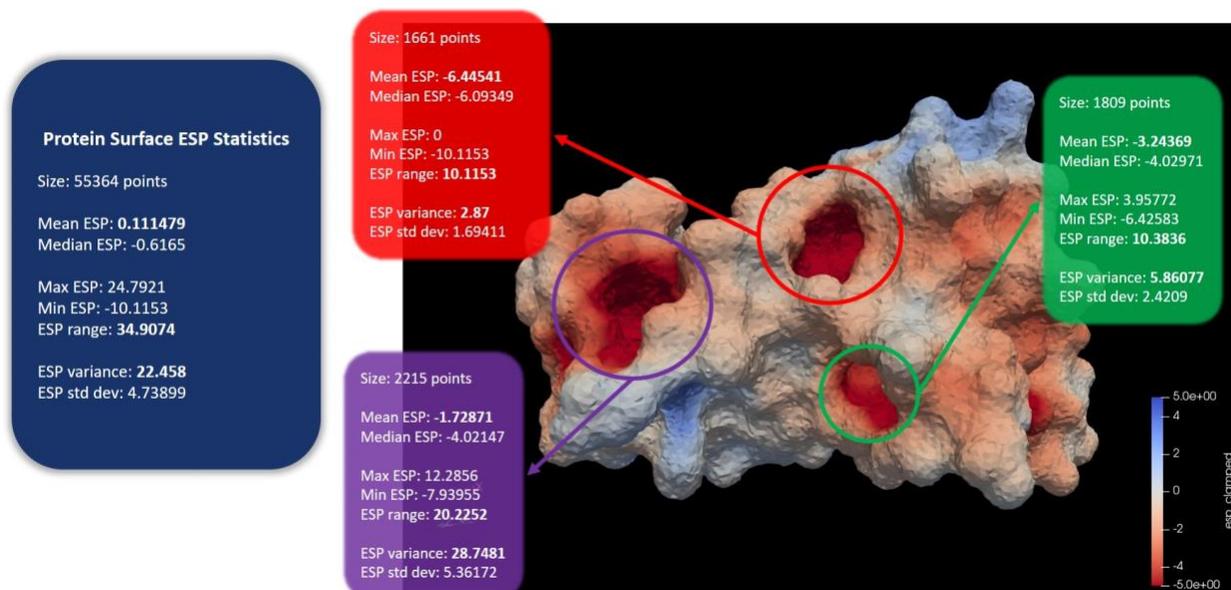
## Acknowledgments

This research was sponsored in part by NSF grant no. CCF-1439079 and Richard T. Cheng Endowment. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Government.

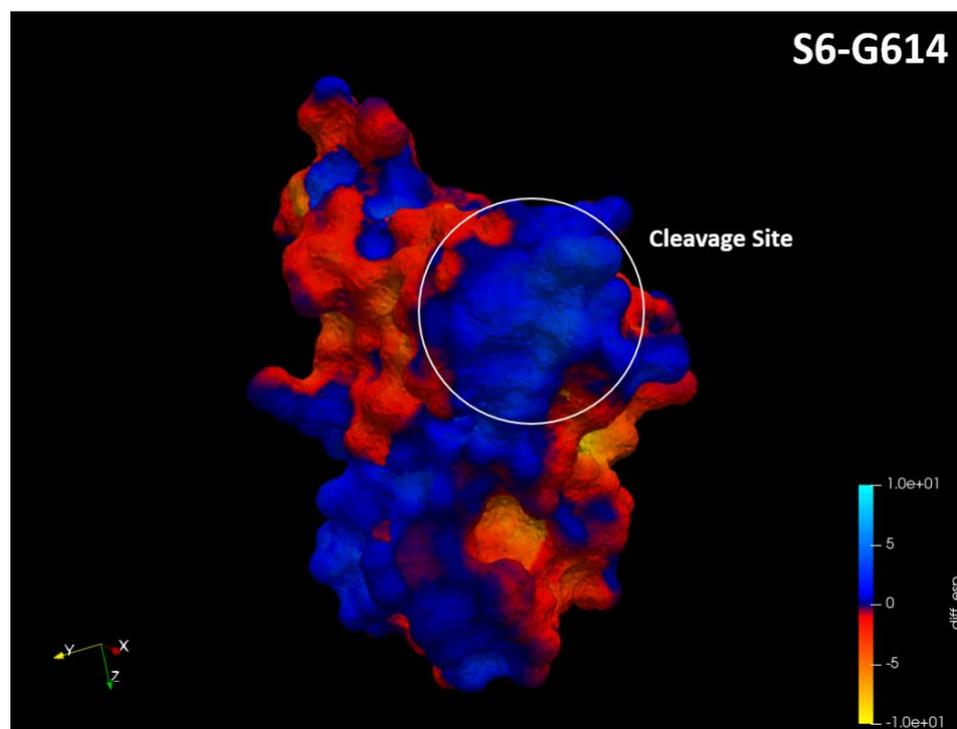
## Appendix: Supplementary Material



**Figure A1:** Shows an angle of the electrostatic potential (ESP) surface of the D614 spike furin cleavage domain (S5-D614 [4]). Electrostatic potential statistics for some cavities are displayed, along with ESP statistics for the entire protein surface.



**Figure A2:** Shows an angle of the electrostatic potential (ESP) surface of the G614 spike furin cleavage domain (S6-G614 [4]). Electrostatic potential statistics for some cavities are displayed, along with ESP statistics for the entire protein surface.



**Figure A3:** Electrostatic potential difference of S6-G614 from S5-D614 (right). Positive differences have a blue and aqua color, while negative differences have a red and yellow color. Blue means that the region is more positive, while red means that the region is more negative. In S6-G614, the cleavage site is more positive than in S5-D614, while nearby regions are more negative.

## References

- [1] Simões, T., Lopes, D., Dias, S., Fernandes, F., Pereira, J., Jorge, J., Bajaj, C., & Gomes, A. (2017). Geometric Detection Algorithms for Cavities on Protein Surfaces in Molecular Graphics: A Survey. *Computer graphics forum : journal of the European Association for Computer Graphics*, 36(8), 643–683. <https://doi.org/10.1111/cgf.13158>
- [2] Coleman, R. G., & Sharp, K. A. (2010). Protein pockets: inventory, shape, and comparison. *Journal of chemical information and modeling*, 50(4), 589–603. <https://doi.org/10.1021/ci900397t>
- [3] Vascon, F., Gasparotto, M., Giacomello, M., Cendron, L., Bergantino, E., Filippini, F., & Righetto, I. (2020). Protein electrostatics: From computational and structural analysis to discovery of functional fingerprints and biotechnological design. *Computational and Structural Biotechnology Journal*, 18, 1774-1789. <https://doi.org/10.1016/j.csbj.2020.06.029>
- [4] Tang, L.; Schulkins, A.; Chen, C.; Deshayes, K.; Kenney, J.S. (2020). The SARS-CoV-2 Spike Protein D614G Mutation Shows Increasing Dominance and May Confer a Structural Advantage to the Furin Cleavage Domain. Preprints , 2020050407. <https://doi.org/10.20944/preprints202005.0407.v1>
- [5] Chrisochoides, N., Fedorov, A., Kot, A., Archip, N., Black, P., Clatz, O., . . . Warfield, S. (2006). Toward Real-Time Image Guided Neurosurgery Using Distributed and Grid Computing. *ACM/IEEE SC 2006 Conference (SC'06)*. <https://doi.org/10.1109/sc.2006.65>
- [6] Archip, N., Clatz, O., Whalen, S., Kacher, D., Fedorov, A., Kot, A., Chrisochoides, N., Jolesz, F., Golby, A., Black, P. M., & Warfield, S. K. (2007). Non-rigid alignment of pre-operative MRI, fMRI, and DT-MRI with intra-operative MRI for enhanced visualization and navigation in image-guided neurosurgery. *NeuroImage*, 35(2), 609–624. <https://doi.org/10.1016/j.neuroimage.2006.11.060>
- [7] Drakopoulos, F., Tsolakis, C., & Chrisochoides, N. P. (2019). Fine-Grained Speculative Topological Transformation Scheme for Local Reconnection Methods. *AIAA Journal*, 57(9), 4007-4018. <https://doi.org/10.2514/1.j057657>
- [8] Chrisochoides, N. (2006). Parallel Mesh Generation. *Lecture Notes in Computational Science and Engineering Numerical Solution of Partial Differential Equations on Parallel Computers*, 237-264. [https://doi.org/10.1007/3-540-31619-1\\_7](https://doi.org/10.1007/3-540-31619-1_7)
- [9] Foteinos, P.A. & Chrisochoides, N. P. (2014). High quality real-time Image-to-Mesh conversion for finite element simulations. *Journal of Parallel and Distributed Computing*, 74(2), 2123-2140. <https://doi.org/10.1016/j.jpdc.2013.11.002>
- [10] Xu, D., Zhang, Y. (2009). Generating Triangulated Macromolecular Surfaces by Euclidean Distance Transform. *PLoS ONE* 4(12): e8140. <https://doi.org/10.1371/journal.pone.0008140>

- [11] Warshel, A., Sharma, P., Kato, M., Xiang, Y., Liu, H. and Olsson, M. (2006). Electrostatic Basis for Enzyme Catalysis. *Chemical Reviews*, 106(8), pp.3210-3235. <https://doi.org/10.1021/cr0503106>
- [12] Xu, J., & Zhang, Y. (2010). How significant is a protein structure similarity with TM-score = 0.5?. *Bioinformatics (Oxford, England)*, 26(7), 889–895. <https://doi.org/10.1093/bioinformatics/btq066>
- [13] Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., ... Baker, N. A. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein science : a publication of the Protein Society*, 27(1), 112–128. <https://doi.org/10.1002/pro.3280>
- [14] Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., . . . Wyles, M. D. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*, 182(4). <https://doi.org/10.1016/j.cell.2020.06.043>
- [15] Majewski, M., Ruiz-Carmona, S., & Barril, X. (2019). An investigation of structural stability in protein-ligand complexes reveals the balance between order and disorder. *Communications Chemistry*, 2(1). <https://doi.org/10.1038/s42004-019-0205-5>