Mathematics-AI for Protein-Protein Interactions

Jiahui Chen

Assistant Professor,
Department of Mathematical Sciences,
University of Arkansas

October 5, 2023

@ Data Science and Statistics Seminar, University of Tennessee

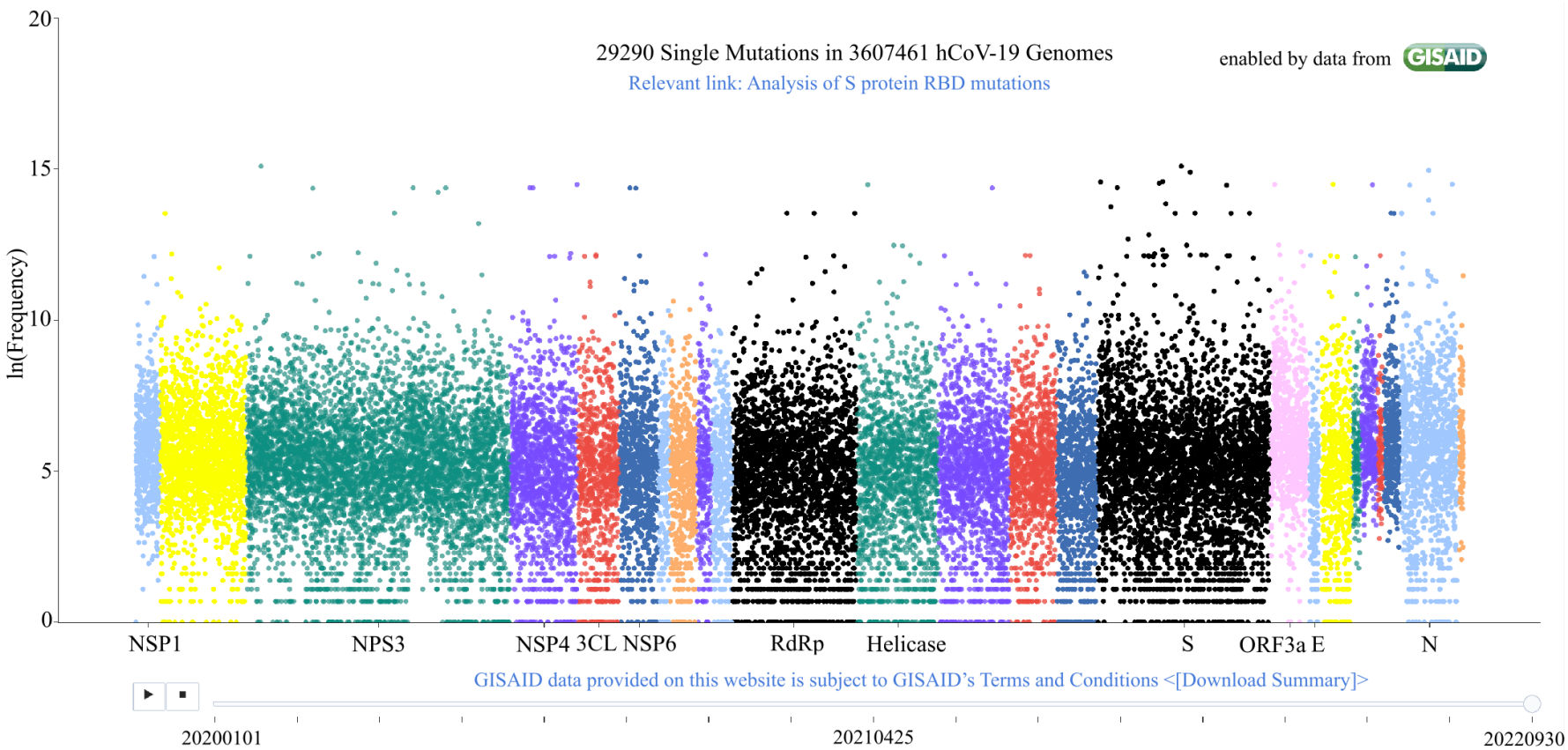
Outline

Outline

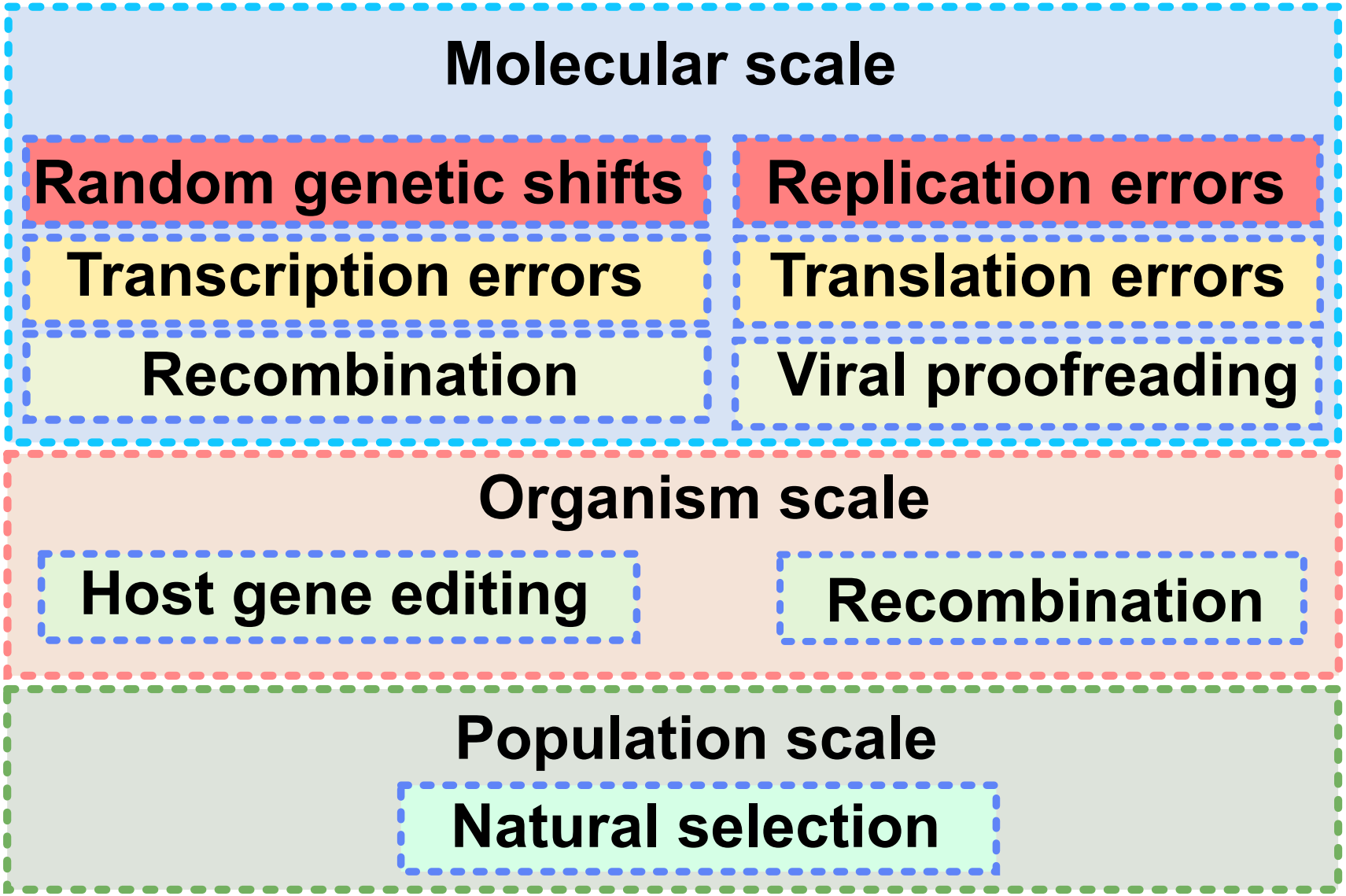
COVID-19 shows the importance of biosciences

The Distribution of SARS-COV-2 Mutations

UNIVERSITY OF ARKANSAS

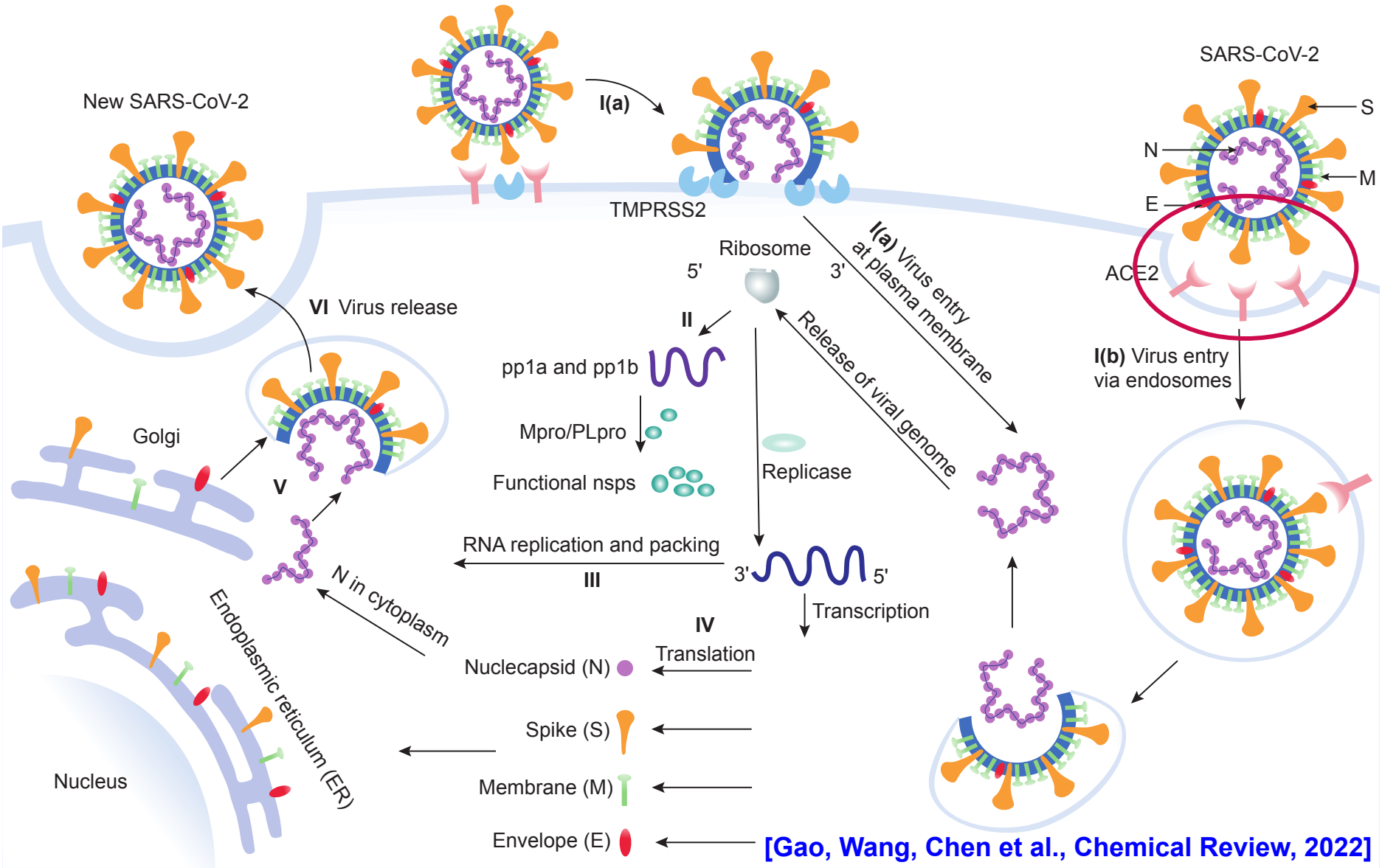


[Chen et al., J Chem Inf Model, 2022]


Molecular scale
Random genetic shifts
Replication errors
Transcription errors
Translation errors
Recombination
Viral proofreading
Organism scale
Host gene editing
Recombination
Population scale
Natural selection

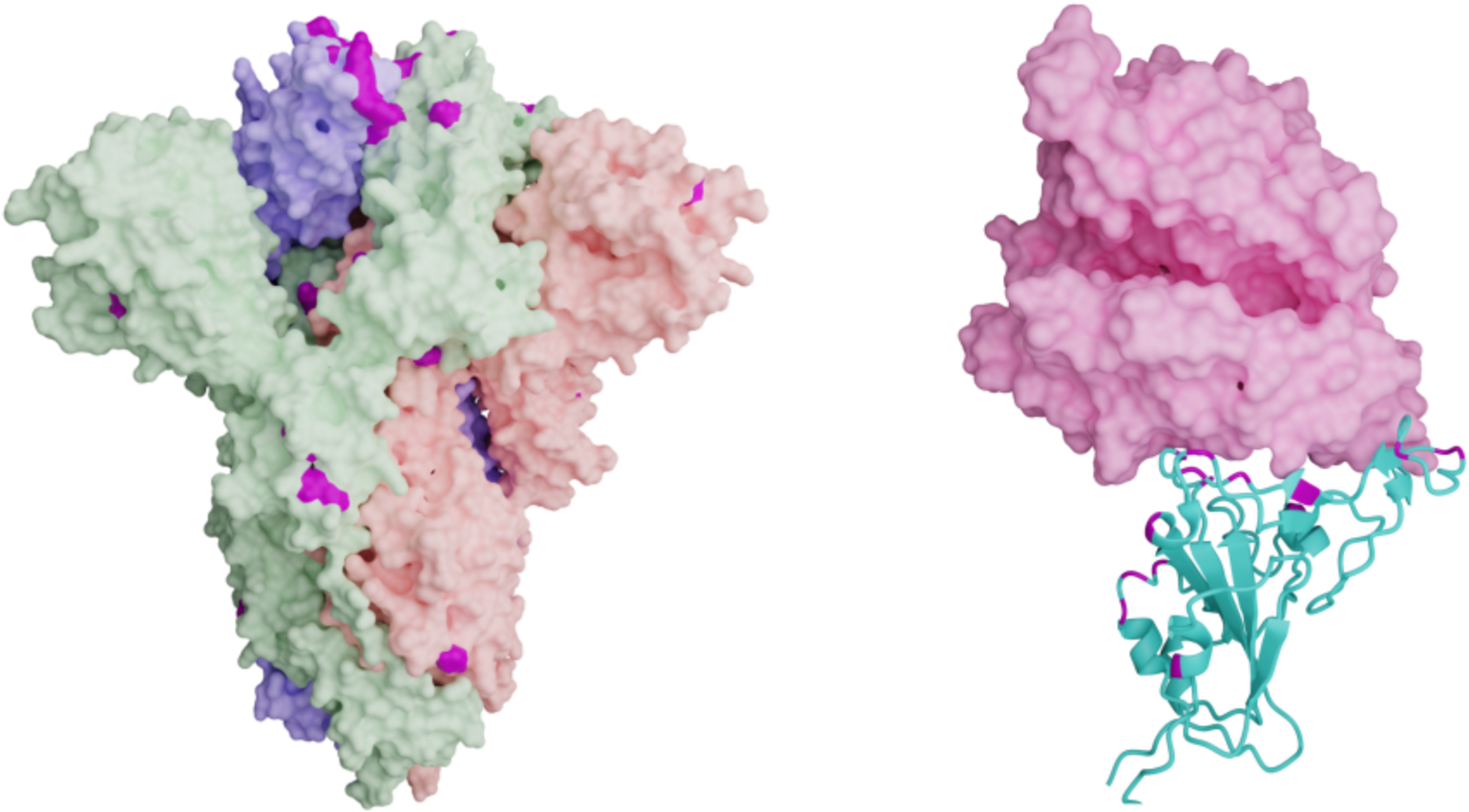
SARS-CoV-2 Life Cycle

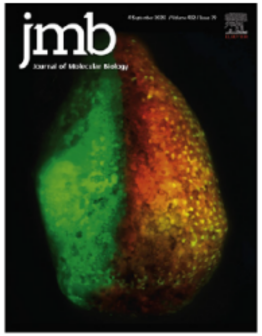
UNIVERSITY OF ARKANSAS



Spike Protein and PPI

UNIVERSITY OF ARKANSAS





Mutations Strengthened SARS-CoV-2 Infectivity

We predicted prevailing SARS-CoV-2 variants to occur at residues 452 and 501

Jiahui Chen[1], Rui Wang[1], Menglun Wang[1] and Guo-Wei Wei[1,2,3]

*1 - **Department of Mathematics,** Michigan State University, MI 48824, USA*
*2 - **Department of Electrical and Computer Engineering,** Michigan State University, MI 48824, USA*
*3 - **Department of Biochemistry and Molecular Biology,** Michigan State University, MI 48824, USA*

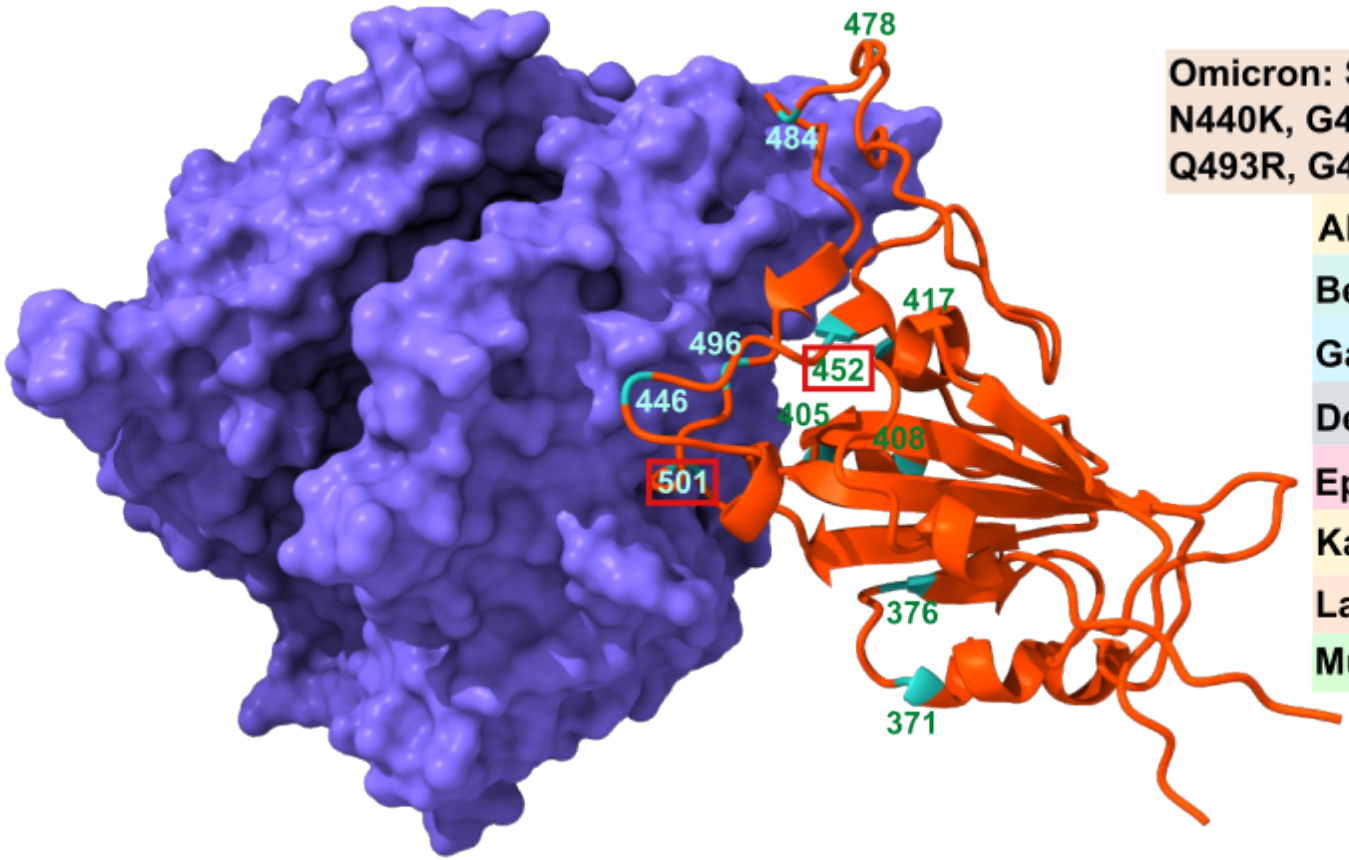
Correspondence to Guo-Wei Wei: *wei@math.msu.edu*
https://doi.org/10.1016/j.jmb.2020.07.009
Edited by Anna Panchenko

Received 4 June 2020;
Received in revised form 9 July 2020;
Accepted 17 July 2020
Available online 23 July 2020

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infectivity is a major concern in coronavirus disease 2019 (COVID-19) prevention and economic reopening. However, rigorous determination of SARS-CoV-2 infectivity is very difficult owing to its continuous evolution with over 10,000 single nucleotide polymorphisms (SNP) variants in many subtypes. We employ an algebraic topology-based machine learning model to quantitatively evaluate the binding free energy changes of SARS-CoV-2 spike glycoprotein (S protein) and host angiotensin-converting enzyme 2 receptor following mutations. We reveal that the SARS-CoV-2 virus becomes more infectious. Three out of six SARS-CoV-2 subtypes have become slightly more infectious, while the other three subtypes have significantly strengthened their infectivity. We also find that SARS-CoV-2 is slightly more infectious than SARS-CoV according to computed S protein-angiotensin-converting enzyme 2 binding free energy changes. Based on a systematic evaluation of all possible 3686 future mutations on the S protein receptor-binding domain, we show that most likely future mutations will make SARS-CoV-2 more infectious. Combining sequence alignment, probability analysis, and binding free energy calculation, we predict that a few residues on the receptor-binding motif, i.e., 452, 489, 500, 501, and 505, have high chances to mutate into significantly more infectious COVID-19 strains.

Predict the key RBD mutation sites in prevailing variants.
Mutations at 452 and 501 in prevailing SARS-CoV-2 variants.



Omicron: S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, **N501Y**, Y505H

Alpha: **N501Y**

Beta: K417N, E484K, **N501Y**

Gamma: K417T, E484K, **N501Y**

Delta: **L452R**, T478K

Epsilon: **L452R**

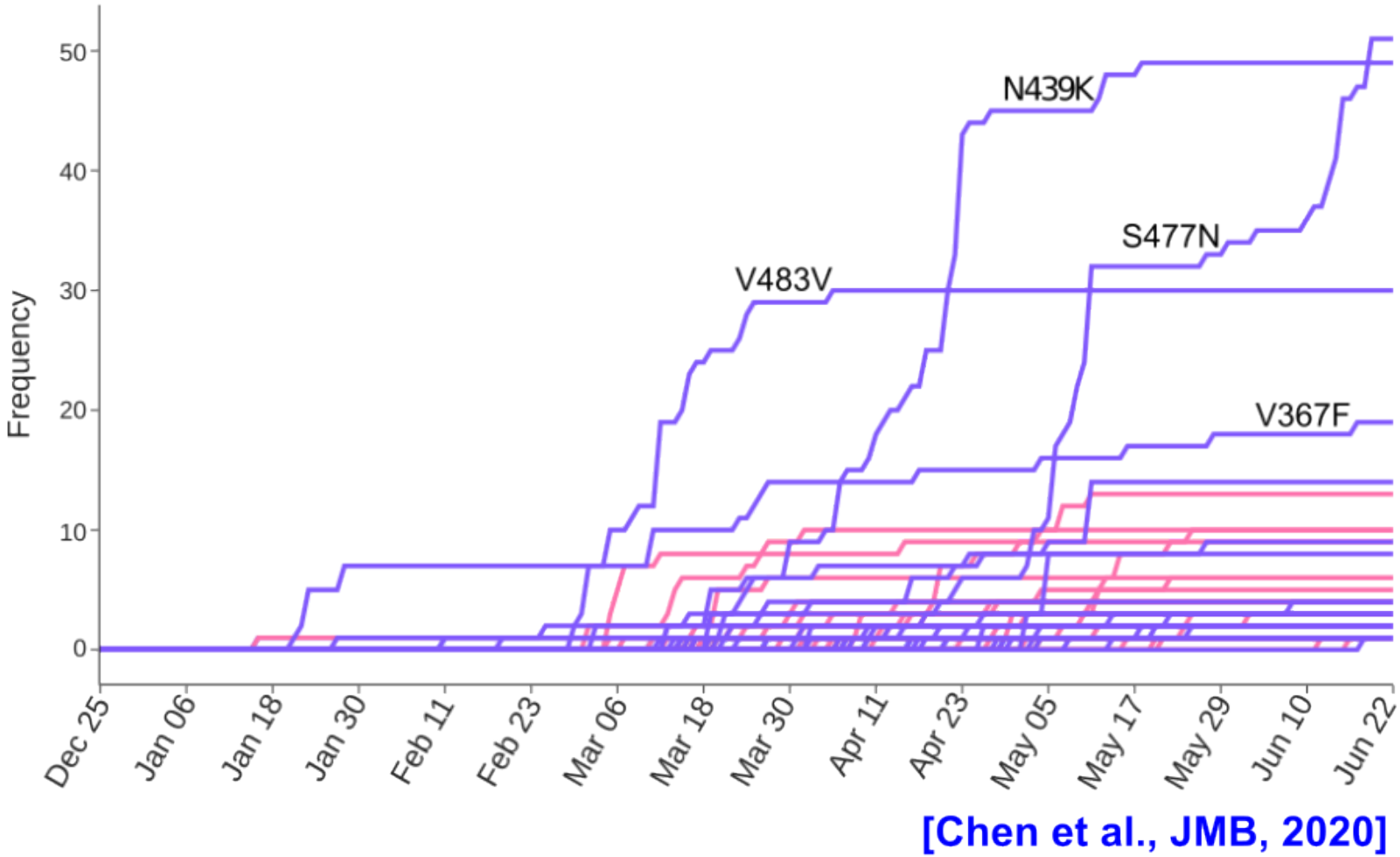
Kappa: **L452R**, E484Q

Lambda: **L452Q**, F490S

Mu: R346K, E484K, **N501Y**

[Chen, et al., JMB, 2020]

Infectivity strengthening mutations (blue) increase fester than infectivity weakening ones (red) over time.



Natural selection favors those mutations that enhance the viral transmission and evolution

Omicron BA.2 (B.1.1.529.2): high potential to becoming the next dominating variant

Jiahui Chen[1] and Guo-Wei Wei[1,3,4*]

[1] Department of Mathematics,
Michigan State University, MI 48824, USA.

[2] Department of Electrical and Computer Engineering,
Michigan State University, MI 48824, USA.

[3] Department of Biochemistry and Molecular Biology,
Michigan State University, MI 48824, USA.

February 11, 2022

WHO comfirmed it on March 26, 2022

Abstract

The Omicron variant of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has rapidly replaced the Delta variant as a dominating SARS-CoV-2 variant because of natural selection, which favors the variant with higher infectivity and stronger vaccine breakthrough ability. Omicron has three lineages or subvariants, BA.1 (B.1.1.529.1), BA.2 (B.1.1.529.2), and BA.3 (B.1.1.529.3). Among them, BA.1 is the currently prevailing subvariant. BA.2 shares 32 mutations with BA.1 but has 28 distinct ones. BA.3 shares most of its mutations with BA.1 and BA.2 except for one. BA.2 is found to be able to alarmingly reinfect patients originally infected by Omicron BA.1. An important question is whether BA.2 or BA.3 will become a new dominating "variant of concern". Currently, no experimental data has been reported about BA.2 and BA.3. We construct a novel algebraic topology-based deep learning model trained with tens of thousands of mutational and deep mutational data to systematically evaluate BA.2's and BA.3's infectivity, vaccine breakthrough capability, and antibody resistance. Our comparative analysis of all main variants namely, Alpha, Beta, Gamma, Delta, Lambda, Mu, BA.1, BA.2, and BA.3, unveils that BA.2 is about 1.5 and 4.2 times as contagious as BA.1 and Delta, respectively. It is also 30% and 17-fold more capable than BA.1 and Delta, respectively, to escape current vaccines. Therefore, we project that Omicron BA.2 is on its path to becoming the next dominating variant. We forecast that like Omicron BA.1, BA.2 will also seriously compromise most existing mAbs, except for sotrovimab developed by GlaxoSmithKline.

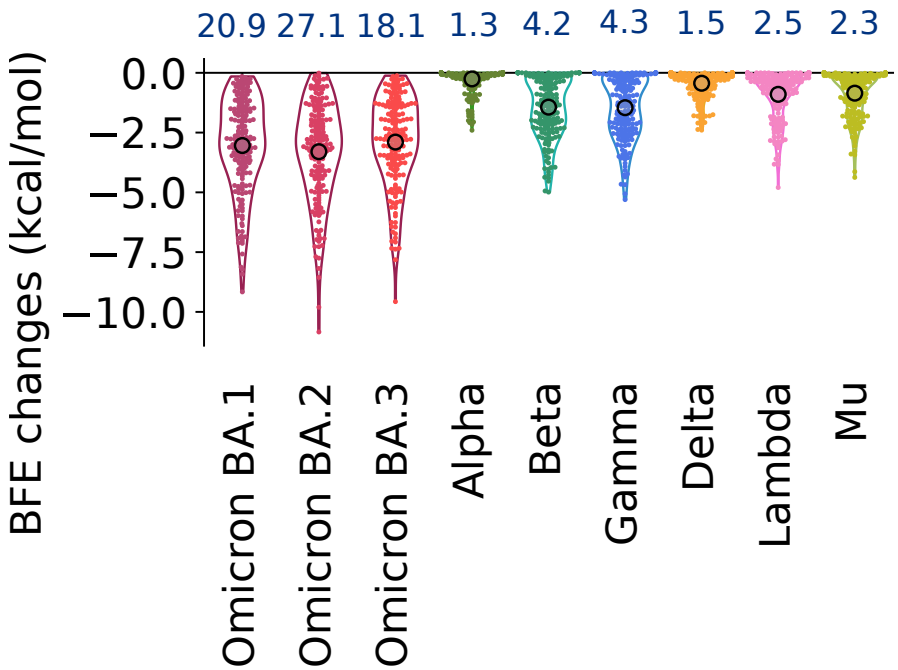
All predictions has been confirmed within 50 days!

Keywords: COVID-19, SARS-CoV-2, Omicron, infectivity, antibody-resistance, vaccine breakthrough,

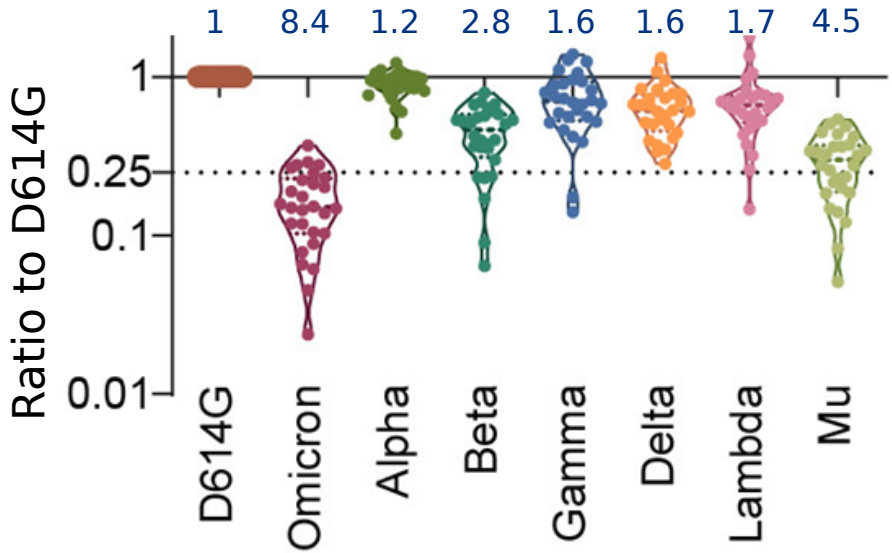
[Chen, Wei, J. Phys. Chem. Lett., 2022]

Omicron Variant Experimental Comparison

UNIVERSITY OF ARKANSAS



[Chen, Wei, J. Phys. Chem. Lett., 2022]



[Wang, Emerg. Microbes Infect., 2022]

Persistent Laplacian projected Omicron BA.4 and BA.5 to become new dominating variants

Jiahui Chen[1], Yuchi Qiu[1], Rui Wang[1], and Guo-Wei Wei[1,2,3*]
[1] Department of Mathematics,
Michigan State University, MI 48824, USA.
East Lansing, MI 48823 USA.
[2] Department of Electrical and Computer Engineering,
Michigan State University, MI 48824, USA.
[3] Department of Biochemistry and Molecular Biology,
Michigan State University, MI 48824, USA.

WHO confirmed it on late June, 2022

May 3, 2022

Abstract

Due to its high transmissibility, Omicron BA.1 ousted the Delta variant to become a dominating variant in late 2021 and was replaced by more transmissible Omicron BA.2 in March 2022. An important question is which new variants will dominate in the future. Topology-based deep learning models have had tremendous success in forecasting emerging variants in the past. However, topology is insensitive to homotopic shape variations in virus-human protein-protein binding, which are crucial to viral evolution and transmission. This challenge is tackled with persistent Laplacian, which is able to capture both the topology and shape of data. Persistent Laplacian-based deep learning models are developed to systematically evaluate variant infectivity. Our comparative analysis of Alpha, Beta, Gamma, Delta, Lambda, Mu, and Omicron BA.1, BA.1.1, BA.2, BA.2.11, BA.2.12.1, BA.3, BA.4, and BA.5 unveils that Omicron BA.2.11, BA.2.12.1, BA.3, BA.4, and BA.5 are more contagious than BA.2. In particular, BA.4 and BA.5 are about 36% more infectious than BA.2 and are projected to become new dominating variants by natural selection. Moreover, the proposed models outperform the state-of-the-art methods on three major benchmark datasets for mutation-induced protein-protein binding free energy changes.

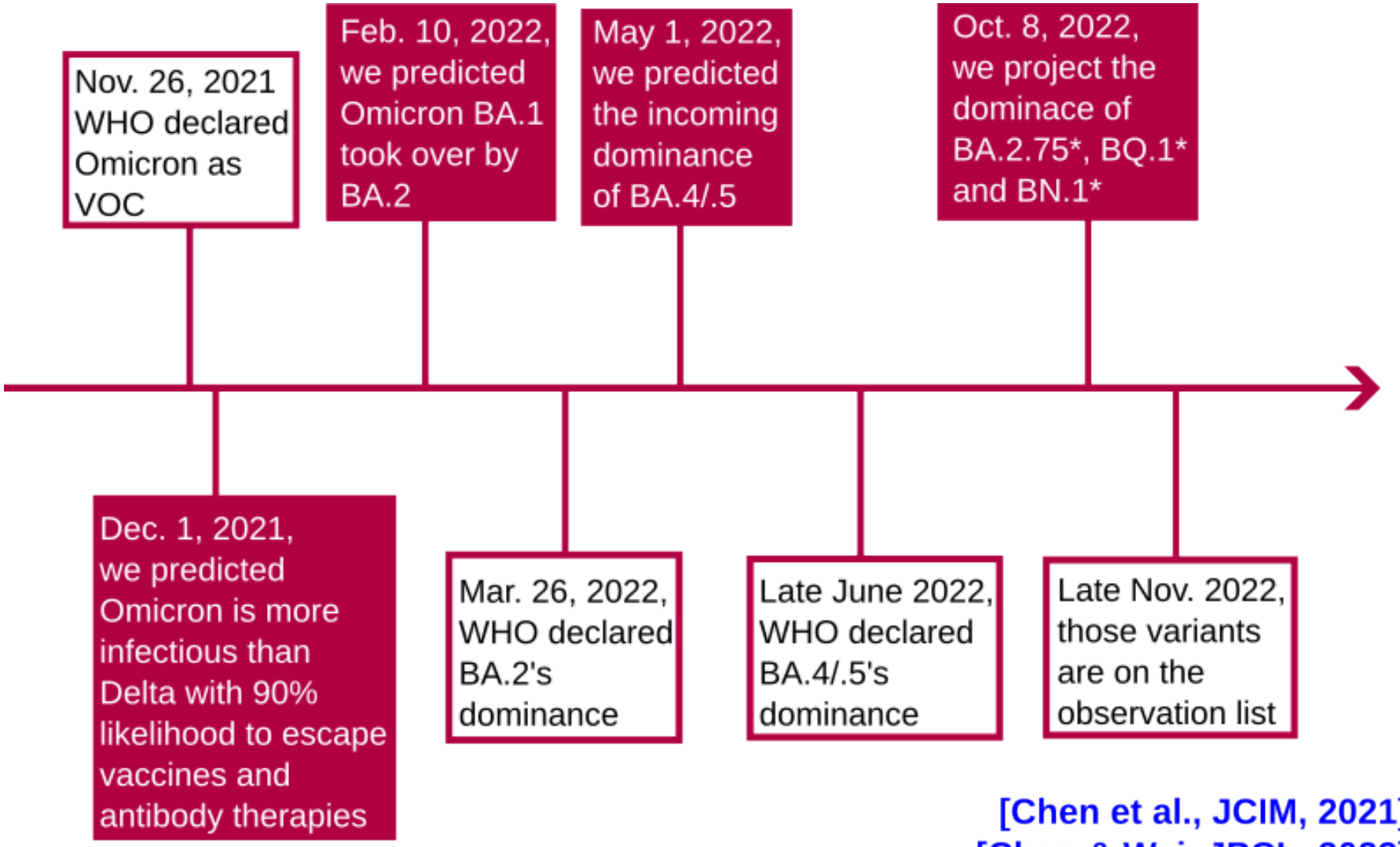
Keywords: SARS-CoV-2, evolution, infectivity, deep learning, persistent Laplacian.

*Corresponding author. Email: weig@msu.edu

[Chen, Qiu, Wang, Wei, Comput. Biol. Med., 2022]

History of Omicron Variant Predictions

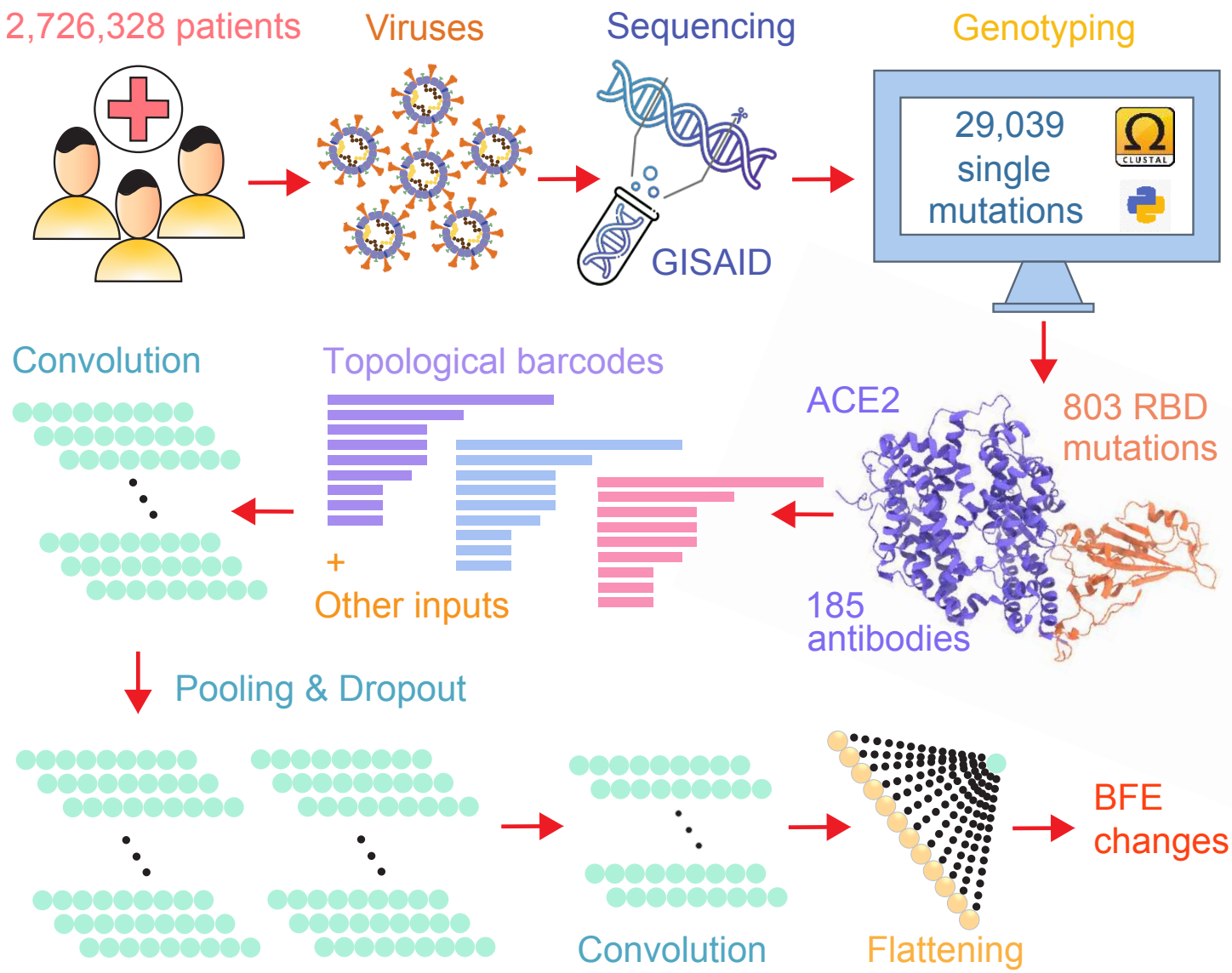
UNIVERSITY OF ARKANSAS



[Chen et al., JCIM, 2021]
[Chen & Wei, JPCL, 2022]
[Chen et al., CBM, 2021]
[Chen et al., JCIM, 2022]

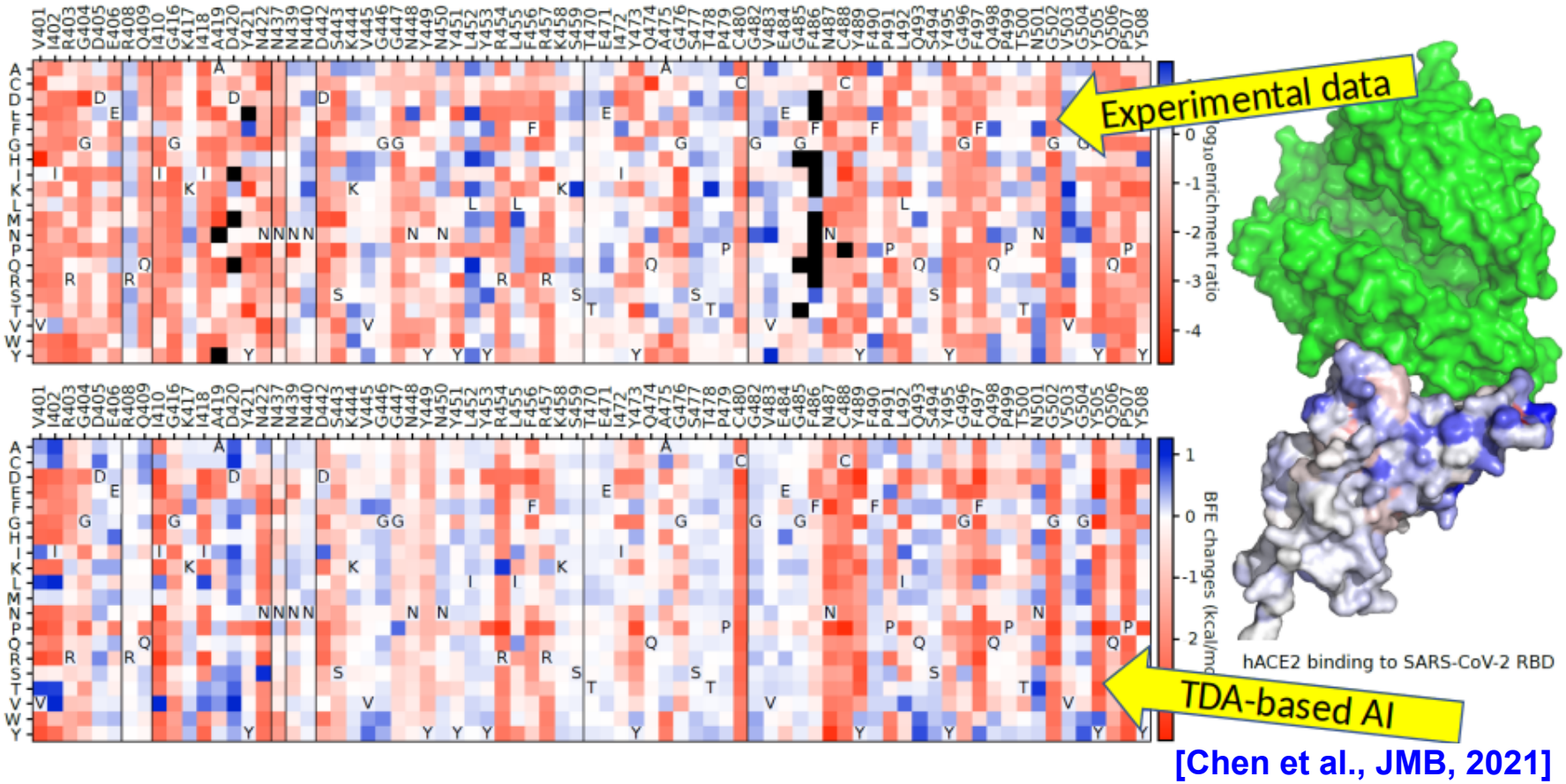
Math-AI Model for Binding Free Energy

UNIVERSITY OF ARKANSAS

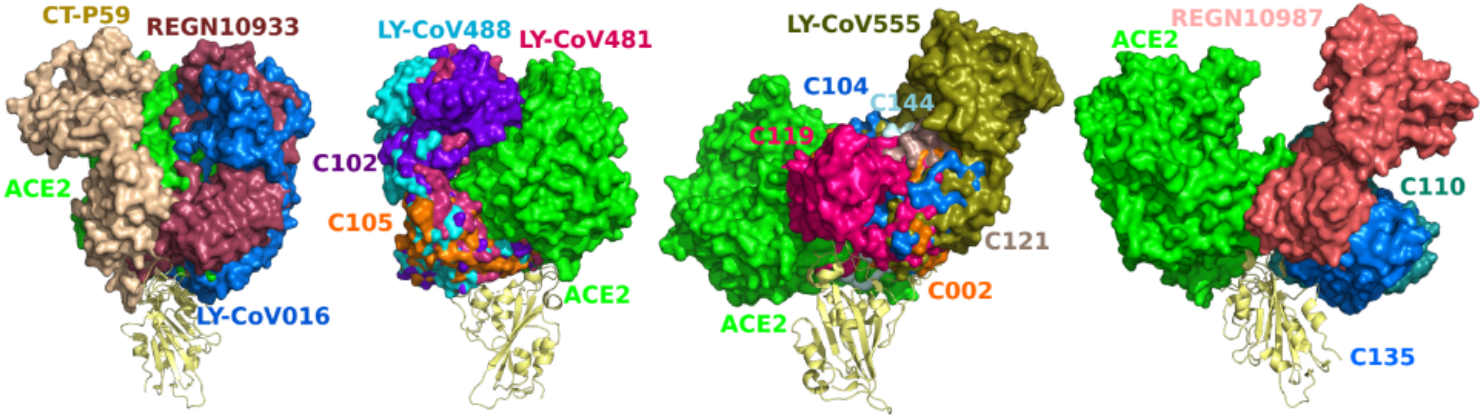


Validation of AI Predictions with Experiments

UNIVERSITY OF ARKANSAS

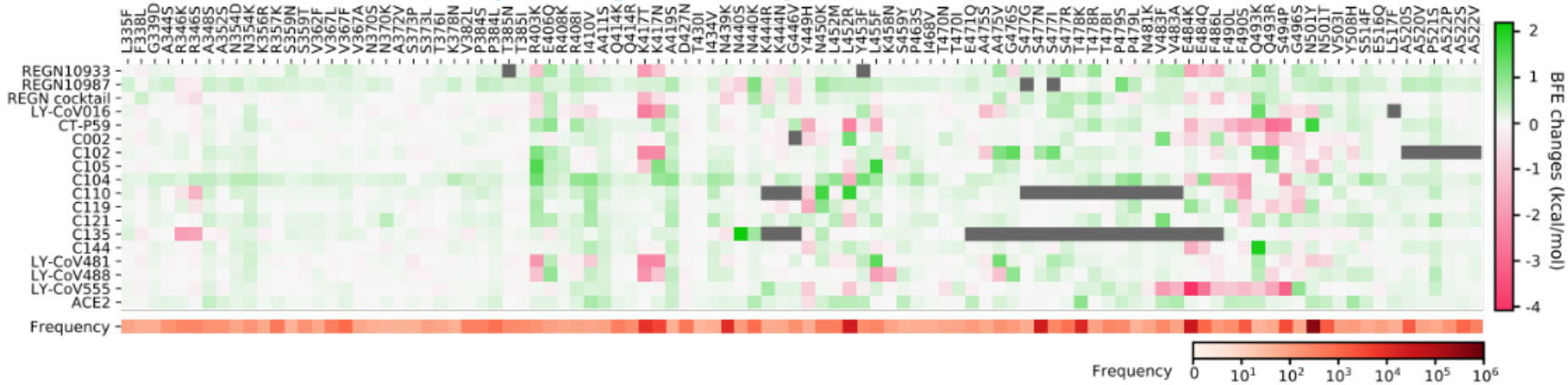


Antibody therapy



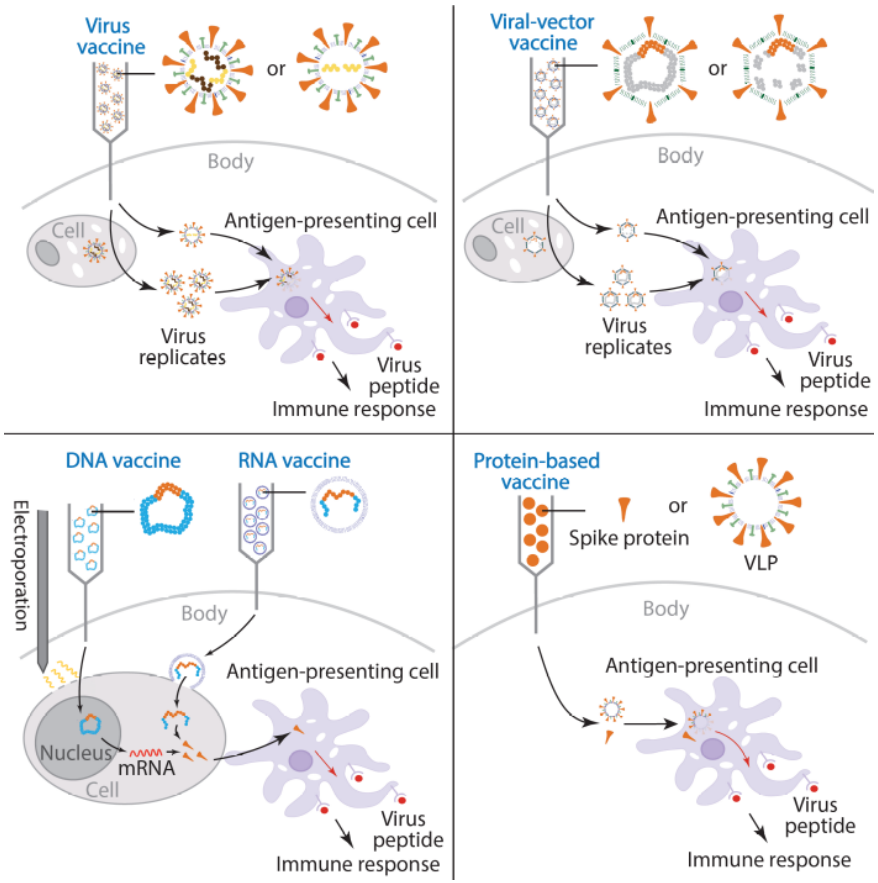
FDA emergency use authorization: REGN10933, REGN10987, LY-CoV016; In clinical trial: CT-P59, C144, C135

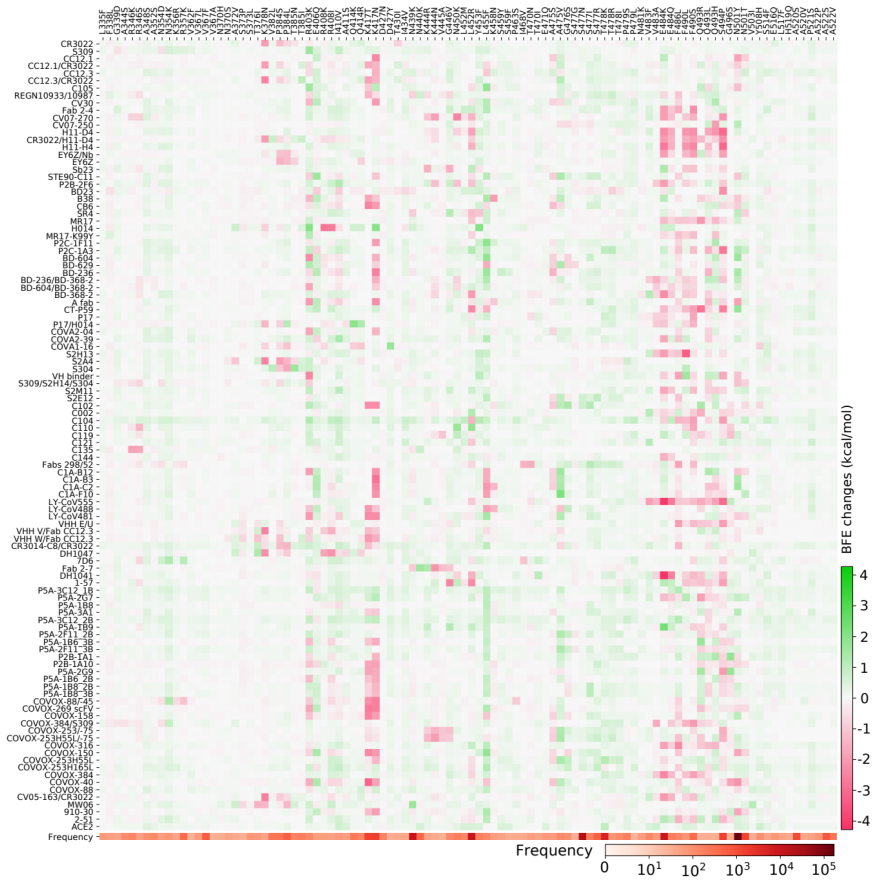
The top 100 most observed out of 712 RBD mutations from 506,768 patient genome sequences



[Chen, et al., JMB 2021]

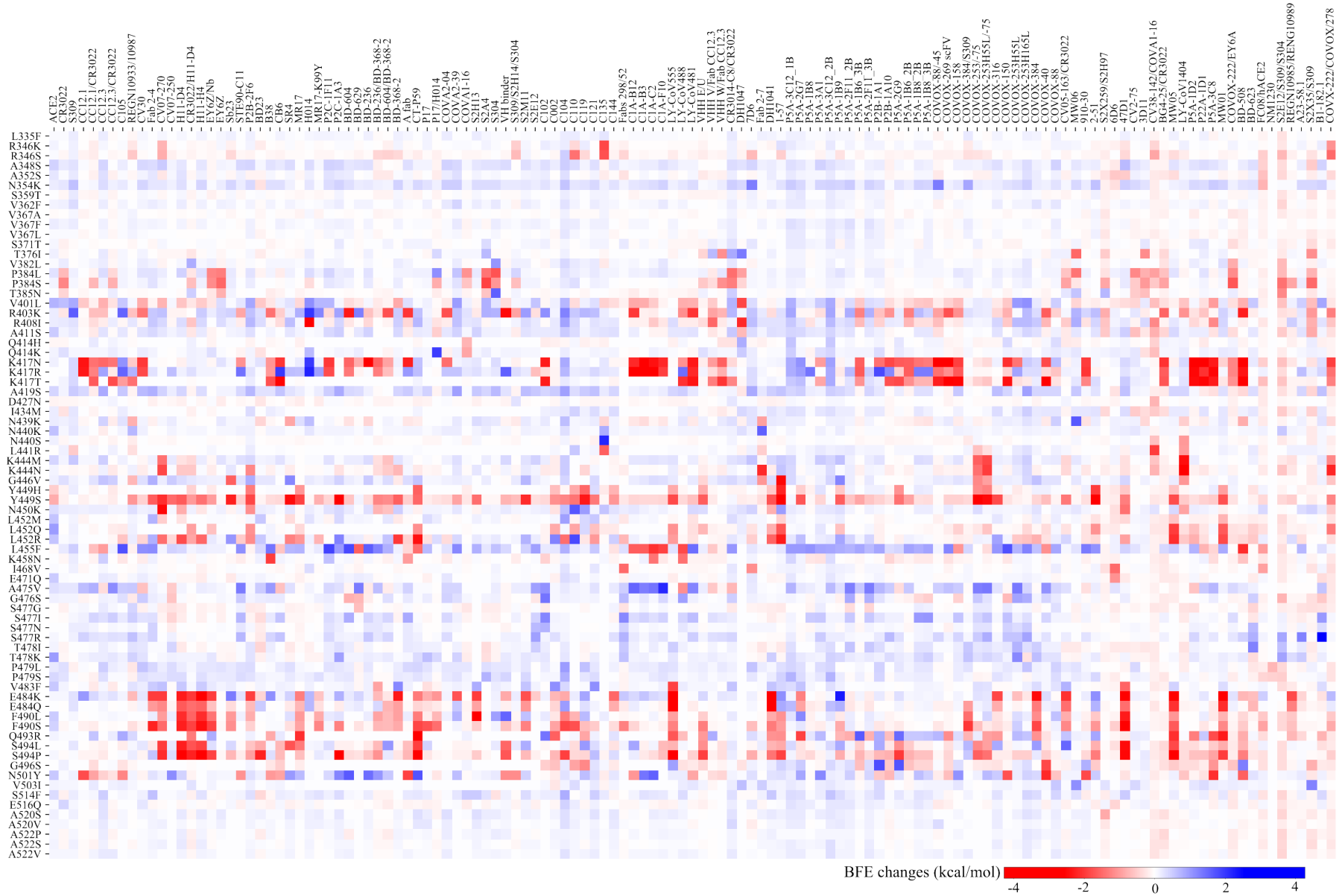
Vaccine efficacy



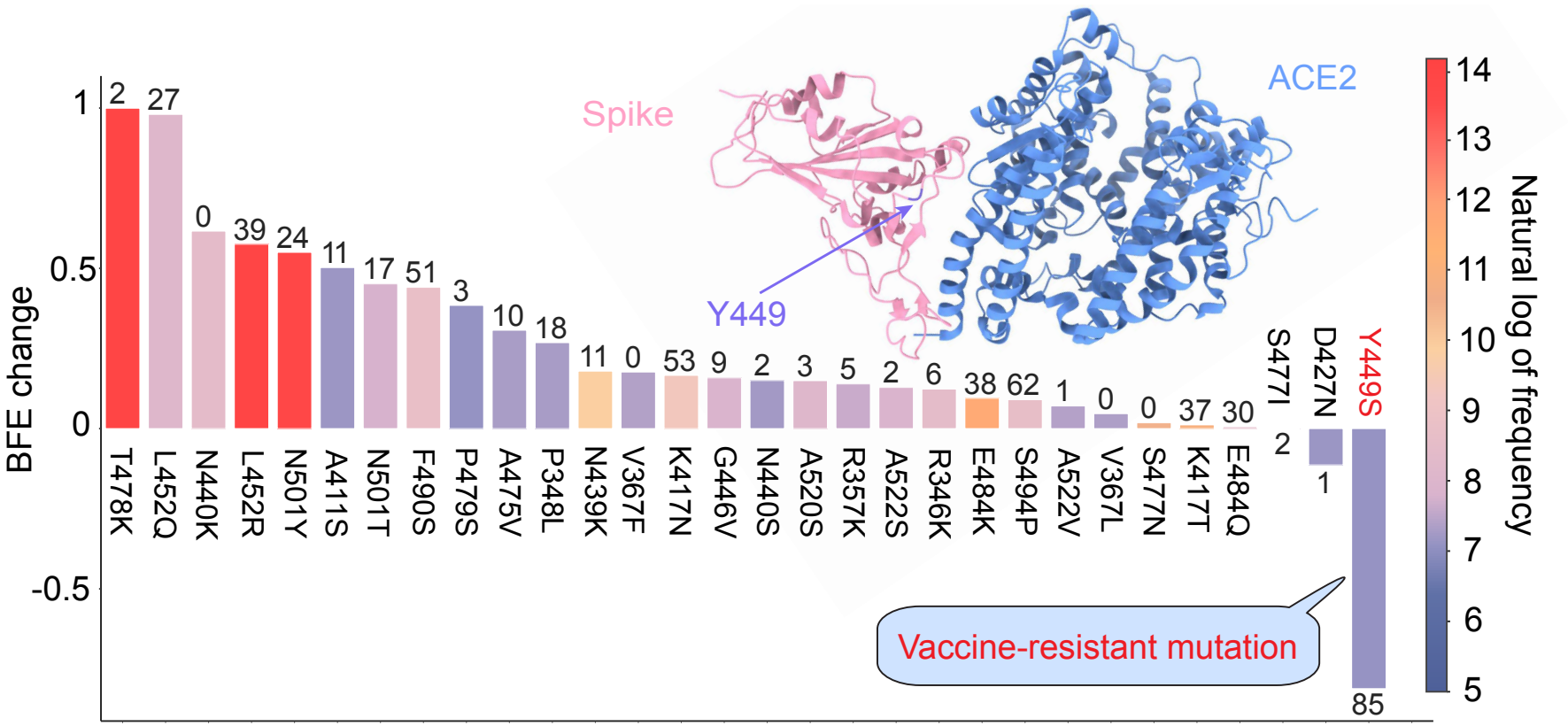


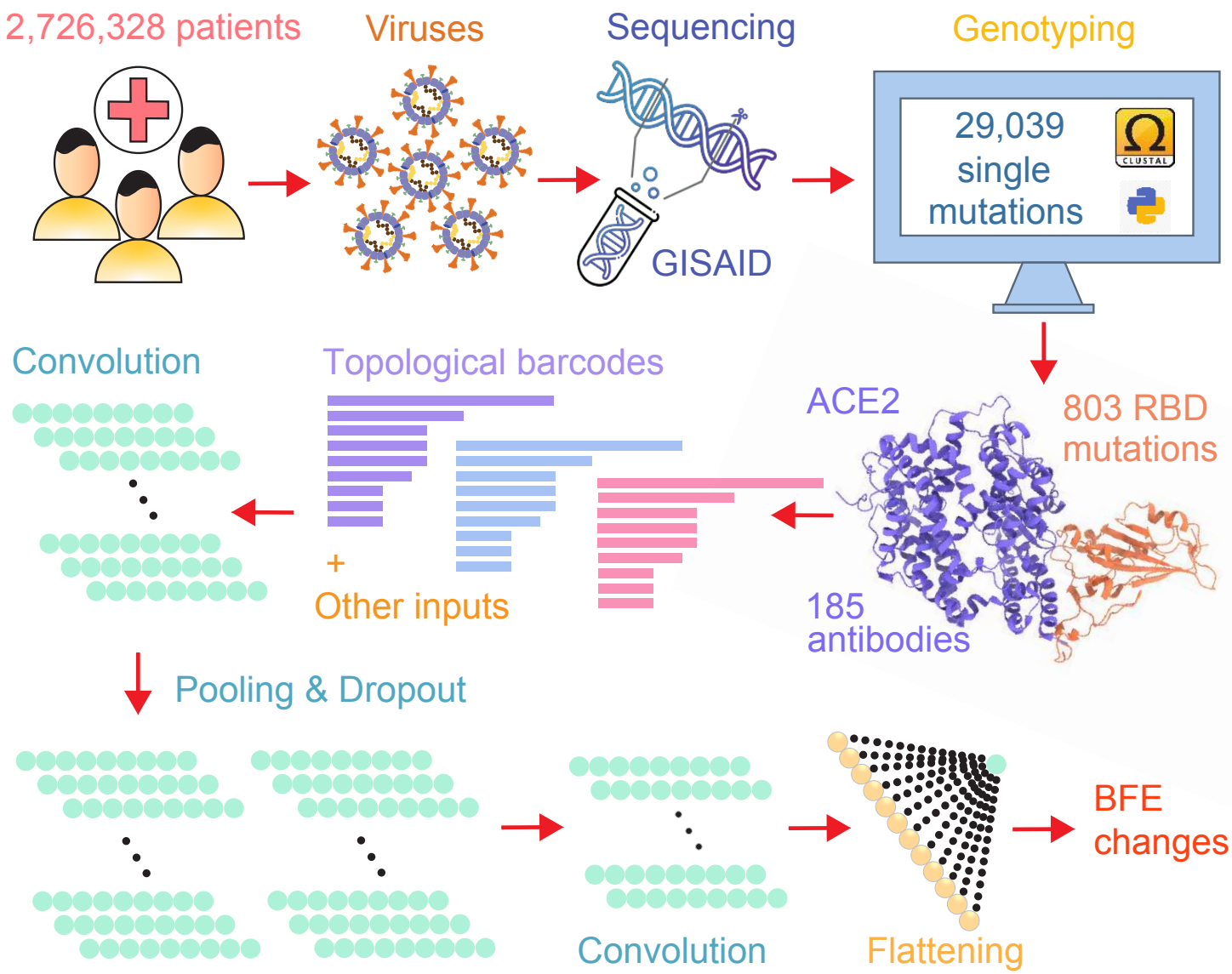
130 Antibodies

UNIVERSITY OF ARKANSAS



Vaccine-Resistant Mutations




2,726,328 patients
Viruses
Sequencing
Genotyping
GISAID
29,039
single
mutations
CLUSTAL
Convolution
Topological barcodes
+
Other inputs
ACE2
803 RBD
mutations
185
antibodies
Pooling & Dropout
Convolution
Flattening
BFE
changes

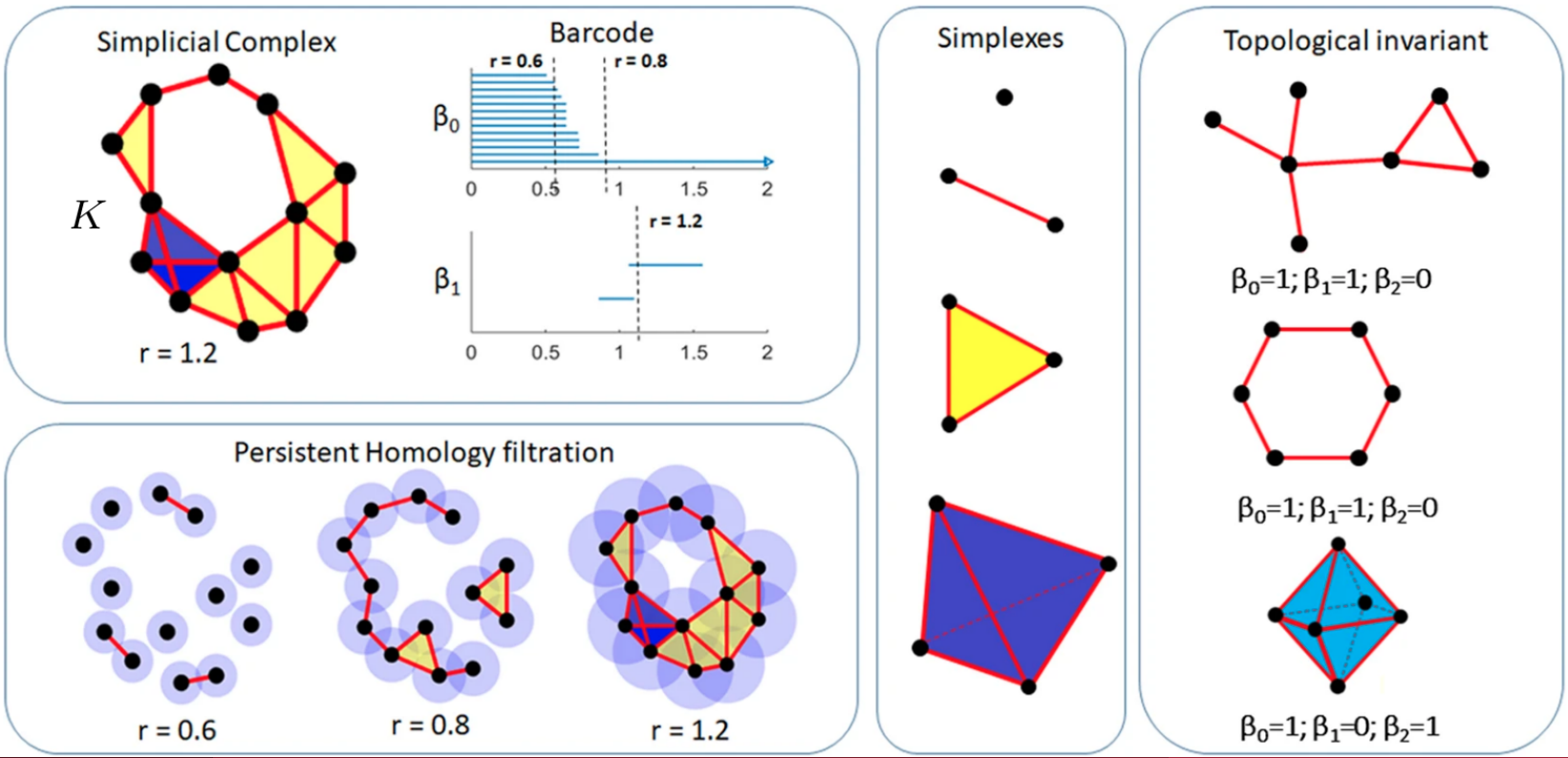
Outline

Persistent Homology

UNIVERSITY OF ARKANSAS

- Simplexes: 0-, 1-, 2-, 3-simplex
- Simplicial complex: K
- k-chain: $\sum_j c_j \sigma_j^k$
- Chain group: $C_k(K)$
- Boundary operator: $\partial_k \sigma_k = \sum_{i=0}^{k} (-1)^i [v_0, \cdots, \hat{v_i}, \cdots, v_q]$
- Homology group: $H_k = \frac{Z_k}{B_k}$, $Z_k = \ker \partial_k$, $B_k = \text{im } \partial_{k+1}$
- Betti number: $\beta_k = \text{Rank}(H_k)$

[Edelsbrunner et al., IEEE 2000, Zomorodian et al., DCG 2005, Carlsson et al., Bull. Am. Math. 2009, Ghrist, Bull. Am. Math. 2008, ...]

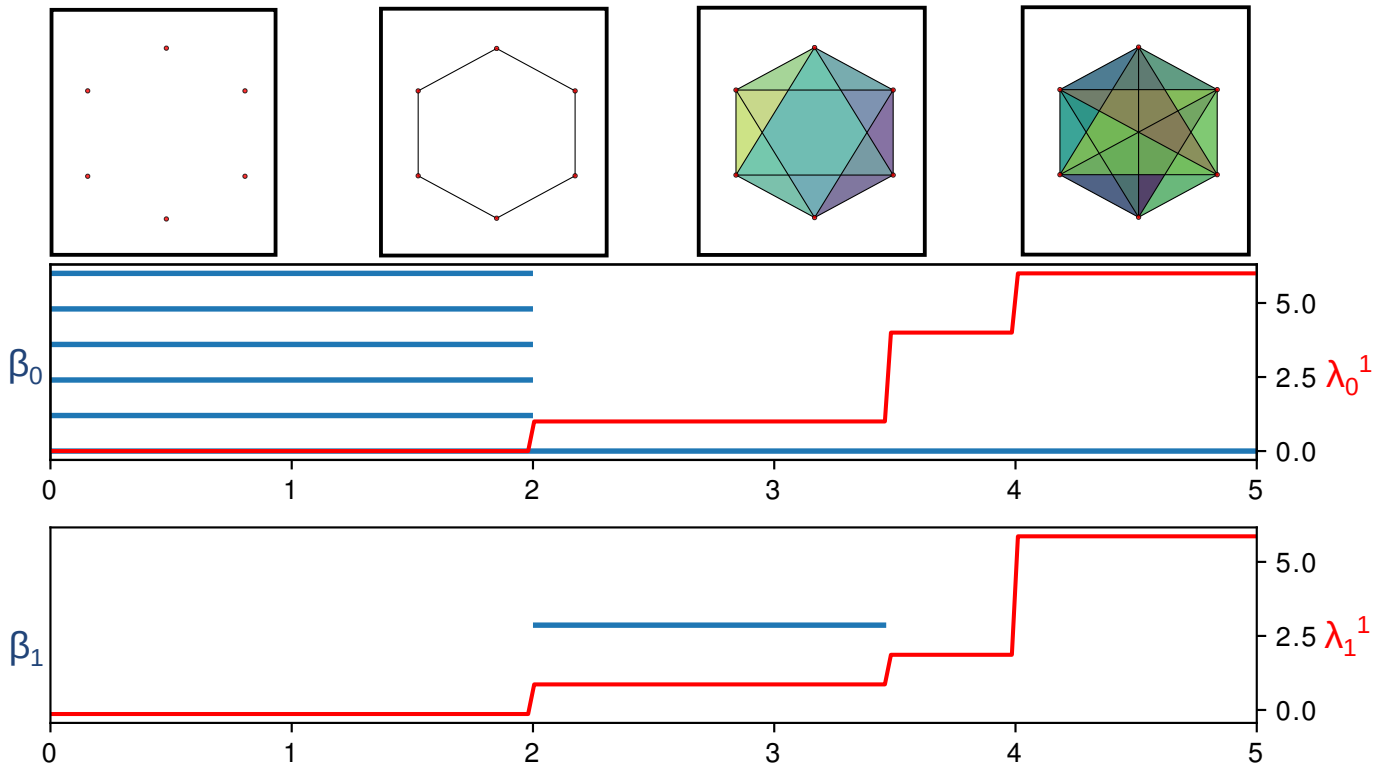


Persistent Laplacian

UNIVERSITY OF ARKANSAS

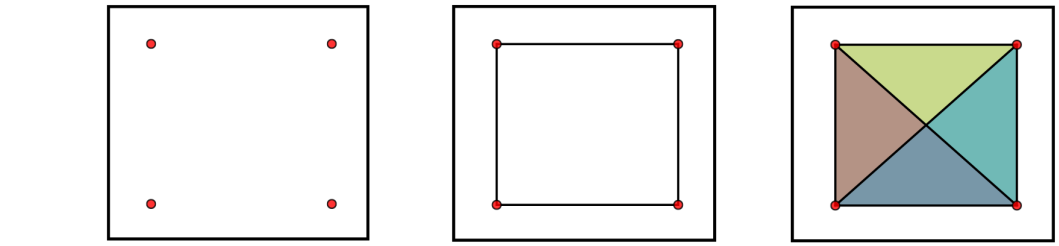
- Coboundary operator: $\partial_k^* : C_{k-1}(K) \to C_k(K)$
- Laplacian: $\Delta_k = \partial_{k+1}\partial_{k+1}^* + \partial_k^*\partial_k$
- Spectrum of Laplacian: $\mathrm{Spec}(\Delta_k) = \{\lambda_k^1, \lambda_k^2, \lambda_k^3, ...\}$

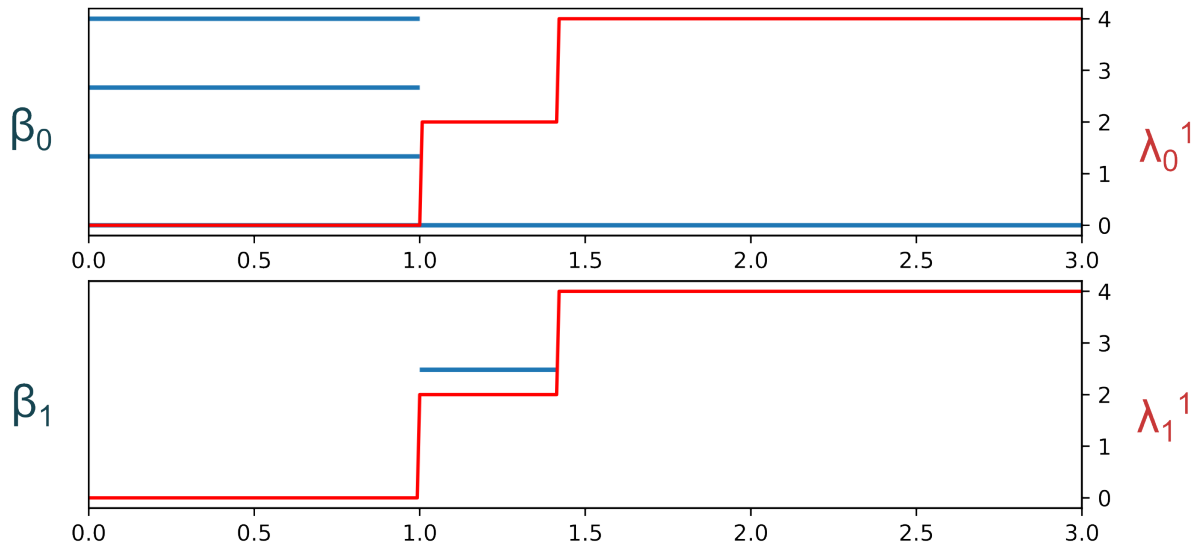
[Wang et al., IJNMBE 2020, **Wang et al., FDS 2021**, **Chen et al., CBM 2022**, ...]



Motivation for Evolutionary de Rham Hodge

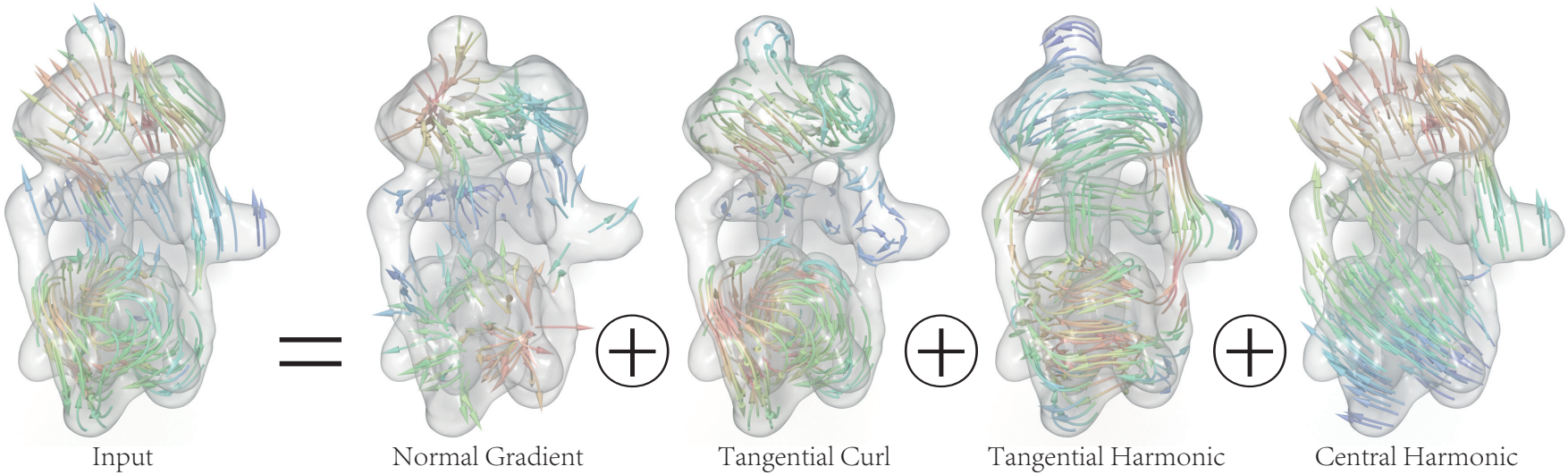
UNIVERSITY OF ARKANSAS



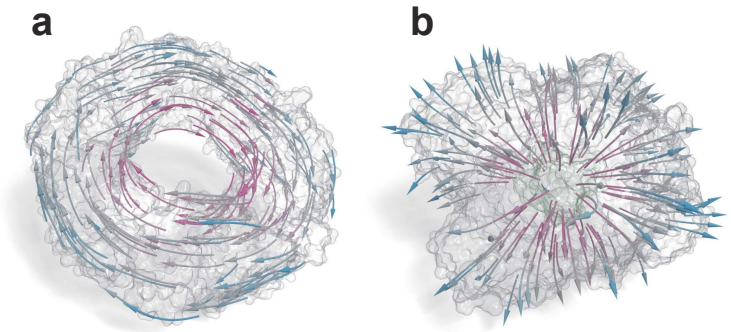


Hodge theory

3-dimensional volumes bounded by 2-manifolds in $\mathbb{R}^3$



- Every cohomology class has a differential form that vanishes under the Laplacian operator of the metric
- Hodge decomposition



Manifolds with boundary, (3-dimensional volumes bounded by 2-manifolds in $\mathbb{R}^3$)

- A differential k-form $\omega^k \in \Omega^k(M)$
- The *differential* operator (i.e., exterior derivative) $d^k : \Omega^k(M) \to \Omega^{k+1}(M)$
- The *Hodge k-star* $\star^k$ (aka Hodge dual) $\star^k : \Omega^k(M) \to \Omega^{3-k}(M)$
- The *codifferential* operators $\delta^k : \Omega^k(M) \to \Omega^{k-1}(M)$, $\delta^k = (-1)^k \star^{4-k} d^{3-k} \star^k$, for $k = 1, 2, 3$

- The *de Rham-Laplace operator*, or *Hodge Laplacian*

$$\Delta^k \equiv d^{k-1}\delta^k + \delta^{k+1}d^k$$

- *de Rham complex*

$$0 \longrightarrow \Omega^0(M) \xrightarrow{d^0} \Omega^1(M) \xrightarrow{d^1} \Omega^2(M) \xrightarrow{d^2} \Omega^3(M) \xrightarrow{d^3} 0$$

- *Bi-directional chain complex*

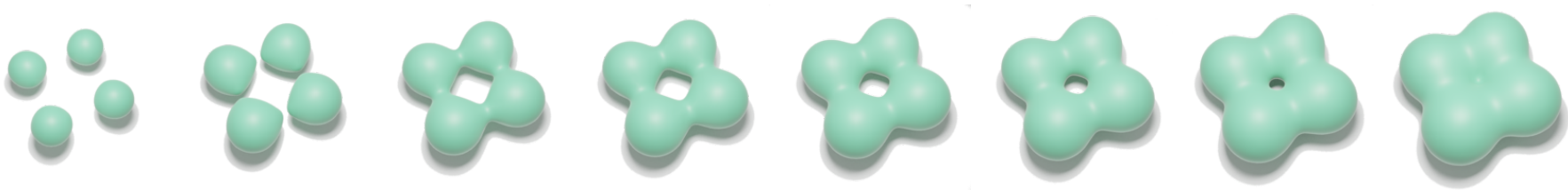
$$\Omega^0(M) \underset{\delta^1}{\overset{d^0}{\rightleftarrows}} \Omega^1(M) \underset{\delta^2}{\overset{d^1}{\rightleftarrows}} \Omega^2(M) \underset{\delta^3}{\overset{d^2}{\rightleftarrows}} \Omega^3(M)$$

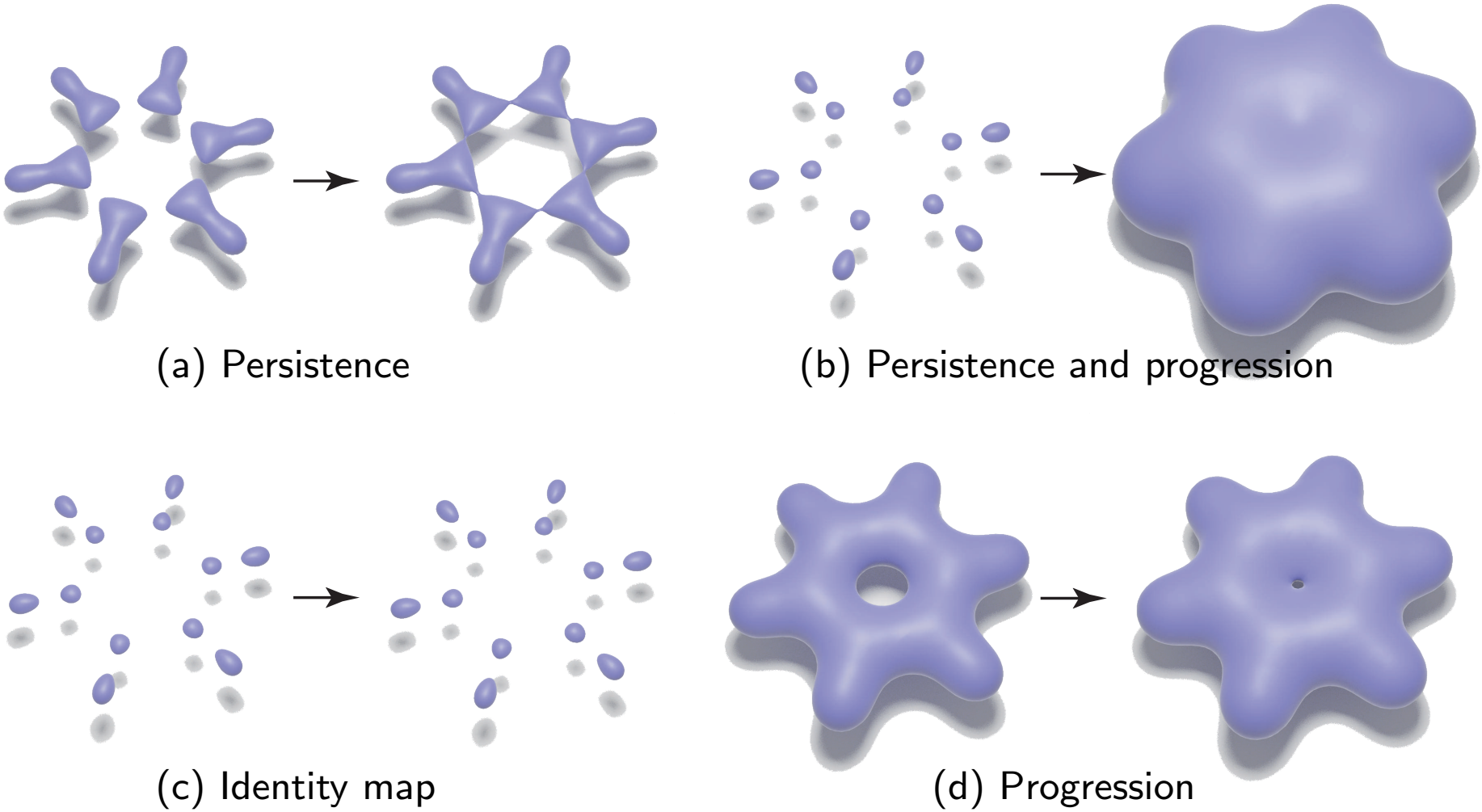
- *de Rham cohomology* $H^k_{dR} = \ker d^k / \mathrm{im}\, d^{k-1}$, and $H^k_{dR} \cong \mathcal{H}^k_\Delta$,

$$\beta_k = \dim \mathcal{H}^k_{\Delta_t} = \dim \mathcal{H}^{3-k}_{\Delta_n}$$

The inclusion map $\mathfrak{I}_{l,l+1} : M_l \hookrightarrow M_{l+1}$.

$$M_0 \overset{\mathfrak{I}_{0,1}}{\longrightarrow} M_1 \overset{\mathfrak{I}_{1,2}}{\longrightarrow} M_2 \overset{\mathfrak{I}_{2,3}}{\longrightarrow} \cdots \overset{\mathfrak{I}_{n-1,n}}{\longrightarrow} M_n \overset{\mathfrak{I}_{n,n+1}}{\longrightarrow} M = M_{c_{\max}}.$$





(a) Persistence

(b) Persistence and progression

(c) Identity map

(d) Progression

Figure 1: Persistence and progression on benzene.

Spectral Representations

UNIVERSITY OF ARKANSAS

- $\{\lambda_{l,i}^T\}$, $\{\lambda_{l,i}^C\}$ and $\{\lambda_{l,i}^N\}$ give the eigenvalues of the Laplacians.
- The multiplicities of the zero eigenvalues in $\lambda_{l,0}^T$, $\lambda_{l,0}^C$, and $\lambda_{l,0}^N$ are associated with Betti numbers β_0, β_1 and β_2, respectively.
- $\lambda_{l,1}^T$, $\lambda_{l,1}^C$, and $\lambda_{l,1}^N$ are the first non-zero eigenvalues

Discrete Exterior Calculus (DEC) by [Desbrun2008]; Finite Element Exterior Calculus (FEEC) by [Arnold2006].

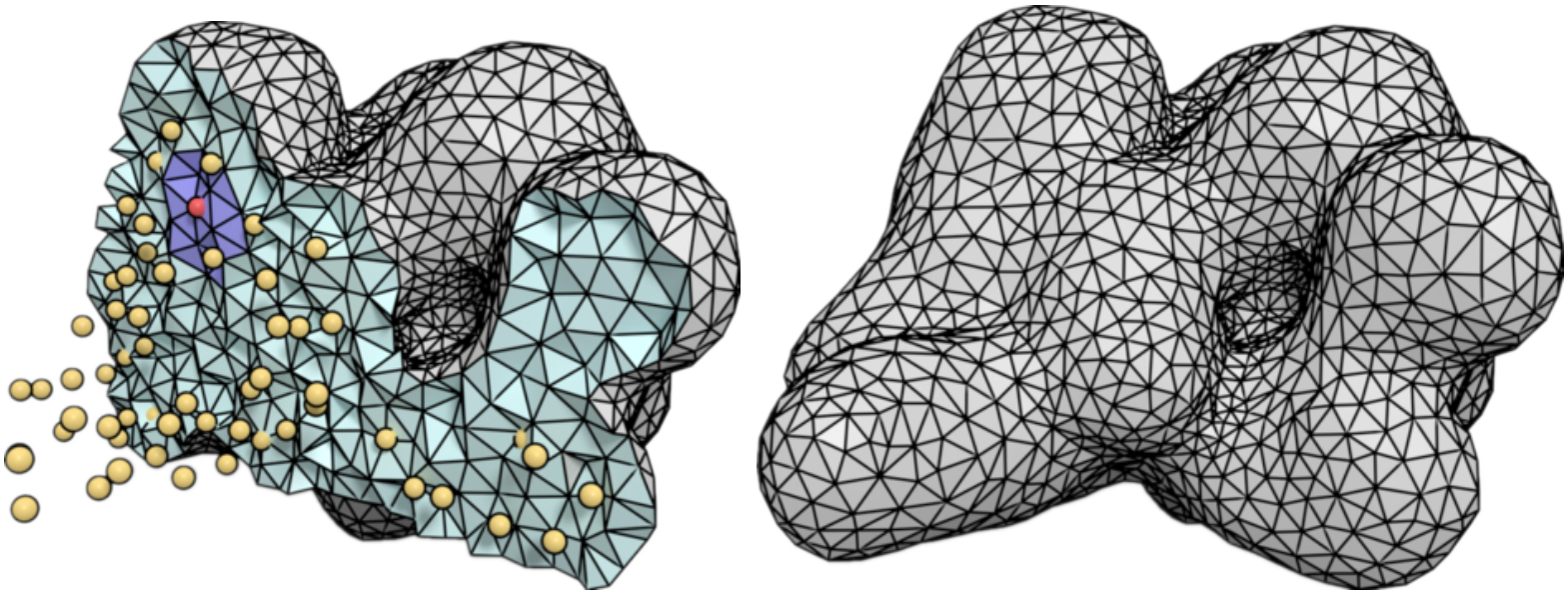


Figure 2: A 3-manifold embedded in 3D Euclidean space is tessellated into a 3D simplicial complex (Delaunay triangulation).

The boundary operator ∂ is defined as

$$\partial\sigma = \sum_{i=0}^{k}(-1)^i[v_0, v_1, ..., \hat{v}_i, ..., v_k],$$

where $\hat{v}_i$ means that the ith vertex is removed and an oriented k-simplex $\sigma = [v_0, v_1, ..., v_k]$.

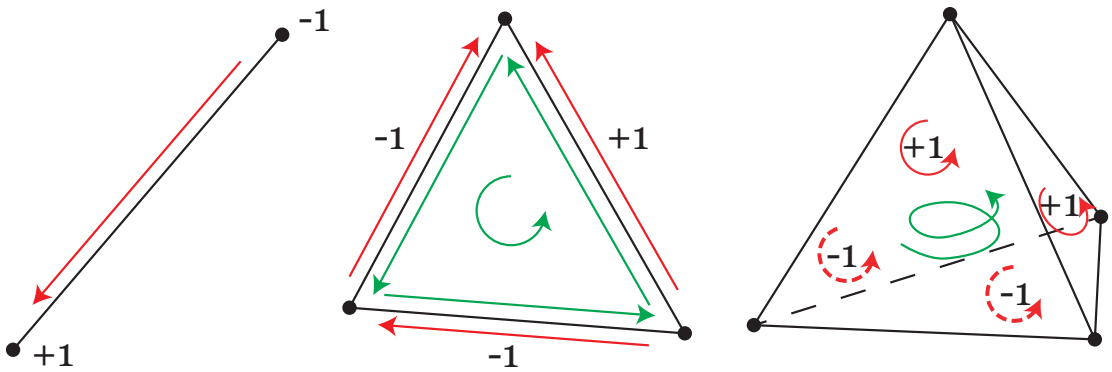


Figure 3: Pre-assigned orientation is colored in red. Induced orientation by ∂ is colored in green.

The discrete Hodge star matrices S_k is just converting primal forms and dual forms by the following equation

$$\frac{1}{|\sigma_k|}\int_{\sigma_k}\omega = \frac{1}{|*\sigma_k|}\int_{*\sigma_k}\star\omega.$$

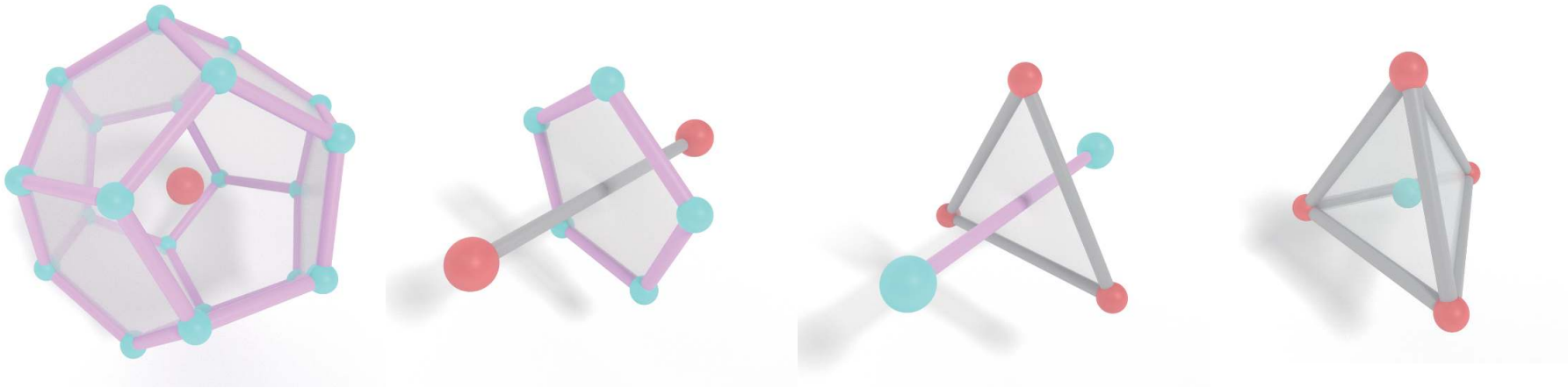


Figure 4: Illustration of the dual (Voronoi diagram) and primal elements (Delaunay triangulation) of the tetrahedral mesh.

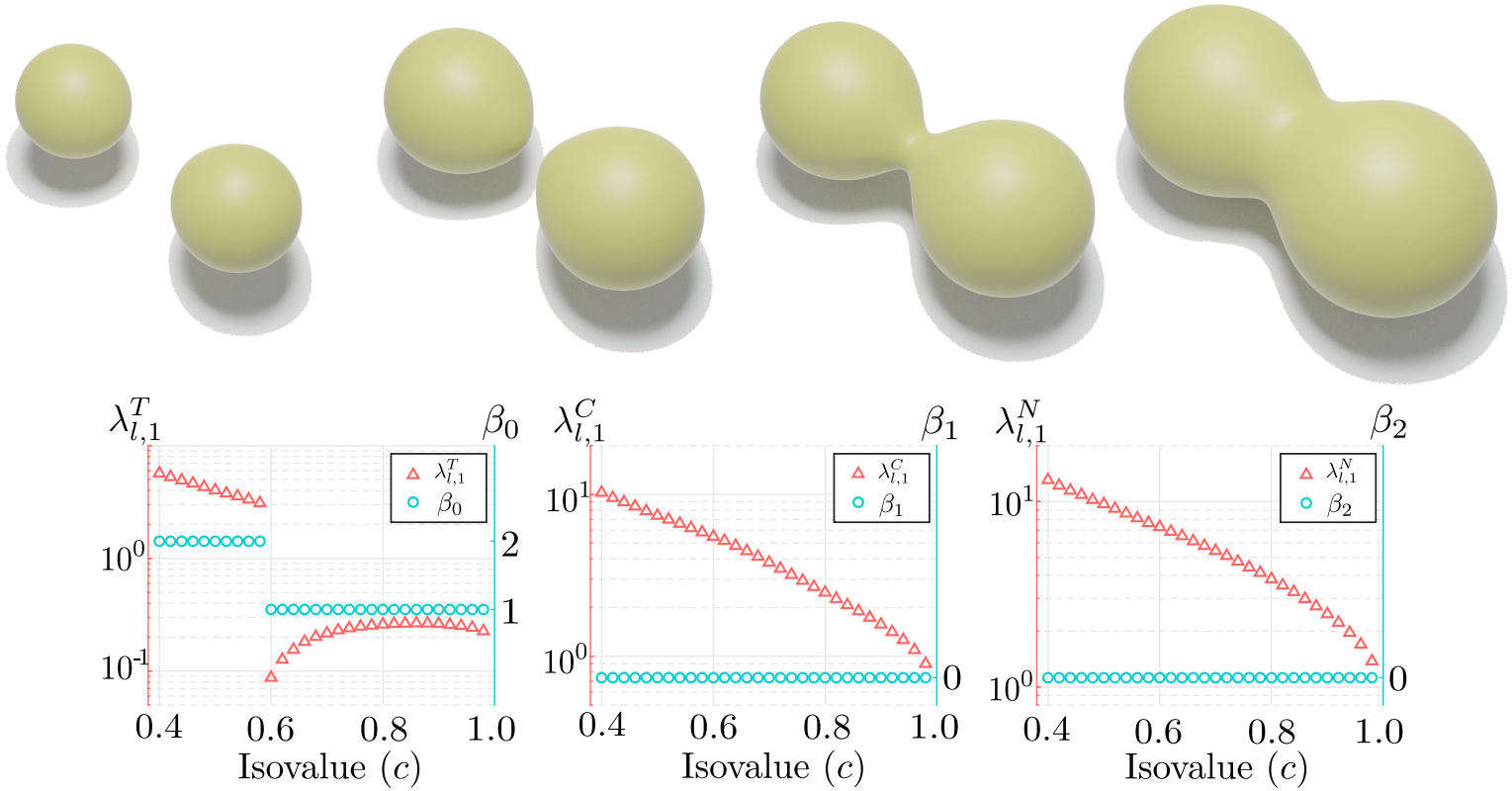
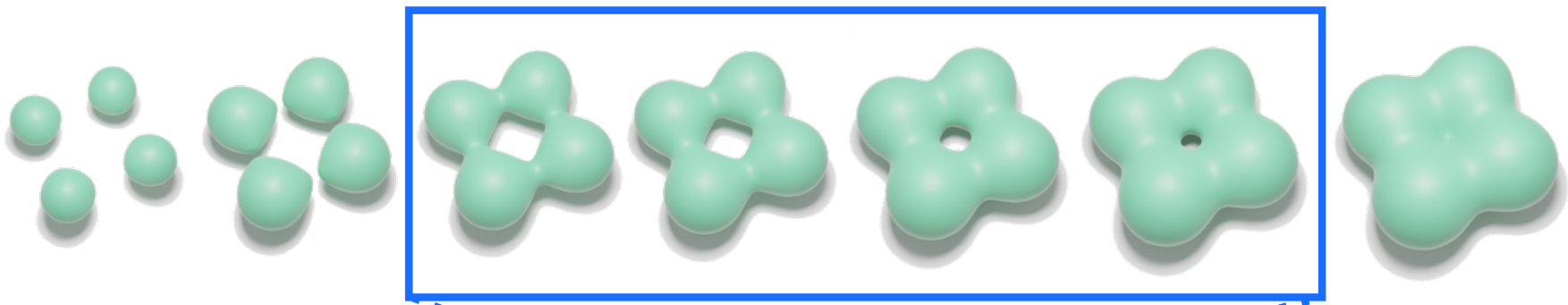
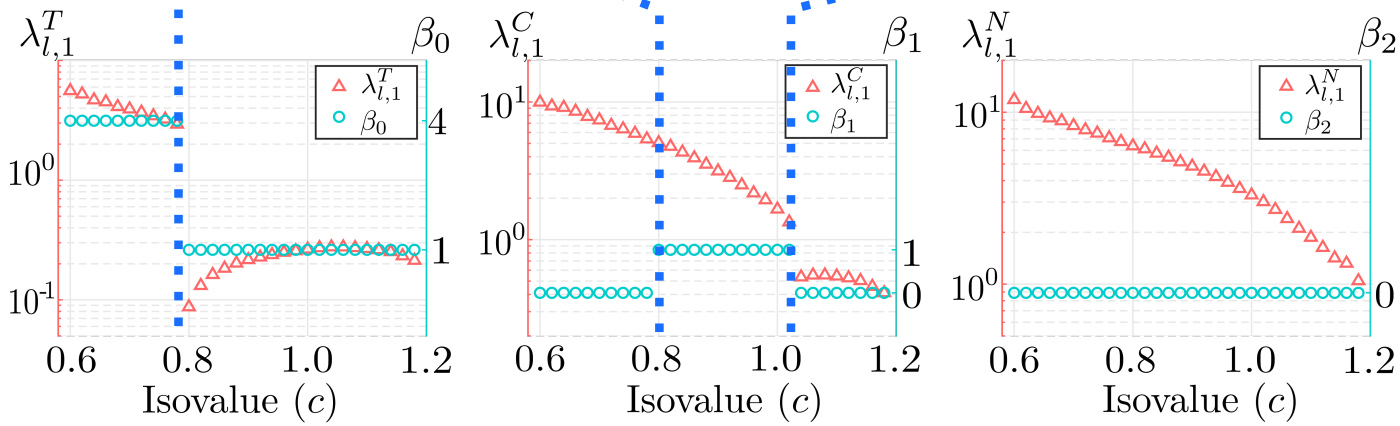


Figure 5: Eigenvalues and Betti numbers vs isovalue (c) of the two-body system with $\eta = 1.19$ and $\max(\rho) \approx 1.0$.

UNIVERSITY OF ARKANSAS

Homotopic shape evolution

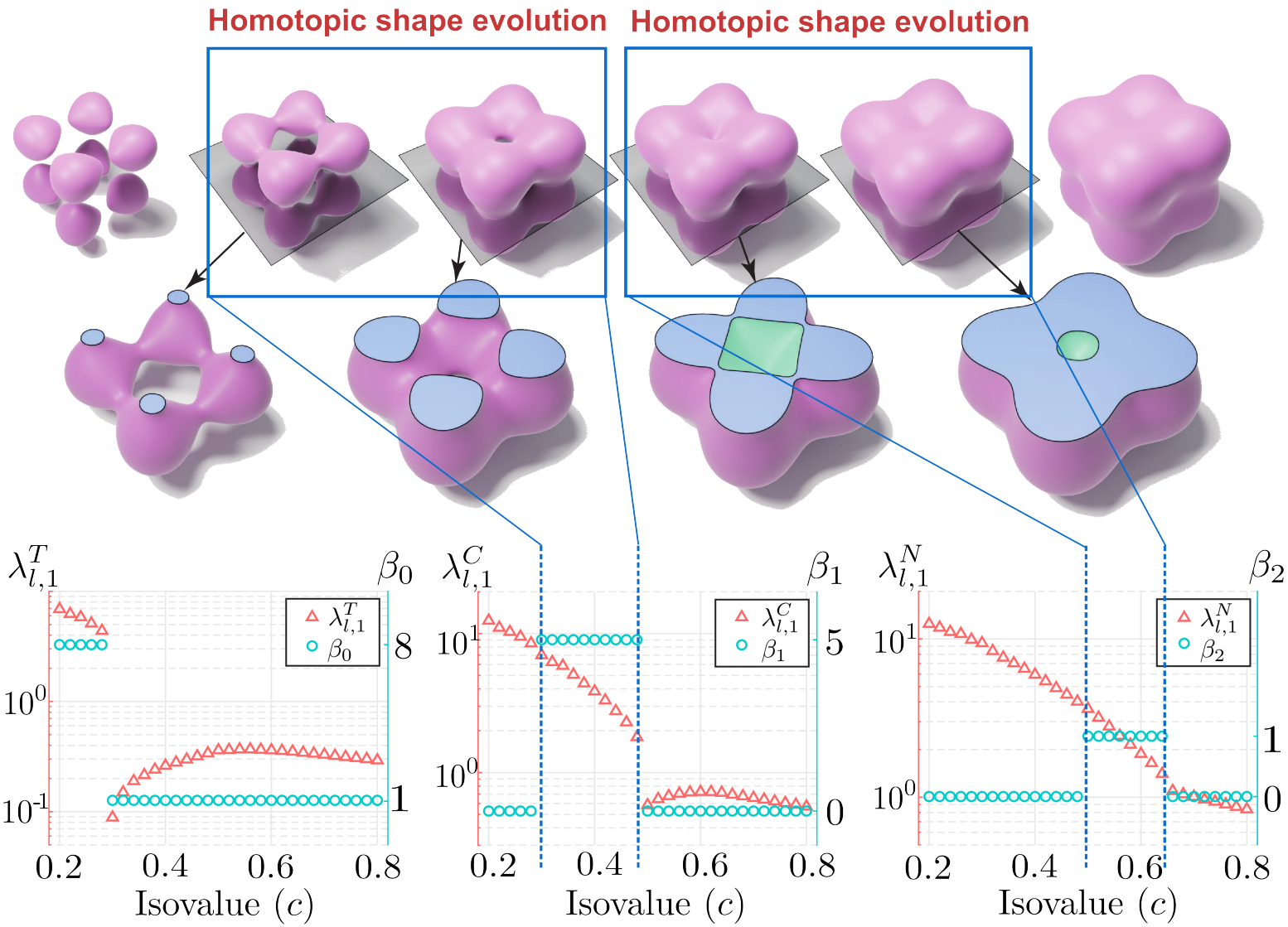




Harmonic (topology) and non-harmonic spectra

[Chen et al., DCDS-B, 2020]

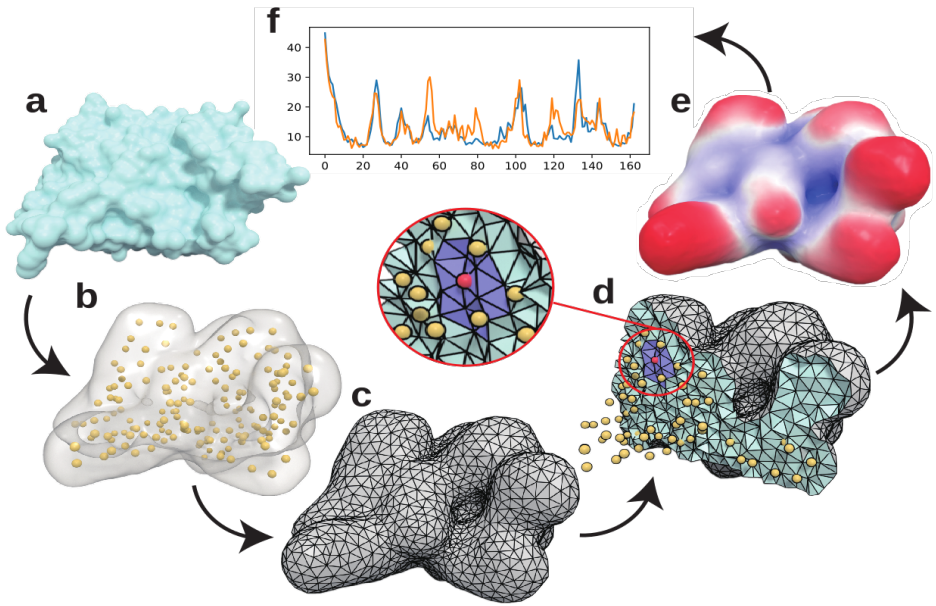
Eight-body System



Harmonic (topology) and non-harmonic spectra

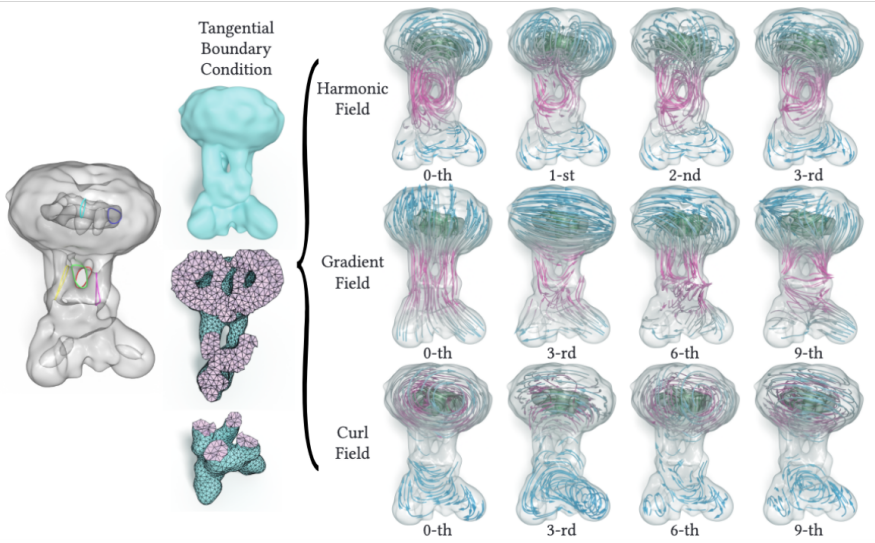
UNIVERSITY OF ARKANSAS

B-factor Prediction



[Zhao et al., Bull. Math. Bio. 2020]

Hodge Decomposition and Mode Analysis



[Zhao et al., Bull. Math. Bio. 2020]

Publication

UNIVERSITY OF ARKANSAS

1. **Chen** et al., *Emerging dominant SARS-CoV-2 variants*, JCIM, 2023
2. **Chen** et al., *A Cartesian FMM-accelerated Galerkin boundary integral Poisson-Boltzmann solver*, JCP, 2023
3. **Chen** et al., *Persistent Laplacian projected Omicron BA. 4 and BA. 5 to become new dominating variants*, CBM, 2022
4. Wilson et al., *Computing electrostatic binding energy with the TABI Poisson-Boltzmann solver*, CIS, 2022
5. **Chen** et al., *Omicron BA.2 (B.1.1.529.2): high potential to becoming the next dominating variant*, JPCL, 2022
6. **Chen** et al., *Mathematical artificial intelligence design of mutation-proof COVID-19 monoclonal antibodies*, CIS, 2022
7. Gao et al., *Methodology-centered review of molecular modeling, simulation, and prediction of SARS-CoV-2*, Chem. Rev., 2022
8. **Chen** et al., *Omicron variant (B.1.1.529): Infectivity, vaccine breakthrough, and antibody resistance*, JCIM, 2022
9. Wang et al., *Emerging vaccine-breakthrough SARS-CoV-2 variants*, ACS infectious diseases, 2021
10. Gao et al., *Perpsectives on SARS-CoV-2 main protease inhibitors*, JMC, 2021
11. Wang et al., *Mechanisms of SARS-CoV-2 evolution revealing vaccine-resistant mutationsin Europe and America*, JPCL, 2021
12. **Chen** et al., *MLIMC: Machine learning-based implicit-solvent Monte Carlo*, CJCP, 2021
13. **Chen** et al., *Revealing the threat of emerging SARS-CoV-2 mutations to antibody therapies*, JMB, 2021
14. Wang et al., *Vaccine-escape and fast-growing mutations in the United Kingdom, the United States, Singapore, Spain, India, and other COVID-19-devastated countries*, Genomics, 2021
15. **Chen** et al., *Prediction and mitigation of mutation threats to COVID-19 vaccines and antibody therapies*, Chemical Science, 2021
16. Wang et al., *HERMES: persistent spectral graph software*, FDS, 2021

Publication

17. Strubbe-Rivera et al., *Modeling the effects of calcium overload on mitochondrial ultrastructural remodeling*, Applied Sciences, 2021
18. Wang et al., *Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants*, Communications biology, 2021
19. **Chen** et al., *SARS-CoV-2 becoming more infectious as revealed by algebraic topology and deep learning*, CIS, 2021
20. **Chen** et al., *Cyclically parallelized treecode for fast computations of electrostatic interactions on molecular surfaces*, CPC, 2021
21. **Chen** et al., *Computing protein pKas using the TABI Poisson-Boltzmann Solver*, JCBC, 2020
22. Wang et al., *Decoding asymptomatic COVID-19 infection and transmission*, JPCL, 2020
23. **Chen** et al., *Mutations strengthened SARS-CoV-2 infectivity*, JMB, 2020
24. Zhao et al., *The de Rham-Hodge analysis and modeling of biomolecules*, BMB, 2020
25. **Chen** et al., *Review of COVID-19 antibody therapies*, ARB, 2020
26. Gao et al., *Repositioning of 8565 existing drugs for COVID-19*, JPCL, 2020
27. Nguyen et al., *Unveiling the molecular mechanims of SARS-CoV-2 main protease inhibition from 137 crystal structures using algebraic topology and deep learning*, Chemical Science, 2020
28. **Chen** et al., *Eveolutionary de Rham-Hodge method*, DCDS-B, 2020
29. **Chen** et al., *On preconditioning the Treecode-accelerated boundary integral (TABI) Poisson-Boltzmann solver*, JCP, 2018
30. Jurrus et al., *Improvements to the APBS Biomolecular Solvation Software Suite*, Protein science, 2018
31. **Chen** et al., *Parallel Computing of the Adaptive N-body Treecode Algorithm for Solving Boundary Integral Poisson-Boltzmann Equation*, Lecture Notes in Computer Science, 2016

Questions

UNIVERSITY OF ARKANSAS

Thank you!