How Math and AI are Revolutionizing Biosciences

Guo-Wei Wei

Mathematics

Biochemistry and Molecular Biology

Michigan State University

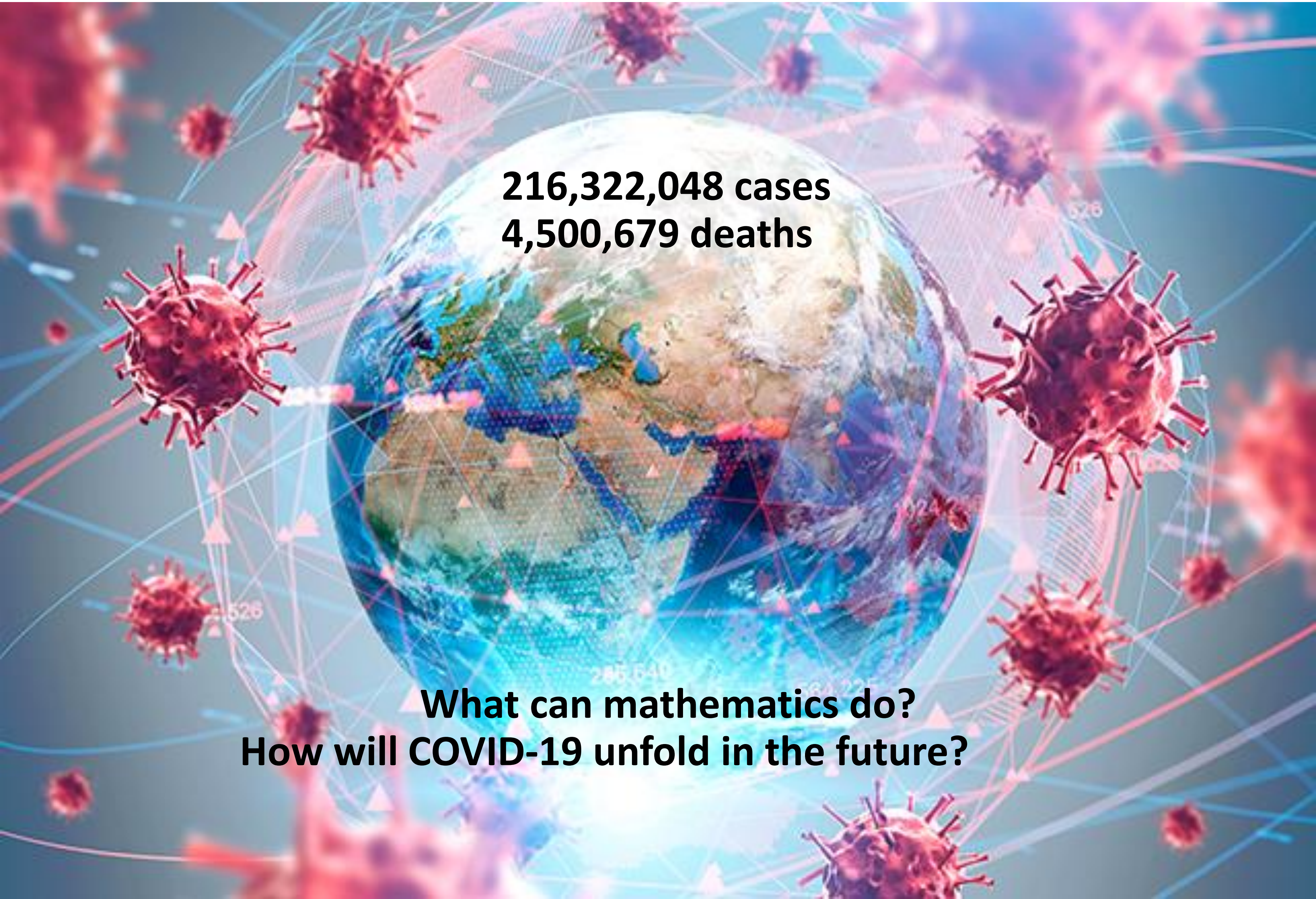
http://www.math.msu.edu/~wei

Beyond TDA - Persistent topology and its applications in data sciences

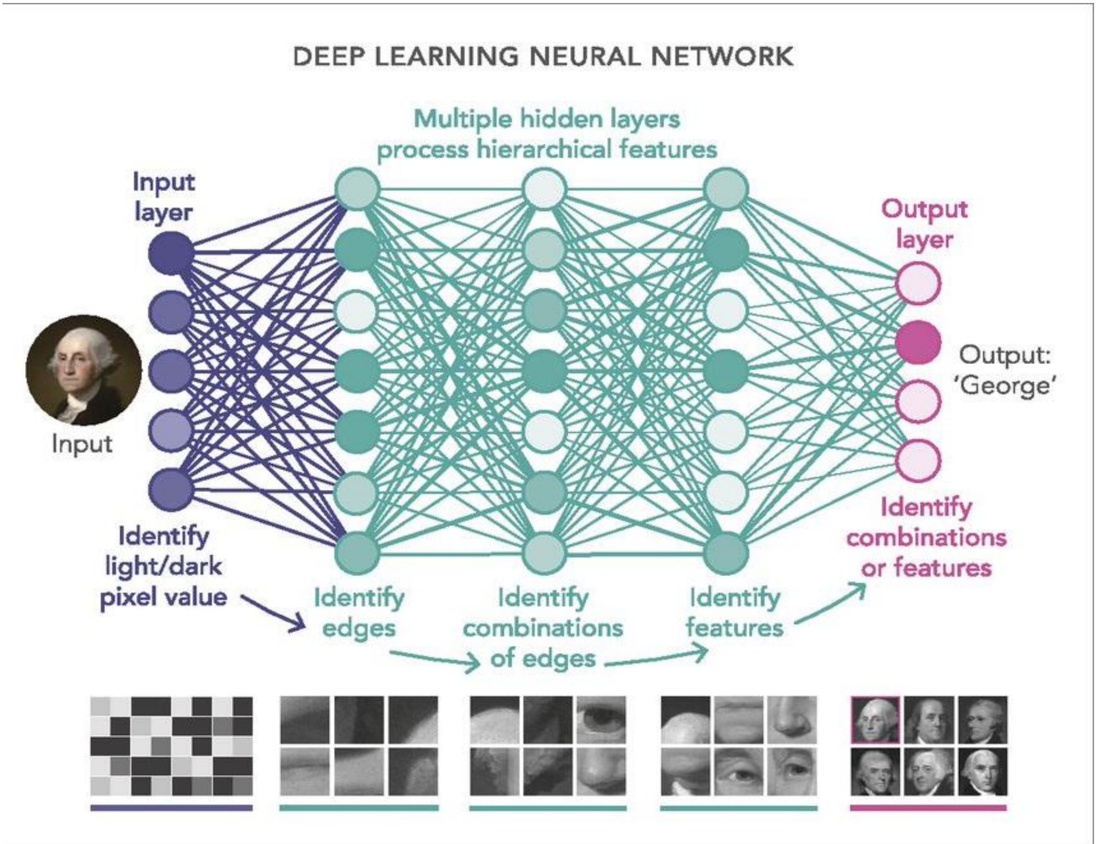
August 28, 2021

Grant support: NIH, NSF, NASA, Pfizer, BMS, MEDC, Georgetown U, COVID-19 HPC Consortium, and MSU Foundation

COVID-19 demonstrates the importance of biosciences

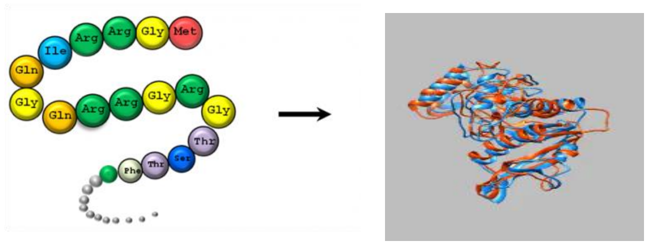


The Promise of AI & Machine Learning

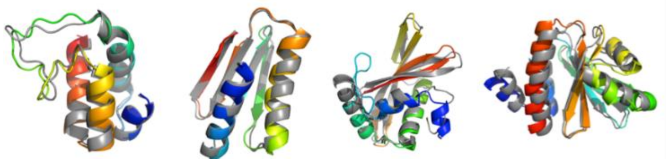


ALPHAFOLD

won 25 of 43 contests and was ranked 1st among 98 competitors in CASP13, Dec. 2018.

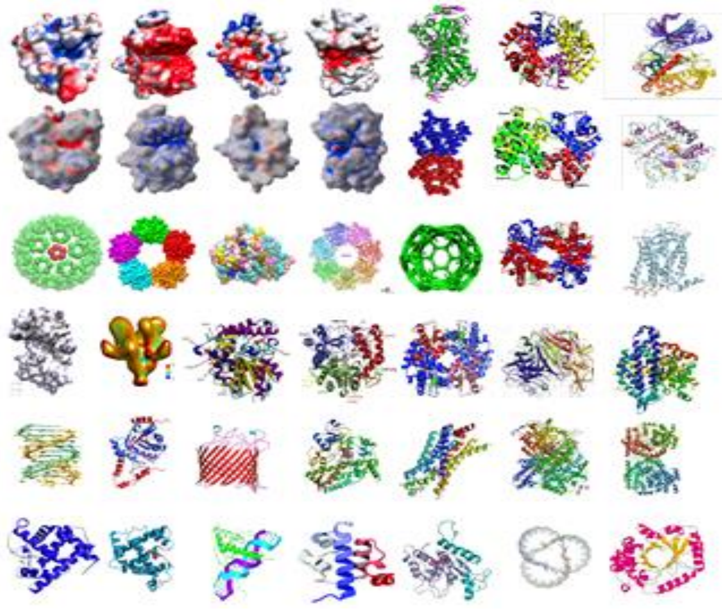


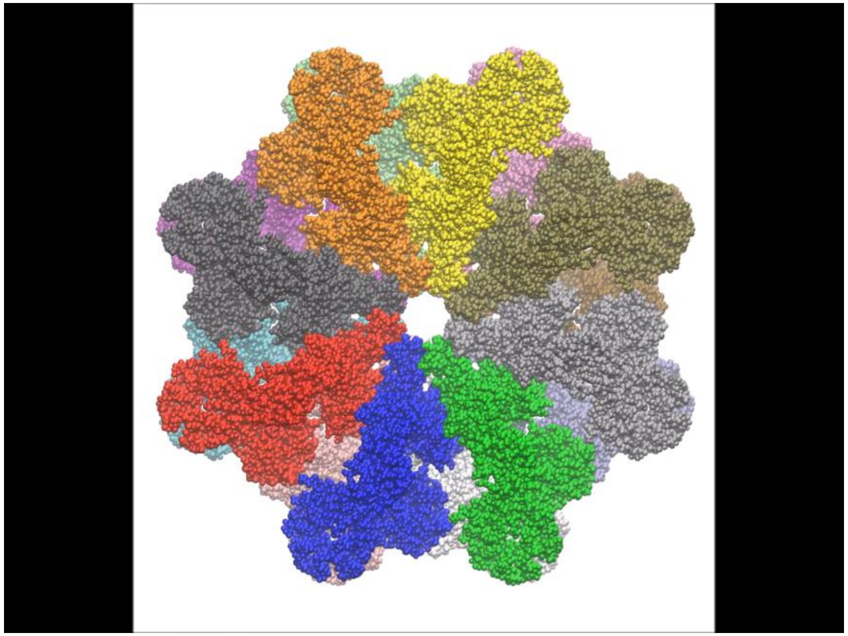
Input | Output



Challenges of AI in biomolecular systems

- **Geometric dimensionality: $\mathbb{R}^{3N}$, where *N* ~5000 for a protein.**
- **Machine learning dimensionality: > $1024^3 m$, where *m* is the number of atom types in a protein.**
- **Molecules have different sizes --- non-scalable.**
- **Complexity: intermolecular & intramolecular interactions**





Two schools of thinking

Given a protein with N atom and an average of n electrons in each atom

Fundamentalism; Mechanistic

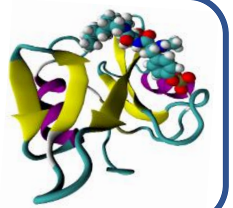
Quantum Mechanics $\mathbb{R}^{3Nn+3N}$

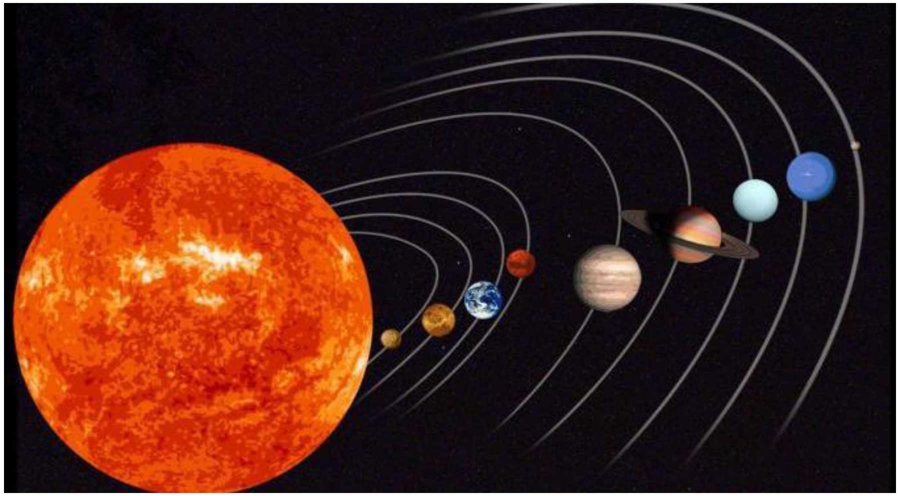
QM/MM $\mathbb{R}^{K}$ $3N< K <3N(n+1)$

Molecular Mechanics $\mathbb{R}^{3N}$

Multiscale Coarse-grain $\mathbb{R}^{M}$ $(3<M<3N)$

Poisson-Boltzmann, PNP, etc. $\mathbb{R}^{3}$





Differentiable Manifold $\mathbb{R}^{2}$

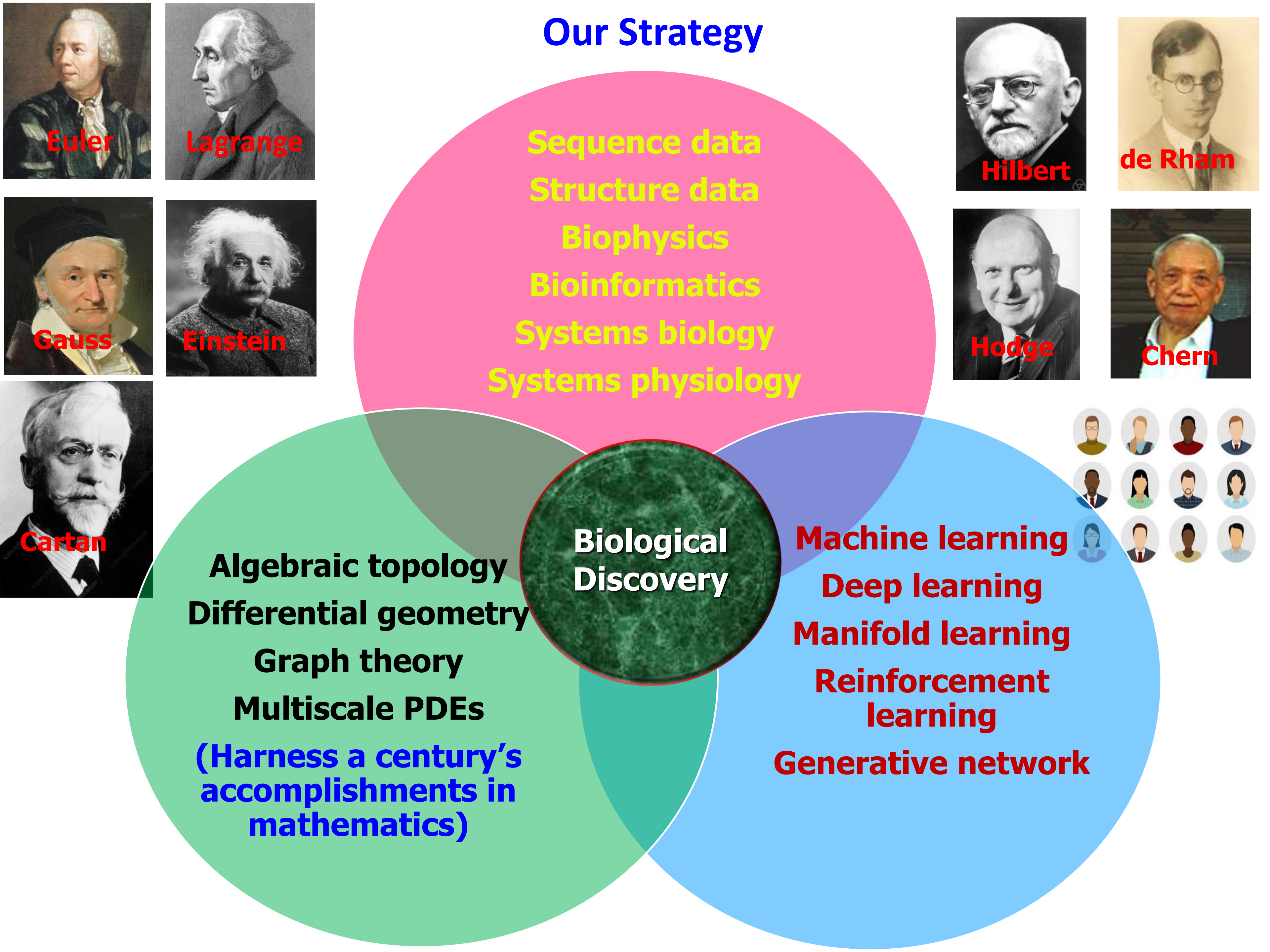
Algebraic Topology $\mathbb{R}^{1}$

Graph Theory $\mathbb{R}^{0}$

Index Theory $\mathbb{R}^{0}$

Reductionism; Data-driven

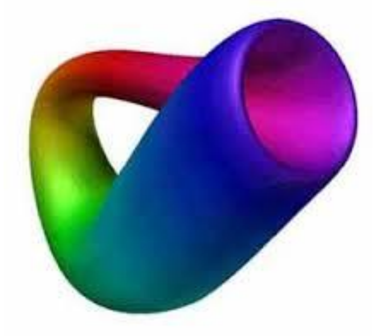
Basic hypothesis: Intrinsic physics lies on low-dimensional manifolds in a high dimensional space

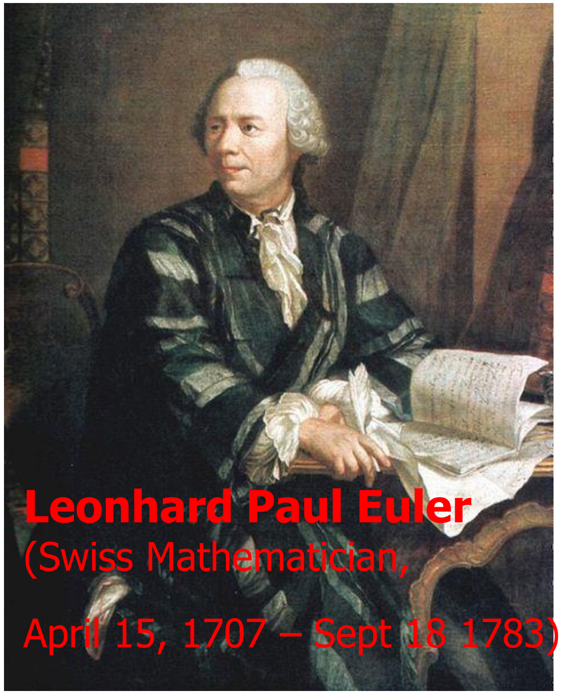

Our Strategy
Euler
Lagrange
Gauss
Einstein
Cartan
Hilbert
de Rham
Hodge
Chern
Sequence data
Structure data
Biophysics
Bioinformatics
Systems biology
Systems physiology
Algebraic topology
Differential geometry
Graph theory
Multiscale PDEs
(Harness a century's accomplishments in mathematics)
Biological Discovery
Machine learning
Deep learning
Manifold learning
Reinforcement learning
Generative network

Classical Topology

Möbius Strips (1858)

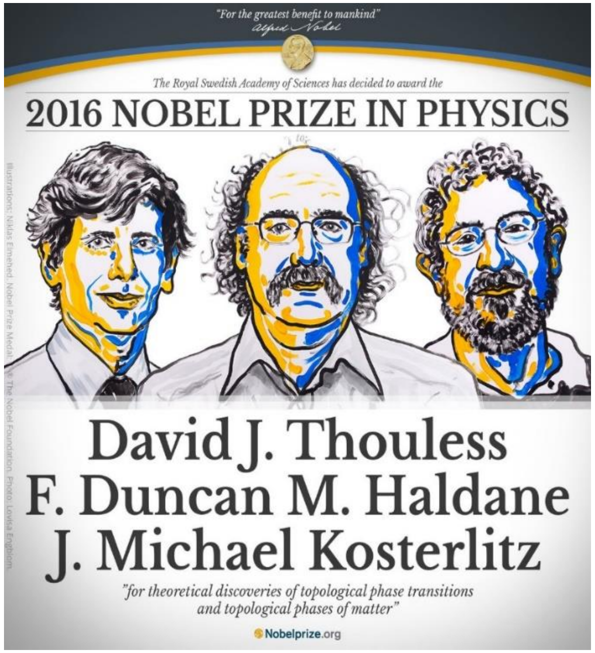
Klein Bottle (1882)



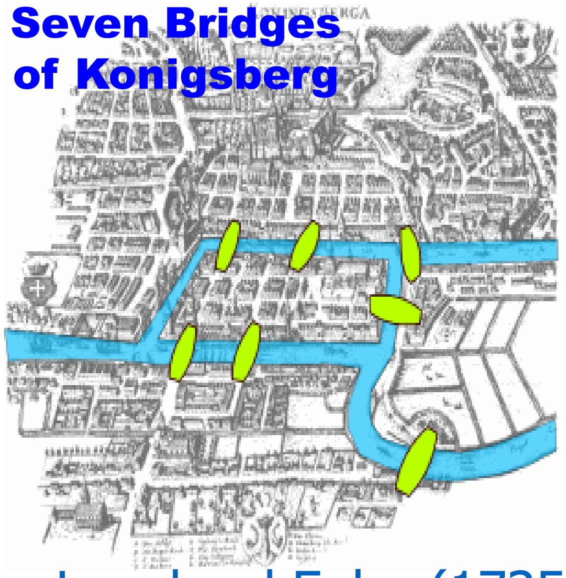


Torus

Double Torus



Augustin-Louis Cauchy,
Ludwig Schläfli,
Johann Benedict Listing,
Bernhard Riemann, and
Enrico Betti



Leonhard Euler (1735)

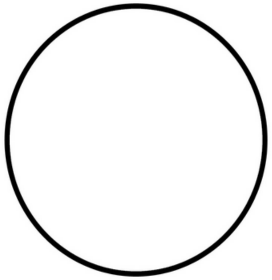
Topological invariants: Betti numbers

β_0 is the number of connected components.
β_1 is the number of tunnels or circles.
β_2 is the number of cavities or voids.

Point

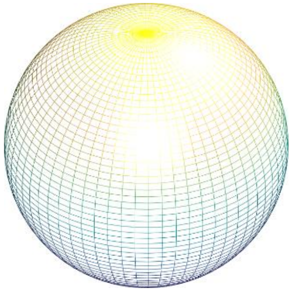
$$\beta_0 = 1 \quad \beta_1 = 0 \quad \beta_2 = 0$$

Circle



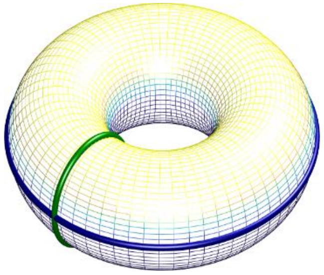
$$\beta_0 = 1 \quad \beta_1 = 1 \quad \beta_2 = 0$$

Sphere



$$\beta_0 = 1 \quad \beta_1 = 0 \quad \beta_2 = 1$$

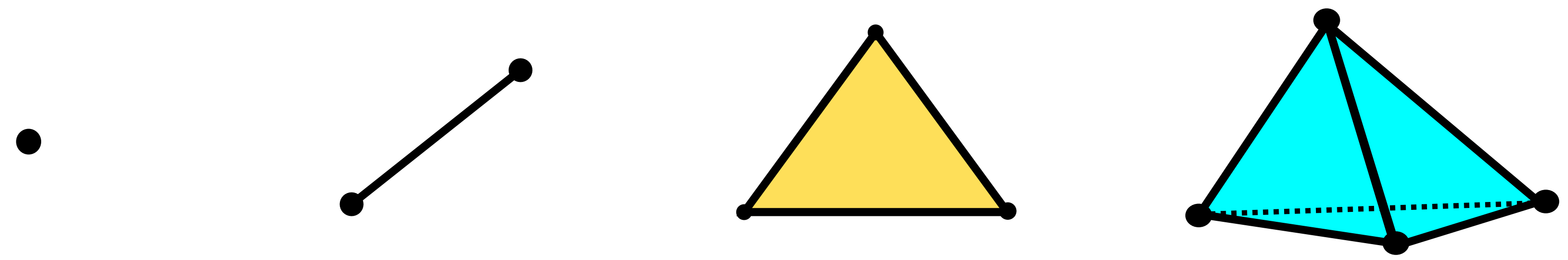
Torus



$$\beta_0 = 1 \quad \beta_1 = 2 \quad \beta_2 = 1$$

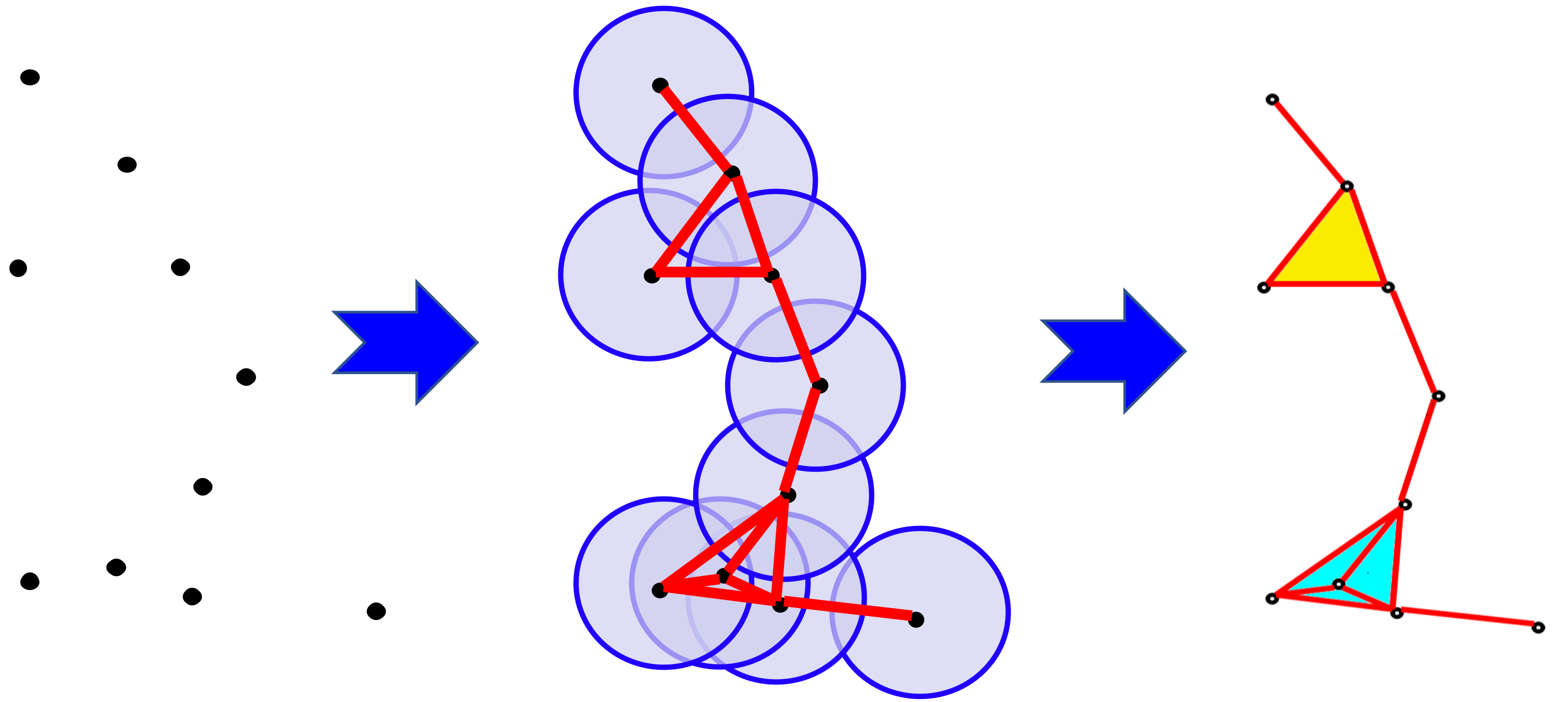
Vietoris-Rips complexes of planar point sets

Simplexes:



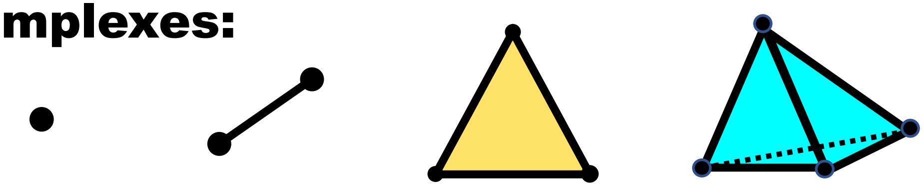
0-simplex *1*-simplex *2*-simplex *3*-simplex

Simplicial complexes of ten points:



Persistent homology

Simplexes:



0-simplex *1*-simplex *2*-simplex *3*-simplex

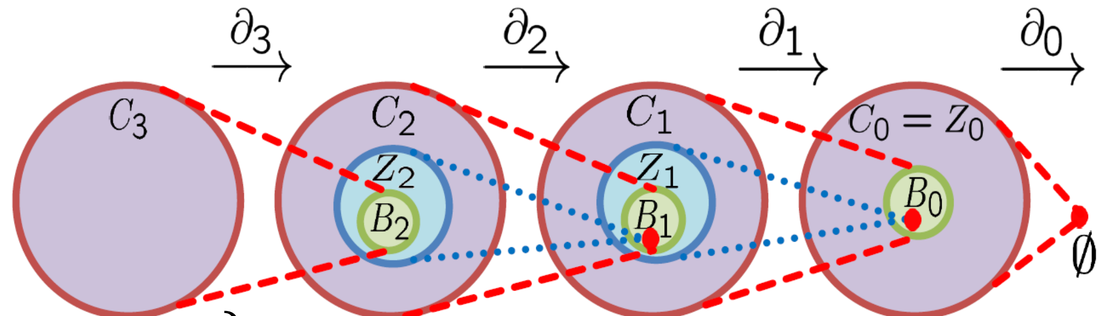
Frosini and Nandi (1999), Robins (1999), Edelsbrunner, Letscher and Zomorodian (2002), Zomorodian and Carlsson (2005), Edelsbrunner and Harer, (2007) Kaczynski, Mischaikow and Mrozek (2004),…

k-chain: $K = \left\{ \sum_j c_j \sigma_j^q \right\}$

Chain group: $C_q(K, \mathbb{Z}_2)$

Boundary operator:

$$\partial_q \sigma^q = \sum_{j=0}^{q} (-1)^j \{ v_0, v_1, \ldots, \widehat{v_j}, \ldots, v_k \}$$

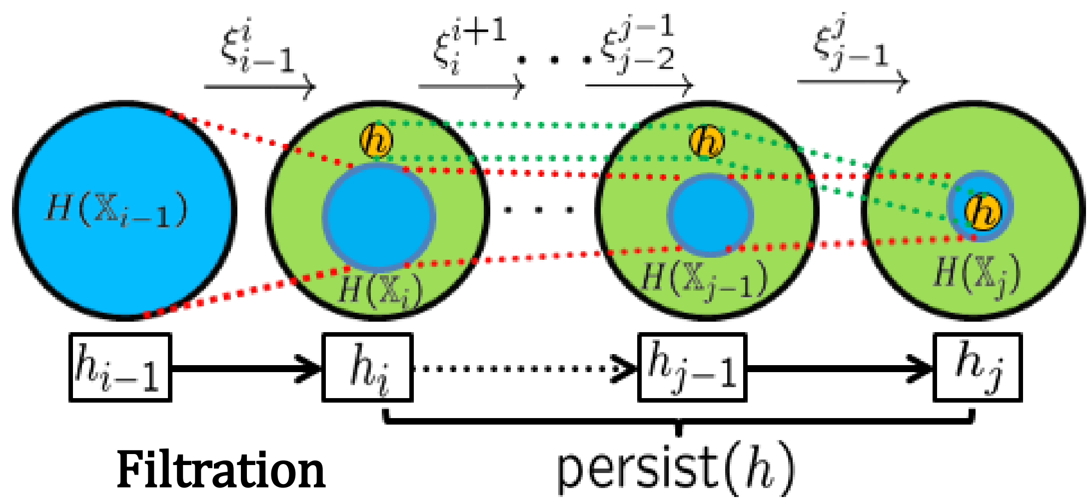


Cycle group: $Z_q = \text{Ker } \partial_q$

Boundary group: $B_q = \text{Im } \partial_{q+1}$

Homology group: $H_q = Z_q / B_q$

Betti number: $\beta_q = \text{Rank}(H_q)$



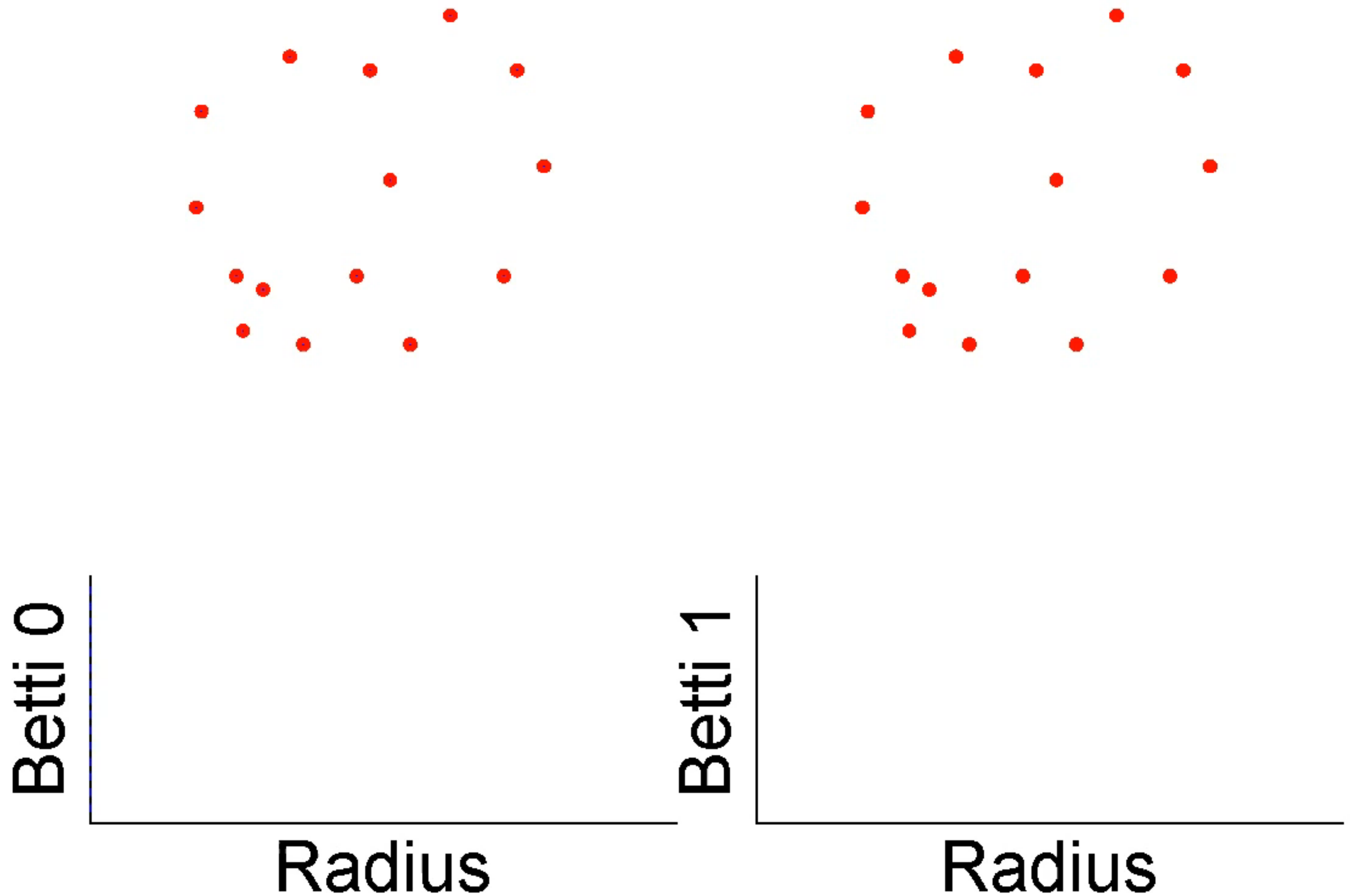
Filtration

Xia, Wei, IJNMBE, 2014;
Xia, Feng, Tong, Wei, JCC, 2015

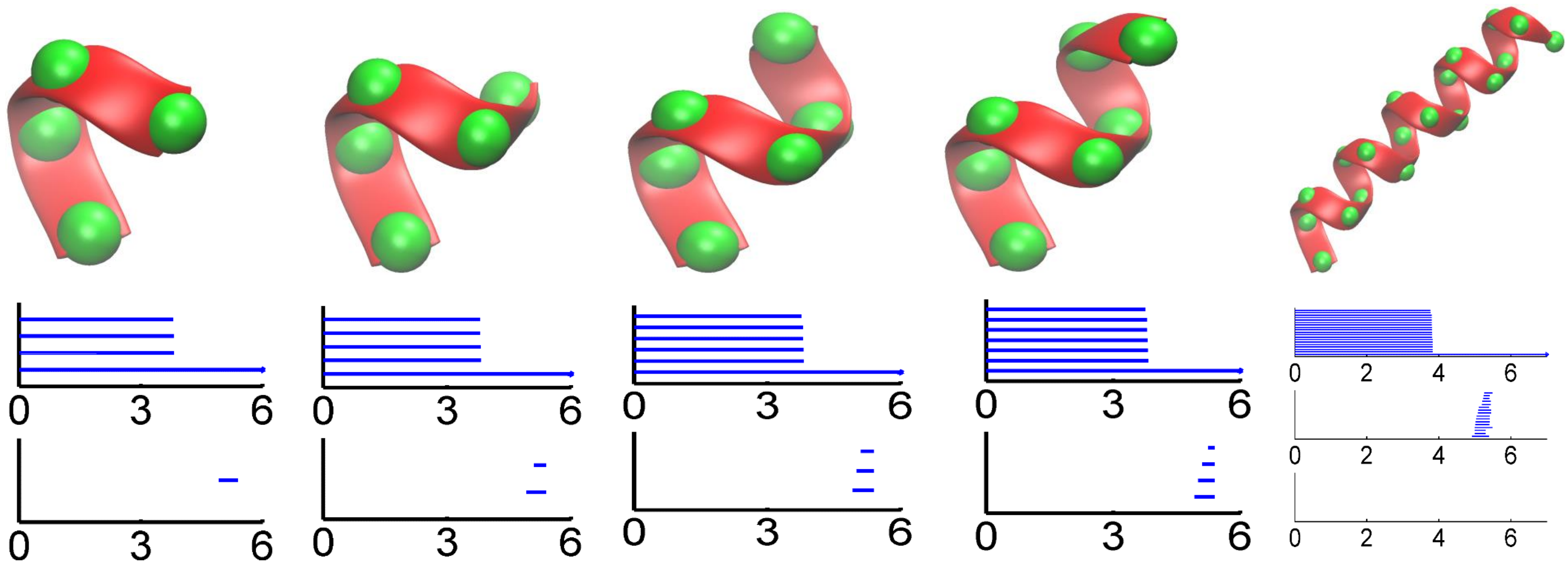
Algebraic Topology

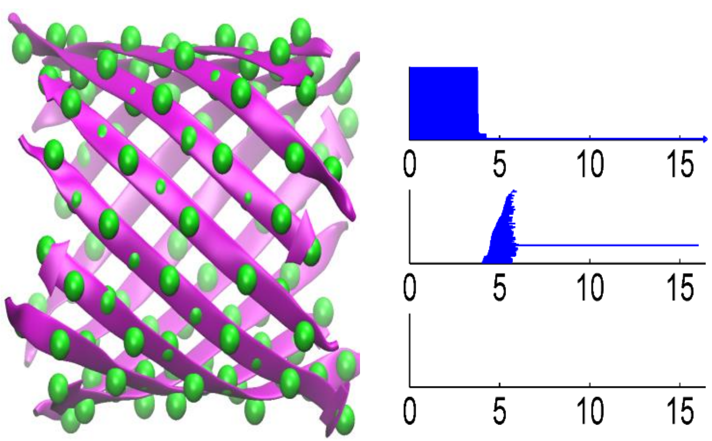
Vietoris-Rips complexes, persistent homology and topological fingerprint

(Xia, Wei, 2014)

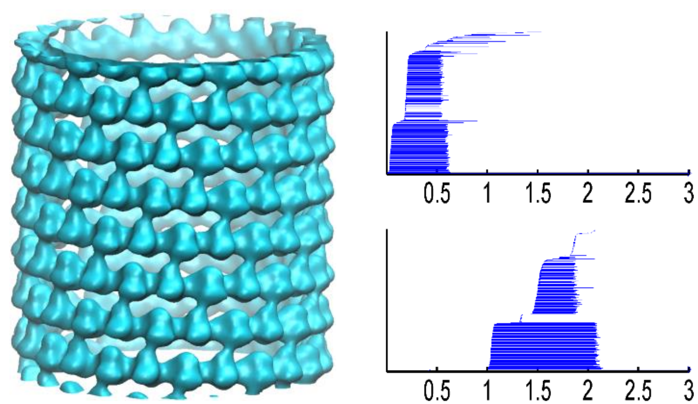


Topological fingerprints of an alpha helix





Beta barrel

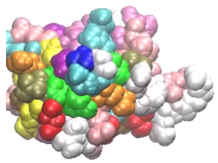


Microtubule

(Xia & Wei, IJNMBE, 2014, 2015)

Algebraic Topology

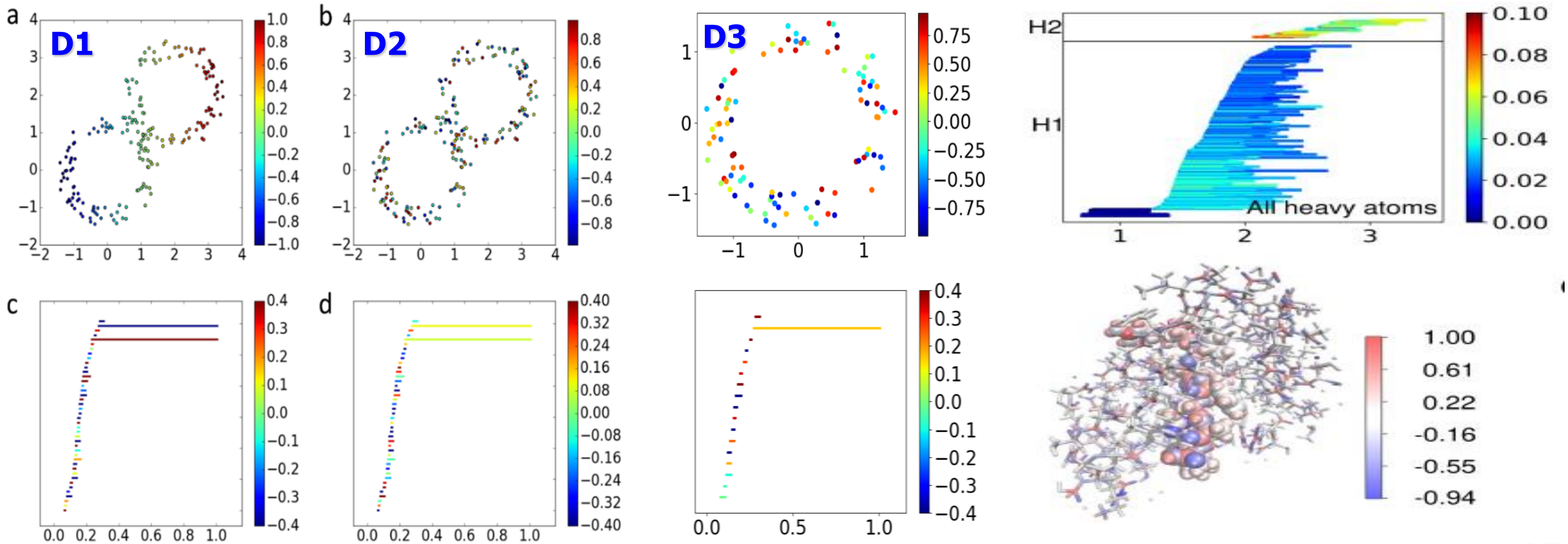
2D persistent homology of protein unfolding (1UBQ)

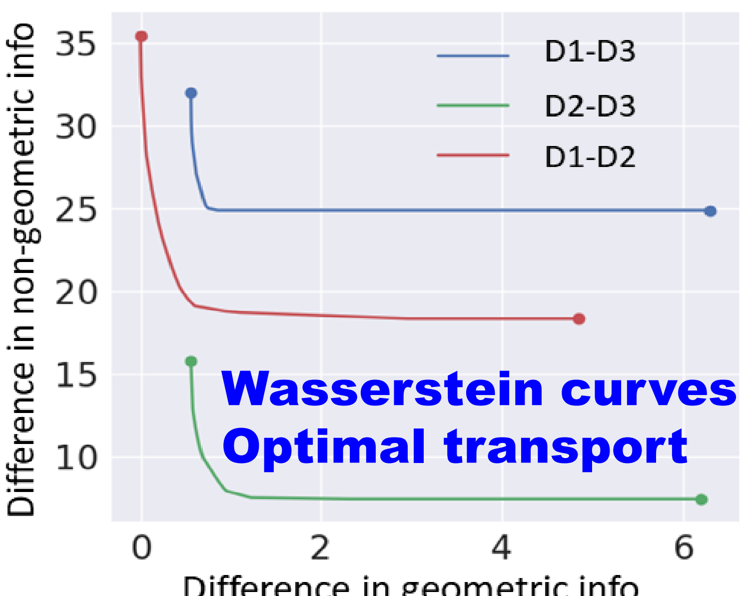


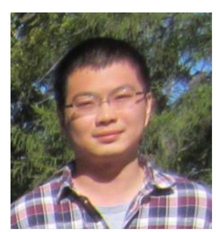
(Xia & Wei, JCC, 2015)

Kelin Xia

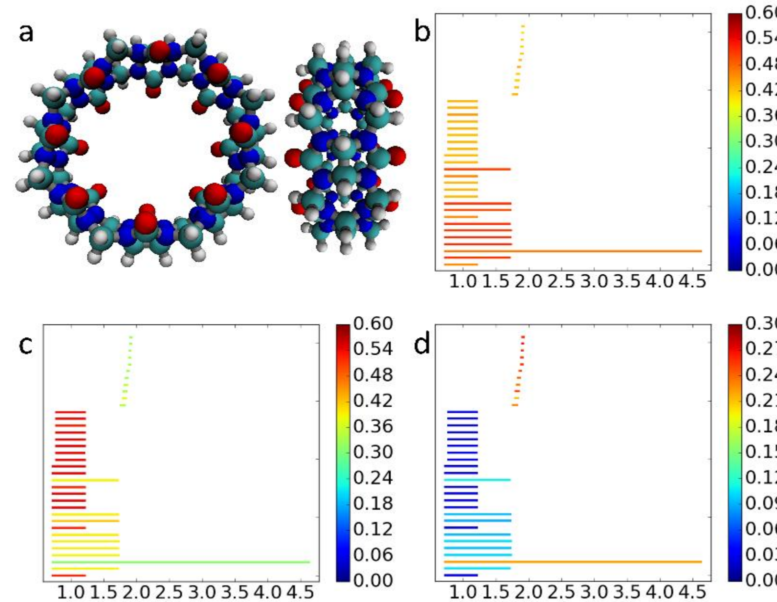
Persistent cohomology
(incorporating non-geometric information in topology)



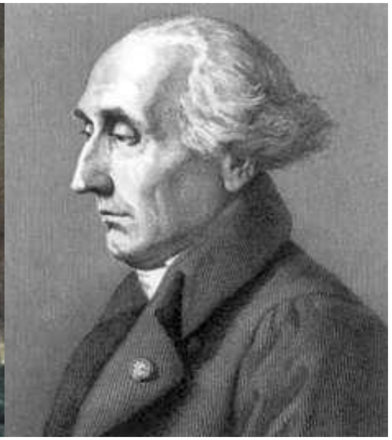




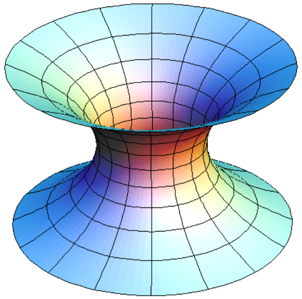
Zixuan Cang
SIAM JMDS
2020

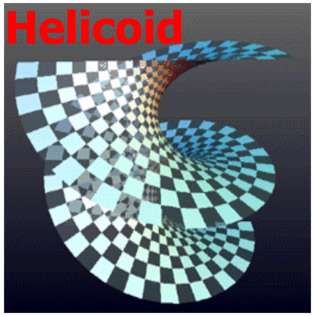


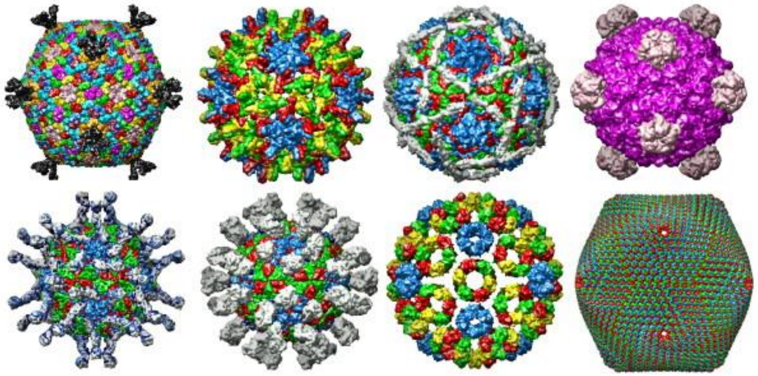
Leonhard P. Euler (Swiss Mathematician, April 15, 1707 – Sept 18 1783



Joseph L. Lagrange (Italian Mathematician, January 25 1736 – April 10, 1813)



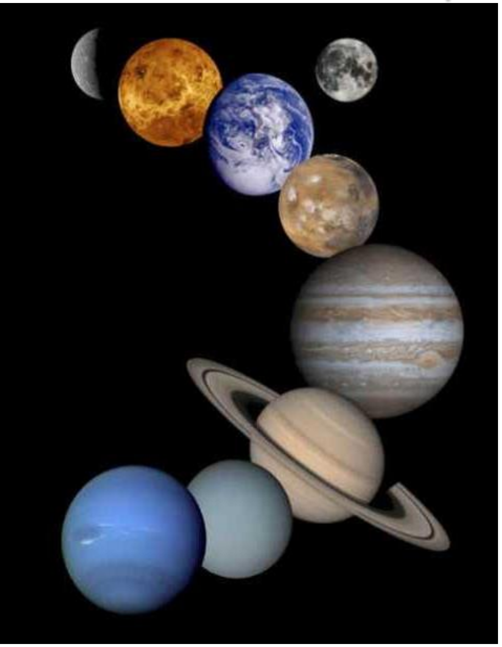


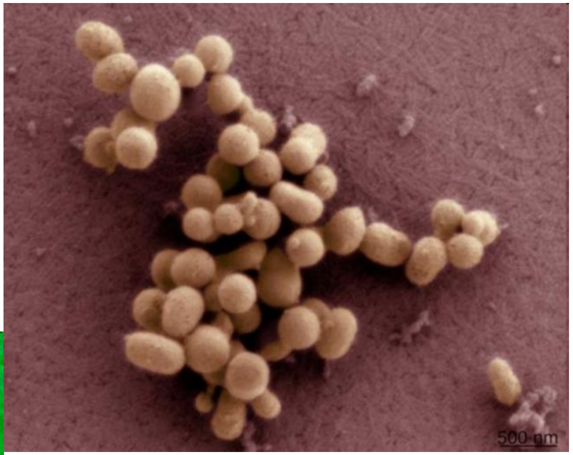


Viral morphology

Minimal Surfaces

A way to minimize energy and maximize stability



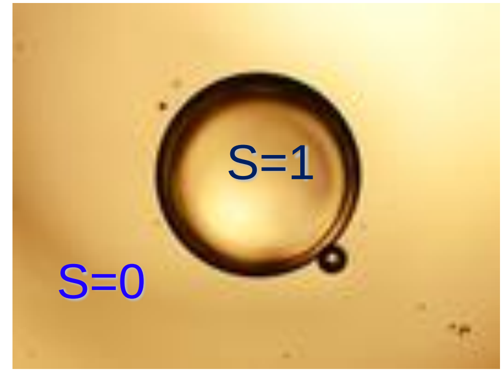


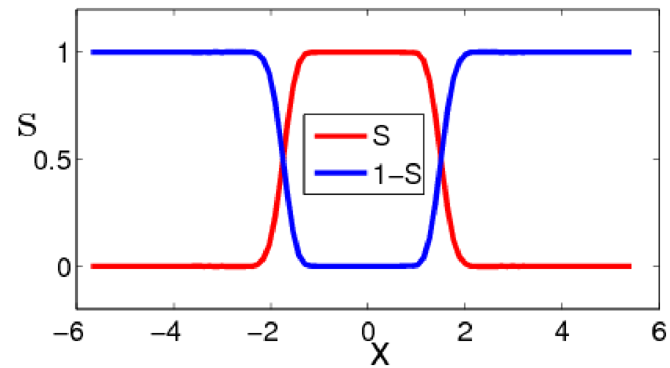
Man-made life, Mycoplasma mycoides

Differential geometry based minimal surface model

$$G = \int \gamma[\text{area}]\, d\boldsymbol{r} \qquad \text{area} = |\nabla S|$$

where G is the surface energy, gamma (γ) is the surface tension, and S is a surface characteristic function:

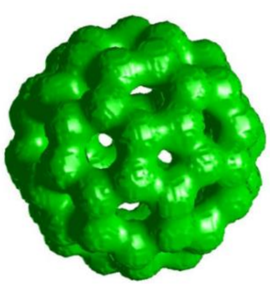


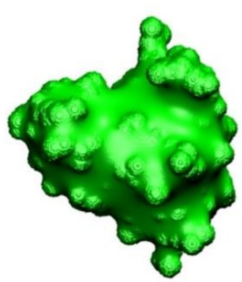


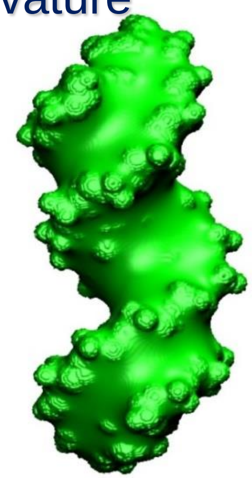
Generalized Laplace-Beltrami flow:

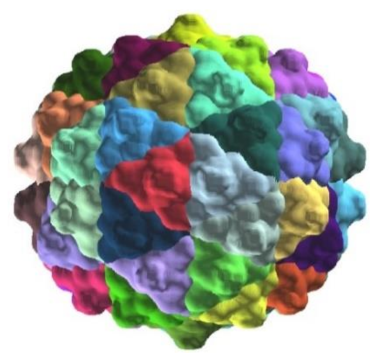
$$\frac{\partial S}{\partial t} = |\nabla S| \left[\nabla \cdot \frac{\gamma \nabla S}{|\nabla S|} \right]$$

Mean curvature





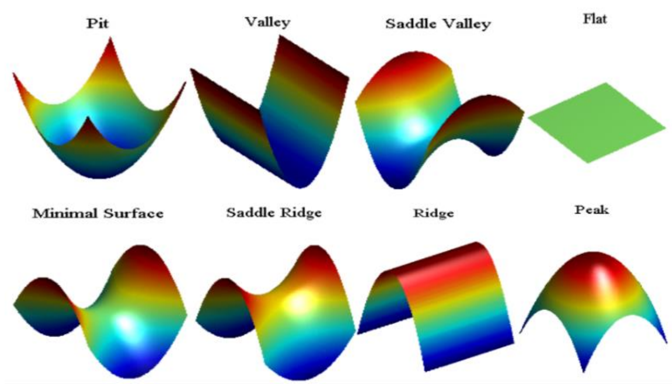




Shan Zhao

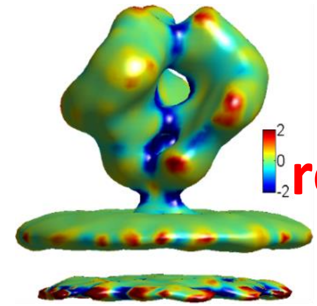
(Bates, Wei, Zhao, 2006; JCC,2008; Zhao, Cang, Tong & Wei, Bioinformatics 2018)

Differential Geometry (Connections & curvature forms)

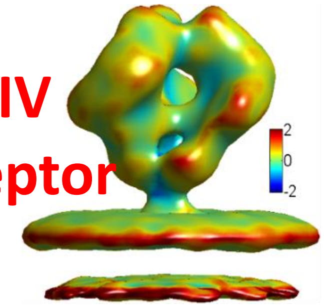


Kelin Xia

Gauss

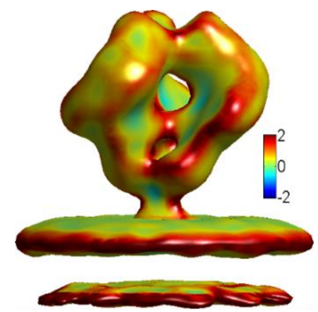


Mean

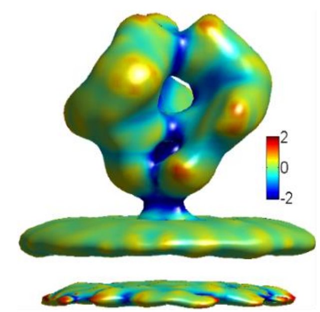


HIV receptor

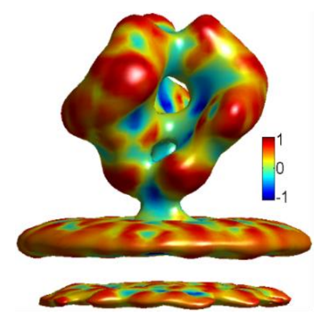
Minimum



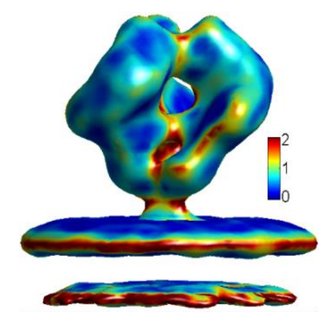
Maximum



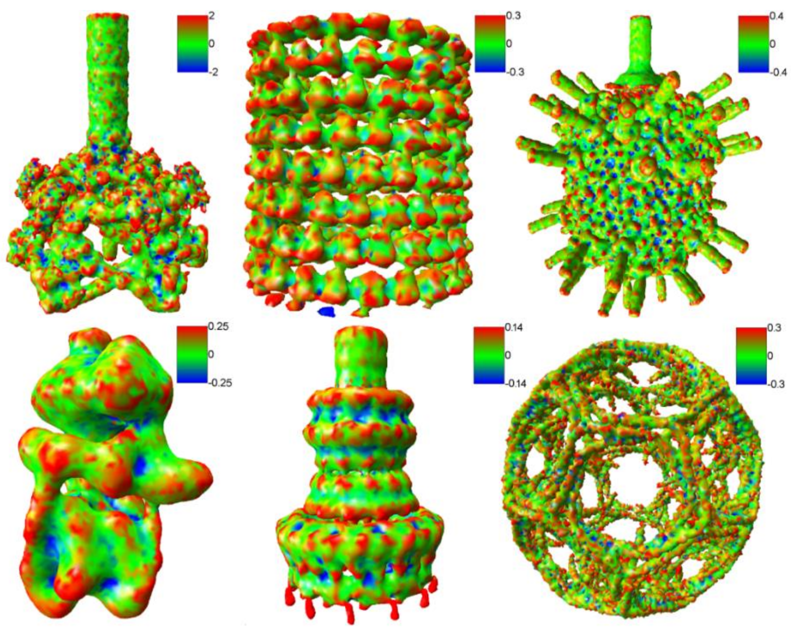
Shape index



Curvedness



Mean curvatures of subcellular structures

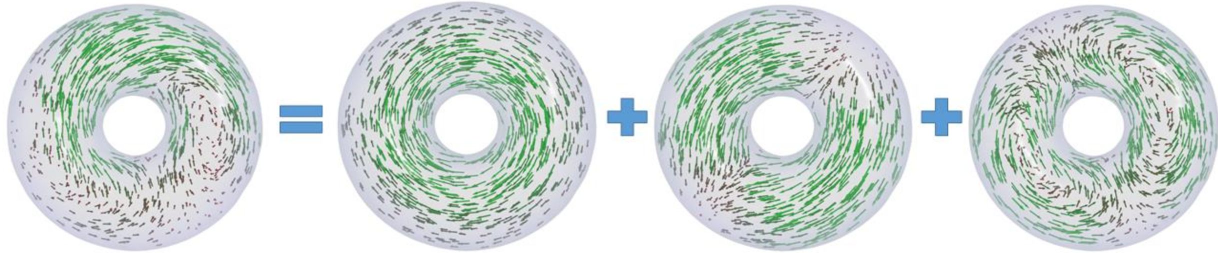


(Feng, Xia, Tong and Wei, JCP, IJNMBI,2012)

De Rham-Hodge theory and discrete exterior calculus

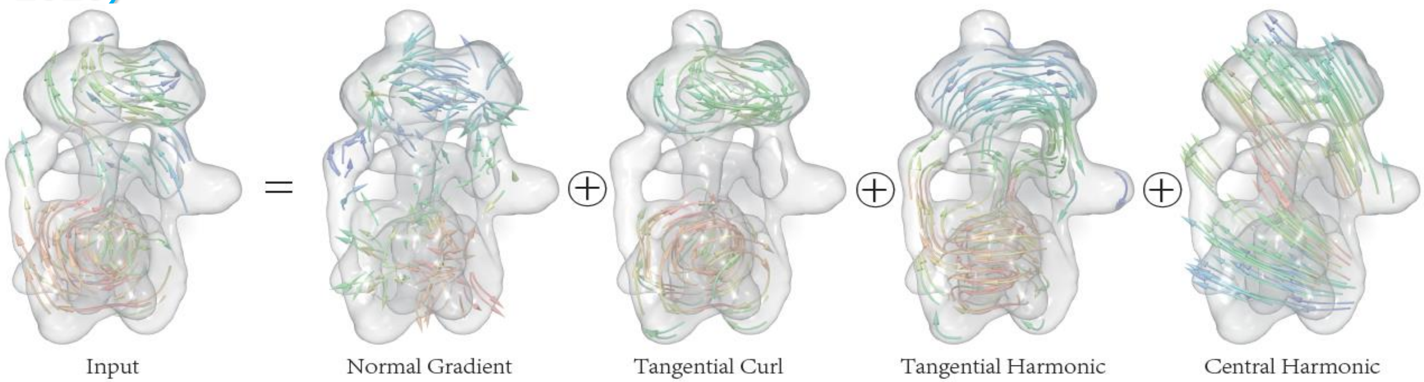
Hodge decomposition:

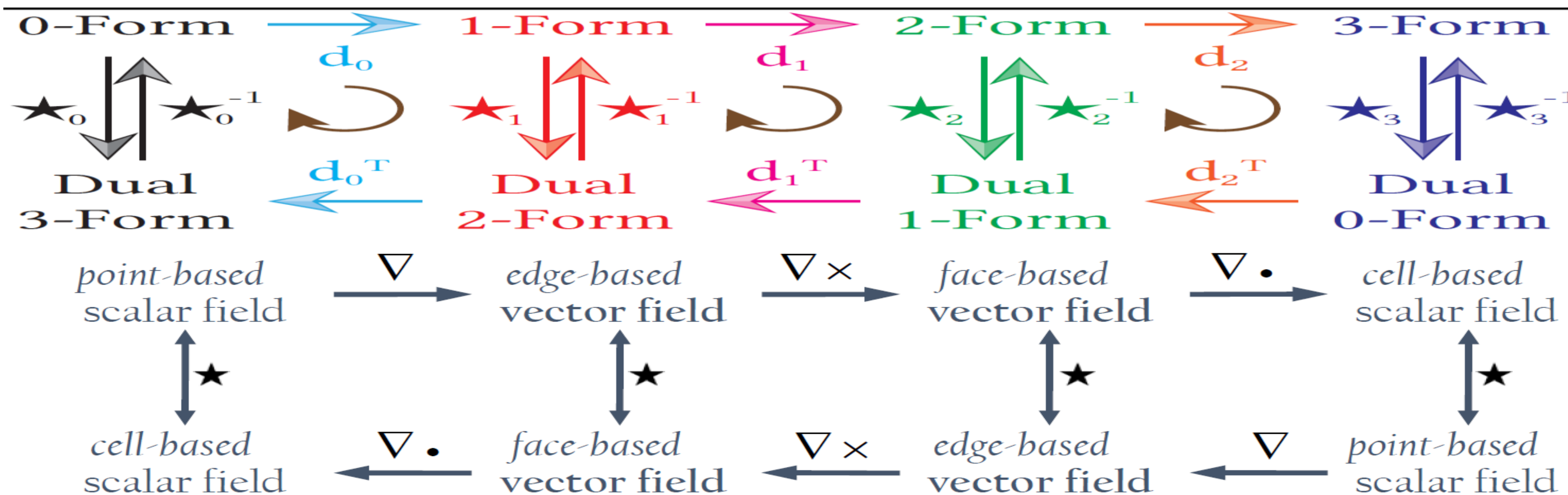
(Zhao, Wang, Chen, Tong & Wei, BMB, 2020)

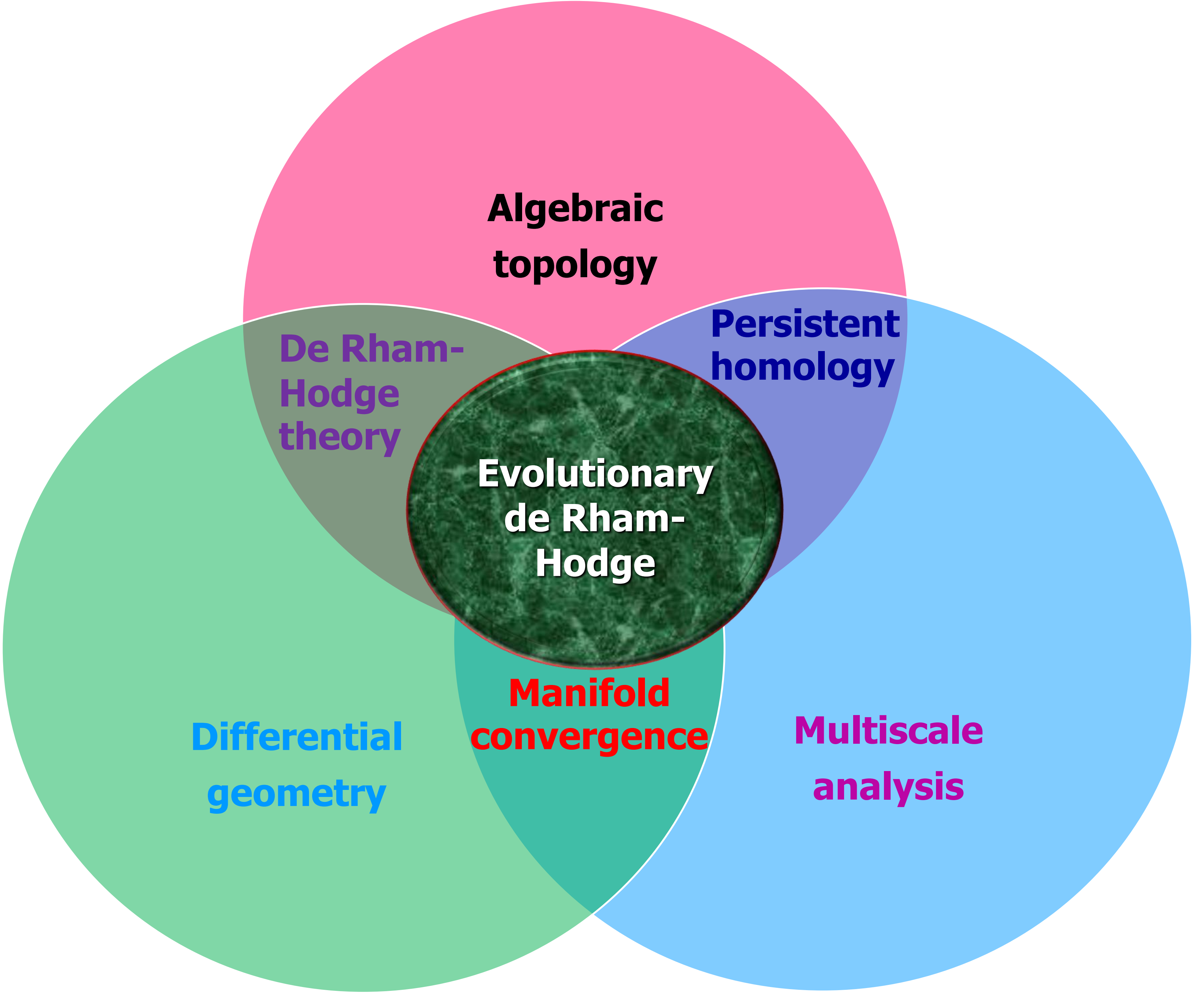


A vector field = Harmonic + curl-free + divergent-free

Cryo-EM data:






Algebraic
topology
De Rham-
Hodge
theory
Persistent
homology
Evolutionary
de Rham-
Hodge
Manifold
convergence
Differential
geometry
Multiscale
analysis

Evolutionary de Rham-Hodge

Filtration of a manifold

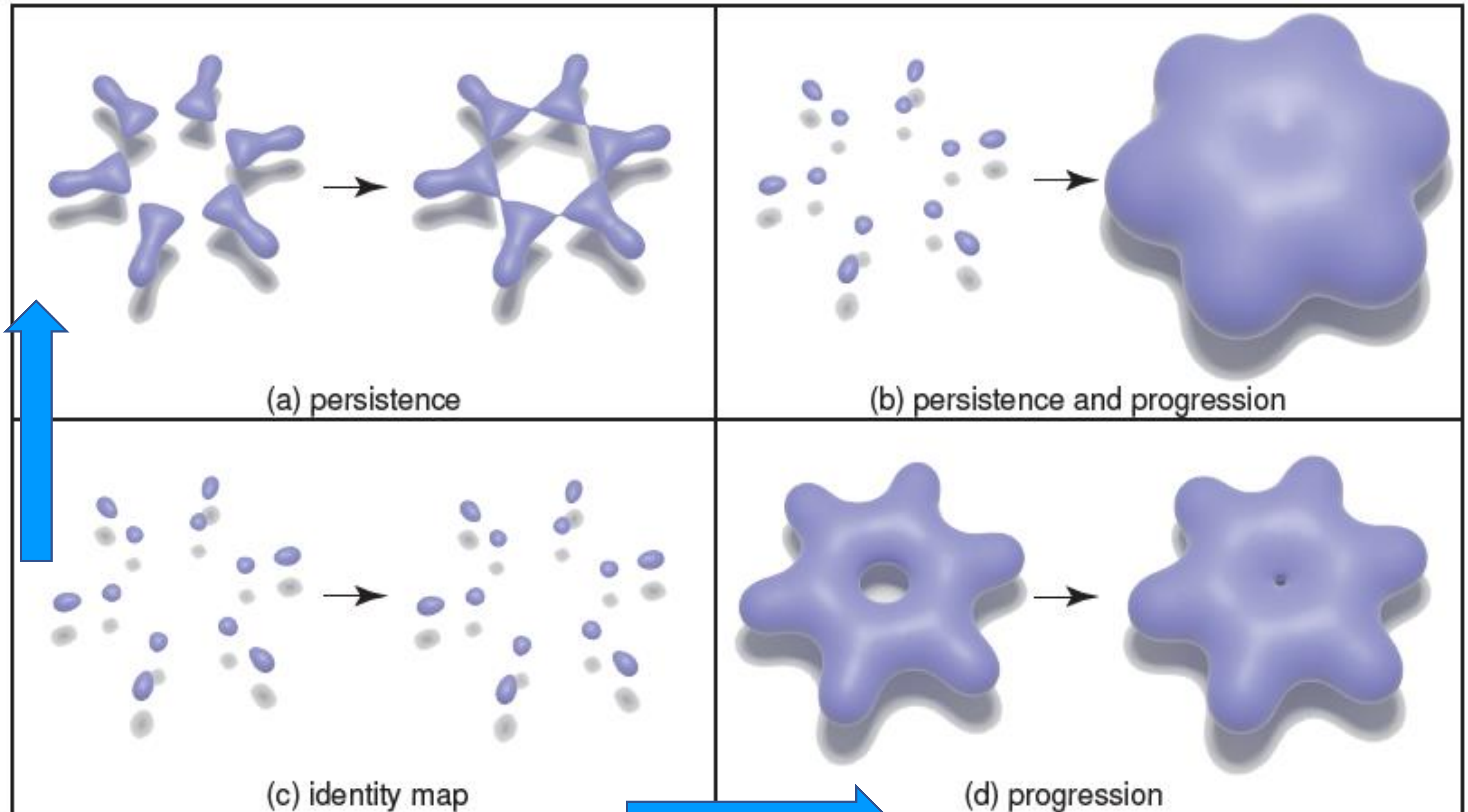
$$M_0 \xrightarrow{\mathfrak{I}_{0,1}} M_1 \xrightarrow{\mathfrak{I}_{1,2}} M_2 \xrightarrow{\mathfrak{I}_{2,3}} \cdots \xrightarrow{\mathfrak{I}_{n-1,n}} M_n \xrightarrow{\mathfrak{I}_{n,n+1}} M$$

De Rham complexes induced by filtration

$$\begin{array}{ccccccc}
\Omega_n^0(M_0) & \xrightarrow{d^0} & \Omega_n^1(M_0) & \xrightarrow{d^1} & \Omega_n^2(M_0) & \xrightarrow{d^2} & \Omega_n^3(M_0) \\
\downarrow \mathfrak{E}_{0,1} & & \downarrow \mathfrak{E}_{0,1} & & \downarrow \mathfrak{E}_{0,1} & & \downarrow \mathfrak{E}_{0,1} \\
\Omega_n^0(M_1) & \xrightarrow{d^0} & \Omega_n^1(M_1) & \xrightarrow{d^1} & \Omega_n^2(M_1) & \xrightarrow{d^2} & \Omega_n^3(M_1) \\
\downarrow \mathfrak{E}_{1,1} & & \downarrow \mathfrak{E}_{1,1} & & \downarrow \mathfrak{E}_{1,1} & & \downarrow \mathfrak{E}_{1,1} \\
\Omega_n^0(M_2) & \xrightarrow{d^0} & \Omega_n^1(M_2) & \xrightarrow{d^1} & \Omega_n^2(M_2) & \xrightarrow{d^2} & \Omega_n^3(M_2) \\
\downarrow \mathfrak{E}_{2,1} & & \downarrow \mathfrak{E}_{2,1} & & \downarrow \mathfrak{E}_{2,1} & & \downarrow \mathfrak{E}_{2,1} \\
\cdots & & \cdots & & \cdots & & \cdots
\end{array}$$

(Chen, Zhao, Tong & Wei, DCDS-B 2020)

Evolutionary de Rham-Hodge



Persistence

Progression (homotopy)

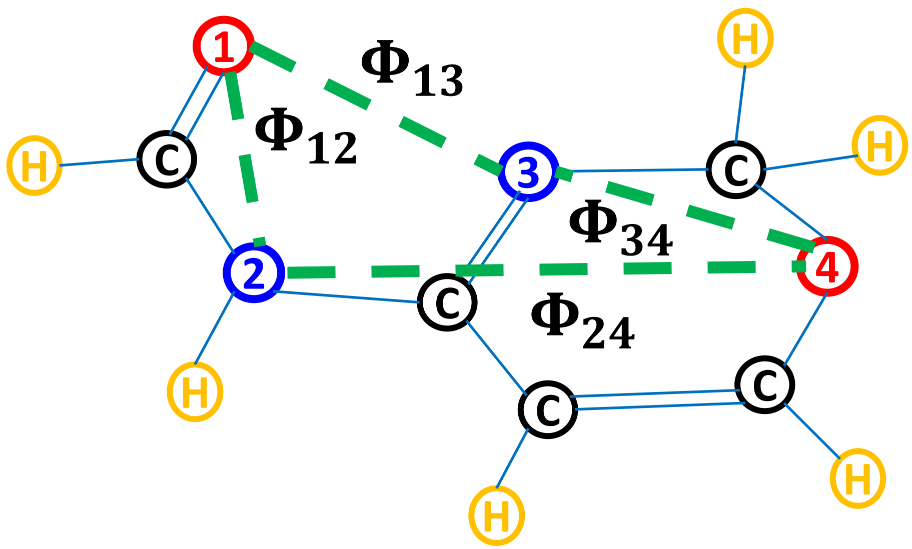
(Chen, Zhao, Tong & Wei, DCDS-B 2020)

Obtain multiscale spectral geometry & persistent topology from *k*-form Hodge Laplacians!

$$\Delta_k^{l,p} = \partial_{k+1}^l d_k^l + d_{k-1}^{l+p} \partial_k^{l+p}$$

Algebraic Graph Theory for Biomolecules

Molecular graph G(V,E)



Adjacency matrix of $G(V_{ON},E)$

$$\begin{pmatrix} 0 & \Phi_{12} & \Phi_{13} & 0 \\ \Phi_{12} & 0 & 0 & \Phi_{24} \\ \Phi_{13} & 0 & 0 & \Phi_{34} \\ 0 & \Phi_{24} & \Phi_{34} & 0 \end{pmatrix}$$

Eigenvalues: $\lambda_1^A, \lambda_2^A, \ldots$

Can one hear the shape of a drum?

Laplacian matrix of $G(V_{ON},E)$

$$\begin{pmatrix} \Phi_{12}+\Phi_{13} & -\Phi_{12} & -\Phi_{13} & 0 \\ -\Phi_{12} & \Phi_{12}+\Phi_{24} & 0 & -\Phi_{24} \\ -\Phi_{13} & 0 & \Phi_{13}+\Phi_{34} & -\Phi_{34} \\ 0 & -\Phi_{24} & -\Phi_{34} & \Phi_{24}+\Phi_{34} \end{pmatrix}$$

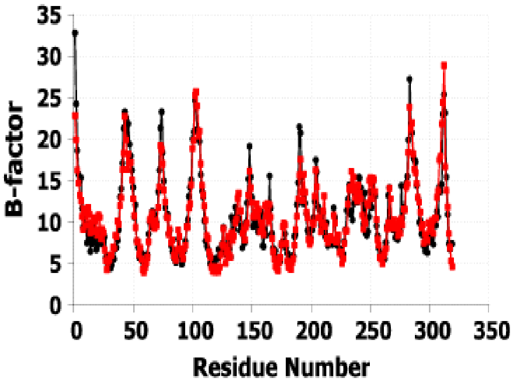
Eigenvalues: $\lambda_1^L, \lambda_2^L, \ldots$

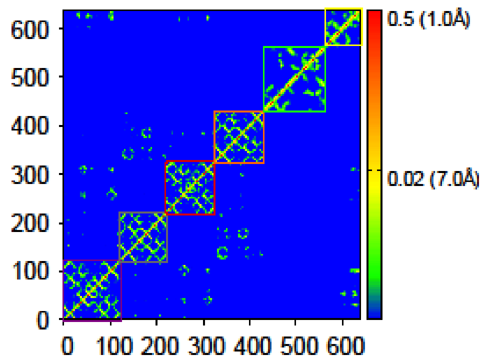
(Nguyen and Wei, JCIM, 2019)

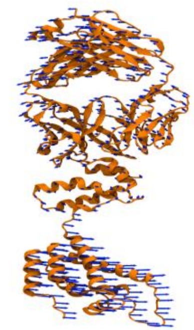
Geometric Graph Theory

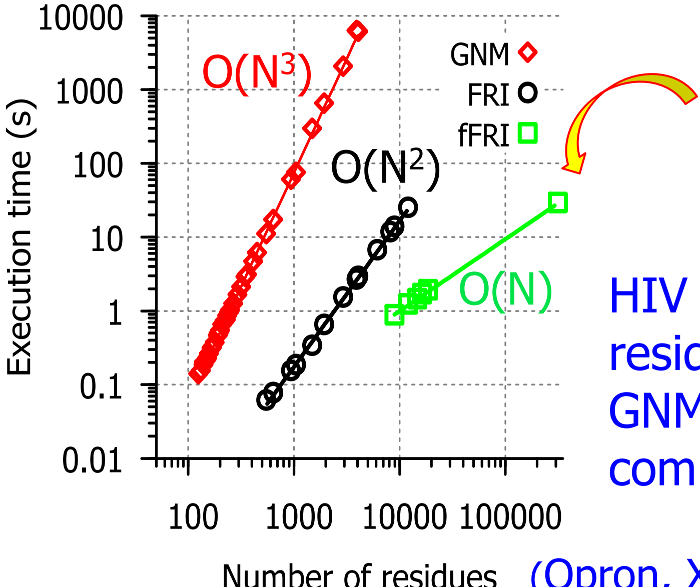
- **Multiscale weighted colored graphs (MWCG)**
- MWCG is about 40% more accurate than Gaussian network model (GNM) in B-factor prediction, based on 364 proteins.

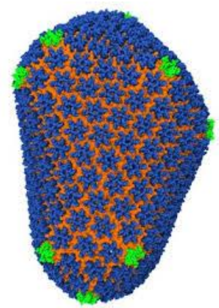
K. Opron











HIV capsid (313,236 residues) would takes GNM 120 years to compute!

$$\Gamma_{ij}(\Phi) = \begin{cases} -\Phi(r_{ij},\eta), & i \neq j, \\ -\sum_{j,j\neq i}^{N} \Gamma_{ij}, & i = j \end{cases}$$

$$\Phi(r_{ij},\eta) = 1, \quad r_{ij} \to 0$$

$$\Phi(r_{ij},\eta) = 0 \quad r_{ij} \to \infty$$

$$\Phi(r_{ij},\eta) = e^{-(r_{ij}/\eta)^{\kappa}}$$

$$B_i^{FRI} = a(\Gamma_{ii}(\Phi))^{-1}$$

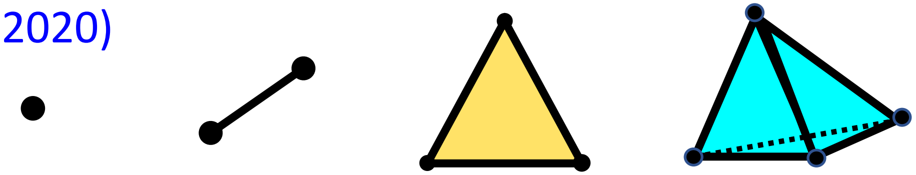
(Opron, Xia and Wei, JCP, 2013; JCP 2014; JCP, 2015; Nguyen, et al, JCIM, 2017, Bramer and Wei, JCP, 2018. Nguyen and Wei, 2018)

Persistent Spectral Graph (Persistent Laplacian)

(Wang, Nguyen, Wei, IJNMBE, 2020)

Rui Wang

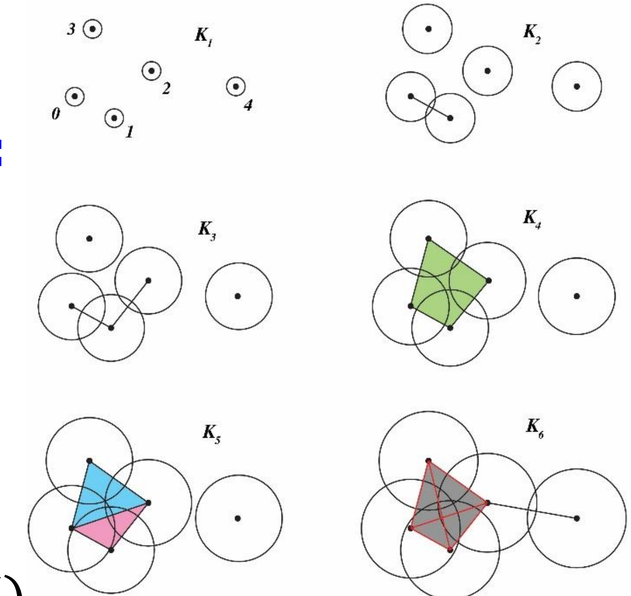
- Simplexes (σ^q):



0-simplex *1*-simplex *2*-simplex *3*-simplex

- K-chain: $K = \left\{ \sum_j w_j \sigma_j^q \right\}$

Software: HERMES



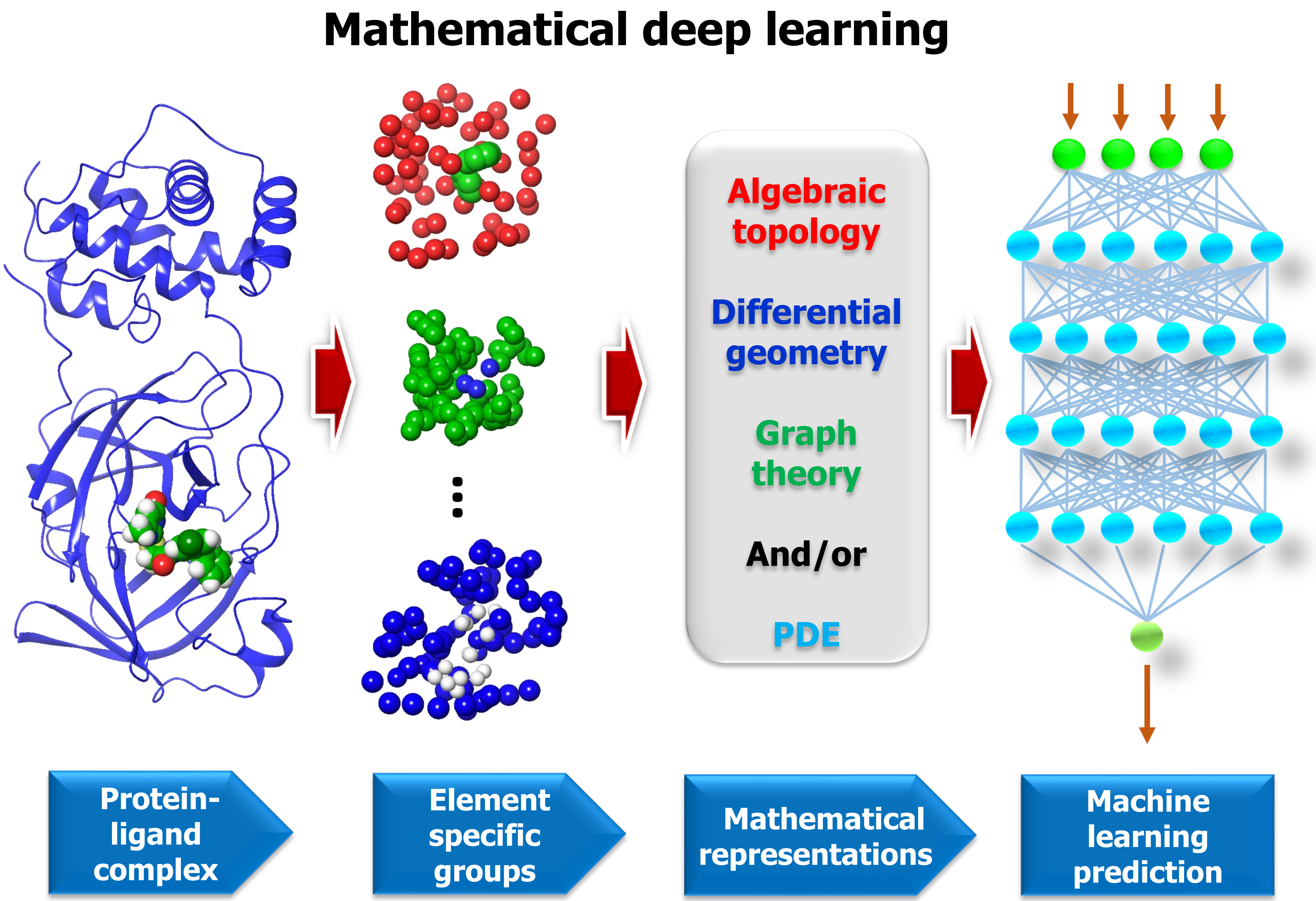
- Chain group: $C_q(K, \mathbb{Z}_2)$
- Boundary operator: $\partial_q : C_q(K) \rightarrow C_{q-1}(K)$

$$\partial_q \sigma^q = \sum_{j=0}^{q} (-1)^j \{v_0, v_1, \dots, \widehat{v_j}, \dots, v_q\}$$

- Adjoint boundary operator: $\partial_q^* : C_{q-1}(K) \rightarrow C_q(K)$
- q-combinatorial Laplacian operator: $\Delta_q = \partial_{q+1}\partial_{q+1}^* + \partial_q^* \partial_q$
- q-combinatorial Laplacian matrix: $\mathcal{L}_q = \mathcal{B}_{q+1}\mathcal{B}_{q+1}^T + \mathcal{B}_q^T \mathcal{B}_q$
- Betti numbers: (Goldberg, Thesis, 2002; Horak, Jost, AIM, 2013; Serrano, Gomze, Arxiv, 2019,…)

Multiscale spectra & topological persistence!

$$\beta_q = \dim(\mathcal{L}_q(K)) - \text{rank}\left(\mathcal{L}_q(K)\right) = \text{\# of zeros eigenvalues of } \mathcal{L}_q(K)$$


Mathematical deep learning
Algebraic topology
Differential geometry
Graph theory
And/or
PDE
Protein-ligand complex
Element specific groups
Mathematical representations
Machine learning prediction

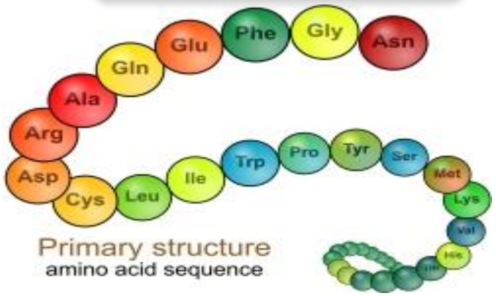
Drug Design Data Resource (D3R) Grand Challenges

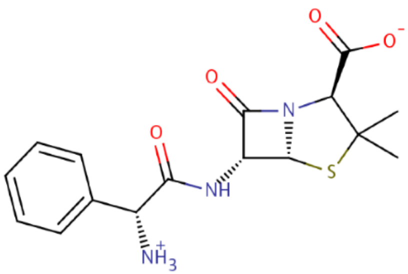
- Funded in part by National Institute of General Medical Sciences
- Hosted at the University of California, San Diego
- Annually since 2015

GC4 (2018 -2019): 55 research groups

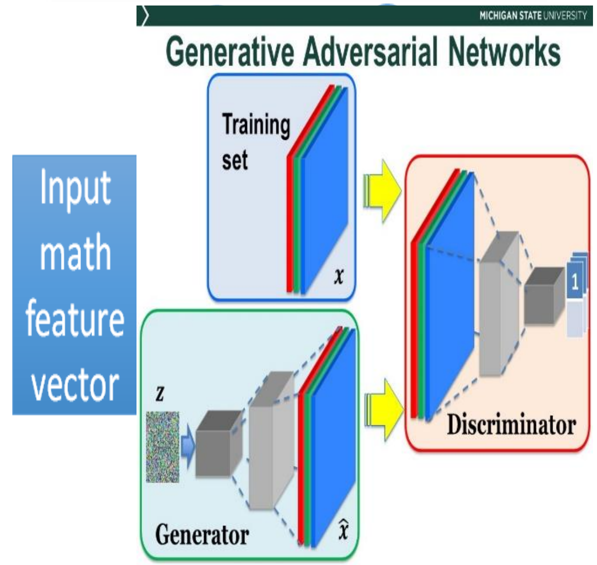
Drug Design Data Resource (D3R) Grand Challenge

Given data

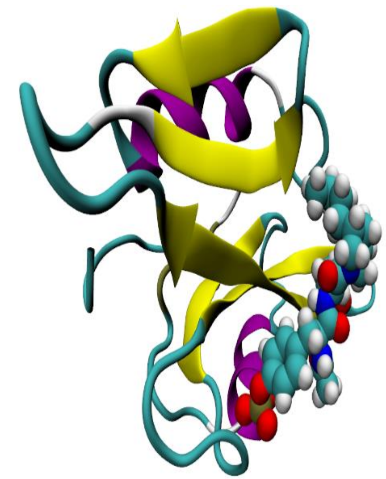


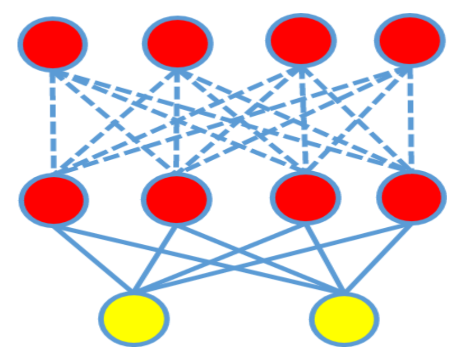


Math based GAN

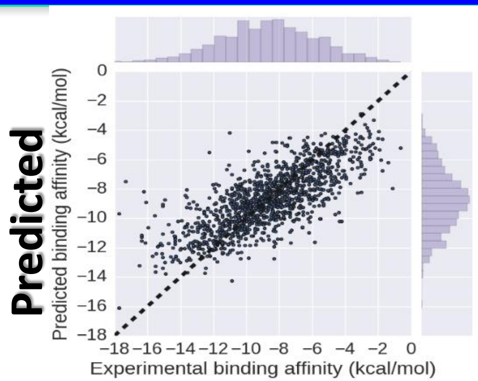


Predicted complex

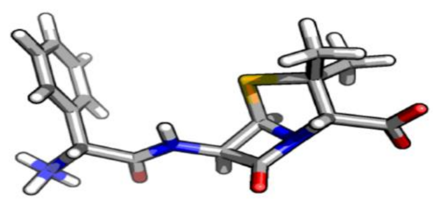




Final predictions to be compared with experiments



Experimental



Drug pose

(Nguyen et al, JCAMD, 2018)

Our performance in D3R Grand Challenges, 2016-2019.

D3R Grand Challenge 4 (2018-2019)

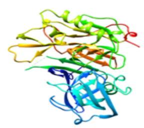
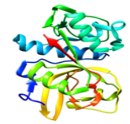
Pose Predictions

BACE Stage 1A
Pose Predictions (Partials)

BACE Stage 1B
Pose Prediction (Partials)

Affinity Predictions

Cathepsin Stage 1
Combined Ligand and Structure Based Scoring
Ligand Based Scoring (No participation)
Structure Based Scoring
Free Energy Set

BACE Stage 1
Combined Ligand and Structure (No participation)
Ligand Based Scoring(Partials) (No participation)
Structure Based Scoring(Partials)(No participation)
Free Energy Set (No participation)

BACE Stage 2
Combined Ligand and Structure
Ligand Based Scoring(No participation)
Structure Based Scoring (Partials)
Free Energy Set

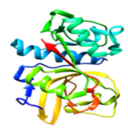
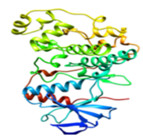
D3R Grand Challenge 3 (2017-2018)

(Nguyen et al, JCAMD, 2018)

Pose Prediction

Cathepsin Stage 1A
Pose Predictions (partials)

Cathepsin Stage 1B
Pose Prediction

Affinity Rankings excluding Kds > 10 μM

Cathepsin Stage 1
Scoring (partials)
Free Energy Set

Cathepsin Stage 2
Scoring (partials)
Free Energy Set

VEGFR2
Scoring (partials)

JAK2 SC2
Scoring (partials)

p38-α
Scoring

JAK2 SC3
Scoring
Free Energy Set

TIE2
Scoring
Free Energy Set 2

ABL1
Scoring (partials)

Active / Inactive Classification

VEGFR2
Scoring (partials)

JAK2 SC2
Scoring (partials)

p38-α
Scoring (partials)

JAK2 SC3
Scoring
Free Energy Set

TIE2
Scoring (partials)
Free Energy Set 1

ABL1
Scoring (partials)

Affinity Rankings for Cocrystalized Ligands

Cathepsin Stage 1
Scoring (partials)
Free Energy Set

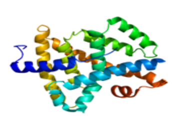
Cathepsin Stage 2
Scoring (partials)
Free Energy Set

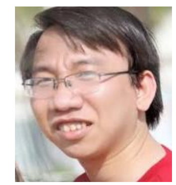
D3R Grand Challenge 2 (2016-2017)

Given: Farnesoid X receptor (FXR) and 102 ligands
Tasks: Dock 102 ligands to FXR, and predict their poses, binding free energies and energy ranking

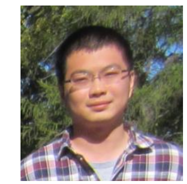
Stage 1
Pose Predictions (partials)
Scoring (partials)
Free Energy Set 1 (partials)
Free Energy Set 2 (partials)

Stage 2
Scoring (partials)
Free Energy Set 1 (partials)
Free Energy Set 2 (partials)

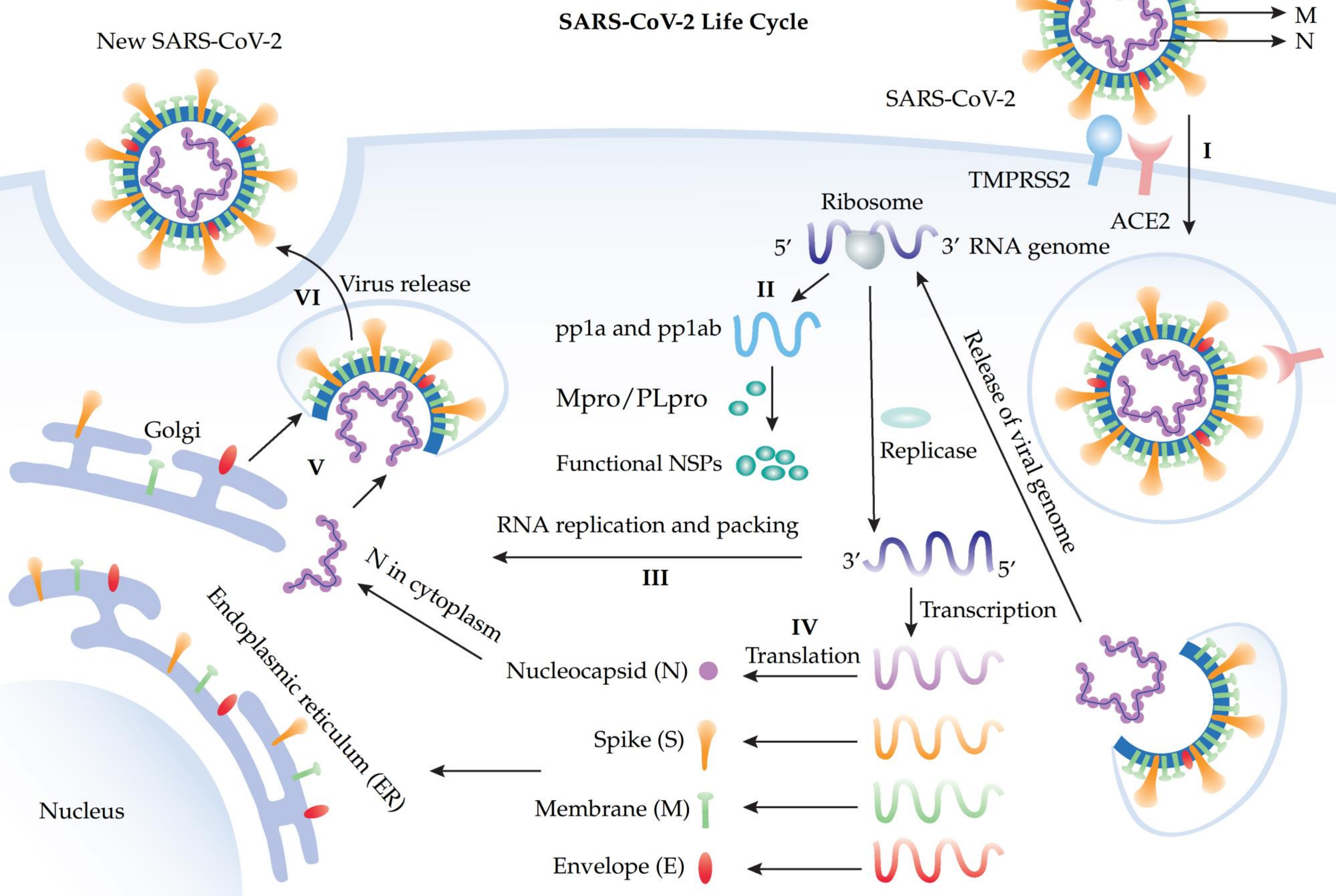



Dr. D Nguyen

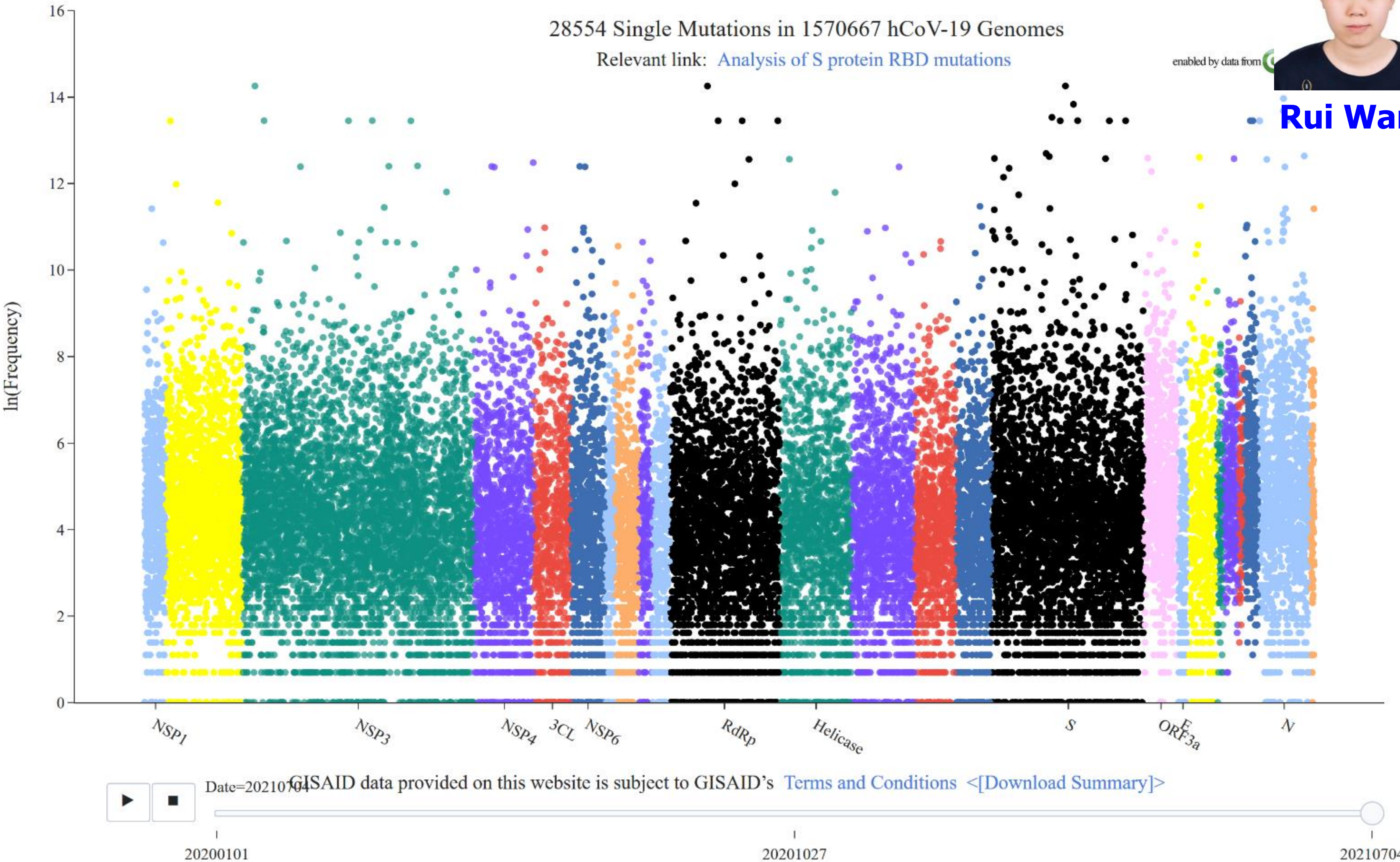

Dr. Kaifu Gao


Dr. Z Cang

Life cycle of SARS-CoV-2 in host cells



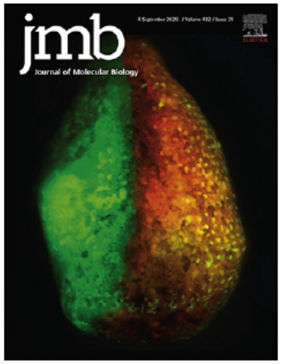
Mutation Tracker



Rui Wang

What governs SARS-CoV-2 transmission and evolution (new variants)?

https://users.math.msu.edu/users/weig/SARS-CoV-2_Mutation_Tracker.html



Mutations Strengthened SARS-CoV-2 Infectivity

We discovered the mechanism of viral transmission and evolution

Dr Jiahui Chen

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

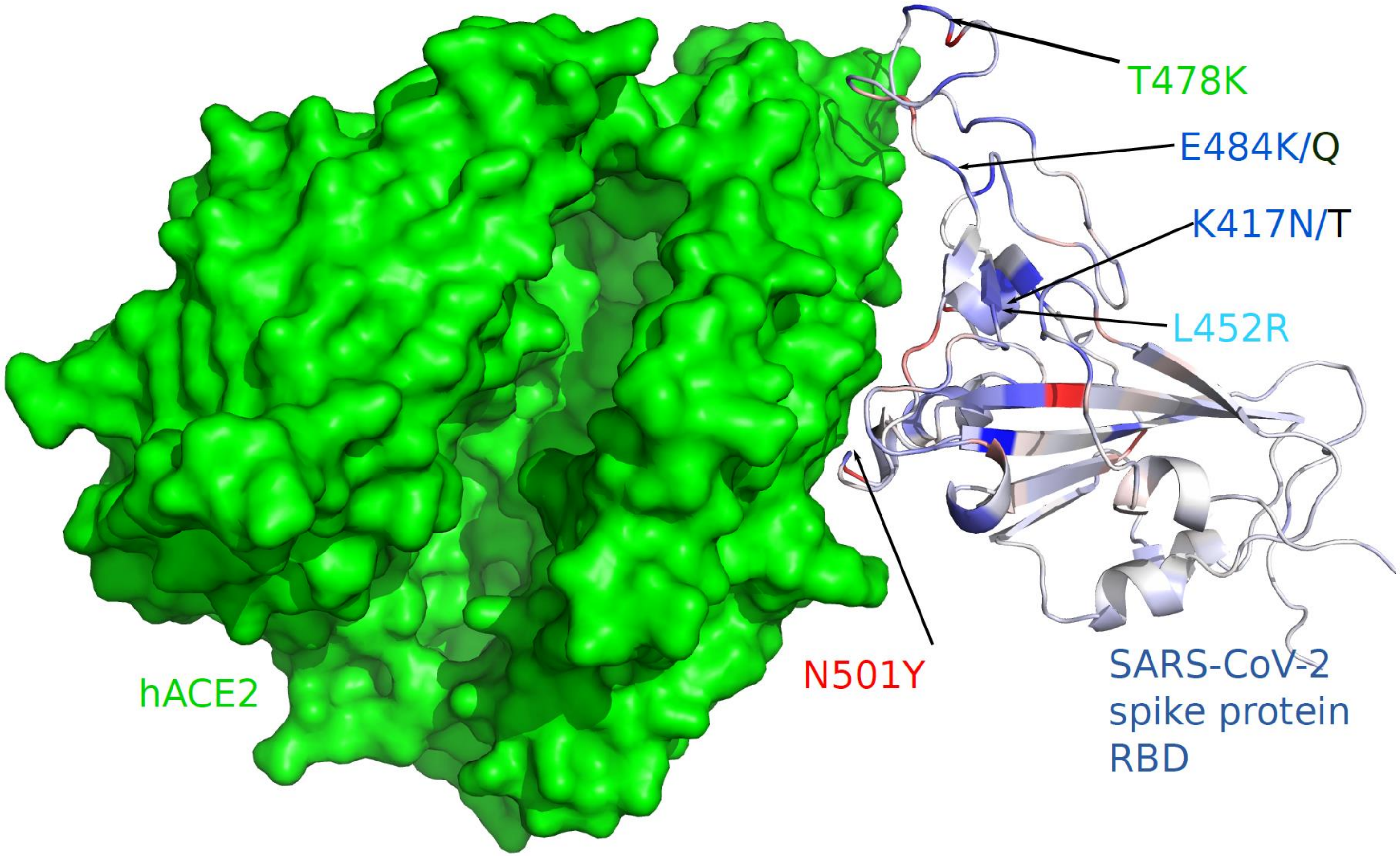
Received 4 June 2020;
Received in revised form 9 July 2020;
Accepted 17 July 2020
Available online 23 July 2020

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.

We predicted key mutation sites in prevailing variants

Mutations at 501 and 452 in prevailing SARS-CoV-2 variants



Alpha (B.1.1.7): N501Y

Beta (B.1.351):
K417N, E484K, N501Y

Gamma (P.1):
K417T, E484K, N501Y

Delta (B.1.617.2):
L452R, T478K

Epsilon (B.1.427):
L452R

Kappa (B.1.617.1):
L452R, E484Q

B.1.1.222: T478K

We discovered the mechanism of viral transmission and evolution

89) of all mutations on the RBD, which potentially increases the complexity of antiviral drug and vaccine development. This global analysis indicates that mutations on the RBD strengthen the binding of S protein and ACE2, leading to more infectious SARS-CoV-2.

We hypothesize that natural selection favors those mutations that enhance the viral transmission and if our predictions are correct, the predicted infectivity strengthening mutations will outpace predicted infectivity weakening mutations over time. Figure 3 illustrates the increase in the frequency of each

strengthening mutations occurred. It is interesting to note that overall, infectivity-strengthening mutations grow faster than infectivity-weakening mutations, which also reveals that SARS-CoV-2 subtypes having infectivity-strengthening mutations are able to infect more people. Specifically, frequencies of S477N, N439K, V483A, and V367F are higher than those of other mutations, indicating these mutations have a stronger transmission capacity.

The SARS-CoV-2 genotypes are clustered into six clusters or subtypes based on their single nucleotide

All experiments, if done correctly, confirm our hypothesis

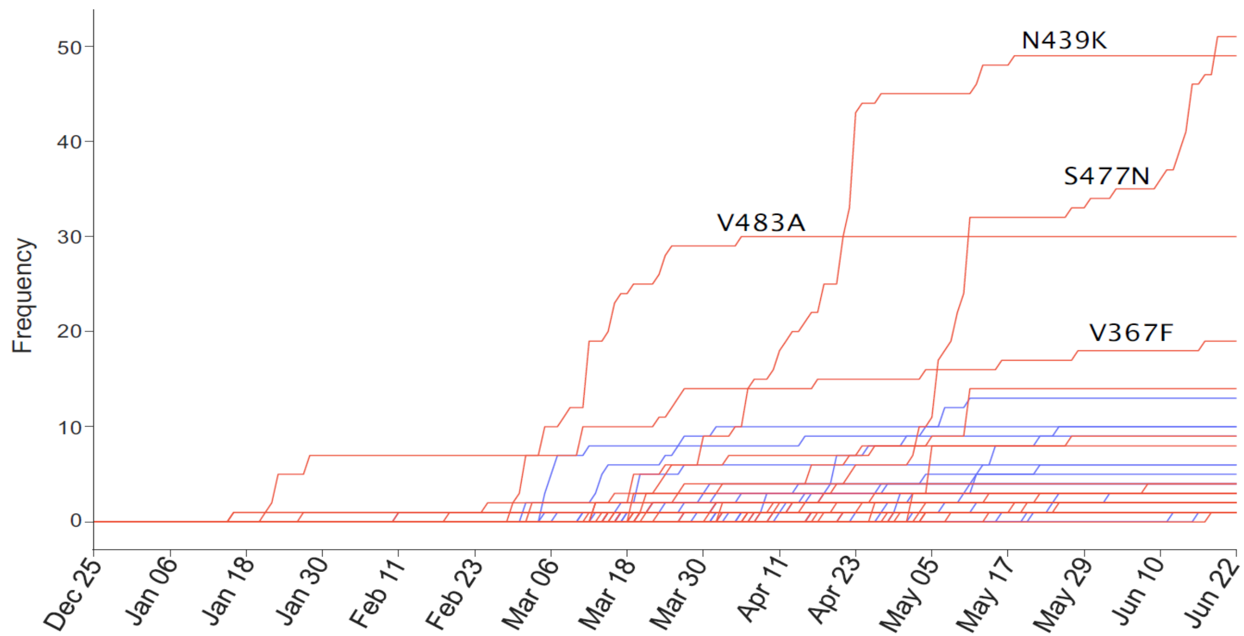
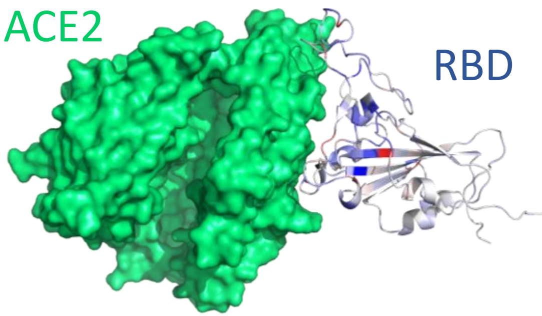


Figure 3. The time evolution of 89 SARS-CoV-2 S protein RBD mutations. The red lines represent the mutations that strengthen the infectivity of SARS-CoV-2 (i.e., $\Delta\Delta G$ is positive), and the blue lines represent the mutations that weaken the infectivity of SARS-CoV-2 (i.e., $\Delta\Delta G$ is negative). Many mutations overlap their trajectories. Here, the collection date of each genome sequence that deposited in GISAID is applied.

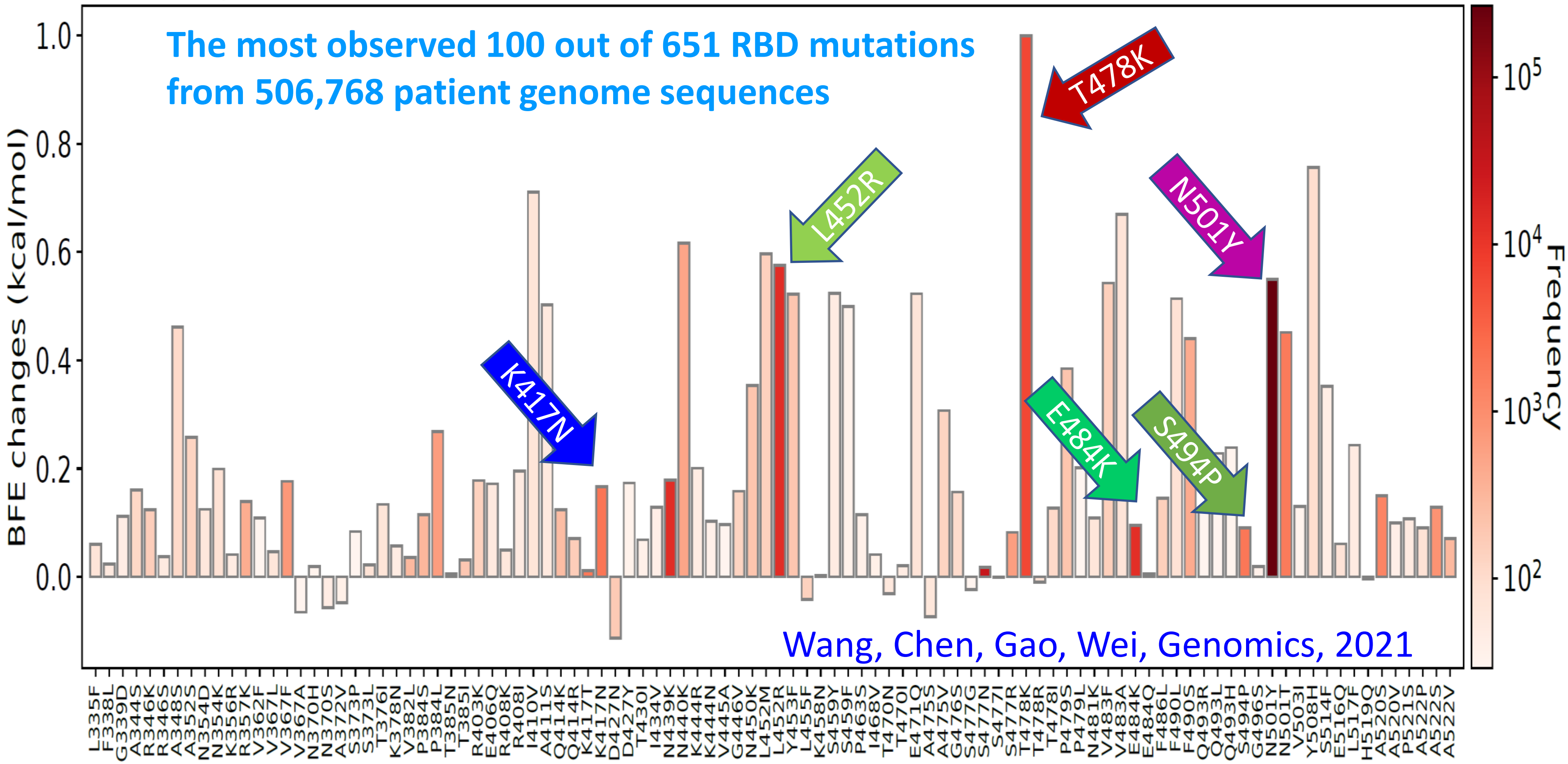
Chen, Wang, Wang, Wei, JMB, July 2020

Mutation-induced binding free energy changes for spike protein-ACE-2 complex (more infectious)

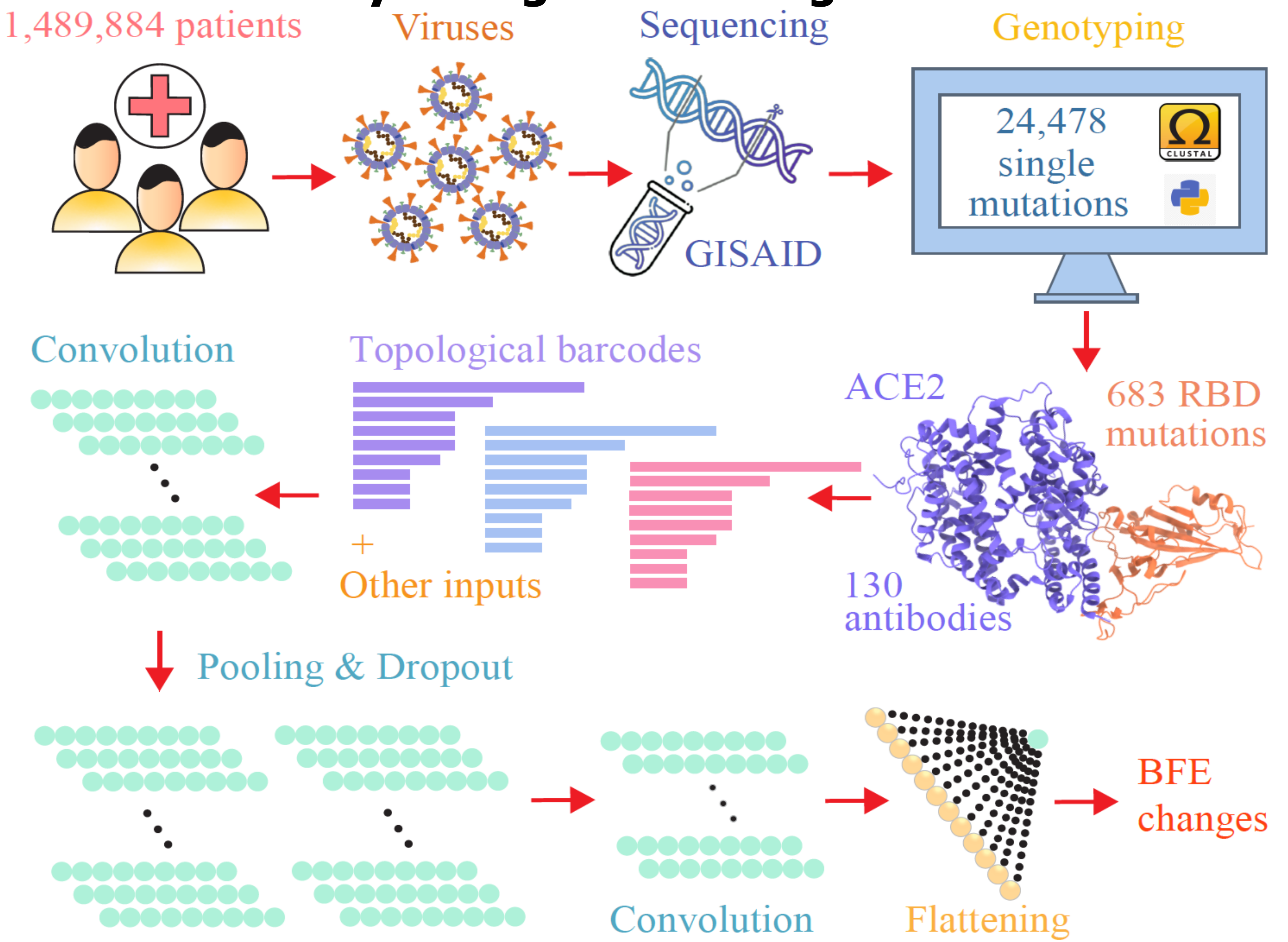


The odd for these 100 most observed mutations to be here accidentally is smaller than one chance in 1.2 nonillion!
(1.2×10^{30})

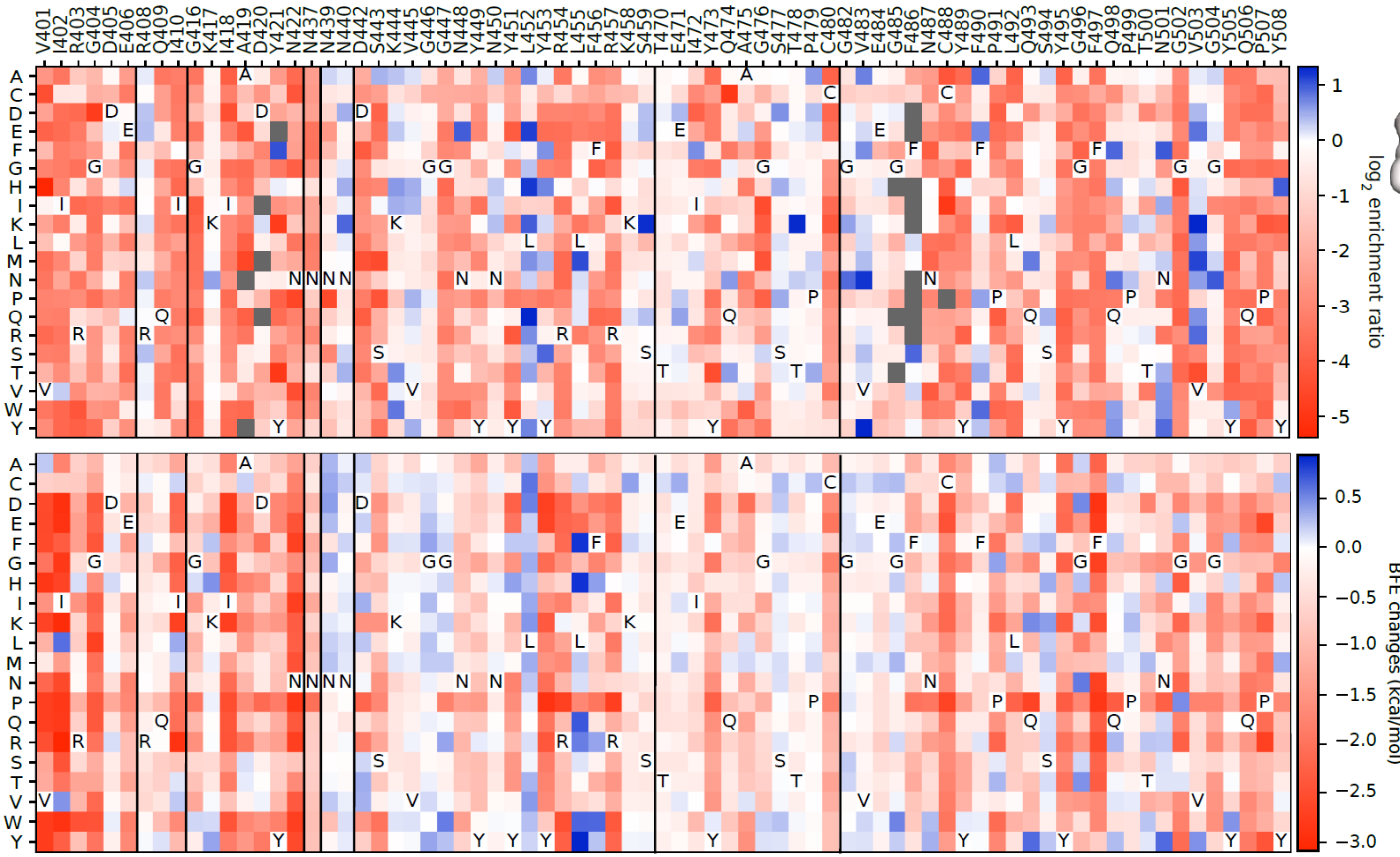
Dr Jiahui Chen

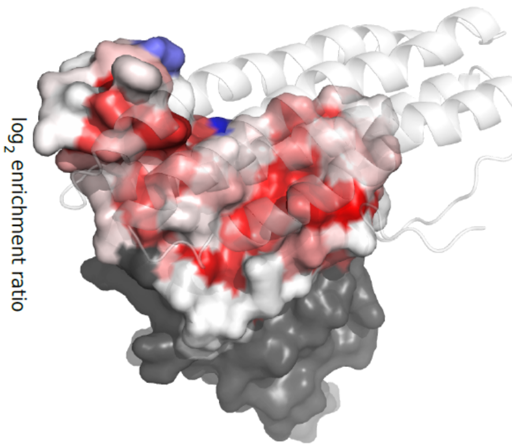


Genome-Math-AI modeling of protein-protein binding affinity changes following mutations



Predictions of deep mutations

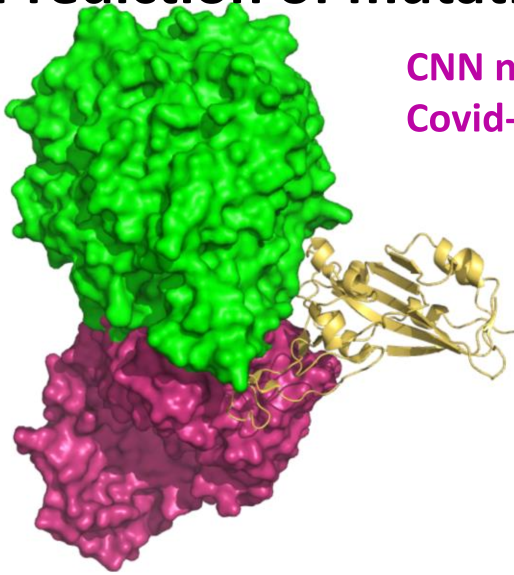




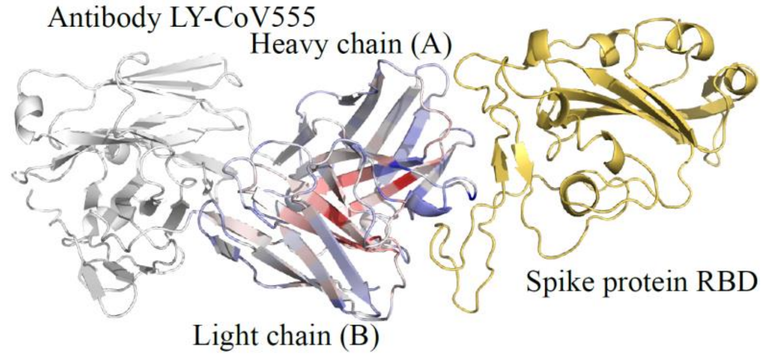
SARS-CoV-2 RBD binding to CTC-445.2

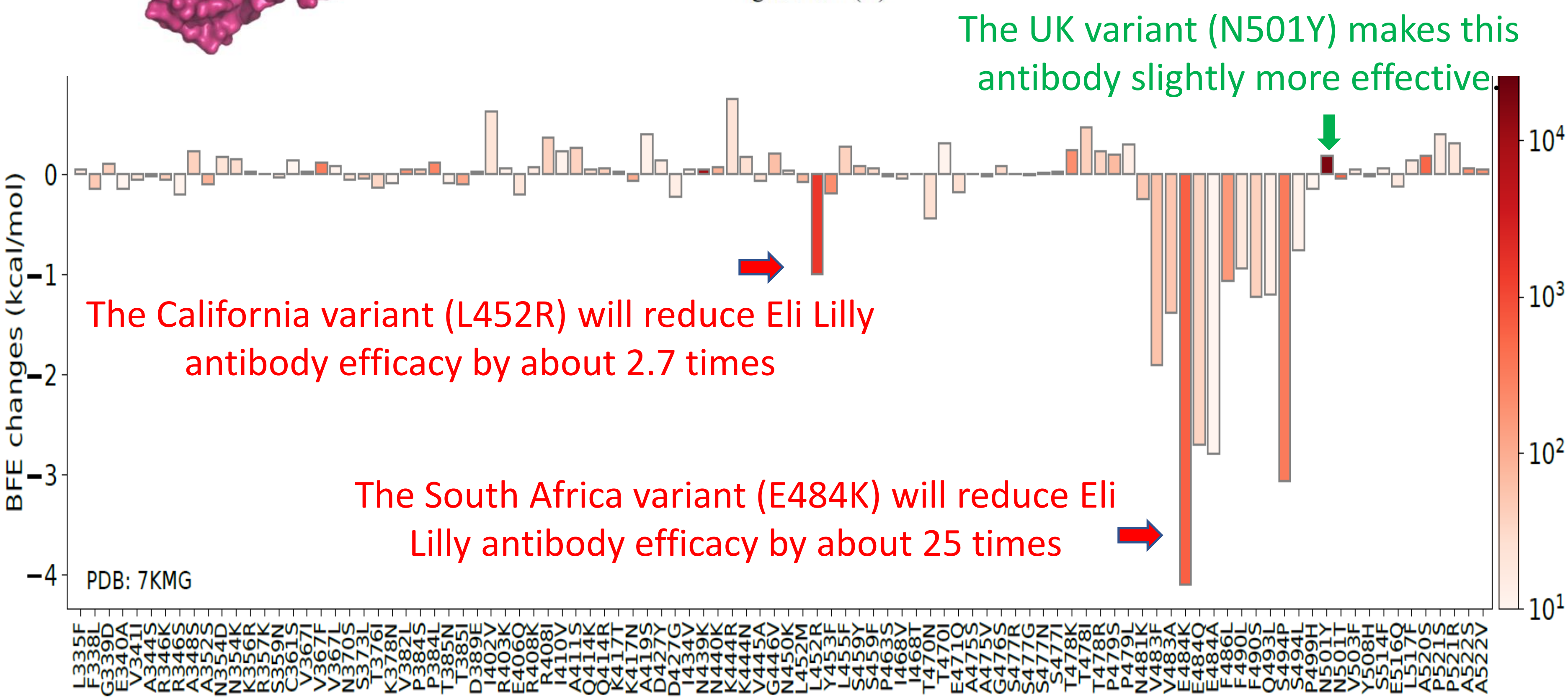
Comparison between experimental data (Top panel, Science, 370(6521): 1208, 2020) and machine learning predicted RBD-mutation-induced BFE changes (Bottom panel) for the SARS-CoV-2 S protein and CTC-445.2 complex. The high similarity between these heatmaps demonstrates the reliability of our machine learning predictions. Our model was extensive validated and trained with tens of thousands of experimental data points.

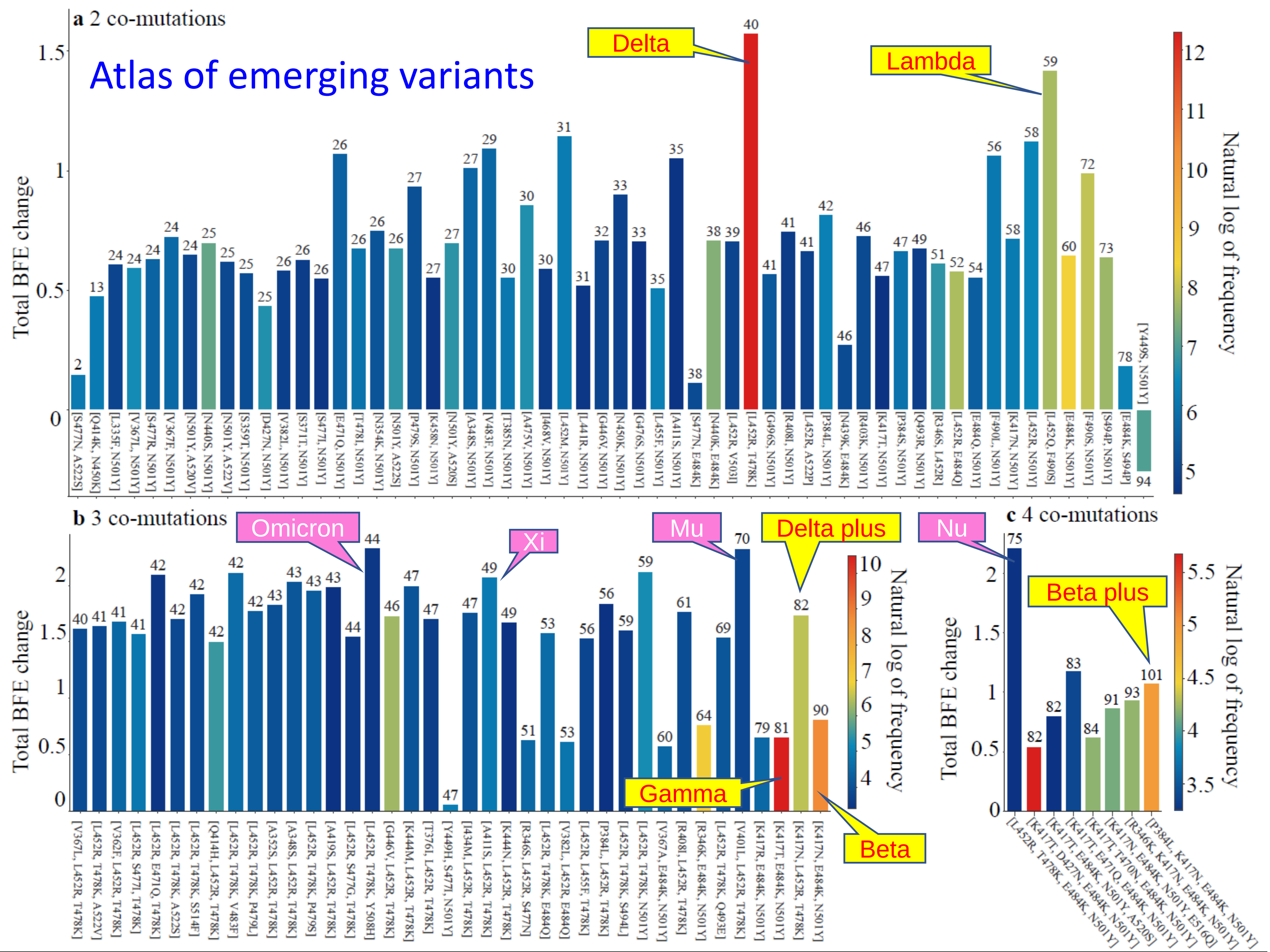
Prediction of mutational impacts to Eli Lilly antibody therapy



CNN news (March 25): US government stops distribution of Eli Lilly Covid-19 antibody treatment due to spread of coronavirus variants







Atlas of emerging variants
a 2 co-mutations
Total BFE change
Natural log of frequency
Delta
Lambda
[S477N, A522S] [Q414K, N450K] [L335F, N501Y] [V367L, N501Y] [S477R, N501Y] [V367F, N501Y] [N501Y, A520V] [N440S, N501Y] [N501Y, A522V] [S359T, N501Y] [D427N, N501Y] [V382L, N501Y] [S371T, N501Y] [S477I, N501Y] [E471Q, N501Y] [T478I, N501Y] [N354K, N501Y] [N501Y, A522S] [P479S, N501Y] [K458N, N501Y] [N501Y, A520S] [A348S, N501Y] [V483F, N501Y] [T385N, N501Y] [A475V, N501Y] [I468V, N501Y] [L452M, N501Y] [L441R, N501Y] [G446V, N501Y] [N450K, N501Y] [G476S, N501Y] [L455F, N501Y] [A411S, N501Y] [S477N, E484K] [N440K, E484K] [L452R, V503I] [L452R, T478K] [G496S, N501Y] [R408I, N501Y] [L452R, A522P] [P384L, N501Y] [N439K, E484K] [R403K, N501Y] [K417T, N501Y] [P384S, N501Y] [Q493R, N501Y] [R346S, L452R] [L452R, E484Q] [E484Q, N501Y] [F490L, N501Y] [K417N, N501Y] [L452R, N501Y] [L452Q, F490S] [E484K, N501Y] [F490S, N501Y] [S494P, N501Y] [E484K, S494P] [Y449S, N501Y]
2 13 24 24 24 24 24 25 25 25 25 26 26 26 26 26 26 26 27 27 27 27 29 30 30 30 31 31 32 33 33 35 35 38 38 39 40 41 41 41 42 46 46 47 47 49 51 52 54 56 58 58 59 60 72 73 78 94
b 3 co-mutations
Total BFE change
Natural log of frequency
Omicron
Xi
Mu
Delta plus
Gamma
Beta
[V367L, L452R, T478K] [L452R, T478K, A522V] [V362F, L452R, T478K] [L452R, S477I, T478K] [L452R, E471Q, T478K] [L452R, T478K, A522S] [L452R, T478K, S514F] [Q414H, L452R, T478K] [L452R, T478K, V483F] [L452R, T478K, P479L] [A352S, L452R, T478K] [A348S, L452R, T478K] [L452R, T478K, P479S] [A419S, L452R, T478K] [L452R, S477G, T478K] [L452R, T478K, Y508H] [G446V, L452R, T478K] [K444M, L452R, T478K] [T376I, L452R, T478K] [Y449H, S477I, N501Y] [I434M, L452R, T478K] [A411S, L452R, T478K] [K444N, L452R, T478K] [R346S, L452R, S477N] [L452R, T478K, E484Q] [V382L, L452R, E484Q] [L452R, L455F, T478K] [P384L, L452R, T478K] [L452R, T478K, S494L] [L452R, T478K, N501Y] [V367A, E484K, N501Y] [R408I, L452R, T478K] [R346K, E484K, N501Y] [L452R, T478K, Q493E] [V401L, L452R, T478K] [K417R, E484K, N501Y] [K417T, E484K, N501Y] [K417N, L452R, T478K] [K417N, E484K, N501Y]
c 4 co-mutations
Total BFE change
Natural log of frequency
Nu
Beta plus
[L452R, T478K, E484K, N501Y] [K417T, D427N, E484K, N501Y] [K417T, E484K, N501Y, A520S] [K417T, E471Q, E484K, N501Y] [K417T, T470N, E484K, N501Y] [K417N, E484K, N501Y, E516Q] [R346K, K417N, E484K, N501Y] [P384L, K417N, E484K, N501Y]
75 82 82 83 84 91 93 101

Algebraic graph

Geometric topology

Differential topology

Algebraic topology

Algebra

Geometric graph

Statistics

Probability

Differential equation

The last frontier of science is biology.

The last frontier of biology is mathematics

Number theory

Algebraic geometry

Differential geometry

Symplectic geometry

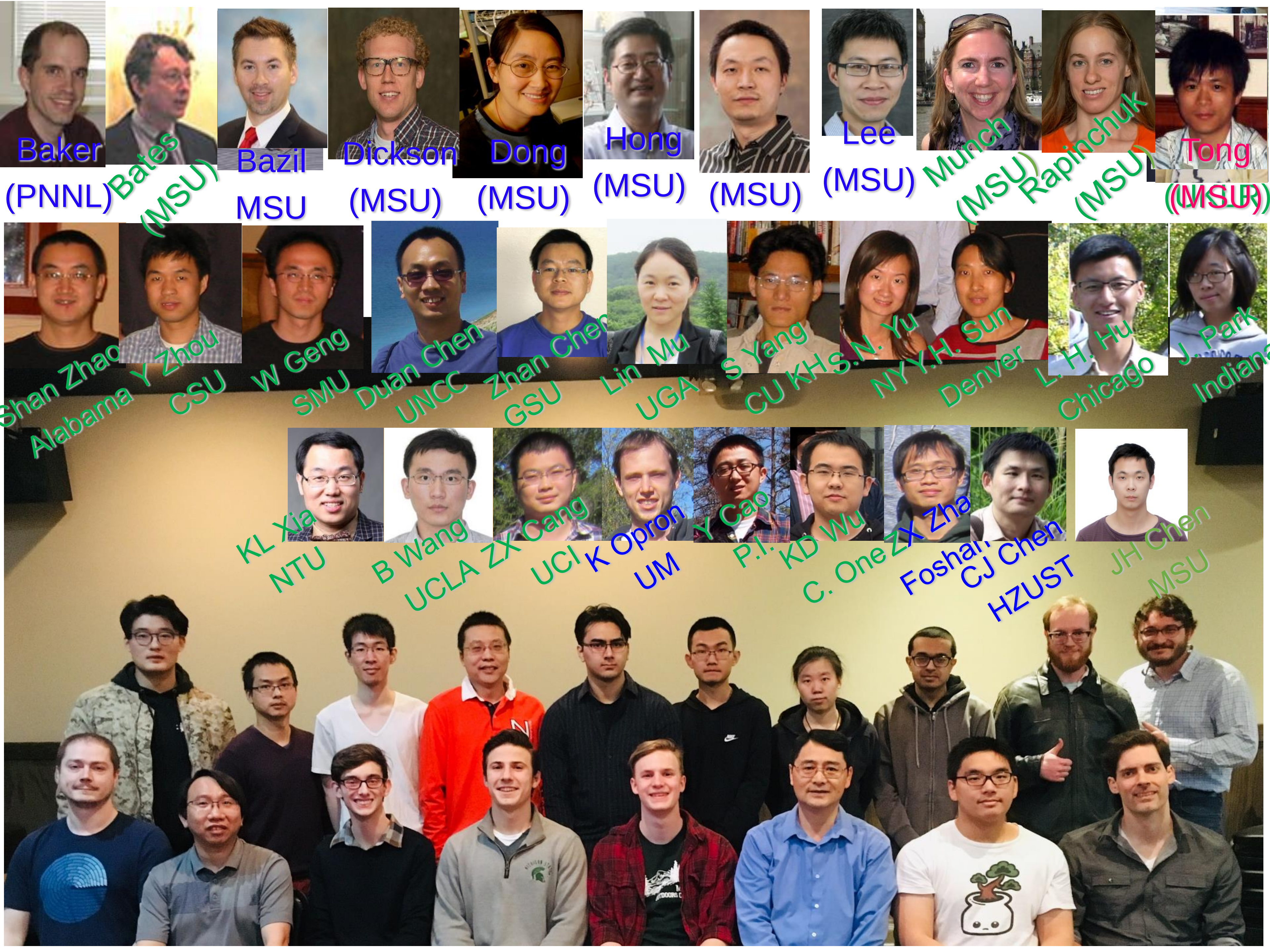
Multiscale analysis

Complex analysis

Harmonic analysis

Real analysis

Stochastic analysis


Baker
(PNNL)
Bates
(MSU)
Bazil
MSU
Dickson
(MSU)
Dong
(MSU)
Hong
(MSU)
(MSU)
Lee
(MSU)
Munch
(MSU)
Rapinchuk
(MSU)
Tong
(MSU)
Shan Zhao
Alabama
Y Zhou
CSU
W Geng
SMU
Duan Chen
UNCC
Zhan Chen
GSU
Lin Mu
UGA
S Yang
CU
KHS.N. Yu
NY
Y.H. Sun
Denver
L. H. Hu
Chicago
J. Park
KL Xia
NTU
B Wang
UCLA
ZX Cang
UCI
K Opron
UM
Y Cao
P.I.
KD Wu
C. One
ZX Zha
CJ Chen
HZUST
JH Chen
MSU