

# ***Topology, Geometry, and AI: New Frontiers in Understanding Biological Systems***

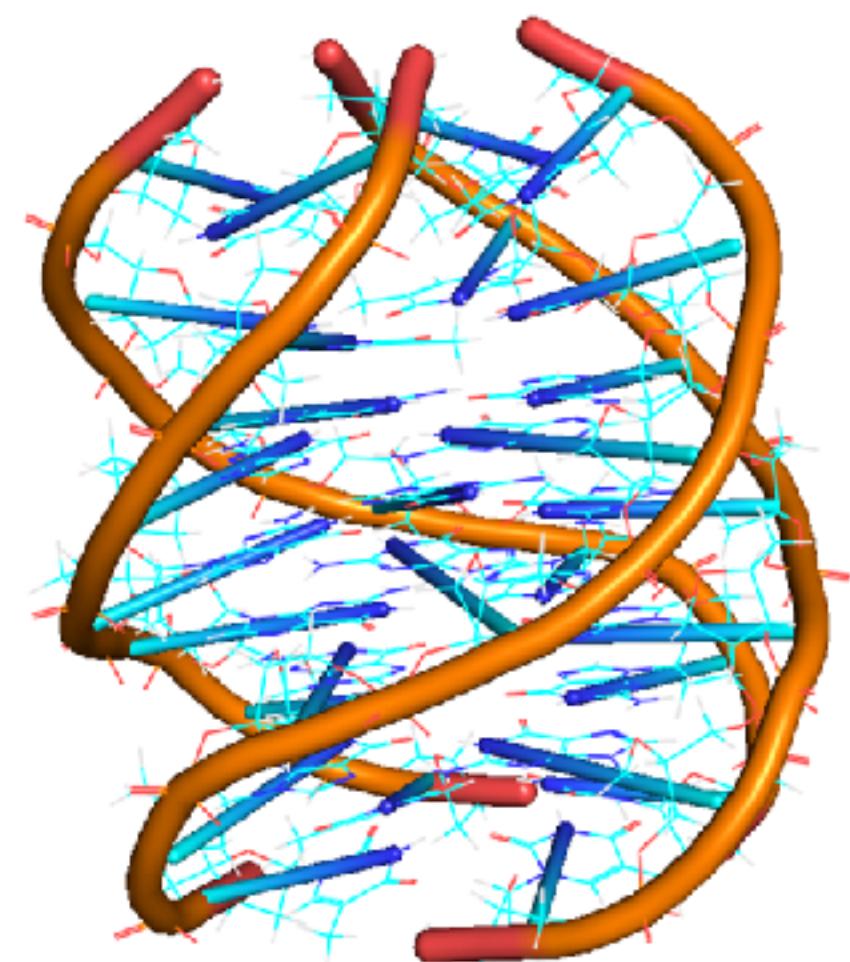
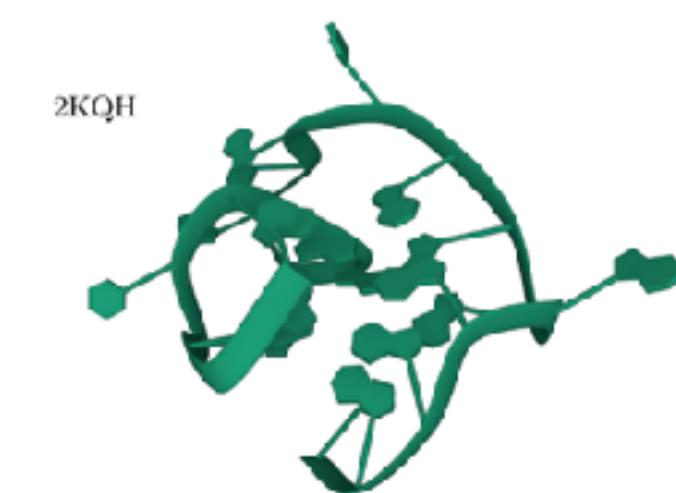
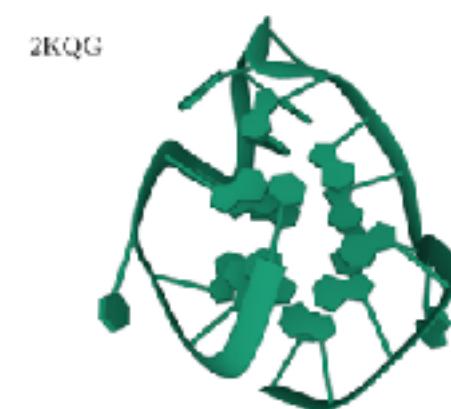
2026 NTU-SU JOINT WINTER SCHOOL AI FOR SCIENCES

Day 4: Data Science in Biomedical and other Sciences

NTU, Singapore

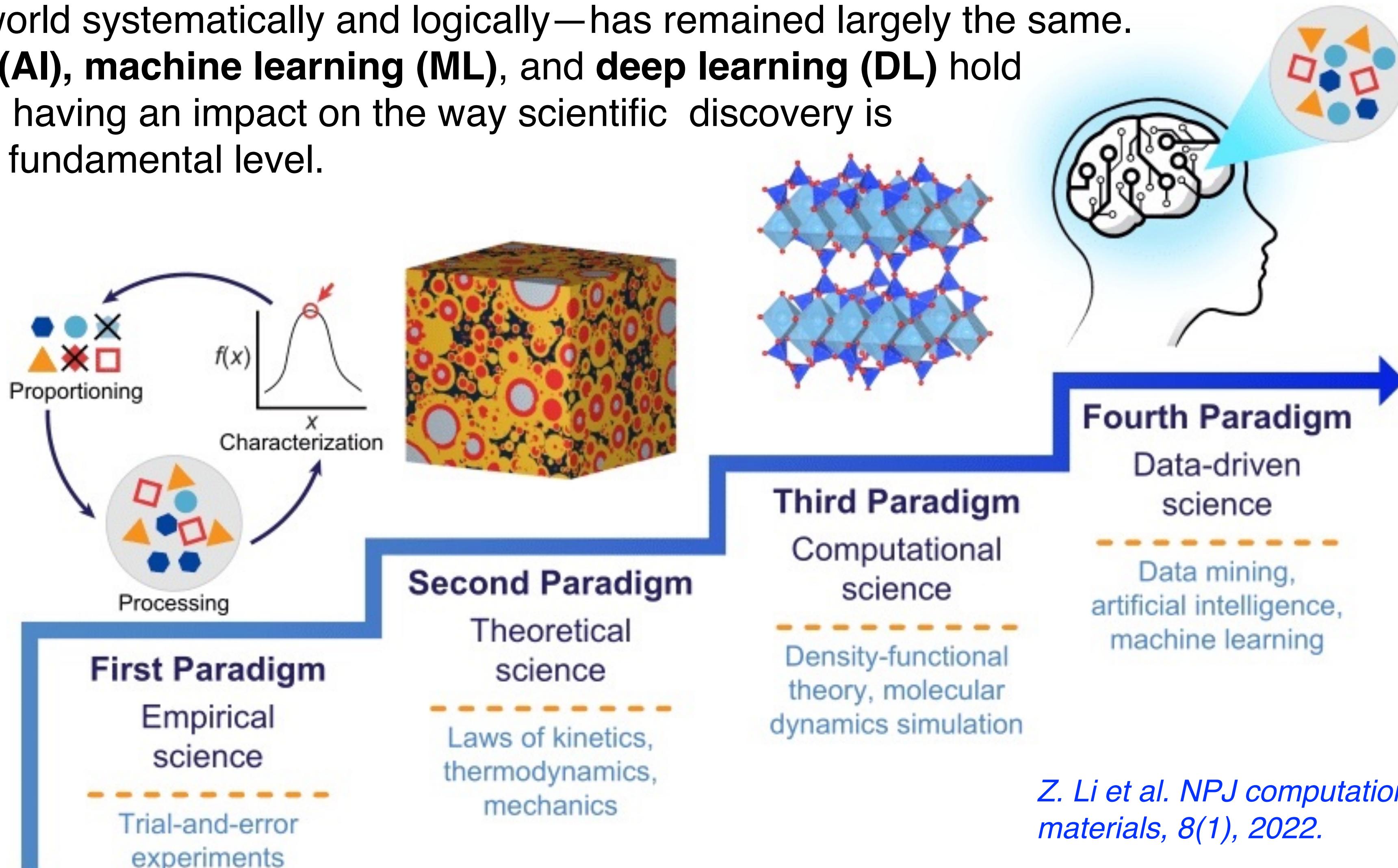
29th Jan 2026

Asst. Prof. Yong Ee Hou  
Division of Physics and Applied Physics, SPMS  
Nanyang Technological University

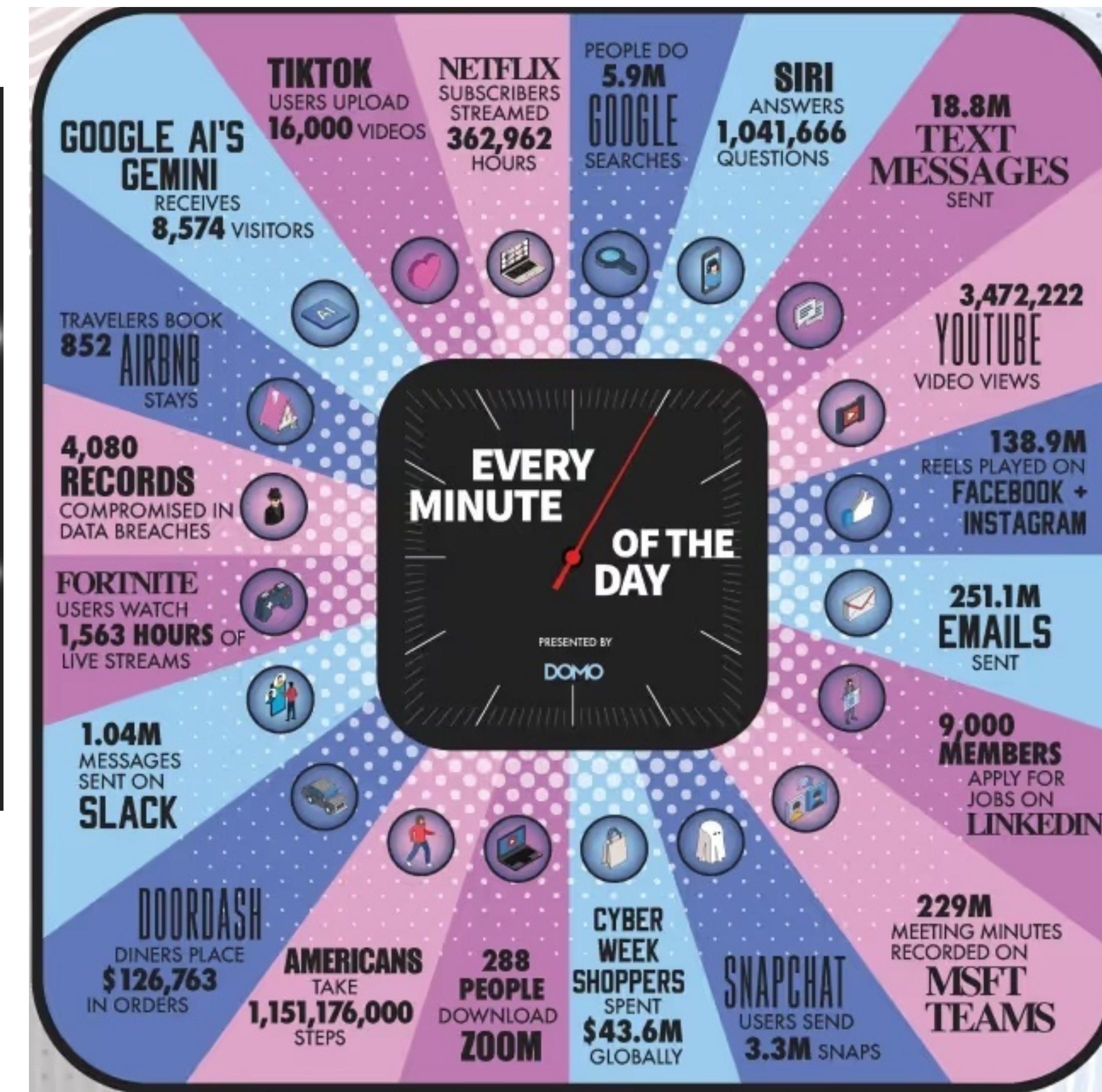
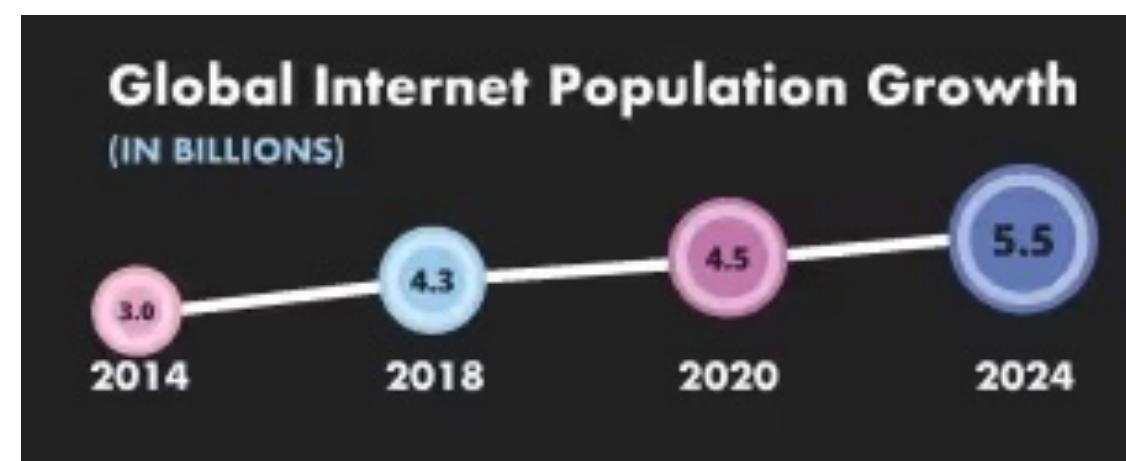
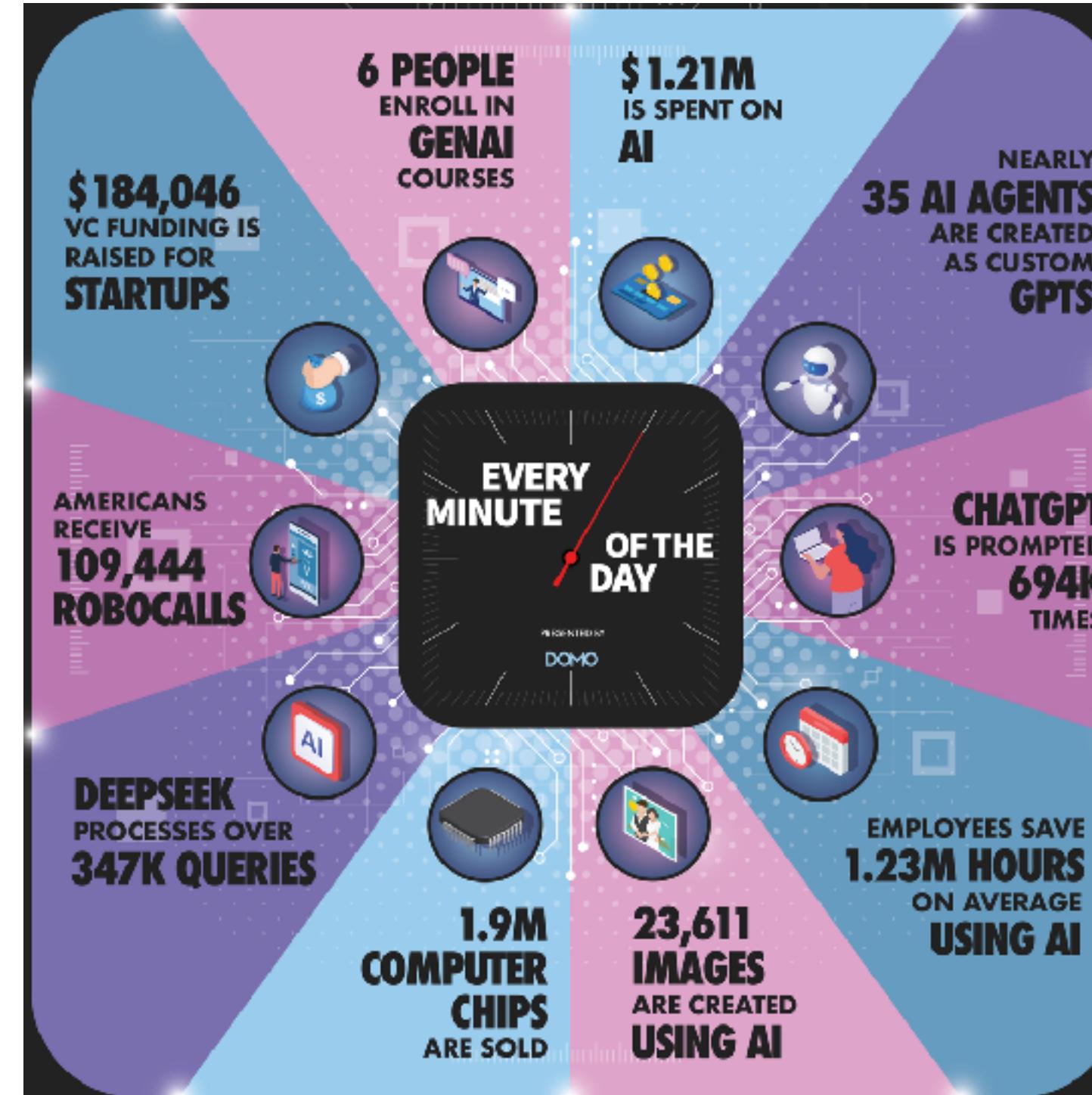


# AI for Science

- For centuries, the method of **discovery**— the fundamental practice of science that scientists use to explain the natural world systematically and logically—has remained largely the same.
- Artificial intelligence (AI), machine learning (ML), and deep learning (DL)** hold tremendous promise in having an impact on the way scientific discovery is performed today at the fundamental level.



# Data is the new oil of modern digital economy...



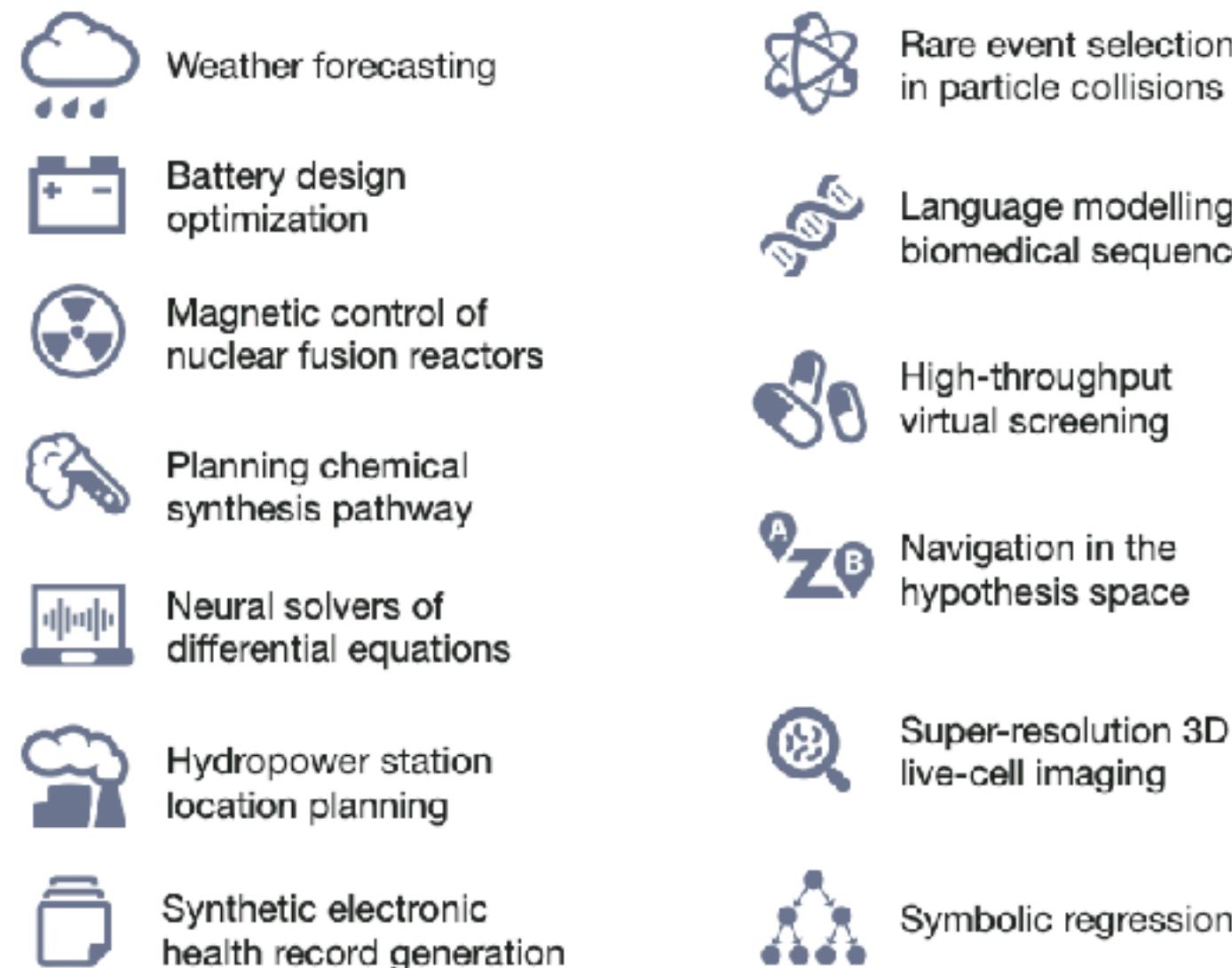
Figs from Domo.com

Asst Prof YONG Ee Hou

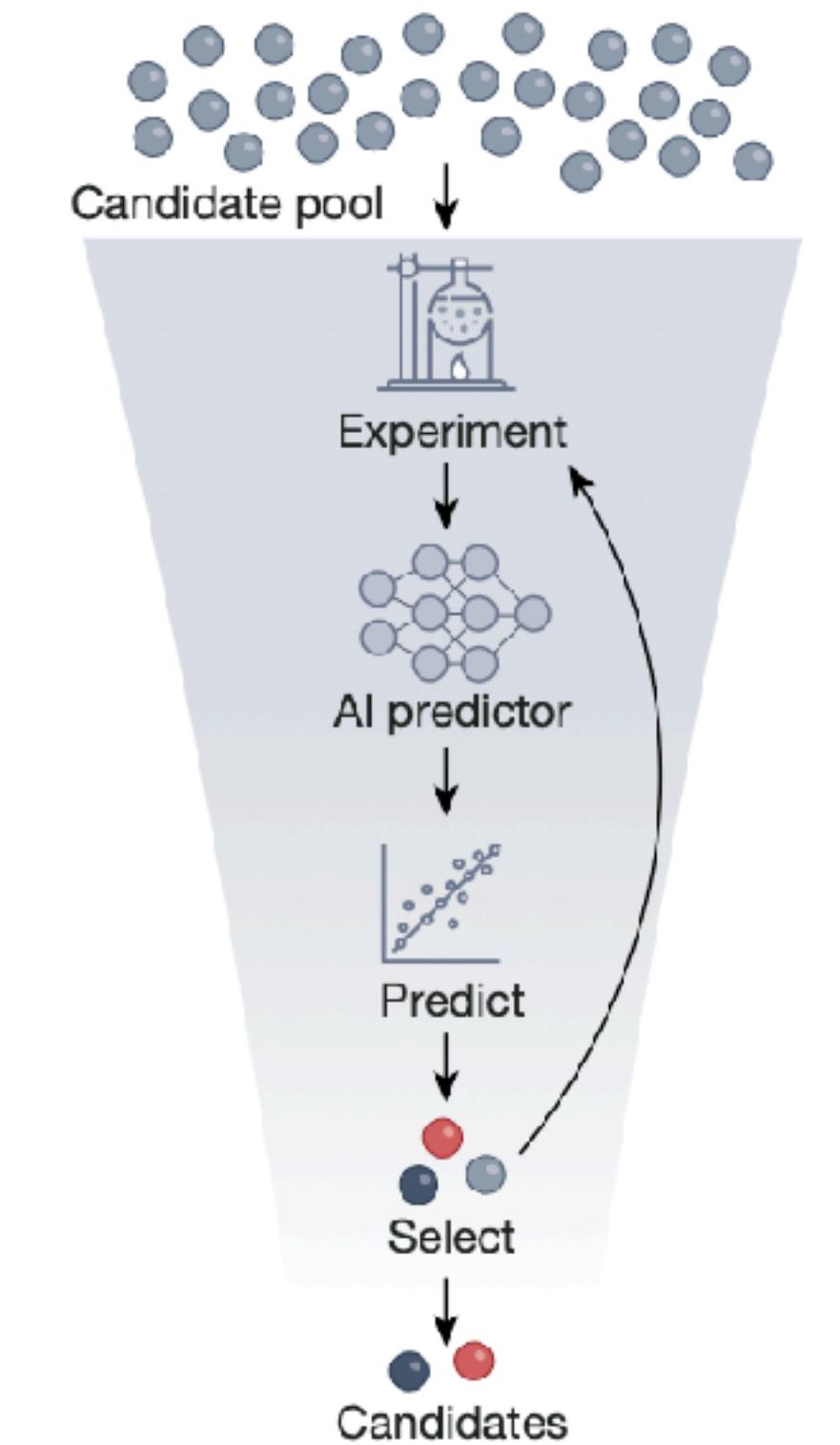
Yong Group: Theoretical Biological & Statistical Physics

# Scientific Discovery in the Age of Artificial Intelligence

Scientific discovery is a **multifaceted** process that involves several interconnected stages, including hypothesis formation, experimental design, data collection and analysis.



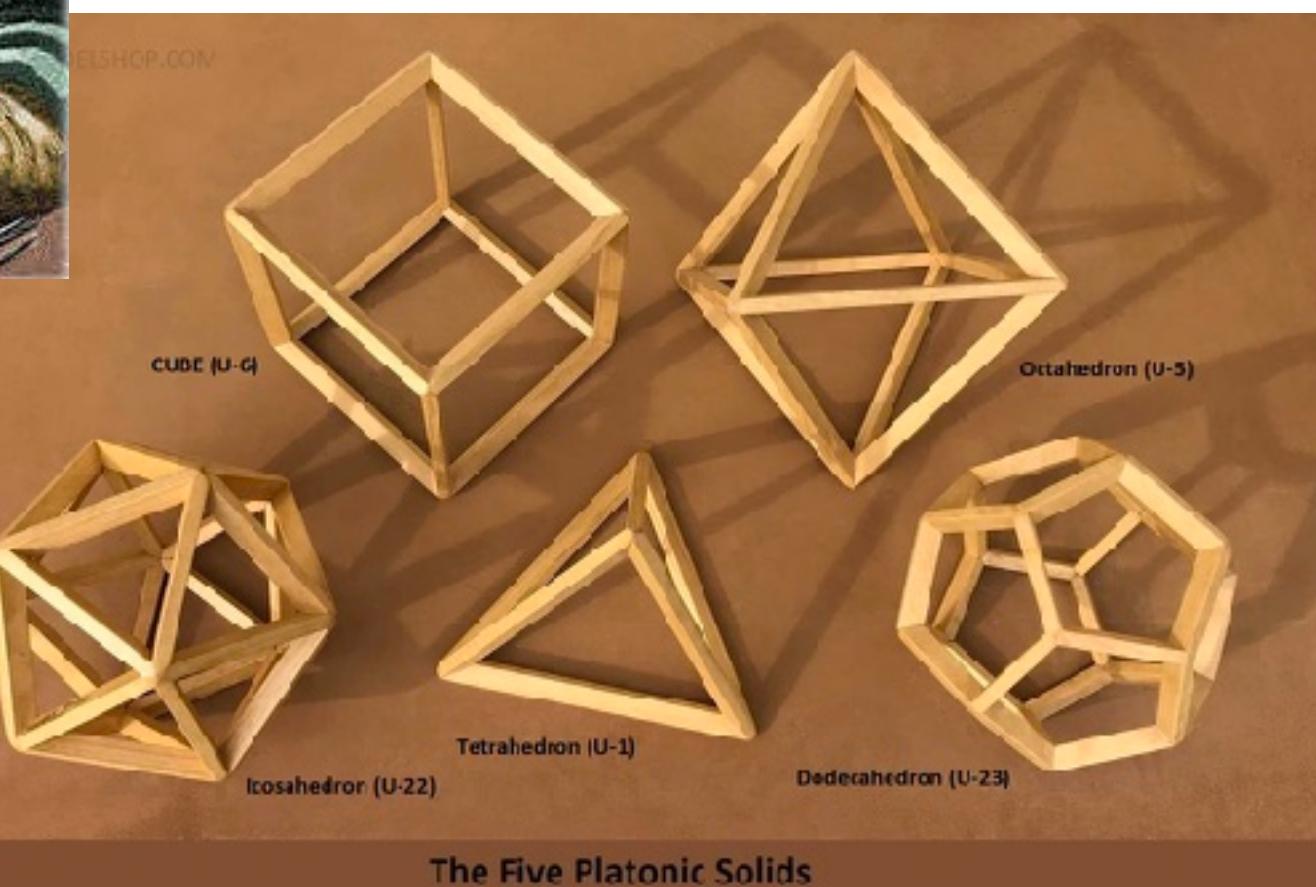
AI is poised to **reshape** scientific discovery by augmenting and accelerating research at each stage of this process.



H. Wang, et al., *Nature* 620, 47–60 (2023).

# Development of Geometry and Topology

Geometry and topology developed to explain forms in nature...



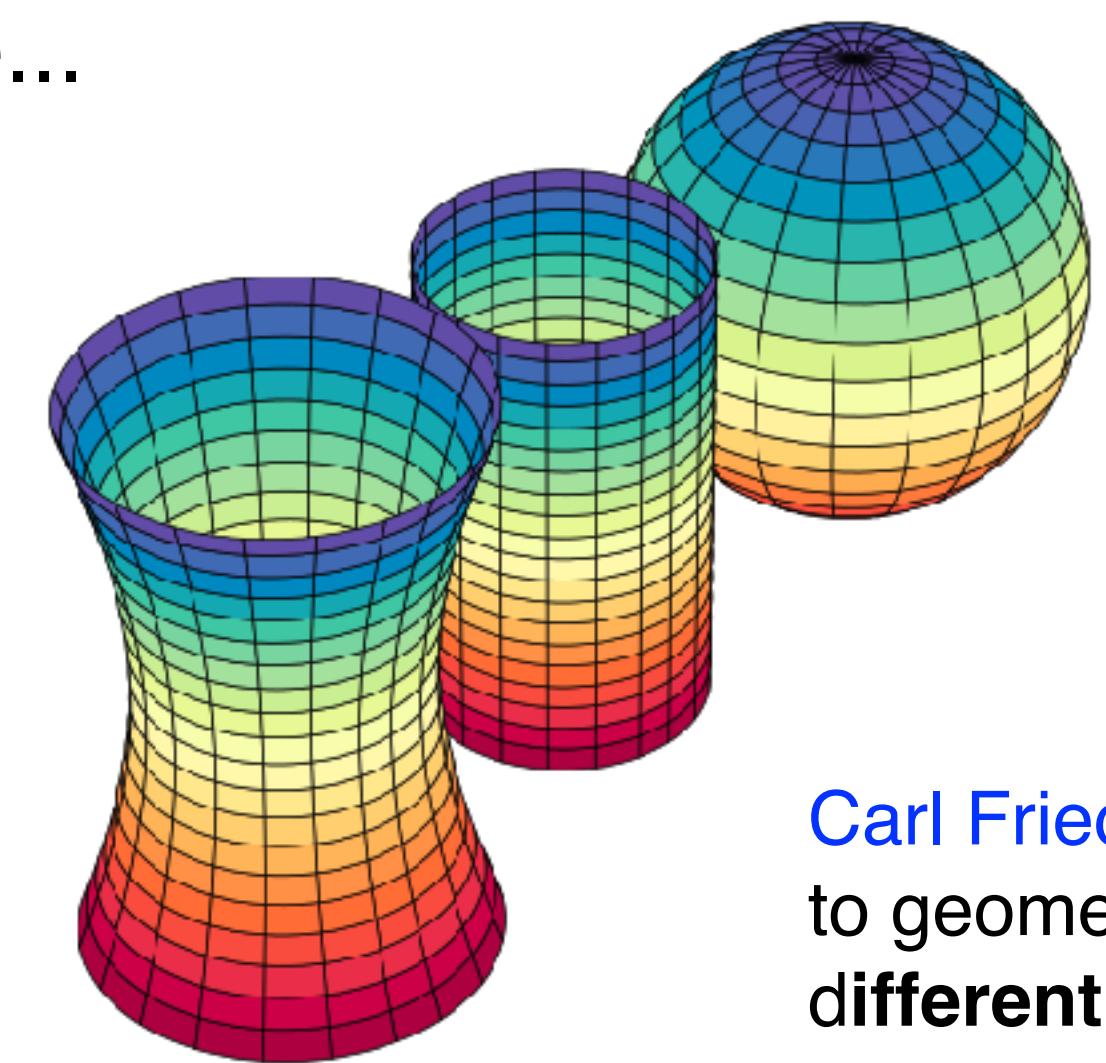
Euclid (~300 BC) was an ancient Greek mathematician, often called the "father of **geometry**"

Leonhard Euler

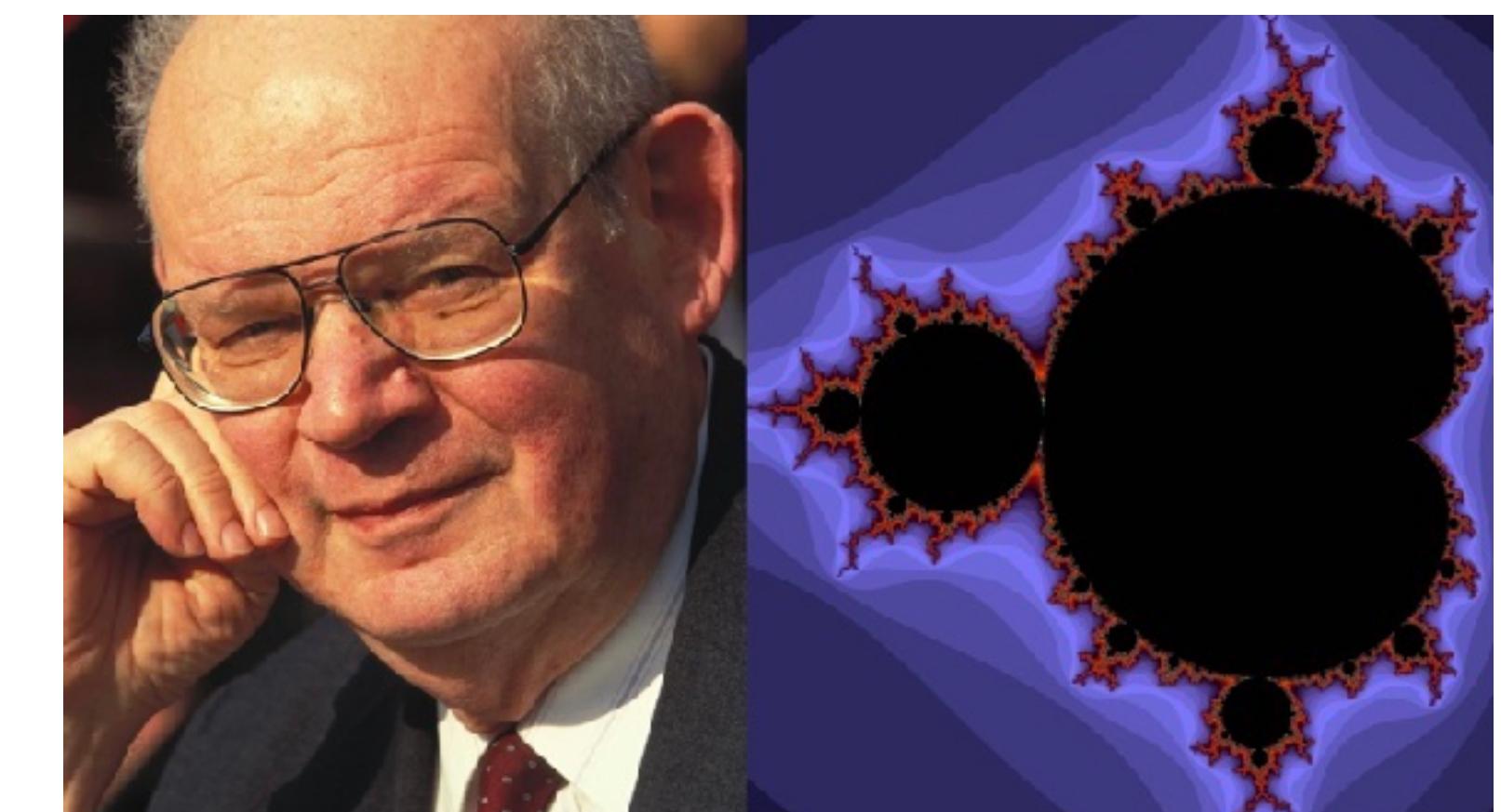


Solved the Königsberg bridge problem (1736) and developed the polyhedron formula ( $V-E+F=2$ ), both of which are considered the earliest applications of **topology**

Images taken online



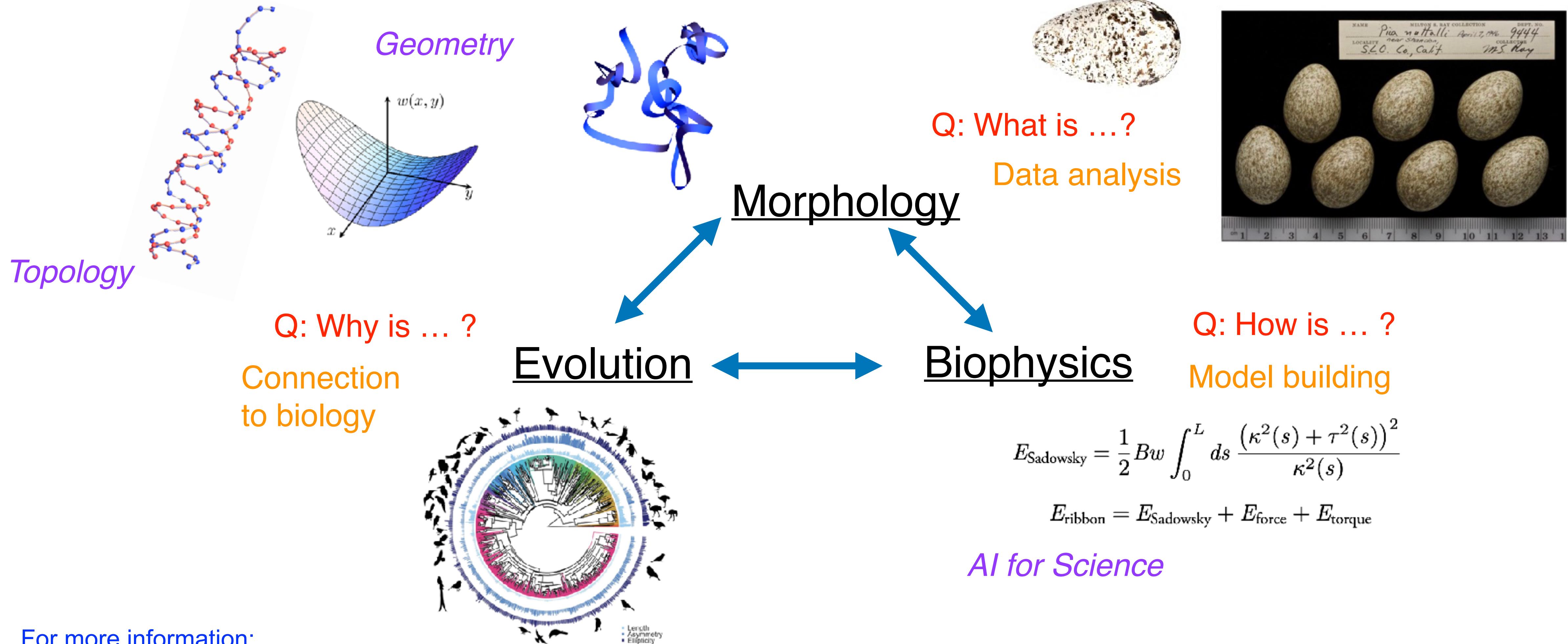
Carl Friedrich Gauss apply calculus to geometry, creating the field of **differential geometry**.



Benoit Mandelbrot is the father of **fractal geometry**, which is the study of irregular, self-similar, and complex patterns.

# Understanding Shape and Form in Nature

Aim: How can we use geometry, topology and AI for scientific discovery?



For more information:  
<https://personal.ntu.edu.sg/eehou>

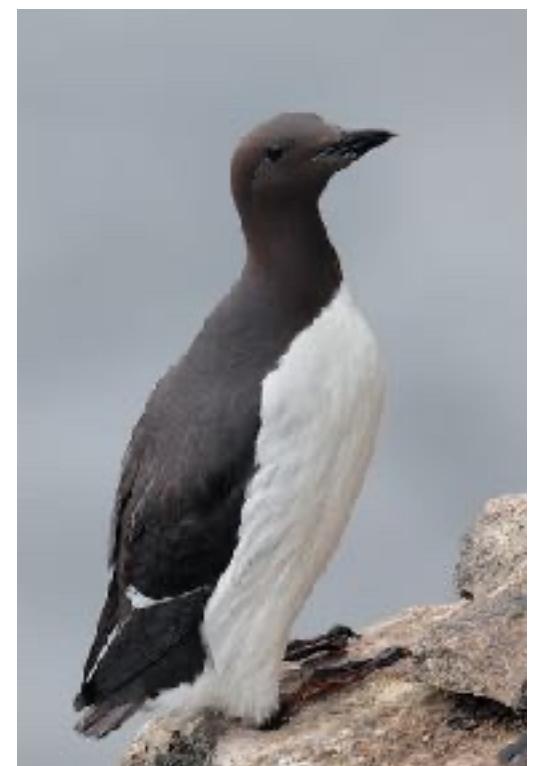
Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

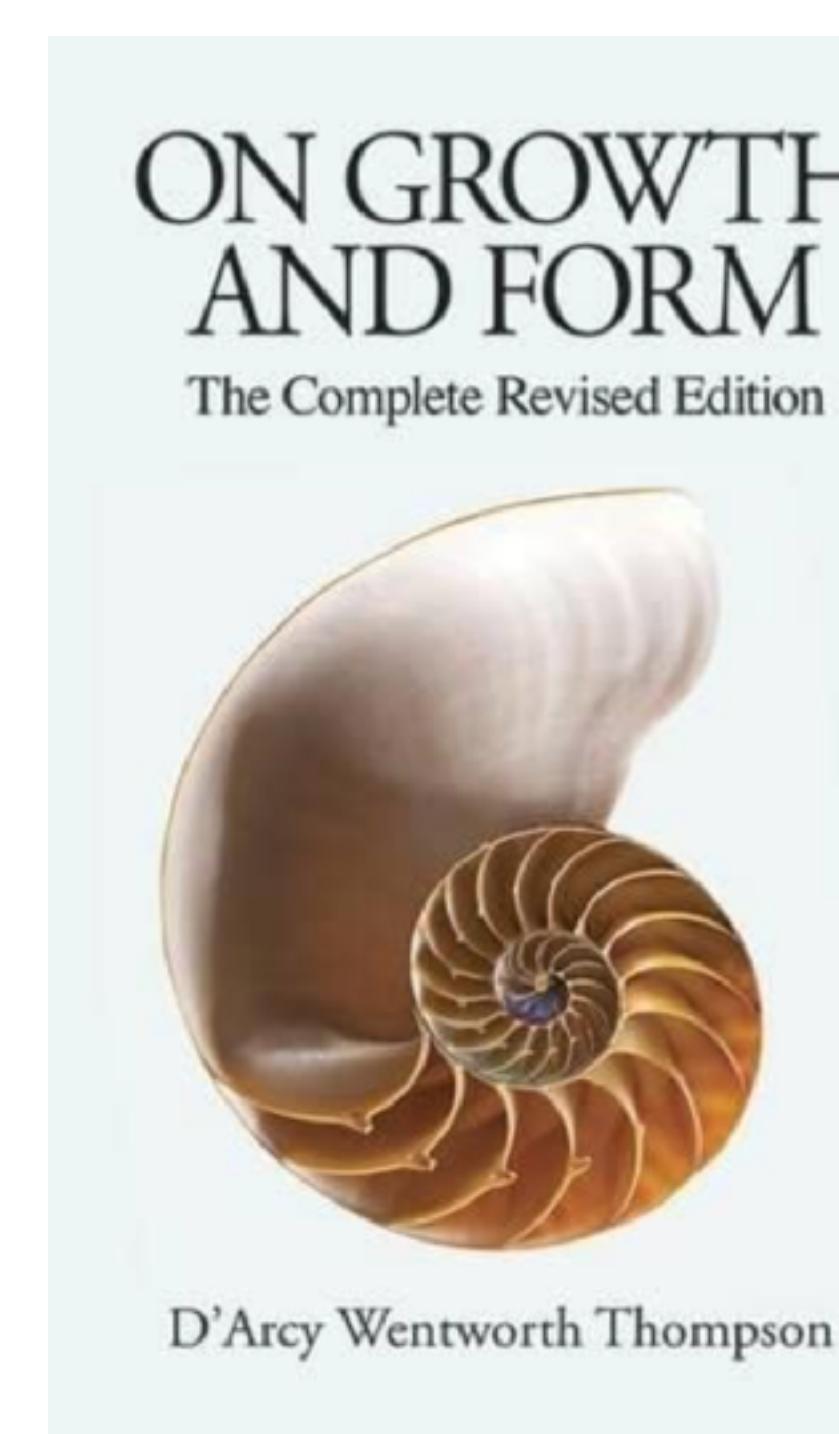
# Geometry of Avian Eggs

The phrase “egg-shaped” evokes the traditional asymmetric form of an avian egg, in fact birds lay eggs with striking **variation in shape**.

1. *What is the shape of an egg?*
2. *How is the egg shaped?*
3. *Why are there so many shapes for bird egg?*



Common Murre  
(Passeriformes,  
i.e., songbirds)



Brown Hawk-owl  
(Strigiformes)

## CH. XV] ON THE SHAPES OF EGGS

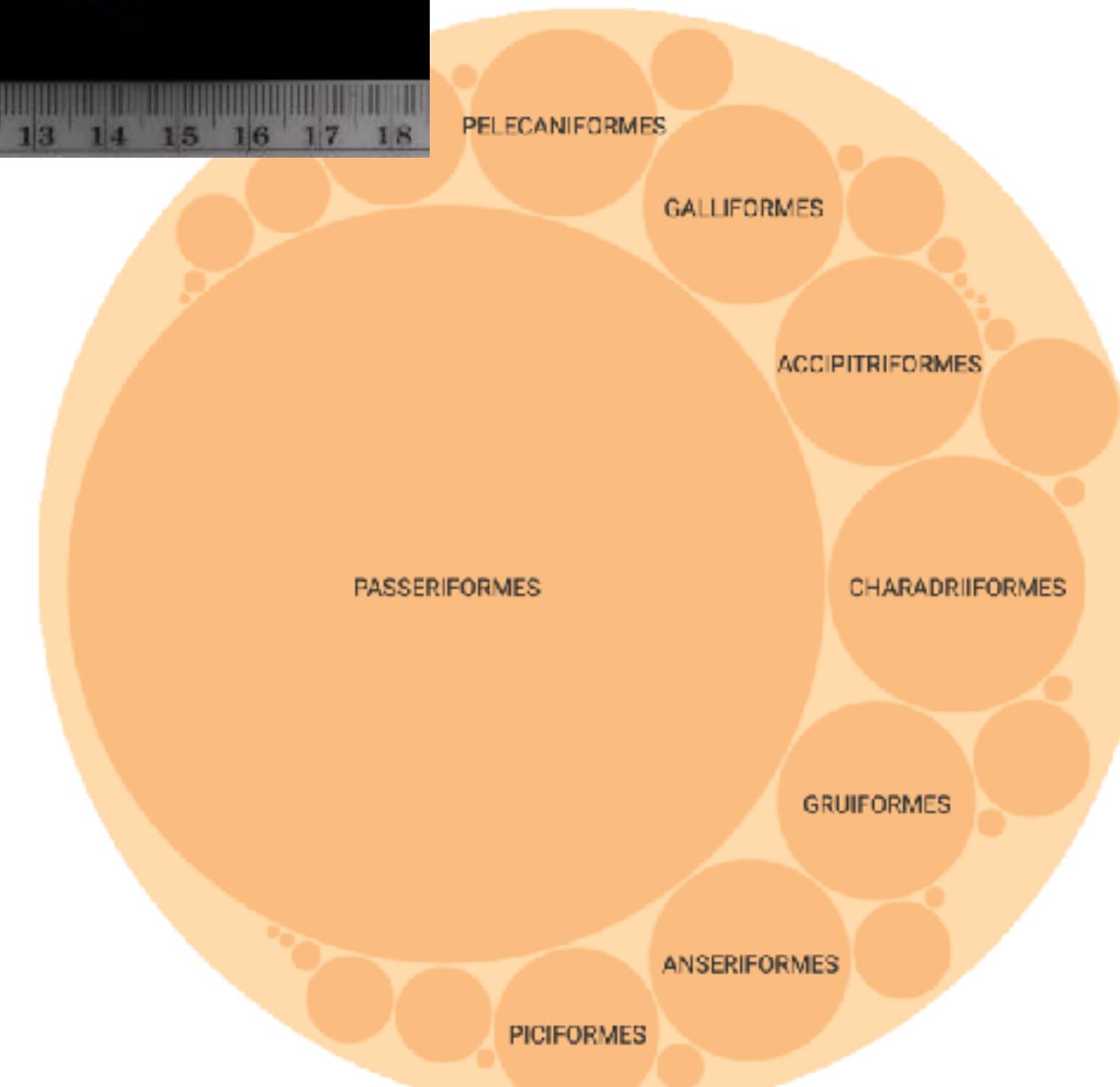
935

who first studied the “flowers” of the coral, and who wrote the *Histoire physique de la mer*; and the specific form as well as the colour and other attributes of the egg have been again and again discussed, and not least by the many dilettanti naturalists of the eighteenth century who soon followed in Marsigli’s footsteps\*.

We need do no more than mention Aristotle’s belief, doubtless old in his time, that the more pointed egg produces the male chicken, and the blunter egg the hen; though this theory survived into modern times† and still lingers on (cf. p. 943). Several naturalists, such as Günther (1772) and Bühle (1818), have taken the trouble to disprove it by experiment. A more modern and more generally accepted

# What is the Shape of an Avian Egg?

Avian egg images supplied by [The Museum of Vertebrate Zoology at Berkeley](#).



- ▶ 49,175 photographs of bird eggs for more than 100 years
- ▶ 37 bird orders
- ▶ 1400 species

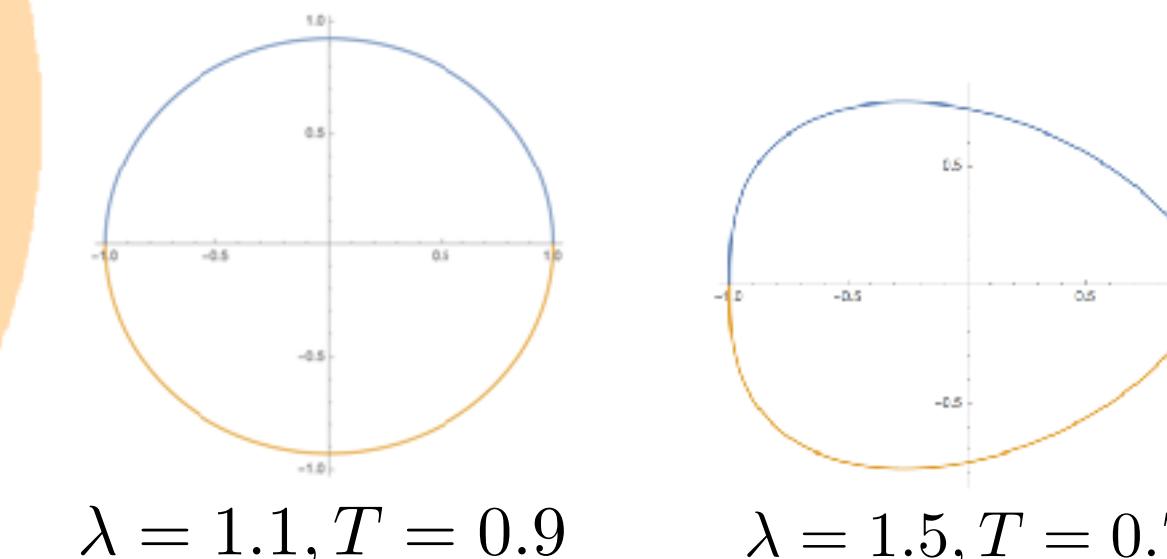
**Baker's model:** projective geometry to develop a simple algebraic equation

[D. Baker, Auk. 119, 2002](#)

$$y = T(1 + x)^{1/(1+\lambda)}(1 - x)^{\lambda/(1+\lambda)}$$

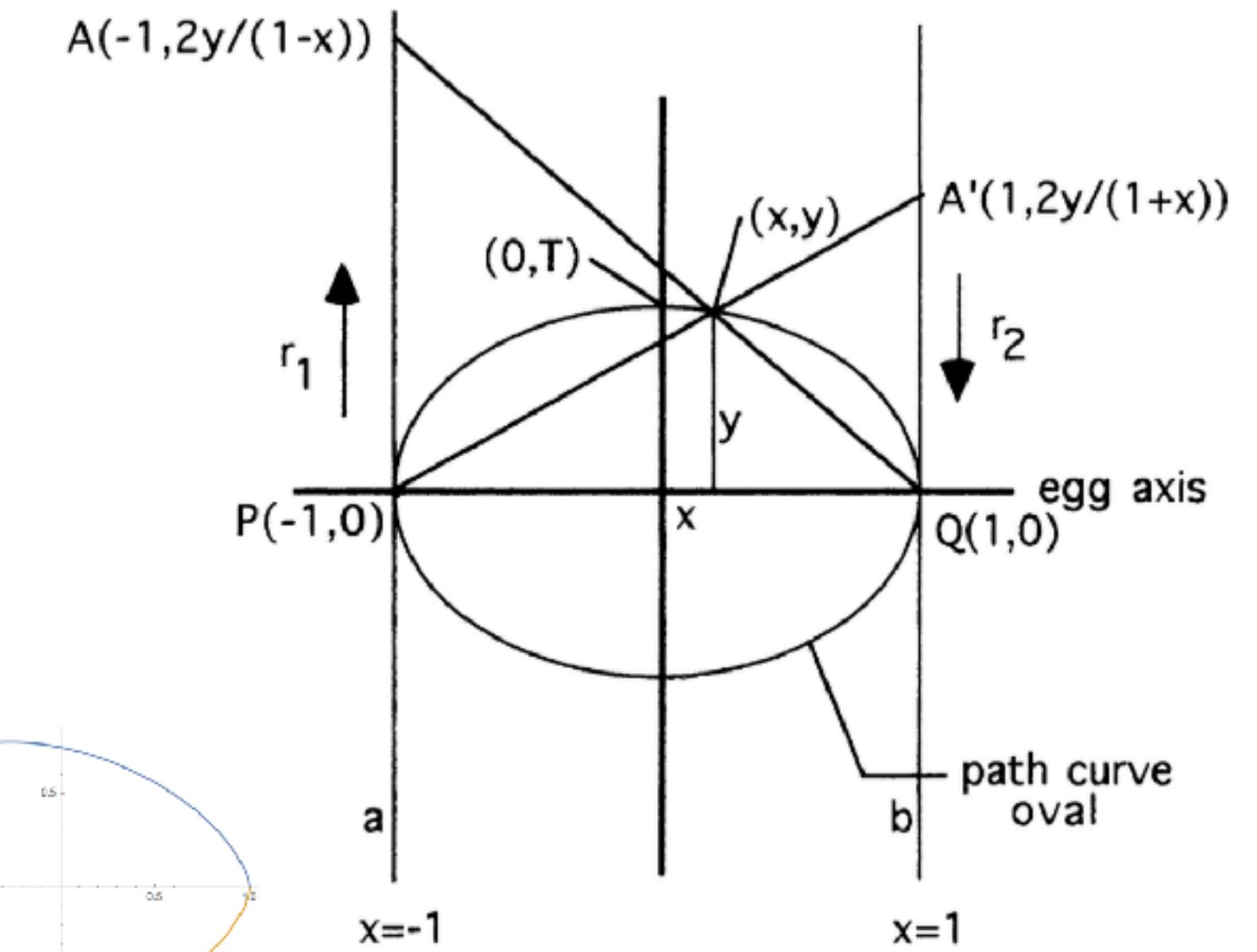
**Asymmetry**  
 $\alpha = \lambda - 1$

**Ellipticity**  
 $E = 1/T - 1$



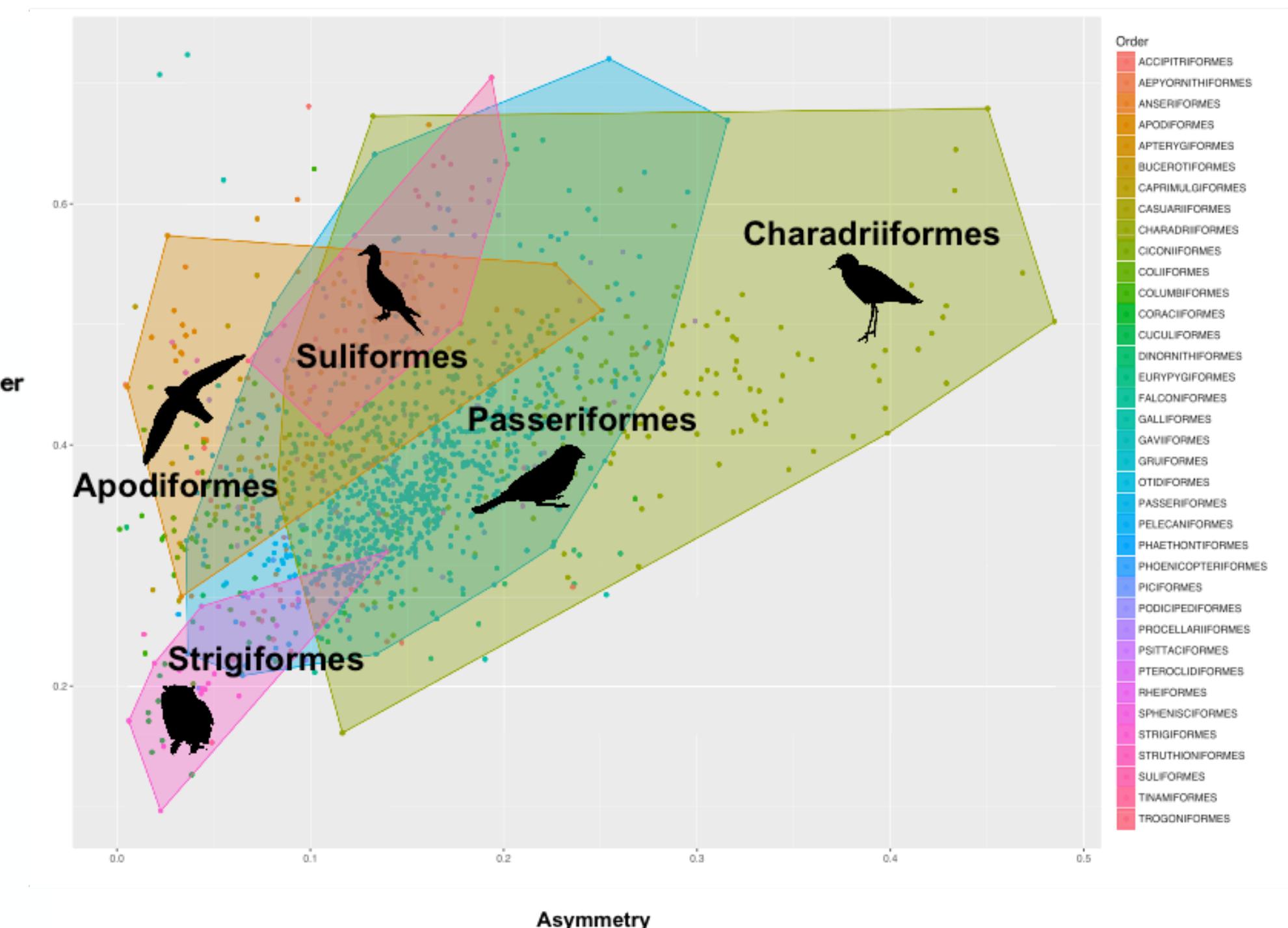
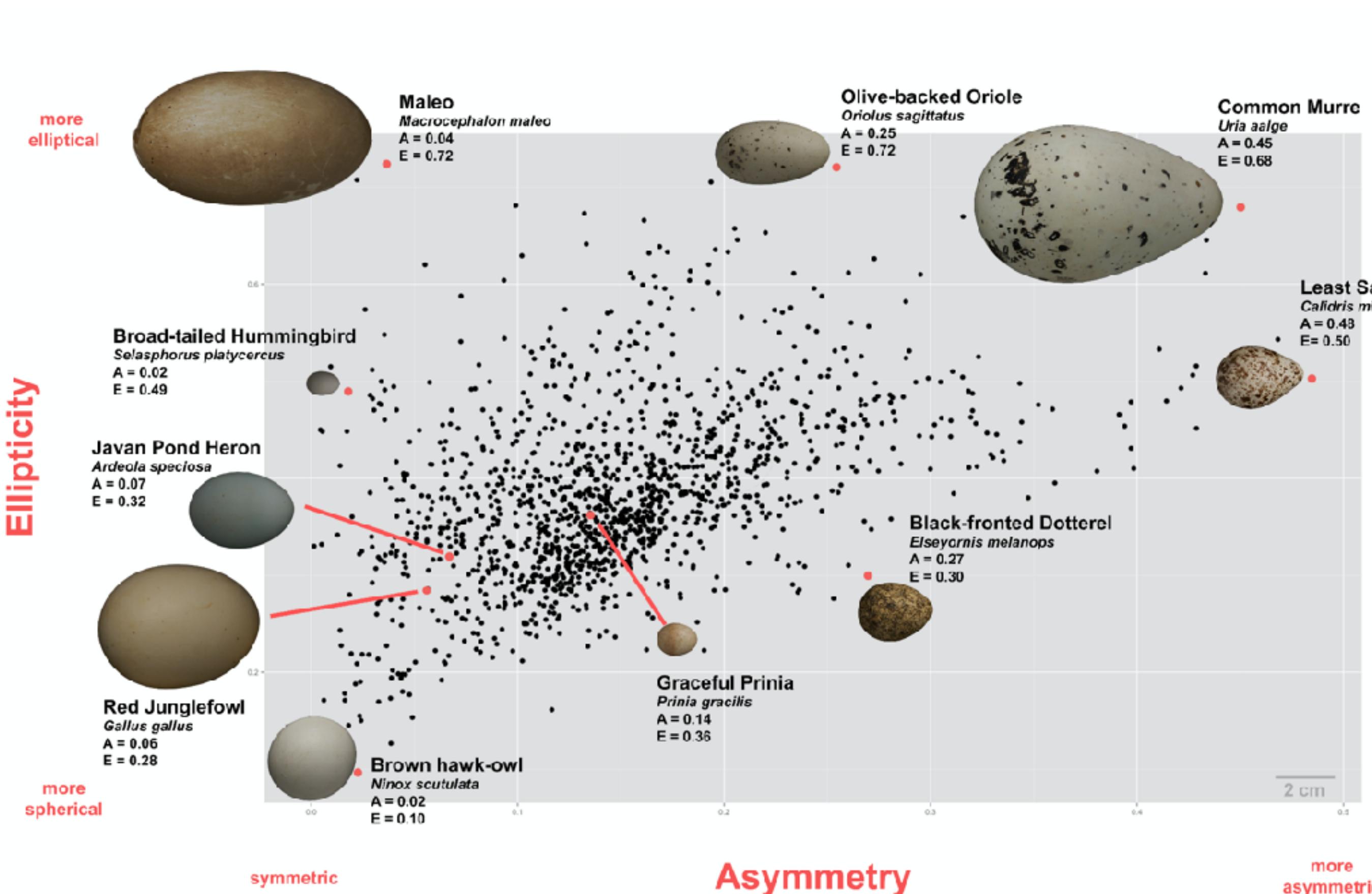
$$\lambda = 1.1, T = 0.9$$

$$\lambda = 1.5, T = 0.75$$



# Avian Egg Morphospace

**Morphological Space:** Occupied morphospace is triangular, with the bounding vertices corresponding to eggs that are symmetric and spherical, symmetric and elliptical, and asymmetric and elliptical.



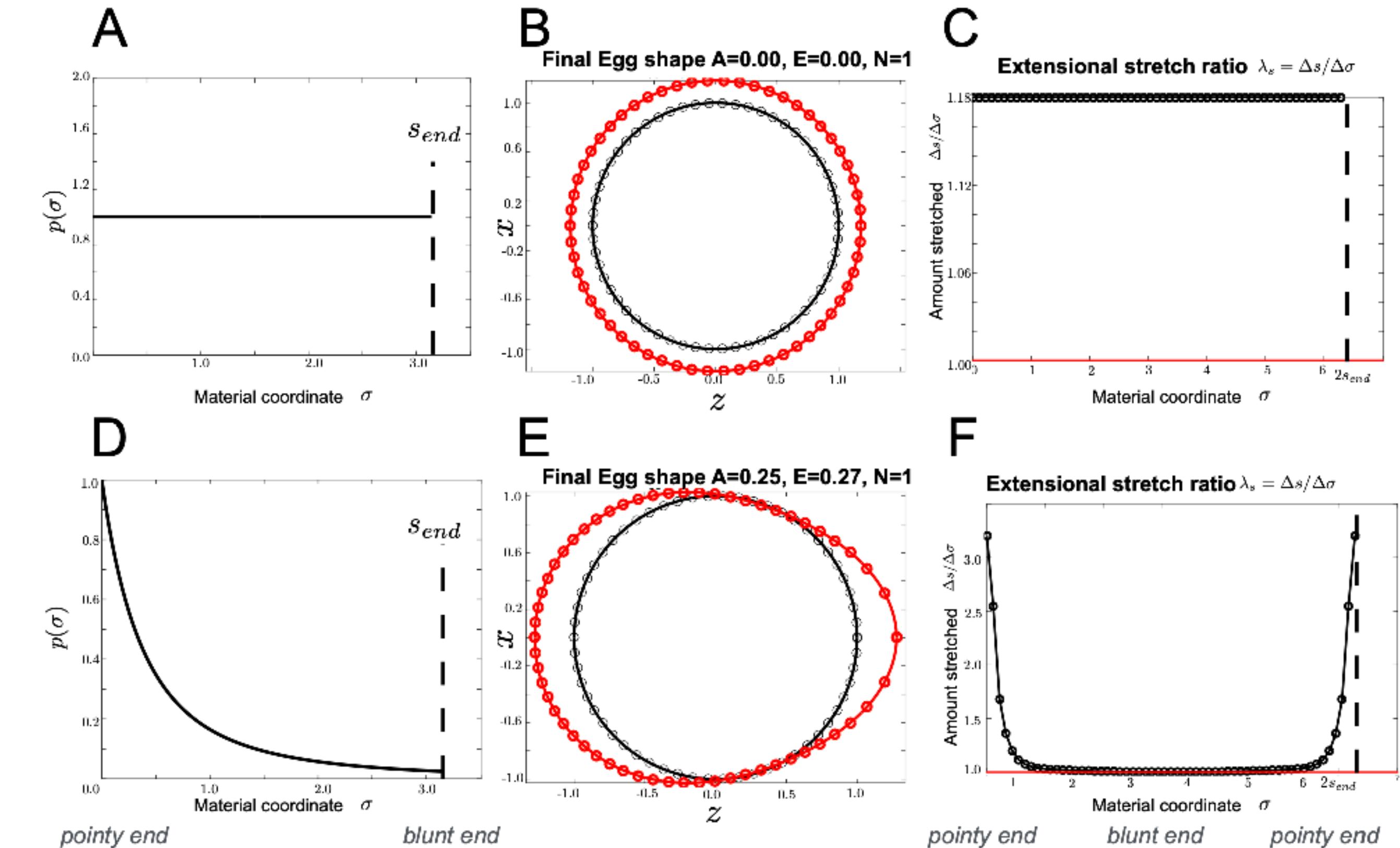
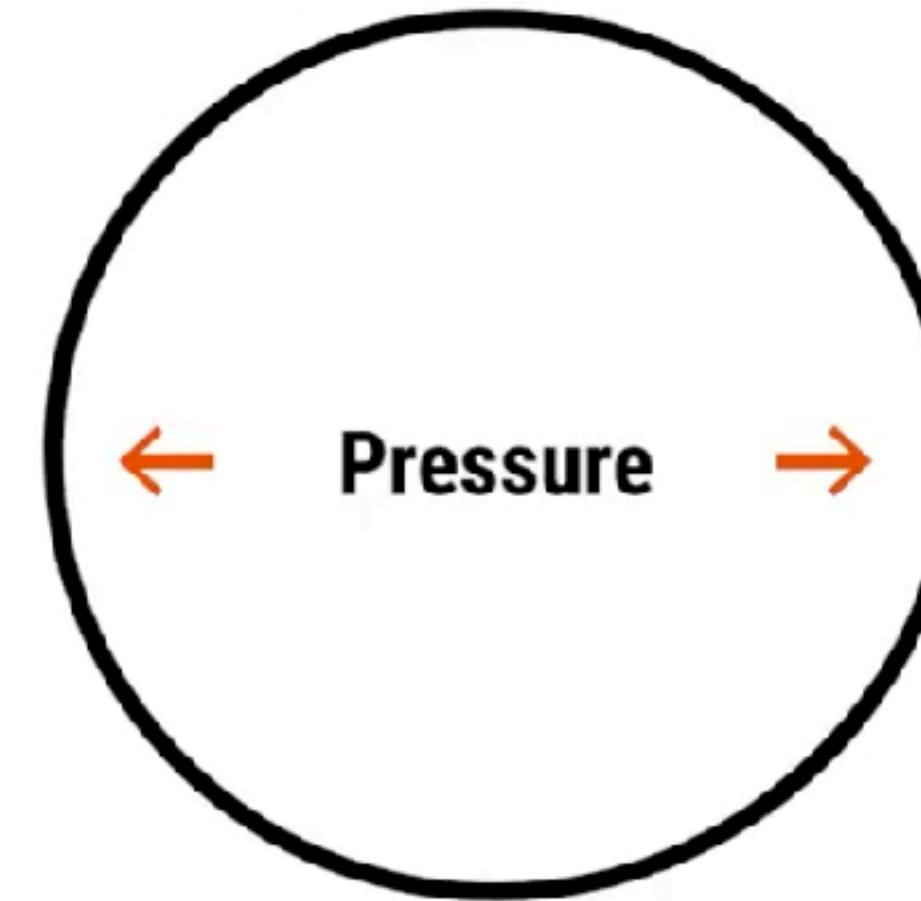
M.C. Stoddard, E.H. Yong, et al., Science 356, 2017

# How is the Avian Egg Shaped?

**Morpho-elasticity approach:** Framework of large deformation membrane theory in which the cell wall is represented as a growing elastic membrane with geometry-dependent elastic properties.

Two parameters in model:

- ratio of azimuthal and axial stiffnesses  $\mu(\sigma)$
- scaled pressure  $p(\sigma)$



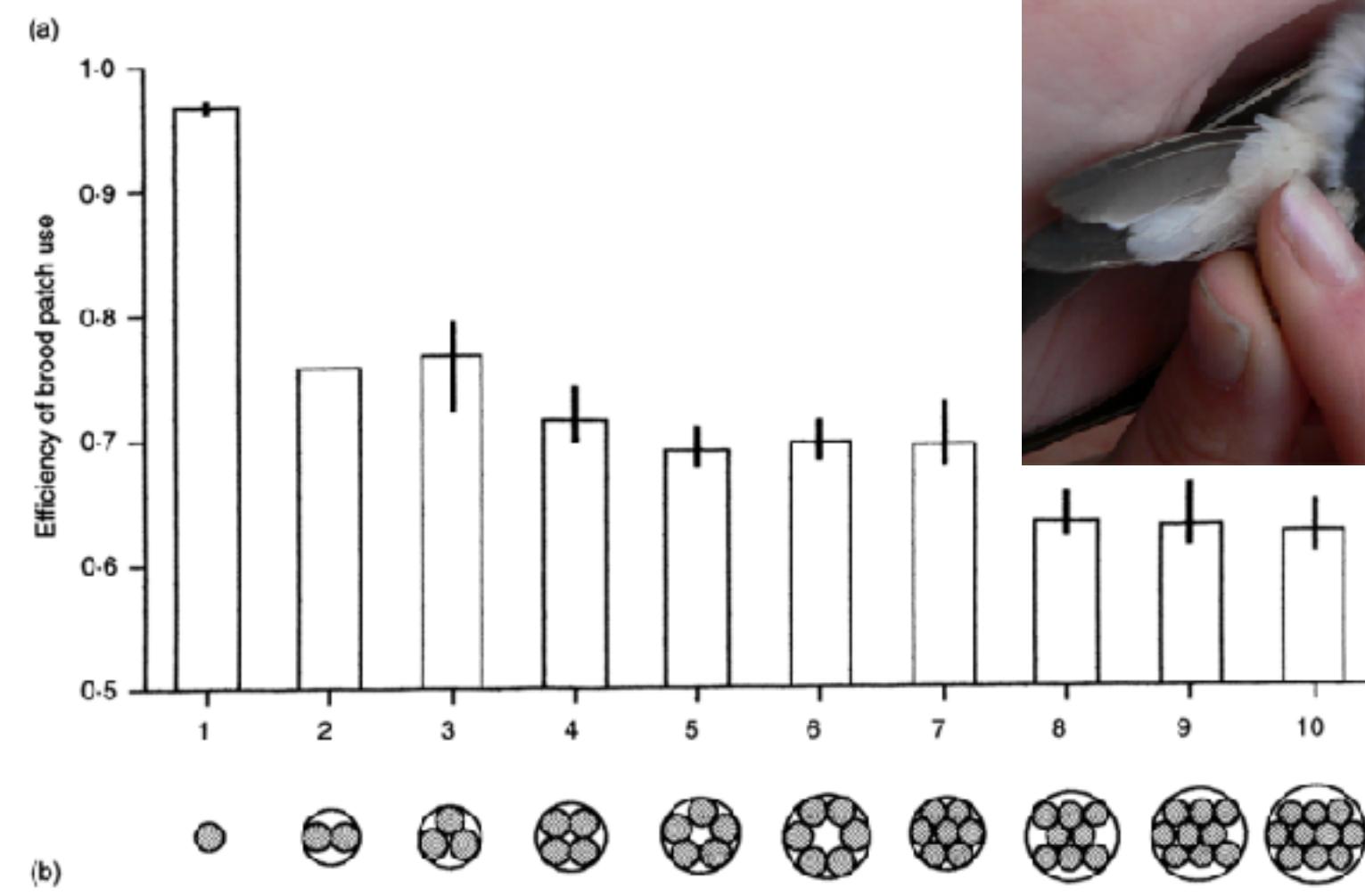
M.C. Stoddard, E.H. Yong, et al, Science 356, 2017

A. Goriely and M. Tabor, J. of Theo. bio. 222(2), 2003

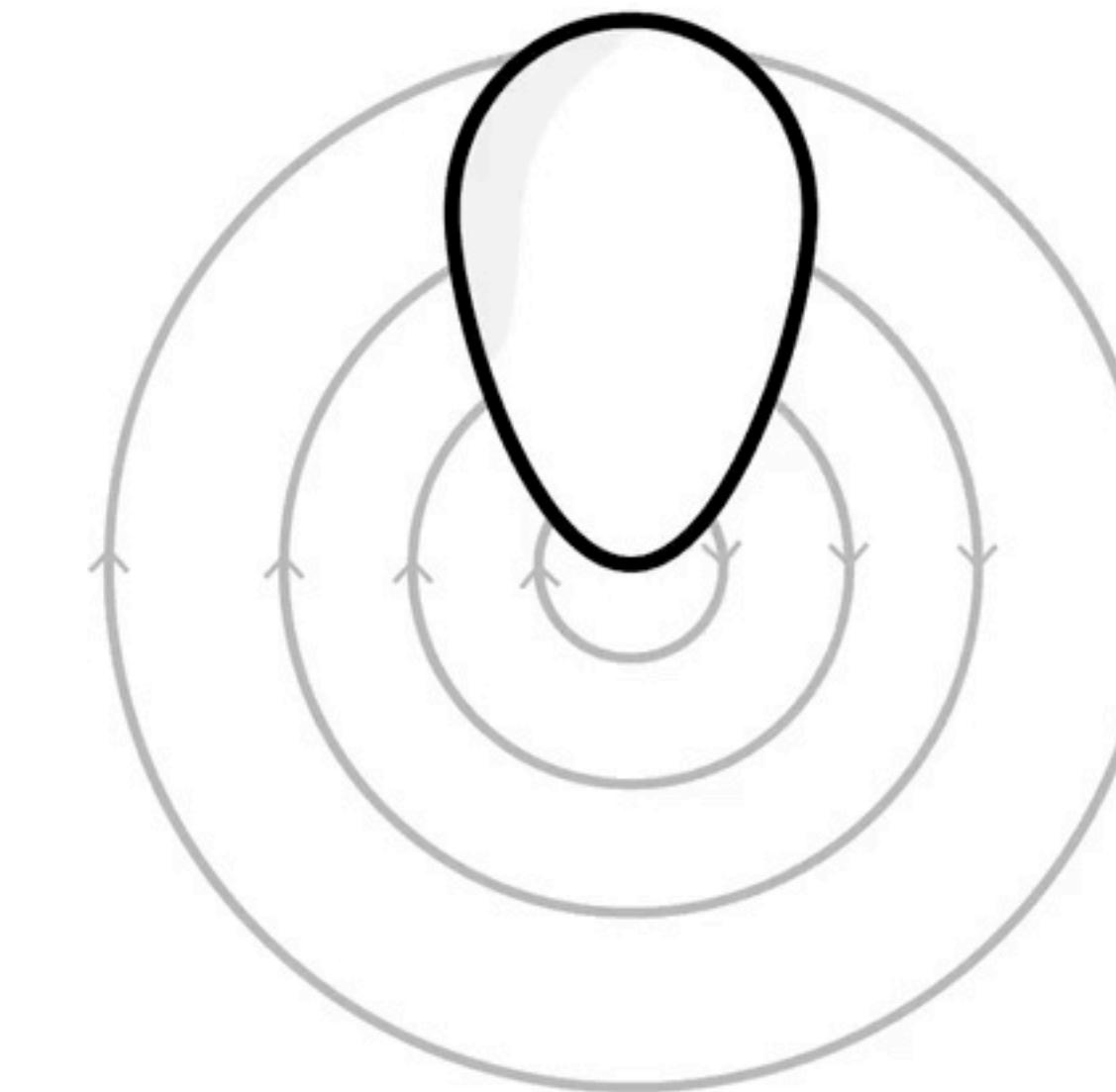
# Why are there so many shapes for bird egg?

Hypotheses for egg shape variation generally relate to differences in **biometric**, **life history**, and **environmental parameters**, such as adult body mass, diet, clutch size, nest location, brood size, chick developmental mode (e.g., precocial), and environmental details (latitude, temperature, and average precipitation)

**Brood patch** is a patch of featherless skin that is visible on the underside of birds during the nesting season



I. Smart, in *Egg Incubation: Its Effects on Embryonic Development in Birds and Reptiles*, C. Deeming, Ed. (Cambridge University Press, New York, 1991)



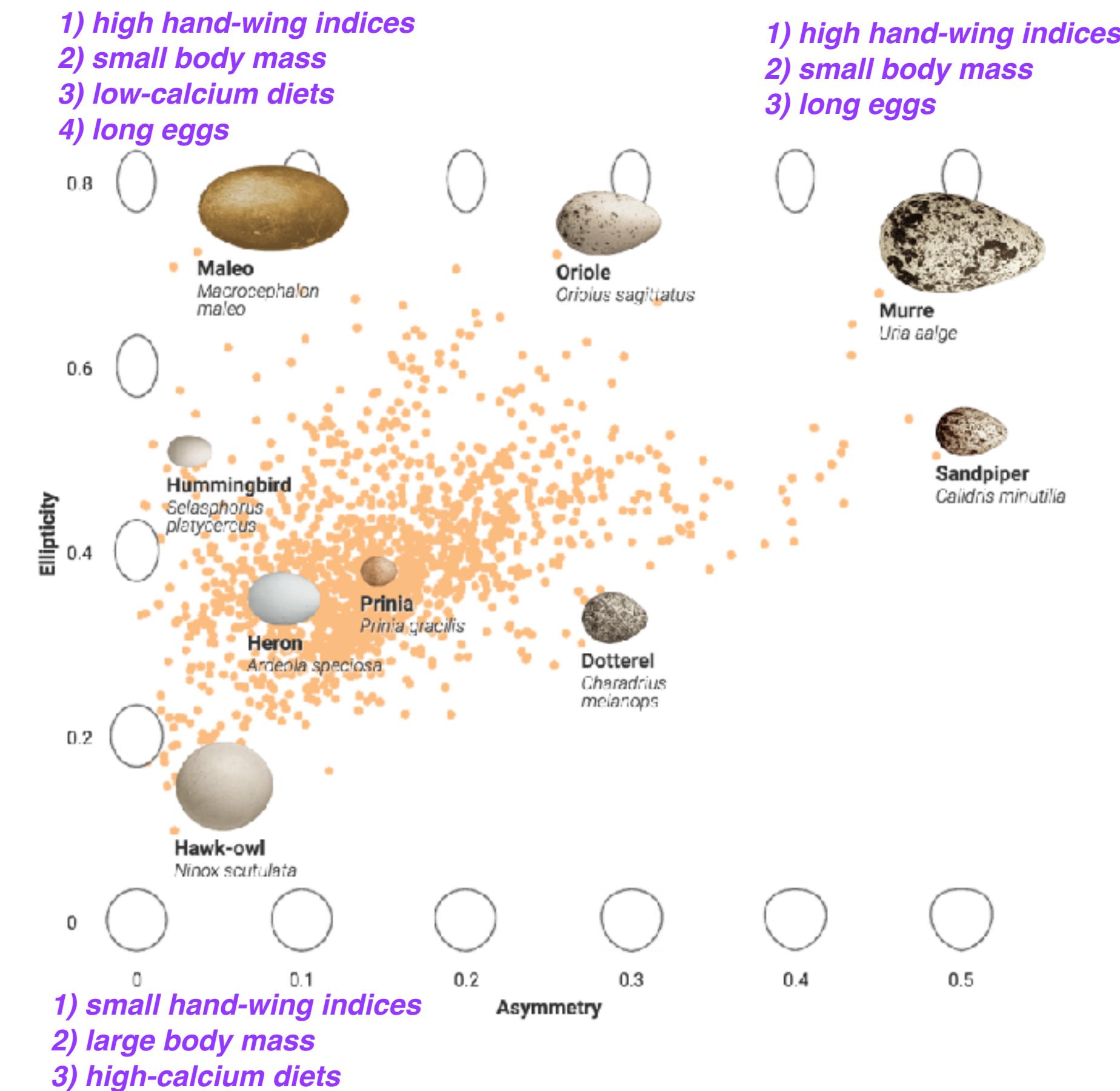
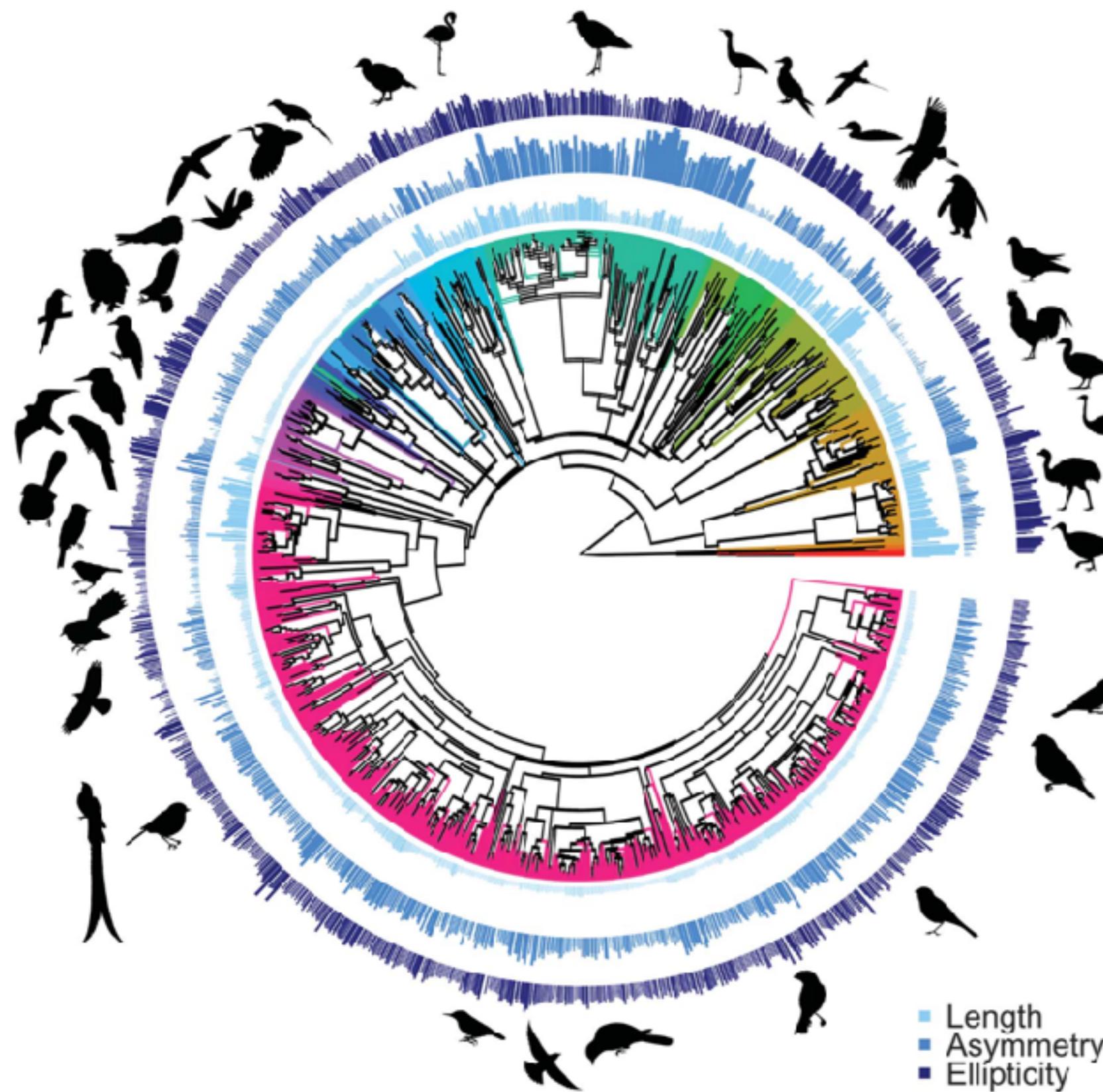
Nest location



Spherical eggs could easily roll off a cliff. Conical eggs, however, roll in a tight circle, making them perfect for cliff-nesting birds.

# Why are there so many shapes for bird egg?

- Assembled a **phylogenetic tree** of 1209 species onto which we mapped average egg length (light blue), asymmetry (medium blue), and ellipticity (dark blue)
- Used phylogenetic trees as a framework for performing **Bayesian mixed-effect models** using R package **MCMCglmm** (Markov chain Monte Carlo generalized linear mixed models)



## Length

Birds tend to lay **long** eggs if

- low clutch sizes
- large body mass
- high-calcium diet
- precocial

## Ellipticity

Birds tend to lay **elliptical** eggs if

- high hand-wing indices
- small body mass
- low-calcium diets
- long eggs

## Asymmetry

Birds tend to lay **asymmetric** eggs if

- high hand-wing indices
- small body mass
- long eggs

J.D. Hadfield, J. of stats. software, 33, 2010

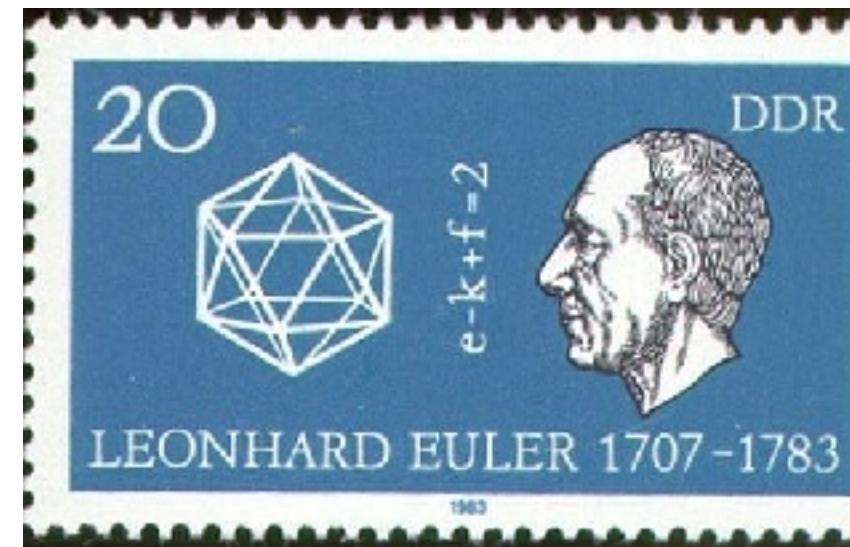
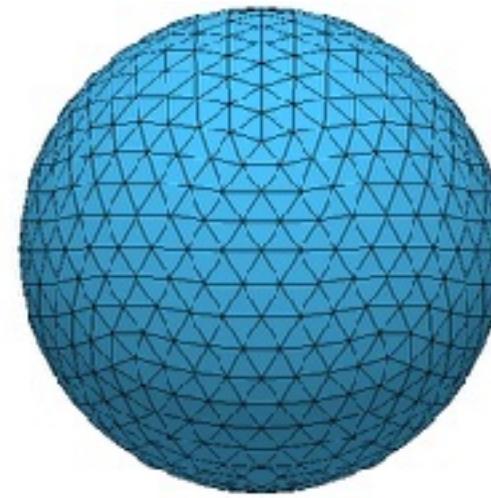
Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

# Euler's Theorem & Geometric Frustration

For a polyhedron with  $V$  = number of vertices ,  $E$  = edges, and  $F$  = faces:

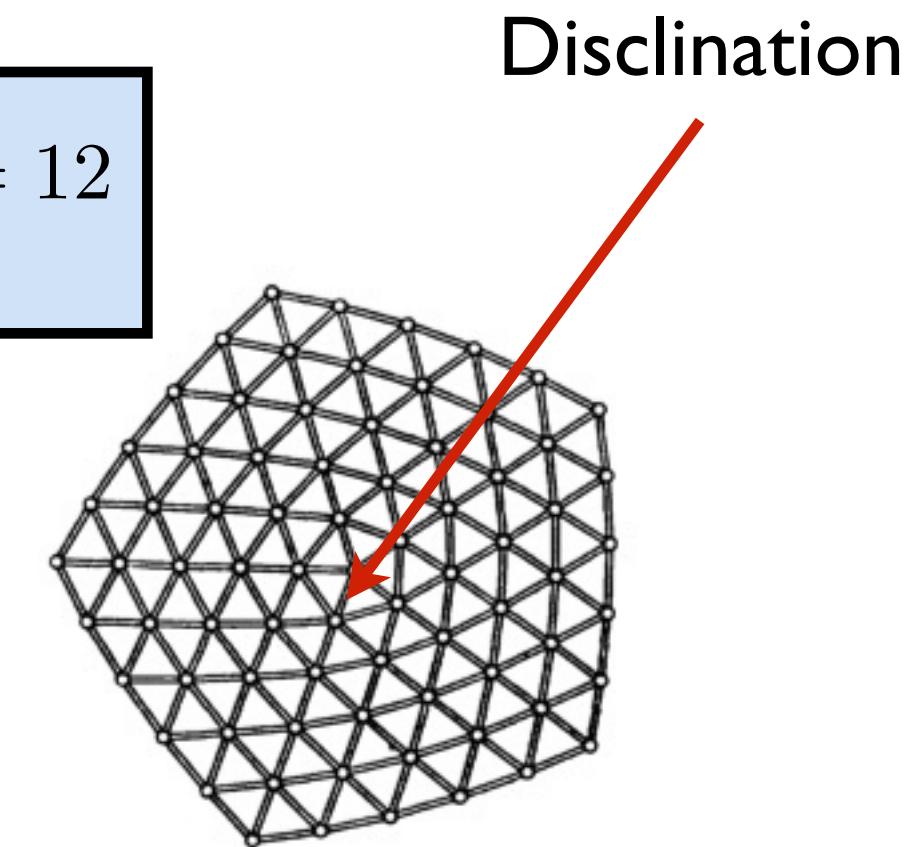
$$\chi = V - E + F = 2$$



Name	Image	Vertices $V$	Edges $E$	Faces $F$	Euler characteristic: $V - E + F$
Tetrahedron		4	6	4	2
Hexahedron or cube		8	12	6	2

Triangulate the surface of a sphere:

$$\sum_z (6 - z) N_z = \sum_z q_z N_z = 12$$



**Table 1.** Different configurations of topological defects.

# distinct defects	Different cases
1	{0, 0, 12}, {4, 0, 0}, {0, 6, 0}
2	{0, 1, 10}, {0, 2, 8}, {0, 3, 6}, {0, 4, 4}, {0, 5, 2}, {1, 0, 9}, {2, 0, 6}, {3, 0, 3}, {2, 3, 0}
3	{1, 1, 7}, {1, 2, 5}, {1, 3, 3}, {1, 4, 1}, {2, 1, 4}, {2, 2, 2}, {3, 1, 1}

J.D. Hadfield, J. of stats. software, 33, 2010

Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

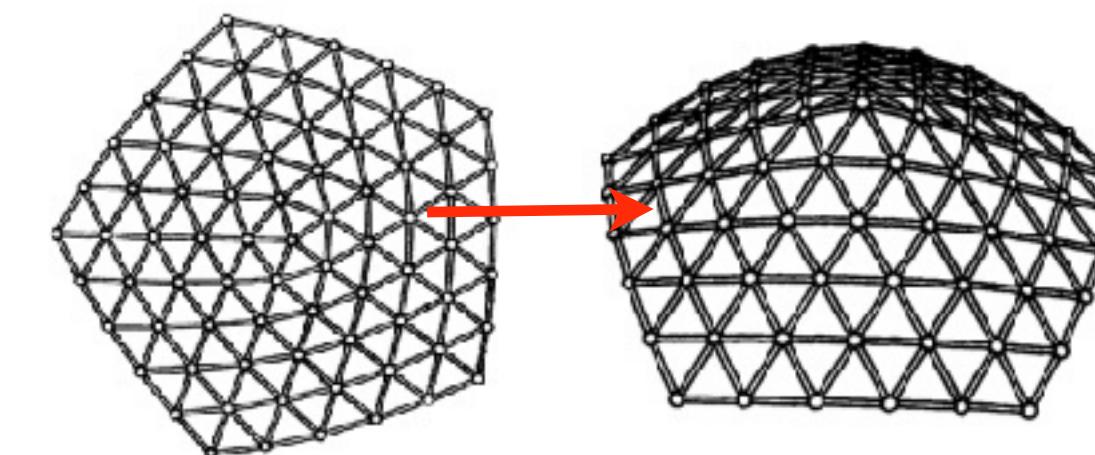
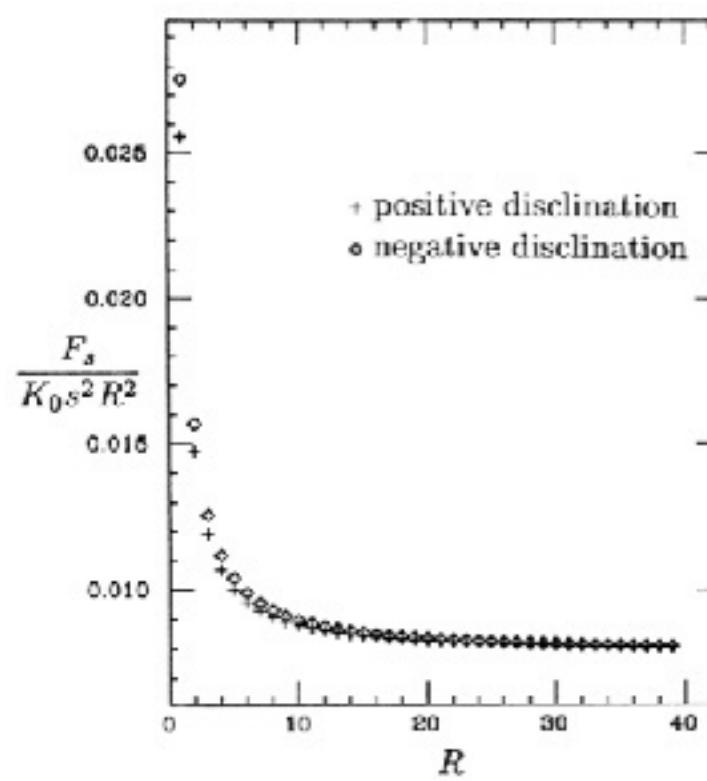
# Buckling Transition & Virus Shapes

## Defects in flexible membranes with crystalline order

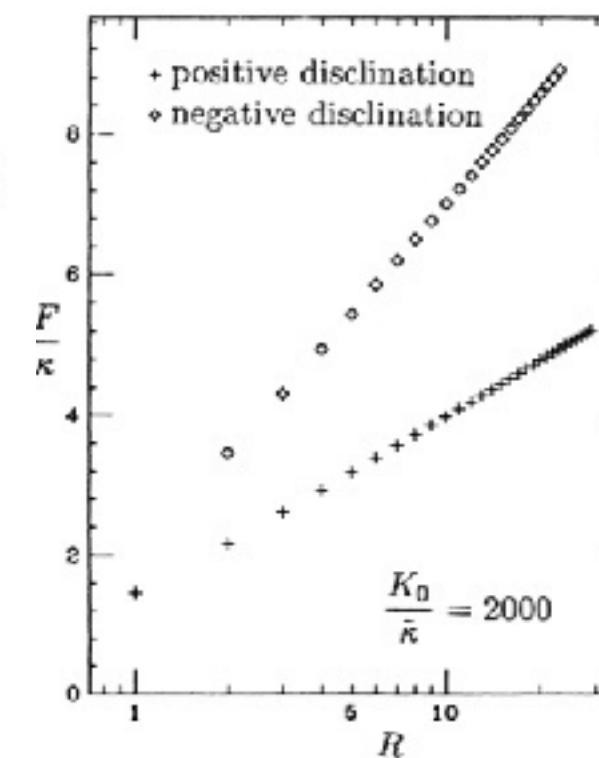
H. S. Seung and David R. Nelson

Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138

(Received 3 March 1988)



$$Ehs^2R^2 \rightarrow |s|\kappa \log(R/a)$$



H.S. Seung and D.R. Nelson, Phys. Rev. A, 38(2), 1988

## Virus shapes and buckling transitions in spherical shells

Jack Lidmar

Department of Physics, Royal Institute of Technology, AlhaNova, SE-106 91 Stockholm, Sweden

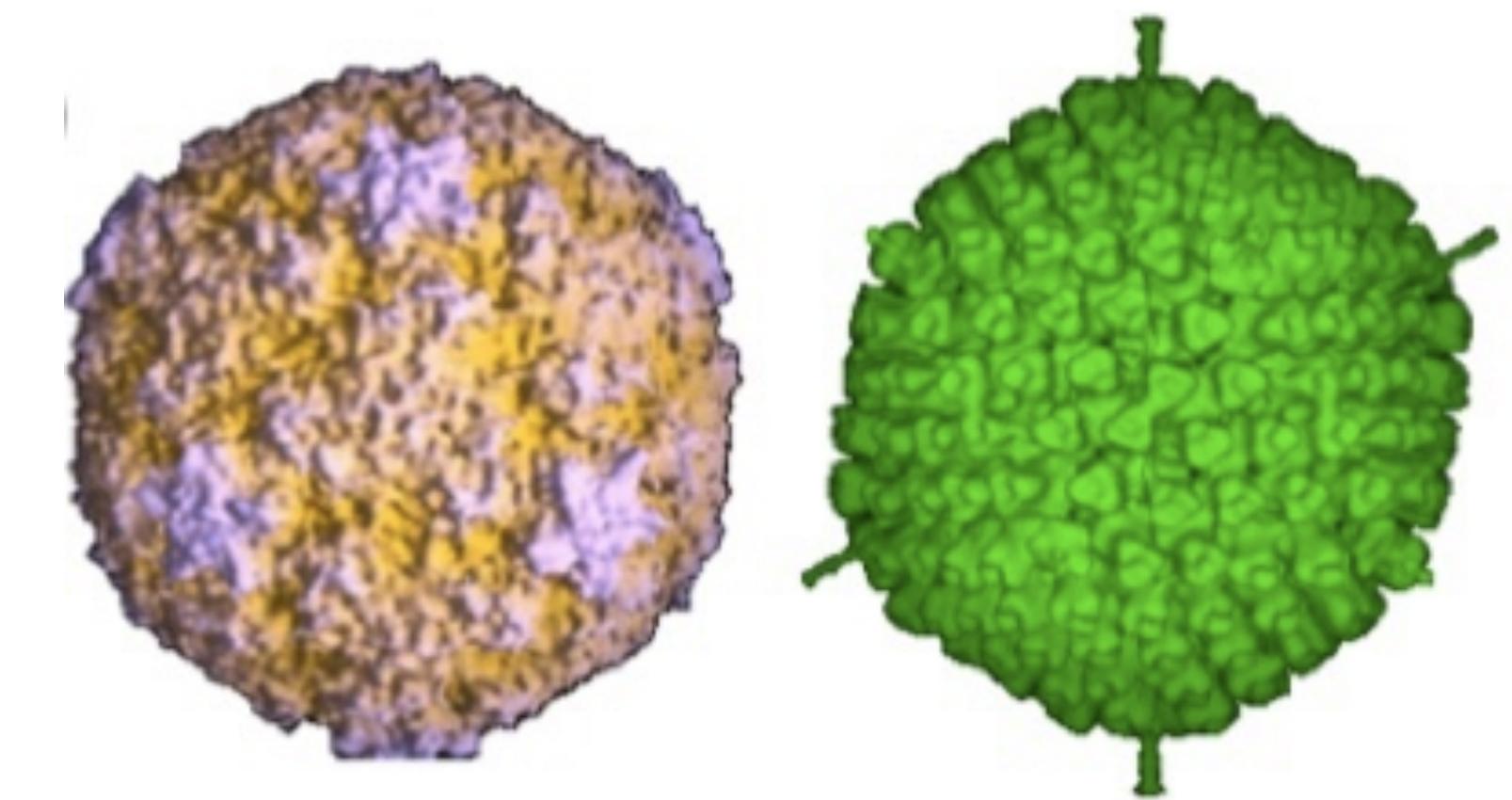
Leonid Mirny

Harvard MIT Division of Health Science and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

David R. Nelson

Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138, USA

(Received 27 June 2003; published 25 November 2003)

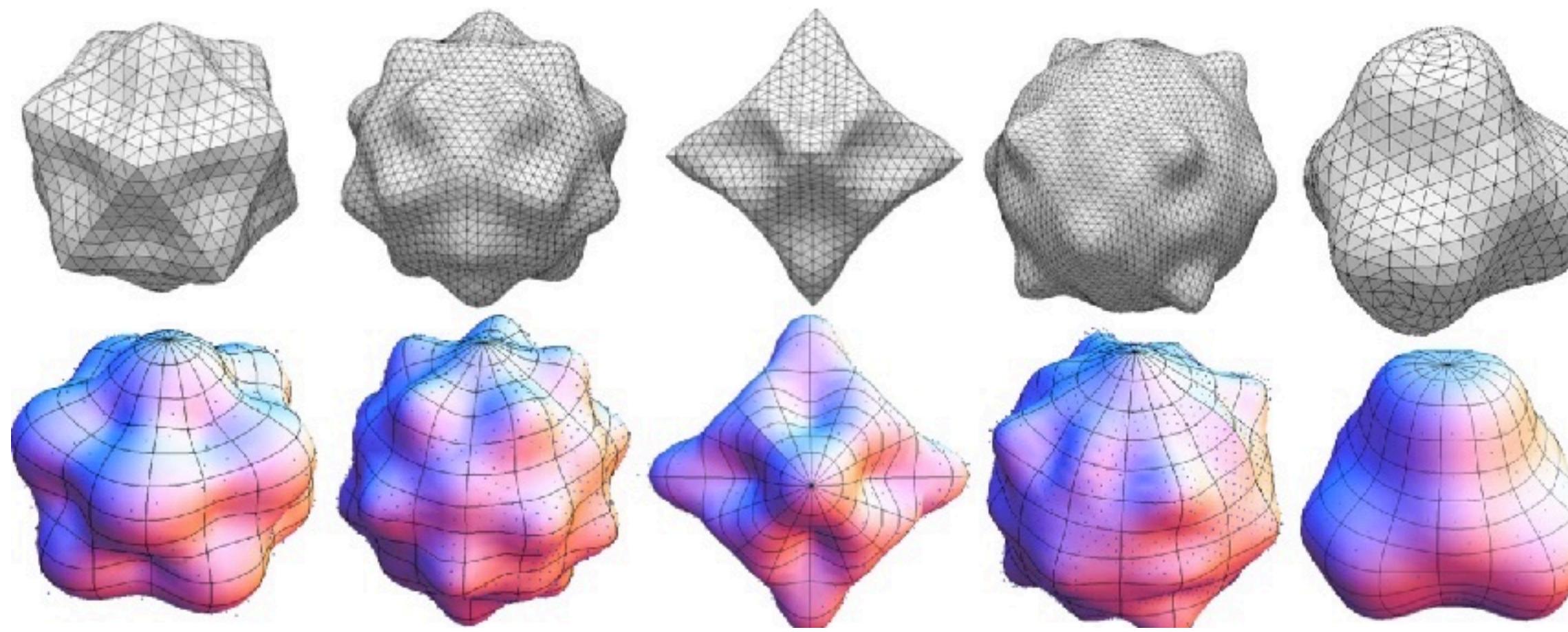


J. Lidmar, L. Mirny & D.R. Nelson, Phys. Rev. E, 68(5), 2003

Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

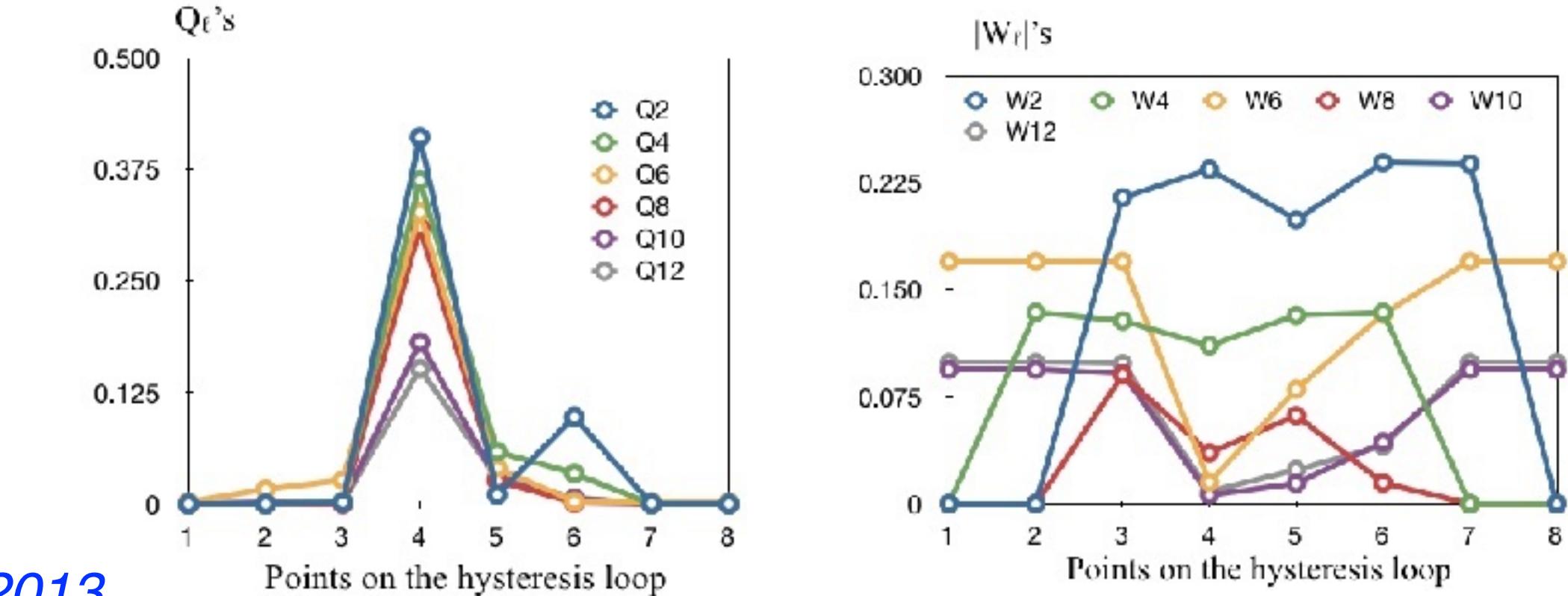
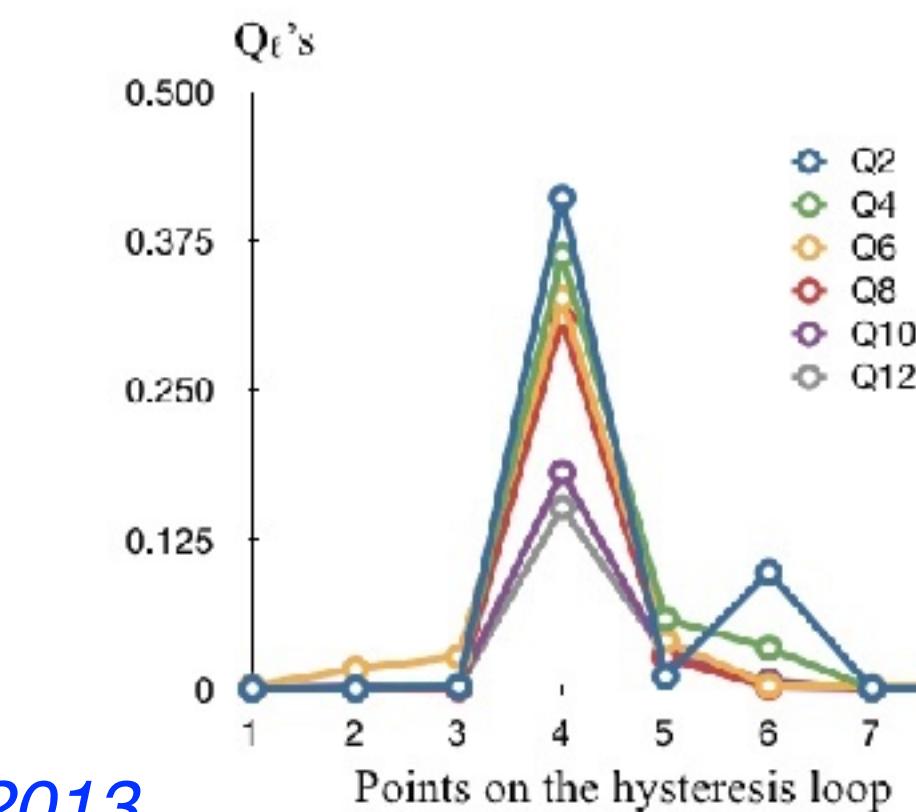
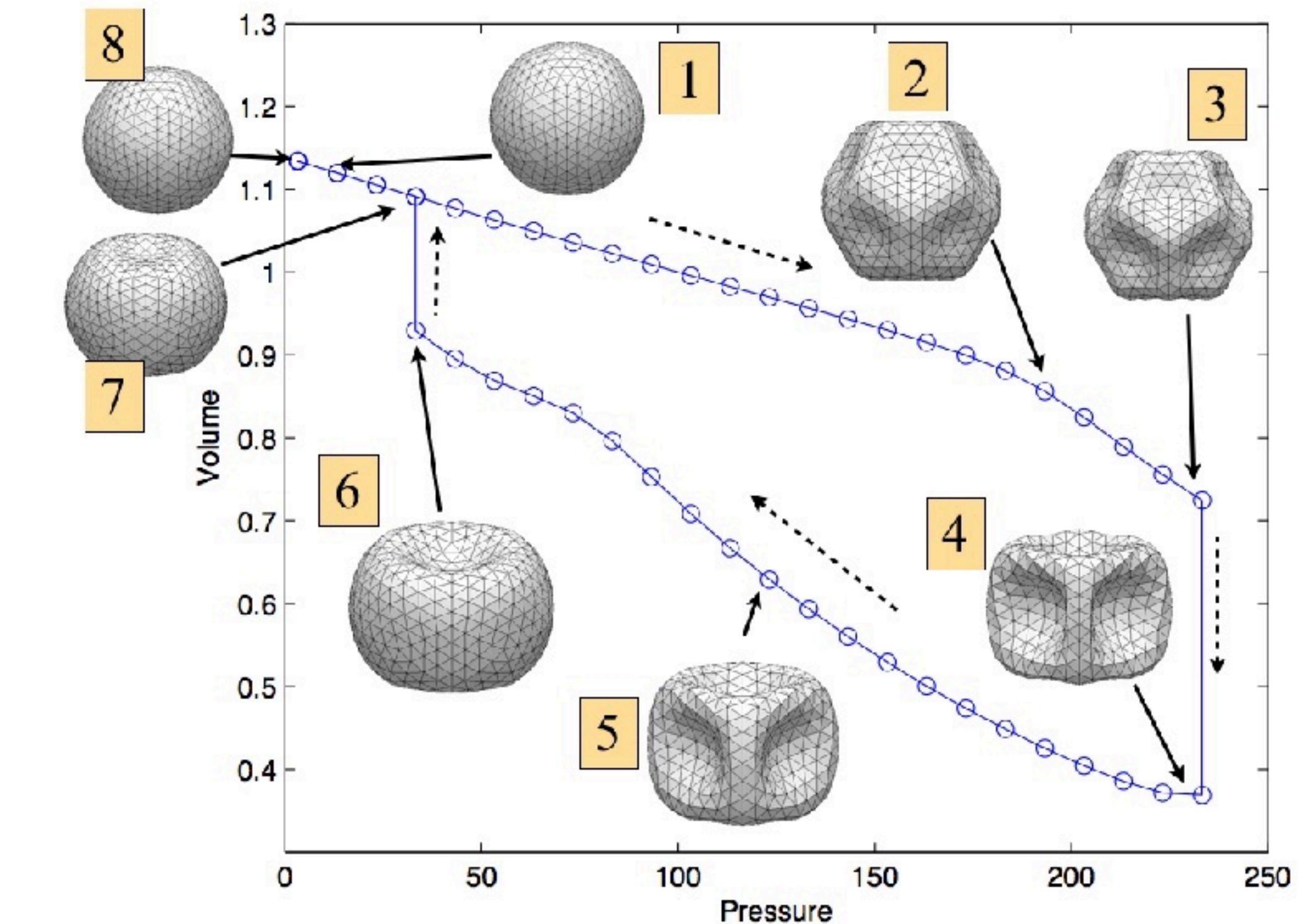
# Elastic Platonic Shells



$$Q_\ell = \left( \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{\ell} |a_\ell^m|^2 \right)^{1/2} \quad \hat{Q}_\ell = Q_\ell / Q_0$$

$$\hat{W}_\ell = \frac{\sum_{m_1+m_2+m_3=0} \binom{\ell}{m_1} \binom{\ell}{m_2} \binom{\ell}{m_3} a_\ell^{m_1} a_\ell^{m_2} a_\ell^{m_3}}{\left[ \sum_m |a_\ell^m|^2 \right]^{3/2}}$$

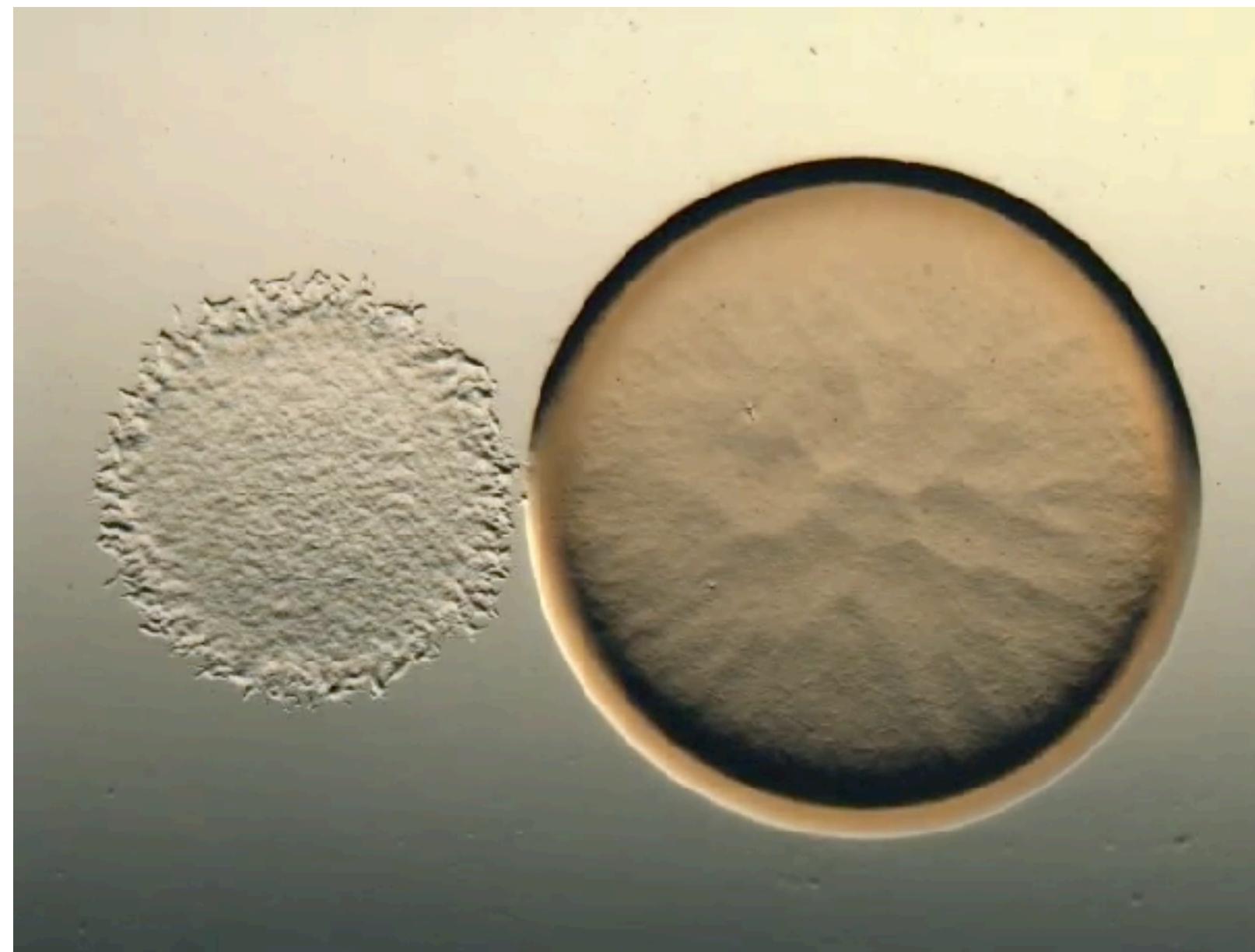
E.H. Yong, D.R. Nelson & L. Mahadevan, Phys. Rev. Letters, 111(17), 2013



# Active Particles to Model Living systems

**Active particles** = “machines” that consume energy to generate own motion.

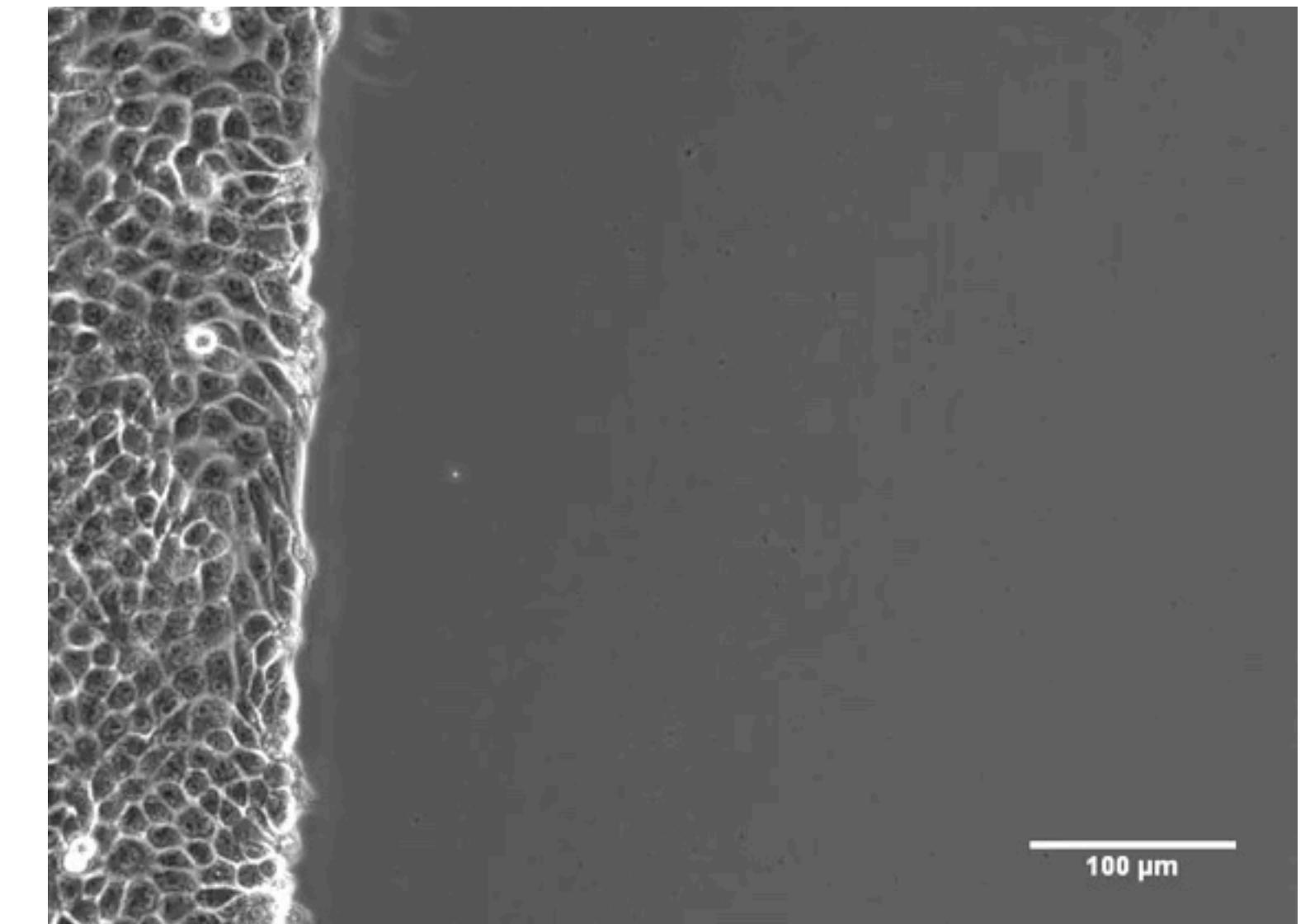
**Active systems** = collection of interacting, self-driven “machines” that **self-organize** on multiple scales in presence of noise and typically, with no leader.



Keane & Berleman, Microbiology 162, 2016



National Geographic, Youtube, 2016



K. Matsuzawa et al. Cell Reports 26, 2018

# Swarming Behaviors in Honeybees

In collaboration with



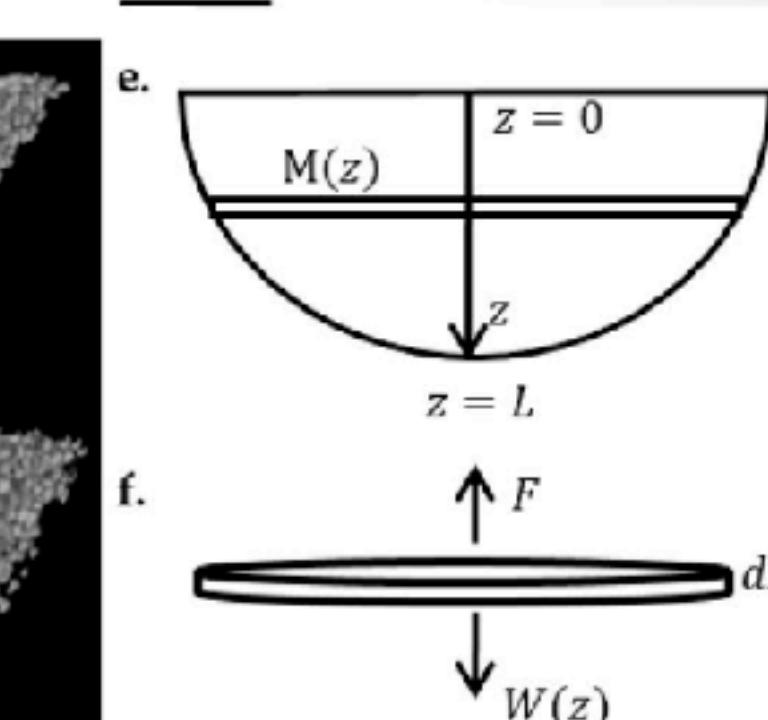
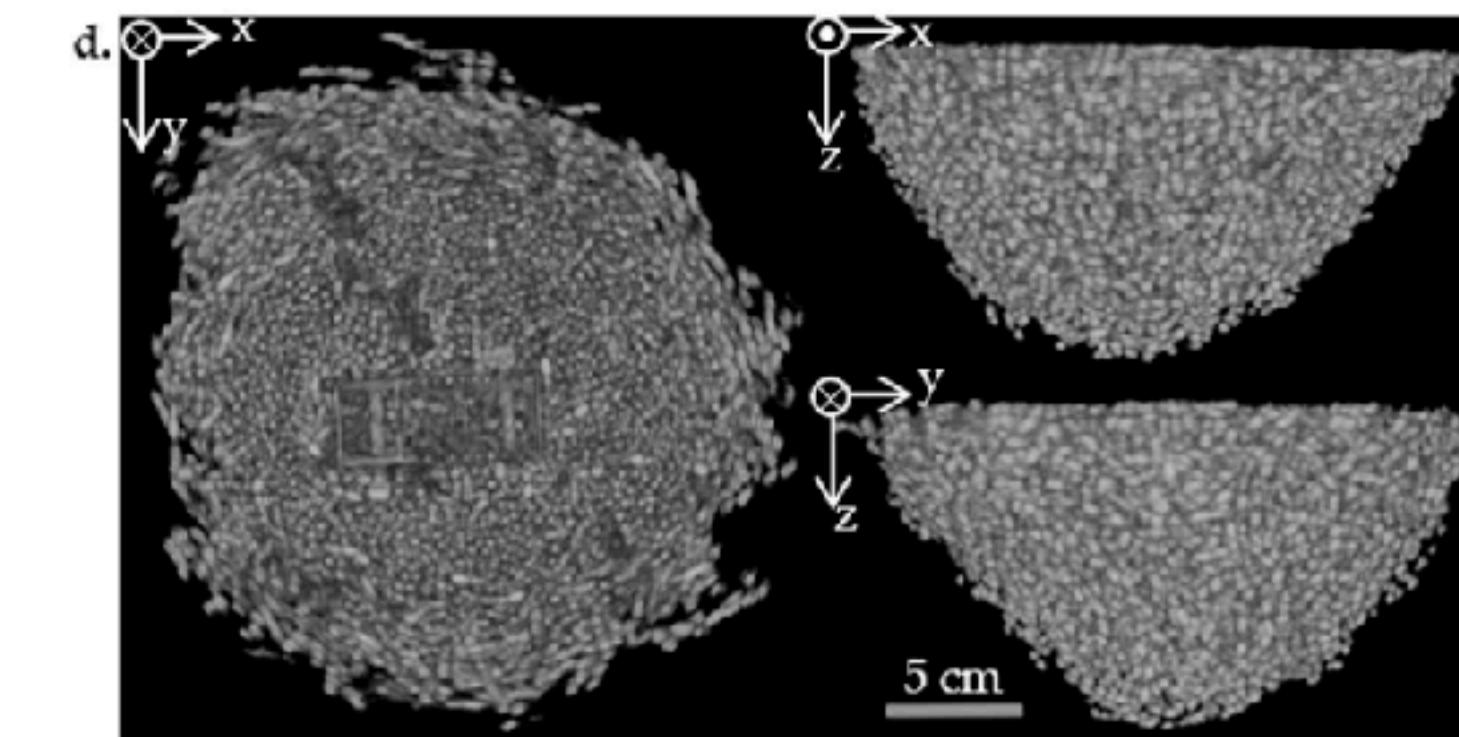
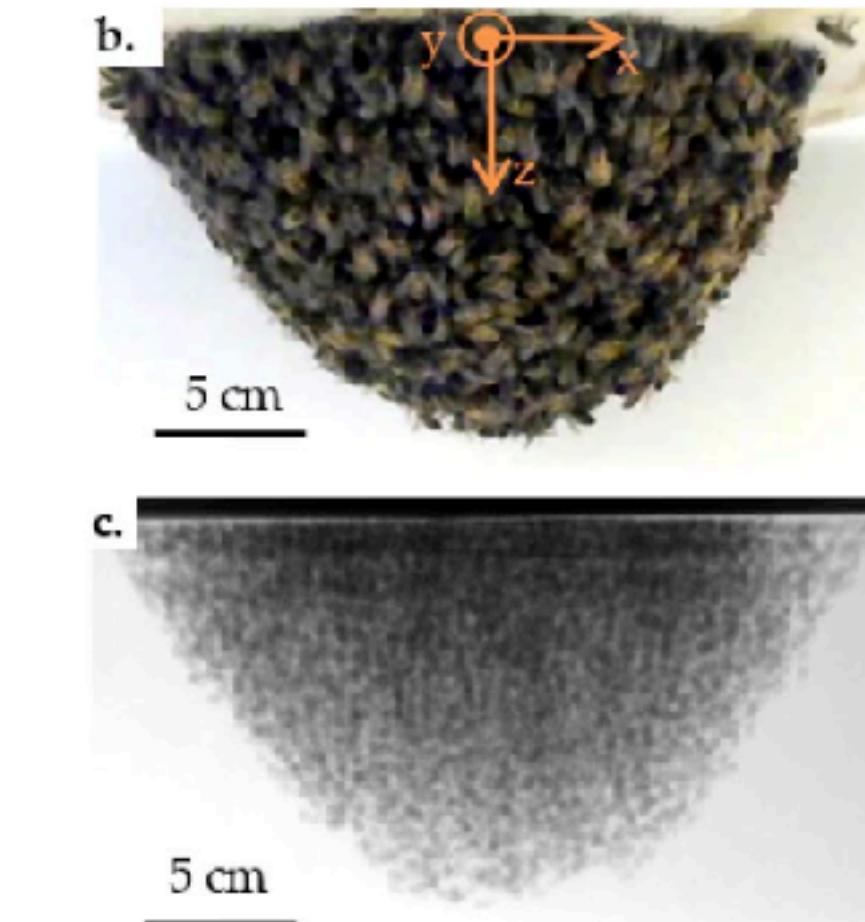
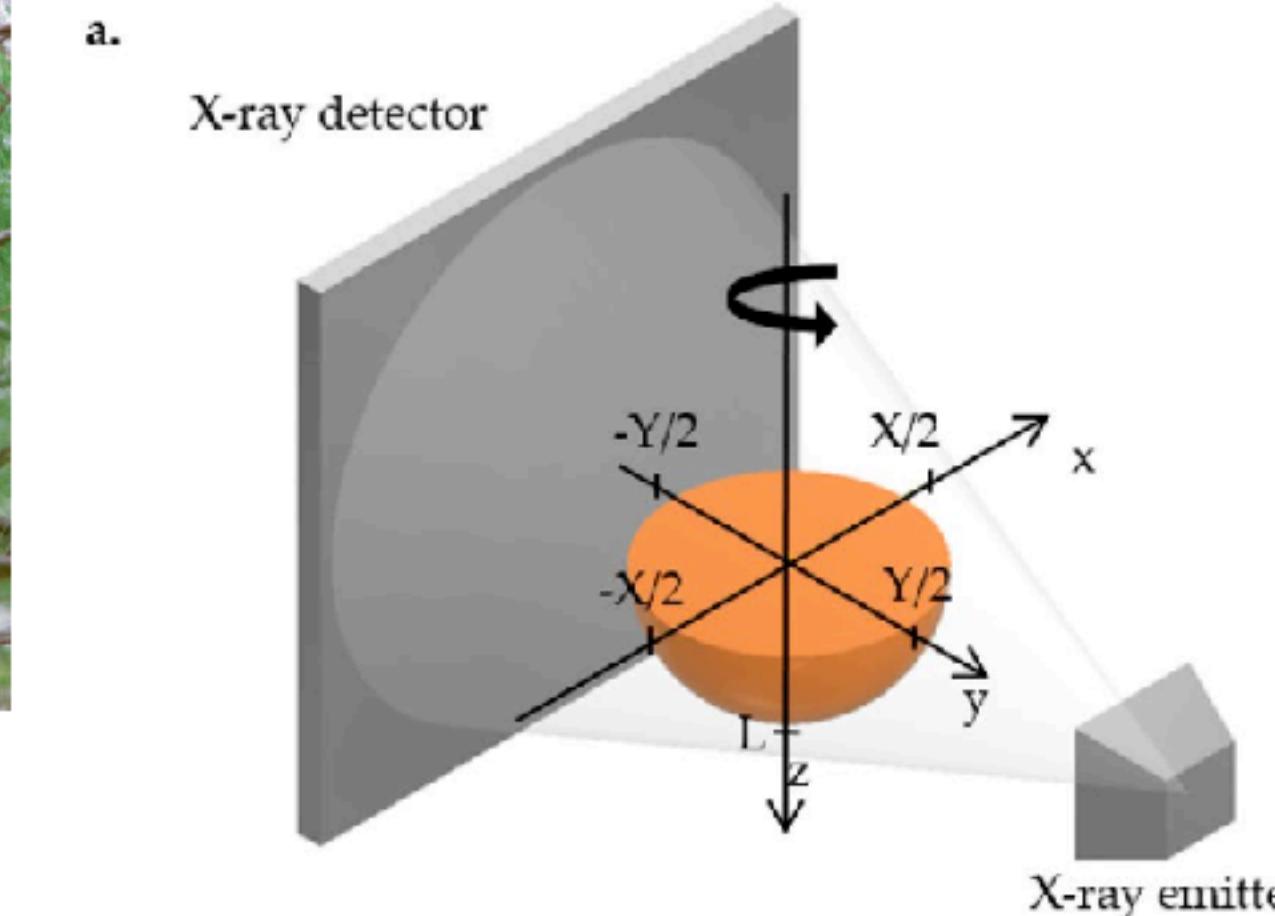
Orit Peleg  
UC Boulder

Bivouacs in nature



- **Bivouac** is temporary cluster formed by Western honey bees (*Apis mellifera*) during colony reproduction.
- Understand self-assembly using agent based modeling to simulate the swarming process.

Mass distribution inside honey bee swarms

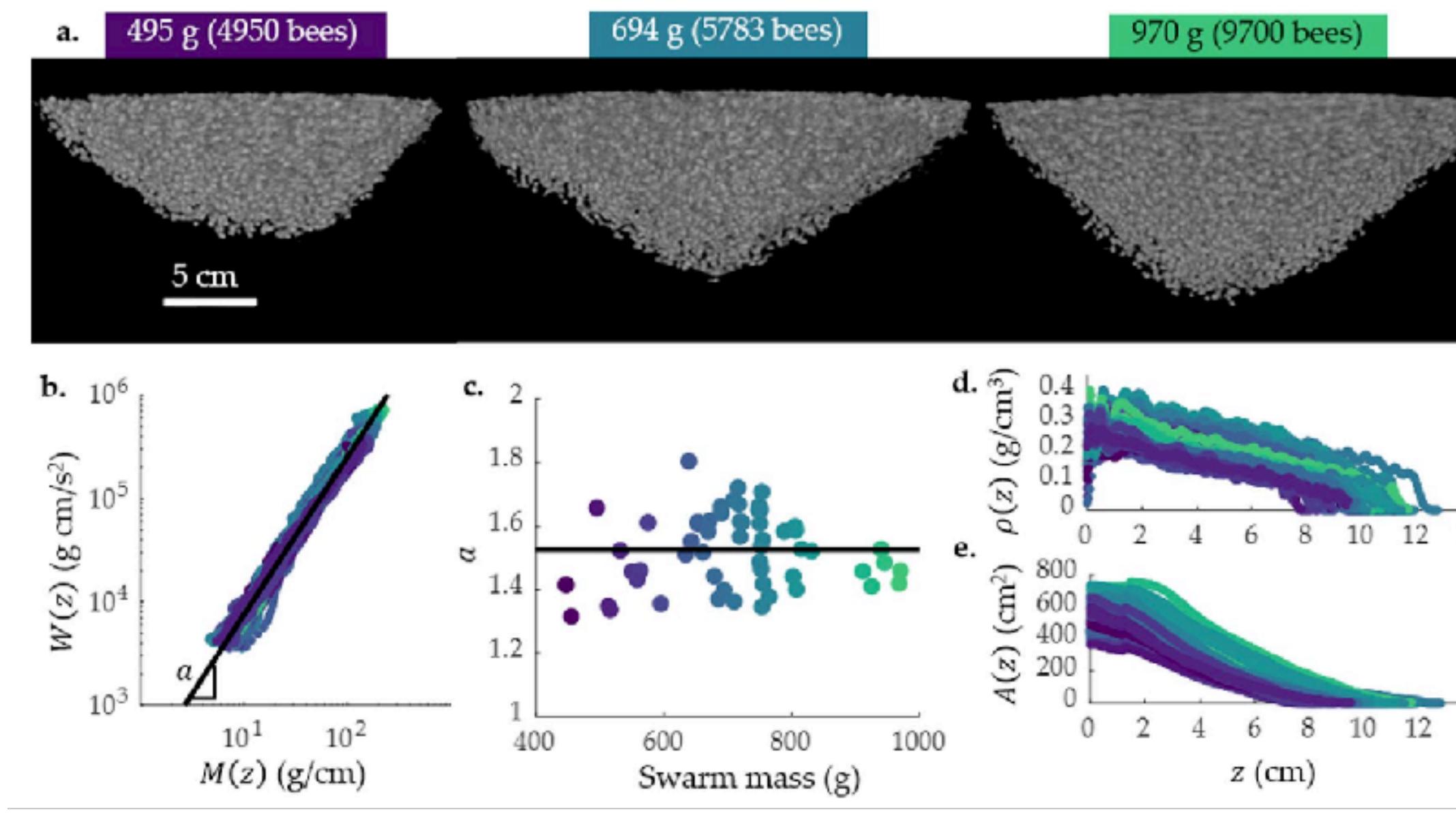


*O. Shishkov et al., Scientific reports, 12(1), 2022*

Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

# Statistical Dynamics of Honeybee Swarms

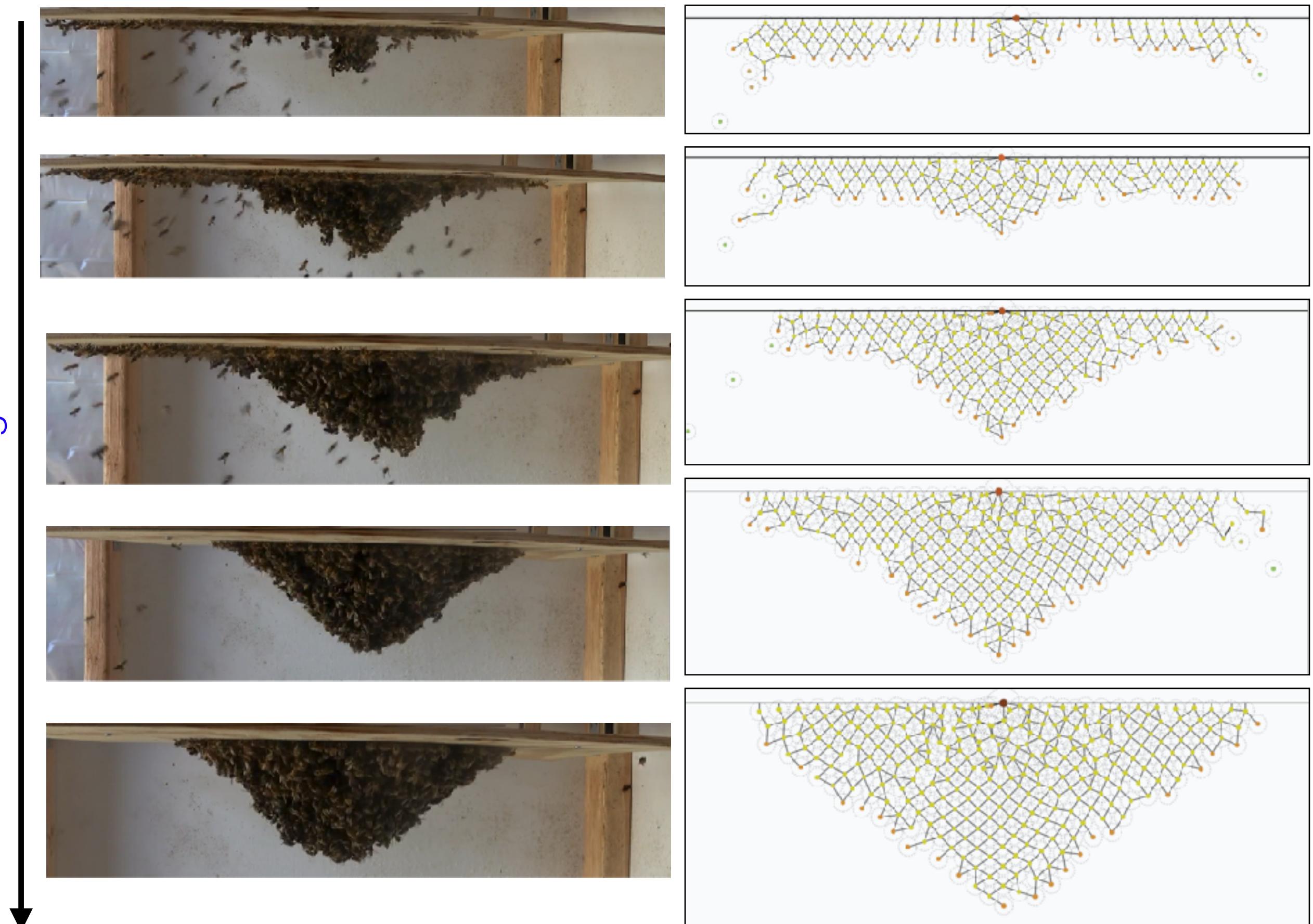


Same power law governs the smallest and largest swarms, with the weight supported by each layer scaling with the mass of each layer to the 1.5 power.

O. Shishkov et al., *Scientific reports*, 12(1), 2022

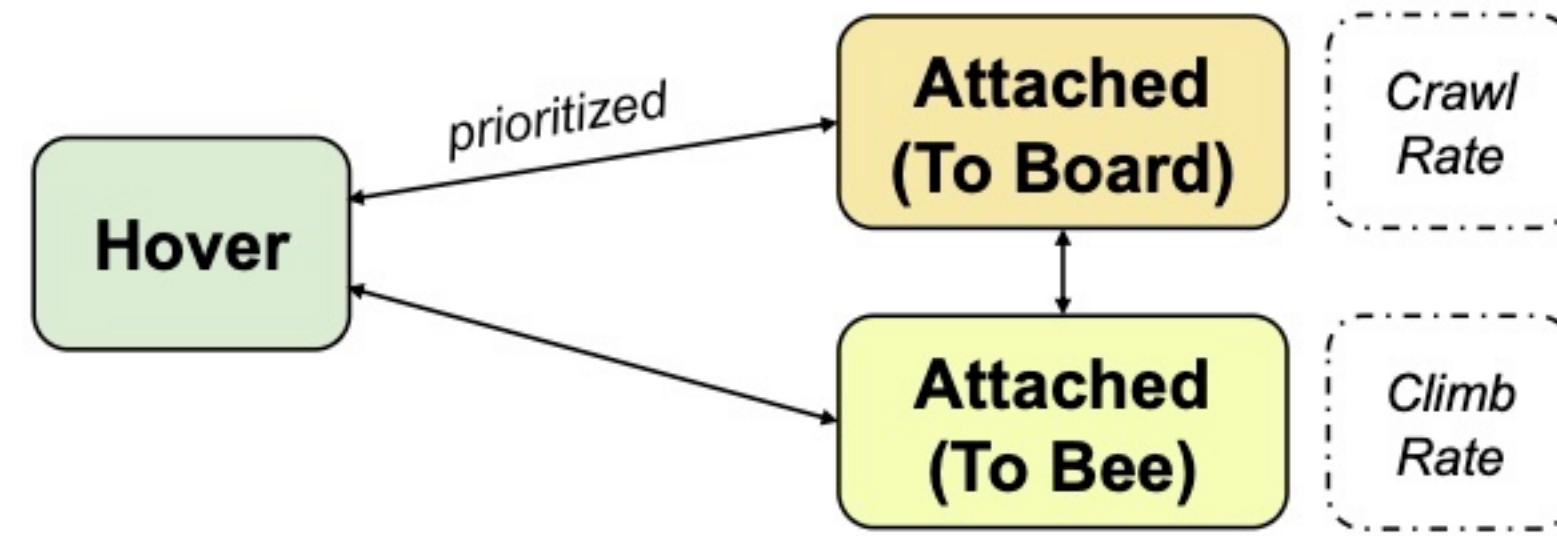
Increasing time

Self-assembly: Swarming of honeybee



Manuscript in preparation

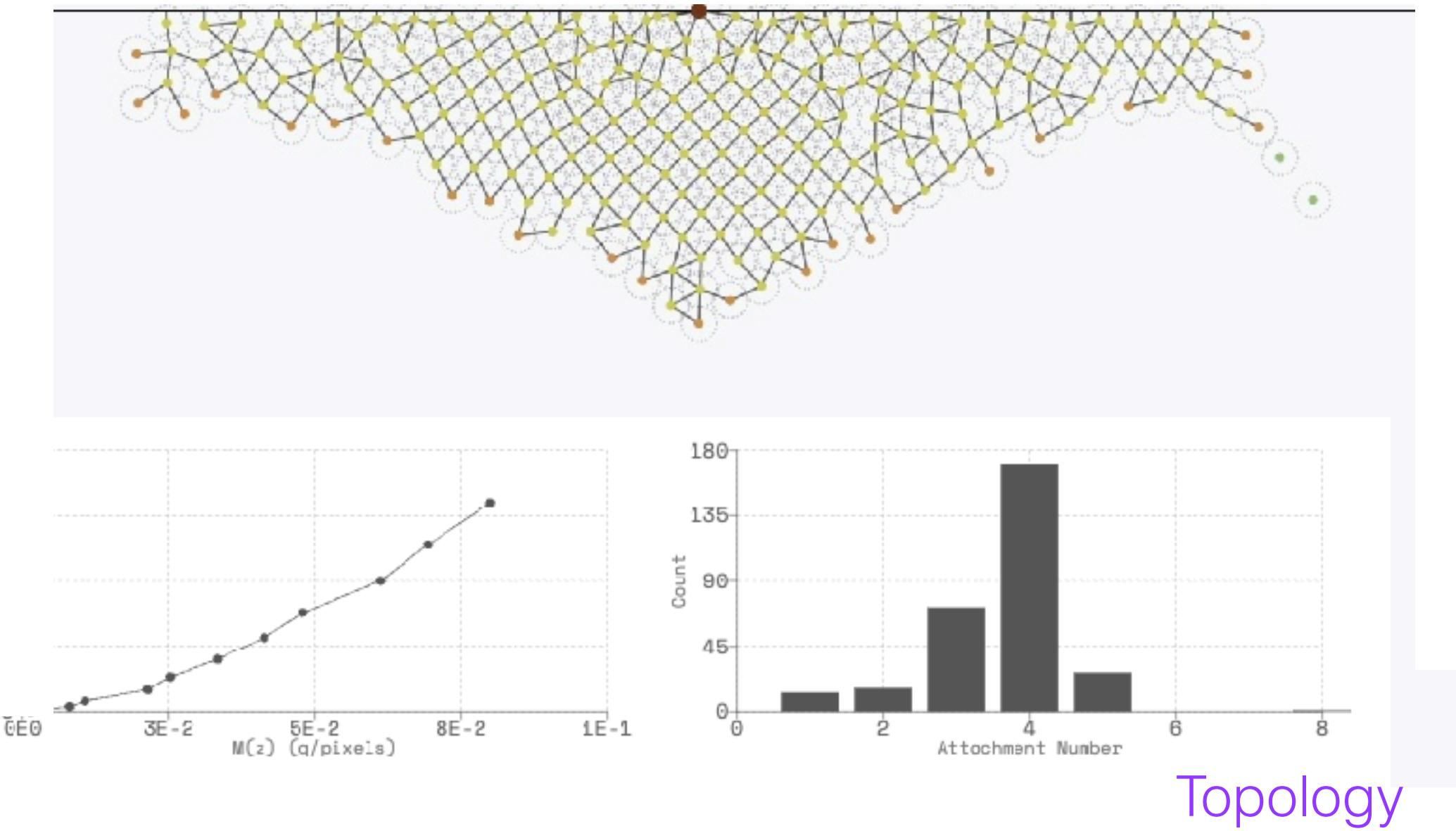
# Computational Model of Honeybee Swarms



Two parameters in model:

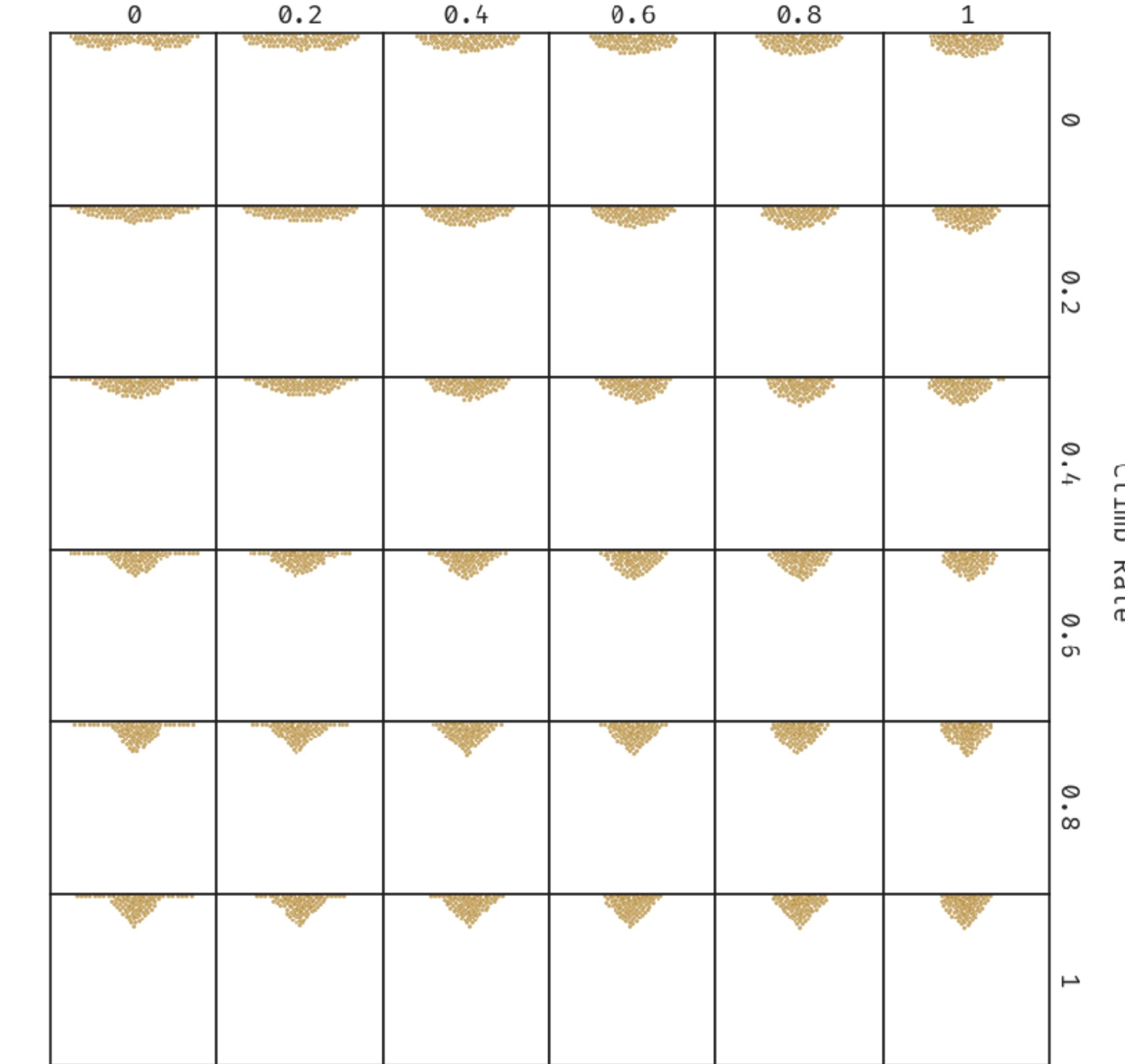
- (1) Crawl rate - tendency to pack
- (2) Climb rate - tendency to move up the swarm

Geometry and topology of complex force network



Topology

Phase Diagram varying 2 parameters

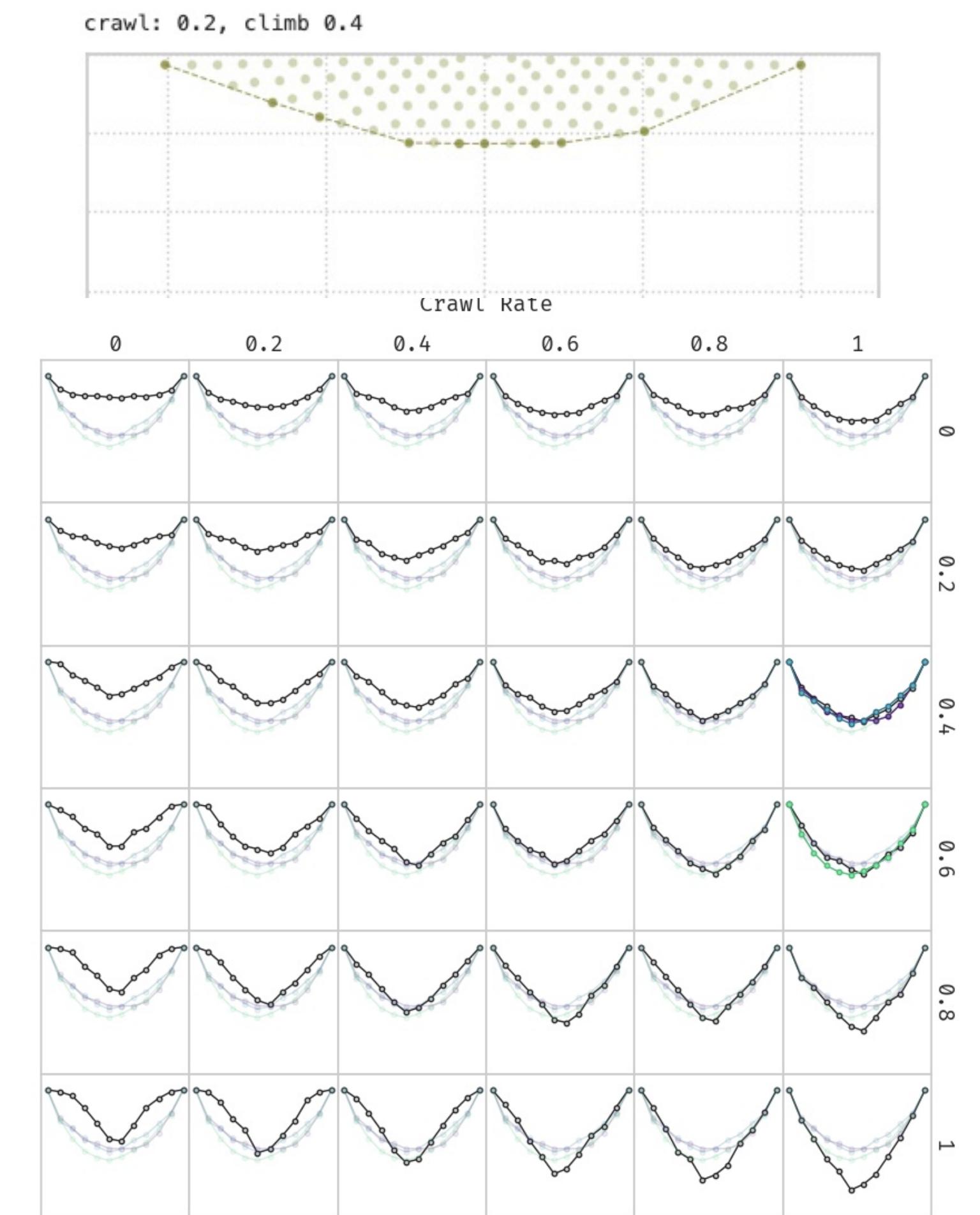
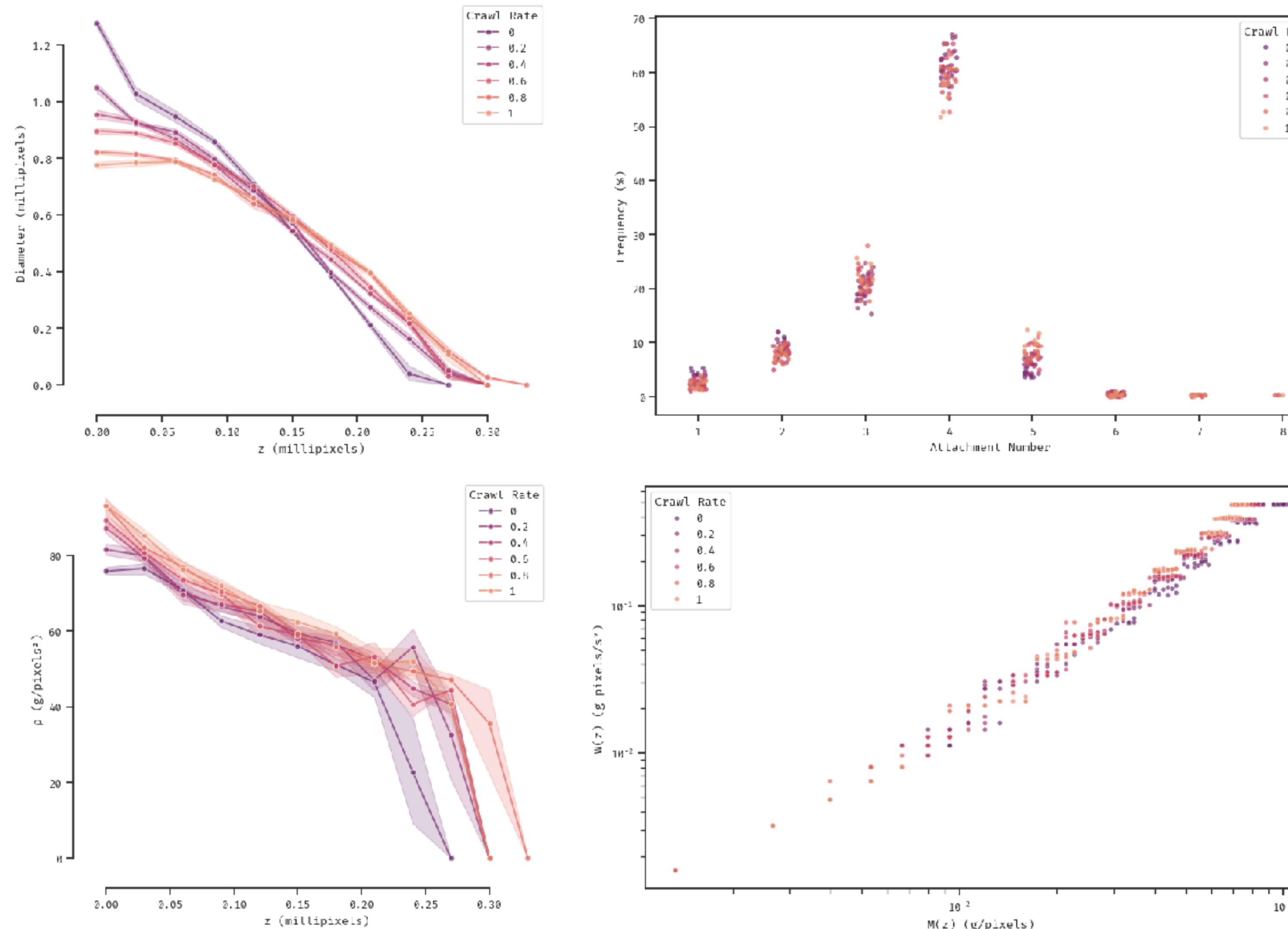


Manuscript in preparation

Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

# Swarming Behaviors in Honeybees



Geometry

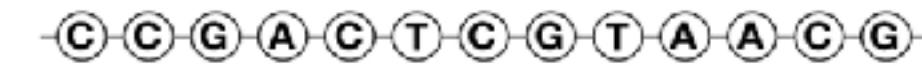
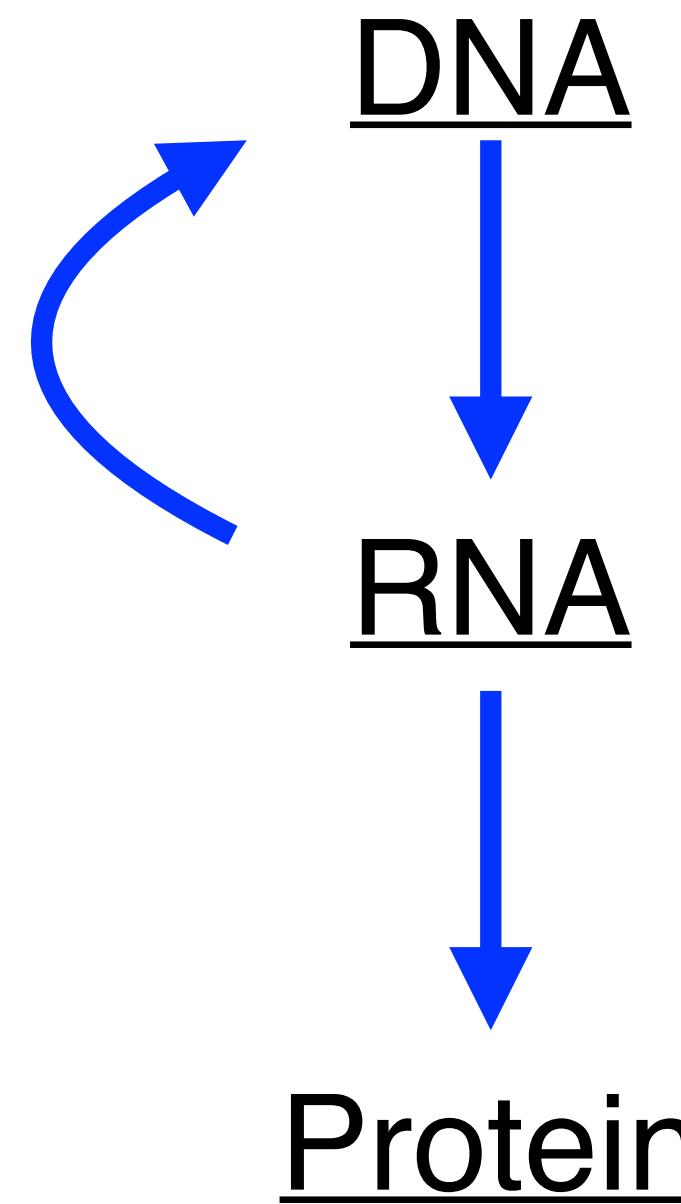
Manuscript in preparation

Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

# Central Dogma of Biology (Physicist's perspective)

"You don't want to mess with your hard disk, but you can have as many copies of RNA as you want"



Information storage  
~hard disk

$L \sim 10^9$  nucleotides  
Complexity  $\sim 4^L$



~floppy disks to give out

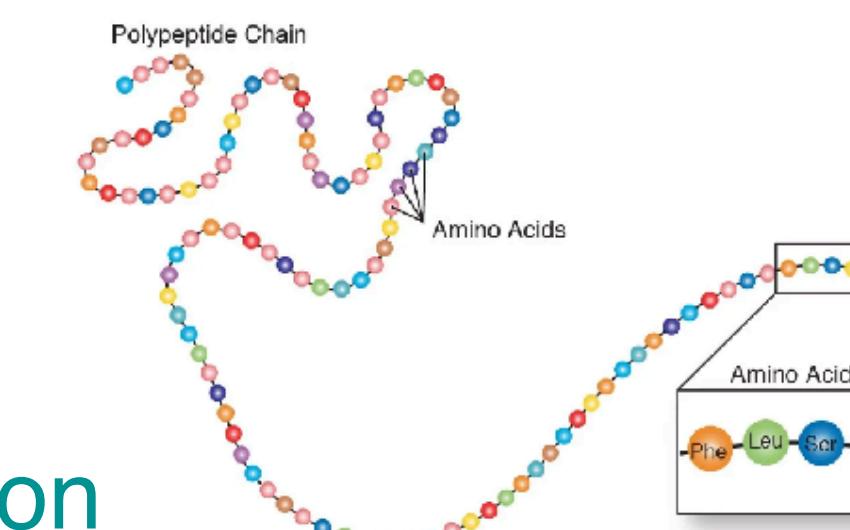
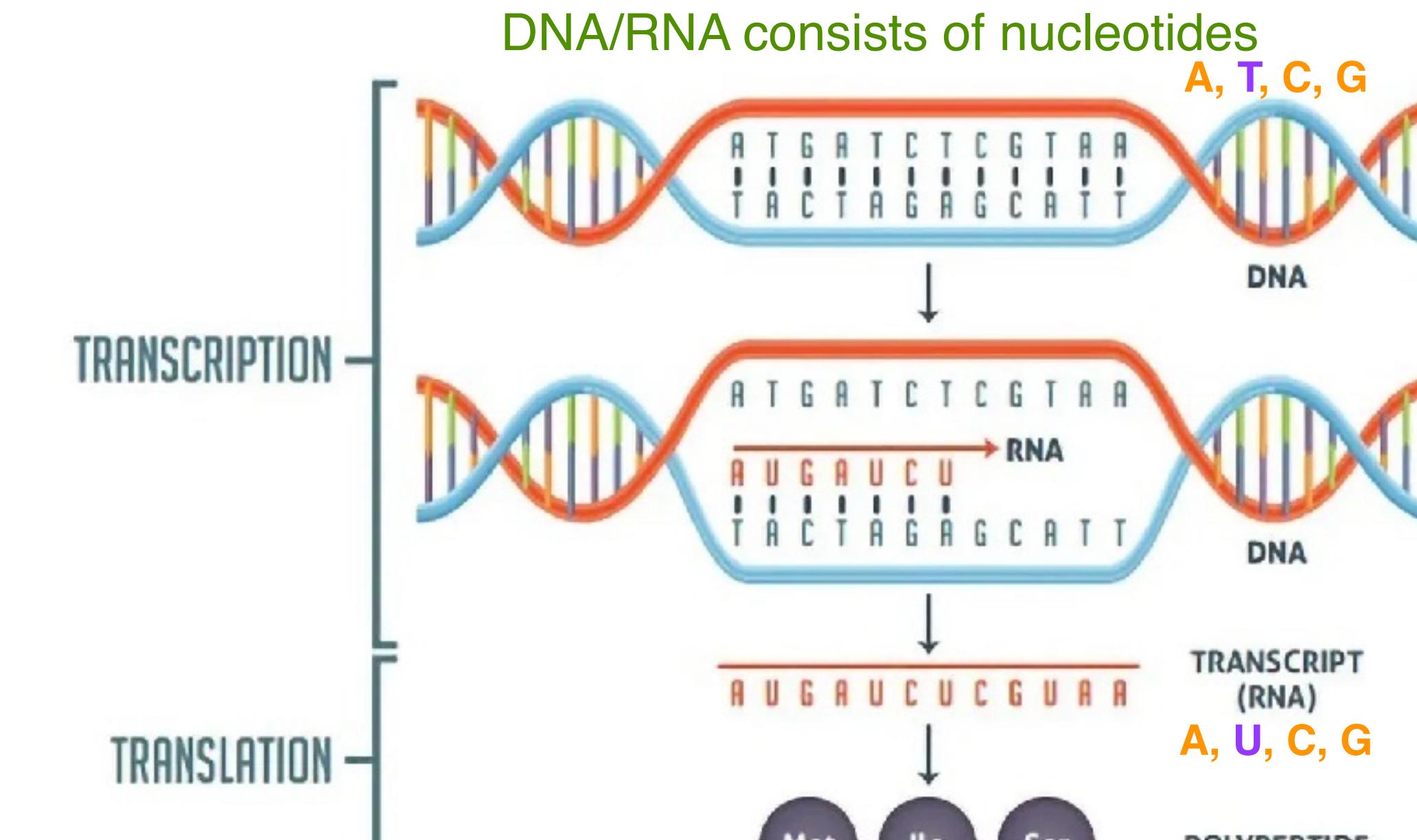
$L \sim 10^3$  nucleotides  
Complexity  $\sim 4^L$

Function (enzyme, motor,  
muscle,...)

$L \sim 500$  amino acids  
Complexity  $\sim 20^L$

RNA/Protein folds into many shape:  
► Different shape → different message/function

Shape: space curve to first approximation



Proteins consist of amino acids

## Amino Acids

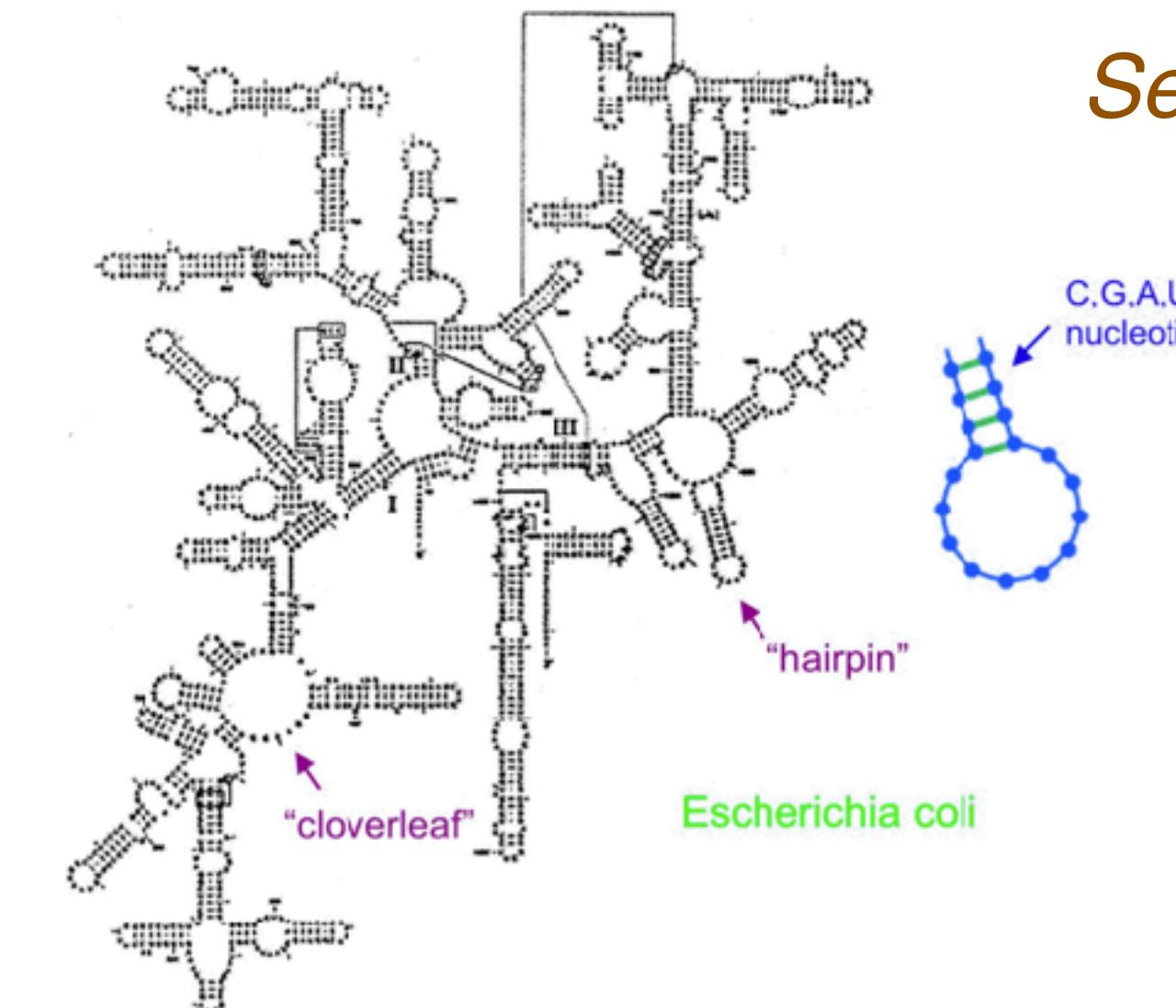
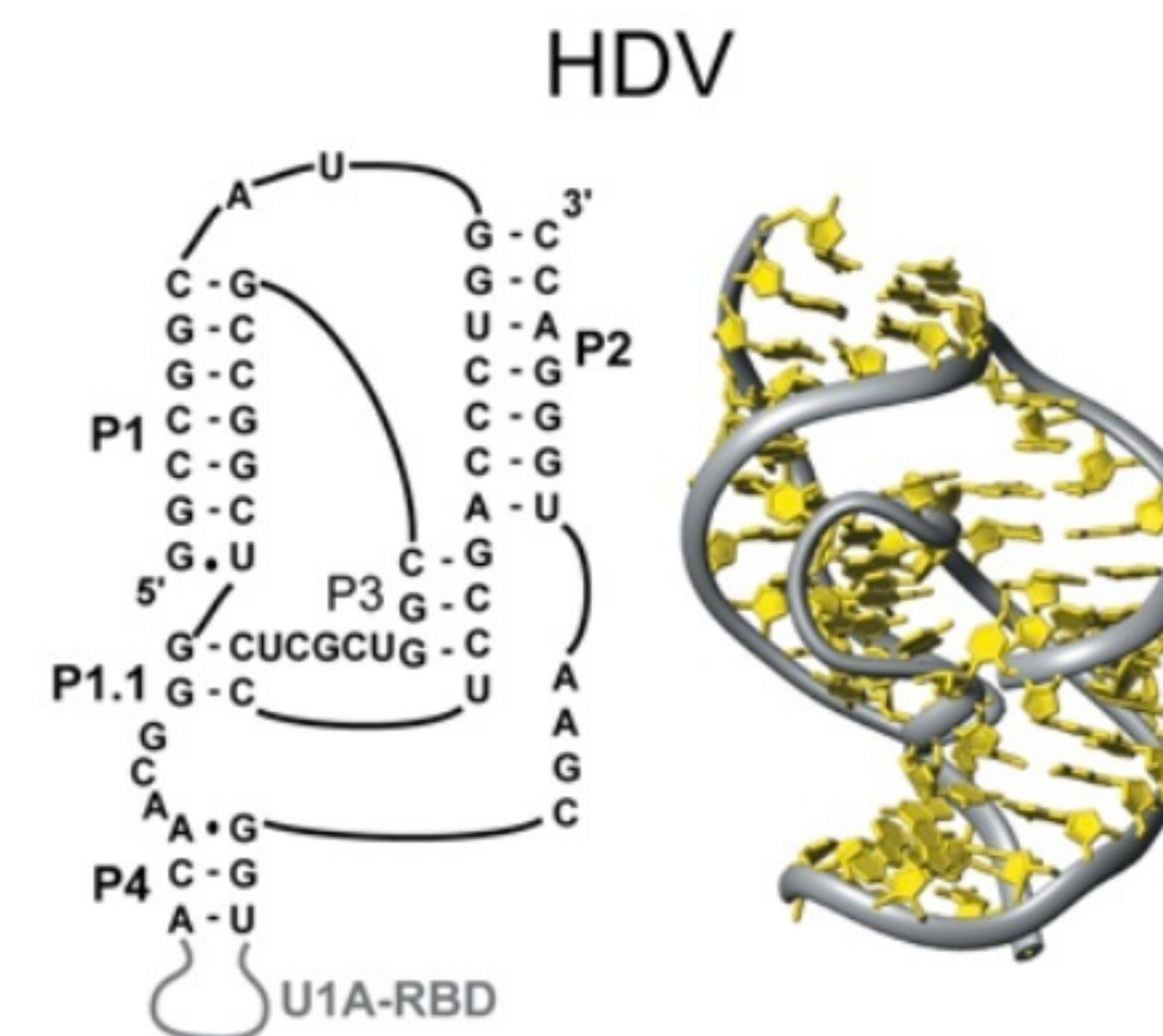
Ala: Alanine	Gln: Glutamine	Leu: Leucine	Ser: Serine
Arg: Arginine	Glu: Glutamic acid	Lys: Lysine	Thr: Threonine
Asn: Asparagine	Gly: Glycine	Met: Methionine	Trp: Tryptophane
Asp: Aspartic acid	His: Histidine	Phe: Phenylalanine	Tyr: Tyrosine
Cys: Cysteine	Ile: Isoleucine	Pro: Proline	Val: Valine

20 amino acids, 9 essential

# *RNA Folding Problem*

- Watson & Crick thought that RNA is "merely" a "passive carrier of information" from DNA to protein ("messenger RNA")
  - Important discovery of few decades (revenge!): RNA plays a crucial enzymatic role.
  - Biochemistry incredibly complicated, but to first approximation: shape of molecule ("lock and key")
  - By the way, biologists working on the origin of life now generally believe that it started with an "RNA world. DNA came later.
  - RNA folding problem: predict an RNA molecule's 3dim functional structure (secondary and tertiary) from its 1D sequence of nucleotides

# *RNA: Given the message, what is its shape?*



# *Sequence space → Shape space*

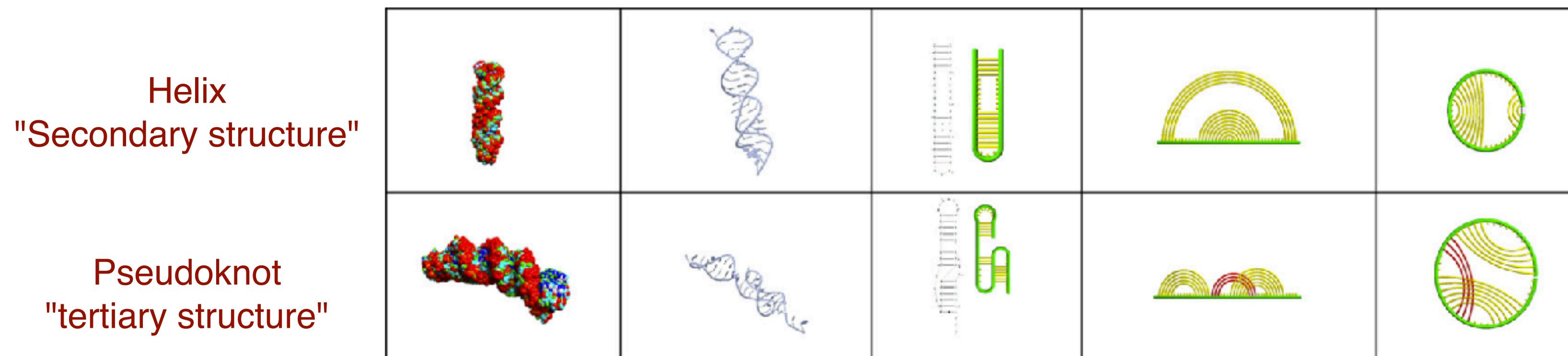
*D.W. Staple & S.E. Butcher,  
PLoS biology, 3(6), 2005*

*A. Zee, KITP Chalk Talk, 2005*

# Topological Classification of RNA Structures

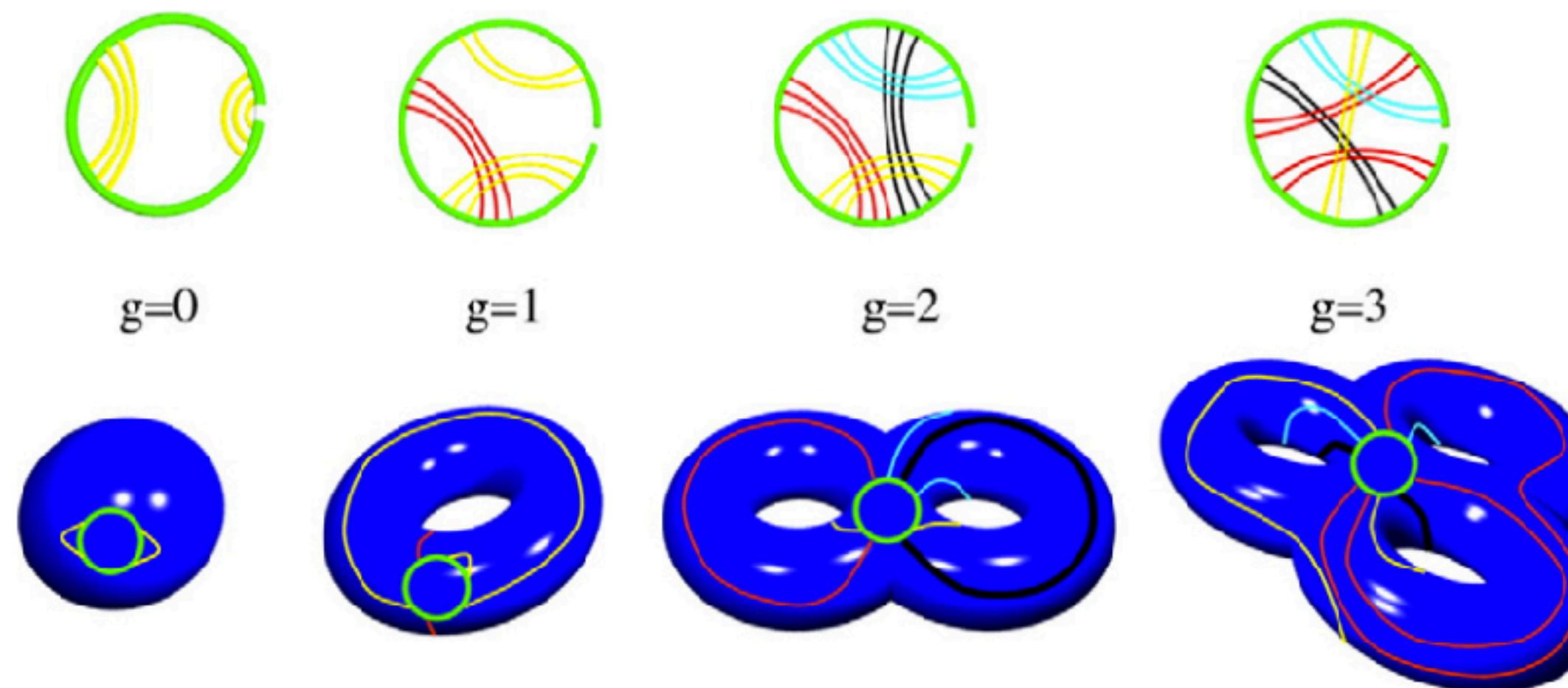
Primary structure of an RNA refer to its chemical sequence, e.g., AUCGGUA

Secondary structures: planar diagrams (could be drawn on a sphere without crossing).



*M. Bon et al. Jour. of Mol. Bio. 379(4), 2008*

Tertiary structures: Diagrams that could be drawn on a torus without crossing (genus=g)



*RNA folding problem has connection to random matrix theory and large N quantum chromodynamics*

*H. Orland & A. Zee, Nuclear Physics B 620 (3), 2002*

# Prediction of RNA Secondary Structure

**Mfold** predicts the optimal and suboptimal secondary structures of single stranded nucleic acids by energy minimization method. In the **nearest-neighbor energy model**, values of free energy contribution are assigned to each closed loop (e.g. stackings, bulge loops, interior loops, multi-loops) in RNA secondary structure, and the total free energy of the structure is calculated as the **sum of all the free energy contribution** from the loops

D.H. Mathews et al., *Journal of molecular biology* 288(5), 1999  
 M. Zuker, *Nucleic acids research*, 31(13), 2003



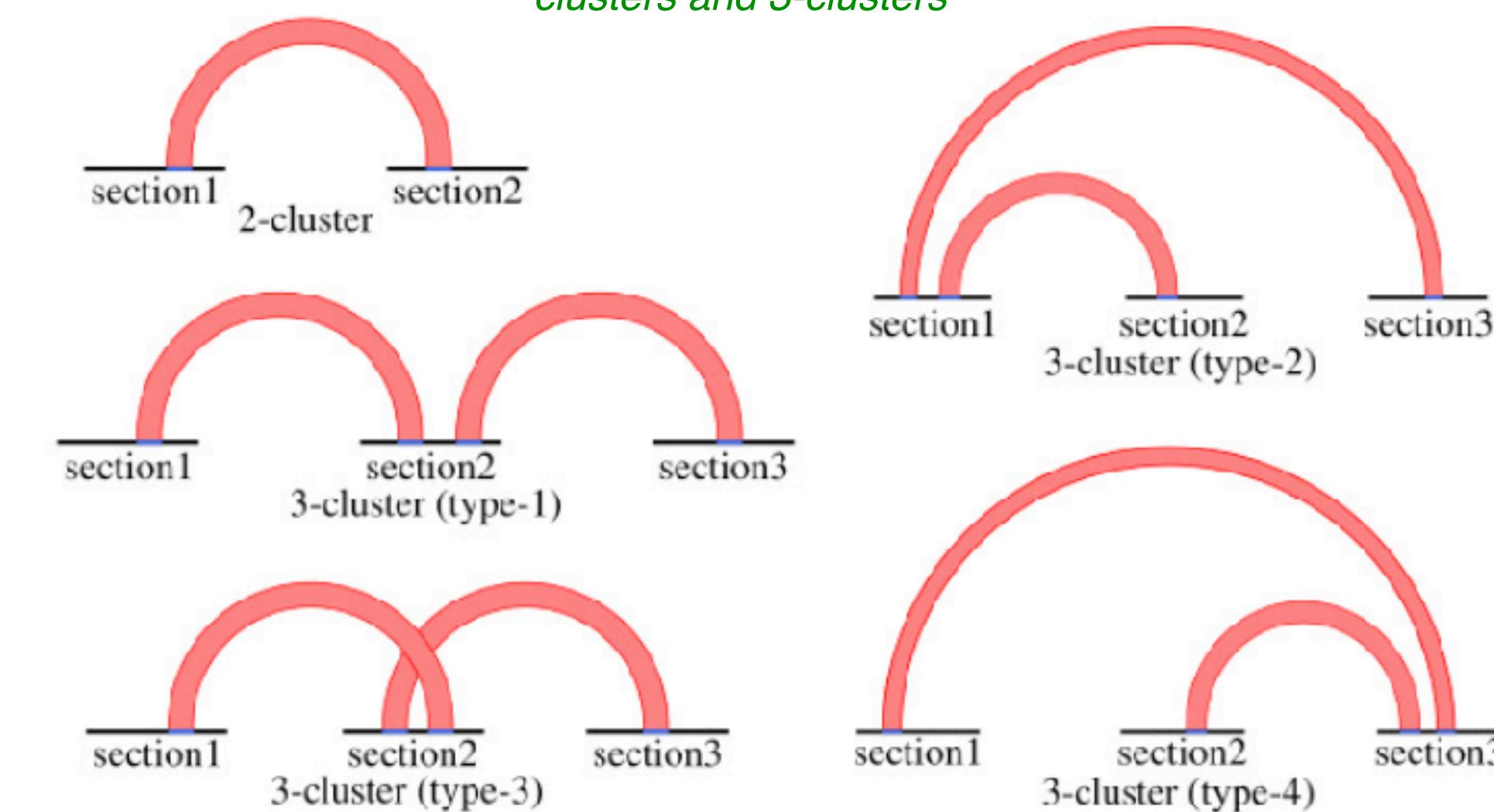
FE contribution of the pseudoknot  
 $= -(\text{hairpin loop (1)}) - (\text{hairpin loop (2)}) + (\text{multi loop (3)}) + (\text{stacking (4)}) + (\text{bulge loop (5)}) + (\text{stacking (6)}) + (\text{multi loop (7)}) + (\text{stacking (8)}).$

*Doesn't work too well with pseudoknots, which are defined by crossing patterns within the global RNA topology, not as isolated elements.*

*We modify energy model!*

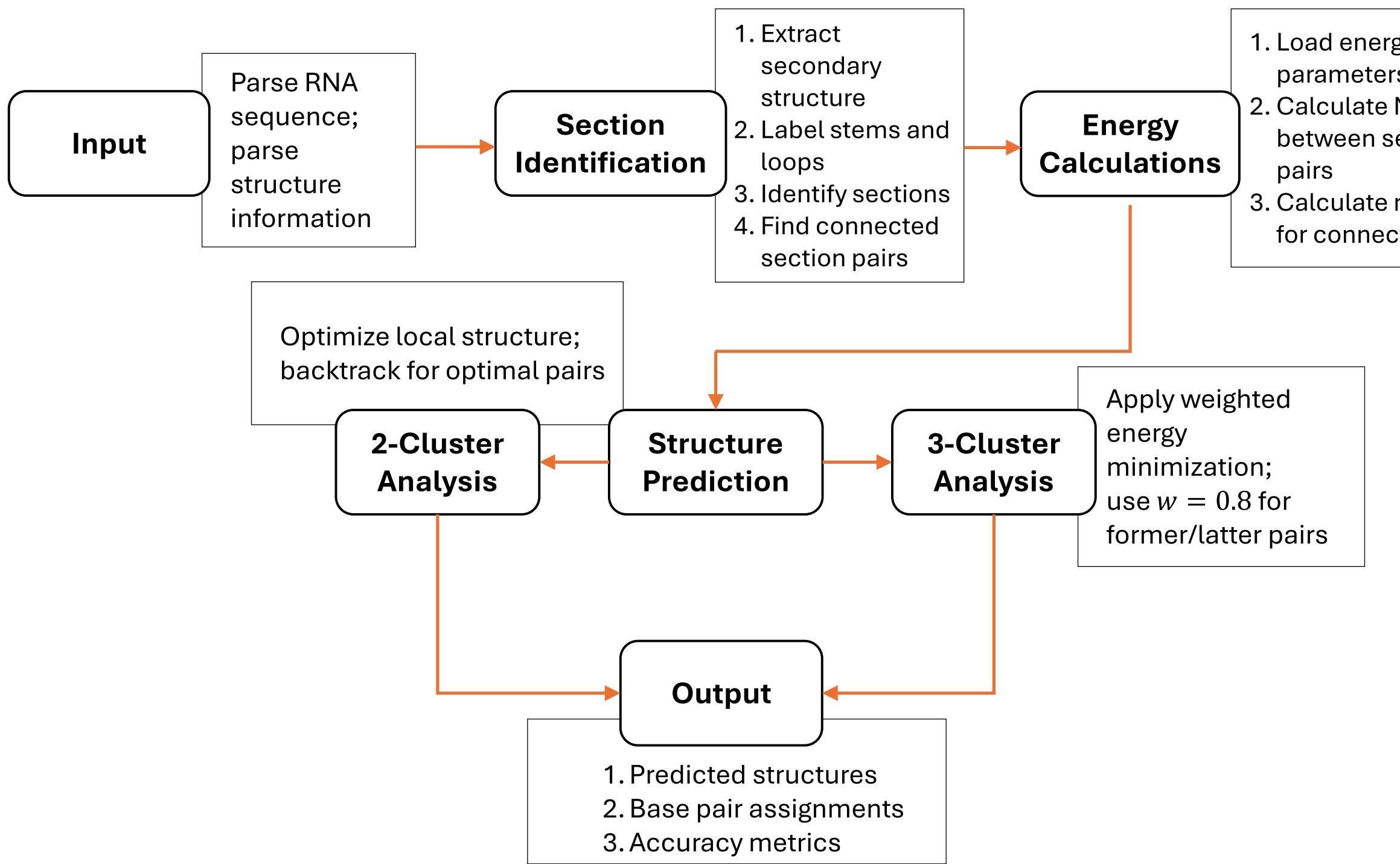
FE contribution of the pseudoknot  
 $= (\text{stacking (4)}) + (\text{bulge loop (5)}) + (\text{stacking (6)}) + (\text{multi loop (7)}) + (\text{stacking (8)}).$

*Topologically distinct pseudoknot structures of both 2-clusters and 3-clusters*



*A section is a contiguous subset of bases in the RNA sequence that are not involved in base pairing,*

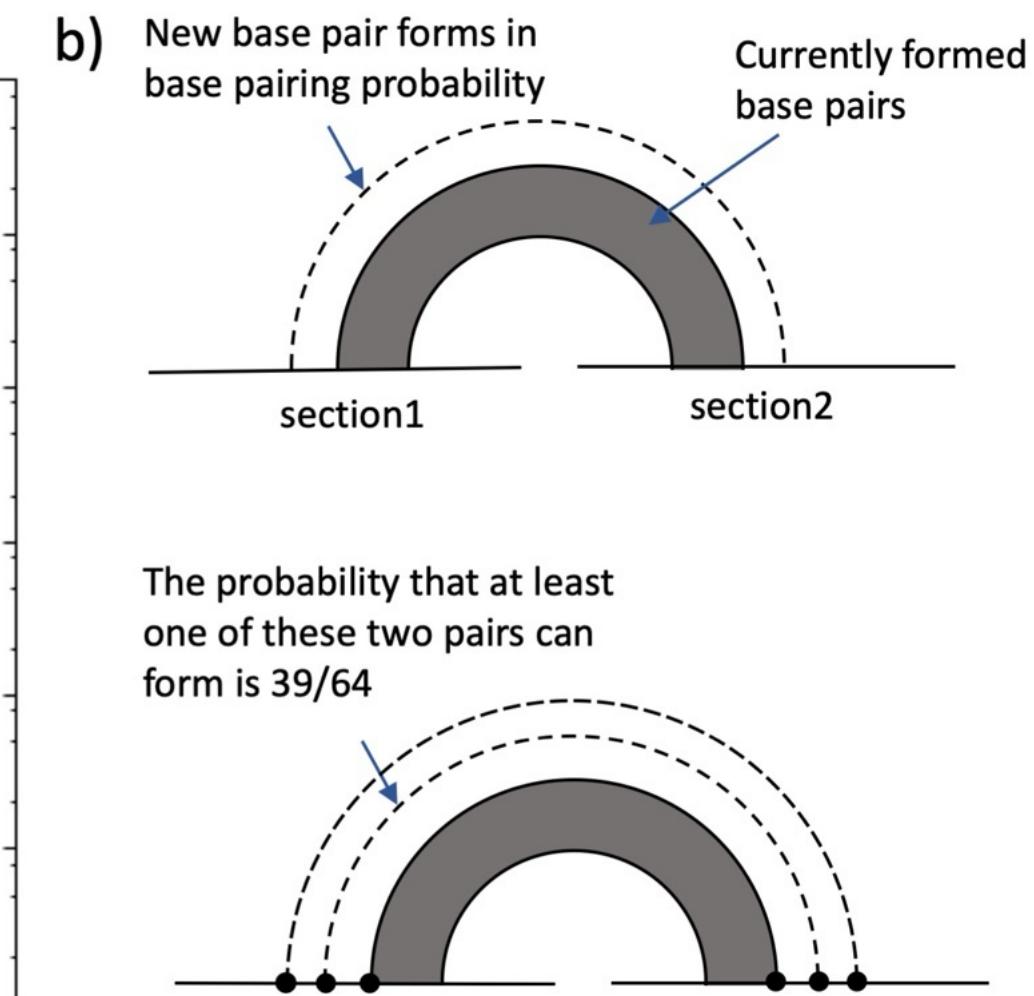
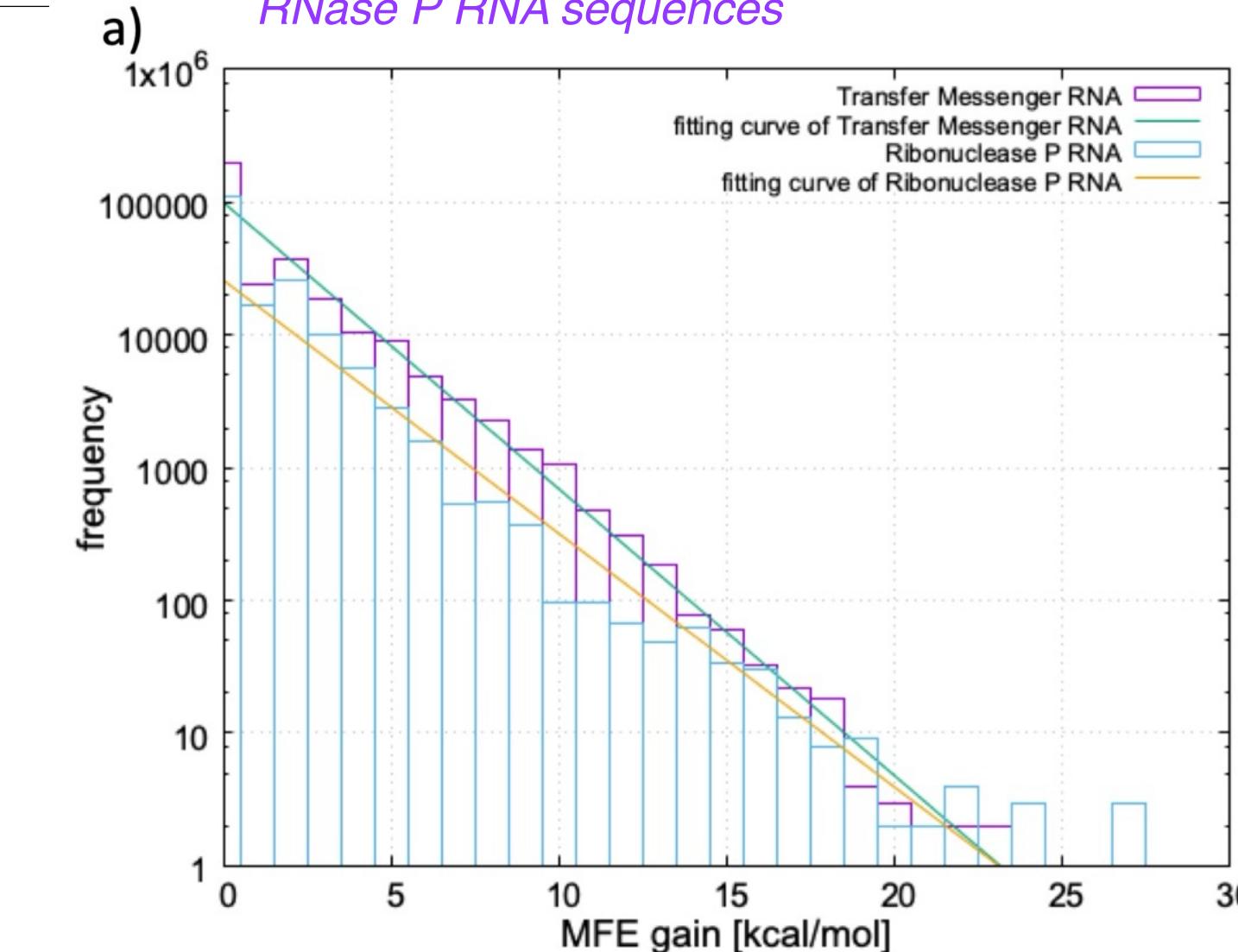
# Hierarchical Analysis of RNA Secondary Structures



> 80% of all 2-clusters demonstrate free energy gain > 0.8 × MFE gain

Parameter	tmRNA	RNase P RNA
The number of all 2-clusters	2,337	534
The number of all 3-clusters	110	1
The number of 2-clusters for which MFE = real FE	1,605 (68.7%)	213 (39.9%)
The number of 2-clusters for which real FE < 0.8 × MFE	1,972 (84.4%)	438 (82.0%)
Sensitivity of base pair prediction for 2-clusters	0.9159	0.9087
PPV of base pair prediction for 2-clusters	0.8400	0.8056

Histogram showing exponential decay of MFE gain frequency for all section pairs in tmRNA and RNase P RNA sequences



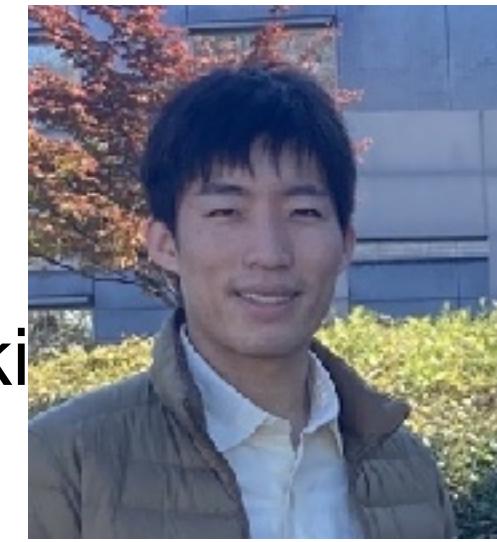
Sensitivity consistently exceeds PPV, indicating our method tends to predict excess base pairs.

Parameter	Former section pair	Latter section pair
Total number	110	110
The number of section pair for which real FE = MFE	55 (50.0%)	10 (9.1%)
The number of section pair for which real FE < 0.8 × MFE	91 (82.7%)	18 (16.3%)
Sensitivity of base pair prediction	0.8343	0.5529
PPV of base pair prediction	0.7164	0.2568

R. Masuki, D. Liew & E.H. Yong, PLOS Comp Bio, accepted

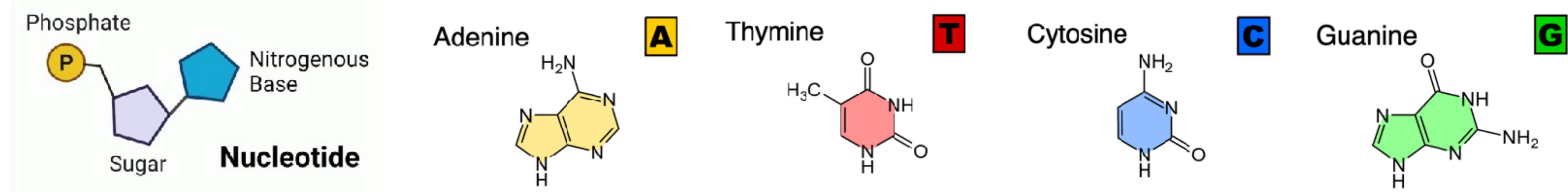
Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics



# DNA – basic building block of life

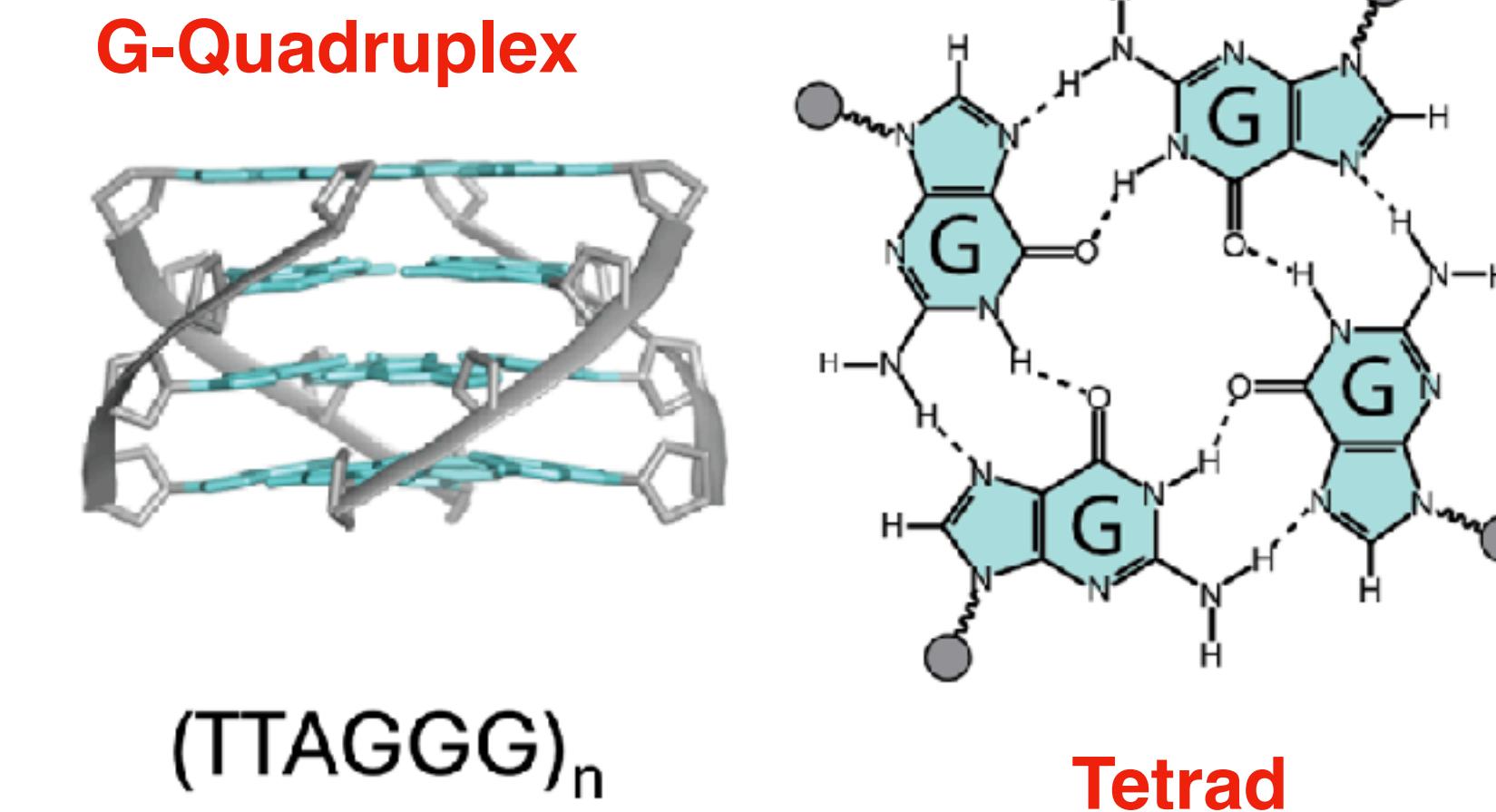
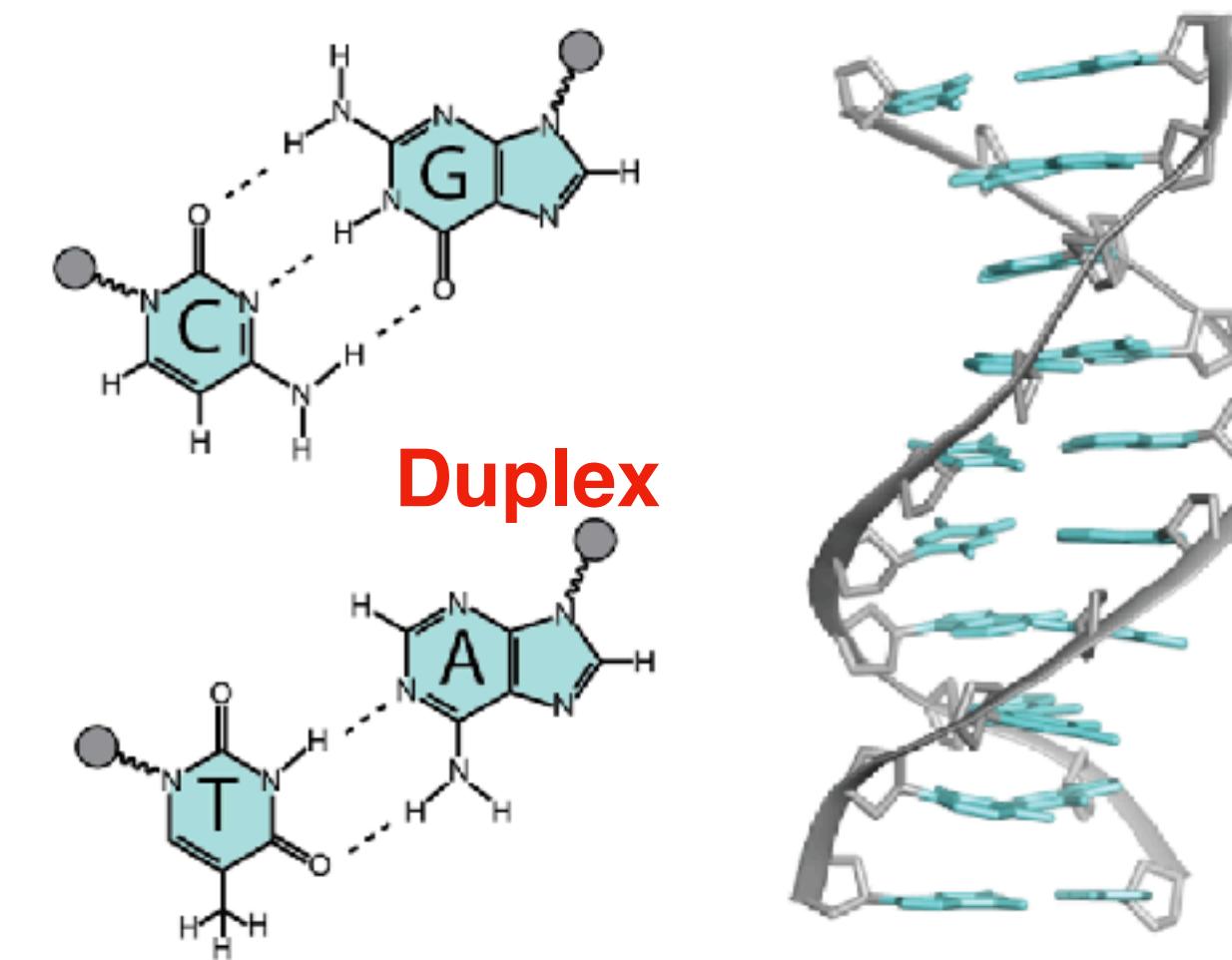
**Deoxyribonucleic acid (DNA)** is made up of four nucleotide bases: adenine (A), thymine (T), cytosine (C), and guanine (G).



*DNA: sequence of  $L$  nucleotides has complexity is  $4^L$*

DNA composed of two polynucleotide chains that coil around each other to form a **double helix**.

Nucleic acids rich in guanine (G-rich) can form stable secondary structures known as **G-quadruplexes**.



Lipps H. J., Rhodes D. Trends. Cell. Biol., 2009  
Bochman et al., Nat. Rev. Genet., 2012

Neidle, S. Nat. Rev. Chem., 2017

Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

# Biological Significance of G-Quadruplex (G4)

**G4 structures** have been observed in many crucial regions of the **human genome** such as telomeres, oncogenic promoters and mutational hot spots, and are related to **critical cellular processes** such as recombination, replication, transcription, and translation.

**Elizabeth Blackburn, Carol Greider and Jack Szostak** received the **Nobel Prize in Physiology or Medicine in 2009** for discovering the process by which chromosomes are protected by telomeres and the enzyme telomerase. Due to its important role in cell biology, telomere has emerged as a promising target for **anticancer drug strategy**.



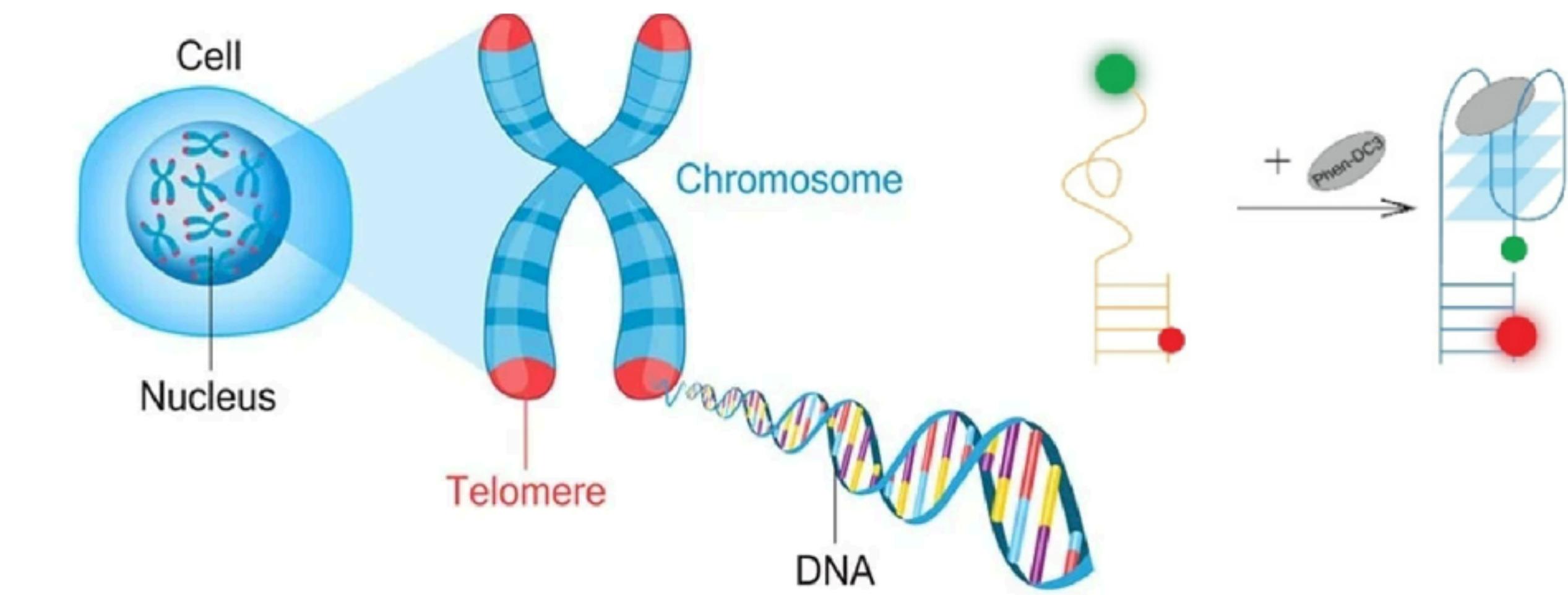
Elizabeth  
Blackburn



Carol Greider

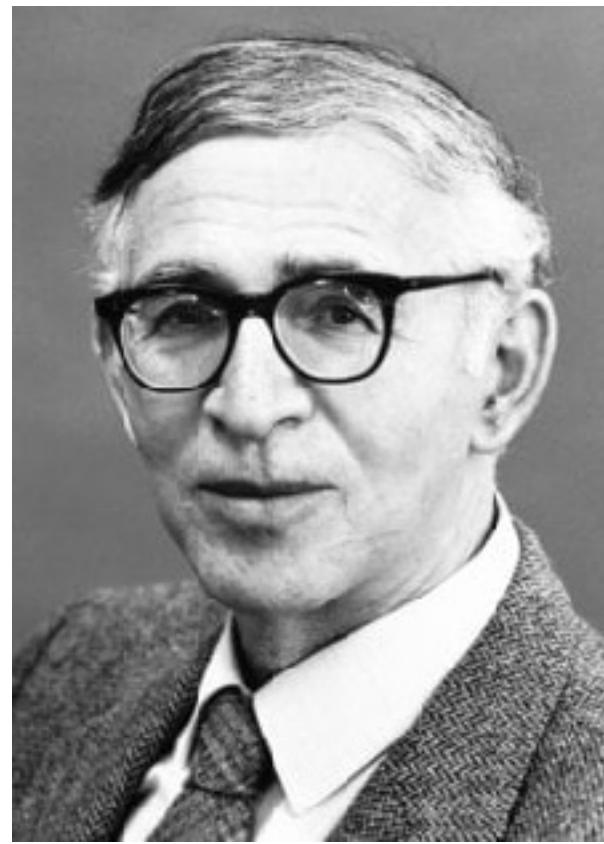


Jack Szostak



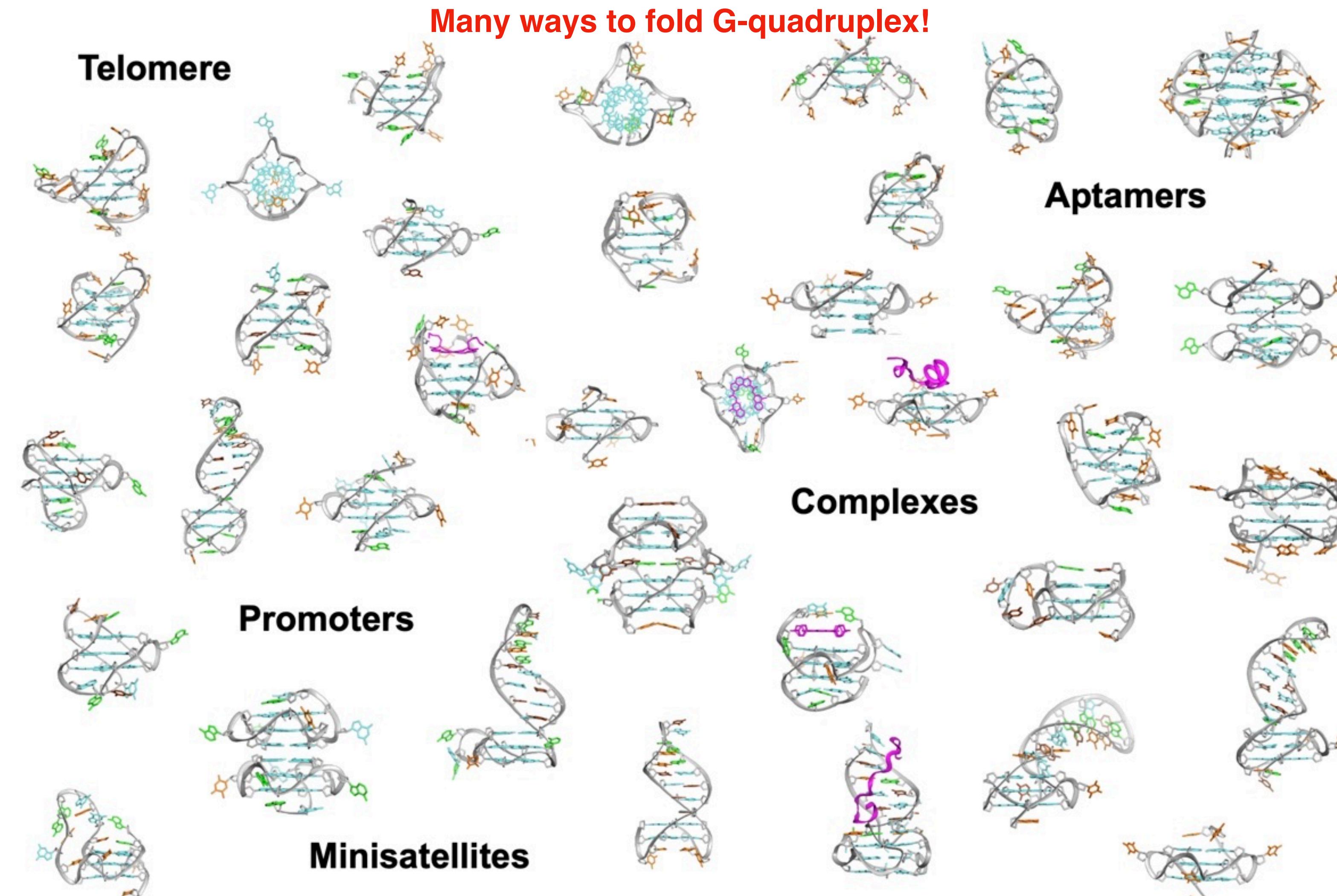
The **G4** structures of the telomere stop the ends of chromosomes from degradation, much like the **plastic tips** on the ends of shoelaces.

# Polymorphism of G-quadruplexes



**Sir Aaron Klug**  
1982 Nobel prize in  
chemistry

*"If G-quadruplexes form so readily in vitro, Nature will have found a way of using them in vivo."*

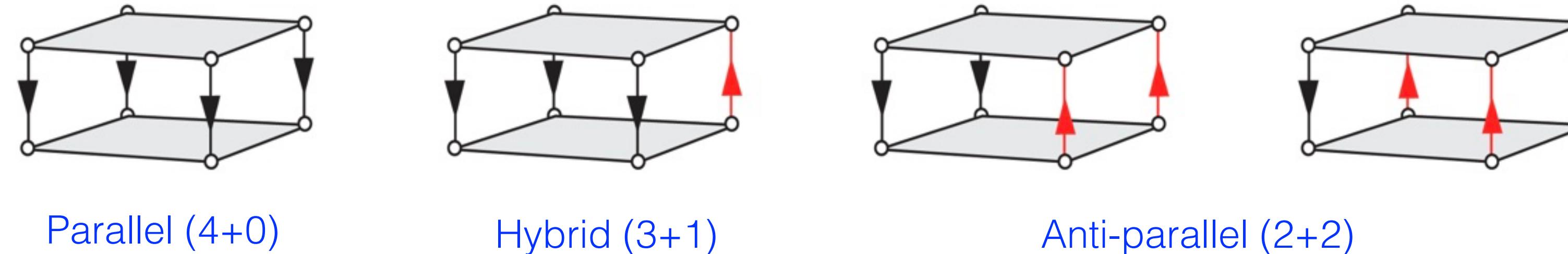


*Bedrat, A., Lacroix, L. and Mergny, J.L., Nucleic acids research, 44(4), 2016*

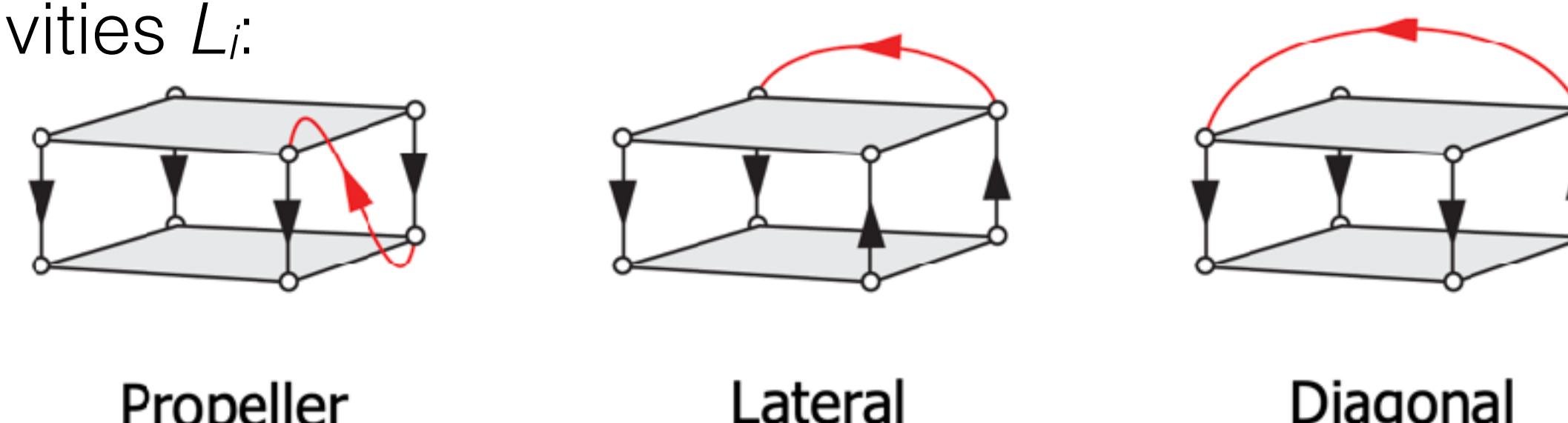
*Karsiotis, A.I., O'Kane, C. and da Silva, M.W., Methods, 64(1), 2013*

# Structural elements of G-quadruplexes

(1) Orientation of the strands:

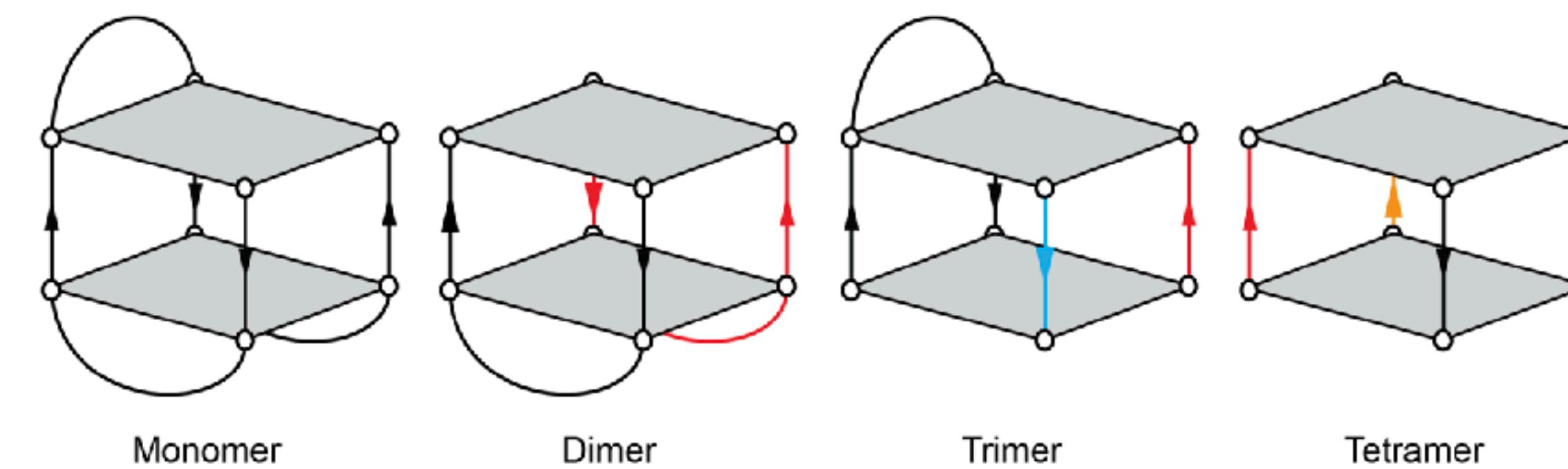


(2) Different loop connectivities  $L_i$ :



*Phan, A.T., Kuryavyi, V. and Patel, D.J., Current opinion in structural biology, 16(3), 2006*

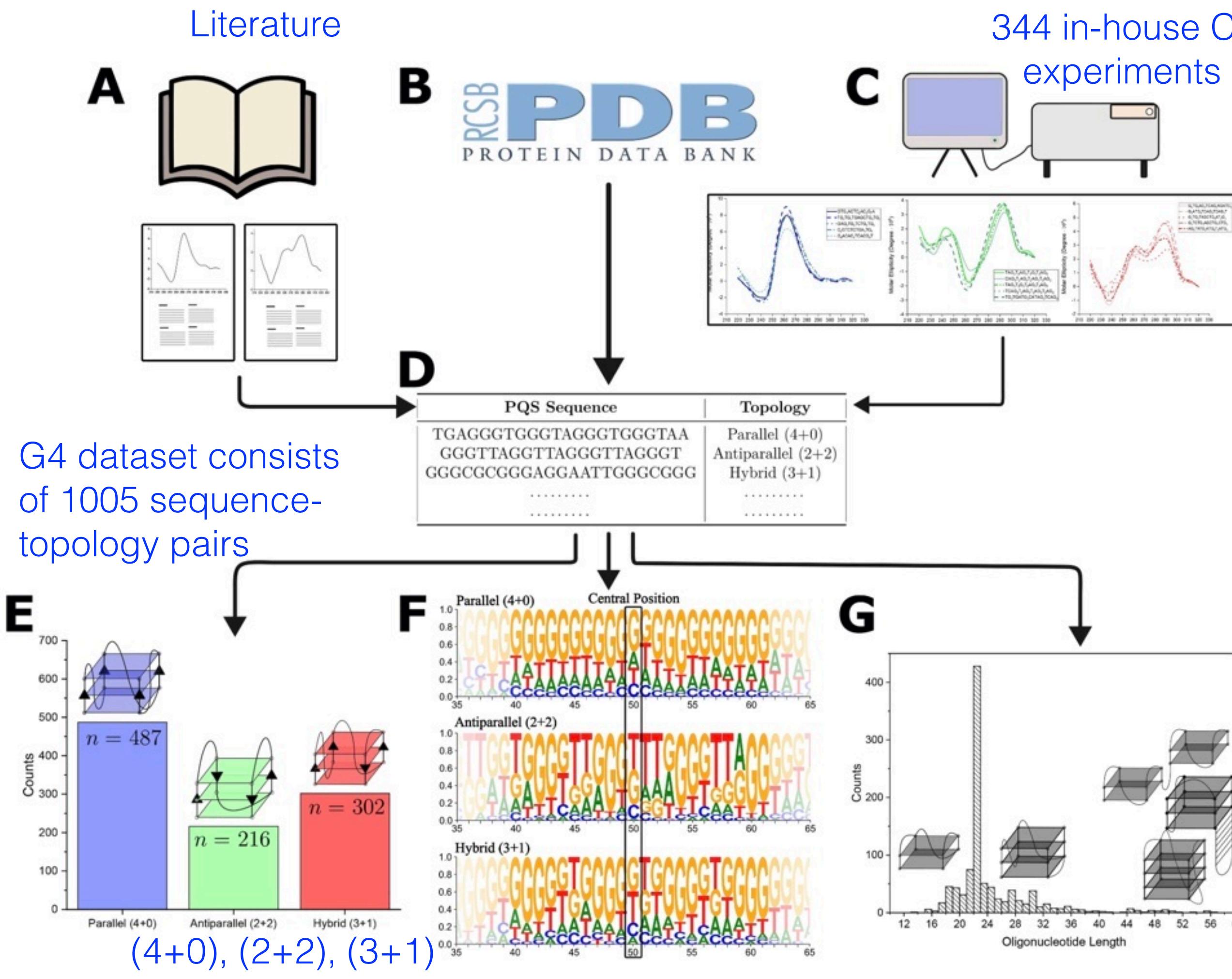
(3) Molecularity



*Phan, A.T., The FEBS journal, 277(5), 2010*

Other factors: Syn/anti glycosidic conformation of guanines, stacking orientation, etc

# Machine Learning Prediction of DNA G4 Folding Topology

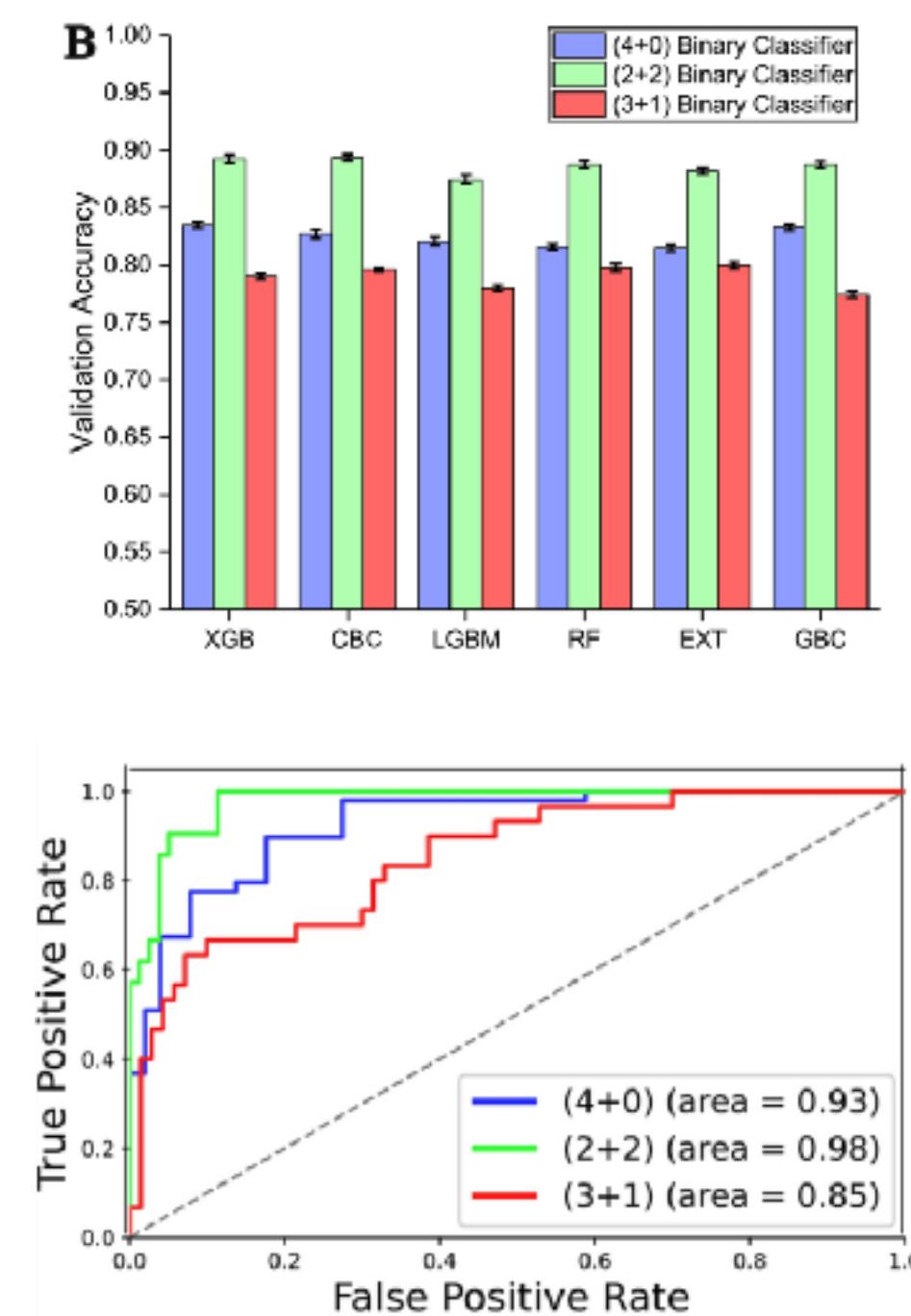
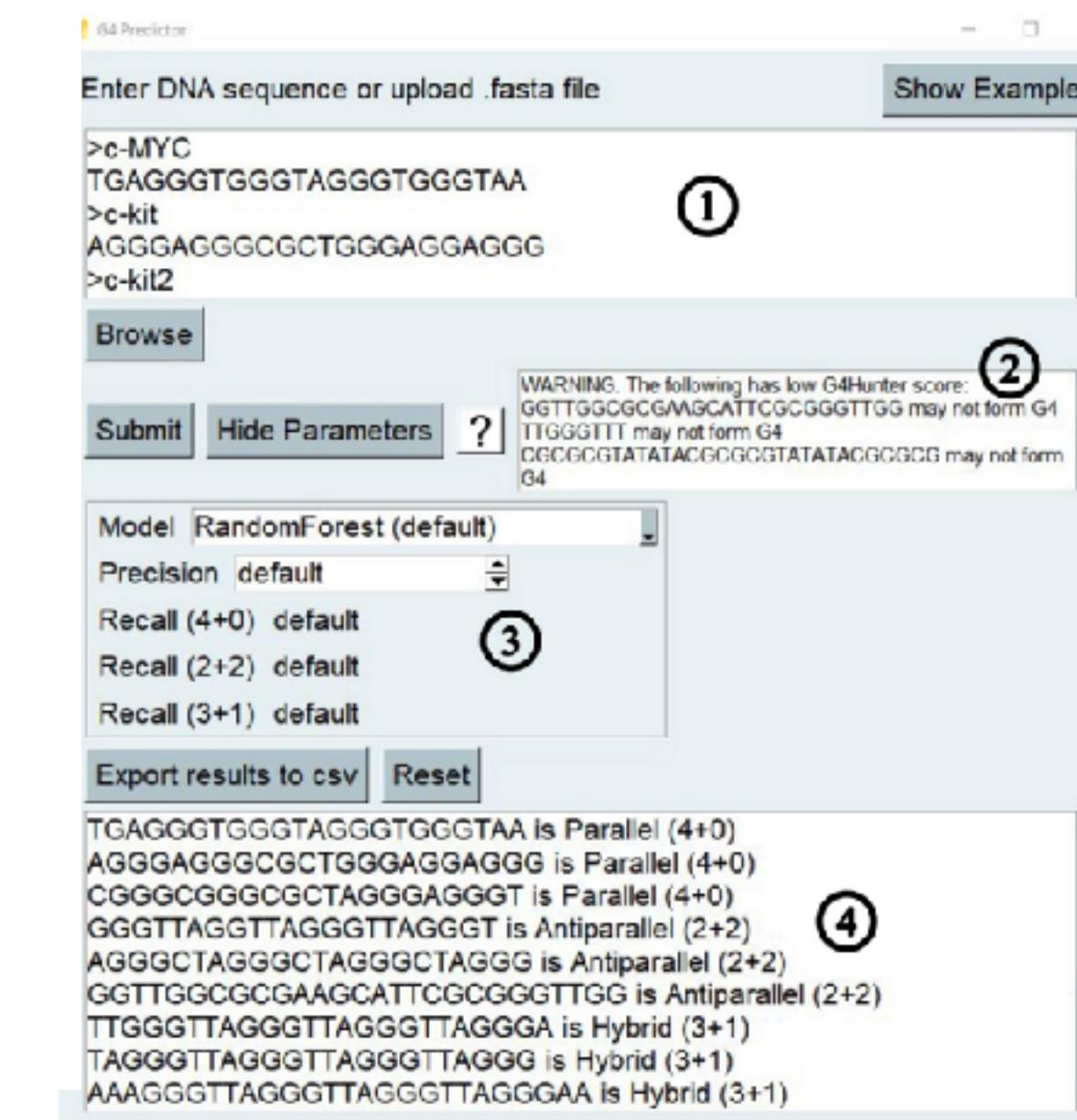


AI for Science

**G4ShapePredictor (G4SP)** is designed to accurately **predict G4 folding topologies** in potassium ( $K^+$ ) buffer based on its primary sequence and is able to incorporate a threshold optimization strategy allowing users to maximise precision.



## G4ShapePredictor



D. Liew, Z.W. Lim, and E.H. Yong, *Scientific Reports*, 14(1), 2024

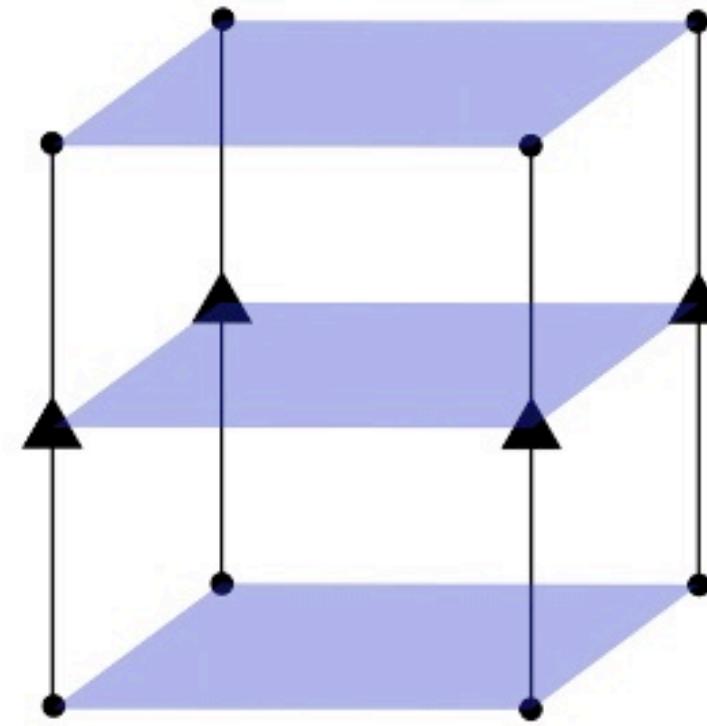
Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

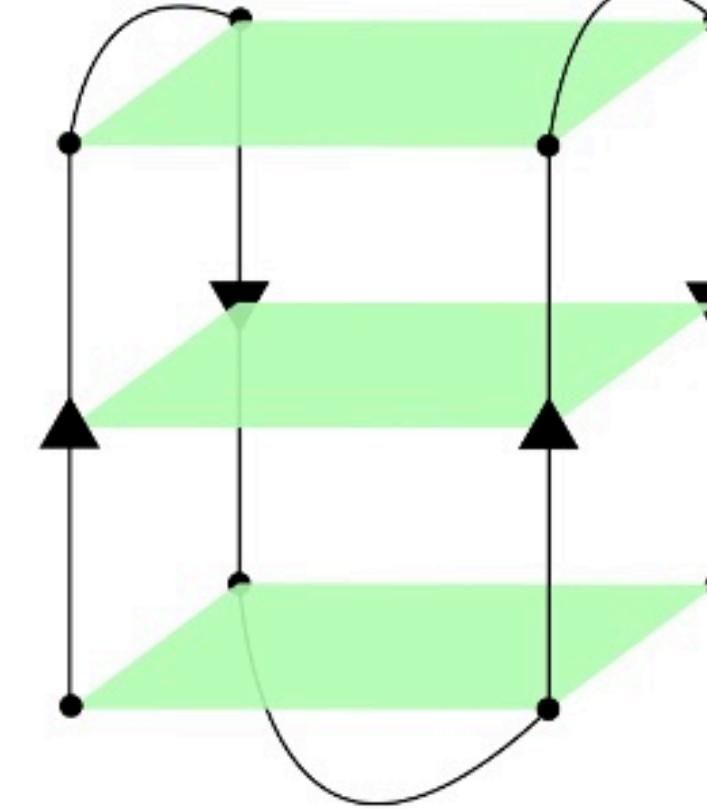
# G4SP suggests Topological Motifs to Identify Structures

*AI for Science*

RGNG<sub>3</sub>N<sub>1-9</sub>G<sub>3</sub>NR

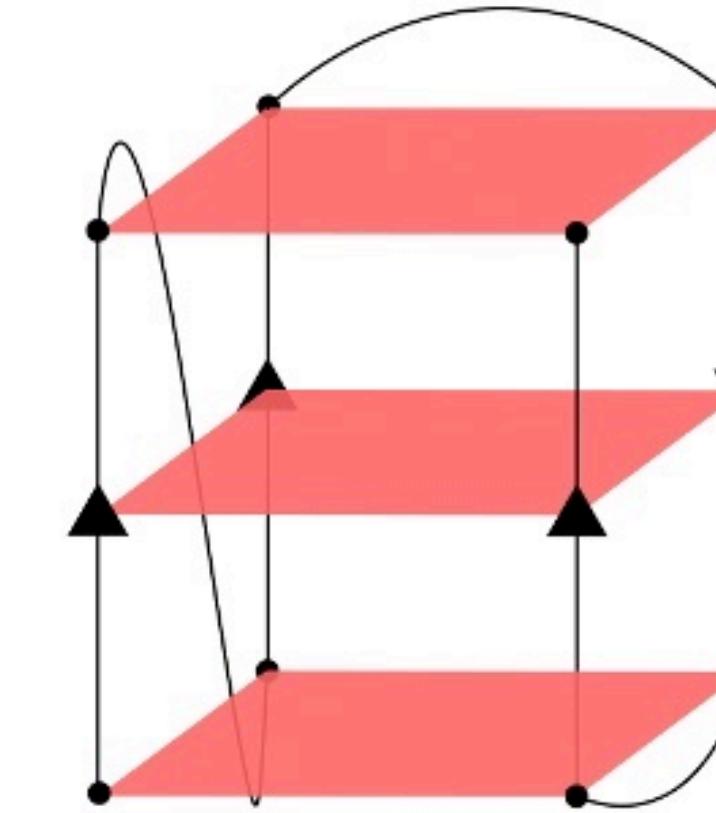


Parallel (4+0)



Antiparallel (2+2)

VG<sub>2</sub>WYN<sub>2-11</sub>MWG<sub>3</sub>



Hybrid (3+1)

WG<sub>2</sub>YNWN<sub>3-10</sub>TWNG

## IUPAC nucleotide code

$$M = \{A, C\}$$

$$W = \{A, T\}$$

$$R = \{A, G\}$$

$$Y = \{C, T\}$$

$$V = \{A, C, G\}$$

$$N = \{A, T, C, G\}$$

D. Liew, Z.W. Lim, and E.H. Yong, *Scientific Reports*, 14(1), 2024

*AI for Science*

Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

# Thermodynamic stability of G-quadruplex

General motif:  $G_m L_1 G_m L_2 G_m L_3 G_m$

Stability reflected by melting temperature  $T_m$

- G-quadruplexes typically have  $T_m$  that are 20-30°C higher than those of comparable double-stranded DNA.
- GC-rich regions tend to have higher  $T_m$  than AT-rich regions.

### Loop length $L$

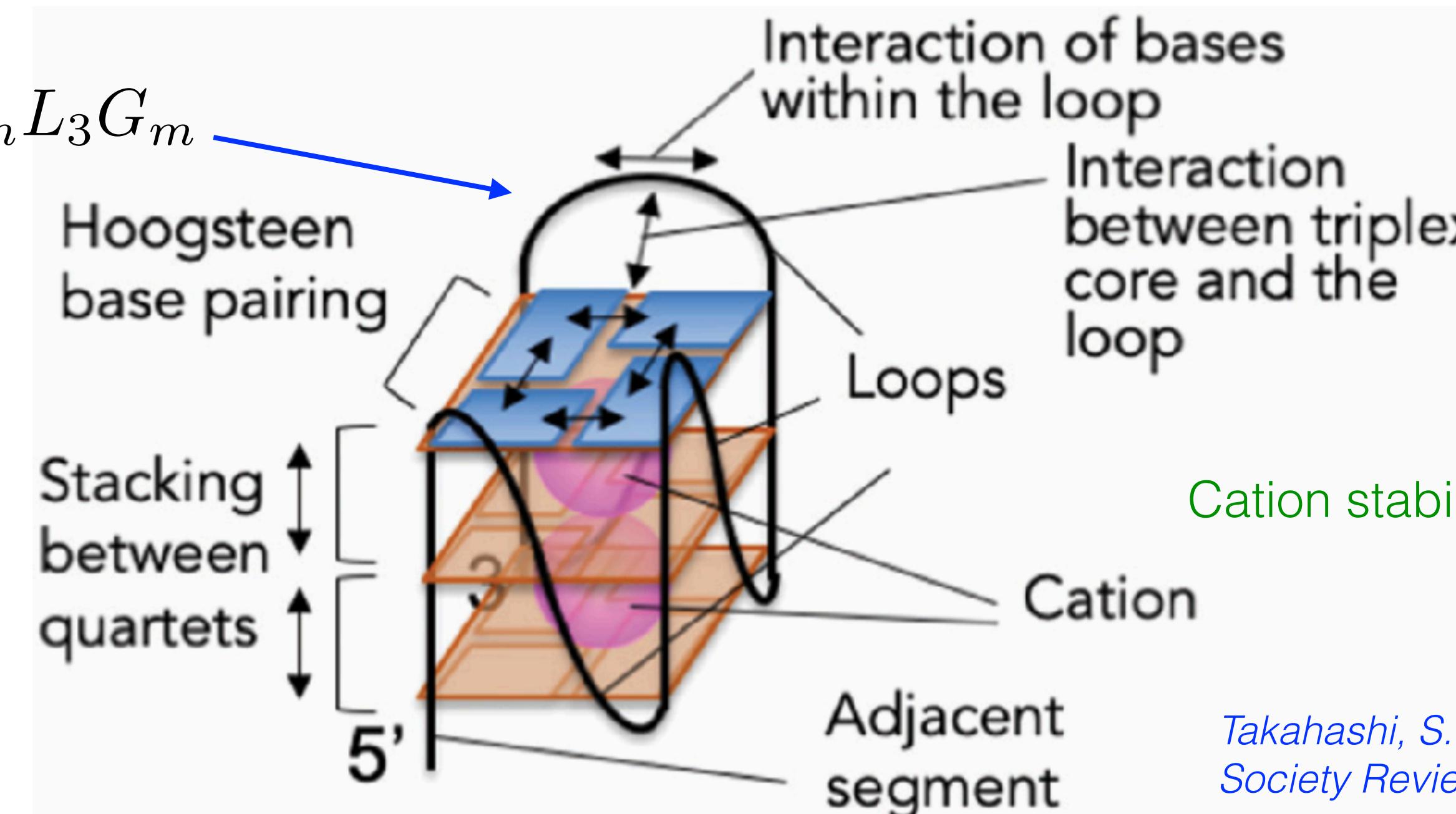
- An increase in loop length is associated with a gradual decrease in stability

### Loop composition

- Each substitution of adenine for thymine in a 1-nt propeller loop lowered G4 melting by 6-8 °C

### Loop length and topology

- Single nucleotide loop (propeller) correlates with high stability in parallel quadruplex
- Long central loop more likely to fold into an antiparallel (2+2) or hybrid-type (3+1) structure whereas short central loop favor parallel (4+0)

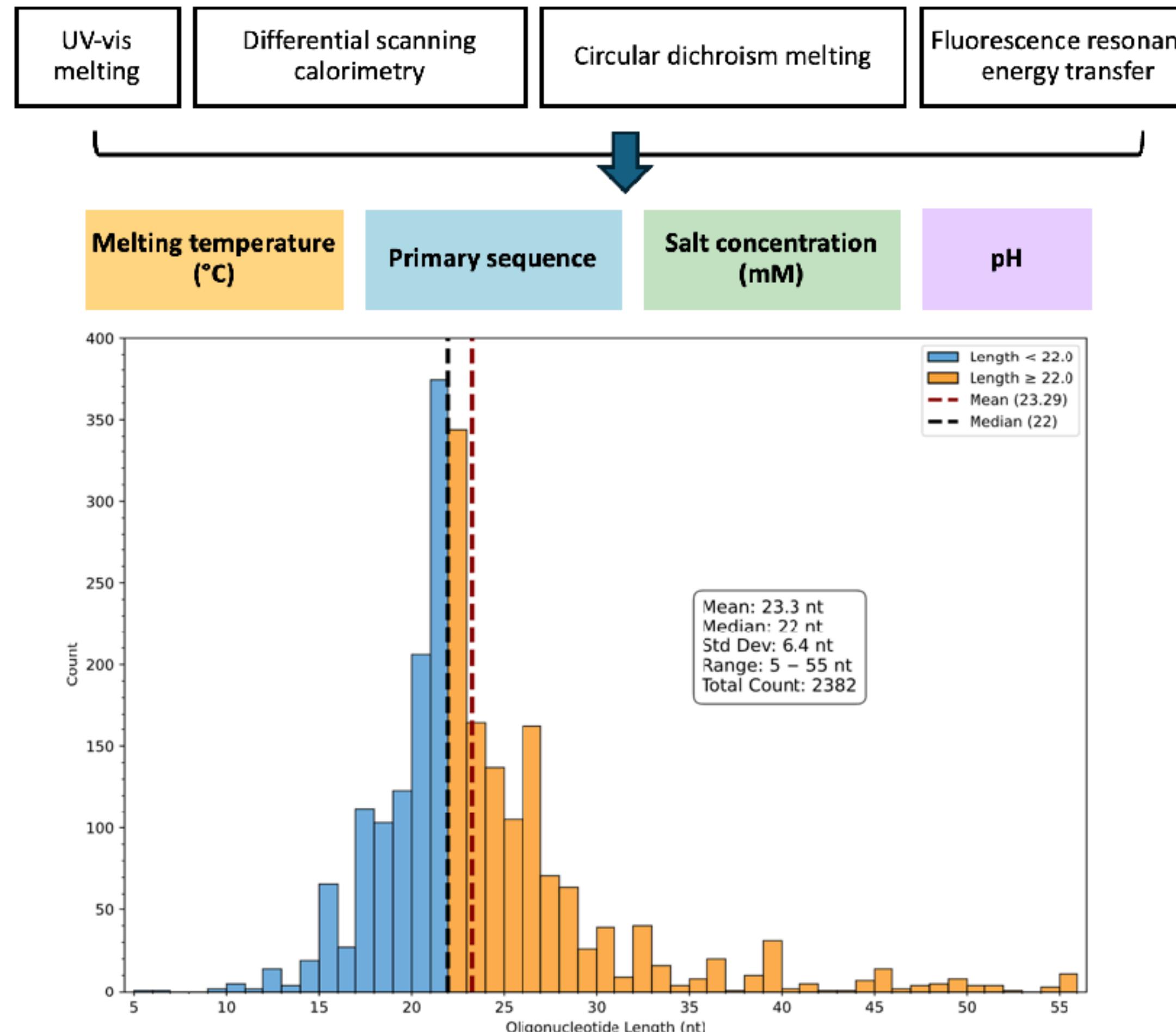


Cation stability:  $K^+ > Na^+ > Li^+$

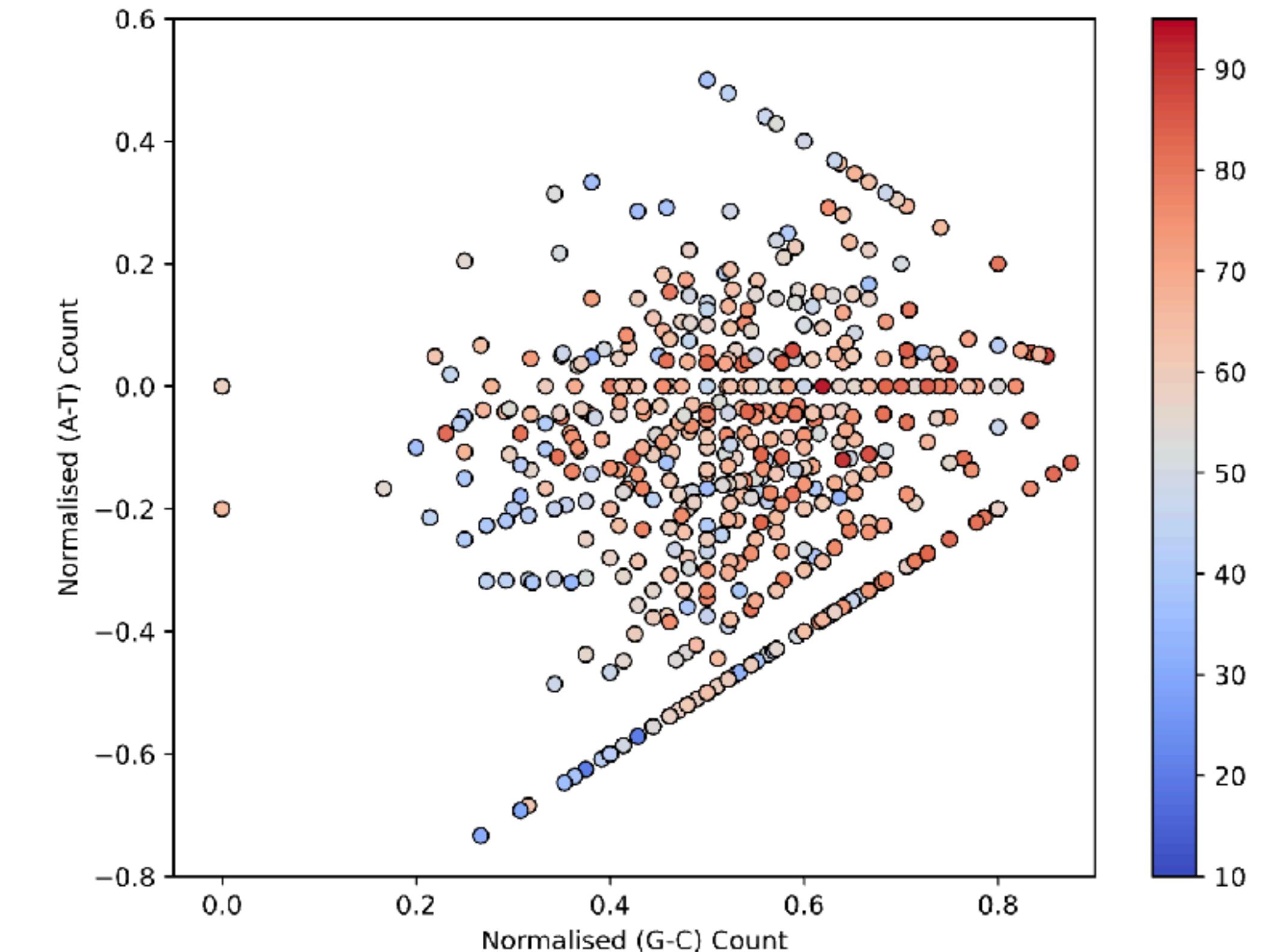
Takahashi, S. and Sugimoto, N., *Chemical Society Reviews*, 49(23), 2020.

Jana, J. and Weisz, K., *ChemBioChem*, 22(19), 2021.

# Curated G4 Dataset for G-Quadruplex Stability Prediction



2382 melting temperature  $T_m$  data points



D. Liew, A.D. Dharmatilleke, E. See, and E.H. Yong, Bioinformatics, 41(10), 2025

Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

# Deep Learning architecture of G4STAB

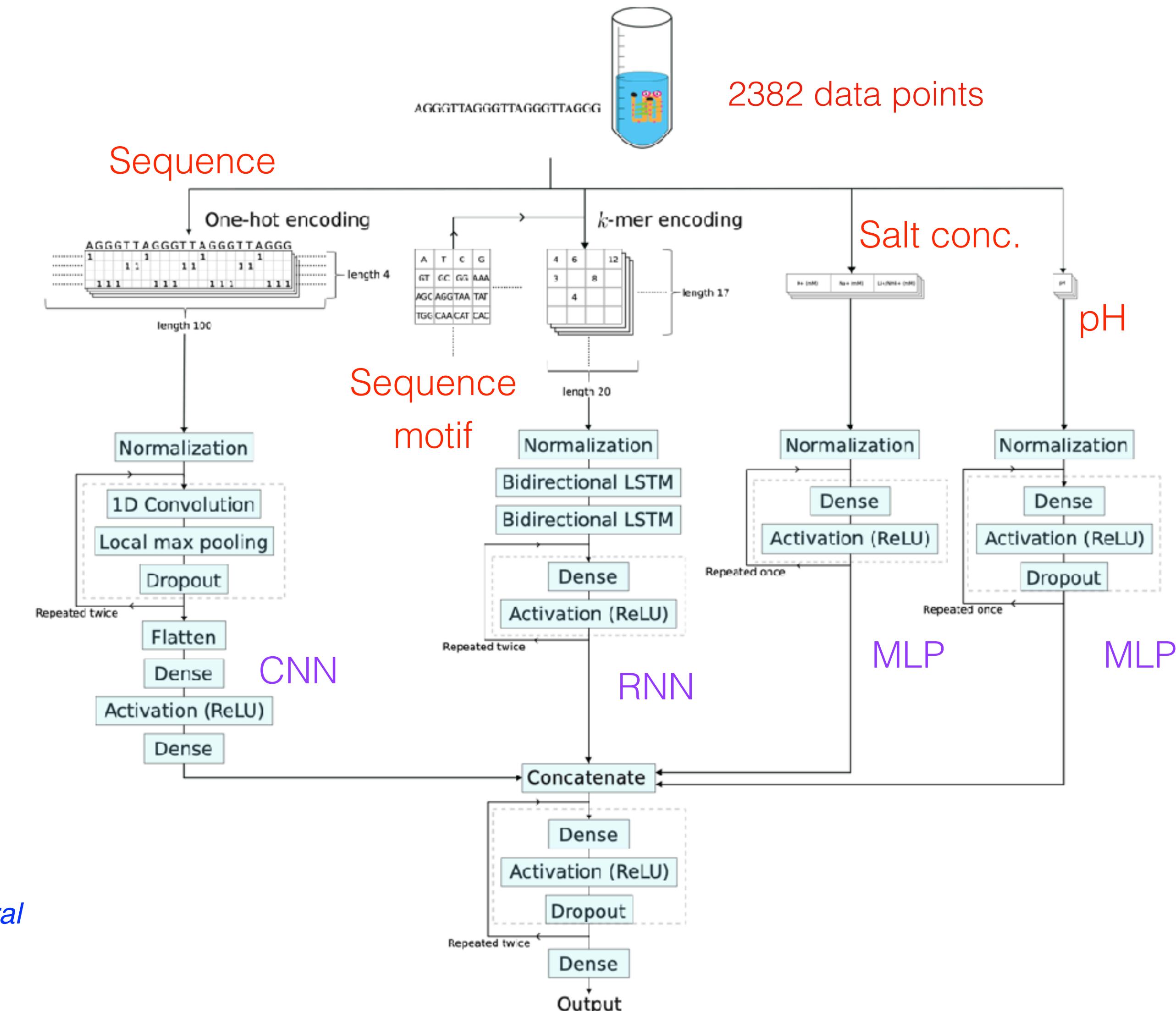
## G4STAB

**Four** parallel input branches processing different feature types through a series of transformations. Features are encoded and transformed separately, before concatenation and passed through dense layers for final prediction

1. **one-hot** encoding — convert categorical data into a binary format
2. **k-mer** encoding — identify and analyze substrings (k-mers) in sequence, e.g., AT, ATT, AGTC, etc
3. **Salt concentration** ( $K^+$ ,  $Na^+$ ,  $Li^+$ )
4. **pH** values

Jin, H., Song, Q. and Hu, X., *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2019*

Snoek, J., Larochelle, H. and Adams, R.P., *Advances in neural information processing systems, 25, 2012*

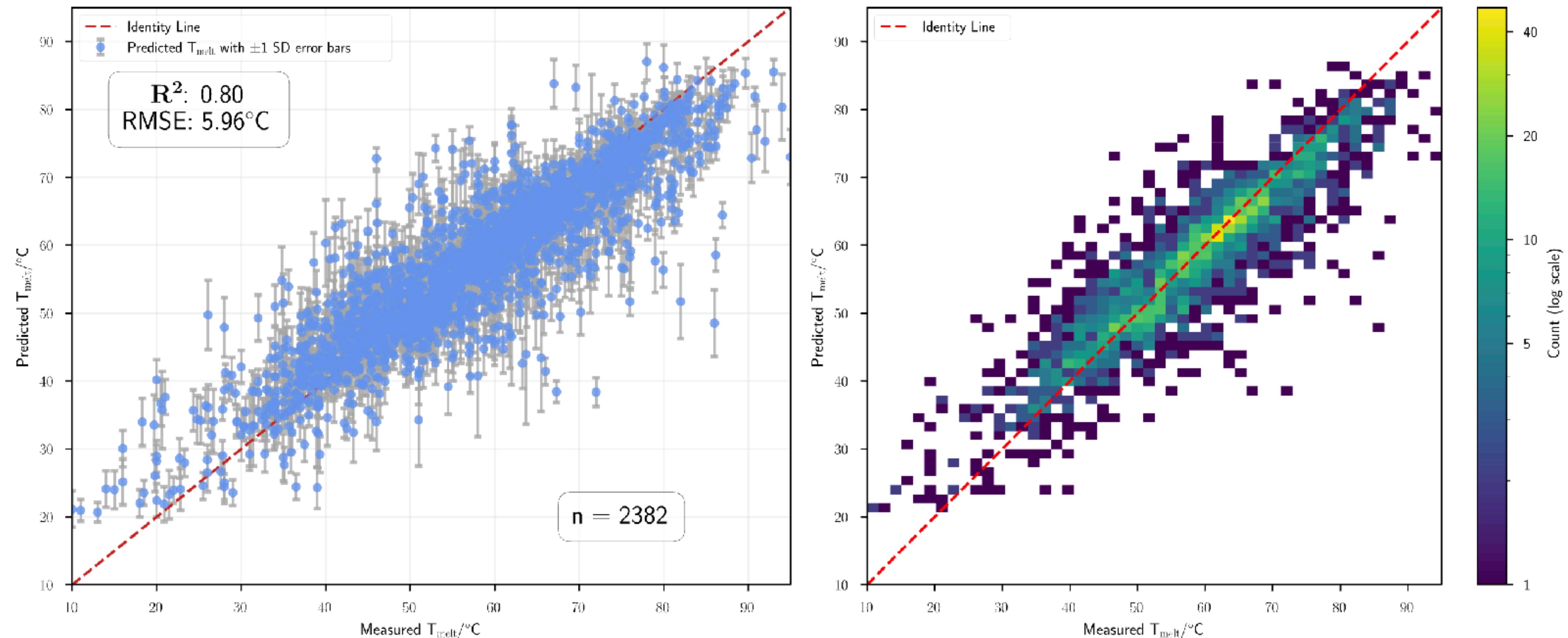


D. Liew, A.D. Dharmatilleke, E. See, and E.H. Yong, *Bioinformatics, 41(10), 2025*



Donn Liew  
PhD student

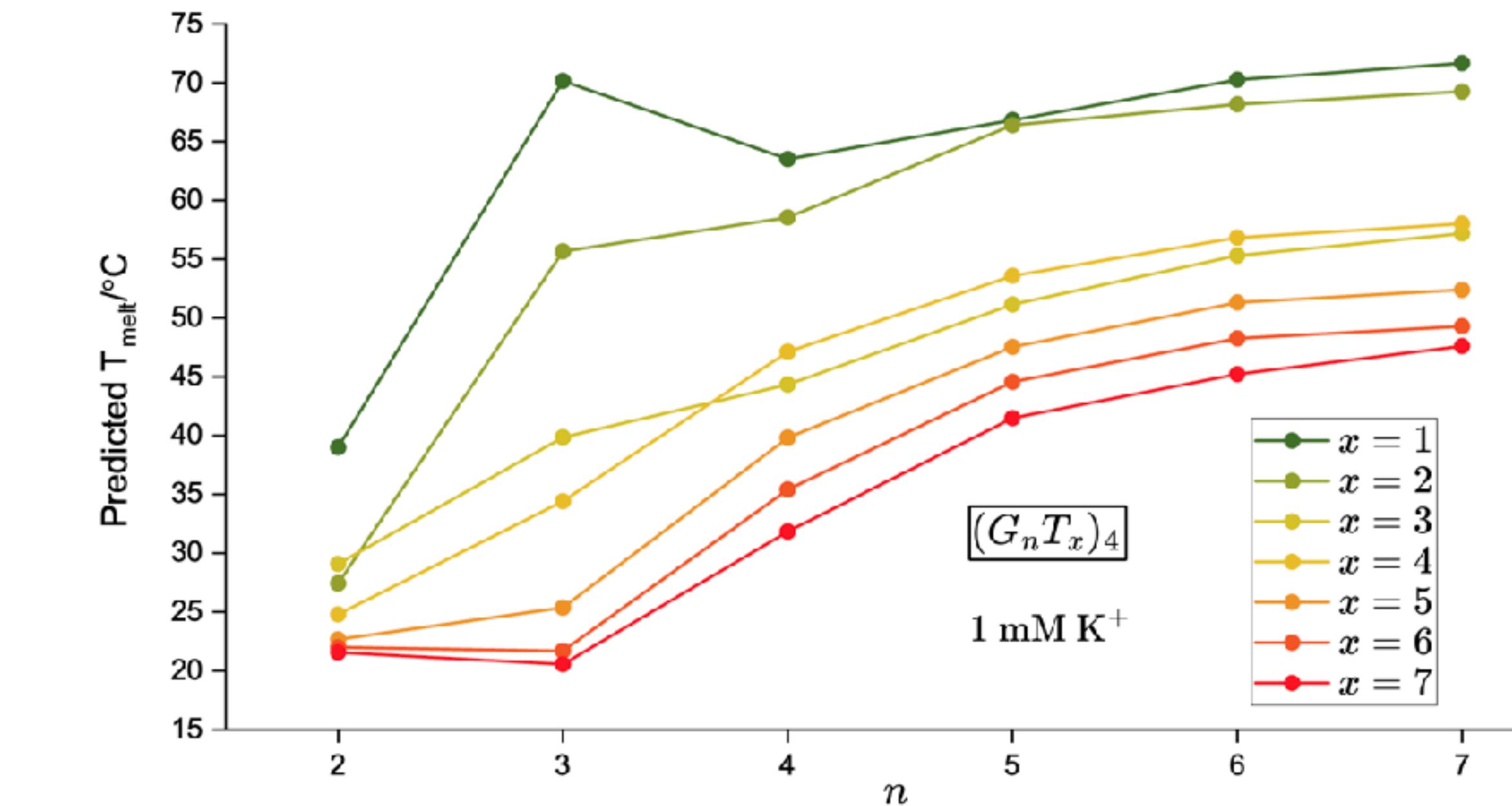
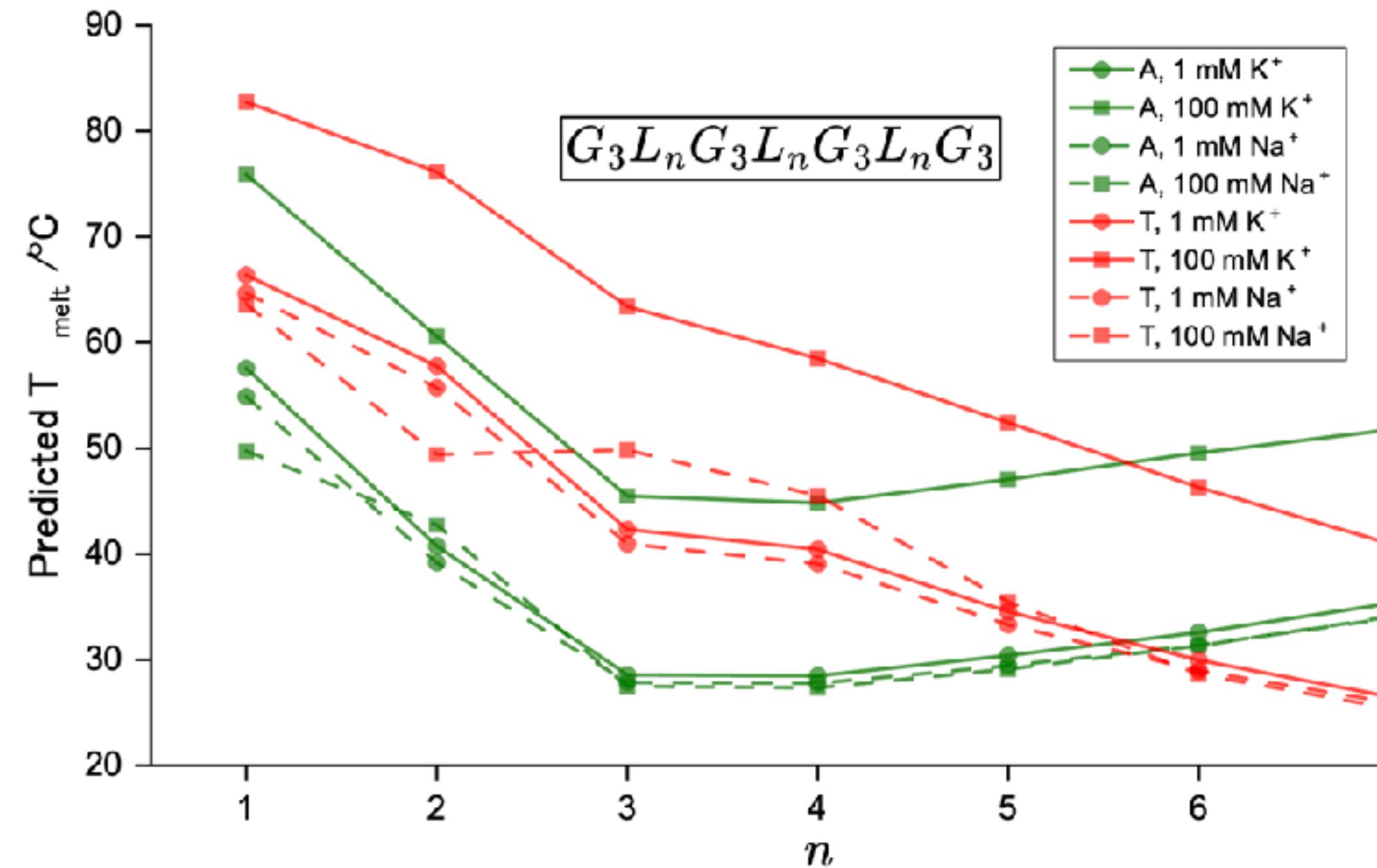
# Accuracy of G4STAB: True vs. Predicted Melting $T$



Evaluation using k-fold cross-validation ( $k=5$ , 200 repetitions)

D. Liew, A.D. Dharmatilleke, E. See, and E.H. Yong, Bioinformatics, 41(10), 2025

# G4STAB captures known G4 stability principles



- Inverse relationship between loop length and stability
- Observation: sequences with A loops level off and slightly increase in stability around loop length of 3-4
- Sequences with T loops consistently decrease in stability with increase loop length

- d(G<sub>n</sub>T)<sub>4</sub> sequence follow a stability order: n=3>7>6>5>4
- d(G<sub>3</sub>T)<sub>4</sub> anomalously stable as reported.
- In contrast, stability for d(G<sub>n</sub>T<sub>2</sub>)<sub>4</sub> consistently increases with G-tract length, n=7>6>5>4>3>2

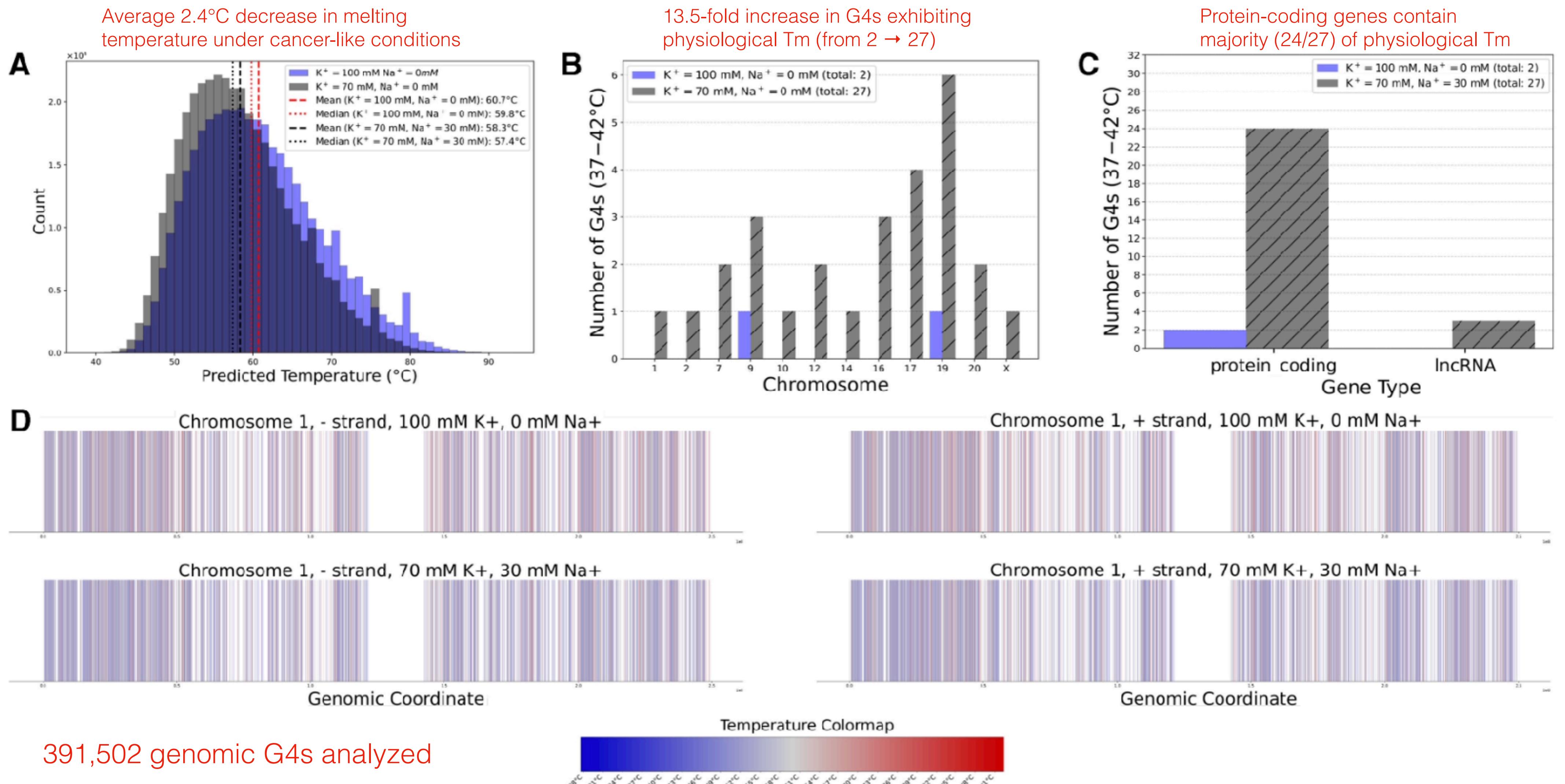
D. Liew, A.D. Dharmatilleke, E. See, and E.H. Yong, Bioinformatics, 41(10), 2025

AI for Science

Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

# G4STAB identifies cancer-like conditions that alter G4 stability



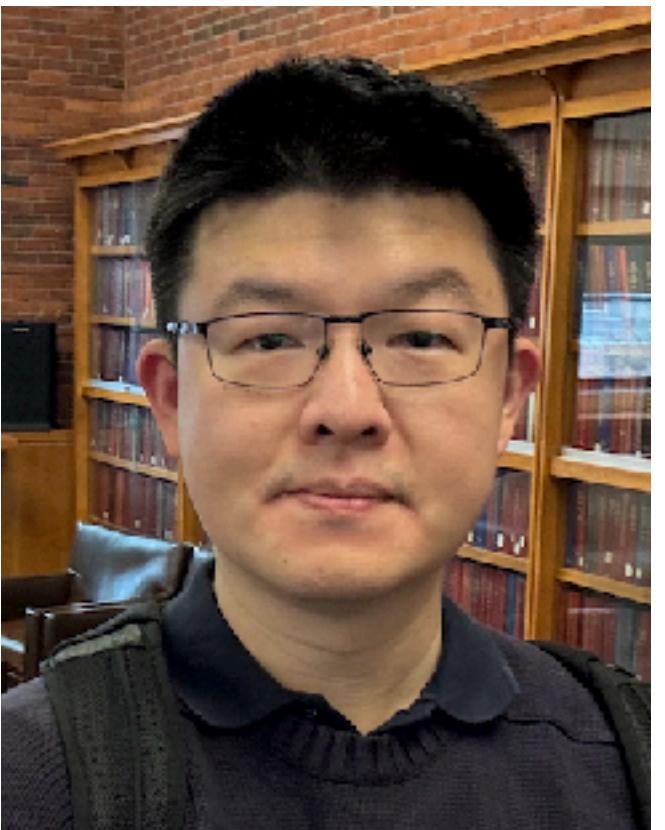
AI for Science

D. Liew, A.D. Dharmatilleke, E. See, and E.H. Yong, Bioinformatics, 41(10), 2025

Asst Prof YONG Ee Hou

Yong Group: Theoretical Biological & Statistical Physics

# Acknowledgement



**Ee Hou YONG**

<https://personal.ntu.edu.sg/eehou/>



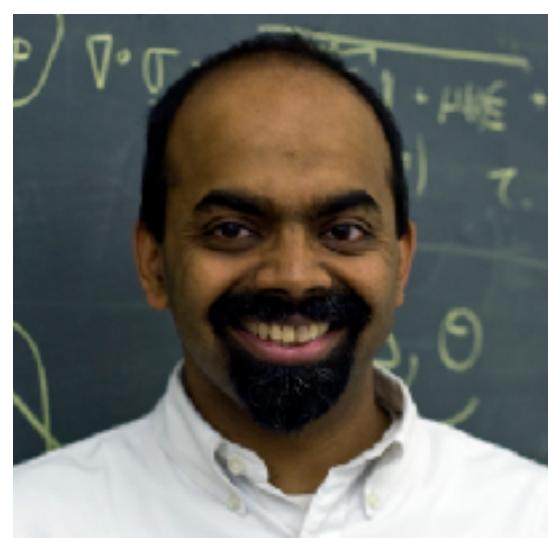
Ministry of Education  
SINGAPORE

MOE Academic Research Fund  
Tier 2 MOE-T2EP50223-0014

MOE Academic Research Fund  
Tier 1 RG140/22

## Thank You! Q&A

### Collaborators:



L. Mahadevan  
Harvard



M. C. Stoddard  
Princeton University



Liang Haiyi  
USTC, China



Orit Peleg  
UC Boulder



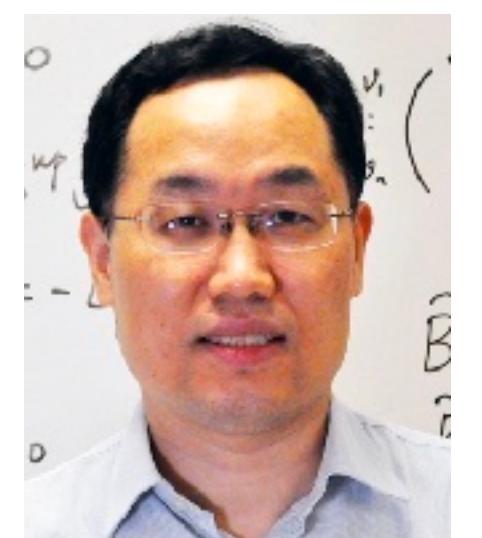
Chng Shu Sin  
NUS



Phan Anh Tuan  
NTU



Tay Wee Peng  
NTU



Xia Kelin  
NTU

### Current and Past Group Members:



Donn Liew



Nuoya Zhang



Farisan Dary



Too Hon Lin



Ryota Masuki

Asst Prof Ee Hou YONG

Theoretical Statistical Physics and Biophysics Group