Overcoming the Multiscale Simulation Challenge for Biomolecular Systems

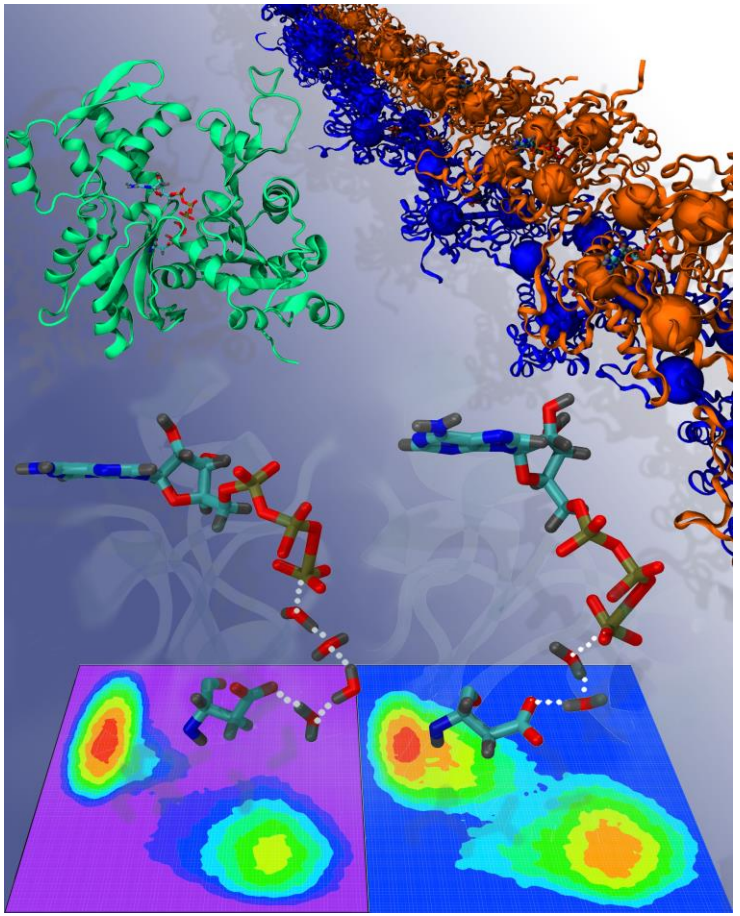
Greg Voth

THE UNIVERSITY OF CHICAGO

CHICAGO CENTER for THEORETICAL ChEMISTRY

Institute for Biophysical Dynamics

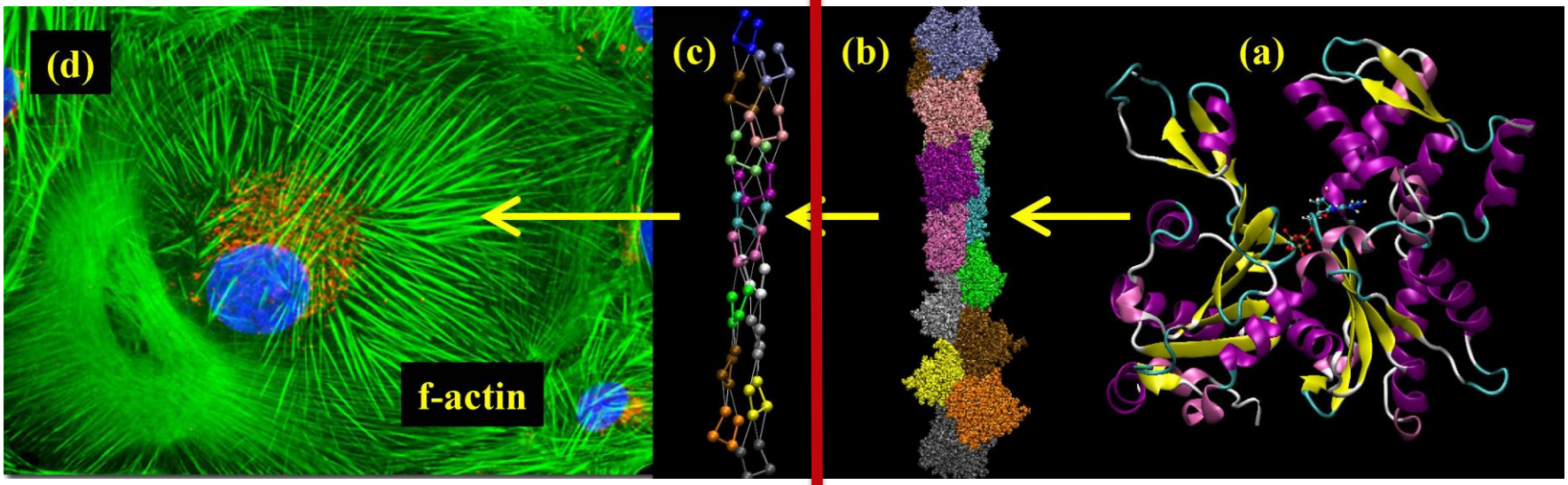
Chicago BIOPHYSICS



THE James Franck INSTITUTE
THE UNIVERSITY OF CHICAGO

Reaching Cellular Scales

Physical-based Computer Simulation at the Scales of Cellular Biology



Higher Scale and Multiscale Simulation

Closely tied to Cellular and Systems Biology, e.g., Various forms of Imaging, Cryo-Electron Tomography, Biochemical Networks, etc

The Concept of "Bottom-up" Coarse-graining

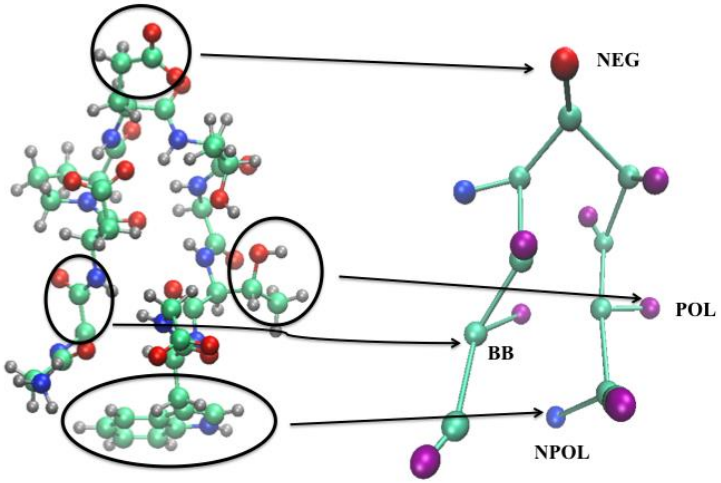
Coarse-Graining can be based on Statistical Mechanics

$$\exp(-\beta F) \propto \int d\mathbf{r} \exp[-\beta V(\mathbf{r})] \qquad (\beta = 1/k_B T)$$

$$\int d\mathbf{r} \exp[-\beta V(\mathbf{r})] \approx \int d\mathbf{R}_{CG} \exp[-\beta V_{CG}(\mathbf{R}_{CG})] \qquad \left(N_{\mathbf{R}_{CG}} \ll N_{\mathbf{r}}\right)$$

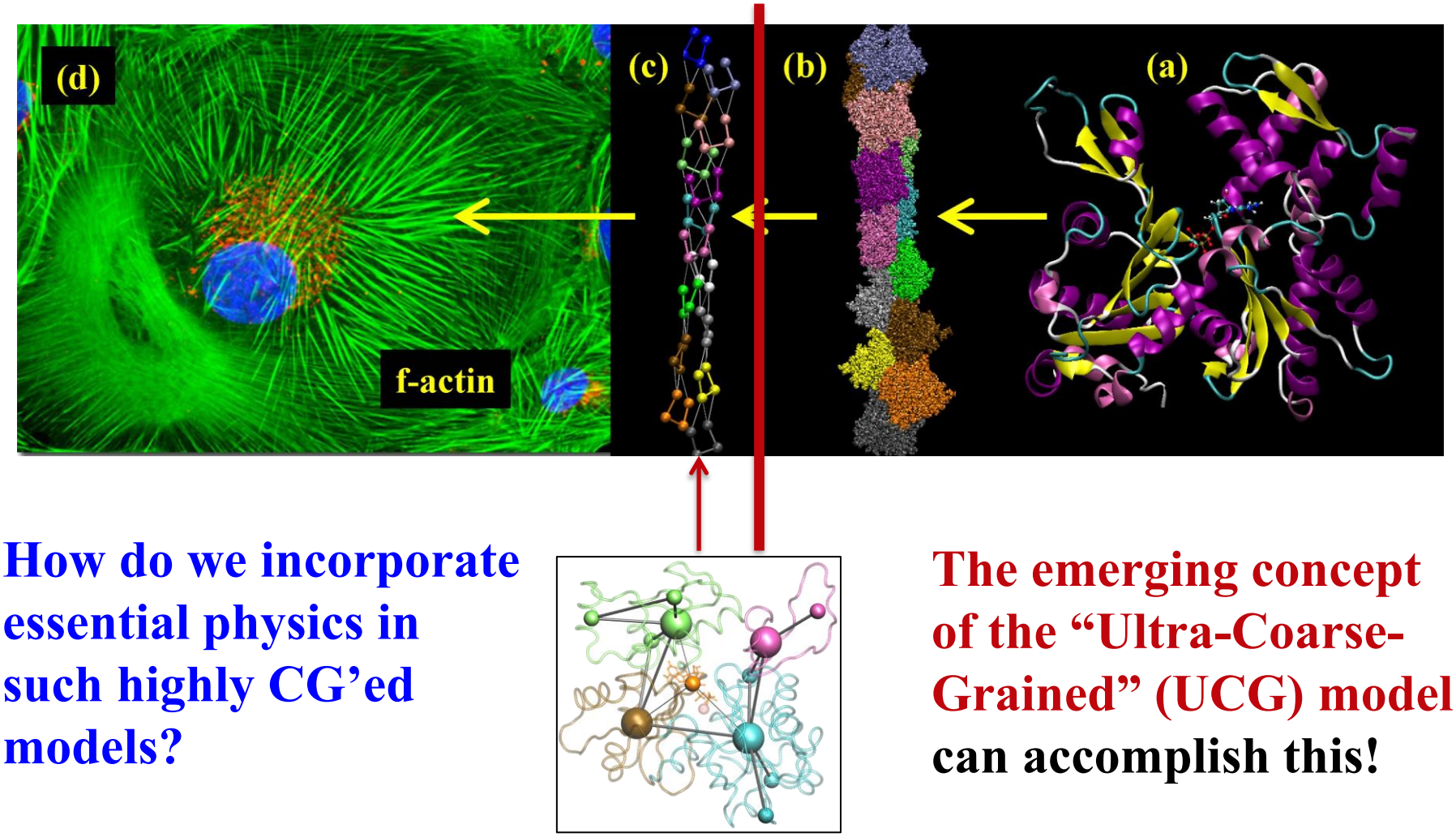
How best to define $\mathbf{R}_{CG}$?

How to determine $V_{CG}(\mathbf{R}_{CG})$?



Shown here is a "high resolution" CG model having some number of CG sites or "beads" per each amino acid residue in the peptide.

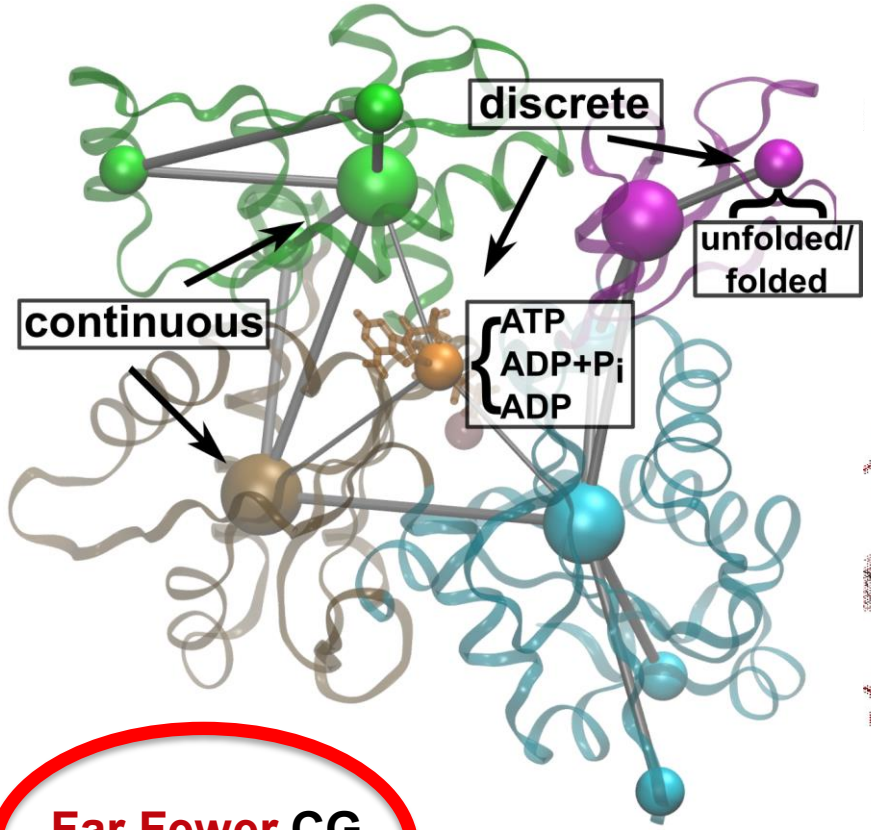
Reaching Cellular Scales

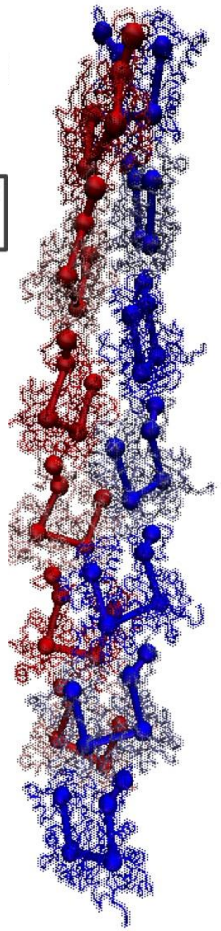


How do we incorporate essential physics in such highly CG'ed models?

The emerging concept of the "Ultra-Coarse-Grained" (UCG) model can accomplish this!

A Step Further and Something Very Different!
Ultra-Coarse-Graining (UCG)*





Far Fewer CG "Particles"

Continuous kinematic movement of CG particles is there *but not enough*

CG Particles must have internal "states"

Dynamic state change *within* the CG particles modulates interactions *between* CG particles

*J. F. Dama, A. V. Sinitskiy, M. McCullagh, J. Weare, B. Roux, A. R. Dinner, and G. A. Voth, "Theory of Ultra Coarse-Graining. I. General Principles", J. Chem. Theor. Comp. **9**, 2466–2480 (2013).

UCG Advantage Þ

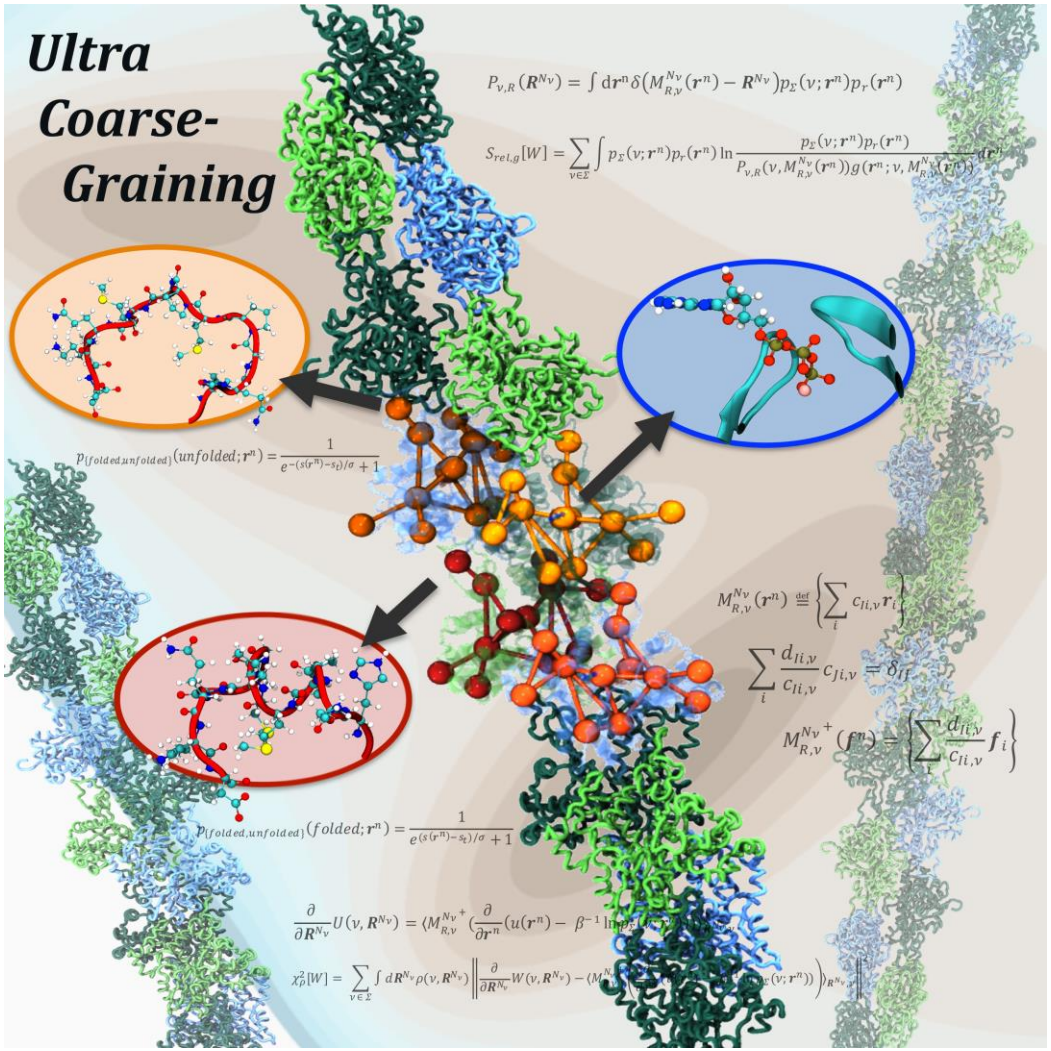
(#UCG States)´ (#UCG Sites) < (#Higher Res CG Sites) < (#Atomic Sites)

Origins of Other Possible "States" in the UCG Sites

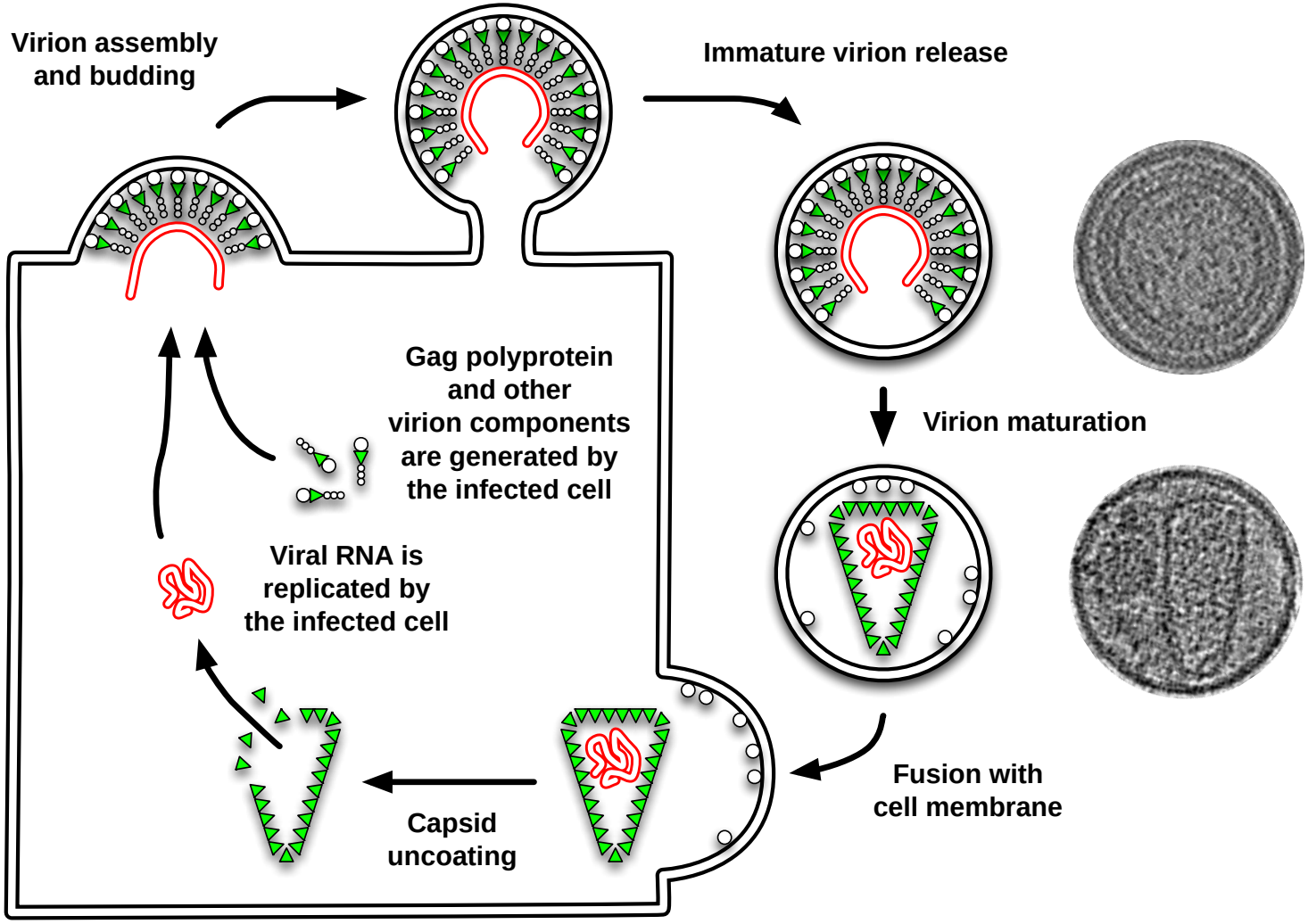
States within UCG "beads"

— physical —
disorder transition
ligand binding
loop folding/unfolding

— chemical —
nucleotide hydrolysis
redox reaction
protonation

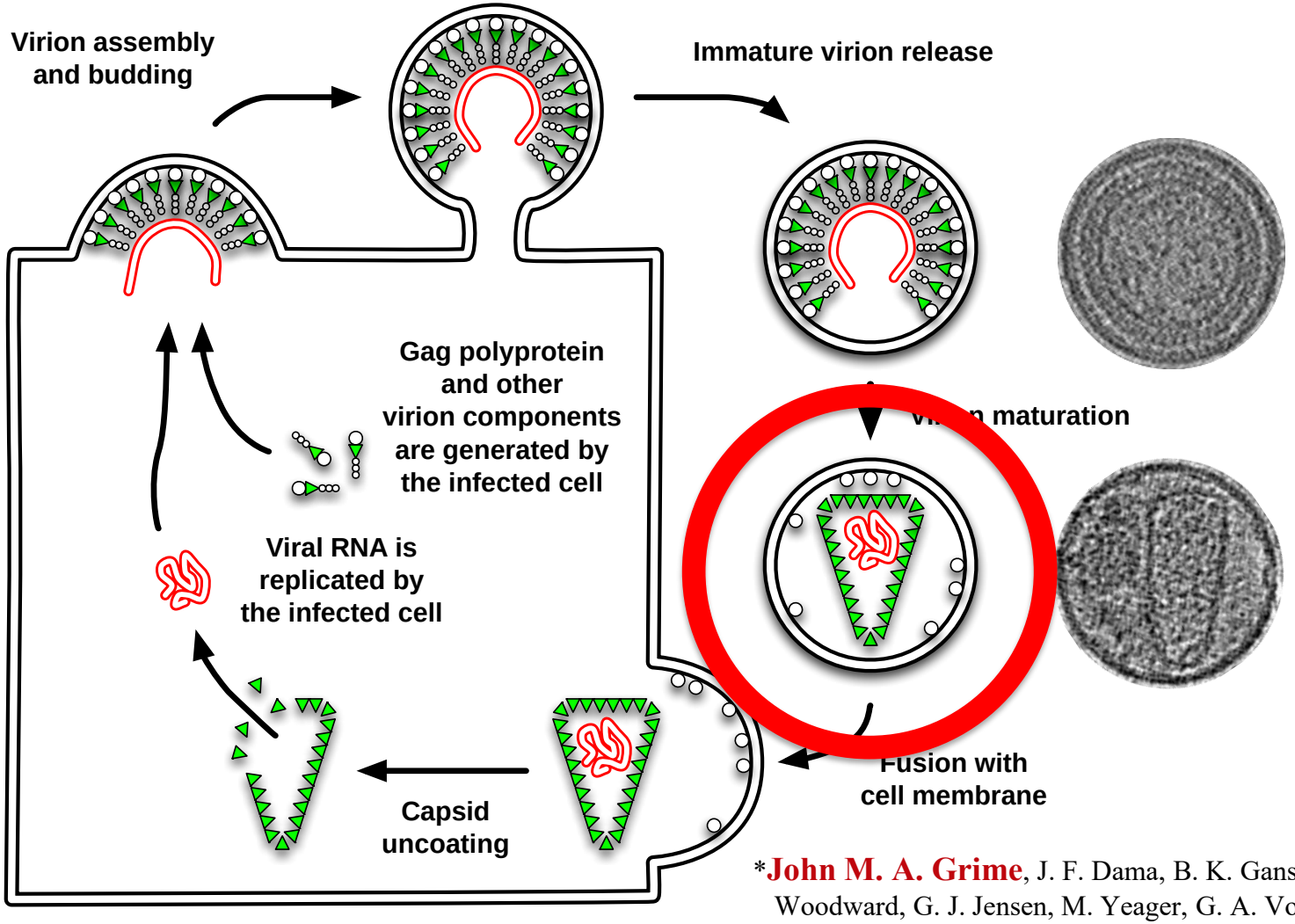


UCG Application: HIV Capsid Assembly



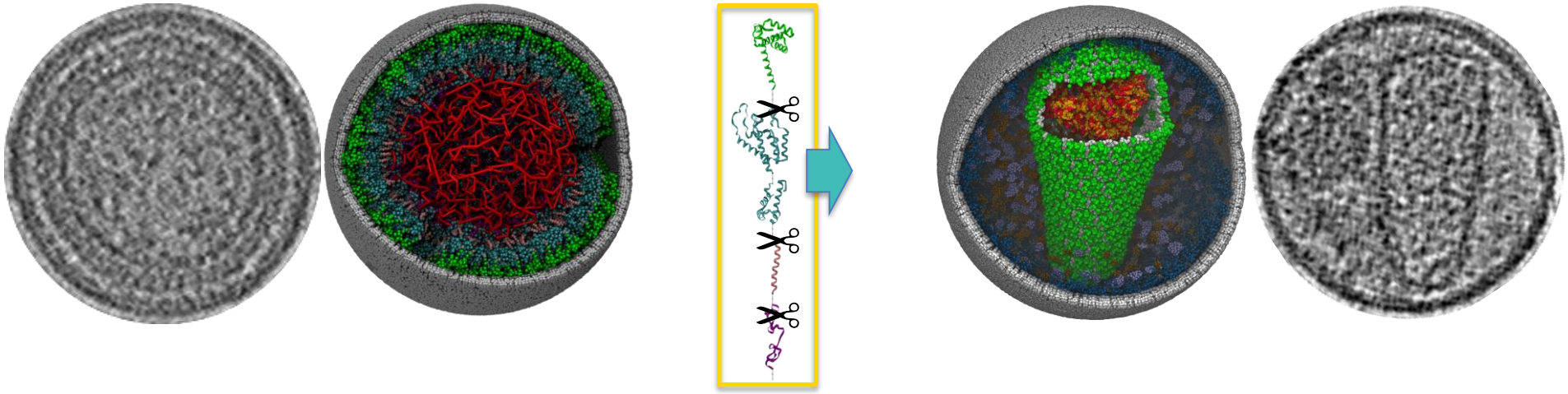
Ganser-Pornillos et al, *Curr. Op. Struct. Biol.* 2008, **18**:203

UCG Application: HIV Capsid Assembly

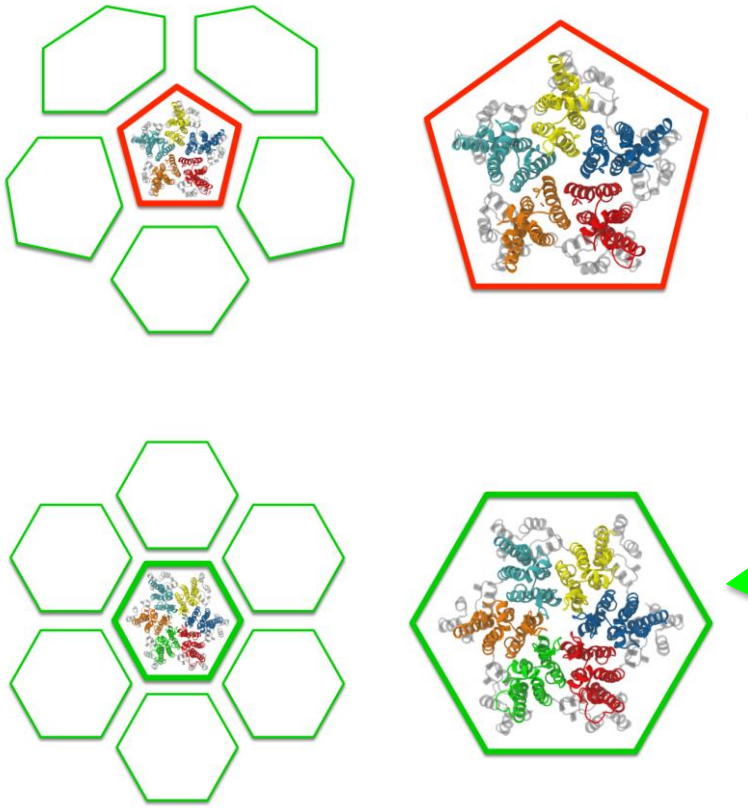


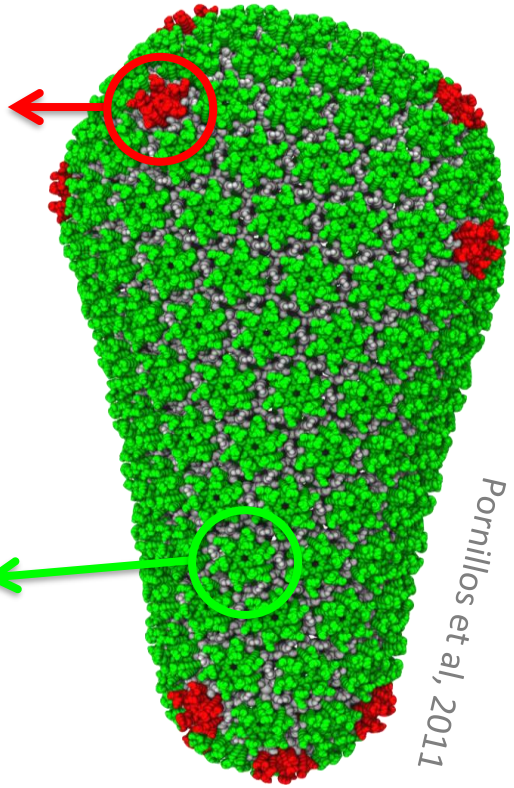
Ganser-Pornillos et al, *Curr. Op. Struct. Biol.* 2008, **18**:203

*John M. A. Grime, J. F. Dama, B. K. Ganser-Pornillos, C. L. Woodward, G. J. Jensen, M. Yeager, G. A. Voth, "Coarse-grained Simulation Reveals Key Features of HIV-1 Capsid Self-Assembly", Nature Comm. **7,** 11568(1-11) (2016).

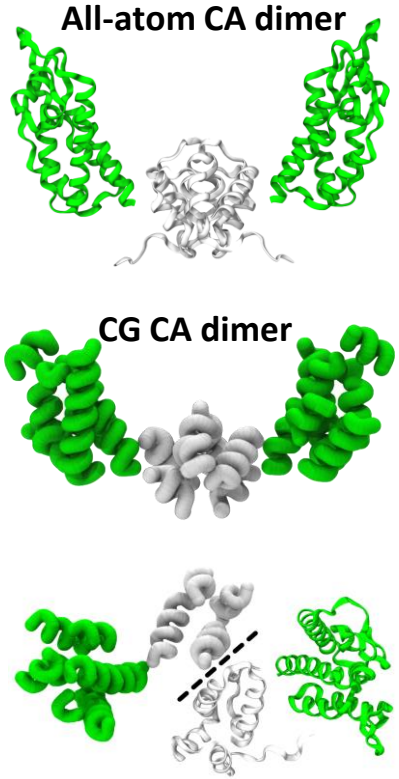


HIV-1 “Maturation”: No Conical Capsid, No Infectivity





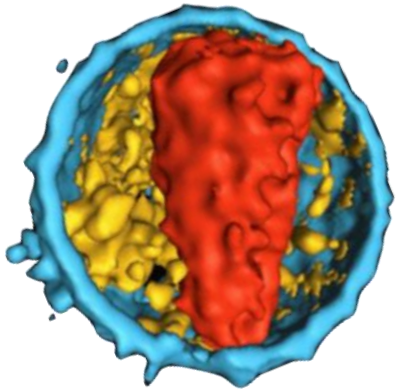
Pornillos et al, 2011



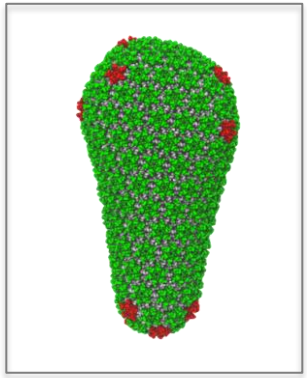
The UCG-MD Problem*

Implicit solvent for large scale ultra-CG: introduces large, dynamic areas of low particle density:

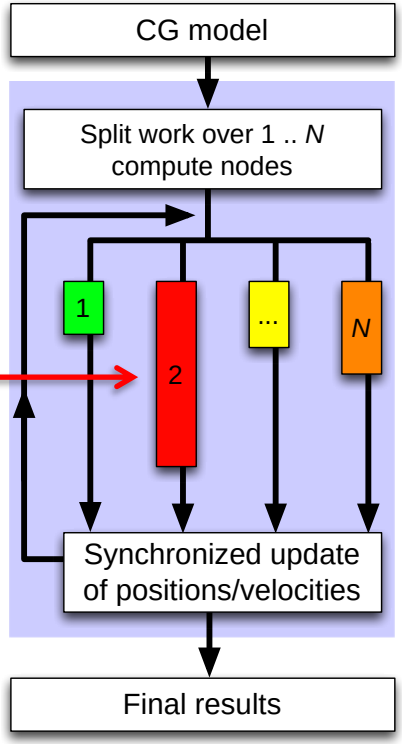
- *Load balancing*: MD simulation proceeds at pace of **slowest node**



HIV capsid[1]



Ultra-CG model



- *Memory requirements*: memory needed even for empty regions of simulation

[1]Tom Goddard 2009, UCSF

*J. M. A. Grime and G. A. Voth, "Highly Scalable and Memory Efficient Ultra-Coarse-Grained Molecular Dynamics Simulations", J. Chem. Theory Comp. **10**, 423–431 (2014).

UCG-MD Challenges*

MD code* for UCG (**John Grime**):

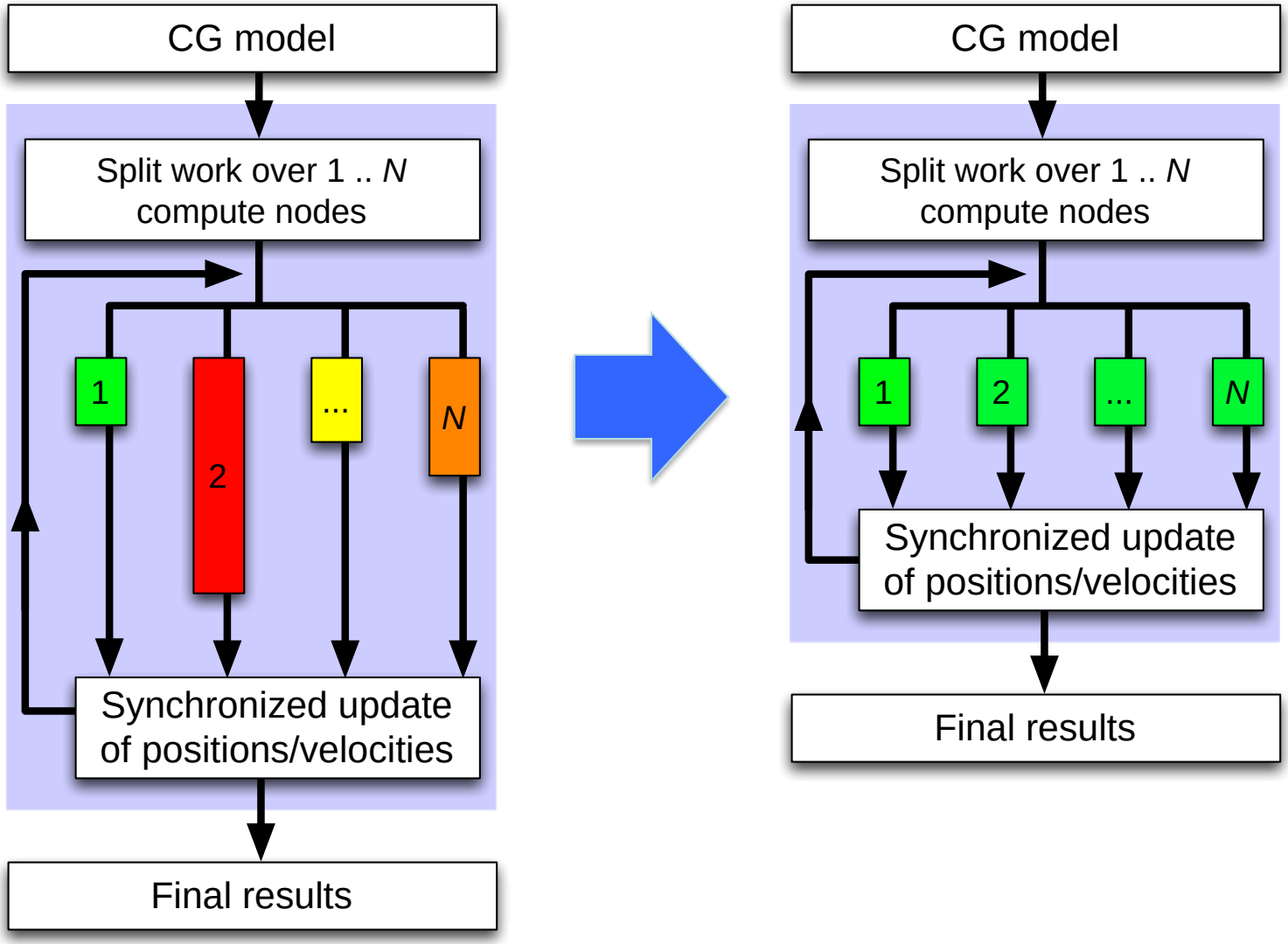
- **Dynamic sparse data** representations
(removes memory barriers for very large systems)

- **Load balancing** via Hilbert space filling curves
(better use of supercomputing resources)

End goal: enable highly dynamic UCG-MD simulations at a **cellular scale**. Motivated by CG biological systems, but in principle widely applicable in materials science.

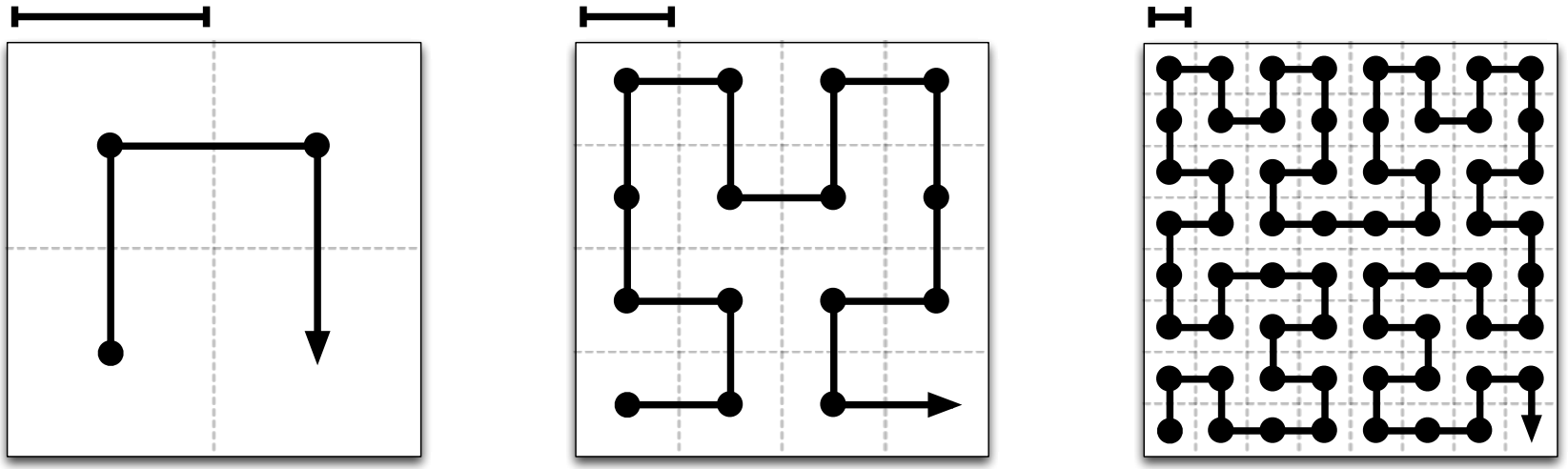
***Much is transitioned to LAMMPS MD code: Also working with NVIDIA on GPU implementations**

Load Balancing



Load Balancing

The space filling Hilbert curve



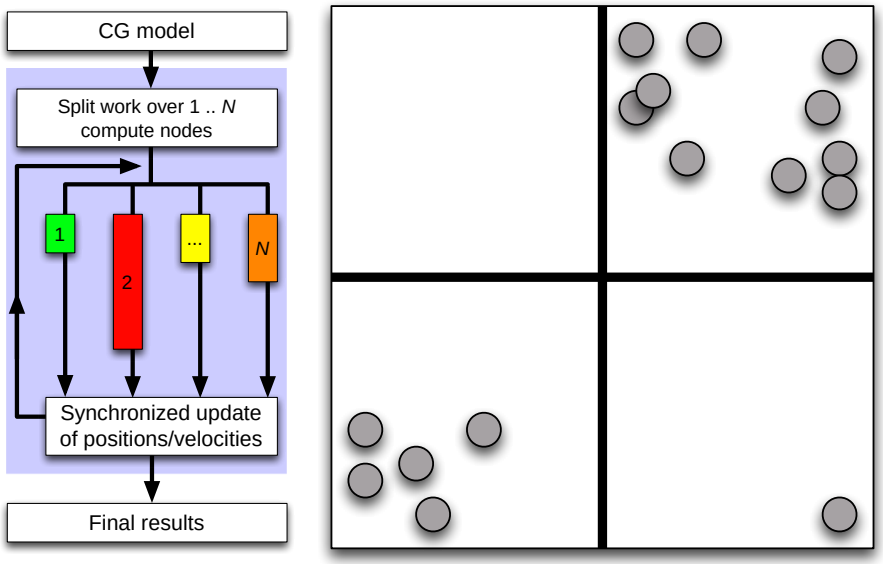
1st order 2nd order 3rd order

Useful properties:

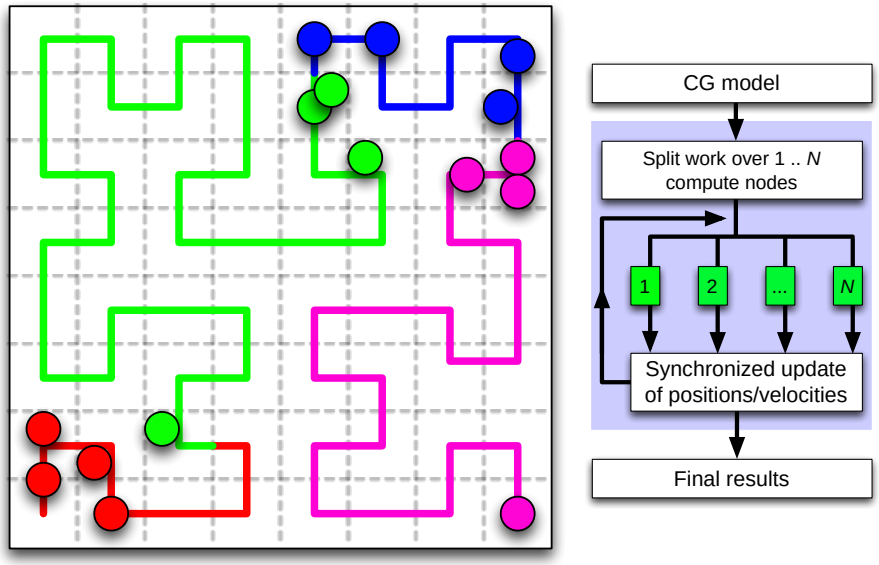
- **Space filling** (high resolution where needed)
- **Dynamic generation** (no precalculation)
- **Locality** (adjacent SFC "indices" also local in the original 2D/3D Cartesian space)

Load Balancing

Example: 16 particle simulation run with four CPUs:



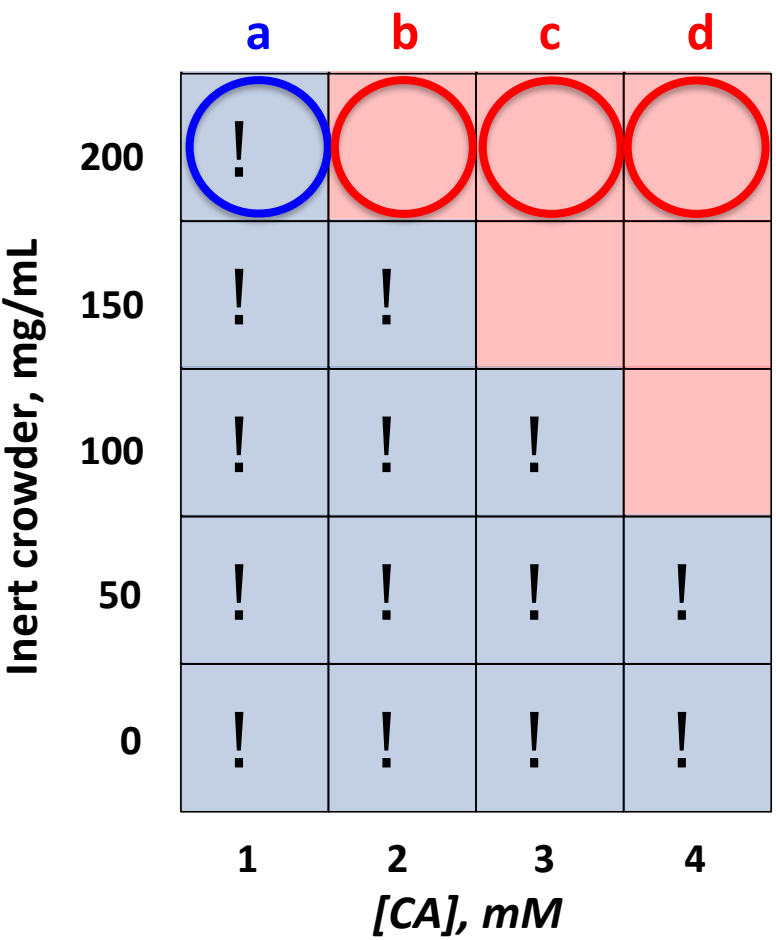
Naïve, uniform spatial decomposition

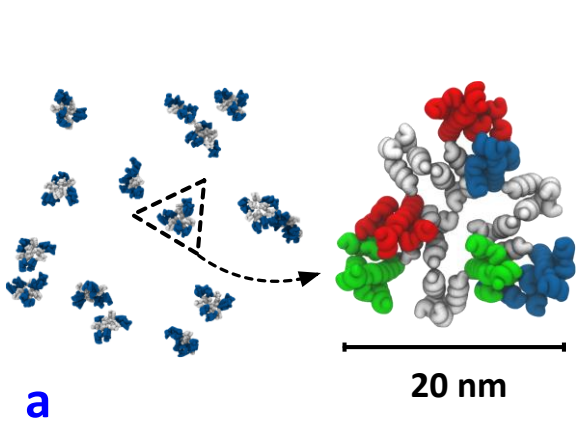


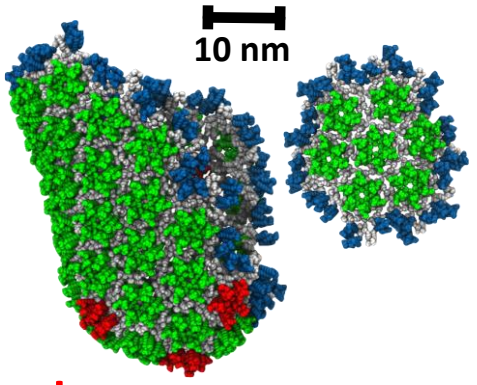
Hilbert curve spatial decomposition

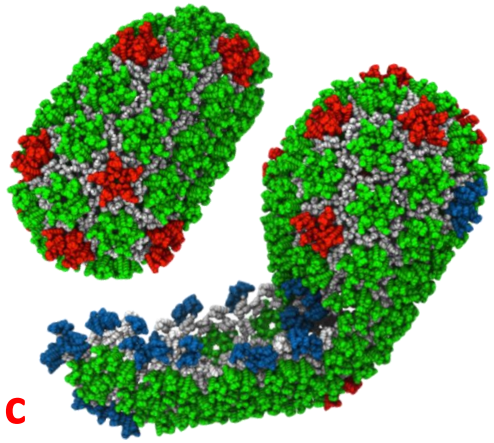
“Basic” CG model: *in vitro* Simulations

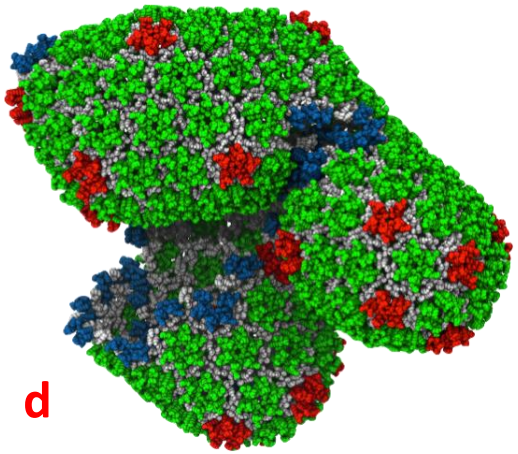
Vary both [CA] and level of molecular crowding up to approx. conditions expected in virion (4mM [CA], 200 mg/mL crowder)

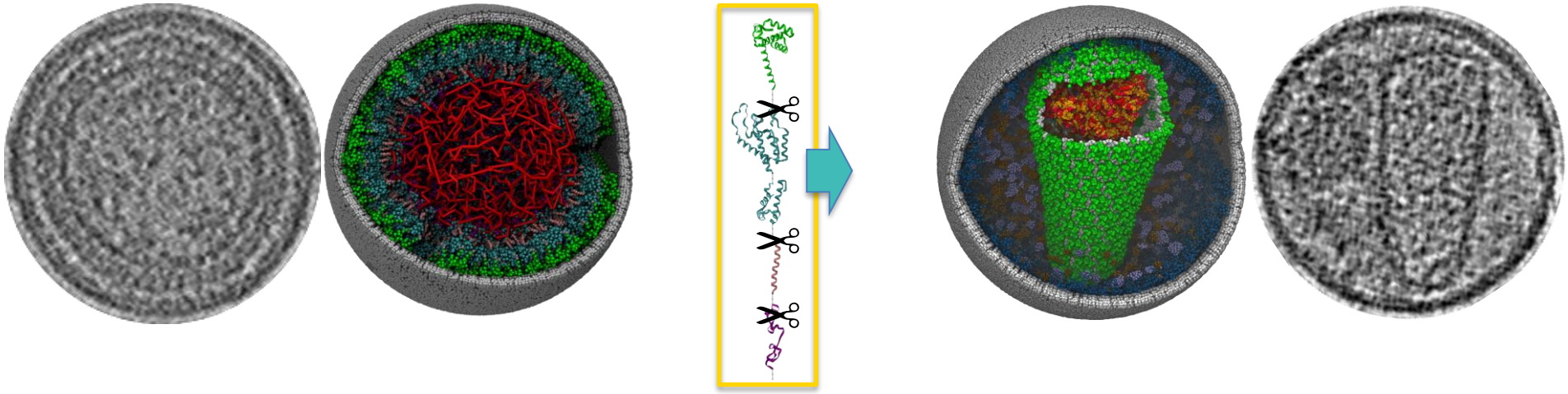




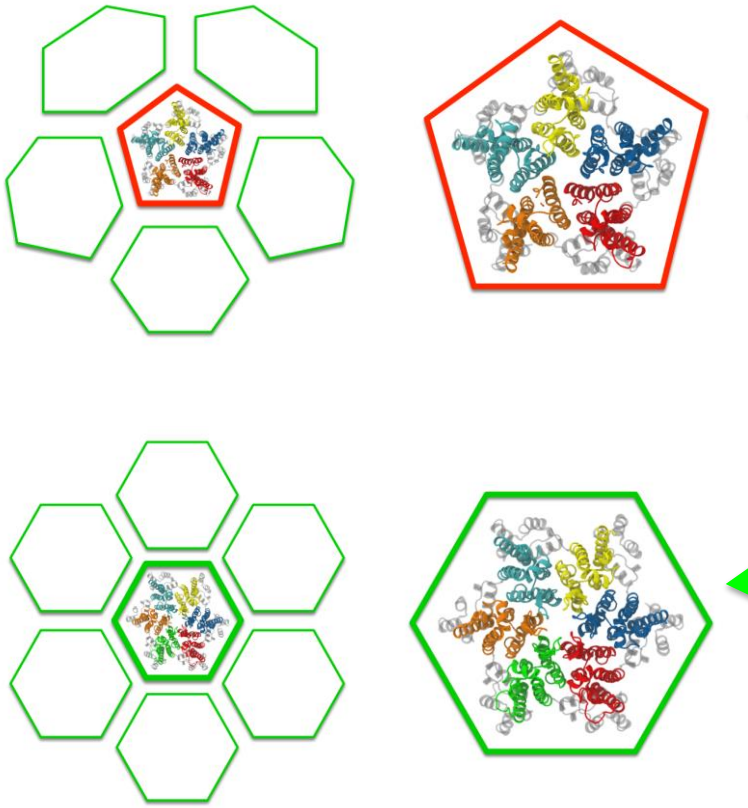


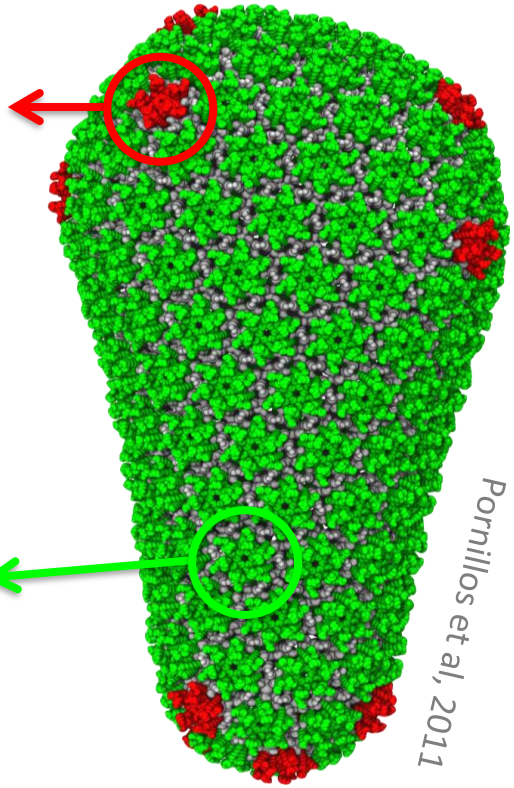




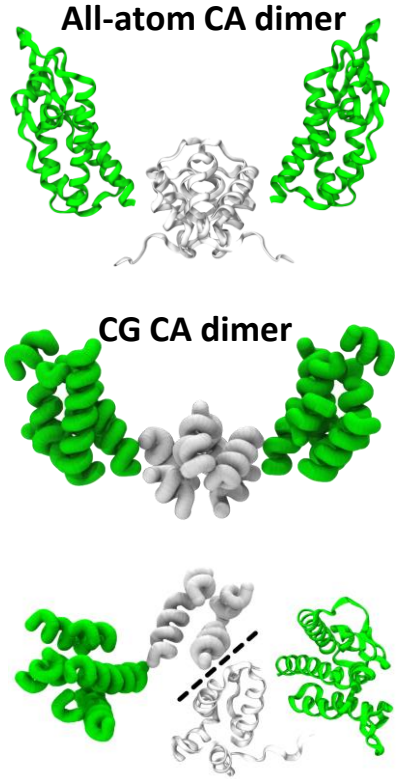


HIV-1 “Maturation”: No Conical Capsid, No Infectivity

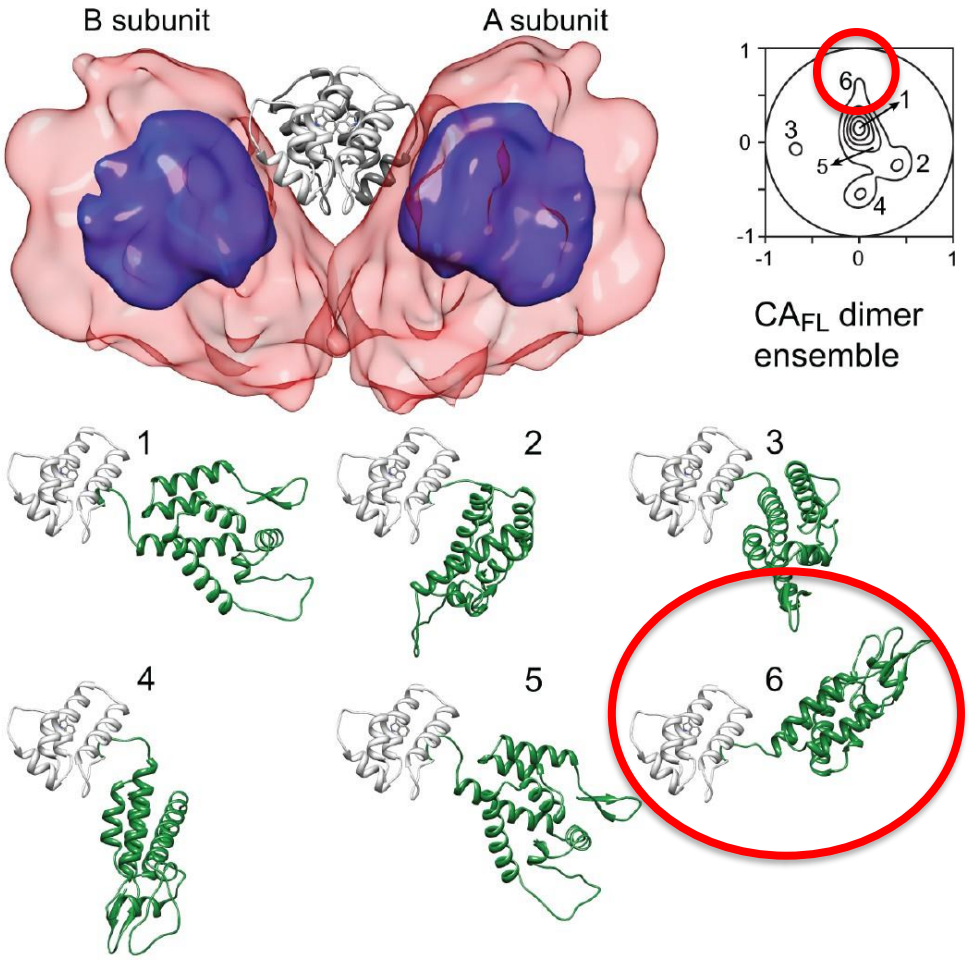




Pornillos et al, 2011



CA Protein Structural Dynamism



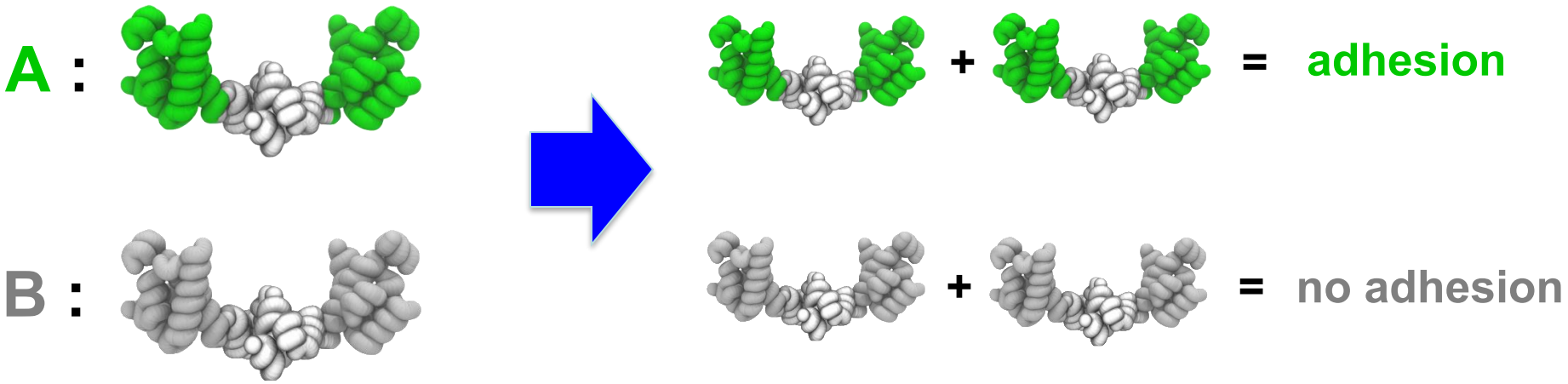
Relatively small amount of CA in "native-style" NTD/CTD conformation in solution (≈ 5%), with domain motions time-correlated (≈ 5 ns) [1].

UCG-style "switching" model for CA …

[1] Deshmukh et al, *JACS* **135**:16133-16147 (**2013**)

UCG-MD Style Model of HIV CA Protein

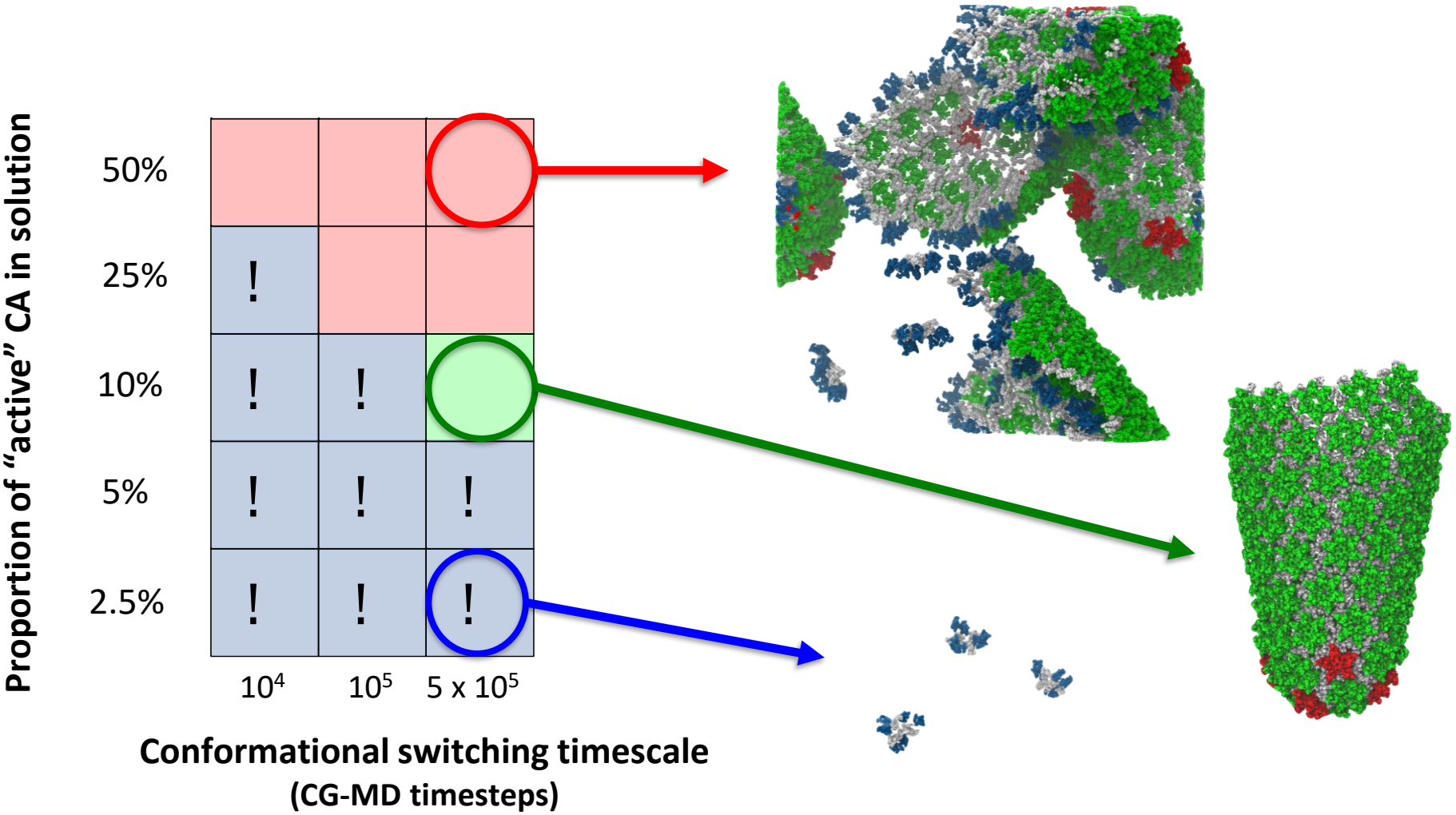
Simple two-component system: **A** and **B** protein dimers with identical internal structure but different interactions:

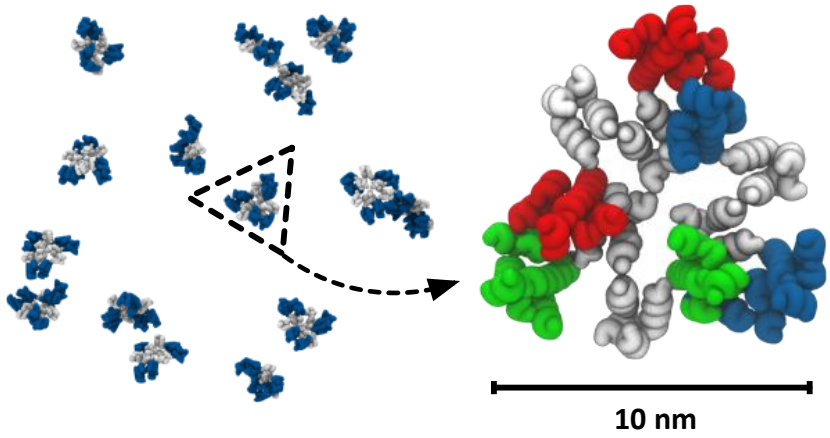


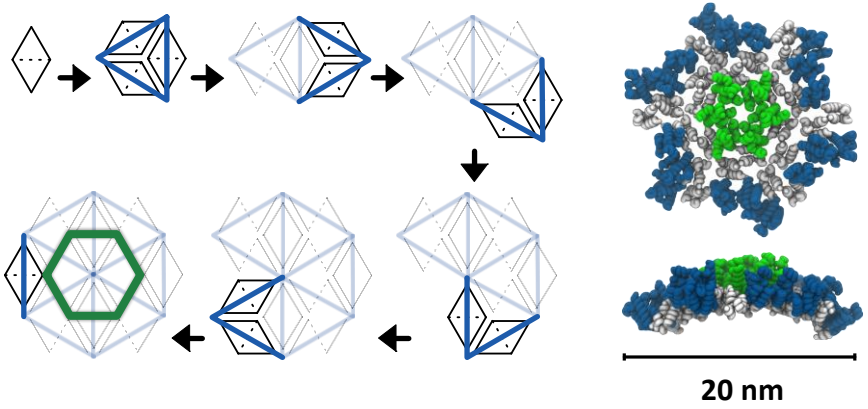
Fixed overall **[CA]** (4mM, 200 mg/mL crowder), ***solution-state*** **proportion of A stochastically (re)assigned with a certain timescale**. Examine effects of conformational heterogeneity on controlled HIV capsid self-assembly …

CA Structural Dynamism in UCG Model

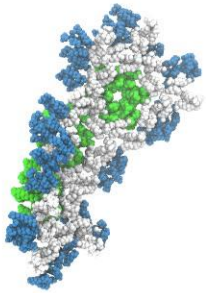
4 mM [CA], 200 mg/mL crowder, vary proportion of "active" (i.e. assembly competent) CA in solution …



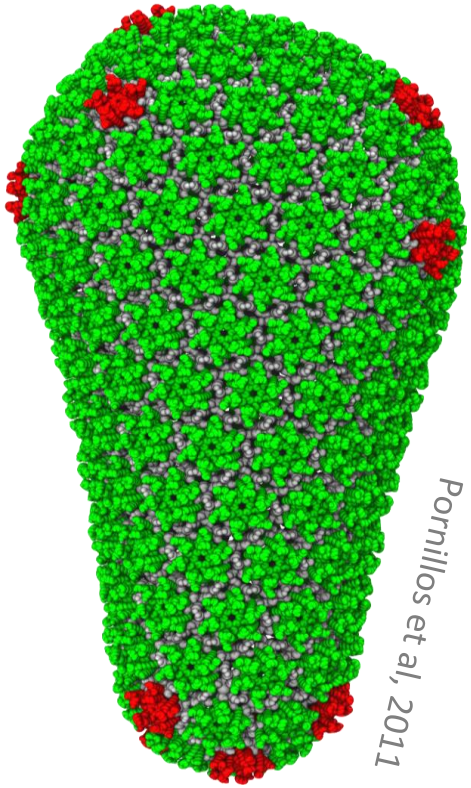




Controlled self-assembly requires UCG-style CA model

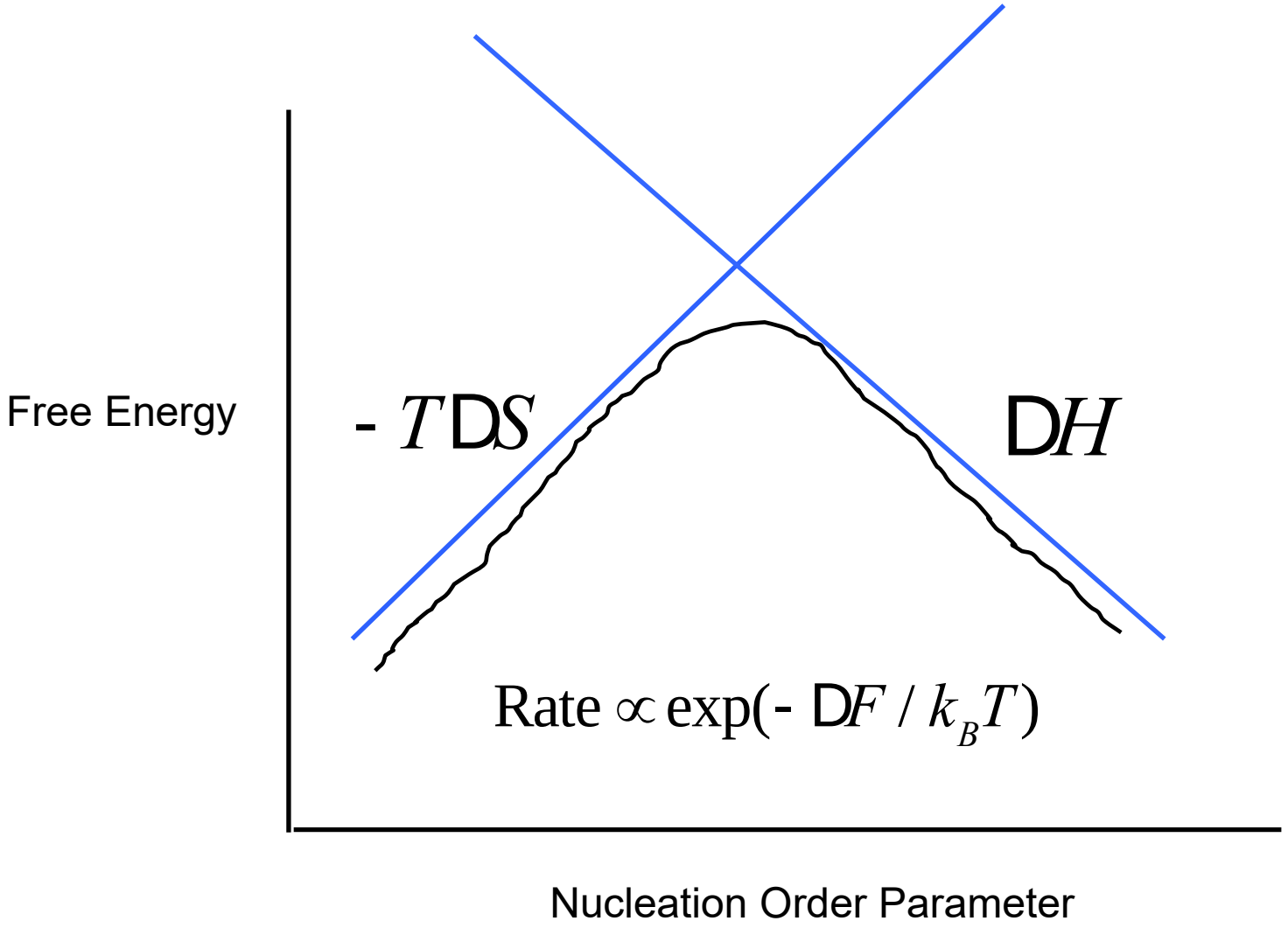


UCG simulation



Pornillos et al, 2011

Nucleation and Growth Phenomenology



Capsid inhibitors (CIs) have recently been reported with pM effective concentrations

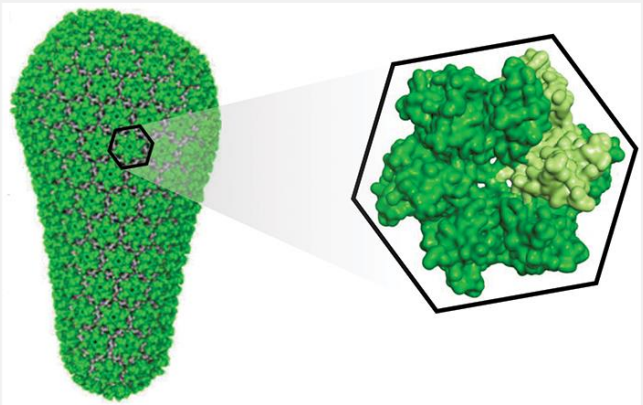
What is the unique mechanism of action for this class of drugs?

Volume 95 Issue 31 | pp. 23-25
Issue Date: July 31, 2017

Conquering HIV's capsid

After a dozen years, researchers have struck upon a molecule that can disrupt an elusive HIV target

By Lisa M. Jarvis



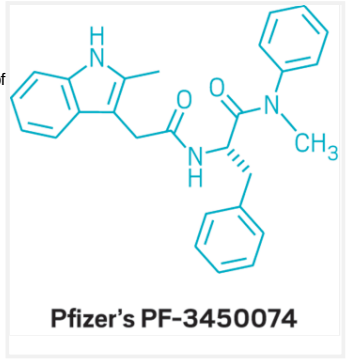
The HIV capsid (left) protects the viral genome so it can be delivered into host cells. Gilead's tool compound, GS-CA1 (light green, right), binds between two capsid proteins in the pinwheel-like hexamer.

Credit: Gilead Sciences

For most of his career at Gilead Sciences, medicinal chemist Winston Tse has lived and breathed one thing. While his peers at other companies hopped from project to project, Tse has spent the past decade obsessing over a single target: the HIV capsid.

HIV's capsid is a complex, protein-rich shell that protects the genetic payload the virus is [redacted] of 1,500 capsid proteins that organize themselves into hexamers and pentamers to form an eggplant-shaped shell. HIV researchers had no close-ups of the full capsid; a crystal structure had captured only the monomeric protein.

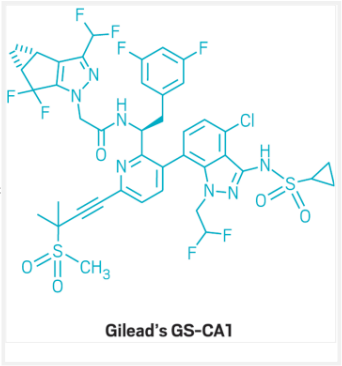
Moreover, scientists weren't—and still aren't—sure how the capsid assembles. Many envision something like a molecular knitting project that begins at the stem end of the eggplant and gets wider as rows of hexamers are added.



Pfizer's PF-3450074

Yet one thing was clear: Those 1,500 proteins need to knit together with just the right geometry and kinetics. "There is a real beauty in how geometrically structured it is," says Tomas Cihlar, vice president of biology at Gilead.

The shell needs to be stable enough to come together during virus maturation but still disassemble to expose its genetic payload once it is inside the host cell. That leads to a "delicate equilibrium in the whole capsid shell, which we thought could really be its Achilles' heel," Cihlar, who conceived of the capsid program back in 2006, adds.



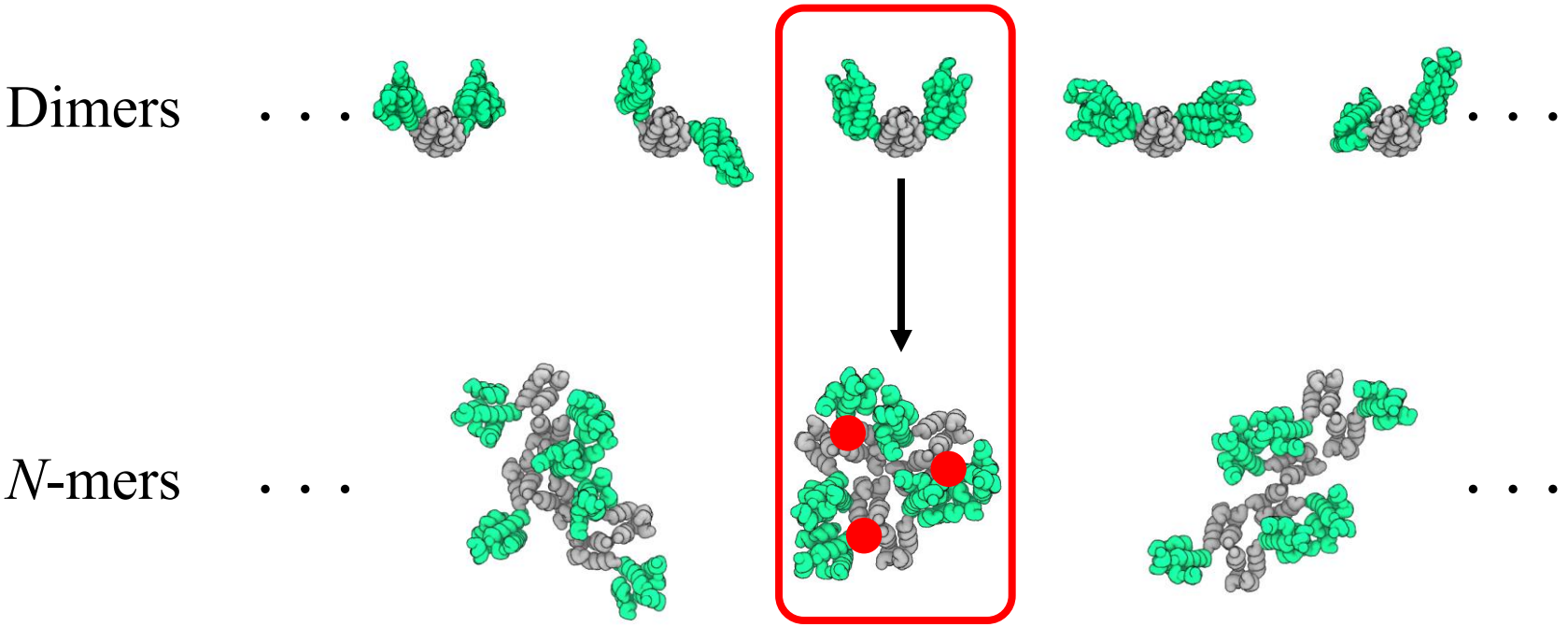
Gilead's GS-CA1

In addition to having limited structural information about the shell, Gilead researchers knew of no molecules that could convincingly bind to the capsid protein. The only clues in the literature were "some really

*A. J. Pak, J. M. A. Grime, A. Yu, and G. A. Voth, "Supercharged Assembly: A Broad-Spectrum Mechanism of Action for Drugs that Undermine Controlled HIV-1 Viral Capsid Formation", J. Am. Chem. Soc. **141,** 10214-10224 (2019).

*Yant et al, A highly potent long-acting small-molecule HIV-1 capsid inhibitor with efficacy in a humanized mouse model", Nature Medicine, **25** 1377–1384 (2019)

CIs stabilize small oligomers*

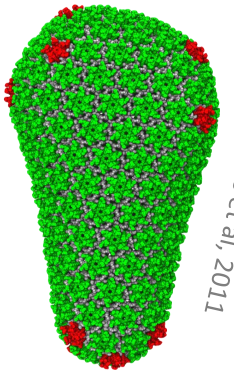


We investigate the effect of **small positive perturbations (~0.5-10 mol%)** to trimer of dimer populations **under conditions that otherwise result in canonical assembly**

*A. J. Pak et al., , J. Am. Chem. Soc. **141,** 10214-10224 (2019).

CIs promote structural pleomorphism…*

…primarily through rapid pentamer incorporation

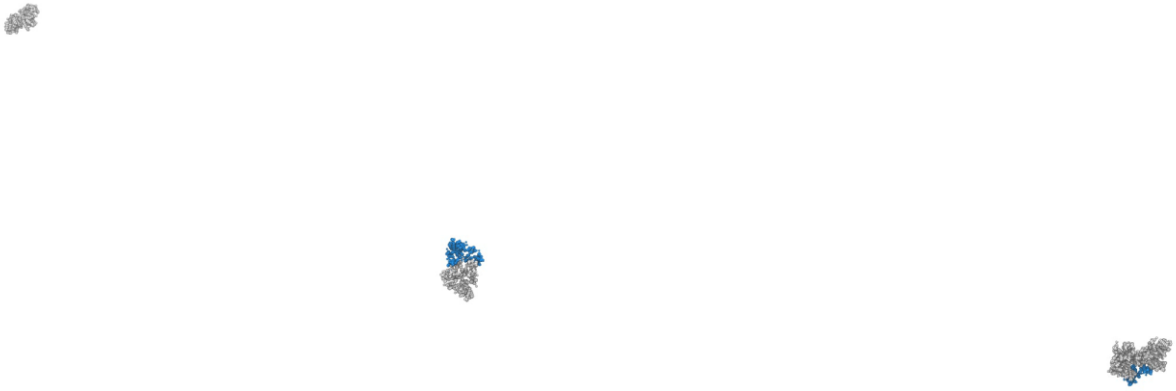


Hexamer

Pentamer

CI-bound Trimer

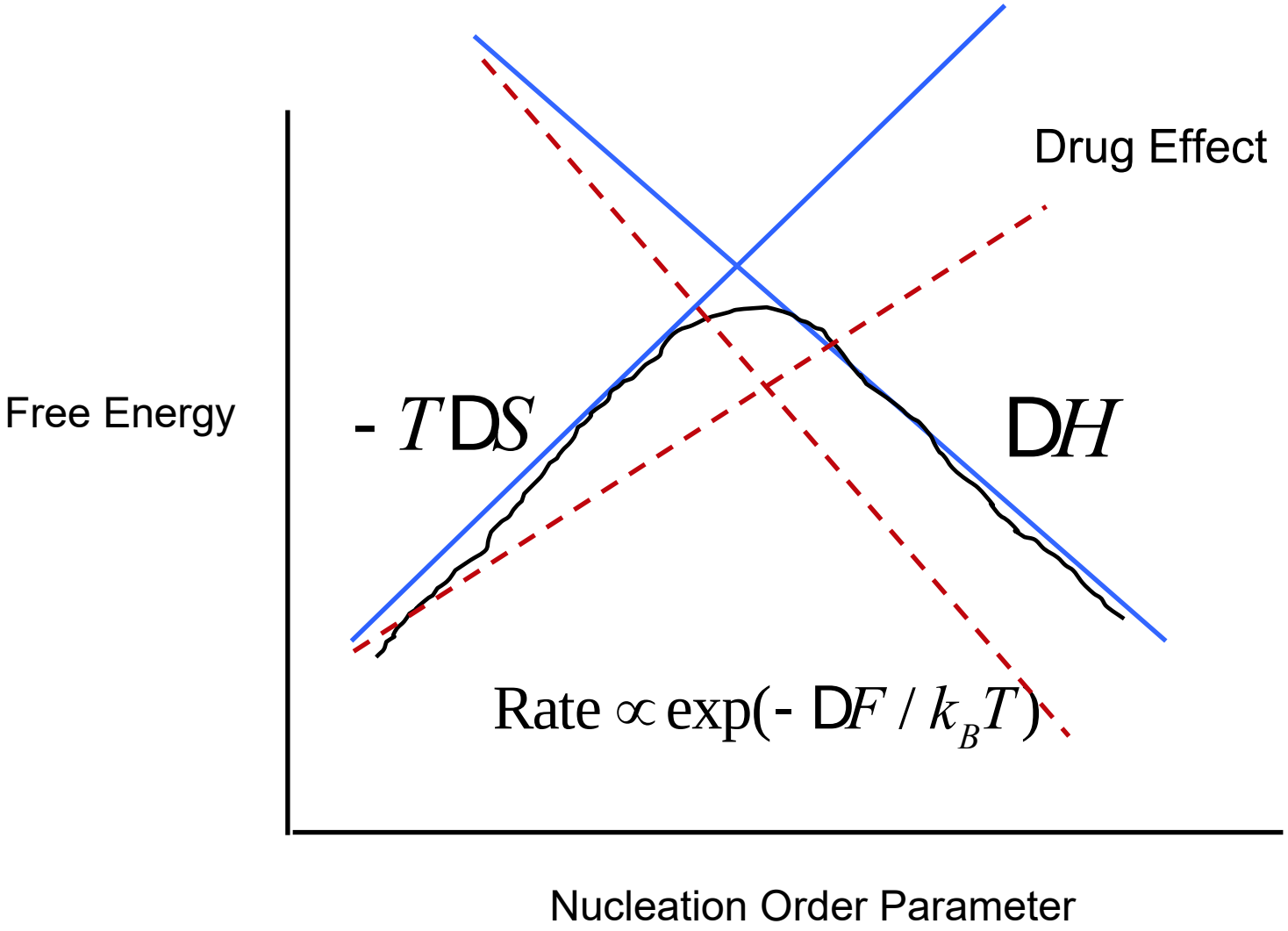
Defective or Incomplete Capsomer



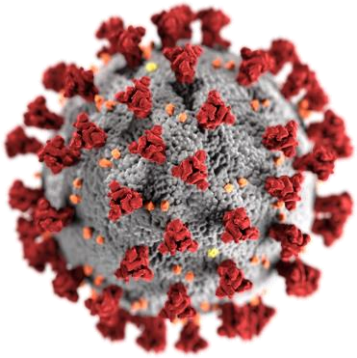
Drug Loading

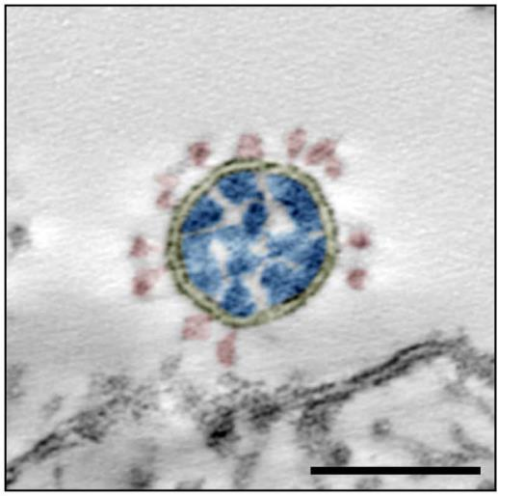
*A. J. Pak et al., , J. Am. Chem. Soc. (2019).

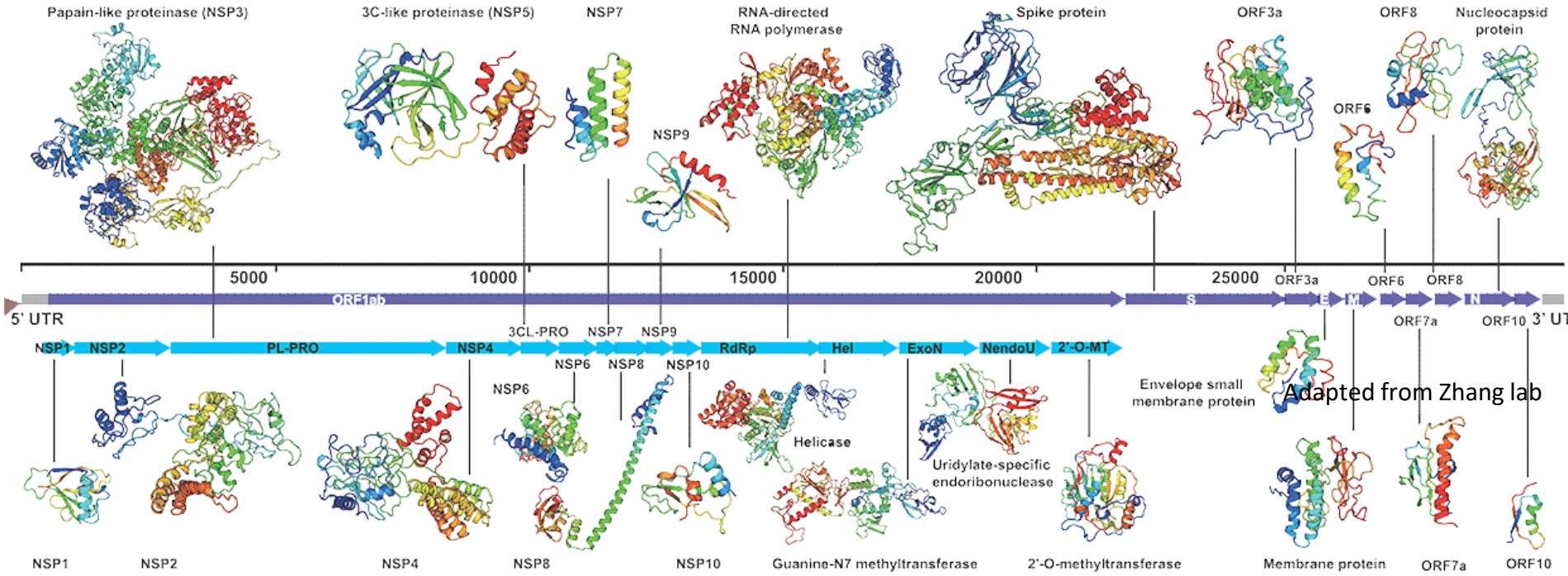
Nucleation and Growth Phenomenology



Coronavirus: The SARS-CoV-2 Virion

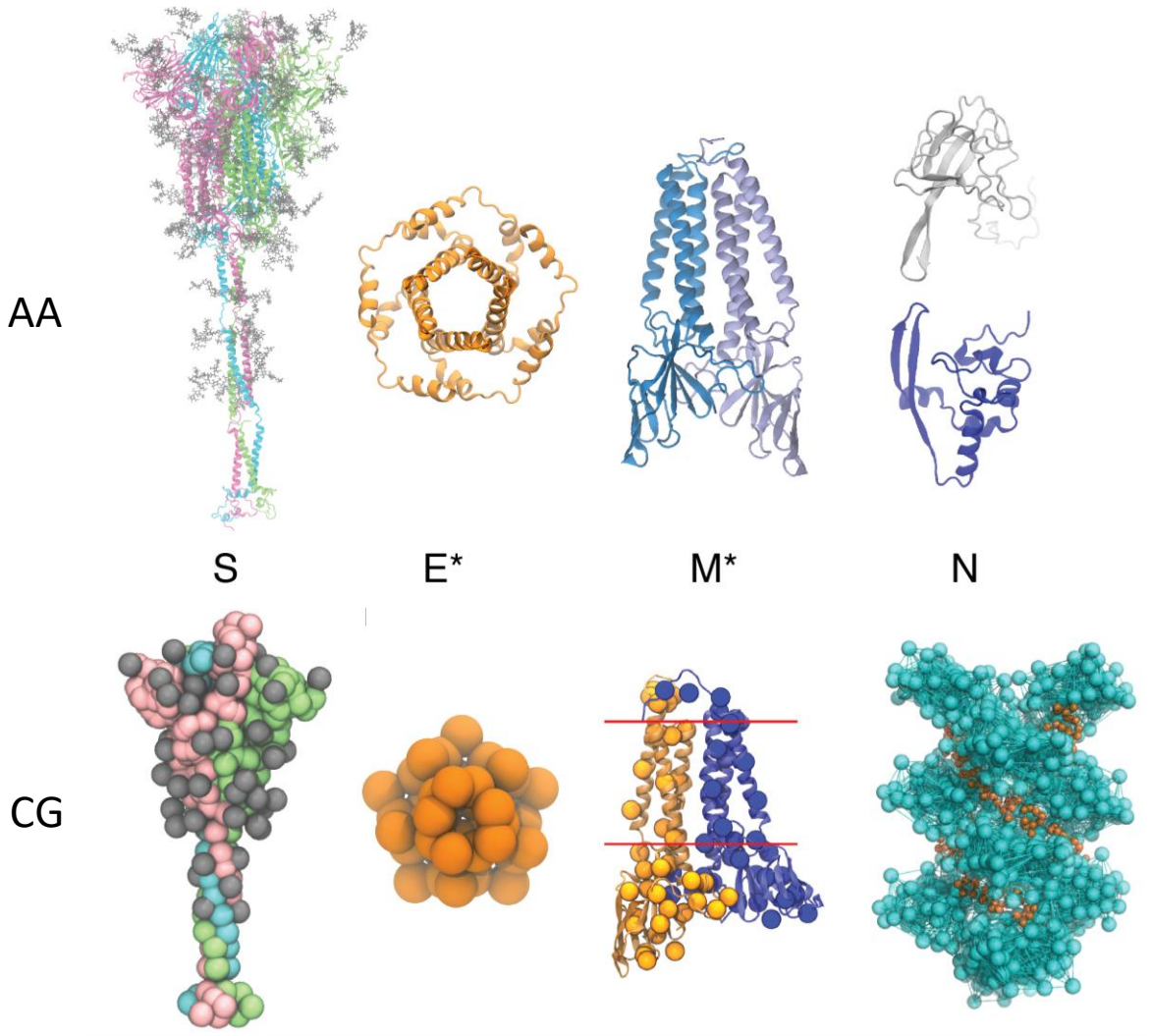


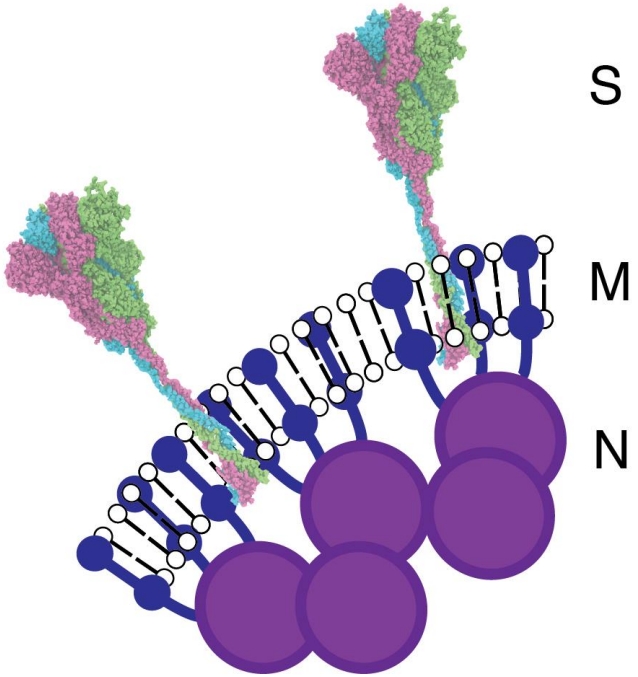




New structures are resolved or simulated nearly every week…

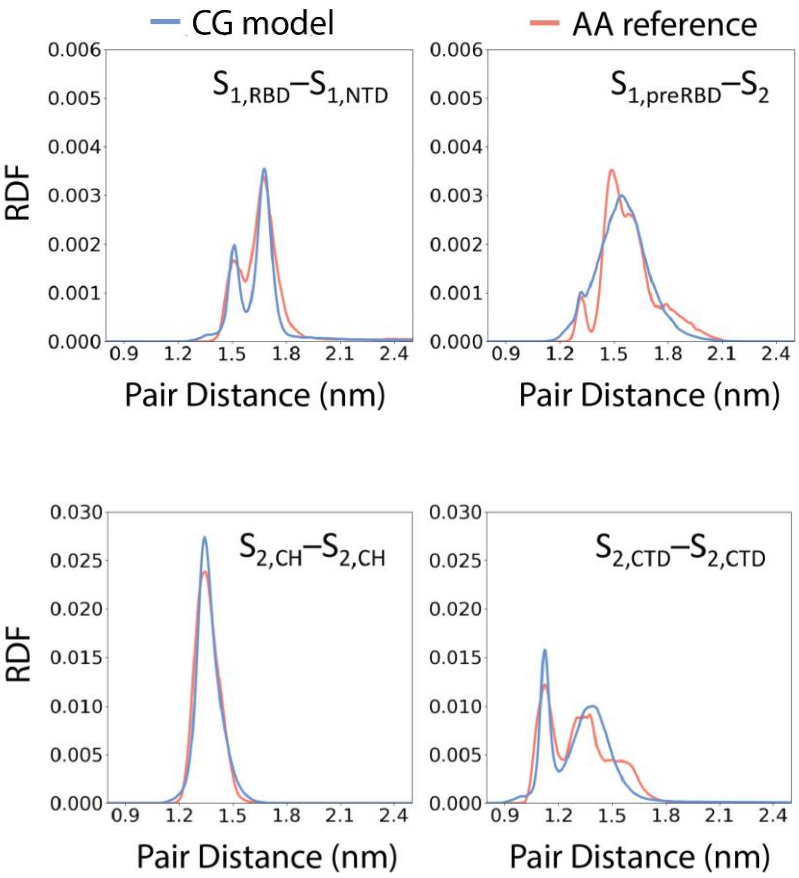
Modeling the Structural Proteins of SARS-CoV-2 Virion Envelope



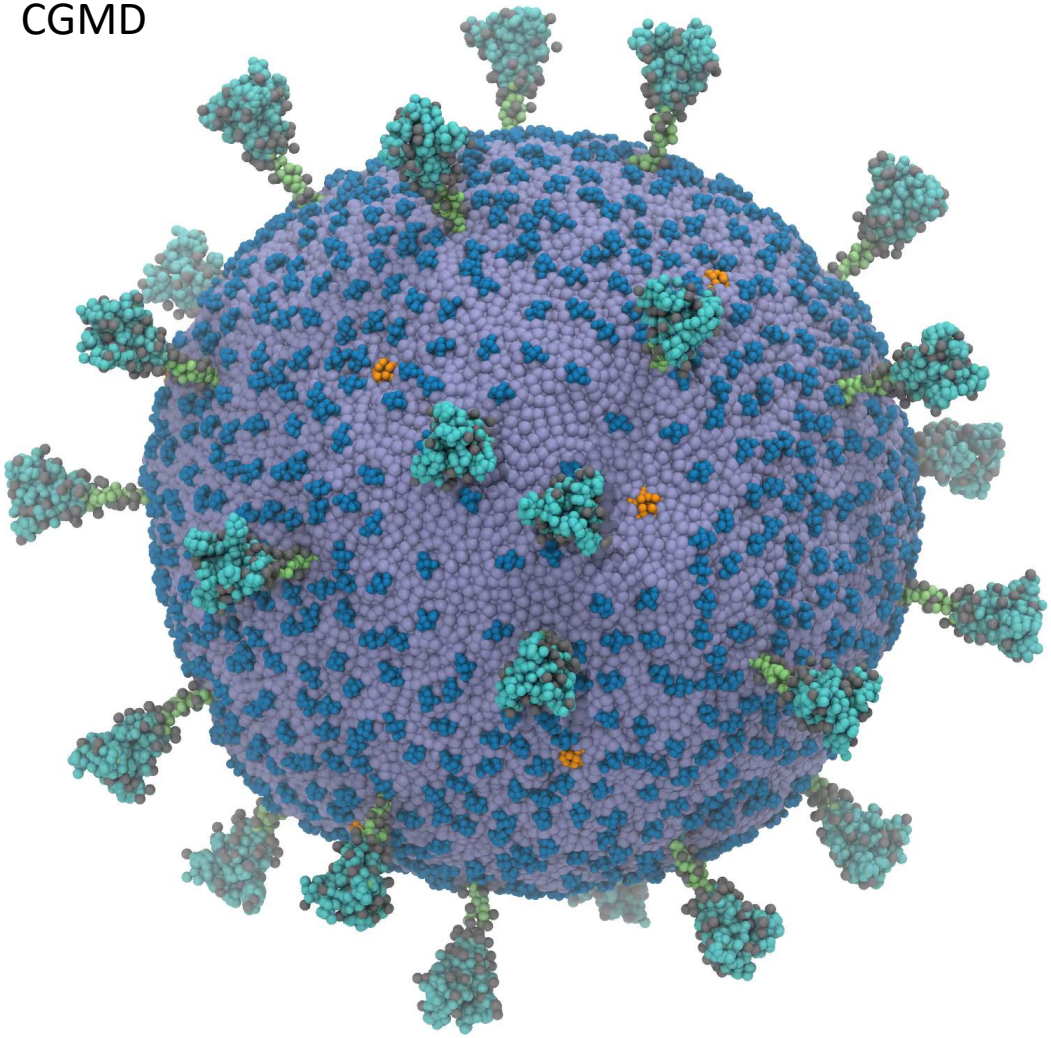


A. Yu et al. *Biophys J* (2021)

A Multiscale CG model of the SARS-CoV-2 Virion

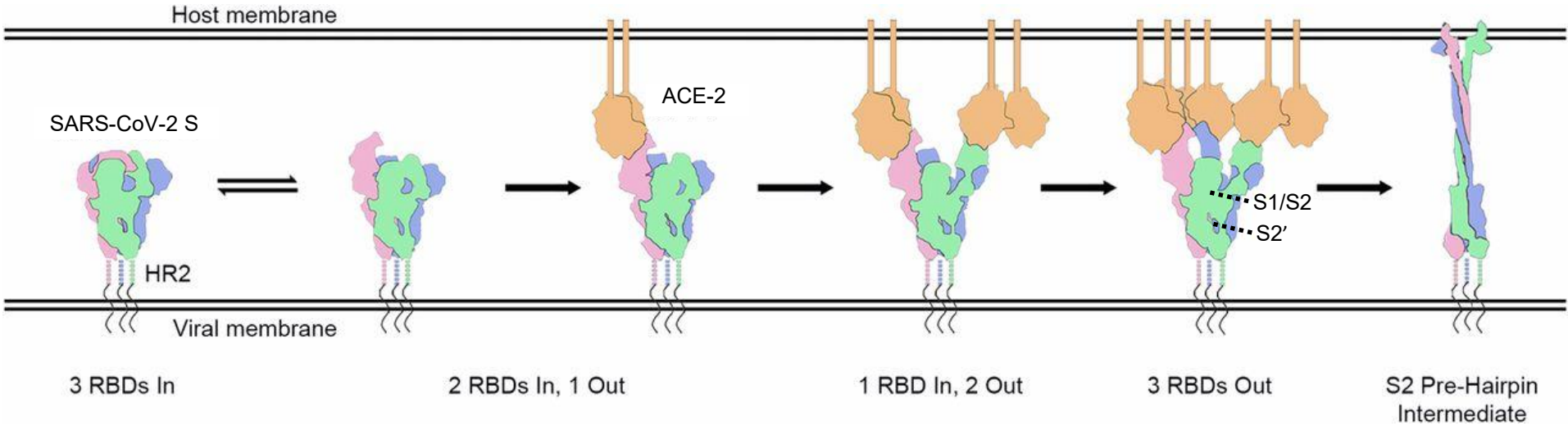


CGMD

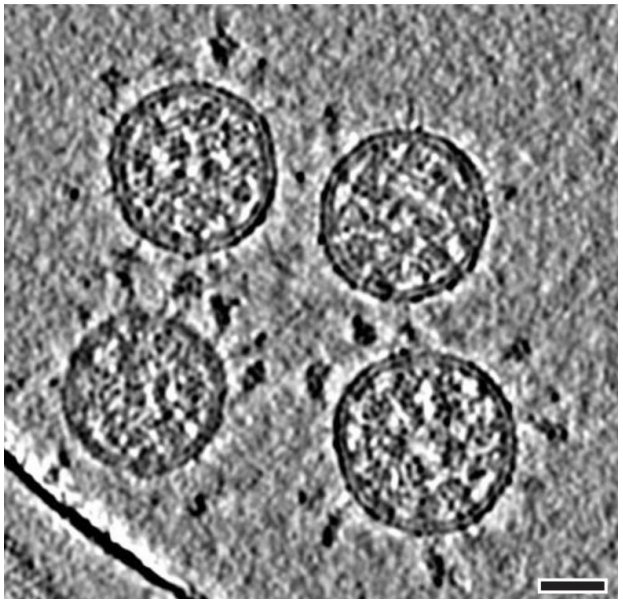


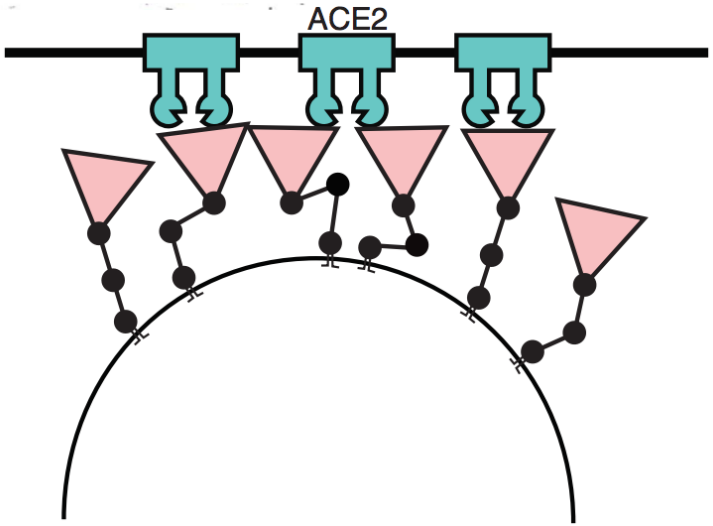
A. Yu et al. *Biophys J* (2021)

SARS-CoV-2 Spike Proteins Mediate Host Cell Entry



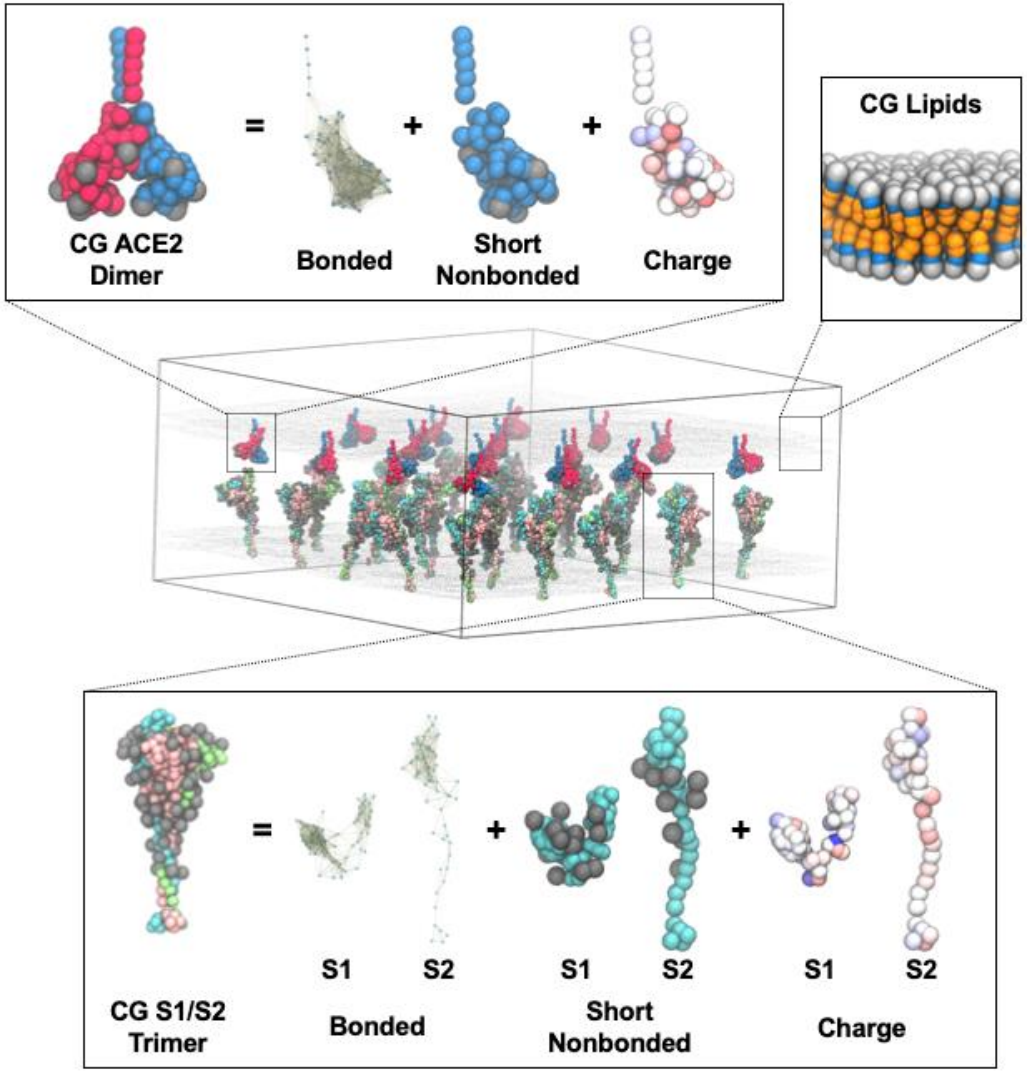
Adapted from Pallesen et al. *PNAS* 2017





Turnova and Beck et al. *Science* 2020

Modeling the Spike/ACE2 Interface



Coarse-grained models:

- S1 and S2 glycoproteins
- ACE2 glycoproteins
- Lipid bilayer

CG Simulation Movies: Two examples shown (11 secs and 50 secs).

Stochastic process in terms of timing (including ACE2/S binding which is reversible) but the general order of events is preserved:

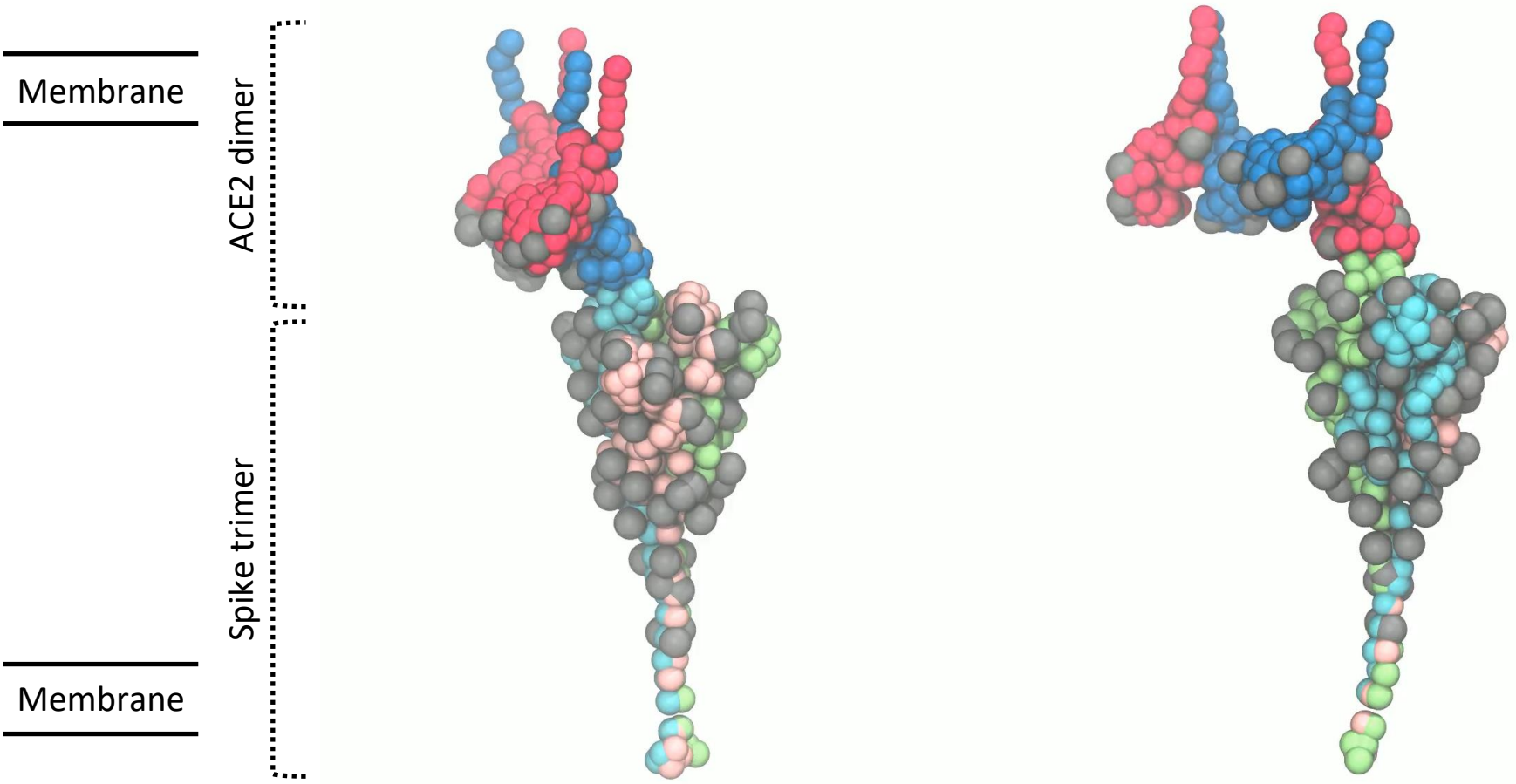
(1) First ACE2 binds

(2) Second ACE2 binds to the counter-clockwise protomer (think of the arrangement of the three S1 as a cyclical lock)

(3) Second ACE2 drags away this protomer

(4) The remaining two protomers dissociate (a combination of thermal fluctuations and “pulling” from ACE2)

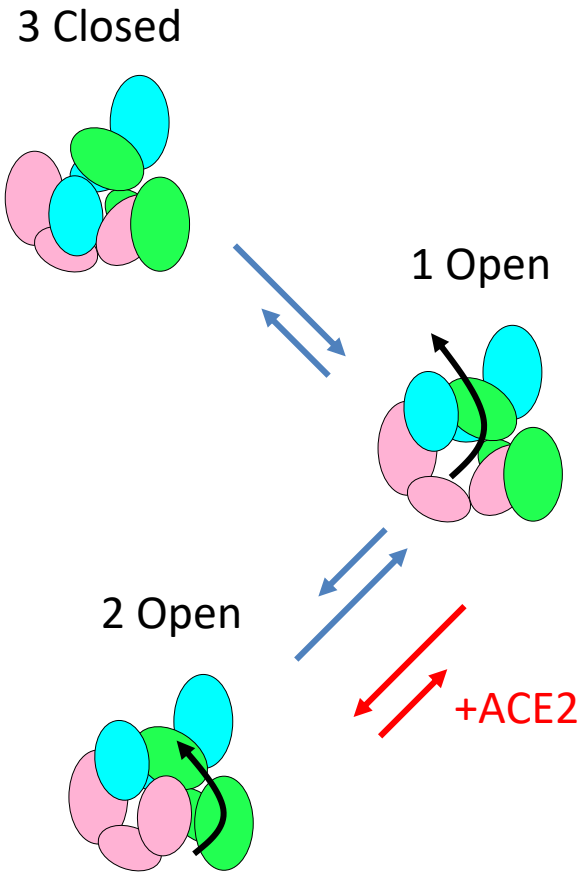
Multivalent ACE2 binding leads to S1 shedding

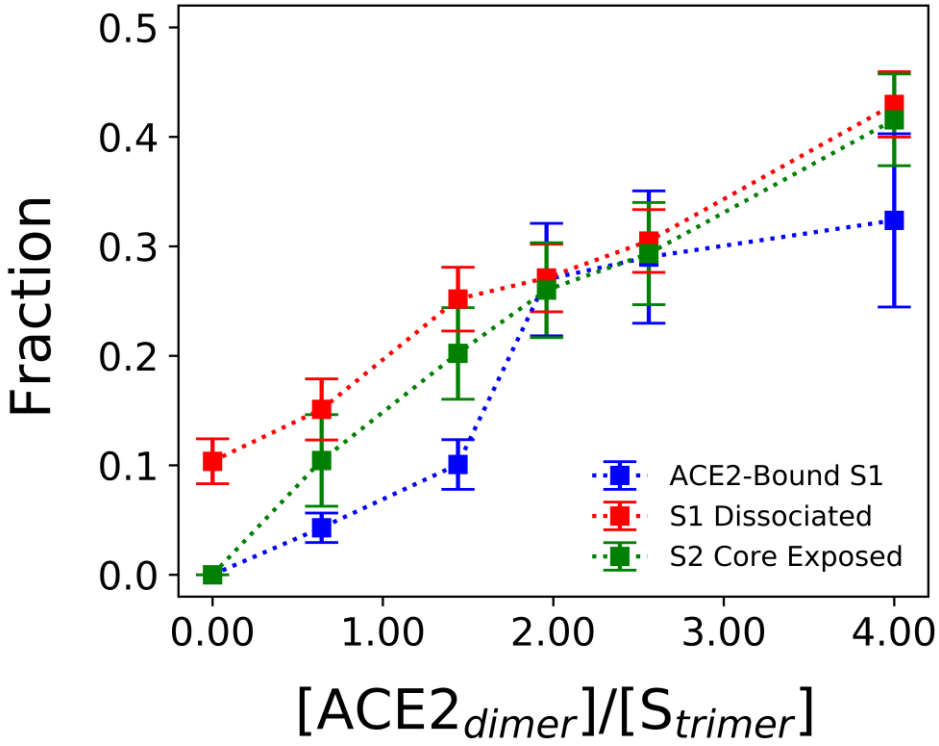


S1 dissociation proceeds processively in the *counter-clockwise* direction (top-view)

(red and blue balls are ACE2; cyan, pink, and green balls are spike; membrane not shown for clarity)

Cooperative binding is necessary for efficient exposure of the S2 trimeric core





A. Pak et al. *to be submitted* (2021)

Acknowledgements

Alvin Yu

Alex Pak

Funding

NSF RAPID

NIH-NIGMS