# Systematic Coarse-graining: Fundamentals and Applications

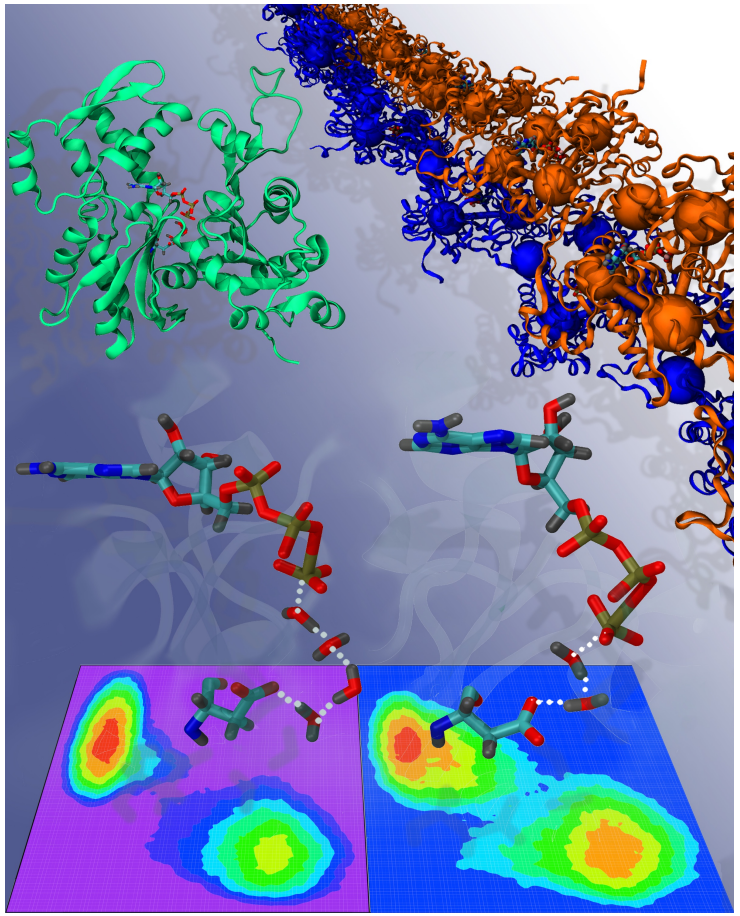

## Greg Voth

CENTER for MULTISCALE THEORY and SIMULATION

THE James Franck INSTITUTE
THE UNIVERSITY OF CHICAGO

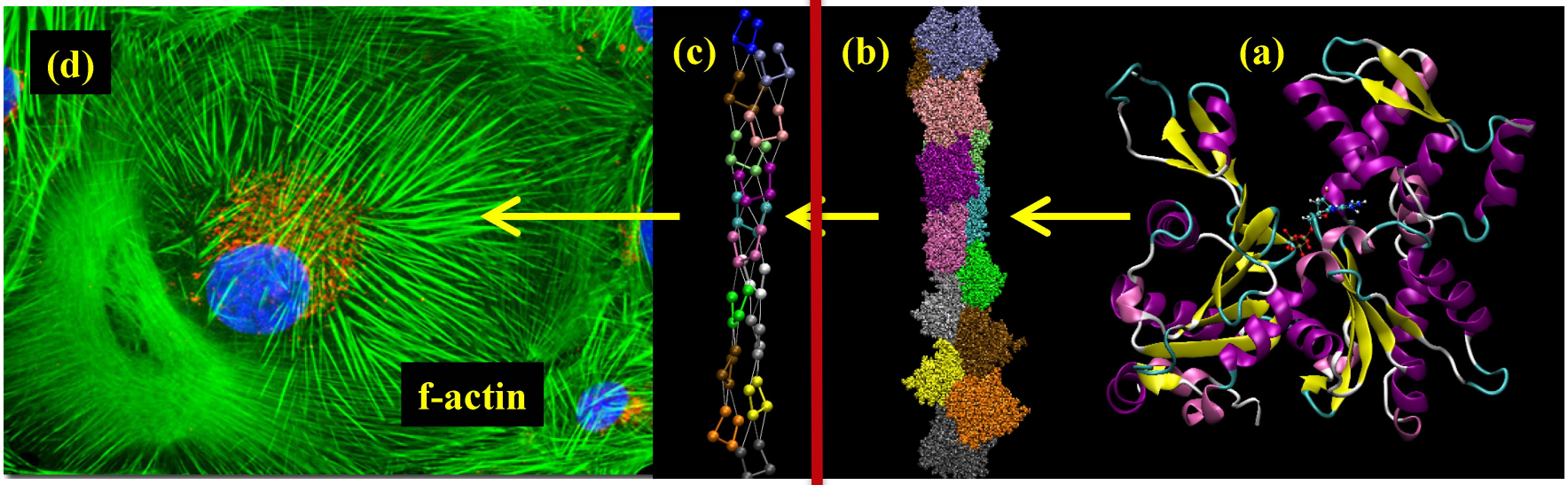
Institute for Biophysical Dynamics
Chicago BIOPHYSICS

THE VOTH GROUP
The University of Chicago | Department of Chemistry

# The Multiscale Challenge

## Physical-based Computer Simulation at the Scales of Cellular Biology




### Higher Scale and Multiscale Simulation

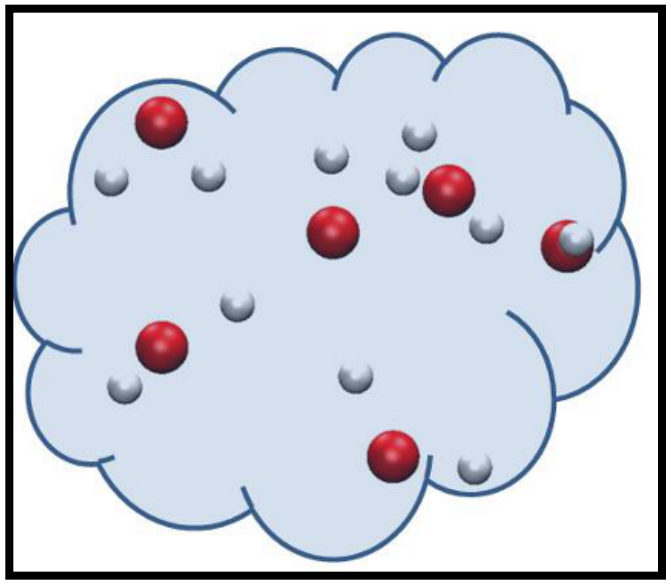
Closely tied to Cellular and Systems Biology, e.g., Various forms of Imaging, Cryo-Electron Tomography, Biochemical Networks, etc

### Atomistic and Near-Atomistic Molecular Dynamics Simulation

Closely tied to Structural and Molecular Biology, e.g., X-ray Crystallography, NMR, Single Particle cryo-EM, Spectroscopy, etc

# "Force" Behind the 2013 Nobel Prize in Chemistry?
# "Coarse-graining Away" of Electronic Structure

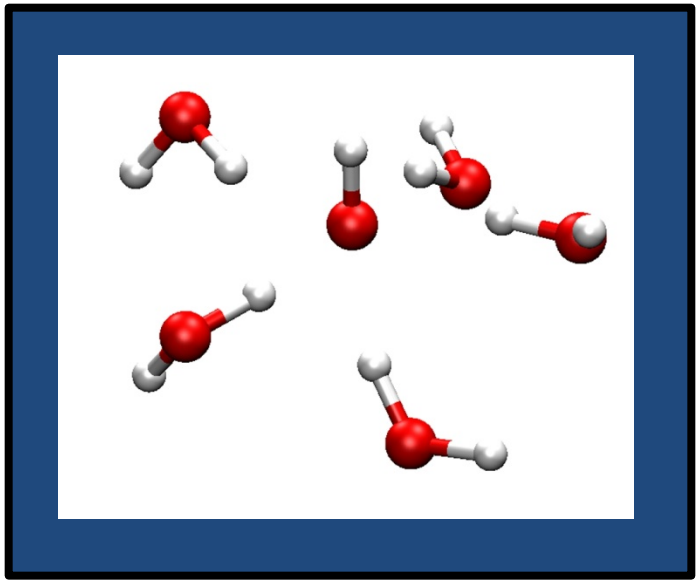
$\mathbf{H}\Psi = E\Psi$



**Electronic degrees of freedom explicitly treated: DFT, MP2, etc...**

**"Coarse-graining" of wave functions to simpler force fields**

$H_{\mathrm{MM}}$



**Particles interact via a simpler "molecular mechanics" forcefield**

# The Concept of "Bottom-up" Coarse-graining

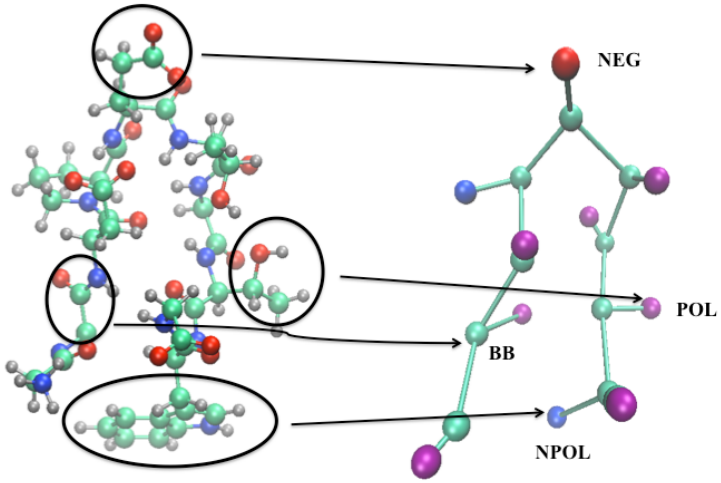
## Coarse-Graining can be based on Statistical Mechanics

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

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

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

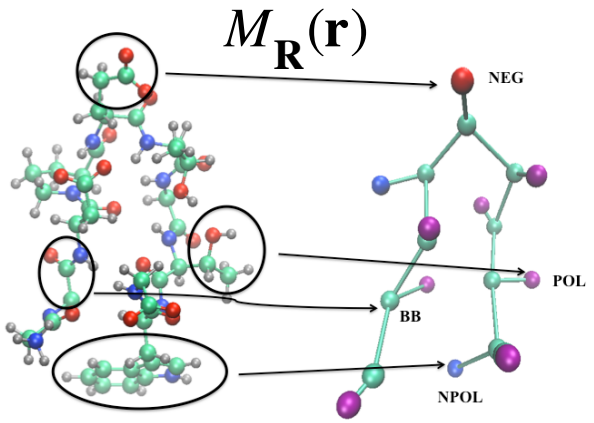
**How to determine $V_{CG}(\mathbf{R}_{CG})$ ?**




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

# Coarse-Graining Consistent with Statistical Mechanics: Mathematical Details

$$\left(N_{\mathbf{R}_{CG}} << N_{\mathbf{r}}\right)$$




**For a given $\mathbf{R}_{CG}$:**

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

$$\int d\mathbf{R}_{CG} \exp[-\beta V_{CG}(\mathbf{R}_{CG})] = \int d\mathbf{r} \exp[-\beta V(\mathbf{r})] \qquad (\beta = 1/k_B T)$$

$$\int d\mathbf{r} \exp[-\beta V(\mathbf{r})] = \int d\mathbf{r} \overbrace{\int d\mathbf{R}_{CG} \delta(M_{\mathbf{R}}(\mathbf{r}) - \mathbf{R}_{CG})}^{=1} \exp[-\beta V(\mathbf{r})]$$

$\rightleftarrows$

Switch Integration Order, Substitute and Subtract RHS

$$\int d\mathbf{R}_{CG} \left[ \exp[-\beta V_{CG}(\mathbf{R}_{CG})] - \int d\mathbf{r} \delta(M_{\mathbf{R}}(\mathbf{r}) - \mathbf{R}_{CG}) \exp[-\beta V(\mathbf{r})] \right] = 0$$

For integral to be *strictly zero* for arbitratry $V(\mathbf{r})$, it follows that...

**(Stat Mech Consistency)**

$\downarrow$

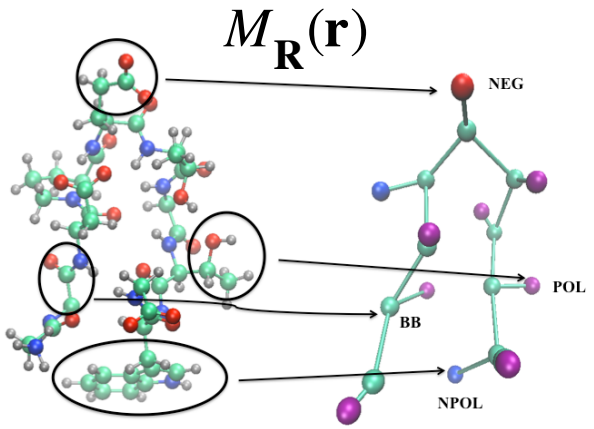
$$\exp[-\beta V_{CG}(\mathbf{R}_{CG})] \equiv \int d\mathbf{r} \delta(M_{\mathbf{R}}(\mathbf{r}) - \mathbf{R}_{CG}) \exp[-\beta V(\mathbf{r})]$$

THE JOURNAL OF CHEMICAL PHYSICS **145**, 044108 (2016)

# On the representability problem and the physical meaning of coarse-grained models

Jacob W. Wagner,[a] James F. Dama,[a] Aleksander E. P. Durumeric, and Gregory A. Voth[b]
*Department of Chemistry, James Franck Institute, Institute for Biophysical Dynamics, and Computation Institute, The University of Chicago, Chicago, Illinois 60637, USA*

# Coarse-Graining Consistent with Statistical Mechanics: Mathematical Details




**For a given** $\mathbf{R}_{CG}$:

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

$$\int d\mathbf{R}_{CG} \exp[-\beta V_{CG}(\mathbf{R}_{CG})] = \int d\mathbf{r} \exp[-\beta V(\mathbf{r})] \qquad (\beta = 1/k_B T)$$

$$\int d\mathbf{r} \exp[-\beta V(\mathbf{r})] = \int d\mathbf{r} \overbrace{\int d\mathbf{R}_{CG} \delta(M_{\mathbf{R}}(\mathbf{r}) - \mathbf{R}_{CG})}^{=1} \exp[-\beta V(\mathbf{r})]$$

$\rightleftarrows$

Switch Integration Order, Substitute and Subtract Right Hand Side

$$\int d\mathbf{R}_{CG} \left[ \exp[-\beta V_{CG}(\mathbf{R}_{CG})] - \int d\mathbf{r} \delta(M_{\mathbf{R}}(\mathbf{r}) - \mathbf{R}_{CG}) \exp[-\beta V(\mathbf{r})] \right] = 0$$

*For integral to be strictly zero for arbitratry* $V(\mathbf{r})$, it follows that... **(Stat Mech Consistency)**

$$\exp[-\beta V_{CG}(\mathbf{R}_{CG})] \equiv \int d\mathbf{r} \, \delta(M_{\mathbf{R}}(\mathbf{r}) - \mathbf{R}_{CG}) \exp[-\beta V(\mathbf{r})]$$

# The Multiscale Coarse-Graining (MS-CG) Variational Approach*

$\vec{\mathbf{r}}$ : **Atomic coordinates**

$\vec{\mathbf{R}}$ : **CG site coordinates**

$\vec{F}_\alpha^{CG}$ : **Exact CG force**

$V_{CG}$ : **Exact CG potential**

$\vec{F}_\alpha$ : **Instantaneous sum of atomic forces acting on the CG site**

$M$ : **Total number of CG sites**

$\chi^2$ : **The residual**

$\langle \ldots \rangle$ **average over configurations**

$$\chi^2\left[\vec{F}^{CG}\right] = \frac{1}{3M}\left\langle \sum_{\alpha=1}^{N_{CG}} \left| \vec{F}_\alpha^{CG}\left(\vec{\boldsymbol{R}}\right) - \vec{F}_\alpha\left(\vec{\boldsymbol{r}}\right) \right|^2 \right\rangle$$

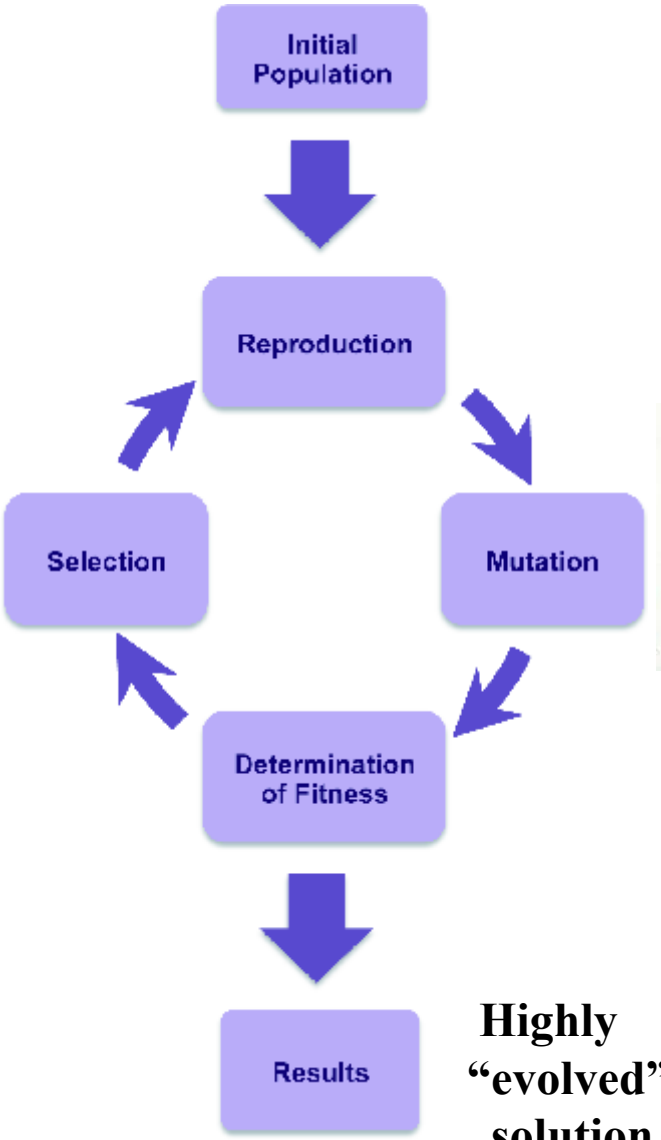
**Find** $\vec{F}_\alpha^{CG}\left(\vec{\boldsymbol{R}}\right)$ **from** $\dfrac{\delta\chi^2\left[\vec{F}^{CG}\right]}{\delta\vec{F}_\alpha^{CG}} = 0$

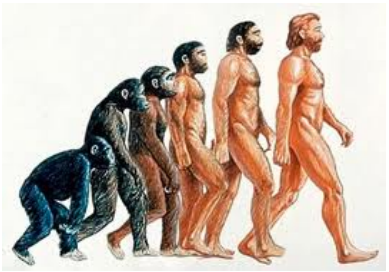
**Proven That:** $\vec{F}_\alpha^{CG}\left(\vec{\boldsymbol{R}}\right) = -\dfrac{\partial V_{CG}\left(\vec{\boldsymbol{R}}\right)}{\partial\vec{R}_\alpha}$

***S. Izvekov and GAV, J. Phys. Chem. B 109, 2469 (2005); J. Chem. Phys. 123, 134105 (2005); W. G. Noid, et al., J. Chem. Phys. 128, 244114 (1-11) (2008); 128, 244115 (1-20) (2008).**

# Force Matching and Genetic Algorithm (aka "Machine Learning")

$$\chi^2 = \left\langle \sum_{j=1}^{N_{CG}} \left| \mathbf{F}_j^{CG} - \mathbf{F}_j^{AA} \right|^2 \right\rangle$$



**Highly "evolved" solution**

??

# The MS-CG Algorithm

**(1) Assume pair wise decomposable radial non-bonded forces:**

$$\vec{F}_\alpha^{CG} \xrightarrow[with]{approximate} \vec{F}_\alpha^{MS} \qquad \vec{F}_\alpha^{MS} = \sum_{\beta\neq\alpha}^{\mathrm{M}} F_{\alpha\beta}^{MS}\left(R_{\alpha\beta};\phi\right)\vec{u}_{\alpha\beta}$$

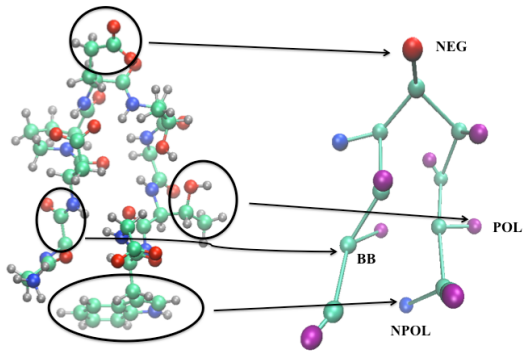
**(2) Expand all types of interactions as a *linear expansion of basis functions*:**

$$F_{\alpha\beta}^{MS}\left(R_{\alpha\beta};\phi\right) = \sum_{d=1}^{N_d}\phi_d f\left(R_{\alpha\beta},\left\{R_1,\ldots R_{N_d}\right\}\right)$$

**(3) Force matching becomes a linear least squares problem:**

$$\chi^2[\phi] = \frac{1}{3M}\left\langle\sum_{\alpha=1}^{M}\left|\vec{F}_\alpha^{MS}\left(\vec{\boldsymbol{R}};\phi\right)-\vec{F}_\alpha\left(\vec{\boldsymbol{r}}\right)\right|^2\right\rangle$$

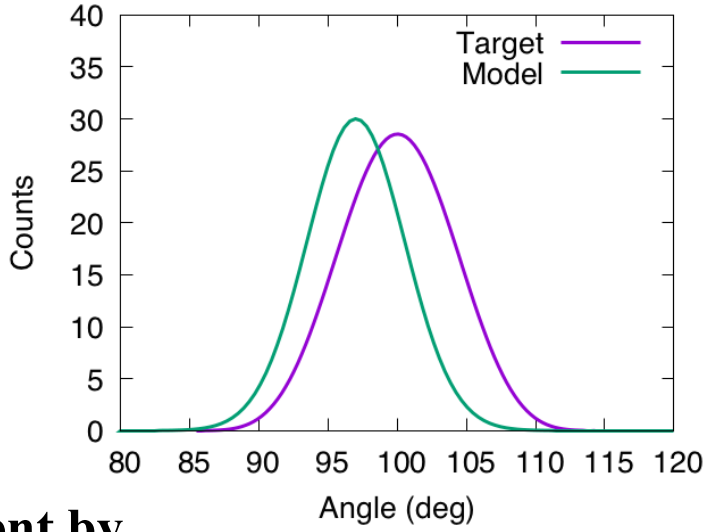
$$\frac{\delta\chi^2[\phi]}{\delta\phi_{d'}} = 0$$



# CG Models by Relative Entropy Minimization*

- **Relative entropy is a measure of "distance" between a model and target distribution**

$$S_{rel} = \int d\mathbf{r} p_T\left(M\left(\mathbf{r}^n\right)\right) \ln \frac{p_T\left(M\left(\mathbf{r}^n\right)\right)}{p_M\left(\mathbf{R}^N\right)}$$




- **Relative entropy can be applied to the canonical distribution**

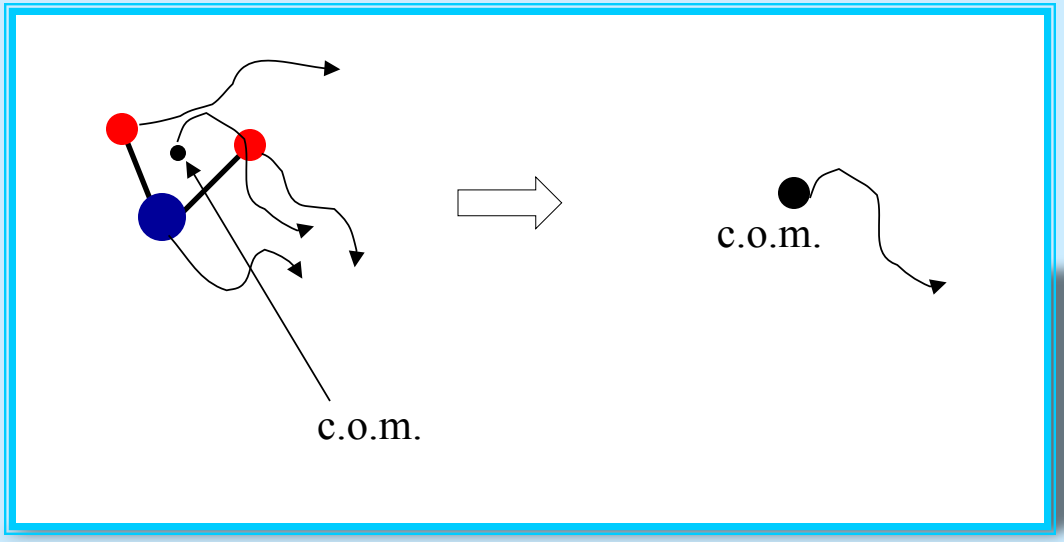
$$S_{rel} = \beta\left\langle U_M - U_T \right\rangle_T - \left(A_M - A_T\right) + \left\langle S_{map} \right\rangle_T$$

- **A CG model can be determined by gradient descent by taking derivatives of the relative entropy with respect to the basis set coeffs.**

$$\frac{\partial S_{rel}}{\partial \lambda_i} = \beta\left\langle \frac{\partial U_M\left(\mathbf{R}^N\right)}{\partial \lambda_i} \right\rangle_T - \beta\left\langle \frac{\partial U_M\left(\mathbf{R}^N\right)}{\partial \lambda_i} \right\rangle_M$$

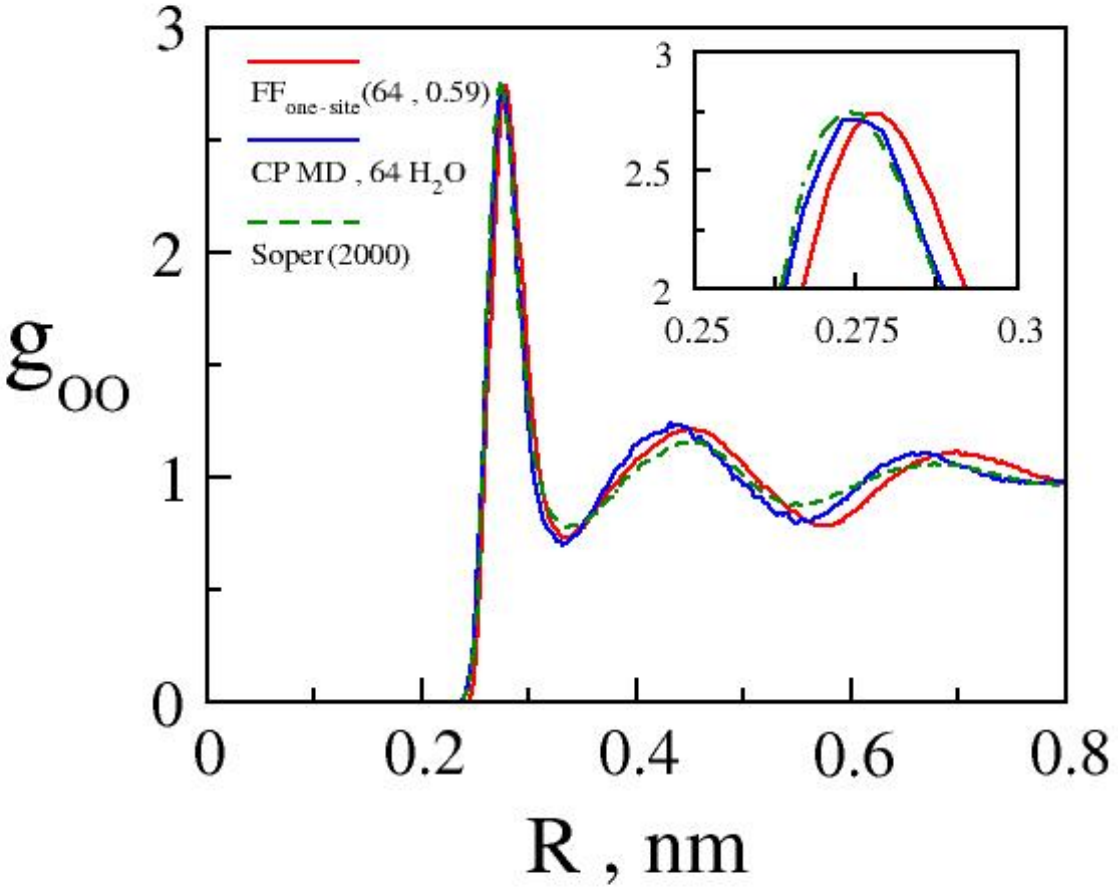
*Scott Shell and co-workers

# Example: One Site CG Water Models



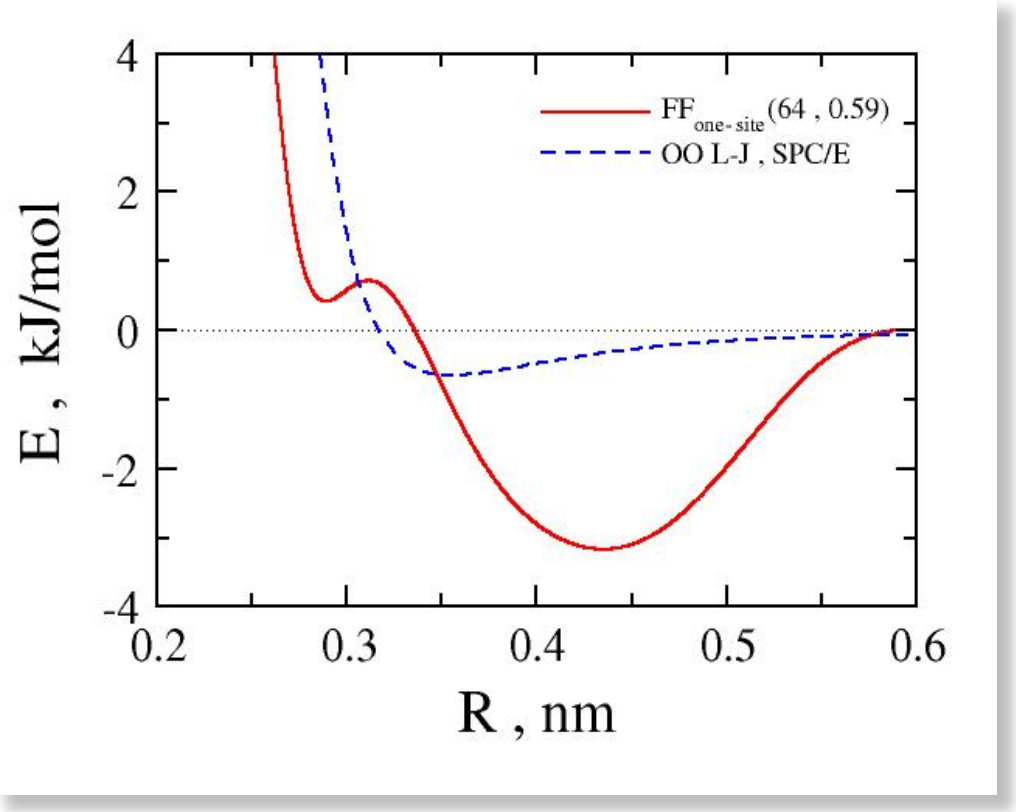

S. Izvekov and G. A. Voth, "Multiscale Coarse-Graining of Liquid State Systems," J. Chem. Phys. **123**, 134105(1-13) (2005).

# Results for One-Site CG Water




**Radial Distribution Function (RDF)**

# ONE-Site MS-CG Water Model




**One Bead CG Water Pair Potential**

$$V_{CG}(\mathbf{R}) = \frac{1}{2}\sum_{i<j} V_{CG,ij}(R_{ij})$$

# MS-CG Effective Potential Free Energy Decomposition*

$$V_{CG}(\mathbf{R}) = \frac{1}{2}\sum_{i<j} V_{CG,ij}(R_{ij},T)$$

$$V_{CG,ij}(R_{ij},T) = E_{CG,ij}(R_{ij},T) - TS_{CG,ij}(R_{ij},T)$$

Energetic/ Enthalpic

Entropic

**Entropy Calculation:** By numerical differentiation

$$-S_{CG,ij}(R,T) = [V_{CG,ij}(R,T+\Delta T) - V_{CG,ij}(R,T)]/\Delta T$$

**$E_{CG,ij}$ obtained through subtraction from MS-CG pair PMF:**

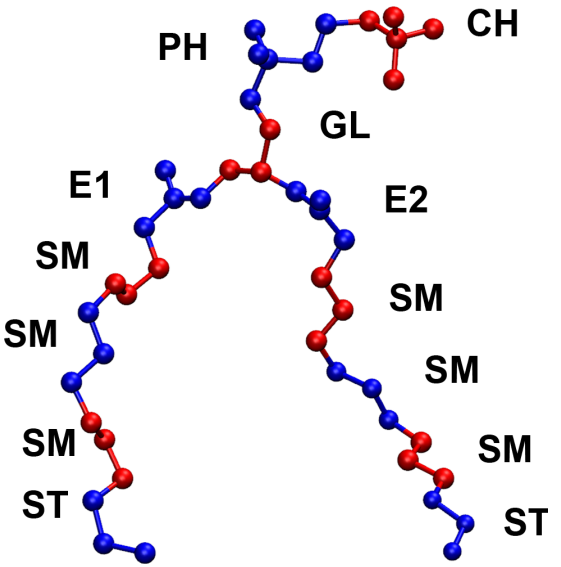
$$E_{CG,ij}(R) = V_{CG,ij}(R,T) - (-TS_{CG,ij}(R))$$

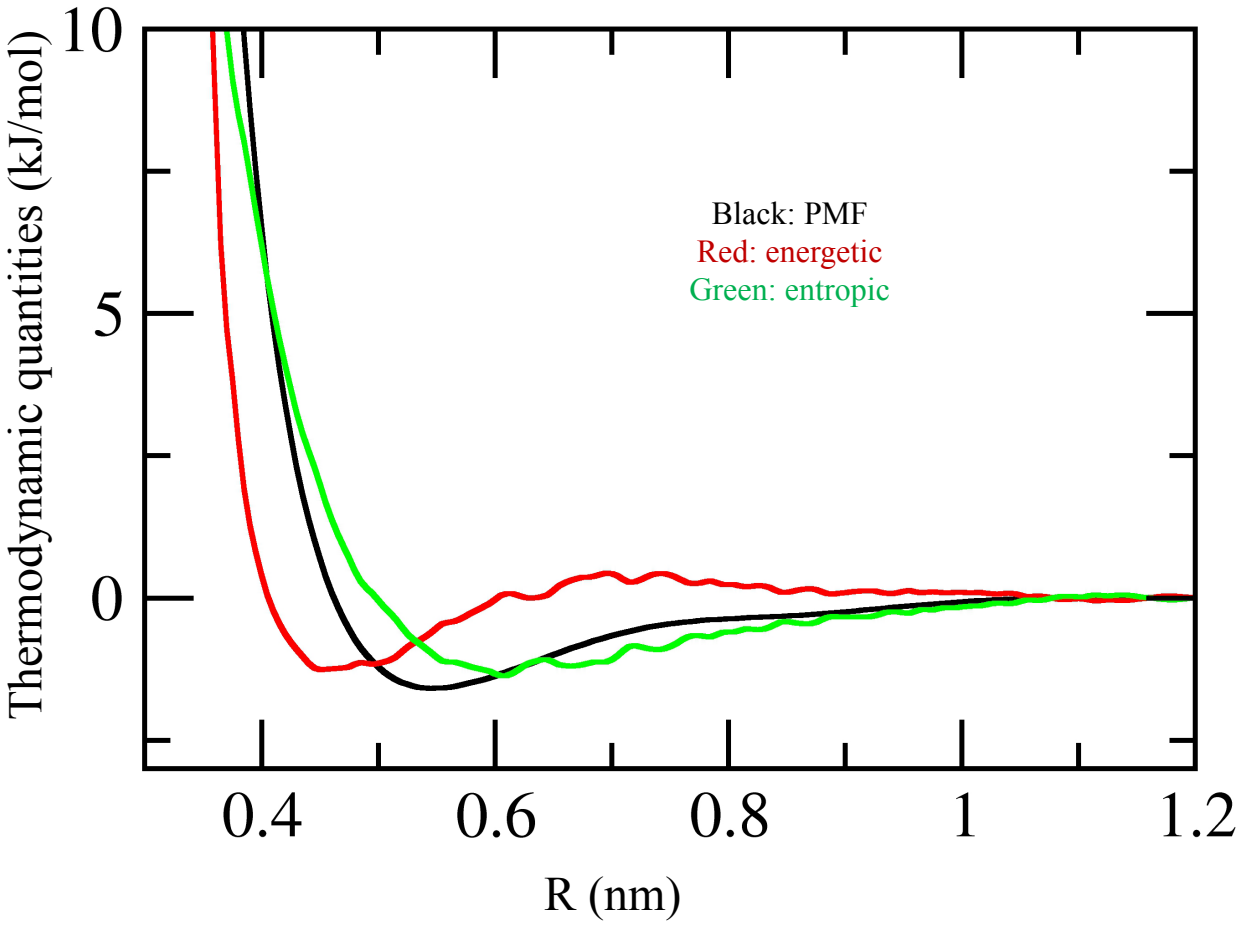
- Entropy is calculated as the slope from a linear fitting.
- Other analytical approximations of PMF can also be implemented.

* L. Lu and G. A. Voth, "The Multiscale Coarse-Graining Method. VII. Free Energy Decomposition of Coarse-Grained Effective Potentials", J. Chem. Phys. **134**, 224107 (2011).

# Coarse-grained DMPC Lipid Bilayer

Goal: To study the CG lipid tail-tail (SM-SM) interaction in a lipid bilayer.



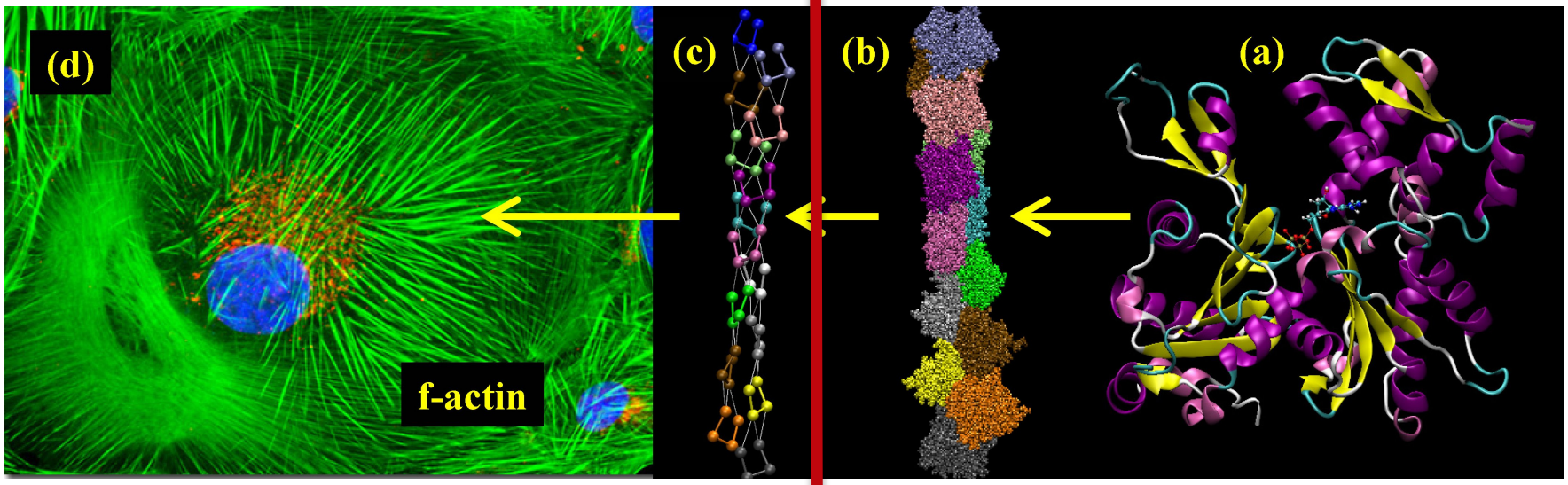




(a) The lipid tail conformational change produces an entropic effect.

(b) The tail excluded volume is increased consequently.

# The Multiscale Challenge

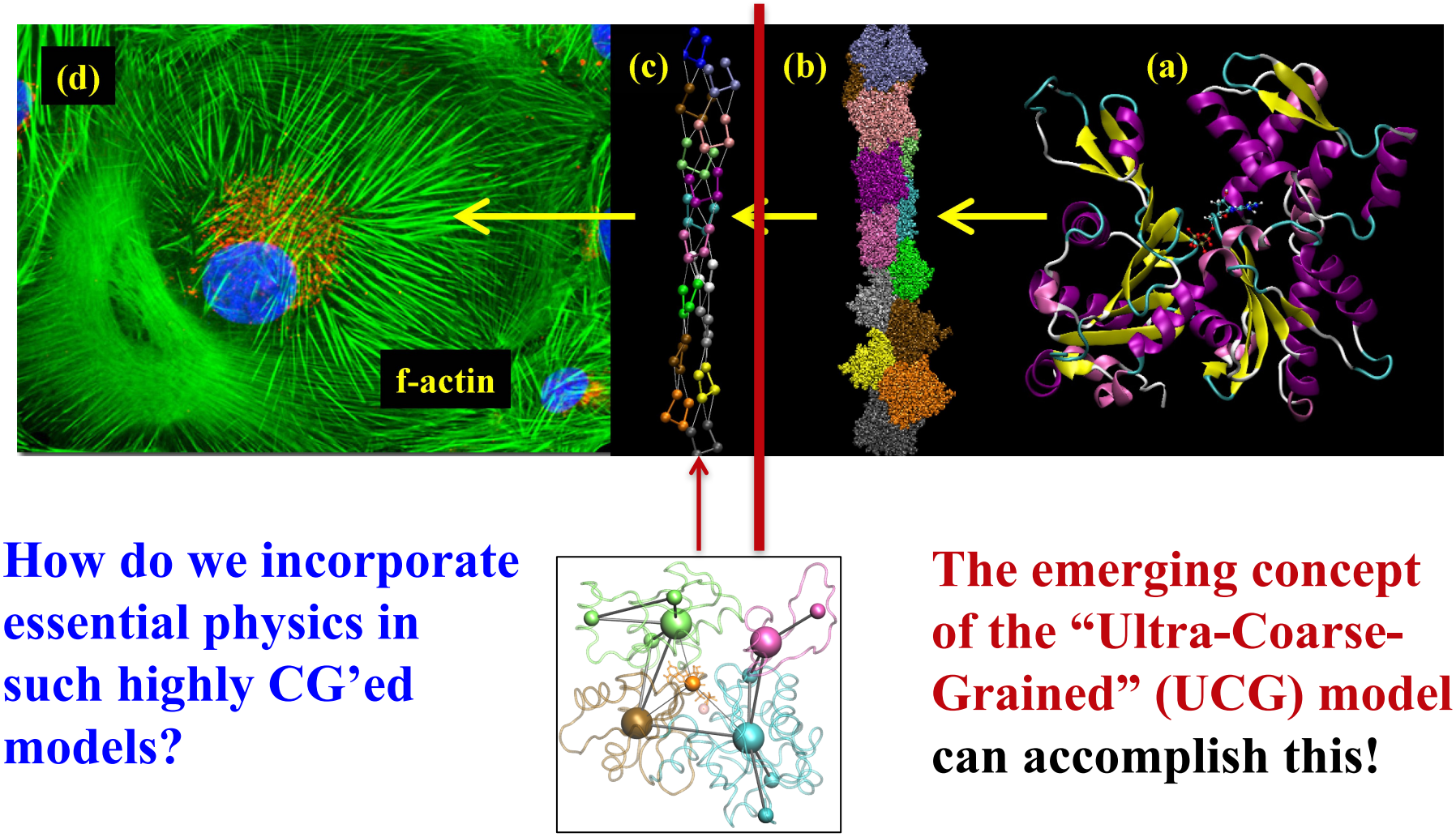
## Physical-based Computer Simulation at the Scales of Cellular Biology




### Higher Scale and Multiscale Simulation

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

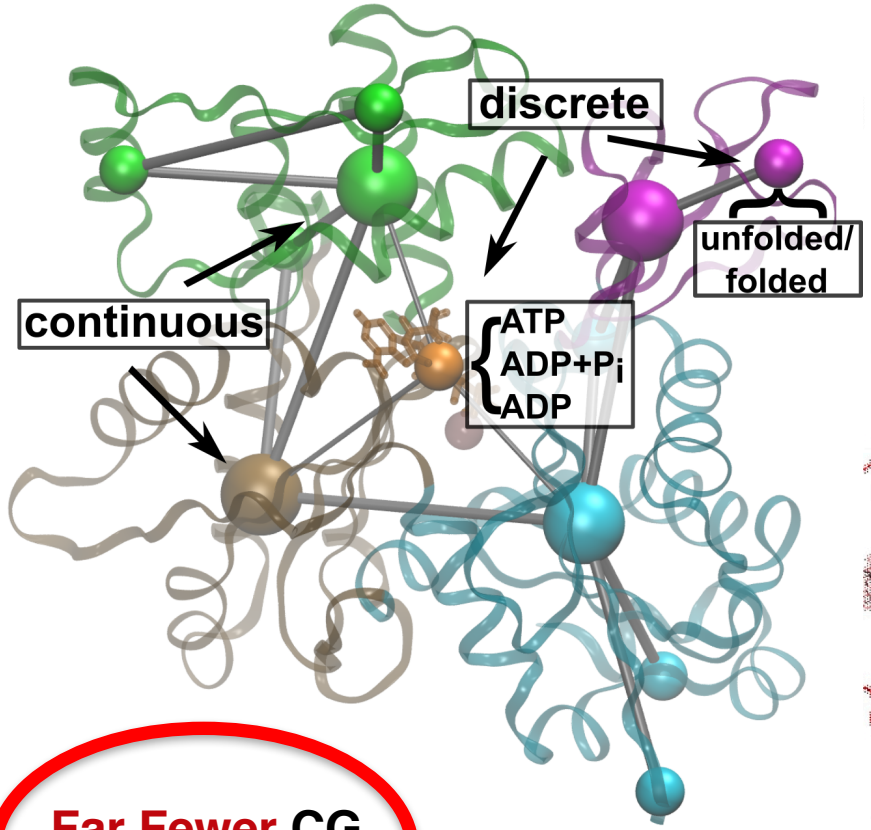
# The Multiscale Challenge

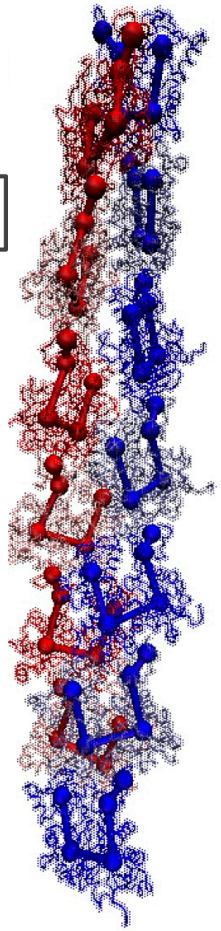



**How do we incorporate essential physics in such highly CG'ed models?**

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

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



**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**

**Far Fewer 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 ⇒**

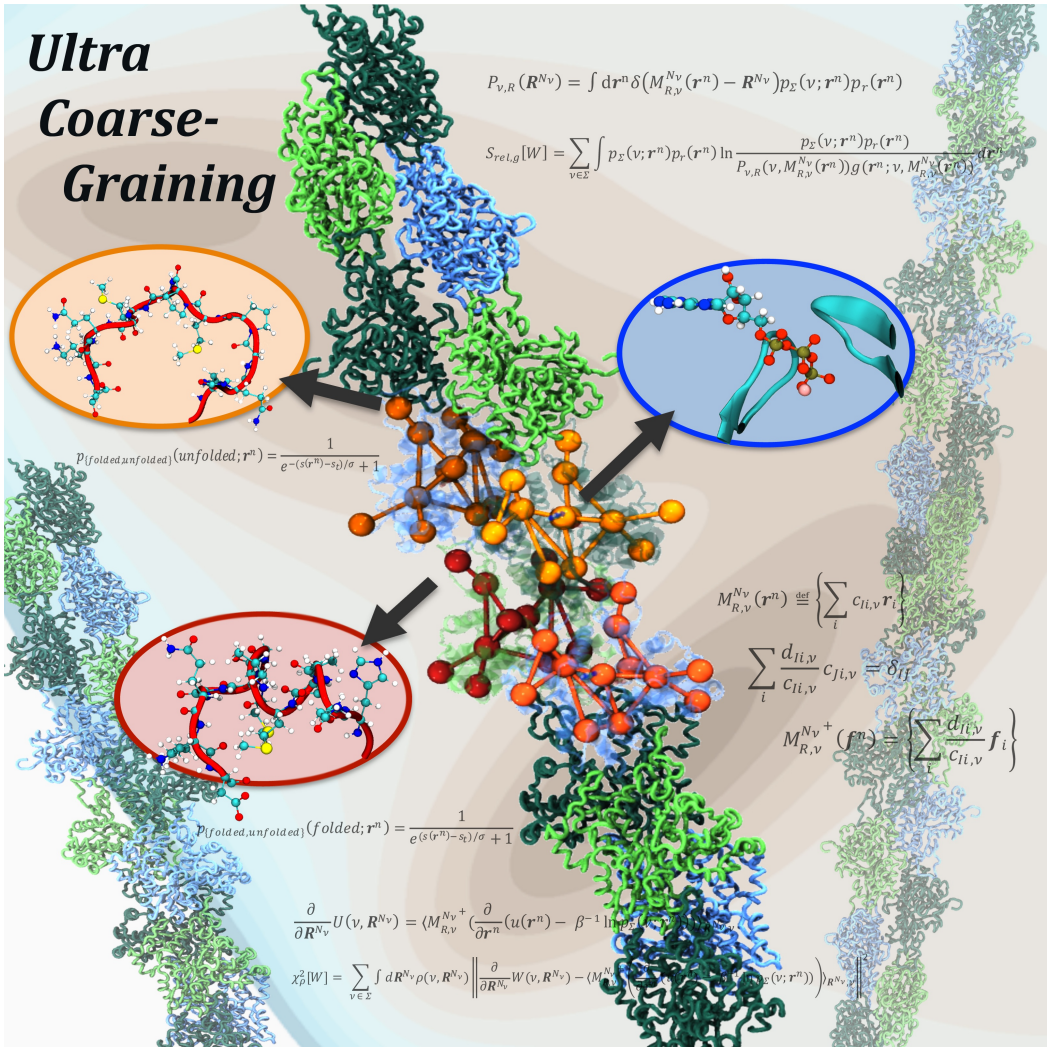
$$(\#\text{UCG States}) \times (\#\text{UCG Sites}) \ll (\#\text{Higher Res CG Sites}) < (\#\text{Atomic Sites})$$

# Origins of Possible "States" in the UCG Sites

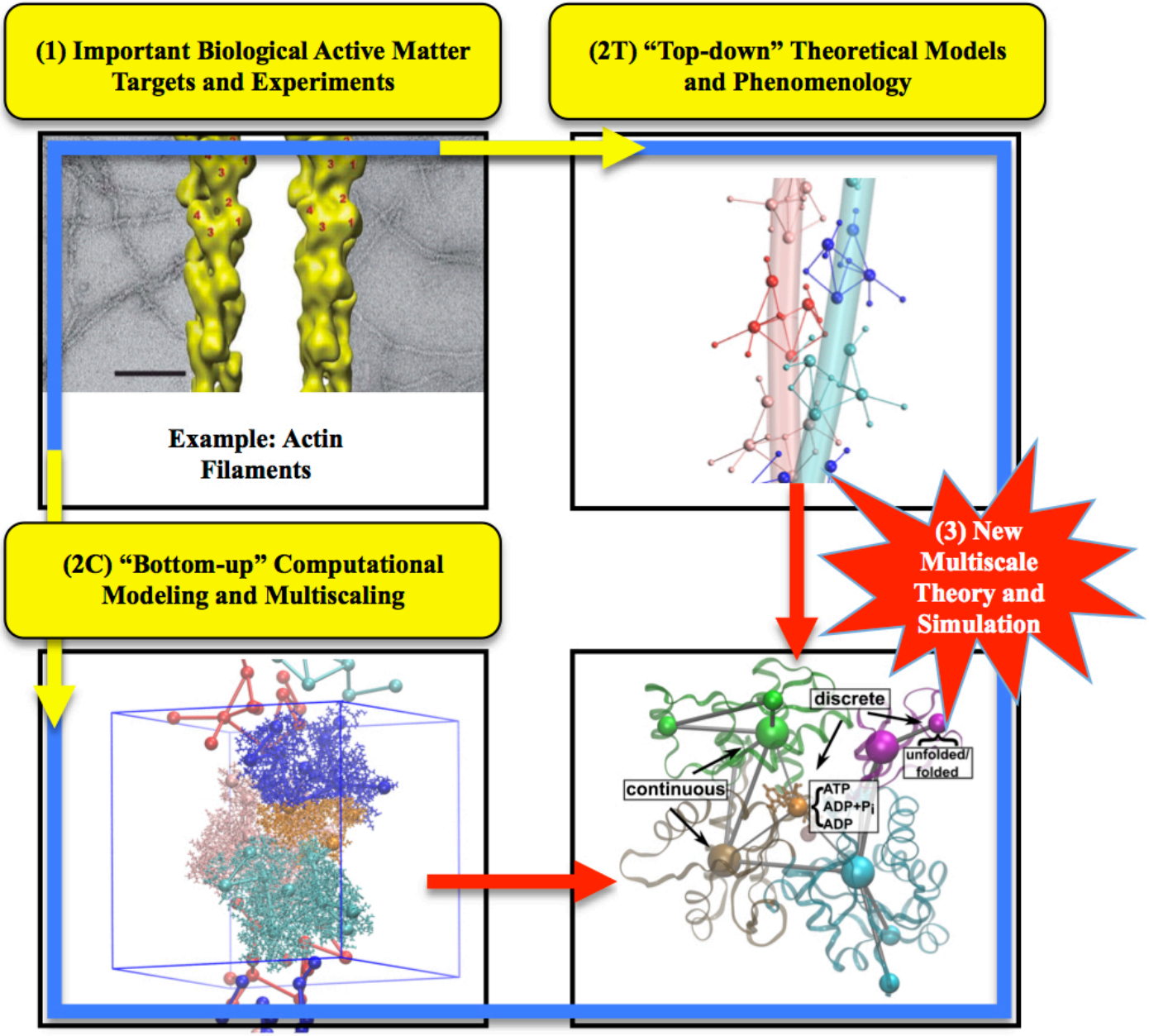
## States within UCG "beads"

**— physical —**
disorder transition
ligand binding
loop folding/unfolding

**— chemical —**
nucleotide hydrolysis
redox reaction
protonation



# Theory Provides New Directions in Multiscale Simulation



# What is the Influence on the UCG Time Evolution Equations from the Internal States?

- There should be some sort of **isomorphism to mixed quantum-classical evolution of nuclear motion on multiple potential energy surfaces**
  - But are there too many states?? *In principle*:

$$\text{Total \# of states} = M^{N_{CG}} \text{ (Yikes!!)}$$

where $M$ is # internal states, $N_{CG}$ is total # of UCG sites

- **No! Dynamics will be in the "decoherence" limit**; No-off diagonal density matrix elements: simpler equations for the remaining diagonal elements
- **Mean field-like solutions:** $\text{Total \# of states} \sim M \times N_{CG}$

# Two Understood Limits of UCG State Dynamics

$$\text{UCG Payoff} \Rightarrow (\#\,\text{UCG States}) \times (\#\,\text{UCG Sites}) \ll (\#\text{Higher Res CG Sites}) < (\#\text{All-Atom Sites})$$

- **Limit I**: States can **change infrequently**, leading to a **surface-hopping style dynamics**
  - Use a **local ansatz** for **rates**:

$$\text{Rate of state switch for } i = k(\{neighs\}_i)$$
$$\text{where } \{neighs\}_i \text{ is local configuration, } k \text{ is UCG rate}$$

- **Limit II**: States can **change frequently**, leading to an **adiabatic and Ehrenfest style dynamics**
  - Use a **local** ansatz for **populations**:

$$\text{Prob of state } \alpha \text{ for site } i = p_{i,\alpha}(\{neighs\}_i)$$
$$\text{where } \{neighs\}_i \text{ is local configuration, } p_{i,\alpha} \text{ is UCG occupation}$$

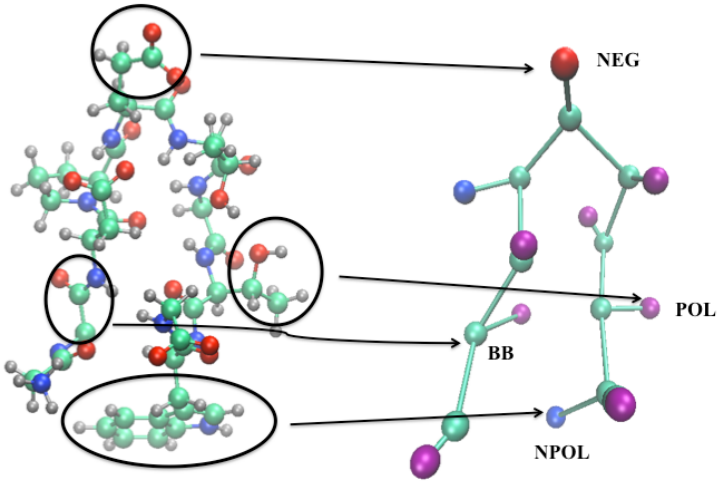
# New Life for "Higher Resolution" Coarse-graining?

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

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

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

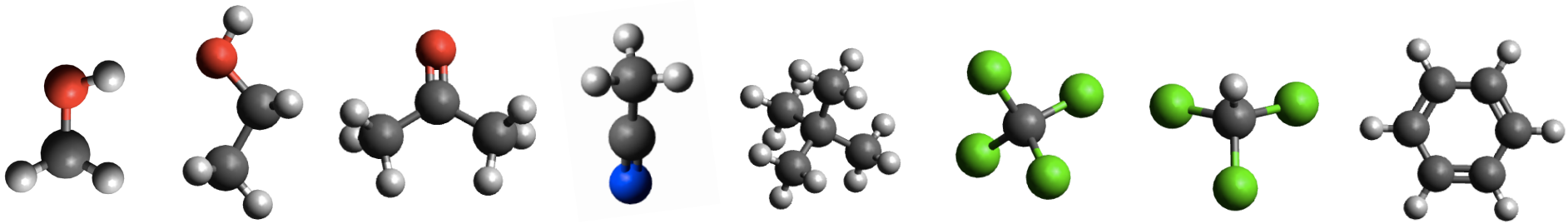
**How to determine $V_{CG}(\mathbf{R}_{CG})$ ?**



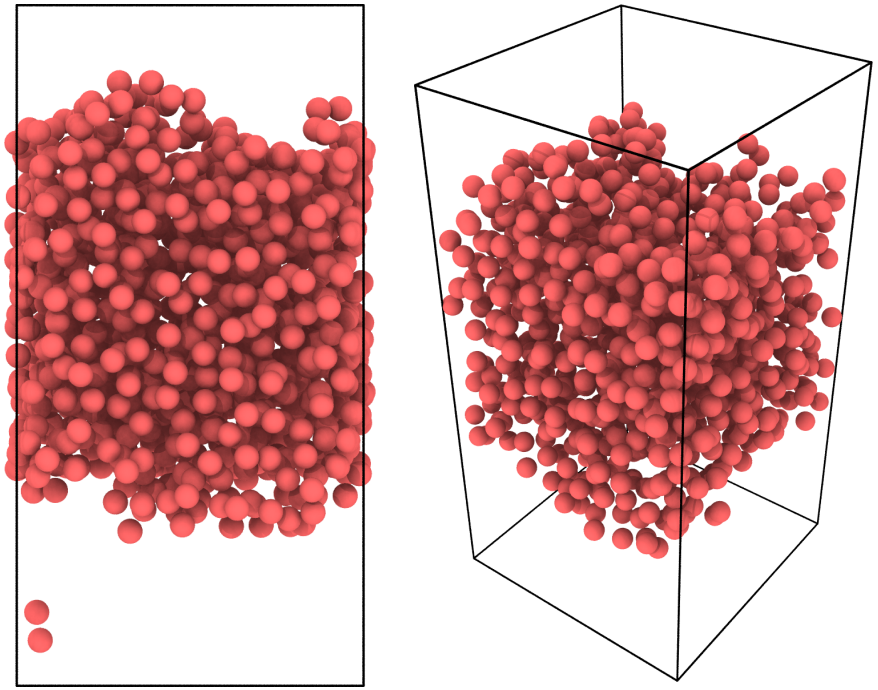

**Shown here is a "high resolution" CG model having some number of CG sites or "beads" per each amino acid residue in the peptide.**

# UCG Model for Liquid/Vapor Interfaces

**Target system**: Include different intermolecular interactions and molecular symmetry between molecules



Methanol Ethanol Acetone Acetonitrile Neopentane $CCl_4$ $CCl_3H$ Benzene

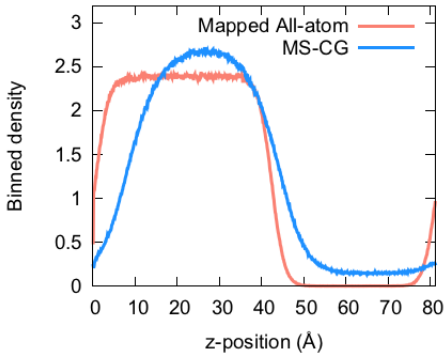


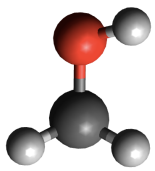
J. Jin, G. A. Voth. Submitted

- **All-atom simulation** is from OPLS/AA for 1,000 molecules
- **Coarse-grained simulation** is from one-site CG model (center-of-mass)

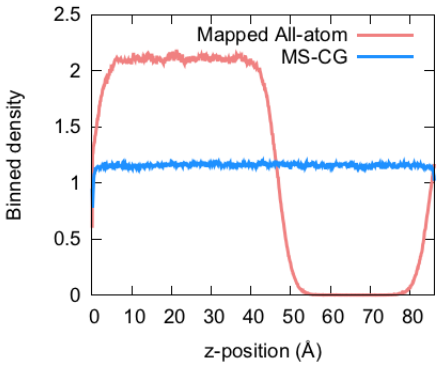
# Slab Density Profiles: All-atom and MS-CG

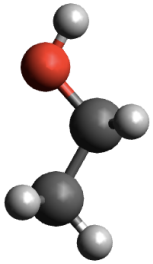
(a) Methanol



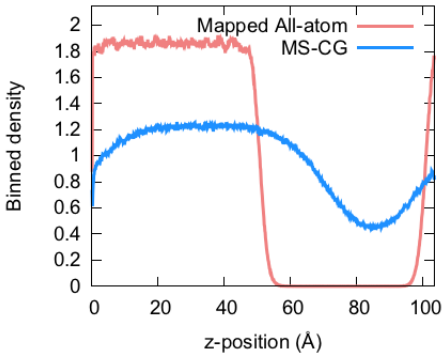



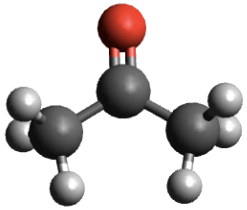
(b) Ethanol



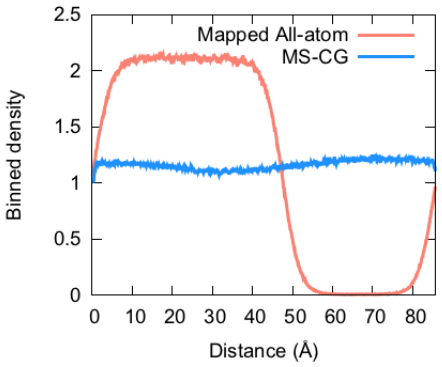



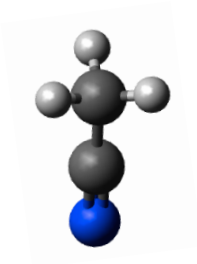
(c) Acetone



(d) Acetonitrile



Longer chain

Linear chain

**MS-CG theory generally fails to describe the interface system**

J. Jin, G. A. Voth. Submitted

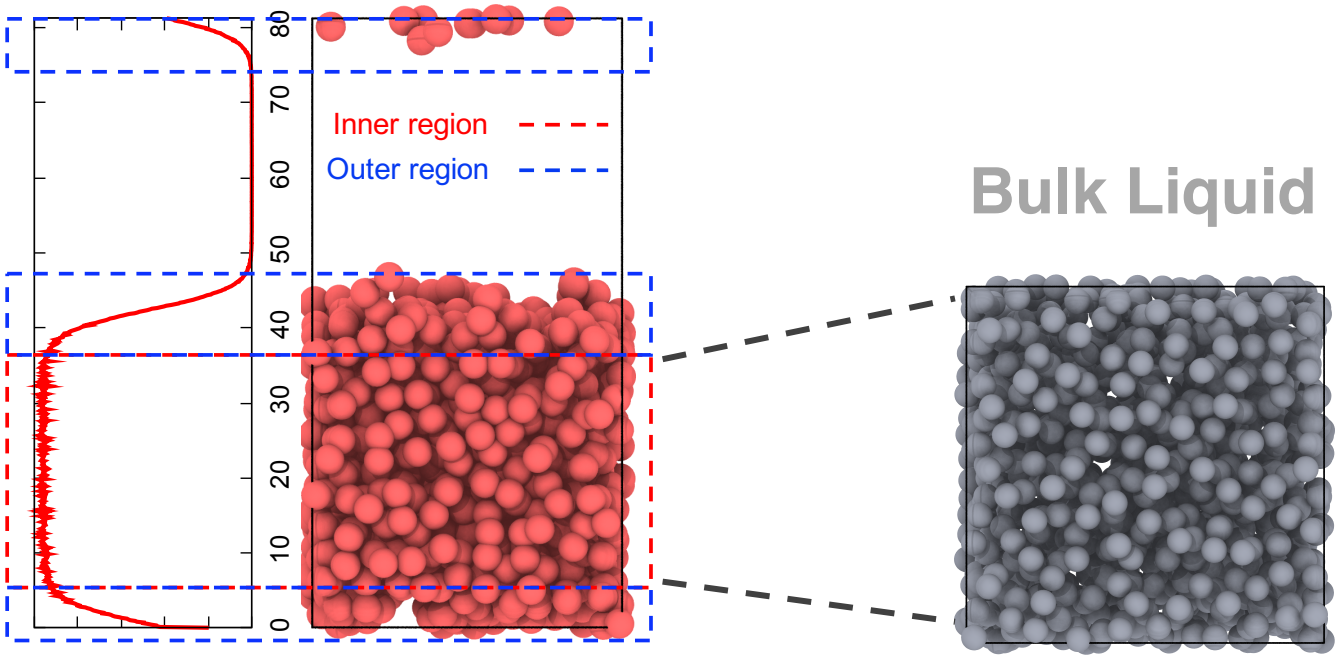
# UCG Model Design: Inner/Outer Regions

Define inner and outer regions based on liquid phase and phase boundary of the system

Vapor Phase

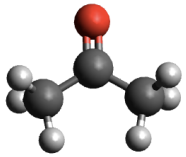
Phase Boundary

Liquid Phase

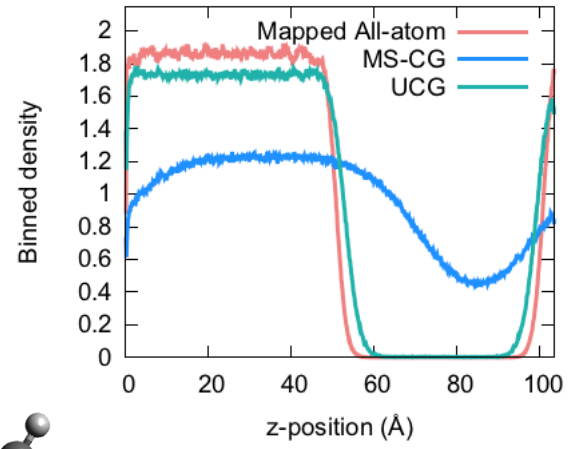



UCG internal states are designed to distinguish denser (inner region) and less dense (outer region) local environment

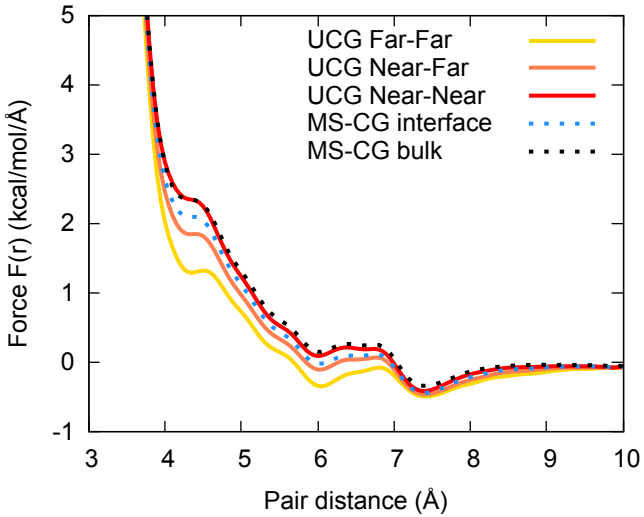
# Liquid/Vapor UCG Model Results

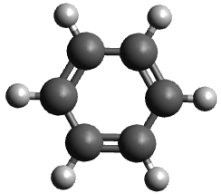


Density Profile

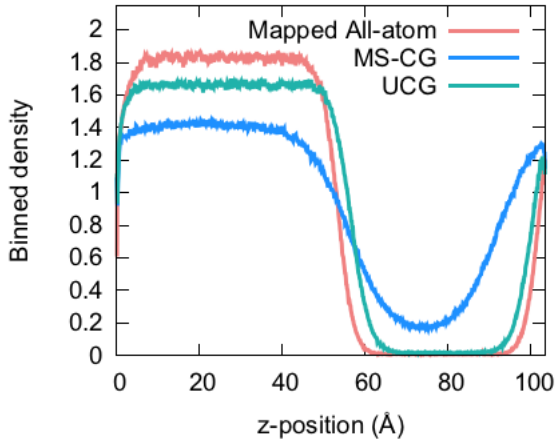



CG Pair Interaction

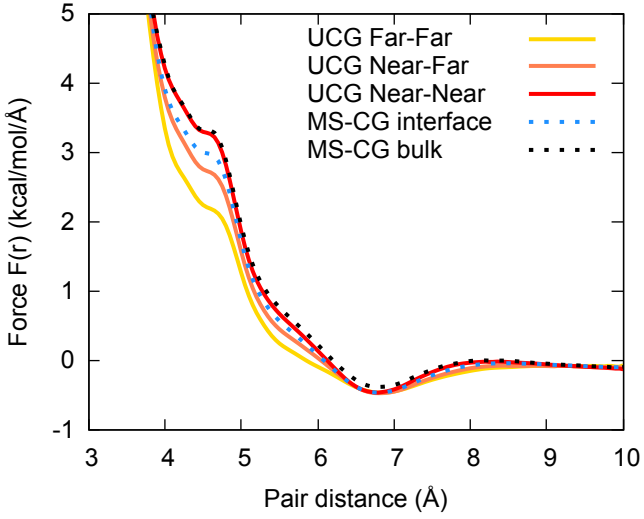


Density Profile




CG Pair Interaction




UCG models provide better density profile

**UCG Near-Near** (liquid phase) interaction is transferable to bulk interaction (**MS-CG bulk**)

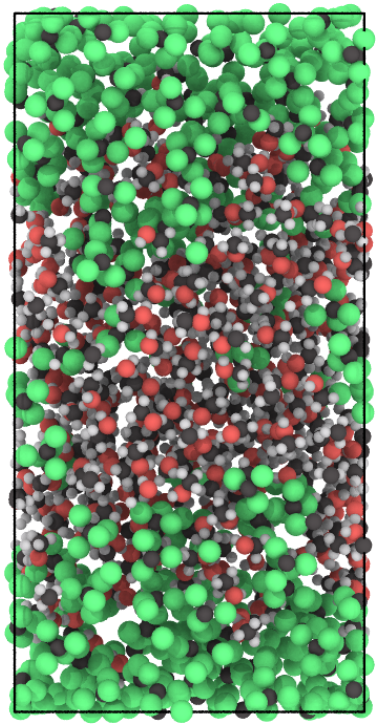
Direct MS-CG to interface (**MS-CG interface**) fails to distinguish interactions between states

J. Jin, G. A. Voth. Submitted

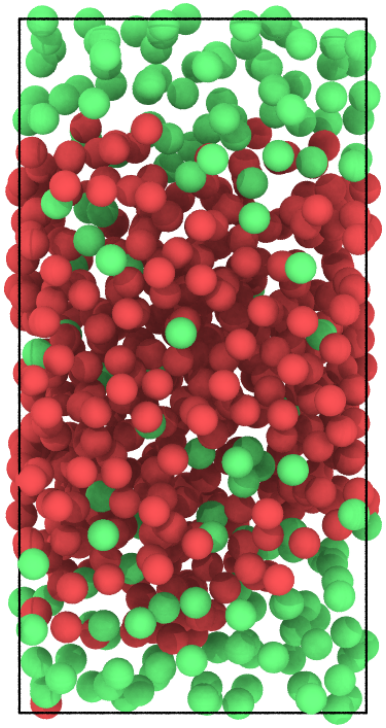
# UCG Model for Liquid/Liquid Interface

- **Liquid/Liquid Interface is challenging**: Heterogeneous system composed of two different liquids
- **Target system**: Methanol/$CCl_4$

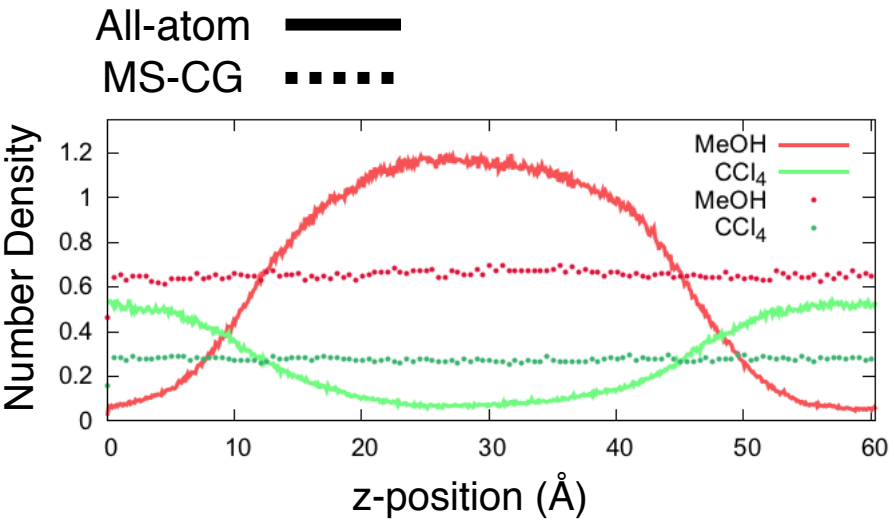
**All-atom**



**One-site CG**



**Slab Density Profile**

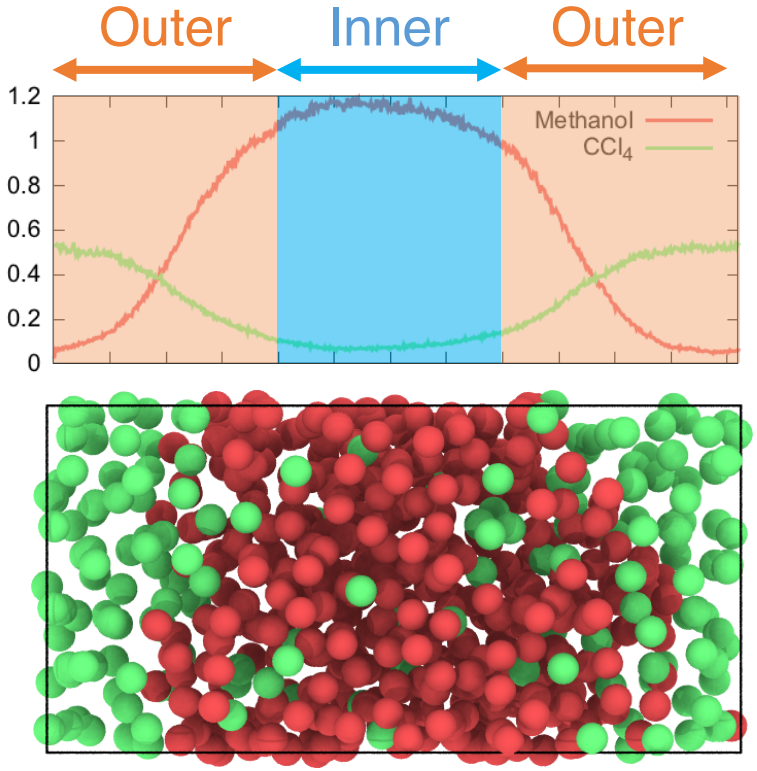


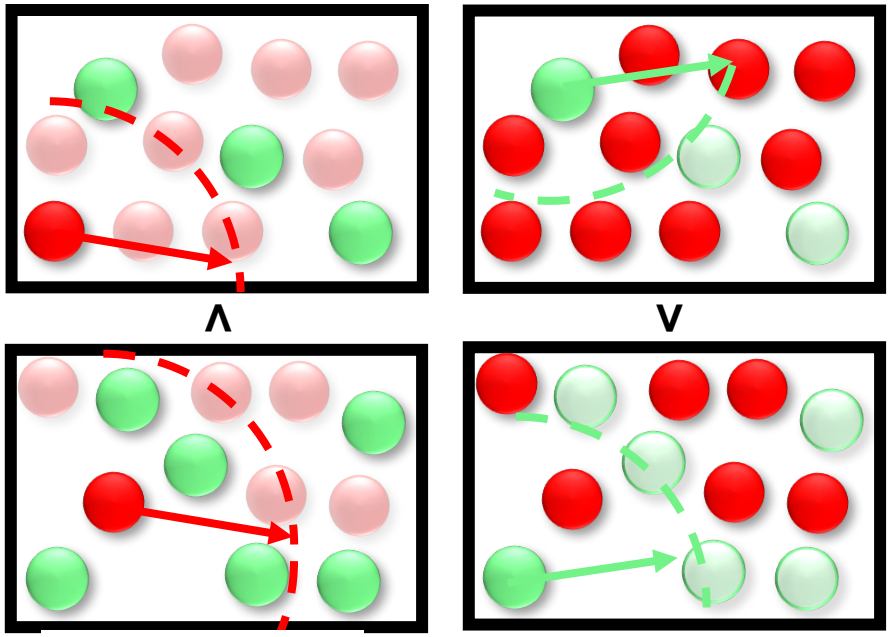

**MS-CG theory fails to capture phase coexistence**

J. Jin, G. A. Voth. In preparation

# UCG Model Design: Liquid/Liquid Interface

Using a **cross-density** as an order parameter to distinguish the different chemical environments



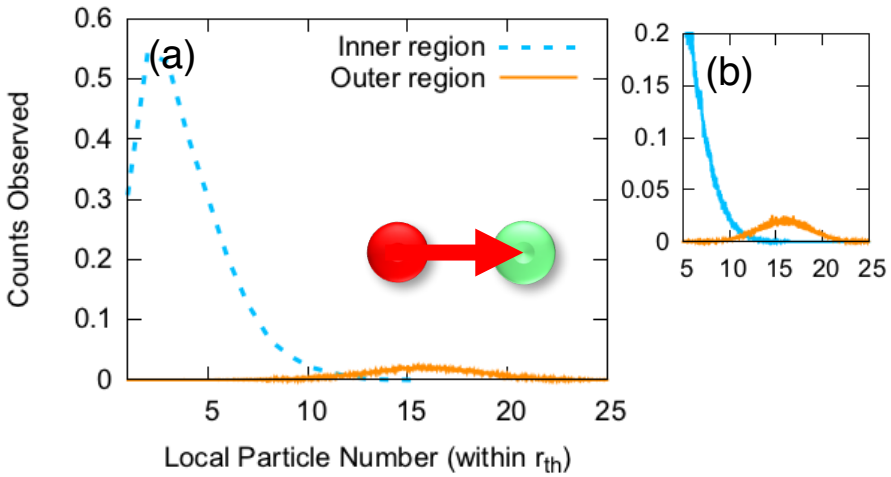

- For MeOH: Local density of neighboring $CCl_4$ near MeOH
- For $CCl_4$: Local density of neighboring MeOH near $CCl_4$

J. Jin, G. A. Voth, Submitted

# UCG Model Design: Local Cross-Density

## 1. Order Parameter for MeOH

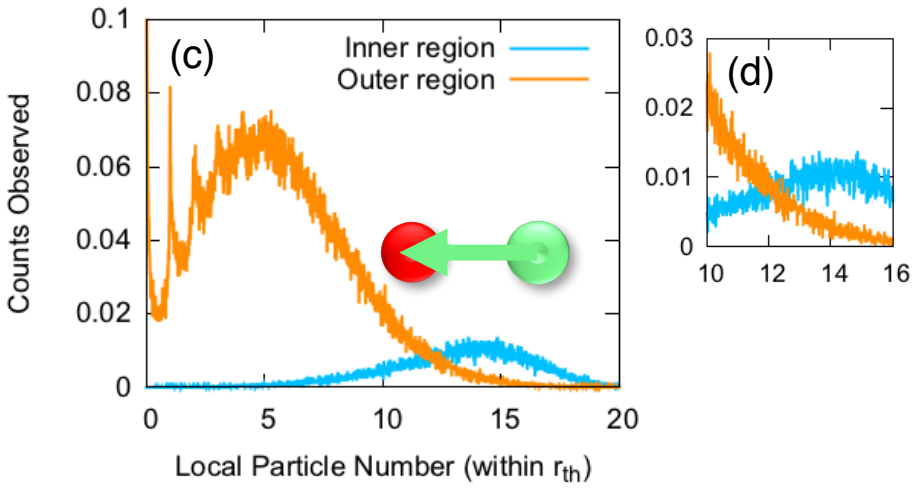
Local density of $CCl_4$ from MeOH




**Inner**: Less denser
**Outer**: Denser

## 2. Order Parameter for $CCl_4$

Local density of MeOH from $CCl_4$



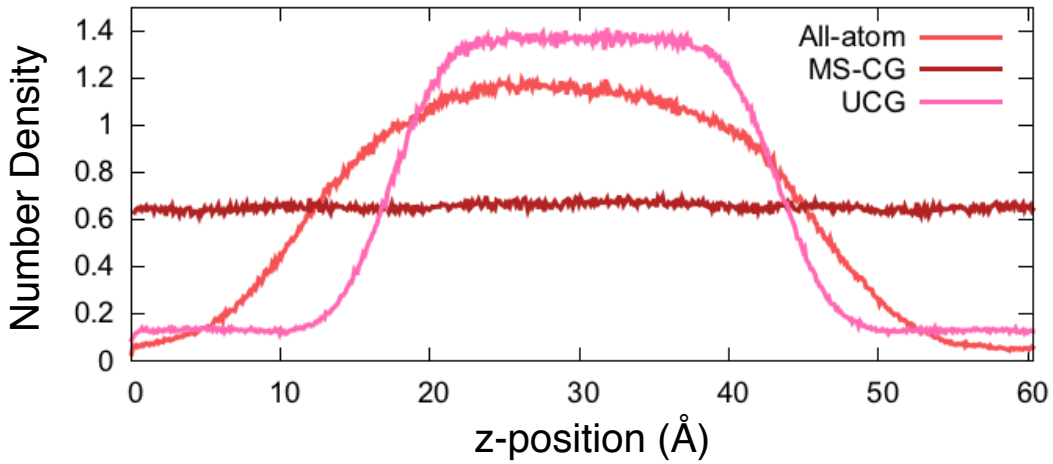

**Inner**: Denser
**Outer**: Less denser

Local cross-density represents a **bimodal character**:
**UCG model** is designed to distinguish two different environments

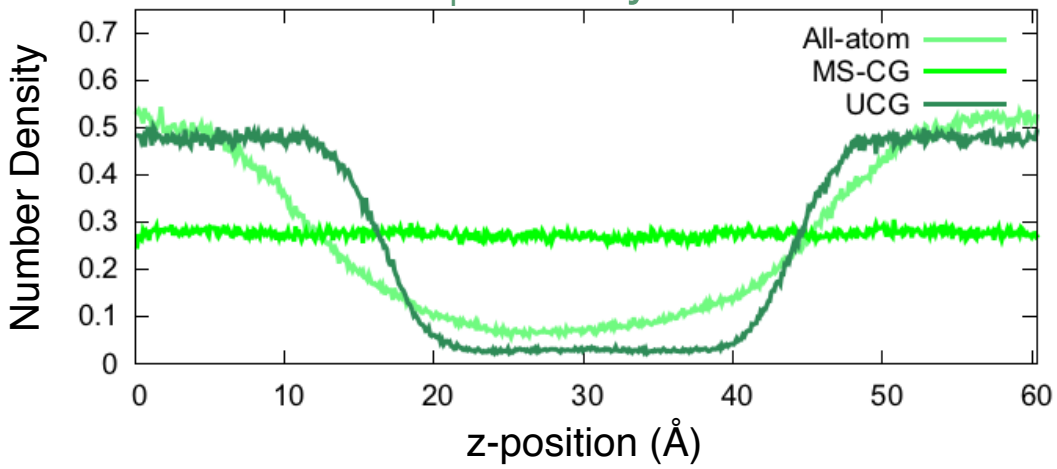
J. Jin, G. A. Voth. Submitted

# Liquid/Liquid UCG Model Results

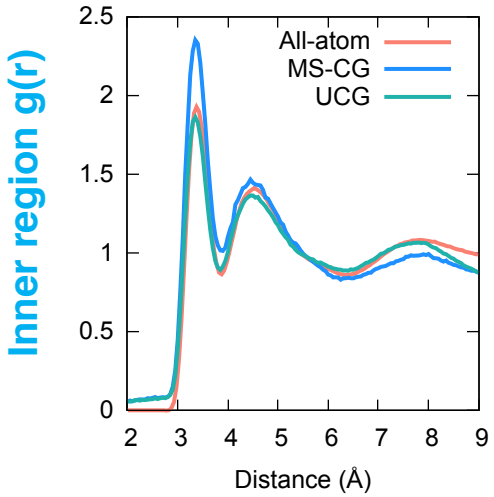
MeOH Density Profile



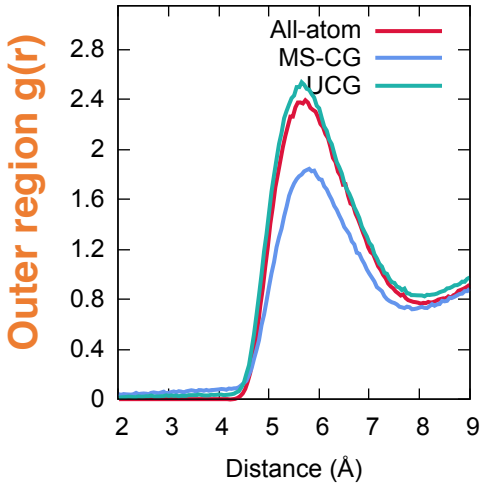

$CCl_4$ Density Profile




MeOH-MeOH RDF

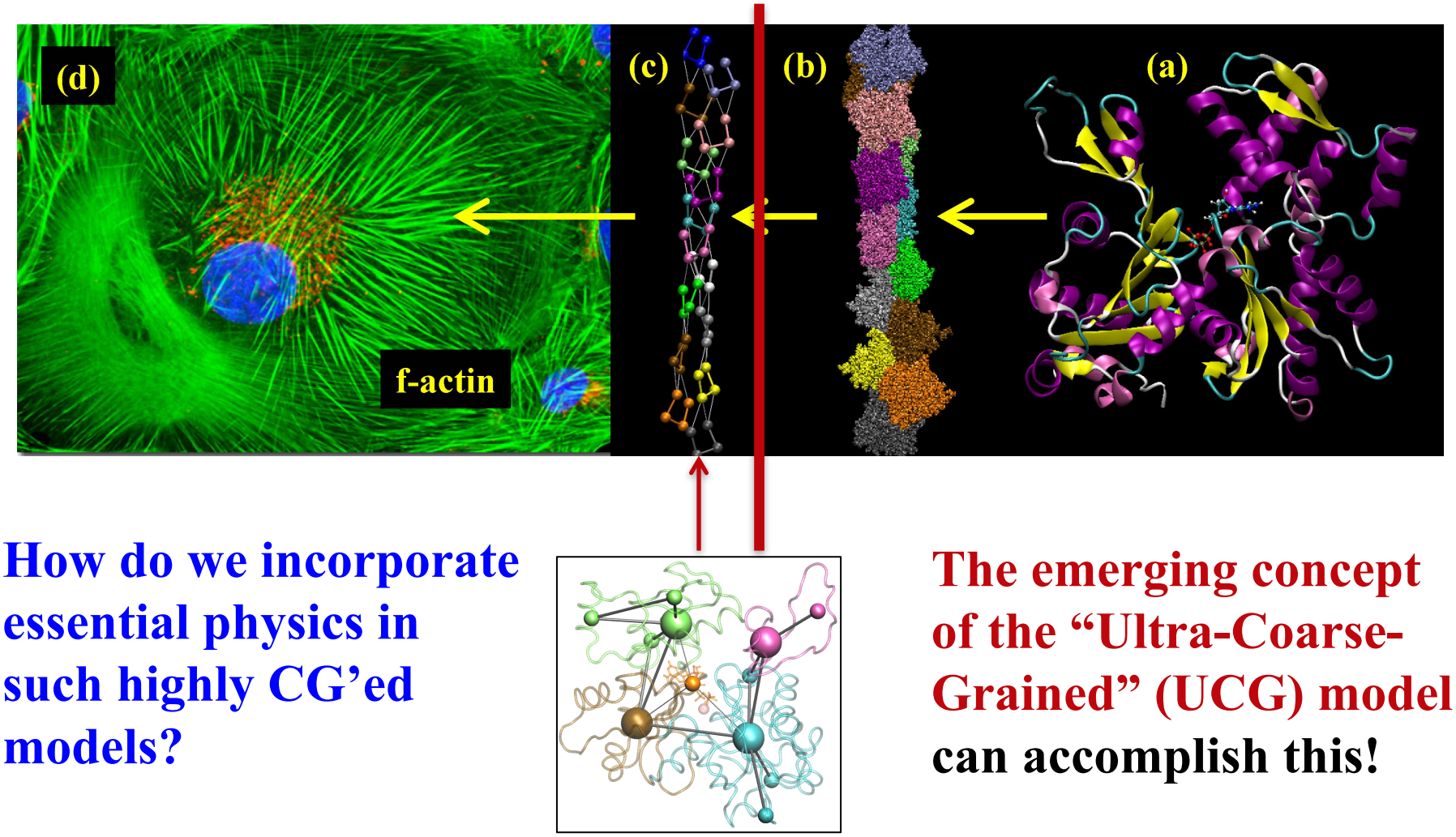



$CCl_4$-$CCl_4$ RDF




UCG model improves the phase coexistence with reproducing the correct structure from the RDFs

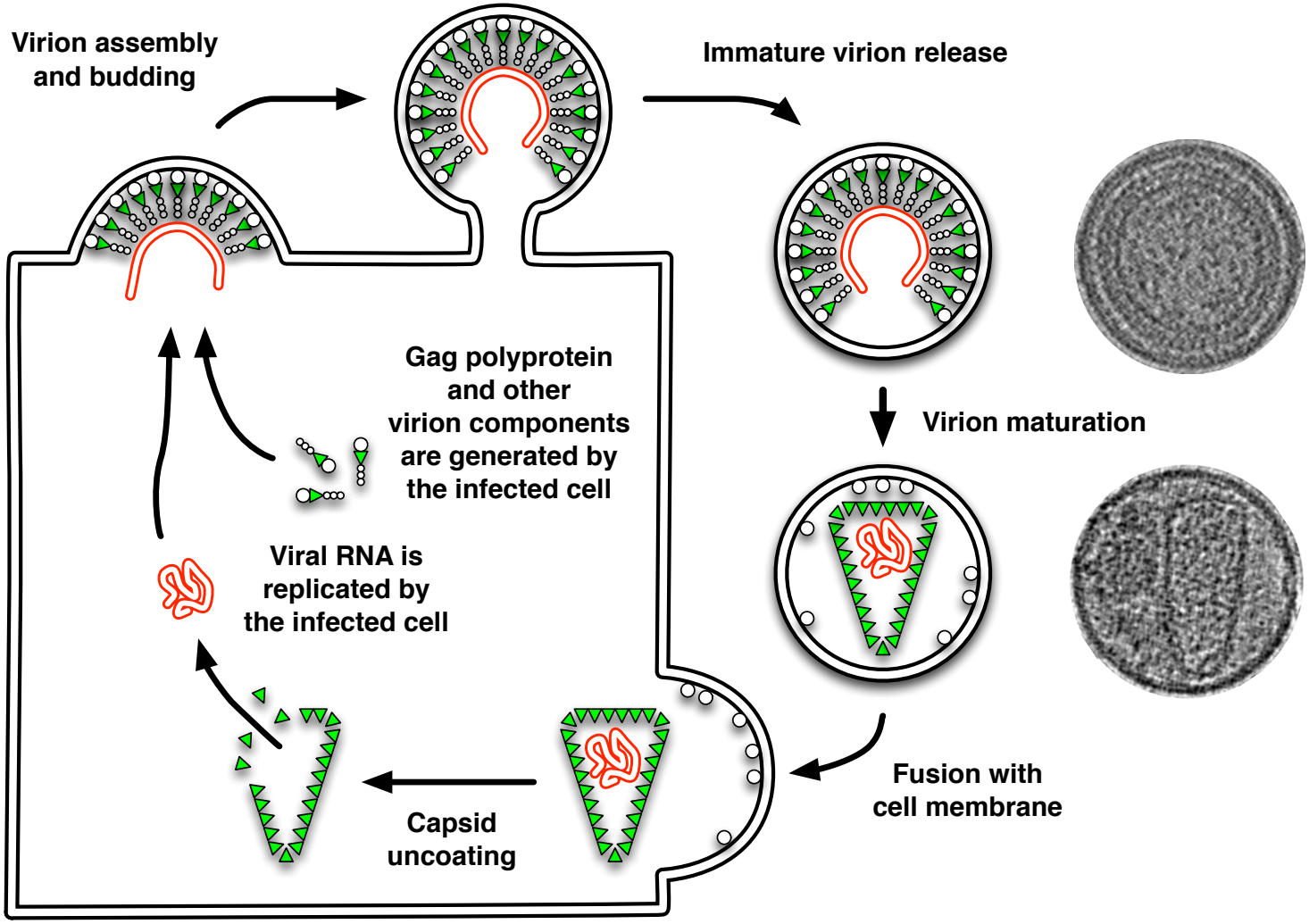
# The Multiscale Challenge




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

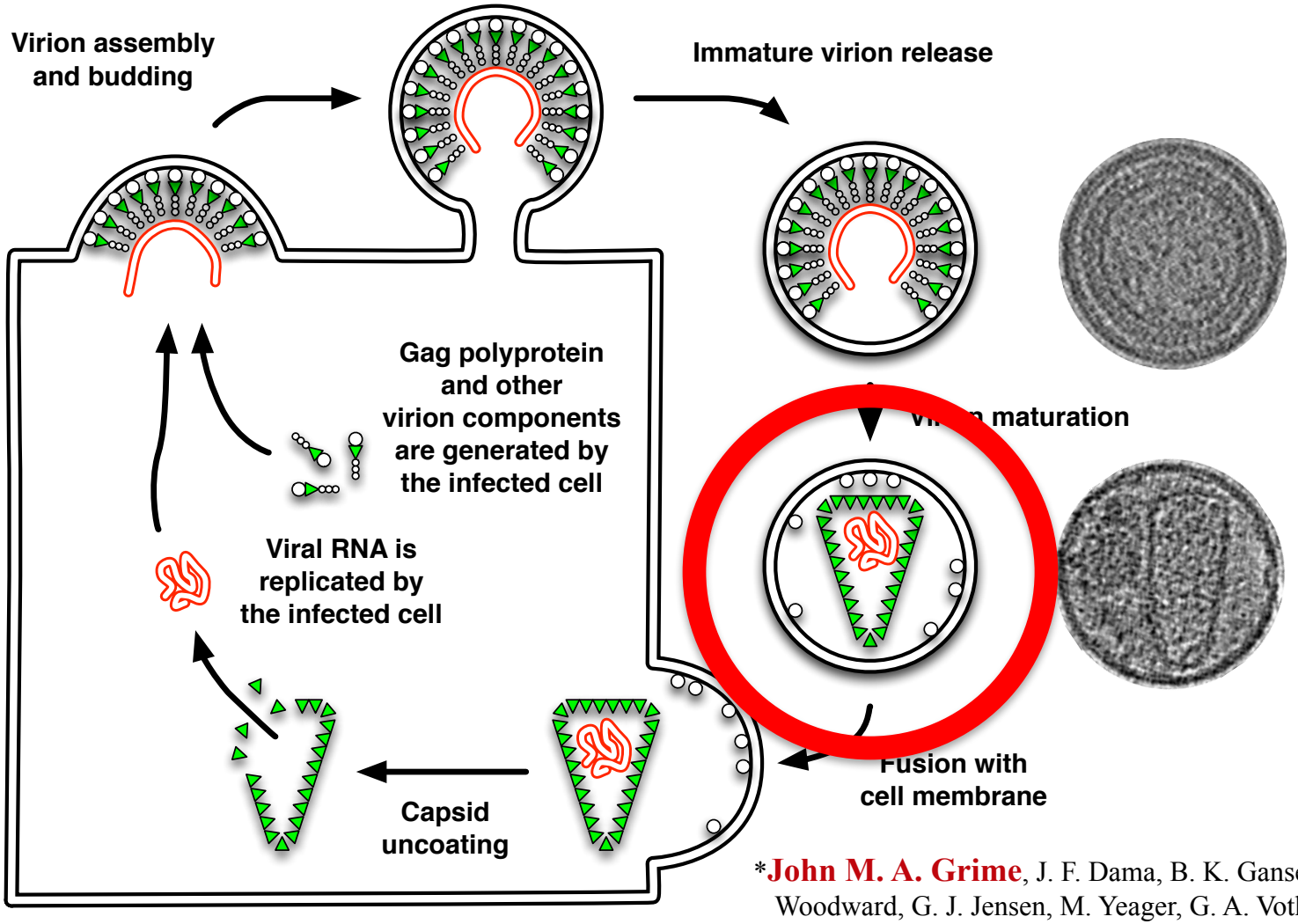
**The emerging concept of the "Ultra-Coarse-Grained" (UCG) model can accomplish this!**

# UCG Application: HIV Capsid Assembly



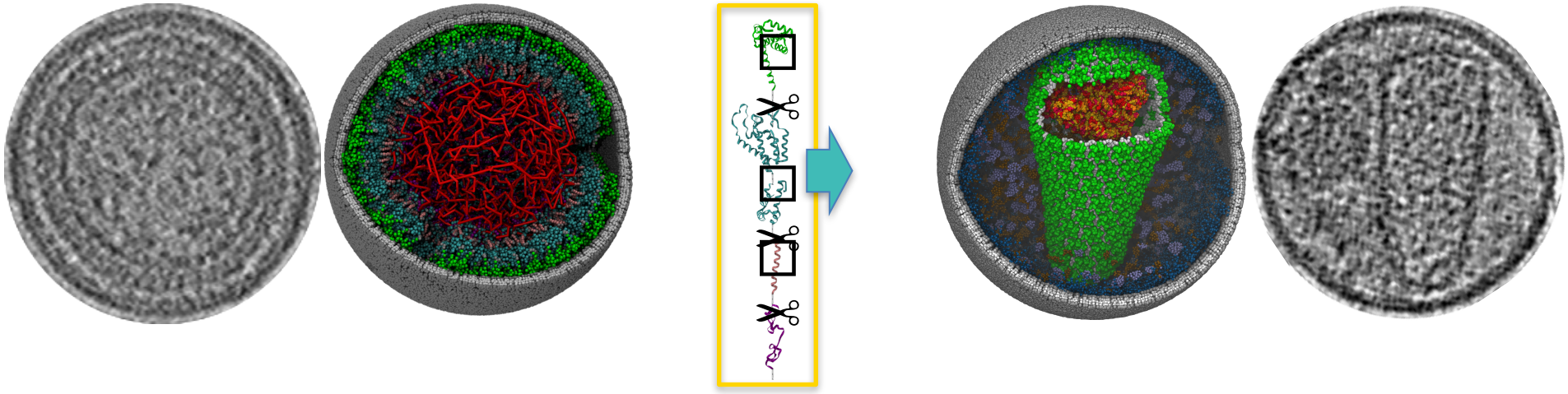

Ganser-Pornillos et al, *Curr. Op. Struct. Biol.* 2008, **18**:203

# Next Application: HIV Capsid Assembly*

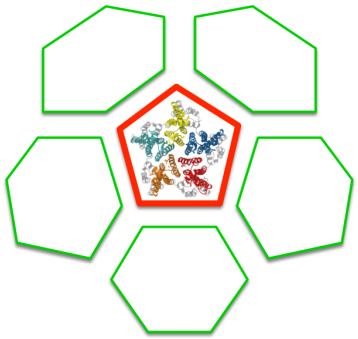


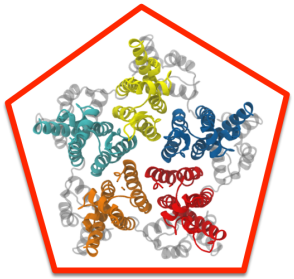

Ganser-Pornillos et al, *Curr. Op. Struct. Biol.* 2008, **18**:203

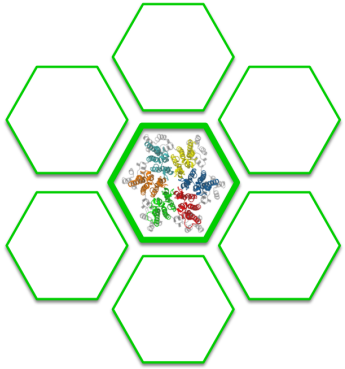
*John M. A. Grime, J. F. Dama, B. K. Ganser-Pornillos, C. L. Woodward, G. J. Jensen, M. Yeager, G. A. Voth, "Coarse-grained Simulation Reveals Key Features of HIV-1 Capsid Self-Assembly", Nature Comm. **7,** 11568(1-11) (2016).

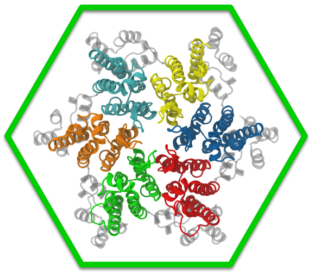


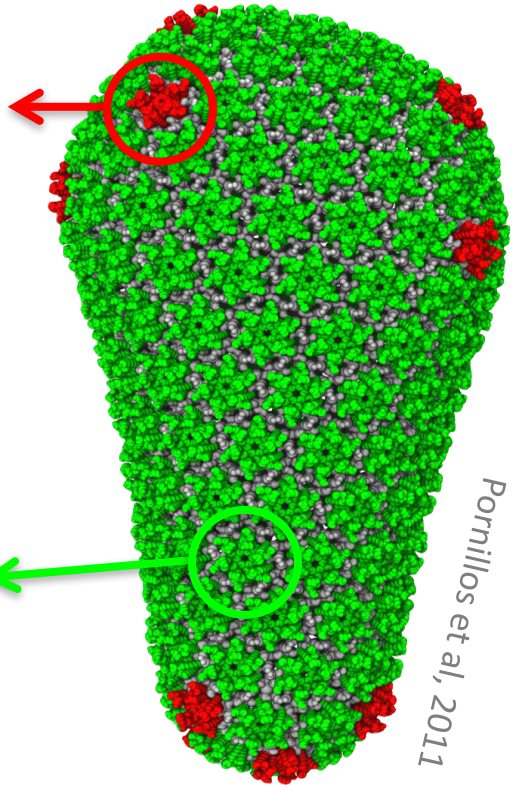
# HIV-1 “maturation”: no conical capsid, no infectivity





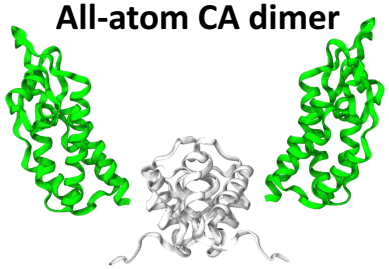




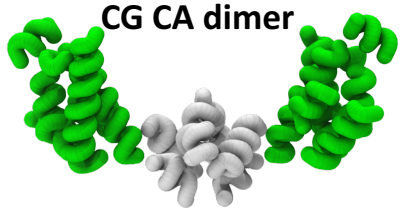


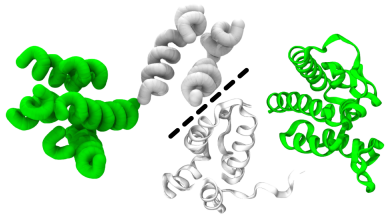
Pornillos et al, 2011

All-atom CA dimer

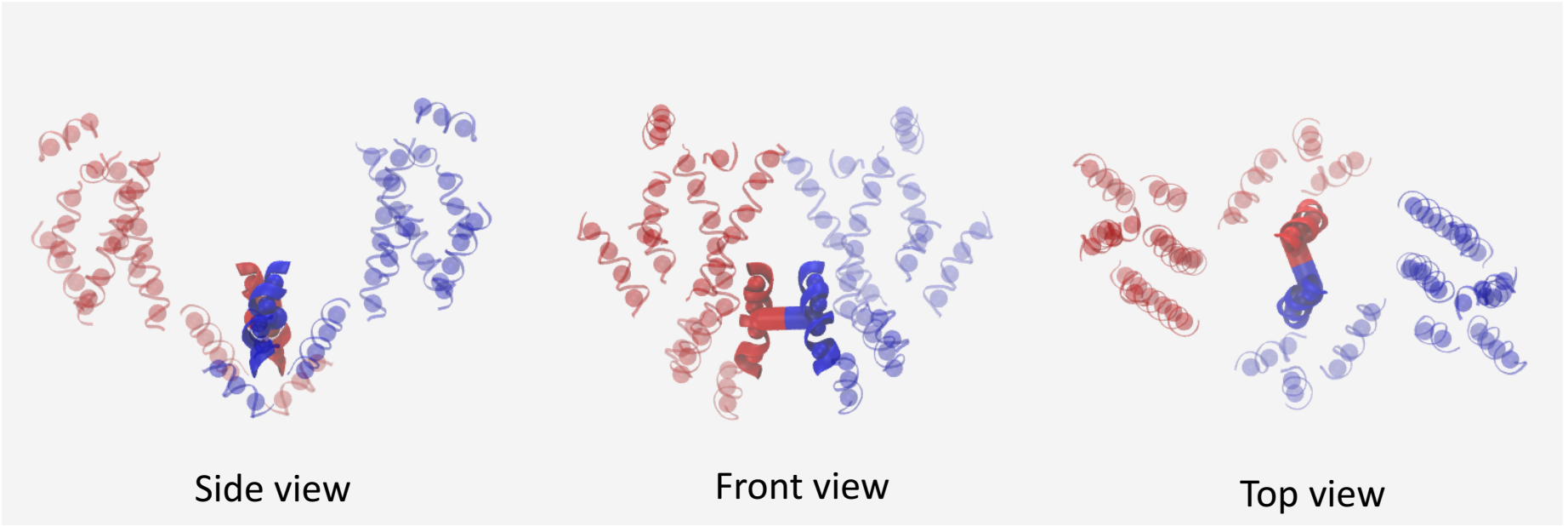


CG CA dimer





# Coarse-grained CA Protein Model: Excluded Volumes by Helices and CTD / CTD Dimer Interface Constraint

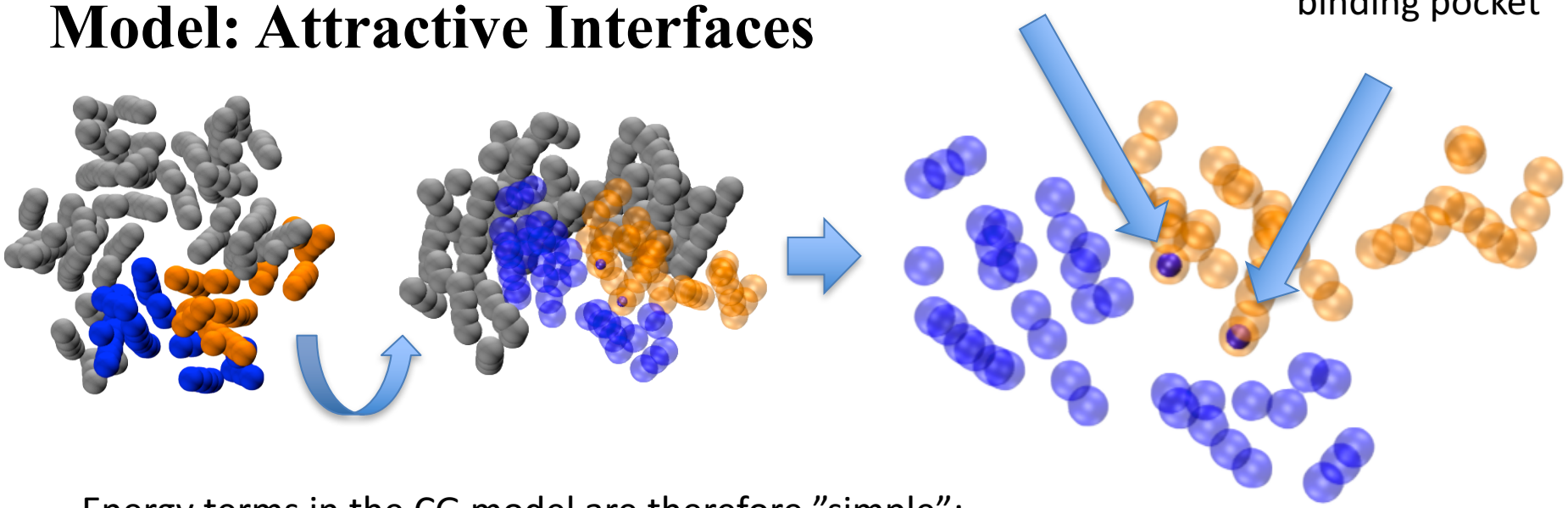



- **Important distance constraint** between carbon betas of residue 185 (helix 9) across the dimer interface; conserved for **2KOD**[1] & **1A43**[2] ( 9.17 Å & 9.18 Å )

[1]: Byeon *et al*, *Cell* 139 (2009)

[2]: Worthylake *et al*, *Acta Cryst. D* 55 (1999)

# Coarse-grained CA Protein Model: Attractive Interfaces




Energy terms in the CG model are therefore ”simple”:

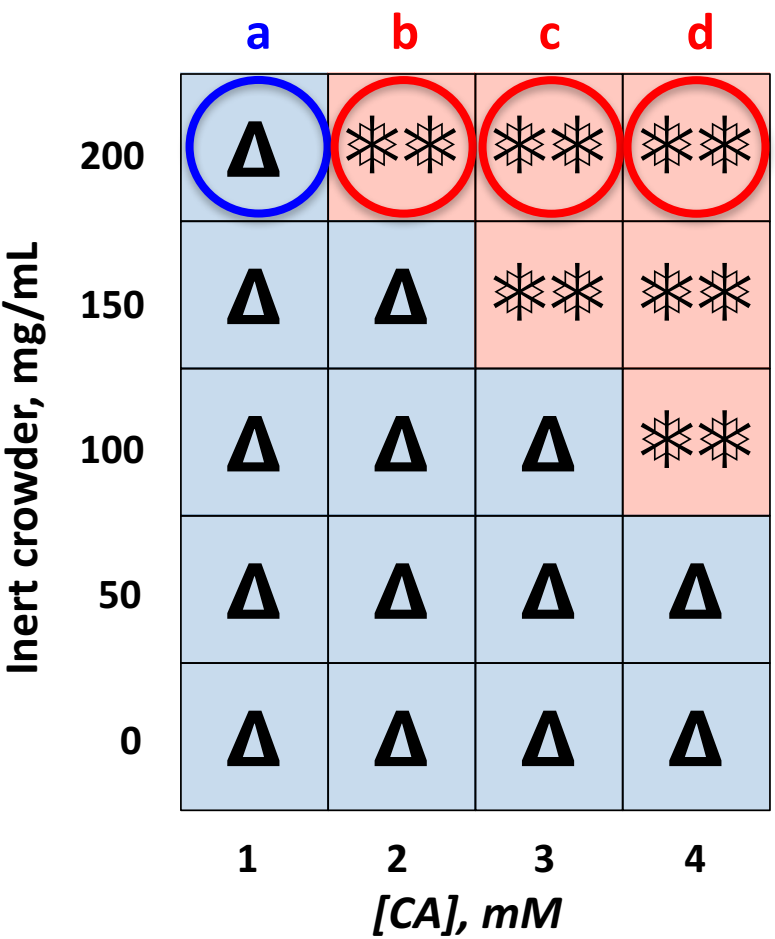
1. **Purely repulsive interactions on almost all CG beads**, relatively insensitive to the precise functional form (Morse, repulsive Lennard-Jones, simple linear repulsion have all been tested)
2. **Simple attractive basin for the binding pockets** (again, apparently insensitive to the functional form: truncated harmonic, Gaussian, simple linear attraction)
3. **Harmonic angle potential** to control CTD pivot motions, parameterized to reproduce range of pivot angles in model capsid structure[1]

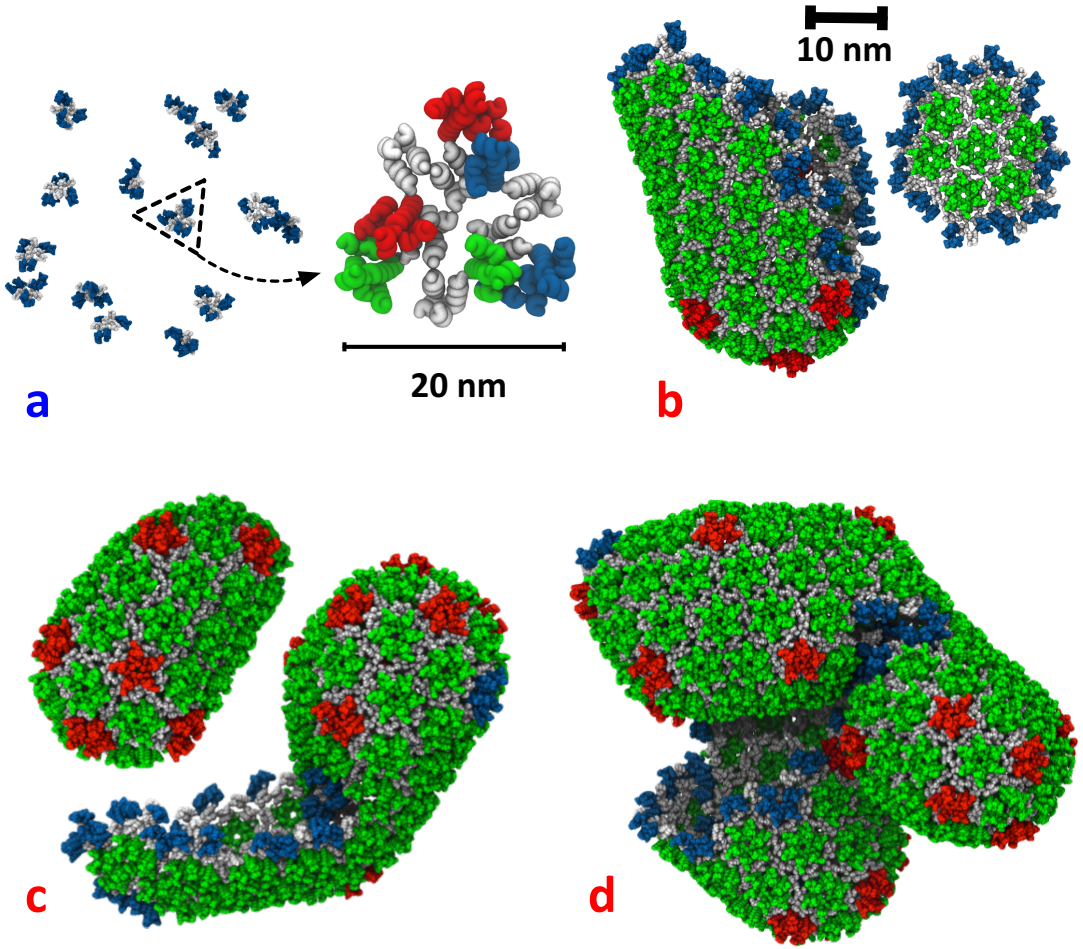
The effects of flexibility in the CTD dimer interface are currently under investigation.

[1] Pornillos et al, Nature 2011; 469:424-428

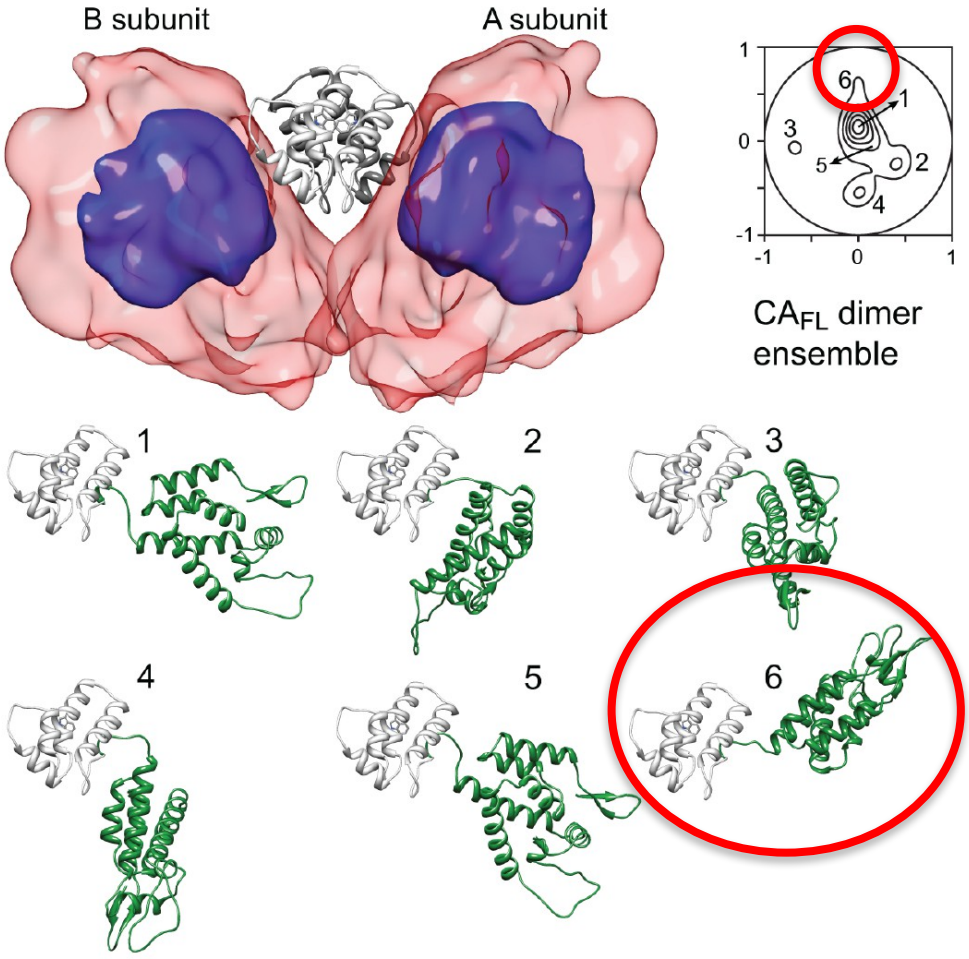
# “Basic” CG model: *in vitro* Simulations

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



# CA Protein Structural Dynamism



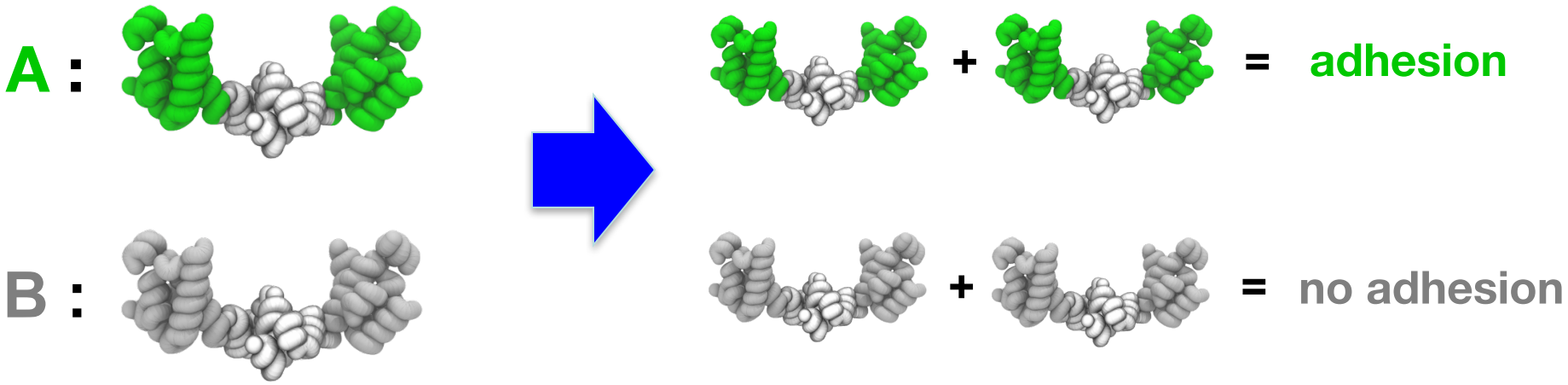

Relatively small amount of CA in "native-style" NTD/CTD conformation in solution (≈ 5%), with domain motions time-correlated (≈ 5 ns) [1].

**UCG-style "switching" model for CA …**

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

# UCG-MD Style Model of HIV CA Protein

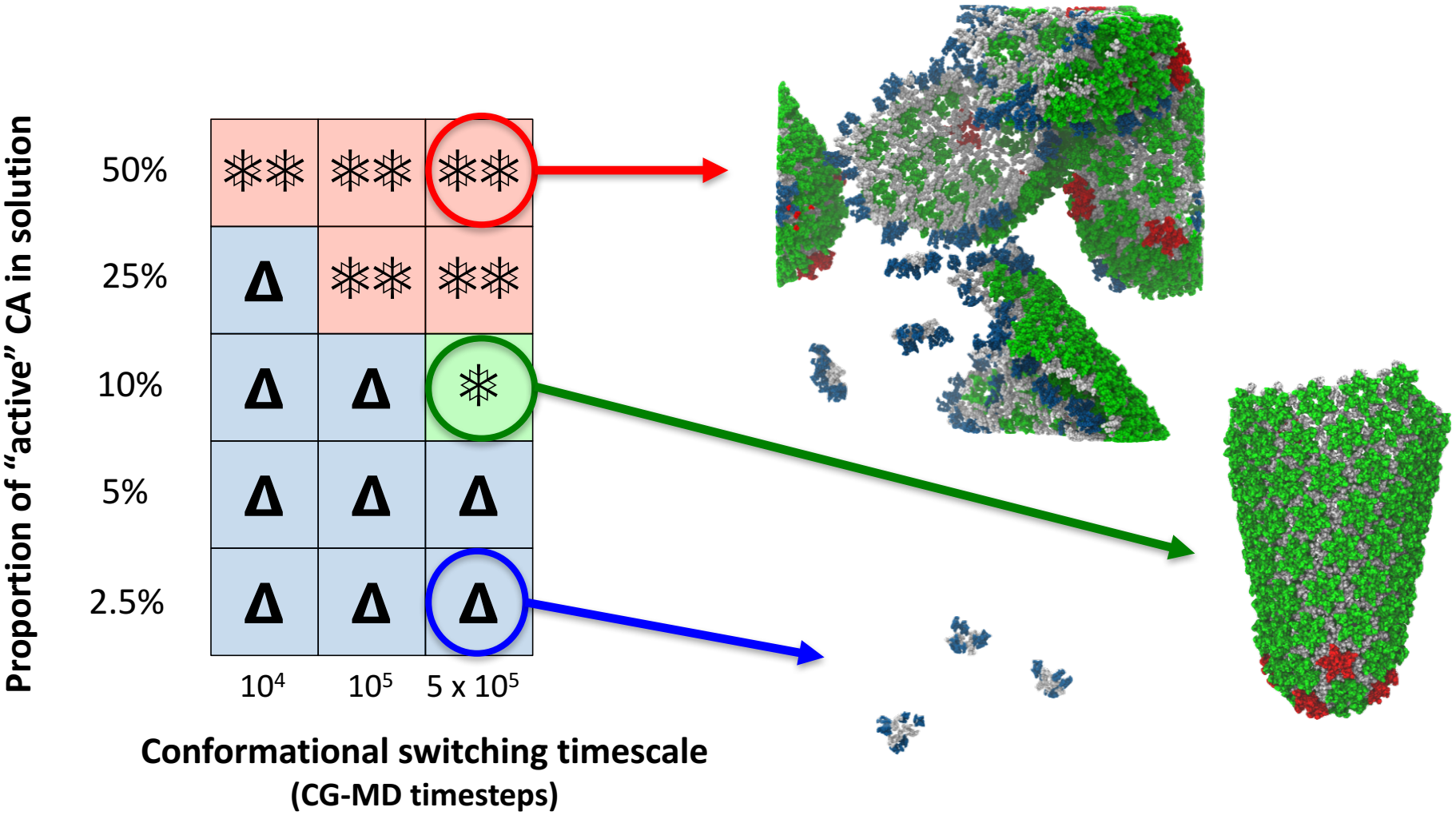
***Simple two-component*** system: **A** and **B** protein dimers with identical internal structure but different interactions:

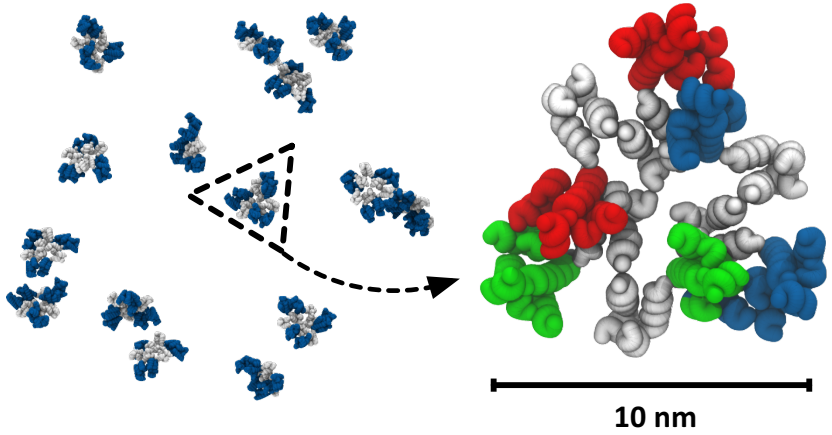


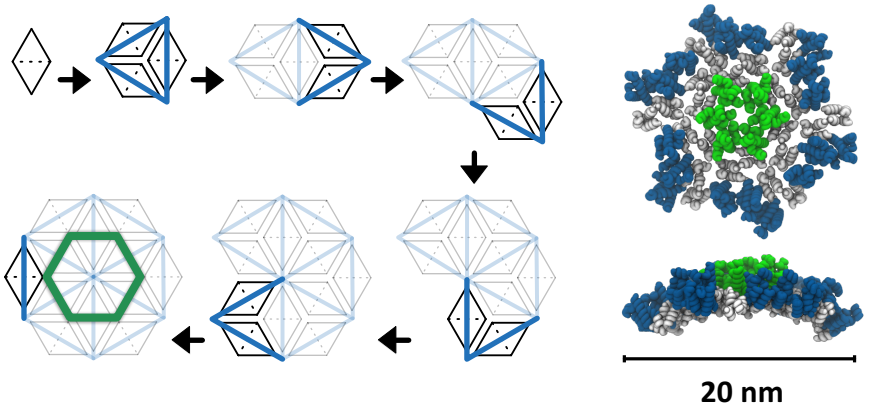

Fixed overall **[CA]** (4mM, 200 mg/mL crowder), ***solution-state*** **proportion of A stochastically (re)assigned with a certain timescale**. Examine effects of conformational heterogeneity on controlled HIV capsid self-assembly …

# CA Structural Dynamism in UCG Model

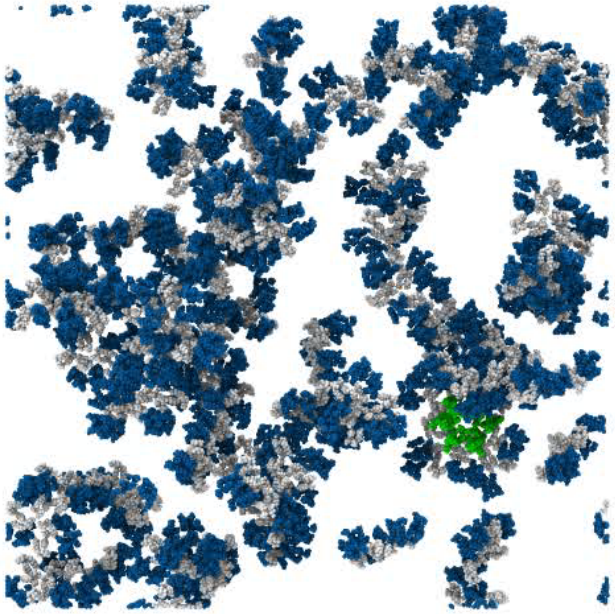
4 mM [CA], 200 mg/mL crowder, vary proportion of "active" (i.e. assembly competent) CA in solution …



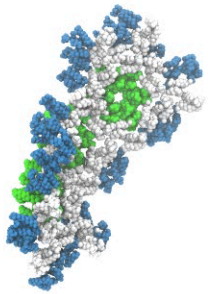


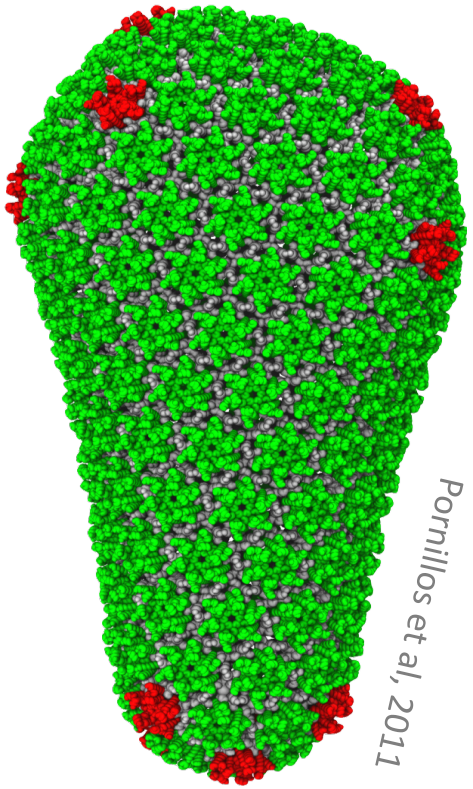

# Controlled self-assembly requires UCG-style CA model



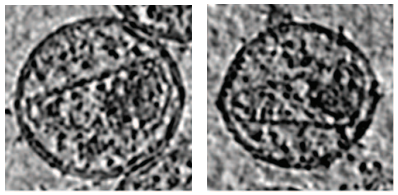
No UCG switching



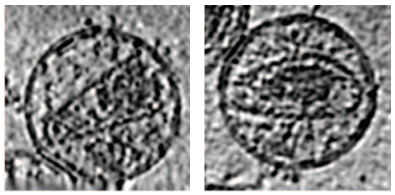
UCG switching

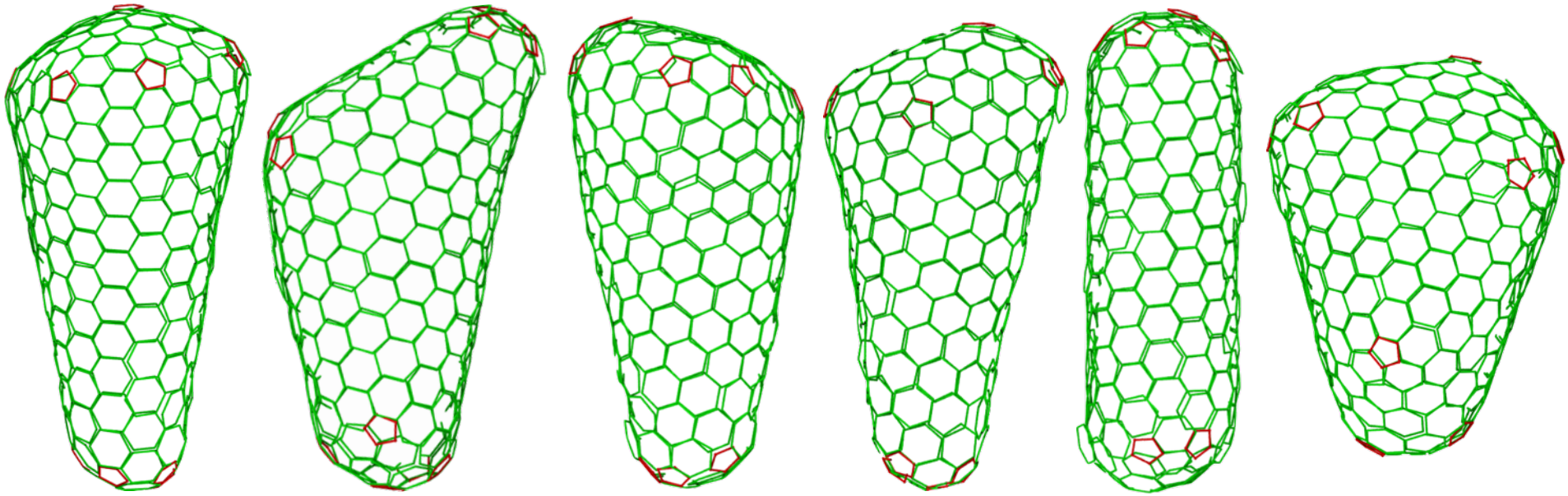


Pornillos et al, 2011



# Recent Cryo-ET[1] used to infer hexagon / pentagon subunits location in *actual in vivo* viral cores:





[1] Mattei, Glass, Hagen, Krausslich, and Briggs. *Science* 354:6318 (2016)

# New Anti-HIV Drug: Gilead *GS-CA1*

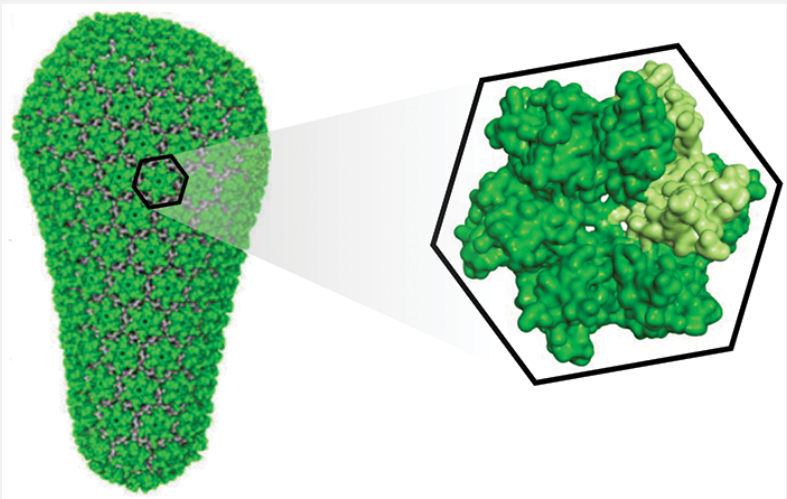


**Volume 95 Issue 31 | pp. 23-25**
**Issue Date: July 31, 2017**

## 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

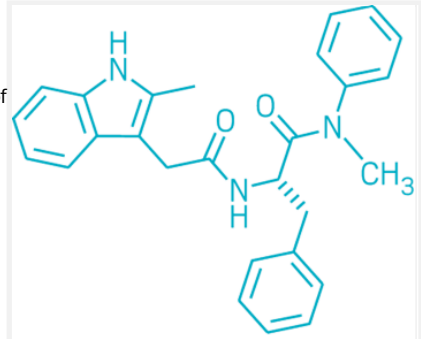
For most of his career at Gilead Sciences, medicinal chemist Winston Tse has lived and breathed one thing. While his peers at other companies hopped from project to project, Tse has spent the past decade obsessing over a single target: the HIV capsid.

HIV's capsid is a complex, protein-rich shell that protects the genetic payload the virus is



Conquering HIV's capsid | July 31, 2017 Issue - Vol. 95 Issue 31 |... http://cen.acs.org/articles/95/i31/Conquering-HIVs-capsid.html?u...

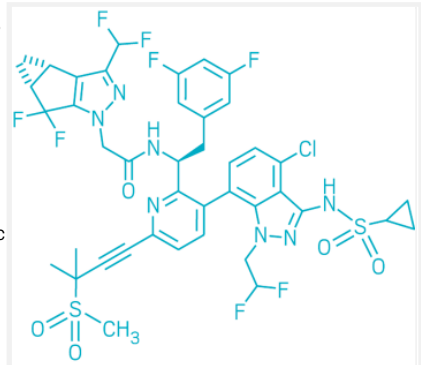
made of 1,500 capsid proteins that organize themselves into hexamers and pentamers to form an eggplant-shaped shell. HIV researchers had no close-ups of the full capsid; a crystal structure had captured only the monomeric protein.



Pfizer's PF-3450074

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.

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.



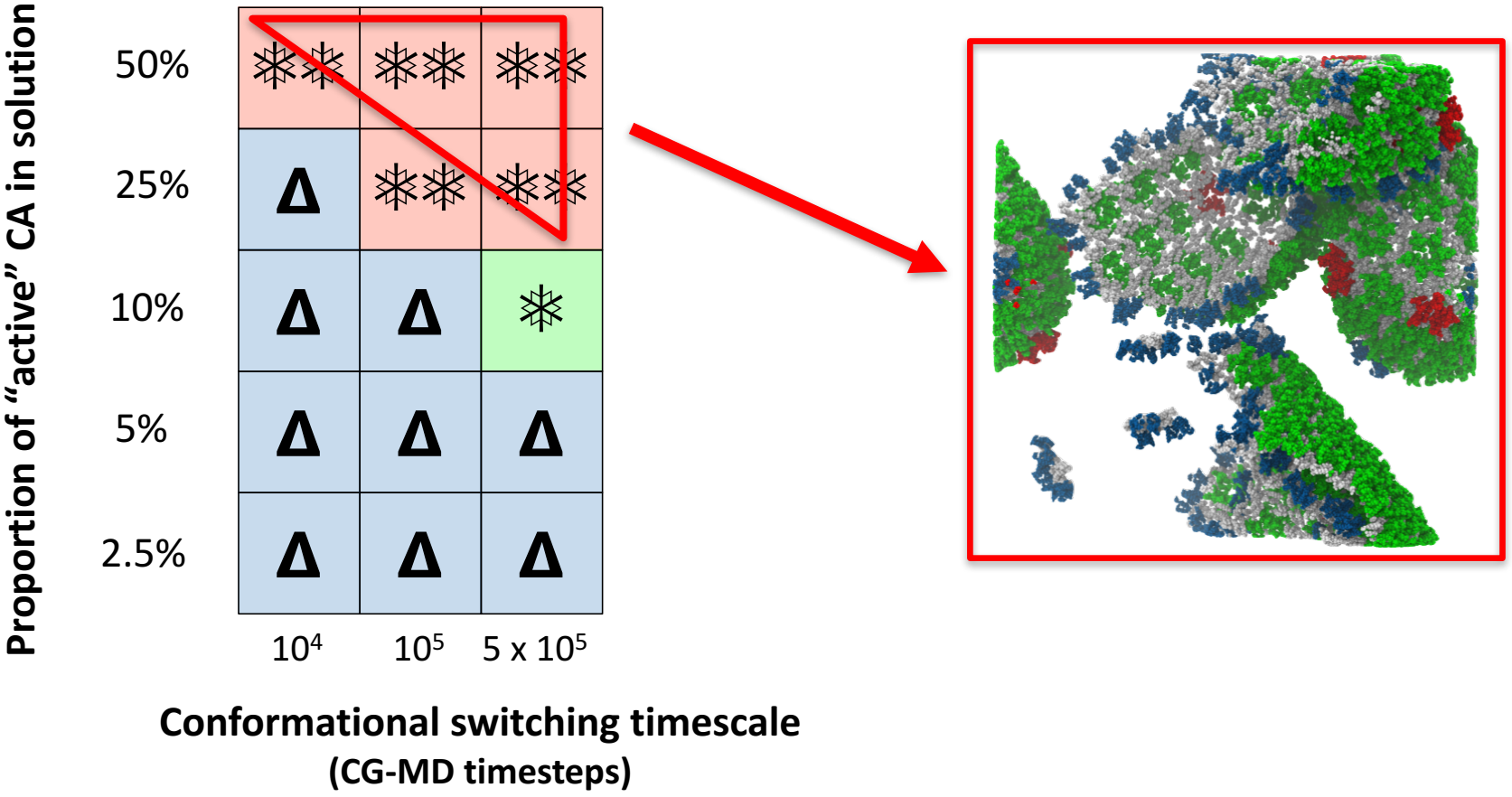
Gilead's GS-CA1

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.

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

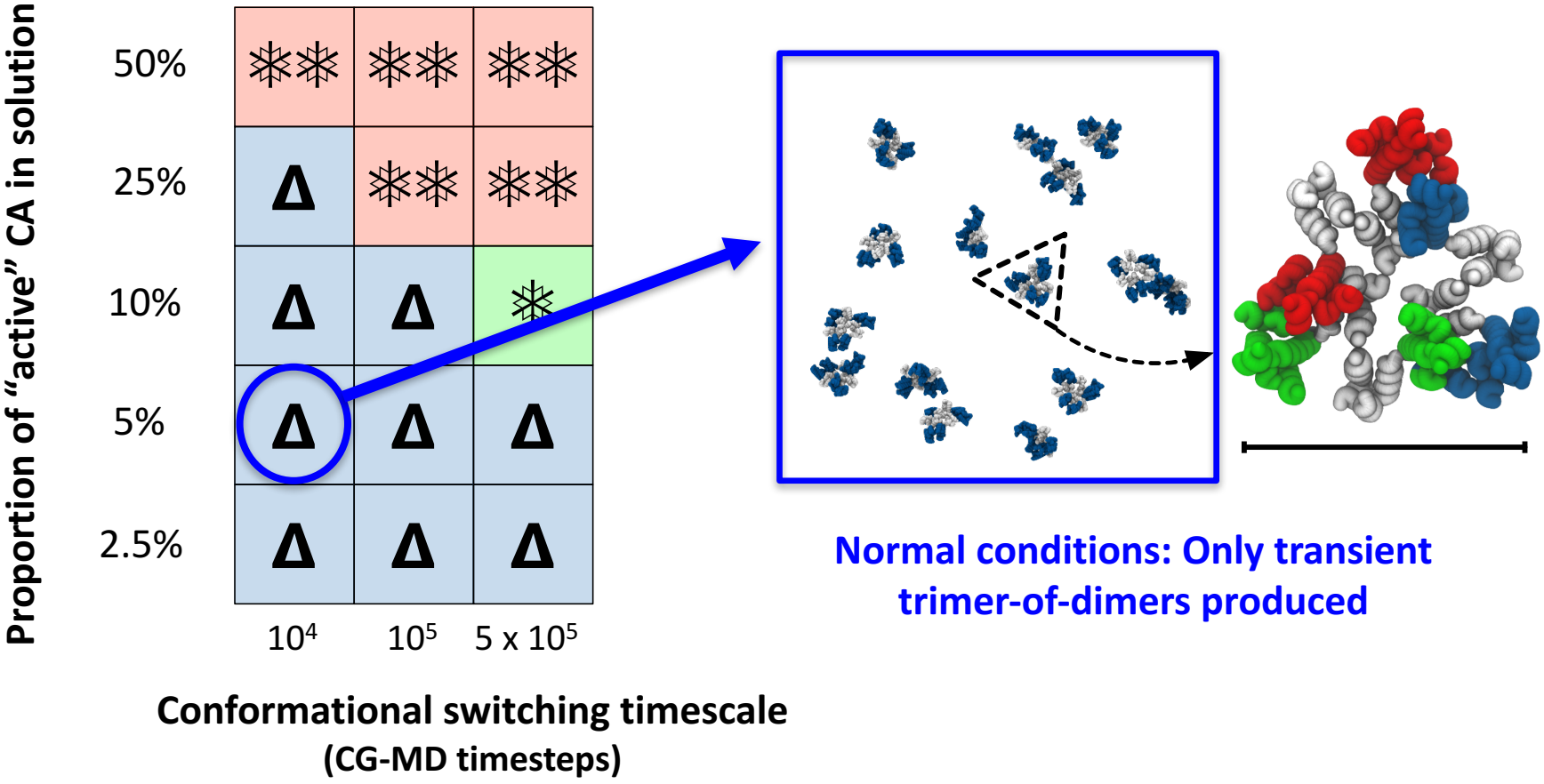
# Gilead: *GS-CA1*

Strong inter-molecular binder: “overloads” controlled CA self-assembly. Effects akin to those in our “uncontrolled” CG simulations …

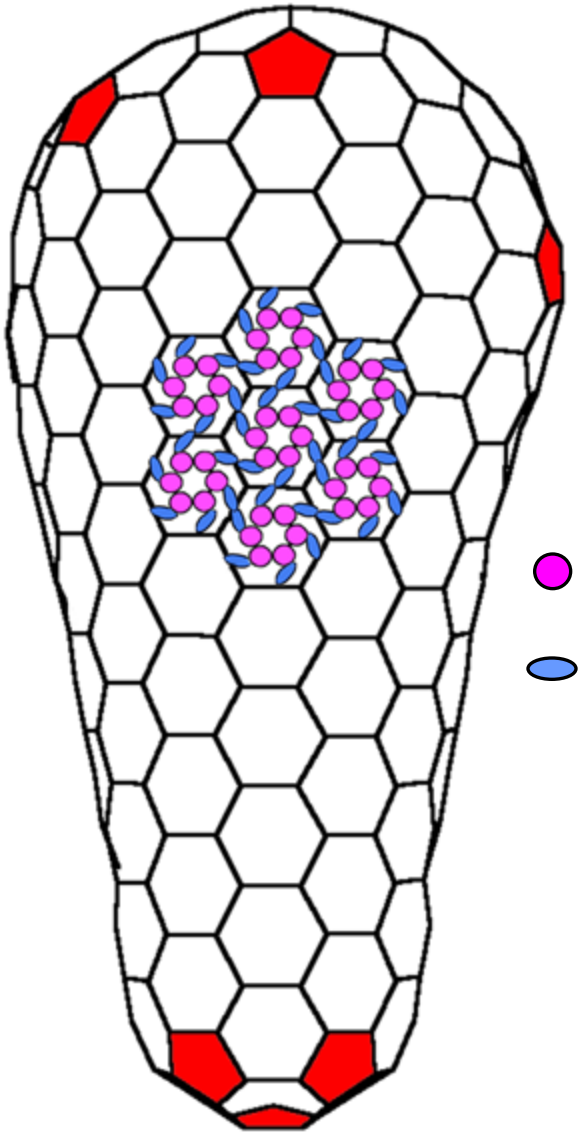


# Gilead: *GS-CA1*

Under conditions that **do not otherwise produce self-assembly**, mimic *GS-CA1* effects via stabilization of initial small number of trimer-of-dimers …



# The Mature HIV-1 Virion: The "Fullerene Cone" Model of the Viral Capsid



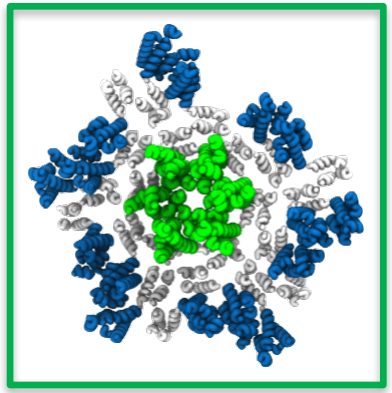
N-Terminal Domain

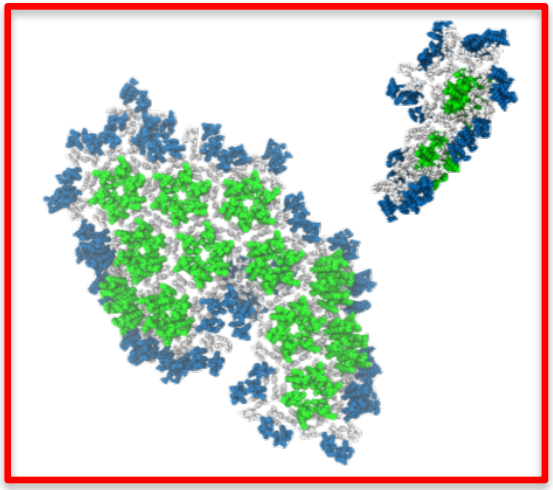
C-Terminal Domain

# Gilead: *GS-CA1*

Under conditions that **do not otherwise produce self-assembly**, mimic *GS-CA1* effects via stabilization of initial small number of trimer-of-dimers …

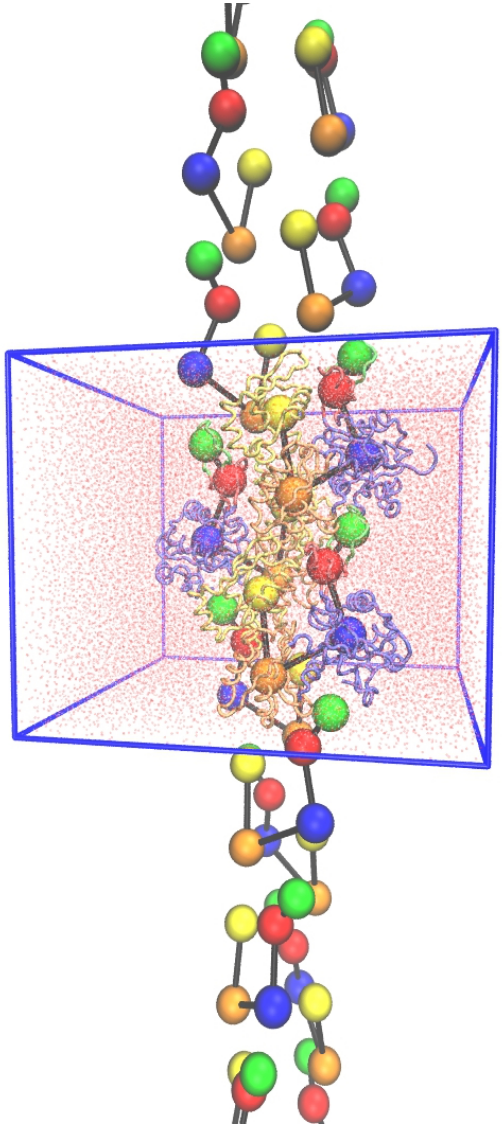
| Initial "stabilized" CA | Result |
|---|---|
| ≈ 0.5% | No effect |
| ≈ 1.0% | No effect |
| ≈ 1.5% | No effect |
| ≈ 2.5% | No effect |
| **≈ 5.0%** | **Single nucleation** |
| **≈ 10.0%** | **Multiple nucleation** |





*Self-assembly process appears sensitive to even small localized "boosts"*

# New Breakthrough for Going Back Downward in Scale:
## Coarse-grained Directed Simulation (CGDS)*



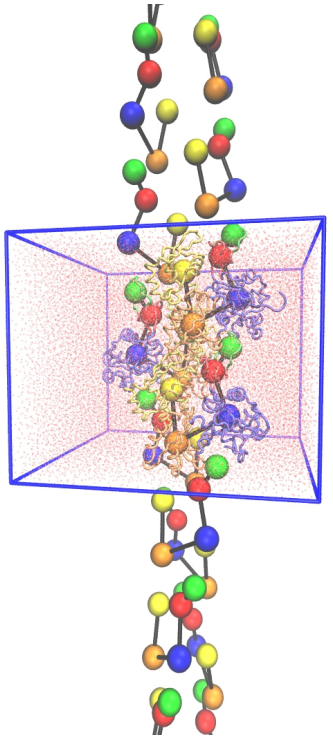
*G. M. Hocky, T. Dannenhoffer-Lafage, and G. A. Voth, “Coarse-grained Directed Simulation”, J. Chem. Theory Comp. **13**, 4593-4603 (2017).

**Goal:** Simulate and restrain subsystem in a way that includes more information about environment

**Ideally:** Simulate subsystem on exact PMF generated by full system

Full system with 3N+3M atoms has coordinates $\vec{r} = (\vec{q}_1, \vec{q}_2)$, subsystem has coordinates has coordinates $\vec{q}_1$. Integrate out $\vec{q}_2$ leaving PMF actin on subsystem:

$$F(\vec{X}) = -k_B T \ln\left(\frac{\int d\vec{q}_1 d\vec{q}_2 \delta(\vec{q}_1 - \vec{X}) e^{-\beta U(\vec{r})}}{\int d\vec{r}\ e^{-\beta U(\vec{r})}}\right)$$

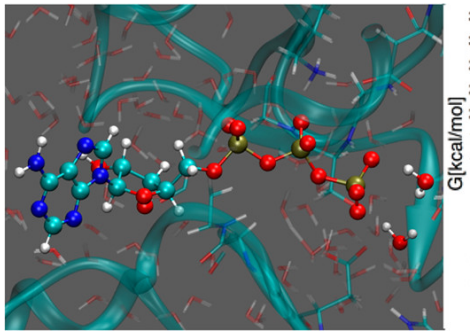


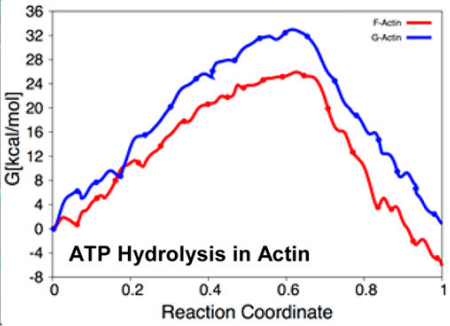
Then the average value of any observable $f$ of the subsystem coordinates ($f(\vec{r}) \equiv f(\vec{q}_1)$) can be recovered just simulating the subsystem:

$$\langle f \rangle = \frac{\int d\vec{r}\, f(\vec{r}) e^{-\beta U(\vec{r})}}{\int d\vec{r}\, e^{-\beta U(\vec{r})}} = \frac{\int d\vec{X}\ f(\vec{X}) e^{-\beta F(\vec{X})}}{\int d\vec{X} e^{-\beta F(\vec{X})}}$$

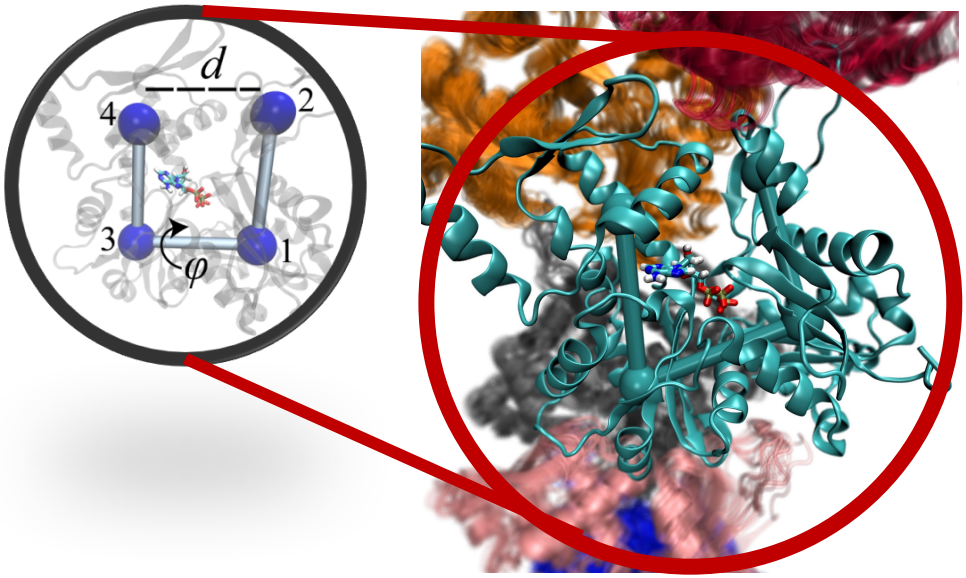
**However: This cannot be done in practice**

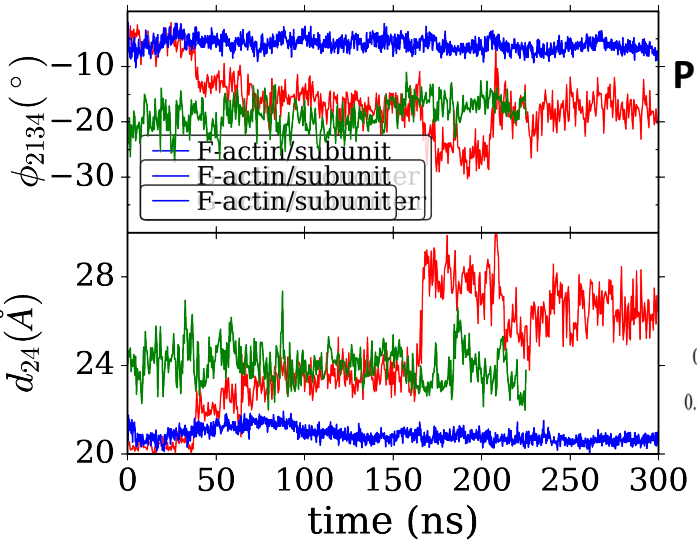
# Actin-Catalyzed ATP Hydrolysis





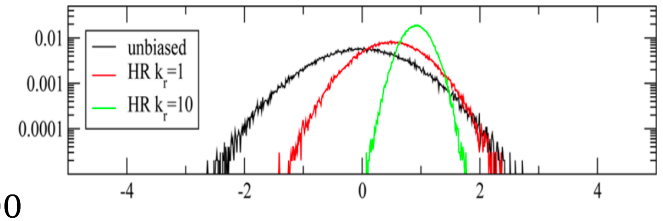

M. McCullagh, M. G. Saunders, G. A. Voth. JACS (2014)
R. Sun, O. O. Sode, J. F. Dama, G. A. Voth. JCTC (2017)

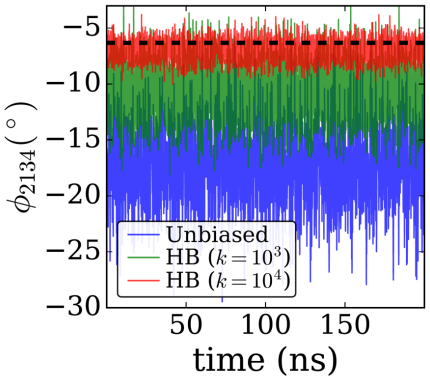



**Problem:** have to restrain actin monomer to get these results

*Harmonic bias shrinks fluctuations and doesn't always reach correct target value!*



# Alternative:

If state of system is well represented by a few coarse-grained observables, can bias these to have same means and fluctuations as in the larger environment

Can introduce this extra information using minimal bias methods that minimize relative entropy between distribution of observables in smaller and larger system. Doing this requires adding linear bias on each observable:

$$I = \int dX P(X) \log(P(X)/P_0(X))$$

…subject to constraints

$$\int dX P(X) = 1 \text{ and } \hat{f}_i = \int dX f_i(X) P(X)$$

$$\Rightarrow P(X) = \frac{e^{-\beta(H(X)+H'(X))}}{\int dX e^{-\beta(H(X)+H'(x))}} \quad \boxed{H'(x) = \lambda f(x)}$$

# Experimentally Directed Simulation (EDS) Method

**Challenge:** how to estimate many Lagrange multipliers for a complex system?

- Stochastic gradient descent. Iteratively minimize squared error:

$$E(\vec{\lambda}) = \sum_i \left( \int f_i(X) P(X, \vec{\lambda}) dX - \hat{f}_i \right)^2$$

$$P(X, \vec{\lambda}) \propto e^{-\beta(H(X) + \sum_i \lambda_i f_i(X))}$$

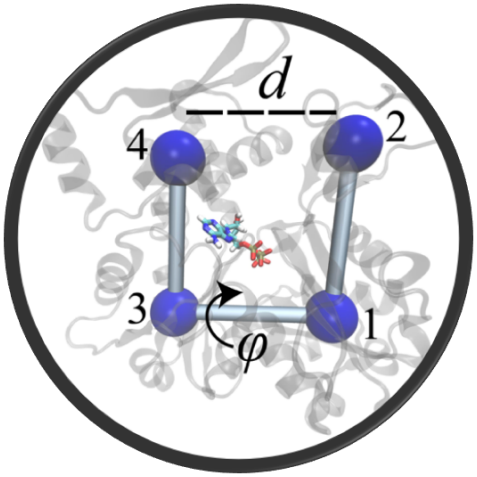
  - For a given set of bias parameters, run for time $\tau$ to compute sample average of $f$

  - Then, choose randomly observable $i$=1…$N$ and:

$$\lambda_i^{t+1} = \lambda_i^t - \gamma_t \frac{\partial E(X)}{\partial \lambda_i}$$

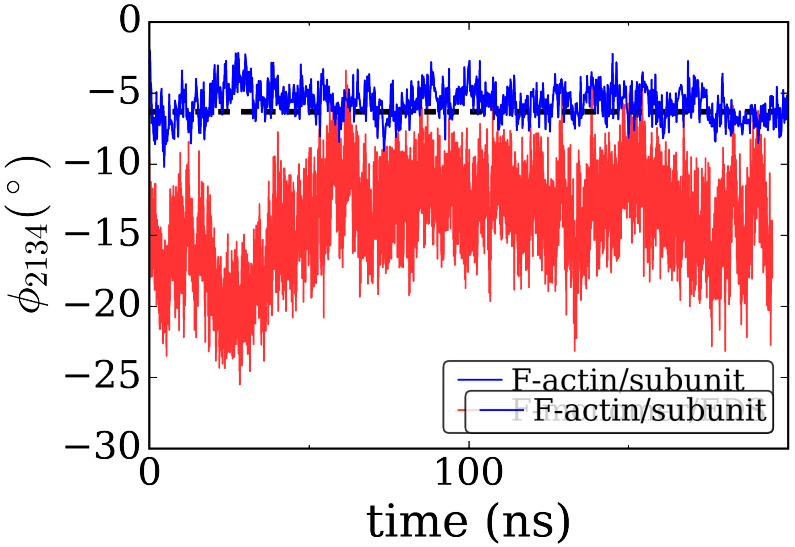
$$\frac{\partial E(X)}{\partial \lambda_i} = -2\beta \langle f(X) - \hat{f} \rangle_t \mathrm{Var}(f(X))$$

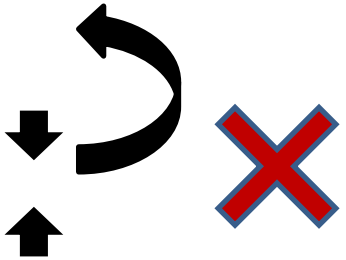
# Application to Actin



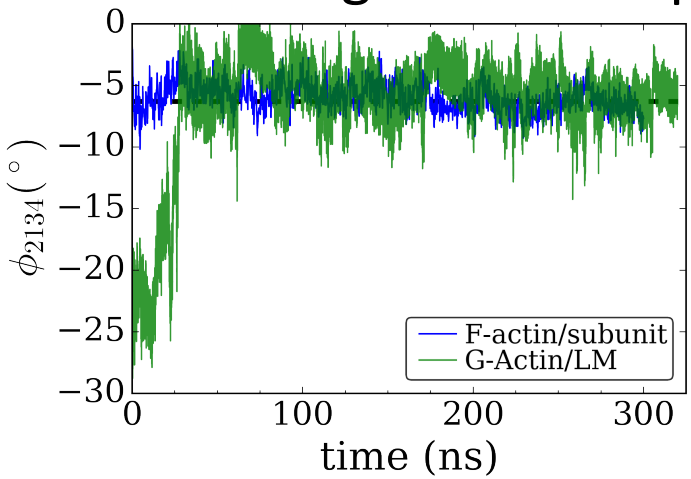

Biasing $\phi, \delta\phi^2, d, \delta d^2$ to match filament values

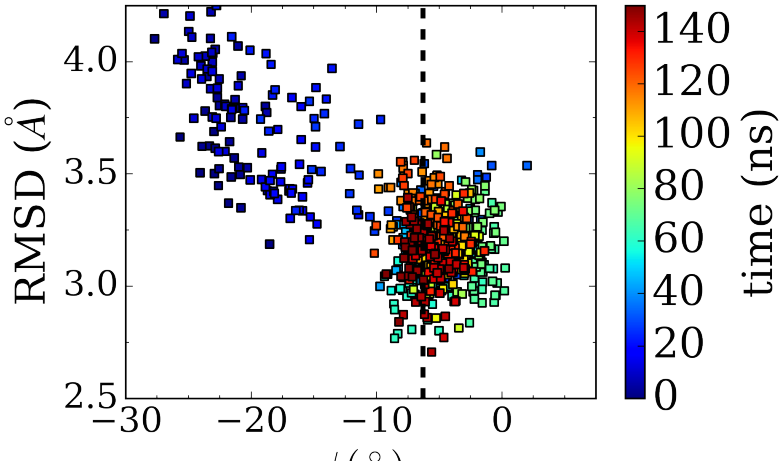
Naïve application failed:



After extensive algorithmic improvements:

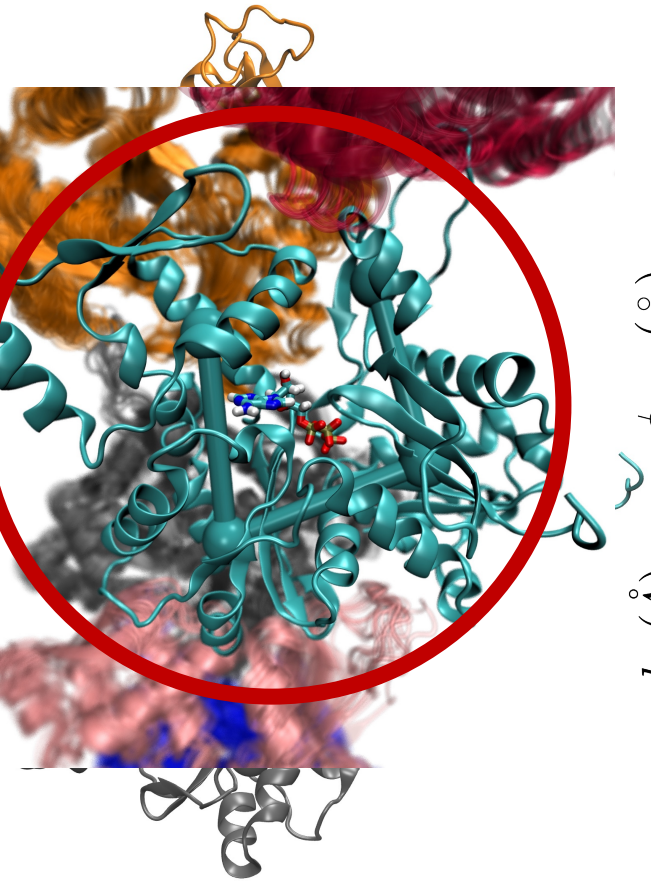


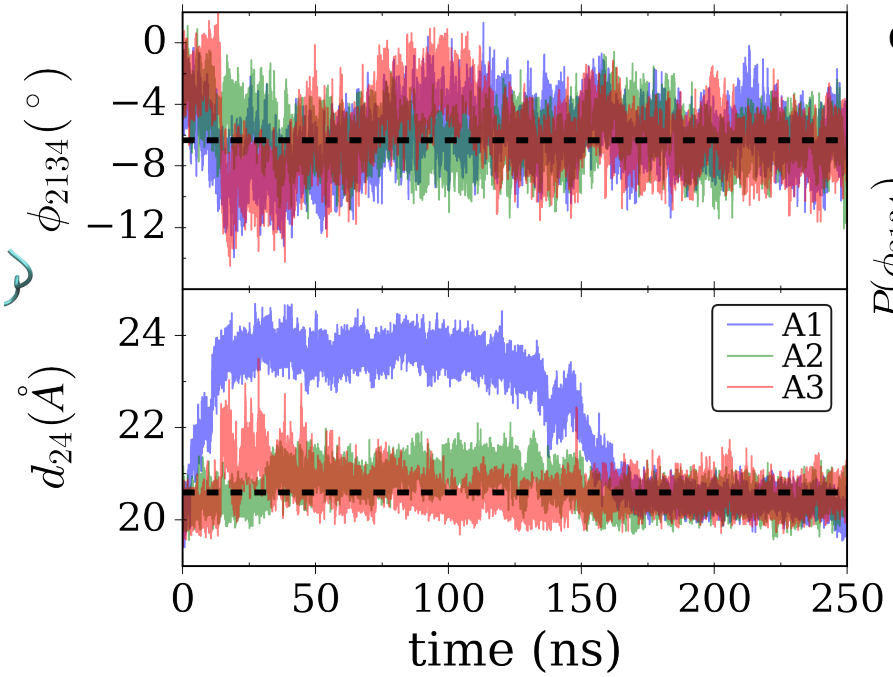
# Larger Subsystems

**In general**: want to simulate smallest possible sub-system

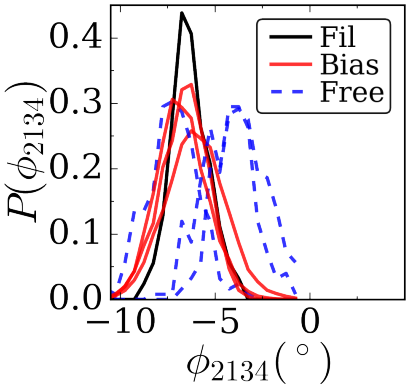
**However**: larger subsystem contains extra context



Biasing only $\phi_1, \phi_2, \phi_3, d_1, d_2, d_3$:

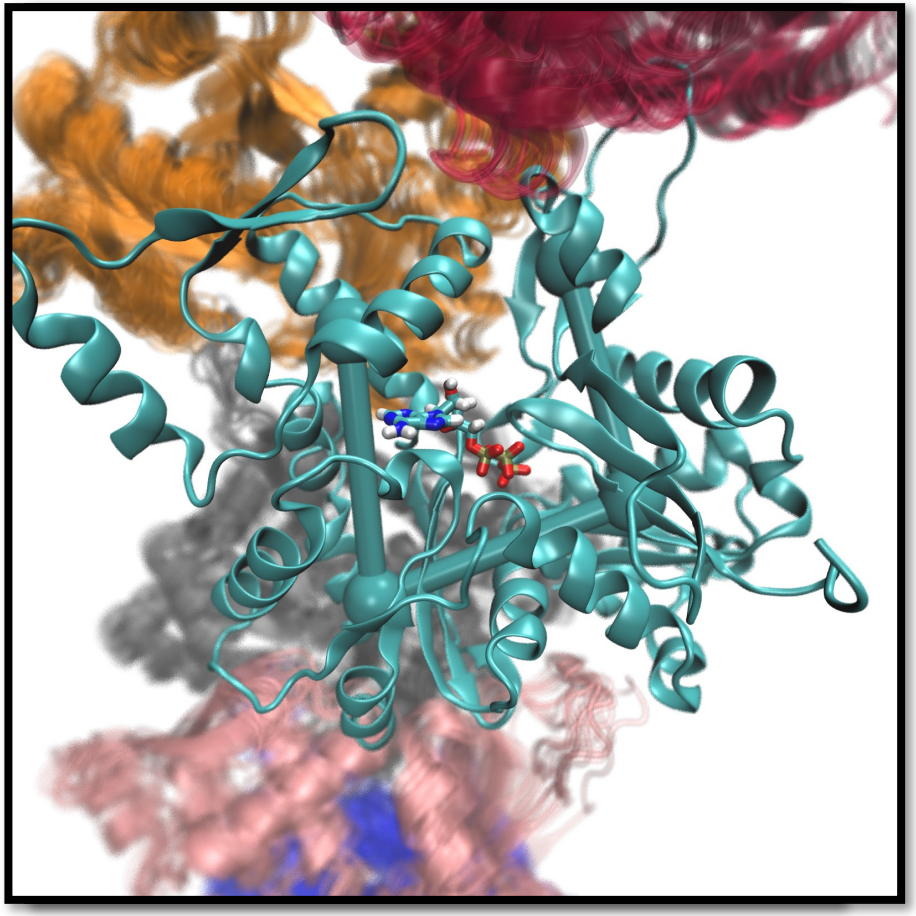



Get variance for free from allosteric contacts:



# CGDS Summary

- Biasing coarse grained observables via learned linear bias parameters is a promising way to represent a subsystem rather than treating a large macromolecular assembly
- Systematically improved learning algorithms, which should apply to any experiment directed simulation
- Available for use now →



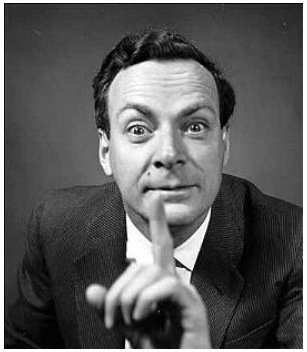
Hocky, Dannenhoffer-Lafage, Voth
*J. Chem. Theory Comput.* (2017)

# Now for a little quantum mechanics….

# Quantum Statistical Mechanics … "Coarse-graining away" the quantum*

*A. V. Sinitskiy and G. A. Voth, "A Reductionist Perspective on Quantum Statistical Mechanics: Coarse-Graining of Path Integrals", J. Chem. Phys. **143**, 094104 (2015).

# Path Integral Formulation

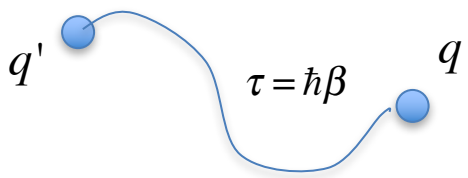


Classical isomorphism (Feynman):

quantum ⟺ classical

$N$ particles

$$\langle q | e^{-\beta H} | q' \rangle$$



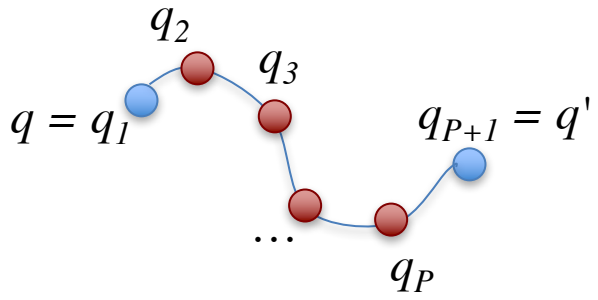

PIMD or PIMC simulations:

Run simulations for large $P$

Chandler, Wolnyes, Berne, Parrinello, Klein, Doll, etc (early 1980's)

$NP$ particles $(P \to \infty)$

$$\left(\frac{mP}{2\pi\hbar^2\beta}\right)^{P/2} \int dq_2 \ldots dq_P e^{-\beta V_P(q, q_2, \ldots q_P, q_{P+1})}$$



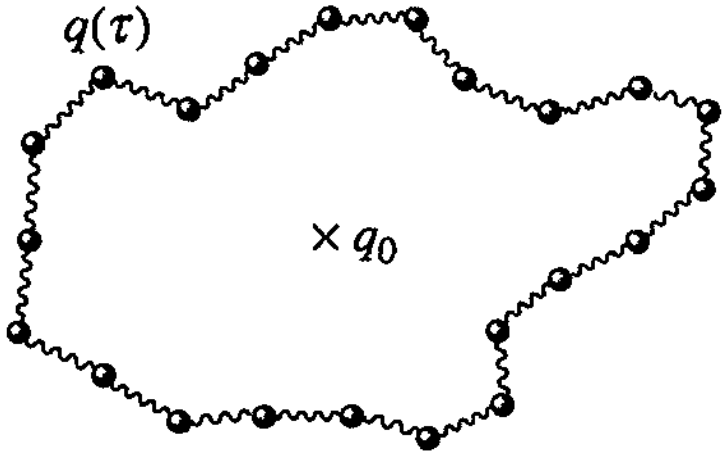

where $$V_P(q, q_2, \ldots, q_P, q') = \sum_{i=1}^{P} \left[ \frac{mP}{2\hbar^2\beta^2}(q_i - q_{i+1})^2 + \frac{V(q_i) + V(q_{i+1})}{2P} \right]$$

# Our Past: Imaginary Time Path Centroids

- PI centroid density: analogue of the Boltzmann density from classical statistical mechanics

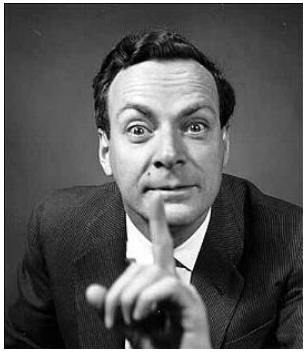
- **How to deal with operators?**

- Center of the cyclic paths: **no off-diagonal elements**.




*J. Cao, G.A. Voth.* (1993) *J. Chem. Phys* and subsequent papers.

# Path Integral Formulation

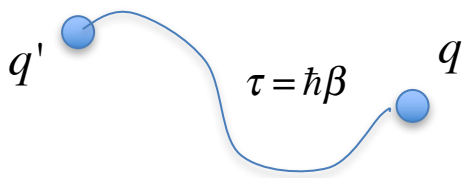


Classical isomorphism (Feynman):

quantum ⟺ classical

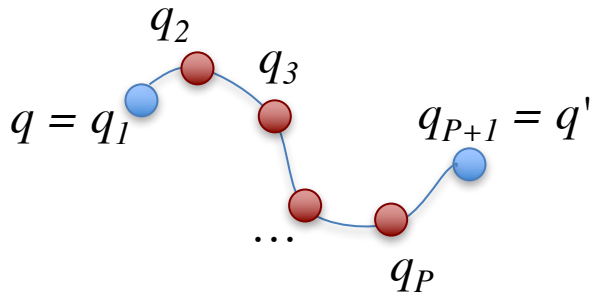
$N$ particles

$$\langle q | e^{-\beta H} | q' \rangle$$




$NP$ particles $(P \to \infty)$

$$\left(\frac{mP}{2\pi\hbar^2\beta}\right)^{P/2} \int dq_2 \ldots dq_P e^{-\beta V_P(q, q_2, \ldots q_P, q_{P+1})}$$



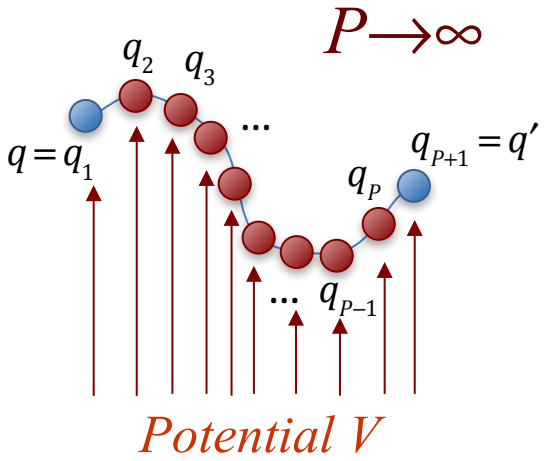

PIMD or PIMC simulations:

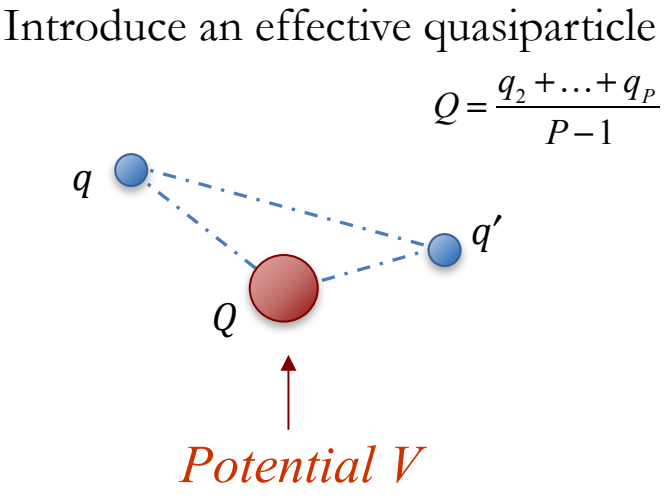
Run simulations for large $P$

Chandler, Wolyes, Berne, Parrinello, Klein, Doll, etc (early 1980's)

where $V_P(q, q_2, \ldots, q_P, q') = \sum_{i=1}^{P} \left[ \frac{mP}{2\hbar^2\beta^2}(q_i - q_{i+1})^2 + \frac{V(q_i) + V(q_{i+1})}{2P} \right]$

# … and so this is what we do here for path integrals




Isomorphic potential $V_P(q,q_2,\ldots,q_P,q')$

Coarse-grained potential $V_{CG}(q,Q,q')$

Strict definition:

$$V_{CG}(q,Q,q')=-\frac{1}{\beta}\lim_{P\to\infty}\left\{\ln\left[\left(\frac{\pi\hbar^2\beta}{m}\right)\left(\frac{mP}{2\pi\hbar^2\beta}\right)^{P/2}\int dq_2\ldots dq_P\, e^{-\beta V_P(q,q_2,\ldots,q_P,q')}\delta\left(Q-\frac{q_2+\ldots+q_P}{P-1}\right)\right]\right\}$$

Sinitskiy and Voth, J. Chem. Phys. **143**, 094104 (2015)

# Weyl Map: Now Momentum Comes in Play

- The Weyl map $W_A$ is a classical-like function corresponding to an arbitrary QM operator $\hat{A}$:

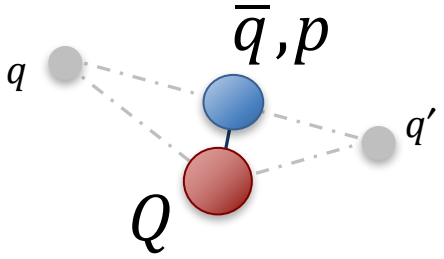
$$\left\langle q' \middle| \hat{A}\left(\hat{p},\hat{q}\right) \middle| q \right\rangle = \int \frac{dp}{2\pi\hbar} e^{\frac{ip\Delta q}{\hbar}} W_A\left(p,\bar{q}\right)$$

Weyl *Zeitschrift für Physik* (1927)

- In our formulas, switch from $q$ and $q'$ to

$$\bar{q} = \frac{q+q'}{2}, \quad \Delta q = q' - q$$

and get rid of $\Delta q$ by integration over it.




- The results: **momentum** explicitly enters our formulas; only **two quasiparticles** are left.

Sinitskiy and Voth, J. Chem. Phys. **143**, 094104 (2015)

# New Perspective on Quantum Statistics

- Now the expectation value of a QM operator $\hat{A}$ at equilibrium and at temperature $T$ can be computed as

$$\left\langle \hat{A}\left(\hat{p},\hat{q}\right)\right\rangle = \frac{\int dP_Q\, dQ\, dp\, d\bar{q}\, e^{-\beta H_{eff}\left(P_Q,Q,p,\bar{q}\right)} W_A\left(\bar{q},p\right)}{\int dP_Q\, dQ\, dp\, d\bar{q}\, e^{-\beta H_{eff}\left(P_Q,Q,p,\bar{q}\right)}} = \left\langle W_A\left(p,\bar{q}\right)\right\rangle_{\text{"classical"}}$$

($\beta = 1/k_B T$; Weyl map)

- The classical two-quasiparticle effective Hamiltonian $H_{eff}$ is

$$H_{eff}\left(P_Q,Q,p,\bar{q}\right) = \left[\frac{P_Q^2}{2M_Q} + \frac{p^2}{2m_{eff}\left(\bar{q}\right)} + V\left(Q\right) + \frac{k_{Q\bar{q}}\left(\bar{q}\right)}{2}\left(Q-\bar{q}\right)^2\right] \times \ldots$$

$$\ldots \left\{1 + O\left(\hbar^3 \frac{\beta^{5/2}}{m^{3/2}} V'''\left(\bar{q}\right)\right)\right\}$$

where

$$k_{Q\bar{q}}\left(\bar{q}\right) = \frac{4m}{\hbar^2\beta^2} \cdot \frac{f^2 \tanh f}{f - \tanh f}, \quad m_{eff}\left(\bar{q}\right) = m \cdot \frac{f}{\tanh f}, \quad f = \frac{\hbar\beta}{2}\sqrt{\frac{V''(\bar{q})}{m}}.$$

Sinitskiy and Voth, J. Chem. Phys. **143**, 094104 (2015)

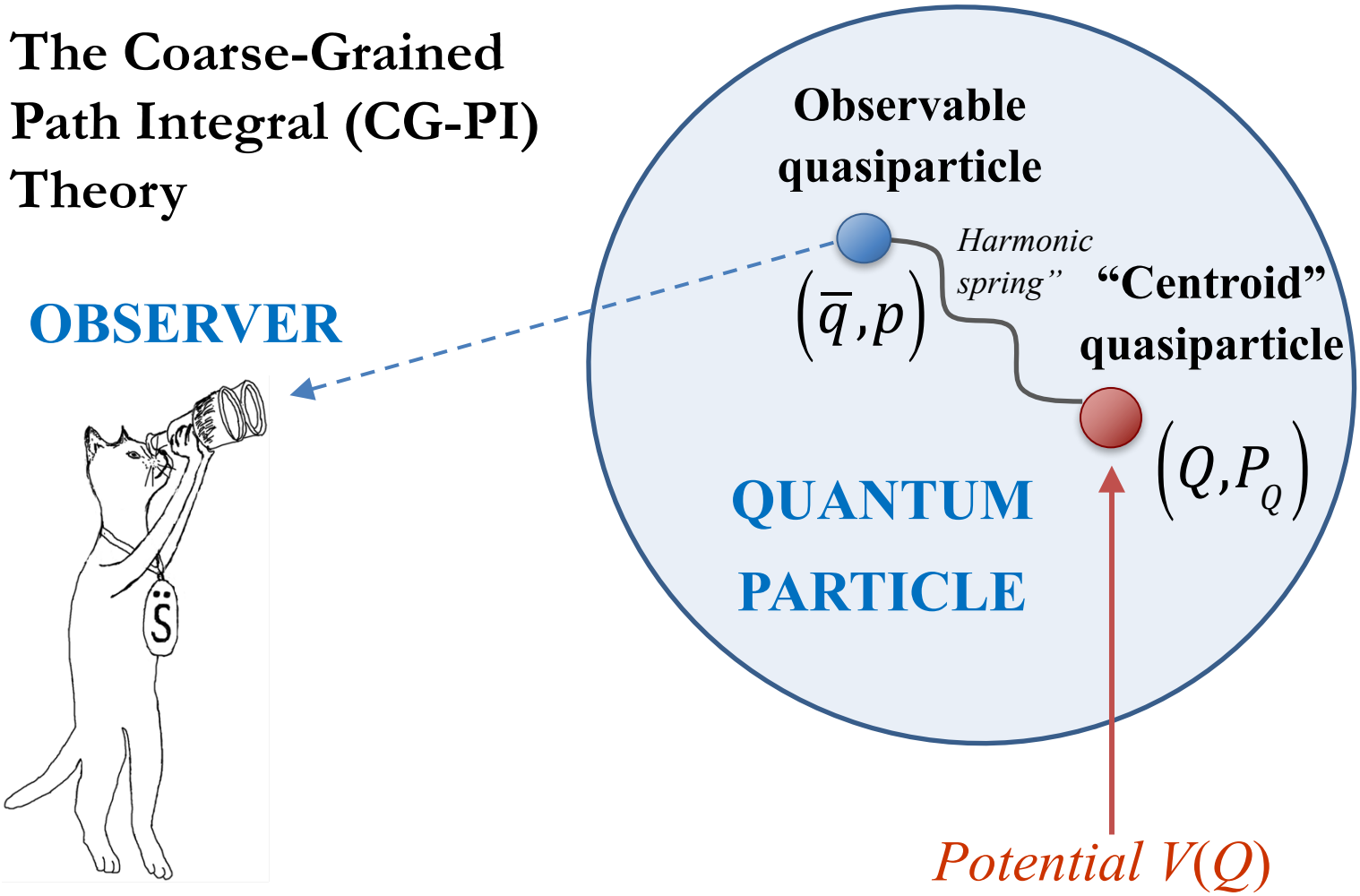
# The Most Simplified Possible Quantum Statistical Mechanics

$$H_{eff}\left(P_Q, Q, p, \bar{q}\right) \cong \frac{P_Q^2}{2M_Q} + \frac{p^2}{2m_{eff}^{const}} + V\left(Q\right) + \frac{k_{Q\bar{q}}\left(\bar{q}\right)}{2}\left(Q - \bar{q}\right)^2$$

Sinitskiy and Voth, J. Chem. Phys. **143**, 094104 (2015)

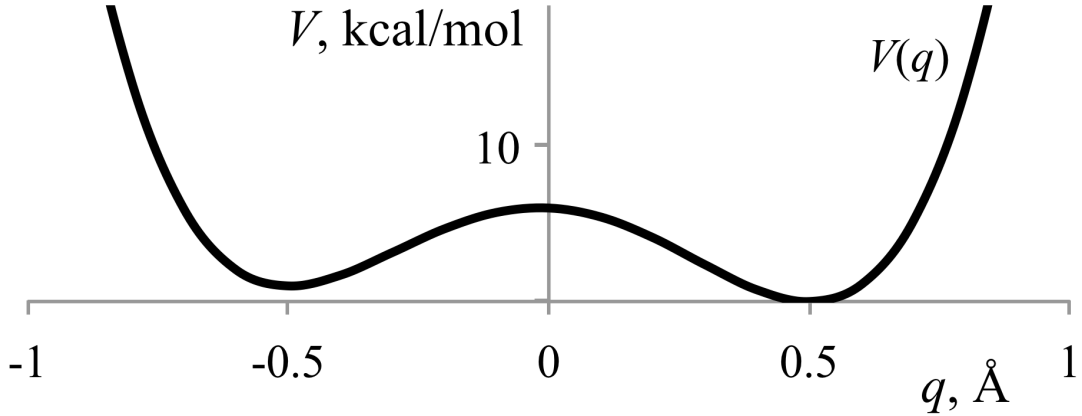
# This is what we have learned about quantum stat mech in its most "reductionist" form:

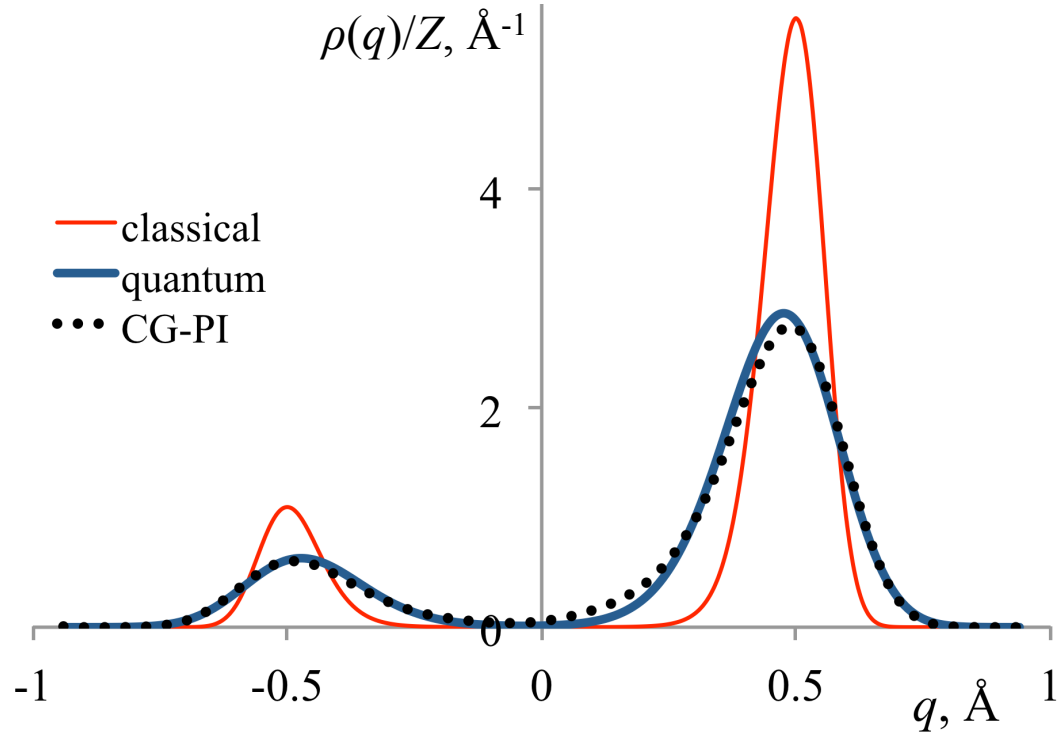
- The Coarse-Grained Path Integral (CG-PI) Theory




Sinitskiy and Voth, J. Chem. Phys. **143**, 094104 (2015)

# Example 2: Tunneling




$$V(q) = \sum_{i=0}^{4} c_i q^i$$

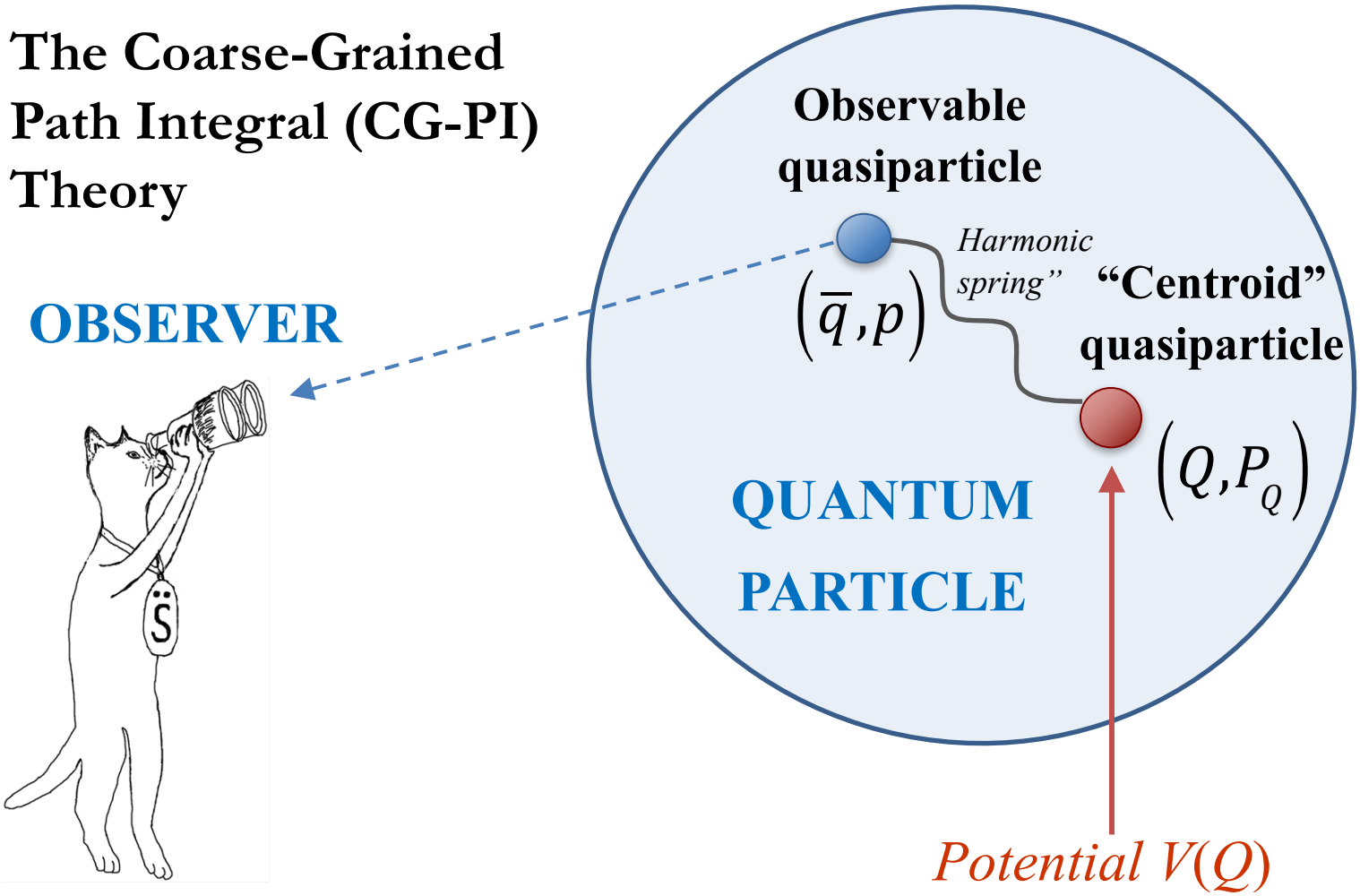
$c_0 = 6, c_1 = -1.5,$

$c_2 = -44, c_3 = 2, c_4 = 88$

- This 1D problem is motivated by hydrogen tunneling through **5 kcal/mol** barrier at **310 K**
- The CG-PI result is very good (vs. classical) and cheap (vs. full quantum)

# This is what we have learned about quantum stat mech in its most "reductionist" form:

- The Coarse-Grained Path Integral (CG-PI) Theory




Sinitskiy and Voth, J. Chem. Phys. **143**, 094104 (2015)

# Summary: For the Future

- **Ultra-Coarse-Graining (UCG): Exciting new capability**
- **Coarse-graining of the quantum mechanics (nuclear motion)**

- **Rigorous "bottom-up" theory for "QM/CG-MM" (JCP 2018)**
- **Quantum theory of MS-CG (qMS-CG) (JCP 2018)**
- **Mesoscopic "non-molecular" coarse-graining (in prep)**
- **Reactive and multi-configurational CG models (in prep)**
- **"On the fly" coarse-graining with quantum electronic structure**

# Acknowledgements


James Dama (Simons Jr Fellow)
John Grime
Alex Pak
Jaehyeok Jin
Aram Davtyan

Anton Sinitskiy

Martin McCullagh (CSU Faculty)
Sergei Izvekov (ARL)
Will Noid (Penn State)
Lanyuan Lu (NTU Singapore)
Jhih-Wei Chu (NCTU Taiwan)
Zhiyong Zhang (USTC)

**Collaborators**

Hans Andersen (Stanford)

Aaron Dinner (UofC)
Benoit Roux (UofC)
Jonathan Weare (UofC)

**Funding**

NSF
NIH-NIGMS