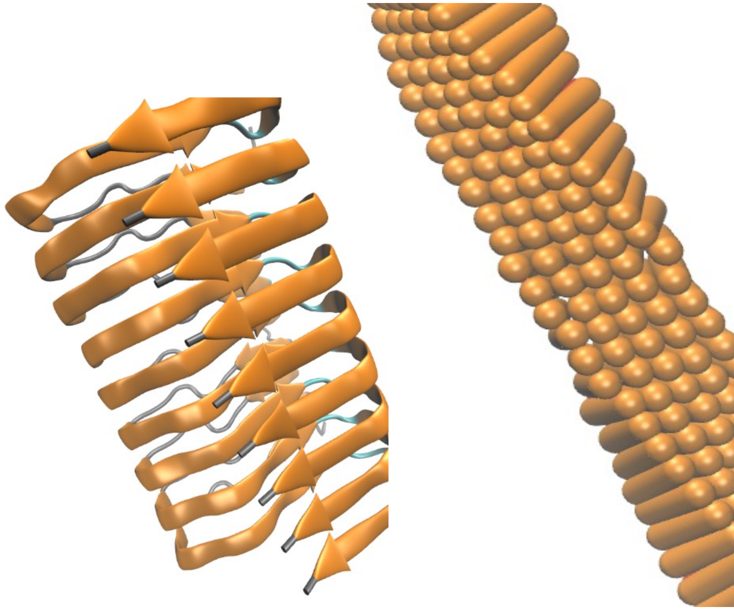
UNIVERSITEIT VAN AMSTERDAM
Faculteit der Natuurwetenschappen,
Wiskunde en Informatica

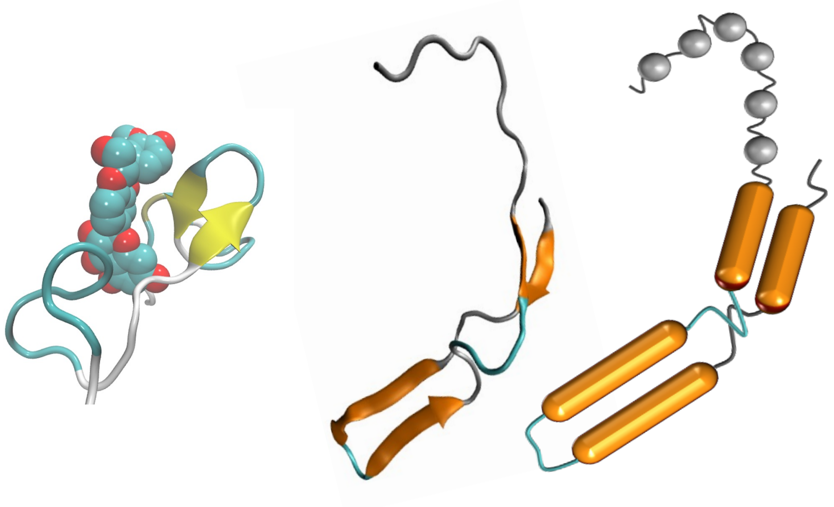
Computational Chemistry
Van 't Hoff Institute for Molecular Sciences
University of Amsterdam

MULTISCALE SIMULATIONS OF
BIOMOLECULAR SYSTEMS



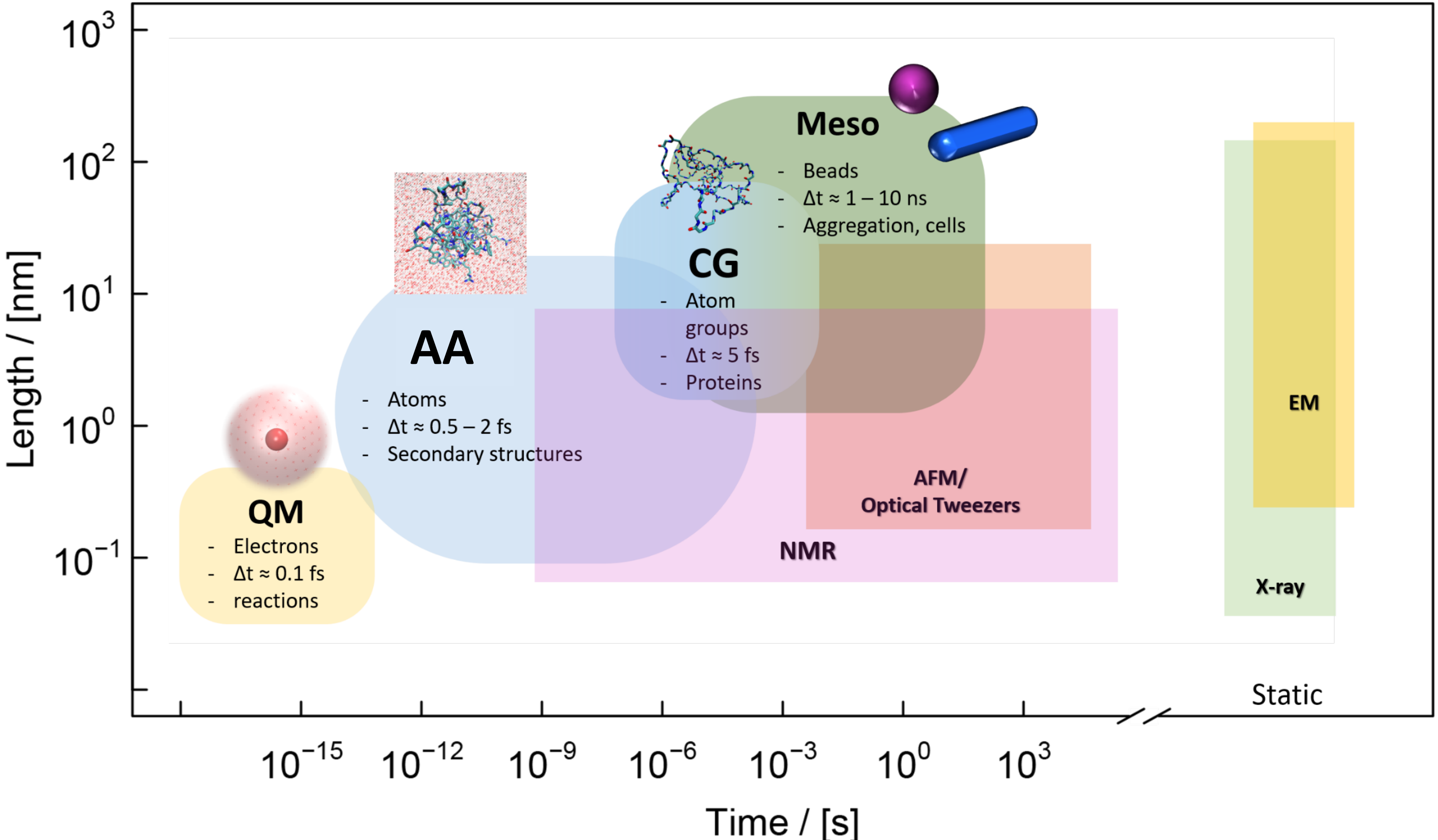
Coarse-graining

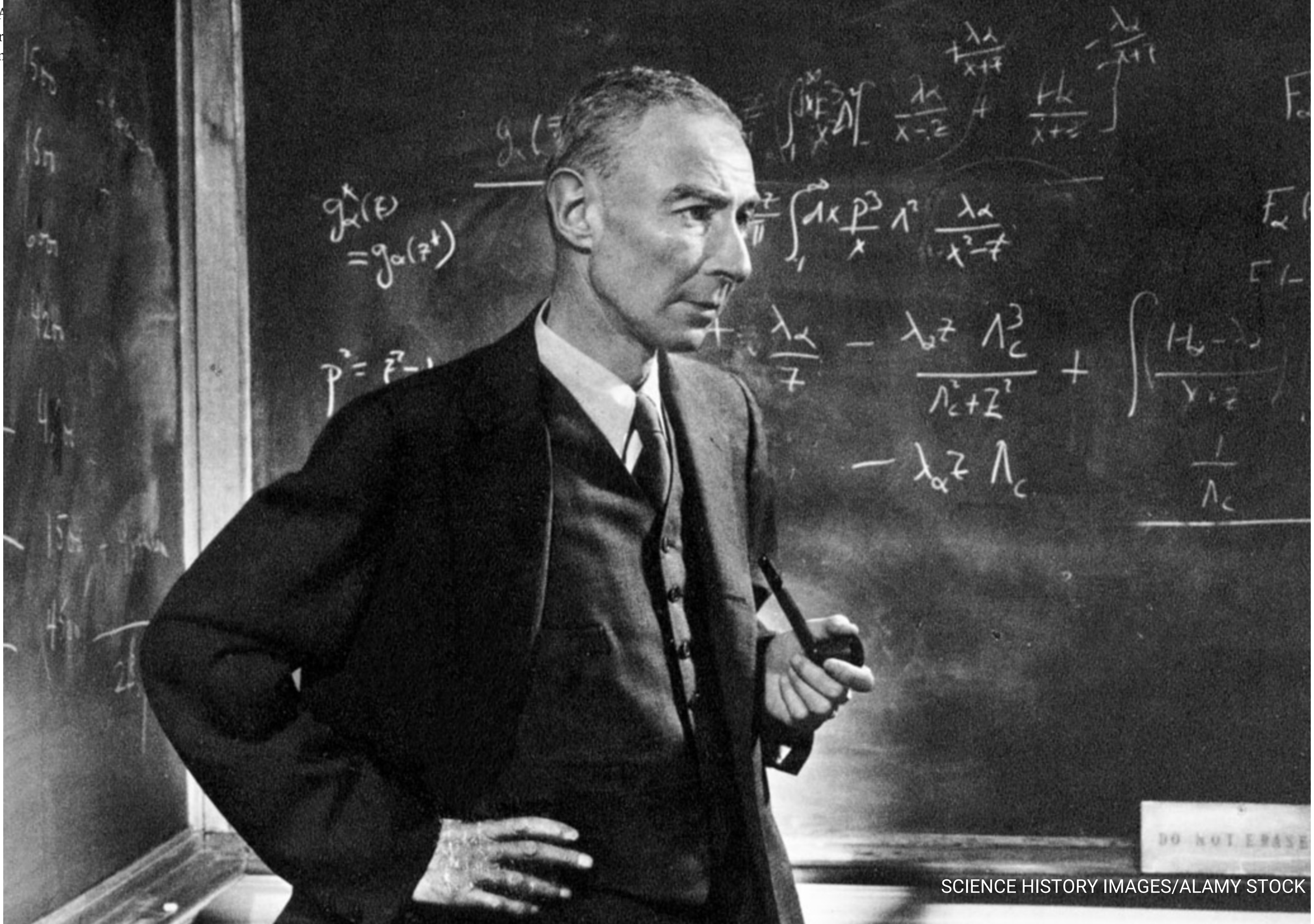
Why? What? How?



ILIE & Caflisch, *Chem. Rev.*, **119**, 6956 (2019)

Why? What?

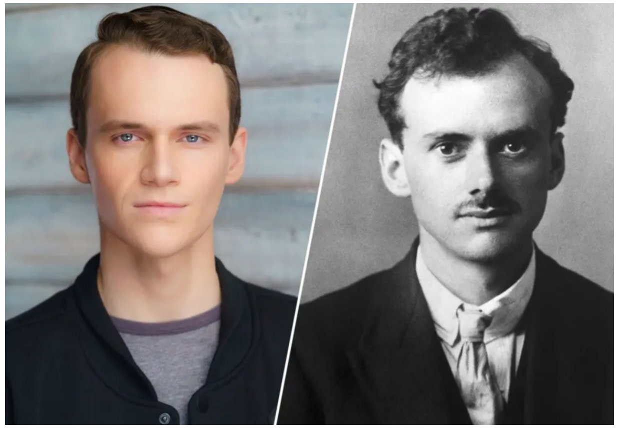




SCIENCE HISTORY IMAGES/ALAMY STOCK

Robert J. Oppenheimer
1904 – 1967
“Father of the atomic bomb”
Born-Oppenheimer approximation (1927)
electronic motion and nuclear motion in molecules can be separated

Richard Feynman
1918 – 1988
1965 – Nobel price *development of quantum electrodynamics*
1959 - designing miniaturized machines



Paul Dirac
1902 – 1984
1933 – Nobel price *discovery of new productive forms of atomic theory*

Erwin Schrödinger
1887 – 1961
1933 – Nobel price *discovery of new productive forms of atomic theory*

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = \hat{H}|\psi(t)\rangle$$

Niels Bohr
1885 – 1962
1922 – Nobel price *investigation of the structure of atoms and the radiation emanating from them*
Developed the model for the atom

Werner Heisenberg
1901 – 1976
1932 – Nobel price *creation of quantum mechanics*

$$\Delta x \Delta p \geq \frac{h}{4\pi}$$

Before CG


Martin Karplus


Michael Levitt


Arieh Warshel

Multiscale models for complex chemical systems

2013

1998

Molecular orbital theory & its use in the calculation of electronic structure of molecules

the era of computing chemists, when hundreds if not thousands of chemists will go to the computing machine instead of the laboratory for increasingly many facets of chemical information, is already at hand.


Robert Mulliken

1966

1940-1950s

Analytical Engine
ENIAC


Niels Bohr


Erwin Schrodinger


Paul Dirac


Werner Heisenberg

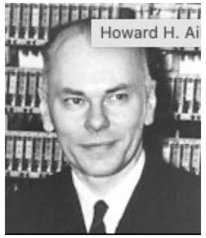
2020

AlphaFold & ML

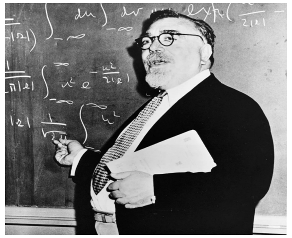

Walter Kohn


John Pople

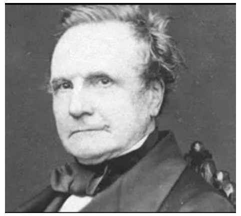
Density functional theory & Computational methods in quantum chemistry



Howard Aiken


Norbert Wiener


Alan Turing


Charles Babbage
(1791-1871)
Analytical Machine

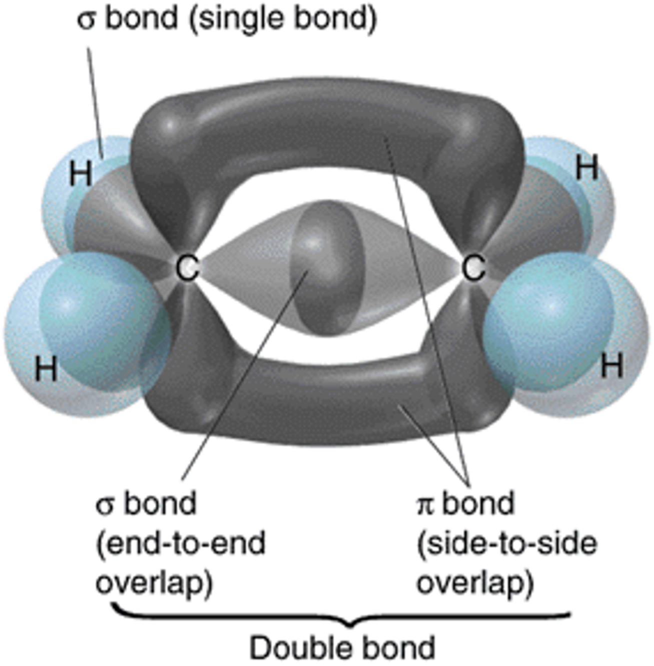
1930s

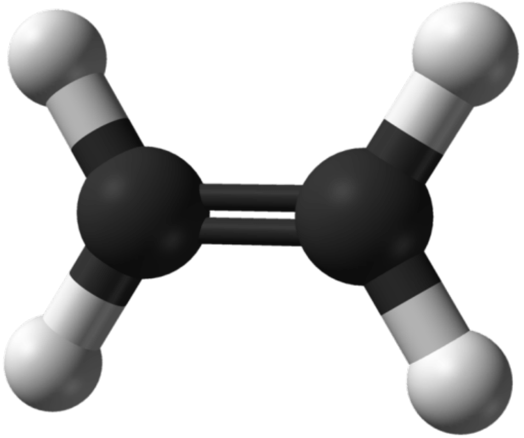
H, He, H_2

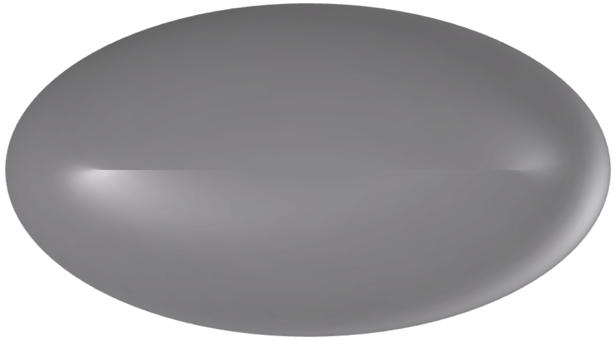
1959

~ 1970

The basics – structural resolution





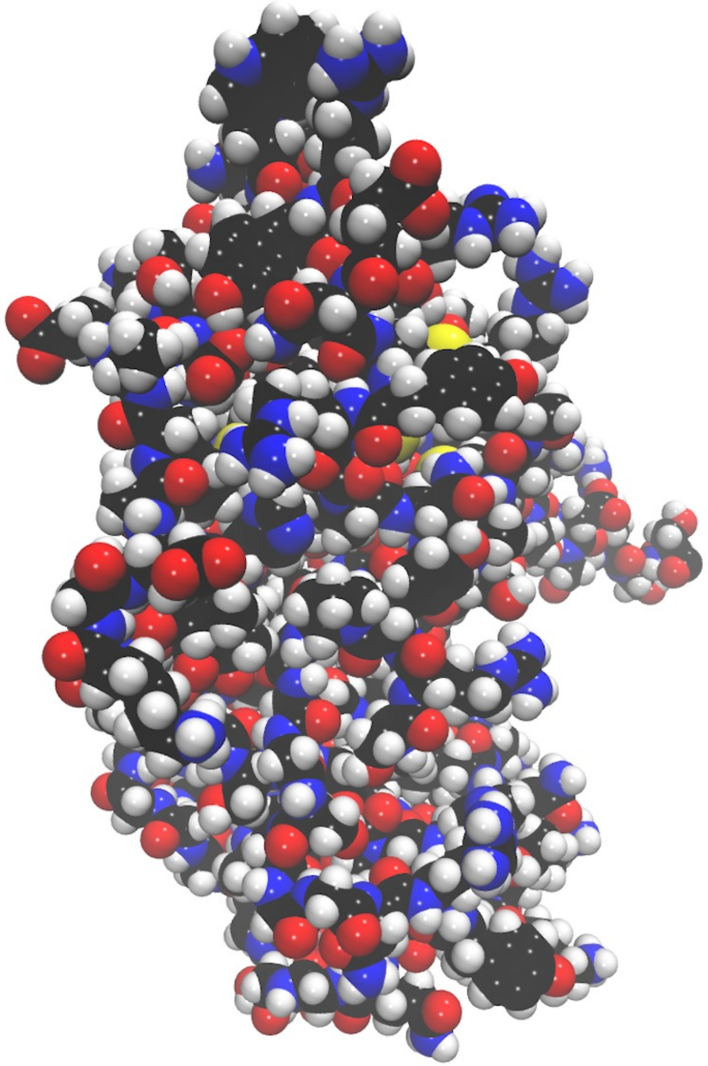


Ethene in electronic structure level representation (orbitals)

Ethene in all-atom molecular mechanics representation (spheres)

Ethene in coarse-grained representation

Molecular Mechanics Representation

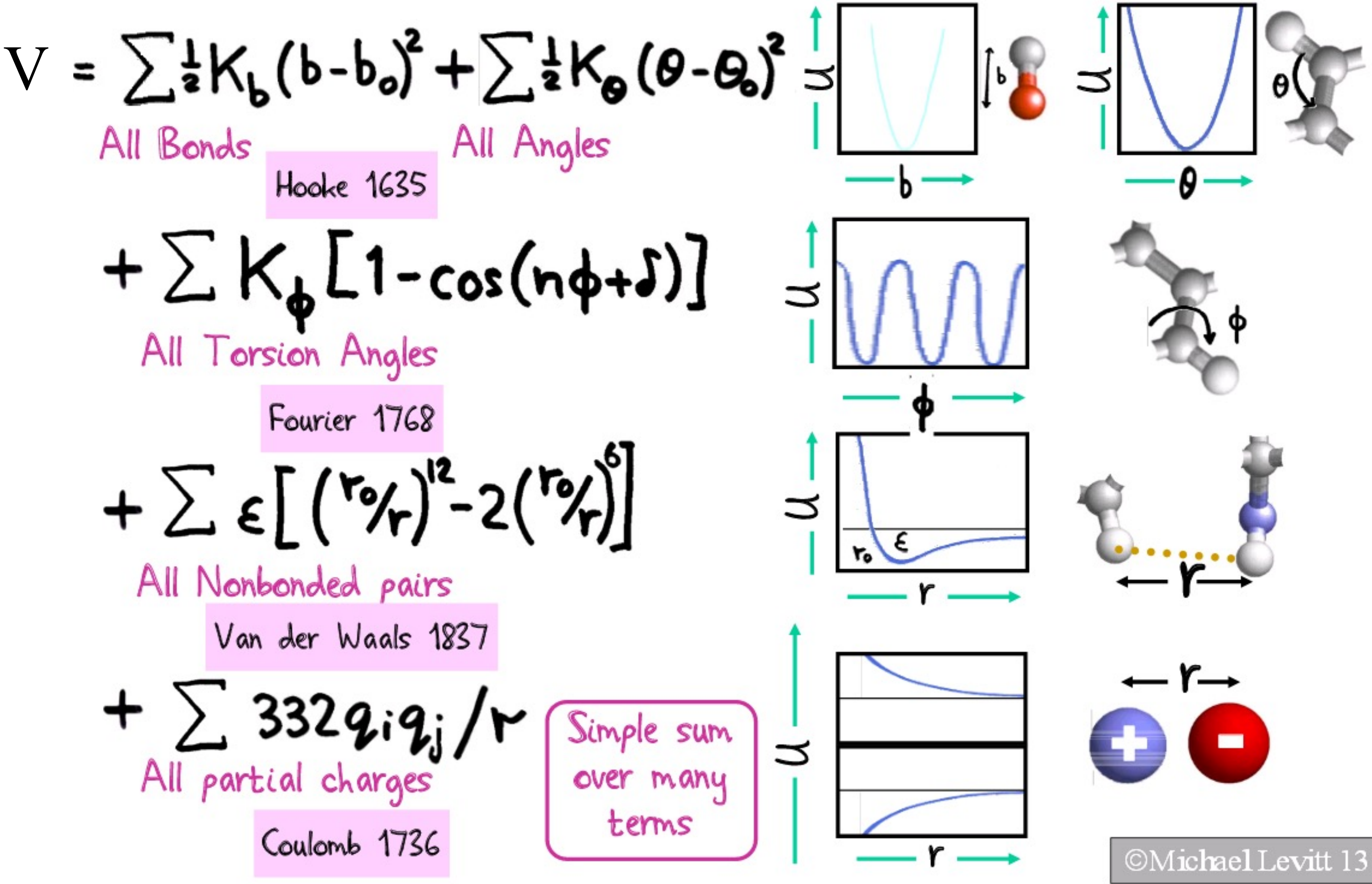


- Atom → sphere
- Biomolecular force field: shapes the collection of spheres into things that look like molecules through so-called **bonded** interactions
- Force field: atom-atom interactions of distal parts through **nonbonded** interactions.
- Electronic structure is in general coarse-grained out and expected to be captured by the force field.

The basics on one slide (classical simulations)

$$F = m\frac{d^2r}{dt^2} = ma$$

$$F = -\frac{dV(r)}{dr}$$



$$V = \sum \tfrac{1}{2}K_b(b-b_0)^2 + \sum \tfrac{1}{2}K_\theta(\theta-\theta_0)^2$$

All Bonds

Hooke 1635

All Angles

$$+ \sum K_\phi[1-\cos(n\phi+\delta)]$$

All Torsion Angles

Fourier 1768

$$+ \sum \varepsilon\left[\left(r_0/r\right)^{12} - 2\left(r_0/r\right)^{6}\right]$$

All Nonbonded pairs

Van der Waals 1837

$$+ \sum 332\, q_i q_j / r$$

All partial charges

Coulomb 1736

Simple sum over many terms

How does it work?

- The force field is an analytical function of the spatial coordinates of the atomic nuclei
- Several different force fields are commonly available (CHARMM, AMBER, OPLS)

$$F = m\frac{d^2r}{dt^2} = ma$$

Given the potential, one can numerically integrate the trajectory of the whole system as a function of time.

$$F = -\frac{dV(r)}{dr}$$

The force F is given by the gradient of the potential *V(r)*.

$$\mathrm{V} = \mathrm{V}_{\mathrm{bonds}} + \mathrm{V}_{\mathrm{angles}} + \mathrm{V}_{\mathrm{dihedrals}} + \mathrm{V}_{\mathrm{nonbonded}}$$

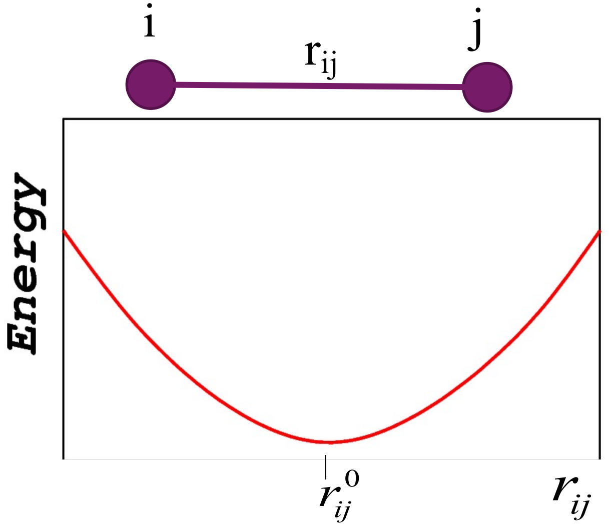
$$V = V_{\text{bonds}} + V_{\text{angles}} + V_{\text{dihedrals}} + V_{\text{nonbonded}}$$

the energy needed to **stretch** a covalent bond between two atoms by Hooke's law for the potential energy stored in a spring

$$V_{bonds}(r_{ij}) = k_{ij}^{B}(r_{ij} - r_{ij}^{0})^{2}$$

k_{ij}^{B} - constant (bond stiffness)

r_{ij}^{0} - equilibrium distance

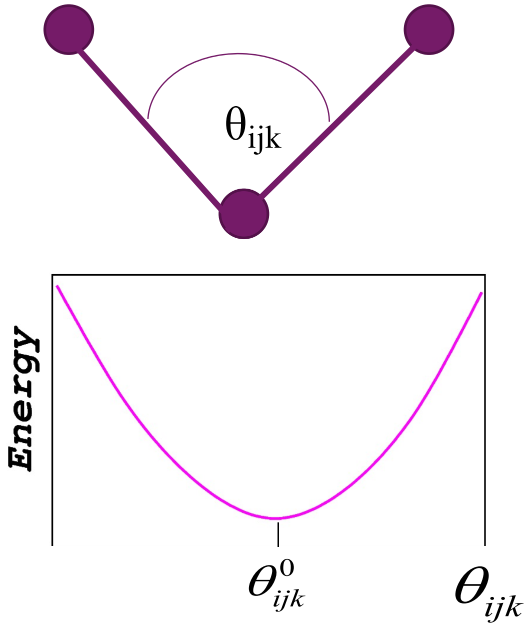


the energy needed to **bend** the angle formed by **two** covalent bonds

$$V_{angles}(\theta_{ijk}) = k_{ijk}^{\theta}(\theta - \theta_{ijk}^{0})^{2}$$

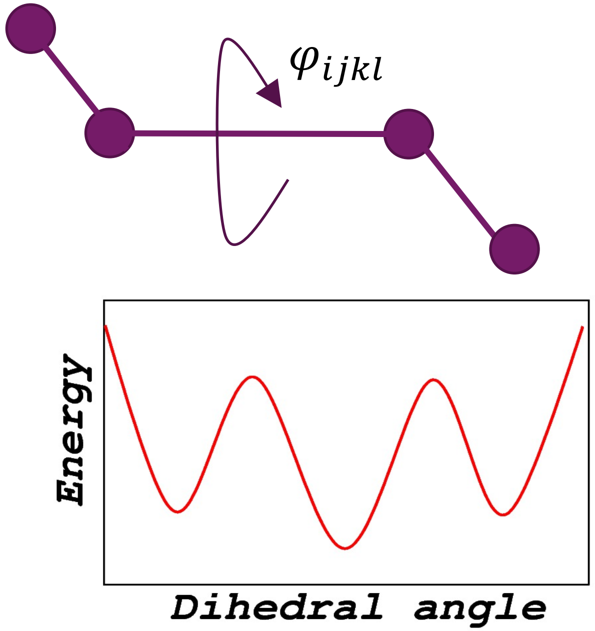
k_{ij}^{θ} - constant (angle rigidity)

θ_{ijk}^{0} - equilibrium angle width



the energy needed to **bend** the dihedral angle formed by **three** covalent bonds

$$V_{dihedrals}(\varphi_{ijkl}) = k_{ijkl}^{\varphi}[1 + \cos(n\varphi_{ijkl} - \delta)]$$



$$V = V_{\text{bonds}} + V_{\text{angles}} + V_{\text{dihedrals}} + V_{\text{nonbonded}}$$

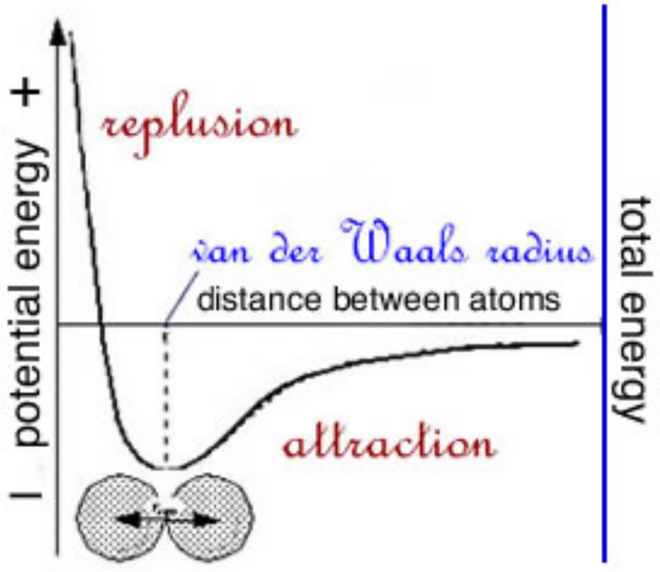
for non-covalently bound atoms

$$V_{\text{vdW}} + V_{\text{electrostatic}}$$

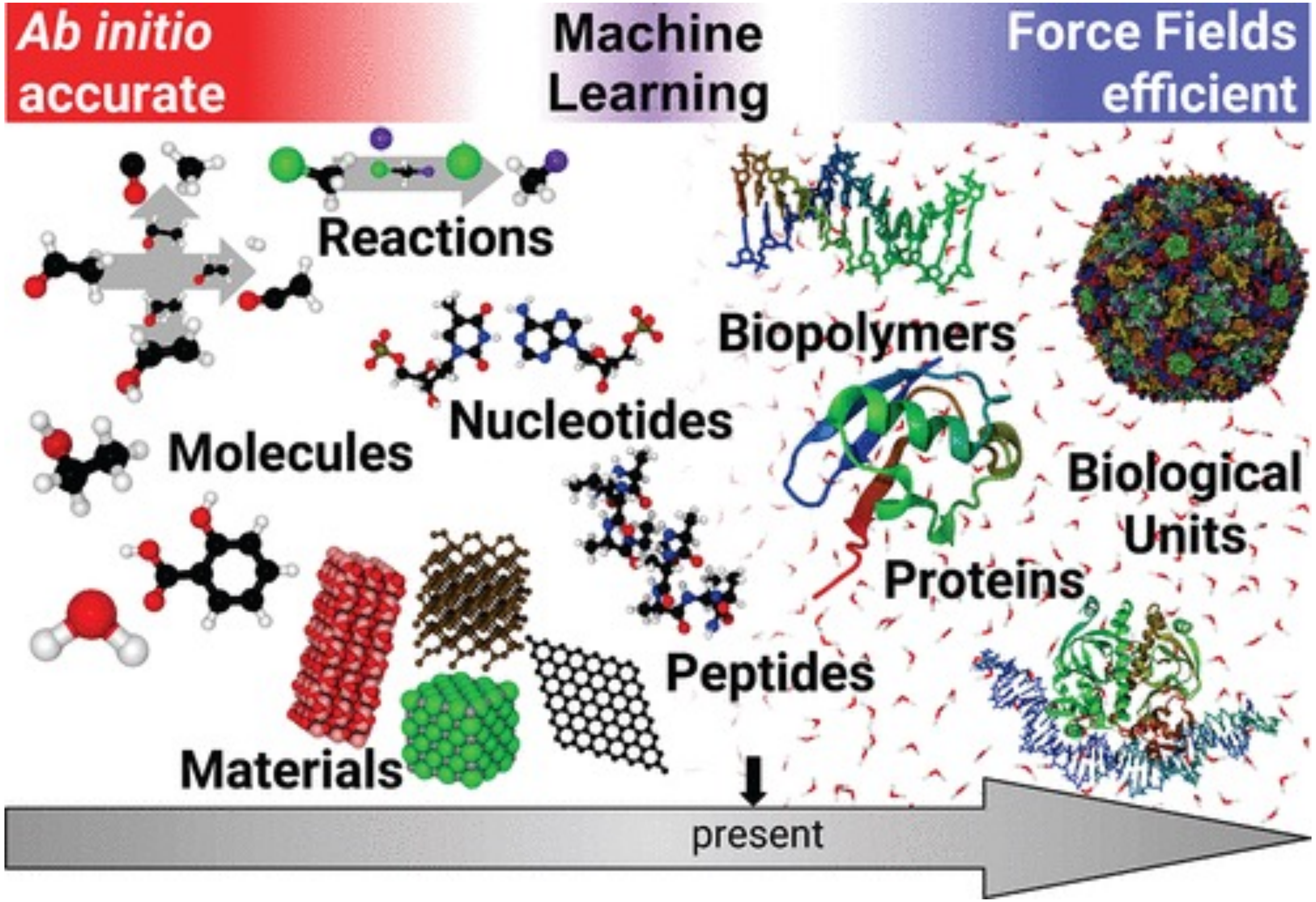
$$V_{vdW} = 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma}{r_{ij}}\right)^{6}\right]$$

$$V_{electrostatic} = \frac{q_i q_j}{\varepsilon_d r_{ij}}$$

ε_d - dielectric constant of the surrounding medium
q_i, q_j - charges

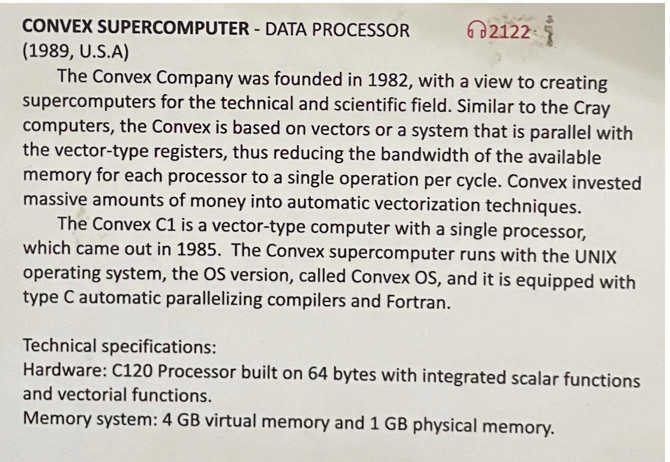


The dream!



Unke et al., *Chem. Rev.*, **121**, 10142 (2021)

Intermezzo – Why?

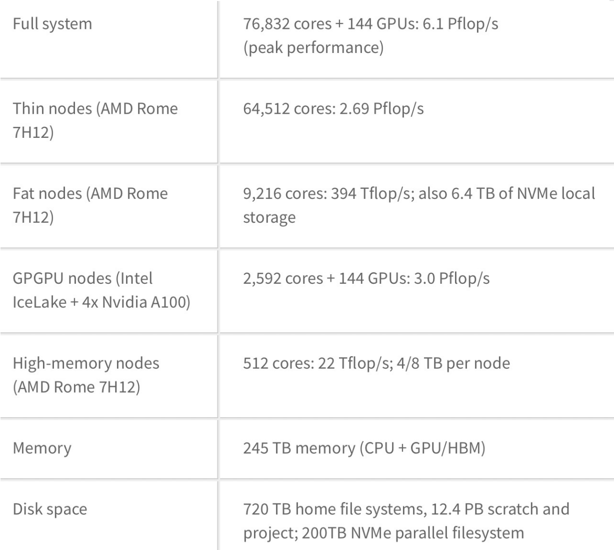


CONVEX SUPERCOMPUTER - DATA PROCESSOR 🎧2122
(1989, U.S.A)

The Convex Company was founded in 1982, with a view to creating supercomputers for the technical and scientific field. Similar to the Cray computers, the Convex is based on vectors or a system that is parallel with the vector-type registers, thus reducing the bandwidth of the available memory for each processor to a single operation per cycle. Convex invested massive amounts of money into automatic vectorization techniques.

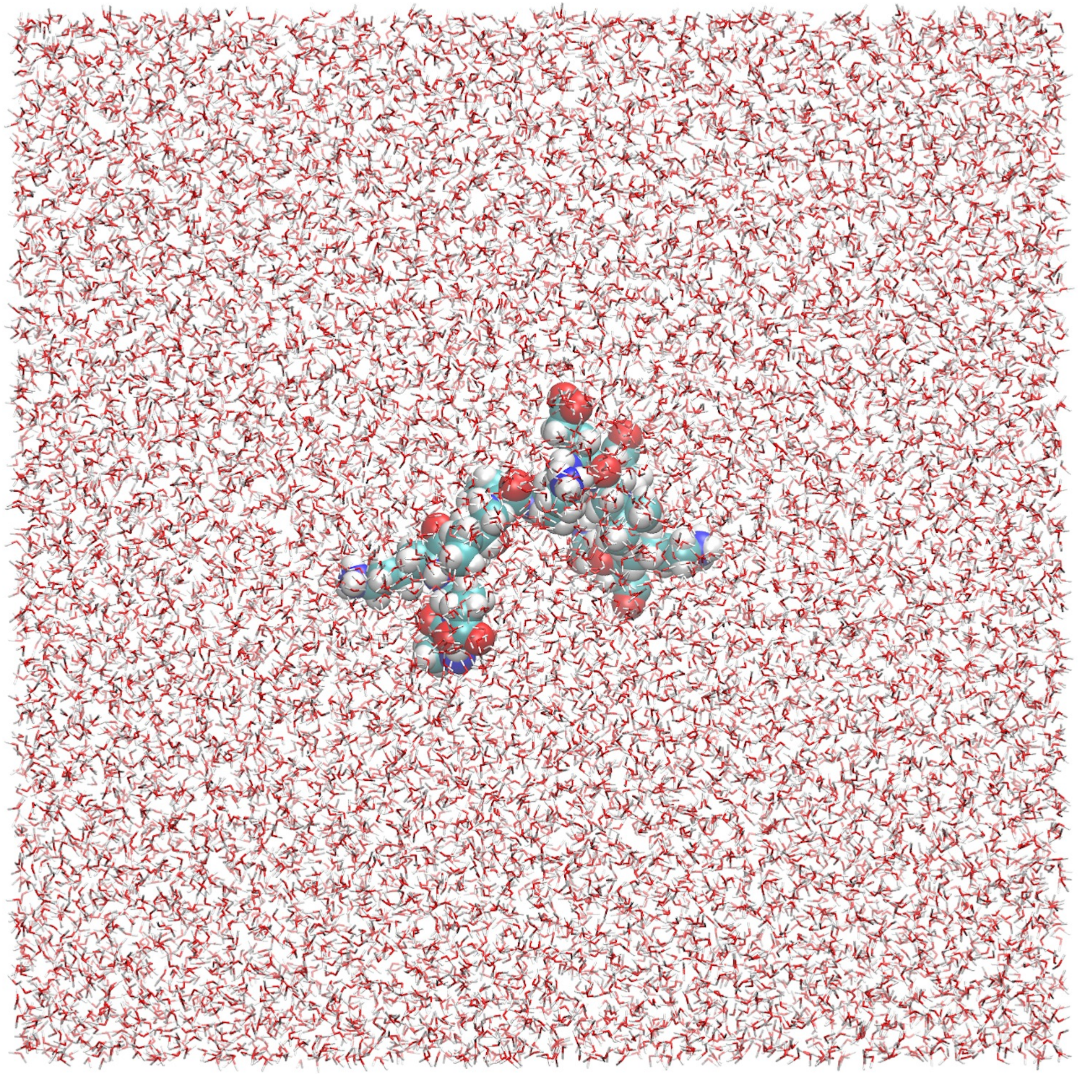
The Convex C1 is a vector-type computer with a single processor, which came out in 1985. The Convex supercomputer runs with the UNIX operating system, the OS version, called Convex OS, and it is equipped with type C automatic parallelizing compilers and Fortran.

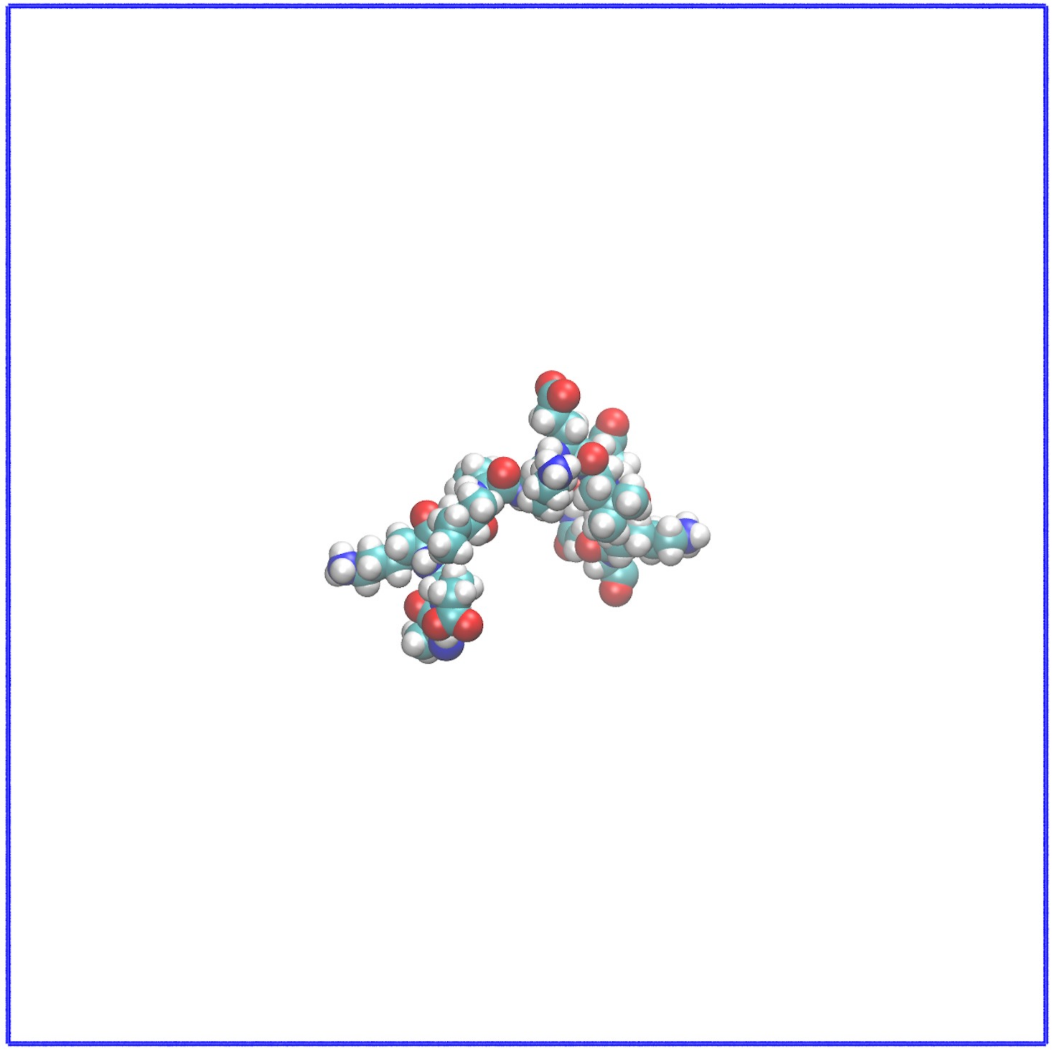
Technical specifications:
Hardware: C120 Processor built on 64 bytes with integrated scalar functions and vectorial functions.
Memory system: 4 GB virtual memory and 1 GB physical memory.



Full system	76,832 cores + 144 GPUs: 6.1 Pflop/s (peak performance)
Thin nodes (AMD Rome 7H12)	64,512 cores: 2.69 Pflop/s
Fat nodes (AMD Rome 7H12)	9,216 cores: 394 Tflop/s; also 6.4 TB of NVMe local storage
GPGPU nodes (Intel IceLake + 4x Nvidia A100)	2,592 cores + 144 GPUs: 3.0 Pflop/s
High-memory nodes (AMD Rome 7H12)	512 cores: 22 Tflop/s; 4/8 TB per node
Memory	245 TB memory (CPU + GPU/HBM)
Disk space	720 TB home file systems, 12.4 PB scratch and project; 200TB NVMe parallel filesystem

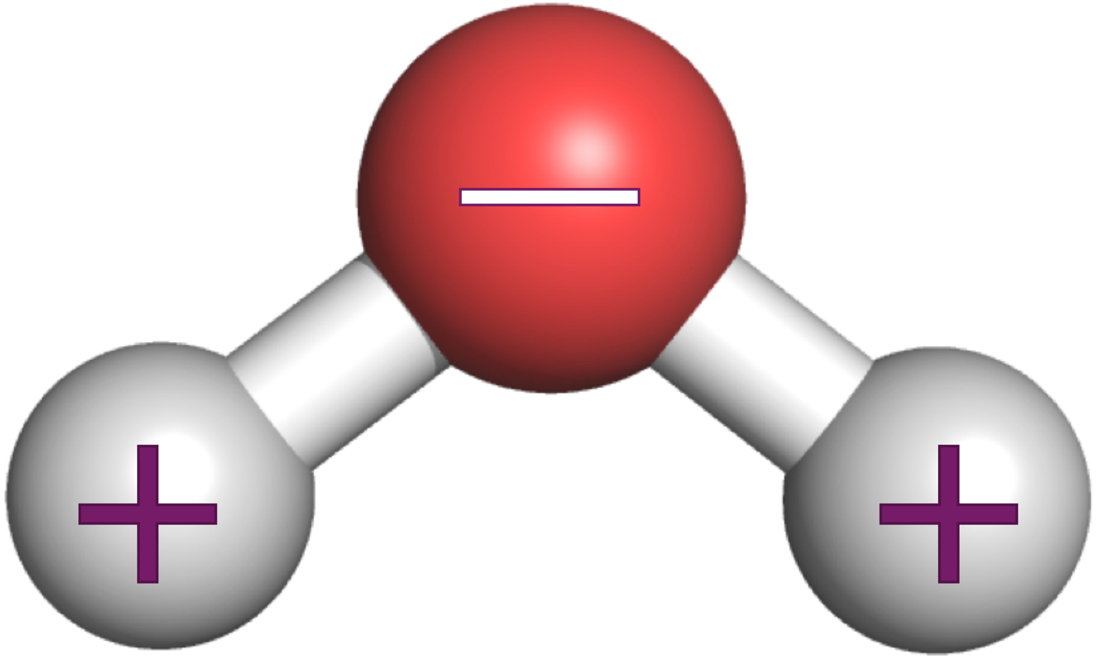
How does it work?



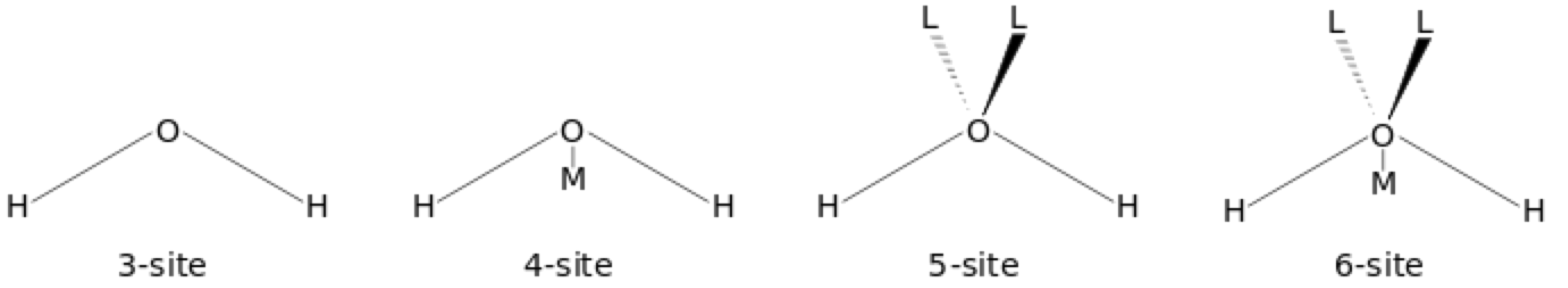


Water models

- **Rigid**
  - Fixed atom positions
  - Only non-bonded interactions
- **Flexible**
  - Atoms on “springs”
  - Include bond stretching and angle bending
  - Reproduce vibration spectra
- **Polarizable**
  - Include specific polarization terms



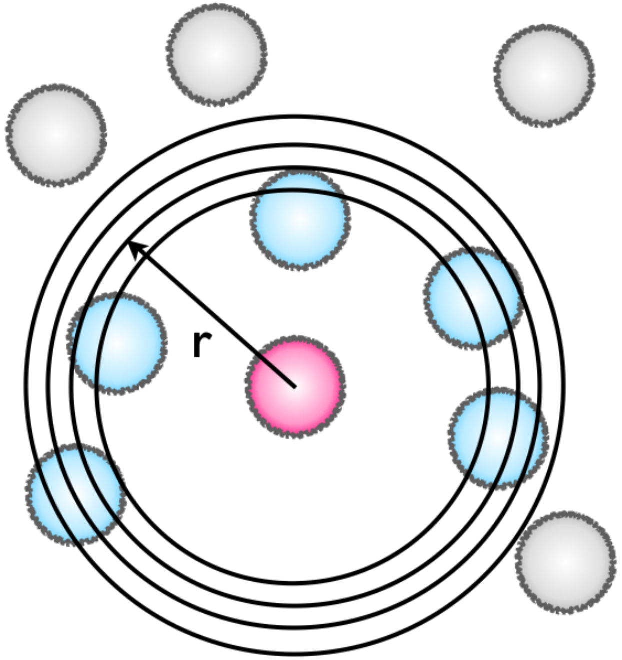
Rigid water models

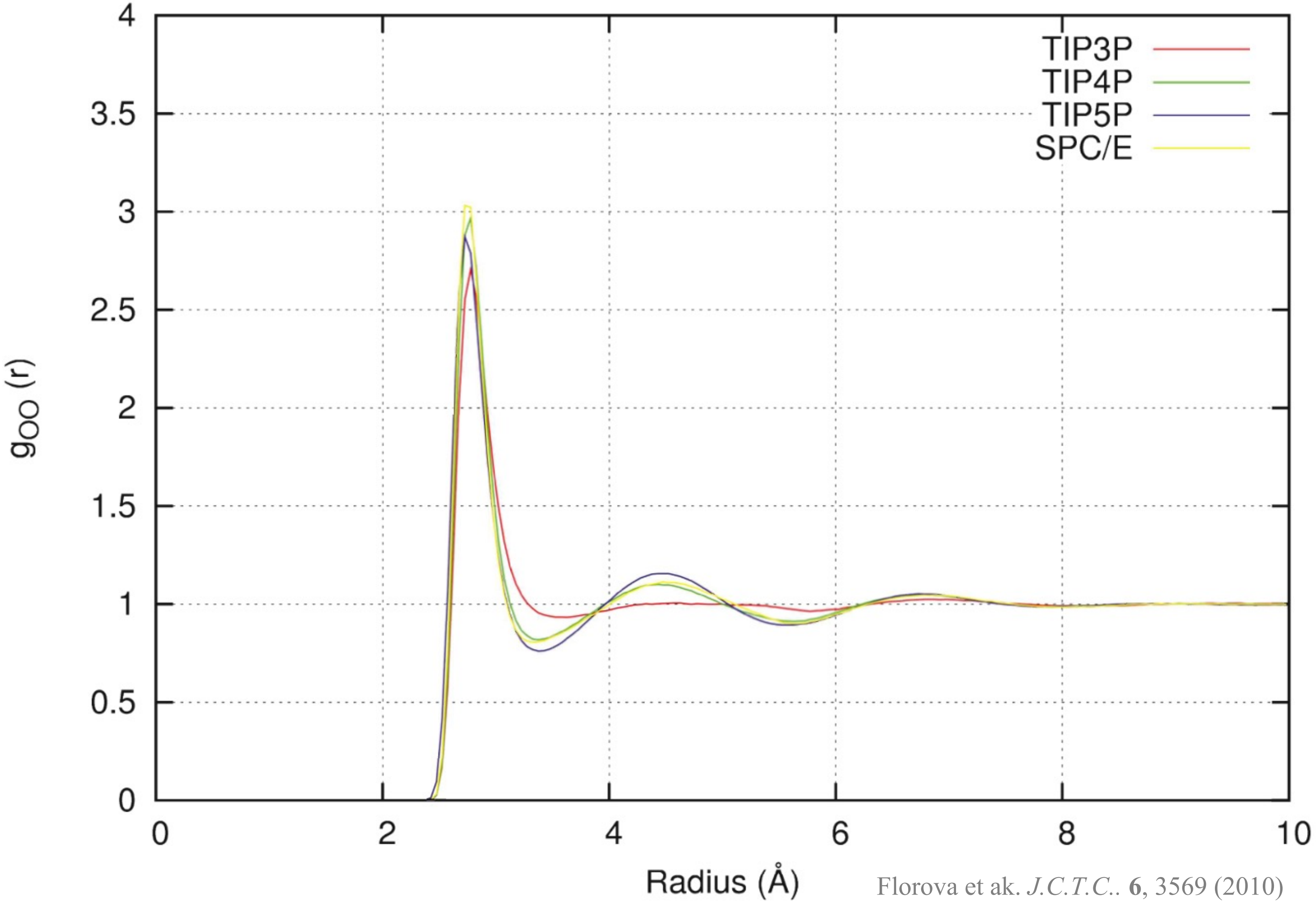


$$V = \sum_{pairs} \left(\underbrace{\frac{A_{LJ}}{r^{12}_{O-O}} - \frac{B_{LJ}}{r^{6}_{O-O}}}_{\text{LJ for O-O}} + \underbrace{k\frac{q_i q_j}{r_{ij}}}_{\text{Coulomb}} \right)$$

r_{O-O}	oxygen-oxygen distances
A_{LJ}, B_{LJ}	Lennard-Jones parameters
r_{ij}	distance between charged sites
k	constant in Coulomb's law

Radial distribution function g(r)



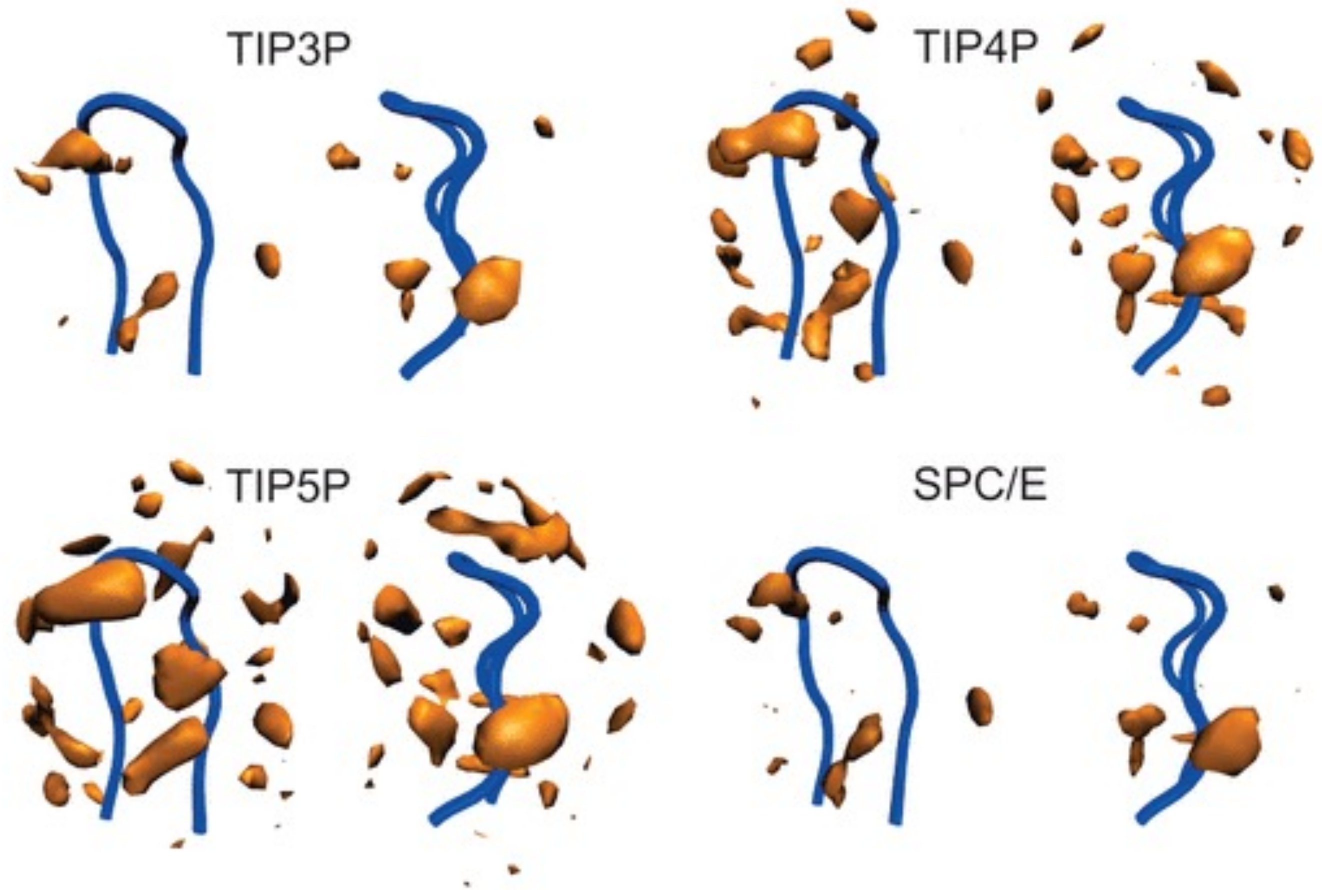


Florova et ak. *J.C.T.C.*. **6**, 3569 (2010)

Small differences in geometry, charges but huge differences in water structure and dynamics

Mark & Nilsson, *J.Phys.Chem. A*, **43**, 9954 (2001)

Example



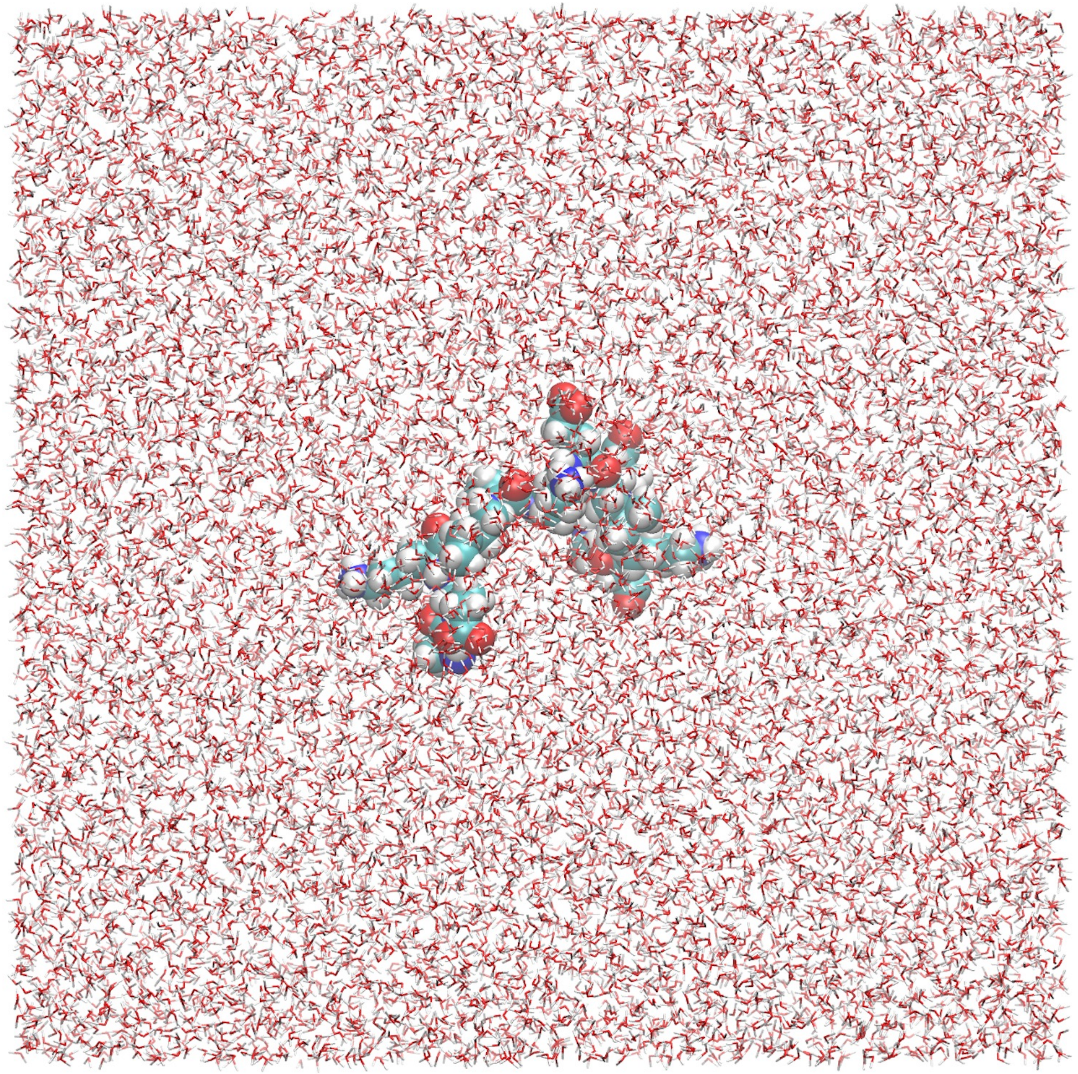
Water density maps

Average number of water molecules in the first two solvation shells

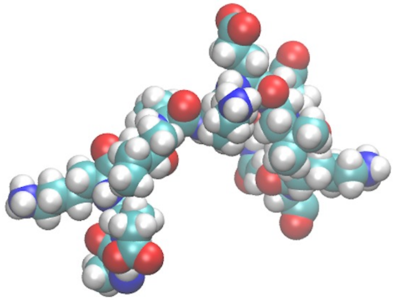
	TIP3P	TIP4P	TIP5P	SPC/E
first shell, <3.4 Å	93 ± 8	93 ± 6	100 ± 8	94 ± 7
second shell, <5.0 Å	185 ± 14	182 ± 12	196 ± 15	186 ± 13
long-residency water molecules	0.0 ± 0.0	2.0 ± 1.3	12.8 ± 4.5	3.7 ± 2.7

Florova et ak. *J.C.T.C.*. **6**, 3569 (2010)

BUT …



Implicit water as dielectric continuum can be so much faster



Implicit solvent models

- Represent solvent and counterions as a continuum → solvent degrees of freedom are taken into account implicitly
- Solvent = high dielectric medium & protein = low dielectric region & spatial charge distribution
- No need of water equilibration
- Much faster than explicit solvent → lower computational cost
- Examples
  - Poisson-Boltzmann model
  - Generalized Born model
  - ABSINTH model (explicit ions)

Implicit solvent models

- seek to approximate is the solute potential of mean force, which determines the statistical weight of solute conformations, and which is obtained by averaging over the solvent degrees of freedom.
- the total free energy = the reversible work performed in two successive steps
  - the particle is inserted in the solvent with zero atomic partial charges
  - the atomic partial charges of the particle are switched from zero to their full values
- **the total solvation free energy corresponds to a sum of non-polar and electrostatic contributions**

$$\Delta G_{\mathrm{solv}} = \Delta G_{\mathrm{polar}} + \Delta G_{\mathrm{nonpolar}}$$

Roux & Simonson *Biophys. Chem.* **78**, 1 (1999)

Implicit solvent models

$$\Delta G_{solv} = \Delta G_{polar} + \Delta G_{nonpolar}$$

$\Delta G_{polar} \downarrow$

Electrostatics
Poisson-Boltzmann
Born

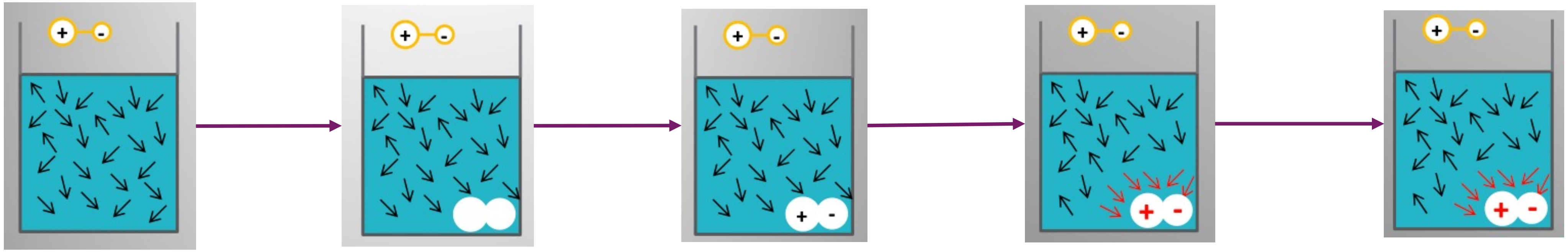
$\Delta G_{nonpolar} \downarrow$

$\Delta G_{cav} + \Delta G_{vdW}$
SASA

ΔG_{cav} – cavity creation within the solvent
ΔG_{vdW} – embedding of the particle into the cavity

Solute shape cavity of vacuum is introduced into the solvent

Solvent molecules reorient and polarize in response to the solute charge density

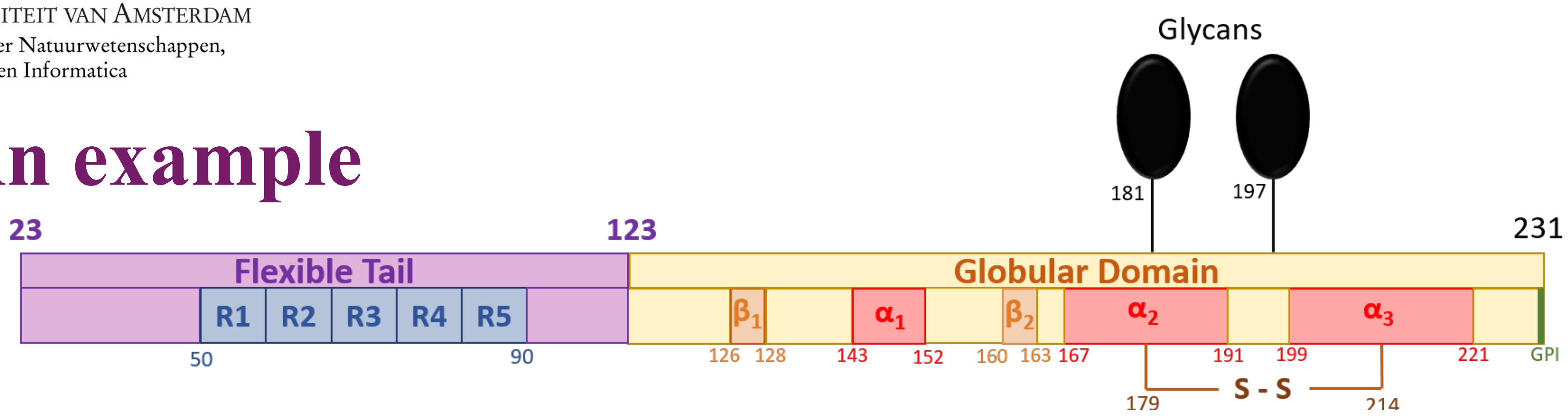


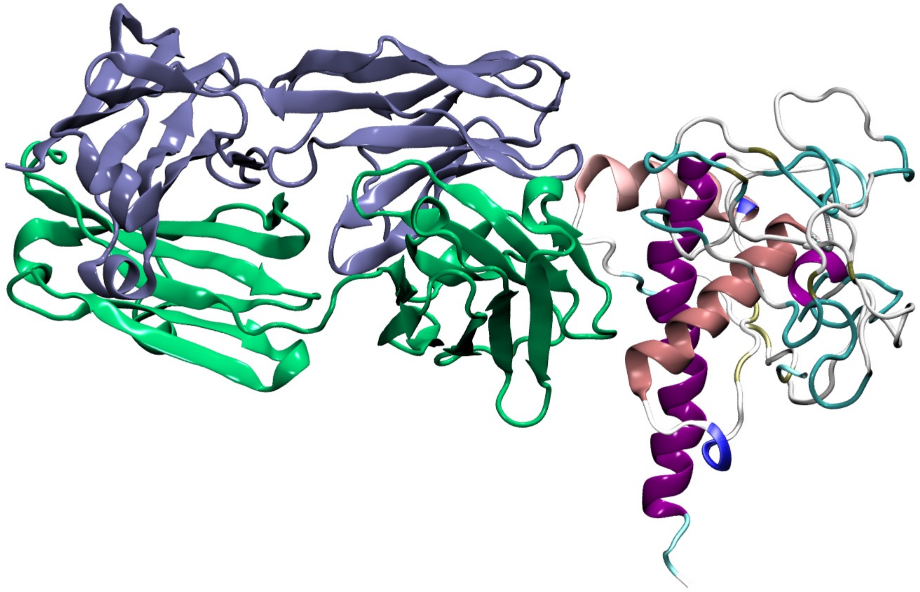
Solute charge density is placed in the solute cavity

Solute polarizes in response to solvent polarization

https://events.prace-ri.eu/event/455/attachments/369/551/solvation_models.pdf

An example





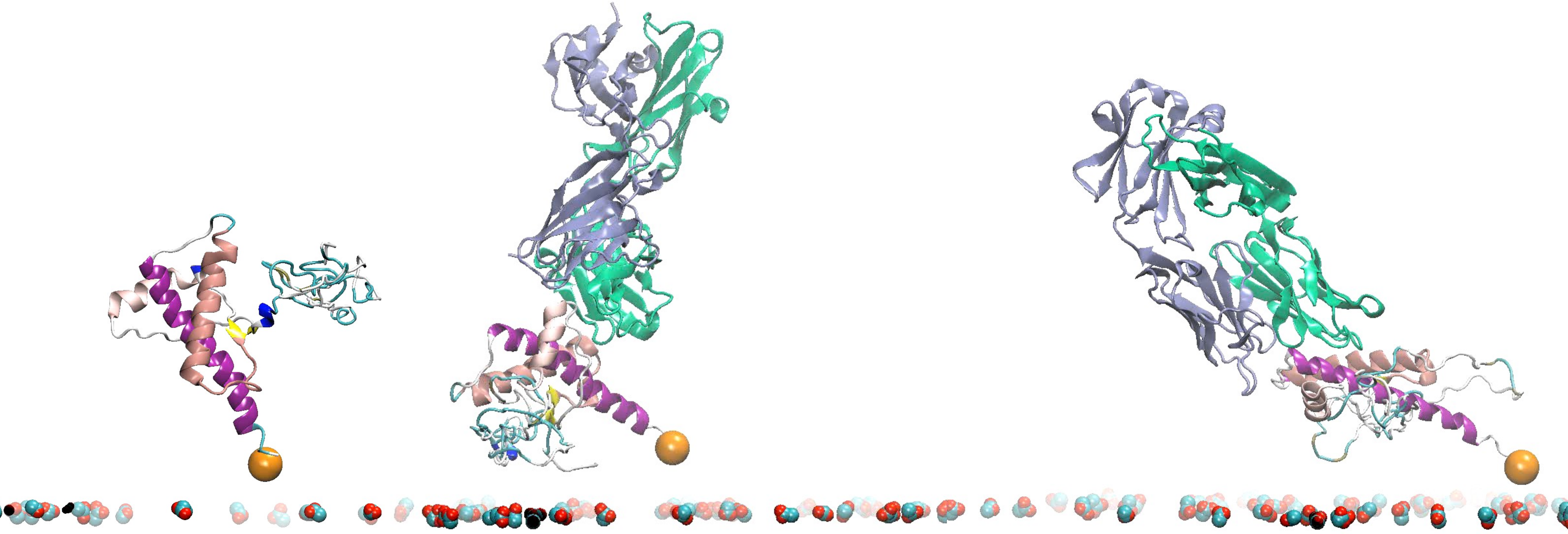
In **explicit solvent** 1.6M atoms
$\Delta t = 2$ fs

In **implicit solvent**
$\Delta t = 5$ fs

Ilie et al. *Biophys. J.* (2022)

Membrane-interaction mechanisms

PrP^C **POM1** **POM6**



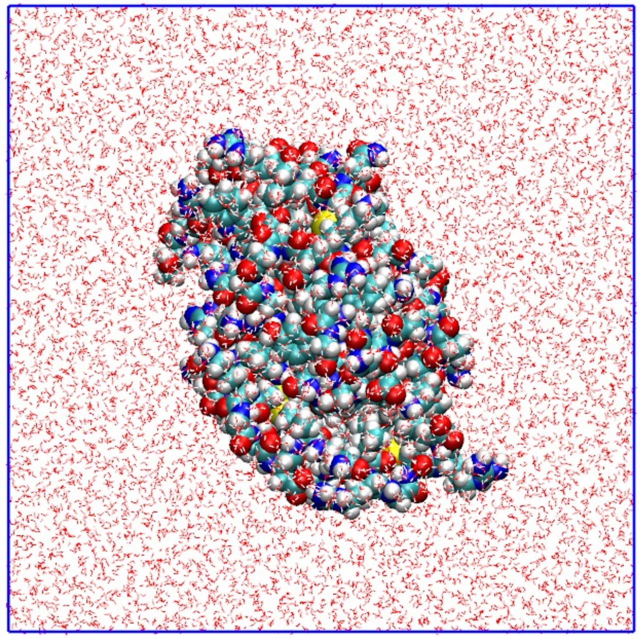
ILIE et al., *Biophys. J.* ***121****, 2813 (2022)*

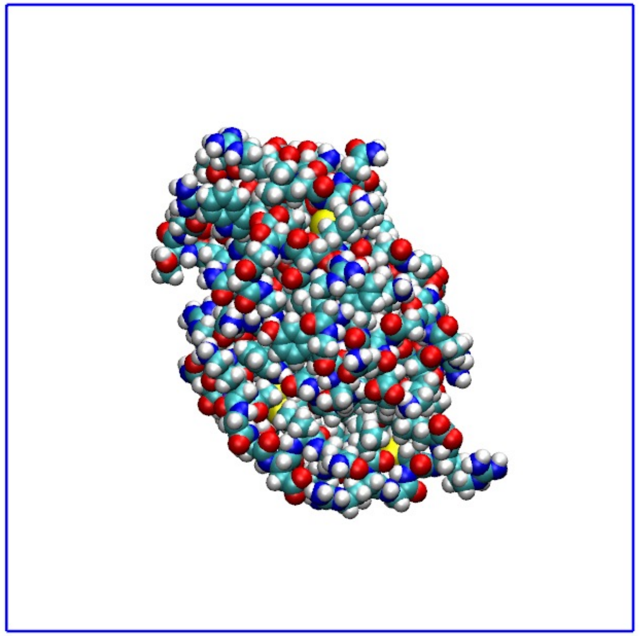
I

1.5

Sp

Cows, Langevin & Brown - Can we do more?





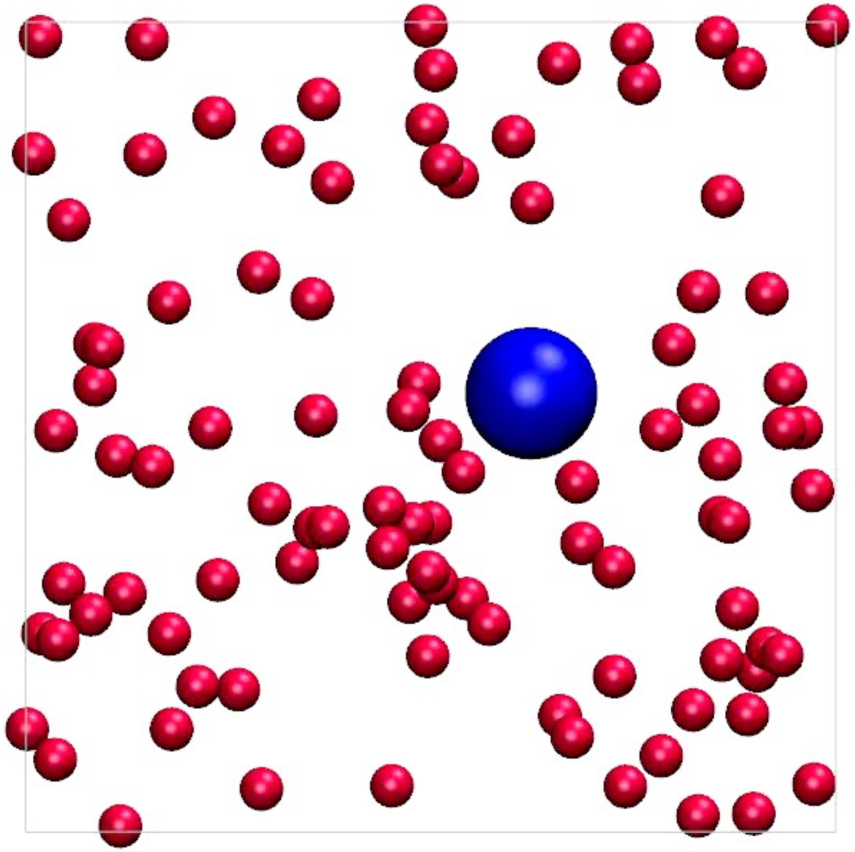
Solvent molecules ➔ stochastic force

< collisions solvent > ➔ friction, ξ

fluctuations around ξ ➔ stochastic Markovian process

(thermal noise of solvent molecules)

Langevin & Brown



$$\boldsymbol{v} = \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{dt}}$$

$$m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{dt}} = -\boldsymbol{F}_D + \boldsymbol{F}_B$$

From hydrodynamics $\boldsymbol{F}_D = 6\pi\eta\boldsymbol{v}r$

$$m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{dt}} = -6\pi\eta\boldsymbol{v}r + \boldsymbol{F}_B$$

Friction force Random force

$$\xi = 6\pi\eta r/m$$

Langevin dynamics

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -6\pi\eta\mathbf{v}r + \mathbf{F}_B$$

$$\mathbf{v}(\mathrm{t}) = \mathbf{v}_0 \mathrm{e}^{-\xi t/m} + \int_0^t \mathrm{d}\tau\, \mathrm{e}^{-\xi(t-\tau)/m}\frac{1}{m}\mathbf{F}_B(\tau)$$

$$\mathbf{v}^2(t \to \infty) \approx 0$$ **false**

At **EQUILIBRIUM**: $\langle v^2 \rangle_{eq} = \frac{kT}{m}$

$$\left\langle \mathbf{F}_B(t) \right\rangle = 0$$ Noise produces no net force

$$\left\langle \mathbf{F}_B(t) \cdot \mathbf{F}_B(t') \right\rangle = 2k_B T\xi\delta(t-t')$$

Fluctuation-dissipation theorem

Balance between “dead” and ”alive”

δ-Dirac function

No correlation between t and t’
due to time scale separation
(short lived collisions)

Langevin dynamics

$$\boxed{m\ddot{\mathbf{r}} = -\xi\dot{\mathbf{r}} + \mathbf{F}(\mathbf{r}) + \sigma\mathbf{R}}$$

$$\sigma = 2k_B T\xi$$

Reduced degrees of freedom

Focus on the important contributions

+ Keeps the contributions to the dynamics of the system

→ Long chain molecules

→ Barrier crossing motions

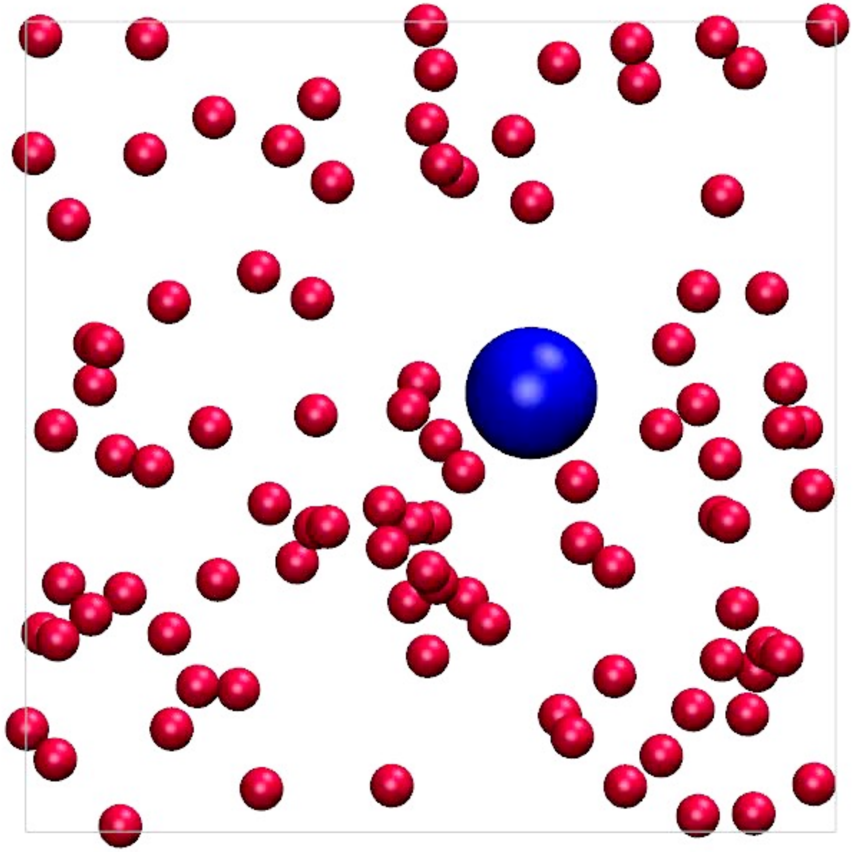
Brownian dynamics

= overdamped Langevin (**a**=0)
- No inertia (high friction)

$$\left.\begin{array}{l} m\ddot{\mathbf{r}} = -\xi\dot{\mathbf{r}} + \mathbf{F}(\mathbf{r}) + \sigma\mathbf{R} \qquad \sigma = 2k_BT\xi \\[2ex] < m\ddot{\mathbf{r}} >_{\Delta t} \cong 0 \end{array}\right\} \quad \boxed{\Delta\mathbf{r} = \frac{\mathbf{F}}{\xi} + \sqrt{\frac{2k_BT\Delta t}{\xi}}\mathbf{R}}$$

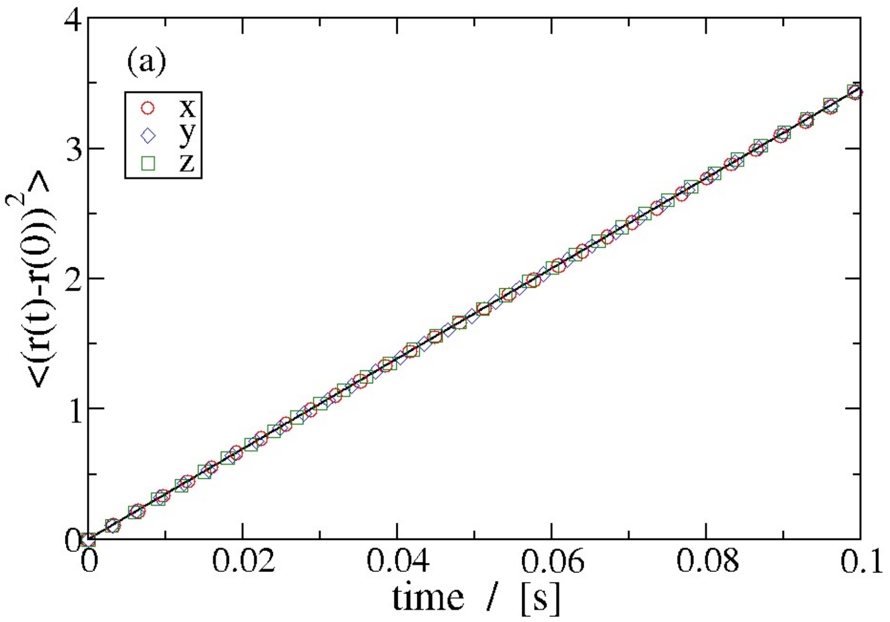
Ermak & McCammon, *J.Chem.Phys.*, **69**, 1352 (1978)

In practice



$$\mathbf{r}(t+\delta t)-\mathbf{r}(t) = \boldsymbol{\mu}^{t}\,\mathbf{F}\boldsymbol{\delta t} + \sqrt{\boldsymbol{\mu}^{t}}\boldsymbol{\Theta}^{\mathrm{t}}(t)\sqrt{2k_B T\delta t}$$

$$\Theta_{\alpha}^{i}\begin{cases}\left\langle\Theta_{\alpha}^{i}(t)\right\rangle = 0\\ \left\langle\Theta_{\alpha}^{i}(t)\Theta_{\beta}^{j}(t')\right\rangle = \delta_{\alpha\beta}\delta_{ij}\delta_{tt'}\end{cases}$$



$$<(\mathbf{r}(t)-\mathbf{r}(0))^2> = 6D_t\delta t$$

I.M. Ilie, W.J. Briels & W.K. den Otter, *J. Chem. Phys.*, **142**, 114103 (2015)
I.M. Ilie, W.K. den Otter & W.J. Briels, *JCP*, **144**, 085103 (2016)

In practice

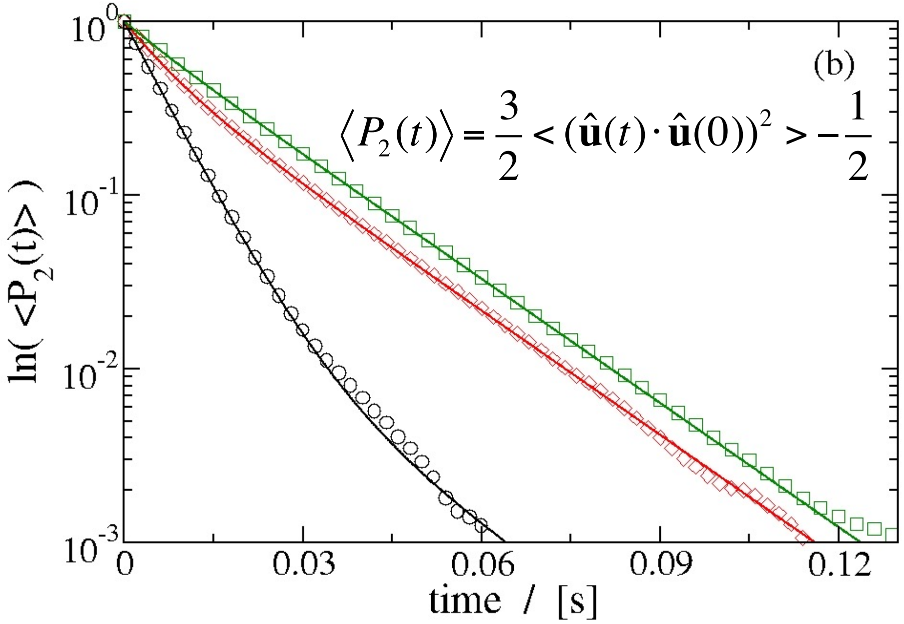
$$\mathbf{q}(t+\delta t)-\mathbf{q}(t)=\mathbf{A}\cdot$$

$$\partial$$

$$\boldsymbol{\mu}^{\mathrm{r}}\boldsymbol{\Theta}^{\mathrm{r}}(t)\sqrt{2k_BT\delta t}$$

$$\downarrow$$

ional friction



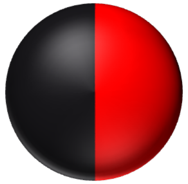
$$\boxed{\mathbf{q}(t+\delta t)-\mathbf{q}(t)=\mathbf{B}\cdot\ \boldsymbol{\mu}^{\mathrm{r}}\cdot\mathbf{A}^{\mathrm{T}}\mathbf{T}\boldsymbol{\delta t}+\mathbf{B}\sqrt{\boldsymbol{\mu}^{\mathrm{r}}}\boldsymbol{\Theta}^{\mathrm{r}}(t)\sqrt{2k_BT\delta t}+\lambda\mathbf{q}}$$

$$q(t+\boldsymbol{\delta}t)=1$$

$$\boldsymbol{\lambda}^2+2\boldsymbol{\lambda}\mathbf{q}(t)\cdot\tilde{\mathbf{q}}(t+dt)+\tilde{\mathbf{q}}^2(t+\boldsymbol{\delta}t)=1$$

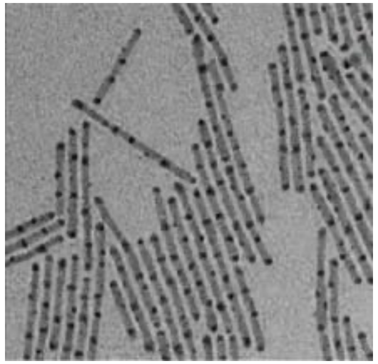
I.M. Ilie, W.J. Briels & W.K. den Otter, *J. Chem. Phys.*, **142**, 114103 (2015)
I.M. Ilie, W.K. den Otter & W.J. Briels, *JCP*, **144**, 085103 (2016)

Intermezzo – the ideal gas and beyond

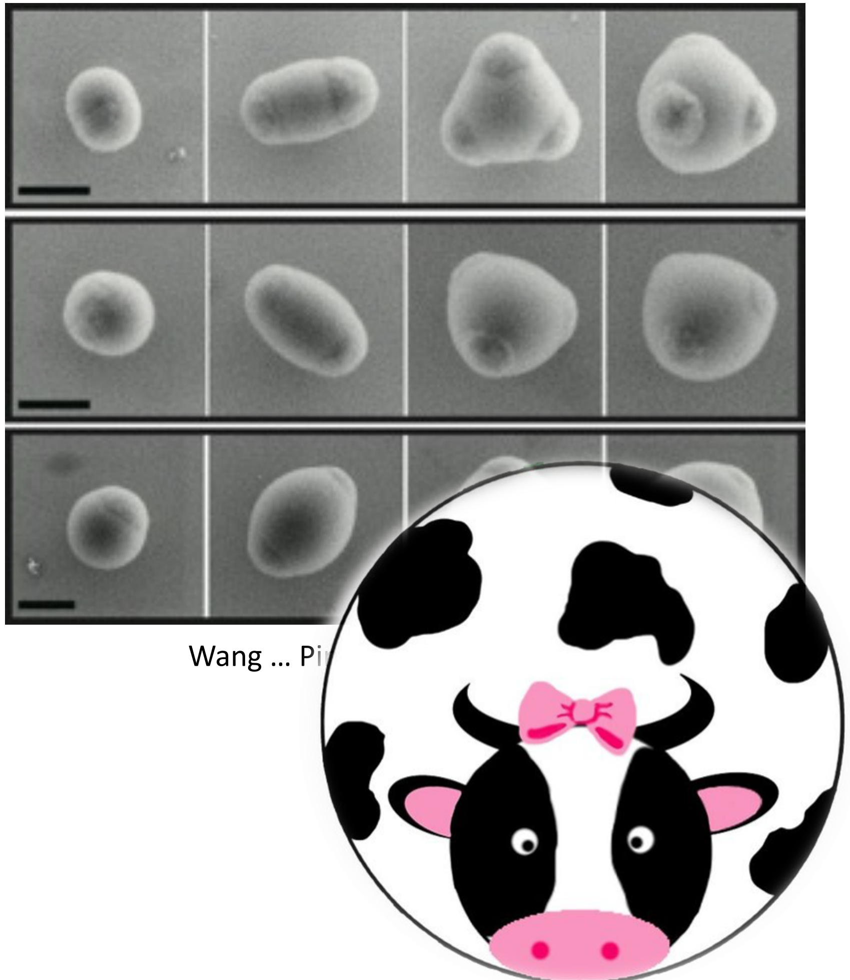


Janus particles

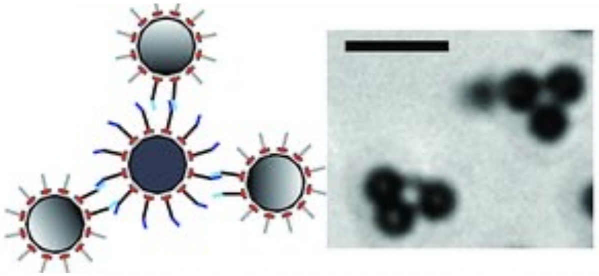
Striped nanospheres



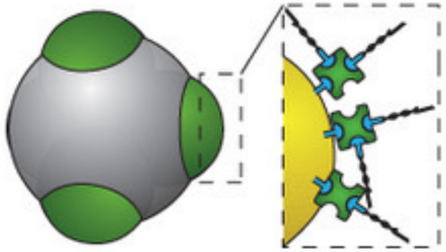
Striped nanorods



Wang … Pi

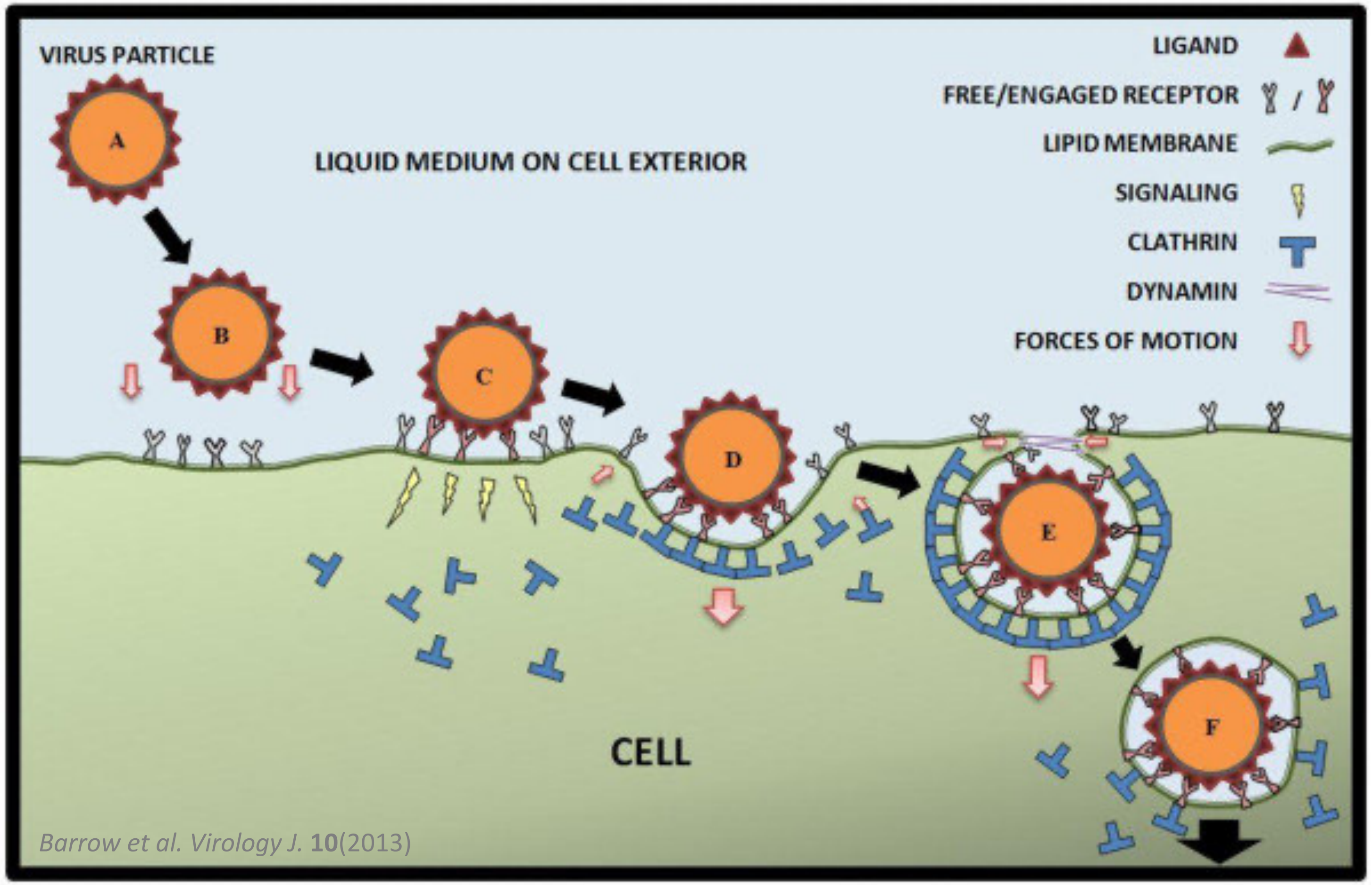


DNA patchy particles

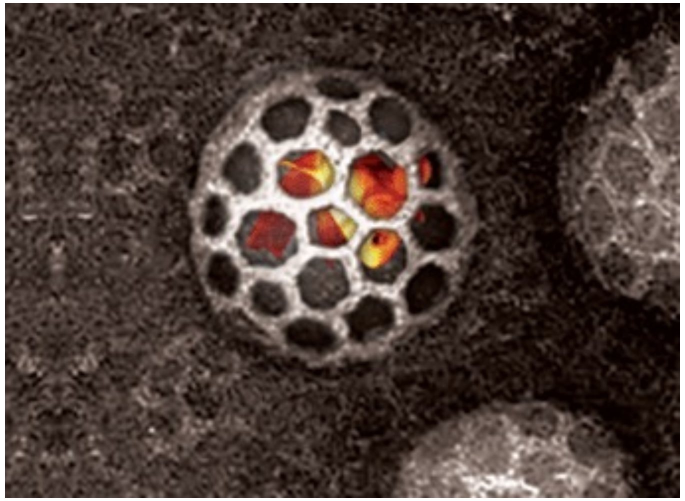


DNA
patches

An example – proteins as patchy particles ☺

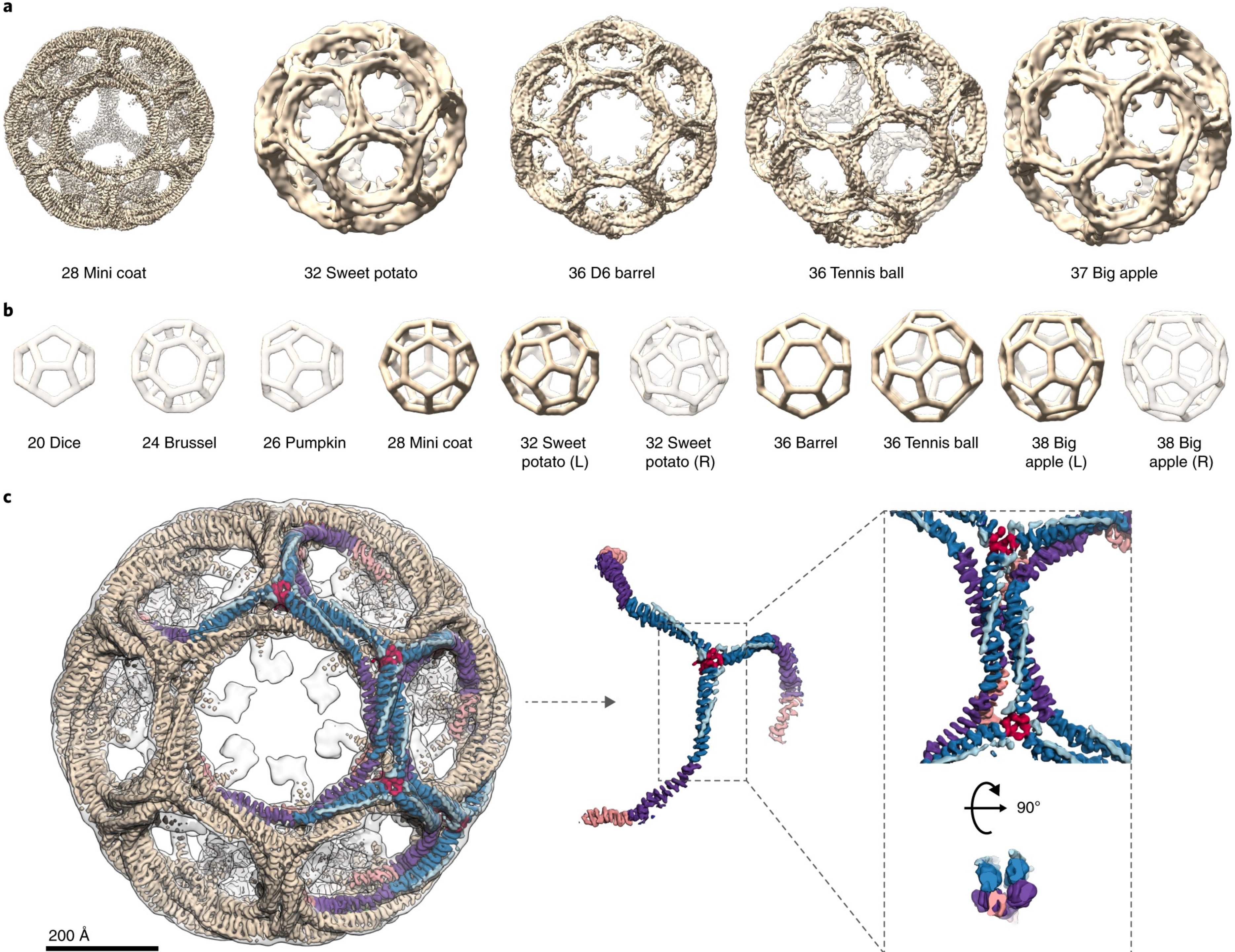


Barrow et al. Virology J. **10**(2013)



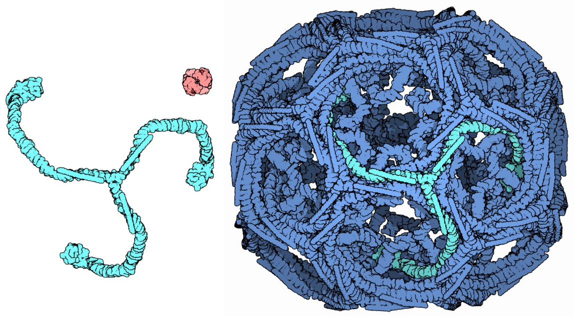
www.endocytosis.org

The model clathrin

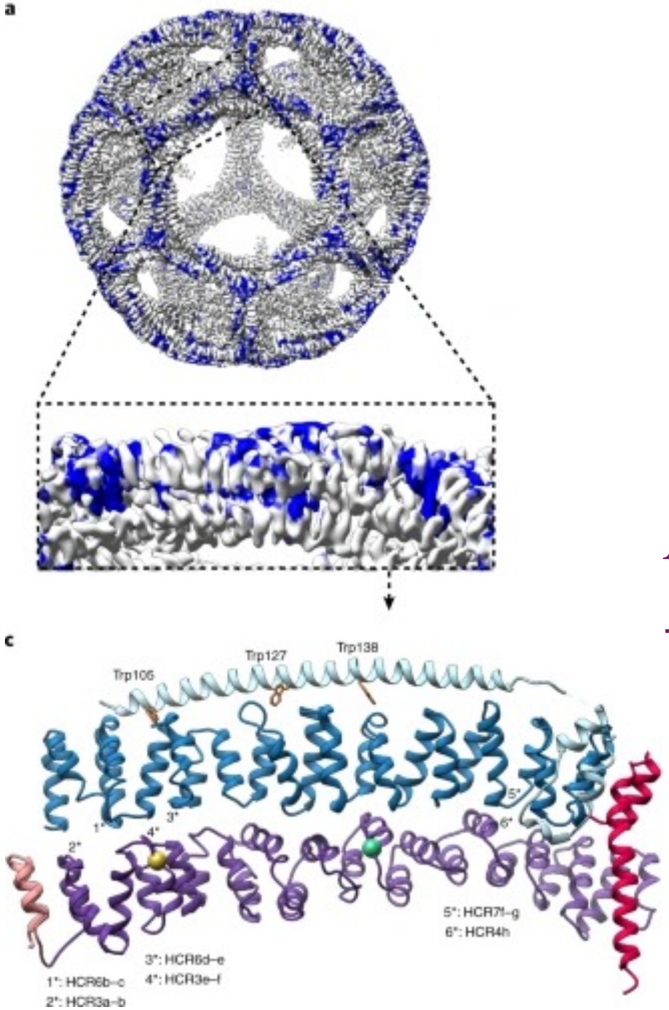


Morris et al. Nat. Struct. Mol. Biol. ***26****, 890 (2019)*

The model



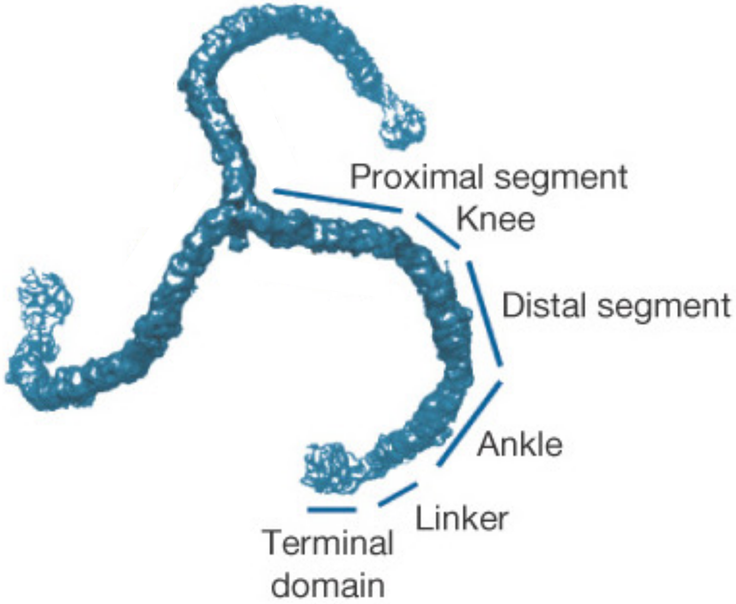
G. T. Johnson & D. Goodsell, RCSB PDB(2007)



5025 AA ➔ 645 kDa

a PIECE!!!!

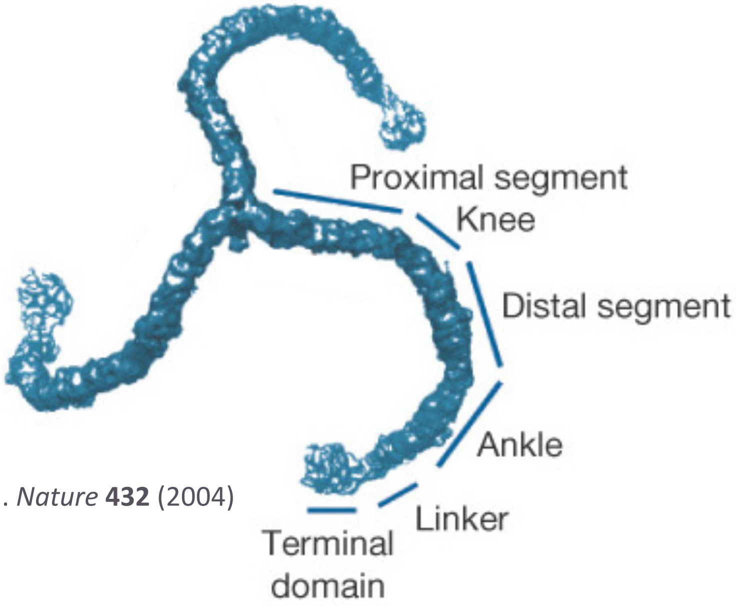
Atomistically impossible
→ coarse graining



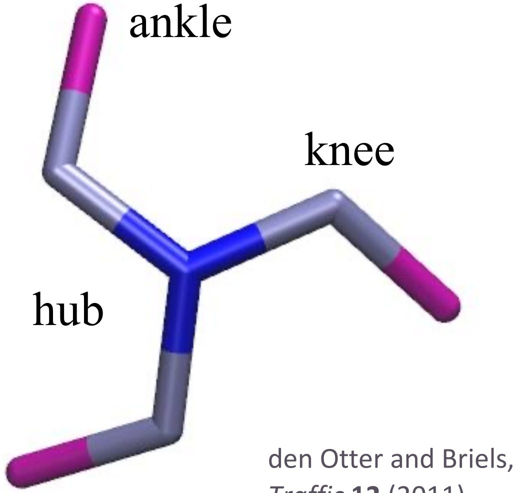
A. Fotin et al. *Nature* **432** (2004)

Morris et al. Nat. Struct. Mol. Biol. ***26****, 890 (2019)*

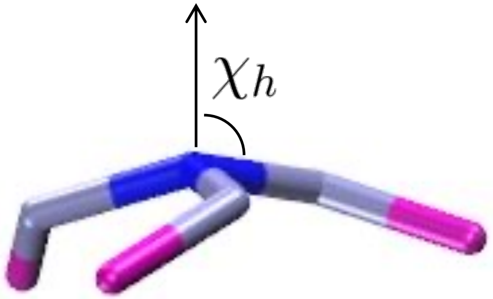
Coarse-grained model: one protein = 1 rigid particle



A. Fotin et al. *Nature* **432** (2004)



den Otter and Briels,
Traffic **12** (2011)



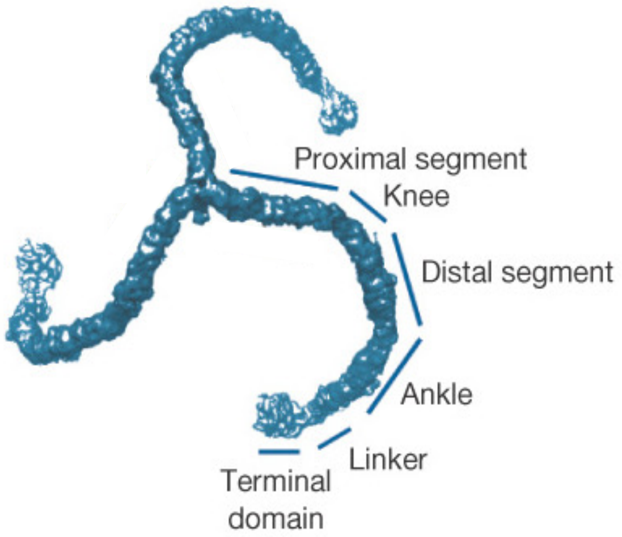
- Slightly curved structure

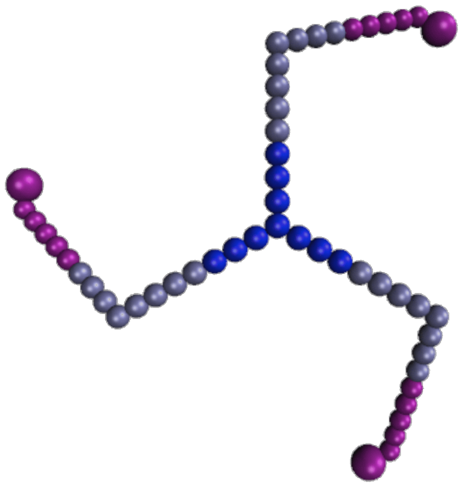
What do we need?

- Friction
- Interaction potential
- A good cluster

$$\mathbf{r}(t+\delta t)-\mathbf{r}(t)=\mathbf{A}\cdot\ \boldsymbol{\mu}^{\boldsymbol{t}}\cdot\mathbf{A}^{\mathrm{T}}\mathbf{F}\boldsymbol{\delta t}+\\ \mathbf{A}\cdot\sqrt{\boldsymbol{\mu}^{\boldsymbol{t}}}\boldsymbol{\Theta}^{\mathrm{t}}(t)\sqrt{2k_B T\delta t}$$

Diffusion





HYDRO++[*]	
Input	
M.W.	645 kDa
η	0.01 Poise
χ	101°
T	20 °C
Output	
D^t	$1.22 \bullet 10^{-7}$ $\mathrm{cm^2/s}$ [**]
D^r	$1.43 \bullet 10^{4}$ $\mathrm{s^{-1}}$

Diffusion matrices $D = \frac{k_B T}{\xi}$

---

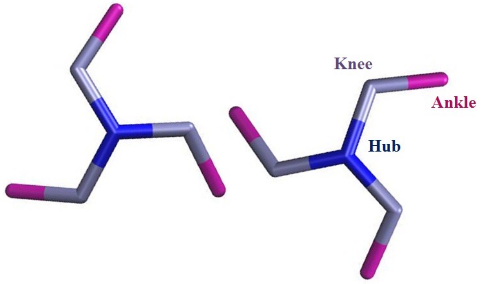
[*] J. García de la Torre, D. Amorós, A. Ortega, *Eur. Biophys. J.* **39**, 381-388 (2010).

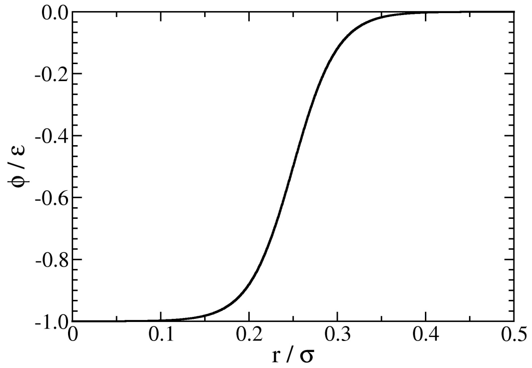
[**] T. Yoshimura, K. Kameyana, S. Maezwa and T.Takagio, *Biochemistry* **20**, 4528 (1991).

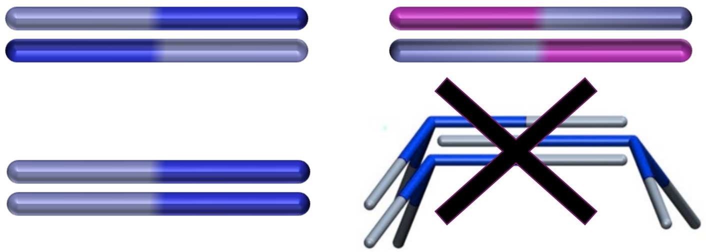
The potential

- interaction between leg segments

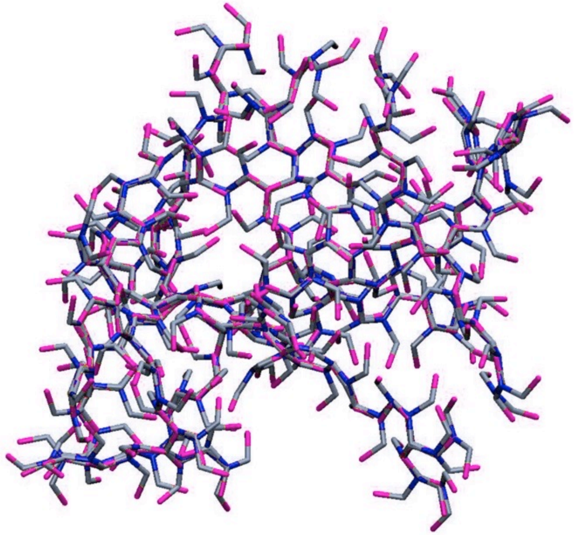
$$\Phi = -\varepsilon \cdot f(r)$$







- Attraction between anti-parallel legs
- Repulsion between parallel legs

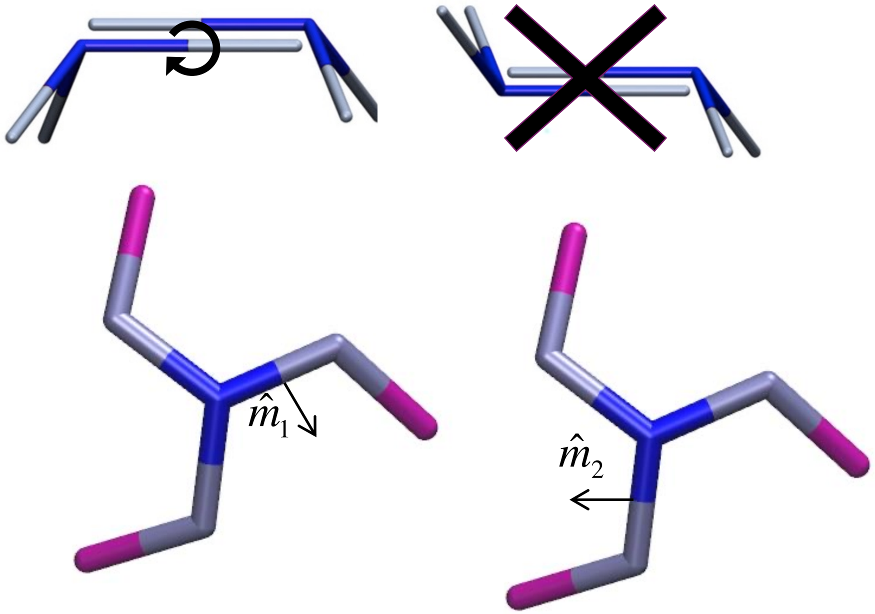


den Otter et al, *Biophys. J.* **99** (2010)

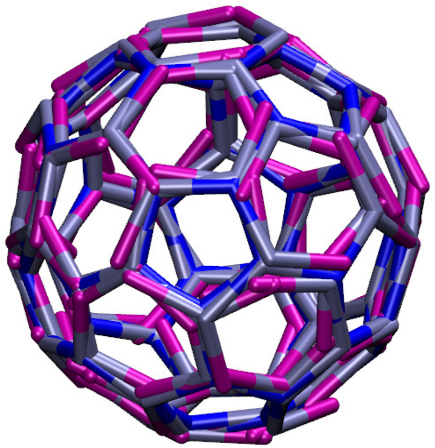
The potential

- interaction between leg segments
- asymmetric binding

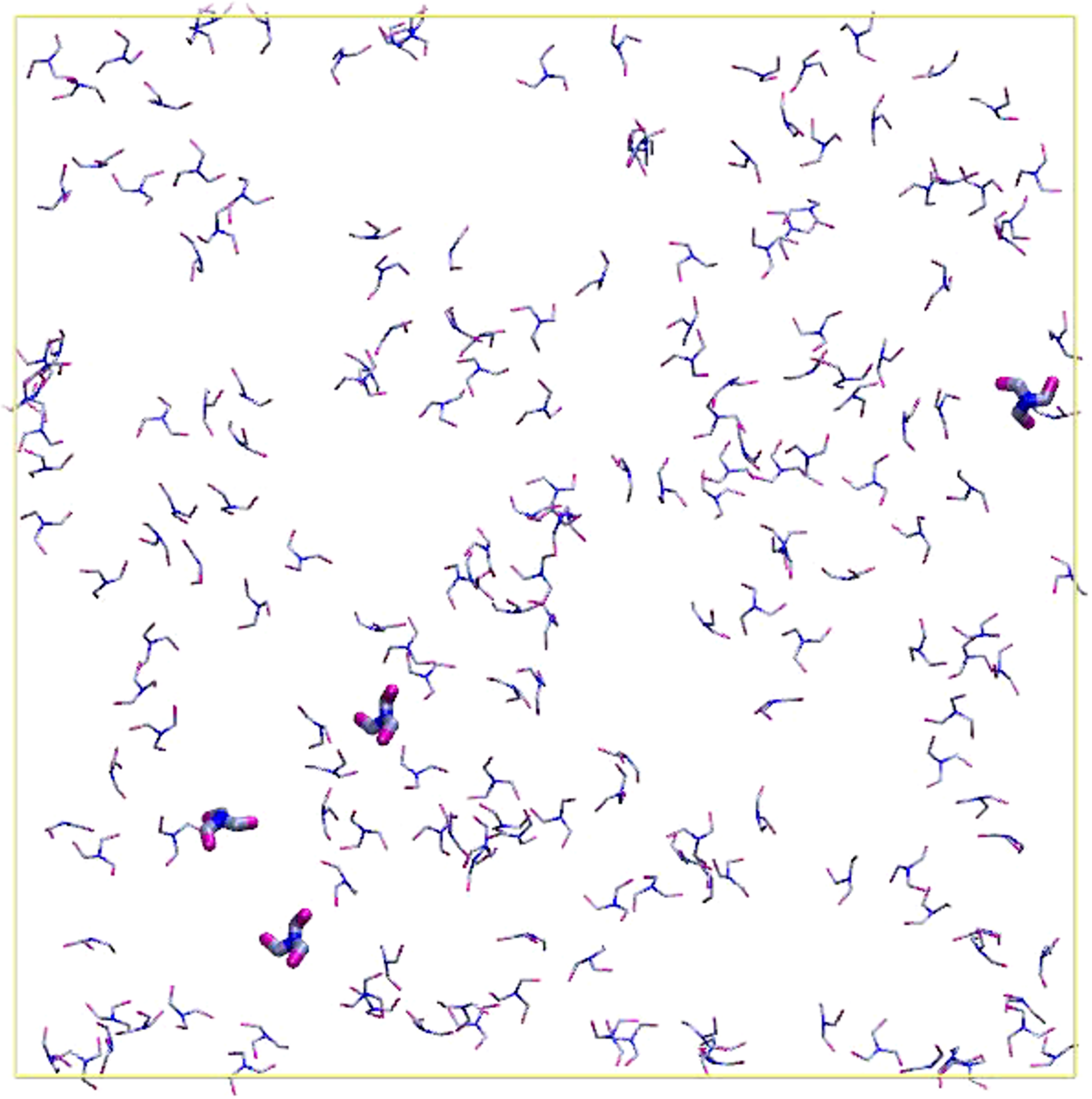
$$\Phi = -\varepsilon \cdot f(r)$$



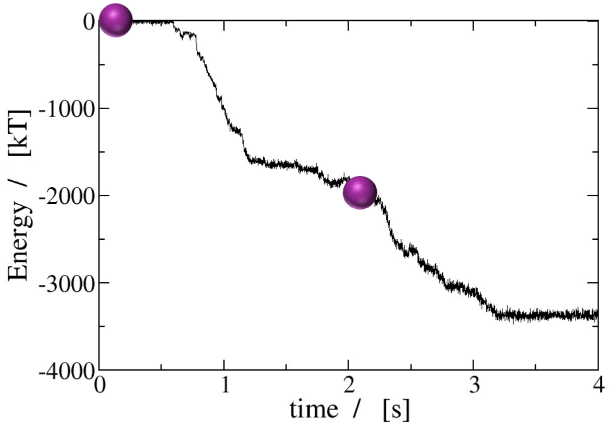
- Rotational asymmetry

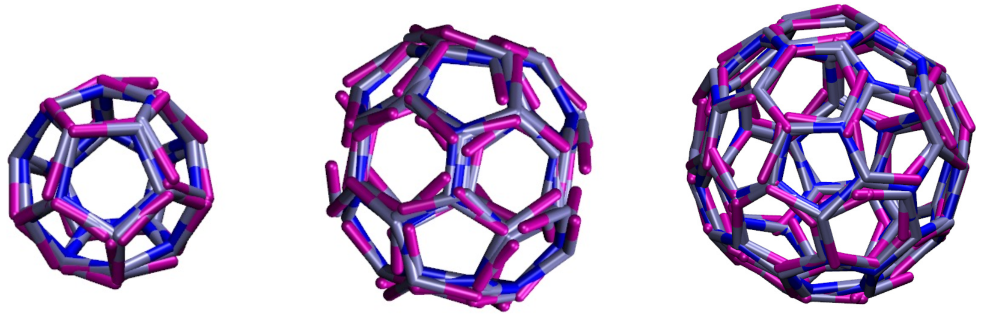


Ilie et al, *J. Chem. Phys.*, **141**, 065101 (2014)



N particles	200
c	100 μg/ml
T	Temp = 20 ˚C
δt	10 ns
timescale	4 s
box length	1 μm





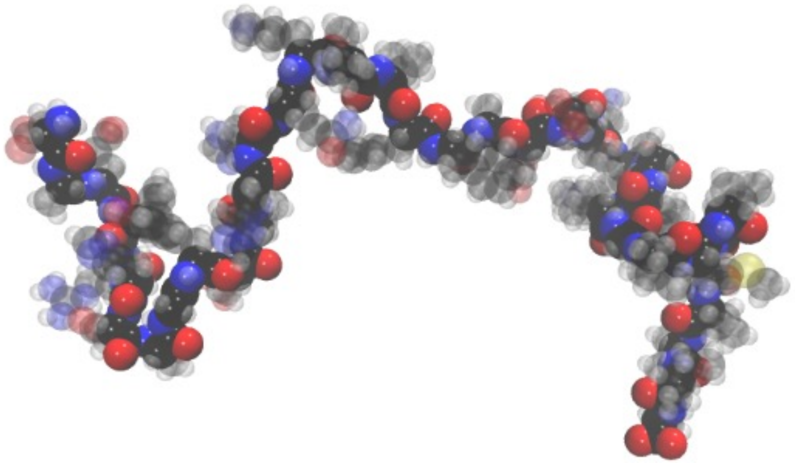
Ilie et al., *J. Chem. Phys.*, **141**, 065101 (2014)

Another example – a sofisticated protein

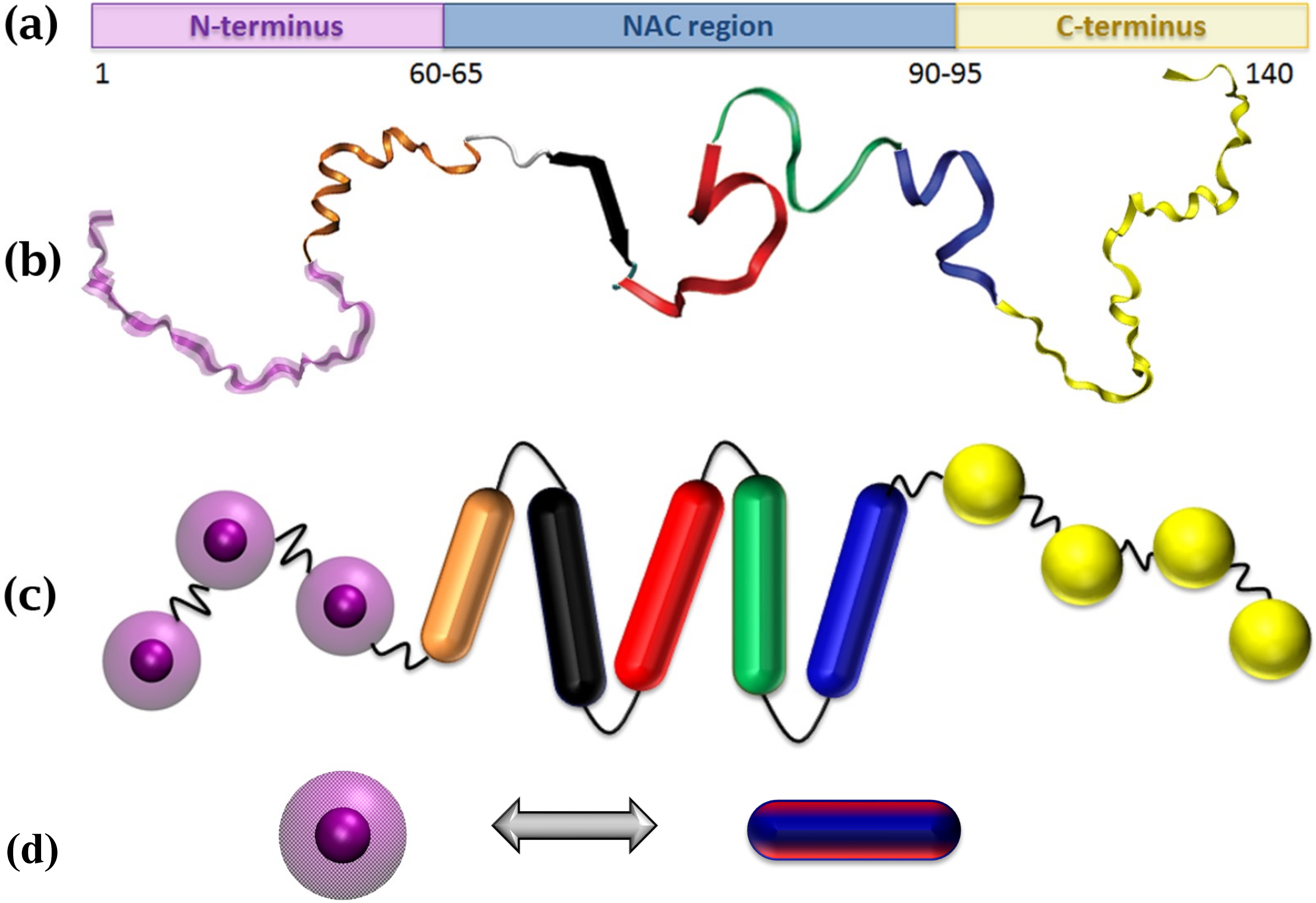
IDP = intrinsically disordered = little or no structure in its native state

1 60-65 90-95 140

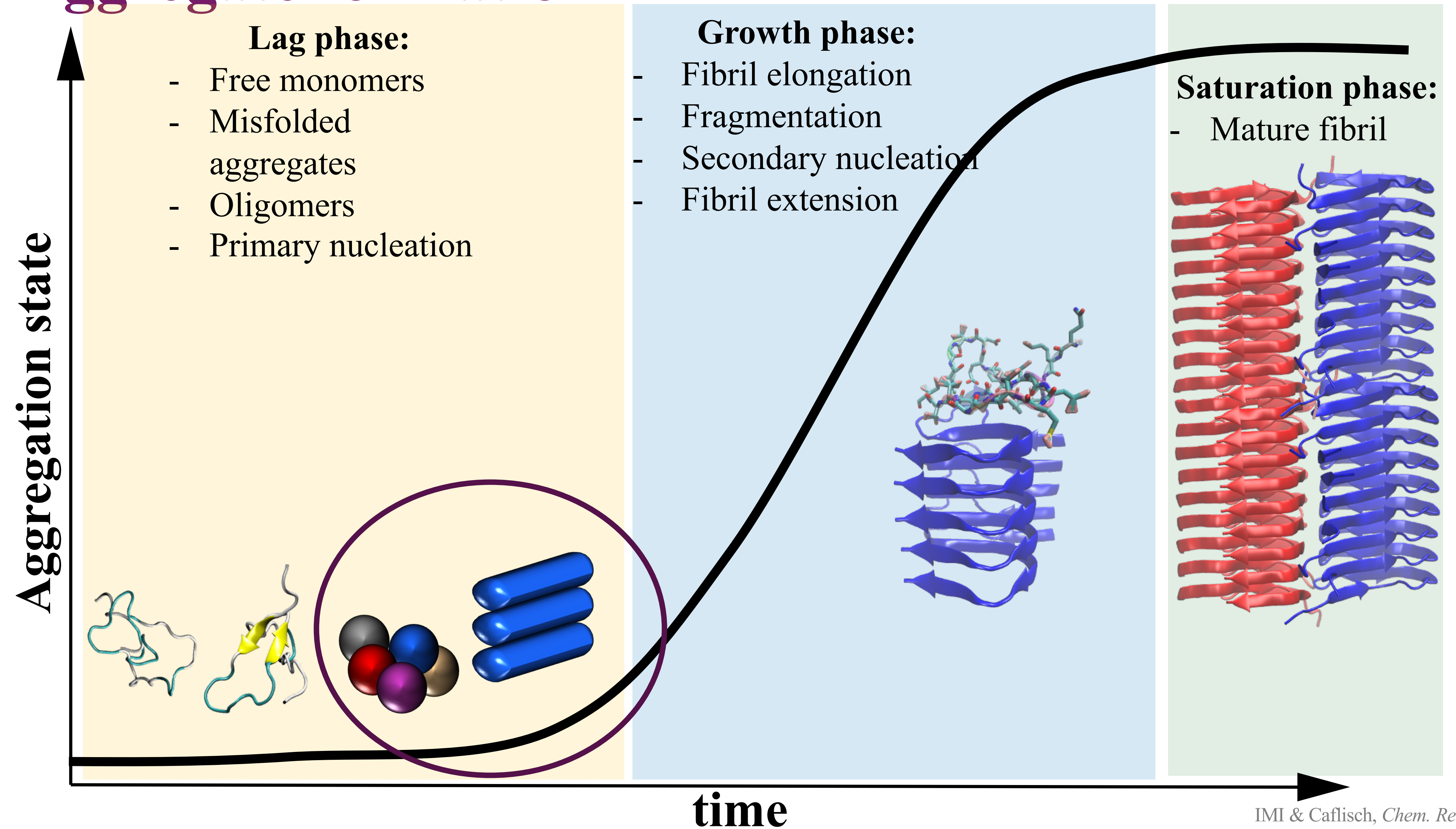
N-terminus	NAC region	C-terminus
- amphipathic (both hydrophilic and lipophilic) - α-helix, disordered	- hydrophobic - building block for α-synuclein aggregates - α-helix, β-sheets, disordered	- highly acidic - negatively charged - disordered



How to CG? What is my research question?



Aggregate formation



Lag phase:

- Free monomers
- Misfolded aggregates
- Oligomers
- Primary nucleation

Growth phase:

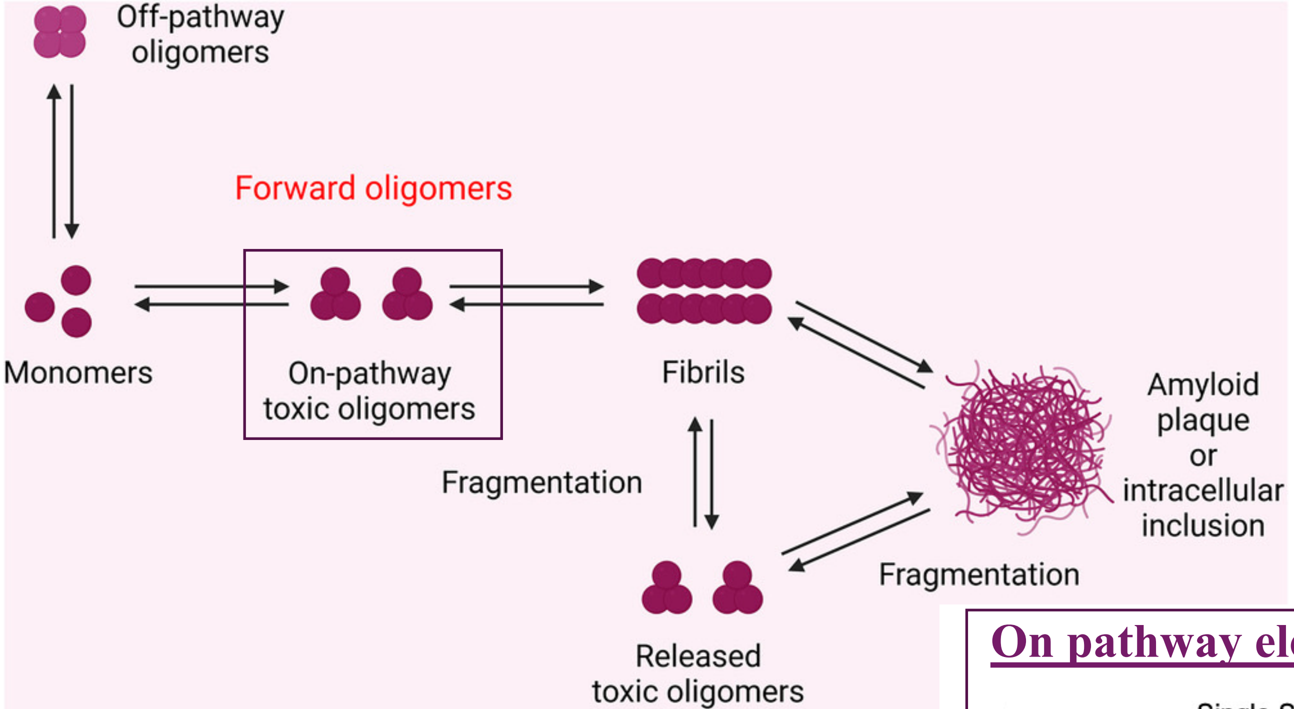
- Fibril elongation
- Fragmentation
- Secondary nucleation
- Fibril extension

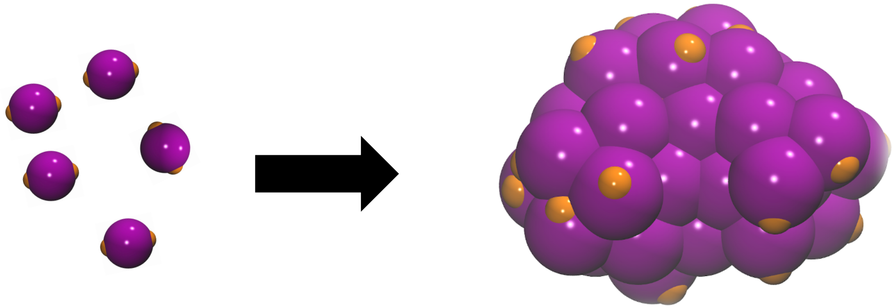
Saturation phase:

- Mature fibril

IMI & Caflisch, *Chem. Rev.*, 2019, **119**, 6956-6993

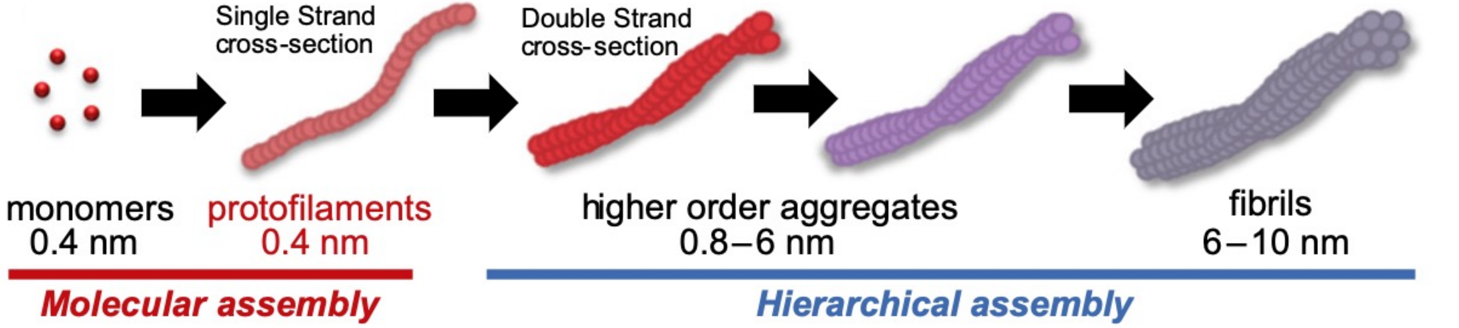
Aggregates = Oligomers





On pathway elongated oligomers = protofibrils

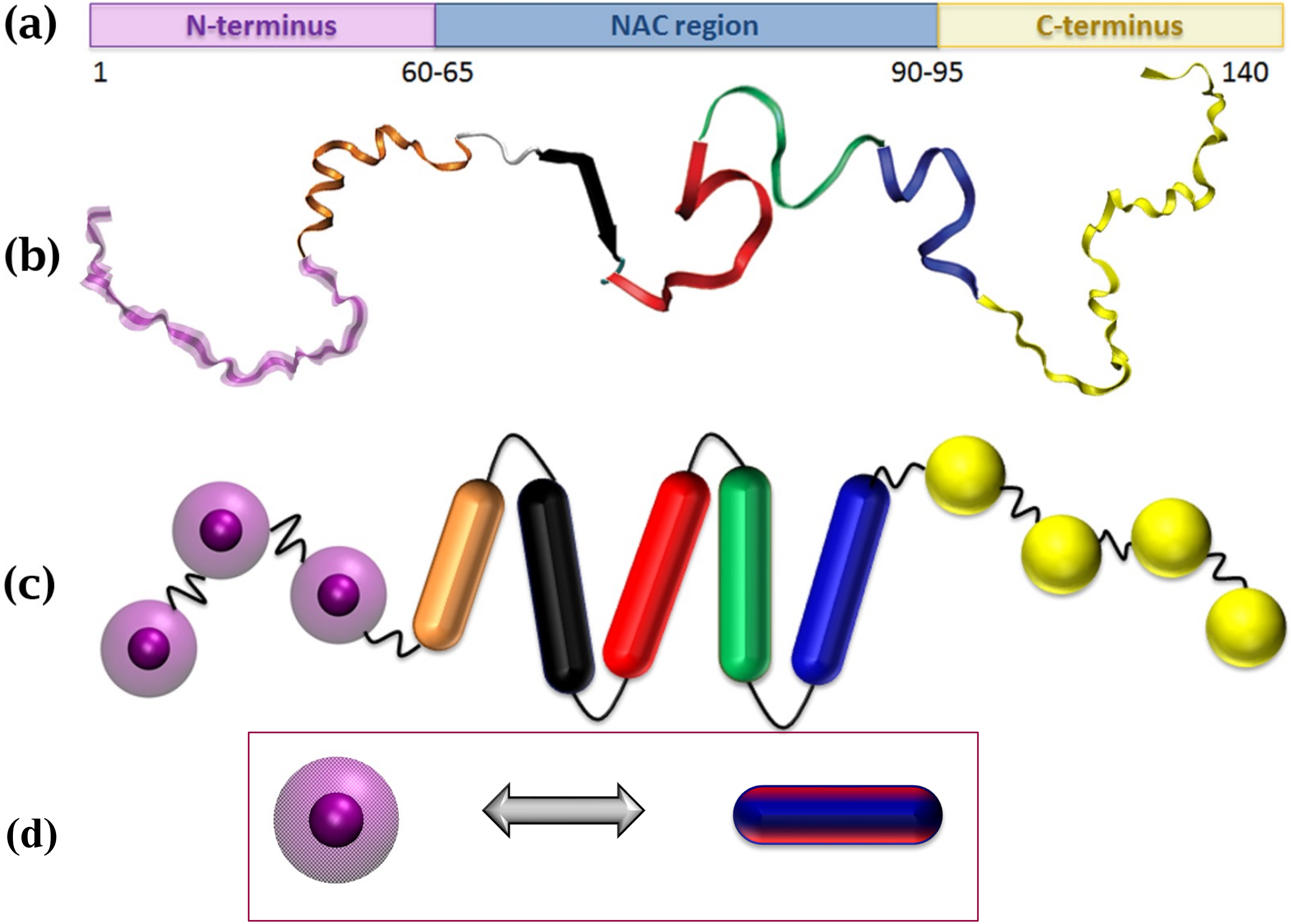
Ruggeri et al. *PNAS*, 2018, **115**, 7231



What are the aggregation mechanisms?

BIGI et al., *Bioessays.*, 2022, **44**, e2200086

What is my research question? Aggregation



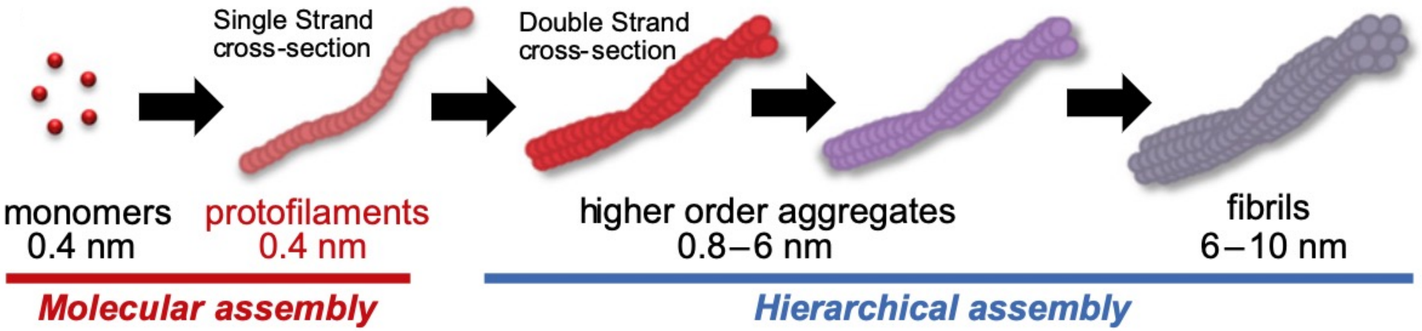
Aggregates = oligomers

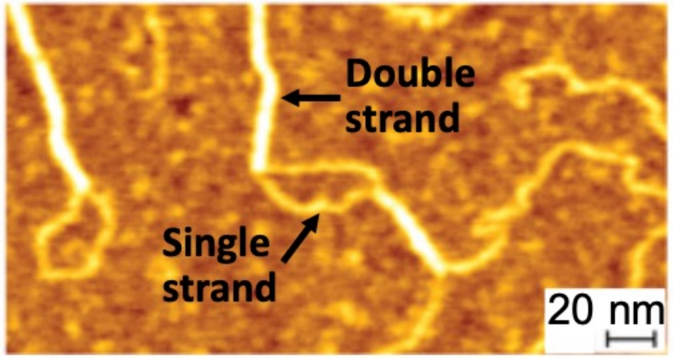
Dr. Simone Ruggeri

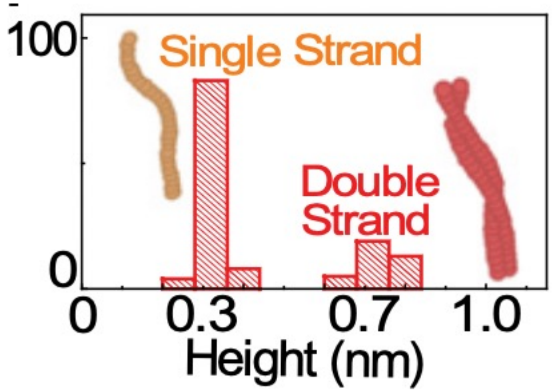
Bernadette Mayer

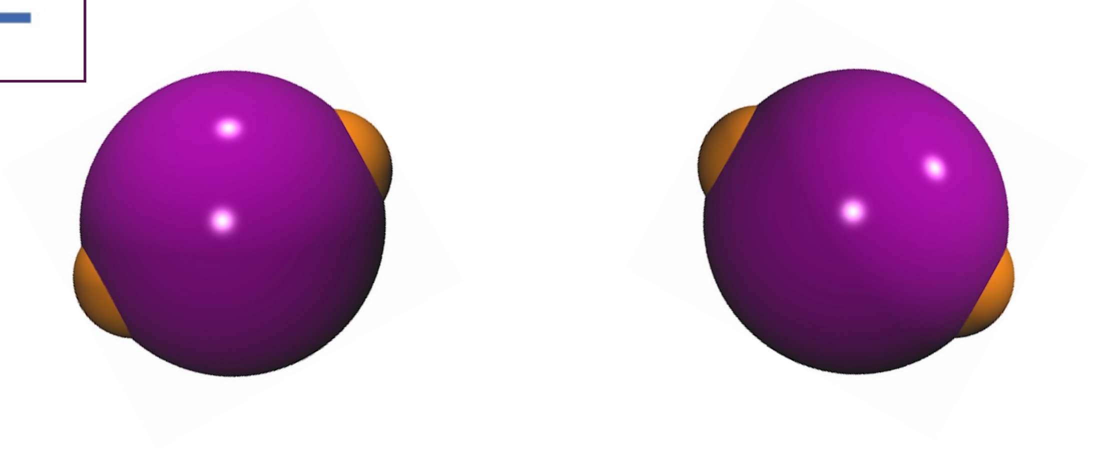
On pathway oligomers

Ruggeri et al. *PNAS*, 2018, **115**, 7231









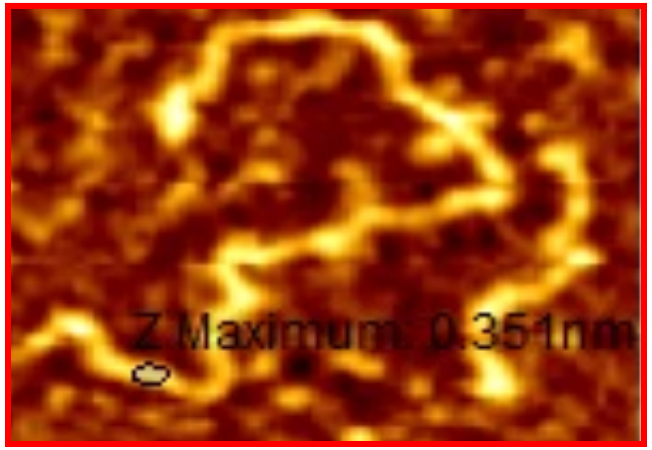
$$\Phi = \sum_{i,j} \Phi_{ex.v}(r_{ij}) + \sum_{i,j} \Phi_{attr}(r_{ij}) + \sum_{i,j} \Phi_{attr}(d_{ij}, \hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j)$$

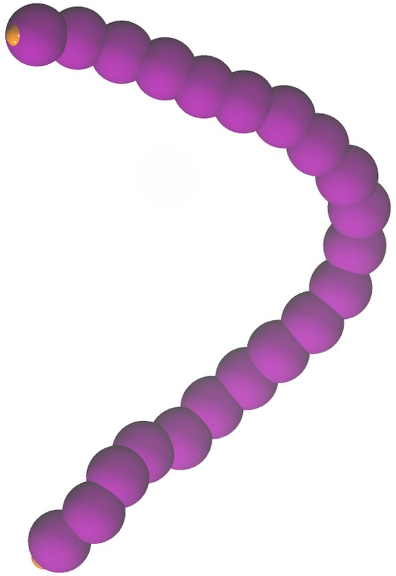
Manuscript in preparation

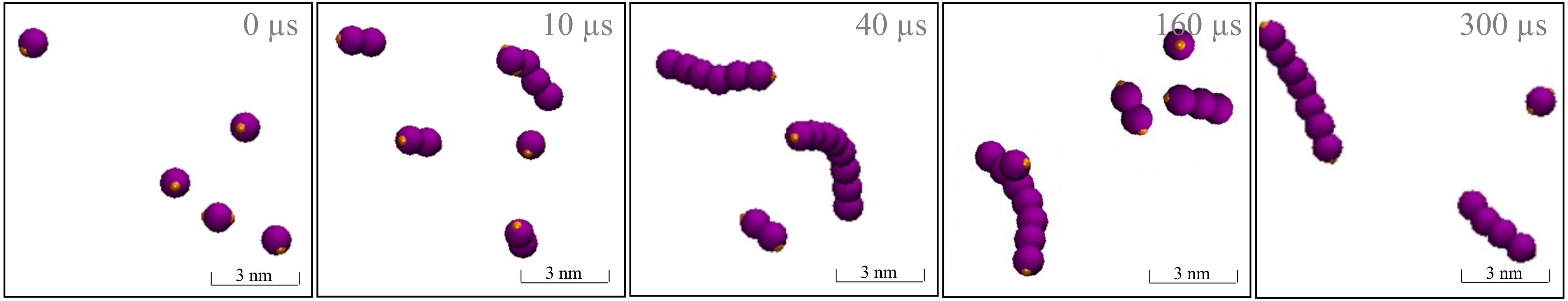
Dr. Simone Ruggeri

Bernadette Mayer

Single-strand protofilaments

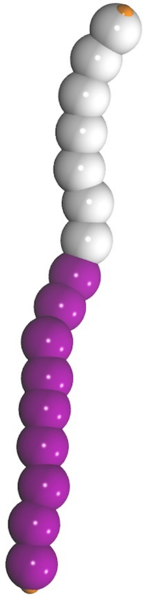




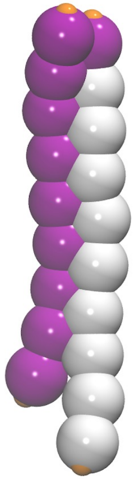


Single-strand mechanisms

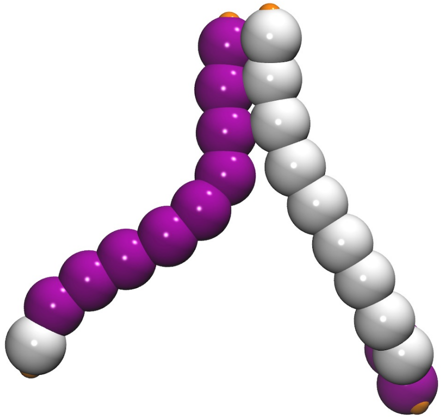
Monomer addition



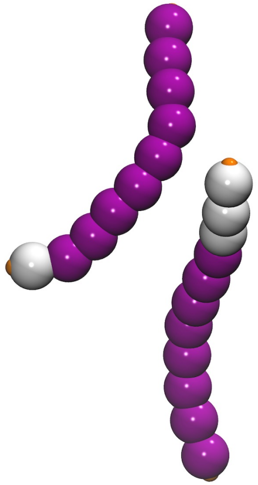
break & self-fold



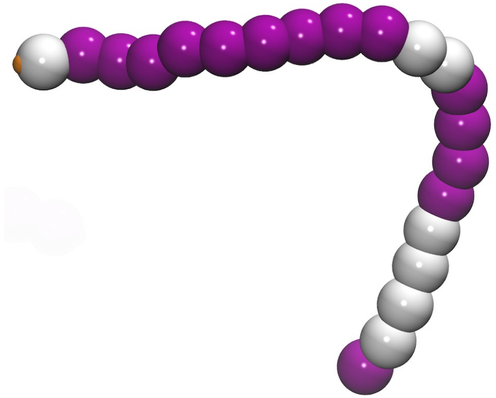
separate



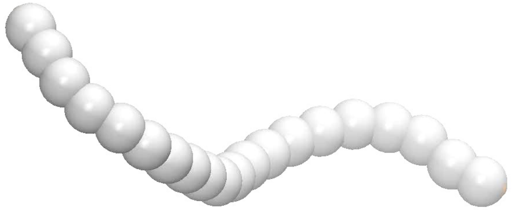
divide



merge



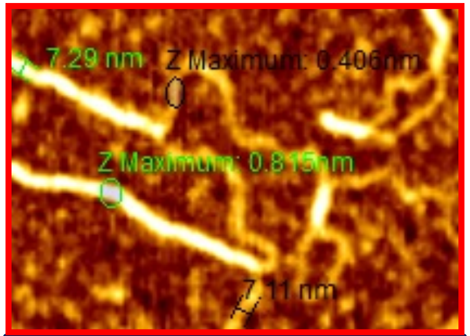
- Driven by specific and non-specific interactions
- The protofilaments are highly dynamic & flexible



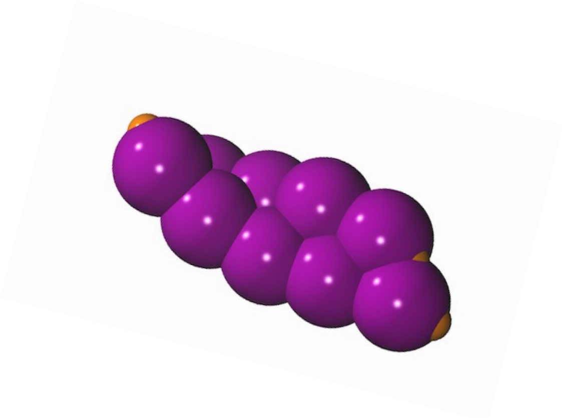

Dr. Simone Ruggeri


Bernadette Mayer

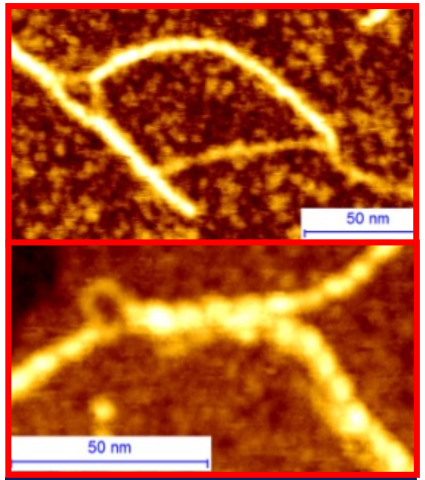
Multi-strand protofilaments



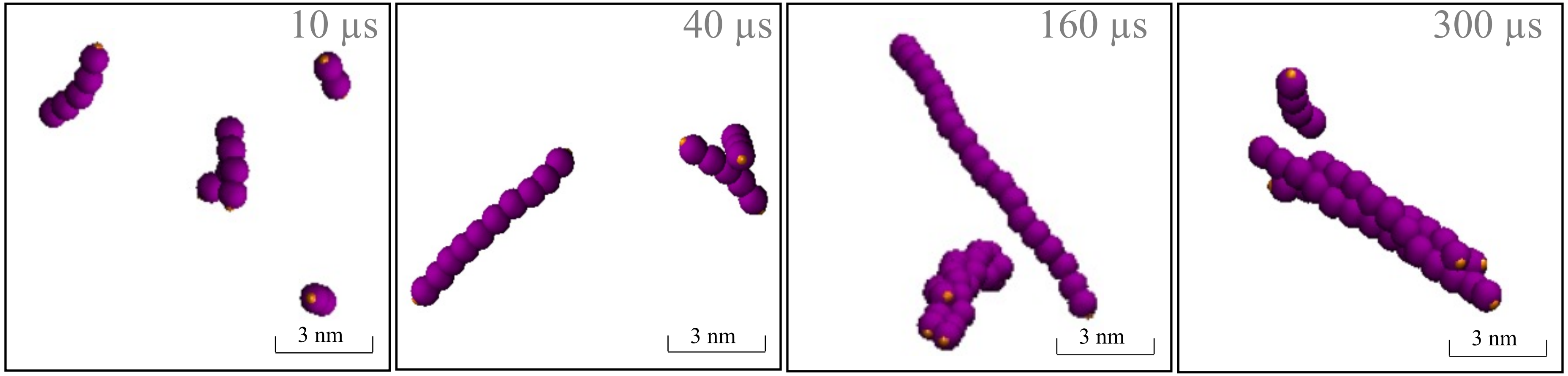
Double strand protofilaments



- short
- transient species – *on-pathway*
- more rigid than single protofilament precursors



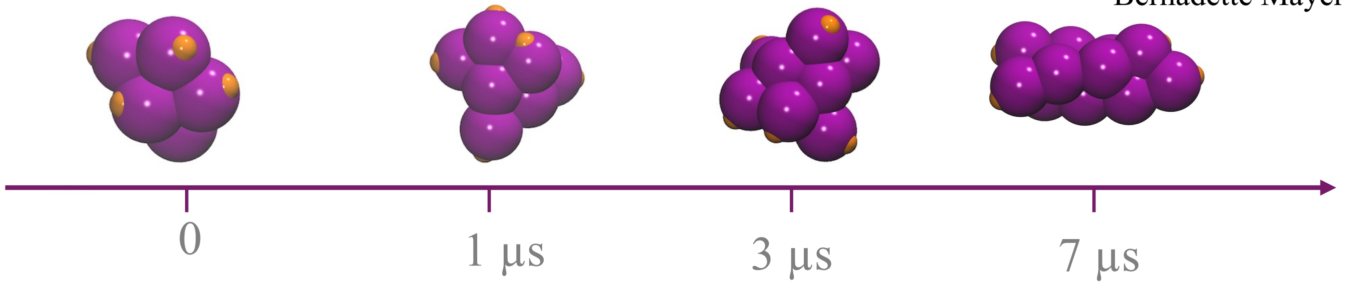
Higher order joining protofilaments



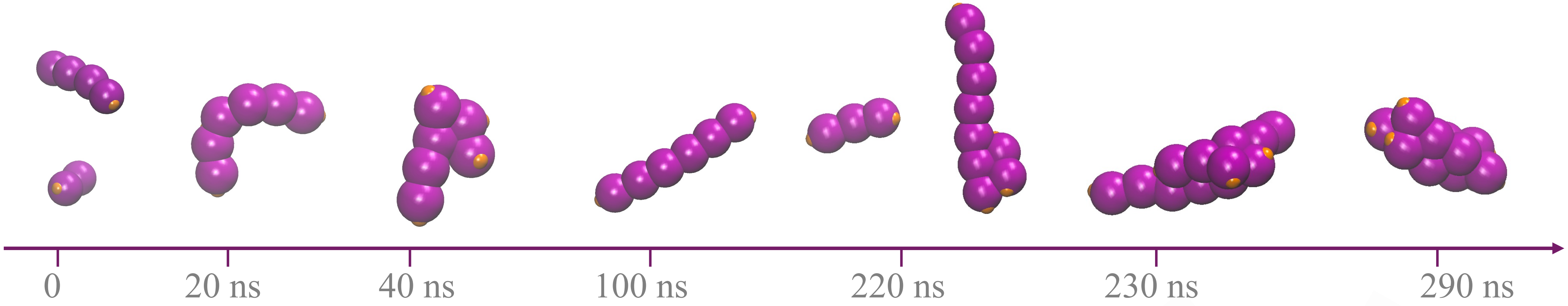
Multi-strand mechanisms


Bernadette Mayer

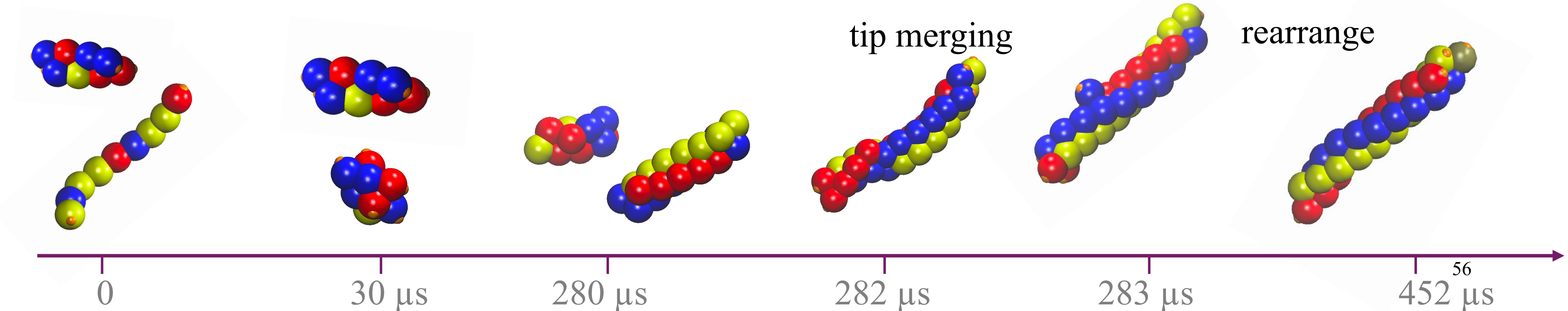
- via spherical oligomers
- driven by non-specific interactions
- stabilized by specific interactions



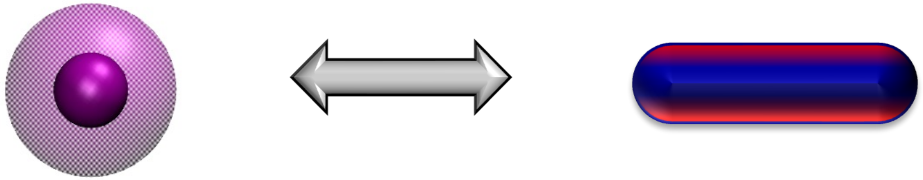
- via breakage of single filaments
- rearrangements



- via tip merging of species
- long rearrangements

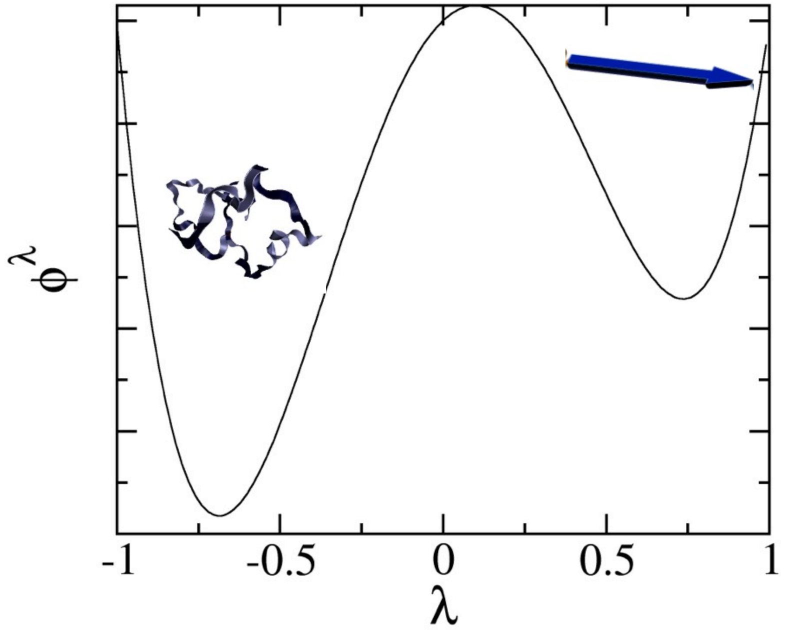


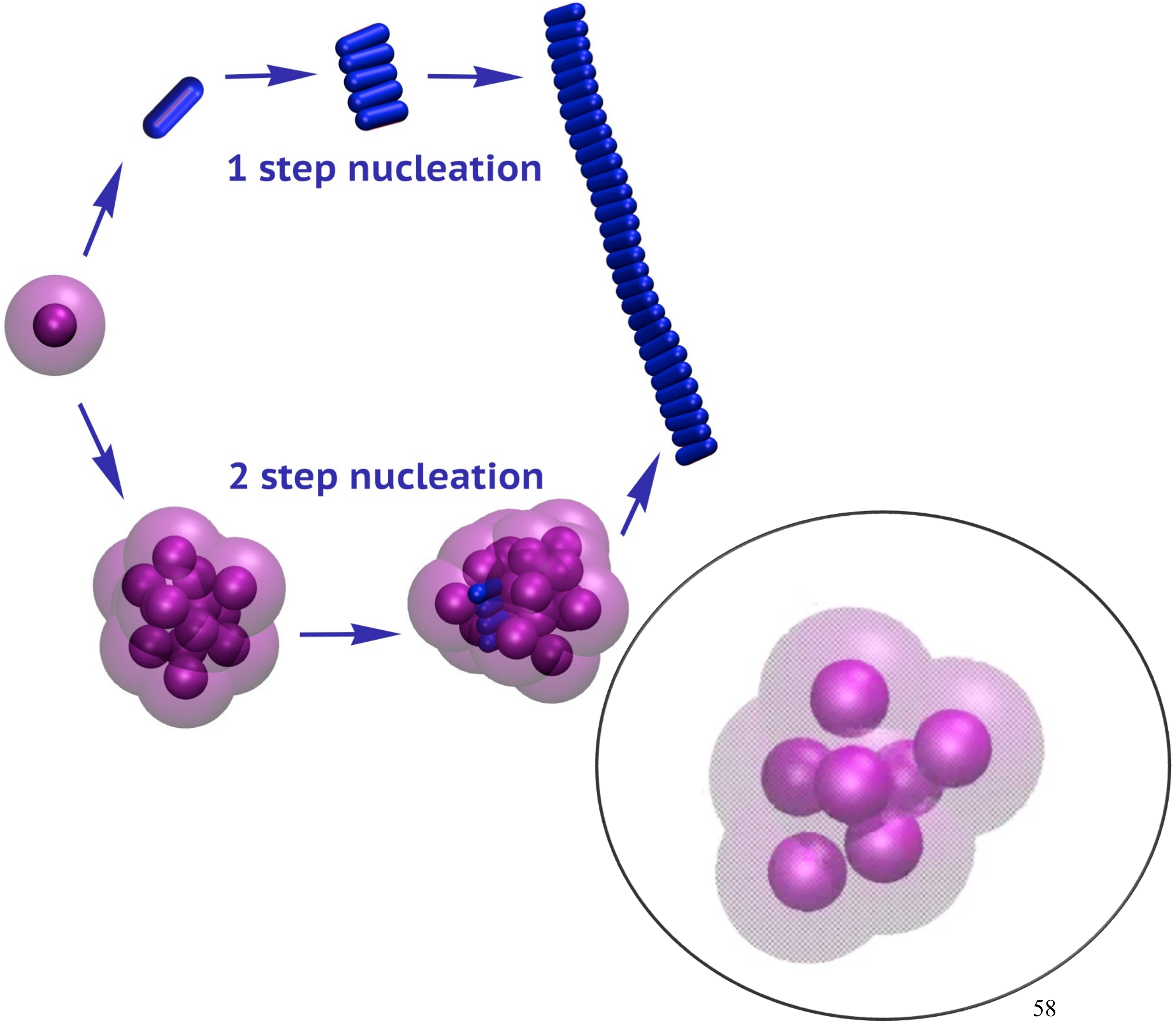
Spherical oligomers & fibrils



- **free**
- favored internally

- **bonded**
- internal + **inter-particle**



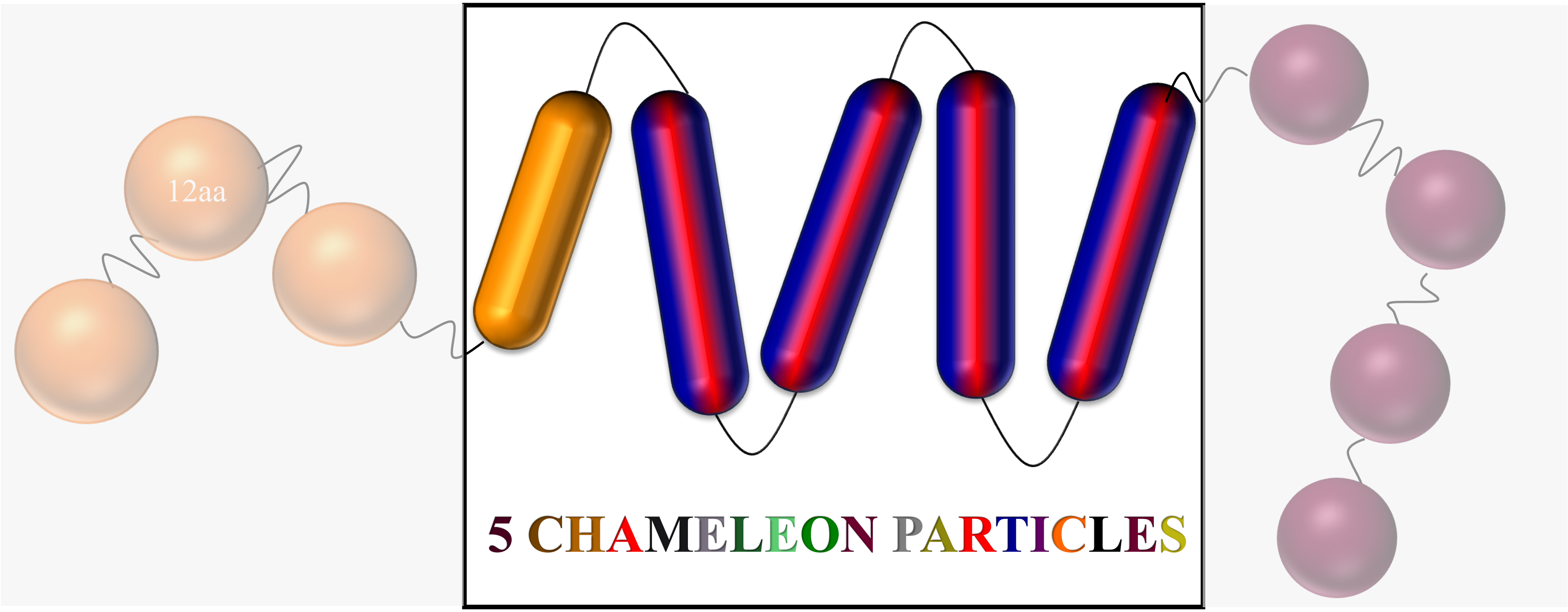


IMI et al., *J. Chem. Phys.*, 2016, **144**, 085103

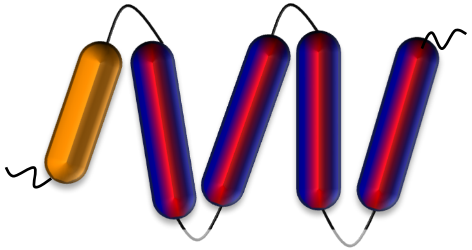
Different type of coarse graining

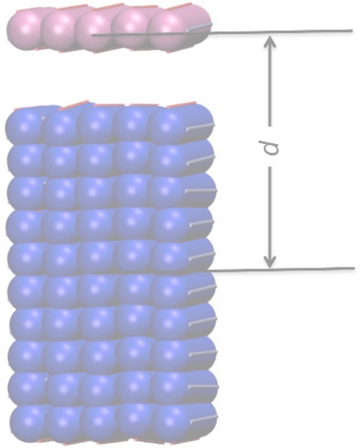
1	60-65	90-95	140

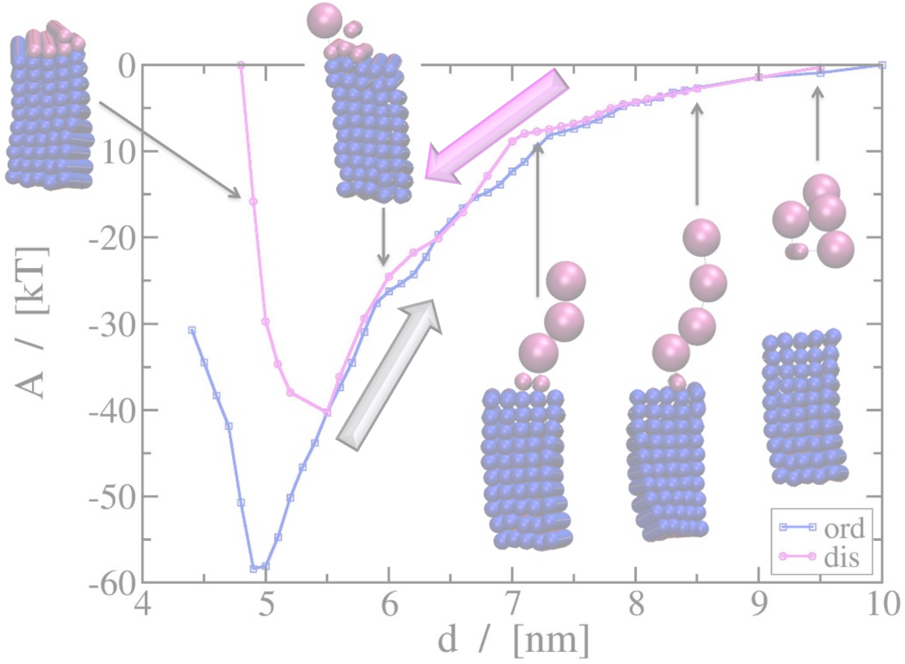
N-terminus	NAC region	C-terminus
- amphipathic (both hydrophilic and lipophilic) - α-helix, disordered	- hydrophobic - building block for α-synuclein aggregates - α-helix, β-sheets, disordered	- highly acidic - negatively charged - disordered

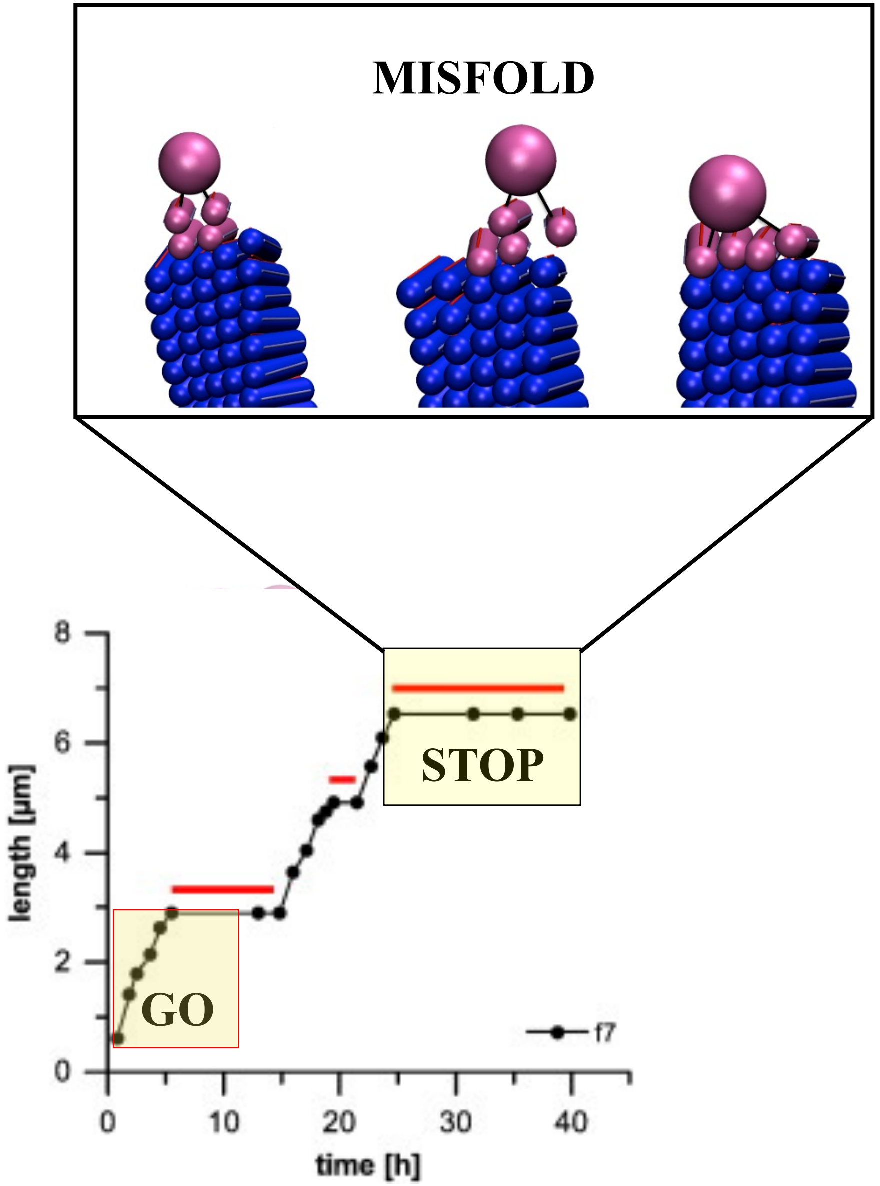


Question: Fibrillar growth



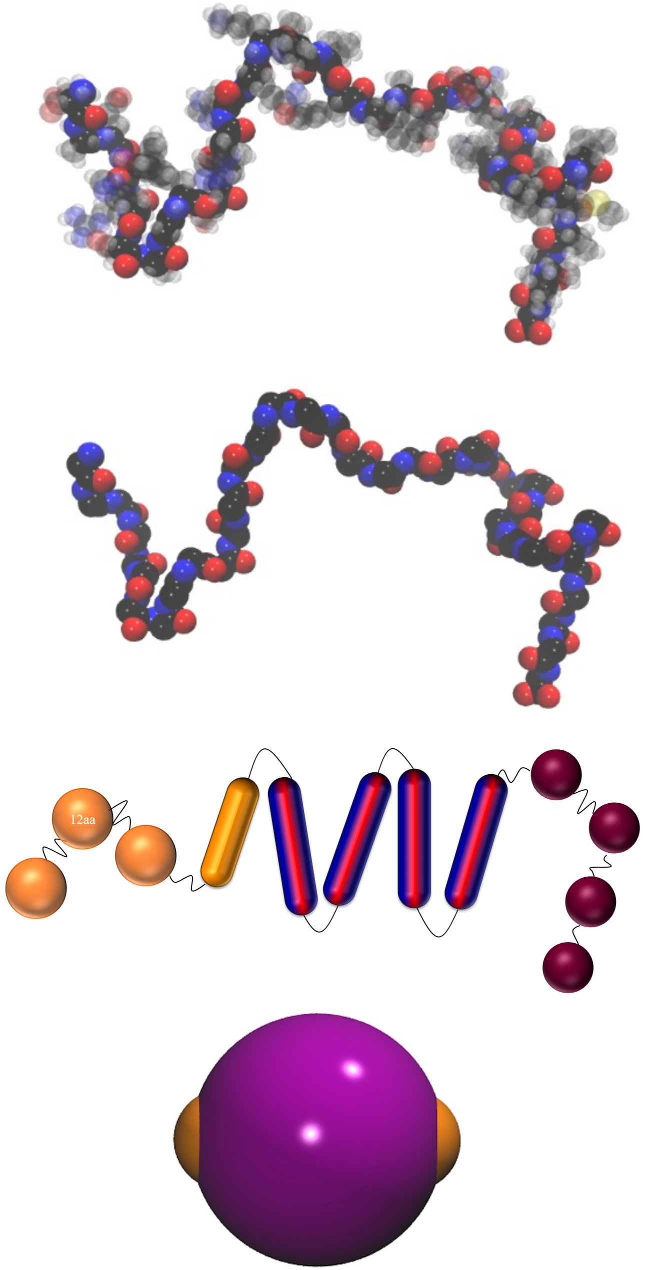






Woerdehoff et al. JMB **427:1428-1435 (2015)

ILIE et al., *J. Chem. Phys.*, **146**, 115102 (2017)


12aa