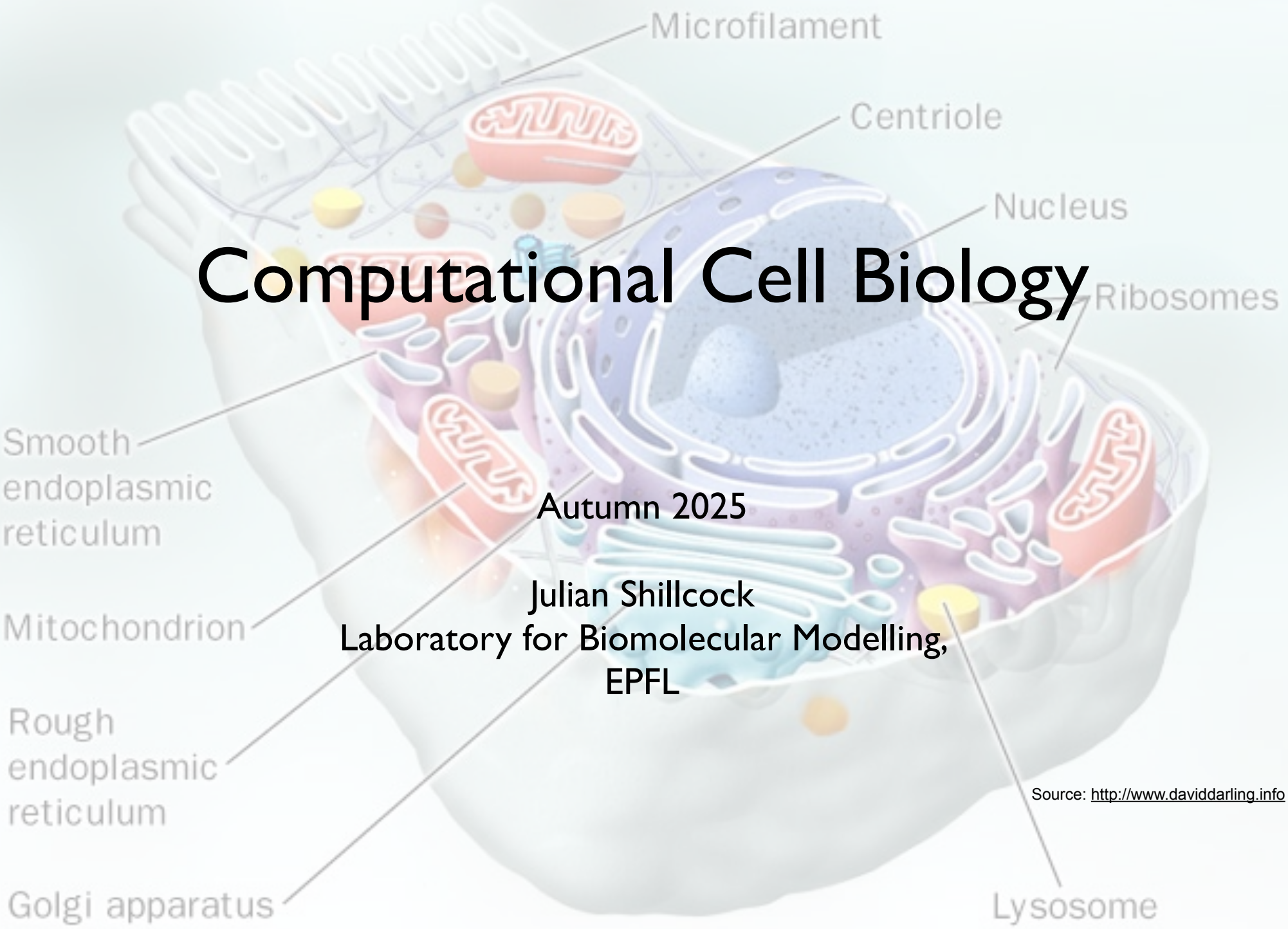
# Computational Cell Biology



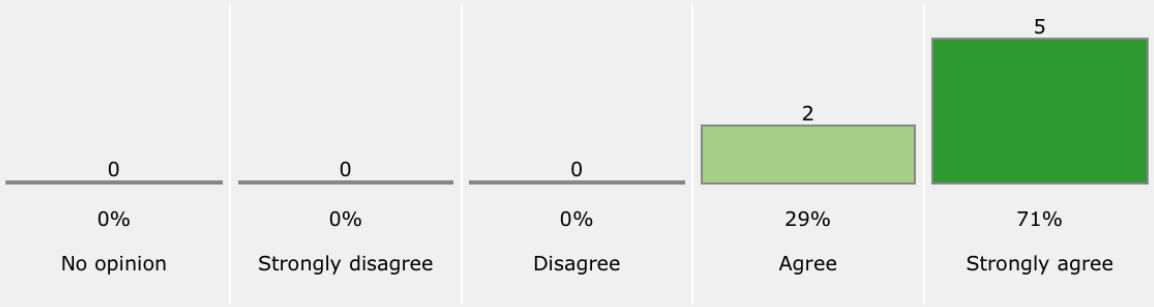

Autumn 2025

Julian Shillcock
Laboratory for Biomolecular Modelling,
EPFL

Source: http://www.daviddarling.info

# Indicative feedback

**The running of the course enables my learning and an appropriate class climate**




- Great prof., great TA, interresting lectures, nice Exercises. Generally a great course, just the workload for 4 Ects. is a lot compared to other courses, it's definitelly more than 30*4 hours. attentative listening is hard without a break.
- I feel like this course is a lot of work for only 4 credits. But I love the way the course is set up!
- I really like the course, and it is very interesting !! However, I think perhaps have less marked assignments with more percentage of the grade would perhaps be better !
- Super interesting course, the outline is clear so we are never lost in a lecture, that is really great :)
- The teacher is kind and he explain really well
- This is an excellent class, super interesting and the work assigned outside class really gets us to enagage with the topic!

tl;dr too much work. Need a break in the lecture.

# Core Concepts

Behind every physically-important equation are mysterious assumptions (e.g., polymers and random walks)

Random walks are everywhere

# Coarse-grained simulation types

All based on integrating some form of Newtonian equations of motion

| | | |
|---|---|---|
| $m.dv/dt = F$ | MD | |
| $m.dv/dt = F^C + F^D + F^R$ | DPD | Finer |
| $m.dv/dt = F^C - m\gamma.v + \sqrt{(2m\gamma k_B T)}.\zeta(t)$ | Langevin | Coarser |
| $0 = F^C - \gamma.v + \sigma.\zeta(t)$ | **Brownian** | |

In Brownian motion, nothing accelerates: everything just diffuses.


Allen, MP, and Tildesley, DJ, Computer Simulation of Liquids, Clarendon Press, Oxford, 1987

Frenkel, D and Smit, B, Understanding Molecular Simulation, Academic Press, 2002

Berendsen, HJC, Faraday Discussions 144:467 (2010)

# Brownian dynamics

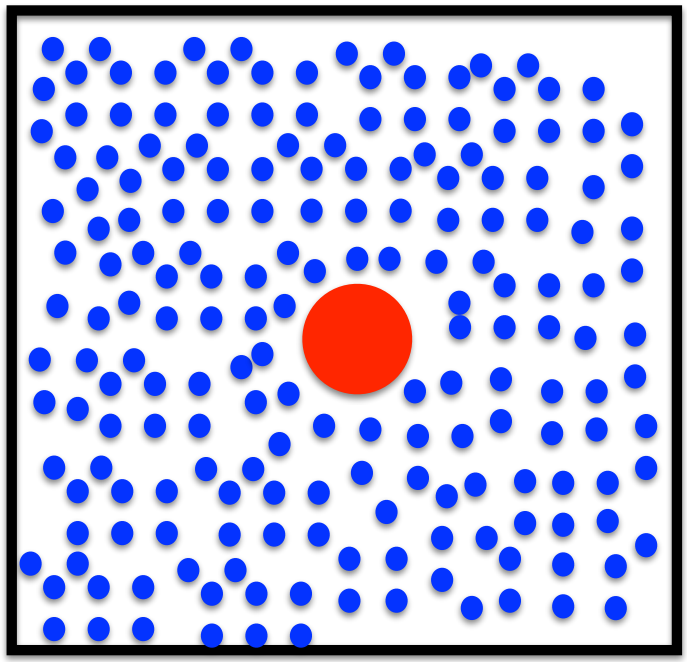
A good example of the usefulness of cg techniques is the diffusion of a nanoparticle.

A Molecular Dynamics simulation of a large* nanoparticle (100 nm - 1 micron) in water would spend >99% of the time integrating the equations of motion of the water molecules.

This is obviously inefficient.

If we could ignore the water and take it into account implicitly we could speed up the simulation enormously.

This is the motivation behind Brownian dynamics.



Not to scale

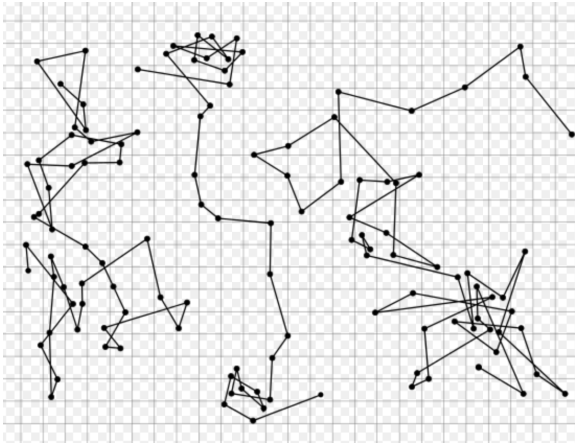
* compared to the water molecules

# History of Brownian Motion

In1827, Robert Brown observed pollen grains in water, and noticed how they moved continuously and very erratically. At first, he thought they were alive….. but pieces of glass and rock also showed similar motion. It was thought to be an experimental artifact.

In 1908, Perrin and others did experiments with more precision: different materials, size, size distribution, fractionation, counting (Perrin received Nobel Prize, 1926, for this work). (Starting with 1.2 Kg he extracted 1/10 gram of particles!)

This helped establish the molecular nature of matter which was still controversial at the time.



Jean-Baptiste Perrin, Les Atomes, 1909

P. Langevin Comptes. Rendues 146:530 (1908)

# Hypothesis or description?

In Brown's time (1820s) up to late 19th C, the existence of molecules was controversial: they were seen as a *calculational* tool to predict macroscopic thermodynamic quantities like pressure, temperature, etc., but they couldn't be *verified*. (cp. electron spin angular momentum). Note that scientists couldn't measure anything directly related to molecules then.

They also didn't know that heat is equivalent to work - Joule 1840-1849

It took a long time before scientists were convinced that Brownian motion was not an effect of external causes (vibrations, temperature differences, illumination, surface tension, microscopic currents, …) but a fundamental physical property of a fluid itself.

Question. How can one exclude that the observed Brownian motion is the result of a) temperature fluctuations in the fluid, b) chemical reactions at the particle's surface, c) microscopic currents, d) external heating gradients, etc?

# Brownian motion and molecules

In Perrin's words (quoted in Haw 2002):

*mean kinetic energy of a colloidal granule is thus equal to that of a molecule'*. He goes on: '*This is*, established by experiment,... *the theorem of the equipartition of kinetic energies. At the same stroke, the kinetic theory of fluids appears rather fortified, and molecules a little more tangible'*.

Building on these first measurements, as already described, Perrin and his students went on to construct a numerical mountain of experimental data [20–23], all of which gave results consistent with each other and with, for instance, other measures of $N$ and with the molecular–kinetic theory. He and his students also perfected direct measurements of displacements (which had presented Henri and Svedberg with such difficulty) to confirm Einstein's diffusion theory. Experiments at a range of solvent viscosities, with a range of different colloidal materials, and under various external conditions, all combined irrefutably to convert the molecular–kinetic theory from a *hypothesis* to a *description* of matter. Through Perrin's and his students' efforts, the colloidal suspension had demonstrated the molecular nature of reality.

M. D. Haw, J. Phys. Cond. Mat. 14: 7769 (2002)

# Think - Pair - Share, 5 mins.

You are given N, d, and p which are the number of steps, step size, and probability of moving right of a 1D random walker.

Write down a function that describes the mean position after N steps, <X>, using just these 3 parameters?

<X(N, d, p)> = ?

Hint. Use the limits of p = 0 and 1 and your intuition about how the mean value should depend on p.

# 1d Discrete Random Walk

A simple, discrete model of a Brownian particle in 1d, that may be symmetric or asymmetric, is the following:

Let a walker start at the origin X = 0, and make a sequence of steps, each of length d, moving right with probability p, and left with 1-p (a symmetric walk has p = 1/2).

What is the mean position $\langle X \rangle$ and its variance $\langle X^2 \rangle - \langle X \rangle^2$ after N steps?

Note that this is identical to the question: if a fair coin is tossed N times, what is the difference between the numbers of heads and tails as N increases?

(Blackboard calculation)

# Random walks and polymers

$\langle X \rangle = Nd\,(2p - 1)$

$\langle X^2 \rangle = (Nd)^2\,(2p - 1)^2 + 4\,N\,d^2\,p(1 - p)$

and so $\langle X^2 \rangle - \langle X \rangle^2 = 4\,N\,d^2\,p(1 - p)$

If $p = 1/2$, then

$$\langle X^2 \rangle = Nd^2$$

Look familiar? $\langle R_{ee}^2 \rangle = Na^2$

The phantom chain model of a polymer *on average* is a random walk in space: any single conformation is not representative of its behaviour, but the average of all its conformations is.

# Coarse-grained simulation types

All based on integrating some form of Newtonian equations of motion

$m.dv/dt = F$ MD

$m.dv/dt = F^C + F^D + F^R$ DPD

$m.dv/dt = F^C - m\gamma.v + \sqrt{(2m\gamma k_B T)}.\zeta(t)$ **Langevin**

$0 = F^C - \gamma.v + \sigma.\zeta(t)$ Brownian


In Langevin dynamics, we keep the acceleration term, and we get more information about the motion than jut $<X^2>$ ~ Time, we get the prefactor.

# Langevin's solution of Brownian Motion

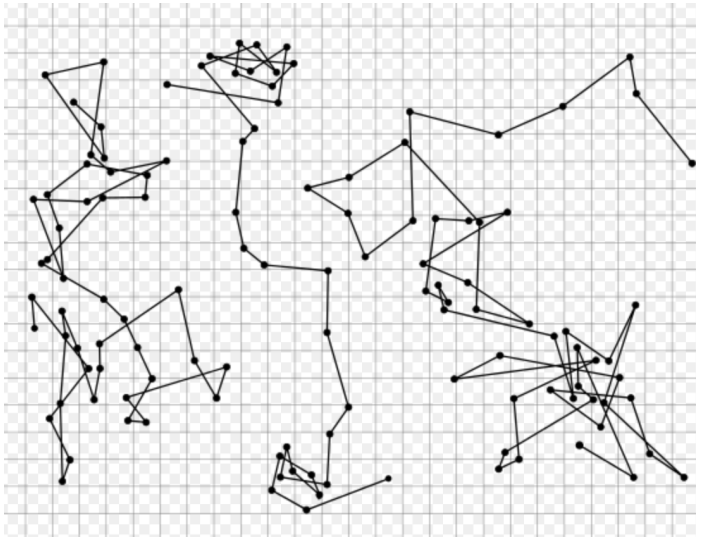
Langevin in 1908 explained Brown's observations starting from the **equipartition theorem** that a particle of mass **m** in equilibrium should have a mean KE of: $\langle 1/2\ mv^2 \rangle = 1/2\ k_B T$

(in 1d, and 3/2 $k_B T$ in three-dimensional space).

He assumed two forces act on a particle of mass **m** in water (where **m** >> $m_W$):

1) a viscous drag force (Stokes' law) ~ $-6.\pi.\eta.a.dx/dt$

   $\eta$ = viscosity
   a = particle radius
   dx/dt = particle velocity



2) a rapidly fluctuating force X(t) subject to:

   $\langle X(t) \rangle = 0$ and $\langle x.X(t) \rangle = 0$

   $\langle X(t)^2 \rangle \neq 0$

Newton's EOM is: $m.d^2x/dt^2 = -6.\pi.\eta.a.dx/dt + X(t)$

# Langevin's solution of Brownian Motion

Langevin's solution for the mean-square displacement of a particle in solution is:

$$\langle x^2 \rangle - \langle x_0^2 \rangle = ( k_B T / 3\pi\eta a)\, t = 2Dt$$

where the diffusion constant is:

$$D = k_B T / 6\pi\eta a$$

The mean-square displacement (MSD) increases *linearly* with time.

In 3D, momentum is conserved in each dimension the particle makes independent random moves in each dimension and the net displacement is the sum of the 3 independent ones:

$$\langle R^2 \rangle = \langle X^2 \rangle + \langle Y^2 \rangle + \langle Z^2 \rangle = 6Dt$$

so in a d-dimensional isotropic space, the mean-square displacement of a RW at time t is:

$$\langle R^2 \rangle = 2dDt$$

NB. D is the 1D diffusion constant which must be multiplied by dimension.

# Langevin's solution of Brownian Motion

Note that $\langle x^2 \rangle \sim t$ means that the particle's instantaneous velocity is not well defined:

Ballistic motion: $x = v.t$, $\quad v(t) = dx/dt = v$

Brownian motion: $\sqrt{\langle x^2 \rangle} \sim \sqrt{t}$, $\quad v(t) = d/dt\ (\sqrt{\langle x^2 \rangle}) = 1/2\ t^{-1/2}$

which caused many problems in the original experiments as they first tried to measure the particles' velocities from graphs of displacement versus time, but eventually used the MSD.

Question: Can one formulate a theory of Brownian motion that does NOT require molecules? i.e., one that still has a wildly erratic path but there are no molecules to *kick* the particle.

Note that the drag force can have the same form for a continuum fluid, but what causes the random force?

Break 5 mins.

# Langevin's equation is an SDE

Langevin's equation is an example of a stochastic differential equation (SDE), in which there is a random term X(t). Each solution of an SDE represents a *different* random trajectory, but their average properties can be calculated if properties of the random function X(t) are defined.

Contrast this with a deterministic differential equation that has a unique solution.

A Langevin equation looks like:

$$dx(t) = x(t + dt) - x(t) = A(x, t)*dt + B(x, t)*X(t)$$

where A(x, t) is called the Drift term, and B(x, t) the diffusion term.

A and B have distinct physical interpretations, and X(t) is a random *noise* term that must be defined more carefully.

Recall Langevin's assumptions:

- diffusing particle is much larger than the water molecules (**separation of scales**)
- many collisions in any measurable time interval
- viscosity of solvent applies a drag to the particle as at macroscopic scale
- solvent/particle collisions are independent ( $\langle x.X(t) \rangle = 0$)

# Integrating a Langevin equation

We have been a bit loose with the function X(t) that represents random solvent collisions.

All we have said is that $\langle X(t) \rangle = 0$, $\langle x.X(t) \rangle = 0$ and $\langle X^2(t) \rangle \neq 0$, but what is X(t)?

How would we integrate an equation that had X(t) in it on a computer?

Consider the deterministic differential equation:

$$dx/dt = A(x, t)$$

we can discretize this for use on a computer:

$$x(t + dt) - x(t) = A(x, t).dt$$

Now consider the Langevin equation (taking $B(x,t) = \sqrt{D}$ = constant):

$$x(t + dt) - x(t) = A(x).dt + \sqrt{D}.X(t)$$

# Integrating a Langevin equation

It turns out that X(t) has to satisfy some strict conditions to be mathematically sensible and computationally useful - see Gillespie reference for details.

A *normal* or *Gaussian random variable* X = N(m, σ), with mean m and variance $\sigma^2$, is one for which X takes a value x with probability:

$$p(X = x) = 1/\sqrt{(2.\pi.\sigma^2)}.\ \exp(\ -(x - m)^2 / 2.\sigma^2\ )$$

And the *only* well-defined, continuous, memory-less, stochastic process (Langevin equation) is:

$$dx(t) = x(t + dt) - x(t) = A(x).dt + \sqrt{D}.\ N(0, 1).\sqrt{dt}$$

where N(0, 1) is the *unit normal random variable* with mean 0 and variance 1. The square-root of dt is crucial.

D.T. Gillespie, The mathematics of Brownian motion and Johnson noise, Am. J. Phys. 64:225 (1996)

# Implementing a Langevin equation

Our original noise term must therefore have the form:

$$X(t) = N(0, 1)\sqrt{dt}$$

and the Langevin equation that we can implement on a computer is:

$$dx(t) = x(t + dt) - x(t) = A(x).dt + \sqrt{D}.\, N(0, 1).\sqrt{dt}$$

where N(0, 1) is a Gaussian random variable with zero mean and unit variance that we sample at each time step to find the next increment dx(t).

Note

1) We cannot ignore dt wrt $\sqrt{dt}$ because the term N(0, 1) is equally often positive and negative which reduces the magnitude of the sum of many random samples.
2) The square root is necessary to reproduce $\langle X^2 \rangle \sim D.T$ for a diffusive process. No other power will do. D is the diffusion constant.
3) This equation forms the basis for the Brownian Dynamics simulation technique
4) Langevin equations can involve correlated randomness not only noise

H. Eugene Stanley, Correlated Randomness Prama J. Physics, 64:645-660 (2005)

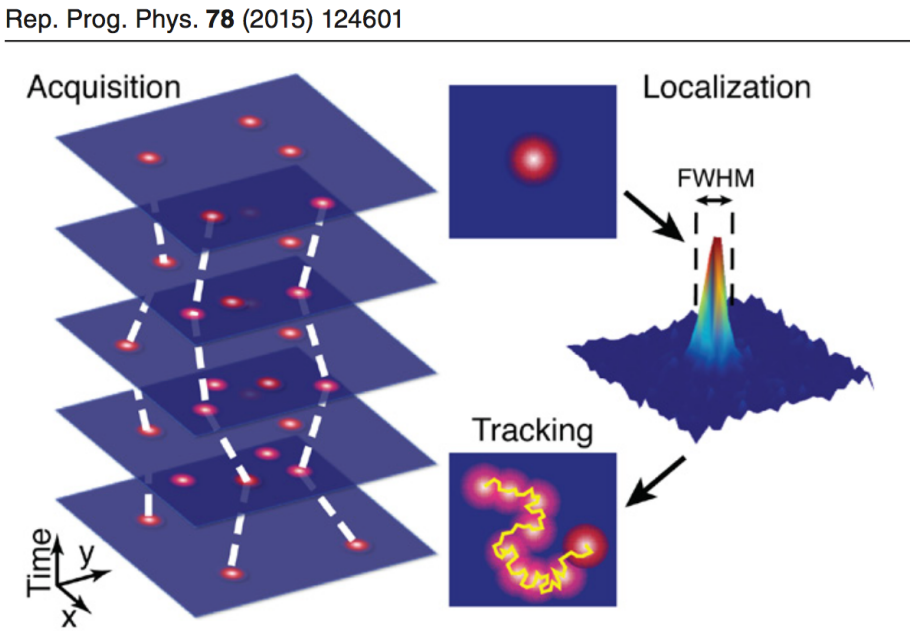
D. L. Ermak, and J. A. McCammon, Brownian Dynamics with Hydrodynamic Interactions, J. Chem. Phys. 69:1352 (1978)

# Measuring Brownian motion

Brown used light microscopy to measure the diffusion of single particles; what experimental techniques are available now?

1) SPT - single particle tracking

Updated version of Brown's method that uses light microscopy to track a single fluorescently-labelled particle, e.g., a quantum dot, as it diffuses in space or on a cell's surface



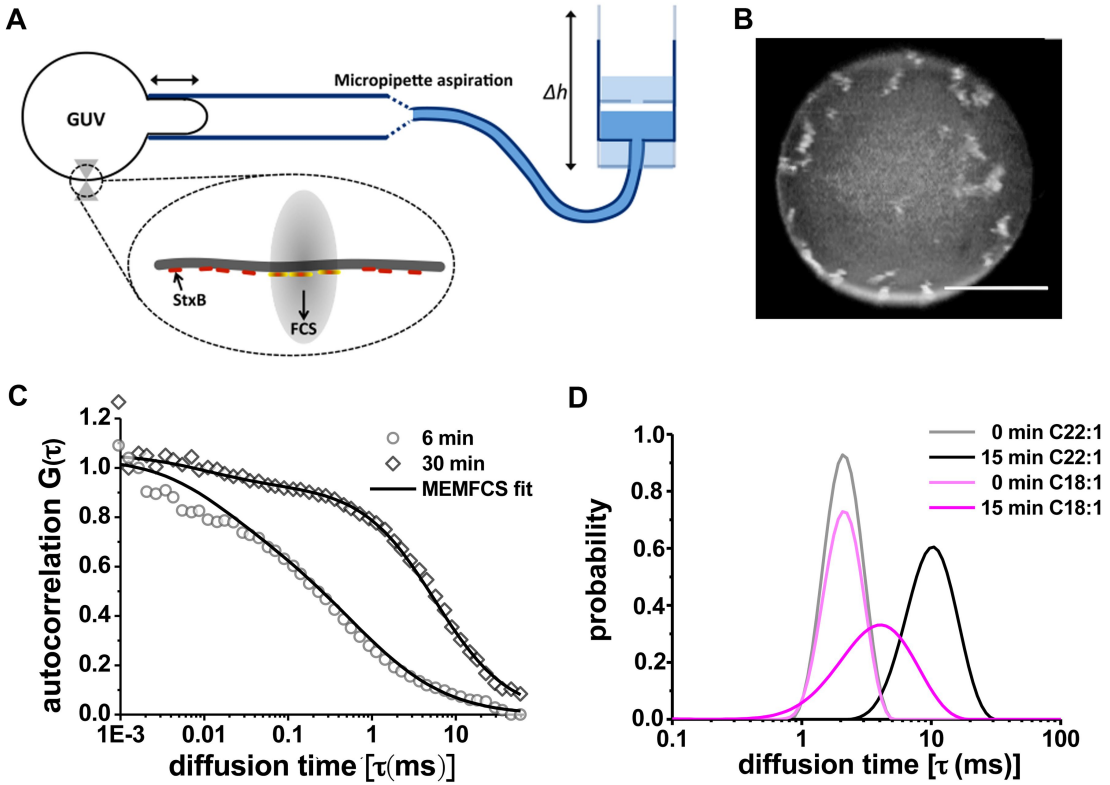

| Pros | Cons |
|---|---|
| complete trajectory | tracks may reflect distinct processes; localisation errors (stuck particle ~ small D), diffusion and sub-diffusion mixed |
| no ensemble averaging | signal is inherently noisy |
| | optical resolution |
| | tracks may have gaps especially for blinking QDs |

C. Manzo and M. F. Garcia-Parajo, A Review of progress in single particle tracking; from methods to biophysical insights. Rep. Prog. Phys. 78:124601 (2015)

# Measuring Brownian motion

## 2) FCS - fluorescence correlation spectroscopy

Laser light is focussed on a spot, and the scattered intensity from the (dilute) fluorescing particles is measured as a function of time and the two-time correlation function is analysed to extract the diffusion coefficient of the particles.



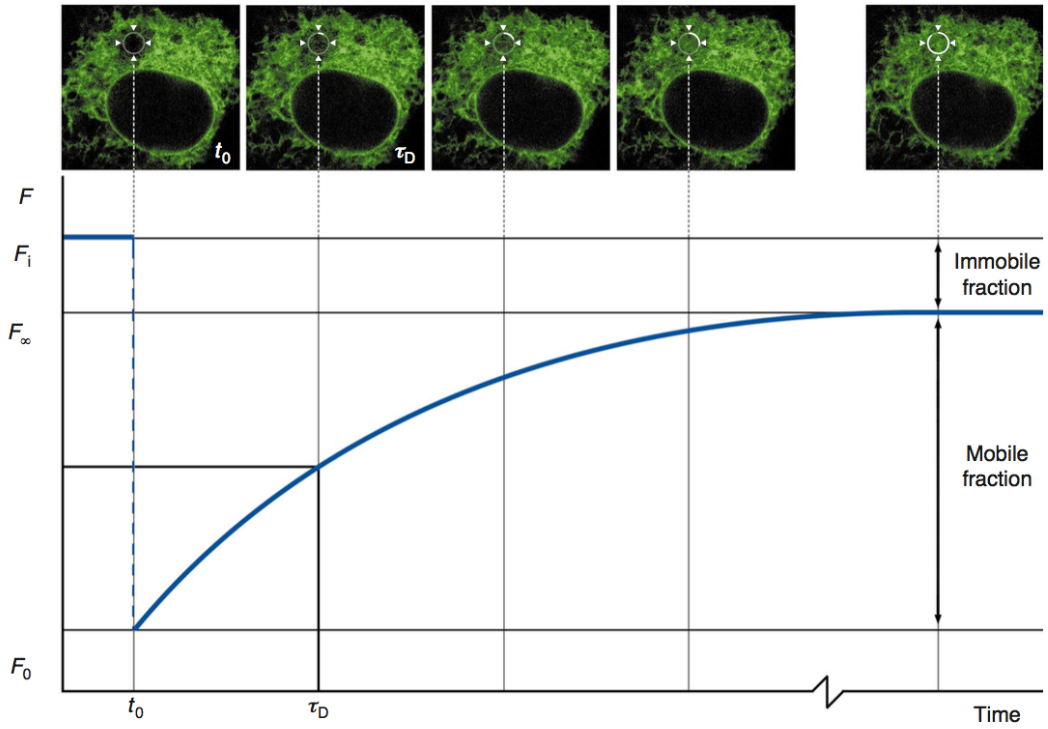

| Pros | Cons |
|---|---|
| Good statistics | Needs a (complex) model to extract diffusion constants |
| | needs a dilute system |

E. L. Elson, Fluorescence Correlation Spectroscopy: Past, Present and Future. Biophys. J. 101:2855 (2011)

# Measuring Brownian motion

## 3) FRAP - fluorescence recovery after photobleaching

A region of membrane containing diffusing dye molecules is irreversibly bleached by intense light, and the gradual recovery of the fluorescence as unbleached dye diffuses back into the region carries information about the particles' diffusion coefficient




Pros

Measure the mobile fraction and diffusion constant

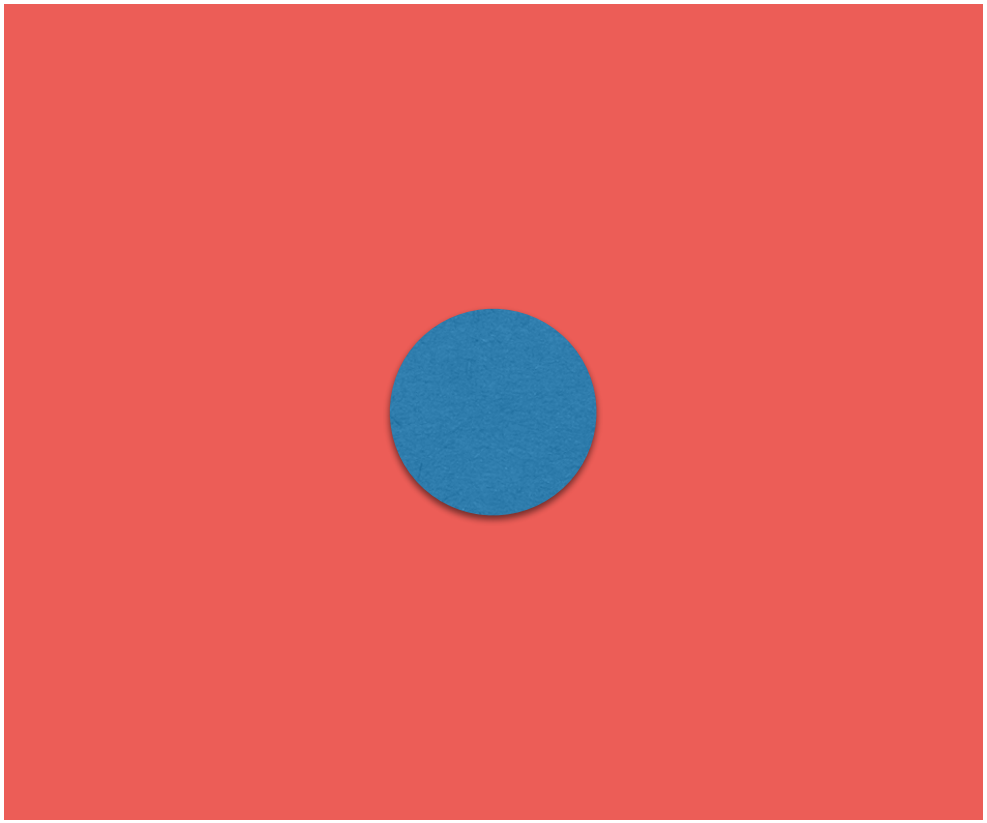
Measures diffusion, reactions, conformational changes

Cons

2D unless complex model used

E. A. J. Reits and J. J. Neefjes, From fixed to FRAP: measuring protein mobility and activity in living cells. Nature Cell Biology 3:E145 (2001).

# FRAP simulation



1) Create a membrane in the initial state
2) Issue a command to change the type and colour of lipids in a circular region
3) Measure the diffusion of the coloured lipids out of/other lipids into the circular region.

# FRAP simulation

1) Create a membrane in the initial state (see today's moodle page for dmpci.frap1)
2) Issue a command to change the type and colour of lipids in a circular region
3) Measure the diffusion of the coloured lipids out of/other lipids into the circular region.

See the commands:

**SelectBeadTypeInCylinder** - makes a "target" out of all beads in a cylinder

**SetTargetDisplayId** - changes the colour of the beads in a target

**ChangeNamedBeadType** - assigns a new bead type to beads in a target

**SetDPDBeadConsInt** - changes the conservative force between two bead types

# Other Random Walks

We have viewed the RW as the track of a particle moving in space, but there are many other applications. We can use it to represent, e.g., the membrane voltage u(t) of a "noisy neuron".

The membrane voltage for a *leaky integrate-and-fire* neuron is:

$\tau . du(t)/dt = -u(t) + R . I(t)$

where $\tau$ is the "time constant" (or memory), R membrane resistance, I(t) current. The voltage u(t) varies and when it crosses a threshold a "spike" is generated and u(t) reset.

We can add noise to the voltage equation to get the Langevin equation:

$\tau . du(t)/dt = -u(t) + R . I(t) + \Gamma(t)$

where the white noise term is as before:

$< \Gamma(t) > = 0$
$< \Gamma(t) . \Gamma(t') > = \delta(t - t')$

Sect. 5.5, W. Gerstner and W. Kistler, Spiking Neuron Models, Cambridge University Press (2002)

# Where do RWs appear in a cell

Now we have a model for RWs, we can see them everywhere

Bulk diffusion in cytoplasm (3d)

Lipid and protein diffusion in membranes (2d)

Ion diffusion through channel proteins (1d)

Actin monomers diffuse and bind to form filaments (3d)

Motor protein diffusion along filaments (1d)

DNA binding proteins, transcription and translation (1d)

(Reaction coordinate in chemical reactions, 1d)

(Membrane potential of a neuron, 1d)

# What use is noise in a cell?

Noise = random thermal motion = unlimited source of energy

Do we have perpetual motion?

In a way, except that it is undirected motion that cannot do useful work without constraints

The cell uses the random thermal noise to create structures of use and to *search* states of interacting particles, or explore (short-distance) space by diffusion, e.g.,

- membranes form spontaneously from amphiphilic lipids in water
- ions flow through a channel in a membrane to do work (but a pump maintains the gradient)
- filaments spontaneously *assemble* but to *disassemble* them requires energy consumption;
- motor proteins pull vesicles along filaments, but ATP is required to make the motion directed
- two chemical reactants will randomly explore possible binding conformations, but to separate them requires expenditure of ATP
- noise allows a system to jump over energy barriers

# Summary

Coarse-graining is an art: some things you get right, but others will be wrong:

Expt: many fast processes (molecular collisions) give rise to slow ones (diffusion)

Simulations: we replace the fast processes by effective forces that recreate the slow evolution of a system from one state to another, e..g, nanoparticle diffusing in water

Random walks are everywhere, and provide a tool for modelling many dynamic processes.

Many processes are the result of many uncorrelated, smaller ones (hierarchy), and the sum of many uncorrelated events is a random process.

Break 10 mins.

# Take-home test 2: Data Management Plans

Funding agencies now expect applications to include a DMP. A DMP specifies what data you will produce, how you will store it and name files, and how other users can find it and use it again.

EPFL provides information on preparing DMPs:

https://www.epfl.ch/campus/library/services/services-researchers/rdm-guides-templates/

A common standard are referred to as the *FAIR* principles:

https://www.epfl.ch/campus/library/wp-content/uploads/2019/09/EPFL_Library_RDM_FastGuide_All.pdf#page=2

# DMP

A DMP is a *living document* that describes the following aspects of your data:

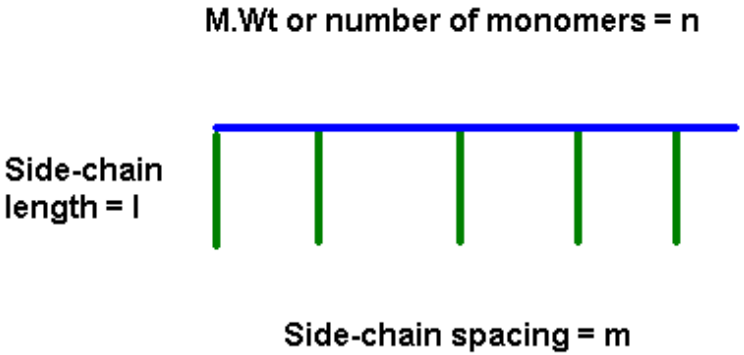
- What data is produced? e.g., fluorescent microscopy images, simulation files, gene sequences, etc.
- What formats are used to store it? e.g., tiff, png, bmp, ascii
- What file/directory naming strategy will be used? e.g., /users/shillcock/my_data
- (What license it will be released under - not relevant in this course)

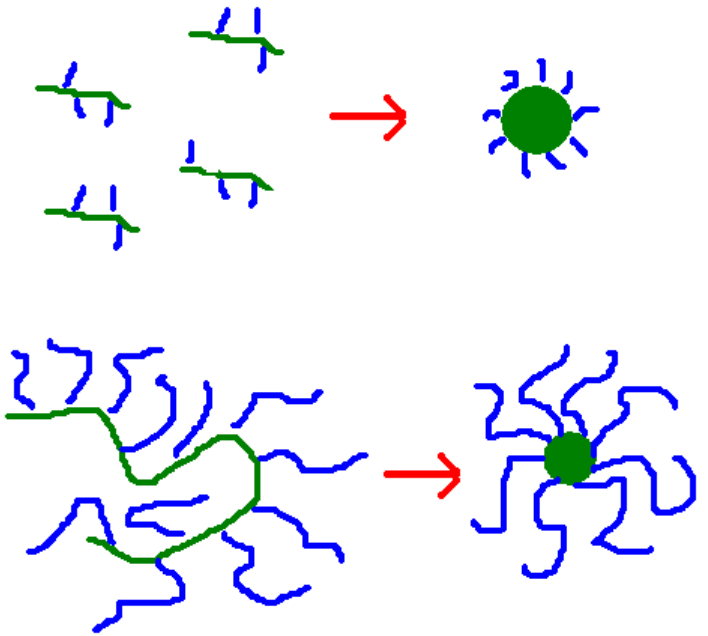
etc

*Findable* - how can a user find the data?
*Accessible* - how is it licensed and released?
*Inter*-operable - does user need special software to read data?
*Reusable* - Long-term storage of data in a repository

Living = continuously updated throughout a project

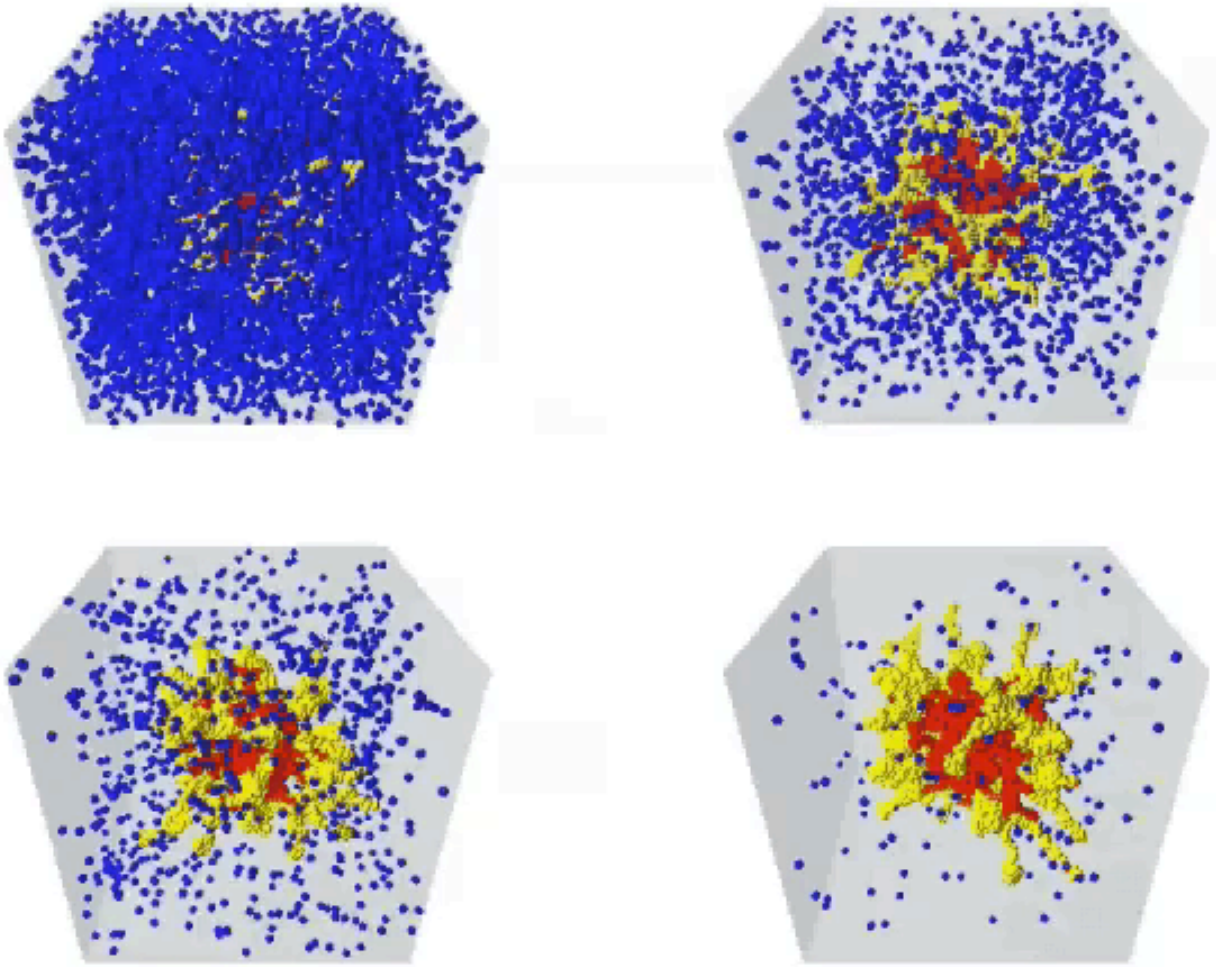
# Comb polymer self-assembly



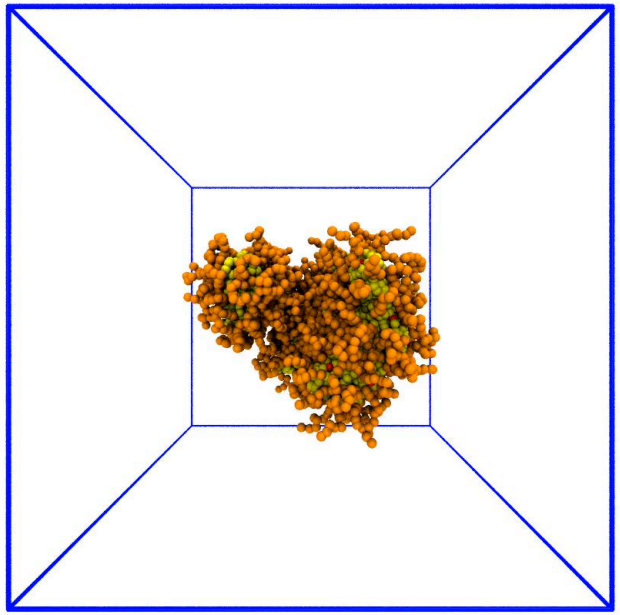
How does the molecular shape affect the aggregate's structure?

e.g., if you cut off every other side-chain, what happens?

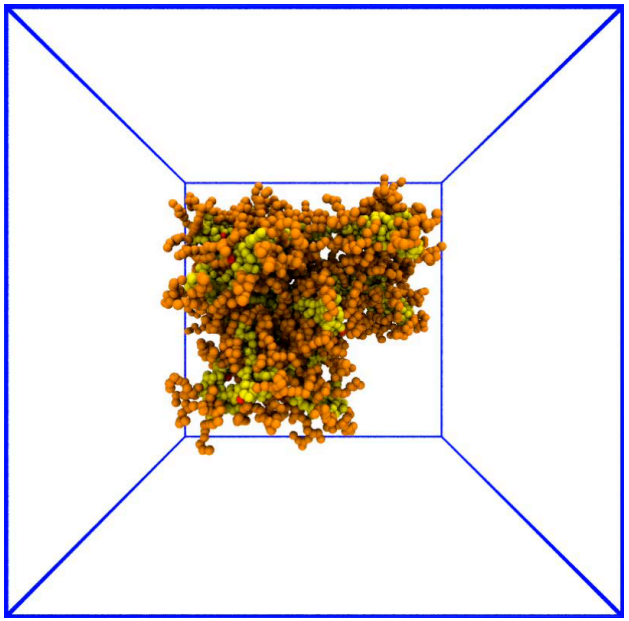


15113 - 15116

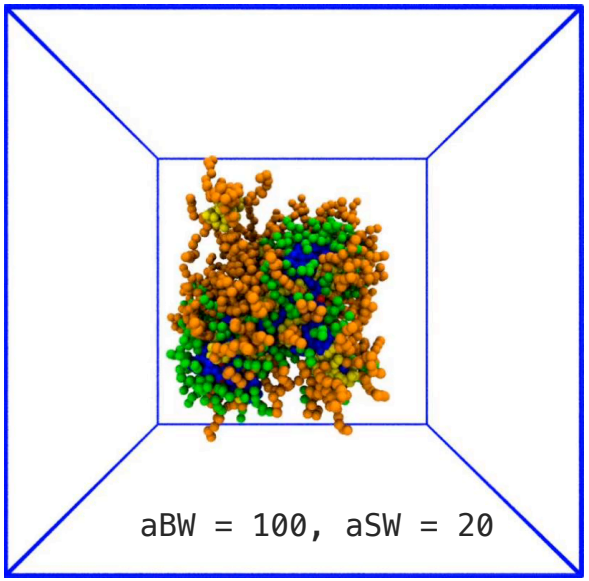
How does the molecular shape affect the aggregate structure?

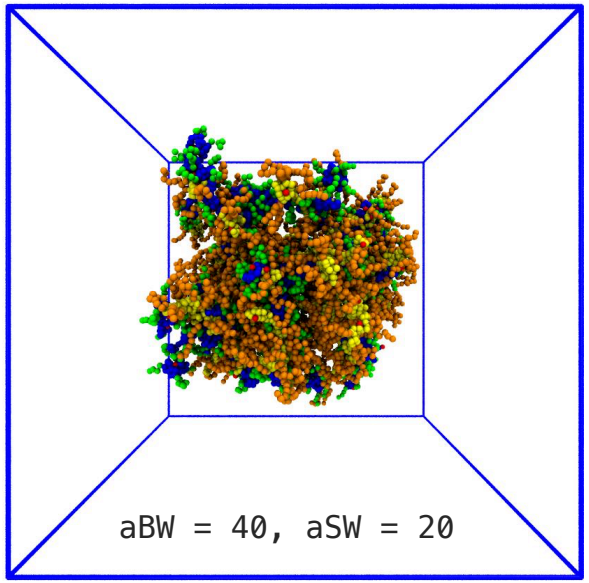


aBW = 70
Polymer Comb 0.0004
" (H1 (8 (B B B (* (S (6 S) S)) B B B)) T1) "



aBW = 40
Polymer Comb 0.0004
" (H1 (8 (B B B (* (S (6 S) S)) B B B)) T1) "



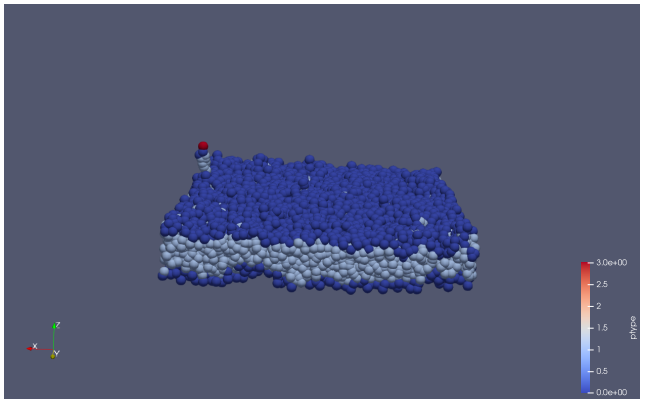




Polymer Comb1  0.0002 " (H1 (8 (B B B (* (S (6 S) S)) B B B)) T1) "
Polymer Comb2  0.0002 " (H1 (8 (B1 B1 B1 (* (S1 S1 S1 S1)) B1 B1 B1)) T1) "

# Ex. 5 Molecular force spectroscopy 1

Select a single lipid:

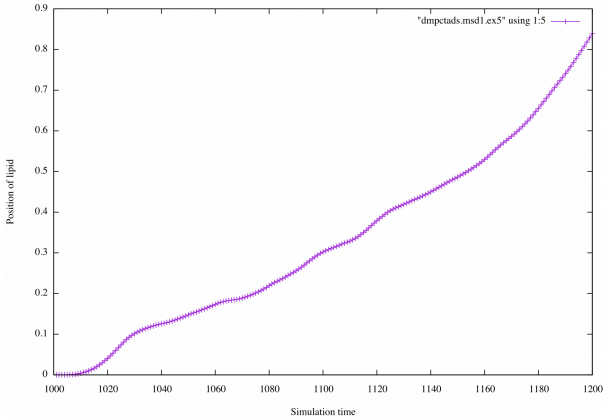
```
Polymer Water   0.978967   " (W) "
Polymer Lipid   0.021013   " (H H (* (T T T T)) H T T T T) "
Polymer Lipid1  0.00002    " (H1 H (* (T T T T)) H T T T T) "
```



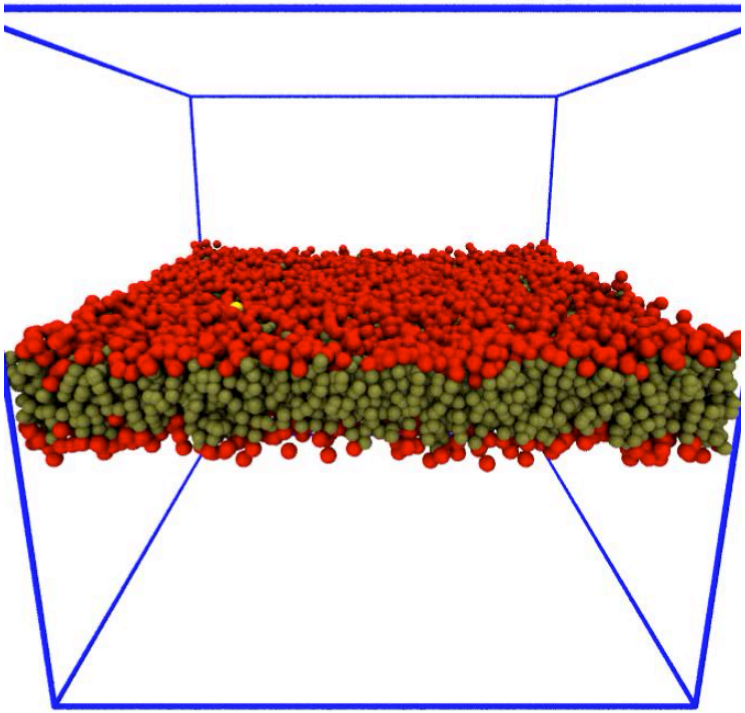
apply a pulling force normal to membrane:

```
Command Comment   1 // Following commands apply a force upwards to the single  lipid with the H1 bead  //

Command SelectBeadTypeInSimBox        1  head  H1

Command ConstantForceOnTarget         100  head  fh   0 0 1   20.0

Command RemoveCommandTargetActivity   1000   fh
```

measure position of lipid and work done on it

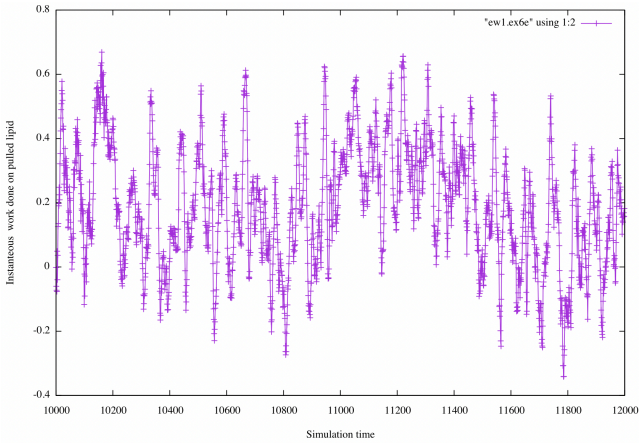


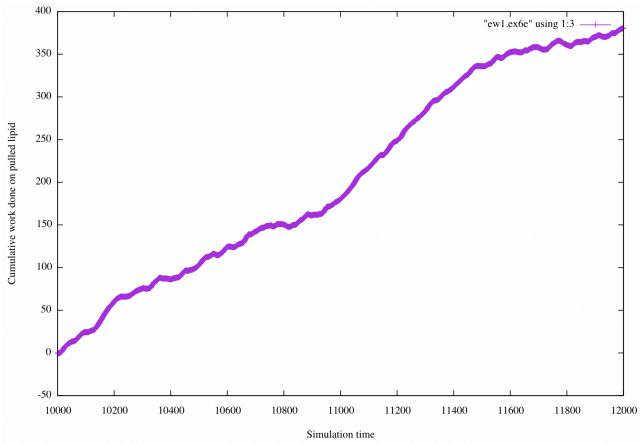

Lipid position in z direction

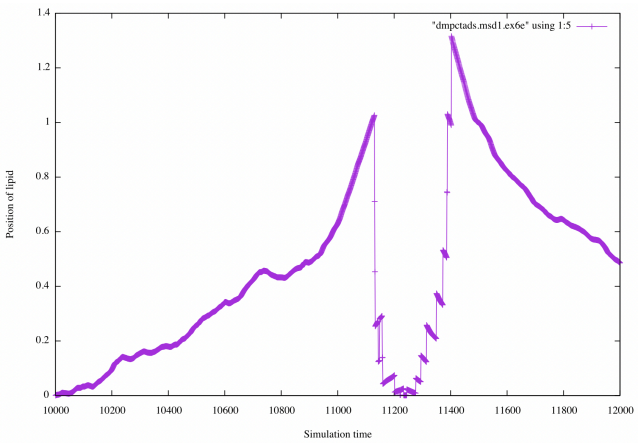


# Ex. 6 Molecular force spectroscopy 2

Apply a force to a single IDP in a droplet and extract it and measure the work done against the force

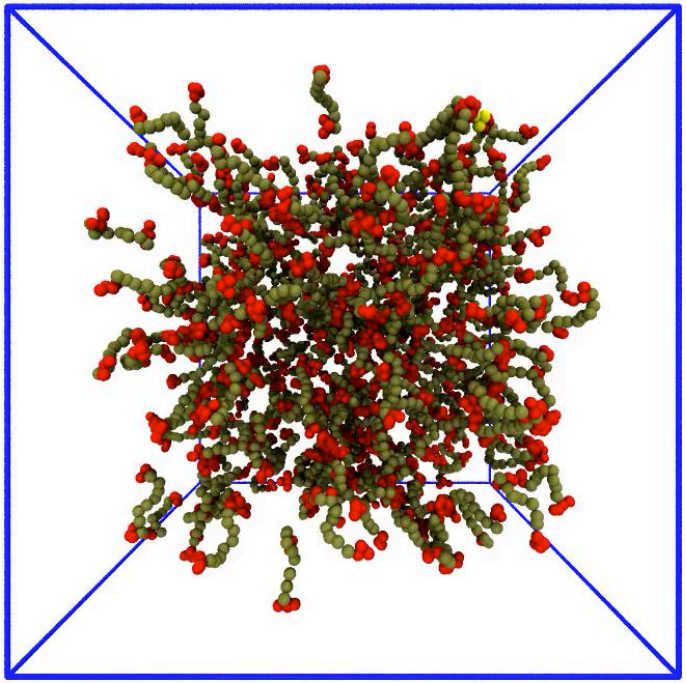



Work done

Position



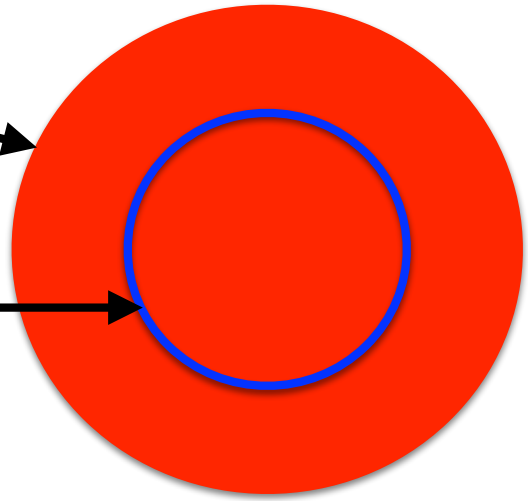
see dmpci.ex6e

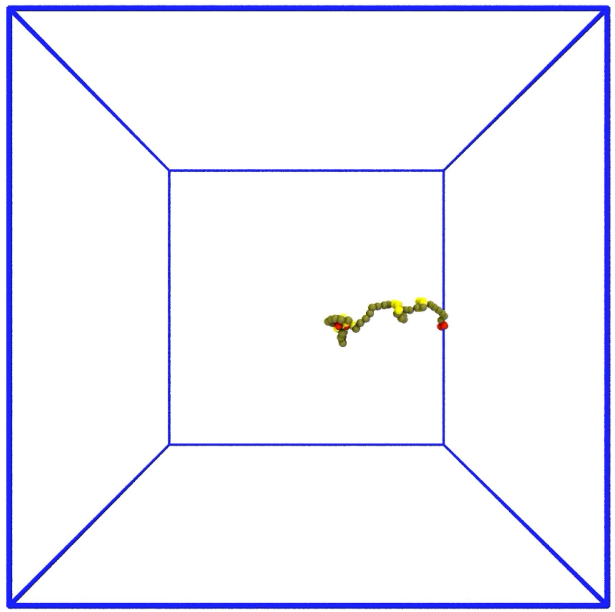
# Ex. 8 Compare the diffusion of a hard sphere and a fluctuating polymer

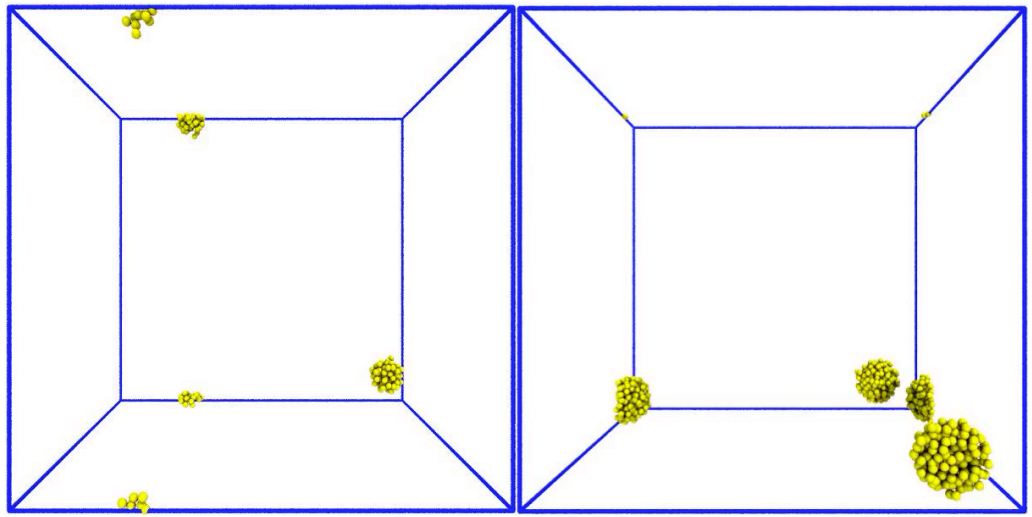
$R_h$ = actual sphere radius

$R_g$ ~ distribution of mass in space

$$M R_g^2 = \int r(m)^2 \, dm$$

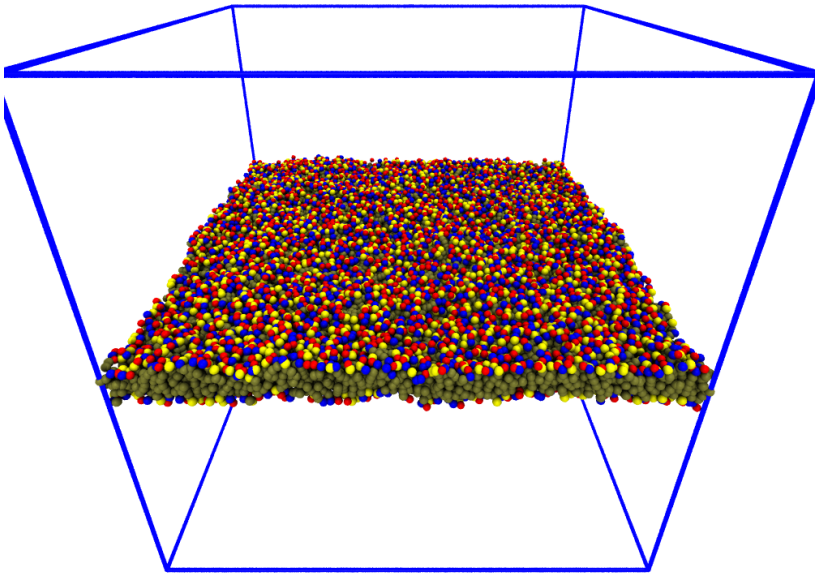




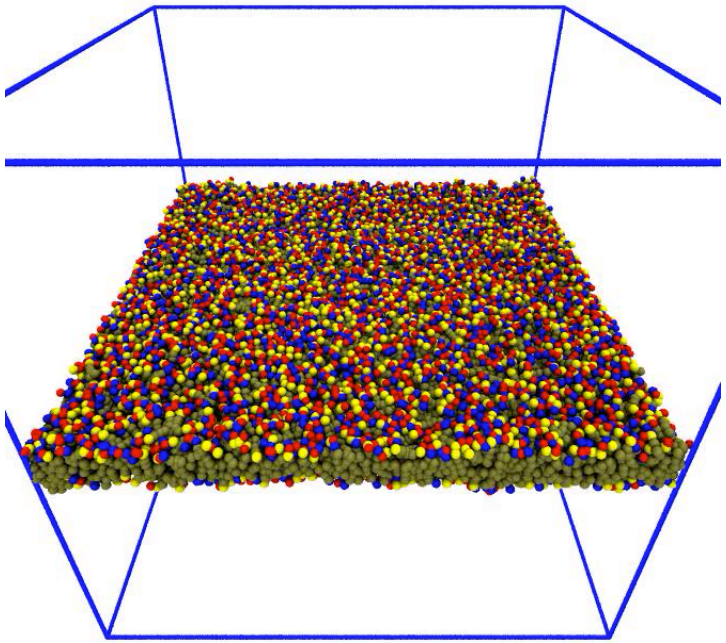


# Ex. 9 Effect of charged headgroups on lipid bilayer stability

Simulate a stable membrane with a lipid whose headgroup can be charged, then set a screened Coulomb repulsion between the headgroups on a fraction of the lipids (from 1 down to ?) and observe the effect on the membrane.



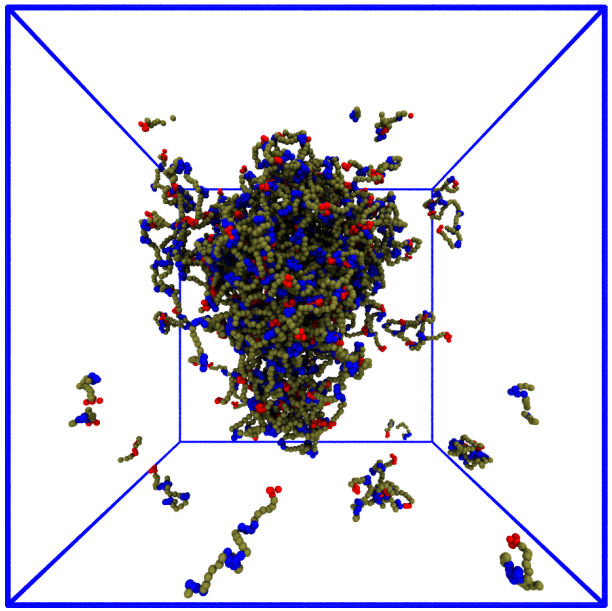
No charge



Charged

```
Polymer Water  0.98787   " (W) "
Polymer Lipid  0.01213   " (PO4 (* (COH T T T T)) NHO T T T) "

Command ChargeBeadByType        10000    1  2.8  2.0
```

# Ex. 10 What are the dense and dilute phase concentrations of a phase separated droplet of a model IDP?



Simple way

assume dense phase is a sphere, measure Rg,
and use density = mass / volume

dilute phase density = # polymers not in
dense phase / (box volume - sphere volume)

Accurate way

Monte Carlo method for estimating dense phase volume

dilute phase density = # polymers not in dense phase / ( box volume - dense volume)