2. BASIC OBSERVABLES AND SIMPLEST MODEL

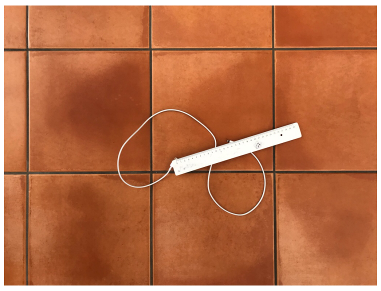
At the microscopic scale of atoms and small molecules, each polymer is different from any other. Thus, at that scale, there is no chance of having a common model that applies to all polymers.

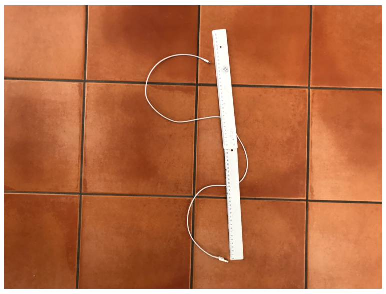
We must thus take a coarse-grained look at polymers.

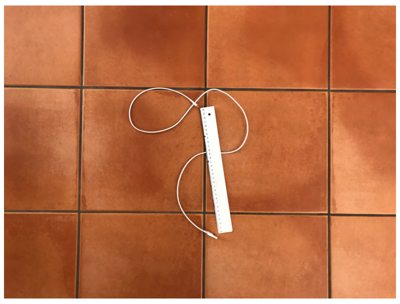
At larger scales, all polymers are 1-dimensional objects.

What are thus the properties of 1D objects that can be measured?

REMINDER: we are trying to do science, which means that we want to measure in a reproducible way.







Here above I threw a more-or-less flexible cable on the ground. Each time it takes a different conformation (it would take the same conformation if it were rigid, but then it would be rather dull).

This means that the individual conformation is not reproducible.

As well, any measure associated to conformations, such as the distance between the two ends as in the pictures, will be different.

No reproducibility, no science?

No! We have just to ask what is reproducible.

EXERCISE !!!

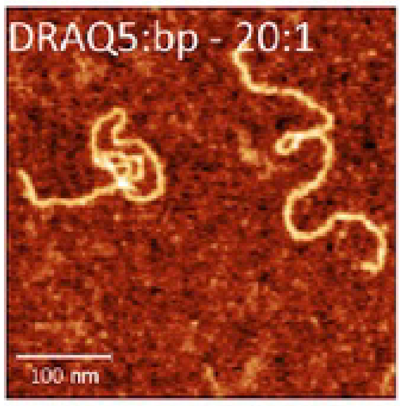
As an exercise, you can take a flexible cable and throw it on the ground 10 times, and measure the end-to-end distance for every conformation. They will all be different, of course. Take then the average. Do it again another ten times, and again take the average. The two averages are not as different. Give the cable to a friend and ask them to repeat the experiment. They should not find values too different from yours. The average is reproducible!

We conclude that for objects that take random configurations only averages are reproducible.

As a reminder, averages of quantities are associated to the moments of the probability distribution of some measurable quantity.

Probability distributions, and as a consequence averages, are reproducible.

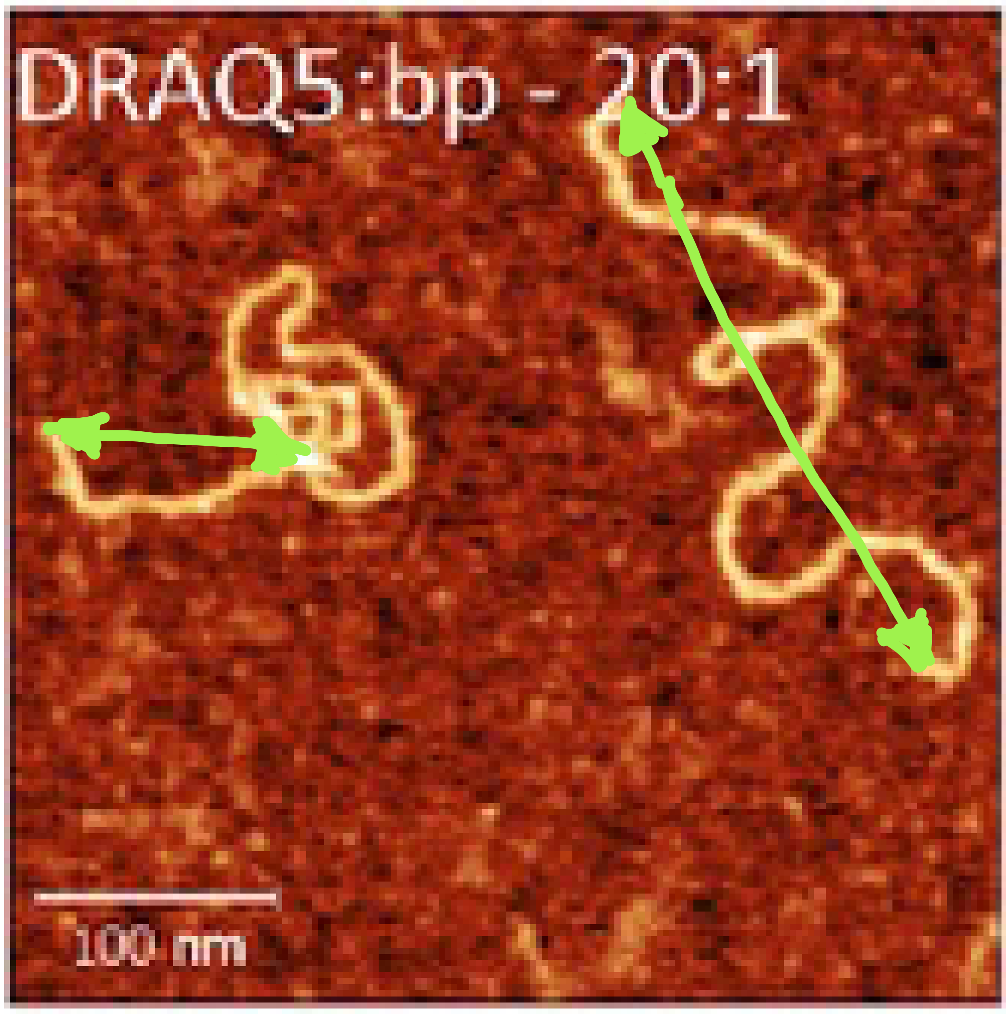
Let's go back to polymers, and use DNA (dsDNA) as a prototype. Images of DNA deposited on a substrate can be obtained using the Atomic Force Microscope (AFM)



From G. Dietler's lab at EPFL

As it can be seen, no two molecules have the same conformation, despite being identical.

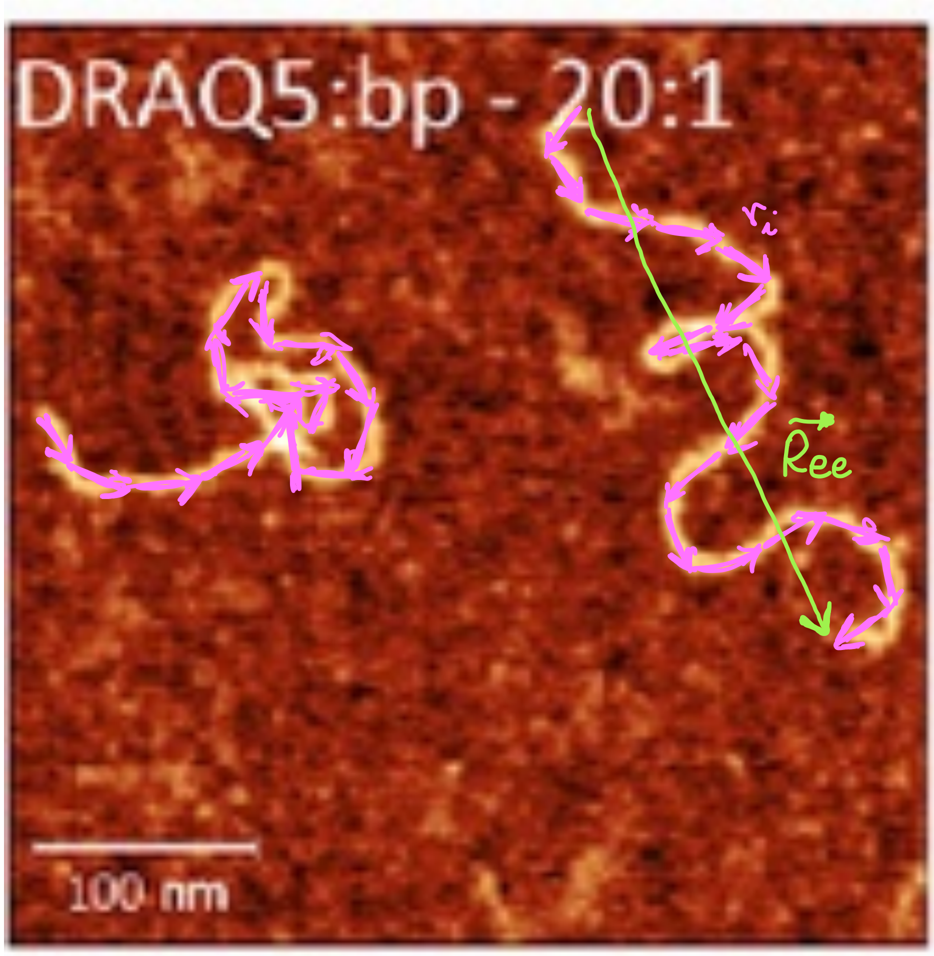
As said, we can nonetheless take the average of some measurable quantity, such as the end-to-end distance.



We can take the average there, or take the histogram $\Rightarrow$ the probability distribution. What we lack, though, is a way to "understand" the results, thus a model.

As usual, we must start from the simplest model.

THE SIMPLEST POLYMER MODEL



We can approximate the polymer as a chain of segments.

Mathematically, the polymer becomes a chain of N vectors, $\{\vec{r}_i\}_{i=1}^{N}$ each of length b:
$|\vec{r}_i| = b \quad \forall i$

Clearly, we expect $b \geq a$
↳ size of a monomer, because we consider monomers as rigid.

We can thus start writing measurable quantities by means of $\{\vec{r}_i\}$: the end-to-end vector $\vec{R}_{ee}$ is

$$\vec{R}_{ee}(\{\vec{r}_i\}) = \sum_{i=1}^{N} \vec{r}_i$$

and the end-to-end distance is

$$R_{ee}(\{\vec{r}_i\}) = |\vec{R}_{ee}(\{\vec{r}_i\})|$$

and at last we have to take its average.

What about the average ?

We can take it in two ways :

1) from experiments :

$$\bar{R}_{ee} = \frac{1}{M}\sum_{i=1}^{M} \left|\vec{R}_{ee}\left(\{\vec{r}_j^{\,i}\}\right)\right|$$

where $\{\vec{r}_j^{\,i}\}_{j=1}^{N}$ is the set of vectors associated to the M^{th} conformation.

2) from the theory :

$$\tilde{R}_{ee} = \underbrace{\int d\vec{r}_1 \dots \int d\vec{r}_N}_{\substack{\text{multiple}\\ \text{integral over all}\\ \text{possible values of the}\\ \text{set of vectors } \{\vec{r}_i\}}} \left|\vec{R}_{ee}\left(\{\vec{r}_i\}\right)\right| \underbrace{P\left(\{\vec{r}_i\}\right)}_{\substack{\downarrow\\ \text{probability distribution of}\\ \text{the set of vectors (and thus of}\\ \text{the corresponding configuration)}}}$$

[remember that $\int d\vec{r}_1 \dots \int d\vec{r}_N \, P(\{\vec{r}_i\}) = 1$]

Here we have thus two problems : we need to know $P(\{\vec{r}_i\})$ and we need to be able to compute the average of $|\vec{R}_{ee}|$.

The first issue is where the choice of the model is crucial.

The second is tricky, and we start from it.

How to take the average of $|\vec{R}_{ee}|$

We recall that

$\vec{R}_{ee} = \sum_{i=1}^{N} \vec{r}_i$, thus $\bar{R}_{ee} = \langle |\vec{R}_{ee}| \rangle = ?$

$\langle \cdot \rangle$ indicates the average taken using the probability distribution

averages are linear operations, they commute with sums. But the absolute value $|\vec{R}_{ee}|$ makes things not working.
To be precise: they nicely work in experiments and simulations, but the absolute value does not help analytical derivations.

What are the alternatives?

1) $\langle \overset{\text{vector}}{\vec{R}_{ee}} \rangle$: This does not work, because if space is isotropic, then $\vec{R}_{ee}$ can be oriented in any direction with the same probability.
Isotropy thus implies that $\langle \vec{R}_{ee} \rangle = 0$, without doing any calculations

2) Since $|\vec{R}_{ee}| = \sqrt{\vec{R}_{ee} \cdot \vec{R}_{ee}} = \sqrt{\vec{R}_{ee}^{\,2}}$, maybe we could compute

$$\langle \vec{R}_{ee}^{\,2} \rangle = \left\langle \left(\sum_{i=1}^{N} \vec{r}_i \right) \cdot \left(\sum_{j=1}^{N} \vec{r}_j \right) \right\rangle$$

Let's do this calculation explicitely.

$$\langle \vec{R}_{ee}^{\,2} \rangle = \left\langle \left(\sum_{i=1}^{N} \vec{r}_i \right) \cdot \left(\sum_{j=1}^{N} \vec{r}_j \right) \right\rangle =$$

$$= \left\langle \sum_{i=1}^{N} \sum_{j=1}^{N} \vec{r}_i \cdot \vec{r}_j \right\rangle = \sum_{i=1}^{N} \sum_{j=1}^{N} \langle \vec{r}_i \cdot \vec{r}_j \rangle =$$

here I split the sum in $i=j$ (N terms) and $i \neq j$ (the remaining $N^2 - N$ terms)

$$= \sum_{i=1}^{N} \langle \vec{r}_i^{\,2} \rangle + \sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N} \langle \vec{r}_i \cdot \vec{r}_j \rangle$$

We set $|\vec{r}_i| = b \quad \forall i$, thus

$$\langle \vec{R}_{ee}^{\,2} \rangle = \sum_{i=1}^{N} b^2 + \sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N} \langle \vec{r}_i \cdot \vec{r}_j \rangle$$

which leads us to the fundamental equation

$$\langle \vec{R}_{ee}^{\,2} \rangle = b^2 N + \underbrace{2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N}}_{} \langle \vec{r}_j \cdot \vec{r}_i \rangle$$

Important equation that is going to be with us for long!

check that this is the same as

$$\sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N}$$

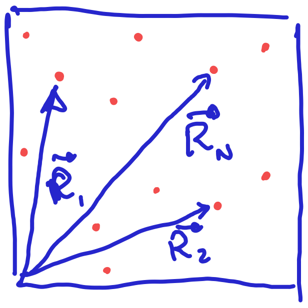
CRUCIAL: THIS equation is always correct, no matter the model we are going to choose. What we do not know, and depends on the model, is the value of $\langle \vec{r}_j \cdot \vec{r}_i \rangle$ for each i,j pair.

Let us thus start and choose the simplest model.

In this, we can be inspired by the simplest gas model: the perfect gas.

In the perfect gas, molecules do not interact, that is, they are completely independent.

To show this, we use the Boltzmann probability:



N identical molecules in volume V

The probability of a given configuration $\{\vec{R}_i\}_{i=1}^{N}$ is

$$P(\{\vec{R}_i\}) = \underbrace{\frac{1}{N!\,\lambda^{3N}} \; \frac{e^{-U(\{\vec{R}_i\})/k_BT}}{Z}}_{\hookrightarrow \text{ see Supplementary Notes \#1}}$$

Z is the partition function, necessary to ensure that

$$\underbrace{\int_V \int_V \cdots \int_V}_{N \text{ integrals}} P(\{\vec{R}_i\})\, d\vec{R}_1 \ldots d\vec{R}_N = 1$$

$$\Rightarrow \qquad Z = \frac{1}{N!\,\lambda^{3N}} \int_V d\vec{R}_1 \ldots \int_V d\vec{R}_N \; e^{-U(\{\vec{R}_i\})/k_BT} = \frac{1}{N!\,\lambda^{3N}}\, Q_N$$

and $U(\{\vec{R}_i\})$ is the interaction potential between the N molecules.

Putting everything together we have $\qquad \beta = 1/k_BT$

$$P(\{\vec{R}_i\}) = \frac{e^{-\beta U(\{\vec{R}_i\})}}{Q_N}$$

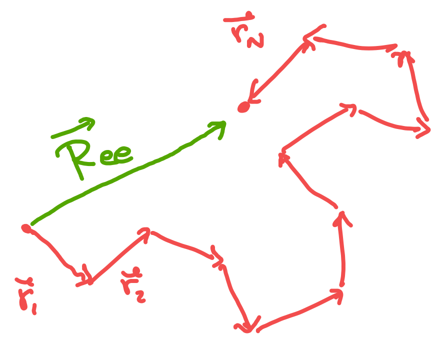
In the perfect gas we have $U(\{\vec{R}_i\})=0$, because there are no interactions whatsoever between molecules.

Thus $P(\{\vec{R}_i\}) = \frac{1}{Q_N}$, which means that all configurations have the same probability.

We can use the same approach here, and set for no interactions between the monomers, apart, of course, for the ones that are necessary to keep monomers in the chain.

In our model, thus, the chain is unbreakable, but otherwise the monomers do not interact.

We can thus draw the chain as



and, since no energy is used to bias the configurations toward any particular shapes, each bond (vector) is completely free to turn around the previous one.

Because of this, the model is also known as FREELY-JOINTED-CHAIN (FJC). Using this model we can compute the average $\langle \vec{r}_j \cdot \vec{r}_i \rangle$.

What is the probability of a given configuration $\{\vec{r}_i\}$?

$$P(\{\vec{r}_i\}) = p(\vec{r}_1)p(\vec{r}_2)\dots p(\vec{r}_N)$$

because each bond is independent (freely-jointed!) from the others, and as a consequence the probability factorizes:

$p(x,y) = p(x)p(y)$ if x,y are independent

For the FJC we do not really need to know the precise form of $p(\vec{r}_i)$. Indeed

$$\langle \vec{r}_j \cdot \vec{r}_i \rangle = \int d\vec{r}_1 \dots d\vec{r}_N \; \vec{r}_j \cdot \vec{r}_i \; p(\vec{r}_1) \dots p(\vec{r}_N) =$$

$$= \underbrace{\int d\vec{r}_1 \, p(\vec{r}_1)}_{1} \underbrace{\int d\vec{r}_2 \, p(\vec{r}_2)}_{1} \dots \int d\vec{r}_i d\vec{r}_j \, (\vec{r}_j \cdot \vec{r}_i) \, p(\vec{r}_i) p(\vec{r}_j) \dots \underbrace{\int d\vec{r}_N \, p(\vec{r}_N)}_{1} =$$

$$= \int d\vec{r}_i \, d\vec{r}_j \; \vec{r}_i \cdot \vec{r}_j \; p(\vec{r}_i) p(\vec{r}_j) = \left[\int d\vec{r}_i \, p(\vec{r}_i) \, \vec{r}_i \right] \cdot \left[\int d\vec{r}_j \, p(\vec{r}_j) \, \vec{r}_j \right] =$$

$$= \langle \vec{r}_i \rangle \cdot \langle \vec{r}_j \rangle$$

$$\Rightarrow \quad \langle \vec{r}_i \cdot \vec{r}_j \rangle = \langle \vec{r}_i \rangle \cdot \langle \vec{r}_j \rangle$$

This result mirrors $\langle xy \rangle = \langle x \rangle \langle y \rangle$ for independent variables

But each $\vec{r}_i$ can be oriented in any direction with the same probability (space is isotropic unless stated otherwise), and as a consequence

$$\langle \vec{r}_i \rangle = 0 \quad \forall i$$

In conclusion: $\langle \vec{r}_i \cdot \vec{r}_j \rangle = 0$ if $i \neq j$ and the fundamental equation

$$\langle \vec{R}_{ee}^{\,2} \rangle = b^2 N + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \underbrace{\langle \vec{r}_j \cdot \vec{r}_i \rangle}_{=0} = b^2 N$$

And defining the average end-to-end distance $\bar{R}_{ee}$ as

$$\bar{R}_{ee} = \sqrt{\langle \vec{R}_{ee}^{\,2} \rangle} = b N^{1/2}$$

we identify that $\bar{R}_{ee}$ increases algebraically with N, with exponent $1/2$. This exponent, moreover, does not depend on dimension of the system: it is the same in $d=1,2,3,4,\dots$

Probability distribution of the end-to-end distribution of the FJC:

We can perform the calculation of $\langle \vec{R}_{ee}^{\,2} \rangle$ without knowing the explicit form of $p(\vec{r}_i)$, just relying on isotropy.
If instead we want to compute the probability distribution of $\vec{R}_{ee}$, namely $P(\vec{R}_{ee})$, we will need more information about $p(\vec{r}_i)$. This notwithstanding, something will be possible at least in the limit $N \gg 1$.

First recall once more that

$$P(\{\vec{r}_i\}) = p(\vec{r}_1)\,p(\vec{r}_2)\,\dots\,p(\vec{r}_N)$$

How do we compute $P(\vec{R}_{ee})$? For every value of $\bar{R}_{ee}$, we sum up the probability of all configurations $\{\vec{r}_i\}$ such that

$$\sum_i \vec{r}_i = \vec{R}_{ee}$$

The mathematical way to do so is by means of a constraint, that we express as $\underset{\uparrow}{\delta}(\vec{R}_{ee} - \sum_i \vec{r}_i)$

Dirac delta distribution

Remember the properties of $\delta(x-x_0)$:

$$\int_{\mathcal{D}} \delta(x-x_0)\,dx = 1 \qquad \text{and} \qquad \int_{\mathcal{D}} f(x)\,\delta(x-x_0)\,dx = f(x_0)$$

$x_0 \in \mathcal{D}$

The Dirac delta can be generalized to higher dimensions :

$$\int_V \delta(\vec{r}-\vec{r}_0)\, d\vec{r} = 1 \qquad \text{and} \qquad \int_V f(\vec{r})\, \delta(\vec{r}-\vec{r}_0)\, d\vec{r} = f(\vec{r}_0)$$

$\vec{r}_0 \in V$

So we can write

$$\underbrace{P(\vec{R}_{ee}) = \int d\vec{r}_1 \int d\vec{r}_2 \ldots \int d\vec{r}_N \; P(\{\vec{r}_i\})\; \delta\Big(\vec{R}_{ee} - \sum_i \vec{r}_i\Big)}_{\text{This is true for any polymer model !!!}}$$

In the case of the FJC we can write

$$P_{FJC}(\vec{R}_{ee}) = \int d\vec{r}_1 \ldots \int d\vec{r}_N \; p(\vec{r}_1) \ldots p(\vec{r}_N)\; \delta\Big(\vec{R}_{ee} - \sum_{i=1}^{N} \vec{r}_i\Big)$$

How do we perform the calculation? We use the Fourier transform :

$$\tilde{P}(\vec{k}) = \int d\vec{R}_{ee}\; e^{-i\vec{k}\cdot\vec{R}_{ee}}\; P(\vec{R}_{ee})$$

with this definition, the anti-transform reads

$$P(\vec{R}_{ee}) = \frac{1}{(2\pi)^d} \int d\vec{k}\; e^{i\vec{k}\cdot\vec{R}_{ee}}\; \tilde{P}(\vec{k})$$

So, let's take the Fourier transform : $\mathcal{F}[P_{FJC}(\vec{R}_{ee})]$

$$\tilde{P}_{FJC}(\vec{k}) = \int d\vec{R}_{ee}\; e^{-i\vec{k}\cdot\vec{R}_{ee}} \int d\vec{r}_1 \ldots d\vec{r}_N \; p(\vec{r}_1) \ldots p(\vec{r}_N)\; \delta\Big(\vec{R}_{ee} - \sum_{i=1}^{N} \vec{r}_i\Big) =$$

$$= \int d\vec r_1 \dots d\vec r_N \, p(\vec r_1) \dots p(\vec r_N) \int d\vec R_{ee} \; e^{-i\vec k \cdot \vec R_{ee}} \; \delta\Big(\vec R_{ee} - \sum_i \vec r_i\Big) =$$

$$= \int d\vec r_1 \dots d\vec r_N \; p(\vec r_1) \dots p(\vec r_N) \; e^{-i\vec k \cdot \sum_{i=1}^{N} \vec r_i} =$$

$$= \int d\vec r_1 \dots d\vec r_N \Big[p(\vec r_1) e^{-i\vec k \cdot \vec r_1}\Big]\Big[p(\vec r_2) e^{-i\vec k \cdot \vec r_2}\Big] \dots \Big[p(\vec r_N) e^{-i\vec k \cdot \vec r_N}\Big] =$$

$$= \int d\vec r_1 \Big[p(\vec r_1) e^{-i\vec k \cdot \vec r_1}\Big] \int d\vec r_2 \Big[p(\vec r_2) e^{-i\vec k \cdot \vec r_2}\Big] \dots \Big[\int d\vec r_N \; p(\vec r_N) e^{-i\vec k \cdot \vec r_N}\Big] =$$

$$= \prod_{i=1}^{N} \int d\vec r_i \Big[p(\vec r_i) e^{-i\vec k \cdot \vec r_i}\Big] = \prod_{i=1}^{N} \tilde p(\vec k)$$

If we assume that all monomers and their connecting bonds are roughly identical, then $p(\vec r_i) = p(\vec r)$ $\forall i$ and

$$\tilde P_{FJC}(\vec k) = \big[\tilde p(\vec k)\big]^N$$

Then we can anti-transform

$$P_{FJC}(\vec R_{ee}) = \frac{1}{(2\pi)^d} \int d\vec k \; e^{i\vec k \cdot \vec R_{ee}} \big[\tilde p(\vec k)\big]^N =$$

$$= \frac{1}{(2\pi)^d} \int d\vec k \; e^{i\vec k \cdot \vec R_{ee}} \; e^{N \underline{\ln \tilde p(\vec k)}}$$

Let's work on this

$$\ln \tilde p(\vec k) = \ln\Big[\int d\vec r \, e^{-i\vec k \cdot \vec r} p(\vec r)\Big] \underset{\substack{\uparrow \\ \text{expand} \\ \text{the exponential}}}{=} \ln\Big[\sum_{l=0}^{\infty} \int d\vec r \, \frac{(-i\vec k \cdot \vec r)^l}{l!} p(\vec r)\Big] =$$

$$= \ln\left[\underbrace{\int d\vec{r}\, p(\vec{r}) \frac{(-i\vec{k}\cdot\vec{r})^0}{0!}}_{=1} + \sum_{\ell=1}^{\infty} \int d\vec{r} \frac{(-i\vec{k}\cdot\vec{r})^\ell}{\ell!} p(\vec{r}) \right] =$$

$$= \ln\left[1 + \sum_{\ell=1}^{\infty} \frac{(-i)^\ell}{\ell!} \int (\vec{k}\cdot\vec{r})^\ell \, p(\vec{r}) d\vec{r} \right]$$

The term for $\ell = 1$ is 0 if space is isotropic:

$$\int \vec{k}\cdot\vec{r}\, p(\vec{r})\, d\vec{r} = \vec{k} \cdot \underbrace{\int \vec{r} p(\vec{r}) d\vec{r}}_{\text{by isotropy } = 0} = 0$$

The sum thus starts from $\ell = 2$.

Moreover, for simplicity, we set $|\vec{r}| = b$ (true in our models, not in general). Then

$$\left\{ \begin{array}{ll} d\varphi & D=2 \\ \sin\theta d\theta d\varphi & D=3 \\ \cdots & \end{array} \right.$$

$$\ln\left[1 - \frac{1}{2} b^2 k^2 \int (\cos\theta_1)^2 p(\{\theta_i\}_{i=1}^{d-1}) \overbrace{J(\{\theta_i\})\, d\theta_1 \ldots d\theta_{d-1}}^{\text{angular volume of integration}} + \right.$$

$$\left. + \sum_{\ell=3}^{\infty} \frac{(-i)^\ell}{\ell!} b^\ell k^\ell \int (\cos\theta)^\ell p(\vec{r})\, d\vec{r} \right] =$$

$$= \ln\left[1 - \frac{1}{2} b^2 k^2 c_2 + \sum_{\ell=3}^{\infty} \frac{(-i)^\ell}{\ell!} b^\ell k^\ell c_\ell \right]$$

Let's plug this back in the anti-transform:

$$P_{FJC}(\vec{R}_{ee}) = \frac{1}{(2\pi)^d} \int d\vec{k}\, e^{i\vec{k}\cdot\vec{R}_{ee}}\, e^{N \ln\left[1 - \frac{1}{2} b^2 k^2 c_2 + \sum_{\ell=3}^{\infty} \frac{(-i)^\ell}{\ell!} b^\ell k^\ell c_\ell\right]}$$

Make a change of variable: $\vec{q} = \sqrt{N} b \vec{k}$ $\qquad d\vec{q} = N^{d/2} b^d d\vec{k}$

$$\Rightarrow P_{FJC}(\vec{R}_{ee}) = \frac{1}{[2\pi b N^{1/2}]^d} \int d\vec{q}\, e^{i\vec{q}\frac{\vec{R}_{ee}}{b\sqrt{N}}}\, e^{N \ln\left[1 - \frac{1}{2N} c_2 q^2 + \sum \frac{(-i)^\ell}{\ell!} \frac{c_\ell}{N^{\ell/2}} q^\ell\right]}$$

In the very large N limit, the terms of the sum decrease faster than the term in q^2, and we can neglect them. Also, for N large enough, the term in q^2 is going to be smaller than 1 (although it depends on q), and we can thus expand

$$\ln(1-x) \simeq -x$$

$$P_{FJC}(\vec{R}_{ee}) \simeq \frac{1}{\left[2\pi\, b N^{1/2}\right]^d} \int d\vec{q}\; e^{i \vec{q}\, \frac{\vec{R}_{ee}}{bN^{1/2}}} \; e^{-\frac{1}{2} c_2 q^2}$$

Rather than doing the full calculation, just remember that the Fourier transform of a Gaussian is also a Gaussian, thus

$$P_{FJC}(\vec{R}_{ee}) = \underbrace{\mathcal{N}_d^{-1}}_{\text{d-dimensional normalisation}} e^{-a \frac{\vec{R}_{ee}^{\,2}}{b^2 N}}$$

and we just need to compute a (which of course is related to c_2)

We already know that

$$\langle \vec{R}_{ee}^{\,2} \rangle = b^2 N$$

Then we compute

$$\langle \vec{R}_{ee}^{\,2} \rangle = \mathcal{N}_d^{-1} \int dx_1 \ldots dx_d \; e^{-\frac{1}{2} a \frac{\overbrace{x_1^2 + \ldots x_d^2}^{=\vec{R}_{ee}^{\,2}}}{b^2 N}} \underbrace{\left(x_1^2 + x_2^2 + \ldots x_d^2\right)}_{\vec{R}_{ee}^{\,2}} =$$

$$= \sum_{l=1}^{d} \mathcal{N}_d^{-1} \int dx_1 \dots dx_d \; e^{-\frac{1}{2}a \frac{x_1^2 + \dots x_d^2}{b^2 N}} \; x_l^2 =$$

$$= \sum_{l=1}^{d} \mathcal{N}_d^{-1} (\mathcal{N}_1)^{d-1} \underbrace{\int dx_l \, x_l^2 \, e^{-\frac{1}{2}a \frac{x_l^2}{b^2 N}}}_{} =$$

does not depend on l, thus the result is $\frac{b^2 N}{a} \mathcal{N}_1^d$

$$= d \, \mathcal{N}_d^{-1} \mathcal{N}_1^d \frac{b^2 N}{a}$$

We also know that

$$\mathcal{N}_d = \int dx_1 \dots dx_d \; e^{-\frac{1}{2}a \frac{x_1^2 + \dots + x_d^2}{b^2 N}} = \left[\int dx \; e^{-a \frac{x^2}{b^2 N}} \right]^d = \mathcal{N}_1^d$$

$$\mathcal{N}_1 \downarrow \; = \sqrt{2\pi \frac{b^2 N}{a}}$$

So we find

$$\begin{tikzcd}
 & b^2 N \arrow[dr] & & \\
\langle \vec{R}_{ee}^2 \rangle = \arrow[ur] \arrow[dr] & & {} \Rightarrow & a = d \\
 & \frac{d}{a} b^2 N \arrow[ur] & &
\end{tikzcd}$$

Finally we have

$$P_{FJC}(\vec{R}_{ee}) \underset{N \gg 1}{\simeq} \left(\frac{d}{2\pi b^2 N} \right)^{d/2} e^{-\frac{d \vec{R}_{ee}^2}{2 b^2 N}}$$