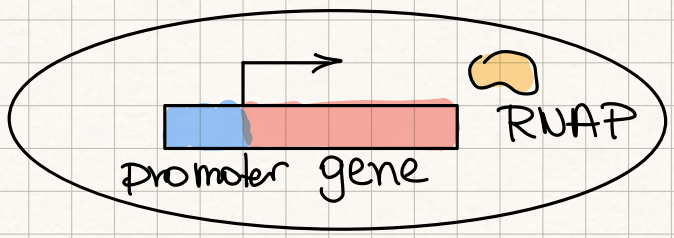
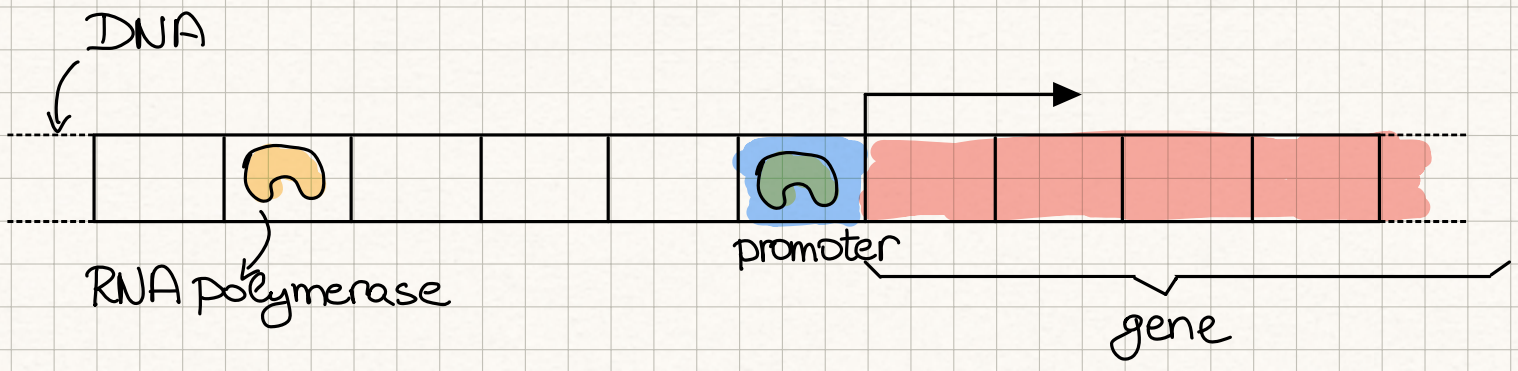
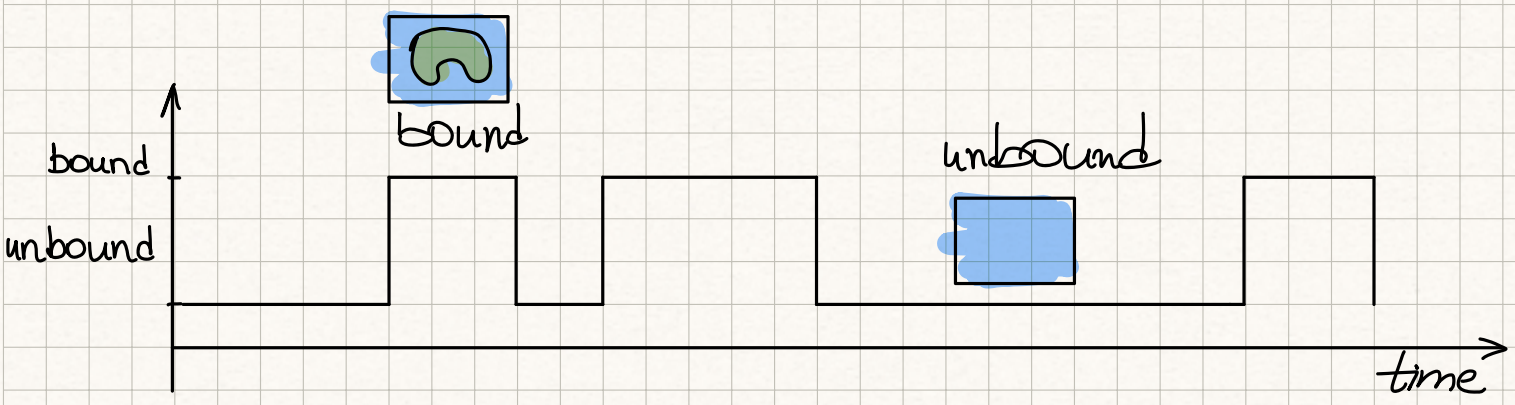
STATISTICAL MECHANICS OF GENE EXPRESSION



RNAP : polymerase





Rate of RNA production = p_{bound} r_0 → rate of transcription when RNAP bound
↳ fraction of time RNAP bound

$$\text{Amount of RNA} = \frac{\text{Rate of RNA production}}{\text{Rate of RNA degradation}}$$

Amount of protein = T × Amount of RNA
↳ in E.coli $T \approx 10^3$, i.e. about a 1000 proteins per mRNA.

So, what we wish to know is:

How does p_{bound} depend on binding strength of RNAP to promoter DNA, and on number of polymerases?

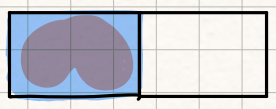
Now we make the key assumption: P_{bound} is determined by Boltzmann's law, or in other words, binding is in equilibrium!

(One can make various arguments why this might be true, but at the end of the day Nature (experiments) are the final arbiter!

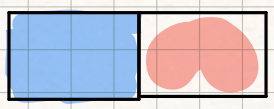
1. One promoter site + one non-specific site

$N_{NS} = 2$

$P = 1$ (# of RNAPs)



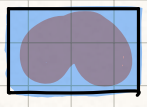
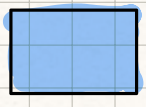
RNAP bound



RNAP not bound

$P_{bound} = ?$

Boltzmann formula

state	probability	energy
	P_{bound}	ε_S / specific
	$P_{not\ bound}$	ε_{NS} / non-specific

specific binding is stronger

$\varepsilon_S < \varepsilon_{NS}$

↓

Joules

$$\text{Boltzmann} \quad \frac{P_{bound}}{P_{unbound}} = \frac{e^{-\varepsilon_S/k_BT}}{e^{-\varepsilon_{NS}/k_BT}}$$

Boltzmann factors

↓

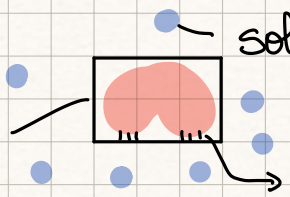
Fundamental formula of Statistical Mechanics

k_BT = thermal energy @ room temperature = 300K

$k_BT = 1.38 \times 10^{-23}\ J/K \cdot 300K = 4 \cdot 10^{-21}\ J$

↳ typical energy of motion of a molecule @ temperature T.

Example:



typical non-specific binding energy

$\varepsilon_{NS} = -2\,k_BT$ ←

$e^{-\varepsilon_{NS}/k_BT} = e^2 \approx 7$

$\frac{P_{bound}}{P_{unbound}} = \frac{e^2}{e^0} \approx 7$

Probability that RNAP is bound is 7 × greater than it not being bound!

General:

$P_{unbound} = 1 - P_{bound}$ and $\frac{P_{bound}}{P_{unbound}} = \frac{e^{-\varepsilon_S/k_BT}}{e^{-\varepsilon_{NS}/k_BT}} = e^{-\Delta\varepsilon_P/k_BT}$

$(\Delta\varepsilon_P \equiv \varepsilon_S - \varepsilon_{NS})$

$\frac{P_{bound}}{1-P_{bound}} = e^{-\Delta\varepsilon_P/k_BT}$

$\Rightarrow P_{bound} = e^{-\Delta\varepsilon_P/k_BT} - P_{bound}e^{-\frac{\Delta\varepsilon_P}{k_BT}}$

$(1 + e^{-\Delta\varepsilon_P/k_BT})P_{bound} = e^{-\Delta\varepsilon_P/k_BT}$

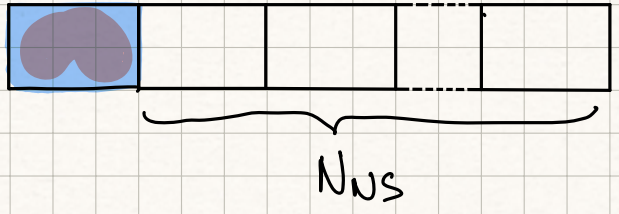
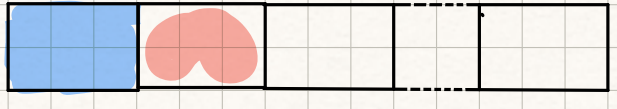
$$P_{bound} = \frac{e^{-\Delta\varepsilon_P/k_BT}}{1 + e^{-\Delta\varepsilon_P/k_BT}}$$

With the numbers in the example ($\Delta\varepsilon_P = -2k_BT$):

$P_{bound} = 7/8$ ($P_{unbound} = 1/8$)

2. One promoter site + $N_{NS} = 5\times10^6$ non-specific binding sites

(Still $P = 1$ i.e. one polymerase)

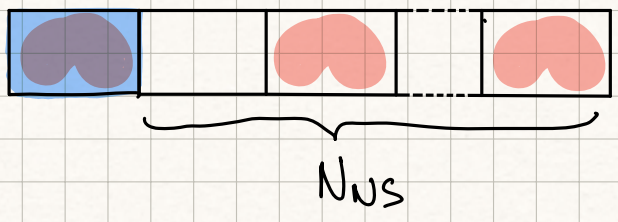
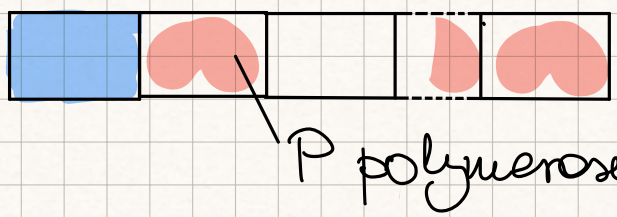
state	energy	multiplicity	weight
	ε_S	1	$1 \times e^{-\varepsilon_S/k_BT}$
	ε_N	N_{NS}	$N_{NS} \times e^{-\varepsilon_{NS}/k_BT}$

$$P_{bound} = \frac{e^{-\varepsilon_S/k_BT}}{e^{-\varepsilon_S/k_BT} + N_{NS}e^{-\varepsilon_{NS}/k_BT}}$$

$$P_{bound} = \frac{\frac{1}{N_{NS}} e^{-\Delta\varepsilon_P/k_BT}}{\frac{1}{N_{NS}} e^{-\Delta\varepsilon_P/k_BT} + 1}$$

after dividing both numerator and denominator by $N_{NS}\, e^{-\varepsilon_{NS}/k_BT}$

3. One promoter site, N_{NS} ($\approx 5\times10^6$) non-specific sites, and $P>1$ (≈ 3000)

state	energy	multiplicity	weight
	$\varepsilon_S + (P-1)\varepsilon_{NS}$	$\frac{N_{NS}!}{(P-1)!(N_{NS}-P+1)!}$	multiplicity $\times$ $e^{-\frac{\text{energy}}{k_BT}}$
	$P\,\varepsilon_{NS}$	$\underbrace{\frac{N_{NS}!}{P!\,(N_{NS}-P)!}}$ See Counting and Probability below!	multiplicity $\times$ $e^{-\frac{\text{energy}}{k_BT}}$

$$P_{bound} = \frac{\text{weight of bound state}}{\text{weight of bound state} + \text{weight of unbound state}}$$

$$P_{bound} = \frac{\text{weight of bound state} / \text{weight of unbound state}}{\text{weight of bound state} / \text{weight of unbound state} + 1}$$

From the states and weights table:

$$\frac{\text{weight of bound state}}{\text{weight of unbound state}} = \frac{\frac{N_{NS}!}{(P-1)!\,(N_{NS}-P+1)!}\, e^{-(\varepsilon_S + (P-1)\varepsilon_{NS})/k_BT}}{\frac{N_{NS}!}{P!\,(N_{NS}-P)!}\, e^{-P\varepsilon_{NS}/k_BT}}$$

$$= \frac{P!\,(N_{NS}-P)!}{(P-1)!\,(N_{NS}-P+1)!}\, e^{-(\varepsilon_S-\varepsilon_{NS})/k_BT}$$

$$= \frac{P}{N_{NS}-P+1}\, e^{-\Delta\varepsilon_P/k_BT}$$

Since $P \ll N_{NS}$ ($P \approx$ few thousand polymerases, $N_{NS} \approx$ few 10^6) we can use a simpler formula: $P/N_{NS}\, e^{-\Delta\varepsilon_p/k_BT}$

Finally, substituting the simpler formula for the ratio of weights:

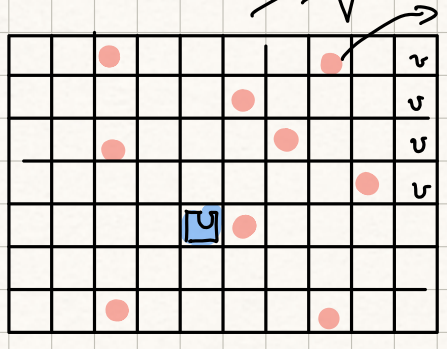
$$P_{bound} = \frac{P/N_{NS}\, e^{-\Delta\varepsilon_p/k_BT}}{P/N_{NS}\, e^{-\Delta\varepsilon_p/k_BT} + 1}$$

For $\Delta\varepsilon_p \approx -3k_BT \Rightarrow e^{-\Delta\varepsilon_p/k_BT} = e^3 \approx 20$ typical binding energy ($\Delta\varepsilon_p$) for an E.coli promoter

$$P_{bound} = \frac{\frac{3\times10^3}{5\times10^6}\cdot 20}{\frac{3\times10^3}{5\times10^6}\cdot 20 + 1} \approx \frac{10^{-2}}{10^{-2}+1}$$

$P_{bound} \approx 1/100$ (fairly small)

Note that this calculation could be used for any binding problem, where P is the # of ligands in solution (i.e. free) that can bind to receptor, N_{NS} is replaced by number of places the ligands can be in solution ($\approx V/v$ where v is the volume taken up by 1 ligand and V the volume of the solution), and $\Delta\varepsilon$ is the ligand-receptor binding energy.



$$P_{bound} = \frac{L/\Omega\, e^{-\Delta\varepsilon_r}}{L/\Omega\, e^{-\Delta\varepsilon_r} + 1}$$

where $\Omega = \frac{V}{v}$ is the # of "solution sites"

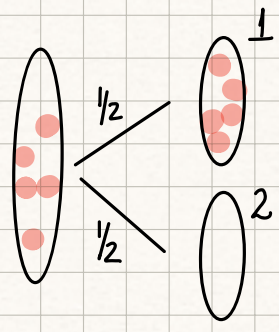
Also: $P_{bound} = \frac{L/(V/v)\, e^{-\Delta\varepsilon_r/k_BT}}{L/(V/v)\, e^{-\Delta\varepsilon_r/k_BT} + 1} = \frac{[L]/K_d}{[L]/K_d + 1}$

$[L] = L/V$ concentration of ligand

$K_d = \frac{e^{\Delta\varepsilon_r/k_BT}}{v}$: dissoc constant.

Counting and probability

1. Cell has 5 proteins. What is the chance that daughter has all 5 / the other daughter has 0)?

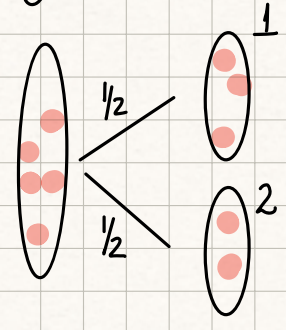


$P_5 = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$

1 1 1 1 1

$P_5 = 1/2 \times 1/2 \times 1/2 \times 1/2 \times 1/2 = \frac{1}{32}$

2. What is the probability that daughter cell 1 gets 3 of the 5 proteins?



1	1	1	2	2
1	1	2	1	2
1	2	1	1	2
2	1	1	1	2
1	1	2	2	1

1	2	1	2	1
2	1	1	2	1
1	2	2	1	1
2	1	2	1	1
2	2	1	1	1

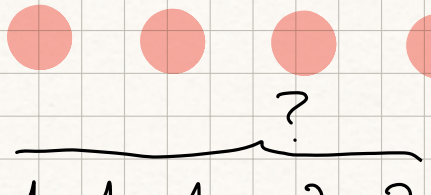
Each of these outcomes has probability

$P_{outcome} = 1/32$

# of desired outcomes = 10

$\Rightarrow P_3 = 10/32$

Counting like this gets hard (to impossible!) when the # of proteins gets large. Instead use factorials ($n! = n(n-1)\ldots$)



: 5 proteins in mother cell.

How many ways are there of assigning 3 to daugher 1 and 2 to daughter 2?

A trick that at first might seem to complicate things. Color each daughter label (1 or 2) with a different color:

1 1 1 2 2 } All 5 labels are now distinguishable

so the # of ways of assigning them to the 5 proteins:

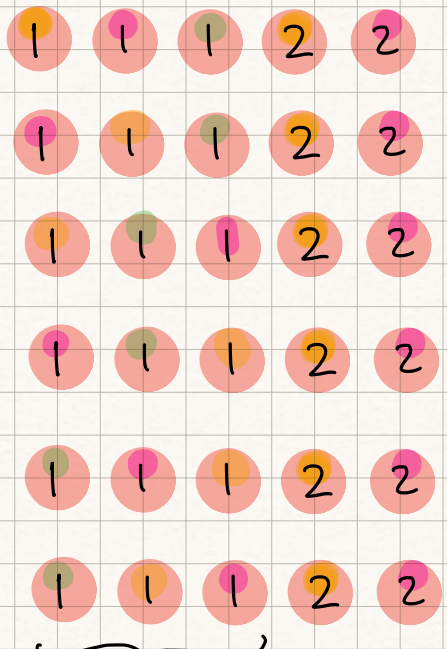
$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$

# of ways of choosing which ● gets assigned 1

# of ways of assigning 1 to a ●

etc.

This way of counting will overcount by a lot since, for example, all of these:



will be counted seperately. (If we ignore the colors on the daughter cell labels (1 and 2), which we wish to do at the end of the calculation, all these assignments are one and the same!)

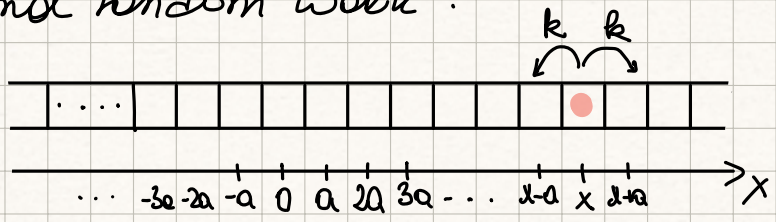
$3! = 6$ permutations of the 3 colors on 1

Therefore, we need to divide by 3! and 2!, accounting for the permutations of the color labels. Final result:

# of ways of assigning 11122 is: $\frac{5!}{3!2!} = \frac{120}{12} = 10.$

DIFFUSION

To describe diffusion mathematically we will use the one-dimensional random walk:



Diffusing particle starts at $x=0$, and hops to the left and to the right a distance a, with rate k. (Assumptions of model!)

What is the probability of finding particle @ position x at time t?

To compute $P(x,t)$ we need an update rule, which tells you:

given $P(x,t)$, i.e., whole probability distribution @ time t, what is the distribution @ time $t+\Delta t$?

$P(x, t+\Delta t) = ?$

↳ We can compute this by noticing that if the particle is at x @ time $t+\Delta t$, it had to be either:

1: @ x at time t
2: @ $x+a$ at time t
3: @ $x-a$ at time t

as long as Δt is tiny! / $\Delta t \ll \frac{1}{k}$ or $k\Delta t \ll 1$)

So:

$$P(x, t+\Delta t) = P(1) + P(2) + P(3)$$

$$= P(\text{particle @ } x \text{ AND no hop}) + P(\text{particle @ } x-a \text{ AND hop})$$

$$+ P(\text{particle @ } x+a \text{ AND hop})$$

Now, rules of probability say that $P(A \text{ AND } B) = P(A) \cdot P(B)$ as long as A and B are independent. We can use this since to hop or not to hop is independent of position (it always happens with rate k).

$P(\text{particle @ } x \text{ AND no hop}) = \underbrace{P(x,t)}_{\text{particle @ } x} \cdot \underbrace{(1-2k\Delta t)}_{\text{no hop}}$

$P(\text{particle @ } x-a \text{ AND hop}) = \underbrace{P(x-a,t)}_{\text{particle @ } x-a} \underbrace{k\Delta t}_{\text{hop to right}}$

$P(\text{particle @ } x+a \text{ AND hop}) = \underbrace{P(x+a,t)}_{\text{particle @ } x+a} \underbrace{k\Delta t}_{\text{hop to left}}$

$$\Rightarrow \boxed{P(x,t+\Delta t) = P(x,t)(1-2k\Delta t) + P(x-a,t)\,k\Delta t + P(x+a,t)\,k\Delta t}$$

↳ Update rule for 1d random walk (model of diffusion)

With a bit more work we can turn this into a differential equation:

$$P(x+a,t) = P(x,t) + a\frac{dP}{dx} + \frac{a^2}{2}\frac{d^2P}{dx^2} + \cancel{\ldots} \quad \text{stop at } 2^{nd} \text{ order}$$

$$P(x-a,t) = P(x,t) - a\frac{dP}{dx} + \frac{a^2}{2}\frac{d^2P}{dx^2} + \cancel{\ldots}$$

$$\frac{P(x,t+\Delta t) - P(x,t)}{\Delta t} = -2kP(x,t) + k\left(P(x,t) + a\frac{dP}{dx} + \frac{a^2}{2}\frac{d^2P}{dx^2}\right) + k\left(P(x,t) - a\frac{dP}{dx} + \frac{a^2}{2}\frac{d^2P}{dx^2}\right)$$

$$\boxed{\frac{dP}{dt}(x,t) = ka^2\frac{d^2P}{dx^2}(x,t)} \quad \text{Diffusion equation!}$$

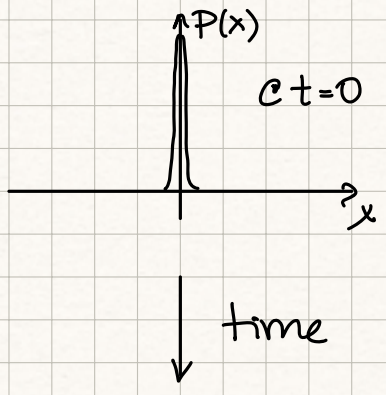
ka^2 has units of $\mu m^2/sec$ and is the diffusion constant D.

Solution of the Diffusion equation: $P(x,t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}}$

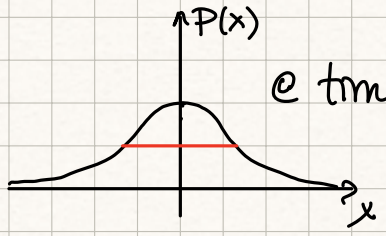
is a Gaussian whose variance: $\langle x^2 \rangle = 2Dt$ grows linearly with time

(The width of the distribution: $\sqrt{\langle x^2 \rangle} = \sqrt{2Dt}$)

Solution in pictures:



Particle starts at $x=0$: therefore $P(x,t)$ is initially very sharply peaked @ $x=0$



Width of the distribution, a.k.a., its standard deviation ($=\sqrt{\text{variance}}$) is

$$\sigma_x = \sqrt{2Dt}$$

σ_x tells us how far from the origin it is likely (with some confidence) to find the particle.

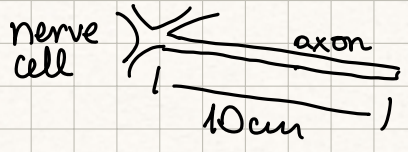
So for a protein in a cytoplasm: $D \approx 10\,\mu m^2/sec$.

Therefore, after time $t = 5$ seconds: $\sigma_x = \sqrt{2Dt}$

$$\sigma_x = \sqrt{100\,\mu m^2}$$

$$\sigma_x = 10\,\mu m$$

It is likely the particle will have diffused away over a distance of $10\,\mu m$. On cellular scales diffusion is pretty fast! (Unless we're looking at our long neurons...



$$\sigma_x = 10\,cm = 10^5\,\mu m$$

$$10^5\,\mu m = \sqrt{2D \cdot t} \Rightarrow t \approx \frac{10^{10}}{20}\,sec$$

$$t \approx 5 \times 10^8\,sec \approx 10\,years!!$$

$\Rightarrow$ Need to wait ~ 10yrs for a protein to diffuse 10cm.