

## DIVERSITY of CHEMORECEPTORS :

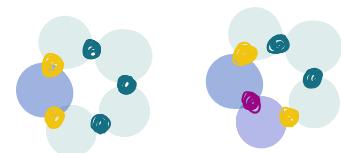
- Organisms evolved diverse receptors. Say  $r_1 \dots r_N = N$  receptors
- each ligand  $L_i$  binds each receptor with  $\neq$  affinity
  - this is useful as a way of encoding odor identity :  
e.g. in rodents, the list of receptor activation codes  
odor identity (mathematically each odor is defined by  
vector of activities  $a_1 \dots a_N$ )  
 $a_i = 0 \text{ or } 1$
  - If each receptor binds odor independently (say binary)  
 $N$  of receptors can encode up to  $2^N$  odor identities.  
→ so with 1000 receptors you can potentially  
encode for  $2^{1000} \sim 10^{300}$  odors (a lot!)  
$$\log_{10}(2^{1000}) = 1000 \underbrace{\log_{10} 2}_{0,3} \approx 300$$
  - 1000 receptors may code for already many odors
  - it looks like many organisms evolved a diversity of receptors  
Assuming diversity is good, how can you generate it?
    - (1) Mutations / selection → need 1 gene per receptor
    - (2) Assembly of receptor from subunits  
→ few genes generate many receptors
  - NOTE For this to work  $a_i$  should not be function of  $a_{j \neq i}$   
→ in this sense it's important that assembling subunits  
generates "true" diversity (mathematically:  $a_{x_i + \beta_j} \neq a_i + \beta_j$ )

• HOW LARGE is THE POTENTIAL DIVERSITY of OCTOPUS CHEMORECTORS?

If all 26 subunits can be combined at will, then you have:

26 homo-pentamers;

+  $26 \times 25$  hetero-pentamers with 2 distinct subunits  
 $\times 2 \left\{ \begin{array}{l} \text{in proportion 4 to 1} \\ = 3+2 \end{array} \right.$



+  $26 \times 25 \times 24$  heteropentamers w 3 ≠ subunits in proportion  
 $\times 2 \left\{ \begin{array}{l} \text{◦ 1-2-2} \\ \text{◦ 1-1-3} \end{array} \right.$



+  $26 \times 25 \times 24 \times 23$  heteropentamers w 4 ≠ subunits  
 $(1-1-1-2)$

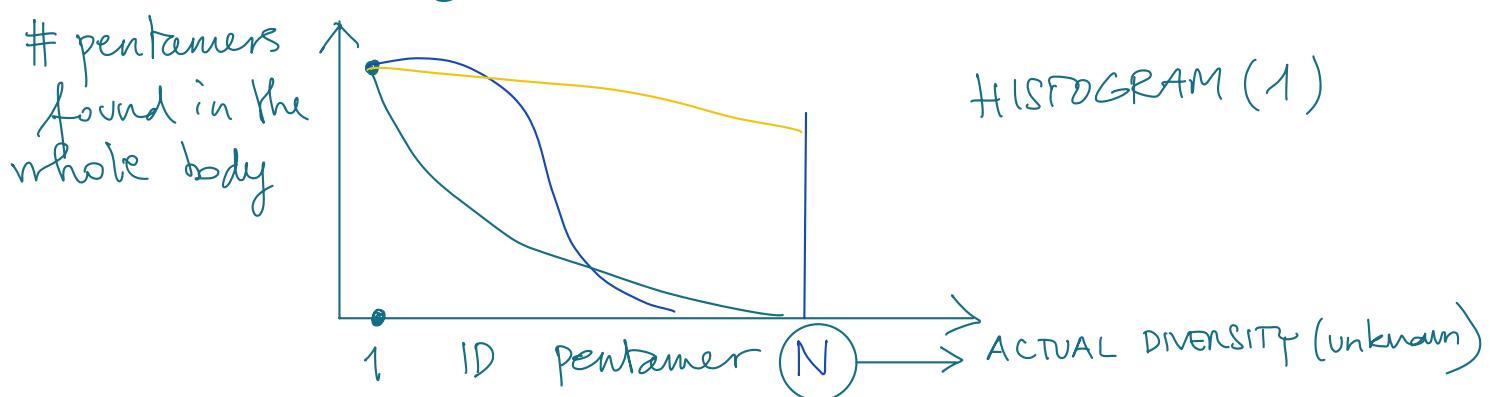
+  $26 \times 25 \times 24 \times 23 \times 22$  heteropentamers w 5 ≠ subunits

$$= \text{GRAND TOTAL} = 8.284.926 \quad (\text{compare with cows} \sim 4k)$$

• WHAT IS THE ACTUAL DIVERSITY we can infer from data?

Suppose that you can sequence each single pentamer in an individual, & then ID the pentamer from the most abundant ( $ID=1$ ) to the least abundant ( $ID=N$ )

Then their histogram may look like any of these:



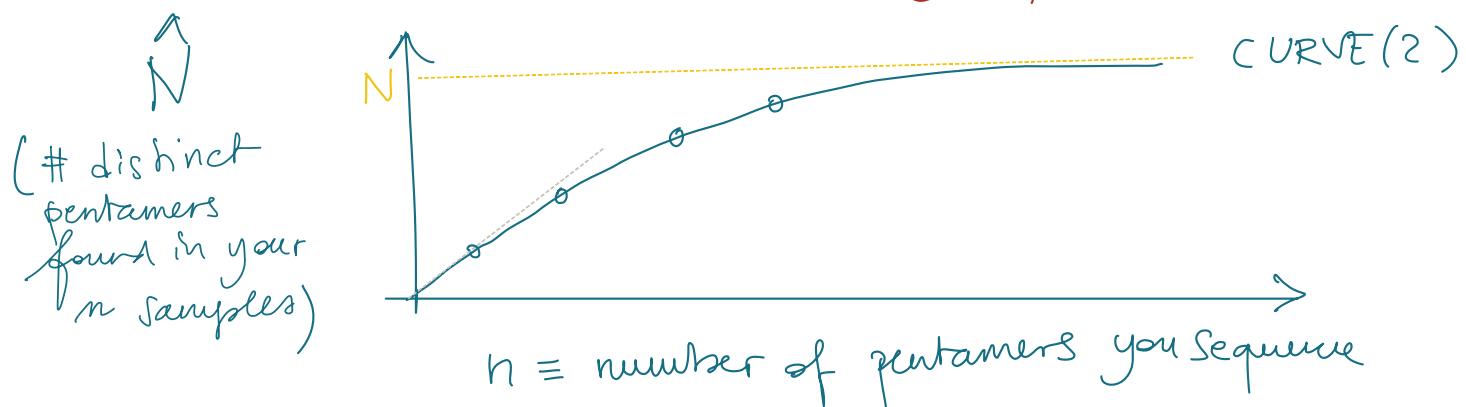
- If you sequence  $n$  random pentamers from the whole animal, you'll find  $\hat{N}$  distinct pentamers:

$\hat{N}$  grows with  $n$ , when  $n$  is small

↓ cannot be larger than the total #  $N$  in the animal

So the curve for sample diversity  $\hat{N}$  as a function of how many receptors you sequence will look like this:

PABLO: can you plot this?



If you sample more & more pentamers, we may be able to see CURVE(2) bend, and infer where it saturates, which provides an estimate of  $N$ .

NOTE: The estimate will be affected by error, because we don't know how HISTOGRAM(1) looks like

# • Does heterogeneity functionally create diversity?

• Molecularly, for homopentamers we know binding site is @ the interface between two subunits

→ suggests that combining ≠ subunits will modify the binding site hence the binding activities.

PABLO: is there a way to confirm with heteropentamers that binding site is @ interface?

• Computationally.

To Do: write Monod-Wyman-Changeux model of allostery for receptor with 5 binding sites

GOAL: make a prediction for null model, where assembly of ≠ subunits does not contribute diversity:

activation homopentamer with 5 subunits of kind i

no diversity from assembly of ≠ subunits

$$\Leftrightarrow \alpha_{\alpha_i + \beta_j} \propto f(\alpha_i + \beta_j)$$

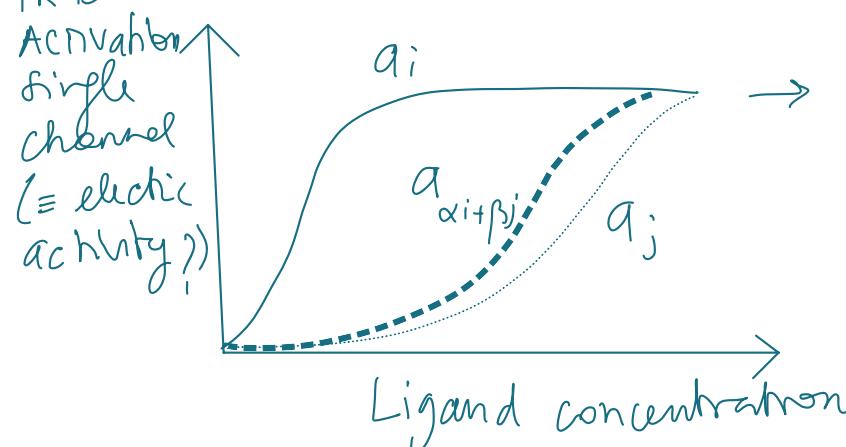
| if f is the same for all ligands

↓  
no diversity

activation of hetero-pentamer with  $\alpha$  subunits of type i &  $\beta$  " " " j

activation heteropent. with 5 subunits of type j

PROB



→ fit  $\alpha, \beta$ , then use to check for other ligands!

PABLO, if you give us the dose-response curve for your 5 ligands, we can see if they are diverse.

Single binding site:  $\theta = \frac{[L]^n}{K_d + [L]^n} = \frac{1}{1 + (\frac{k_A}{[L]})^n}$

Hill-Monod eqn (basically Langmuir)

fraction of receptors that are bound to ligand ( $\equiv$  probability that 1 ligand will bind the receptor)

ligand concn. product half occupancy

Hill coeff

$n > 1 \rightarrow$  cooperative binding } related to how binding sites cooperate.  
 $n < 1 \rightarrow$  inhibition

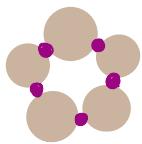
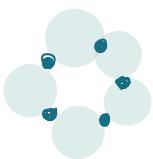
Or

$$Y = \frac{[L]^n}{K_d + [L]^n}$$

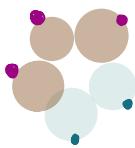
fraction of occupied binding sites

ligand conc in  $\mu M$

$K_d$  dissociation constant : concentration  $\theta$  which half of binding sites are occupied (indicative of binding affinity)



null:



NULL:

no cooperativity  $\leftrightarrow n=1$  (Monod)

$$\theta_A = \frac{[L]}{K_d + [L]} \rightarrow \text{does not depend on \# binding sites.}$$

$\rightarrow$  total # binding sites occupied

$\bar{n}_A = \frac{n_A \theta_A}{n_A}$   $\rightarrow$  linearly depends on  
# binding sites.

$$\theta_B = \frac{[L]}{K_d^B + [L]} \rightarrow n_B \theta_B = \bar{n}_B$$

$$\rightarrow \bar{n} = \bar{n}_A + \bar{n}_B = n_A \theta_A + n_B \theta_B$$

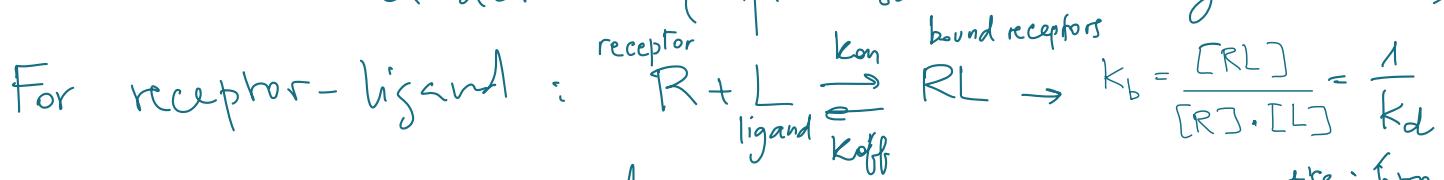
Activation of receptor =  $A(\bar{n})$

Review: Hill-Monod can be derived as consequence of Law of Mass Action



$$@ \text{equilibrium } K = \frac{\alpha_C^c \alpha_D^d}{\alpha_A^a \alpha_B^b} = \text{constant}$$

$\alpha$  activities (often substituted by concentration)



Total # receptors is conserved

$$\rightarrow [R] + [RL] = [R_T] \rightarrow [R] \left( 1 + \frac{[L]}{K_d} \right) = [R_T]$$

$$\frac{[R][L]}{K_d} (*)$$

$$\frac{[R]}{[R_T]} = \frac{1}{1 + [L]/K_d}$$

$$K_d = \frac{k_{off}}{k_{on}} : \text{from kinetics of reaction : } \frac{d[RL]}{dt} = k_{on} [R][L] - k_{off} [RL]$$

$$\rightarrow \text{again @ equilibrium } \frac{[RL]}{[R][L]} = \frac{k_{on}}{k_{off}} = \frac{1}{k_d}$$

True diversity  $\leftrightarrow \alpha_{\alpha i + \beta j} \neq \alpha a_i + \beta a_j$

↑  
activation homopentamer;  
↓ activation homopentamer;  
activation heteropentamer  
with  $\alpha$  subunits<sub>i</sub> +  $\beta$  subunits<sub>j</sub>  
 $\& \alpha + \beta = 5$

$$a_i = A(\theta_i)$$

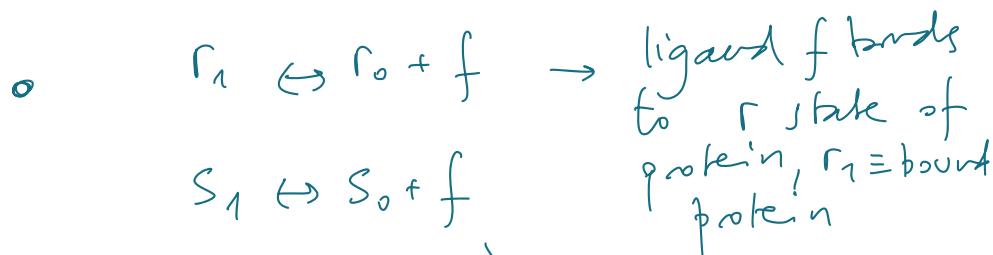
$$\theta_i = \frac{1}{1 + \left( \frac{k_A^i}{[L]} \right)^{n_i}}$$

# MONOD-WYMAN-CHANGEVX model

2 states of protomer ( $\equiv$  single binding site I think), S & r

- $r_0 \leftrightarrow s_0$  no ligand,

$$\left\{ \frac{s_0}{r_0} \right\} = l' \quad \text{isomerization constant}$$



$\downarrow$  bonds to s state w  
 $\neq$  affinity

for each "protomer"  
for example in a lattice,  
on surface of protein

Free energy transition is assumed to be proportional to fraction of protomers that are bound:

$$\Delta F = \varepsilon - \eta(r) \Rightarrow l' = e^{\beta \Delta F} = e^{\beta \varepsilon} = l'^{(r)} e^{-\beta \eta(r)}$$

! original modulation that comes from ligand binding

$\rightarrow$  predicts that depending on  $\lambda$  you can get graded response

↙ all or none response

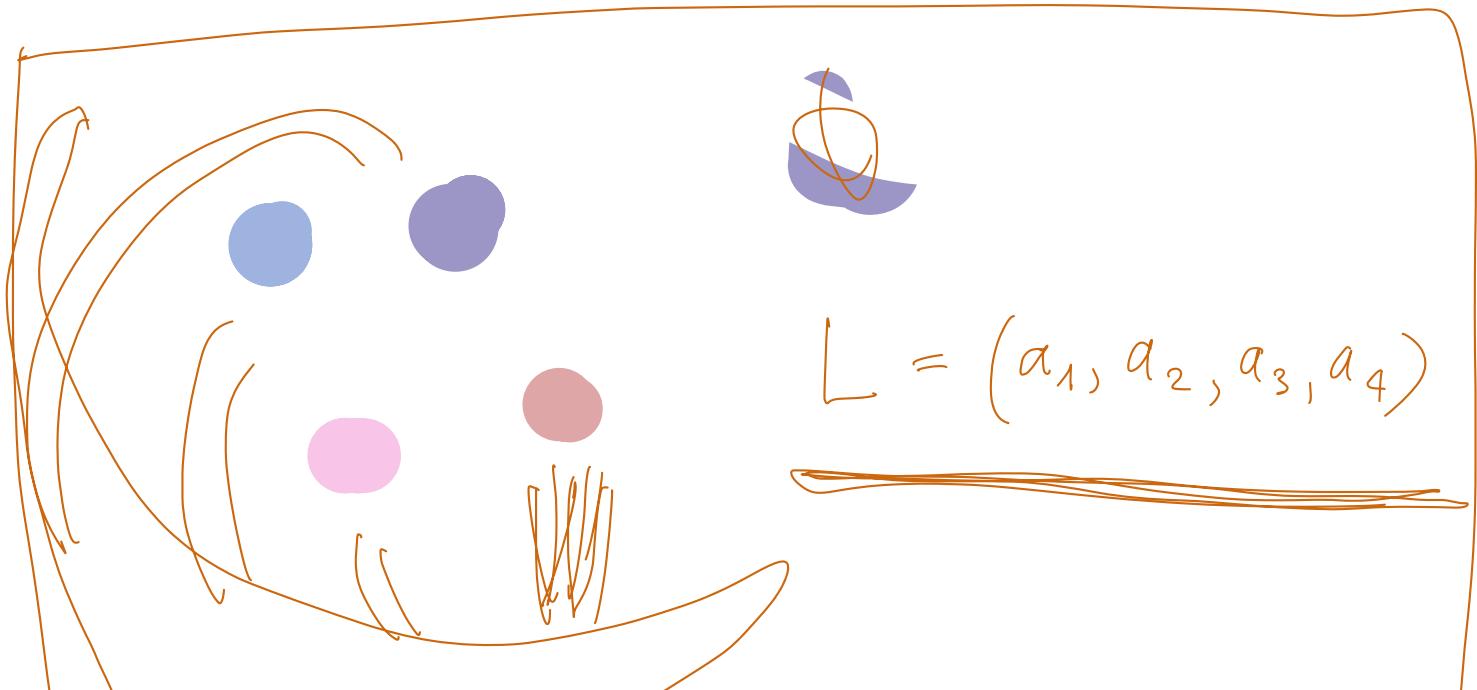
- Nicotinic receptor makes only 1 kind of heteropentamer
  - but big difference: only 1 subunit makes homopentam.
  - & the ones that assemble together CANNOT make " "

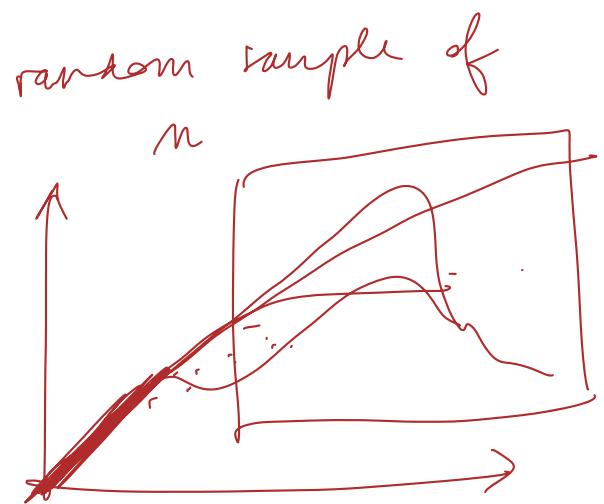
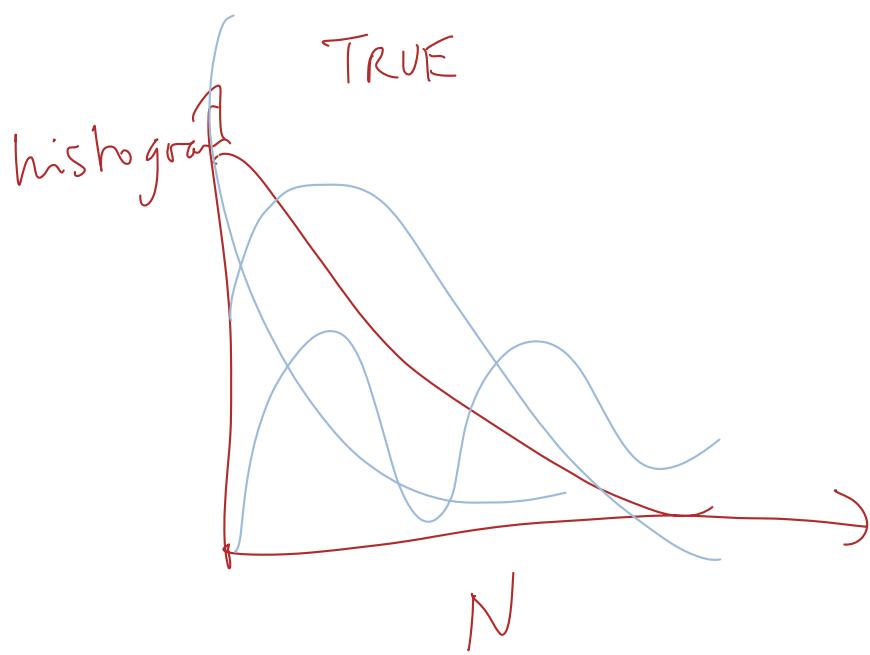
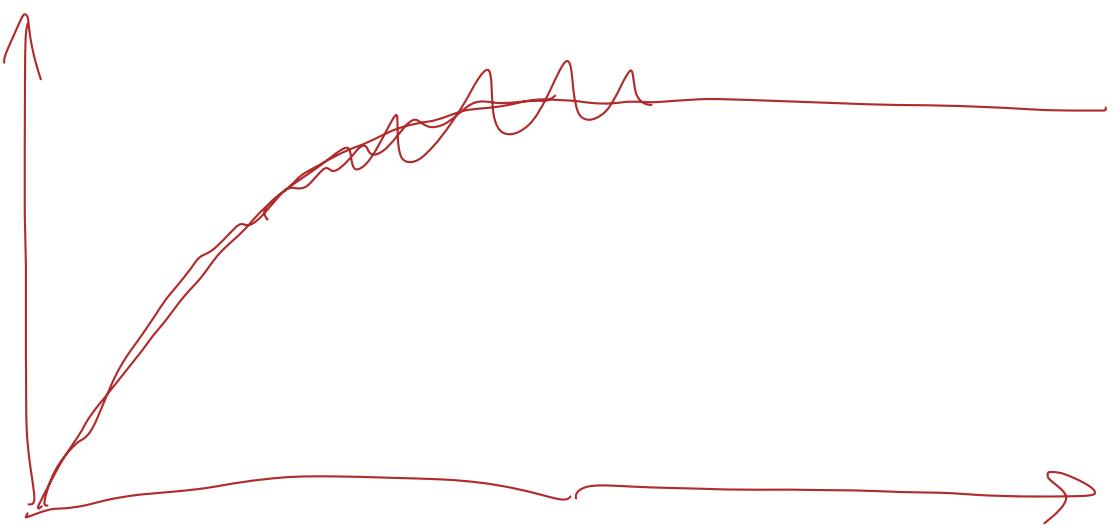
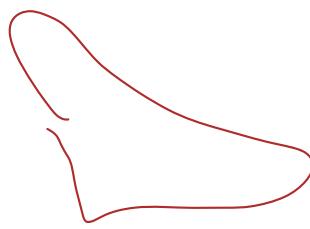
Reconstituted in HeLa cells

- May be possible to get structural biology (cryo-em)
  - × global expression levels of 2 subunits
  - × cryo-em all of the channels → cryo-em & do structure of all of them could be
    - ① only specific heteromer
    - ② " " + 2 homomers
    - ③ a lot of ≠ stuff

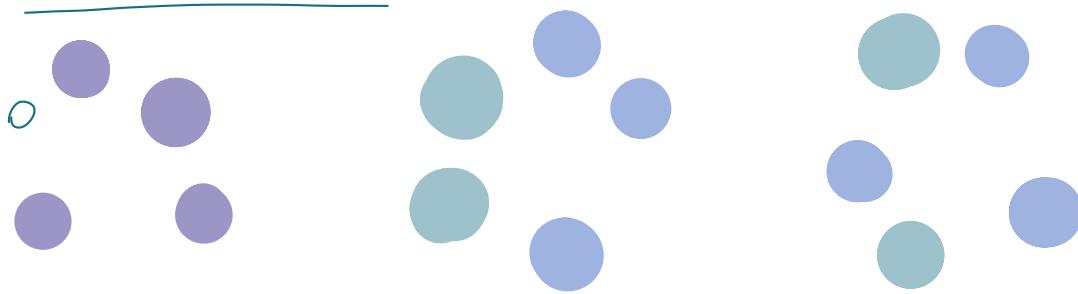
If its ① → easy !  
 ② → not clear if this heterogeneity is  
 within all cells  
 ③ → mess

- Another technique pull down from the otopus arm all receptors that have specific subunit.



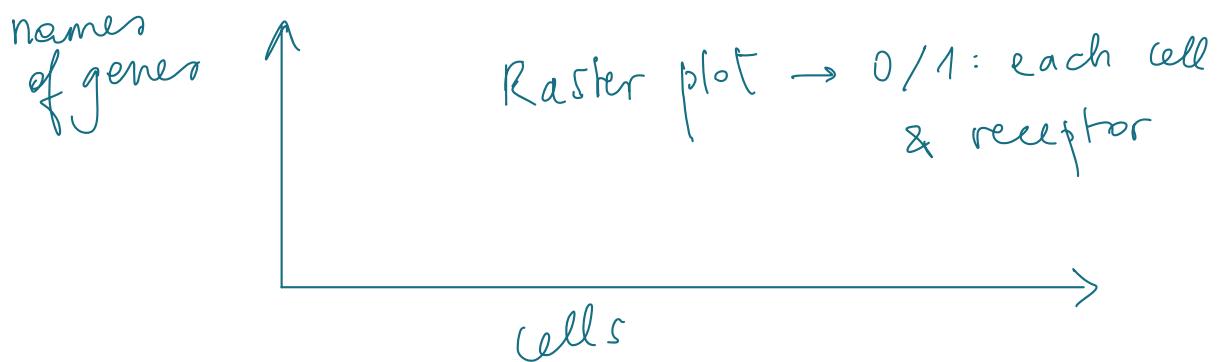


SQUID → 4 receptors instead of 26



### DATA SET:

from the 400 → know exactly what subunits compose each of them.



### WRITE DOWN OUR THINKING.

Next week → first dataset

≈ 2 months → more

---

GENOMICALLY for many receptors:

- what receptors
- morphology → available

SPECIFIC SHARKS: → cat shark most reliable

- where receptors are

- what do they bind

Q: are big whites just use better receptors?  
prob not.

→ PREDICT WHERE OPTIMAL RECEPTION LOCATION

→ Branson says "not located  
everywhere".

Measure of real flow speed & fish speed in the field  
for cat shark & dog shark!

$$\underset{\xi}{\operatorname{argmax}} \quad L(m_1=1 | \xi) L(m_2=1 | \xi) p(\xi) =$$

$$\int p_1 p(\phi_1 | \xi) d\phi_1 = \int l(\phi_1) \delta(\phi_1 - \tilde{\xi}_1(\xi)) d\phi_1 =$$

$$= l(\tilde{\xi}_1)$$

$$= \underset{\xi}{\operatorname{argmax}} \quad l(\tilde{\xi}_1) l(\tilde{\xi}_2)$$

THINKING:

cell types  
- not necessarily single channel

$$L \leftrightarrow (a_1, a_2, a_3)$$

"glomerulus"

$a_1 \dots a_k$  = activity of  $k$  receptors, function of ligand identity & concentration.

$p(a_i | L_j)$  = probab. that channel  $i$  opens, given ligand concentrat.  $L$  of a given ligand  $j$ .

. if  $a_i(L_k) = f(a_1^{(k)} \dots a_m^{(k)})$   $\rightarrow$   $a_i$  provides no additional

To Do show that  $M1(L, a_1 \dots a_m) = M1(L, a_1 \dots a_m)$

when  $a_i(L_k) = f(a_1^{(k)} \dots a_m^{(k)})$   $a_i$

Agnes

$f \equiv$  monod model

$\rightarrow$  dependent  $\leftrightarrow$  parameters Monod =  
 $i$

= function parameters Mon  
 $j, k$

To Do Thjo: fit  $\tilde{K}_c, k_o$  for the 3 cell  
types & 11 Ligands

Note: for homopentamer 2  $\rightarrow$  no activation.

DEF

$$I(a;c) = \sum_{a \in A} \sum_{c \in C} P(a;c) \log \frac{P(a,c)}{P(a)P(c)} = \sum_c P(a,c) \log \frac{P(a,c)}{P(c)} - \sum_{ac} P(a,c) \log \frac{P(c|a)}{P(c)}$$

$$= \sum_{\bar{c}} P(\bar{c}) \underbrace{\sum_a p(a|\bar{c}) \log p(a|\bar{c})}_{-H(a|\bar{c})} - \sum_a P(a) \log P(a) \underbrace{\sum_{\bar{c}} P(\bar{c}|a)}_1$$

$L$  = integer that labels identity of  $N$  ligands

$b$  = binary words with  $K$  digits

$H(b) = \log_2 2^k = k \equiv \# \text{ receptors} \rightarrow \text{no!} \supset \text{have to count the profiles} \supset \text{actually see. This changes}$

depending on binding affinities.

$$H(b|L) = -\sum_{j=1}^N p(j) \sum_{i=1}^{2^k} \underbrace{p(b_i|j)}_{\text{probability of receptor binding } b_i \text{ given ligand } j} \log_2 p(b_i|j) =$$

### ① Deterministic binding with constant profile

Given  $j \rightarrow \supset \text{activate always the same pattern. Single } b^*$

$$H(b|L) = -\sum_{j=1}^N p(j) \left( \sum_{b_i \neq b^*} 0 \times \cancel{\log_2 0} + \sum_{b_i = b^*} 1 \times \cancel{\log_1 1} \right) = 0$$

$\rightarrow MI = H(b)$  (max) because once the ligand binds, you know what the profile of activation is. It's always  $b^*$ .

But this also means that the only profile  $\supset$  for is  $b^*$

$$\rightarrow H(b) = -\sum_{b_i \neq b^*} 0 \times \log_2 0 - \sum_{b_i = b^*} 1 \times \log_1 1 = 0$$

$$MI(L, b) = H(b) - H(b|L) = H(b) - \underbrace{H(L|b)}_I = 0 \quad \text{in both cases} \checkmark$$

$$\lim_{x \rightarrow 0} x \log x = \lim_{x \rightarrow 0} \frac{\log x}{1/x} = \lim_{x \rightarrow 0} \frac{x}{-\frac{1}{x^2}} = \lim_{x \rightarrow 0} x = 0$$

### ② Deterministic binding, each ligand has its own profile $\rightarrow b_i$ also has $N$ possible observed values only

$$MI(L, b) = H(b) - H(b|L) = 0 \quad \text{TRIVIAL}$$

### ③ Now same as ②, but $N \gg M \rightarrow \# \text{ receptors limits MI}$

$$MI(L, b) = \log_2 2^k - H(b|L) = K \quad \text{capacity}$$

④ Now same as ③ but  $b$  = binary word with  $K$  digits, but one of the digit is a function of the others  $a_R = f(a_1 \dots a_K)$

$$p(b) = p(\bar{b}) \times \delta(a_R = a_R(\bar{b})) \rightarrow H(b) = - \sum_{\substack{b=\bar{b} \\ b=\bar{b}!a_R}}^1 \frac{1}{2^{k-1}} \log_2 \frac{1}{2^{k-1}} - \sum_{\substack{b=\bar{b} \\ \text{there are } 2^{k-1}}}^0 0 = k-1$$

$\uparrow$   
 $\frac{1}{2^{k-1}}$       1 if  $a_R = a_R(\bar{b})$   
                0 otherwise

→ you don't see all of the  $2^K$  possible  $b_s$ , only  $2^{k-1}$

→  $H(b) = \log_2 2^{k-1} = k-1 \rightarrow \text{capacity is reduced.}$

→  $a_R$  is useless

So the easiest way to think about this, is that if  $a_R = a_R(\bar{b})$

$$H(b|L) = 0 \quad \text{but} \quad H(b) < K \rightarrow MI(b, L) < K$$

⑤ Can extend to the fact that  $L = j \in (1, \dots, N)$

where  $N = \# \text{concentration bins} \times \# \text{identities}$

→ each binary word can code for 1 ligand @ 1 concentration

→  $a_R = f(a_1 \dots a_K)$  means that receptor

$(R)$  binds a ligand @ given concentration through a complex (non linear) combination of what all other receptors do with the same ligand @ same concentration.

For each odor & concentration:

$$a_R = \psi(\phi_R) \rightarrow \vec{\phi}_R = \psi^{-1}(a_R)$$

↓ parameters (2 or 3 param.)  
 model

if  $\vec{\phi}_R = g(\vec{\phi}_1 \dots {}_{(R)} \dots \vec{\phi}_K)$

$$\rightarrow a_k = \varphi(g(\overset{\rightharpoonup}{\phi_1} \cdots \underset{(R)}{\wedge} \cdots \overset{\rightharpoonup}{\phi_k})) = \varphi(g(\overset{\wedge}{\varphi'(a_1)} \cdots \underset{(R)}{\wedge} \cdots \overset{\wedge}{\varphi'(a_k)})$$

i.e. if  $\phi_{\text{HETEROmer}} = \text{Ligand}(\phi_{\text{HOMO1}}, \phi_{\text{HOMO2}})$

& if  $\delta$  is the SAME for all ligands, then  
Ligand  
the heteromer brings no additional info.

$\phi = (\text{probab. to be open w/o Ligand}; \text{dissociation rate while open; } \phi_0 \quad \phi_1 \quad \phi_2) \quad \Rightarrow \quad \phi = \phi_0 + \phi_1 + \phi_2$

$$(\phi_1^H, \phi_2^H) = f(\phi_1^{HoMo1}, \phi_2^{HoMo1}, \phi_1^{HoMo2}, \phi_2^{HoMo2})$$

SIMPLIFY to binary :  $a_R = \text{sign}(\phi_R - \phi_{y_2})$

$$\phi_1^R = f(\phi_1^1 \dots \phi_1^k)$$

$$\phi_1^{\text{HETEN}_0} = f(\phi_1^{\text{HOMO}_1}, \phi_1^{\text{HOMO}_2})$$

