AladDNA Modeling

1 Constitutive expression

Proteins A, B, C and D, are used to control the expression of next level. As are constantly produced, we will consider its expression as the zero input response of our system. We will assume that their concentrations have reached their balanced values before light inputs are given to the organism.

$$gA + RNA_{pol} \underset{k_{-1_A}}{\overset{k_{1_A}}{\rightleftarrows}} gA \cdot RNA_{pol}$$

$$gA \cdot RNA_{pol} \xrightarrow{k_{mA}} gA + RNA_{pol} + mA$$

$$mA \xrightarrow{k_{p_A}} mA + A$$

$$mA \xrightarrow{d_{mA}} \emptyset$$

$$A \xrightarrow{d_A} \emptyset$$

All these reactions must be expressed in algebraic language in order to work with their information in numerical terms. Using Mass Action Kinetics, we will express the presence of each protein as a balance between its appearance and disappearance, due to its involvement in different processes.

$$\frac{dProtein}{dt} = Protein_{appearance} - Protein_{disappearance} \tag{1}$$

Degradations are considered as first order reactions. Production rates that gobern each process are the same for all proteins, as they are preceeded by the same promotor. However, this could be changed by using promotors with different strength. Thus, each protein could have an individualized production rate (transcription + translation).

$$[\dot{A}] = k_{P_A} \cdot [mA] - d_A[A] \longrightarrow \begin{bmatrix} \dot{mA} \\ \dot{A} \end{bmatrix} = \begin{bmatrix} -d_{mA} & 0 \\ k_{P_A} & -d_A \end{bmatrix} \begin{bmatrix} mA \\ A \end{bmatrix} + \begin{bmatrix} k_{mA} c_{gA} \\ 0 \end{bmatrix} \tag{2}$$

As genes encoding proteins A, B, C and D, follow the same transcription pathway, we will

write their equations taking A as reference.
Expressing each concentration as a "x" variable:

$$[gA] = x_1 \qquad [RNA_p] = x_2 \qquad [gARNA_p] = x_3$$
$$[mA] = x_4 \qquad [A] = x_5$$

Where A is E-PIF6.

$$\dot{x_1} = -k_{1_A} \cdot x_1 \cdot x_2 + k_{-1_A} \cdot x_3 + k_{mA} \cdot x_3 \tag{3}$$

$$\dot{x_2} = 0$$

The expression above means that $[RNA_{pol}]$ inside the cell is high enough to consider it in excess. Therefor, its variation due to this reaction is nearly 0, i.e

$$x_2 \approx c_{RNA_p} \tag{4}$$

$$\dot{x_3} = k_{1_A} \cdot x_1 \cdot x_2 - (k_{-1_A} + k_{mA}) \cdot x_3 \tag{5}$$

$$\dot{x_4} = k_{mA}) \cdot x_3 - d_{mA} \cdot x_4 \tag{6}$$

$$\dot{x_5} = k_{p_A} \cdot x_4 - d_A \cdot x_5 \tag{7}$$

Notice that total amount of genes does not vary. What changes is the number of gene copies which is being transcribed. There must be a balance between active and inactive genes, i.e:

$$\dot{x_1} + \dot{x_3} = 0 \rightarrow x_1 + x_3 = c_{nA} \rightarrow Copy\, number\, of\, gene\, A$$

Each process is goberned by a rate with information about its kinetics, i.e: rates are directly proportional to the speed of the reaction. As a consequence of all different events involved in genetical expression, some dynamics can be considered much faster than others.
Slower reactions will be conditioning the final dynamic of the gene expression. Quassy Steady State Analysis let us assume that faster ractions have already happened while slower ones are still in process. Mathematically, this means rates several orders higher, allowing us for simplificating several expressions.

$$\dot{x_1} + \dot{x_3} = 0 \rightarrow x_1 + x_3 = c_{nA} \rightarrow Copy\, number\, of\, gene\, A$$

$$\frac{1}{k_{1_A}} \cdot \dot{x_3} = \frac{k_{-1_A} + k_{mA}}{k_{1_A}} \cdot x_3 - x_1 \cdot x_2; \qquad 0 = -x_1 \cdot x_2 + \frac{k_{-1_A}}{k_{1_A}} \cdot x_3;$$

$$x_3 = \frac{x_1 \cdot x_2}{\frac{k_{-1_A} + k_{mA}}{k_{1_A}}}$$

Moreover, all the genes which are not being transcripted, must be free. Then

$$c_{nA} = x1 + x3 \tag{8}$$

Using expressions (2) and (6):

$$x_3 = \frac{(c_{nA} - x_3) \cdot c_{RNA_{pol}}}{\frac{k_{-1_A} + k_{mA}}{k_{1_A}}} = \frac{c_{nA}}{\frac{1}{c_{RNA_{pol}}} \cdot \frac{k_{-1_A}+k_{mA}}{k_{1_A}}} - \frac{x_3}{\frac{1}{c_{RNA_{pol}}} \cdot \frac{k_{-1_A}+k_{mA}}{k_{1_A}}} =$$

$$= \frac{c_{nA}}{(\frac{1}{c_{RNA_{pol}}} \cdot \frac{k_{-1_A}+k_{mA}}{k_{1_A}}) \cdot (1 + \frac{1}{\frac{1}{c_{RNA_{pol}}} \cdot \frac{k_{-1_A}+k_{mA}}{k_{1_A}}})} = \frac{c_{nA}}{1 + \frac{1}{\frac{1}{c_{RNA_{pol}}} \cdot \frac{k_{-1_A}+k_{mA}}{k_{1_A}}}}$$

From equations (7), (4) and (5):

$$\dot{x_4} = k_{mA} \cdot \frac{c_{nA}}{1 + \frac{1}{\frac{1}{c_{RNA_{pol}}} \cdot \frac{k_{-1_A}+k_{mA}}{k_{1_A}}}} - d_{mA} \cdot x_4 = \overline{k_{mA}} \cdot c_{nA} - d_{mA} \cdot x_4$$

$$\dot{x_5} = k_{p_A} \cdot x_4 - d_A \cdot x_5$$

As A is being constantly produced, it can be assumed that when the trasncription of the second level starts, initial concentration of A will be balanced. Thus:

$$\dot{x_4} = 0 \longrightarrow 0 = \overline{k_{mA}} \cdot c_{nA} - d_{mA} \cdot x_4; x_{4_{eq}} = \frac{\overline{k_{mA}} \cdot c_{nA}}{d_{mA}}$$

$$\dot{x_5} = k_{p_A} \cdot \frac{\overline{k_{mA}} \cdot c_{nA}}{d_{mA}} - d_A \cdot x_5 = \overline{k_A} \cdot c_{nA} - d_A \cdot x_5$$

Proceeding equally with B, C and D, the system which describes the first level:

$$\begin{bmatrix} \dot{A} \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} -d_A & 0 & 0 & 0 \\ 0 & -d_B & 0 & 0 \\ 0 & 0 & -d_C & 0 \\ 0 & 0 & 0 & -d_D \end{bmatrix} \cdot \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} + \begin{bmatrix} \overline{k_A} \cdot c_{nA} \\ \overline{k_B} \cdot c_{nB} \\ \overline{k_C} \cdot c_{nC} \\ \overline{k_D} \cdot c_{nD} \end{bmatrix} \tag{9}$$

Taking the assumption that initial conditions are the equilibrium values of these concentrations, they are:

$$[A]_{eq} = \frac{\overline{k_A} \cdot c_{nA}}{d_A} \qquad [B]_{eq} = \frac{\overline{k_B} \cdot c_{nB}}{d_B}$$

$$[C]_{eq} = \frac{\overline{k_C} \cdot c_{nC}}{d_C} \qquad [D]_{eq} = \frac{\overline{k_D} \cdot c_{nD}}{d_D}$$

Where $\overline{k}$ for each variable is:

$$\overline{k} = \frac{k_p \cdot k_m}{1 + \frac{1}{c_{RNAp}} \cdot \frac{k_{-1} + k_m}{k_1}} \cdot \frac{1}{d_m} \tag{10}$$

2 Regulated expression: $2nd$ level

2.1 Proteins production

When A (E-PIF6), B (PhyB-VP16), C (Gal4-KDronpa) and D (NDronpa-VP16) are produced, they interact with the operator binding sites OBD1 (in gE, gF and gG) and OBD2 (in gH, gI and gJ). Thus, proteins produced in previous levels, regulate the expression of the following ones, achieving our ***library of Binding Domains***
Light induced transcription, is due to the activacion of proteins B and D, whose structures change according to a certain light wavelenght. Theoretically, only if B* (activated) binds A (which is already attached to the genes), proteins E, F and G will be produced. On the other hand, it occurs similarly with D*, C and proteins H, I and J.
The symbol "+", makes reference to those genes which can be transcripted because recombinases have not bind to them.

$$A + gE^+ \underset{k_{-2_A}}{\overset{k_{2_A}}{\rightleftarrows}} gE^+ \cdot A$$

$$gE^+ \cdot A + RNA_{pol} \underset{k_{-3_A}}{\overset{k_{3_A}}{\rightleftarrows}} gE^+ \cdot A \cdot RNA_{pol}$$

$$gE^+ \cdot A \cdot RNA_{pol} \xrightarrow{k_{mE_b}} gE^+ + RNA_{pol} + mE$$

$$B \underset{k_{760_{deact}}}{\overset{k_{650_{act}}}{\rightleftarrows}} B^*$$

$$gE^+ \cdot A + B* \underset{k_{-4_A}}{\overset{k_{4_A}}{\rightleftarrows}} gE^+ \cdot A \cdot B^*$$

$$gE^+ \cdot A \cdot B^* + RNA_{pol} \underset{k_{-1_E}}{\overset{k_{1_E}}{\rightleftarrows}} gE^+ \cdot A \cdot B^* \cdot RNA_{pol}$$

$$gE^+ \cdot A \cdot B^* \cdot RNA_{pol} \xrightarrow{k_{mE}} gE^+ \cdot A \cdot B^* + RNA_{pol} + mE$$

$$mE \xrightarrow{k_{p_E}} mE + E$$

$$mE \xrightarrow{d_{mE}} \emptyset$$

$$E \xrightarrow{d_E} \emptyset$$

To study the behaviour of the blue-light toggle, we must consider the tetramerization that could take place when D changes to D*. In this step, there are two possible pathways:

D becomes D* and binds to C (already fixed to genes H, I and J); or D becomes to D_4*, reducing the amount of D*, and so does the gH, gI and gJ transcription rates.

$$C + gH+ \underset{k_{-2_C}}{\overset{k_{2_C}}{\rightleftarrows}} gH^+ \cdot C$$

$$gH^+ \cdot C + RNA_{pol} \underset{k_{-3_C}}{\overset{k_{3_C}}{\rightleftarrows}} gH^+ \cdot C \cdot RNA_{pol}$$

$$gH^+ \cdot C \cdot RNA_{pol} \xrightarrow{k_{mH_b}} gH^+ \cdot C + RNA_{pol} + mH$$

$$D \underset{k_{390_{deact}}}{\overset{k_{490_{act}}}{\rightleftarrows}} D^*$$

$$D \underset{k_{390_{mono}}}{\overset{k_{490_{tetra}}}{\rightleftarrows}} D_4*$$

$$gH^+ \cdot C + D^* \underset{k_{-4_C}}{\overset{k_{4_C}}{\rightleftarrows}} gH^+ \cdot C \cdot D^*$$

$$gH^+ \cdot C \cdot D^* + RNA_{pol} \rightleftharpoons gH^+ \cdot C \cdot D^* \cdot RNA_{pol}$$

$$gH^+ \cdot C \cdot D^* \cdot RNA_{pol} \xrightarrow{k_{mH}} gH^+ \cdot C \cdot D^* + RNA_{pol} + mH$$

$$mH \xrightarrow{k_{p_H}} mH + H$$

$$mH \xrightarrow{d_{mH}} \emptyset$$

$$H \xrightarrow{d_H} \emptyset$$

$$[\dot{m}H] = \frac{(P_{Hb} + P_H \cdot [D^*]) \cdot [C] \cdot [gH^+]}{1 + M_H \cdot [C] + N_H \cdot [C] \cdot [D^*]} - d_{mH} \cdot [mH] \tag{11}$$

$$[\dot{H}] = k_{P_H} \cdot [mH] - d_H \cdot [H] \tag{12}$$

"P" constants express transcription processes, being P_b the basal term, and P the light

activated one. Moreover, constants M and N, represent the consume of necessary reactives for transcription, being M the basal parameter, and N the light induced one.

$$P_b = k_{m_b} \cdot \frac{k_3 \cdot c_{RNA_p}}{k_{-3} + k_{m_b}} \cdot \frac{k_2}{k_{-2}}$$

$$P = k_m \cdot \frac{k_1 \cdot c_{RNA_p}}{k_{-1} + k_m} \cdot \frac{k_4}{k_{-4}} \cdot \frac{k_2}{k_{-2}}$$

$$M = (1 + \frac{k_3 \cdot c_{RNA_P}}{k_{-3} + k_{m_b}}) \cdot \frac{k_2}{k_{-2}}$$

$$N = (1 + \frac{k_1 \cdot c_{RNA_P}}{k_{-1} + k_{m_b}}) \cdot \frac{k_2 \cdot k_4}{k_{-2} k_{-4}}$$

In Biology, processes with different dynamics may occur simultaneously, as it happens with RNA and proteins. Although proteins depend on the fluctuation of RNA levels, that variations can be considered as constant (quasy stationary state) since because proteins' dynamics is much slower than RNA's. As proteins are our reference system in this case, we will assume a mRNA concentration that will not vary with time.

$$k_{p_E} = \frac{(P_{Eb} + P_E \cdot [B^*]) \cdot [A] \cdot [gE^+]}{d_{mE} \cdot (1 + M_E \cdot [A] + N_E \cdot [A] \cdot [B^*])} \tag{13}$$

$$k_{p_H} = \frac{(P_{Hb} + P_H \cdot [D^*]) \cdot [C] \cdot [gH^+]}{d_{mH} \cdot (1 + M_H \cdot [C] + N_H \cdot [C] \cdot [D^*])} \tag{14}$$

Being (13) valid for F y G, and (14) for I y J.
On the other hand, we will suppose that change induced by light is brief. Because of that:

$$[B^*] = [B]_{eq}$$

In D, we should take into account the fact that D (NDronpa) when activated could tetramerize.

$$[D^*] = k_{490} \cdot [D]_{eq}$$

We could simplify considerating that $P_E \cdot [B^*] \cdot [A]$ is the transcription rate induced by light for the promoter used,and $P_{Eb} \cdot [A]$ is the basal production, the addition of both can

be expressed as the constant C_{1E} that expresses the transcription of gene E.

$$[\dot{E}] = \frac{C_{1E} \cdot [gE^+]}{d_{mE} \cdot (1 + M_E \cdot [A] + N_E \cdot [A] \cdot [B^*])} - d_E \cdot [E] \tag{15}$$

$$[\dot{H}] = \frac{C_{1H} \cdot [gH^+]}{d_{mH} \cdot (1 + M_H \cdot [C] + N_H \cdot [C] \cdot [D^*])} - d_H \cdot [H] \tag{16}$$

2.2 Recombinase action

Proteins G and J, activated by red and blue light respectively, could form dimers. These dimers, are a kind of enzyme which can inactivate some DNA regions. Introducing those recombinase recognition sites, between genes E, F, G, H, I and J, we can solve the problem of the second pulse. Enzyme G recognises sites of genes H, I and J. Thus, when red light is activated, recombinase G prevents the blue-toggle production, ensuring that if the second pulse is blue, the other part of the circuit will not be activated. However, this makes our circuit a single use device, but consequences of simultaneous production, could put in risk its safety. Similarly to the notation used before, as genes with a "+" are those which can be transcripted, then genes followed by a "-", have been modified by enzymes, cancelling their expression. Genes from levels 1 and 3, do not have any distinction symbol, because there are not recognition sites for recombinases in those genes.
In order to recombinate one gene from (+) to (-), two enzyme-dimmers must bind to it. Then, the stoichiometry of this reaction is **2 : 1** (bindings : recombination).

$$G + G \underset{k_{-2_G}}{\overset{k_{2_G}}{\rightleftarrows}} G_2$$

$$2G_2 + gH^+ \xrightarrow{k_{3_G}} 2G_2 \cdot gH^+$$

$$2G_2 \cdot gH^+ \xrightarrow{k_{4_G}} 2G_2 + gH^-$$

As the gene copy number will not be constant because of recombinases action, then $[g^+]$ will be determined by the following expressions:

$$[gH^+] = \frac{c_n H - [gH^-]}{1 + k'_{3_G} \cdot [G]^4} \tag{17}$$

With

$$[g\dot{H}^{+}] = -k_{G_H} \cdot [G]^4 \cdot [gH^{+}] \tag{18}$$

Where constants k'_{3_G} , and k_{G_H} are:

$$k'_{3_G} = \frac{k_3}{k_4} \cdot \frac{{k^2}_2}{4 \cdot {k^2}_{-2}}$$

$$k_{G_H} = k_{3_G} \cdot \frac{k^2_{2_G}}{4 \cdot k^2_{-2_G}}$$

If blue light is firstly activated, recombinase phyC31 will be responsible of the reaction $g^{+} \longrightarrow g^{-}$ in genes E, F and G, until all of those genes are inactive, i.e g^{-}. It is exactly the same, but with recombinase bxb1 and genes H, I and J, when red light is given. However, because of basal production of recombinases, all equations must be used, whichever light has been chosen. Thus, if recombinases are not produced, then $[g^{+}] = c_n$

3 Output production

This third and final level, is controlled by proteins E, F, H and I, as it was the previous level by A and C. All these biomolecules have an structure which interacts with complex B* or D*, depending on if they have PIF6 or NDronpa. It is because of the reliability of the experimental demonstration, that our circuit ends in this level. Good experimental results must be achievable in order to show that our model works. However, more complex circuits could be designed by adding more regulatory proteins to our toggles. This would increase the number of possible outputs, just by adding more light pulses to control proteins production.

$$E + g\alpha \underset{k_{-2_E}}{\overset{k_{2_E}}{\rightleftarrows}} g\alpha \cdot E$$

$$g\alpha \cdot E + RNA_{pol} \underset{k_{-3_E}}{\overset{k_{3_E}}{\rightleftarrows}} g\alpha \cdot E \cdot RNA_{pol}$$

$$g\alpha \cdot E \cdot RNA_{pol} \xrightarrow{k_{m\alpha_b}} g\alpha \cdot E + RNA_{pol} + mH$$

$$g\alpha \cdot E + B^* \underset{k_{-4_E}}{\overset{k_{4_E}}{\rightleftarrows}} g\alpha \cdot C \cdot B^*$$

$$g\alpha \cdot E \cdot B^* + RNA_{pol} \underset{k_{-1_\alpha}}{\overset{k_{1_\alpha}}{\rightleftarrows}} g\alpha \cdot E \cdot B^* \cdot RNA_{pol}$$

$$g\alpha \cdot E \cdot B^* \cdot RNA_{pol} \xrightarrow{k_{m\alpha}} g\alpha \cdot E \cdot B^* + RNA_{pol} + m\alpha$$

$$m\alpha \xrightarrow{k_{p_\alpha}} m\alpha + \alpha$$

$$m\alpha \xrightarrow{d_{m\alpha}} \emptyset$$

$$\alpha \xrightarrow{d_\alpha} \emptyset$$

The final production of the chosen output, is exactly the same as it was in the second level, but now transcription will depend on the second light pulse. The time when last pulse must be applyed, is a critical parameter, specially if the colour changes from red to blue or viceversa. This is because recombinases must have finished their work, preventing the other toggle's production. Otherwise, two outputs may will be produced.

$$[\dot{m}\alpha] = \frac{(P_{\alpha_b} + P_\alpha \cdot [B^*]) \cdot [E] \cdot c_{n_\alpha}}{1 + M_\alpha \cdot [E] + N_\alpha \cdot [E] \cdot [B^*]} - d_{m\alpha} \cdot [m\alpha] \tag{19}$$

$$[\dot{m}\beta] = \frac{(P_{\beta_b} + P_\beta \cdot [D^*]) \cdot [F] \cdot c_{n_\beta}}{1 + M_\beta \cdot [F] + N_\beta \cdot [F] \cdot [D^*]} - d_{m\beta} \cdot [m\beta] \tag{20}$$

$$[\dot{m}\gamma] = \frac{(P_{\gamma_b} + P_\gamma \cdot [B^*]) \cdot [H] \cdot c_{n_\gamma}}{1 + M_\gamma \cdot [H] + N_\gamma \cdot [H] \cdot [B^*]} - d_{m\gamma} \cdot [m\gamma] \tag{21}$$

$$[\dot{m}\Omega] = \frac{(P_{\Omega_b} + P_\Omega \cdot [D^*]) \cdot [I] \cdot c_{n_\Omega}}{1 + M_\Omega \cdot [I] + N_\Omega \cdot [I] \cdot [D^*]} - d_{m\Omega} \cdot [m\Omega] \tag{22}$$

$$[\dot{\alpha}] = k_{P_\alpha} \cdot [m\alpha] - d_\alpha[\alpha] \tag{23}$$

$$[\dot{\beta}] = k_{P_\beta} \cdot [m\beta] - d_\beta[\beta] \tag{24}$$

$$[\dot{\gamma}] = k_{P_\gamma} \cdot [m\gamma] - d_\alpha[\gamma] \tag{25}$$

$$[\dot{\Omega}] = k_{P_\Omega} \cdot [m\Omega] - d_\alpha[\Omega] \tag{26}$$

When steady state has been reached:

$$[\alpha]_{eq} = \frac{(P_{\alpha_b} + P_\alpha \cdot [B^*]) \cdot [E] \cdot c_{n_\alpha} \cdot k_{p_\alpha}}{(1 + M_\alpha \cdot [E] + N_\alpha \cdot [E] \cdot [B^*]) \cdot d_{m_\alpha} \cdot d_\alpha} \tag{27}$$

$$[\beta]_{eq} = \frac{(P_{\beta_b} + P_\alpha \cdot [D^*]) \cdot [F] \cdot c_{n_\beta} \cdot k_{p_\beta}}{(1 + M_\beta \cdot [F] + N_\beta \cdot [F] \cdot [D^*]) \cdot d_{m_\beta} \cdot d_\beta} \tag{28}$$

$$[\gamma]_{eq} = \frac{(P_{\gamma_b} + P_\gamma \cdot [B^*]) \cdot [H] \cdot c_{n_\gamma} \cdot k_{p_\gamma}}{(1 + M_\gamma \cdot [H] + N_\gamma \cdot [H] \cdot [B^*]) \cdot d_{m_\gamma} \cdot d_\gamma} \tag{29}$$

$$[\Omega]_{eq} = \frac{(P_{\Omega_b} + P_\Omega \cdot [D^*]) \cdot [I] \cdot c_{n_\Omega} \cdot k_{p_\Omega}}{(1 + M_\Omega \cdot [I] + N_\Omega \cdot [I] \cdot [D^*]) \cdot d_{m_\Omega} \cdot d_\Omega} \tag{30}$$