Supplementary Document to "BioJazz: *in silico* evolution of cellular networks with unbounded complexity using rule-based modeling"

Song Feng^{1,&}, Julien F Ollivier^{2,&}, Peter S Swain^{3,*}, and Orkun S Soyer^{1,*}

*Corresponding authors. [&]These authors contributed equally to this work.

1. The energy framework implemented in ANC for modelling allosteric regulation

In particular, the Arrhenius equation gives the kinetic rate of the R-T transition as: $k_{RT} = k_{R\dagger} = Ae^{-\Delta G_{R\dagger}/kT}$, with "†" denoting the transition state, A denoting the Arrhenius constant, and kT being the product of Boltzmann's constant and temperature {DePaula 2006}. Similarly, $k_{TR} = k_{T\dagger} = Ae^{-\Delta G_{T\dagger}/kT}$. The equilibrium distribution of the R and T states will be governed by the equilibrium constant K_{RT} , which is given by k_{RT}/k_{TR} , where $K_{RT} = Ae^{(-\Delta G_{R\dagger} + \Delta G_{T\dagger})/kT} = Ae^{-\Delta G_{RT}/kT}$.

In ANC, modifiers are assumed to contribute independently to the free energy of each conformational state, R and T, allowing us to formulate the free energy difference between these two states ($\Delta G'_{RT}$) in a given domain with N modifiers as:

$$\Delta \mathbf{G}_{RT}' = \Delta \mathbf{G}_{RT} + \sum_{i=1}^{N} \left(\Delta \mathbf{G}_{T}^{(i)} - \Delta \mathbf{G}_{R}^{(i)} \right), \tag{1}$$

where $\Delta G_T^{(i)}$ and $\Delta G_R^{(i)}$ give the effect of the *i*th modifier on the free energies of the R and T states. While $\Delta G_{RT}'$ could be evaluated via equation [1], this requires assignment of the $\Delta G_T^{(i)}$ and $\Delta G_R^{(i)}$ values. Instead of doing this, we can exponentiate [1] and thus equivalently define the effect of each modifier on the overall equilibrium distribution between the R and T states. To do so, we defined the relation of the equilibrium constant of the domain without any modifiers (K_{RT}) to that with modifiers (K'_{RT}) as:

$$\frac{k'_{RT}}{k'_{TR}} = K'_{RT} = K_{RT} \prod_{i=1}^{N} \Gamma_i, \qquad (2)$$

where $\Gamma_i = e^{-(\Delta G_T^{(i)} - \Delta G_R^{(i)})/kT}$ denotes the effect of the *i*th modifier on the equilibrium distribution between the R and T states. The Γ_i relate to the altered kinetic rate constants in the presence of the *i*th modifier in the following manner:

$$k'_{RT} = k_{RT} \prod_{i=1}^{N} (\Gamma_i)^{\Phi_i}, \qquad (3)$$

$$k'_{TR} = k_{TR} \prod_{i=1}^{N} (\Gamma_i)^{(\Phi_i - 1)},$$
(4)

with the parameter Φ_i describing the proportional effects of the *i*th modifier on the R-T transitions. To simplify the implementation of this approach, all modifiers acting on different reactive sites of a domain are assumed to employ the same Φ value (i.e. $\Phi_i = \Phi_j$ ($i \neq j$) for all reactive sites in one domain) {Ollivier 2010}.

2. Calculation of different scores in ultresensitivity fitness function

If we define y_{min} and y_{max} as the minimum and maximum values of the response during the interval from a change in the signal to steady-state, then the response amplitude for each of the signal ramp-ups (indicated with a '+' sign) and ramp-downs (indicated with a '-' sign) is calculated as:

$$\Delta y_{i+} = y_{i+_{max}} - y_{i+_{min}}$$
(5)
$$\Delta y_{i-} = y_{i-_{max}} - y_{i-_{min}}$$
(6)

where the subscripts denote the the corresponding ramps in the input signal. With these measurements, the amplitude score (S_{amp}) is given as the normalized amplitude of the change in response to the second ramp-up and ramp-down signals:

$$S_{amp} = \frac{(\Delta y_{2+} + \Delta y_{2-})/2}{y_{total}}$$
(7)

with y_{total} being the maximum possible response (i.e. the concentration difference between a fully active and fully inactive output protein), and acts as a normalization factor ensuring S_{amp} to be between 0 and 1.

To quantify the ultrasensitivity of the system, we use the difference between the amplitudes of the responses to the second ramp-up/ramp-down signals, and the first choice/third choice. We first define the difference in the response to the different ramp-up and ramp-down signals as:

$$S_{u1} = \frac{(\Delta y_{2+} + \Delta y_{2-})/2}{(\Delta y_{1+} + \Delta y_{1-})/2} \quad (8) \qquad \qquad S_{u3} = \frac{(\Delta y_{2+} + \Delta y_{2-})/2}{(\Delta y_{3+} + \Delta y_{3-})/2} \quad (9)$$

and the ultrasensitivity score (S_{ult}) is then calculated as:

$$S_{ult} = \sqrt{\left(\frac{S_{u1}}{r_u + S_{u1}}\right) \cdot \left(\frac{S_{u3}}{r_u + S_{u3}}\right)} \tag{10}$$

In equation [12], r_u is a user-defined scaling parameter that ensures the two ratios S_{u1} and S_{u3} (and thus the ultrasensitivity score) is between 0 and 1.

Besides the amplitude and ultrasensitivity scores, we also define a complexity score (S_{com}) . It is plausible to assume that networks are under selection to minimize their

energetic burden to the cell, and this score allows us to capture network complexity. The complexity score is given by;

$$S_{com} = \frac{r_c}{r_c + C} \tag{11}$$

where C is the sum of the total number of rules, proteins, domains, and reactive sites in the ANC model and r_c is a user-defined scaling parameter for scaling the complexity score S_{com} between 0 and 1.

3. Calculation of fitness function quantifying adaptive response

The key aspects of adaptive response dynamics are that the system shows an initial response to the input $(\Delta O_{max}^{+/-} \neq 0)$ and that the steady state value of the output returns to its pre-input level, irrespective of the level of the input. In other words, after a sustained change in input (e.g. a step change), the output should initially respond but ultimately settle back to its original steady state: $\Delta O_{ss}^{+/-} \approx 0$. Therefore, the adaptation fitness *w* can be configured as:

$$w = \sqrt{\frac{\Delta O_{max}^+}{C} \times \frac{K}{K + \Delta O_{ss}^+}} \cdot \sqrt{\frac{\Delta O_{max}^-}{C} \times \frac{K}{K + \Delta O_{ss}^-}}$$
(12)

where *C* is a normalization factor to scale $\Delta O_{max}^{+/-}$ and $\Delta O_{ss}^{+/-}$ in to [0,1], and *K* is a threshold parameter. By imposing such a selective pressure, it is possible to evolve an increased response sensitivity ($\Delta O_{max}^{+/-}$) and reduced adaptive error ($\Delta O_{ss}^{+/-}$), and so achieve networks with an adaptive response.