Genetic Switches between two population with regards to mRNA and proteins applying Markov Chain Stochastic Model Check.

Qin He[1;2;3;4] RubinWang[5] XiaochuanPan[5] Jukka Peltola[2;3] Susanna Narkilahti[3] Jari Hytinnen[4]
1.Signal Processing Department,Tampere Universities
2.Medicine and Health Technology department,Tampere Universities, 3.TAYs Tampere,Finland
4.Neural-Dyamics Lab, East China University of Science and Technology Shanghai, China
5. Biophysics Department,Tampere Universities

Abstract

As Arc, one virus-like gene, crucial for learning and memory, was discovered by researchers in neurological disorders fields, Arc mRNA's single directed path and allowing protein binding regional restric-tively is a potential investigation on helping shuttle toxic proteins responsible for some diseases related to memory deficiency. To study especially the transform between mRNA and proteins, the switching function of the phenotypes, 'normals' multiplying populations and 'persisters', resilient to stress instead of multiplying is of our interest. Mean time to switching (MTS) is calculated explicitly quantifying the switching process in statistical methods combining Hamiltonian Markov Chain(HMC).The model derived from predator and prey with type

II functional response studies the mechanism of normals with intrin-sic rate of increase and the persisters with the instantaneous discovery rate and converting coefficients. During solving the results, since the numeric method is applied for the 2D approximation of Hamil-tonion with intrinsic noise induced switching combining geometric minimum action method. In the application of Hamiltonian Markov Chain, the behavior of the convertion (between mRNA and proteins through 6 states from off to on) is described with probabilistic conditional logic formula and the final concentration is computed with both Continuous and Discret Time Markov Chain(CTMC/DTMC) through Embedding and Switching Diffusion.The MTS, trajectories and Hamiltonian dy-namics demonstrate the practical and robust advantages of our model on interpreting the switching process of genes (IGFs, Hax Arcs and etc.) with respects to memory deficiency in aging process which can be useful in further drug efficiency test and disease curing.

1 Introduction

In cell biology, non-equilibrium stochastic process is of interest since the observation of experimental results are becoming of higher res-olution, studying the molecules both with imaging and expression data are often conducted in both single and population (thousand) order, which basically described in stochastic process whether on a discrete or continuous scale with status changes either genotypi-cally or phenotypically. Many problems are thus studied related to status switching, including cell regulatory networks[1], signal re-sponse on excitability and inhibition[2], (convinced by translational and transcriptional burst of expression for instances.), metastability among populations, (binding of ligands and proteins, forming of polymerases and etc.).In this paper, we focus on the interaction among genes, mRNA, proteins and etc. To be more specific, while the switching problem among molecules can be studied on geno-type, including sequencing for single RNA, alignments and binding considering condons and etc, we stay on the switching with expres-sion (concentration) only, which is simplified as modified population problem using Lotka-Volterra equations[3] of two populations only. Thus, rather than the competitor model(for instances, cell bifurcations.), we applied simulation of switching on predator model. The model is based on the following basic assumptions: Prey population (promoters) is fed with enough food all the time while the predator population of the predator(the persisters) depends on the size of prey(promoters).

In our paper, we mainly study the interaction of DNA and its interaction with the associated proteins.(Clinical data of Hax1 and HS_1 is downloaded from Ensmbl gene database[4]). On one hand, the switching model is calculated under the large deviation theory(LDT)[5] combining the least actions. The Markov chain[6] consider the states of the 2D coordinates (x; y) of mRNA numbers and protein numbers referencing the distribution of x, which follows the order O(1) while P_X follows the time scale on O(1/e) and guaran-teeing the variant of LDT hold with the transform of the expressions in single population. Only considering the process of diffusion case, we study the binding of hax1 with simple switching between on and off status under its interaction with HS_1 seen as in the constant environment, i.e. the closed system at mean field. The dimer which can be cancelled out connect the binding between two single population. On the other hand, one numeric method is applied to solve the problem, making compare with the stochastic process[7] on ap-proximation equation of the mean switching time(MST) with the transform between two status (we studied the switching time with four situations, both multiplicative and asymptotic of single population and the binding and degradation between two population.) Again, this method is also calculated based on the Hamiltonians. We give out the MST with respect to N/Nc denoting N as the population number of interest and Nc as the threshold of certain status(either of that population or the other population). Since our study only based on data in the process of transforming in the constant environment, extinction is not considered in this paper. To study both intrinsic and extrinsic noise with the exciting and inhibiting bursts is the potential topic in the future. In the following contents, the first chapter is the proposition of the model, based on least action with LDT and MTS approximation with one stochastic differential equation (SDE) [8]separately; And the second chapter gives numeric experiments based on Hamilton Markov Chain[9] computation of the expression data of hax1 and HS1; In the last chapter, the model is described in the normal logic formula with both probabilistic condition model[10] and the results are analysed with both hamiltonian, realization size, convergence, the rewards computation taking the CTMC as Poisson process[11] and the reachability computation with the transfer kernel of switching diffusion[12] through DTMC. In the appendix, there also includes the complete proof of model with action S based on Hamilton not only based on the explicit equation in this paper. Some descriptive Statistics and pre-computation based on the data can be accessed through link in availability. As the process related to motor coordination and func-tion, the Hax's function in regulation, B cell's signal transduction can be further studied with more data considering its excitability and metastability functions with stimulation of drugs for instance in the future as well. And one computation applying DTMC with linear regression on previous work is made as the further extension of the model.

2. Proposed Model

Molecular interactions are studied on phenotypic data of the mRNA and its associated protein in this paper, especially the trajectory of the production of hax1 and HS_1 with interaction with each other through least action method combining diffusion process[10]. Furthermore, in solving the equation, one stochastic differentiation equation approximates the analytic solution and calculation of MST[11] based on converging with Hamiltonian quantities, finding three convergence points through eigenvalue of position quantities as well as satisfying $H = 0$ and $H_q = 0$ where $q(P_X ; P_Y)$ are momentum quantities. In the 3rd subsection, the transition is illustrated with belief graph first and then convert ratio are utilized in computing the discrete embedding of the continuous temporal logic. As comparison, the third subsection compute the discretized time markov chain as the approximation considering it as a hybrid systems.

utilizing the first equation of the four. Furthermore, $S(\phi)'s$ upper bound can also be obtained through defining a minimizing sequences $(T_k, \psi_k)_{k \in N}$ with the following rescaling process: For every $k \in N$ let: $\lambda_k(\alpha) = max(\lambda(\phi(\alpha), \phi'(\alpha)), \frac{1}{k}), \alpha \in [0,1], B_k(\alpha) = \int_0^{alpha} \frac{1}{\lambda_k} da, \alpha \in [0,1], T_k(\alpha) = B_k(1), \psi_k(t) = \phi(B_k^{-1}(t)), t \in [0, T_k]$ Specifically, the inverse of B_k is approximated with the Brownian standard σ_x satisfying the $\alpha'(t) = \lambda_k(\alpha(t))$ and thus $\frac{1}{k} \leq \alpha'(t) \leq |\lambda_k|_{inf} \leq \inf$ holds for all $t \in [0, T_k]$ with the absolute continuity of $\alpha(t)$. And thus, the ψ_k is continuous in the whole time sequence $(0, T_k)$, enabling the inverse process: $t = t(\alpha) = G_k(\alpha)$ with $dt = d\alpha/\lambda_k$ and $\phi'(alpha) = \psi_k'(t)G_k'(\alpha) = \frac{\psi_k(t)'}{\lambda_k(\alpha)}$. Thus,

$S_{T_k}(\phi_k) = \int_0^{T_k} \frac{L(\phi, \phi'\lambda_k)}{\lambda_k} d\alpha$

leading to the upper bound switching the integrate and limitation with $k- > \inf$, and with the proof in appendix B(in another work with landscape model) fufilling the first order and second order conditions: $\phi' = \frac{H_\theta(\phi, \theta)}{\lambda}$ is negative definite during the θ maximizing process: $\frac{L(\phi, \lambda * \phi')}{\lambda} = sup_{\theta \in R^n}(< phi', \theta > - \frac{H(\phi, \theta)}{\lambda})$ and guaranteeing them both fufiled by $\theta = \theta(\phi, \phi')$ with the second equation, so that upperbound here is the same as the integrands of the lower bound as well as holds the $\theta = 0$ when the $\lambda = 0$ is satisfied, and therefore: $\frac{L(\phi, \lambda * \phi')}{\lambda} = < \phi' theta > - \frac{H(\phi, \theta)}{\lambda} = < \phi', \theta >, \theta = \theta(\phi, \phi')$

The calculation can be found completely in Appendix B.

2.1 switching model with least action

First of all, we consider the dynamics of population of the interaction involved systems as diffusion[12], and thus the Hamiltonian $H(x, \theta)$ is computed with the minimization of action (quasi-potential)[13] instead of some other methods, for instance WKB[14]. With the Lagrangian denoted with respect to Hamiltonian according to LDT: $L(x,y) = sup_{\theta \in R^n}(< y, \theta > -H(x, \theta)) = < y, \theta(x,y) > -H(x, \theta(x,y))$ Due to the maximizer $\theta(x,y)$ being implicitly defined by $H_\theta(x, \theta(x,y)) = y$, we calculate the action from quasi-potential: $V(x1, x2) = inf_{T>0} inf_{\psi \in C_{x1}^{x2}(0,T)} S_T(\psi) = inf_{T>0} inf_{\phi \in C_{x1}^{x2}(0,1)} inf_{\psi \in C_\phi(0,T)} S_T(\psi) = inf_{\phi \in C_{x1}^{x2}(0,1)} S(\phi)$

So that for any $\phi \in C(0,1)$ the action $S(\phi)$ is given by the equivalent four formula:

$S(\phi) = inf_{T>0} inf \phi \in C_\phi(0,T) * S_T(\psi)$

$S(\phi) = sup_{\theta:[0,1] -> R^n}^{H(\phi, \theta)=0} \int_0^1 < \phi', \theta > d\alpha$

$S(\phi) = \int_0^1 < \phi', \hat{\theta}(\phi, \phi') > d\alpha$

$S(\phi) = \int_0^1 \frac{L(\phi, \lambda * \phi')}{\lambda} d\alpha, \lambda = \lambda(\phi, \phi')$

Note that L(x, y) is the Lagrangian associated with the Hamiltonian $H(x, \theta)$ with function $\theta(x,y)$and $\lambda(x,y)$ are implicitly defined for all $x \in D$ and $y \in R^n/0$ as the unique solution $(solution(\theta, \lambda) \in R^n X [0, \inf)$ of the system possessing zero value when $\phi' = 0$ or $\lambda(\phi, \phi') = 0$ setting the integrands to zero with: $H(x, \theta) = 0, H_\theta(x, \theta) = \lambda y, \leq lambda$

where the lower bounds for $S(\phi)$ is directly achieved :

$S(\phi) = inf_{T>0} inf \phi \in C_\phi(0,T) S_T(\psi)$

$\geq \int_0^1 sup_{H(\phi, \theta)=0}^{\theta \in R^n} < \phi', \theta > d\alpha$

$\geq \int_0^1 < \phi', \theta(\phi, \phi') > d\alpha.$

2.2 diffusion Approximation with numerical methods on the convert ratio referencing bacteria sensoring and MTS on difference mapping

As to study the switching model interpreting the process explicitly, we thus combine the deterministic[15] background of the switching between on and off and give out one stochastic model based on the explicit (ordinary differential equation) ODE of the numbers of mRNA and proteins. Although the final model(referencing the quorum sensing model of bacteria in changing environment[16]) removes the dimers but it is used in the first place while cancelled out the in the quasi steady state according to its far more faster production and degradation rate comparing to transcription and translation.(Simplified mechanism sees Appendix A). Start from the bistability of the metastability[17] of the two state model, with the absorbing boundary conditions, $\rho_0(x*, t) = 0$ and the identification of mean transition rate with principal eigen value λ_ϵ^0, the quasi-stationary approximation of $\rho_n(x,t) = C_0 * \exp(-\lambda_\epsilon^0 t) * \phi_\epsilon^0(x, n)$. Furthermore, with the quasi-potential satisfying $\Sigma_{m \in 0,1} S_n(x)(A_{n,m}(x) + \phi_0'(x)\delta_{n,m})F_m(x) = 0$, $H = 0.5 * (g_1^2 p_1^2 + g_2^2 p_2^2) + p_1 * \phi_1 + p_2 * \phi_2$ where p_i is the momentum conjugate to the generalized coordinate x_i, where $g_i = \sqrt{(S_i^2 f_i + X_i/\tau i)}$ (For more specific study of the ϕ_1 and ϕ_2 as the interacted diffusive speed, most studies applies WKB equations.) Since we focus on the transform between two status of the two populations, mRNA (of HS-1) X_n and proteins Hax1 Y_n as the system.(with dimer Z of production rate k_{XY} and degradation rate k_P)[18] and the degradation rate of HS_1 and hax1, as K_X and K_Y, separately. From the original ODES:

$\frac{dZ}{dt} = k_{XY}XY - k_Z Z$

$\frac{dX}{dt} = -k_{XY}XY + k_Z Z - k_X X + V_X * \frac{Z}{(K_X + Z) + X_0}$

$\frac{dY}{dt} = -k_{XY}XY + k_Z Z + V_Y \frac{Z}{(K_X + Z)} + V_Y * \frac{Z}{(K_Y + Z)} + Y_0 - k_Y X$

where X_0 and Y_0 are the initial volumes or baseline volumes of these two populations and with instant volume as V_X and V_Y and due to the zero value of $\frac{dP}{dt}$, the term of P can be replaced through:

$$P = \frac{k_{XY}}{k_P} XY$$

$$\frac{dX}{dt} = -k_X X + V_X * \frac{k_{XY}}{k_P} XY + X_0 = -k_X X + V_X * \frac{1}{(1+\frac{k_X k_P}{k_{XY}})} + X_0$$

$$\frac{dY}{dt} = -k_Y Y + V_Y * \frac{k_{XY}}{k_P} XY + Y_0 = -k_Y Y + V_Y * \frac{1}{(1+\frac{k_Y k_P}{k_{XY}})} + Y_0$$

Considering the transform of X(Upstream only), in the first step as degradation as the first term of right of the upper formula, the degradation part of X with k_X which can be interpreted as the Poisson process and rewrite into $-\frac{\mu 1}{exp^{p}X}$, and in the second term, the coefficient of degradation part of X , $C1$ is denoted as $\frac{V_X * k_X}{\mu 1}$. Mean while with the assumption of continuous Markov chain, where the convert ratio of Y is n, the $\frac{k_{XY}}{k_P} * X * Y$ is equivalent to $(\frac{Y}{X+Y})^n$ so that the whole degradation part becomes $\frac{C1 * \mu 1}{1+(\frac{Y}{X+Y})^n} exp^{p_X}$, $\frac{C2 * \mu 2}{1+(\frac{X}{X+Y})^n} exp^{p_Y}$ (*1) and the final transform rate of mRNA number X and proteins Y are:

$\frac{C1}{1+(\frac{Y}{X+Y})^n}(exp^{p_X} - 1) - \mu 1 * X(exp^{-p_X} - 1)$ and $\frac{C2}{1+(\frac{X}{X+Y})^n}(exp^{p_Y} - 1) - \mu 2 * Y * (exp^{-p_Y} - 1)$ where, the coefficient of degradation part of Y $C2$ denoted as $\frac{V_Y * k_Y}{\mu 2}$. as the reciprocal of the other population ratio. And as Y stands for the number of the proteins, X for the number of the mRNA separately with m and n as their translation and transcription rate. With the total sum of the system mocule numbers assumed as X+Y, we have the Hamiltonian:

$$\frac{C1}{1+(\frac{Y}{X+Y})^n}(expP_X - 1) - \frac{\mu 1 * X}{(exp - P_X - 1)} + \frac{C2}{1+(\frac{X}{X+Y})^n}(expP_Y - 1) - \frac{\mu 2 * Y}{(exp - P_Y - 1)}$$

,where p_X, p_Y are calculated setting H = 0 and $H_\theta = 0$, and convertion rate which can be calculated as $\frac{dy}{dx}$ specifically here, letting the first term equals the second and third equals the fourth term. (Complete see Appendix B) Note that: each single DNA population($hax1$ and HS_1) has its own degradation rate when considering about its mRNA computation, and the other population's protein is taken as the intake, promoting its population when as normals binding onto the according site of persisters, activating it. Vice versa, thus, the two populations have similar structured formula describing each degradation and population under the dual interacted population. The mean switching time is calculated based on the solution of the SDE: $z' = z + \sqrt{(\frac{N_c}{N})} * \sqrt{1+2*\epsilon \cdot z^2} * \eta$,

where $N_c = \frac{1}{\ }$ and η $N(0, \delta)$ is the white noise with correlator $<\eta(\tau)\eta(\tau')> = \delta(\tau - \tau')$. Note that it is the span of the master equation in powers of the inverse population size N^{-1} re-scaling with $\tau = 2 * *t/N$, and $z = x_1 - x_2$ ranges over the interval [-1, 1][20], leading to the solution $\tau_{\pm} = \frac{2\pi b}{1-2c} * cot(\pi)$. Thus, we have the algorithm:

Input: maximum time scale size T, mRNA numbers y, proteins x, maximum steps Steps, tolerance Tol, parameters of the sensing model(coefficients of convertion c1, c2, transcription and translation rate m, n, degradation rate k1, k2, formation coefficients mu1, mu2, diffusion rate b1, b2), dt as each time increasement

Initialize: maximum time scale, T, maximum step number steps, tolerance Tol, numbers of mRNA after the first diffusion process xhat if necessary, initialized as one random the number in the first status and thus we start with the largest interval to cover higher possibilities, i.e. [x(0),x(0)+1,...,x[1]-1].

for do

xhat(length(xhat))< x(i+1): **repeat**

record the size T, time t, steps Steps - steps +1

set the sequence according to size T (the interval for mRNA numbers) x(1), x(2),...x(T) and generate the population number of proteins according data distribution, y(1), y(2),...y(T). T initialized as the X(i+1)-X(i).

consider Hamilton Markov (Hierarchical)[19]

if *xhat* exists(iterated from previous status) **then**

: segment the interval into several sub-sequences (X0 as the new current status, X1 as the previous status.)

end if

Note that: As we only consider up streaming, down regulation into those before the previous status is not included.

function dynamics Inputs: x and fitted y (or $x0, y0$ or $x1, y1$)

calculate degradation term $w1$ and $w2$ according to the (*1)

calculate px, py, dx, dy, conversion rate, H_θ and H_c, s, according to Appendix 2

predict multiplied mRNA and protein numbers xhat, yhat, and other Hamiltonians.

calculate updated gamma, delta

calculate tolerance for further stopping criteria as the residule of gamma and cell numbers with: $tol = abs(Gamma - gamma)/gamma, toll = mean(abs(xhat - x1)./x1 + abs(yhat - y1)./y1)$

Output: H, Hhat, HthetaX, HthetaY, HxX, HxY, HamilX, HamilY, HamilXhat, HamilYhat, xhat, yhat, sX, sY, px, py, pxhat, pyhat, actionratio, delta, gamma, Delta, Gamma, cr1, c1, c2, crhat, c1hat, c2hat, tol, toll, Txc, Tyc, Txchat, Tychat

concatenate results:

if X0, X1 exist **then**

xhat = [X0hat, X1hat]

end if

do similar prediction regenerate the mRNA numbers X according to Y with **function dynamics** again for comparison. Results are with postfix 'L'

store the quantities of 'successful' moves with smaller tolerance and action for either from mRNA or protein numbers.

if satisifies the configuration condition **then**

$Tol = min(tol * toll, tolL * tollL)$

else

Fail++

end if

2.3 Stochastic Model on the Uncertainty of Reasoning Switching Model

Stochastic model checking is a method for calculating the likelihood of the occurrence of certain events during the execution of a system while linear-time properties specify the traces to be exhibited through a transition system, including the atomic propositions AP, as the subset of $(2^{AP})^w$, satisfaction relation and etc. Although the definition of such language with LT property is composed by infinite word, the transition systems fulfilled the LT property is of finite states. When the state is visited transitional, each time a state is visited, we adversarial pick a transition distribution that respects the interval constraints, and take a probabilistic step according to the chosen distribution.

2.3.1 Probabilistic Uncertainty Conditional Model

Considering the formal semantic system, the normal logic formula are utilized to construct the belief graph,with the agents defined as the molecules in the transfer process (mRNA and proteins in this paper.) As the sensing model is highly sensitive to external factors, it is suitable for the agents to be modeled as epistemic states which is based on the prior of the given[22] transfer instead of the deterministic probability. That is reasoning the actions in the presence of sensing dynamically with regards to epistemic agents [23] rather than the whole convert system.

And, with each action states S(a, p) defined, the belief graph is supported by reasoning bidirectionally about both the transcription and translation to ensure trust and safety in the interaction (as shown in Appendix A).

2.3.2 Model Check for Stochastic Models Combining Continuous Time Markov Chain with Embedding in Reward Computation

The logic applied on a probabilistic notion regards to the belief graph is based on the trust which is reflected by the reliability and predictability. Specifically, the language of the stochastic models used for computing CTMC is the Continuous Stochastic Logic(CSL) developed and extended by some research[24.]

A CTMC is a tuple C = (S,s,R,L) where S is the finite set of states, s is the initial state; R is $S*S \to R_{\geq 0}$ is the transition rate matrix; L: $S \to 2^{AP}$ is a labelling function which assigns to each state s € S is the set L(s) of atomic propositions valid in the state. Instead of the case of DTMCs, a fixed set of atomic propositions AP is applied, the transition rate matrix R assigns rates to each pair of states in the CTMC, used as parameters of the exponential distribution. A transition can only occur between states s and s' if R(s,s)>0, representing the probability of this transition being triggered within t time-units equals $1\text{-}e^{-R(s,s')t}$. Time spent in state s, before such transition occurs, is exponentially distributed with rate E(s), where:

E(s) = sum(R(s,s'))

where E(s) is known as the exit rate of state s.

The embedded DTMC of a CTMC, is the probability of each state s' transitioned from the precious s, independent of the time, defined as:

Emb(C) = (S,s,$P^{emb(C)}$,L) where for s,s' € S:

,$P^{emb(C)}$(s,s') = { R(s,s') /E(s), if E(s) ≠0

1 if E(s) =0 and s =s'

0 otherwis

where the behavior of the CTMC in the alternative way remains in a state s delayed and exponentially distributed with rate E(s) and transit with Pemb(C)(s,s').

The infinitesimal generator matrix for the CTMC C=(S,s,R,L) is the matrix Q: S*S->R defined as:
Q(s,s') = R(s,s'), if s is not s'
- ∑s'' ≠s, R(s,s') otherwise
The CTMC stores the transition from s to s' in ratio format instead of the possibility in DTMC[25].
However, the probability measures Prs on ∑PathC(s) as the unique measure such that Prs(C(s)) = 1 and for any cylinder C(s,I,..,In-1,sn,I',s'), Prs(C(s, I,…,In-1,Sn, I',s')) equals:
Prs(C(s,I,…,In-1,sn))= Prs(C(s,I,…,In-1,sn))*P1emb(C1)(sn,s')(e-E(sn)*inf I' -e-E(sn)*sup I')
In our case, such model check as with PCTL, we can easily derive the path formulae for the states between S0 and S7 separately with 6 time intervals I= [t0,ti] :
P~p[◊Iφ] = P~p[true UI φ],
P~p[□Iφ] = P~p[exist UI φ],
φ = 'transit
Stands for the probability that a transition occurs in time interval I=[t0, ti],
And thus, For determing the least solution,ProbC,(s,φi,U[0,t],ψ)
=∫∑Pemb(C)(s,s')*E(s)*e-E(s)*x*ProbC(s', φi,U[0,t], ψ)
=ProbC(φ,U[t,œ])
=Prob{ProbC(s, φ,U[0,t'-t], ψ), if s|=φ
0 otherwise
And define the rewards function a CTMC D=(S,s,R,L), the semantics is defined as:
S |= R~r[I= t] □ ExpC(s,XI=t)~r

2.3.3 Model Check for stochastic models combining Discrete Time Markov Chain with Switching Diffusion in Reachability Computation

In the second application of model check, the continuous dynamics described by switching diffusions is studied with reachability and dually safety properties on DTMC[21] Compared with the MC on continuous time domain, DTMC is defined with a fixed, finite set of atomic propositions used to label states. The DTMC D is a tuple similar as CTMC (S,s,P,L), where S is a finite set of states; s is the initial states; P: S*S->[0,1] is the transition probability matrix where $\sum_s$ P(s,s') = 1 for all s€S where L(s) of atomic propositions are valid.
max{dq(t1,t2)}<=Kd*|t2-t1|, where dq(t1,t2) pseudally defined as dq(t1,t2) = sqrt(E[(Xq(t2)- Xq (t1))^2]), and K>=12 is the Dudley metric universal constant.
Let h defined larger than 0 be a sampling time and the mean E and the covariance C to simulate a nomal distribution N(x| E,C). Then, the discreete kernel is T((A,q_j),(x,q_i)) =
$\int_A$N(x|eF(qi)*hx,gamma(I,h)) dx*e-Āh if qi = qj
$\int_A$(∫N(x|Eqi,x(s),C qi,x(s)*) dx*lambda*Āij/Āi*Āi*h*e(-Āh) if qi = qj,
Where gamma(i,t) = $\int_A$(eF(qi)*(t-m))*G(qi)*G(qi)T(eF(qi)*(t-m))Tdm,
Ω Āi,Āi,t(s) = (Āi-Āi)*e$^{(}$Āj*s-Āi*t-Ai*s$^{)}$/(eĀj*s- eĀj*s),
$E_{qi,x(s)}$ = $e^{F(qi)s}e^{F(q)(h-s)}$x,
$C_{qi,x(s)}$ = $e^{F(qi)s}$ gamma(i,t) $e^{F(q)(h-s)}$x+ gamma(j,h-t),
0=<eps<=1-$e^{-Āi*h}$- Āi*h*$e^{--Āi*h}$
With the events on t€I A^n ={X(t) €S| P~p[1,] = P~p[true U^I φ]},
B^n={X(t) €S| P~p[$◊^I$φ] = P~p [exist U^I φ],
Psafe(X,S,I) = limP(A^n^B^c), Preach(X,S^c,I) =1- limP(A^n^B^c)
Tdx(z1, z2) = T(z1,z2), if z1,z2€S_{dx}
1-$\sum_{zj€Sdxr}$zT(z), if z1€$S_{dx,z2/,}$ z2€S
1, if z1,z2€ φ,
0, if z1€ φ, z2€S_{dx}
Continuous kernel proof see Appendix B.
To compute the reachability/safety properties, we introduce the scheme based on Discrete Time Markov Chain(DTMC) which discretize the state space to approximate the Discrete Time Markov Process(DTMP) results from the original switching diffusion process H, a tuple H=(Q, K, F,G,W,^), where Q = {q1,…,q|Q|} is the set of discrete modes instead of the matrix in CTMC and Y=(X, α) its solution. For any q €Q, call Xq the solution of the SDE:
dXq(t) = F(q)*Xq(t)dt+G(q)*dW(t) (*)
In this section we assure that Xq is a ui-dimensional, zero mean Gaussig,an process (GP). Xq is almost surely bounded within the interval I by Assumption.

Set h = min{2(-n)/(2*sqrt(2)*K^2*Kd),2 $^{-n}$} and €n = $2^{(-n/2)}$, where n€N, and Kd is a constant such that for any t1, t2€I

3 results and discussion

0.1 To have a clearer understanding of the switching process combining the binding with increasing and decreasing speed both of *hax1* and HS_1, the two population are regarded as promoters and resistors both when activating and deactivating each other's production.(The coarse process can be briefly described as in 1, and it is briefly introduced in the previous chapter.)

3.1 Finding three critical points and explore their stability

As we have data (see Appendix C) of 15 status in all both for hax1 and HS_1 with their different cell numbers taken as X and Y in our model. For reward computation for their Markov chain, we precompute the their Hamiltonians, Action Potentials, mean switching time and related dynamics in the form (see availability), and the 6 upstreaming status, which is the focus of the experiment application of our model. Using the pre-computation results, we are able to discuss about some practical problems about the current model. There are three groups of quantities studied combining the action potential as well as Hamiltonian inspired by bacterial quorom sensoring, 'momentum and cell numbers', 'MTS with the SDE', and 'corresponding Hamiltonians', of each transform status in Appendix A. First, we use the taylor expansion to simplify the four ODE achieved in Appendix B: to

$dx = \frac{C1}{1+(\frac{y}{x+y})^m} P_x - \mu 1 * x * P_X)$

$dy = \frac{C2}{1+(\frac{x}{x+y})^n} P_y - \mu 2 * y * P_Y)$

$dP_X = \frac{C2*m*(\frac{x}{x+y})^{m-1}}{(1+(\frac{x}{x+y})^m)^2}(P_Y - 1) - \mu 1 * P_X - \mu 1$

$dP_Y = \frac{C1*n*(\frac{y}{x+y})^{n-1}}{(1+(\frac{y}{x+y})^n)^2}(P_X - 1) - \mu 2 * P_Y - \mu 2$ Note that our model here simplify the origin model where $ci = \frac{ai}{bi}$, with $b_i = 1$ as the burst size of protein i, $\frac{x}{(x+y)} = \frac{x}{K_2*(x+y)}$ as $k_2 = 1$ is the dissociation constants standing for gene x binding on y's protein binding site. Regarding x and y as leading order variable, we apply phase analysis to consider the solution's stability around the three zero-energy points, which achieved through setting dx, dy, dP_X and dP_Y all to zero and combine the Hamiltonian's special case when H= 0 (and $H_P = 0$): P1(x, y, $\frac{\mu 2*X}{C1}$, $\frac{\mu 1*Y}{C2}$, where x and y are the solution of $x = \frac{C1}{\mu 1*(1+(\frac{y}{x+y})^n)}$ and $y = \frac{C2}{\mu 2*(1+(\frac{x}{x+y})^m)}$, P2(x, y, 0, 0), where x and y are the solution of $x = \frac{C1}{\mu 1(1+(\frac{y}{x+y})^n} = -y = -\frac{C2}{\mu 1(1+(\frac{x}{x+y})^m}$, and P3(0,0,0,0) As P_X and P_Y are either zero or formula can be replaced by x and y around those three convergence points. We here, consider the analysis on x and y as following: denote $dx = f(x,y)$ and $dy = g(x,y)$, and the we try to find x* and y* satisfy the $f(x,y) = 0$ and $g(x,y) = 0$ as well as holding the zero-energy points for their momentum. Thus with approximation: $dx = f_X(x*, y*)(x - x*) + f_Y(x*, y*)(y - y*)$, and $dy = g_X(x*, y*)(x - x*) + g_Y(x*, y*)(y - y*)$. We have $A = \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}$

$= \begin{bmatrix} -\mu 1 * P_X & \frac{C2*m*(\frac{x}{x+y})^{m-1}}{(1+(\frac{x}{x+y})^m)^2} * P_X \\ -\mu 2 * P_Y & \frac{C1*n*(\frac{y}{x+y})^{n-1}}{(1+(\frac{y}{x+y})^n)^2} * P_Y \end{bmatrix}$

where there exists the a>0, b>0 for the eigenvalue λ:

$$\lambda^2 + a * \lambda + b = 0 \tag{1}$$

$$a = -(f_X + g_Y)|_{(x*, y*)} \tag{2}$$

$$b = |A| \tag{3}$$

utilizing the first equation of the four. Furthermore, $S(\phi)'s$ upper bound can also be obtained through defining a minimizing sequences $(T_k, \psi_k)_{k\in N}$ with the following rescaling process: For every $k \in N$ let: $\lambda_k(\alpha) = max(\lambda(\phi(\alpha), \phi'(\alpha)), \frac{1}{k}), \alpha \in [0,1], B_k(\alpha) = \int_0^{alpha} \frac{1}{\lambda_k} da, \alpha \in [0,1], T_k(\alpha) = B_k(1), \psi_k(t) = \phi(B_k^{-1}(t)), t \in [0, T_k]$ Specifically, the inverse of B_k is approximated with the Brownian standard σ_x satisfying the $\alpha'(t) = \lambda_k(\alpha(t))$ and thus $\frac{1}{k} \leq \alpha'(t) \leq |\lambda_k|_{inf} \leq inf$ holds for all $t \in [0, T_k]$ with the absolute continuity of $\alpha(t)$. And thus, the ψ_k is continuous in the whole time sequence $(0, T_k)$, enabling the inverse process: $t = t(\alpha) = G_k(\alpha)$ with $dt = d\alpha/\lambda_k$ and $\phi'(alpha) = \psi_k'(t)G_k'(\alpha) = \frac{\psi_k(t)'}{\lambda_k(\alpha)}$. Thus,

$S_{T_k}(\phi_k) = \int_0^{T_k} \frac{L(\phi, \phi'\lambda_k)}{\lambda_k} d\alpha$

leading to the upper bound switching the integrate and limitation with $k->$ inf, and with the proof in appendix B(in another work with landscape model) fufilling the first order and second order conditions: $\phi' = \frac{H_\theta(\phi,\theta)}{\lambda}$ is negative definite during the θ maximizing process: $\frac{L(\phi,\lambda*\phi')}{\lambda} = sup_{\theta \in R^n}(< phi', \theta > -\frac{H(\phi,\theta)}{\lambda})$ and guaranteeing them both fufiled by $\theta = \theta(\phi, \phi')$ with the second equation, so that upperbound here is the same as the integrands of the lower bound as well as holds the $\theta = 0$ when the $\lambda = 0$ is satisfied, and therefore: $\frac{L(\phi,\lambda*\phi')}{\lambda} =< \phi' theta > -\frac{H(\phi,\theta)}{\lambda} =< \phi', \theta >, \theta = \theta(\phi, \phi')$

The calculation can be found completely in Appendix B.

so that point(x*, y*) is the convergence points. Thus, we discuss about the stability of the three points as following: we denote $X = (1 + (\frac{x}{x+y})^m)$ and $Y = (1 + (\frac{x}{x+y})^n)$, compute the a and b as: $a = -(f_X + g_Y)|_{(x*,y*)} = \mu 1 * P_X + \frac{C2*m(\frac{x}{x+y})^{m-1}}{(1+(\frac{x}{x+y})^m)^2} * P_Y, b = |A|_{(x*,y*)} = \mu 2 * P_X * P_Y * \frac{C1*n(\frac{x}{x+y})^{n-1}}{(1+(\frac{x}{x+y})^n)^2} - \mu 1 * P_X * P_Y * \frac{C2*m(\frac{x}{x+y})^{m-1}}{(1+(\frac{x}{x+y})^m)^2}$

[1] for $P1(x, y, \frac{\mu 2*X}{C1}, \frac{\mu 1*Y}{C2})$, where x and y are the solution of $x = \frac{C1}{\mu 1*(1+(\frac{x}{x+y})^n)}$ and $y = \frac{C2}{\mu 2*(1+(\frac{x}{x+y})^m)}$,

$a = \mu 1 * \frac{\mu 2*x}{C1} + \frac{C2*m(\frac{x}{x+y})^{m-1}}{(1+(\frac{x}{x+y})^m)^2} * \frac{\mu 1*y}{C2} = [illegible] + [illegible]$

$= [illegible]$ As in out model, [illegible] and [illegible], we have [illegible] and thus [illegible] while denominator [illegible], which gives out [illegible],

$b = \mu 2 * P_X * P_Y * \frac{C1*n(\frac{x}{x+y})^{n-1}}{B^2} - \mu 1 * P_X * P_Y * \frac{C2*m(\frac{x}{x+y})^{m-1}}{A^2}$

$= [illegible] * (A^2 \frac{x}{x+y})^{n-1} \mu 1 * C2 * m - (B^2 \frac{x}{x+y})^{m-1} \mu 2 * C1 * n), = [illegible]$

nominator $= \frac{dy}{dx} - \frac{dx}{dy}$ while denominator > 0

Thus, P1 is stable if and only if $\frac{dx}{dy} > \frac{dy}{dx}$, here means the production of hax1 is faster than HS_1.

[2] for P2(x, y, 0, 0), where x and y are the solution of $x = \frac{C1}{\mu 1(1+(\frac{x}{x+y})^n} = -y = -\frac{C2}{\mu 1(1+(\frac{x}{x+y})^m}$,and $P_X = P_Y = 0$

Thus, for P2, a = b = 0. It's unstable.

[3] for P3(0, 0, 0, 0), same as P2, a = b = 0 and it's unstable.

3.2 HMC dynamics

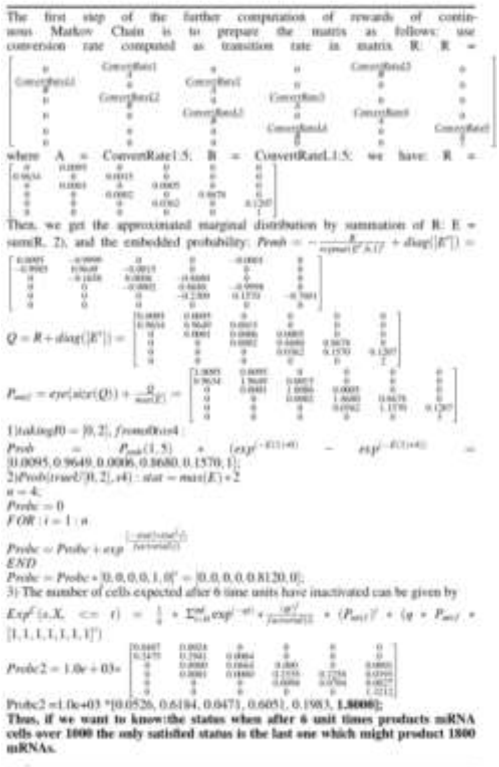
In the second part here, with regard to the detailed behavior of mRNA and protein dnamics[26], we look into their momentum and numbers with 6 status(only the firs 2(a)-2(c) and the last 2(d)-2(f) transition xamples of the origin 3groups*5transition statuses figures) in all are studied detailedly while the whole data based on 15 status. As we only investigated the positive direction, the red ones(top left) the application on clinical data while green one(top right) in the larger scaled simulation with more transition status(blue dashed line is the predicted dynamics)..Note that[27] the persisters and normals are the roles they take in the whole process(considering from bifurcation to catastrophe and extinction) where here they can be all considered as promoters as their numbers both grows in this process until the last status as their interaction in constant environment is of our main interest as we mentioned before. Generally, with small change studied in one status, the trend is more significant than the larger scale transition. For instance, the green simulation are always more sensitive to the momentum change and shows them more significantly on the cell trajectory comparing to the red clinical transition(we manually break one clinical status into sub-status in simulations.)

Specifically, in the 1− > 2 transition, the production of the HS 1 is slightly
faster than hax1 with the accelerate from faster to slower as well as the hax1 0 s momen-tum decreases from fast to slow while HS 0 1 s momentum increases from fast to slow similarly. The larger scaled simulation show the trend similarly but with larger momen tum difference and thus gives out the curve trajectory instead of straight line in the top left figure; On contrary, in 2− > 3 transition, both the clinical application and larger simulation give totally the same behavior according to the dynamics , where proteins products faster than mRNA but with similar acceleration . Other transition can be similarly analysis. Note that from the 4− > 5 of the larger scale simulation, there starts to show the switching where the protein changes into persisters with degradation instead of production which can be both detected from cell numbers figure in the top left and momentum figures in the right bottom although the fewer status contained

3.3 Conclusion

chain random walk with its faster convergence. As in 2(g) and 2(h), the convergence(variation to mean) of the markov chain hamilton is in blue line and the red line for clinical data and simulation on more possible transition status, giving different convergence but similar phase interval(according to 2(i)), interestingly. The last status transition converge the worst followed by the fist transition. And the result simulated with more markov chain status converges better than the clinical results. And according to the convert rate, the mRNA to Protein transfer ratio should be the highest when starting, and goes especially lower in the last two status which is in assistance to the protein binding as we cut off the process around the convergence point where the two population has reached metastability F M [29].According to the simulation result, the protein has gone through the switching process changing from normals to persisters and back to normals(bursts in optimal time in 2(h) might also due to the switch.).

Mean while, as the second population providing food(protein) to the other's binding site and either activiate or deactivate it, it works as the extrinsic noice induced the excitability or exhibition of the other gene. Here, as we choose hax1 and HS 1 , they work as promoters for each others. One noticable computation is the reward computaiton based on stochastic model selection which is useful in predict the possible status of the cell numbers easily with precomputation. And we can consider correct the transition matrix with simualted clinical tested results to improve the prediction as well. On the other hand, the most important calculation action potential is easier to be achieved through Hamilton as we proved with geometric minimum action and stochastic approximation. Ohter methods can cover Hamilton Jacobian matrix, WKB and etc.[28]. As we also improve the algorithm with adding hierarchical markov in calculating number of cells in different status only record successful move according to the tolerance based on action potential and residual of



The first step of the further computation of rewards of continuous Markov Chain is to prepare the matrix as follows: use conversion rate computed as transition rate in matrix R: R = [illegible]

where A = ConvertRate1:5; B = ConvertRateL1:5; we have: R = [illegible]

Then, we get the approximated marginal distribution by summation of R: E = sum(R, 2), and the embedded probability: Pemb = [illegible] + diag(|E'|) = [illegible]

$Q = R + diag(|E'|) =$ [illegible]

$P_{emb} = eye(size(Q)) + \frac{Q}{max(E)} =$ [illegible]

1) taking t0 = [0, 2], from 0 to t4 :

Prob = $P_{emb}(1,5)$ * [illegible] = [0.0095, 0.9649, 0.0006, 0.8680, 0.1570, 1];

2) Prob(travel [0, 2], s4) : stat = max(E) * 2

n = 4;

Probc = 0

FOR i = 1 : n

Probc = Probc + exp [illegible]

END

Probc = Probc * [0, 0, 0, 0, 1, 0]' = [0, 0, 0, 0, 0.8120, 0];

3) The number of cells expected after 6 time units have inactivated can be given by

[illegible]

[1, 1, 1, 1, 1, 1, 1]')

Probc2 = 1.0e+03* [illegible]

Probc2 =1.0e+03 *[0.0526, 0.6184, 0.0471, 0.6051, 0.1983, **1.8000**];

Thus, if we want to know:the status when after 6 unit times products mRNA cells over 1000 the only satisfied status is the last one which might product 1800 mRNAs.

prediction numbers both, the convergence of the algorithm is guaranteed. And

further research can be conducted on the whole process from bifurcation to catastrophe and extinction as well. Problem with multi population is also possible. As hax1 is observed to have function in signaling and regulating of genes especially in learning systems and motor related brain function, this switching model study related to its binding might help to predict the cell numbers and production or degradation rate especially later with further study into both with promoters and persisters as to test different drug and their efficiency

on the aging process related disease.

In the computation of reachability, the approximation with DTMC[30] mainly compute the kernels of Brownian with shift, finally discretize the original switching diffusion process. As the DTMC gives out the kernel with probability instead of the ratio, it is then convenient to be written into transition matrix Pdx which is discretized from Sdx on finite space state and gives out the reachability with error I/h*(Kdx+exp(-2^n-2^n/2+1)). As the proof in Appendix, the error bounds with Lipschitz constants converged with prominent K= mh1+Lh2. The computed result N*K*dxis here 0.453 with N = 6, m = L = 2,dx = 0.002 and h1 = 0.001, and h2 = ceil(h2*N)=1.71. The result is

Since the final result of the continuous process is not of probability range thus we normalize it with P = ratio/sum(ratio) and the DTMC approximation shown in the figure is the approachability(1-safety.) The result of the ttest tested continuous embedded matrix and h =0,p = 0.0515,ci = -0.7212 , 0.0029,stats = struct with stat: -2.2103(df: 10), do not reject the hypothesis that the two process. As the final safety consistently for two methods gives highest concentration for the last state showing the example computation's direction from off to on, although there is the slight difference that the forth in the continuous process is relative lower comparing to its other five states as well as the one in the discrete process states. The DTMC gives strictly increasing concentration from off to on during the 6 states.

As the switching diffusion is a commonly used model in genetic field, not only useful in the transmission of different molecules but also can be derived into analytical models giving straight transfer information abou tsome process with either concentration[31] change or energy change.

Reference

Freedman,H.I.,Dynamics of simple gene-network motifs subject to extrinsic fluctua-tions,1980

Noufe H. Aljahdaly,Analytical Solutions of a Modified Predator-Prey Model through a New Ecological Interaction,2019

ensembl,https://www.ensembl.org/index.html

Ingo Lohmar, Baruch Meerson, Switching between phenotypes and population extinction, Racah Institute of Physics,2018

Marta Kwiatkowdka, Gethin Norman, David Parker, Stochastic Model Checking, Oxford, 2019

David Martínez-Rubio, Varun Kanade, Patrick Rebeschini, Decentralized Cooperative Stochastic Bandits, Oxford, 2005

Paul C Bressloff, Stochastic swiching in biology: from genotype to phenotype, Department of Mathematics, University of Utah, 2017

Baier, Christel, Principles of Model checking, Massachusetts Instititute of Technology,2008

Cheng Lv, Xiaoguang Li, Fangting Li, Tiejun Li, Constructing the Energy Landscape for Genetic Switching System Driven by Intrinsic Noise, Peking University, 2014

C.W. Gardiner, Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences, 4th ed, Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences, 4th ed.,Springer, 2009

Alberto Finzi, Thomas Lukasiewicz, Game-Theoretic Agent Programming in GologLuca Cardelli, Marta Kwiatkowowska, Robustness Guarantee.Yasar Demirel, in Non-equalibrium Thermodynamics (Third Edition), 2014

Matthias Heymann, Eric Vanden-Eijnden, Geometric Minimum Action Metod: A Least Action Principle on the Space of Curves, Courant Institute, 2007

Approximate Model Checking of Stochastic Hybrid Systems: Alessandro Abate, Joost-Pieter Katoen, John Lygeros, Maria Prandini.

Marta Kwiatkowska, Gethin Norman, Stochastics Model Checking

An experimental investigation on the health monitoring of concrete structures using piezoelectric transducers at various environmental temperatures.

Author Robert Balson Dingle, Their Derivation and Interpretation, Academic Press, 1973

J.C., Gittins, Keble College, Oxford, 2010

Ali, Isra, Alfarouk, Khalid O., Reshkin, Stephan J., Ibrahium, Muntaser E., Doxycy-cline as Potential Anti-cancer Agent, Anti Cancer Agents, 2018

G.S. Skone, Irina Voiculescu, Stratagems for effective function evaluation in computa-tional chemistry, Oxford, 2010

Zahid Ur Rehman, Quorum-Quenching Bacteria Isolated From Red Sea Sediments Reduce Biofilm Formation, KAUST,2016

Peter Minary, Charlotte M Deane, Explorong peptide/MHC detachment process using hiearchical natural move Monte Carlo, Oxford, 2015Biancalani, E. Giamperi, A. Bazzani, G. Catellani, and A. Maritan, Phys.Soc, Jpn., Physics, 2015

D.M. Roma, R.A. P'Flanagan, A.E. Ruckenstein, A.M. Sengupta, R., Optimal path to epigenetic switching, Rarah Institute of Physics,2015

C HASTE : incorporating a novel multi-scale spatial and temporal algorithm into a llarge-scale open source librarty B Y M IGUEL O. B ERNABEU 1 , R AFEL B ORDAS 1 , P RAS P ATHMANATHAN 1 ,J OE P ITT -F RANCIS 1 , J ONATHAN C OOPER 1 , A LAN G ARNY 2 , D AVIDJ. G AVAGHAN 1 , B LANCA R ODRIGUEZ 1 , J AMES A. S OUTHERN 1,3 AND J ONATHAN P. W HITELEY 1, *

Chris Holmes Leonhard Held Bayesian Ausiliary Variable Models for Binary and Multinomiak Regression

Michael Betancourt and Mark Girolami Hamiltonian Monte Carlo for Hierarchical Model

rffany Amariuta, Ynag luo, Steven Gazal, Emma E. Davenport, Bryce van de Geijn, Kazuyoshi Ishigaki, Alkes L. Price, IMPACT: Genomic Annotation of Cell-State-Specific Regulatory Elements Inferred from the Epigenome of Bound Transcription Factors.

Alessandro Abate, Joost-Pieter Katoen, Alexandru Mereacrem Quantitative Automata Model Checking of Autonomous Stochastic Hybrid Systems

Previous work : https://github.com/dashboard

Acknowledgments

Although I have graduated with excellence graduates awards,I am still working in the university currently cooperated with university hospital. Thanks to all the supervisors and colleagues I have worked with.

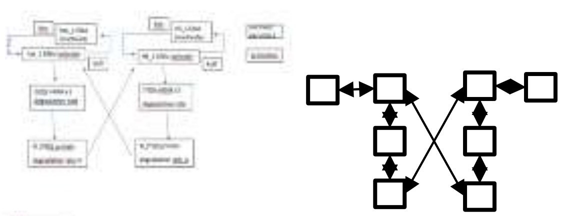
Availability

Please find supplementary data, code and experimental results in this link:
https://github.com/dashboard

Appendix

Figure

S0 = hax1_DNA^inactive ^ hax1 _DNA^low_concentration
S1 = hax1_DNA^active ^ hax1 DNA^high_concentration
S2 = hax1_mRNA ^¬degradate
S3 = hax1_protein^¬degradate
S4 = HS1_DNA^inactive ^ HS1 _DNA^low_concentration
S5 = HS1_mRNA^active ^ HS1 DNA^high_concentration
S6 = HS1_mRNA^¬degradate
S7 = HS1_protein^¬degradate
Transition rate: P/P



Figures 1, figure2

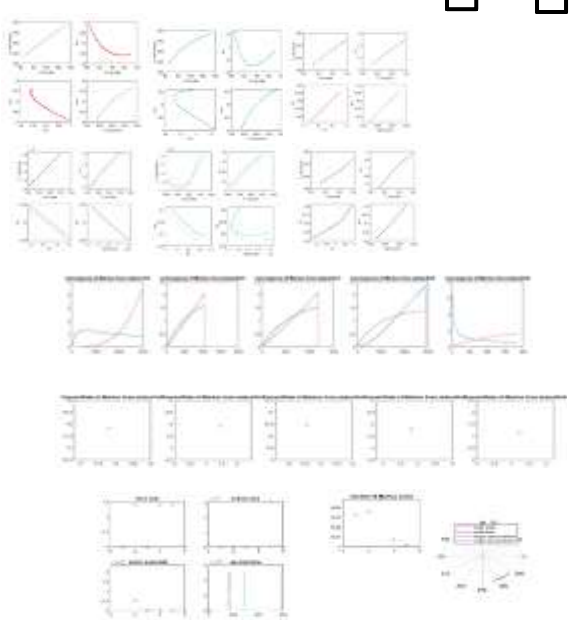


Figure 3 (a-j)

B Proof

As our assumption of diffusion process is based on the proposition of the geometric quasi-potential:

$$V(x1,x2) = 2 * inf \int_{\gamma} |b|_{a} sin^2(\frac{1}{2})\eta dx \quad (4)$$

which later can be used in computing quasi-potential under the case without SDE: $V(x1,x2) = inf_{\phi \in C_{x1}^{x2}(0,1)} S(\phi) with S(\phi) = sup_{H(\phi,\theta)=0}^{\theta:[0,1]->R^n} \int_0^1 < \phi', \theta > d\alpha$ As we have proved the most important invariance of the key four representation of the action S through the same value of upper bound and lower bound of S. The only supplementary required is after the rewrite of V(x1, x2) with S, the inf is computed with the lower-semicontinuity of S and the compactness of certain functions: [illegible]

A. linear regression method on data, (Y = B*X) :
-0.00114 -6.69764 1.11999 4345.69955
3.25582 NA 0.33371

And the analysis summary :
#Coefficients: (1 not defined because of .66singularities)
Estimate Std. Error t value Pr(>|t|)
#(Intercept) -0.00114 0.060137 88 0.001425

#as.matrix(X)V2   1.11999   0.003344   66.79
0.0001425 **
#as.matrix(X)V4   3.25582   0.000215   271.46
0.0001425   **
#as.matrix(X)V5 -0.44712   0.000600   1.02
0.0001425 **
#as.matrix(X)V6   0.33371   0.287314   2e10**2
0.0001425 **

#Residual standard error: 0.231 on 456 degrees of freedom
#Multiple R-squared: 1,   Adjusted R-squared: 0.9946
F-statistic: 1.21e+02 on 2 and 456 DF, pvalue 254
The estimated effect of V2,V4, V5, V6 on the convert is 1.11999 3.25582 , -0.44712, 0.3371.

This means that for every 1% increase in V2 on X, there is a correlated 1.11% decrease in the incidence of Y. Similar to V4, V5, V6

The standard errors for these regression coefficients are very small, and the t-statistics are very large (66.79,271.46 , 1.02 and 200 , respectively). The p-values reflect these small errors and large t-statistics. And for both parameters, there is almost zero probability that this effect is due to chance.(** gives the variance in 0.001)

Auxiliary method to resampled the makorv transition matrix in the mRNA to protein switching model which is P1 before.

1.0095	0.0095	0	0	0	0
0.9634	1.9649	0.0015	0	0	0
0	1.0000e-04	1.0006	5.0000e-04	0	0
0	0	2.0000e-04	1.8680	0.8678	0
0	0	0	0.0362	1.1570	0.1207
0	0	0	0	0	3

And Augmented into P2:

4.8461	2.0063	0	0	0	0
0.9379	0.6672	2.0242	0	0	0
0	2.0063	0.6836	4.2907	0	0
0	0	2.0242	2.3113	6.7428	0
0	0	0	4.2907	3.6798	2.5908
0	0	0	0	6.7428	1.1108

h =0   0   0   0   0   0
p   =[0.3667,   0.4350,0.2222,   0.1495,0.1030, 0.8709]

C.code

```
MaxStep = 50;
```

```
W = eye(size(z));

X = z;

v = var(X)+0.000000001;

for steps = 1:MaxStep

    temp = ones(size(v))./v;

    Vtemp = inv(X'*inv(W)*X+temp);

    V = Vtemp*eye(size(Vtemp))*Vtemp';

    L = chol(abs(V),'lower');

    S = V*X';

    B = S*inv(W)*z;

  % observations for normal with W

    H = X.*S;

    CW = H./(W-H);

    CW(isnan(CW))=0;

    m = X.*B;

    m = m-CW.*(z-m);

    qtemp = CW.*(CW+1);

    qtemp(isnan(qtemp)) = 0;

  % draw Z from truncated normal

    q = sqrt(qtemp*eye(size(qtemp))*qtemp');

    Z = X;

    R = X;

  %Z(:,i) = mvnpdf(X,mean(m,2)',q);

    temp = mean(m,2)';

  for i = 1:size(X,1)

    Z(:,i) = normpdf(X(:,i),temp(i),q(i,i));

  end

  %update B

    B = B +((Z-X)./W).*S;

    B(isinf(B)) = 0;
```

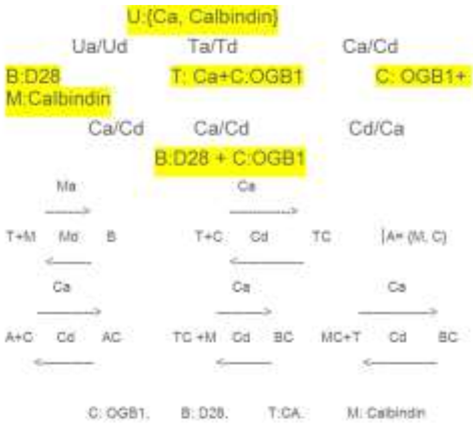
```
%update beta

    beta = B + L*T;

%observations for logistics

    m = beta.*X;

for i = 1:size(X,1)

    temp = makedist('Logistic','mu',mean(m(:,i)),'sigma',abs(std(m(:,i))));

    Z(:,i) = pdf(temp, Z(:,i));

    R(:,i) = Z(:,i)- m(:,i);

end

%sampling lambda

    Y = normpdf(Z,0,1);

    Y = Y.^2;

    Y = 1+(Y-sqrt(Y.*(4*R+Y)))./(2*R);

    lambda = Z;

for i = 1:6

    Ztemp = R(:,i).*Y(:,i);

    Ztemp2 = R(:,i)./Y(:,i);

    lambda(:,i) = Ztemp;

    lambda(Z(:,i)>(ones(size(Y(:,i)))./(1+Y(:,i)))) =

  if mean(lambda(:,i)) > 4/3

    Z(:,i) = rightmost(Z(:,i),lambda(:,i));

  else

    Z(:,i) = leftmost(Z(:,i),lambda(:,i));

  end

end

end

function Z = rightmost(U,lambda)
        X = exp(-0.5*lambda);
        Z = X;
        % squeezing
        for t = 1:length(X)-1
        Z(t) = Z(t) - (t+1)^2*X(t)^(t+1)^2-1;
        t = t+1;
        Z(t) = Z(t) + (t+1)^2*X(t)^(t+1)^2-1;
  end
        Z(Z<U)= 0;
End

function Z = leftmost(U,Lambda)
        H = 0.5*log(2)+2.5*log(pi)-2.5*log(Lambda)-repmat(pi^2,6,1)./(2*Lambda)+0.5*Lambda;
        lU = log(U);
        X = exp(-pi^2/(2*Lambda));
        Z = X;
        K = Lambda/pi^2;
        % squeezing
        for t = 1:length(X)-1
        Z(t) = Z(t) - K(t)^(t^2-1);
        t = t+1;
        Z(t) = Z(t) + K(t)^(t^2-1);
  end
        Z((reshape(H,6,1)+reshape(log(Z),6,1))<reshape(lU,6,1))= 0;
End
```

D. Previous work

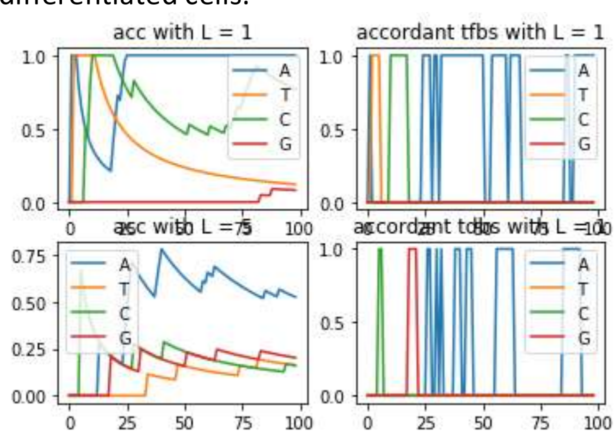
1.Markoc chain of CaMKII circuit with regards to cognitive systems diseases especially about the MC and GC networks around hippocampus and dendrate gyrus. Please see one of my current work report following:

https://www.overleaf.com/read/znrwdjxppyzs

Mainly, I simplify one calciumodulin Activation by Calcium Transients of postsunaptic dendritic spines which finally constructed by 12 equations, first four as the clobe, nlobe binding to medium concentrated Ca and the second module as the similar clobe and nlobe but binding to high concentated Ca binding sites followed by the fast binding of kinase II and the indicator OGB1. As it is the simplified version, its only based on the four independent binding site.

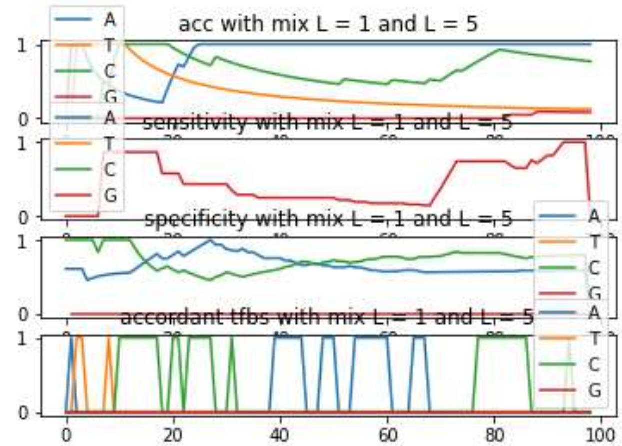


2. Cell reprogramming is usually a time lapse studied through gene expression data and DNA sequence data. Through computing the reprogramming rate, it is shown that more reprogramming happen under the condition of inhibition of DNA methylation or the knowckdown of somatic transcription factors. In the first model, one fixed-variable-order Bayesian tree is constructed for the identification of transcription factor binding sites(TFBSs), while in the second model, the focus is on the expression data of the cell and transcription factors. This is mainly based on the process that, the promoters of ESCs can not only bounded by their own products but also activate other pluripotent genes and inhibit lineage specific genes. Thus, a markov model is applied to induce the reconstruction of transcription regulation in embryonic stem cell states, by the ectopic expression of factors and reprogram the differentiated cells.



To realize the detection of binding site of DNA transcrition factor, wIth homogeneous order 1 and 5 VOM(0.65,10), I first get the binding sites with the highest accuracy prediction peaks.

X = {A,T,C,G}, and thus the d = 4. And using the mix model with pruning on KL scaled by threshold c on log scaled odds , the bayesian tree is for specifically to predict those at binding site(pruned level less than 3).

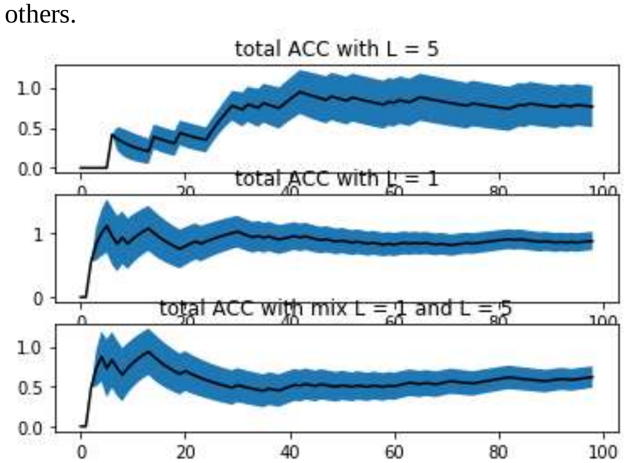


For the foreground dataset, L = 1, and each DNA sequence from X is predicted directly through the one before it while using the most easiest critieria, chossing the one with largest probability.
For the background dataset, L = 5, and each DNA sequence from X is predicted through the previous five sequences before it while using the frequency of the respective subsequencies combining Bayes Theory. And again, the highest peak frequency is chosen to be the biding site with its probability approximated:

$$\hat{P}\left(x_j = x_j \middle| x_{j-L}^{j-1} = x_{j-L}^{j-1}, \Omega_k\right) = \frac{n_k\left(x_{j-L}^{j}\right)}{n_k\left(x_{j-L}^{j-1}\right)},$$

$$P(x_1^N) = \prod_{j=1}^{N} P\left(X_j = x_j \middle| X_{j-L_j}^{j-1} = x_{j-L_j}^{j-1}\right),$$

As the results, generally, the KL converges faster with L = 1 and the mix model of order 1 and 5. In addition, the model is accumulated instead of step-wise, leading to the convergence not rising with the iteration. It is obvious that, on some specific TFBSs, the prediction is even higher than others.



Algorithm: Stepwise Markov Ising model with lineage tree
1. The first step is to configure the epigenetic states of equencies through their expression, whehter "close" of "open" on any temporary cell state. (In our data, there are 64 cell states in all.)

2. The second step is to check the final state of each lineage tree, as the expression of modules maybe conflicted with each other, influenced by other cell's transcriptional regulatory network which might lead to cell death(we denoted as 0) while ϵ denoted as the last state to the last states, which making 0 and n(66th) as the absorbing states.

3. The transition matrix is built according to reference [2]:

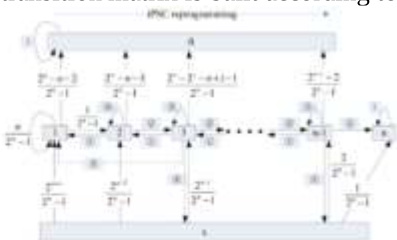
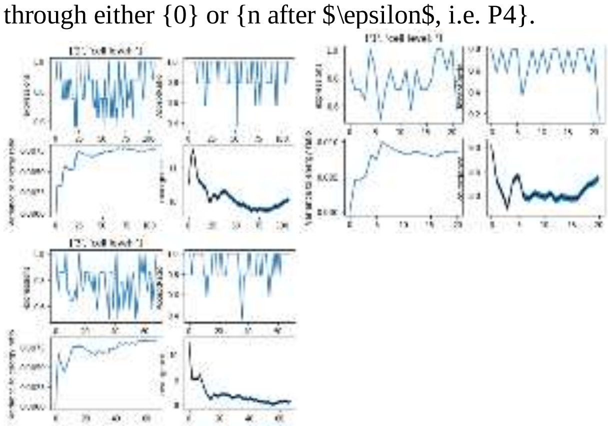


Figure 2 Reprogramming Markov Chain

Note that in the first step, The transition states predicted through the highest probability. And in the second step, either probability(0 or ϵ and further n) over 0.5 will lead to the end of the markov process.

Basically, it is based on the Ising model, choosing direction among {P1,P2,P3,P5} instead of the original random spin choise from {-1,1} and the final fate of the cell is predicted through either {0} or {n after ϵ, i.e. P4}.



Among the 200 simulations, first level cell expression takes fewest amounts while the highest convergence.(Although the variance to energy ratio are quite similar for the three levels.).

Generally, the Accept ratio is higher after the expression level goes to the lowest which is either when inhibition of DNA methylation exhists or the knowckdown of somatic transcription factors occurs.

[1]Novel Markov model of induced pluripotency predicts gene expression changes in reprogramming Zhirui Hu, Minping Qian, Michael Q Zhang*

[2]Identification of transcription factor binding sites with variable-order Bayesian networks, I. Ben-Gal,∗, A. Shani1, A. Gohr, J. Grau, S. Arviv1, A. Shmilovici, S. Posch, and I. Grosse