Modelling, Kinetics and Computation of bio or biomimetic process

A Thesis
Submitted in partial fulfilment for the Degree of

Master of Science

as a part of Integrated Ph.D. programme

(Material Science)

By

Brijesh Saraswat

CHEMISTRY AND PHYSICS OF MATERIALS UNIT

JAWAHARLAL NEHRU CENTRE FOR ADVANCED SCIENTIFIC RESEARCH

(A DEEMED UNIVERSITY)

JAKKUR, BANGALORE-560064 INDIA

APRIL 2019

To my Grand Parents

Declaration

I hereby declare that the matter embodied in the thesis entitled "***Modelling, Kinetics and Computation of bio or biomimetic processes***" is the result of investigation carried out by me at Chemistry and Physics of Materials Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India under the supervision of **Dr Meher K Prakash** and that it has not been submitted elsewhere for the award of any degree or diploma.

In keeping with the general practice in reporting scientific observations, due acknowledgement has been whenever the work described in based on findings of other investigators.

_____

Brijesh Saraswat

CERTIFICATE

I hereby certify that the matter embodied in this thesis entitled "***Modelling, Kinetics and Computation of bio or biomimetic processes"*** has been carried out by Mr Brijesh Saraswat at Chemistry and Physics of Material Science Unit, Jawaharlal Nehru Centre of Advanced Scientific Research, Bangalore, India under my supervision and that is not submitted elsewhere for the award of any degree or diploma.

_____

Dr Meher K Prakash
(Research Supervisor)

Acknowledgements

First of all, I want to thank my supervisor Dr Meher K Prakash for his constant guidance and motivation all through the research work. The work which I have done, would not be possible without his help and advice. He always provides me with an enormous amount of freedom in the workplace. It's always been a great pleasure and opportunity for me to work with him.

I am thankful to Prof. Udaykumar Ranga and Prof. Subi George for posing very interesting theoretical problems based on their experiments, and valuable discussion.

I am grateful to all professors of CPMU, TSU for their valuable and fruitful teachings. It has been a pleasure to learn from all of them.

I am very grateful to all the members of the Computational Biology Group who helped me in many ways during this project, and for all the fun we had in the last one year. My sincere thanks to seniors Sruthi, Malay for helping me in coding, correcting my spoken English and other technical issues. I would like to thank my labmate Himanshu and Ayan (always exited for food) for having a jocular discussion and for their company for snacks (momos and samosa).

My heartfelt thanks for Dr Somalya for his PJ's having a broad discussion about the craziest and funny things (cricket, football, politics and science) in the world.

I am profoundly grateful to Dr Sandhya and Dr Antrip for arranging a house party at their apartment and encouraging me to focus.

The life outside the lab was made enjoyable by the good presence of my integrated PhD batch-mates. I thanks all my batch mates for making life easy and cheerful. Raagya, Pragya for providing their decorated notes for exams and always being there when need, Shashank who always have suggestion for good movies and book, Ashutosh who has been a good room-mate for 2 years, Nijita for her crazy behaviour after sunset, Tarandeep for all memorable moments like in organic chemistry exam and discussion in cab. Shivani for calculating all bills of trips and Dinners in our initial days of JNCASR. Harshit, Geetika, Sushmita, Aditi for helping in Chemistry course work.

I would like to thank batch-mates Amit, Shubham, Ankit for making feel that I am not far from Haryana, Chavi, Rashi, Irine, Sohali and Jindal Di for helping me to understand Biological systems.

I am grateful to my Integrated PhD seniors (Rajendra, Narendra, Lakshya, Sukanya, Jankyi) for their guidance about course work and experiments.

My heartfelt thanks for Dr Sutanuka, Shrestha, Vijeta, Shikha for helping me in learning Biological and chemical expects of my MS projects.

I am grateful to senior Pawan Bhaiya, Anand Bhaiya, Uttam Bhaiya, Abhiroop Bhaiya for sharing their pragmatical knowledge.

I would like to thanks to Navneet, Sajith, Krishna for teaching me volleyball.

I want to use this opportunity to thanks my friends from school Aman, Bhandari, Samriddhi to encourage me to study science. My undergraduate friends Aditya, Sachin, Mishra Ji, PK for encouraging and helping with resources to join JNCASR.

I would like to thank my parents and relatives for their blessings and scoldings, my brother and sister for their support and encouragement without all this reaching this level would not easy.

Last I cannot forget to acknowledge JNCASR and staff of JNCASR (specially Complab people Chandan, Uday) for their help in setting my workstation and resolving the problems with the workstation.

List of Abbreviations

RD	Reaction Diffusion
MD	Molecular Dynamics
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
HIV	Human Immunodeficiency Virus
ART	Antiretroviral Therapy
TAT	Trans-Activator of Transcription
LTR	Long Terminal repeat
GFP	Green Fluorescent Protein
ATP	Adenosine Triphosphate
ADP	Adenosine Diphosphate
CK	Creatine Phosphokinase
HK	Hexokinase
PCr	Phosphocreatine
Cr	Creatine
G	Glucose
G6P	Glucose-6-phosphate

Summary

Chemical or physical interactions among the molecular species in living cells define their intricate functioning. Thermodynamics and kinetics are used to understand how feasible these biological processes are and how fast they happen. Kinetic modeling is a well established framework. Here in chapter 1, we talk about some of fundamental and regularly used models to develop our understanding toward biological system and their behavior. Stochastic, deterministic and Michaelis-Menten models which will be used in the other chapters of this thesis.

One of the biggest problems while treating a HIV patient through anti retroviral therapy is the unpredictable latency of the virus, which leaves the drugs ineffective. Previous studies have shown that TAT protein plays a positive role in deciding virus' active and latency state. In chapter 2, we studied the active-latent cycle of HIV virus through the cooperativity of TAT protein in feedback circuit. Our simulations shows, greater the cooperativity (stronger promoter) higher the chances of the virus going in latency state and staying in latency state for longer time. This observation is interesting for comparing the latency across HIV subtypes.

In living cells, in addition to the covalently bonded polymers such as proteins or DNA, there are other polymers that form dynamically, purely by non-covalent associations. The polymerization is driven mainly by ATP, ADP, GTP, without which the polymer may become thermodynamically unstable. Polymerization can thus be controlled by changing the availability of, for example, ATP in the solution. In Chapter 3 we study fuel driven association, dissociation phenomenon and how their structural properties can be controlled using biofuels like ATP, ADP.

List of Figures

List of Tables

Contents

Chapter -1

Introduction

1.01 Chemical Kinetics for Biological systems

Living cells embody complex interaction networks with a millieu of reactions among the large numbers and types of molecules. Several distinguishing characteristics of living systems are that they are far from equilibrium, are inherently multi-component, with several important species, many of which are in low copy numbers. Chemical reactions are the backbone to the functioning of living cells. From basic reactions such as the synthesis of ATP from glucose to the synthesis of proteins or replication of DNA, most the important phenomena are chemically driven. Diffusion of the molecules, physical or chemical interactions among them and the timescales of the processes regulate the functioning of the system. Understanding, the mechanism behind these functions, is necessary both for learning about the basic biology and for research in areas like drug discovery, biotechnology. We illustrate examples of the necessity for understanding reactions in a few biological phenomena below -

Unicellular organisms such as *E. coli* show an interesting phenomenon called chemotaxis. On average, they drift towards nutrients (chemoattractants) or away from toxins (chemorepellents). But this drift evolves out of a biased random walk, in which, if conditions appear favorable, based on the density of the nutrient molecules that are sensed by the bacterium and the random change in direction (tumble) will be delayed. This biased random walk eventually leads bacteria to chemoattractant. While this phenomenon may be described as a bacterial 'strategy', the decision making is driven by sensing molecules in low copy numbers. The mechanisms of nutrient sensing by (un)binding, the timescales over which chemical modifications such as methylation that follow it last to constitute a memory, and the mechanisms by which the signals drive the motor proteins are all inherently molecular in nature. Developing a molecular understanding of any of these steps, and understanding how fast biological systems respond and adapt to environmental changes[1][2][3] requires modelling the phenomenon and following the kinetics of the different processes.

There are several aspects of the formation of patterns in developmental biology that hinge on the gradients of chemicals, commonly known as morphogens. Morphogen gradients guide the chemical species to diffuse and react, eventually generating patterns required for cell differentiation. Alan Turing explained these self-regulating pattern systems in 1952[4] by using the reaction-diffusion (RD) model during the development of plants and animals. Dynamic patterns in strips of fish *(Pomacanthus imperator)* arising from the interactions between pigments in fish skin, were one of the first observations which were also predicted by the RD model. The patterns on zebrafish were explained by RD models much later in 2009[5][6].

At the microscopic level, the interaction between molecules like protein, nucleic acids, DNA, RNA defines the functioning of the cell. In such systems, low concentrations of the molecular species give rise to thermal fluctuation in the system (gene transcription, signalling). To understand such a system, one needs models which are an intrinsically random, with predictable average properties. Stochastic models are used to study gene regulatory circuits in cells[7], to explain stochastic stepping motors observed in myosin, kinesin on F-tubulin and microtubules which hydrolyse ATP molecules, etc.

Development of drugs for infectious or non-infectious diseases is extremely challenging - from the design of molecules which can bind their enzymatic targets to planning their controlled release. Binding kinetics of the drugs to the enzymatic targets and their affinities tell us about the efficacy of the drugs against blocking the access to their critical enzymes in pathogens and in quantifying their role as antibiotics [8]. Different drugs and therapies are being developed to cure cancer. Developing efficient treatment strategies require deterministic and stochastic models to describe the growth of cancer cells and the effect of therapies on their growth. Cancer starts with localized tumour growth, and may metastasize by spreading to other regions in the body through the bloodstream. Some of the models compare how effective are different therapies are to treat cancer cells and the rate at which tumour spread under the influence of different therapies and drugs[9][10]. The utility of chemical reaction models, with Michaelis Menten kinetics[11], is very important for understanding the hydrolysis of cellulose to glucose in the conversion of lignocellulosic biomass (dry plant waste) to biofuel[12].

Kinetic modelling is needed in several instances such as adaptive bacterial metabolism, kinetics of enzymes, chemical reactions inside cells especially in individual compartments.

Enzymes enhance all these essential reactions rates in a biological system. Kinetic study of these reaction tells us about the thermodynamic state of a system like what is driving these reactions, in which direction reactions will proceeding. Another example of the importance of kinetic study in a biological system is for drug development and discovery. It tells how long drug will bind to the enzyme and what will be the rate of its release under a different condition such as pH, temperature.

Thus whether it is for understanding basic cell biology biology or for the development of drugs and strategies for disease treatment, there is a very significant role to understanding the key chemical species that are involved, modeling their physical or chemical interactions and quantifying their biological effects.

1.02 Kinetic modelling

To understand the feasibility or progress of reaction pre-knowledge of thermodynamics and kinetics is required. The field of chemical kinetics originated in the second half of the nineteenth century from the work of several pioneering scientists such as **Arrhenius** and **Van't Hoff**. Thermodynamics tells us the direction in which reaction proceeds with a change in energy and entropy, and explains about driving force for a reaction. Processes which are favored thermodynamically occur spontaneously ($\Delta G<0$). Kinetics is more detailed, it tells us about the conditions under which processes happen and at which rates. Thermodynamic relations can be derived from the kinetics. Kinetic theories and models evolved from unimolecular to describing interactions in multicomponent reactions. With initial conditions of the system and these differential equations, the evolution of spatial and temporal patterns of the different molecular species can be defined.

Law of mass action: The most fundamental kinetic models are based on the law of mass action in which the rate of the reaction is proportional to the activities of chemical reacting species.

Michaelis-Menten Model: Most chemical reactions have high activation energy and in living systems enzymes act as catalysts and accelerate the reactions by many orders of magnitude by stabilizing the products or transition state. However, as with any catalyst, the enzymes themselves remain unaltered after the reaction. Thus, new approaches were required to eliminate explicit dependence of the enzymes from the mass action models. These phenomena were first modelled in 1913 by Leonor Michaelis, and Maud Menten (known as Michaelis-Menten model) while studying hydrolysis of sucrose into glucose and fructose. In this model, enzyme binds to the substrate to form complex product and further

to the product and release enzyme-free again for binding another substrate. In the overall reaction of the substrate (*S*) binding to enzyme (*E*) and eventually forming product (*P*) via the intermediate (*ES*),

$$S + E\text{-} \rightleftharpoons ES \rightarrow P + E$$

under the assumptions that the initial concentration of the product is zero, and the concentration of intermediate complex reaches steady state, an effective reaction rate for the product

$$\frac{d[P]}{dt} = -\frac{V_{max}[S]}{Km + [S]}$$

k_1,k_2,k_3 are forward, reverse and the final rate of reaction and V_{max} is the maximum velocity of reaction given as V_{max} = k_3E; Km = $k_1/(k_2+k_3)$ is the Michaelis-Menten constant.

Hill equation model. Hill equation was first developed by A. V. Hill in 1910 to describe the equilibrium relation between oxygen and haemoglobin. Hill equation is a model for the cooperative binding of multiple ligands on the receptor to produce the product. This model is extensively used to analyse drug receptor relationship and enzyme activity.

$$nS + E\text{-} \rightleftharpoons nSE \rightarrow nP + E$$

$$\frac{d[P]}{dt} = \frac{V_{\max}[S]^n}{K_m^n + [S]^n}$$

Michaelis-Menten model is a special case of the Hill model with *n*=1.

These mathematical models such as these have been used over a century for successfully modelling the kinetics of chemical reactions. The differential equations with multiple components have been routinely solved for the specific initial conditions, in different condition like pressure, temperature and presence of catalysts/enzymes.

Deterministic Models: Solving the coupled differential equations for the chemical kinetics in most conditions are fairly straight forward. A solution of these differential equations gives a trajectory at the systems level. Due to their predictability, or rather lack of any random components, the predictions of these models are smooth, continuous and have no noisy signatures in them. They may be easily mapped to the ensemble level measurements. Deterministic models are efficient in terms of computational power and time. Although depending upon the initial conditions, and some times sensitivity to them, different chaotic patterns can emerge. However, the nature of the modeling requirements have changed

especially past 2 decades. As more intricate experiments at the single molecule level or low copy number level started revealing interesting phenomena, newer approaches were required to model and perform computations.

Reaction-diffusion model: Most of the standard reaction and mass action models assume the system is well mixed and all components homogenously spread across the reaction medium. However this is not true especially in living systems, where the molecules may be synthesized at one place and reaction happens elsewhere. The spatial drift of these species is mostly spontaneous, driven by diffusion. When either the numbers of molecules are low or the timescale of diffusivity is comparable or slower than the reactions, it becomes important to simultaneously consider both the diffusion and the reaction. These are typically modeled using a Reaction-Diffusion (RD) formalism simultaneously studying both the effects. The spatial distribution of the molecules and their reaction kinetics is extremely important in several biological reactions, pattern formation, etc. In 1952 Alan Turing proposed this model to describe the pattern formation in a natural system using this model. This model shows the pattern formed by oscillating chemical reactions and diffusion in the system.

Stochastic Models: Much of the standard reaction modeling and rate constants relies on having Avagadro number of molecules. When the number of molecules are fewer, it becomes important to make a distinction about whether a pair of molecules reacted or not rather than remain with the probability that the reaction occurred. For example, in many cellular processes such as gene expression, with a single DNA and very few promoter molecules, the outcomes are expected to be stochastic. Exactly under the same conditions the gene may or may not get expressed in different cells, or it may happen at a very different timescale.

This inherent difference resulting in spiky outputs with integral increments in the number of output molecules are modeled stochastically. In stochastic models, kinetics of the system is studied on the probability-based selection of event, among a multitude of options. Randomly events are chosen from a system based on concentrations and the rates of reaction. This random selection of events from the system takes care of fluctuation observed in the system. Because of the intrinsic random nature of the model, every repetition of the calculation gives a different trajectory with the same kinetic parameters, which is not the case with the deterministic model. On the otherhand a stochastic models recognize the fact that all reactions are not happening continuously when the numbers are low but there are certain reaction events, separated by long times with no reaction.

Stochastic models require more computational time as compared to the deterministic model.

Molecular dynamics: One may wonder if the methods of solving differential equations may be too much behind advanced computational methods such as Molecular Dynamics (MD). MD is a standard computational technique to study the time evolution of the molecular system at the atomic level. The interatomic interaction potentials between bonded and non-bonded atoms are defined. The forces on atoms arising from these forces, and their time trajectory when they have an average thermal energy are studied using numerical algorithms using Newtonian mechanics. MD simulations are commonly used for studying proteins, lipid membranes, nucleotides and their interactions among themselves or with other molecules. MD can be used for a virtual *in silico* screening of the drug candidates for a rational design of antibiotics, for example. Large scale computational resources are employed to model atom level details of how interactions happen and the mechanisms that drive biological phenomena can be unravelled. Since these methods deal with low copy numbers as well, the results are stochastic. While classical MD itself does not lead to bond formation or breaking, it can be effectively combined with quantum mechanical calculations in hybrid formulations known as QM/MM to address this short coming. In a sense MD or QM/MM appears to address all the factors that need to factored into modelling. However the extremely fine level of details that are included in MD making the study of large systems or for longer times forbiddingly expensive in computations. Most state-of-the-art computations reach a few microns in size, and a few microseconds in simulation time. This limitation leaves a huge gap in addressing most systems-level biological phenomena.

1.03 Focus of the thesis

Chemical reations in homogenous, heterogenous, at low copy numbers and with poor diffusivity are extremely important for understanding some of the critical processes like gene expression, pattern formation in living systems. As newer experimental challenges appear with extremely fine level of detail and knowledge of living cells, newer models which can capture these phenomena are required. This requires an incorporation of all the relevant details, computation of the effects, comparison with the experiments for gaining an understanding into the functioning of living cells.

Although some of the tools of chemical kinetics appear to be older than a century, it is clear that a different way of treating them, with low molecular concentrations is extremely relevant today to understand some of the cutting edge experiments.

In this thesis, we demonstrate modeling and computation in two such studies - one in understanding one of the critical steps in the disease biology of HIV and another for understanding how driven self-assembly may be controlled.

1.04 References

1. Sourjik, Victor, and Ned S. Wingreen. "Responding to chemical gradients: bacterial chemotaxis." *Current opinion in cell biology* 24.2 (2012): 262-268.
2. Hansen, Clinton H., Robert G. Endres, and Ned S. Wingreen. "Chemotaxis in Escherichia coli: a molecular model for robust precise adaptation." *PLoS computational biology* 4.1 (2008): e1.
3. Min, Taejin L., et al. "Chemotactic adaptation kinetics of individual Escherichia coli cells." *Proceedings of the National Academy of Sciences* 109.25 (2012): 9869-9874.
4. Turing, Alan Mathison. "The chemical basis of morphogenesis." *Bulletin of mathematical biology* 52.1-2 (1990): 153-197.
5. Nakamasu, Akiko, et al. "Interactions between zebrafish pigment cells responsible for the generation of Turing patterns." *Proceedings of the National Academy of Sciences*106.21 (2009): 8429-8434.
6. Kondo, Shigeru, and Takashi Miura. "Reaction-diffusion model as a framework for understanding biological pattern formation." *science* 329.5999 (2010): 1616-1620.
7. Vilar, José MG, Călin C. Guet, and Stanislas Leibler. "Modeling network dynamics: the lac operon, a case study." *The Journal of cell biology* 161.3 (2003): 471-476.
8. Tonge, Peter J. "Drug–target kinetics in drug discovery." *ACS chemical neuroscience* 9.1 (2017): 29-39.
9. Jain, Siddhartha. "Kinetic model for designing a cancer therapy." *Cancer Cell International* 2.1 (2002): 13.
10. Kolokotroni, Eleni A., et al. "Studying the growth kinetics of untreated clinical tumors by using an advanced discrete simulation model." *Mathematical and Computer Modelling*54.9-10 (2011): 1989-2006.
11. Kolomeisky, Anatoly B. "Michaelis–Menten relations for complex enzymatic networks." *The Journal of chemical physics* 134.15 (2011): 04B613.
12. Bansal, Prabuddha, et al. "Modeling cellulase kinetics on lignocellulosic substrates." *Biotechnology advances* 27.6 (2009): 833-848.

Chapter-2

Understanding the origins of latency in HIV

2.01 Introduction

Human immunodeficiency virus type 1 (HIV-1) is a pandemic that affects millions of people worldwide. Despite its incredible genetic diversity[1,2] antiretroviral therapies (ART) have been successful in suppressing HIV[3,4] to undetectable levels. However, the treatment has to be continued for life as interrupting ART leads to a rebound. One major obstacle to a complete treatment of HIV has thus been the presence of a latent reservoir[5-10] of infected cells, which may get reactivated. In fact, upon infection of $CD4^+$ T cells, HIV stochastically goes into either the active or the latent state. Conceptually the onset of latency and re-activation may be driven by the environmental cues or by autonomous programs[11]. Genetic noise is a common phenomenon[12] resulting in different phenotypes from identical genetic constitution, and the role of stochastic noise with low copy numbers of the different species was studied[11]. It is now established that the factors underlying the viral latency and reactivation are hardwired in HIV[13,14] and can be independent of the activity of the host cell.

Further, at the molecular level, the role played by TAT positive-feedback circuitry in reactivation as well as in enabling and sustaining latency has been clearly demonstrated using a minimal synthetic circuit in which HIV TAT amplifies the expression from HIV LTR[13]. TAT is recruited to the HIV promoter by binding to TAR, a stem-loop RNA encoded by the viral mRNAs. The phosphorylation of a variety of proteins within the elongating transcription complex, follow as a consequence of this interaction between TAT and TAR. Computational models were developed analogous to the two-state transcription models[15,16] studied earlier. These computational models could demonstrate the decoupling of HIV activity from the cellular state[13] as well as studied the reactivation from the latent state[14]. The latter study also discussed the role of the TAT self-cooperativity in reactivation from latency, with different Hill coefficients *H*. However, the available data

did not necessitate or support the possibility of a variable *H*. While the TAT positive-feedback was identified to be an important factor, the role of the strength of self-cooperativity in the establishment, sustenance of latency and reactivation from it in synthetic constructs or in naturally occuring viral variants was not studied.

In this work, we use combined computational models to study the two-step effect of how the promoter strength influences the degree of self-cooperativity, which inturn affects the viral latency. Our motivation for the studying effects of the promoter strength was two-fold: having the promoter strength as a tunable parameter may allow one to observe newer patterns in latency characteristics as well as to relate these learnings to infer the differences in the latency of the different HIV subtypes. HIV subtype C with a global prevalence of about 50% commonly found in Africa, India and parts of China is characteristically different from the well studied HIV subtype B which found in the infections in America and Europe. A notable difference between the two subtypes, among several other factors, is the presence of a strong promoter in subtype C (3- or 4-NF□B binding units, compared to the 2 in HIV subtype B)[17]. Intriguingly, despite this strong promoter HIV subtype C stays more latent compared to subtype B. In our exploration of the effect of promoter strength, we find interesting and seemingly paradoxical results of the stronger promoter leading to early and longer, which we interpret using the computational model.

2.02 Model

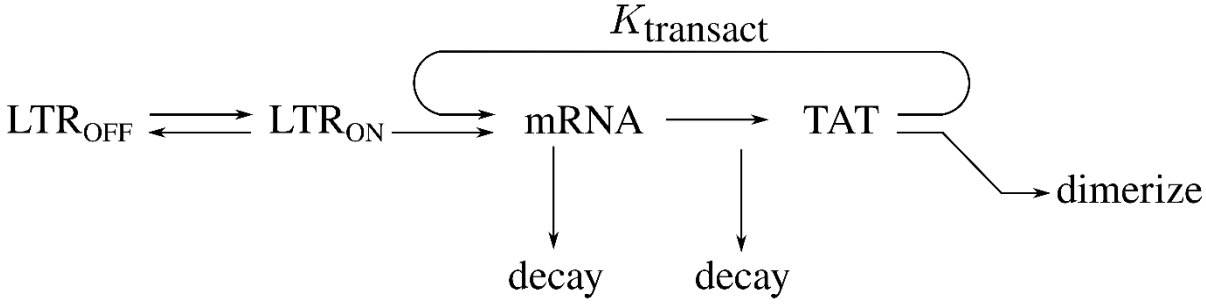

Figure 2.1 A schematic of how the gene expression of HIV TAT protein occurs is shown. The schematic is adapted from Razooky et al. [14]. GFP is to track the host cell activity. HIV is integrated into the host genome, and TAT expression is supposed to be representative of the complete HIV transcription. Further details of the gene expression can be inferred from the description of the equations given below.

The circuit that represents the gene-expression and positive feedback in TAT expression is shown in Figure 2.1. Biochemical reactions and rates of reactions were taken from Razooky[14]. They were adapted by adding dimerisation of TAT molecules, as well as adding a self-cooperative TAT based viral expression. The equations are given below.

LTR → LTR + mRNA	k_{basal}= 1e-5
LTR + TAT → LTR + TAT + mRNA	$k_{transact}$
mRNA → TAT	k_{TAT}= 0.00132
mRNA → GFP	k_{GFP}= 0.5
TAT + TAT → TAT_TAT	k_{dimer}= 0.1188x1e-3
mRNA → x	k_{mRNA_decay} = 4.8x1e-5
TAT → x	k_{TAT_decay}= 4.3x1e-5
GFP → x	k_{GFP_decay}= 0.301x1e-5
TAT_TAT→ x	k_{TAT_TAT}=4.3x1e-5

In the above set of reactions, the first reaction represents transcription from LTR (Long terminal repeats) promoter to form mRNA at the base rate. The second reaction represents an enhanced transcription rate (greater than base rate) from LTR due to positive feedback from TAT (Trans-Activator of Transcription) protein. Reaction three represent the translation of mRNA; information encoded in mRNA is decoded by the ribosome to form a specific protein. In the third reaction, we assume that one mRNA is translated to one TAT protein and it decays as translation finishes. The fourth reaction is same as of the third reaction, instead of TAT protein GFP protein is produced by decoding mRNA. GFP protein is used to mark cells and identify which cells are active and which are inactive. The fifth reaction represents the oligomerisation (dimerisation) of TAT molecule, or this TAT molecule may be useful for some other function in the cell, but it does not help in the transcription process of the cell. Last four reactions in the system shows the degradation of mRNA, TAT, GFP, TAT_TAT molecules in the cell. Transcription rate for a cooperative model of HIV was calculated using the Hill equation. Initial velocity for transcription was kept constant to calculate the transcription rate for different cooperative (Hill coefficient) values in the cell.

$$k_{transact} = \frac{a}{Km^H + TAT^H}$$

Where a is the maximum velocity of reaction and K_m is Michaelis- Menten constant and H is Hill coefficient.

TAT molecule binds with a promoter and gives mRNA and TAT; again free TAT molecule can bind to the promoter; this cycle is positive feedback cycle[14]. TAT positive feedback rate was calculated using the Hill coefficient model[21]. By using the Hill coefficient (H) to the feedback cycle. There can be more than one site possible on promoter for TAT binding and activating promoter to produce mRNA. H=0 means there is no site for TAT to bind on the promoter, production of mRNA is only because of the basal rate. H=1 means one TAT molecule can bind to the promoter and enhance the production rate of mRNA production. For H=2, two TAT molecules with the promoter and increases the rate of production of mRNA.

$$\frac{d[TAT]}{dt} = \frac{V_{\max}[TAT]^H}{k_m^{TAT} + [TAT]^H}$$

We derived the co-operativity parameters H=1 and H=2 for the case of 3 or 4 NFκB promoter sites from the experimental data provided to us by the group of our experimental collaborator Prof. R. Udaykumar [S. Chakraborty, PhD Thesis, JNCASR]. Stochastic simulation of chemical reaction (Gillespie Stochastic Simulation Algorithm) is used to model the biochemical system of HIV.

TAT molecule binds with a promoter and gives mRNA and TAT, again TAT molecule can bind to promoter this cycle is positive feedback cycle [14]. TAT positive feedback rate was calculated using the Michaelis - Menten model[21]. By using the Hill coefficient (H) to the feedback cycle. There can be more than one site possible on promoter for TAT binding and activating promoter to produce mRNA. H=0 means there is no site for TAT to bind on the promoter, production of mRNA is only because of the basal rate. H=1 means one TAT molecule can bind to the promoter and enhance the production rate of mRNA production. For H=2, two TAT molecule binds with the promoter and increases the rate of production of mRNA.

2.03 Simulation

Gillspie algorithm: Since the copy numbers involved in the gene expression are extremely low, the modelling of host cell activation and HIV expression was performed as a stochastic model using Gillespie Algorithm[20][22]. Some of the details are as follows.

In this system, the N number of chemical species of each type are uniformly distributed in fixed volume V at any time t. These N species can interact through M specified channels of chemical reactions at a constant temperature. Quantity propensity function is calculated for each reaction α_i, and total propensity function is $\alpha_0=\sum\alpha_i$.

a) Generate two uniformly distributed between zero and – one to make a decision on which reaction will happen and the other for when it will happen
b) Calculate the time for the next reaction to take place at $(t + \tau)$ using random number r_1. Where

$$\tau = (1/\alpha 0) * \ln(\frac{1}{r1})$$

c) The chemical reaction is selected from a random number and propensity function $\alpha_i/\alpha_0<r_2<\alpha_{i+1}/\alpha_0$.
d) Species molecular number following this reaction are updated at time $t + \tau$.
e) Iterate

2.04 Results

Cycling between active and latent states: Stochastic simulation was performed to check for the bi-stable state of the virus with a positive feedback cycle as there is low reactant concentration. An initial number of TAT molecules were taken 5, the number of GFP was 25000 and number of Promoter was 1 and rest all species were initially zero (concentration adopted from Razooky et al. [14])

In the absence of a dimerization term, the TAT concentration monotonously increased, regardless of the promoter strength without showing any signs of latency (data not shown). TAT molecule is also found in dimer form and is non-transcriptional[18]. We added a dimerisation of TAT molecule term in the previous set of equations to consider the transformation of transcriptional TAT molecule to non-transcriptional TAT molecules inside the cell. The rate of dimerisation of TAT is assumed to be about ten times slower than the rate of formation of TAT from mRNA. Simulation of the biochemical reaction was done with the additional reaction of dimerisation of TAT molecule for Hill coefficient H=1,2 and 3. H=0 was not considered because for H=0 there will be no cooperativity from TAT molecules toward LTR promoter. Adding dimerization resulted in the HIV activity cycling between latent to active states as shown in Figure 2.2

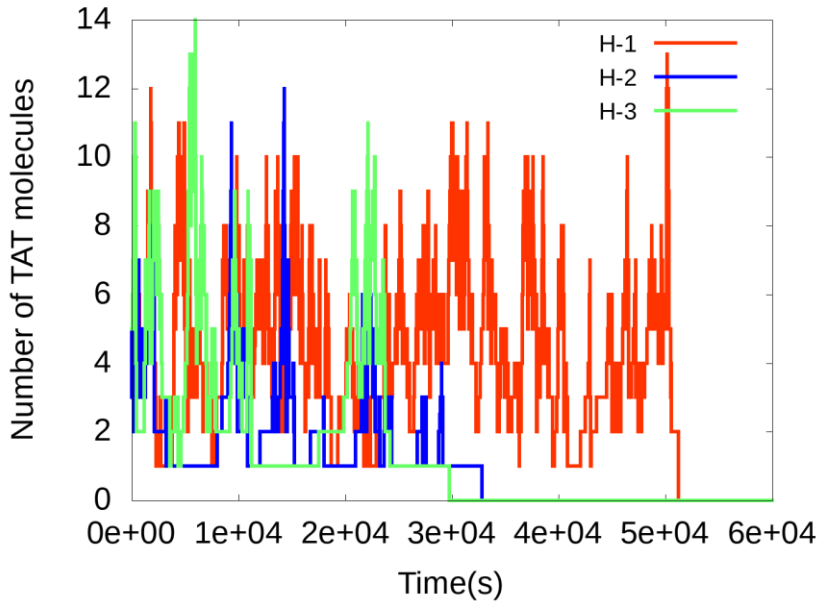

Figure 2.2 Cycling in TAT levels was observed after adding dimerisation of TAT. The number of TAT molecules are less than cutoff then the virus is in latency state (sleep state) if the number of TAT molecules is greater than cutoff than it is the inactive state.

After considering non-transactivating TAT molecules in the system, the noisy nature of graph shows bi-stable states of the virus, which can be defined by considering ‘x’ number of TAT molecules as a cut-off for two states. If the number of TAT molecules is greater than ‘x’ then the virus is an active state; otherwise, it is in latency (inactive or sleep) state. Here also we observe more production of TAT molecules for higher cooperativity. We can see that the virus goes in longer latency cycle for higher cooperativity (strong promoter)

system. The maximum value for TAT molecules in a cell for H=1 is 13 while for H=2 is 28.

Effect of cooperativity on latency state: Normalised time distribution of latency state was plotted for Hill coefficient H = 1 and 2. Chances of finding a virus in long latency decreases as latency time increases for both strong (H=2) and weak (H=1) promoter.

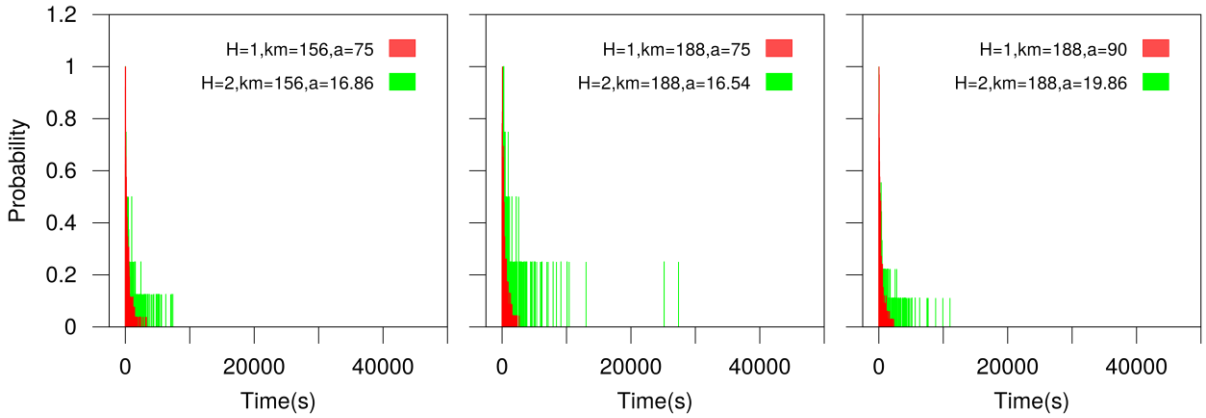

Figure 2.3 Normalized time distribution of latency state H=1(Red) and H=2(Green) for a different set of parameters. The cut-off for the virus to be in latency state is two TAT molecules. Samples were recorded for 60 days.

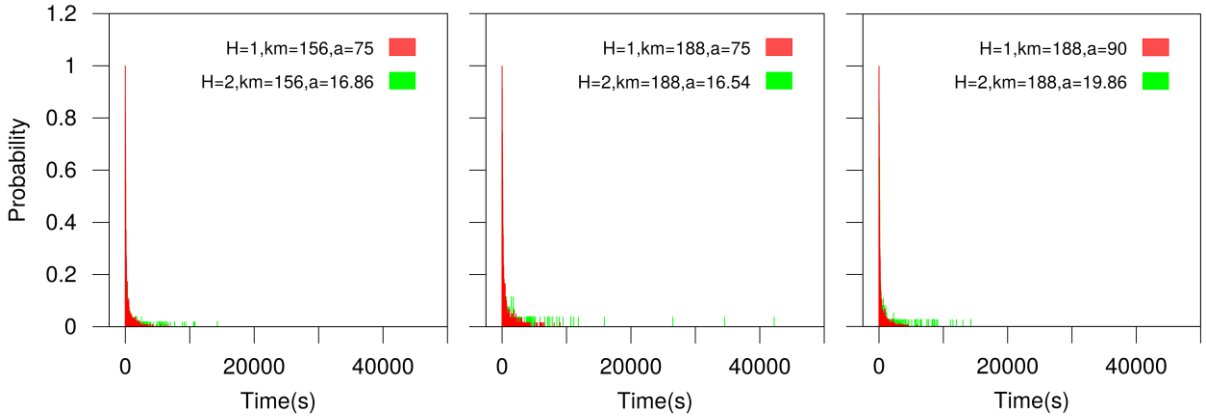

Figure 2.4 Normalized time distribution of latency state of H=1(Red) and H=2(Green) for the different set of parameters, five TAT molecules as a cut-off for the virus for latency state. The sample was recorded for 60 days.

From Figures 2.3 and 2.4, longer latency state period for virus hiving strong promoter (H=2) is more as compared to the virus having weak promoter (H=1) with two TAT molecules as a threshold for switching from latency to active state. It can be seen that for the strong promoter (H=2) results in many more long-lived latency states than a weak

promoter (H=1). With a five TAT molecules threshold, and within the simulated data, there is not much difference in latency state for strong (H=1) and weak (H=2) promoter(from Figure 5).

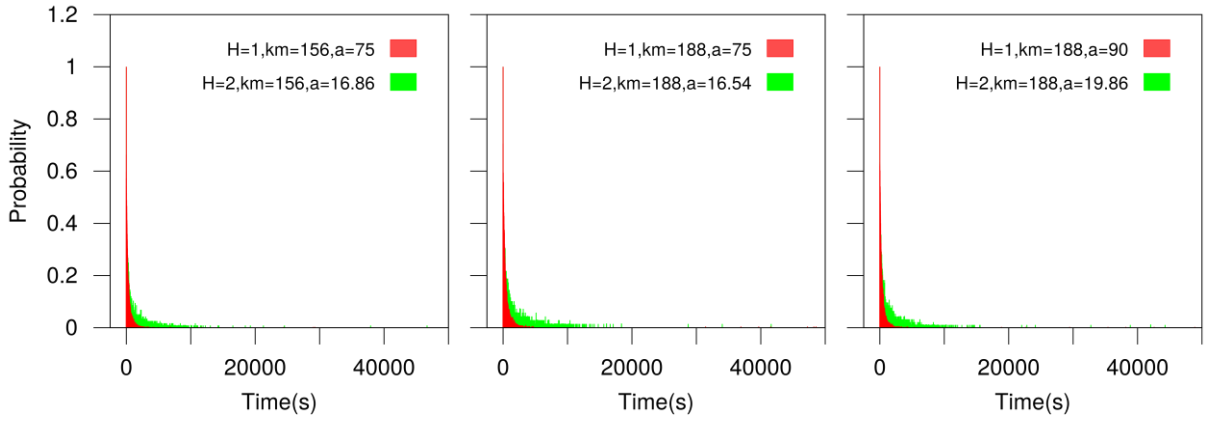

Figure 2.5 Normalized time distribution of latency state for cut-off as two TAT molecules. 5000 number of samples were taken for distribution. The sample was run recorded for 15 days.

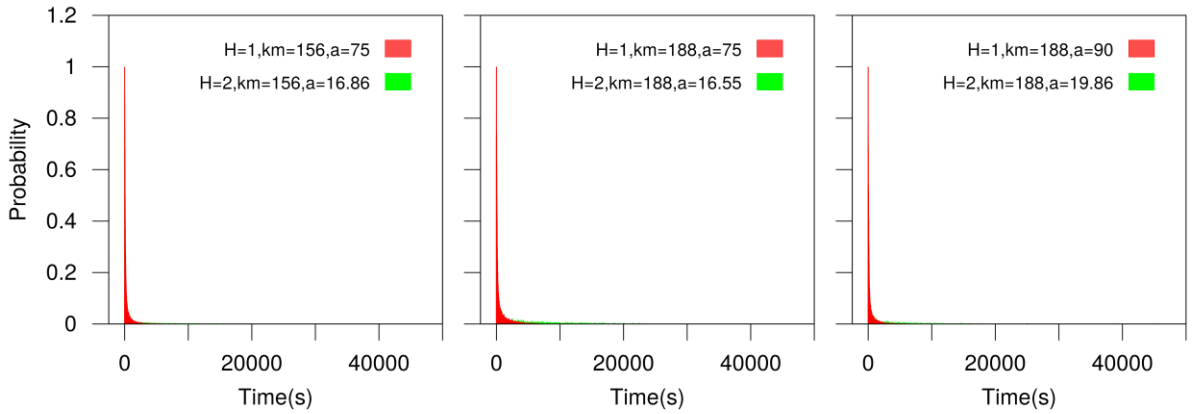

Figure 2.6 Normalized time distribution latency state for cut-off as five TAT molecules. 5000 number of samples were taken for distribution. Samples were recorded for 15 days.

For two TAT molecules as a cut-off for latency state, for H=2, long latency states are observed as compared to H=1, while for five TAT molecules as a cut-off for latency state there is no difference in latency state for H=1 and H=2 which shows that five TAT molecules as a cut-off for latency are not a good parameter.

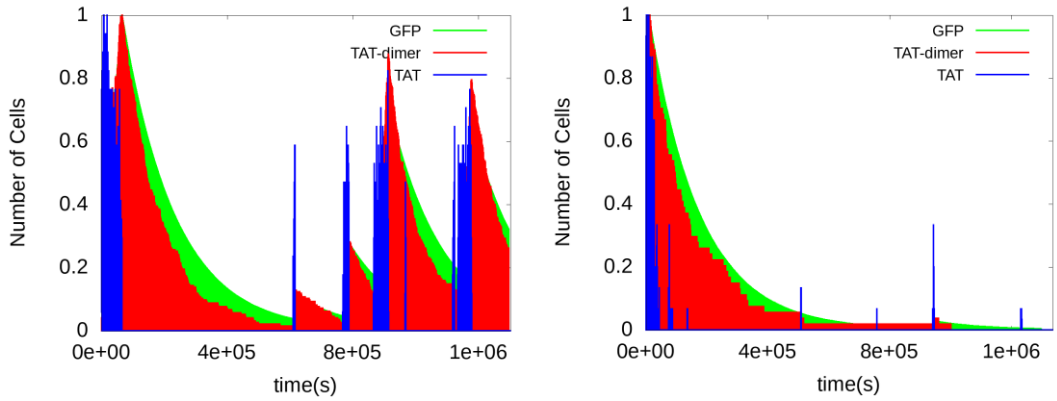

Figure 2.7 Normalized molecules (respect to the maximum value) with time. These two graph shows how GFP, transactivating TAT, and non-Transactivating TAT molecules changes as the virus goes to latency. Left graph is for H=1, and the right graph is for H=3.

From Figure 2.7, Graph on the left side shows that all the transactivating TAT molecules go to zero as virus goes to latency state, and GFP and non-transactivating TAT molecule growth becomes negative (net GFP and non-transactivating molecules starts decaying). Left graph is for H=1, and the right graph is for H=3 cooperativity.

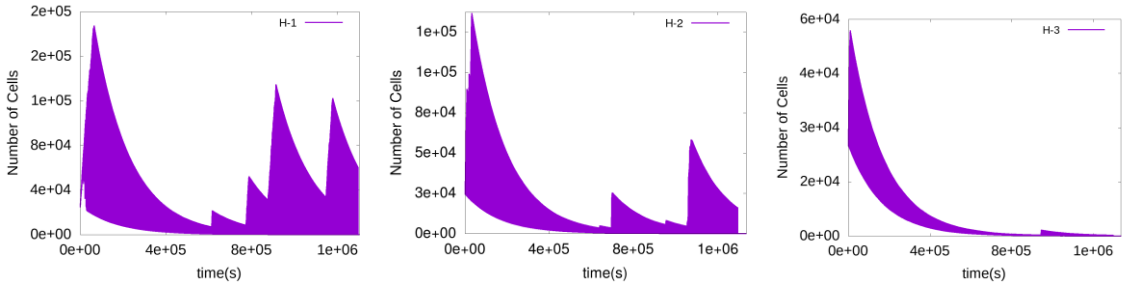

Figure 2.8 Growth of GFP molecule is plotted for self co-operative value H=1,2 and 3.

From Figure 2.8, Left side graph shows growth of GFP molecules when co-operativity is H=1. H=1 virus goes in the latency cycle after 65000 seconds (18 hrs approx.). Centre Graph shows growth when co-operativity is H=2. For H=2 virus goes in latency state after around 40,000 seconds (11 hrs approx.). Right side graph shows the growth of the molecule when co-operativity is H=3. For H=3, the virus goes to latency after around 15,000 seconds (4 hrs approx.). For Higher co-operativity virus goes in latency state faster. Peaks in the above graph show how many times a virus came to an active state. For higher co-operativity (H=3) it goes in longer latency state as compared to low co-operativity (H=2 and 1).

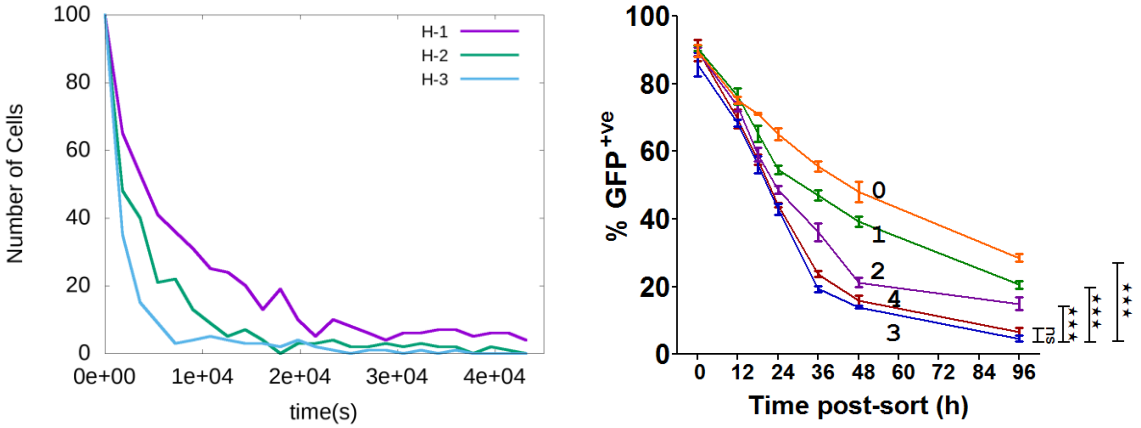

Figure 2.9 Number of active cells present at a different time is plotted for cooperativity H=1,2,3. Left side is from our simulations right side has been adapted from the PhD thesis [23] of S. Chakraborty, MBGU, JNCASR. Simulations show similar qualitative pattern of an early onset of latency with strong promoters.

To study how the promoter strength of the virus helps it to go in fast latency. A Hundred cells sample was taken for studying the cooperative effect of latency cycle of the virus. Fig-10 shows the variation of a number of the active cell with time for different cooperativity. Initially, all virus cells were taken in an active state for all cooperative parameters (H=1,2 and 3). Left side image is predicted the behaviour of promoter strength on latency cycle and the right image is experimentally observed the behaviour of latent cells. In Experimentally observed results weak, medium and strong promoters are represented as 0,1 (weak),2 (medium) and 3,4 (strong) whereas in computational model promoter strength are represented by cooperativity 1, 2 and 3 represent weak, medium and strong promoters. The number of active cells presents in the system at any point of time decrease exponentially with increase in time. For weak promoter (H=1), fall in active cell quantity is slow as compared to the decrease in active cell number for the medium and strong promoter. From here, we can conclude that stronger the promoter in HIV, faster it will go to latency state from active state.

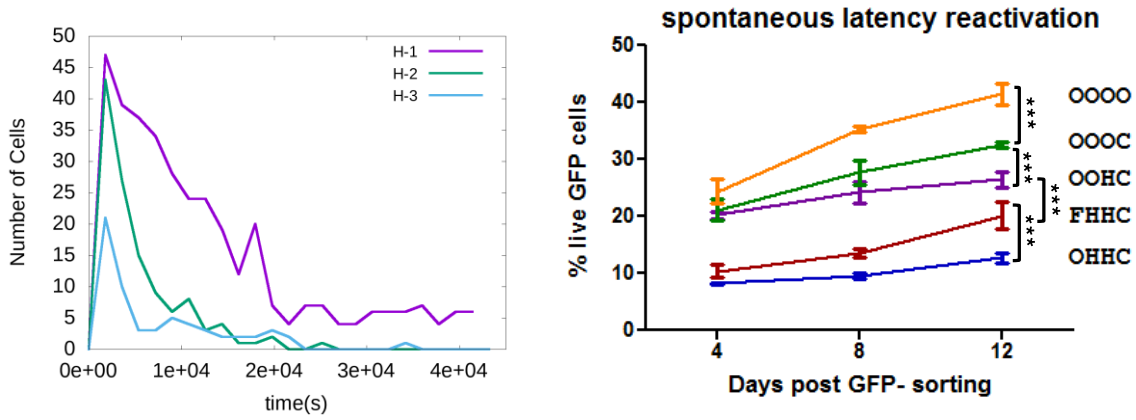

Figure 2.10 Comparison of the spontaneous reactivation in (A) our simulations (B) experiments (adapted from S. Chakraborty, PhD Thesis[23], MBGU, JNCASR)

To study how often latent cells becomes active, and the effect of promoter strength on the latency period of the cell. Again sample of hundred cells was taken which were in latency state to check the promoter strength effect on latency cycle. Figure 2.10 shows the number of cells active at a point of time. At time t_0 all cells are in latency state and as time increases

cells come out of latency and some of them again go back to latency. Left graph is experimentally observed the behaviour of cells which shows cells with weak promoter come out of latency state in a large amount as compared to cells with strong promoters. This experimental observation is verified by the computational model of HIV (described above). The computational model simulation shows lower cooperativity higher the chances of the virus to come out of latency state and higher the cooperativity lower are a change for the virus to come out of latency state.

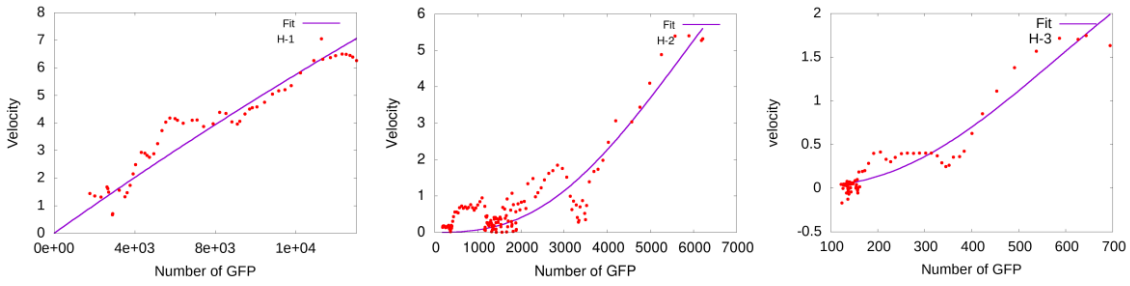

Figure 2.11 Hill coefficient curves calculated from the results our simulations, for H=1,2,3. It was reassuring to see the Hill coefficients H=1,2 were recovered.

Figure 2.11 is a plot of the velocity of GFP molecule with a number of GFP molecule. Velocity is calculated as a change in GFP molecules amount with time. Hill coefficient is

fitted on the velocity versus a number of molecules graph. Fitted value for hill coefficient was found to be 1, 2.66 and 2.57 for H=1,2 and 3.

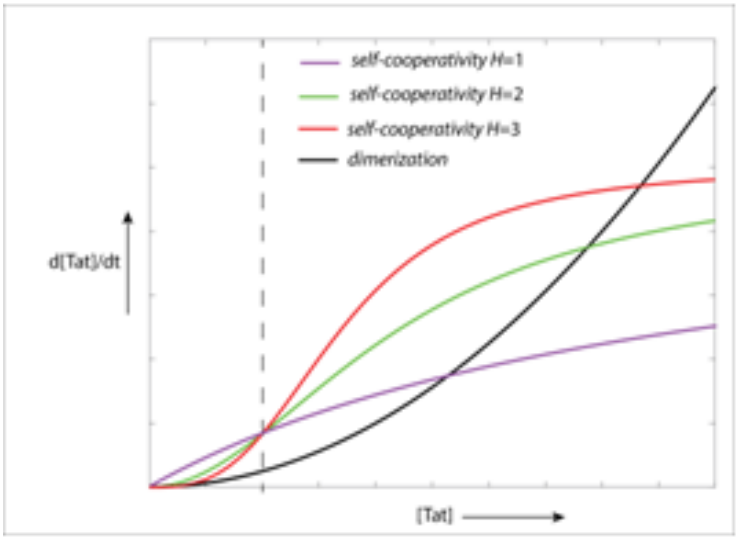

Figure 2.12 Schematic of the competing factors. The two major factors for maintaining the levels of transactivating Tat are its loss due to dimerization and its formation due to cooperativity. The dotted line separates the regions where the cooperativity plays opposite roles. When Tat exceeds or falls below this threshold, H=3 switches from being the dominant to the lagging factor.

Analyzing the theoretical model, a two step association was made, one of the promoter strength with the self-cooperativity, and second of the self-cooperativity with the latency. Interestingly the overall stochastic modeling we used captures these differences in latency. We tried to decipher how the cooperativity contributes to latency. Two critical aspects of our mathematical model were the self-cooperative Tat production and the dimerization of Tat. The balance between these positive and negative factors contributes towards the availability of monomeric Tat which provides a positive feedback for further Tat production. We schematically compare these positive and negative factors in Figure 2.12. We notice a qualitatively different behavior between the rate of Tat production below and above a certain Tat concentration (shown in dotted line), with the rate of self-cooperative Tat production higher for H=1 below the threshold and for H=3 above the threshold. Thus, at low Tat concentrations near latency, the self-cooperativity plays a counter productive role and tends to maintain the latency. At higher concentrations of Tat, the rate of disappearance of Tat due to dimerization also accelerates as the square of the Tat concentration and thus the self-cooperativity becomes self-limiting. Evidence for the dimerization of Tat has been reported earlier[4], and we employed it in this model. However,

the interpretation of the mathematical model which captures the effects of cooperativity on latency could be different. Unfolding or degradation of Tat, or its secretion into cytoplasm are all factors which have a first order dependence on the Tat concentration. Hence all these effects can be combined into a net rate of loss of Tat which is first order in its concentration. However, a second order effect allows the possibility of the rate of disapperance of monomeric Tat to be lower than that due to self-cooperativity at low concentrations and eventually leading to a crossover in the behavior. This second order effect had been interpreted as the dimerization in this work, however it is not restrictive.

2.05 Conclusion

Establishment of latency and differences between the subtypes have been elusive. Here we could demonstrate that with dimerisation, TAT can result in stable cycling between the latent and active states. Further, the present work also suggests that the self-cooperativity level seems to promote the existence of long-lived latency states, in certain kinetic parameter space.

2.06 References

1. Lynch, Rebecca M., et al. "Appreciating HIV type 1 diversity: subtype differences in Env." *AIDS research and human retroviruses* 25.3 (2009): 237-248.
2. Taylor, Barbara S., et al. "The challenge of HIV-1 subtype diversity." *New England Journal of Medicine* 358.15 (2008): 1590-1602
3. Cohen, M. S., et al. "Prevention of HIV-1 Infection with Early Antiretroviral Therapy. N Engl J Med." (2011).
4. Cohen, Myron S., et al. "Antiretroviral therapy for the prevention of HIV-1 transmission." *New England Journal of Medicine* 375.9 (2016): 830-839.
5. Chun, Tae-Wook, et al. "Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection." *Nature*387.6629 (1997): 183.
6. Finzi, Diana, et al. "Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy." *Nature medicine* 5.5 (1999): 512.
7. Finzi, Diana, et al. "Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy." *Science*278.5341 (1997): 1295-1300.

8. Siliciano, Janet D., et al. "Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells." *Nature medicine* 9.6 (2003): 727.
9. Siliciano, Robert F., and Warner C. Greene. "HIV latency." *Cold Spring Harbor perspectives in medicine* 1.1 (2011): a007096.
10. Strain, M. C., et al. "Heterogeneous clearance rates of long-lived lymphocytes infected with HIV: intrinsic stability predicts lifelong persistence." *Proceedings of the National Academy of Sciences* 100.8 (2003): 4819-4824.
11. Arkin, Adam, John Ross, and Harley H. McAdams. "Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected Escherichia coli cells." *Genetics* 149.4 (1998): 1633-1648.
12. Raser, Jonathan M., and Erin K. O'shea. "Noise in gene expression: origins, consequences, and control." *Science*309.5743 (2005): 2010-2013.
13. Razooky, Brandon S., et al. "A hardwired HIV latency program." *Cell* 160.5 (2015): 990-1001.
14. Razooky, Brandon S., and Leor S. Weinberger. "Mapping the architecture of the HIV-1 Tat circuit: A decision-making circuit that lacks bistability and exploits stochastic noise." *Methods*53.1 (2011): 68-77.
15. Kepler, Thomas B., and Timothy C. Elston. "Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations." *Biophysical journal* 81.6 (2001): 3116-3136.
16. Paulsson, Johan. "Summing up the noise in gene networks." *Nature* 427.6973 (2004): 415.
17. Bachu, Mahesh, et al. "Multiple NF-κB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance." *Journal of Biological Chemistry* 287.53 (2012): 44714-44735.
18. Tosi, Giovanna, et al. "Highly stable oligomerization forms of HIV-1 Tat detected by monoclonal antibodies and requirement of monomeric forms for the transactivating function on the HIV-1 LTR." *European journal of immunology* 30.4 (2000): 1120-1126.
19. Weinberger, Leor S., et al. "Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity." *Cell* 122.2 (2005): 169-182

20. Erban, Radek, and S. Jonathan Chapman. "Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions." *Physical biology* 6.4 (2009): 046001.
21. Sanft, Kevin R., Daniel T. Gillespie, and Linda R. Petzold. "Legitimacy of the stochastic Michaelis–Menten approximation." *IET systems biology* 5.1 (2011): 58-69.
22. Gillespie, Daniel T. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions." *Journal of computational physics* 22.4 (1976): 403-434.
23. Sutanuka Chakraborty (Synergy between enhanced transcriptional strength of the viral promoter and the Tat-autoregulatory circuit accelerates latency-establishment in HIV-1C, MBGU, JNCASR (2018)

Chapter-3

Fuel driven self-assembly and its modulation

3.01 Introduction

Polymer science is a field that studies the kinetic formation and degradation of polymers from monomers, their statistical and mechanical properties. The field of polymerisation reaction engineering continues to grow in prominence as the use of polymers in daily life is increasing. Polymerisation is a process in which monomer units linked by chemical reaction or interaction (like covalent bonds) to form long-chain structures. There are two rudimentary classes in which polymers divided synthetic polymers and biological polymers. Synthetic polymers originating from petroleum products or designed by engineers or scientists such as PVC, nylon, Teflon are well known. Biological polymers exist in nature and can be extracted from living systems. Carbohydrates, proteins, nucleic acids, lipids, rubber, silk are some of the naturally occurring polymers.

Different from these polymers are other naturally occurring structures such as microtubules, which are formed by non-covalent interactions among monomers. The non-covalent bond allows the polymerization to be externally modulated by changing factors which either favor or disfavor the formation of these self-assembled polymers. Inspired by these, scientists are interested in dynamical polymerization, where the monomers are not covalently bonded, but rather are held together because of dynamic non-covalent bonds. These polymers are known as supramolecular polymers. Studies on controlled and dynamical self-assembly are important for engineering, synthetic biology in addition to gaining basic understanding into biological systems. Structural control and complexity of polymer plays an important role in their functioning[1] (electronics, biomedical devices). Structural control in polymers is achieved by control radical polymerisation method (CPR)[2]. In biological systems this structural control or reorganisation is usually achieved by fuel like ATP, ADP, GTP[3]. Tuning the ATP and ADP levels in the solution affects the polymerization state.

The underlying mechanisms of polymerization may be explained by isodesmic or cooperative models[6][7]. In isodesmic model, addition of monomer on the stack is governed by a single equilibrium constant, whereas in cooperative model two equilibrium constants one for nucleation and other for elongation of the polymer[6][7] are required. After formation of the critical size of nucleus forms, elongation of the polymer will take place on the nucleus. Stochastic or deterministic models may be used depending on the context in which the fuel or monomeric units are limited or are in abundance.

Inspired by biological systems, there have been several recent experimental studies which controlled and tuned the temporal structure of the supramolecular assembly with fuels like ATP, ADP[4],[5] etc. Molecules like naphthalene diimide (NDPA) with receptor Zinc(II) dipicolylethylenediamine which binds to the phosphate group of ATP or ADP to form one dimensional (1-D) self-assembly. These self-assembled systems are formed because of π-π interaction, hydrophobic interaction due to phosphate group between base units. The system is highly dynamic. With a change of ATP or ADP concentration in solution, structural changes in the assembly can be fueled and realized. NDPA forms a helical structure when it binds to ATP or ADP, with minor differences. When ATP binds to NDPA, it forms a right-handed helical NDPA-ATP assembly ((P)-NDPA-ATP), and ADP bound NDPA forms a left-handed helical NDPA-ADP assembly ((M)-NDPA-ADP)[4]. The change in the orientation of the helix with ATP or ADP was observed in the CD spectra. Right-hand helix gives positive signature CD spectrum, and left-hand helix gives negative signature CD spectra[4][8].

The degree of polymerization and length of the self-assembled structures depends on the availability of the fuel, and the duration for which it is available. While the self-assembly with NDPA has already been reported[4], we study the kinetic aspects of it to see how the theoretical knowledge can be useful for having a control on the polymerization, for example by modulating the duration and intensity of the ATP availability. This part of the work has two goals – to quantify the availability of ATP in a solution by using coupled reactions which generate it and to understand how it affects the polymerization.

3.02 Model

We model the ATP driven self-assembly in two steps. We first model the coupled reactions that generate or consume ATP to create the appropriate circumstances for self-assembly.

We then couple the equations for the self-assembly with this ATP generation. ATP-ADP concentration in a system can be controlled using coupled enzymatic reactions. Creatine kinase and hexokinase are proteins classified as transferase enzymes. Creatine kinase (CK)[9] and hexokinase (HK) are two enzymes catalytic are used to control ATP-ADP concentration. Creatine kinase enzyme reaction is a reversible reaction; it converts ADP to ATP by taking phosphate group from phosphocreatine. This enzyme reaction is present in mitochondria where ATP is generated from phosphocreatine (PCr) through taking a phosphate group from PCr[9]. Conversion of creatine (Cr) to phosphocreatine (PCr) by utilising one phosphate group from adenosine triphosphate (ATP) and releasing adenosine diphosphate is observed in muscles tissues, brain tissue.

$$PCr + ADP + CK\text{-} \rightleftharpoons Cr + ATP + CK$$

Hexokinase is an enzyme that phosphorylates six carbon glucose to hexose phosphate by transferring a phosphate group from ATP to glucose and resulting in ADP as a waste product[10]. Phosphorylation of glucose is the first step for the conversion of glucose to pyruvate. This reaction helps to trap glucose inside the cell due to the negative charge on the phosphate group and helps in metabolising glucose.

$$Glucose + ATP + HK\text{-} \rightleftharpoons G6P + ADP + HK$$

These two enzyme driven reversible reactions can each be modeled by Michaelis-Menten and coupled to understanding how depending on the ATP or ADP that is required for a certain self-assembly, the starting conditions may be tuned.

Simple Michaelis-Menten model for the enzymatic reactions

In the case of enzyme catalysis, the reaction rate increases by order of magnitude. Substrate interacts with enzyme and binds to its active site to form an enzyme-substrate complex. This complex degrades into the product and regenerates enzyme for further reaction. The Michaelis-Menten model first explained this phenomenon.

$$S + E \rightleftharpoons ES \rightarrow P + E$$

k_f and k_r reaction rates in the forward and reverse directions. k_0 is the final rate of reaction from complex to the product. This step is the rate-limiting step[11].

Applying the law of mass action on the above reaction.

$$\frac{d[S]}{dt} = -k_f[E][S] + k_r[ES]$$

$$\frac{d[E]}{dt} = -k_f[E][S] + (k_r + k_0)[ES]$$

$$\frac{d[ES]}{dt} = k_f[E][S] - (k_r + k_0)[ES]$$

$$\frac{d[P]}{dt} = k_0[ES]$$

We assume that the concentration of the substrate is much higher than the enzyme, which is usually the case. Under that assumption, *ES* complex reaches steady state very fast and will remain constant until a significant amount of substrate is consumed. Therefore we can assume *ES* complex concentration to be a constant.

$$\frac{d[ES]}{dt} = 0$$

which gives $[ES] = k_f[E][S]/(k_r + k_0)$. Since the total amount of enzyme, complexed and uncomplexed, is constant, we represent it as

$E_t=E_0+ES$, where E_0 is free enzyme at any point of time and *ES* is the complexed form.

Solving for the above equations we get

$$\frac{d[P]}{dt} = \frac{k_0[Et][S]}{\frac{k_r + k_0}{k_f} + [S]}$$

Where $k_0[E_t]$ is maximum velocity V_{max} and $\frac{kr+k0}{kf}$ is Michaelis-Menten constant K_m.

$$V = \frac{V_{max}[S]}{k_m + [S]}$$

V is the velocity of reaction at any time.

Bi-substrate Michaelis-Menten model for Creatine kinase (CK)

While demonstrated above the general principles of how a single substrate bound to enzyme is catalyzed, as it can be seen, the reactions using the creatinine and hexokinases involve two substrates that need to be bound to the enzyme before the reaction can happen.

$$PCr + ADP + CK \leftrightharpoons Cr + ATP + CK$$

Conceptually there are rates associated with the binding of the substrates, depending on whether they bind first or second. The schematic in Figure 3.1 shows the list of all possibilities, where the order of binding of the two substrates can change as can the order of the release of the two products.

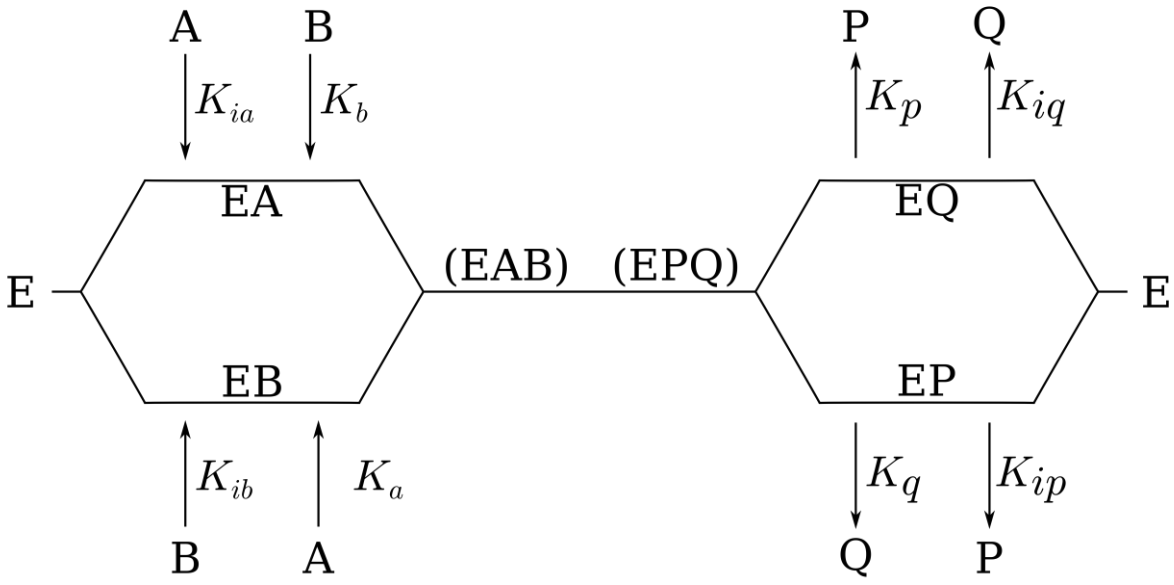

Figure 3.1 The figure illustrates the schematic the four different options in which the intermediate complex may be formed or get unbound depending on the order in which the two substrates can get bound or get released. The notations used for the different molecular species are explained in the text below. The figure has been adapted from Morrison etal ref 13.

Creatine kinase reaction can be modelled using random Bi-Bi model for the Michaelis-Menten equation. In the schematic given above A, B, P, Q and E represent ADP, phosphocreatine, creatine, ATP and enzyme. Conversion of EAB and EPQ is the rate

determining step, and other steps reach rapid equilibrium. EAP and EBQ do not form in this set of coupled chemical reactions.

K_{ia}, K_{ib}, K_{ip}, K_{iq} are dissociation constants for enzyme and reactant A, B, P, Q.

$$k_{ia} = \frac{[E][A]}{[EA]} \qquad [EA] = \frac{[E][A]}{k_{ia}}$$

$$k_{ib} = \frac{[E][B]}{[EB]} \qquad [EB] = \frac{[E][B]}{k_{ib}}$$

$$k_{ip} = \frac{[E][P]}{[EP]} \qquad [EP] = \frac{[E][P]}{k_{ip}}$$

$$k_{iq} = \frac{[E][Q]}{[EQ]} \qquad [EQ] = \frac{[E][Q]}{k_{iq}}$$

K_a, K_b, K_p, K_q are dissociation constants for the enzyme intermediate EA, EB, EP, EQ and reactant B, A, Q, P.

$$k_b = \frac{[EA][B]}{[EAB]} \qquad [EAB] = \frac{[EA][B]}{k_b}$$

$$k_a = \frac{[EB][A]}{[EAB]} \qquad [EAB] = \frac{[EB][A]}{k_a}$$

$$k_p = \frac{[EQ][P]}{[EPQ]} \qquad [EPQ] = \frac{[EQ][P]}{k_p}$$

$$k_q = \frac{[EP][Q]}{[EPQ]} \qquad [EPQ] = \frac{[EP][Q]}{k_q}$$

Substituting values of EA, EB, EP, EQ in the above equations

$$[EAB] = [E][A][B]/k_a k_{ib} \qquad [EAB] = [E][A][B]/k_{ia} k_b$$

$$[EPQ] = [E][P][Q]/k_{iq} k_p \qquad [EPQ] = [E][P][Q]/k_q k_{ip}$$

Assuming that $K_a K_{ib} = K_{ia} K_b$ and $K_p K_{iq} = K_{ip} K_q$ are equal since [EAB] and [EPQ] complex can form in two ways. The velocity of the reaction is the rate of change of reactant or product. The velocity of reaction in forward and reverse directions can be expressed as $V_f = k_f[EAB]$ where k_f is the conversion of [EAB] complex to the product and $V_r = k_r[EPQ]$. The net reaction velocity is a difference of a forward and reverse velocities.

Assuming E_t is the total amount of enzyme present in the system in all forms, $[E_t]=[E]+[EA]+[EB]+[EP]+[EQ]+[EAB]+[EPQ]$, the velocity of reaction from total enzyme can be given as

$$V = \frac{k_1[E_t][EAB]}{[E_t]}$$

$k_1[E]$ is the maximum velocity of the reaction in the forward direction.

$$V = \frac{V_{maxf}[EAB]}{[E_t]}$$

The ratio of [EAB] to $[E_t]$ can be calculated using the above equations.

$$\frac{[EAB]}{[E_t]} = \frac{[EAB]}{[E]+[EA]+[EB]+[EP]+[EQ]+[EAB]+[EPQ]}$$

Substituting value of [EAB],[EPQ],[EA],[EB],[EP],[EQ].

$$\frac{[EAB]}{[E_t]} = \frac{[E][A][B]}{[E]+[EA]+[EB]+[EP]+[EQ]+[EAB]+[EPQ]}$$

$$\frac{[EAB]}{[E_t]} = \frac{[A][B]}{k_{ia}k_b + k_b[A] + k_a[B] + [A][B] + \frac{k_{ia}k_b[P]}{k_{ip}} + \frac{k_{ia}k_b[Q]}{k_{iq}} + \frac{k_{ia}k_b[P][Q]}{k_p k_{iq}}}$$

Now velocity of reaction in the forward direction can be written as

$$V_f = \frac{V_{maxf}[A][B]}{k_{ia}k_b + k_b[A] + k_a[B] + [A][B] + \frac{k_{ia}k_b[P]}{k_{ip}} + \frac{k_{ia}k_b[Q]}{k_{iq}} + \frac{k_{ia}k_b[P][Q]}{k_p k_{iq}}}$$

Similarly, velocity for the backward reaction will be

$$V_r = \frac{V_{maxr}[P][Q]k_{ia}k_b/k_p k_{iq}}{k_{ia}k_b + k_b[A] + k_a[B] + [A][B] + \frac{k_{ia}k_b[P]}{k_{ip}} + \frac{k_{ia}k_b[Q]}{k_{iq}} + \frac{k_{ia}k_b[P][Q]}{k_p k_{iq}}}$$

Net velocity of reaction for creatine kinase Enzymatic reaction will be $V = V_f - V_r$

$$V = \frac{V_{maxf}[A][B] - \frac{V_{maxr}[P][Q]k_{ia}k_b}{k_p k_{iq}}}{k_{ia}k_b + k_b[A] + k_a[B] + [A][B] + \frac{k_{ia}k_b[P]}{k_{ip}} + \frac{k_{ia}k_b[Q]}{k_{iq}} + \frac{k_{ia}k_b[P][Q]}{k_p k_{iq}}}$$

Bi-substrate Michaelis-Menten model for Hexokinase (HK)

$$Glucose + ATP \leftrightharpoons ADP + G6P$$

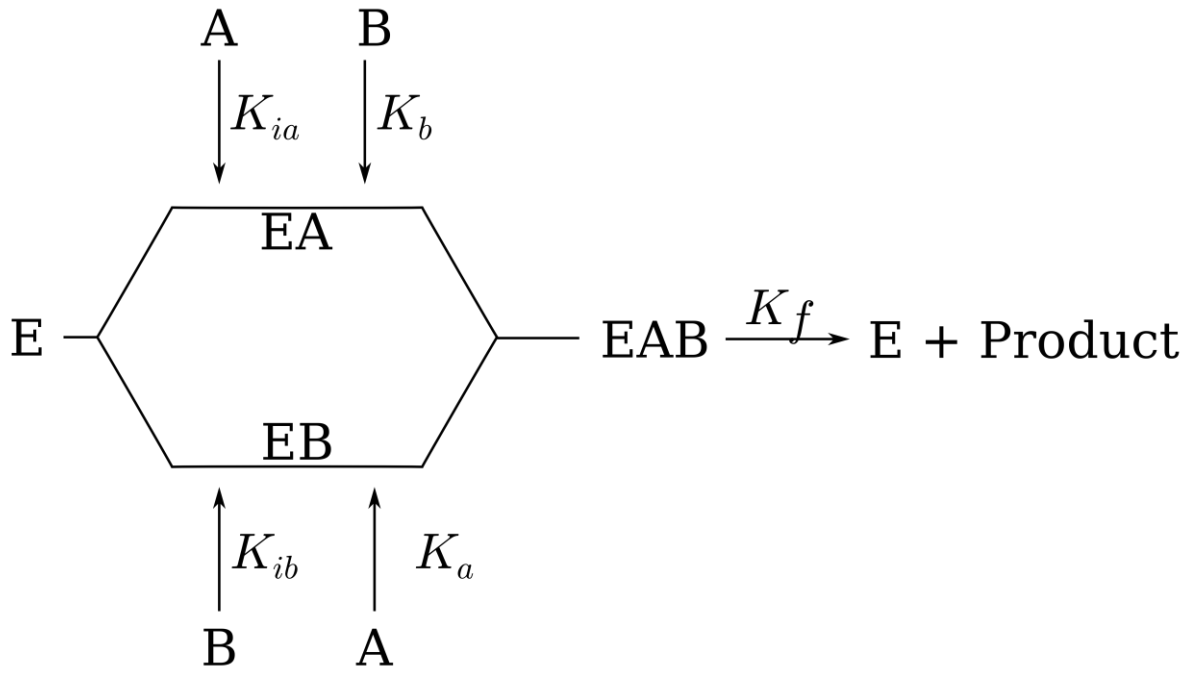

Figure 3.2. A schematic of the bi-substrate reaction mechanism involving hexokinase, adapted from C Tsai ref 14. The notation for each of the species is given below.

Conversion of glucose to glucose-6-phosphate happens in the presence of Hexokinase enzyme. ATP or Glucose can bind to the enzyme to form complex intermediate product simultaneously. Binding substrate for hexokinase reaction can be represented by A (ATP), B (Glucose) and Enzyme by E. Both the substrates A and B can bind to enzyme E. Hexokinase reaction can be modelled using random Bi-Bi model for Michaelis-Menten equation. Where both substrates can bind to the enzyme simultaneously. K_{ia}, K_{ib} are dissociation constant for enzyme E and reactant Aand B.K_a and K_b are dissociation constant for Enzyme intermediate EA and EB and reactant A and B.

$$k_{ia} = \frac{[E][A]}{[EA]} \qquad [EA] = \frac{[E][A]}{k_{ia}}$$

$$k_{ib} = \frac{[E][B]}{[EB]} \qquad [EB] = \frac{[E][B]}{k_{ib}}$$

$$k_a = \frac{[EB][A]}{[EAB]} \qquad [EAB] = \frac{[EB][A]}{k_a}$$

$$k_b = \frac{[EA][B]}{[EAB]} \qquad [EAB] = \frac{[EA][B]}{k_b}$$

$$[EAB] = \frac{[E][A][B]}{k_a k_{ib}} \qquad [EAB] = \frac{[E][A][B]}{k_{ia} k_b}$$

The velocity of the reaction is the rate of change of reactant or product which can be represented as $V = k_f[EAB]$. [EAB] is an intermediate step of hexokinase reaction. Conversion of [EAB] to the product is rate limiting step of reactithe on. Other enzyme intermideate steps reaches equilibrium or steady state rapidly. Assuming $K_{ia}K_b=K_{ib}K_a$, the intermediate complex [EAB] can be reached from two paths, and both having the same probability. Multiplying and dividing velocity expression by total Enzyme concentration.

$$V = \frac{k_f[EAB][E_t]}{[E_t]}$$

$K_f[E_t]$ is the maximum velocity for reaction.

$$\frac{d[A]}{dt} = V = \frac{V_{max}[EAB]}{[E_t]}$$

The concentration of total enzyme in the system is the sum of all free enzyme molecules, bind to a single substrate and with two substrates (an intermediate complex).
$[E_t]=[E]+[EA]+[EB]+[EAB]$

$$\frac{[EAB]}{[E_t]} = \frac{[EAB]}{[E] + [EA] + [EB] + [EAB]}$$

Substituting values for [EA],[EB],[EAB] in above equation.

$$\frac{[EAB]}{[E_t]} = \frac{\frac{[E][A][B]}{k_{ia}k_b}}{[E] + \frac{[E][A]}{k_{ia}} + \frac{[E][B]}{k_{ib}} + \frac{[E][A][B]}{k_{ia}k_b}}$$

$$\frac{[EAB]}{[E_t]} = \frac{[A][B]}{k_{ia}k_b + k_b[A] + k_a[B] + [A][B]}$$

Substituting the value of [EAB]/[Et] in the expression of velocity expression

$$\frac{d[P]}{dt} = V = \frac{V_{max}[A][B]}{k_{ia}k_b + k_b[A] + k_a[B] + [A][B]}$$

Polymerization model

A cooperative nucleation elongation model for polymerisation, in which the rate of nucleation and elongation are different, was used in this study. According to this assumption, the formation of nuclei is not thermodynamically favourable. After reaching critical nucleus, elongation of the polymer takes place by addition of monomeric units to the nuclei[12].

$$Z_1 + Z_1 \rightleftharpoons Z_2$$
$$Z_2 + Z_1 \rightleftharpoons Z_3$$
$$.$$
$$Z_{i-1} + Z_1 \rightleftharpoons Z_{i+1}$$
$$.$$
$$Z_n + Z_1 \rightleftharpoons Z_{n+1}$$

In this polymerization model noted above, first reaction is the nucleation step (Z_2) with the addition of two monomers (Z_1), followed by elongation steps. The rate of formation of nucleation is small as compared to elongation. The ordinary differential equation for above mention nucleation and elongation of the polymer

$$\frac{dZ_1}{dt} = a_n Z_1 \sum_{i=1}^{N-1} Z_i + b_n \sum_{i=2}^{N} Z_i$$

$$\frac{dZ_2}{dt} = a_n Z_1^2 - aZ_1Z_2 - b_n Z_2 + bZ_3$$

$$\frac{dZ_i}{dt} = aZ_1(Z_{i-1} - Z_i) + b(Z_{i+1} - Z_i)$$

$$\frac{dZ_n}{dt} = aZ_1Z_{n-1} - bZ_n$$

a_n and b_n represents nucleus association and dissociation rate in the polymerisation reaction. Elongation association and dissociation rate are represented by a and b. Z_i is the concentration of polymer of length *i* monomeric units

Coupled ATP generation and polymerization

Fuel driven selfassembly of NDPA polymer modelling was done using a deterministic model since the reactions were performed in bulk quantities. Hexokinase and creatine kinase reaction will control the concentration of ATP and ADP. ATP-ADP will get associated with NDPA molecule to form monomeric units and can follow further polymerisation steps.

$$PCr + ADP \leftrightharpoons Cr + ATP$$
$$Glucose + ATP \leftrightharpoons ADP + G6P$$
$$NDPA + ATP \leftrightharpoons NDPA - ATP(Z_1)$$
$$NDPA + ADP \leftrightharpoons NDPA - ADP(Z_1)$$
$$Z_1 + Z_1 \rightleftharpoons Z_2$$
$$Z_2 + Z_1 \rightleftharpoons Z_3$$
$$.$$
$$Z_{i-1} + Z_1 \rightleftharpoons Z_{i+1}$$
$$.$$
$$Z_n + Z_1 \rightleftharpoons Z_{n+1}$$

NDPA-ATP or NDPA-ADP molecules units behaves as monomeric units. The polymerization rates depend on the concentration of ATP. For the purpose of this work, we derived the initial conditions (Table 3.1) and rates from the experimental work (Table 3.2).

3.03 Results

Controllable ATP pulses: ATP/ADP binds to NDPA to form polymer stacks and acts as fuel for the formation of the polymer. Availability of ATP/ATP defines the configurable orientation of polymer helix and switching between ATP and ADP changes the orientation of the polymer. Hexokinase transfers one phosphate from ATP molecule to glucose and converts ATP to ADP. Conversion of ATP to ADP and binding to NDPA molecule changes the configuration of polymer stack from right helix to left helix. Creatine kinase transfers a

phosphate to ADP to form ATP from phosphorylation. Because of the opposing factors from hexokinase and creatine kinase enzymes, a pulse of ATP is generated which lasts with a certain intensity over a period of time. The intensity and duration of the pulse can be tuned and to achieve a control on polymerization as can be seen from Figure 3.3.

Species	Concentration (mM)	Species	Concentration (mM)
Hexokinase Enzyme	0.8	Glucose	30.0
Creatine kinase Enzyme	2.0	Glucose-6-P	0.0
NDPA	0.05	PCr	0.15
ADP	0.075	Cr	0.0
ATP	0.00		

Table 3. 1. Initial concentration of chemical species that were adopted from [4].

Rate constant	*Value*	*Rate constant*	*Value*
v1	*0.408*	k_p	*6.1*
v2	*0.222*	k_{ip}	*15.6*
v3	*0.24928*	k_q	*0.48*
k_a	*0.05*	k_{iq}	*1.2*
k_{ia}	*0.17*	k_{ha}	*0.1044*
k_b	*2.9*	k_{hia}	*0.5956*
k_{ib}	*8.6*	k_{hb}	*0.0497*
k_{Ia}	*0.17*	k_{Ib}	*20.6*

Table 3.2. Rate constants for Michaelis-Menten equation are adopted from [13],[14].

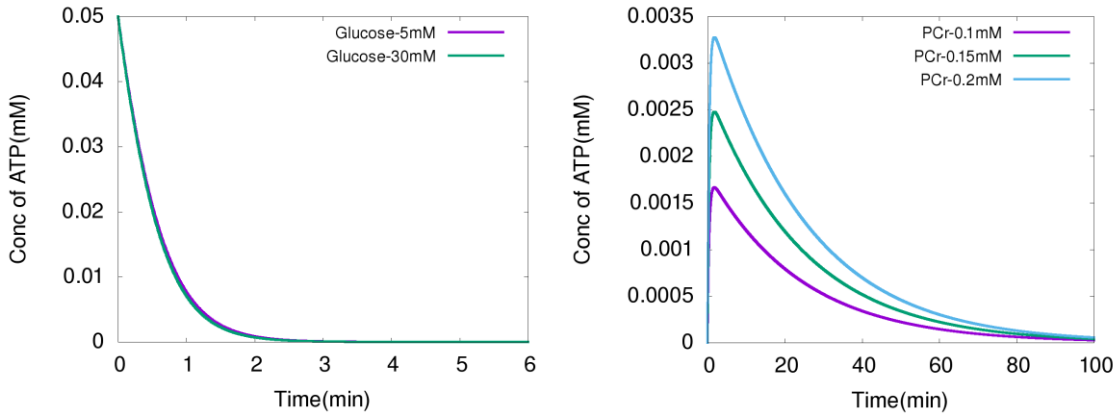

Figure 3.3 Variation of ATP concentration with (A) a change in Glucose concentration. (B) with a change in PCr concentration. The intensity and duration of these ATP pulses can be inferred from these graphs.

As initial concentration of glucose varies from 5mM to 30mM, the rate of conversion of ATP to ADP did not change. From Figure 3.3A we conclude that all available hexokinase enzyme in the system is occupied by glucose already at low concentration like 5mM. From the ATP pulses shown in Figure 3.3, the intensity and duration of the pulses were inferred. with the idea that these two parameters can be used to simultaneously tune the polymer length and time for which it remains intact. To modulate the lifetime of ATP in the solution, we need to optimise the concentration of PCr and Glucose in the system to have ATP in desire amount and for desired time. A dependence of the ATP pulse intensity on a simultaneous variation of glucose and PCr is shown in Figure 3.4. There is no or insignificant change in ATP concentration with an increase in glucose concentration at any concentration of PCr. Production of ATP increases with increase in PCr concentration in solution irrespective of Glucose concentration.

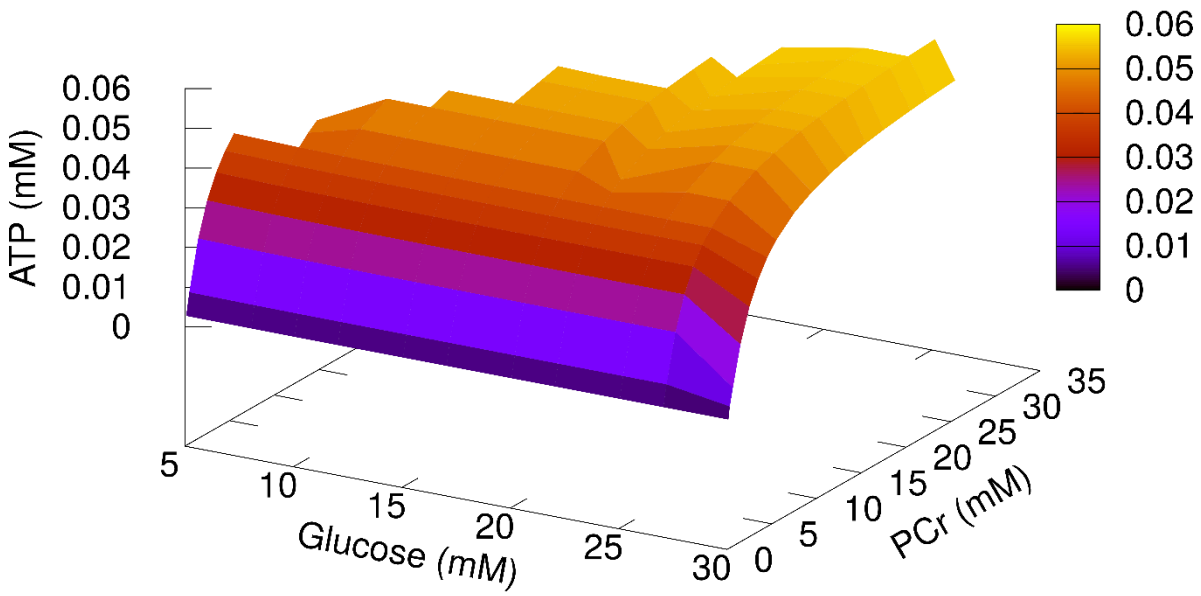

Figure 3.4 A variation of maximum ATP concentration obtained in the pulse with a simultaneous change in Glucose and PCr concentrations.

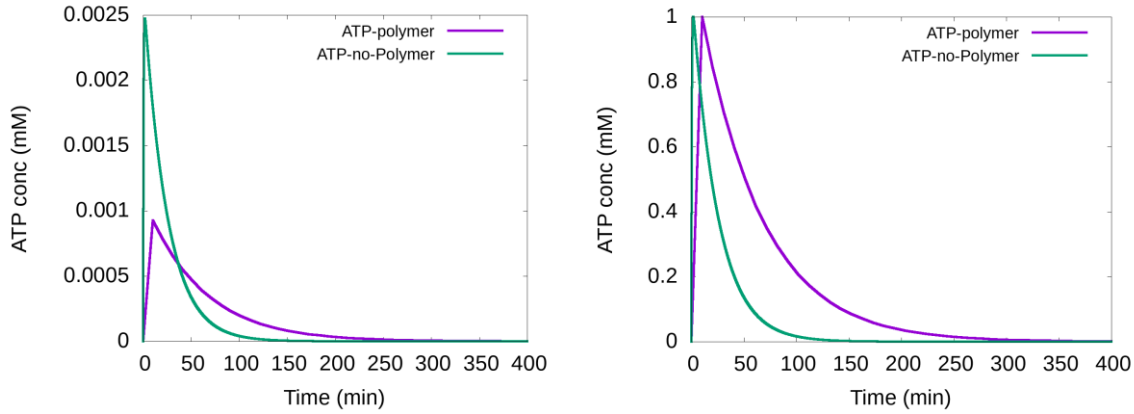

Figure 3.5. A variation of ATP concentration in the absence and presence of a simultaneous polymer formation which affects its concentration. The un-normalized (A) and normalized (B) graphs to illustrate a small shift in the peak position, but without any other significant differences.

Controlling polymerization

As initially all NDPA molecules will bind with ADP molecules to form monomers and then monomers will aggregate to form polymers or stack. Change in ADP or ATP concentration will have a direct effect on the polymer configuration. As the binding affinity of NDPA molecule with ATP is more as compared to ADP[8].

In the polymerisation of NDPA-ADP monomer, some of free ADP molecules will be converted to ATP molecule by transferring Phosphate group to ADP. This change in ATP concentration is more when there is no polymerization and low when ADP is bind to NDPA molecule (Figure 3.5). The rate of conversion of ADP to ATP is more when there is no polymerization is taking place and low when polymers are forming and dynamic.

Using these pulses of ATP, length and rate of polymerization (Figure 3.6) can be controlled. NDPA binds with ADP to form monomer and monomers get the stack to form a polymer. Distribution of NDPA-ADP polymer is an asymmetric gaussian distribution. With the increase in time monomers are being consumed to increase the length of polymer and more monomers are formed. This polymerisation process increases the mean length of polymer distribution with an increase in time.

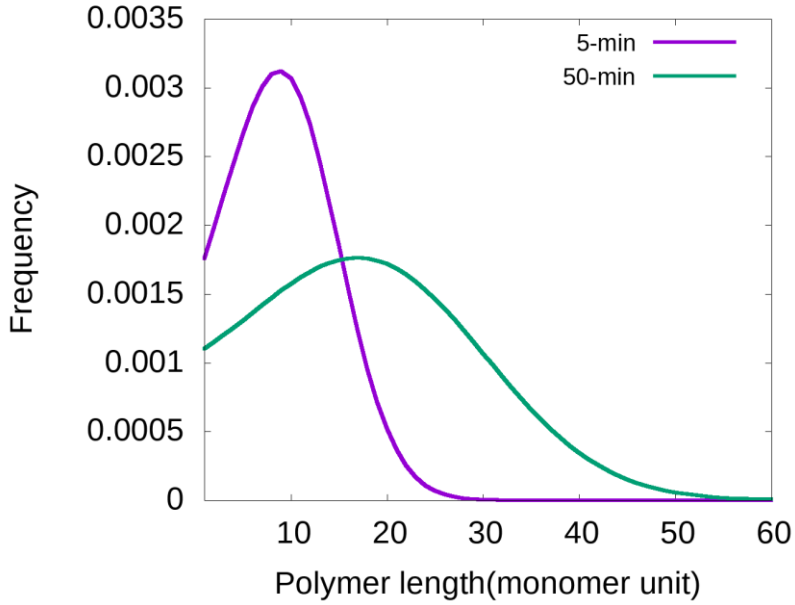

Figure 3.6 NDPA-ADP polymer length distribution at different time points during the simulation.

3.04 Conclusion

We have performed detailed calculations mapping the ATP production from the coupled enzymatic reactions. Tuning the concentrations of the different molecular species allowed a variable intensity and pulse of ATP, which can be coupled with the self-assembly NDPA-ADP and NDPA-ATP polymers. Helicity of polymer can be changed with control change in ATP or ADP concentration, which in turn was controlled by PCr or glucose.

3.05 References

1. Lutz, Jean-Francois, et al. "From precision polymers to complex materials and systems." *Nature Reviews Materials*1.5 (2016): 16024.
2. SZWARC, Michael. "'Living'polymers." *Nature* 178.4543 (1956): 1168
3. Dhiman, Shikha, Aritra Sarkar, and Subi J. George. "Bioinspired temporal supramolecular polymerization." *RSC Advances* 8.34 (2018): 18913-18925.
4. Dhiman, Shikha, Ankit Jain, and Subi J. George. "Transient helicity: fuel-driven temporal control over conformational switching in a supramolecular polymer." *Angewandte Chemie International Edition* 56.5 (2017): 1329-1333.
5. Mishra, Ananya, et al. "Biomimetic temporal self-assembly via fuel-driven controlled supramolecular polymerization." *Nature communications* 9.1 (2018): 1295.
6. Smulders, Maarten MJ, et al. "How to distinguish isodesmic from cooperative supramolecular polymerisation." *Chemistry–A European Journal* 16.1 (2010): 362-367.
7. Martin, R. Bruce. "Comparisons of indefinite self-association models." *Chemical reviews* 96.8 (1996): 3043-3064.
8. Kumar, Mohit, Narendra Jonnalagadda, and Subi J. George. "Molecular recognition driven self-assembly and chiral induction in naphthalene diimide amphiphiles." *Chemical Communications* 48.89 (2012): 10948-10950.
9. McLeish, Michael J., and George L. Kenyon. "Relating structure to mechanism in creatine kinase." *Critical reviews in biochemistry and molecular biology* 40.1 (2005): 1-20.
10. Copley, Martha, and Herbert J. Fromm. "Kinetic studies of the brain hexokinase reaction. A reinvestigation with the solubilized bovine enzyme." *Biochemistry* 6.11 (1967): 3503-3509.
11. Sanft, Kevin R., Daniel T. Gillespie, and Linda R. Petzold. "Legitimacy of the stochastic Michaelis–Menten approximation." *IET systems biology* 5.1 (2011): 58-69.
12. Markvoort, Albert J., et al. "Fragmentation and coagulation in supramolecular (co) polymerization kinetics." *ACS central science* 2.4 (2016): 232-241.
13. Morrison, J. F., and Elizabeth James. "The mechanism of the reaction catalysed by adenosine triphosphate-creatine phosphotransferase." *Biochemical Journal* 97.1 (1965): 37-52.
14. Tsai, C. Stan, and Q. Chen. "Purification and kinetic characterization of hexokinase and glucose-6-phosphate dehydrogenase from Schizosaccharomyces pombe." *Biochemistry and Cell Biology* 76.1 (1998): 107-113.

Future perspectives

In this thesis we explored the role of chemical kinetics in biological or biomimetic systems. Chemical kinetics typically comes across as a very established field, that may not be of much relevance in the current research literature. However, as we could show with the two examples in this work, a requirement to understand infection cycles of a difficult to treat disease like HIV or to controlled fuel driven self-assembly are contemporary experimental challenges with potential implications in health and in bioengineering respectively. Addressing these concerns required a through understanding of the conditions, development of the minimal models which capture the systems level effects, computing the consequences and interpreting them. Some of the obvious conceptual extensions of these, each of which will require a newer understanding of the respective systems are the understanding of latency in other infections such as by mycobacterium tuberculosis and a controlled generation of spatio-temporal patterns arising from fuel driven self-assembly. These are some of the immediate challenges we look forward to.