MATHEMATICAL MODELING OF GENE EXPRESSION SYSTEM IN EUKARYOTES: FROM GENE INDUCTION TO PROTEIN SYNTHESIS

THESIS SUBMITTED TO THE UNIVERSITY OF PUNE

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY

By

SUCHETA A. GOKHALE

Under the guidance of

DR. CHETAN GADGIL

Scientist

CSIR-National Chemical Laboratory

August, 2013

MATHEMATICAL MODELING OF GENE EXPRESSION SYSTEM IN EUKARYOTES: FROM GENE INDUCTION TO PROTEIN SYNTHESIS

THESIS SUBMITTED TO THE UNIVERSITY OF PUNE

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY

By

SUCHETA A. GOKHALE

PGS/Ph.D./1455

University of Pune, India

Under the guidance of

DR. CHETAN GADGIL

Scientist

CSIR-National Chemical Laboratory

Place of Work: Chemical Engineering and Process Development Division,

CSIR-National Chemical Laboratory, Pune – 411008, India

August, 2013

Certificate by the Guide

CERTIFIED that the work incorporated in the thesis "**Mathematical modeling of gene expression system in eukaryotes: From gene induction to protein synthesis**" submitted by Ms. Sucheta A. Gokhale was carried out by the candidate under my supervision/guidance. Research material as has been obtained from other sources has been duly acknowledged in the thesis.

Date

Dr. Chetan Gadgil
Supervisor/Research Guide
University of Pune (BUTR/Science/243/181)
&
Scientist
Chemical Engineering and Process Development Division
CSIR-National Chemical Laboratory
Dr. Homi Bhabha Road, Pashan,
Pune – 411008
&
Scientist
CSIR-Institute of Genomics and Integrative Biology
Near Jubilee Hall, Mall Road
Delhi – 110007

Declaration by the Candidate

I declare that the thesis entitled "**Mathematical modeling of gene expression system in eukaryotes: From gene induction to protein synthesis**" submitted by me for the degree of Doctor of Philosophy is the record of work carried out by me during the period from December, 2008 to June, 2013 under the guidance of Dr. Chetan Gadgil and has not formed the basis for the award of any degree, diploma, associateship, fellowship, titles in this or any other University or other Institution of higher learning.

I further declare that the material obtained from other sources has been duly acknowledged in this thesis.

Date

Sucheta A. Gokhale
PGS/Ph.D./1455
Pune University, Pune
&
Senior Research Fellow
Chemical Engineering and Process Development Division
CSIR-National Chemical Laboratory
Dr. Homi Bhabha Road, Pashan
Pune – 411008

Dedicated to
My parents, elder brother,
and my husband

Acknowledgement

I take the opportunity to thank all those who supported, encouraged, and guided me in the entire journey of my Ph.D.. First and foremost, I thank my research guide, Dr. Chetan Gadgil. He not only guided me for scientific difficulties but general discussions have also been enlightening. His teaching to take the responsibility and ownership of own work, right from the beginning of my Ph.D., helped me to be self-critical, and think and work independently. He gave me appropriate guidance and at the same time freedom to explore things. The scientific discussions with him have always been stimulating. I am also thankful to him for patiently going through the drafts of manuscripts and thesis. In addition to the scientific skills, he also emphasized on having good soft skills. An important tact I learned is the elevator talk. On the whole, working with him has been the best part of my Ph.D. My gratitude to him is ever-lasting.

In addition to my home institute, my research has gained a lot from visits to CSIR-IGIB, India. It was a great experience to work with researchers and scientists at CSIR-IGIB. I am greatly thankful to Prof. Dr. Samir Brahmachari for his scientific guidance and also for giving me motivation and vision to think about larger questions. In spite of his busy schedule as Director General of CSIR he always gave me quality time for discussions. I also thank Dr. Beena Pillai for her inputs. The discussions with Dr. Pillai have been very inspiring. I would like to thank Dr. Vinod Scaria for his help. I am thankful to Dr. Manoj Hariharan with whom I started the microRNA work. It was a good opportunity to work with Manoj as I could learn many new things, especially presentation skills. I would also like to thank my friends Reema, Rakesh, and Bharat for their great help and support during my stay at CSIR-IGIB.

I thank Prof. Dr. J.K. Pal, University of Pune and Dr. Pranay Goel, IISER-Pune for evaluating my progress.

I whole-heartedly thank all my teachers for their guidance. Especially, I thank Dr. Phadke, who is associated with Ramnarain Ruia college, Mumbai, not only for his scientific teaching that kept alive my inquisitiveness but also for teaching *how to learn*. I will always be grateful to him for the thoughts and approach he developed within me.

In my institute I would like to thank the Director Dr. S. Pal, Head of Division Dr. V. Ranade for their support in my research work. I thank Dr. B.D. Kulkarni and Dr. P. Doshi

for evaluating my periodic progress. I would like to thank Dr. L. Narlikar and Dr. K. Joshi for their inputs and help during the work. I am thankful to Dr. M. Gadgil for her guidance. I am thankful to all the scientists who taught us the various courses. I would like to thank Mrs. Puranik and Mrs. Kolhe from Students Academic Office for their support in administrative work.

Research lab would not have been such a good place to work without my friends Vinaya, Prachi, Shraddha, Avinash, Priyanka, Dimpal, Mudita and Charudatta. They have always been supportive, considerate and great companions to me. It has been great fun to have mind-boggling scientific discussions with them. On a tiring day, a brief session of games in lab with them had been very refreshing. In addition, I would like to thank Kasturi, Megha and Hemangi for their help and guidance in administrative processes.

I thank CSIR for funding. I would also like to thank anonymous reviewers of our scientific communications for their suggestions that improved the quality of my work.

Back home, I am deeply indebted to my parents for their efforts to develop me as a person. They always encouraged me to pursue what I want. They have always been supportive and cheered me up whenever I felt low. I also have special regards for my elder brother. He has been role model for me in many respects. I would also like to thank my sister-in-law for always supporting me during my Ph.D.. I am greatly thankful to my husband, Vivek for his constant, unconditional support. He patiently listened to my queries, gave me important inputs and also helped me with technicalities during thesis writing. He is the best company to celebrate even small achievements. I would also like to extend my appreciation to my in-laws for being patient and supportive.

Contents

List of Tables

List of Figures

Abbreviations

Acronym	Expansion
ACF	Auto-correlation function
ARE	AU rich element
Cdc14	Cell division cycle 14 protein
CDK	Cycling dependent kinase
CF	Cleavage factor
CPSF	Cleavage and polyadenylation specificity factor
CstF	Cleavage stimulatory factor
CV	Coefficient of variation
eEF	Eukaryotic elongation factor
eIF	Eukaryotic initiation factor
eRF	Eukaryotic release factor
FACS	Fluorescent activated cell sorting
FRAP	Fluorescent recovery after photobleaching
GMP	Guanosine monophosphate
GT	Guanylyl transferase
HAT	Histone acetyl transferase
IL-10	Interleukin-10
IRES	Internal ribosome entry site
lncRNA	Long non coding RNA
MCM7	Mini-chromosome maintenance complex component 7
miRNA	microRNA
MPSA	Multiple parameter sensitivity analysis
mRNA	Messenger RNA
MT	Methyl transferase
NMD	Non-sense mediated decay
NPC	Nuclear pore complex
PARN	poly(A) ribonuclease
PIC	Pre-initiation complex
piRNA	Piwi-interacting RNA
PSE	Proximal sequence element
RBP	RNA binding protein
RISC	RNA induced silencing complex
RTP	RNA 5’ triphosphate
SCA	Spinocerebellar ataxia
siRNA	Small interfering RNA
SL1	Selectivity factor 1
SNAPc	snRNA activating protein complex

snRNP	Small nuclear ribonuleoprotein
SWI/SNF	Switch/Sucrose Non Fermenting
TBP	TATA Binding Protein
TFB	Transcription factor B protein
TFE	Transcription factor E protein
TNF	Tumor necrosis factor
tRNA	Transfer RNA
UTR	Untranslated region
Xist	X-inactive specific transcript

Abstract

Gene expression is fundamental to all biological processes. A majority of the biological processes are regulated at gene expression level. Two major bio-molecules, RNA and proteins are the products of gene expression. Broadly, gene expression is divided into two core processes, transcription and translation. Transcription leads to synthesis of RNA molecules and translation results in synthesis of protein molecules. Though broadly divided into these two steps, each step itself is a multistep reaction requiring a large number of protein and RNA factors. Interestingly, the machinery to synthesize RNA and protein is itself made up of RNA and protein components thus inherently having feedback control. Additionally, each of the steps in gene expression is tightly controlled by various regulators. Such a precise control and co-ordination between the steps of gene expression is required for synthesis of desired protein and RNA at specific times, concentrations and locations in a cell. The presence of feedback controls in gene expression confers interesting properties such as bistability, hysteresis, oscillations that have physiological implications in developmental decisions, biological time keeping etc. Though gene expression is extensively studied using both experimental and theoretical approaches, many details of gene expression and its regulation are still unknown. Investigation of regulators of gene expression and their properties is an active area of research.

An important property of gene expression is stochasticity. Even a clonal population of cells shows phenotypic variation. The inherent random nature of reaction occurrence and small number of molecules of each component of the gene expression machinery result in cell-to-cell variation or noise in gene expression. Many experimental and theoretical studies have analysed various aspects of noise in gene expression such as types of noise, its origin, and effects on population. Genome wide single cell measurement studies have revealed the global structure of noise in the protein molecules. It has been observed that the noise in the protein concentration decreased with increase in protein abundance. However, the relationship was observed to be true only for low and intermediate-abundance proteins. Highly expressed proteins showed a minimum constant low level of

noise. The observed noise floor has been attributed to slow varying extrinsic fluctuations. In addition to these studies, there are some studies investigating the important step in gene expression that maximally affects the noise in protein level. However, the conclusion reached by different studies is different.

In the present study certain aspects of regulation and stochasticity in gene expression are studied. Regulation at transcriptional level has been studied using a mathematical model for auto-regulated synthesis of transcription factor TATA Binding Protein. Regulation at post-transcriptional level by microRNA has been studied through development and analysis of a comprehensive model. A detailed model of gene expression was developed and used to investigate sources of noise contributing to the observed noise floor. Global stochastic sensitivity analysis was performed to quantitatively estimate the relative contribution of major steps in gene expression to noise.

Transcription factor TATA Binding Protein (TBP) is a general transcription factor required by all the three RNA polymerases for transcription initiation. Therefore, TBP in turn is required for its own synthesis as well. In addition to being a crucial transcription factor, TBP mutants having low DNA binding affinity have been identified in case of neurodegenerative disease spinocerebellar ataxia type 17 (SCA17). Another interesting property of TBP is its maternal inheritance. A mathematical model of TBP was developed to study its auto-regulated synthesis. Using the model, the effect of variation in initial conditions and parameters on the viability of cell was explored. The model predicted that the TBP system is bistable, with one stable steady state corresponding to zero-TBP state implying unviable cells, while other high-TBP state corresponding to normal physiological level of TBP. The model prediction of low DNA binding affinity leading to cell death has implication in understanding the role of TBP mutants in neuro-degeneration. The model prediction of requirement of minimum amount of initial TBP for cell viability suggests an explanation for the observed maternal inheritance of TBP mRNA. The effect of presence of TBP dimer in buffering TBP from perturbation was explored using the model. It was observed that TBP dimer can help to buffer against free TBP perturbation only under certain condition and the relative concentration of TBP dimer to TBP-DNA complex is an important determinant of the buffering capability.

The next level of regulation studied is post-transcriptional regulation by microRNA. miRNAs are small, single stranded RNA molecules regulating gene expression at post-transcriptional level. miRNA has been generally considered to repress the target protein. However, in recent years some studies have reported unexpected increase in target protein level in the presence of miRNA regulation. In this study a detailed model of miRNA mediated regulation was developed to explore possible effects of miRNA regulation on the target protein level. The twelve reaction rate parameters of the model were grouped into four dimensionless numbers that were observed to be sufficient to predict miRNA regulatory effects on the steady state of the target protein. Certain conditions were identified under which the presence of miRNA can result in higher target protein level. A majority of the experimental observations of increased target protein level were explained in terms of the model framework such that they are no more unexpected. The effect of miRNA regulation on a steady state distribution of target protein was explored. It was observed that miRNA mediated regulation did not change the nature of the steady state distribution of target protein.

In addition to the study of miRNA effect on the steady state protein level, a simple method was developed to incorporate dynamic effects of intronic miRNA mediated regulation into existing mathematical models of cellular processes. The method was used to modify a widely used protein based mathematical model of cell cycle. Such incorporation of additional regulation was observed to improve the model performance as the predictions of the modified model were found to be closer to experimental observations. The universality of the method was tested by comparing miRNA mediated regulation with analogous protein mediated post-transcriptional regulation and the method was shown to be amenable for use in a wide range of conditions.

A detailed generic model of eukaryotic gene expression process was developed to study the noise in protein molecules. Previously developed models of gene expression could not explain the observed saturation of noise without explicit additive noise term. The detailed model developed in this study was used to investigate whether the gene expression-extrinsic and cell-intrinsic sources of noise can explain the observed saturation of noise. It was observed that only the sources considered in the detailed model

did not explain the saturation. The time scale of regulator fluctuations was observed to be an important factor for saturation. It was observed that the fluctuations with time scale slower than that of the proteins of interest can show saturation of noise.

Different studies investigating the important step determining the noise in protein have identified different steps in gene expression to be important and the knowledge about their relative contribution is ambiguous. In this study, a global stochastic sensitivity analysis was performed to estimate relative contribution of major steps in gene expression to noise in the steady state protein level. Two measures of noise, viz., coefficient of variation and Fano factor were used. Interestingly, it was observed that, the two measures showed differential sensitivity to parameters. Thus the study suggested that previous results regarding sensitivity are required to be reanalyzed using both the measures of noise. The study highlighted the fact that the obtained sensitivity should be attributed to the particular measure of noise and not to the generic variability or noise.

In summary, the present study contributed to advance the knowledge about certain aspects of eukaryotic gene expression. In this study, mathematical models of process of gene expression and its regulation were developed. Through simulations and analyses, the study provided explanations for certain unintuitive observations. It also generated falsifiable hypotheses that can be experimentally tested to gain deeper insight into the processes.

List of Publications

Gokhale, S., Hariharan, M., Brahmachari, S., Gadgil, C. (2012). "A simple method for incorporating dynamic effects of intronic miRNA mediated regulation." Mol. BioSyst. 8(8): 2145-2152.

Gokhale, S. A. and Gadgil, C. J. (2012). "Analysis of miRNA regulation suggests an explanation for 'unexpected' increase in target protein levels." Mol. BioSyst. 8(3): 760 - 765.

Gokhale, S., Nyayanit, D., Gadgil, C. (2011). "A systems view of the protein expression process." Systems and Synthetic Biology. 5.3-4: 139:150.

Gokhale, S. A., Roshan, R., Khetan, V., Pillai, B., Gadgil, C. J. (2010). "A kinetic model of TBP auto-regulation exhibits bistability." Biology direct 5(1): 50.

1 Introduction and overview

Diverse biological processes ranging from bacterial movement towards a nutrient source to the development of a multi-cellular organism are all regulated at gene expression. Gene expression can be considered as a fundamental aspect of all biological processes. It is a highly regulated multistep process that leads to the synthesis of two major bio-molecules, RNA and proteins. Gene expression is broadly divided into two core processes, transcription and translation. The first step (transcription) leads to synthesis of RNA molecules using DNA as the template, and the second step (translation) results in synthesis of protein molecules using RNA as the template. Precise control of the gene expression machinery and coordination of the processes in gene expression are required for the production of RNA and protein molecules at specific times, concentrations and locations in a cell.

Initially gene expression was explained by the central dogma of molecular biology as residue by residue transfer of sequential information from DNA to RNA to protein (Crick 1958). It was later revised to include some more routes of sequential information transfer such as from RNA to DNA and from RNA to RNA (Crick 1970). However, it is well known that gene expression is not only the residue by residue sequential information transfer but it is a highly regulated complex process. Protein molecules were initially considered as the key regulators of gene expression. However, with the discovery of small non-coding RNA (Fire, Xu et al. 1998) and long non coding RNA (lncRNA) molecules (Kapranov, Cawley et al. 2002; Carninci, Kasukawa et al. 2005), a new layer of RNA mediated regulation has emerged in the last couple of decades. The dynamic interactions of RNA and protein molecules form an intricate regulatory network which results in diverse but defined outcomes. Considering regulatory molecular interactions in addition to sequential information transfer, the information transfer in the gene expression process can be represented as Figure 1-1, where all the three components interact with each other.

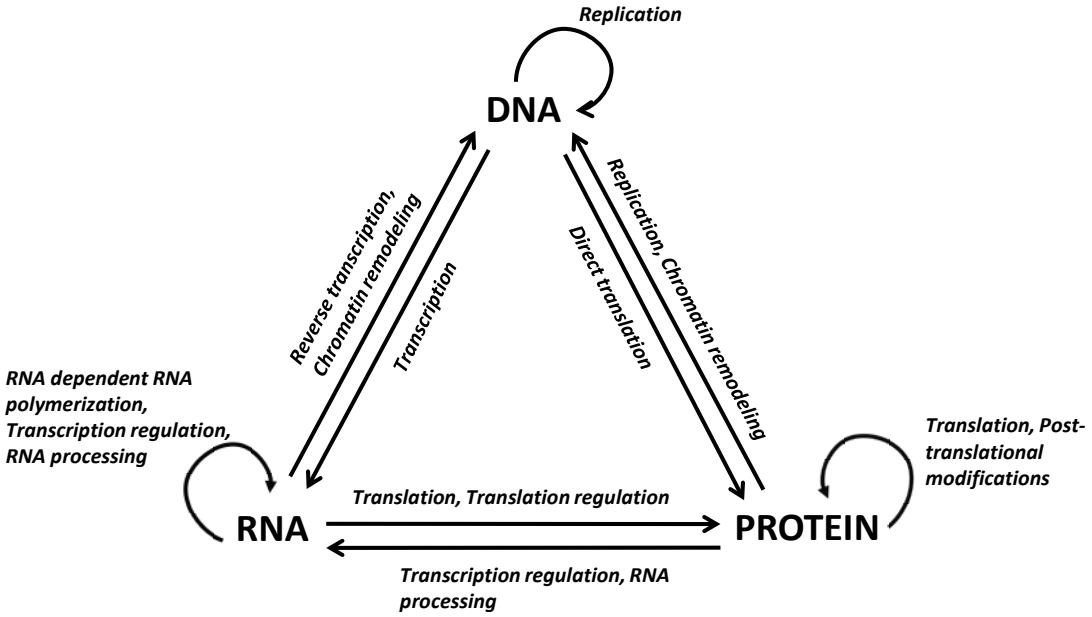


Figure 1-1 – Residue by residue sequential information transfer routes and regulatory interactions in the gene expression process

Though major processes of gene expression are similar in prokaryotes and eukaryotes, different locations of synthesis of mRNA (the nucleus) and protein (the cytoplasm), presence of multiple chromosomes, presence of introns in RNA are some of the factors that make gene expression more complex in eukaryotes. In addition, much larger number of regulatory protein and RNA molecules further increases the complexity due to combinatorial effects.

Gene expression has been extensively studied using both experimental and theoretical approaches yet many qualitative and quantitative details are unknown. Investigation of regulators of gene expression and their properties is an active area of research. In the present study, some aspects of the gene expression process and its regulation are studied using mathematical modeling, simulation, and analysis. Regulation at transcriptional level has been studied by developing a mathematical model of auto-regulatory synthesis of transcription factor TATA binding protein (TBP). Post-transcriptional regulation by microRNA has been studied through the development of a detailed model. A comprehensive model for gene expression has been developed to identify sources of intrinsic and extrinsic noise that result in variability in protein at high abundance level.

Global stochastic sensitivity analysis was performed to estimate the relative contribution of major steps in gene expression to the noise at the steady state of protein.

1.1 Overview of the gene expression process in eukaryotes

All the processes from gene induction to protein degradation that result in a defined spatio-temporal expression of mRNA and protein molecules in a cell comprise gene expression. Broadly, it can be divided into seven major sub-processes viz., (1) chromatin remodeling, (2) transcription, (3) RNA processing, (4) transport of RNA from nucleus to cytoplasm, (5) translation, (6) degradation of mRNA, and (7) degradation of protein. These major sub-processes of gene expression are represented in Figure 1-2.

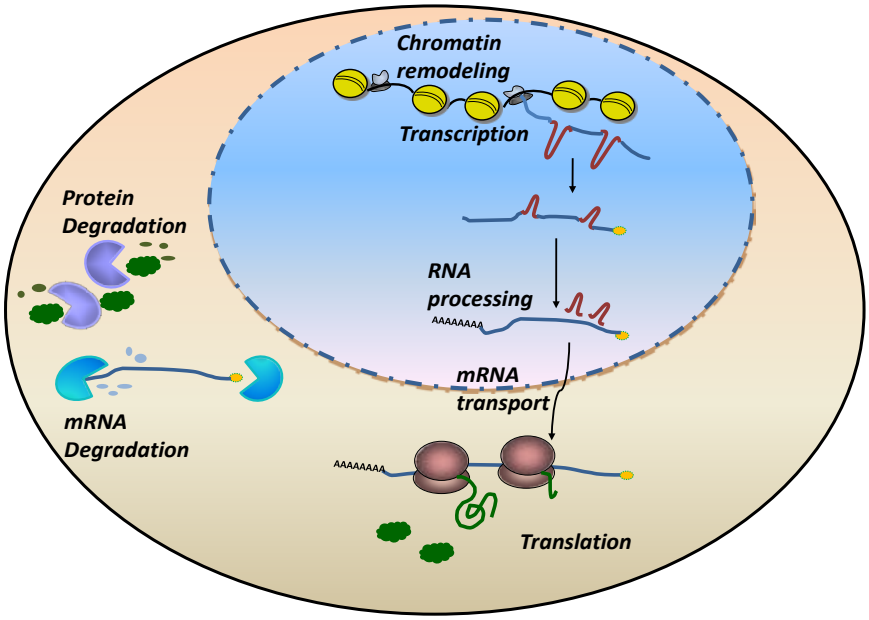


Figure 1-2 – Diagrammatic representation of major sub-processes in gene expression in a eukaryotic cell

The major sub-processes include chromatin remodeling, transcription, RNA processing, mRNA transport, translation, degradation of mRNA, and degradation of protein.

Though, gene expression is broadly divided into these seven sub-processes, each sub-process itself is a multistep reaction requiring a large number of RNA and protein factors. Interestingly, the machinery for RNA and protein synthesis itself consists of RNA and protein factors, thus inherently contains feedback control. Each of these sub-processes

has been studied in detail. Table 1-1 lists some representative experimental studies and reviews for each of the sub-process.

Table 1-1 – Representative literature references for major sub-processes in eukaryotic gene expression process

Process	Literature reference
Chromatin remodeling and Histone modification	(Strahl and Allis 2000; Fry and Peterson 2001; Biran and Meshorer 2012)
Transcription	(Tjian 1996; Thomas and Chiang 2006; Venters and Pugh 2009)
RNA processing	(Padgett, Grabowski et al. 1986; Hastings and Krainer 2001; Johnson and Vilardell 2012)
mRNA transport	(Stewart 2007; Noble and Wente 2010; Oeffinger and Zenklusen 2012)
Translation	(Pain 1996; Gebauer and Hentze 2004; Groppo and Richter 2009)
mRNA degradation	(Beelman and Parker 1995; Houseley and Tollervey 2009; Balagopal, Fluch et al. 2012)
protein degradation	(Baumeister, Walz et al. 1998; Ravid and Hochstrasser 2008; Clague and Urbe 2010)

1.1.1 Chromatin remodeling and histone modification

Chromatin remodeling to unpack and expose the regulatory region of DNA to transcription factors and transcription machinery can be considered as a first step of gene expression. It starts with binding of gene specific regulator(s) to the regulatory region of a gene. The regulator then recruits chromatin remodeling protein complexes, and histone modifying enzymes to the site. During chromatin remodeling the highly organized structure of chromatin is unpacked by sliding or peeling off histone proteins from DNA with the help of chromatin remodeling complexes. These complexes are ATP dependent

helicases such as SWI/SNF, ACF, CHRAC (Peterson 2002). Histone modifying enzymes such as Histone acetyl transferase (HATS), deacetylase, lysine methyl transferase carry out post-translational modifications on specific amino acid residues of histone proteins (Kouzarides 2007). Such post-translational modifications result in changing the affinity of histone protein for DNA making it either dissociate from or tightly bind to DNA to either induce or repress transcription (Hirose 1998; Fry and Peterson 2001). The process of chromatin remodeling and histone modification makes the regulatory region of DNA competent for binding of transcription factors and RNA polymerase.

1.1.2 Transcription

Transcription results in synthesis of RNA molecules. It is divided into three phases, viz., initiation, elongation, and termination. Transcription initiation is comprised of assembly of transcription factors and subunits of RNA polymerase at the promoter site to form pre-initiation complex (PIC). There are three types of RNA polymerase molecules that are required for transcription of different types of genes. RNA polymerase II is required for most protein coding genes which lead to synthesis of messenger RNA (mRNA). Once the PIC is formed RNA synthesis can take place. The process of transcription initiation is completed by local opening of double helix DNA strand at the promoter site (Tjian 1996; Thomas and Chiang 2006). The next phase of transcription is elongation. During elongation RNA polymerase slides along the DNA and synthesizes RNA molecule by polymerization reaction forming phosphodiester bonds between nucleotides (Kugel and Goodrich 2000; Saunders, Core et al. 2006). During elongation, RNA polymerase adds nucleotides to the growing RNA chain depending on the sequence complementarily. It can backtrack to remove the added incorrect nucleotide. Once the polymerization is complete transcription is terminated. During termination the transcript is released and the RNA polymerase subunits dissociate from the DNA (Kerppola and Kane 1991; Richardson and Roberts 1993). The process of transcription results in the synthesis of precursor messenger RNA (pre-mRNA).

1.1.3 RNA processing

The pre-mRNA synthesized during transcription has to be processed to form mature mRNA molecule. Generally RNA processing occurs co-transcriptionally. RNA

processing includes 5'-capping, splicing of introns, and 3'-polyadenylation. 5' capping is carried out once around 20-30 nucleotide long transcript is synthesized. During capping a guanosine monophosphate (GMP) moiety is attached to the 5' end of the transcript. GMP is then methylated at N7 position. The entire process of 5' capping is carried out with the help of three enzymes; RNA 5' triphosphatase (RTP), guanylyl transferase (GT) and, methyl transferase (MT).

Splicing itself consists of multiple reactions in which two exon sequences are fused together and intron sequences in between the exons are removed. Splicing is catalysed by an enzyme complex which consists of five small nuclear ribonucleoprotein complexes (snRNPs) U1, U2, U4, U5 and U6 and many non-snRNP proteins mainly of SR family (Padgett, Grabowski et al. 1986; Hastings and Krainer 2001).

During 3' polyadenylation, around 200 adenosine residues are added at the 3' end of the transcript. It is carried out by a protein complex that includes cleavage and polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CstF), cleavage factor I and II (CF I and CF II), and poly(A) polymerase (Proudfoot, Furger et al. 2002).

1.1.4 RNA transport

Once the mature mRNA molecules are formed, they are transported from nucleus to cytoplasm. Multiple protein factors bind to mature mRNA to form transport competent ribonucleoprotein (mRNP) complex. The mRNP molecules travel in the nucleus through restricted chromatin-free zones. Such nuclear movement of mRNA is termed as channeled diffusion (Noble and Wente 2010). The mRNP molecules are transported to cytoplasm through nuclear pore complex (NPC) with the help of Mex67:Mtr2 protein heterodimer. Another protein Dbp5 remodels the mRNP by removing Mex67 to prevent its return to the nucleus (Stewart 2007). Transport of mRNA from nucleus to cytoplasm completes the nuclear portion of gene expression.

1.1.5 Translation

Similar to transcription, translation is also divided into three phases – initiation, elongation, and termination. In the initiation phase of translation small (40s) ribosomal

subunit, methionine-transfer RNA (Met-tRNA), and eukaryotic initiation factors (eIFs) assemble at 5' end of mRNA forming a 43s initiation complex. The complex slides along mRNA to scan for the initiation codon. Once the initiation codon is located, 60s subunit of ribosome binds to the complex and can start polypeptide synthesis. Such cap dependent translation initiation is predominant in eukaryotes. However, cap-independent translation initiation can take place in case of certain mRNA molecules having internal ribosome entry sites (IRES) (Kelen, Beyaert et al. 2009). During elongation, mRNA slides through the ribosomal machinery. Eukaryotic elongation factors (eEFs) help to recruit amino acid loaded tRNA (changed tRNA) to the site of amino acid addition. A peptide bond is formed between successive amino acid residues using energy from GTP. Once the stop codon is encountered translation is terminated. Termination is mediated by eukaryotic release factors (eRF). During termination the ribosomal subunits and the polypeptide chain are released from the mRNA (Pain 1996; Groppo and Richter 2009).

1.1.6 mRNA degradation

mRNA degradation occurs mainly in the cytoplasm. However, a small fraction of mRNA molecules is degraded in the nucleus through non-sense mediated decay (NMD). The mRNA molecules containing non-sense codon are degraded through NMD (Chang, Imam et al. 2007). Cytoplasmic degradation of mRNA involves deadenylation, decapping and 5'-3' exonucleolytic cleavage. Generally mRNA degradation starts with shortening of polyA tail or deadenylation which is carried out by multiple enzymes for instance, poly(A) ribonuclease (PARN). After removal of polyA tail, 3' exonucleolytic cleavage can also take place (Beelman and Parker 1995). 5' cap of mRNA is removed with the help of decapping enzyme (Guhaniyogi and Brewer 2001). 5'-3' exonucleolytic cleavage of mRNA is carried out by multiple enzymes mainly XRN1 (Houseley and Tollervey 2009).

1.1.7 Protein degradation

Protein molecules which are to be degraded are tagged by ubiquitin. Ubiquitination is carried out with the help of three enzymes, viz., E1, Ub-activating enzyme, E2, Ub-conjugating enzyme and E3, Ub-protein ligase. The tagged proteins are degraded via proteasomal degradation pathway. The Ub-tagged proteins associate with the proteasomal

complex where the protein is unfolded and translocated to the catalytic core of proteasome. Before translocation ubiquitin is removed with the help of deubiquitinylating enzymes. Inside the proteasomal core the protein is degraded into 7-9 mer peptides (Baumeister, Walz et al. 1998; Ravid and Hochstrasser 2008).

1.2 Regulation of gene expression

The sub-processes in gene expression can occur simultaneously and are highly interconnected and coordinated with the help of multiple protein and RNA regulators (Komili and Silver 2008). For instance, nuclear pore complex has been shown to have role in the regulation of transcription (Casolari, Brown et al. 2004). Splicing factors are involved in transport of mRNA to cytoplasm (Hieronymus, Michael et al. 2004). Proteasomal proteins have been shown to bind to chromatin and regulate transcription (Auld, Brown et al. 2006). In addition to regulation of expression of generic RNA or protein molecules, in certain cases these RNA and protein factors regulate their own expression at various stages of gene expression and form feedback loops. Some examples of the feedback regulatory motifs are summarized in Figure 1-3.

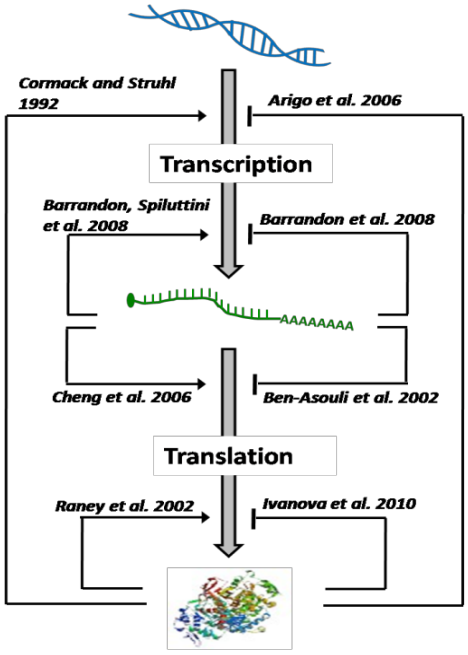


Figure 1-3 – Examples of RNA and protein mediated auto-regulation at different stages of gene expression

Regulation of gene expression can broadly be divided as transcriptional regulation, post-transcriptional regulation, and post-translational regulation. Transcriptional regulation includes regulation of all the processes that lead to synthesis of RNA molecules. Post-transcriptional regulation includes regulation of processes downstream of transcription up to translation. Post-translational regulation includes regulation of the processes once the peptide chain is formed. Two classes of regulation viz., transcriptional regulation and post-transcriptional regulation are considered in the present study. The large class of post-translational regulation, which includes enzymatic modification of protein molecules such as methylation, phosphorylation etc (Seo and Lee 2004; Deribe, Pawson et al. 2010), is not included in this study.

1.2.1 Transcriptional regulation

It includes regulation of gene activation, chromatin remodeling, histone modification and all the three phases of transcription. Protein molecules such as chromatin remodelers, transcription factors are widely known transcriptional regulators. In addition, small non-coding and long non-coding RNA molecules are known to regulate gene expression at transcriptional level. For instance, Xist (X-inactive specific transcript) RNA recruits chromatin remodeling enzymes that help to inactivate X chromosome (Barrandon, Spiluttini et al. 2008). FC RNA and B2 RNA competitively bind to RNA polymerase and repress transcription (Kwek, Murphy et al. 2002). Transcriptional regulation is considered as the key regulatory mechanism to determine differential gene expression. It is known to be responsible for major cellular decisions such as progression through cell cycle (Cho, Huang et al. 2001), development and differentiation (Fong, Cattoglio et al. 2012; Park, Kim et al. 2013). Improper transcriptional regulation has been shown to result into diseases (Lee and Young 2013). It has been stated that in addition to the gene number, morphological and behavioural complexity can be related to protein complexes that regulate gene expression. For instance yeast codes for ~300 transcription factors while human codes for around 3000 transcription factors (Levine and Tjian 2003). These transcription factors interact among each other forming a complex interaction network (Lee, Rinaldi et al. 2002). The positive and negative regulatory interactions among the

regulators form regulatory motifs which have interesting properties such as bistability, oscillations, hysteresis (Cherry and Adler 2000; Angeli, Ferrell et al. 2004). These properties are responsible for the observed phenotypic effects in developmental cell fate decisions (Ferrell and Machleder 1998), circadian clocks (Reppert and Weaver 2001) etc. The transcriptional regulatory motifs having these properties have been studied experimentally and theoretically (Hasty, McMillen et al. 2001) and have been used to construct synthetic circuits (Gardner, Cantor et al. 2000; Shis and Bennett 2013) with desirable properties.

1.2.2 Post-transcriptional regulation

The next level of regulation is post-transcriptional regulation. It includes regulation of RNA processing, transport to the cytoplasm, and translation. A large number of RNA Binding Proteins (RBPs) are known to be post-transcriptional regulators (Glisovic, Bachorik et al. 2008). For instance, protein p32 have been shown to regulate splicing by inhibiting essential splicing factors ASF/SF2 (Petersen-Mahrt, Estmer et al. 1999). Many eukaryotic elongation factors (eEF) and release factors (RF) are known for regulation of translation (Gebauer and Hentze 2004; Kelen, Beyaert et al. 2009). In case of RNA regulators an important class of small RNA molecules called microRNA (miRNA) is known to regulate mRNA activity at post-transcriptional level by various mechanisms (Pillai, Bhattacharyya et al. 2007). Multiple long non-coding RNA molecules are also known to regulate gene expression at post-transcriptional level (Yazgan and Krebs 2007). Post-transcriptional regulation is mainly important for fine tuning of gene expression (Ying and Lin 2005; Sevignani, Calin et al. 2006). It is also advantageous when quick response is required. Similar to transcriptional regulation, post-transcriptional regulation has been shown to be important in cellular processes such as differentiation and development and in diseases (Alvarez-Garcia and Miska 2005; Costa 2005).

The two levels of regulation are not independent but are coordinated through interactions between transcriptional and post-transcriptional regulators (Martinez and Walhout 2009) forming a complex dynamic network. Identification and characterization of the motifs in such networks is an active area of research.

1.3 Gene expression at a single cell level

The knowledge about the processes and regulation of gene expression is obtained from biochemical and genetic studies performed using a population of cells. With the advances in visualization and imaging technologies studies at microscopic level are possible (Larson, Singer et al. 2009). Using technique like Fluorescent-Activated Cell Sorting (FACS) protein levels in individual cells of a population can be quantified (Newman, Ghaemmaghami et al. 2006). MS2 system, single cell microarrays can be used to quantify mRNA (Holstege, Jennings et al. 1998; Golding, Paulsson et al. 2005). Kinetics of certain processes in single cells can be studied using technique like fluorescence recovery after photobleaching (FRAP) (Darzacq, Shav-Tal et al. 2007). Study of gene expression at single cell and single molecule level has revealed many interesting properties of dynamics of RNA and protein synthesis and their steady state distribution. It has been observed that synthesis of mRNA and protein is not a continuous phenomenon but occurs in bursts. Figure 1-4 represents mRNA and protein bursts of engineered luciferase under control of *Bmal1a* promoter observed in a single cell trace (Suter, Molina et al. 2011).

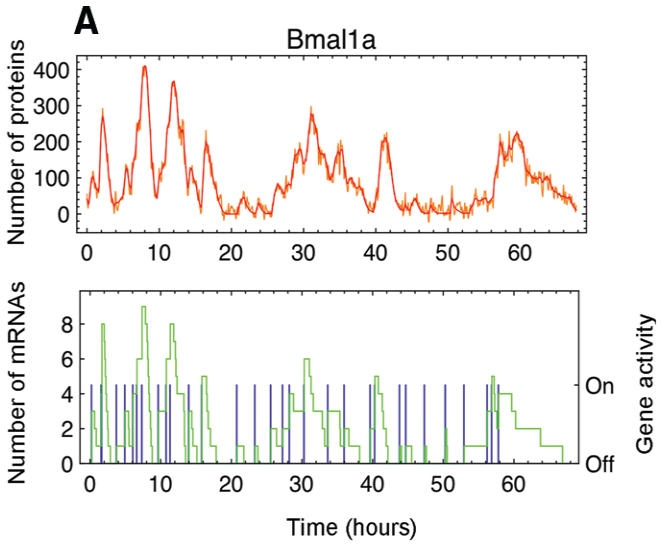


Figure 1-4 – Bursts of mRNA and protein synthesis

Luminescence trace (orange), protein copy number (red), mRNA copy number (green) and gene activity (blue). From Suter, D. M., N. Molina, et al. (2011). "Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics." Science 332(6028): 472. Reprinted with permission from AAAS.

Quantitative measurements in single cells have shown that many of the mRNA and protein molecules are present in low numbers. For instance, it is reported that fission yeast cells contain 1-10 copies of most mRNA molecules (Marguerat, Schmidt et al. 2012). The protein number in yeast cells ranges from lesser than 50 to more than 10^6 (Ghaemmaghami, Huh et al. 2003). Median protein abundance in yeast cell was reported to be around 2500 (Milo, Jorgensen et al. 2010). It is suggested that many essential proteins and regulatory factors are present at low molecular abundance (Ghaemmaghami, Huh et al. 2003). Such low reactant numbers leads to fluctuations in reaction rates and molecule numbers. The resulting stochasticity in gene expression leads to phenotypic variability even within a clonal population of cells. For instance, a study by Novick and Weiner showed a highly variable and random production of β-galactosidase in individual cells of a population of *E. coli* (Novick and Weiner 1957).

The variation in population or noise is classified into two types, intrinsic noise and extrinsic noise. Intrinsic noise is due to the inherent random nature of occurrence of reaction events, while extrinsic noise is due to the variation in the reactant numbers. In case of gene expression, variation in the abundance of molecules such as transcription factors, RNA polymerase, proteasomal degradation machinery contributes to extrinsic noise.

The single cell measurement experiments such as those carried out using FACS on a population of cells result in a distribution of protein or mRNA numbers. It is illustrated in Figure 1-5. The nature of the distribution of a population is an important property and can sometimes reveal presence of a specific type of regulatory motif.

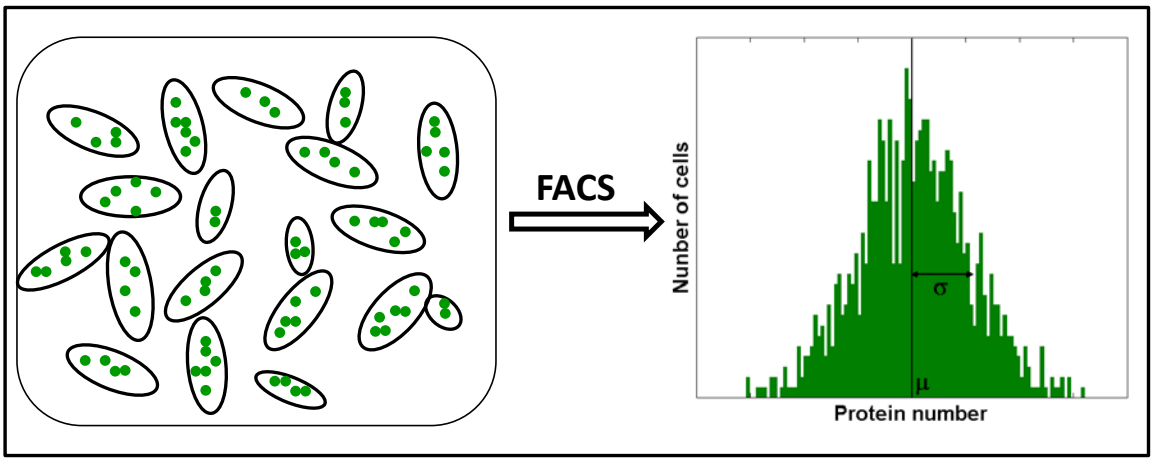


Figure 1-5 – Distribution of protein molecule number obtained by single cell measurement experiment

Measurement of the protein content in individual cells of a cell population, using techniques such as FACS results in a histogram of frequency of cells having a particular number of protein molecules. It indicates the protein distribution with mean (μ) and standard deviation (σ).

Using the distribution of protein number, noise can be expressed using different quantitative measures. Two measures of noise are commonly used viz., Coefficient of Variation (CV) and Fano factor. Coefficient of variation is the standard deviation of the population normalized to the mean (σ/μ). CV is a useful measure to compare noise between two populations when the mean is different. Fano factor is defined as the ratio of variance of the population to mean (σ^2/μ). Fano factor is particularly useful when obtained distribution is Poisson, in which case the Fano factor is 1 as mean and variance are equal. In such cases deviation of Fano factor from 1 indicates deviation from Poisson distribution.

The stochastic nature of gene expression results in population heterogeneity. Therefore, a bimodal or multimodal population can be observed in an isogenic population. Interestingly, stochasticity can result in bimodality even in the absence of deterministic bistability (To and Maheshri 2010). Population heterogeneity can have both advantageous and disadvantageous effects (Raj and van Oudenaarden 2008). Phenotypically different cells can perform different functions. The phenotype of cells can change stochastically and in fluctuating environment can confer fitness to the overall population (Thattai and Van Oudenaarden 2004; Kussell and Leibler 2005). However,

noise can also be disadvantageous under some conditions and is minimized in case of certain proteins that are part of a multi-protein complex or deletion of which is lethal to an organism (Fraser, Hirsh et al. 2004).

Many experimental (Blake, Kærn et al. 2003; Raser and O'Shea 2004; Volfson, Marciniak et al. 2005) and theoretical (Paulsson 2004; Paulsson 2005) analyses have been carried out to study various aspects of noise in gene expression. Yet there are some uncertainties about certain factors such as contribution of various processes of gene expression to the noise at steady state protein level, regulation of noise by various regulators etc. These questions are required to be answered. In addition to the experimental studies, theoretical analysis is an effective approach to understand properties of noise.

1.4 Mathematical modeling

Mathematical modeling of biological processes involves representation of biological phenomena in a framework suitable for the processes to be studied and analysis using appropriate mathematical theory to address proposed questions.

Advances in experimental techniques such as imaging, transcriptome and proteome measurements, single cell measurements, real time kinetic measurements, high throughput sequencing are generating enormous amount of data. To understand the underlying processes using these data quantitative approach is necessary. Using the knowledge obtained by studying various small parts of a cellular system, it has been realized that wiring of the parts forming the whole system confers very different emergent properties to the whole system. These properties cannot be understood by studying the parts but the entire circuit is required to be studied. Mathematical modeling helps to bridge the gap between the data and our knowledge at molecular level, and the observations of biological phenomena at macroscopic level. Firstly, clear representation of a process in a mathematical framework helps to resolve the ambiguities in our descriptive knowledge about the process. Mathematical modeling and analysis is useful for generating certain falsifiable hypotheses and guiding suitable experimental designs. The iterative process of mathematical analysis and experimental study leads to improved

understanding about a process. Predictive power of mathematical models can be used to reduce the experimental efforts.

The present work aims to study the process of gene expression and its regulation using mathematical modeling, simulation, and analysis. Depending upon the system and the underlying question a suitable mathematical approach has been used.

1.5 About the thesis

The thesis describes the development of mathematical models, simulations and analyses of the processes and regulation of gene expression. The properties and the effect of transcriptional and post-transcriptional regulation are studied. Certain aspects of the noise in gene expression are explored in the present study.

A general transcription factor called TATA Binding Protein (TBP) is required by all the three polymerases for transcription. TBP is therefore necessary for its own synthesis as well, forming a positive feedback loop. In addition to being a crucial transcription factor, it has been found to be important in a neurodegenerative disease where DNA binding affinity of TBP was observed to be affected. To understand the properties of the positive feedback regulation and the role of TBP in the disease, a mathematical model of auto-regulatory synthesis of TBP was developed and analysed using deterministic approach. By studying the effect of variation of the model parameters on the viability of the cells, the model provided a possible explanation for the role of TBP in neuro-degeneration. The model provided an explanation for parental inheritance of TBP through the analysis the effect of initial conditions on cell viability.

The next level of regulation analysed in the present study is post-transcriptional regulation by microRNA. These molecules are generally known to repress the target protein level. However, in the last few years some studies have reported ‘unexpected’ increase in the target protein level even in the presence of miRNA. A detailed mathematical model of miRNA mediated regulation was developed and analysed using deterministic approach. Using steady state analysis, certain conditions in terms of reaction rate parameters were identified under which miRNA can result in increased

target protein levels. A majority of experimental observations of increased target protein level were explained in terms of the model framework. The effect of miRNA mediated regulation on the steady state distribution of target protein was explored using stochastic simulations.

In addition to the study of effect of miRNA on the steady state of target protein, the effect of miRNA on the dynamics of target protein was analysed. A simple method to incorporate the dynamic effects of intronic miRNA mediated regulation into existing mathematical models of cellular processes was developed. The method was used to modify existing mathematical model of cell cycle. The performance of the modified model was observed to be improved.

Certain aspects of stochasticity in the gene expression processes were investigated in this study. It has been observed that noise in the steady state protein level decreases with increasing abundance. However, at high protein abundance the noise saturates to a level and all high-abundance proteins show a same low level of noise. The experimentally observed noise floor has been attributed to noise sources extrinsic to gene expression. A detailed model of gene expression process was developed to identify the sources of extrinsic noise contributing to the observed saturation of noise. Using stochastic simulations the effect of the timescale of fluctuations on saturation of noise was studied.

The knowledge about the important step in gene expression, that maximally affects noise at steady state protein level is ambiguous. In this study, a global stochastic sensitivity analysis was carried out to quantitatively estimate the relative contribution of major steps in gene expression to the noise at steady state protein level. Global sensitivity analysis using different measures of noise demonstrated differential sensitivity of these measures to different parameters.

The thesis is divided into five chapters. This chapter gives an overview of gene expression, its regulation, stochasticity in gene expression, and requirement of mathematical analysis. Chapter 2 describes the study of mathematical modeling and analysis of auto-regulatory synthesis of TBP. The third chapter is divided into two sections. Section 1 details the study of mathematical modeling of miRNA mediated

regulation to explore its effect on the steady state of target protein. Section 2 describes the method to incorporate dynamic effects of intronic miRNA mediated regulation into existing mathematical models. Chapter 4 summarizes the work related to the study of noise in gene expression. It is divided into two sections. First section describes the work of development of a detailed model of gene expression to investigate the sources of extrinsic noise resulting in the saturation of noise at high protein abundance. Second section describes the global stochastic sensitivity analysis to estimate the relative contribution of major steps in gene expression to the noise at steady state protein level. The overall conclusion and future directions are discussed in the fifth chapter.

In summary, the present study explored certain properties of regulation and stochasticity in gene expression using mathematical modeling approach. By analyzing the developed models using appropriate mathematical framework such as deterministic and stochastic, the study provided explanations for certain unintuitive observations and generated some falsifiable hypotheses that can be experimentally tested.

2 Mathematical model of transcriptional regulation by the transcription factor TATA Binding Protein

2.1 Introduction

TATA Binding Protein (TBP) is a ubiquitously expressed general transcription factor. It binds to a regulatory sequence in the promoter region called TATA box. The TATA box is a conserved sequence TATAAT that occurs at about 30 to 40 base pairs upstream of the transcription start site. Once bound, it recruits other transcription factors and subunits of RNA polymerase complex to start transcription initiation. In case of the other class of promoters called TATA-less promoters TBP is known to bind to regulatory region through interaction with other proteins called tethering factors (Pugh and Tjian 1991). The promoters of genes transcribed by RNA polymerase I are known to be TATA-less promoters. In such case TBP binds to regulatory region as a subunit of the selectivity factor 1 (SL1). The RNA polymerase II promoters resulting in synthesis of mRNA are both TATA-containing and TATA-less promoters. Transcription from both the promoters requires TBP as a subunit of TFIID. The other class of RNA polymerase II promoters is snRNA promoters, which are TATA-less promoters. TBP binds to the proximal sequence element (PSE) of these promoters as a part of snRNA activating protein complex, $SNAP_c$ (Hernandez 1993). The RNA polymerase III promoters are TATA-less promoters. TBP is known to be a part of the initiation complex at these promoters as a subunit of TFIIB (White, Jackson et al. 1992; Tjian 1996). Experimental study by Cormack and Struhl have shown that depletion or inactivation of TBP results in rapid decrease in transcription by all the three RNA polymerases in yeast (Cormack and Struhl 1992), indicating the requirement of TBP by all the three RNA polymerases for transcription initiation. The promoter for the gene of TBP protein, in diverse species of vertebrates, itself is a TATA-less promoter (Chalut, Gallois et al. 1995; Ohbayashi, Schmidt et al. 1996). However, irrespective of presence of TATA box, TBP is crucial for transcription initiation. This suggests the requirement of TBP for its own transcription as well, implying an underlying positive auto-regulation of TBP for its own synthesis.

X-ray crystallographic studies have shown that TBP is a saddle shaped molecule. Through its concave surface it interacts with the minor groove of DNA and through its convex surface it interacts with other protein molecules (Nikolov, Chen et al. 1996). It has been shown that free TBP monomers when not bound to DNA forms TBP dimers

(Coleman, Taggart et al. 1995). The interaction for dimerization occurs through the concave surface of TBP. Therefore, the dimer is incapable of binding to DNA. Dimeric form of TBP is quite stable having an equilibrium dissociation constant in the nanomolar range (Coleman and Pugh 1997). TBP dimerization competes with TBP-DNA binding. Additionally, slow dissociation of dimers could be a rate limiting step in availability of TBP monomers for DNA binding. Thus dimerization can be considered as a mechanism for negatively auto-regulating DNA binding activity to prevent unregulated gene expression by limiting access to DNA. TBP dimers have been shown to be more stable than monomers and thus dimerization of TBP is also suggested to prevent TBP from degradation (Jackson-Fisher, Chitikila et al. 1999).

TBP has been found to play an important role in a neurodegenerative disorder spinocerebellar ataxia type 17 (SCA17). N-terminal domain of TBP contains a stretch of polyglutamine (polyQ) ranging in size from 29-42 residues. Increase in the length of the CAA/CAG composite repeat for glutamine beyond 42 has been shown to cause SCA17 (van Roon-Mom, Reid et al. 2005). Experimental study by Freidman and Wang et al (Friedman, Wang et al. 2008) have reported that polyQ expansion reduced in-vitro DNA binding affinity of TBP indicating the role of altered affinity in neurodegenerative diseases. In addition to the role of TBP in diseases, another interesting property of TBP is its maternal inheritance. A study by Edelmann et al have shown that sea urchin oocytes contained high levels of TBP mRNA that were required for embryogenesis (Edelmann, Zheng et al. 1998). Such maternal contribution is crucial for cell viability in early events of development as the cells are transcriptionally inactive.

In this study, these observed properties of TBP were investigated using mathematical modeling. The TBP-DNA binding has been previously modeled (Coleman and Pugh 1997) but the auto-regulatory synthesis and effect of dimerization has not been studied. There are known transcription factors which dimerize and bind to DNA (Halvorsen, Nandabalan et al. 1990) or in some cases monomeric form binds to multiple sites in the regulatory region of a gene (Cranz, Berger et al. 2004). Mathematical model of these motifs of transcriptional regulation has been developed (Keller 1995; Verma, Rawool et al. 2006). The regulation of TBP differs from other known transcription factors as it binds

to DNA as a monomer at one site on the DNA, however free TBP can exists as a dimer. Thus study of the regulatory motif of TBP is important. In this study a mathematical model of TBP considering auto-regulatory synthesis of TBP and negative regulation by dimerization was developed. The effect of variation in the initial conditions, and the parameters on the steady state of TBP was analysed. The role of dimer in buffering the TBP system against perturbation was investigated. Using the model we could suggest some explanations for biological properties of maternal inheritance and the effect of reduced TBP-DNA binding affinity on cell viability.

2.2 Methodology

2.2.1 Model development

The biological interactions of TBP considered in the model are represented in Figure 2-1.

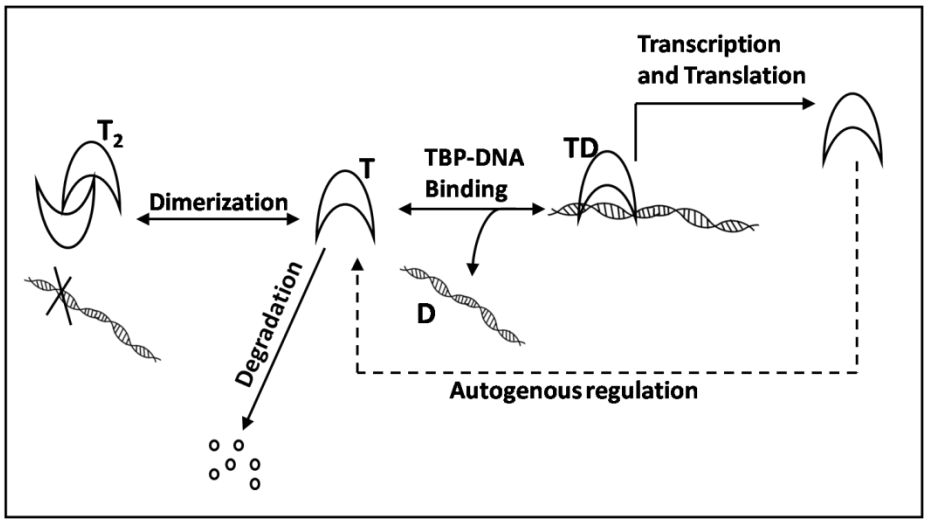


Figure 2-1 – A schematic representation of the biological interactions involved in auto-regulation of TBP

The arc shapes represent TBP molecules. TBP (T) can bind to DNA (D) to form TBP-DNA (TD) complex which initiates TBP synthesis. Free TBP can dimerize (T_2). TBP dimers cannot bind to DNA. Monomeric TBP can undergo degradation.

Binding of TBP to DNA was considered as a reversible reaction following mass action kinetics. Similarly, TBP dimerization was also considered as a reversible reaction having mass action kinetics. In the reaction of TBP synthesis details of mechanism of

transcription and translation were not included. It was considered as a single reaction by lumping transcription and translation into one step. Such approximation has been previously used in models of transcriptional regulation (Keller 1995; Isaacs, Hasty et al. 2003). Transcription and translation are known to be non-linear processes. Therefore, the reaction of TBP synthesis was assumed to follow Hill kinetics (Becskei, Seraphin et al. 2001). The use of Hill kinetics did not imply any particular mechanism of TBP synthesis but only represents a non-linear, saturating nature of protein synthesis. The rate of reaction was assumed to be dependent upon the concentration of TBP-DNA complex. It was assumed that TBP had the same affinity for all TBP binding sites. Therefore, the concentration of TBP-bound TBP-promoter was considered as a constant fraction of the total TBP bound DNA. A basal synthesis term (k_0) was included in the rate expression for TBP synthesis. However, TBP being crucial for its own synthesis, analysis was performed by setting the value of k_0 to zero. The reaction for degradation of TBP was assumed to be first order mass action kinetics. Knowing that TBP dimer is stable than monomeric TBP (Jackson-Fisher, Chitikila et al. 1999), degradation of TBP dimer was not considered in the model. The reactions used in the model are listed in Table 2-1.

Table 2-1 – Reactions used in the mathematical model of TBP

Reaction	Rate equation
$T + T \rightarrow T_2$	$k_1 \times [T]^2$
$T_2 \rightarrow T + T$	$k_2 \times [T_2]$
$T + D \rightarrow TD$	$k_3 \times [T] \times [D]$
$TD \rightarrow T + D$	$k_4 \times [TD]$
$\Phi \rightarrow T$	$k_0 + \frac{k_5 \times [TD]^{k_6}}{k_7^{k_6} + [TD]^{k_6}}$
$T \rightarrow \Phi$	$k_8 \times [T]$

The differential equations for the four components viz., free TBP (T), TBP dimer (T_2), TBP-DNA complex (TD) and unbound DNA (D) were formulated considering the rates of synthesis, degradation and conversion. The mass balance equations are given as

$$\frac{d[T]}{dt} = 2\times k_2 \times [T_2] - 2\times k_1 \times [T]^2 + k_4 \times [TD] - k_3 \times [T] \times [D] + \left(k_0 + \frac{k_5 \times [TD]^{k_6}}{k_7^{k_6} + [TD]^{k_6}} \right) - k_8 \times [T] \quad (2.1)$$

$$\frac{d[T_2]}{dt} = k_1 \times [T]^2 - k_2 \times [T_2] \quad (2.2)$$

$$\frac{d[D]}{dt} = k_4 \times [TD] - k_3 \times [T] \times [D] \quad (2.3)$$

$$\frac{d[TD]}{dt} = k_3 \times [T] \times [D] - k_4 \times [TD] \quad (2.4)$$

The concentration of total TBP binding sites [D_0] is given as

$$[D] + [TD] = [D_0] \quad (2.5)$$

The reaction rate parameters for dimerization and TBP-DNA binding were obtained from model by Pugh et al (Coleman and Pugh 1997). Affinity of TBP for TATA-containing and TATA-less promoters was assumed to be the same. The molecular abundance for TBP reported for different cell types such as yeast, mammalian cell, and sea urchin egg cell covers a wide range of 2000 to $2*10^6$ (Edelmann, Zheng et al. 1998; Borggrefe, Davis et al. 2001). Variation in cell sizes from 1 μm radius for yeast cell to around 50 μm radius for sea urchin egg cell resulted in cell volume range of $4.17*10^{-15}$ to $5.22*10^{-10}$ litres assuming spherical cells. Thus the concentration range for TBP was calculated to be 10^{-5} to 10^{-8} M. The number of genes expressed in a cell represents the minimum number of TBP binding sites. Studies have shown a large range for number of expressed genes in different cell types, from ~3000 (Velculescu, Zhang et al. 1997) in yeast cell to ~10000 in mammalian cells (Lewin, Krebs et al. 2009). In mammalian cells the maximum number of TBP binding sites was reported to be ~80,000 (Denissov, van Driel et al. 2007). Hence the physiological concentration range for TBP binding sites was calculated to be 10^{-5} to

10^{-9} M. The details of concentration ranges for TBP and DNA are given in Appendix Ia. In the model, a mammalian cell condition with ~25000 TBP molecules and ~25000 TBP binding sites was considered. Reaction rate parameter for TBP degradation was obtained from known half life of wild type TBP (Jackson-Fisher, Chitikila et al. 1999). Specific rate for TBP synthesis was set to a value to obtain the physiological concentration of TBP. The parameters are given in Table 2-2.

Table 2-2 – Parameter values used in the mathematical model of TBP

Parameter value	**Reference**
$k_1 = 1*10^5\ M^{-1}s^{-1}$	(Coleman and Pugh 1997)
$k_2 = 1*10^{-3}\ s^{-1}$	(Coleman and Pugh 1997)
$k_3 = 2*10^5\ M^{-1}s^{-1}$	(Coleman, Taggart et al. 1995; Coleman and Pugh 1997; Weideman, Netter et al. 1997)
$k_4 = 4*10^{-4}\ s^{-1}$	(Weideman, Netter et al. 1997)
$k_5 = 5*10^{-13}\ Ms^{-1}$	From (Schmidt and Schibler 1995) and k_8
$k_7 = 1.25*10^{-8}M$	Half of D_0
$k_8 = 7.4*10^{-5}\ s^{-1}$	(Jackson-Fisher, Chitikila et al. 1999)
$k_6 = 2$	Assumed
$k_0 = 5*10^{-15}\ Ms^{-1}$	Assumed ($0.01*k_5$)
$[D_0] = 2.5*10^{-8}\ M$	(Denissov, van Driel et al. 2007)

The ordinary differential equations (2.1-2.5) were solved numerically using the stiff differential equation solver `ode15s` of Matlab version 7.6.0.324 (The Mathworks, Natick, USA). The effect of variation of initial molecular concentration was analysed. In addition, the effect of variation of specific reaction rates for TBP-DNA binding affinity (k_3), TBP synthesis (k_5), and Hill coefficient (k_6) on the steady state of TBP was investigated.

2.2.2 Steady state solution

The concentration of each component of the model remains constant at steady state. Thus, the equations (2.1-2.4) were equated to zero and solved to obtain a function in terms of free TBP (T). Further, the system of 4 equations (2.1-2.4) was reduced to a system with 3 equations (2.8-2.10) using the conservation for total TBP binding sites (equation 2.5). The equations were non-dimensionalized using $[D_0]$ as reference concentration and $1/k_8$ as reference time (equations 2.11-2.13). κ_i are corresponding dimensionless parameters. Analytical expressions for the steady state levels were obtained. For a condition of zero basal synthesis of TBP, $k_0 = 0$ Ms^{-1}, the expression for critical TBP-DNA binding affinity (κ_3_c) and critical TBP-synthesis (κ_5_c) was obtained. κ_3_c and κ_5_c represented the values below which there existed only one steady state for TBP system. The analysis was performed using Mathematica version 7.0.1.0 (Wolfram Research, Champaign, USA).

2.2.3 Simulations for dimerization effect

Two systems, one original system with dimerization reaction and other system without dimerization reaction, were compared to study the effect of presence of TBP dimer. A hypothetical system was considered in which the specific rate for dimerization reaction (k_1) was set to zero. Removing the reaction for dimerization reduced the total TBP concentration by the amount equivalent to TBP dimer. The high-TBP steady state for the two systems was considered to be initial condition. The two systems were perturbed by decreasing the free TBP concentration by 10% of its steady state concentration. The response time, defined as the time required for a system to reach 99% of its steady state value was compared. 99% of the steady state level corresponds to 90% of recovery from the perturbation. The effect of presence of dimer was studied by varying reaction rate parameters, k_5 and k_7 and TBP binding sites D_0 (in range of 10^{-5} to 10^{-9}M), resulting in range of TBP concentration as that observed in different physiological conditions. In addition to the numerical analysis, perturbation analysis was performed using analytical method.

2.3 Results and discussion

2.3.1 TBP system showed presence of multiple steady states

At steady state, the concentration of all the components remains constant. The equations (2.1-2.4) were solved by equating each to zero. The non-linear function *f* in terms of single variable, free TBP (T) was obtained.

The function *f* is given as,

$$f([T]) = (k_8 \times [T] - k_0) \times \left(k_7{}^{k_6} + \left(\frac{[T] \times [D_0]}{\frac{k_4}{k_3} + [T]} \right)^{k_6} \right) - k_5 \times \left(\frac{[T] \times [D_0]}{\frac{k_4}{k_3} + [T]} \right)^{k_6} = 0 \qquad (2.6)$$

When basal TBP synthesis was set to zero, i.e. $k_0 = 0$ Ms^{-1} the expression is given as,

$$f([T]) = \left(k_8 \times [T]\right) \times \left(k_7{}^{k_6} + \left(\frac{[T] \times [D_0]}{\frac{k_4}{k_3} + [T]} \right)^{k_6} \right) - k_5 \times \left(\frac{[T] \times [D_0]}{\frac{k_4}{k_3} + [T]} \right)^{k_6} = 0 \qquad (2.7)$$

The equations for steady state concentration of TBP (equations 2.6 and 2.7) were plotted against equivalent total TBP concentration to obtain the solutions of the function ($f = 0$). Figure 2-2 shows the graph of function value *f* against total TBP concentration for the two equations 2.6 and 2.7.

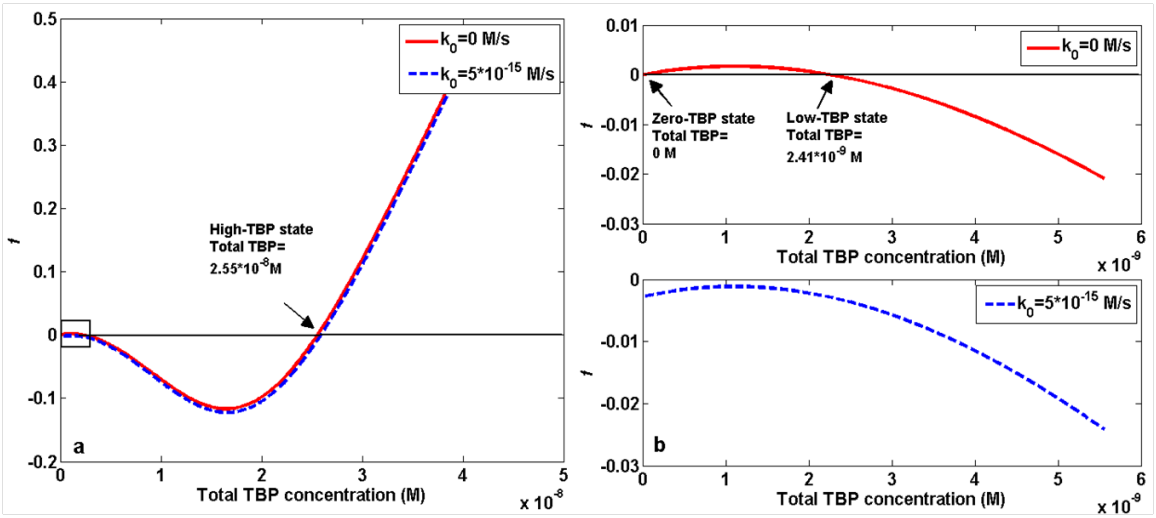


Figure 2-2 – Graphical solution for TBP steady states

(a) The function of TBP concentration described by equations 2.6 and 2.7 was plotted against corresponding total TBP concentration. (b) Magnified view of the region enclosed in the square in 2a.

In the presence of basal TBP synthesis ($k_0 > 0$) only one steady state was observed. However, knowing the requirement of TBP for its own synthesis, it was assumed that no TBP synthesis occurs in the absence of TBP. Thus basal synthesis rate, k_0 was set to zero to represent biologically relevant condition. In the absence of basal synthesis, presence of three steady states was observed. The concentration of total TBP corresponding to these steady states was observed to be 0 M, $2.41*10^{-9}$ M, $2.55*10^{-8}$ M. These states were referred to as a zero-TBP state, a low-TBP state and a high-TBP state respectively. The numerical values of non-zero steady states were observed to change with variation in parameter values. However, the highest concentration state was referred to as a high-TBP state while the other non-zero TBP state was referred to as low-TBP state. From the linear stability analysis, it was observed that the low-TBP state was unstable while the zero-TBP state and the high-TBP state were stable steady states. The TBP concentration for the single steady state observed in the system with basal TBP synthesis ($k_0 > 0$) was equivalent to the high-TBP state. The critical value of basal synthesis, k_{0_c} below which there were three steady states was found out to be $0.008*k_5$ i.e. $4*10^{-15}$ Ms^{-1}. Conditions for multistability with respect to basal TBP synthesis (k_0) and Hill co-operativity

coefficient (k_6) were identified using analytical expression for steady state levels of TBP. To identify the conditions for multistability the TBP system was reduced to a system with three variables viz., TBP (T), TBP dimer (T_2) and TBP-DNA complex (TD) using the conservation equation (2.5).

The differential equations for three-variable system are,

$$\frac{d[T]}{dt} = -2\times k_1\times[T]^2 + 2\times k_2\times[T_2] - k_3\times[T]\times([D_0]-[TD]) + k_4\times[TD] + \frac{k_5\times[TD]^{k_6}}{k_7^{k_6}+[TD]^{k_6}} - k_8\times[T] \quad (2.8)$$

$$\frac{d[T_2]}{dt} = k_1\times[T]^2 - k_2\times[T_2] \quad (2.9)$$

$$\frac{d[TD]}{dt} = k_3\times[T]\times([D_0]-[TD]) - k_4\times[TD] \quad (2.10)$$

The system of three variables (equations 2.8-2.10) was non-dimensionalized using $[D_0]$ as reference concentration and $1/k_8$ as reference time. The dimensionless parameters (κ_1 to κ_7) are given in Appendix Ib. The non-dimensionalized equations are as follows.

$$\frac{d[T]}{dt} = -2\times\kappa_1\times[T]^2 + 2\times\kappa_2\times[T_2] - \kappa_3\times[T]\times(1-[TD]) + \kappa_4\times[TD] + \frac{\kappa_5\times[TD]^{k_6}}{\kappa_7^{k_6}+[TD]^{k_6}} - [T] \quad (2.11)$$

$$\frac{d[T_2]}{dt} = \kappa_1\times[T]^2 - \kappa_2\times[T_2] \quad (2.12)$$

$$\frac{d[TD]}{dt} = \kappa_3\times[T]\times(1-[TD]) - \kappa_4\times[TD] \quad (2.13)$$

Conditions for existence of multiple steady states were found out by analyzing the expression for steady state TBP level. Following conditions were considered for different κ_0 and κ_6 values. The conditions are summarized in Table 2-3.

Table 2-3 – Conditions for existence of multiple steady states

	$\kappa_0 = 0$	$\kappa_0 \neq 0$
$k_6 = 2$	3 real positive steady states ($\kappa_5 > 2\kappa_7^2\kappa_4/\kappa_3$ and $\kappa_7 < 1$)	1 real positive steady state ($\kappa_0 > \kappa_{0\_c}$) 3 real positive steady states ($\kappa_0 < \kappa_{0\_c}$)
$k_6 = 1$	2 real positive steady states ($\kappa_5 > \kappa_7\kappa_4/\kappa_3$)	1 real positive steady state ($\kappa_0 << \kappa$ and $\kappa_5 > \kappa_7\kappa_4/\kappa_3$)

The details of these conditions are given as follows.

1. $\kappa_0 = 0$, $k_6 = 2$

Three steady states were observed.

$$T_{ss1} = 0$$

$$T_{ss2} = \frac{\kappa_5 - 2\kappa\kappa_7^2 - \sqrt{\kappa_5^2 - 4\kappa^2\kappa_7^2 - 4\kappa\kappa_5\kappa_7^2}}{2(1+\kappa_7^2)}$$

$$T_{ss3} = \frac{\kappa_5 - 2\kappa\kappa_7^2 + \sqrt{\kappa_5^2 - 4\kappa^2\kappa_7^2 - 4\kappa\kappa_5\kappa_7^2}}{2(1+\kappa_7^2)}$$

Where, $\kappa = \kappa_4/\kappa_3$

It was evident that multiple real, non-negative steady states would exist when $\kappa_5 > 2\kappa\kappa_7^2$. Simplification of expression for T_{ss2} led to

$$T_{ss2} = \frac{\kappa_5 - 2\kappa\kappa_7^2 - \sqrt{(\kappa_5 - 2\kappa\kappa_7^2)^2 - 4\kappa^2\kappa_7^2(1-\kappa_7^2)}}{2(1+\kappa_7^2)}$$

It was observed that $\kappa_7 < 1$, would lead to a non-negative steady state. Therefore, in this case 3 non-negative real steady states were observed.

2. $\kappa_0 \neq 0$, $\kappa_0 < \kappa_{0\_c}$, $k_6 = 2$

Three steady states were observed.

3. $\kappa_0 = 0$, $k_6 = 1$

Two steady states were observed.

$$T_{ss1} = 0$$

$$T_{ss2} = \frac{\kappa_5 - \kappa\kappa_7}{1+\kappa_7}$$

The condition, $\kappa_5 > \kappa\kappa_7$ would result in two real steady states.

4. $\kappa_0 \neq 0$, $k_6 = 1$

Two steady states were observed.

$$T_{ss1} = \frac{\kappa_0 + \kappa_5 - \kappa\kappa_7 + \kappa_0\kappa_7 - \sqrt{4\kappa\kappa_0\kappa_7(1+\kappa_7) + (\kappa_0 + \kappa_5 - \kappa\kappa_7 + \kappa_0\kappa_7)^2}}{2(1+\kappa_7)}$$

$$T_{ss2} = \frac{\kappa_0 + \kappa_5 - \kappa\kappa_7 + \kappa_0\kappa_7 + \sqrt{4\kappa\kappa_0\kappa_7(1+\kappa_7) + (\kappa_0 + \kappa_5 - \kappa\kappa_7 + \kappa_0\kappa_7)^2}}{2(1+\kappa_7)}$$

For the conditions $\kappa_0 << \kappa$ and $\kappa_5 > \kappa\kappa_7$, the term in the square root $(\kappa_0+\kappa_5-\kappa\kappa_7+\kappa_0\kappa_7)$ results in positive term. Simplifying the expression for T_{ss1},

$$T_{ss1} = \frac{(\kappa_0 + \kappa_5 - \kappa\kappa_7 + \kappa_0\kappa_7) - \sqrt{(\kappa_0 + \kappa_5 - \kappa\kappa_7 + \kappa_0\kappa_7)^2 + 4\kappa_0(1+\kappa_7)\kappa\kappa_7}}{2(1+\kappa_7)}$$

This indicated that T_{ss1} would be the negative steady state and T_{ss2} would be non-negative steady state.

2.3.2 Minimum amount of initial TBP is required for cell viability

It was observed that starting from different initial conditions the system reached one of the two stable steady states depending upon the initial condition. Figure 2-3 shows a phase plane plot of TBP-DNA and free TBP concentration.

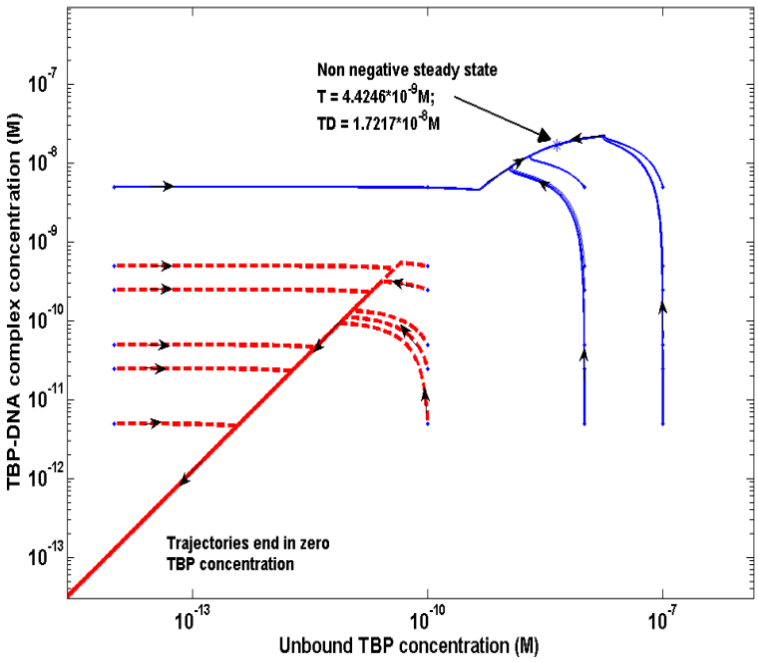


Figure 2-3 – Phase plane plot of TBP-DNA complex vs. free TBP

Bistable nature of the system and the requirement of minimum amount of TBP is illustrated by the phase plane plot. Depending upon the initial conditions the system reached either of the stable steady states.

It was evident that below a certain total TBP concentration all the trajectories led to the zero-TBP stable steady state, while above a certain total TBP level system reached to the high-TBP steady state. A physiological state resulting in low total TBP concentration of lesser than ~0.1 nanomolar will be in the zone of attraction of the zero-TBP state and would result in cell death. The model predicted that a certain minimum amount of TBP was needed to initially start its own synthesis. This property of positive auto-regulatory synthesis explained the reason for observed maternal inheritance of TBP (Edelmann,

Zheng et al. 1998) similar to other proteins involved in core processes of gene expression (Payer, Saitou et al. 2003; Farley and Ryder 2008).

In the phase plane plot, the trajectories of TBP-DNA complex and free TBP were observed to merge into a line. Relatively fast reaction of TBP-DNA binding compared to slow reaction of TBP synthesis and degradation led to such nature of trajectories suggesting rapid equilibrium of TBP-DNA complex with free TBP. The line represented the pseudo-equilibrium concentrations of TBP-DNA and free TBP. The ratio of the concentrations of free TBP and free DNA to TBP-DNA complex ([T]*[D]/[TD]) for values corresponding to the straight line region of the trajectories was found to be equal to the equilibrium dissociation constant for TBP-DNA binding.

2.3.3 High-TBP steady state was found to be sensitive to variation in reaction parameters

To examine the sensitivity of steady states to variation in parameters, the steady state concentration of TBP was calculated at different values of reaction rate parameters. The sensitivity was determined qualitatively and quantitatively. To qualitatively test the sensitivity, change in the number of steady states and their stability was examined. For quantitative determination of sensitivity the relative change in the value of high-TBP steady state to a change in parameter value was calculated. Sensitivity to variation in the three parameters viz., specific rate of TBP-DNA binding (k_3), TBP synthesis (k_5), and the Hill coefficient (k_6) was examined.

2.3.3.1 Sensitivity to TBP-DNA binding affinity (k_3)

Figure 2-4 shows a plot of total TBP steady state concentration vs. the specific rate of TBP-DNA binding.

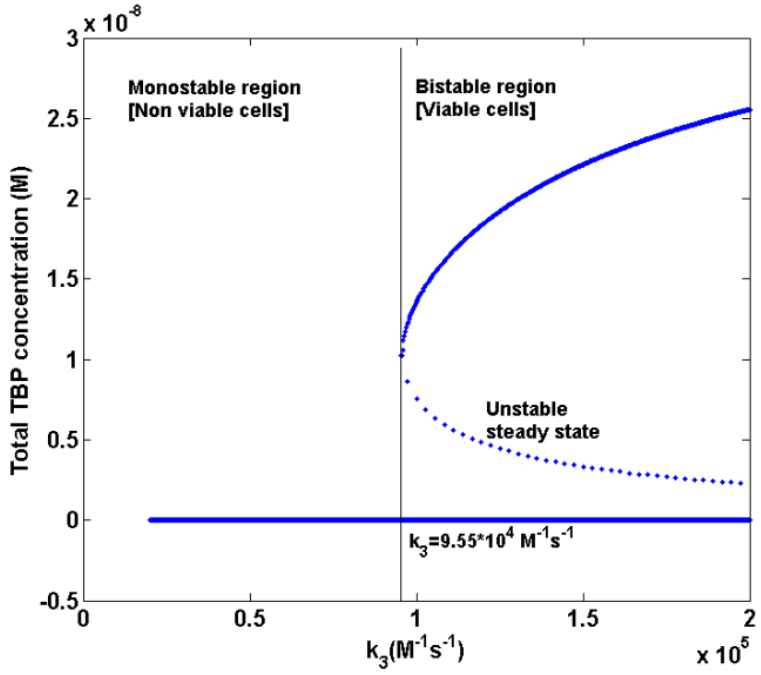


Figure 2-4 – Bifurcation plot for specific rate of TBP-DNA binding (k_3)

Parameter k_3 was varied in simulations. Other parameters values were kept constant at values as reported in Table 2-2.

It was observed that for a value of k_3 greater than $9.556*10^4$ $M^{-1}s^{-1}$ three steady states existed. However, below the bifurcation value (k_3_c) only one stable steady state corresponding to zero-TBP state was observed, indicating unviable cells. Thus only for the values of k_3 above the critical values, k_3_c the system could reach either of the stable steady states depending upon the initial conditions. At the bifurcation value of k_3, the high-TBP state and low-TBP state were observed to have the same value. Therefore, equating the expressions for the two steady states, analytical expression for the critical value of k_3 (k_3_c) was obtained.

$$\kappa_{3\_c} = \frac{2\left(\kappa_4\kappa_5\kappa_7^{\,2} + \sqrt{\kappa_4^{\,2}\kappa_5^{\,2}\kappa_7^{\,2} + \kappa_4^{\,2}\kappa_5^{\,2}\kappa_7^{\,4}}\right)}{\kappa_5^{\,2}} \qquad (2.14)$$

Using the parameter values given in Table 2-2, numerical value for k_3_c was calculated to be $9.58*10^4$ $M^{-1}s^{-1}$, which was comparable to the value obtained from numerical simulations.

A large variation in sensitivity of high-TBP state to variation in k_3 was observed, for different conditions tested. Details are given in Appendix Ic. Overall, the high-TBP steady state was observed to be sensitive to variation k_3. For the reference parameter set (Table 2-2), it was observed that for 50% decrease in k_3 value there was ~47% reduction in the high-TBP steady state concentration.

Several mutants of TBP with altered DNA binding affinity have been reported (Strubin and Struhl 1992). Naturally occurring such mutants are known to be associated with cancer and neurodegenerative diseases (van Roon-Mom, Reid et al. 2005; Friedman, Wang et al. 2008). The model predicted that the high-TBP steady state was sensitive to variation in the binding affinity of TBP to DNA. It indicated that with around 50% reduction in the affinity than the physiological value, the system moves to a regime where zero-TBP state is the only stable steady state. This suggested that at low affinity of TBP for DNA, the cell may not be able to survive irrespective of the initial TBP concentration. This prediction has an implication in understanding the role of low DNA affinity mutant of TBP in neurodegenerative disease.

2.3.3.2 Sensitivity to specific rate of TBP synthesis (k_5)

Figure 2-5 shows a plot of total TBP steady state concentration vs. the specific rate of TBP synthesis.

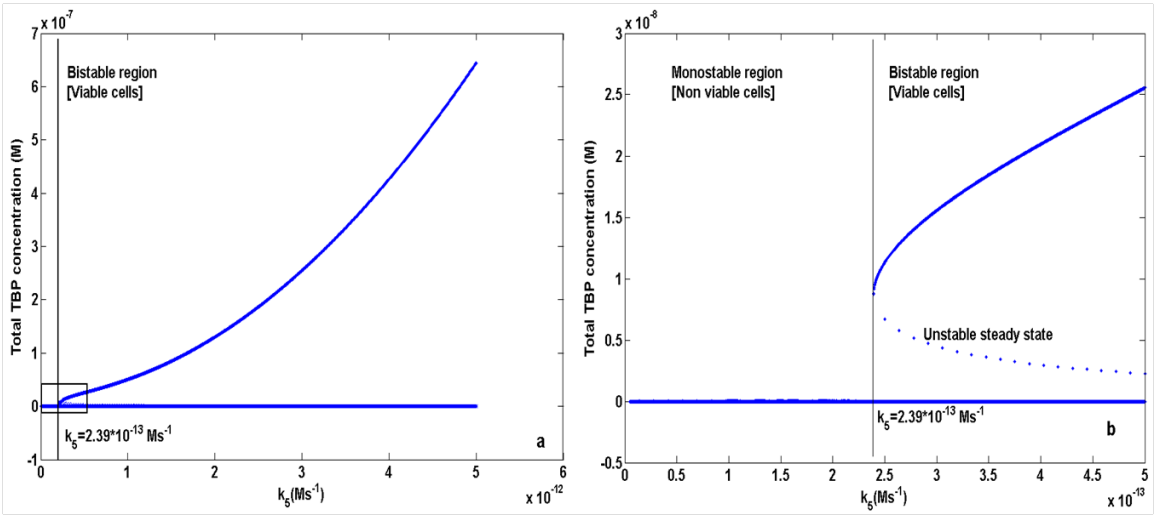


Figure 2-5 – Bifurcation plot for specific rate of TBP synthesis (k_5)

(a) Parameter k_5 was varied in the simulations. Other parameters values were kept constant at values as reported in Table 2-2. (b) Magnified view of the region enclosed in the square in 5a.

Similar to the specific rate of TBP-DNA binding affinity, only one stable state corresponding to zero-TBP state was observed up to a critical value of TBP synthesis (k_5_c) of 2.394*10^{-13} M/s. For the values above this critical value two stable steady states corresponding to high-TBP and zero-TBP and a low-TBP unstable steady state were observed. At the critical value of k_5 the high-TBP and the low-TBP states were observed to have same value. Equating the expressions for these two steady states the analytical expression for k_5_c was obtained.

$$\kappa_5\_c = 2\left(\kappa\kappa_7^{\,2} + \sqrt{\kappa^2\kappa_7^{\,2} + \kappa^2\kappa_7^{\,4}}\right) \tag{2.15}$$

Using the parameters values the value of k_5_c was found to be 2.39*10^{-13} M/s, same as that obtained from numerical simulations.

The high-TBP steady state was observed to be slightly more sensitive to variation in k_5 than k_3. For instance, for 50% reduction in k_5 there was ~55% reduction in the high-TBP steady state concentration.

Both the specific rate of TBP-DNA binding affinity (k_3) and TBP synthesis (k_5) were simultaneously varied to test the effect on the steady states. Figure 2-6 shows the effect of variation of two parameters on the number of steady states. The two regions, one corresponding to monostable regime of zero-TBP steady state and the other corresponding to bistable regime of high-TBP state and zero-TBP state were observed.

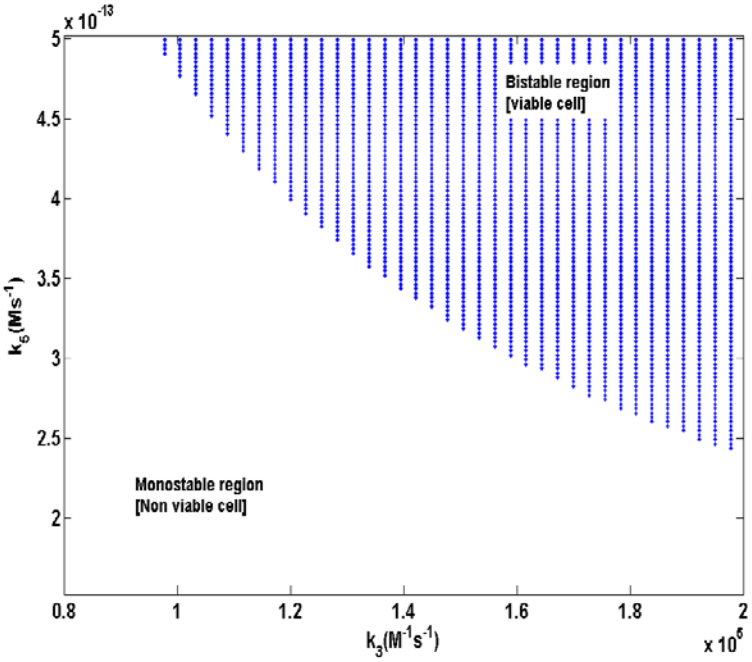


Figure 2-6 – Graph showing the effect of simultaneous variation of two reaction rate parameters k_3 and k_5 on the number of steady states of TBP system

Parameter k_3 and k_5 was simultaneously varied in simulations. Other parameters values were kept constant at values as reported in Table 2-2.

2.3.3.3 Sensitivity to the Hill co-operativity coefficient (k_6)

Similar to the other parameters, the two regimes of monostable state and bistable state were observed (Figure 2-7).

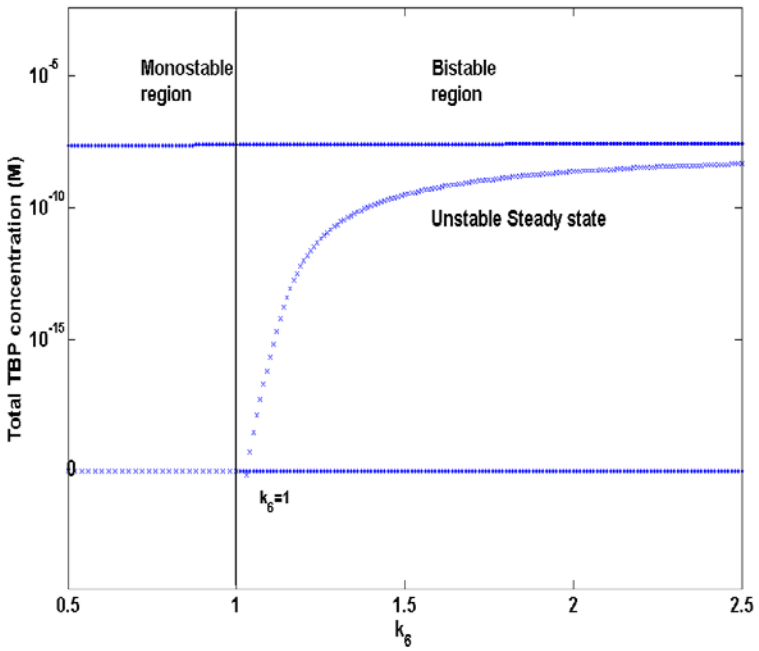


Figure 2-7 – Bifurcation plot for Hill co-operativity coefficient (k_6)

Parameter k_6 was varied in simulations. Other parameters values were kept constant at values as reported in Table 2-2.

However, for the value of k_6 below 1 there were two steady states corresponding to zero-TBP state and high-TBP state. In this case the steady state corresponding to high-TBP state was observed to be the stable steady state. In contrast to sensitivity to other parameters, the high-TBP steady state was observed to be less sensitive to variation in k_6. For 50% reduction in k_6 the value of high-TBP steady state concentration decreased only by ~9%.

2.3.4 TBP dimer can help to buffer the system against perturbations only under certain conditions

It is known that under physiological conditions free TBP exists in dimeric form. Using the developed model the role of TBP dimer in buffering the system against transient perturbations in TBP level for a range of physiological concentrations of TBP molecules and TBP binding sites was explored. From expression 2.6 it was clear that presence or absence of dimer does not change the number of steady state in the system. However, the total TBP concentration was changed by the amount equivalent to the level of dimer, $[T_2]$

= $(k_1/k_2)*[T]_{ss}^2$. To investigate the effect of presence of dimer, the original reaction system was compared with another system without dimerization reaction. The response time of the system with dimer was compared with the system without dimer after a transient perturbation of 10% decrease in free TBP level than the high-TBP state concentration. In addition to numerical simulations, perturbation analysis was performed using analytical method and the ratio of the response time obtained using both the methods was observed to be in agreement (correlation coefficient = 0.99). The details of analytical perturbation study are given in Appendix Id. Figure 2-8 shows the graph of ratio of response time in the presence of dimer to that in the absence of dimer as a function of dimer concentration relative to TBP bound DNA concentration.

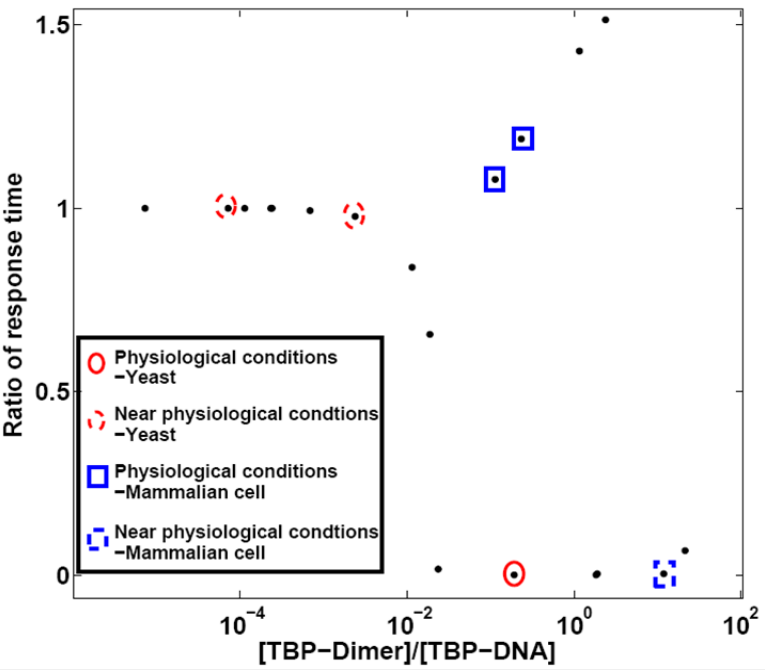


Figure 2-8 – The graph of ratio of response time in the presence of dimer to that in the absence of dimer vs. the ratio of TBP dimer concentration to TBP-DNA complex concentration

Unannotated points represent conditions that are not likely to be physiological either in yeast or mammalian cell.

It was observed that for certain conditions the ratio was around one indicative of no effect of presence of the dimer, while under certain conditions the ratio was observed to be much lesser than 1 indicating that the presence of dimer reduced the response time. It was observed that for the typical yeast cell condition and a near mammalian physiological

condition, the presence of TBP-dimer reduced the ratio of response time. However, for typical mammalian physiological condition and near yeast physiological conditions presence of dimer did not significantly change the response time. This indicated that the dimer did not always help to buffer the system against perturbations. It was observed that the relative concentration of TBP dimer and TBP-DNA complex was an important factor to determine the response against perturbations. When the relative concentration of TBP-dimer to TBP-DNA complex was higher, for instance, near physiological conditions for mammalian cells, dissociation of dimer helped the system to restore the free TBP steady state level. While when the TBP dimer concentration was less than TBP-DNA complex concentration, for instance, near physiological conditions for yeast, dissociation of TBP-DNA complex would help to restore the free TBP steady state level. Hence, in such conditions the response time for both the two systems was observed to be almost same. From the graph it was observed that when the relative concentration of the dimer was high (TBP-dimer/TBP-DNA > 10), presence of dimer helped the system to restore the free TBP level. However, when the relative concentration was low (TBP-dimer/TBP-DNA < 0.01) presence of dimer was not important for to buffer against perturbation. Under certain physiological conditions, where the relative concentration was similar, it was not observed to be good indicator to determine the response against transient TBP perturbation. The dissociation of TBP from DNA is important as it is related to deactivation of genes.

2.4 Conclusion

In this study, a deterministic kinetic model for auto-regulatory synthesis of TBP was developed. A non-linear regulation of TBP bound DNA for TBP synthesis and a negative regulation of TBP dimer to control accessibility of TBP to DNA was considered in the model. Thus the model contained both the positive and negative regulation of TBP. The effect of variation of initial conditions on the steady state of TBP was explored. It was observed that a certain minimum amount of TBP was required to reach a physiological TBP level. Additionally, the effect of variation of reaction rate parameters was explored. It was observed that high-TBP steady state was sensitive to variation in parameters. Reduction in specific rate of TBP-DNA binding moved the system from bistable regime

to a monostable regime with zero-TBP steady state. This indicated that at low values of TBP-DNA binding affinity the cell may not be able to survive irrespective of initial TBP level. Using the model, the effect of presence of dimer in buffering the system against perturbation was investigated. It was observed that the dimer may not always help to buffer against perturbations but the buffering ability was found to be dependent on the relative concentrations of TBP dimer and TBP-DNA complex. Using the mathematical model, the study provided explanations for certain biological observations and made testable predictions.

3 Mathematical modeling and analysis of post-transcriptional regulation by microRNA

3.1 Introduction

RNA molecules are now well known important regulatory factors. After the report by Fire, Xu et al (Fire, Xu et al. 1998), small RNA strands have emerged as important regulators of gene expression. There are various types of small RNA molecules such as small interfering RNA (siRNA), Piwi-interacting RNA (piRNA), and microRNA (miRNA), differentiated based on their biogenesis pathway (Ghildiyal and Zamore 2009). These different types of RNA molecules interact with their target mRNA to regulate protein expression.

miRNA is one important class of non-coding RNA initially reported in *C. elegans* by Lee, Rhonda et al (Lee, Rhonda et al. 1993). miRNAs are around 21 nucleotide long, single stranded RNA molecules (Carthew 2006). Nearly half of the human genome is estimated to be regulated by miRNA (Sonenberg and Hinnebusch 2009). It was estimated that nearly 1% of the genes in vertebrates code for miRNA (Lim, Glasner et al. 2003). These molecules are synthesized either as independent transcripts from microRNA genes or from intronic region of protein coding genes. The microRNA molecules derived from intronic region are called intronic miRNA. Around 40% of the known miRNA are found in the intronic region of genes (Rodriguez, Griffiths-Jones et al. 2004).

microRNA biogenesis is a multistep process. miRNA genes are generally transcribed by RNA polymerase II (Lee, Kim et al. 2004). However, RNA polymerase III is also known to transcribe certain miRNA genes (Borchert, Lanier et al. 2006). Transcription of miRNA genes results in the synthesis of primary-miRNA (pri-miRNA). The pri-miRNA is a double stranded structure containing a hairpin loop. In the nucleus, the pri-miRNA is cleaved by microprocessor complex Drosha-DGCR8/Pasha to form a precursor hairpin structure called pre-miRNA. The pre-miRNA is exported from the nucleus to the cytoplasm by Exportin-5. Once in the cytoplasm, pre-miRNA is cleaved by an RNase called Dicer which is in complex with the double stranded RNA binding protein TRBP, resulting in miRNA duplex without the loop. Out of the two strands, miRNA strand with less stable base pair at its 5’ end is selected while the other strand is degraded. RNA-induced silencing complex (RISC) is loaded onto the selected strand resulting in mature

functional miRNA (Carthew 2006; Ghildiyal and Zamore 2009; Winter, Jung et al. 2009).

The mature miRNA is generally known to bind to 3' UTR of target mRNA to regulate protein synthesis. However, many studies have reported binding to 5' UTR (Richter 2008). Different mechanisms have been reported for post-transcriptional regulation by miRNA. By binding to the target mRNA, miRNA can repress protein synthesis by blocking translation at initiation phase or post-initiation phase by ribosome run-off. The miRNP complex can degrade nascent polypeptide resulting in decreased protein level. Another mechanism is translocation of mRNA to a translationally inactive cellular location called processing body or P-body. miRNA also leads to degradation of target mRNA and can catalyze multiple rounds of degradation (Hutvagner and Zamore 2002; Pillai, Bhattacharyya et al. 2007; Richter 2008). Due to these mechanisms of action miRNA is generally known as repressor of protein expression. However, recently some studies have reported unexpected increased target protein expression even in the presence of miRNA (Vasudevan, Tong et al. 2007; Ghosh, Soni et al. 2008; Ma, Liu et al. 2010).

In this study of miRNA mediated post-transcriptional regulation, a comprehensive mathematical model of miRNA mediated regulation was developed in order to investigate whether the detailed model can explain unintuitive observations of increased target protein. Using the model, certain conditions were identified which can result in increased target protein level even in the presence of miRNA mediated regulation. In addition to the study of effects on the steady state of target protein, the effect of dynamic regulation on the target protein was explored. A new method to incorporate intronic miRNA mediated regulatory effects into existing mathematical models was developed. In addition to its specific application to cell cycle reaction network, universality of the method was also examined.

This chapter is divided into two sections. In section 1, the study of the effect of miRNA regulation on steady state target protein is described. In section 2, development of simple method to incorporate intronic miRNA regulation into existing mathematical models is described.

3.2 Section 1: Analysis of the effects of microRNA mediated regulation suggested explanations for increased target protein level

3.2.1 Introduction

miRNA are generally known as repressors of protein expression. However, in last few years some studies have reported unexpected increase in target protein level in the presence of miRNA. For instance, miR369-3 was shown to up-regulate translation of TNF-α through association with an AU-rich element (Vasudevan, Tong et al. 2007). Another study by Ghosh, Soni et al have shown increased target protein expression due to binding of miR34 to 3' UTR of its target actin transcript (Ghosh, Soni et al. 2008). Up-regulation of IL-10 was reported in the presence of miR4661 due to inhibition of target mRNA degradation (Ma, Liu et al. 2010). miR10a was shown to enhance the translation of mRNA encoding ribosomal proteins by interacting with their 5' UTR (Orom, Nielsen et al. 2008). The 'unexpected' increase in the level of target protein NFIA was attributed to the activity of miR223 (Lu, Buchan et al. 2010). These are some experimental studies that report increased protein level in the presence of miRNA regulation. However, these studies could not identify specific mechanism of the observed unintuitive activating effect of miRNA. Previously developed mathematical models of miRNA regulation focused on the repressive effect and hence did not include an analysis of activating effect of miRNA. In this study a detailed mathematical model of miRNA mediated regulation was developed and analysed to identify conditions for observed activating effect of miRNA.

There are several mathematical models of miRNA mediated regulation that consider different levels of details of the mechanism. The summary of processes considered in different models is given in Table 3-1. mRNA-miRNA binding has been considered as either reversible (Khanin and Higham 2007; Levine, Ben Jacob et al. 2007) or irreversible (Zhdanov 2008; Vohradsky, Panek et al. 2010; Cuccato, Polynikis et al. 2011). One of the mechanism of miRNA regulation is translocation of target mRNA into a translationally inactive cellular location called P-bodies. This process has been considered only in the model by Levine et al (Levine, Ben Jacob et al. 2007) by representing mRNA in processed state. Translocation of mRNA to P-bodies has been

shown to be reversible. The mRNA can again be brought to the cytoplasm where it can be translated actively (Bhattacharyya, Habermacher et al. 2006). However, to our knowledge return of mRNA to cytoplasm has not been considered in the previous models. Translation by the complex has been modeled as occurring with equal efficiency (Levine, Ben Jacob et al. 2007) or unequal efficiency (Khanin and Higham 2007) or zero (Zhdanov 2008; Cuccato, Polynikis et al. 2011). The degradation of mRNA has been modeled as either an implicit function of miRNA level (Khanin and Higham 2007; Levine, Ben Jacob et al. 2007); or explicitly as either a linear function (Zhdanov 2008) or hyperbolic function (Vohradsky, Panek et al. 2010) of the miRNA concentration. The model by Cuccato, Polynikis et al (Cuccato, Polynikis et al. 2011) considered four different mathematical expressions for mRNA degradation, viz., first order, higher order, hyperbolic and Hill kinetics. In addition to these generic models of miRNA regulation there are several other models that study miRNA regulation in the context of specific cellular process, for instance, (Tsang, Zhu et al. 2007; Xie, Yang et al. 2007; Nandi, Vaz et al. 2009).

Table 3-1 – Summary of mathematical models of miRNA mediated regulation

miRNA model reference	**(Levine, Ben Jacob et al. 2007)**	**(Khanin and Higham 2007)**	**(Zhdanov 2008)**	**(Vohradsky, Panek et al. 2010)**	**(Cuccato, Polynikis et al. 2011)**	**This model**
mRNA synthesis	Unregulated	Unregulated	Unregulated	Unregulated as well as regulated	Unregulated	Unregulated
miRNA synthesis	Unregulated	Unregulated	Regulated by protein	Unregulated	Unregulated	Unregulated as well as regulated by protein
mRNA miRNA binding	Reversible	Reversible	Irreversible	Irreversible (Not explicitly considered)	Irreversible	Reversible

Role of P-body	By considering mRNA in processed state	Not considered	Not considered	Not considered	Not considered	By considering complex which can return either mRNA or miRNA
Translation by complex	At the same rate as mRNA translation	At lower rate than mRNA translation	No translation	Not considered	Not considered	At lower rate than mRNA translation
Degradation of the complex	At a rate more than that of mRNA	At a rate more than that of mRNA	miRNA concentration dependent	miRNA concentration dependent saturating mRNA degradation	miRNA concentration dependent 4 different mechanisms	At a rate more than that of mRNA
Recovery of miRNA	Not considered	Considered	Not considered	Not considered	Not considered	Considered
Recovery of mRNA	Not considered	Not considered	Not considered	Not considered	Not considered	Considered

In this study a detailed model of miRNA regulation was developed to include several experimentally reported details of miRNA action. The model included four major steps, (1) reversible binding of miRNA and mRNA to form a complex, (2) different translational efficiency of free and miRNA bound mRNA (3) catalytic mode of miRNA regulation (4) selective (non-stoichiometric) return of mRNA from the complex. Steady state analysis was performed to explore the effect of miRNA regulation on the target protein level. From the analysis of the model four dimensionless numbers were identified by grouping twelve reaction rate parameters that were sufficient to determine the regulatory effect of miRNA on target protein. The ranges of the dimensionless numbers were identified such that under those conditions miRNA regulation could result in increased target protein level. Majority of the experimental findings of increased target protein could also be explained in the framework of the model. Stochastic simulations

were performed to explore the effect of miRNA regulation on the steady state distribution of the target protein. It was observed that the nature of the steady state distribution of target protein remained unchanged in the presence of miRNA. In addition to the regulatory effect of miRNA, the four dimensionless numbers were observed to be sufficient to determine the relative steady state noise in the target protein level.

3.2.2 Methodology

3.2.2.1 Model development

The known biological interactions of mRNA and miRNA were represented as chemical reactions. The reaction system used in the model is summarized in Figure 3-1.

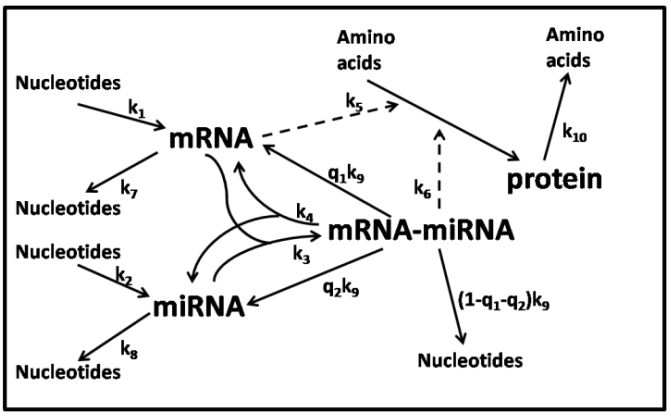


Figure 3-1 – Reaction system for miRNA mediated post-transcriptional regulation

Parameters k_1 to k_{10} represent specific rates for reactions. Parameters q_1 and q_2 are the probabilities of free mRNA and free miRNA return from the complex. Solid arrows represent conversion reactions and dashed arrows represent catalytic reactions.

Synthesis of mRNA (*m*) and miRNA (*mi*) was assumed to be zero order with rates k_1 and k_2 respectively. mRNA and miRNA were considered to bind reversibly to form mRNA-miRNA complex (*mmi*). k_3 and k_4 were specific rates of binding and dissociation. Both mRNA and the complex were considered to undergo translation to form protein (*p*) with different efficiencies, k_5 and k_6 respectively. The complex could selectively return mRNA or miRNA with probabilities q_1 and q_2 respectively. All the four components were assumed to undergo first order degradation with specific rates of degradation k_7, k_8,

k_9 and k_{10} for mRNA, miRNA, the complex and protein respectively. The reactions are summarized in Table 3-2.

Reactions of synthesis, degradation and conversion were considered to write mass balance of each of the four components. From the reaction system, the mass balance for the rate of change of the four components was written as,

$$\frac{d(m)}{dt} = k_1 - k_3 \times m \times mi + k_4 \times mmi - k_7 \times m + q_1 \times k_9 \times mmi \quad (3.1)$$

$$\frac{d(mi)}{dt} = k_2 - k_3 \times m \times mi + k_4 \times mmi - k_8 \times mi + q_2 \times k_9 \times mmi \quad (3.2)$$

$$\frac{d(mmi)}{dt} = k_3 \times m \times mi - k_4 \times mmi - k_9 \times mmi \quad (3.3)$$

$$\frac{d(p)}{dt} = k_5 \times m + k_6 \times mmi - k_{10} \times p \quad (3.4)$$

The parameter values were taken from a previously developed model for sRNA regulation by Shimoni, Friedlander et al (Shimoni, Friedlander et al. 2007), except for k_2, k_6, q_1, and q_2. The reaction rate constant for miRNA synthesis (k_2) was considered to be 5 times faster than that for mRNA synthesis rather than 50 times as considered for sRNA model. The specific rate for translation by mRNA-miRNA complex (k_6) was considered to be lower than translation by mRNA (k_5), but the effect of variation of this parameter was studied by varying the value in wide range including higher values. The values of parameters q_1 and q_2 were assumed such that $q_1 + q_2 \leq 1$. The reference parameter values are listed in Table 3-2.

Table 3-2 – Reactions, corresponding rate expression and parameter values for the mathematical model of miRNA mediated regulation

Reaction		**Rate equation**	**Parameter Value**
mRNA synthesis	$\varphi \rightarrow m$	k_1	k_1 = 0.02 (molecule/sec)
miRNA synthesis	$\varphi \rightarrow mi$	k_2	k_2 = 0.1 (molecule/sec)

mRNA-miRNA binding	*m + mi → mmi*	k_3 ×m× mi	k_3 = 1 (molecule^{-1}sec^{-1})
complex dissociation	*mmi → m + mi*	k_4 × mmi	k_4 = 0.02 (sec^{-1})
Translation by mRNA	*φ → p*	k_5 × m	k_5 = 0.01 (sec^{-1})
Translation by the complex	*φ → p*	k_6 × mmi	k_6 = 0.001 (sec^{-1})
mRNA degradation	*m → φ*	k_7 × m	k_7 = 0.002 (sec^{-1})
miRNA degradation	*mi → φ*	k_8 × mi	k_8 = 0.0025 (sec^{-1})
Complex degradation	*mmi → φ*	k_9 × mmi	k_9 = 0.002 (sec^{-1})
Protein degradation	*p → φ*	k_{10} × p	k_{10} = 0.001 (sec^{-1})
mRNA returning to cytoplasm	*mmi → m*	q_1 ×k_9 × mmi	q_1= 0.2
miRNA returning to cytoplasm	*mmi → mi*	q_2×k_9 × mmi	q_2 = 0.4

3.2.2.2 Steady state analysis

The equations (equations 3.1-3.4) were non-dimensionalized using $1/k_7$ as reference time and k_1/k_7 as reference concentration. Parameters κ_i are corresponding non-dimensional parameters listed in Table 3-3.

Table 3-3 – Non-dimensionalized parameters for the mathematical model of miRNA mediated regulation

Parameter with dimension	**Dimensionless parameter**
mRNA synthesis k_1 (molecule/sec)	$\kappa_1 = \frac{k_1}{k_1/k_7} \times \frac{1}{k_7} = 1$
miRNA synthesis k_2 (molecule/sec)	$\kappa_2 = \frac{k_2}{k_1/k_7} \times \frac{1}{k_7}$
mRNA-miRNA binding k_3 (molecule^{-1}sec^{-1})	$\kappa_3 = k_3 \times \frac{k_1}{k_7} \times \frac{1}{k_7}$

complex dissociation k_4 (sec^{-1})	$\kappa_4 = \frac{k_4}{k_7}$
Translation by mRNA k_5 (sec^{-1})	$\kappa_5 = \frac{k_5}{k_7}$
Translation by the complex k_6 (sec^{-1})	$\kappa_6 = \frac{k_6}{k_7}$
mRNA degradation k_7 (sec^{-1})	$\kappa_7 = \frac{k_7}{k_7} = 1$
miRNA degradation k_8 (sec^{-1})	$\kappa_8 = \frac{k_8}{k_7}$
Complex degradation k_9 (sec^{-1})	$\kappa_9 = \frac{k_9}{k_7}$
Protein degradation k_{10} (sec^{-1})	$\kappa_{10} = \frac{k_{10}}{k_7}$

At steady state as the rate of change of component level with respect to time is zero, the four equations (equations 3.1 to 3.4) were equated to zero. The expressions for steady state level of the four components were obtained by solving the set of non-linear equations. The steady state analysis was performed using Mathematica version 7.0.1.0 (Wolfram Research, Champaign, USA). Through the steady state analysis four dimensionless numbers (*a*, *b*, *c*, and *d*) were obtained by grouping twelve reaction rate parameters. The ratio (*r*) of the steady state level of protein in the presence of miRNA to that in the absence of miRNA was obtained as a measure of regulatory effect of miRNA. The protein ratio values were calculated for ranges of the dimensionless numbers *a*, *b*, *c*, and *d* using Matlab version 7.6.0.324 (The Mathworks, Natick, USA). All the dimensionless numbers were varied in a wide range to include both higher and lower values than the typical reported physiological value.

3.2.2.3 Stochastic simulations

To examine the effect of presence of miRNA on the steady state distribution of target protein exact stochastic simulations were performed using Gillespie algorithm (Gillespie 1976). For the simulations reaction rate parameters were varied in combinations of k_3-k_8 or k_1-k_2-k_3 or k_1-k_2-k_8 or k_5-k_6 in order to keep the values of the dimensionless number constant which in turn results in constant protein ratio (*r*). One such set contains 23 parameter combinations. Three to four such sets were simulated for each *r* value by varying dimensionless numbers and hence the reaction rate parameters, ensuring that each of the dimensionless number was varied. The details of the parameter combinations are given in Appendix IIa. Simulations were performed for 5 different ratio values for final time of 50000 seconds. Initial conditions were set to the corresponding deterministic steady state for each of the parameter combinations. Sample of 100000 simulation runs was generated to obtain statistics. Coefficient of variation (CV), referred here as noise, defined as ratio of standard deviation to mean of the population (σ/μ) was calculated for each parameter combination.

3.2.3 Results and discussion

3.2.3.1 Regulatory effect of miRNA can be determined using four dimensionless numbers

At steady state the concentration of components remains constant. Thus each of the equations 3.1 – 3.4 was equated to zero and the set of non-linear equations was solved analytically to obtain the expressions for steady state levels of the four components.

The steady state expression in terms of dimensionless parameters for free mRNA in the presence of miRNA is given as,

$$m_{ss-mir} = \frac{2}{1+\chi+\sqrt{1+\chi^2+2\delta}} \tag{3.5}$$

Where, $\chi = \frac{\kappa_3\kappa_9(\kappa_2(1-q_1)-1(1-q_2))}{\kappa_8(\kappa_4+\kappa_9)}$ and $\delta = \frac{\kappa_3\kappa_9(\kappa_2(1-q_1)+1(1-q_2))}{\kappa_8(\kappa_4+\kappa_9)}$ (3.6)

The reaction rate parameters were arranged into three dimensionless numbers. These numbers are given as,

$$a = \kappa_2(1-q_1)\,,\ b = (1-q_2)\,,\ c = \frac{\kappa_3\kappa_9}{\kappa_8(\kappa_4+\kappa_9)} \tag{3.7}$$

From the model equation (3.3), the steady state expression for *mmi* was obtained,

$$mmi_{ss-mir} = \left(\frac{\kappa_3}{\kappa_4+\kappa_9}\right)\times m_{ss-mir}\times mi_{ss-mir} \tag{3.8}$$

Using the expression for steady state of *mmi* and model equation (3.2), the expression for steady state of *mi* is given as,

$$mi_{ss-mir} = \frac{\kappa_2}{\kappa_3\times m_{ss-mir} - \kappa_4\left(\frac{\kappa_3}{\kappa_4+\kappa_9}\right)\times m_{ss-mir} + \kappa_8 - q_2\kappa_9\left(\frac{\kappa_3}{\kappa_4+\kappa_9}\right)\times m_{ss-mir}} \tag{3.9}$$

Using the expressions (3.8) and (3.9) and model equation (3.4), the expression for steady state target protein in the presence of miRNA is given as,

$$p_{ss-mir} = \frac{\kappa_5\times m_{ss-mir} + \kappa_6\left(\frac{\kappa_3}{\kappa_4+\kappa_9}\right)\times m_{ss-mir}\times mi_{ss-mir}}{\kappa_{10}} \tag{3.10}$$

The expression was simplified to obtain the steady state expression for protein in terms of reaction rate parameters,

$$P_{ss-mir} = \frac{2\kappa_5 + \dfrac{2\kappa_2\kappa_3\kappa_6}{\kappa_8(\kappa_4+\kappa_9) + \dfrac{2\kappa_3\kappa_9(1-q_2)}{1+\gamma+\chi}}}{\kappa_{10}(1+\gamma+\chi)} \tag{3.11}$$

Where, $\gamma = \sqrt{1 + \dfrac{{\kappa_3}^2{\kappa_9}^2(\kappa_2 - 1 - q_1\kappa_2 + q_2)^2}{{\kappa_8}^2(\kappa_4+\kappa_9)^2} + \dfrac{2\kappa_3\kappa_9(\kappa_2+1-q_1\kappa_2-q_2)}{\kappa_8(\kappa_4+\kappa_9)}}$

The parameters γ and χ can be expressed in terms of the three dimensionless numbers as,

$$\gamma = \sqrt{1 + c^2(a-b)^2 + 2c(a+b)} \text{ and, } \chi = c(a-b) \tag{3.12}$$

The expression (3.11) was simplified using the dimensionless numbers and rearranged to give the simplified expression for steady state level of protein in the presence of miRNA,

$$P_{ss-mir} = \frac{2\kappa_5}{\kappa_{10}(1+\gamma+\chi)} + \frac{2\ c\ \kappa_2\kappa_6}{\kappa_9\kappa_{10}(2bc+1+\gamma+\chi)} \tag{3.13}$$

In the absence of miRNA, the steady state protein level would depend upon the synthesis, degradation of protein and the steady state of mRNA. Therefore, the protein level in the absence of miRNA can be given as, $\frac{k_5}{k_{10}}.\frac{k_1}{k_7}$, where the term k_1/k_7 represents the steady state level of mRNA. In terms of dimensionless parameters the steady state level of protein in the absence of miRNA is given as,

$$P_{ss} = \frac{\kappa_5}{\kappa_{10}} \tag{3.14}$$

The regulatory action of miRNA was defined as the ratio (*r*) of steady state level of protein in the presence of miRNA to that in the absence of miRNA. Therefore, from expression (3.13) and (3.14) the ratio is given as,

$$r = \frac{P_{ss-mir}}{P_{ss}} = \frac{2}{1+\gamma+\chi} + \frac{2c \times \kappa_2\kappa_6}{\kappa_9\kappa_5(2bc+1+\gamma+\chi)} \tag{3.15}$$

The fourth dimensionless number was defined as,

$$d = \frac{\kappa_2\kappa_6}{\kappa_9\kappa_5} \tag{3.16}$$

Thus the 12 reaction rate parameters were grouped into four dimensionless numbers *a*, *b*, *c*, and *d*. Therefore, the ratio was expressed in terms of these four dimensionless number as,

$$r = \frac{P_{ss-mir}}{P_{ss}} = \frac{2}{1+\gamma+\chi} + \frac{2cd}{2bc+1+\gamma+\chi} \quad (3.17)$$

From the expression for ratio (equation 3.17) it was evident that only the values of these four dimensionless numbers and not the individual reaction rate parameters were important in determining the relative change in the steady state level of target protein. Therefore, the four dimensionless numbers were found to be sufficient to determine the regulatory action of miRNA regulation. From equations (3.7) and (3.16), it was clear that different combinations of reaction rate parameters can result in the same values of the dimensionless numbers, and therefore lead to the same value of the ratio. Change in multiple reaction rate parameters thus may not always lead to change in the observed regulatory effect. A change in a parameter can compensate for change in other parameter resulting in the same protein ratio value. Additionally, similar regulatory effect can be obtained by controlling different reactions. The effect of simultaneous change in multiple parameters values can thus be predicted using the dimensionless numbers.

All the four dimensionless numbers could not be explained in terms of the ratios of simple physical processes. The dimensionless number *a* can be interpreted as a measure of relative synthesis rate of miRNA and mRNA. *b* can be considered as a fraction of miRNA in P-body that is not returned. The dimensionless number *d* can be interpreted as a combined measure of relative translational efficiency and relative stability of the complex and free mRNA.

3.2.3.2 Identification of conditions for 'unexpected' activating effects of miRNA

From equation 3.17, it was clear that value of *r* lesser than 1 would indicate lesser level of target protein in the presence of miRNA than that observed in the absence of miRNA. Therefore, $r < 1$ indicated the commonly observed repressive effect of miRNA. The value of *r* equal to one would indicate equal level of target protein level, suggesting no effect of miRNA. The condition $r > 1$ indicated higher level of target protein in the presence of miRNA, which has been reported in some recent experimental studies. Therefore, the expression for the ratio of steady state protein levels (equation 3.17) was analysed for different limiting cases of the dimensionless numbers to find out conditions where $r > 1$,

i.e. the steady state level of protein in the presence of miRNA was higher than that in the absence of miRNA. These conditions are summarized in Table 3-4.

1. $c = 0$, $k_9 \neq 0$

The value of c can be zero when k_3 is zero and/or k_8 or k_4 have very large value. Under these conditions, from equations (3.12) and (3.17), the ratio r reduced to 1 indicating equal steady state level of protein even in the presence of miRNA.

2. $c = 0$, $k_9 = 0$

Under the condition, d is undefined. Therefore the expression for steady state protein level in the presence of miRNA was obtained by using equation (3.11).

$$P_{ss-mir}\,|_{\kappa_9=0} = \frac{\kappa_5}{\kappa_{10}} + \frac{\kappa_2\kappa_3\kappa_6}{\kappa_8\kappa_4\kappa_{10}} \tag{3.18}$$

Using the above expression (equation 3.18) and equation (3.14) the expression for ratio was calculated.

$$r\,|_{\kappa_9=0} = 1 + \frac{\kappa_2\kappa_3\kappa_6}{\kappa_8\kappa_4\kappa_5} \tag{3.19}$$

The expression indicated that under the condition of $k_9 = 0$, the protein level in the presence of miRNA is higher than that in the absence of miRNA.

3. $a = 0$

This case would be valid at $k_2 \sim 0$ or $q_1 \sim 1$ which means very low miRNA synthesis or near complete return of mRNA. The expression for the ratio is given as

$$r\,|_{a=0} = 1 + \frac{dc}{1+c} \tag{3.20}$$

Under this condition the protein level in the presence of miRNA would be more than that in the absence of miRNA but it would reach to a certain maximum value (1+*d*).

4. $b = 0$

The ratio of proteins is given as

$$r|_{b=0} = \left(\frac{1}{1+ac}\right)(1+dc)$$

The above condition would be valid when $q_2 \sim 1$, implying complete return of miRNA. In this condition, the steady state protein level in the presence of miRNA can be lower, equal or higher than that in the absence of miRNA depending upon the values of k_2 and *c*.

Table 3-4 – Limiting conditions of dimensionless numbers resulting in lower, equal or higher target protein level

***c* = 0**		***a* = 0 (*b* = 1)**	***b* = 0 (*a* = k_2)**
$k_9 \neq 0$	**$k_9 = 0$**		
$r = 1$ Equal protein	$r = 1 + \frac{\kappa_2 \kappa_3 \kappa_6}{\kappa_8 \kappa_4 \kappa_5}$ Equal or higher protein	$r = 1 + \frac{dc}{1+c}$ Equal or higher protein	$r = \left(\frac{1}{1+ac}\right)(1+dc)$ Lower, equal or higher protein

In addition to the limiting conditions, the effect of variation of the values of the dimensionless numbers on the steady state of target protein was examined. Figure 3-2 shows $\log_{10}(r)$ value as a function of the dimensionless numbers.

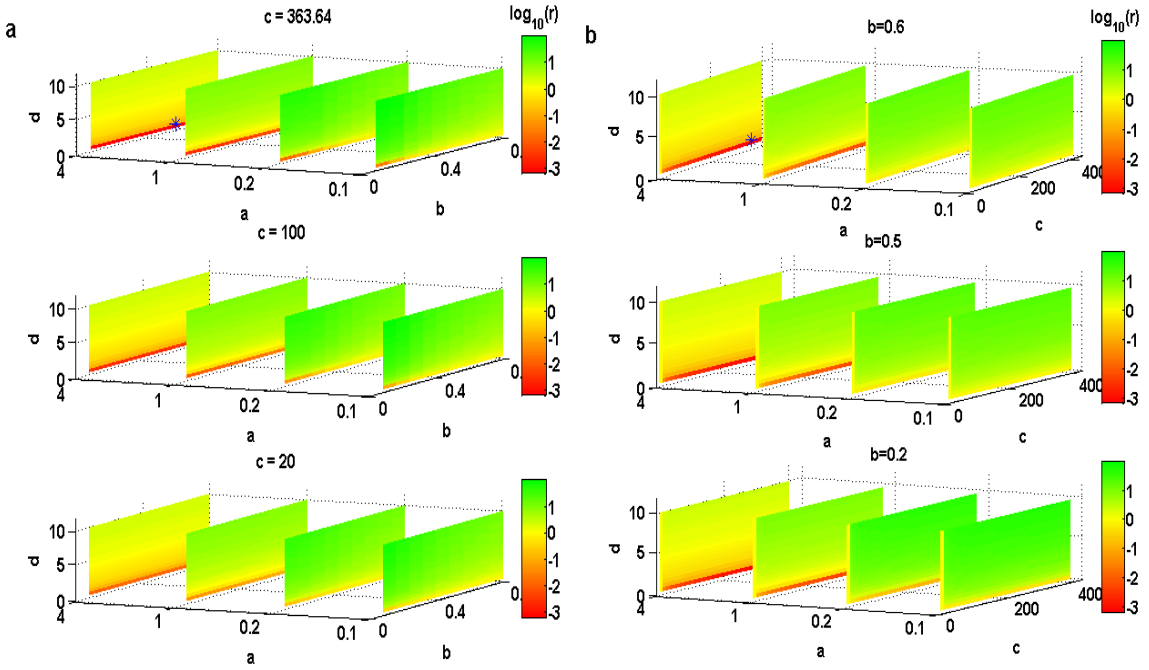


Figure 3-2 – Regulatory effect of miRNA as a function of four dimensionless numbers

(a) $\log_{10}(r)$ is plotted as a function of *a*, *b*, and *d* at three different values of *c*. (b) $\log_{10}(r)$ is plotted as a function of *a*, *c*, and *d* at three different values of *b*. Physiological normal condition (*a* = 4, *b* = 0.6, *c* = 363.64, and *d* = 0.5) are indicated by asterisks (*).

The red region in the Figure 3-2 indicated a region where the ratio was lesser than 1. Yellow region indicated the parameter region where miRNA had no effect and the green region indicated parameter region, where the level of target protein is higher in the presence of miRNA. It was observed that the ratio was sensitive to variation in *d* over a wide range. The ratio was sensitive to variation in *a* and *c* in a low value range. It was not observed to be sensitive to variation in *b*. The values of the four dimensionless numbers *a*, *b*, *c*, and *d* for the typical set of parameter values (Table 3-2) were 4, 0.6, 363.54, and 0.5 respectively. It was observed that for these values, the ratio was less than 1, consistent with the commonly observed repressive effect of miRNA on a target protein. From Figure 3-2, values of dimensionless numbers can be obtained that can result in a particular value of protein ratio. An iso-surface can be drawn which will predict the combinations of the dimensionless numbers corresponding to the selected ratio value.

For the conditions where the steady state level of target protein remained same or increased in the presence of miRNA, the dynamics of the systems was studied. For instance, under the condition, k_9 equal to zero, the steady state protein level remained same in the presence of miRNA. Time course simulation was performed to find out the effect of miRNA mediated regulation on the dynamics of target protein. Figure 3-3 shows the time course profile of target protein in the presence and absence of miRNA. It was observed that, as predicted the same steady state was reached however the steady state was reached much later in the presence of miRNA (Figure 3-3). Such effect on dynamics of target protein could be important in the cellular context.

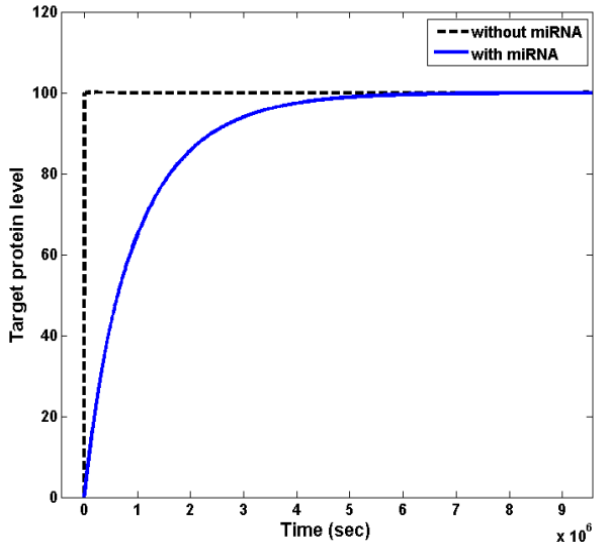


Figure 3-3 – Effect of miRNA regulation on the transient of target protein resulting in the same steady state level

Parameters – $k_9 = 0$, $k_6 = 0$, other parameter values as stated in Table 3.2; Initial condition – All components at zero level.

Experimental observations of increased target protein level were analysed in the framework of the model. miRNA are known to bind to the 3' UTR of the target mRNA. The 3' UTR contains AU rich elements (ARE), which are generally known to be signals for mRNA degradation (Vasudevan, Tong et al. 2008; von Roretz and Gallouzi 2008). In certain cases miRNA can bind near to the ARE thereby masking the degradation signal in

turn protecting target mRNA from degradation (Vasudevan, Tong et al. 2008). In a study by Ma, Liu et al, it was observed that miR-4661 binds to ARE of IL-10 mRNA and prevents binding of RNA binding proteins required for degradation, resulting in protection of IL-10 mRNA (Ma, Liu et al. 2010). In case of the model, it can be interpreted as a decrease in the reaction rate parameter for *mmi* degradation, k_9. The limiting case, k_9 equal to zero was shown to be sufficient for equal or higher target protein level. Decrease in k_9 results in increase in *d* and decrease in *c*, both of which have been shown in the model to lead to increase in the target protein level. Such information about the target site in 3'UTR and the regulatory effect of miRNA is useful for bioinformatics studies related to miRNA target site prediction to predict the potential effect of miRNA having a particular target site.

Increased translational efficiency was another suggested mechanism of miRNA mediated up-regulation (Jopling, Yi et al. 2005; Ghosh, Soni et al. 2008; Orom, Nielsen et al. 2008; Jangra, Yi et al. 2010). In the model, this can be represented as increased translation by complex, k_6. Increase in k_6 leads to increase in *d* and hence the target protein concentration increases. In a study by Mortensen et al (Mortensen, Serra et al. 2011), activation of translation was observed in response to regulation by xlmiR16. The study also reported that the level of total mRNA remained unchanged. The authors stated that the probable mechanisms, such as increased mRNA stability or increased translation, could not be determined. The observed condition was analysed in the framework of the model.

The condition of equal total mRNA both in the presence and absence of miRNA can be represented as,

$$m_{ss} = m_{ss-mir} + mmi_{ss-mir} \tag{3.21}$$

Protein ratio is,

$$r = \frac{P_{ss-mir}}{P_{ss}} = \frac{\kappa_5 \times m_{ss-mir} + \kappa_6 \times mmi_{ss-mir}}{\kappa_5 \times m_{ss}} \tag{3.22}$$

Substituting the value of m_{ss} from equation (3.21) in equation (3.22),

$$r = \frac{P_{ss-mir}}{P_{ss}} = \frac{\kappa_5 \times m_{ss-mir} + \kappa_6 \times mmi_{ss-mir}}{\kappa_5 \times m_{ss-mir} + \kappa_5 \times mmi_{ss-mir}} \quad (3.23)$$

Therefore, to satisfy the condition of translational activation with same total mRNA level, k_6 must be greater than k_5. Thus from the analysis, it was identified that increased translation by the complex, and not increased mRNA stability can justify the observation of increased target protein level.

Thus, a majority of the experimental observations of increased target protein level were explained in terms of the model framework. The model developed in this study contained many of the known biological details of miRNA regulation such as different translational efficiency of *mmi* complex, selective return of mRNA, that are not considered in the previous mathematical models. However, there are certain limitations to the model. Though selective return from P-body was considered, a separate P-body compartment was not considered in the model. Additionally, separate intra P-body reactions of mRNA, miRNA and the complex were not included due to lack of knowledge about their interactions and state in the P-body. The rate of mRNA or miRNA return was linked to the rate of complex degradation in the model. However, this relationship did not allow return of either component in the absence of complex degradation. Some modifications are possible to improve the model.

3.2.3.3 Relative noise can also be determined using four dimensionless numbers

To examine the effect of miRNA mediated regulation on the steady state distribution of target protein stochastic simulations were performed. In the stochastic simulations, reaction rate parameters were varied by keeping the value of dimensionless numbers the same, as stated in the Methodology section. It ensured the same ratio (*r*) of protein levels. It was observed that the mean of the stochastic simulations remain unchanged for the same set of dimensionless numbers, irrespective of the values of reaction rate parameters. Interestingly, the relative noise was also observed to be unchanged for a particular set of dimensionless numbers, indicating that the four dimensionless numbers could determine the relative noise in the steady state protein level. Additionally, it was observed that the

parameter k_{10} did not appear in any dimensionless number. This indicated that changes in the stability of the protein will not affect either the relative mean or the relative noise level of the steady state target protein distribution.

The mean of the means and the noise values calculated from the simulations is shown in Figure 3-4 as a graph of $\log_{10}$(noise ratio) vs. $\log_{10}$(protein ratio). A linear relationship of noise ratio (*nr*) and protein ratio (*r*) was observed. The relative noise was observed to decrease with increase in the protein ratio. The relation between the noise ratio (*nr*) and the ratio (*r*) can be given as $nr = r^{-0.2}$ with correlation coefficient of 0.92. Such a result is expected from one-parameter distribution like Poisson distribution. However, in case of Poisson distribution, the log plot of the data would show a linear relationship with slope -0.5, as shown by dashed line in Figure 3-4. From the graph it was clear that the steady state distribution is not Poisson, though the exact nature could not be identified.

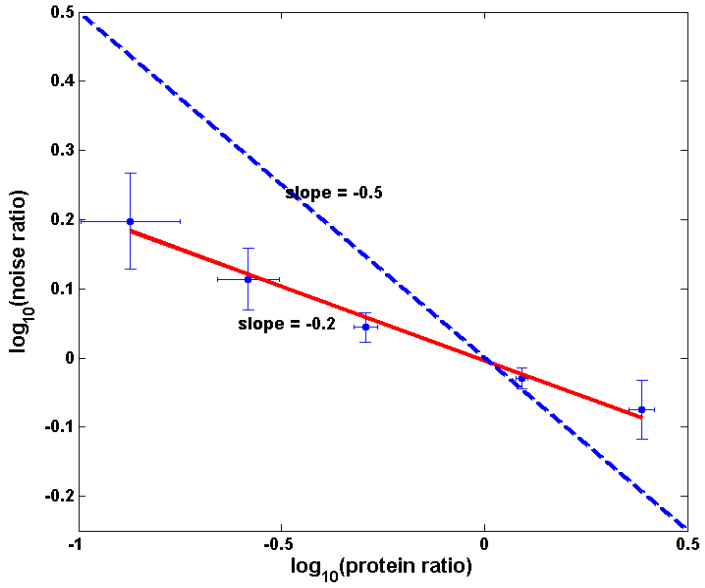


Figure 3-4 – Variation of relative noise with relative mean

The error bars show standard deviation. The solid red line shows a linear fit with slope -0.2. The dashed blue line has slope -0.5, an expected fit for a Poisson distribution.

To identify whether miRNA regulation changed the nature of the steady state distribution of target protein, the steady state distributions of target protein in the presence and absence of miRNA were qualitatively compared using Quantile-quantile plot (Q-Q plot).

Figure 3-5 shows a Q-Q plot of steady state distribution of target protein in the absence of miRNA vs. that in the presence of miRNA. A comparison of the target protein distribution with and without miRNA regulation showed that the nature of the distribution remained unchanged.

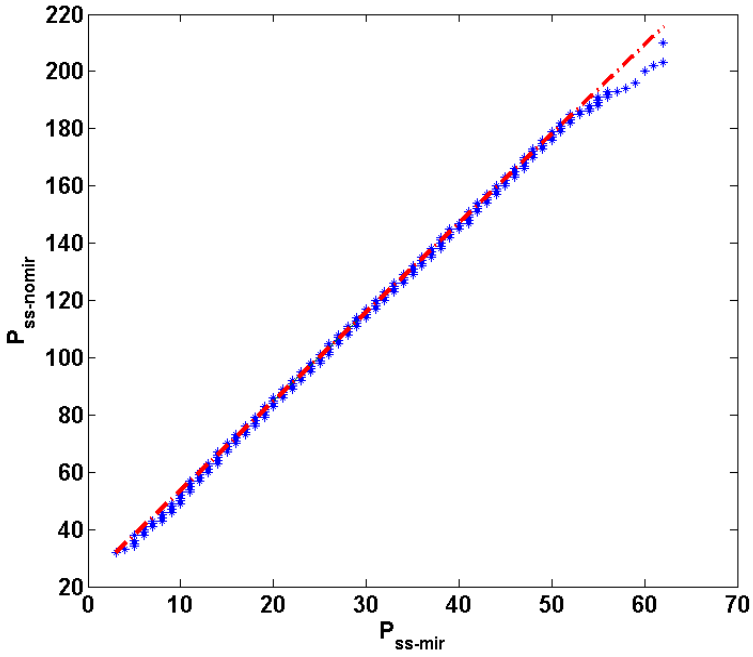


Figure 3-5 – Q-Q plot of the steady state protein distribution in the absence of miRNA vs. that in the presence of miRNA

3.2.4 Conclusion

The study illustrated development of detailed model of miRNA regulation. The previously developed mathematical models were extended to include known details of regulation such as different translational efficiencies of free and bound mRNA, selective return of mRNA or miRNA. Using the steady state analysis, quantitative criteria for regulatory effect of miRNA were identified in terms of four dimensionless numbers obtained from twelve reaction rate parameters, thus showing the utility of dimensionless numbers in the analysis of biological systems. It is possible to obtain another set of dimensionless numbers that can lead to same or better insights in the process. Using the developed model, the unexpected experimental observations of increased target protein level could be explained just on the basis of the known facts about the regulation without

incorporating additional regulatory processes. The effect of miRNA regulation on the steady state distribution of target protein was explored. The nature of distribution of target protein was observed to remain unchanged in the presence of miRNA. It was observed that the four dimensionless numbers were sufficient to determine the relative noise level in the target protein. Additionally, protein stability was not observed to affect either the relative steady state or the relative noise.

Using the model certain falsifiable predictions about the exact mechanism for observed experimental increased protein level, the effect of protein degradation rate on the relative noise level, and steady state distribution of target protein were made. These predictions can be experimentally tested which will lead to better understanding of one of the important mode of regulation.

3.3 Section 2 – Development of a simple method to incorporate intronic miRNA mediated regulatory effects into existing mathematical models

3.3.1 Introduction

In addition to the effects on the steady state level, the effects on the dynamics of target protein are important in all cellular processes. The discovery of the first miRNA *lin-4* was from the observation of *C. elegans* mutants that failed to show the ability to control the timing of specific post-embryonic developmental events (Lee, Rhonda et al. 1993). miRNA are now known as crucial regulatory players in majority of processes such as differentiation (Carrington and Ambros 2003), development of organisms (Alvarez-Garcia and Miska 2005), and cell cycle (Norbury and Nurse 1992; Vasudevan, Tong et al. 2008), where temporal regulation of reaction network is necessary. For the processes where the molecular circuits have to be sequentially activated and deactivated, the regulation also needs to be dynamic. One class of miRNA called intronic miRNA, has a unique positional advantage for dynamic regulation. It has been experimentally shown that many of the intronic miRNA are expressed along with their host genes (Rodriguez, Griffiths-Jones et al. 2004; Baskerville and Bartel 2005; Liu, Papagiannakopoulos et al. 2007). The synthesis of intronic miRNA is synchronized with that of the host gene. Therefore, the expression of the target protein can be fine tuned by simultaneously

turning down the expression of antagonistically functioning genes (Ying and Lin 2004; Barik 2008) as the host mRNA synthesis begins. Hence it is important to study intronic miRNA mediated regulatory effects on the dynamics of cellular processes.

As described in the section 1, there are several mathematical models that study the effect of miRNA mediated regulation on target protein dynamics (Khanin and Vinciotti 2008; Zhdanov 2009). These models are generic models of miRNA regulation that are not specific for particular process, additionally these models do not specifically focus on intronic miRNA. In addition to the generic model there are some models that study the effect of miRNA in specific cellular processes (Xie, Yang et al. 2007; Aguda, Kim et al. 2008; Nandi, Vaz et al. 2009; Wei, Yan et al. 2011). These models extend the existing models for the particular process by adding a minimum of one component (miRNA) and one reaction to incorporate miRNA effect. However, such addition results in changes in the network connectivity and stoichiometry.

Knowing the co-expression of intronic miRNA with the host gene, it was hypothesized that the use of a host protein as a proxy for intronic miRNA might be an effective and simpler way to incorporate the dynamic regulatory effects. This method is easy to implement as the network connectivity remains unchanged. To test the efficacy of this method, regulatory effects of intronic miRNA were incorporated into a mathematical model of cell cycle reaction network. Cell cycle is one such process having several genes that are regulated by intronic miRNA or are host genes for intronic miRNA. Some examples of intronic miRNA active in cell cycle are listed in Table 3-5.

Table 3-5 – Examples of intronic miRNA active in cell cycle

miRNA	Target	System Studied	Reference
hsa-let-7f	RAS	In lung tumors	(Johnson, Grosshans et al. 2005)
hsa-let-7f	MYC	In lymphoma cells	(Chang, Yu et al. 2007)
hsa-miR-106b-25	E2F1-3	Impair TGF beta-dependent cell cycle arrest and apoptosis in gastric cancer	(Petrocca, Visone et al. 2008)

hsa-miR-106b	p21	In human mammary epithelial cells controls cell cycle progression	(Ivanovska, Ball et al. 2008)
hsa-miR-133	RHOA	controls cardiac hypertrophy	(Carè, Catalucci et al. 2007)
hsa-miR-214	PTEN	in human ovarian cancer	(Yang, Kong et al. 2008)
hsa-miR-93	E2F1-3	impair TGF beta-dependent cell cycle arrest and apoptosis in gastric cancer	(Diaz, Silva et al. 2008)

There are several mathematical models for cell cycle (Novak and Tyson 1997; Obeyesekere, Knudsen et al. 1997; Hamada, Tashima et al. 2009; Ferrell Jr, Tsai et al. 2011). These models consider the major cyclin-CDK reaction network in cell cycle progression and focus on protein based regulation such as regulation by transcription factors and post-translational regulation. Therefore, addition of RNA mediated regulation can improve the model performance. In this study the generic model of cell cycle developed by Tyson group (Novak and Tyson 2004) was modified to include regulatory effect by intronic miRNA hsa-miR-25 using host protein as proxy for miRNA, on two target proteins. It was observed that such incorporation improved that model performance as the predictions of the modified model were closer to the experimental observations.

To examine the universality of the method, miRNA mediated regulation was compared with analogous protein mediated post-transcriptional regulation. In addition to these regulatory motifs without feedback, the two types of regulation were compared in the presence of positive and negative feedback. It was observed that the target protein profiles were similar in majority of the conditions for the two types of regulation, indicating the generic nature of the method.

3.3.2 Methodology

3.3.2.1 Incorporation of hsa-miR-25 in cell cycle model

The model for cell cycle developed by Tyson group (Novak and Tyson 2004) was used to incorporate the effects of intronic miRNA mediated regulation. The model contained a protein component called TFB. The gene of which is a host gene for intronic miRNA hsa-

miR-25. The reaction network model also contained two protein targets of the intronic miRNA, viz., TFE and cdc14. The expression pattern of intronic miRNA hsa-miR-25 was reported to be similar to that of the host gene, MCM7 (Rodriguez, Griffiths-Jones et al. 2004; Liu, Papagiannakopoulos et al. 2007; Petrocca, Visone et al. 2008), referred to as TFB in the model. Therefore, the protein product of the host gene TFB was used as a proxy for intronic miRNA levels. The regulation was modeled as host protein concentration dependent reduction in the target protein production rate. The rate expression for target protein synthesis was modified by using a term similar to competitive inhibition kinetics such that the rate would depend upon host protein level implying dependency on intronic miRNA levels.

$$Rate\ with\ miRNA\ effect = \frac{Rate\ without\ miRNA\ effect}{1+\frac{[Protein_{Host}]}{km}} \tag{3.24}$$

Hence to represent the dynamic concentration dependent repressive action of the intronic miRNA, following multiplicative term was included in the rate expression.

$\frac{1}{1+\frac{[TFB]}{km}}$ where, *km* is a constant representing repressive action.

The rate expressions for synthesis of target proteins cdc14 and TFE in the basal model of cell cycle were modified to include the miRNA effect as follows.

For target protein cdc14,

$$Rate\ expression\ for\ cdc14\ synthesis\ with\ miRNA\ effect = \frac{Rate\ expression\ for\ cdc14\ used\ in\ Tyson\ model}{1+\frac{[TFB]}{km_1}} \tag{3.25}$$

For target protein TFE,

$$\text{Rate expression for TFE synthesis with miRNA effect} = \frac{\text{Rate expression for TFE used in Tyson model}}{1+\frac{[TFB]}{km_2}} \quad (3.26)$$

For large values of km_1 compared to TFB, the expression for rate of cdc14 synthesis (equation 3.25) reduces to the form used in original Tyson model. Similarly for large values of km_2 (km_2 >> TFB), expression (3.26) reduces to the one used in Tyson model.

The response of the system to the incorporated inhibitory effects of miRNA was evaluated by performing the time course simulations for varying values of km_1 and km_2 in range of 10^{-3} to 10^{3}. All the other parameter values and initial conditions were used as given in the Tyson model. The time course profiles of the components in the original and modified model were compared to study the effects of miRNA regulation. Since quantitative data was available only for cyclins, the predictions of both the original and modified models were compared with experimental observations for cyclin A, B and E. For other components the predictions of the modified model and the unmodified model were compared only to each other.

3.3.2.2 Model for incorporation of intronic miRNA regulation through host protein

To compare the effect of miRNA mediated regulation and protein mediated regulation on the target protein dynamics, a miRNA system identical to that considered for sRNA (Shimoni, Friedlander et al. 2007) was considered. The model contained four components viz., mRNA (*m*), miRNA (*mi*), mRNA-miRNA complex (*mmi*), and protein (*p*). The model includes zero order synthesis of mRNA and miRNA with specific rates of synthesis k_1 and k_2 respectively. Binding of mRNA and miRNA was considered to be reversible with k_3 as specific rate of binding and k_4 as specific rate of dissociation. Protein synthesis (k_5) was considered to be first order with the rate proportional to free mRNA. Degradation of all the four components was considered to be first order with specific rates of degradation k_7, k_8, k_9 and k_{10} for mRNA, miRNA, mRNA-miRNA complex, and protein, respectively. Considering synthesis, degradation, and conversion reactions, the differential equations for the four components are,

$$\frac{d(m)}{dt} = k_1 - k_3 \times m \times mi + k_4 \times mmi - k_7 \times m \tag{3.27}$$

$$\frac{d(mi)}{dt} = k_2 - k_3 \times m \times mi + k_4 \times mmi - k_8 \times mi \tag{3.28}$$

$$\frac{d(mmi)}{dt} = k_3 \times m \times mi - k_4 \times mmi - k_9 \times mmi \tag{3.29}$$

$$\frac{d(p)}{dt} = k_5 \times m - k_{10} \times p \tag{3.30}$$

Parameter values same as that given in the Shimoni, Friedlander et al model; were used for simulations. The differential equations were solved numerically to obtain the target protein profiles.

To compare the effect of intronic miRNA mediated regulation and host protein mediated regulation on the target protein dynamics in a generic reaction network, two reaction systems were formulated. Figure 3-6 shows the intronic miRNA mediated regulation (Figure 3-6a) and host protein mediated regulation (Figure 3-6b).

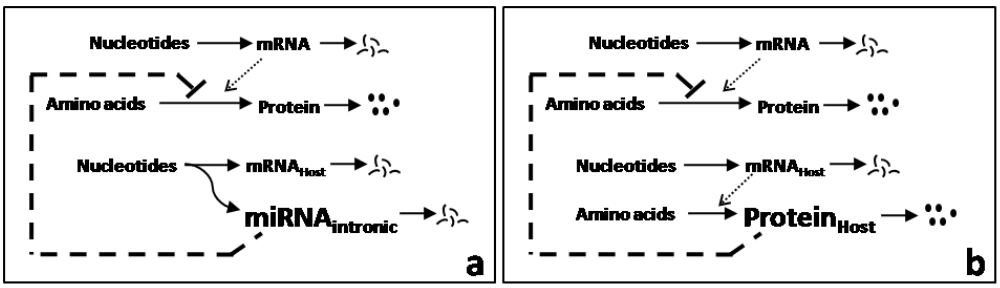


Figure 3-6 – Intronic miRNA and host protein mediated regulatory motifs

(a) Unregulated intronic miRNA mediated regulation (b) and host protein mediated inhibition. Inhibitory action of miRNA is modeled by equations 3.27-3.30 and corresponding protein mediated inhibition is modeled by equation 3.31.

In case of intronic miRNA mediated regulation the value of specific rate of miRNA synthesis was considered to be the same as that of the host mRNA synthesis i.e. $k_1 = k_2$.

To incorporate host protein mediated inhibition, a terms similar to competitive inhibition was used in the rate expression for target protein.

$$Rate\ of\ Protein\ synthesis = \frac{k\_syntheis \times [mRNA_{Target}]}{1 + \frac{[Protein_{Host}]}{k_i}} \tag{3.31}$$

Where, k_i is a constant representing the repressive action of the host protein. In this case the value of k_i was set in order to get almost the same steady state value of target protein as that obtained by miRNA regulation. Starting with different initial conditions, dynamics of the target protein in both the systems was compared.

3.3.2.3 Model for miRNA mediated auto-regulatory loop

In addition to comparison of simple miRNA and protein mediated regulation, these two types of regulatory modes were compared in the presence of positive and negative feedback loops (Figure 3-7). In this case, regulation by a general miRNA and analogous post-transcriptional regulation by a protein were compared. Figure 3-7a shows miRNA mediated positive and negative feedback regulation motifs. Target protein negatively regulating miRNA indicates miRNA mediated positive feedback regulation while, target protein positively regulating miRNA indicated negative feedback regulation. Figure 3-7b shows protein mediated positive and negative feedback regulation.

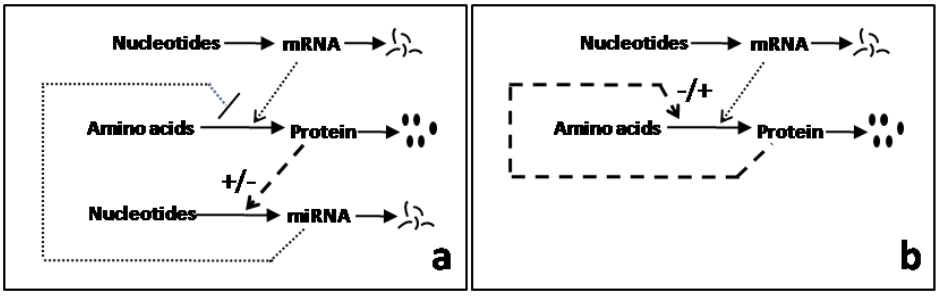


Figure 3-7 – miRNA and protein mediated regulation in the presence of feedback regulation

Comparison of (a) miRNA mediated (b) and protein mediated regulation as components of negative and positive feedback networks. Target protein effect on miRNA synthesis in (a) is modeled using equations 3.32 and 3.33; and on its own synthesis in (b) using equations 3.34 and 3.35.

To represent the regulation by target protein, the rate expression for miRNA synthesis reaction was modified by incorporating target protein concentration dependent synthesis term. For negative auto-regulation where the target protein positively regulates miRNA synthesis the rate expression was modified as,

$$Rate\ of\ miRNA\ synthesis = k_2 + \left(k_{2\_\max} - k_2\right)\frac{[p]}{kp + [p]} \tag{3.32}$$

In this case the minimum (basal) rate of miRNA synthesis is k_2 and the maximum rate at very large value of *p* is k_{2_max}.

In case of positive auto-regulation, where the target protein negatively regulates miRNA synthesis the rate expression was modified as,

$$Rate\ of\ miRNA\ synthesis = k_{2\_min} + \frac{\left(k_2 - k_{2\_\min}\right)}{1 + \frac{[p]}{kp}} \tag{3.33}$$

Therefore, in this case the maximum (basal) rate is k_2 and the minimum rate at very large values of p is k_{2_min}.

3.3.2.4 Model for protein mediated auto-regulatory loop

miRNA mediated regulation was compared to analogous protein mediated post-transcriptional regulation. For protein mediated negative feedback loop (Figure 3.7b), the rate expression for protein synthesis was represented as,

$$Rate\ of\ protein\ synthesis = \left(k_{5\_\min} + \frac{\left(k_5 - k_{5\_\min}\right)kp}{kp + [p]}\right)[m] \tag{3.34}$$

In the presence of high protein level, k_{5_min}**m* was the minimum rate at which protein synthesis would take place. In the presence of very low protein level, protein synthesis would take place at basal rate k_5**m*.

For protein mediated positive feedback loop the rate expression for protein synthesis was,

$$Rate\ of\ protein\ synthesis = \left(k_5 + \left(k_{5\_max} - k_5 \right) \frac{[p]}{kp + [p]} \right) [m] \quad (3.35)$$

The maximum rate of protein synthesis is k_{5_max}*m.

The maximum reaction rate constant in case of positive feedback (k_{2_max}-k_2) and (k_{5_max}-k_5) was set to five times the value of k_2 and k_5 respectively. The minimum reaction rate constant (k_{2_min} and k_{5_min}) was set to 1/5 times the values of k_2 and k_5 respectively. In each case the *kp* values were set in order to get 2 fold up or down regulation of steady state level of target protein. 2 fold up and down regulation was also tested for 10 fold higher and lower reaction rate constants.

The two types of regulations were compared qualitatively by observing the target protein profiles. Quantitative comparison was done by measuring the rise time, defined as the time required by target protein to first reach 90% of the steady state value.

Time course simulations were performed using stiff ode solver, `ode15s`, of Matlab version 7.6.0.324 (The Mathworks, Natick, USA) and SBtoolbox2 of Matlab.

3.3.3 Results and Discussion

3.3.3.1 Incorporation of concentration dependent inhibitory effects of intronic miRNA improved the model performance

In addition to the protein mediated regulation, studies have revealed importance of miRNA mediated regulation in cellular systems. Therefore, it is necessary to modify the existing mathematical models of cellular processes to incorporate miRNA regulation to understand its effect on dynamics of cellular processes. However, incorporation of such additional regulation would imply reformulation of existing reaction network to include new components (regulator) and reaction (regulatory effect) which in turn leads to changes in stoichiometry and rates. Therefore a simpler method was developed to incorporate intronic miRNA mediated dynamic regulatory effects in cell cycle model. The widely used Tyson model of cell cycle was used to incorporate regulatory effects of intronic miRNA hsa-miR-25 on two targets TFE and cdc14. Figure 3-8 shows the reaction system used in the model. In addition to the reactions used in the original model,

the regulation by hsa-miR-25 on the two target proteins TFE and cdc14 is shown in the reaction network.

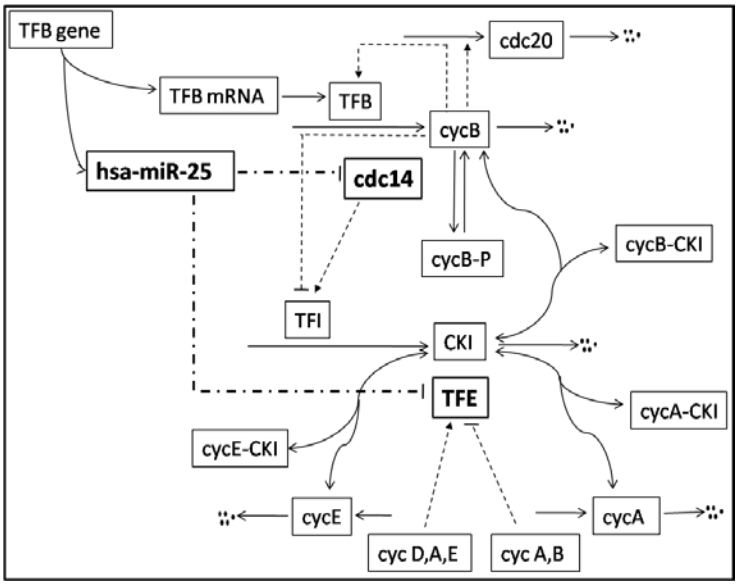


Figure 3-8 – Regulation by intronic miRNA hsa-miR-25 in the reaction network model of cell cycle

TFB gene is host gene for hsa-miR-25, which has two targets cdc14 and TFE. The cell cycle reaction network model is adapted from B. Novak and J. J. Tyson.

The host gene for intronic miRNA is MCM7, referred to as TFB in the model. Both the host protein and the targets are known to be important components governing the cell cycle dynamics (Rodriguez, Griffiths-Jones et al. 2004; Liu, Papagiannakopoulos et al. 2007; Petrocca, Visone et al. 2008). As intronic miRNA is co-expressed with the host gene, profiles of hsa-miR-25 and TFB were assumed to be similar. The effect of miRNA mediated inhibition was modeled as host protein concentration dependent decrease in the rate of synthesis of target proteins. The effect of varying the extent of inhibition by changing the inhibition constants km_1 and km_2 was studied. The effect of intronic miRNA mediated regulation was qualitatively analysed by observing the time course profiles of cyclin proteins. Figure 3-9 shows the time course profiles for cyclin A, B and E for different values of km_1 and km_2.

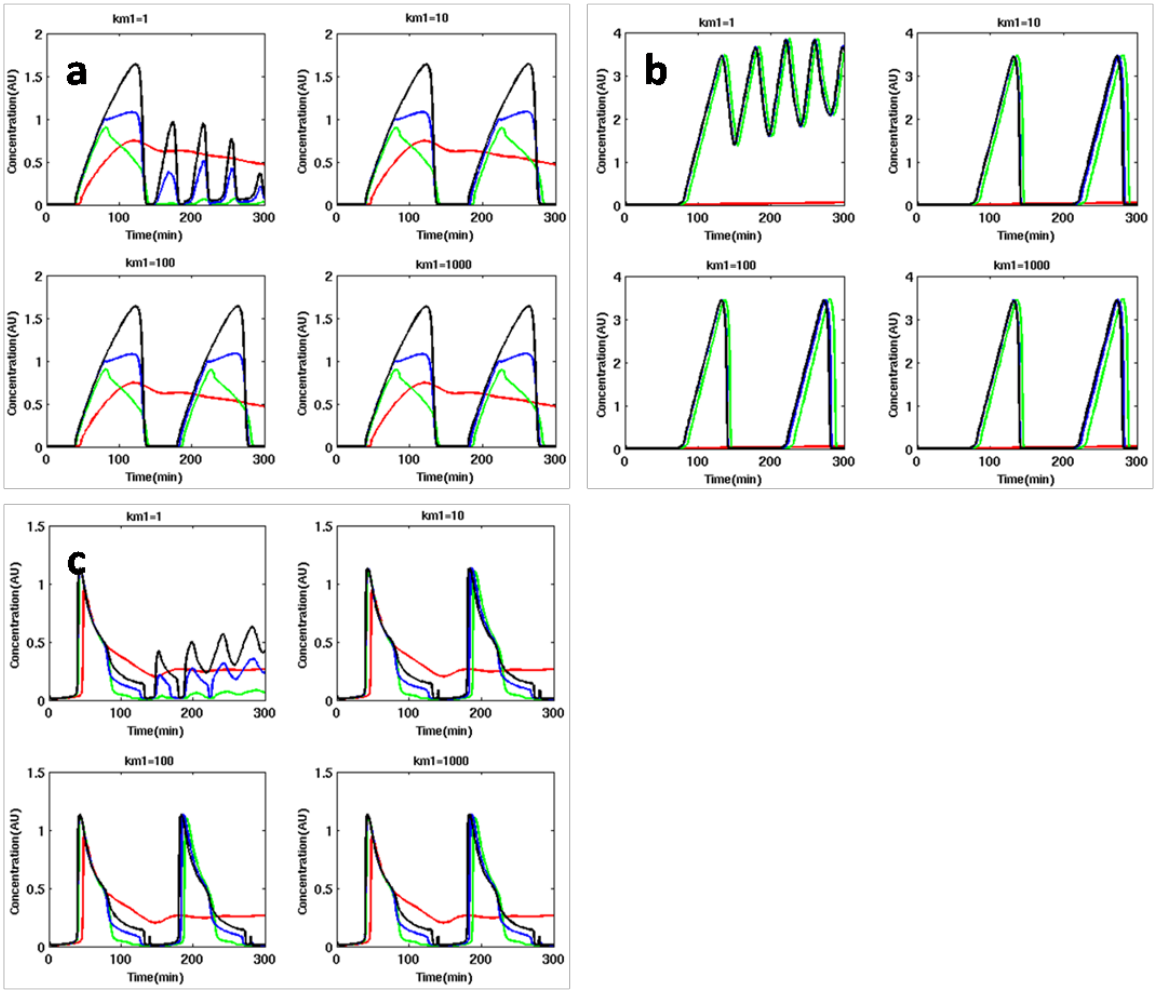


Figure 3-9 – Effect of changing inhibition parameters km_1 and km_2 on time course profiles for cyclins

Time course profiles of (a) CycA, (b) CycB, and (c) CycE. In all the three figures km_2 = 0.01 (red line), km_2 = 0.1 (green line), km_2 = 1 (blue line), and km_2 = 10 (black line).

For high values of parameters as expected the cyclin profiles were identical to the unmodified models. From Figure 3-9 it was observed that the model accuracy was more sensitive to changes in km_1 than km_2. The parameter values, km_1 = 10 and km_2 = 1 were observed to be resulting in sustained oscillations representing cell cycles of the same duration as that of unmodified model. These values were considered as the best set of parameters and were analysed further.

In addition to the proxy protein concentration dependent inhibition, the effect of constant average inhibition of target protein synthesis was examined. However, the time course profiles of the components were not in accordance with the experimental observations. This indicated that, concentration dependent inhibition was required to appropriately simulate miRNA regulatory effects. Details are given in Appendix IIb.

Figure 3-10 shows the cyclin profiles for modified and unmodified model for the best set of parameters (km_1 = 10 and km_2 = 1). The concentration profiles of other 16 components remained almost unchanged and are given in Appendix IIc.

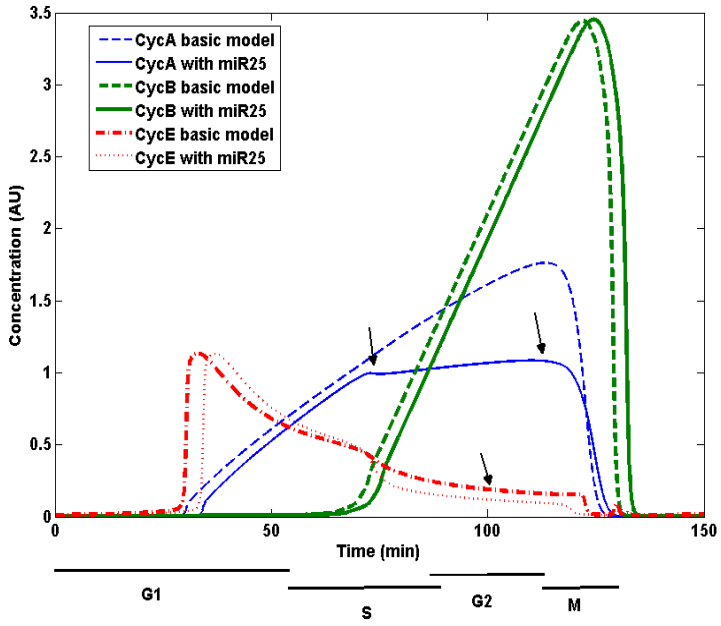


Figure 3-10 – Comparison of time course profiles of cyclin proteins for original cell cycle model and model with hsa-miR-25 regulation

Time course profiles of cyclin A, B, and E for 1 cell cycle duration. Parameter values used were, km_1 = 10; km_2 = 1. The arrows point to the observed improvements in the cyclin profiles.

Experimental studies have stated that cyclin A binds to cdk2 and cdc2 resulting in two distinct peaks of cyclin A kinase activities one appearing in S phase and the other in G2 phase. Cyclins A and B accumulate and reach a maximal level before mitosis at which they are degraded. Cyclin A is activated earlier than cyclin B and destroyed before the

peak of MFP (Norbury and Nurse 1992; Pagano, Pepperkok et al. 1992). Experiments have shown that cyclin A kinase activity appears at the start of S phase, continuous to increase throughout S phase and peaks in G2 phase (Koff, Giordano et al. 1992). In the simulations with the modified model, cyclin A was observed to increase after 40 min (late G1 phase) in all the simulations. It showed a slight decrease in S phase and again increased in G2 phase. Cyclin A decreased in M phase. Thus cyclin A profile simulated using the model with hsa-miR-25 effect was in agreement with experimental observations.

In case of cyclin E, experimental studies have reported its requirement at G1/S transition. Cyclin E activity peaks at late G1 and early S. The maximum amount of activity was 4-8 times greater than the minimum activity during the cell cycle (Koff, Giordano et al. 1992). In the simulations with unmodified model, cyclin E showed a peak before 50 min (towards late G1 phase) and then reduced to approximately 1/5 of its peak level. In case of simulations with modified model, for approximately 8 times reduction in cyclin E values was observed. The simulated cyclin E profiles were observed to be in better agreement with the experimental studies.

In this model, regulation by only one intronic miRNA was incorporated. Incorporation of regulation by other miRNA regulators would further improve the model performance. The simple function used in this case to represent regulatory effect of intronic miRNA considered only inhibitory effect. In order to incorporate activating effects, a different function and more parameters will be required.

3.3.3.2 Host protein mediated inhibition adequately represents dynamics of intronic miRNA mediated regulation

The method of using host protein as a proxy for incorporating regulatory effects of intronic miRNA was observed to be effective in case of cell cycle model. To examine the universality of this method, the effect of host protein mediated inhibition (equation 3.31) and intronic miRNA mediated regulation (equations 3.27-3.30) on target protein dynamics was compared for different initial conditions for a generic reaction network.

Figure 3-11 shows the time course profiles of target protein for five different initial conditions.

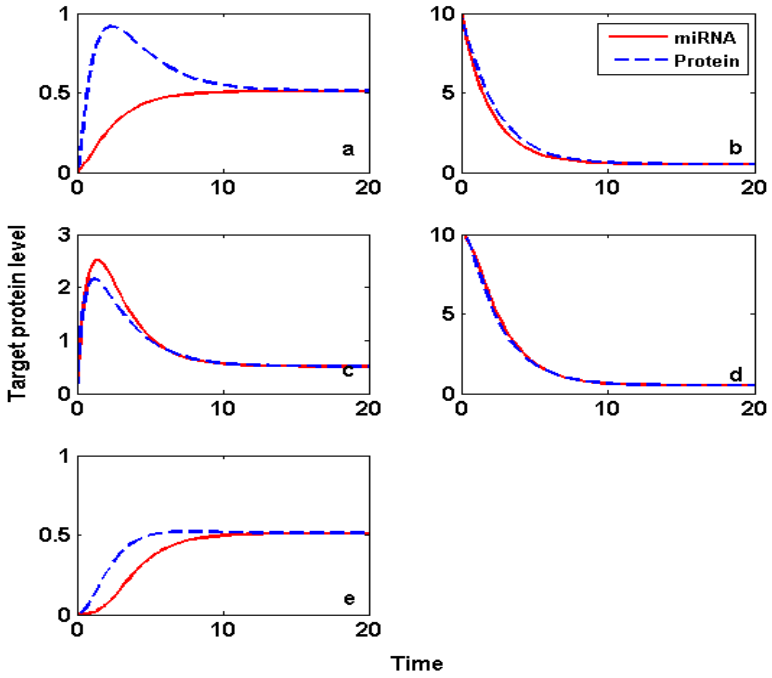


Figure 3-11 – Time course profile of target protein with intronic miRNA regulation and with host protein mediated inhibition

Initial conditions – (a) all components at zero level. (b) Target protein at unregulated steady state, all other components at zero level. (c) Target mRNA at unregulated steady state level, all other components at zero level. (d) Target mRNA and protein at unregulated steady state, all other components at zero level. (e) Regulator molecule (miRNA and host protein) at steady state, all other components at zero level.

The target protein profiles in case of both accurate (with miRNA) and simplified (host protein as proxy) models were observed to be similar in four out of five initial conditions. For the conditions where all components start from zero level, the target protein profiles were observed to differ significantly. However, in biological systems such a situation when both the regulator and regulated components are at zero concentration is unlikely to occur. For conditions where the regulator molecules (Figure 3-11e) or the regulated protein (Figure 3-11b, c, d) were at their unregulated steady state level the target protein dynamics were similar in both types of regulations. For the two conditions (c) and (e) the

observed target protein profiles were similar in case of inhibition by host protein and miRNA mediated regulation. Under conditions (b) and (d) target mRNA, and target mRNA and protein were initially kept at respective unregulated steady states. For these conditions the target protein profiles for the two types of regulations were almost identical. This indicated that the host protein mediated inhibition affected the dynamics of target protein in a way similar to that of miRNA mediated regulation. Thus it was observed that intronic miRNA mediated effect can be incorporated by using the host protein mediated inhibition in generic unregulated networks.

From the simulations it was observed that host protein mediated inhibition affected the dynamics of target protein in the similar way as that observed in case of miRNA regulation. Simulations were performed to explore the effect on the dynamics of target protein when the inhibition was removed. Under these conditions the target protein profiles were observed to be qualitatively similar. The details are given in Appendix IId.

3.3.3.3 Protein mediated feedback can show similar dynamics to that with miRNA mediated feedback

In the cellular systems, various layers of regulation are coordinated through interactions of regulatory molecules. Regulatory molecules such as transcription factors or miRNA do not work in isolations but form a complex interaction network. Under such conditions, protein regulators can regulate miRNA synthesis and miRNA can regulate the expression of transcription factors forming feedback loops. Therefore, in addition to the comparison of regulation by intronic miRNA and host protein without any feedback regulation; these two mechanisms of regulation were compared in the presence of feedback motifs. The dynamics of target protein was compared in the presence of miRNA and protein mediated positive and negative feedback loops. Two conditions were examined where, target protein positively (Aguda, Kim et al. 2008) or negatively (Stark, Brennecke et al. 2005) regulates miRNA synthesis. In each case analogous protein mediated post-transcriptional regulation was considered. The dynamics was quantitatively compared by considering rise time in the two types of regulations.

In case of negative auto-regulatory motif, two fold down regulation of steady state level of target was protein was considered. Two different initial conditions were considered (Figure 3-12 a and b). Qualitatively the target protein profiles were observed to be similar for the two types of regulations. For the initial conditions, where both the target and regulator molecule were set to zero initial level (Figure 3-12a), the rise time in case of miRNA mediated negative feedback was about 1.8 times lesser than that for protein mediated negative feedback. However, as discussed the condition of all components starting at zero level is unlikely to occur. For the second initial condition, where the target mRNA was set to its unregulated steady state level and regulator was set to zero level (Figure 3-12b), the rise time in case of miRNA mediated feedback was about 1.25 times lesser than protein mediated feedback.

Similar to the condition of negative feedback, 2 fold up-regulation of steady state level of the target protein was considered for positive feedback. The same initial conditions were examined (Figure 3-12 c and d). For both the initial conditions, significant difference in the rise time was not observed for miRNA and protein mediated feedback regulation. The target protein profiles were observed to be qualitatively and quantitatively similar.

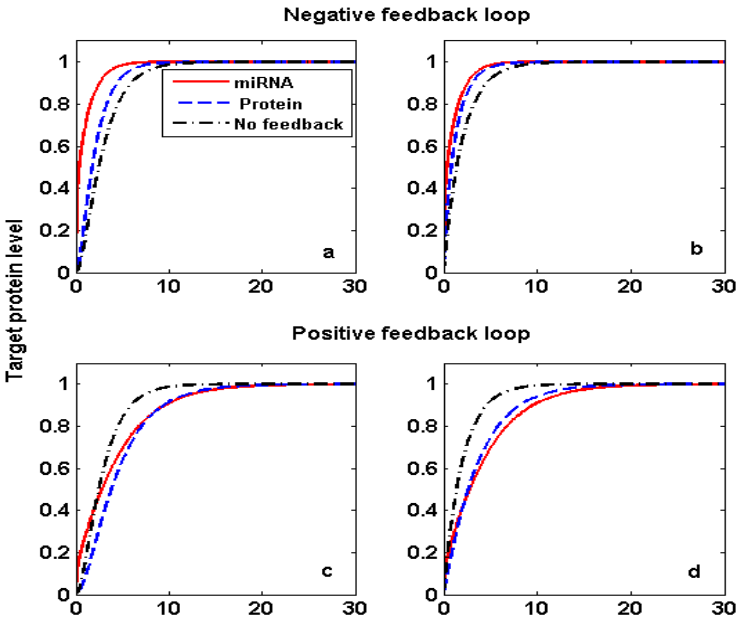


Figure 3-12 – Time course profiles of target protein with miRNA regulation and protein mediated regulation

The level of target protein is normalized with respect the maximum value. (a) and (b) for negative auto-regulatory loop. (c) and (d) for positive auto-regulatory loop. Initial conditions – (a) and (c) all components at zero level (b) and (d) Unregulated steady state level of target mRNA, other components at zero level.

It was observed that protein mediated feedback can result in similar target protein profiles as that obtained by miRNA mediated feedback. It is known that negative feedback speeds up the response while positive feedback slows down the response (Rosenfeld, Elowitz et al. 2002; Alon 2007). In this study it was observed that, the negative feedback loop mediated through miRNA had lesser rise time than the protein mediated negative feedback. However, significant difference was not observed between the two types of regulations in case of positive feedback motif. Overall, protein mediated post-transcriptional regulation was observed to have similar effects on the target protein dynamics as miRNA mediated regulation both without and with feedback loops. A similar study comparing sRNA and transcription factor based motifs in bacteria by Mitarai and Anderson et al showed that parameters can be adjusted to get same rise time (Mitarai, Andersson et al. 2007). The observations in the present study agree with this observation. However, the observation of similar rise time questions the advantage of having miRNA based regulation. A major advantage that has been suggested is metabolic

cost (Mitarai, Andersson et al. 2007). Depending upon the cellular environmental conditions and the fluctuations either of the motif is optimal in terms of energy cost, though both motifs produce similar response. Another advantage of miRNA mediated regulation is target prioritization. One miRNA can prioritize expression levels of various mRNA targets depending upon different binding specificities. Such prioritization has not been reported for protein mediated regulation, such as for RNA binding proteins.

From the comparison of miRNA mediated and protein mediated regulation, the method of using protein proxy was observed to be generic and can be used in the existing mathematical models. Such incorporation would not represent any biological mechanism of regulation but would help to incorporate the observed regulatory effect.

3.3.4 Conclusion

In this study, a simpler method to incorporate miRNA mediated dynamic regulatory effects into existing mathematical models of cellular processes was developed. The predictions of the modified model were found to be in better agreement with the experimental observations. Concentration dependent inhibition was found to be better method than using a constant average inhibition. In this study, regulation by only one intronic miRNA was incorporated. Incorporation of regulation by other miRNA components would improve the predictive capability of the model. In this model, a simple function capable of capturing only the inhibitory effects of miRNA was used. Additionally, there are some reports of different expression patterns of intronic miRNA and the host gene (Aboobaker, Tomancak et al. 2005; Isik, Korswagen et al. 2010). To represent such cases certain modifications (and more parameters) will be required in the model. In addition, to incorporate regulation of non-intronic miRNA additional parameters would be required. The universality of the method was verified by comparing miRNA mediated regulation with analogous protein mediated post-transcriptional regulation. In addition, the two types of regulations were compared in presence of positive and negative feedback motifs. It was observed that protein mediated post-transcriptional regulation can adequately represent miRNA mediated regulation in terms of target protein profiles, indicating that this method can be used to incorporate the observed regulatory effects of miRNA.

4 A comprehensive model of gene expression to explore the noise in the protein molecules

4.1 Introduction

Gene expression is central to all biological processes. It consists of multiple steps such as gene activation, transcription, translation, and mRNA and protein degradation. Each of these steps itself is a multistep process. All the steps in gene expression are tightly regulated by various RNA and protein factors to produce diverse but defined outcomes. The coordinated action of several such factors governs the concentration of all the protein and mRNA components in a cell. The regulation of gene expression by the products of gene expression viz., RNA and protein components inherently imply presence of feedbacks. Presence of feedback loops can results in unintuitive behaviors such as multi-stability, oscillations, hysteresis etc. Mathematical modeling, simulation and analysis is an effective approach to study such systems.

There are various types of mathematical models of gene expression. A majority of them are specific to one of the processes such as transcription, mRNA transport and do not include details of multiple processes. It is very difficult to capture the vast amount of details of protein production process in a single mathematical model. As most mathematical models are developed for a specific purpose, for instance to analyze a particular sub-process, or to help suggest explanation for a particular experiment, this level of detail is usually not required. In addition, development of detailed generic model is challenging as there are multiple processes specific only for a certain set of genes. However, in order to understand the contribution of different sub-processes to the overall dynamics of gene expression, their inclusion in a comprehensive model is necessary.

An important property of gene expression is stochasticity. Isogenic population of cells is known to show cell-to-cell phenotypic variation as shown by Spudich and Koshland Jr (Spudich and Koshland Jr 1976) in case of bacterial chemotaxis. Genome wide single cell measurement studies have shown that noise in the steady state protein level decreases as the protein abundance increases. However, this relationship was observed to be true only for low-abundance and intermediate-abundance proteins. Highly expressed proteins show a constant minimum level of noise. This noise floor has been attributed to slow varying extrinsic fluctuations such as fluctuations in the levels of ribosomes, polymerases etc.

Current simple models of gene expression cannot explain the noise saturation behavior without inclusion of an added noise term.

In addition to this, the knowledge about the contribution of each step in gene expression to noise in steady state protein level is ambiguous. Different studies have suggested different steps such as translation or transcription to be important in determining the variability in protein number. Therefore, there is a need to quantitatively estimate the relative contribution of steps in gene expression to the noise in protein level.

In this study, a detailed and generic model of gene expression was developed in order to examine whether, the sources of gene expression-extrinsic and cell-intrinsic noise considered in detailed model can explain saturation of noise. The effect of time scale of fluctuation on the saturation of noise was also studied. From the study the time scale of extrinsic fluctuation was found to be important. It was observed that extrinsic fluctuations having slowest time scale of fluctuation can contribute to the observed saturation of noise. Global sensitivity analysis was performed to quantitatively measure the relative contributions of steps of gene expression on the steady state noise in protein level using 3 different models of gene expression and two measures of noise, CV and Fano factor. Interestingly, the two measures of noise were observed to be sensitive to different parameters.

The chapter is divided into two sections. Section 1 describes the development of comprehensive model of gene expression and, stochastic simulations to understand the noise-protein abundance relationship. Section 2 describes the global stochastic sensitivity analysis of 3 different models of gene expression using two different measures of noise.

4.2 Section 1 – Development of a comprehensive model of gene expression to investigate extrinsic noise sources contributing to noise saturation

4.2.1 Introduction

Gene expression being of key importance has been extensively studied using both experimental and theoretical approaches. A large number of mathematical models of gene expression of different types such as deterministic, stochastic, continuous, or discrete

have been developed. There are some reviews of mathematical models of gene expression (Smolen, Baxter et al. 2000; Hasty, McMillen et al. 2001) that discuss various approaches available and used for mathematical modeling of gene expression, gene regulation and gene networks. As a first step towards the development of a comprehensive model, we have reviewed previously developed mathematical models of processes in gene expression (Gokhale, Nyayanit et al. 2011). We have divided gene expression into five major sub-processes viz., chromatin remodeling and histone modification, transcription, post-transcriptional processing and RNA degradation, translation, and protein degradation. The review consists of brief description of these processes, their regulation by both protein and RNA regulators, with examples of each. It also includes feedback regulation of each process by protein and RNA regulators with known examples. Table 4-1 summarizes examples for processes in gene expression with their regulation.

Table 4-1 – Literature references for processes in gene expression with regulation by protein and RNA regulators

Regulator	**Process**				
	Chromatin remodeling and Histone modification	**Transcription (Initiation, Elongation and Termination)**	**Post-transcriptional processes (Capping, Poly-adenylation, Splicing, Interference and RNA degradation)**	**Translation (Initiation, Elongation and Termination)**	**Post-translational modification (ubiquitination and Protein degradation)**
Protein	(Peterson 2002; Kouzarides 2007)	(Kerppola and Kane 1991; Thomas and Chiang 2006)	(McCracken, Fong et al. 1997; Petersen-Mahrt, Estmer et al. 1999; Guhaniyogi and Brewer 2001; Rozenblatt-Rosen, Nagaike et al. 2009)	(Monnier A., Belle R. et al. 2001; Sans M.D., Xie Q. et al. 2004; Wang X. and Proud C.G. 2008)	(Vervoorts J., Luscher-Firzlaff J. et al. 2006)

Protein Feedback	(Yang X.O., Angkasekwinai P. et al. 2009)	(Cormack and Struhl 1992; Arigo, Carroll et al. 2006)	(Cheng C., Yaffe M.B. et al. 2006)	(Raney A., Law G. L. et al. 2002; Onouchi H., Nagami Y. et al. 2005; Ivanova I. P., Loughrana G. et al. 2010)	(Hutti J.E., Turk B. E. et al. 2007; Jessica E. Hutti 2007; Noula Shembade1 2010)
RNA	(Volpe, Kidner et al. 2002; Barrandon, Spiluttini et al. 2008; Gonzalez, Pisano et al. 2008)	(Brantl and Wagner 2002; Kwek, Murphy et al. 2002; Mattick and Makunin 2006)	(Bartel 2004; Storz, Altuvia et al. 2005; Ghildiyal and Zamore 2009)	(Wang, Iacoangeli et al. 2002)	None to our knowledge
RNA Feedback		(Barrandon, Spiluttini et al. 2008)	(Xie, Kasschau et al. 2003)	(Li, Vilardell et al. 1996; Ben-Asouli, Banai et al. 2002)	None to our knowledge

It was observed that the reports of RNA mediated feedback regulation of individual steps of gene expression process were much fewer than reports of protein mediated feedback regulation. With more studies at transcriptome level, more such RNA mediated regulatory motifs will be revealed. Mathematical models for five major sub-processes with regulation by protein and RNA regulators are summarized in Table 4-2.

Table 4-2 – Literature references for mathematical models for gene expression and its regulation at each step by protein and RNA

Regulator	Process				
	Chromatin remodeling and Histone modification	**Transcription (Initiation, Elongation and Termination)**	**Post-transcriptional processes (Capping, Poly-adenylation, Splicing, Interference and RNA degradation)**	**Translation (Initiation, Elongation and Termination)**	**Post-translational modification (Ubiquitination and Protein degradation)**
Protein	(Kim H. D and O'Shea E.K 2008; Luca Mariani 2010)	(Kugel and Goodrich 2000)	(Cao and Parker 2001; Singh, Yang et al. 2007)	(Nayak S. , Siddiqui J.K. et al. 2011)	(Holzhütter and Kloetzel 2000; Peters, Janek et al. 2002; Luciani, Kesmir et al. 2005)
Protein Feedback	(Sedighi M. and Sengupta A. M 2008; Narula J., Smith A. M. et al. 2010)	(Bernard S., Cajavec B. et al. 2006; Rajala T., Hakkinen A. et al. 2010)	Model needed	(Bar N.S. 2009; De Silvaa E., Krishnana J. et al. 2010)	(Lee J., Choi K. et al. 2010)
RNA	Model needed	Model needed	(Levine, Ben Jacob et al. 2007)	Model needed	Model needed
RNA Feedback	Model needed	Model needed	(Aguda, Kim et al. 2008)	Model needed	Model needed

More examples of mathematical models with details of type of model, and main output of the model are given in Appendix IIIa.

Some of the mathematical models consider the overall process of gene expression but at lesser detail while some models focus on a part of the whole process in great detail.

Majority of these models include one or more lumped parameter or reaction to represent sub-processes in gene expression. Depending upon the degree of details the models can be qualitatively classified. Figure 4-1 gives a visual representation of the sub-processes considered in different mathematical models and the degree of details considered.

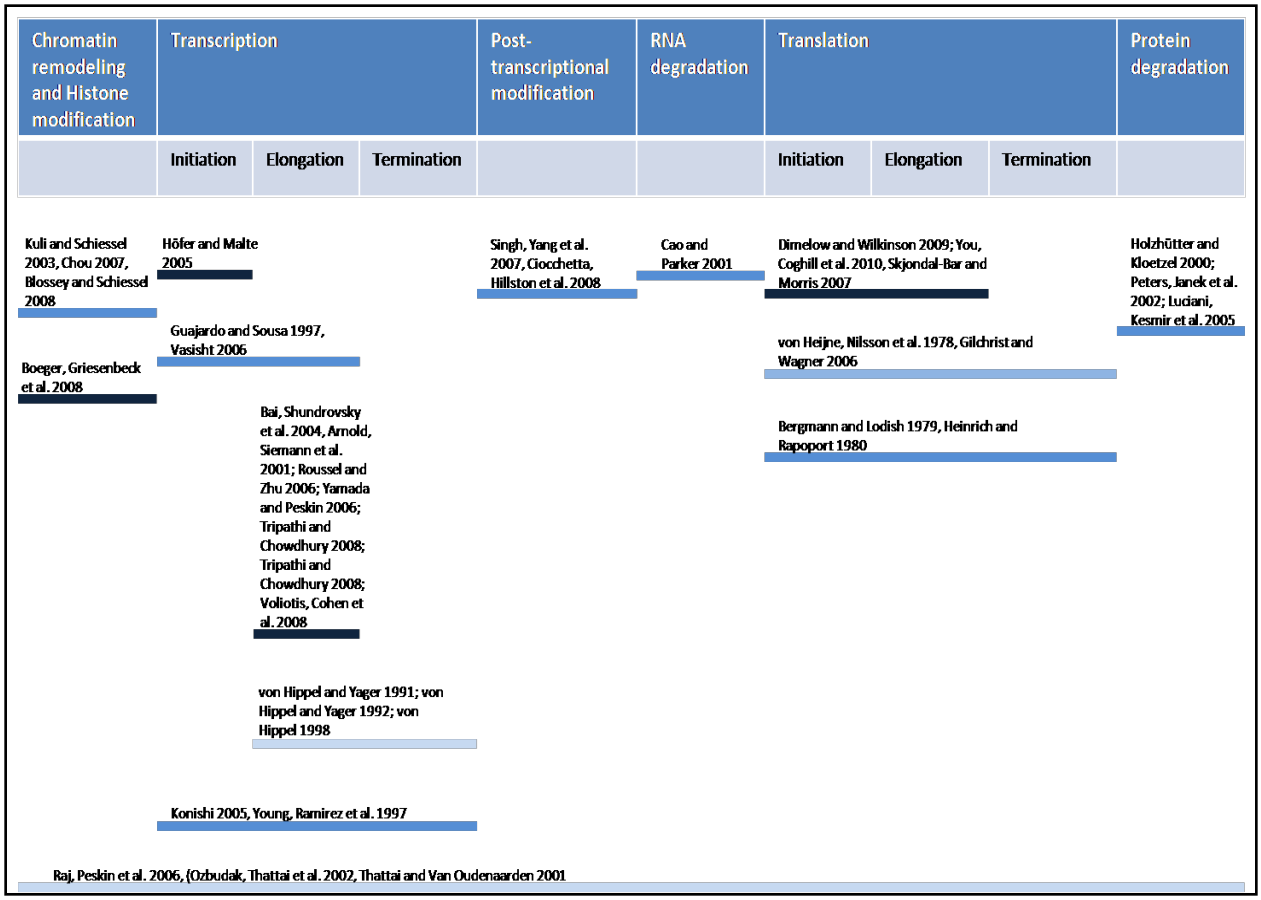


Figure 4-1 – Diagrammatic representation of the sub-processes considered in different mathematical models and the degree of details

Some of models for the protein production process organized in terms of breadth and level of detail included. The intensity of the blue bar qualitatively indicates the relative level of details included in the model.

It was observed that the number of mathematical models developed for transcription and translation were much higher in number than the models for RNA processing, degradation, and protein degradation. A detailed model for mRNA transport from nucleus

to cytoplasm has not yet been developed. These are some of the sub-processes where development of mathematical model and analysis is required. In addition, a comprehensive and generic model of gene expression is required to be developed in order to understand the relative contribution of sub-processes to the overall dynamics of gene expression. Having the models of sub-processes, a comprehensive model of gene expression can be developed by using modular approach. A mathematical model for each sub-process can be considered as a module with defined input and output. Different such modules can be linked together to develop a detailed model of entire pathway of gene expression from gene activation to protein degradation. In this study a detailed model of gene expression in eukaryotes was developed using reaction frameworks from previously developed mathematical models and experimental data.

In addition to the interesting properties such as oscillations and switching observed at a population level, many other significant properties like bimodality are evident in single cell studies of a population. After the first report of fluctuations in autocatalytic reactions (Delbruck 1940), in last couple of decades there have been numerous experimental and theoretical studies focusing on different aspects of stochasticity in gene expression such as its origin, consequences (Raser and O'Shea 2005), types of noise (Volfson, Marciniak et al. 2005), and methodology to separately measure different sources of noise (Swain, Elowitz et al. 2002). In addition to these studies focusing on one or a few genes, development in single cell measurement technique led to some proteome wide single cell measurement studies in *E. coli* (Taniguchi, Choi et al. 2010) and yeast *Saccharomyces cerevisiae* (Bar-Even, Paulsson et al. 2006; Newman, Ghaemmaghami et al. 2006). These studies have revealed the global structure of noise in steady state protein level. The studies have shown that the noise in protein steady state level has inverse relation with protein abundance. However, the inverse relation was observed only for proteins with low and intermediate abundance. In case of high-abundance proteins the noise did not decrease below a certain low value, an observation which could not be explained by previously developed simpler models. A theoretical study by Paulsson (Paulsson 2004; Paulsson 2005) has shown that at very high expression level, the intrinsic noise reduces to a level lower than the extrinsic noise. Therefore, the study has theoretically shown that such genes should have the same noise level dominated by extrinsic noise. The observed

noise floor has been attributed to slow varying fluctuations in the level of polymerases, ribosomes etc (Taniguchi, Choi et al. 2010), which are extrinsic to the developed models. In addition to the genome wide studies, early studies (Rosenfeld, Young et al. 2005; Volfson, Marciniak et al. 2005; Sigal, Milo et al. 2006) examined dynamics of fluctuations in certain protein expression. These studies examined the auto-correlation of fluctuation and suggested that fluctuation time scales larger than cell cycle duration are important for population variability, while fast varying fluctuations may average away and contribute little to cell-to-cell variation. A theoretical study by Shahrezaei et al (Shahrezaei, Ollivier et al. 2008) suggested that the extrinsic noise in gene expression is coloured and the correlation time equivalent to cell cycle is important for population variability.

In this study a detailed model of gene expression was developed. The model considered certain sources of cell-intrinsic noise in gene expression such as gene activation due to binding of regulator protein, binding of RNA polymerase, and competition by other mRNA components for translation machinery. The detailed model was used to investigate whether the sources of noise considered in the model can explain saturation of noise. The effect of addition of extrinsic fluctuations of different time scales on the variability at high protein abundance was studied. The relationship between the time scale of extrinsic noise and that of effecter protein was identified. It was observed that extrinsic fluctuations added using slowest varying protein can show saturation of noise at high abundance. The time scale of simulation compared to that of the fluctuation was also observed to be an important factor that determined the observed saturation.

4.2.2 Methodology

4.2.2.1 Model development

A detailed generic model of eukaryotic gene expression processes was developed to include possible details of the process. Some of the reactions were adapted from previously developed mathematical models and some were obtained from the experimental studies. The parameter values were obtained from mathematical models, experimental studies, and databases. The complete model contained 41 components and 42 reactions. The details of each reaction are discussed in Results and Discussion section.

4.2.2.2 Parameter values and sampling

The parameters for specific rate of mRNA synthesis, protein synthesis, mRNA and protein degradation were obtained from genome wide population level experimental study by Schwanhausser, Busse et al (Schwanhausser, Busse et al. 2011) for mouse fibroblast cells. It was observed that the ranges of these parameters cover a wider range than that observed for parameters in yeast. Four parameters, one each for mRNA synthesis, protein synthesis, mRNA degradation, and protein degradation, constitute one parameter set. One such parameter set comprising these four parameters corresponds to a particular value of mean protein abundance. The entire dataset contained 4247 proteins, and therefore 4247 such sets with all the four reported parameter values in one parameter set. The parameter sets were selected such that they represent the experimentally observed protein abundance proportion in the total data set. The data sets were grouped according to the protein abundance as shown in Table 4-3. Random sample from each abundance group was selected such that the proportion from each abundance group was identical to that observed in the experimental data. Total 500 parameter sets were selected. For a sample of total 1000 data points, two samples of 500 parameter sets were obtained ensuring unique parameter sets.

Table 4-3 – Proportion of protein in each abundance range and corresponding number of data points selected to generate a sample of 500 data points (parameter sets)

Protein abundance (molecules/cell)	# data points (parameter sets)	% data points	# sample data points selected
10^2	391	9.23	46
10^3	1470	34.70	173
10^4	1684	39.75	199
10^5	617	14.56	73
10^6	74	1.74	9

From the parameter sets, specific mRNA degradation rate for detailed model (k_7 - 42-reaction model) was obtained by minimizing the percent relative error between the steady state total mRNA level of 4-reaction model and 42-reaction model using Matlab function `fmincon`. The other parameters values are listed in Table 4-4 and the details are described in Results and Discussion section.

4.2.2.3 Stochastic simulation

Gillespie's exact stochastic simulation algorithm (Gillespie 1976) was used for simulation. The detailed model of 42 reactions is a stiff system. From the reaction rate parameter values (Table 4-4) it was observed that the reactions such as, gene activation, inactivation, mRNA synthesis were slow, while the reactions such as binding of eukaryotic translation initiation factors (eIF) were sometimes two orders of magnitude faster. The molecular abundance of eIFs was also orders of magnitude higher than other low abundance components such as RNA polymerase, regulator molecules. Stochastic simulations of the entire model would require longer time, as faster reactions would lead to a smaller increment in time at each step. Therefore, to reduce the time required for simulation, the reactions were separated into two groups viz., slow reactions (Reactions, R1 to R7, R22, R23, R26, R27, R30, R31, R34, R35, R38, R40 and R42) and fast reactions (the remaining reactions) depending upon the reaction rate parameter value and the molecular abundance. As the fast reactions contained components with high abundance deterministic simulation would be an appropriate approximation. Therefore, the fast reactions were simulated numerically using deterministic kinetics in Matlab version 7.6.0.324 (The Mathworks, Natick, USA) using `ode15s`. The steady state output values of the two components *c1* and *eIF4G*, which were involved in slow reactions, were given as input to the slow reactions. The slow reactions were simulated using in-house developed Fortran codes for Gillespie's exact stochastic simulation algorithm (Gillespie 1976). Parallel version of the codes was used to run parallel simulations. The simulations were performed for final time of $3*10^5$ min, i.e. more than ten times the time scale obtained with slowest reaction rate parameter. A sample of 1000 runs was used to generate statistics. Coefficient of variation, (CV) defined as the ratio of standard deviation to mean (σ/μ), was used as a measure of noise.

4.2.2.4 Calculation of timescale of fluctuation

To determine the timescale of fluctuation, autocorrelation function (ACF) for each run of the time course data for 500 proteins simulated using 4-reaction model was obtained. Autocorrelation function is defined as a set of autocorrelation coefficients arranged as a function of separation in time or time lags. The ACF was calculated using Matlab function obtained from Mathworks file exchange. The average ACF for each protein was calculated from the 1000 ACF values obtained for each run. Mixing time (τ_m), defined as the time required for ACF to reach half of its initial value, was calculated for the 500 proteins. Mean ACF and mixing time are illustrated in Figure 4-2.

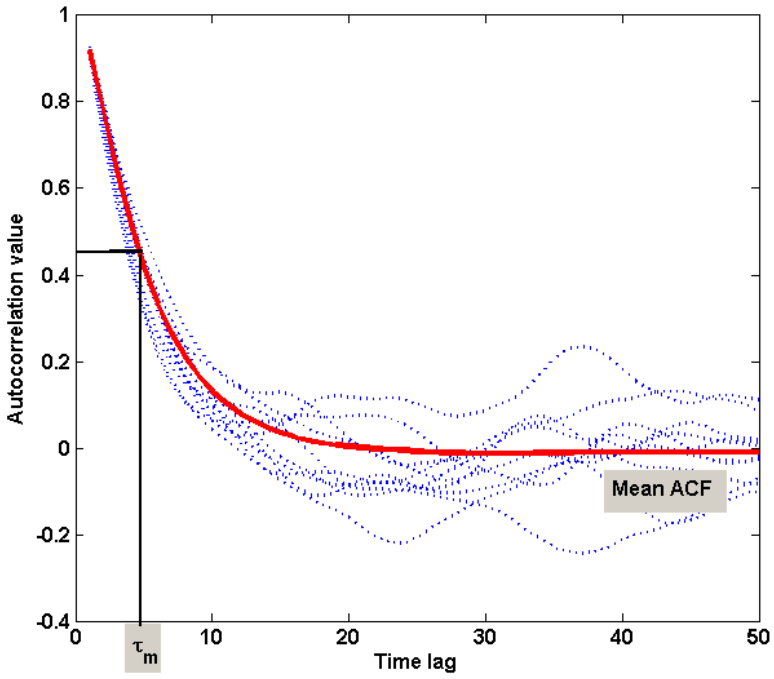


Figure 4-2 – Graph of autocorrelation values vs. time lags

Each blue dotted line is an auto correlation function (ACF) for one run of one parameter set (or protein). Red solid line is mean ACF of all the 1000 runs for one parameter set (or protein). τ_m is the time at which ACF reduces to half of its initial value.

Mixing time was used as a measure of time scale of fluctuation. The effect of addition of fluctuations of different mixing time on the noise at steady state protein level was studied.

4.2.2.5 Definition of a measure to quantify observed saturation of noise

A quantitative measure was required to measure and compare the observed saturation. In the graph of steady state $\%CV^2$ vs. mean protein abundance, two abundance regions were defined one with range 10^2-10^5 and other > 10^5. A linear equation was fit to each of the abundance region, using least square fit method of Matlab. The ratio of the slopes, of region $>10^5$ to that of abundance region 10^2-10^5 was calculated. The ratio was used as measure to determine the extent of saturation. Figure 4-3 graphically represents the measure used to quantify saturation of noise.

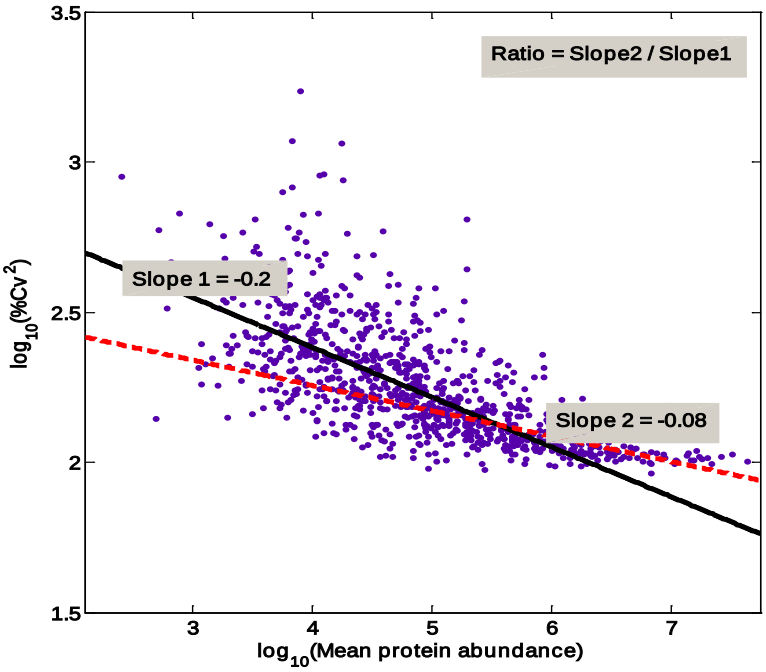


Figure 4-3 – Figure illustrating the calculation of ratio of slopes

Black solid line represents a linear fit to the data points in the $\log_{10}$(Mean protein abundance) region of 2 – 5, with slope, Slope1. Red dashed line represents a linear fit to the data points in the $\log_{10}$(Mean protein abundance) region > 5, with slope, Slope 2. The ratio of two slopes, Slope2/Slope1 was used as measure of saturation.

For comparison with the experimental observation, the data of running median given in the Figure 2g (gated population) of the published study by Newman, Ghaemmaghami et al (Newman, Ghaemmaghami et al. 2006) was extracted using Datathief software. The entire data from the experimental study could not be extracted using Datathief due to multiple overlapping data points. The obtained data of running median was used to calculate slopes. Slopes for regions 10^2 to 10^3 and 10^3 to 10^4 were calculated as there were no data points above abundance of 10^5. The ratio of the slopes was found 0.22.

4.2.3 Results and Discussion

4.2.3.1 A detailed mathematical model of eukaryotic gene expression

Starting with the previously developed (Thattai and Van Oudenaarden 2001) widely used 4-reaction model of gene expression a detailed model of gene expression process was developed. The 4-reaction model (Figure 4-4) contained zero order mRNA synthesis, first order synthesis of protein with the rate of reaction dependent upon mRNA level, and first order degradation of mRNA and protein with the rate of reactions dependent upon mRNA and protein levels respectively.

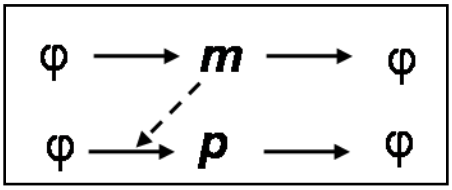


Figure 4-4 – 4-reaction model of gene expression

The model contains zero order mRNA synthesis, first order synthesis of protein catalysed by mRNA, and first order degradation of mRNA and protein.

The 4-reaction model has been extended to a 6-reaction model, for instance in, (Raj, Peskin et al. 2006) to include reactions of gene activation and gene inactivation (Figure 4-5). In this model, mRNA synthesis is considered to be a first order reaction with rate dependent on the level of active gene.

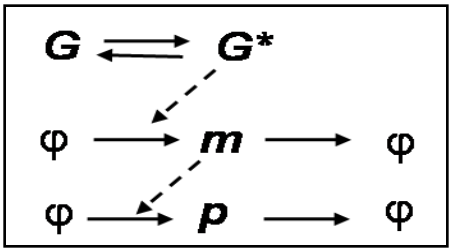


Figure 4-5 – 6-reaction model of gene expression

In addition to the 4 reactions of mRNA and protein synthesis and degradation, the reactions for gene activation and deactivation are considered in the model.

Modifications of these models have been carried out to include details of certain steps. For instance, the model by Blake et al (Blake, Kærn et al. 2003) considers two steps of gene activation to form a pre-initiation complex. Using such different models for sub-processes in gene expression and experimental data, a detailed model of gene expression was developed. The model contained 6 major processes viz., gene activation, transcription, mRNA transport, translation, mRNA degradation and protein degradation. The detailed model contained 41 components and 42-reactions. The reaction rate expressions and parameter values are given in Table 4-4.

Gene activation – Gene activation can be considered as the first step for gene expression. The process consists of action of chromatin remodelers and histone modifying enzymes which slide or peel off histones from chromatin making the regulatory region of gene accessible for transcription factors and RNA polymerase machinery. The process of gene activation is thus a multistep process. In addition, the modifications on histone and sliding of histone to specific position are different from gene to gene and also differ for different physiological conditions. In the model all the steps which make the promoter accessible to RNA polymerase machinery, were considered into one reaction. The reaction was represented as

$$R1: Gene + Reg \xrightarrow{k_1} GeneActive$$

In this reaction a representative regulatory protein (*Reg*) was considered to bind to a gene, to make the gene compatible for RNA polymerase binding. In this case, *Reg* does

not imply any specific protein but is considered as a generic regulatory protein. Therefore, *Reg* could be replaced by the particular regulator for the study of any specific gene where that regulator is known to be required. The transcription factor, TATA-binding protein (TBP) is known to be essential for transcription. Hence, in this study, to calculate abundance of *Reg* per gene, abundance of TBP was used. In case of yeast cell, there are ~20,000 TBP molecules per cell (Milo, Jorgensen et al. 2010). Considering the average number of genes expressed in yeast to be around 4500 (Milo, Jorgensen et al. 2010), availability of *Reg* per gene was found to be 4.4 molecules. Therefore, total *Reg* abundance was considered to be 4 molecules per gene. However, depending upon the regulatory protein under study *Reg* parameter can be changed.

The activated gene was modeled as getting deactivated upon dissociation of regulator. The reaction was represented as,

$$R2: GeneActive \xrightarrow{k_2} Gene + Reg$$

A different deactivating protein was not considered in the reaction of gene deactivation. Addition of deactivating protein is possible with the corresponding change in the reaction rate parameter.

In this model, total *Reg* amount was considered as constant. Thus binding of *Reg* to *Gene* would change the free *Reg* amount. The effect of addition of *Reg* distribution on the noise in steady state protein distribution was studied. The effect of addition of reactions of regulator protein synthesis and degradation on the steady state noise of effecter protein was also investigated in detail.

In this case *Gene* was considered to have only one copy. However, diploid or multiple copies of genes can be considered. In a simple manner, parameter of *Gene* molecular abundance can be changed from one to the desired copy number and the reaction rate parameter can be adjusted to a value to obtain the same rate of gene activation. To account for the source of variability due to presence of alleles of a gene, different components such as *Gene1, Gene2* can be considered, each leading to synthesis of the same type of *mRNA*. Such framework would be useful in order to separate the allele specific intrinsic noise contribution from the total noise at protein level.

Transcription – Pre-initiation complex formation was considered as the first step in transcription initiation process. There are two pathways of assembly of RNA polymerase machinery on the regulatory region of gene called the sequential assembly pathway and RNA polymerase holoenzyme pathway (Thomas and Chiang 2006). In this model subunits of RNA polymerase were not considered as separate components. RNA polymerase was considered as a single component binding to active gene. Such assumption was equivalent to RNA polymerase holoenzyme pathway, in which all the subunits assemble to form the polymerase holoenzyme which then binds to promoter of a gene. RNA polymerase was considered to bind reversibly to the activated gene to form a complex called *GeneActive_RNApol*. The complex was considered to catalyze mRNA synthesis. During transcription pre-mRNA is synthesized which is processed by splicing of introns, 3' polyadenylation and 5' capping to form mature mRNA. Splicing and capping is known to occur co-transcriptionally (Singh and Padgett 2009). Therefore, in this model, a single reaction of mRNA synthesis was assumed to result in mature mRNA synthesis. The reactions for transcription were represented as,

$$R3: GeneActive + RNApol \xrightarrow{k_3} GeneActive\_RNApol$$

$$R4: GeneActive\_RNApol \xrightarrow{k_4} GeneActive + RNApol$$

$$R5: GeneActive\_RNApol \xrightarrow{k_5} GeneActive + RNApol + cap\_mRNA\_AAAn$$

In this case, once the mRNA was synthesized, RNA polymerase was considered to dissociate from the *GeneActive_RNApol* complex, as occurs in transcription termination. The gene was considered to remain in activated form after RNA polymerase dissociation to account for transcription re-initiation event.

Similar to *Reg* abundance, it is known that there are ~30000 RNA polymerase molecules in yeast cell (Borggrefe, Davis et al. 2001). Hence, total *RNApol* abundance is considered to be 7 molecules per cell. In this model, non-specific binding of regulatory molecules or polymerase to DNA was not considered.

There are mathematical models which consider nucleotide wise increase in length of mRNA, look-ahead and backtracking features of RNA polymerase (Bai, Shundrovsky et al. 2004; Voliotis, Cohen et al. 2008). These details are possible to incorporate if the

model is used for a certain known protein. In such case knowing the length of mRNA, reactions of type

$$GeneActive\_RNApol \xrightarrow{k\_elongation} GeneActive\_RNapol\_mRNA_i$$

where, *i* varies from 1 to *n*, can be incorporated. In such framework, termination can be considered once mRNA of length *n* nucleotides is formed. The reported speed of RNA polymerase for instance, ~2 kb/min (Ardehali and Lis 2009) can be used to obtain the value of elongation reaction rate parameter. In case of this generic model these details of nucleotide wise polymerization were not included.

In this model, binding of multiple RNA polymerase molecules on one gene, promoter clearance, and RNA polymerase traffic on DNA were not considered. These parameters are dependent on the gene length and speed of RNA polymerase. Alternative splicing being gene and condition specific phenomenon was not considered in this model.

mRNA Transport – After the splicing and other co-transcriptional or post-transcriptional modifications the mature mRNA is transported to cytoplasm. Certain protein factors such as Mex67 are required for the active transport of mRNP complex. However, some studies have shown that there is no accumulation of mRNP around nuclear pore (Mor and Shav-Tal 2010). In experimental study by Audibert, Weil et al (Audibert, Weil et al. 2002) mRNA transport has been reported to be a first order reaction. Therefore, in this model mRNA transport was considered to be a first order reaction with rate of reaction dependent only upon the mature mRNA level in the nucleus. The reaction was represented as,

$$R6: cap\_mRNA\_AAAn \xrightarrow{k_6} mRNA$$

mRNA degradation – mRNA degradation can take place through multiple pathways. Initially, 5’ cap and 3’ poly(A) tail are removed by decapping enzyme and deadenylase respectively. The remaining RNA strand is then degraded by endo- and exo- nucleases (Beelman and Parker 1995). The detailed model for mRNA degradation was developed by Cao and Parker (Cao and Parker 2001). The model considered 60-mer polyA tail which was deadenylated in units of 10-mer. The detailed mechanism of mRNA

degradation is important for the comprehensive model. As all the deadenylation and de-capping reactions are catalysed by highly processive enzyme these were represented as irreversible reactions. Thus, in the context of the mathematical model the first deadenylation reaction would effectively remove the functional mRNA molecules. Therefore, the reaction of mRNA degradation was represented as a single first order degradation reaction in this model.

$$R7 : mRNA \xrightarrow{k_7} \phi$$

The model accounts for competition and protection by ribosome as only free mRNA can undergo degradation reaction. Binding of other RNA binding proteins that protect mRNA from degradation can thus be easily incorporated in this model. To include degradation due to RNA binding proteins, additional component such as modified mRNA having higher degradation efficiency can be included. However, conversion to modified mRNA would be equivalent to effective removal of translationally active mRNA unless the reaction is reversible.

In this model nuclear mRNA degradation was not considered.

Translation – A detailed quantitative model of translation is developed by You, Coghill et al (You, Coghill et al. 2010) and Dimelow and Wilkinson (Dimelow and Wilkinson 2009). The reactions and parameters from the model by Dimelow and Wilkinson were used in this model. However, the model by Dimelow and Wilkinson focused on protein synthesis by one type of mRNA and did not consider presence of other mRNA in a cell. Under physiological conditions, multiple mRNA molecules each coding for different proteins are present in a cell simultaneously. Therefore, the translation machinery is shared by all the mRNA molecules, reducing the effective concentration of translational apparatus available for each mRNA. Therefore, addition of multiple mRNA components would be an appropriate representation to account for low molecule number effect. Thus in this model, competition by other mRNA components (*mRNAr*) for translation machinery was considered. The reactions for mRNA translation were represented as follows.

$$R8 : eIF2GDP + eIF2B \xrightarrow{tk_{1f}} eIF2GDP\_eIF2B$$

$$R9: eIF2GDP\_eIF2B \xrightarrow{tk_{1r}} eIF2GDP + eIF2B$$

$$R10: eIF2GDP\_eIF2B \xrightarrow{tk_{2f}} eIF2GTP + eIF2B$$

$$R11: eIF2GTP + eIF2B \xrightarrow{tk_{2r}} eIF2GDP\_eIF2B$$

$$R12: eIF2GTP + tRNA \xrightarrow{tk_{3f}} eIF2GTP\_tRNA$$

$$R13: eIF2GTP\_tRNA \xrightarrow{tk_{3r}} eIF2GTP + tRNA$$

$$R1\ \ : eIF1 + eIF3 + eIF5 \xrightarrow{tk_{4f}} eIF1\_eIF3\_eIF5$$

$$R1\ \ : eIF1\_eIF3\_eIF5 \xrightarrow{tk_{4r}} eIF1 + eIF3 + eIF5$$

$$R1\ \ : eIF2GTP\_tRNA + eIF1\_eIF3\_eIF5 \xrightarrow{tk_{5f}} MFC$$

$$R1\ \ : MFC \xrightarrow{tk_{5r}} eIF2GTP\_tRNA + eIF1\_eIF3\_eIF5$$

$$R18: MFC + eIF1A + r40s \xrightarrow{tk_{6f}} c1$$

$$R19: c1 \xrightarrow{tk_{6r}} MFC + eIF1A + r40s$$

$$R20: eIF4E + eIF4G \xrightarrow{tk_{7f}} eIF4E\_eIF4G$$

$$R21: eIF4E\_eIF4G \xrightarrow{tk_{7r}} eIF4E + eIF4G$$

$$R2\ \ : eIF4E\_eIF4G + mRNA + PABP \xrightarrow{tk_{8f}} eIF4E\_eIF4G\_mRNA\_PABP$$

$$R23: eIF4E\_eIf4G\_mRNA\_PABP \xrightarrow{tk_{8r}} eIF4E + eIF4G + mRNA + PABP$$

$$R2\ \ : eIF4E\_eIF4G + mRNAr + PABP \xrightarrow{tk_{8f}} eIF4E\_eIF4G\_mRNAr\_PABP$$

$$R25: eIF4E\_eIF4G\_mRNAr\_PABP \xrightarrow{tk_{8r}} eIF4E\_eIF4G + mRNAr + PABP$$

$$R26: eIF4E\_eIF4G\_mRNA\_PABP + eIF4A + eIF4B \xrightarrow{tk_{9f}} c2$$

$$R27: c2 \xrightarrow{tk_{9r}} eIF4E\_eIF4G\_mRNA\_PABP + eIF4A + eIF4B$$

$$R28: eIF4E\_eIF4G\_mRNAr\_PABP + eIF4A + eIF4B \xrightarrow{tk_{9f}} c2r$$

$$R29: c2r \xrightarrow{tk_{9r}} eIF4E\_eIF4G\_mRNAr\_PABP + eIF4A + eIF4B$$

$$R30: c1 + c2 \xrightarrow{tk_{10f}} c3$$

$$R31: c3 \xrightarrow{tk_{10r}} c1 + c2$$

$$R32: c1 + c2r \xrightarrow{tk_{10f}} c3r$$

$$R33: c3r \xrightarrow{tk_{10r}} c1 + c2r$$

$$R34: c3 + eIF5B \xrightarrow{tk_{11f}} c3\_eIF5B$$

$$R35: c3\_eIF5B \xrightarrow{tk_{11r}} c3 + eIF5B$$

$$R36: c3r + eIF5B \xrightarrow{tk_{11f}} c3r\_eIF5B$$

$$R37: c3r\_eIF5B \xrightarrow{tk_{11r}} c3r + eIF5B$$

$$R38: c3\_eIF5B + r60s \xrightarrow{tk_{12f}} r40s\_mRNA\_r60s + eIF2GDP + tRNA + eIF1 + eIF3 + eIF5 + eIF1A + eIF4E + eIF6G + PABP + eIF4A + eIF4B + eIF5B$$

$$R39: c3r\_eIF5B + r60s \xrightarrow{tk_{12f}} r40s\_mRNAr\_r60s + eIF2GDP + tRNA + eIF1 + eIF3 + eIF5 + eIF1A + eIF4E + eIF6G + PABP + eIF4A + eIF4B + eIF5B$$

$$R40: r40s\_mRNA\_r60s \xrightarrow{k_8} r40s + mRNA + r60s + protein$$

$$R41: r40s\_mRNAr\_r60s \xrightarrow{k_9} r40s + mRNAr + r60s + proteinr$$

To calculate *mRNAr*, total mRNA content of a cell was considered. There are different reports regarding the total mRNA content in yeast cell. The value ranges from ~12,000 to 60,000 (Von Der Haar 2008; Zenklusen, Larson et al. 2008). The highest number of mRNA molecules of one type was observed to be ~3000 molecules/cell (Schwanhausser, Busse et al. 2011). We have considered 30000 molecules of other mRNA species (*mRNAr*) as a value near mean of the reported values. This ensured that even with addition of one more type of mRNA the total mRNA content would remain within the reported range.

Similar to mRNA elongation, codon wise amino acid addition can be incorporated in case of a known protein.

Protein degradation – Protein degradation is a highly regulated process. There are multiple post-translational mechanisms which mark the protein for degradation. Ubiquitination is one important post-translational modification associated with protein degradation. Once a protein molecule is committed for degradation, it is degraded by the 26s proteasome complex into peptides. The process of tagging for degradation and degradation varies from protein to protein depending upon the external and internal

stimulus for degradation, and environmental and physiological conditions (Ravid and Hochstrasser 2008). Therefore, the effective half life of protein was considered to represent protein degradation reaction in the generic model. The reaction for protein degradation was represented as a first order reaction.

$$R42: protein \xrightarrow{k_{10}} \phi$$

Thus, the complete model for eukaryotic gene expression comprised of 42 reactions and 41 components was developed. Figure 4-6 diagrammatically represents the major processes considered in the model, source mathematical models and experimental data for reactions and parameter values[1].

[1] The parameter values for specific rate of mRNA and protein synthesis and degradation were obtained from experimental study by Schwanhausser, Busse et al. A corrigendum has been published in February, 2013 Schwanhausser, B., D. Busse, et al. (2013). "Corrigendum: Global quantification of mammalian gene expression control." Nature **495**(7439): 126-127., stating systematic underestimation of protein molecular abundance and derived specific translation rate due to erroneous scaling factor. For quantitative analysis the simulations are required to be performed with new parameter set. However, as the error is systematic, the relative timescale of fluctuation of regulators and effecter protein would remain same. Therefore, the qualitative nature of the results would remain similar.

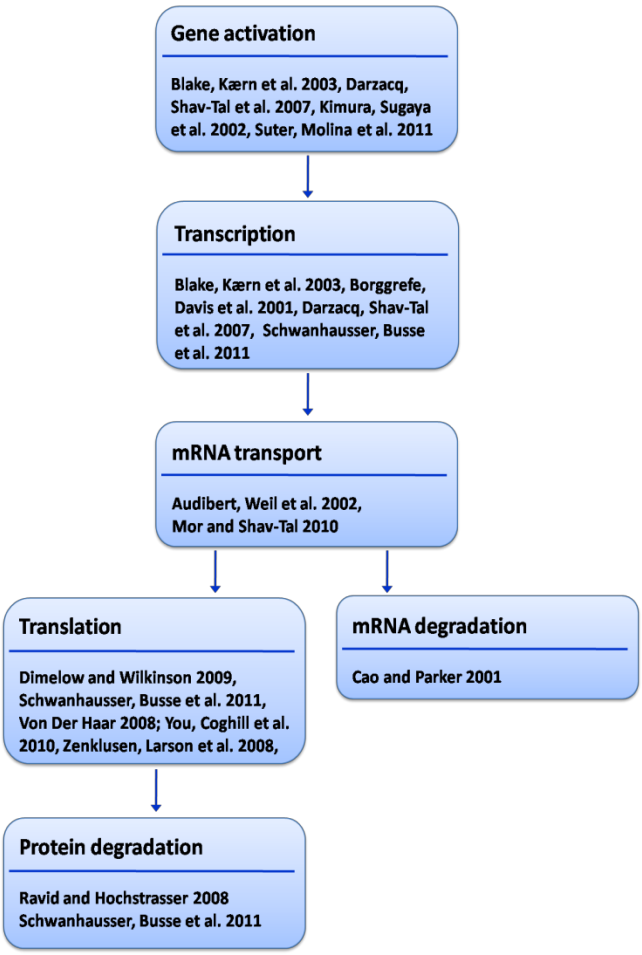


Figure 4-6 – Major steps considered in detailed model with the source mathematical models and experimental data for reactions and parameter values

The rate expressions used for each of the 42 reactions and the parameter values are summarized in Table 4-4.

Table 4-4 – The reaction rate expressions and parameter values used in the model

Rate expressions and parameter and molecular abundance values for reactions R8 to R39 were used from the model by Dimelow and Wilkinson.

	Rate expression	**Reaction rate parameter values and Reference**
R1	$k_1 \times Gene \times Reg$	k_1, 0.025 – 0.0005 molecule^{-1}min^{-1} (Suter, Molina et al. 2011)

R2	$k_2 \times GeneActive$	k_2, 1 – 0.002 min^{-1} (Suter, Molina et al. 2011)
R3	$k_3 \times GeneActive \times RNApol$	k_3, 0.18 $molecule^{-1}min^{-1}$ (Blake, Kærn et al. 2003)
R4	$k_4 \times GeneActive\_RNApol$	k_4, 0.3420 min^{-1} (Blake, Kærn et al. 2003)
R5	$k_5 \times GeneActive\_RNApol$	k_5, 0.0054 – 40.98 min^{-1} (Schwanhausser, Busse et al. 2011)
R6	$k_6 \times cap\_mRNA\_AAAn$	k_6, 12 min^{-1} (Cao and Parker 2001)
R7	$k_7 \times mRNA$	k_7,$7.2*10^{-3}$ – $2.85*10^{-4}$ min^{-1} (Schwanhausser, Busse et al. 2011)
R8	$tk_{1f} \times eIF2GDP \times eIF2B$	tk_{1f}, 0.0065 $molecule^{-1}min^{-1}$
R9	$tk_{1r} \times eIF2GDP\_eIF2B$	tk_{1r}, 7356 min^{-1}
R10	$tk_{2f} \times eIF2GDP\_eIF2B$	tk_{2f}, 504 min^{-1}
R11	$tk_{2r} \times eIF2GTP \times eIF2B$	tk_{2r}, 0.0830 $molecule^{-1}min^{-1}$
R12	$tk_{3f} \times eIF2GTP \times tRNA$	tk_{3f}, 0.0219 $molecule^{-1}min^{-1}$
R13	$tk_{3r} \times eIF2GTP\_tRNA$	tk_{3r}, 4134 min^{-1}
R14	$tk_{4f} \times eIF1 \times eIF3 \times eIF5$	tk_{4f}, $4.52*10^{-7}$ $molecule^{-2}min^{-1}$
R15	$tk_{4r} \times eIF1\_eIF3\_eIF5$	tk_{4r}, 1158 min^{-1}
R16	$tk_{5f} \times eIF2GTP\_tRNA \times eIF1\_eIF3\_eIF5$	tk_{5f}, 0.0374 $molecule^{-1}min^{-1}$
R17	$tk_{5r} \times MFC$	tk_{5r}, 5220 min^{-1}
R18	$tk_{6f} \times MFC \times eIF1A \times r40s$	tk_{6f}, $2.03*10^{-7}$ $molecule^{-2}min^{-1}$
R19	$tk_{6r} \times c1$	tk_{6r}, 438 min^{-1}

R20	$tk_{7f} \times eIF4E \times eIF4G$	tk_{7f}, 0.0046 molecule^{-1}min^{-1}
R21	$tk_{7r} \times eIF4E\_eIF4G$	tk_{7r}, 6552 min^{-1}
R22	$tk_{8f} \times eIF4E\_eIF4G \times mRNA \times PABP$	tk_{8f}, 2.6337*10^{-7} molecule^{-2}min^{-1}
R23	$tk_{8r} \times eIF4E\_eIF4G\_mRNA\_PABP$	tk_{8r}, 4008 min^{-1}
R24	$tk_{8f} \times eIF4E\_eIF4G \times mRNAr \times PABP$	tk_{8f} as above
R25	$tk_{8r} \times eIF4E\_eIF4G\_mRNAr\_PABP$	tk_{8r} as above
R26	$tk_{9f} \times eIF4E\_eIF4G\_mRNA\_PABP \times eIF4A \times eIF4B$	tk_{9f}, 4.6934*10^{-7} molecule^{-2}min^{-1}
R27	$tk_{9r} \times c2$	tk_{9r}, 1398 min^{-1}
R28	$tk_{9f} \times eIF4E\_eIF4G\_mRNAr\_PABP \times eIF4A \times eIF4B$	tk_{9f} as above
R29	$tk_{9r} \times c2r$	tk_{9r} as above
R30	$tk_{10f} \times c1 \times c2$	tk_{10f}, 0.0804 molecule^{-1}min^{-1}
R31	$tk_{10r} \times c3$	tk_{10r}, 522 min^{-1}
R32	$tk_{10f} \times c1 \times c2r$	tk_{10f} as above
R33	$tk_{10r} \times c3r$	tk_{10r} as above
R34	$tk_{11f} \times c3 \times eIF5B$	tk_{11f}, 0.0643 molecule^{-1}min^{-1}
R35	$tk_{11r} \times c3\_eIF5B$	tk_{11r}, 456 min^{-1}
R36	$tk_{11f} \times c3r \times eIF5B$	tk_{11f} as above
R37	$tk_{11r} \times c3r\_eIF5B$	tk_{11r} as above
R38	$tk_{12f} \times c3\_eIF5B \times r60s$	tk_{12f}, 0.0808 molecule^{-1}min^{-1}
R39	$tk_{12f} \times c3r\_eIF5B \times r60s$	tk_{12f}, as above

R40	$k_8 \times r40s\_mRNA\_r60s$	k_8, $5*10^{-4}$ – 558.48 min^{-1} (Schwanhausser, Busse et al. 2011)
R41	$k_9 \times r40s\_mRNAr\_r60s$	k_9, average k_8 value, 279.24 min^{-1}
R42	$k_{10} \times protein$	k_{10}, $2.36*10^{-2}$-$2.01*10^{-6}$ min^{-1} (Schwanhausser, Busse et al. 2011)

4.2.3.2 Addition of the molecular mechanisms corresponding to extrinsic sources of noise considered in the detailed model did not explain noise saturation

Previously developed simpler models of gene expression viz., 4-reaction and 6-reaction models showed a single decreasing trend of steady state protein noise and did not explain saturation of noise at high protein abundance. The noise floor was attributed to fluctuations extrinsic to these models. The detailed model developed in this study considered certain sources of extrinsic noise such as binding of regulator protein, RNA polymerase, and translational machinery. It also considered competition by other mRNA components in a cell. The detailed model was used to investigate whether these sources of noise explain the saturation of noise. Using the parameter values obtained from the study Schwanhausser, Busse et al, stochastic simulations of the detailed model were carried out as described in Methodology section. Figure 4-7 shows the graph of steady state $\%CV^2$ vs. mean protein abundance. However, in the graph only a single decreasing trend of noise with protein abundance was observed.

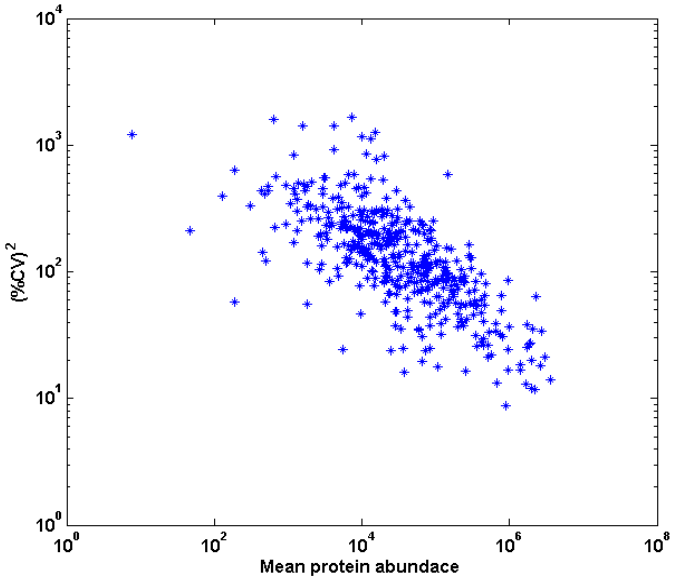


Figure 4-7 – Steady state %CV^2 vs. mean protein abundance for the detailed model of gene expression

Graph represents 890 data points.

As compared to the 4-reaction model of gene expression the overall noise level at high protein abundance was observed to be higher in 42-reaction model. However, the sources of noise considered in the detailed model were not found to be responsible for the noise floor observed at high protein abundance.

It has been theoretically shown that addition of distribution of parameters would lead to noise floor (Taniguchi, Choi et al. 2010). Stochastic simulations were performed using 4-reaction model of gene expression using different types of distributions such as normal, exponential, and gamma, for specific rate of mRNA synthesis. To perform these simulations, a distribution with 1000 data points having the same mean as the value of parameter in the corresponding original parameter set, was generated. One parameter value from the distribution was used for one run of the simulation. Thus 1000 different parameter values were used for sample of 1000 runs. Figure 4-8 shows the graphs of steady state %CV^2 vs. mean protein abundance for different types of distributions.

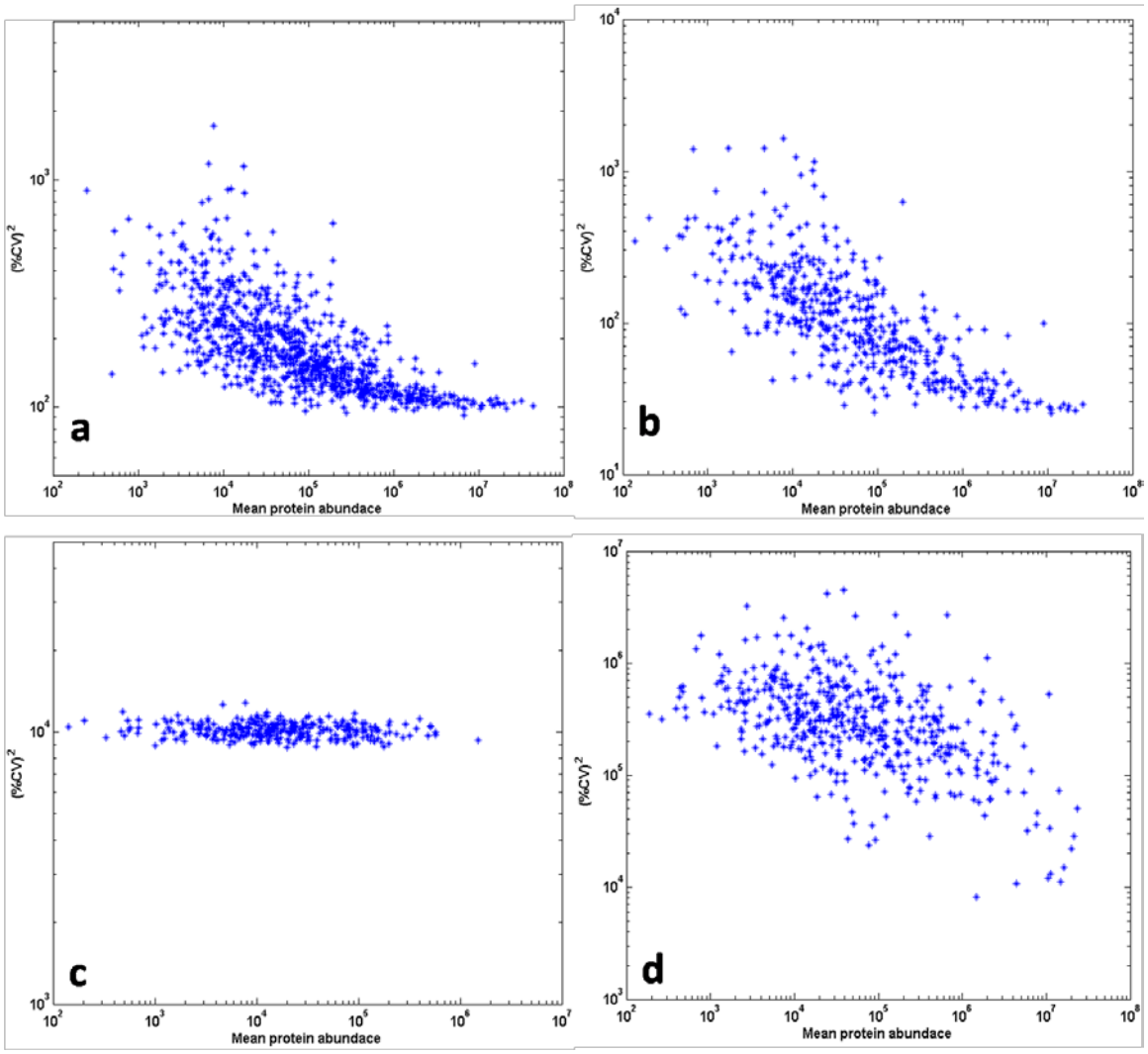


Figure 4-8 – Steady state %CV^2 vs. mean protein abundance for the 4-reaction model of gene expression with distribution of specific rate of mRNA synthesis

(a) normal distribution with standard deviation 0.1% of mean parameter value (1000 data points), (b) normal distribution with standard deviation 0.05% of mean parameter value (500 data points), (c) exponential distribution (300 data points), (d) gamma distribution (500 data points)

It was observed that the nature and the level of noise saturation were determined by the nature of parameter distribution. In case of normal distribution with standard deviation 0.1% of the mean value (Figure 4-8a), %CV^2 saturation was observed at 100. Similarly, in case of normal distribution with standard deviation 0.05% of the mean value (Figure 4-8b), %CV^2 saturation was observed at 25. In case of exponential distribution (Figure

4-8c), mean and standard deviation are equal, resulting in saturation at 10^4. In case of gamma distribution (Figure 4-8d), a saturating trend was not observed. In case of normal distribution with standard deviation of 0.1% and 0.05% of the mean value, noise trend was observed to be qualitatively similar to the experimentally observed data. As mentioned in the Methodology section, as a quantitative measure of saturation, ratio of the slopes was calculated. The ratio values were found out to be 0.5 and 0.77 for normal parameter distribution with standard deviation of 0.1% and 0.05% of the mean, respectively.

In case of 42-reaction model, *Reg* (regulator protein considered in reaction R1) was considered as a parameter to add extrinsic noise. *Reg* being a regulator protein, a 4-reaction model of gene expression was used to obtain a distribution with the same mean as the previously defined *Reg* value. The resultant steady state distribution of *Reg* with 1000 abundance values was used in the detailed model to simulate a sample of 1000 runs. Similar to the simulations of 4-reaction model with parameter distribution, one *Reg* abundance value from *Reg* distribution was used for one run of simulation. The steady state distribution of protein obtained using 4-reaction model of gene expression was shown to be gamma distribution (Shahrezaei and Swain 2008). Thus using *Reg* distribution can be considered equivalent to using a gamma distribution of parameters. Figure 4-9 shows the graph of steady state $\%CV^2$ vs. mean protein abundance for detailed model with *Reg* distribution as extrinsic source of noise. In this case a saturating trend could not be observed. However, $\%CV^2$ did not decrease below 10^3, in contrast with the 42-reaction model without parameter distribution (Figure 4-7).

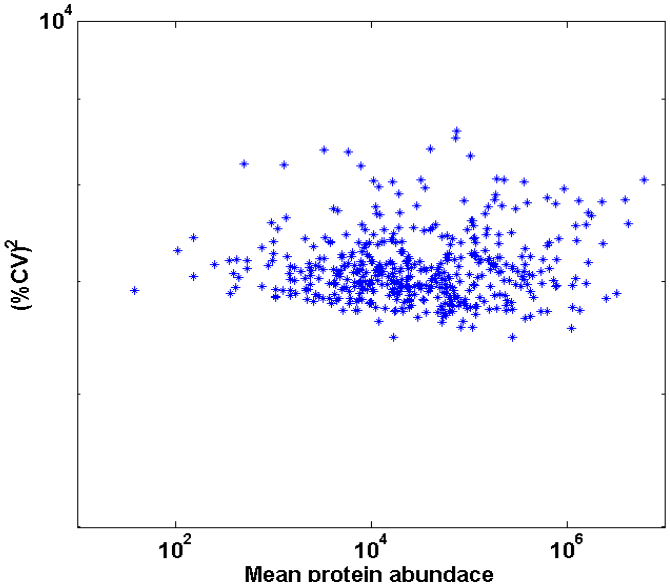


Figure 4-9 – Steady state %CV^2 vs. mean protein abundance for the detailed model of gene expression with distribution of *Reg*

The distribution of *Reg* was obtained using 4-reaction model of gene expression. The graph shows 450 data points.

In a cellular system, all the regulatory proteins are synthesized and degraded through the same gene expression process. Therefore, instead of the resultant steady state distribution, 4-reactions of regulator mRNA and protein synthesis and degradation were incorporated in the detailed model. It was thought that the complete 46 reaction model would show saturation of noise, due to addition of gene expression extrinsic noise. However, to our surprise, in this case a single decreasing trend of noise was observed. Figure 4-10 shows the graph of steady state %CV^2 vs. mean protein abundance.

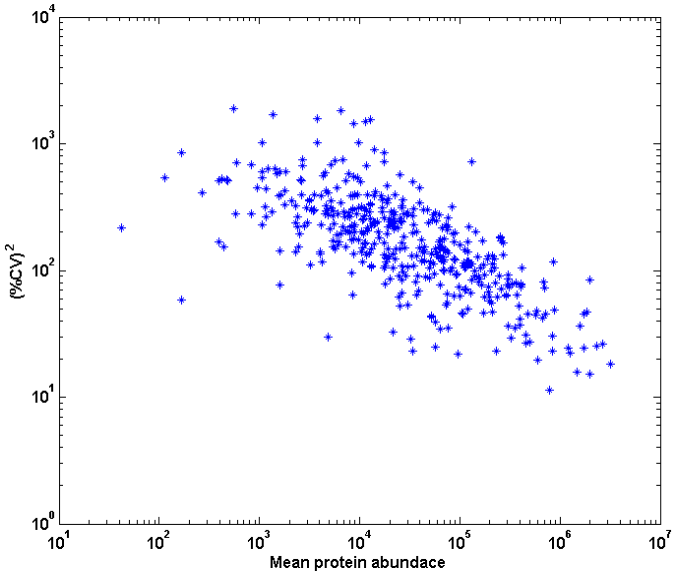


Figure 4-10 – Steady state %CV^2 vs. mean protein abundance for the detailed model of gene expression with additional 4 reactions for Reg synthesis

The graph shows 440 data points.

In the condition, where steady state distribution of *Reg* was used as a source of extrinsic noise, different but constant value of regulator was used for each run for the entire time of the simulation. In this case the value of *Reg* changed during the time course of simulation for each run. It was clear that the fluctuations produced in the 4-reactions of regulator synthesis and degradation were fast compared to the fluctuations of high abundance protein and averaged away contributing insignificant to the total noise at high protein abundance, as suggested in previous studies.

4.2.3.3 Slowest varying fluctuation could explain saturation of noise at high protein abundance

It has been suggested that fluctuations of time scale comparable to cell cycle duration are important for population variability (Rosenfeld, Young et al. 2005). Therefore, to investigate the effect of time scale of fluctuation, proteins with different mixing time (τ_m) were used as *Reg*. Three different types of protein with τ_m equal to cell cycle duration, slightly greater than 2-cell cycle duration and around 30 times cell cycle duration were selected. The protein with 30-cell cycle duration was a slowest fluctuating protein in the

dataset. Figure 4-11 shows the time course profiles for the three proteins used as *Reg*, for visual comparison of time scale of mixing. In case of protein with τ_m equal to 30 cell cycle duration the protein levels were not observed to be mixing for a long time (Figure 4-11c).

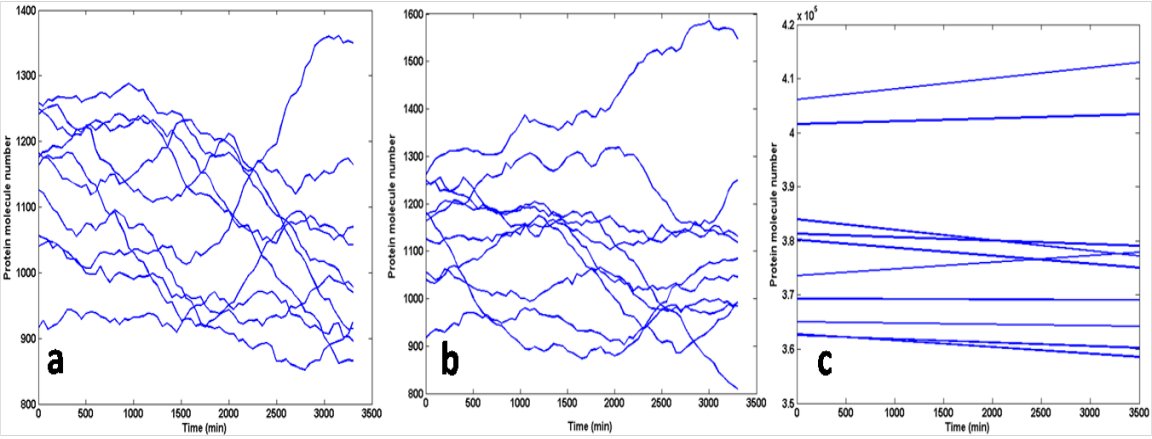


Figure 4-11 – Time course profiles of regulator protein for 10 runs

(a) *Reg* with τ_m equal to one cell cycle duration (b) *Reg* with τ_m greater than cell cycle duration (c) *Reg* with τ_m around 30 cell cycle duration

Instead of detailed model, an 8-reaction model of gene expression was used (Figure 4-12). In this reaction system 4-reaction model of gene expression was used for both the effecter protein and the regulator. Steady state distribution of each of the 4 components viz., regulator mRNA (*Regm*), regulator protein (*Regp*), effecter mRNA (*m*) and effecter protein (*p*) was used as initial condition, in order to investigate the effect of the added extrinsic noise on the final distribution of effecter protein. The reaction rate parameter of mRNA synthesis was changed such that the mean value of new pseudo-first order reaction rate <*k'*Regp*> was numerically equal to the previous zero order reaction rate parameter. This ensured same steady state mean values of effecter mRNA and protein.

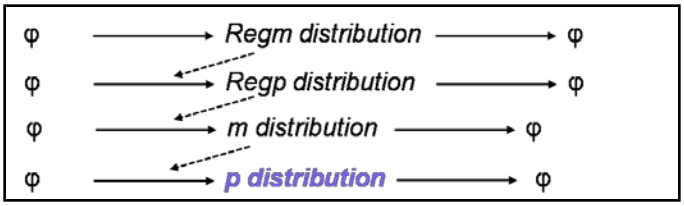


Figure 4-12 – Reaction system considering 4-reaction model of gene expression each for Reg protein and effecter protein

The simulations were initially carried out for 1 cell cycle duration (~1650 min) (Schwanhausser, Busse et al. 2011). However, a single decreasing trend of noise was observed in case of all the three types of *Reg*.

It has been reported in experimental study that, even in the presence of cell division events, the protein level continues to increase or decrease to reach its steady state level, spanning multiple cell cycle durations (Kondo, Mori et al. 1997; Elowitz and Leibler 2000). Considering cell division to be synchronized and deterministic event, its contribution was not considered in this study. In addition, even in synchronized gated population of cells, variability was observed, indicating cell-intrinsic source of variability.

Simulations with 8-reaction model were performed for longer duration. It was observed that the noise increased as the duration of simulation was increased and the value reached almost a constant level after around 20 cell cycle duration. This indicated that though the mean value was the same, the distribution reached steady state after around 20 cell cycle duration. Figure 4-13 shows the graph of steady state $\%CV^2$ vs. mean protein abundance at the end of 30 cell cycle duration for three types of *Reg*.

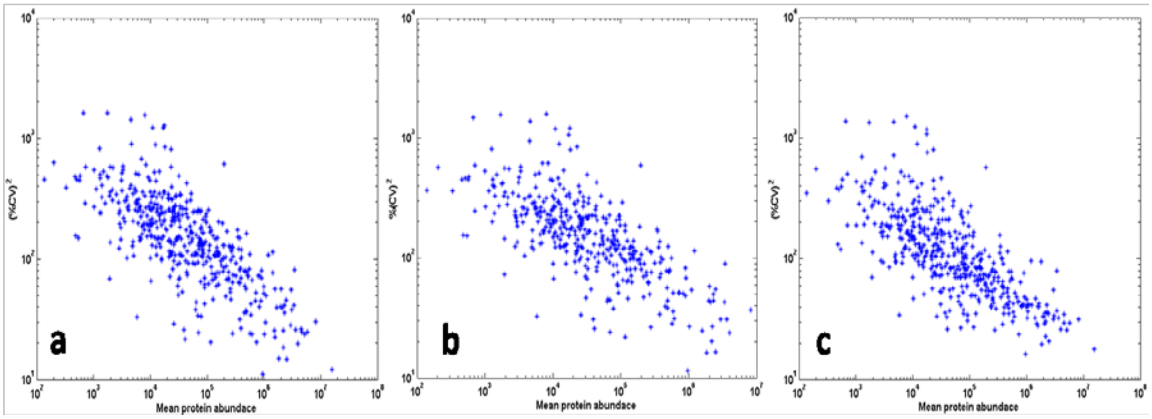


Figure 4-13 – Steady state %CV^2 vs. mean protein abundance for the 8-reaction system

(a) *Reg* with τ_m equal to cell cycle duration (b) *Reg* with τ_m greater than two cell cycle duration (c) *Reg* with τ_m 30 cell cycle duration (500 data points for each *Reg*)

A single degreasing trend was observed for *Reg* with τ_m equivalent to cell cycle and that greater than cell cycle (Figure 4-13 a and b). In case of slowest fluctuating protein as *Reg*, slightly lesser decrease in noise was observed at high protein abundance (Figure 4-13c). A clear saturating trend was not observed however, %CV^2 did not decrease below 15. Hence, it was evident that fluctuation time scale compared to cell cycle duration or greater than cell cycle duration was not the source of variability at high expression level. Such fluctuations are faster compared to other slow varying effecter proteins and may average away during the time period of simulations, producing little variability. This also indicated that the time of measurement of noise as compared to the mixing time was also an important determinant of observed variability. The simulation with slowest varying protein as *Reg*, were continued for time duration of around 30 times that of mixing time. It was observed that the saturation was more clearly evident in the simulations for longer time (Figure 4-14). The ratio was found out to be 0.85 at the time duration of five times τ_m. The observed saturation was qualitatively similar to the experimental observation. However, the ratio was significantly greater than that for experimental data. Larger parameter sample set would lead to improvement in linear fit. The running median can also be calculated to clearly determine the trend and the ratio of slopes obtained from running median data can be compared with the experimental observation.

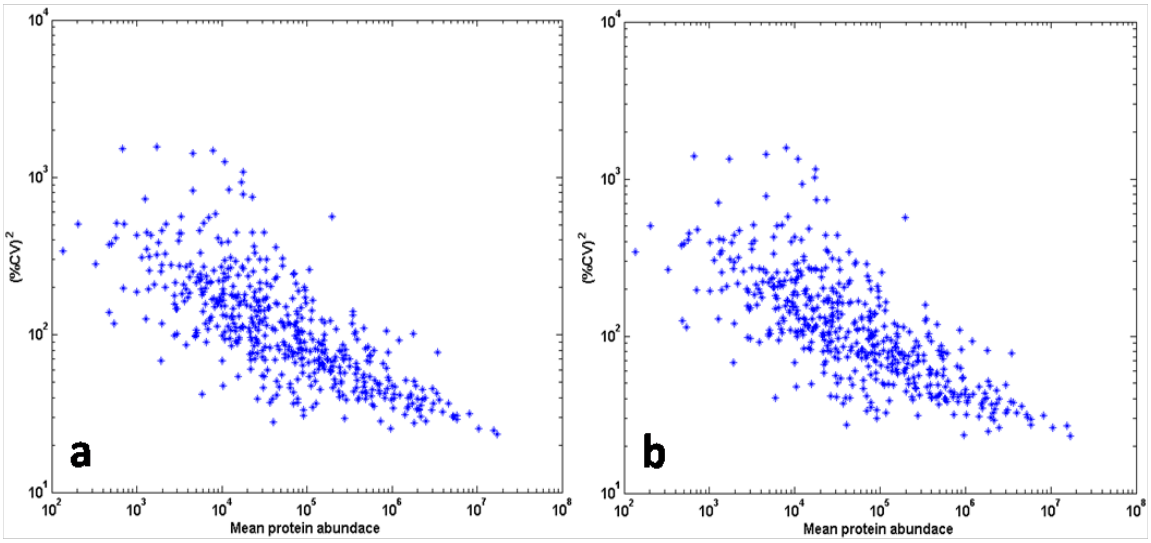


Figure 4-14 – Steady state %CV^2 vs. protein abundance for the 8 reaction system with slowest fluctuating protein as *Reg*

For the final time (a)5 times τ_m (b) 30 times τ_m (500 data points)

The ratio value was observed to increase with longer simulation duration. The ratio was found out to be 0.92 at simulation time 30 times that of τ_m (Figure 4-14 b). As suggested in the previous studies, the slow fluctuations may average away during the time period required for slower varying proteins to reach the steady state. The time of measurement of noise was also an important factor determining the observed saturation as during the long time of simulation the slowest varying fluctuations could also average away resulting in no saturation. The ratio of slopes was measured for different time points. Figure 4-15 shows the value of ratio of slopes for different time points compared to τ_m.

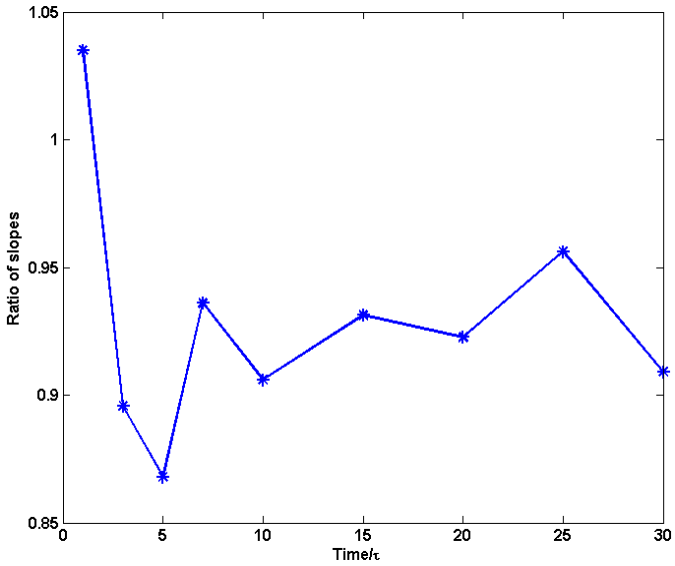


Figure 4-15 – Graph showing ratio of slopes at different time duration as compared to the mixing time (τ_m) of *Reg*

It was observed that the ratio decreased till the time period up to 5 times of τ_m after which the ratio was observed to increase. The minimum value of the ratio was observed at time 5 times of τ_m. During the simulations, the slowest varying fluctuations were also observed to average to some extent. This indicated that, the time of measurement of the noise was also an important factor determining the observed noise floor.

It was observed that the model with regulator as slowest varying protein could show noise floor. The results are qualitatively similar to the 4-reaction model with imposed extrinsic noise, using parameter distribution. The 4-reaction model with distribution of specific rate of mRNA synthesis showed saturation of noise. However the ratio of slopes was greater than that for experimental data (2.3 times the ratio of slopes of experimental data). It was observed that only the 8-reaction model with slowest varying extrinsic noise could show saturation of noise closer to that showed by 4-reaction model with parameter distribution (1.7 times the ratio for 4-reaction model). However, the ratio value was much higher than that observed for experimental data (4 times). Addition of multiple such types of regulation would increase the gene extrinsic noise leading to closer match to the observed extrinsic noise limit. Thus, the mixing time of *Reg* compared to that of the

effecter protein and not to the cell cycle duration was found to be important in determining saturation of noise.

In this study, it was observed that extrinsic fluctuations having the slowest time scale of fluctuation can contribute to observed saturation of noise to some extent. As suggested in the previous studies the fluctuations in the concentrations of RNA polymerase, ribosomes can contribute to the observed saturation provided the time scale of fluctuations is slower than the high abundance effecter proteins. Multiple such gene expression-extrinsic but cell-intrinsic sources of noise can lead to saturation of noise with the noise floor at a value observed in single cell global protein measurement experiments. Some mathematical models have coupled deterministic gene expression with stochastic cell division event (Marathe, Bierbaum et al. 2012). Consideration of gene expression as deterministic would be a good approximation for high-abundance proteins. These mathematical models have shown variability in populations due to stochastic cell division and random partitioning of proteins (Thattai and Van Oudenaarden 2001; Marathe, Gomez et al. 2012), suggesting one source of noise contributing to the observed extrinsic noise limit. To explain variability in gated population of cells, cell-intrinsic sources such as the one studied in this model are required to be incorporated.

In this study we have proposed one possible mechanism that can contribute to the extrinsic noise leading to saturation of noise at high abundance. Cell intrinsic noise of slowest time scale of fluctuation that is considered to be extrinsic to gene expression, can show saturation at high protein abundance level.

4.2.4 Conclusion

As a first step towards the development of a detailed model of gene expression, previously developed mathematical models of processes and regulation of gene expression were reviewed. Starting from a simple model of gene expression a detailed model of gene expression was developed using a modular approach. The entire model was divided into six modules each for one major sub-process. Such modular arrangement facilitates the use of a single module independent of other reactions. The model contains multiple simplifications of the processes and modifications to include details of reactions

are possible such as incorporation of details of mRNA transport, nuclear mRNA decay, and mRNA and protein degradation. However, many processes in gene expression are gene specific and vary with environmental and physiological conditions. In order to make the model generic such details could not be included. For the use of the model to study a specific protein such details can easily be included in any of the modules of the model.

As the detailed model considered certain cell-intrinsic sources of noise, it was used to investigate whether these sources can explain saturation at high protein abundance. It was observed that, though these noise sources contribute to total noise, these were not responsible for observed saturation. It was observed that, fluctuations slower than the effecter proteins in a cell could show saturation at high protein abundance. In this study we have proposed one possible source of gene-extrinsic but cell-intrinsic source of noise that can contribute to the extrinsic noise leading to saturation of noise at high protein abundance.

4.3 Section 2 – Global sensitivity analysis of gene expression revealed differential sensitivity of steps in gene expression to different measures of noise

4.3.1 Introduction

There are multiple experimental and theoretical studies investigating various aspects of noise in gene expression. In order to explore the sources of noise in gene expression, some studies have focused on finding a step in gene expression that maximally influences the noise in protein level. However, the observations regarding most important step reported in these studies are ambiguous.

One of the initial theoretical studies by Thattai and Oudenaardan (Thattai and Van Oudenaarden 2001) and supporting experimental study by Ozbudak, Thattai et al (Ozbudak, Thattai et al. 2002) using *Bacillus subtilis* compared transcription and translation and suggested that the noise in protein level was influence by translation much more than transcription. On the other hand the study using yeast cells by Blake et al (Blake, Kærn et al. 2003) stated that as opposed to prokaryotes, transcription with re-initiation affects the noise in protein level in case of eukaryotic cells. In both the studies

Fano factor, defined as a ratio of variance to mean, was used as measure of noise. Another experimental study using yeast cells by Raser and O'Shea (Raser and O'Shea 2004) investigated variability in population in terms of Fano factor by considering mRNA levels. The study compared gene activation and transcription under three different conditions of specific rates of gene activation, deactivation and transcription. Infrequent gene activation and high transcription was observed to produce larger variability in population in terms of mRNA level. In addition to these experimental studies, theoretical studies by Kierzek, Zaim et al (Kierzek, Zaim et al. 2001) on prokaryotic system compared transcription and translation and showed that transcription initiation frequency affects noise in steady state protein level more than translation initiation frequency. In this study coefficient of variation was used as a measure of noise. Another study by Komorowski, Mi kisz et al (Komorowski, Mi kisz et al. 2009) examined the contributions of regulatory factor to gene expression noise and stated that repression at translational level results in more noise than repression at transcription. In contrast to the previous study, Fano factor was used as measure of noise. Table 4-5 summarises the major output of these studies. It is evident that these studies have compared different processes in gene expression and have also used two different measures of noise viz., coefficient of variation and Fano factor. Contradictory observations regarding an important step in gene expression have been reported in these studies. Due to the comparison of different processes in gene expression and use of different measures of noise theses studies cannot be compared. This suggests a need for systematic measurement of relative contributions of each step in gene expression for different measures of noise.

Table 4-5 – Studies analysing gene expression to identify important step determining noise in protein

Reference	Process having more contribution to noise	Organism	Measure of noise
(Thattai and Van Oudenaarden 2001; Ozbudak, Thattai et al. 2002)	Translation	Prokaryote	Fano Factor

(Blake, Kærn et al. 2003)	Transcription with re-initiation	Eukaryote	Fano factor
(Kierzek, Zaim et al. 2001)	Transcription	Prokaryote	Coefficient of variation
(Komorowski, Mi kisz et al. 2009)	Translation	Not specified	Fano Factor

Sensitivity analysis is a method to measure the effect of parameter variation on the output of a system. In the stochastic framework there are different methods for stochastic sensitivity analysis. These consider sensitivity analysis based on density function sensitivity using different approaches such as Fisher information matrix (Gunawan, Cao et al. 2005; Komorowski, Costa et al. 2011), spectral polynomial chaos expansion (Kim, Debusschere et al. 2007), and path-wise derivative approach (Sheppard, Rathinam et al. 2012). All these studies consider sensitivity to infinitesimally small change or perturbation in parameter. Similar to control coefficients in deterministic systems, a mathematical framework for stochastic control analysis has been recently developed (Rocco 2009; Kim and Sauro 2010). It uses linear noise approximation to calculate sensitivity of infinitesimally small change in parameter value to noise to find out stochastic sensitivity coefficients for mean concentration, coefficient of variation and covariance. Thus almost all the studies considered local sensitivity except the study by Degasperi and Gilmore (Degasperi and Gilmore 2008) which considered Morris method and examined the sensitivity in wider parameter range using histogram distance measure.

It is known that in case of gene expression, the parameter values for each of the major steps span 2 to 3 orders of magnitude range. Thus to examine sensitivity of these steps on the steady state noise in protein level global sensitivity analysis was thought to be appropriate in order to investigate relative contribution of each step over such a wide range of values. Multiple global sensitivity analysis methods such as multiple parameter sensitivity analysis (MPSA), partial rank correlation coefficient (PRCC), Morris method, weighted average of local sensitivities (WALS) are available (Zi 2011). Among these Multiple Parameter Sensitivity Analysis (MPSA) (Hornberger and Spear 1981) is one method widely used for biological systems (Cho, Shin et al. 2003; Zi, Cho et al. 2005). In

this study, MPSA was used to perform global sensitivity analysis. Three models of gene expression viz., 4-reaction model, 6-reaction model and a 42-reaction detailed model were used for sensitivity analysis. In case of 4-reaction model the relative contribution of transcription, translation, mRNA degradation, and protein degradation on steady state noise in protein was investigated. In case of 6-reaction model, additional reactions of gene activation and gene deactivation were use. The same six reactions were examined in case of the detailed model of gene expression. In this study the sensitivity to two measures of noise, viz., coefficient of variation, and Fano factor was calculated, in order to compare differential contribution of parameters to these measures.

It was observed that the two measures of noise were indeed sensitive to different parameters. In addition, comparison of 4-reaction, 6-reaction and detailed model showed that addition of gene activation, gene deactivation reaction changed the relative contribution of reactions to noise in steady state protein level. Analysis of analytical expressions for local sensitivity of coefficient of variation and Fano factor for gene expression revealed the differential contribution of steps to these measures of noise. The numerically observed differential sensitivities could be explained from the analytical expression.

4.3.2 Methodology

4.3.2.1 Models and parameter ranges

Global sensitivity analysis was performed using three models of gene expression viz., 4-reaction model, 6-reaction model and 42-reaction model in order to explore whether addition of reactions leads to change in relative contribution of each step to noise at steady state protein level. As a representative parameter range of eukaryotic gene expression, the parameter ranges for specific rates of transcription, translation, mRNA degradation and protein degradation were obtained from Schwanhausser, Busse et al (Schwanhausser, Busse et al. 2011). The parameter range for gene activation and deactivation were obtained from experimental study by Suter, Molina et al (Suter, Molina et al. 2011). In the Schwanhausser, Busse et al dataset, majority of the parameter values were observed centered around a mean value and very few lie at the boundaries. Consideration of such outlier parameter values can result in altered sensitivity than what

is observed for a normal parameter range. Therefore, instead of considering the complete range for each parameter, log transformed mean ± 2 standard deviation range was considered. Log transformation was required as the parameter ranges covered two to three orders of magnitude. More than 95% of the data points were covered in the selected range. The parameter ranges are listed in Table 4-6.

Table 4-6 – The parameter ranges used in gene expression models

Parameter	Range of parameter values
Gene activation (min^{-1})	$1*10^{-1} – 2*10^{-3}$
Gene inactivation (min^{-1})	$1 – 2*10^{-3}$
Transcription reaction rate constant (min^{-1}, for 4-reaction model)	$4.33*10^{-3} – 2.16*10^{-1}$
mRNA degradation reaction rate constant (min^{-1})	$2.98*10^{-3} – 4.80*10^{-4}$
Translation reaction rate constant (min^{-1})	$3.97*10^{-2} – 7.86$
Protein degradation reaction rate constant (min^{-1})	$1.83*10^{-3} – 3.82*10^{-5}$

4.3.2.2 Multiple Parameter Sensitivity Analysis (MPSA)

MPSA also called as regionalised sensitivity analysis was used to perform global sensitivity analysis. In MPSA, model output for randomly generated parameter data set is evaluated. The parameter sets are classified into two classes by comparing their output with the output of the reference parameter set. Parameter sensitivity is determined by comparing the cumulative distribution function curves for each parameter sets using Kolmogorov-Smirnov test.

To perform MPSA, a parameter sample was obtained using Latin hypercube sampling method. It is more efficient method than random sampling. It has been suggested that random sample may require larger sample to cover entire range of data. To generate a sample of *n* data points within a certain range using Latin hypercube sampling, the range is divided into *n* intervals of equal size. One value is randomly selected from each

interval depending upon the probability density in the interval. This ensures that parameter ranges are evenly covered. Figure 4-16 shows plots of parameter sample generated using Latin hypercube sampling and random sampling.

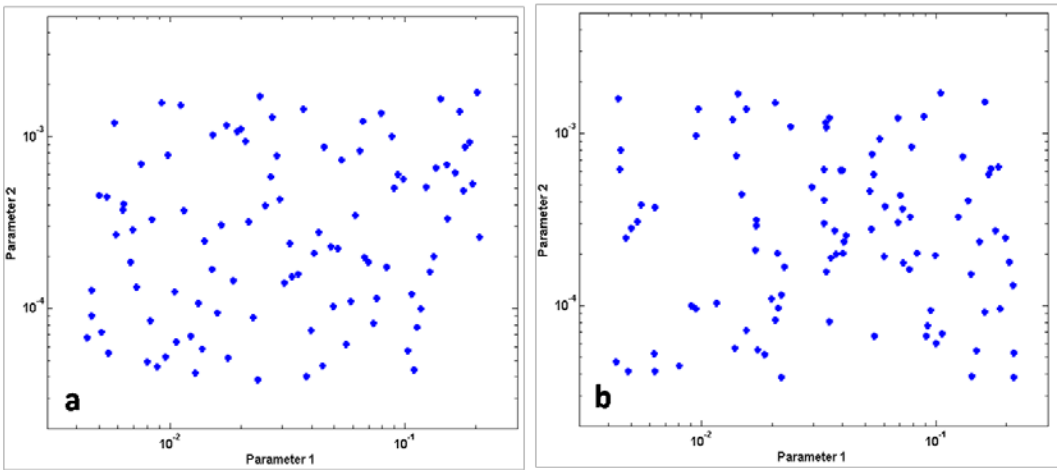


Figure 4-16 – Parameter samples generated using Latin hypercube sampling and random sampling

A sample of parameters generated using (a) Latin hypercube sampling (b) random sampling. A more uniform sample of parameters can be generated using Latin hypercube sampling method.

From Figure 4-16 it was evident that the parameter sample generated using Latin hypercube method (Figure 4-16a) was uniformly distributed over entire parameter range as compared to random sample (Figure 4-16b).

Matlab function `lhsdesign` was used to generate parameter sample. Each parameter set contained 4 parameters (mRNA synthesis, protein synthesis, mRNA degradation and protein degradation) for 4-reaction model, and 6 parameters for 6-reaction and 42-reaction model (gene activation and deactivation in addition to previously stated 4 parameters), within the parameter ranges given in Table 4-6. For other parameters in the detailed model such as gene copy number, *Reg* abundance, reaction rate parameters for binding of translation factors, parameter ranges could not be obtained and hence not considered for sensitivity analysis. One sample of parameter sets contained 100 parameter sets. Five such samples were generated ensuring different random seed for

each sample. Parameter sensitivity was obtained by considering the average sensitivity of 5 samples.

In case of specific reaction rate of mRNA synthesis for 6-reaction model, the parameter sample was generated in the same range as that for 4-reaction model. Knowing the corresponding values of gene activation and deactivation reaction rate parameters the steady state active gene level was calculated and used to set the specific rate of mRNA synthesis so as to obtain the same rate as that would be obtained in case of 4-reaction model. Thus the reaction rate parameter for mRNA synthesis in case of 6-reaction model was calculated as,

$$k_{transcription6\_raction} = \frac{k_{transcription4\_reaction}}{\frac{k_{geneActivation}}{(k_{\text{geneActivation}} + k_{\text{geneDeactivation}})}}$$

In case of 42-reaction model, the sample of specific rate of mRNA synthesis was generated in the same range as that for 4-reaction model. An equilibrium level of *GeneActive_RNApol* was used to obtain the specific rate of mRNA synthesis. To obtain an equilibrium level of *GeneActive_RNApol* first four reactions (R1 to R4, Section 1) of were considered to be in equilibrium. Thus reaction rate parameter for mRNA synthesis in case of 42-reaction model was calculated as,

$$k_{transcription\_42reaction} = \frac{k_{transcription\_4reaction}}{[GeneActive\_RNApol]}$$

In case of specific rate of mRNA degradation for 42-reaction model, the range was obtained from previously generated sample of 1000 parameter sets (Section 1).

The output was defined as the square of difference between %CV (or Fano factor) of reference parameter set and that of the sample parameter sets.

$$output_i = (\%\,\mathrm{CV}_i - \%\,\mathrm{CV}_{\text{ref}})^2$$

In case of each sample of 100 parameter sets, first parameter set was considered to be a reference set. A threshold value, defined as the ratio of summation of all output values to the number of parameter sets in the sample was obtained for each sample.

$$Threshold = \frac{\sum output_i}{sample\ size}$$

Depending upon the value of output for a parameter set being greater or lesser than the threshold, the parameter sets were classified into two groups. For each parameter the cumulative distribution functions for the two groups were obtained. The cumulative distribution functions were compared using Kolmogorov-Smirnov (KS) test at significance level of 5%. KS test calculates the maximum vertical distance between the two cumulative distribution function curves. Sensitivity of an output is determined by the distance between the two distributions. Figure 4-17 illustrates the comparison of cumulative distribution function curves based on vertical distance.

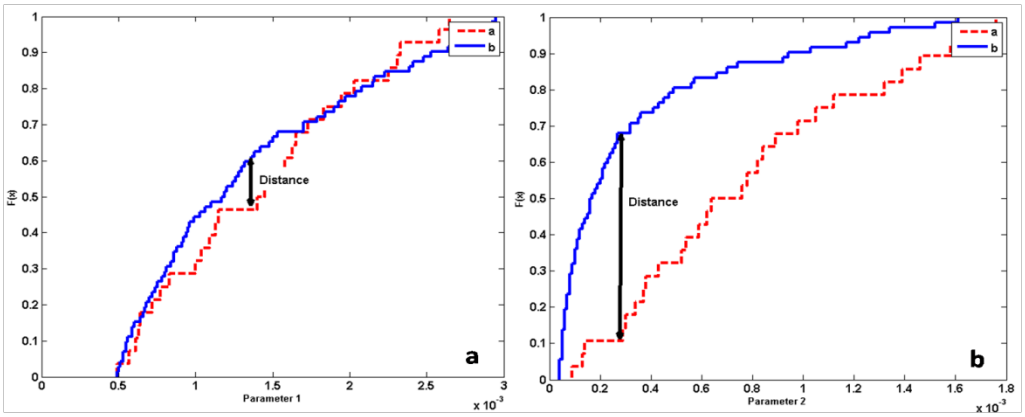


Figure 4-17 – Cumulative distribution function curves

Cumulative distribution function curves for the two groups obtained by comparison of sample output values with the predefined threshold value, for two parameters, (a) Parameter 1 and (b) Parameter 2. Maximum vertical distance between the two curves is calculated using KS test. The maximum vertical distance in case of parameter 2 is more compared to that of parameter 1 indicating significantly different distribution functions. The output is observed to be sensitive to Parameter 2.

To measure the sensitivity of each parameter a score was defined by considering the test statistics *k* weighted by the negative logarithm of p-value. The score is defined as, $k*(-log_{10}(p))$, where *k* is KS test statistics and *p* is p-value of the test. The sensitivity score was calculated for each sample and the average score of the 5 samples was used as a measure of sensitivity.

4.3.2.3 Analytical expression for local sensitivity

To understand the contribution of reaction rate parameter to sensitivity of steady state CV and Fano factor analytical expressions for local sensitivity were obtained. Local sensitivity can be defined as the change in the model output relative to infinitesimally small change in parameter value i.e. $\delta output / \delta p_i$. Using the 4-reaction model of gene expression analytical expression for steady state protein CV and Fano factor was obtained by following the framework by (Thattai and Van Oudenaarden 2001). Sensitivity coefficients were calculated by obtaining the partial derivatives of CV and Fano factor with respect to every parameter as, $\delta CV / \delta p_i$ and $\delta Fano\ Factor / \delta p_i$. Relative sensitivity coefficients $\frac{\delta output}{\delta p}.\frac{p}{output}$ were calculated to compare the sensitivity values at different parameter values. In the 4-reaction model, parameters k_1, k_2, k_3 and k_4 were defined as specific rates of mRNA synthesis, mRNA degradation, protein synthesis, and protein degradation, respectively. The expressions for sensitivity coefficients were obtained using Mathematica version 7.0.1.0 (Wolfram Research, Champaign, USA).

4.3.3 Results and Discussion

4.3.3.1 Sensitivity analysis using 4-reaction model showed maximum sensitivity of CV for transcription while that of Fano factor to translation

For 4-reaction model of gene expression, MPSA was performed for %CV and Fano factor as measures of noise. Figure 4-18 shows cumulative distribution functions of %CV for 4 parameters, transcription, mRNA degradation, translation, and protein degradation.

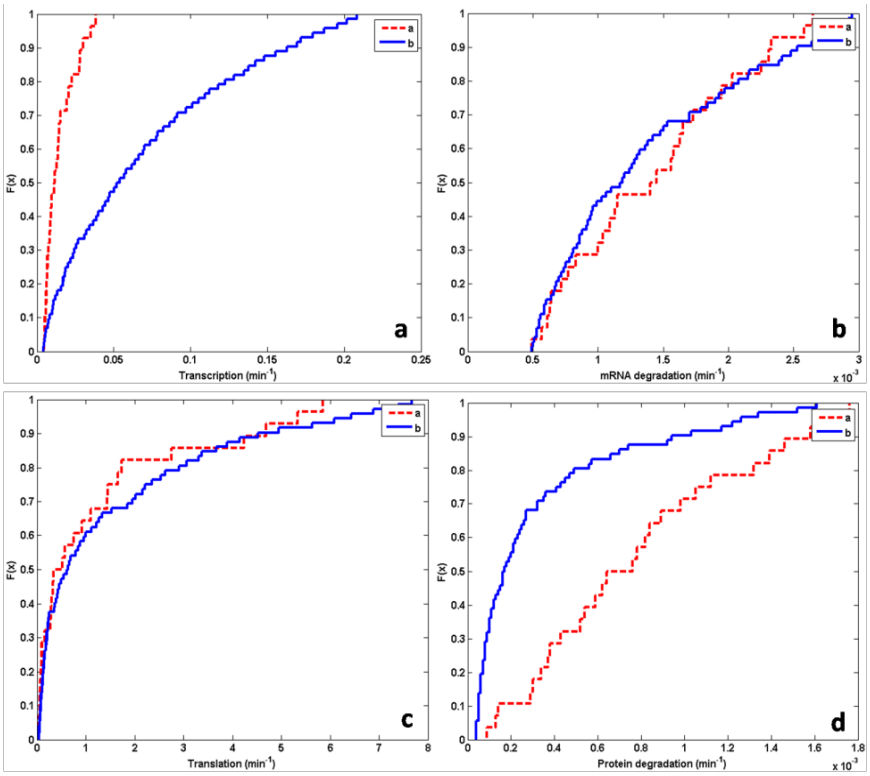


Figure 4-18 – Cumulative density function of steady state protein %CV of two groups for one sample of parameter sets for 4-reaction model of gene expression

(a) transcription, (b) mRNA degradation, (c) translation, and (d)protein degradation.

It was observed that the maximum distance between the cumulative distribution functions for two parameters mRNA degradation (Figure 4-18b) and translation (Figure 4-18c) was much lesser as compared to that for transcription (Figure 4-18a) and protein degradation (Figure 4-18d). It indicated that, the cumulative distribution functions of two groups for transcription and protein degradation were significantly different. As sensitivity is reflected by the distance between the two distributions, %CV was observed to be sensitive to these two parameters. On the other hand %CV was not observed to be sensitive to mRNA degradation and translation. Sensitivity scores are summarized in Table 4-7. From the sensitivity scores it was evident that, transcription had highest sensitivity score, 4.22 and thus has maximum influence on steady state protein CV.

Similar to MPSA for %CV, sensitivity was calculated for Fano factor. To our surprise, the sensitivity of Fano factor for the four parameters was different than that observed for

%CV. Figure 4-19 shows the cumulative distribution function of Fano factor for four parameters.

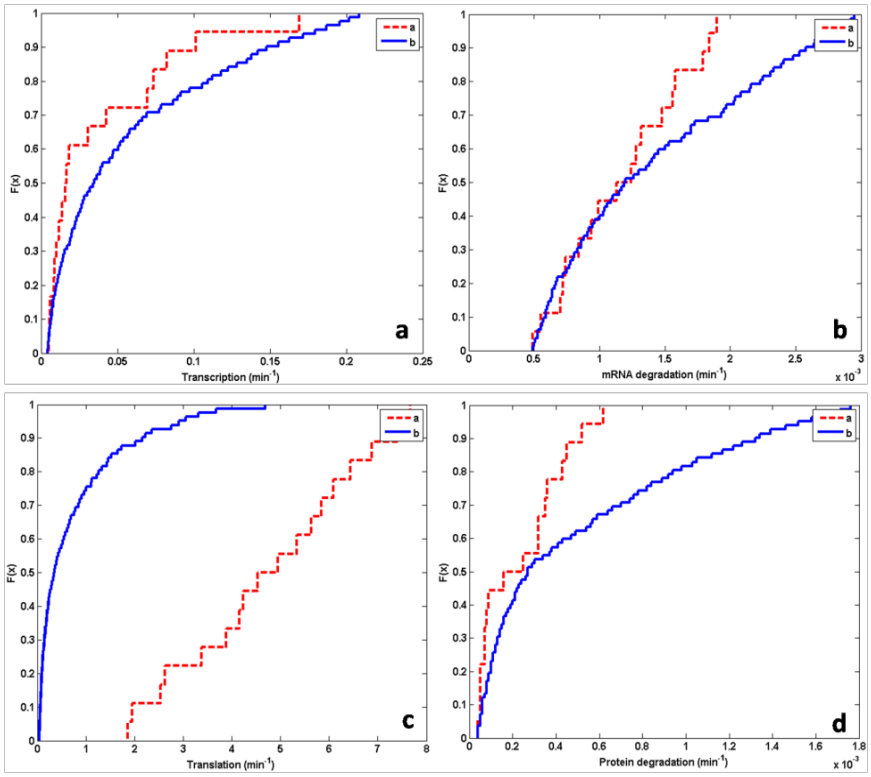


Figure 4-19 – Cumulative density function of steady state protein Fano factor of two groups for one sample of parameter sets for 4-reaction model of gene expression

(a) transcription, (b) mRNA degradation, (c) translation, and (d)protein degradation.

In this case the cumulative distribution functions of transcription (Figure 4-19a) for the two groups were observed to be similar as opposed to that observed in case of %CV. The cumulative distribution functions of translation for two groups were observed to be significantly different in this case (Figure 4-19c). Therefore, Fano factor was observed to be most sensitive to translation. The sensitivity score for Fano factor (Table 4-7), was observed to be around 10 times higher (8.67) than that for other parameters. These results are in agreement with the previously reported sensitivity of Fano factors for transcription and translation (Thattai and Van Oudenaarden 2001).

Table 4-7 – Average sensitivity score of steady state protein %CV and Fano factor for 4-reaction model of gene expression

Parameter	% CV average score	Fano factor average score
Transcription	4.22	0.11
mRNA degradation	0.10	0.81
Translation	0.09	8.67
Protein degradation	1.95	0.31

Global sensitivity analysis of gene expression using 4-reaction model revealed that the two measures of noise were in fact sensitive to different parameters. It was observed that %CV was most sensitive to transcription while Fano factor was most sensitive to translation, indicating differential sensitivity of these measures of noise to different parameters.

The sensitivity of these two measures of noise was calculated using Morris method to confirm the obtained sensitivity using MPSA. The details are given in Appendix IIIb. The sensitivity results obtained using Morris method were in agreement with the sensitivity results obtained using MPSA.

4.3.3.2 Sensitivity analysis using 6-reaction model showed maximum sensitivity of CV for protein degradation and maximum sensitivity of Fano factor to translation

To examine whether addition of reactions changes their relative contribution to noise, sensitivity was calculated using 6-reaction model of gene expression. The model contains reactions of gene activation and gene deactivation reactions in addition to the 4 reactions of mRNA and protein synthesis and degradation. It was observed that addition of gene activation and gene deactivation reaction indeed changed the relative sensitivity to variation in parameters. Figure 4-20 shows the cumulative distribution functions of %CV for six reaction rate parameters.

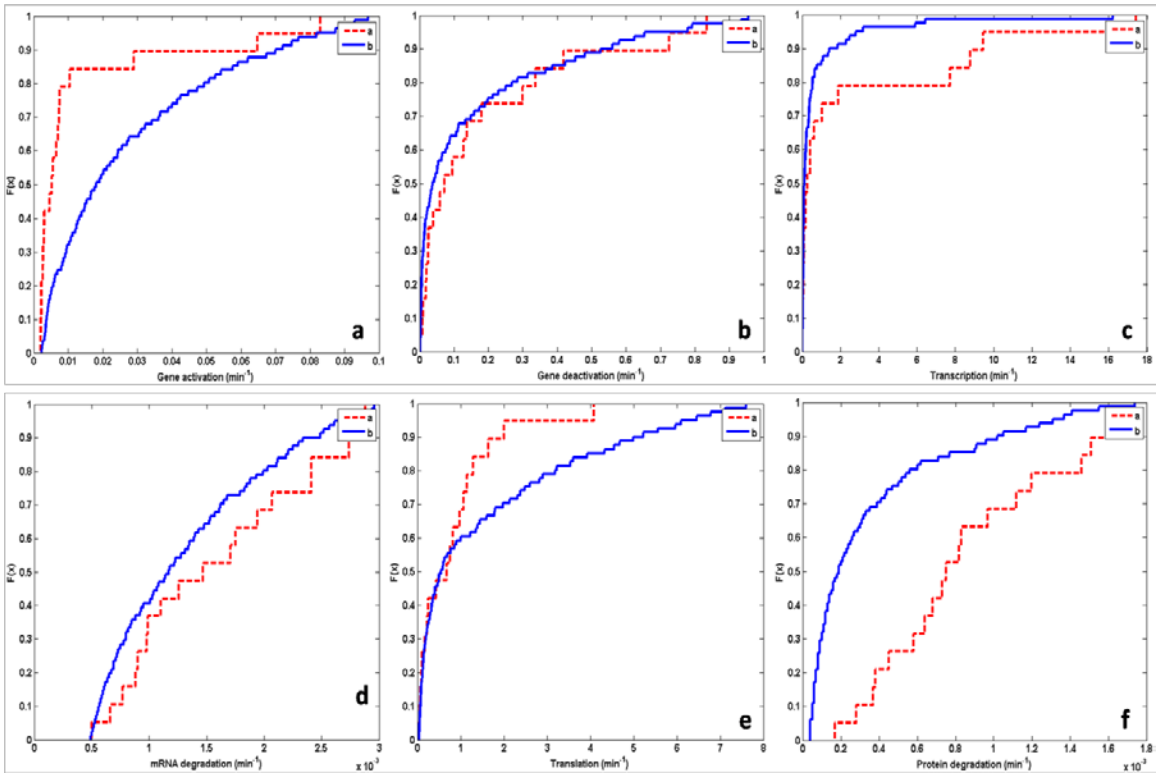


Figure 4-20 – Cumulative density function of steady state protein %CV of two groups for one sample of parameter sets for 6-reaction model of gene expression

(a) gene activation, (b) gene deactivation, (c) transcription, (d) mRNA degradation, (e) translation, and (f) protein degradation.

It was observed that the cumulative distribution functions for two groups were considerably different in case of gene activation (Figure 4-20a) and protein degradation (Figure 4-20f). However, in contrast to the 4-reaction model, the cumulative density functions were observed to be similar for transcription reaction rate parameter (Figure 4-20c). From the sensitivity scores (Table 4-8) it was evident that, CV was maximally sensitive to protein degradation. Gene activation was also observed to affect CV at protein steady state level. The other parameters were observed to have much lesser influence of CV as compared to protein degradation and gene activation.

Similar to 4-reaction model, sensitivity of Fano factors for the six parameters was calculated. In this case as well, Fano factor was observed to be sensitive to different

parameters than CV. Figure 4-21 shows graphs of cumulative distribution function of Fano factor for the six parameters.

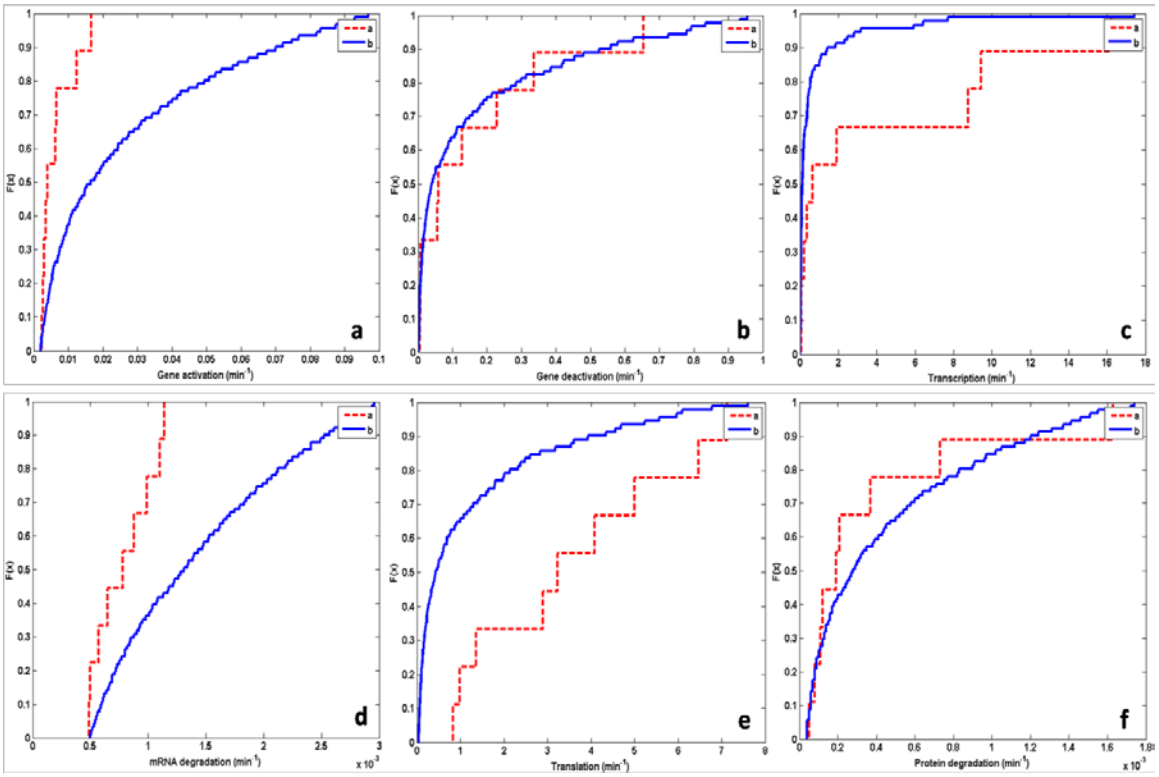


Figure 4-21 – Cumulative density function of steady state protein Fano factor of two groups for one sample of parameter sets for 6-reaction model of gene expression

(a) gene activation, (b) gene deactivation, (c) transcription, (d) mRNA degradation, (e) translation, and (f) protein degradation.

It was observed that the cumulative distribution functions for the two groups were different for gene activation (Figure 4-21a), transcription (Figure 4-21c), mRNA degradation (Figure 4-21d) and translation (Figure 4-21e). It was evident from the graphs that, in contrast to the similar cumulative distribution functions of CV for translation, the functions were different in case of Fano factor. Table 4-8 summarized the sensitivity scores for %CV and Fano factor.

Table 4-8 – Average sensitivity score of steady state protein %CV and Fano factor for 6-reaction model of gene expression

Parameter	% CV average score	Fano factor average score
Gene activation	2.13	1.14
Gene deactivation	0.08	0.16
Transcription	0.25	1.38
mRNA degradation	0.08	0.59
Translation	0.06	2.02
Protein degradation	3.75	0.10

From the sensitivity scores, Fano factor was observed to be most sensitive to translation, similar to that observed for 4 reaction model. It was also observed to be sensitive to gene activation and transcription. With addition of two reactions of gene activation and deactivation, the relative contribution of reactions to CV and Fano factor was observed to be changed. However, translation remained as important step determining the Fano factor in case of both 4-reaction and 6-reaction models of gene expression.

In a study by Raser and O'Shea (Raser and O'Shea 2004), influence of gene activation, deactivation, and mRNA synthesis on variability of mRNA was examined for three different conditions using Fano factor as a measure of noise. In this study, we have examined sensitivity of %CV and Fano factor at steady state level of mRNA for four parameters viz., gene activation, gene deactivation, mRNA synthesis and mRNA degradation. Figure 4-22 and Figure 4-23 show graphs of cumulative distribution function for %CV and Fano factor respectively.

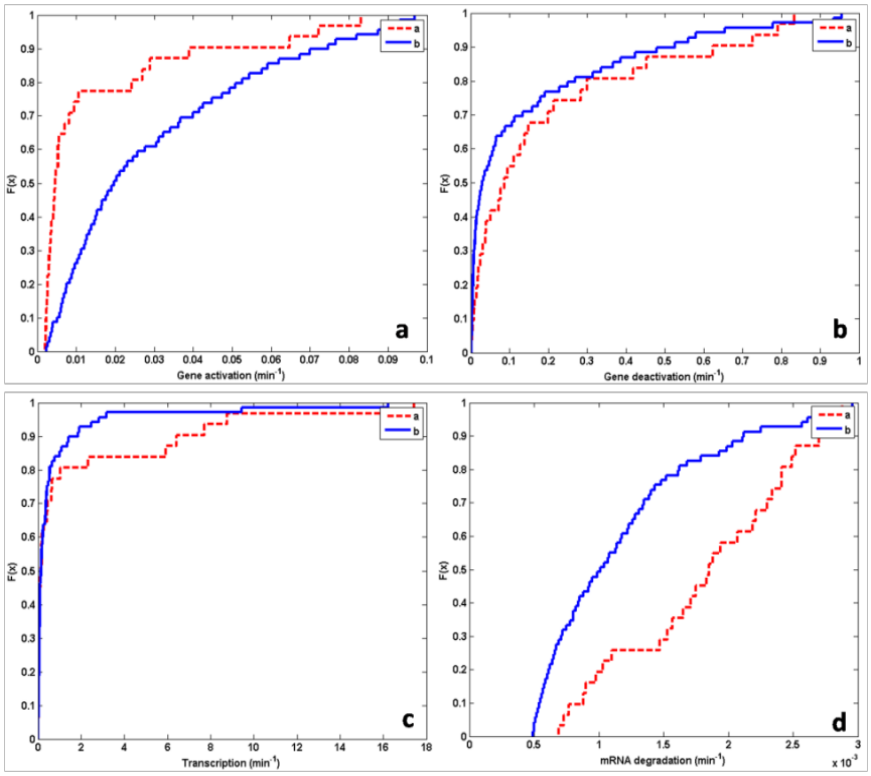


Figure 4-22 – Cumulative density function of steady state mRNA %CV of two groups for one sample of parameter sets for 6-reaction model of gene expression

(a) gene activation, (b) gene deactivation, (c) transcription, (d) mRNA degradation.

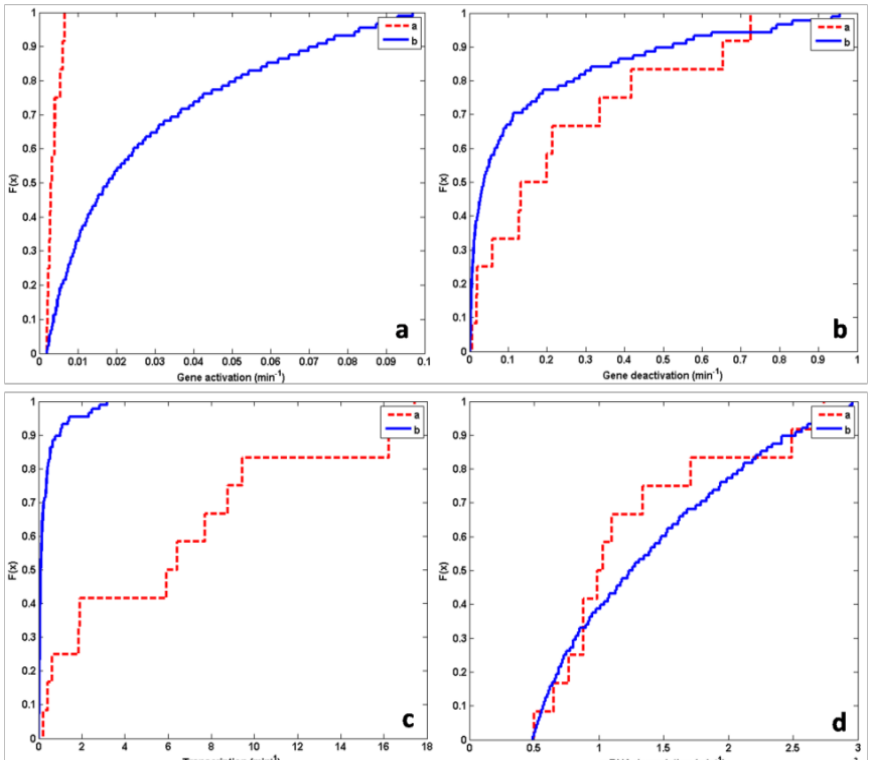


Figure 4-23 – Cumulative density function of steady state mRNA Fano factor of two groups for one sample of parameter sets for 6-reaction model of gene expression

(a) gene activation, (b) gene deactivation, (c) transcription, (d) mRNA degradation.

Similar to the previous cases, CV and Fano factor were observed to be sensitive to different parameters. It was observed that the cumulative distribution functions were different in case of gene activation for both CV and Fano factor (Figure 4-22a and Figure 4-23a). On the other hand, cumulative distribution functions in case of transcription were observed to be similar for CV (Figure 4-22c) while those were different for Fano factor (Figure 4-23c). The cumulative distribution function curves for mRNA degradation were different for CV (Figure 4-22d) but similar in case of Fano factor (Figure 4-23d).

Table 4-9 – Average sensitivity score of steady state mRNA %CV and Fano factor for 6-reaction model of gene expression

Parameter	% CV average score	Fano factor average score
Gene activation	2.08	4.39
Gene deactivation	0.09	0.46
Transcription	0.23	4.52
mRNA degradation	1.08	0.18

From the sensitivity scores (Table 4-9) it was observed that gene activation was observed to be most important to determine CV while transcription was most important to determine Fano factor at steady state mRNA level. Gene activation was also found to be important for Fano factor. mRNA degradation was observed to have least influence on Fano factor in contrast to CV where it was observed to be an important parameter.

In this study, sensitivity of noise to mRNA degradation was also explored in addition to the previously considered three parameters of gene activation, gene deactivation and transcription. From the study it was predicted that mRNA degradation was an important parameter to determine CV at steady state level of mRNA.

4.3.3.3 Sensitivity analysis using 42-reaction model shows maximum sensitivity of CV for protein degradation and maximum sensitivity of Fano factor to gene activation

MPSA was performed on previously developed detailed (42-reaction) model of gene expression. In the detailed model, additional reactions such as binding of regulator protein, RNA polymerase, mRNA transport, binding of translation machinery were considered. The parameter *Reg* abundance can also be changed and sensitivity of noise to the parameter can be examined. However, in this study, the six parameters same as those analysed in case of 6-reacttion model were used for sensitivity analysis.

Figure 4-24 and Figure 4-25 shows graphs of cumulative distribution functions for %CV and Fano factor at steady state protein level.

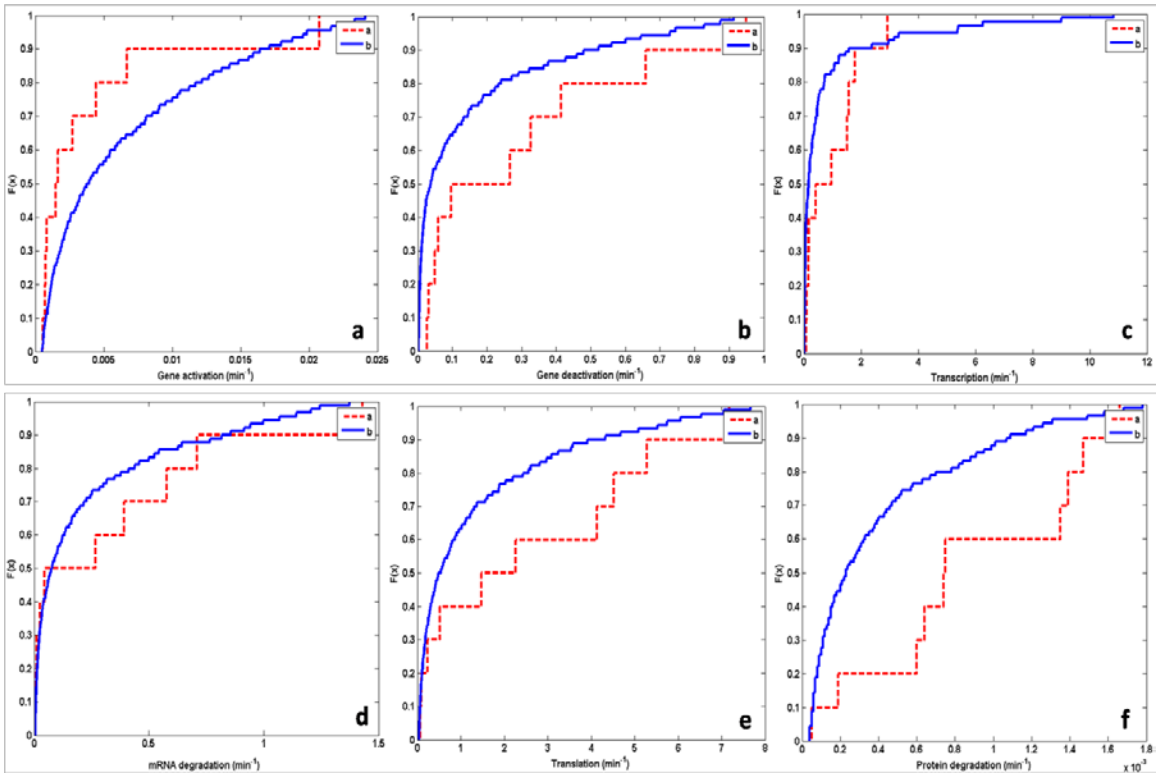


Figure 4-24 – Cumulative density function of steady state protein %CV of two groups for one sample of parameter sets for 42-reaction model of gene expression

(a) gene activation, (b) gene deactivation, (c) transcription, (d) mRNA degradation, (e) translation, and (f) protein degradation.

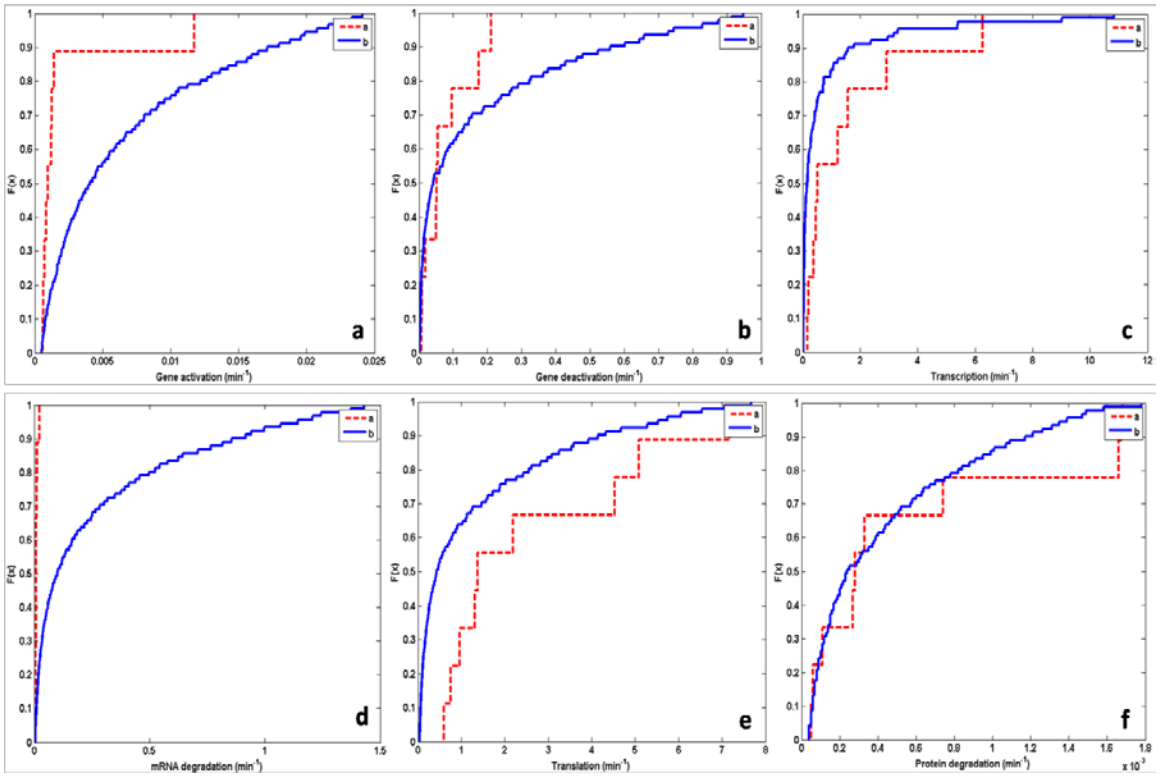


Figure 4-25 – Cumulative density function of steady state protein Fano factor of two groups for one sample of parameter sets for 42-reaction model of gene expression

(a) gene activation, (b) gene deactivation, (c) transcription, (d) mRNA degradation, (e) translation, and (f) protein degradation.

It was observed that in case of %CV, the cumulative distribution functions were significantly different in case of protein degradation (Figure 4-24f). The cumulative distribution function curves were observed to be different to some extent in case of gene activation (Figure 4-24a). In contrast to 6-reaction model where cumulative distribution functions of CV for translation were observed to be similar, in this case the functions were observed to be different to some extent (Figure 4-24e). In case of Fano factor, similar to 6-reaction model, the functions for gene activation were observed to be different (Figure 4-25a). However, in this case the functions for translation were observed to be similar (Figure 4-25e). Significant difference was observed in case of mRNA degradation (Figure 4-25d). Table 4-10 summarized the sensitivity scores of steady state protein %CV and Fano factor. It was evident that CV was most sensitive to protein degradation, while Fano factor was observed to be most sensitive to gene activation.

Table 4-10 – Average sensitivity score of steady state protein %CV and Fano factor for 42-reaction model of gene expression

Parameter	% CV average score	Fano factor average score
Gene activation	0.37	0.95
Gene deactivation	0.28	0.19
Transcription	0.30	0.47
mRNA degradation	0.15	0.91
Translation	0.34	0.33
Protein degradation	0.73	0.37

In contrast to both 4-reaction and 6-reaction models, Fano factor was not observed to be sensitive to translation. In case of detailed model of gene expression, the specific rate of mRNA degradation was affected by the specific rate of translation, as only free mRNA can undergo degradation. Thus observed sensitivity Fano factor to mRNA degradation can be due to the influence of translation reaction rate parameter in determining the mRNA degradation reaction rate parameter. Sensitivity of CV and Fano factor at steady state mRNA level was also calculated. Table 4-11 summarizes the average sensitivity scores.

Table 4-11 – Average sensitivity score of steady state mRNA %CV and Fano factor for 42-reaction model of gene expression

Parameter	% CV average Score	Fano factor average score
Gene activation	0.19	0.16
Gene deactivation	0.21	0.12
Transcription	0.72	0.39
mRNA degradation	3.03	2.39

It was observed that both CV and Fano factor at steady state mRNA level were most sensitive to mRNA degradation. In contrast to 6 reaction model, in which CV and Fano factor were observed to be sensitive to gene activation in this case transcription was observed to be important. In the detailed model, gene activation was considered to be dependent on the parameter of regulator protein (*Reg*) abundance. Thus, sensitivity to *Reg* abundance is also required to be studied.

From the global sensitivity analysis for two measures of noise using three different models of gene expression, it was clear that CV and Fano factor were sensitive to different parameters. It was observed that addition of reactions changed the relative contribution of reactions to noise in steady state level of mRNA and protein.

4.3.3.4 Analytical expressions for sensitivity of CV and Fano factor explain differential contribution of parameters to different measures of noise

To understand the contribution of each reaction rate parameter to CV and Fano factor analytical expression for these two measures of noise was obtained using the 4-reaction model of gene expression. The expressions for CV and Fano factor at steady state level of protein in terms of reaction rate parameters were given as,

$$CV = \sqrt{\frac{k_2 k_4 (\mathrm{k}_2 + \mathrm{k}_3 + \mathrm{k}_4)}{k_1 k_3 (\mathrm{k}_2 + \mathrm{k}_4)}} \tag{4.1}$$

and

$$Fano\ Factor = \frac{k_2 + k_3 + k_4}{k_2 + k_4} \tag{4.2}$$

where, k_1, k_2, k_3 and k_4 are specific rates of mRNA synthesis, mRNA degradation, protein synthesis, and protein degradation respectively. From the analytical expressions for CV (equation 4.1) and Fano factor (equation 4.2) it was observed that all the four parameters contribute to steady state CV. However, specific rate for transcription did not appear in the analytical expression for Fano factor. Therefore, variation in transcription

reaction rate constant would not affect Fano factor at steady state protein level indicating no sensitivity of Fano factor to transcription. In the numerical simulations for global sensitivity, transcription was observed to have least influence on Fano Factor (Table 4-7). The observed sensitivity value can be due to simultaneous variation in parameters performed for global sensitivity.

To examine the contribution of these parameters to sensitivity, analytical expression for local sensitivity was obtained. Though local sensitivity is useful to investigate the sensitivity in the vicinity of a particular parameter value, it is very informative in order in investigate the contribution of reaction rate parameters to the sensitivity. Analytical expressions for sensitivity of CV with respect to every parameter were obtained. The expressions are given as,

$$\textit{Sensitivity of CV to transcription} = \frac{\delta CV}{\delta k_1} = -\frac{k_2 k_4 (k_2 + k_3 + k_4)}{2k_1^{\,2} k_3 (k_2 + k_4)\sqrt{\dfrac{k_2 k_4 (k_2 + k_3 + k_4)}{k_1 k_3 (k_2 + k_4)}}} \tag{4.3}$$

$$\textit{Sensitivity of CV to mRNA degradation} = \frac{\delta CV}{\delta k_2} = \frac{\dfrac{k_2 k_4}{k_1 k_3 (k_2 + k_4)} - \dfrac{k_2 k_4 (k_2 + k_3 + k_4)}{k_1 k_3 (k_2 + k_4)^2} + \dfrac{k_4 (k_2 + k_3 + k_4)}{k_1 k_3 (k_2 + k_4)}}{2\sqrt{\dfrac{k_2 k_4 (k_2 + k_3 + k_4)}{k_1 k_3 (k_2 + k_4)}}}$$

(4.4)

$$\textit{Sensitivity of CV to translation} = \frac{\delta CV}{\delta k_3} = \frac{\dfrac{k_2 k_4}{k_1 k_3 (k_2 + k_4)} - \dfrac{k_2 k_4 (k_2 + k_3 + k_4)}{k_1 k_3^{\,2} (k_2 + k_4)}}{2\sqrt{\dfrac{k_2 k_4 (k_2 + k_3 + k_4)}{k_1 k_3 (k_2 + k_4)}}} \tag{4.5}$$

$$\textit{Sensitivity of CV to protein degradation} = \frac{\delta CV}{\delta k_4} = \frac{\dfrac{k_2 k_4}{k_1 k_3 (k_2 + k_4)} - \dfrac{k_2 k_4 (k_2 + k_3 + k_4)}{k_1 k_3 (k_2 + k_4)^2} + \dfrac{k_2 (k_2 + k_3 + k_4)}{k_1 k_3 (k_2 + k_4)}}{2\sqrt{\dfrac{k_2 k_4 (k_2 + k_3 + k_4)}{k_1 k_3 (k_2 + k_4)}}}$$

(4.6)

Comparing the expressions for sensitivity of CV to mRNA degradation (equation 4.4), translation (4.5) and protein degradation (4.6) it was observed that, the expressions for sensitivity to mRNA degradation and protein degradation contained additional positive terms $\frac{k_4(\mathrm{k}_2+\mathrm{k}_3+\mathrm{k}_4)}{k_1k_3(\mathrm{k}_2+\mathrm{k}_4)}$ and $\frac{k_2(\mathrm{k}_2+\mathrm{k}_3+\mathrm{k}_4)}{k_1k_3(\mathrm{k}_2+\mathrm{k}_4)}$ in the numerator respectively. Therefore, the sensitivity of CV to mRNA degradation and protein degradation would be greater than that for transcription at given values of parameters. Numerical sensitivity results (Table 4-7) were observed to be in qualitative agreement with the observed analytical local sensitivity. Generally mRNA molecules are known to be less stable than protein molecules, indicating higher numerical values of mRNA degradation reaction rate constant than protein degradation. Therefore, comparing the expression for sensitivity to mRNA degradation and protein degradation, it can be inferred that sensitivity of CV to mRNA degradation would be lesser than that for protein degradation for majority of the cases, as observed in numerical sensitivity analysis. However, only in case of a very stable mRNA and unstable protein it may not hold true.

Analytical expressions for sensitivity of Fano factor to these four parameters were obtained. As evident from the expression of Fano factor, transcription did not contribute to Fano factor at steady state protein level. Therefore, sensitivity of Fano factor to transcription was observed to be zero. The analytical expressions for sensitivity of Fano factor to other three parameters was given as,

$$\textit{Sensitivity of Fano factor to mRNA degradation} = \frac{\delta \textit{Fano factor}}{\delta k_2} = \frac{1}{k_2+k_4} - \frac{k_2+k_3+k_4}{(k_2+k_4)^2} \qquad (4.7)$$

$$\textit{Sensitivity of Fano factor to translation} = \frac{\delta \textit{Fano factor}}{\delta k_3} = \frac{1}{k_2+k_4} \qquad (4.8)$$

$$\textit{Sensitivity of Fano factor to protein degradation} = \frac{\delta \textit{Fano factor}}{\delta k_4} = \frac{1}{k_2+k_4} - \frac{k_2+k_3+k_4}{(k_2+k_4)^2} \qquad (4.9)$$

The sensitivity of Fano factor to mRNA and protein degradation (expressions 4.7 and 4.9) was observed to be same. Comparing these expressions with expression for

sensitivity of Fano factor to translation, it was observed that a negative term $\frac{k_2 + k_3 + k_4}{(k_2 + k_4)^2}$ was present in sensitivity to mRNA and protein degradation. Therefore, qualitatively it can be inferred that sensitivity to translation would be greater than that for mRNA and protein degradation. Numerical simulations results (Table 4-7) were observed to be in agreement with the observed qualitative nature of relative sensitivity. From the expression for sensitivity of Fano factor to translation it was observed that, change in Fano factor values due to change in translation would depend only upon the value of mRNA and protein degradation and not on value of translation.

As local sensitivity examines the behavior of model in the neighbourhood of a parameter value, the absolute values of change in output at different parameter values can differ. In addition, local sensitivity considers variation only in one parameter while other parameters are considered to be constant. Therefore, absolute value of local sensitivity can change depending upon the values of other parameters. To compare local sensitivity at different parameter values, relative sensitivity coefficient, $\frac{\delta output}{\delta p}.\frac{p}{output}$ was used.

From expressions 4.1 and 4.3, the relative sensitivity coefficient of CV to transcription, $\frac{\delta CV}{\delta k_1}.\frac{k_1}{CV}$ was calculated to be -0.5. This indicated that, at any parameter values, Δ increase in k_1 would result in 0.5 Δ decrease in CV. Therefore, the relative change in CV for change in transcription at any parameter values would be same, indicating that values of other parameters did not affect.

The relative sensitivity coefficient of CV to translation was given as,

$$\frac{\delta CV}{\delta k_3}.\frac{k_3}{CV} = \frac{-(k_2 + k_4)}{2(k_2 + k_3 + k_4)} \qquad (4.10)$$

From the expression (4.10) it was observed that relative sensitivity coefficient of CV to translation did not depend upon transcription. The relative change in CV at a particular value of translation was observed to be dependent on values of mRNA and protein

degradation as well. It can be inferred that at very low translation compared to mRNA or protein degradation, for Δ increase in mRNA or protein degradation there would be 0.5Δ decrease in CV. While at high translation, mRNA and protein degradation would not affect and the change in CV would depend upon translation. From the physiological parameter ranges it was observed that, the numerical values of translation reaction rate constant being order of magnitude greater than that of mRNA and protein degradation, these processes had little effect of relative change in sensitivity. Figure 4-26 shows the plot of relative sensitivity coefficient as a function of translation, mRNA degradation and protein degradation.

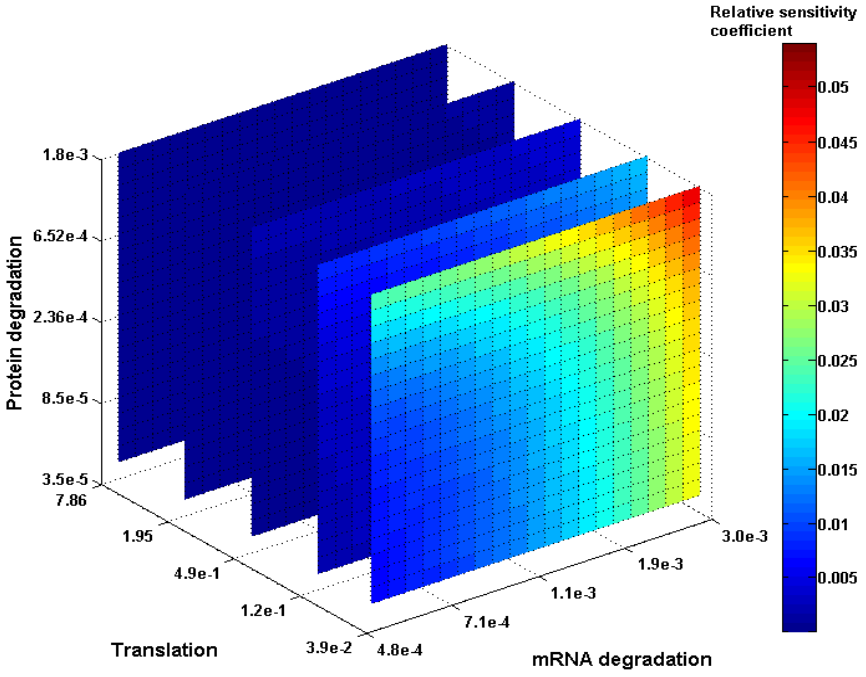


Figure 4-26 – Plot of relative sensitivity coefficient of CV to translation as a function of translation, mRNA degradation, and protein degradation

From the plot it was observed that, translation had maximum effect on relative sensitivity coefficient. Comparing mRNA degradation and protein degradation, mRNA degradation had more effect on relative sensitivity coefficient.

Similarly, to investigate the effect of other parameters on the sensitivity coefficient of Fano factor to translation the expression for the relative sensitivity coefficient was obtained. The expression is given as,

$$\frac{\delta Fano\ factor}{\delta k_3} \cdot \frac{k_3}{Fano\ factor} = \frac{k_3}{k_2 + k_3 + k_4} \tag{4.11}$$

The nature of the function (equation 4.11) was observed to be saturating. At high values of translation (k_3) the relative sensitivity coefficient should approach one. Figure 4-27 shows plot of relative sensitivity coefficient as a function of translation, mRNA degradation and protein degradation.

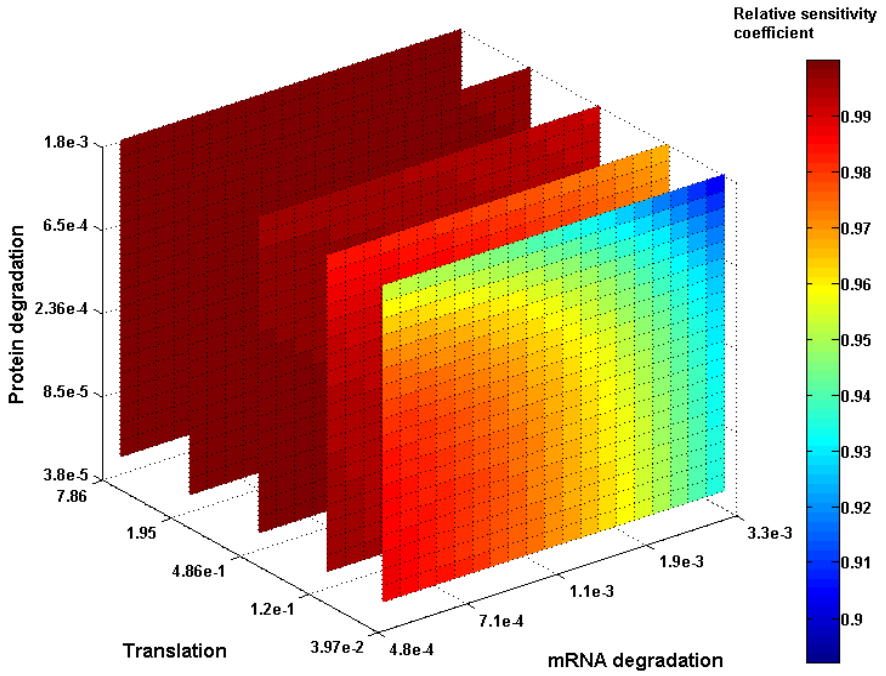


Figure 4-27 – Plot of relative sensitivity coefficient of Fano factor to translation as a function of translation, mRNA degradation, and protein degradation

It was evident that for translation reaction rate parameter values greater than $4.86*10^{-1}min^{-1}$, the relative sensitivity coefficient was not affected by mRNA degradation and protein degradation. The value was observed to reach 0.99 at higher values of specific

rate of translation. Comparing mRNA and protein degradation, mRNA degradation was observed to have more effect than protein degradation, similar to that observed in case of relative sensitivity coefficient of CV to translation.

In summary, from the global and local sensitivity analysis using two different measures of noise it was observed that the two measures of noise were sensitive to different parameters. Additionally, the sensitivity to one parameter was observed to be affected by the values of other parameters as well. The observed differential sensitivity of CV and Fano factor to parameters could be explained from analytical expressions for sensitivity.

4.3.4 Conclusion

To examine the relative contribution of major steps in gene expression to noise in steady state protein level, global stochastic sensitivity analysis was performed using MPSA. In contrast to majority of the previous experimental and theoretical studies that consider only one measure of noise, in this study sensitivity was determined for both coefficient of variation and Fano factor. 3 models of gene expression with differing degree of details were used for the study. It was observed that addition of different reactions changed the relative contribution to the noise in steady state protein level. Thus the level of abstraction can affect the observed relative sensitivity.

Previous theoretical study on first order reaction network has shown that the two measures lead to contrary conclusions about noise (Gadgil, Lee et al. 2005). Therefore, both the measures of noise are required to be examined. From the observed differential sensitivity of the two measures of noise it is clear that the sensitivity estimated using a particular measure of noise, should not be attributed to the generic variability or ‘noise’. Thus, the previous analyses determining sensitivity need to be reanalyzed to explore the sensitivity for other measure of noise.

In summary, by comparing different measures of noise and models of gene expression considering different levels of abstractions, the study highlighted differential contribution of parameters to the two measures of noise and provided a comparative view of sensitivity of noise at steady state protein level over wide parameter ranges.

5 Overall conclusion and Future directions

The study of biological phenomena has changed from the investigation of one gene to exploring the emergent properties of a system as a whole. Development of new experimental techniques of growing cell cultures, visualization, and quantification has advanced the investigation of large systems at single molecule resolution. These experiments create enormous amount of data that needs to be analysed. Theoretical analysis is an effective approach to analyse large amount of data and study complex biological processes. Development of mathematical model marks an important step in the iterative process of experimental analysis and theoretical analysis.

The present study involves investigation of some aspects of gene expression and its regulation. It includes the study of regulation of gene expression at transcriptional and post-transcriptional level and of stochasticity in gene expression.

In the first part, a mathematical model of transcriptional regulation by transcription factor TBP was developed for auto-regulatory synthesis of TBP. The model prediction of requirement of minimal amount of TBP for cell viability has implication in understanding the observed maternal inheritance of TBP. The relationship between low DNA binding affinity of TBP and cell viability is useful to gain insight into the role of the observed low DNA binding affinity mutants of TBP in neurodegenerative diseases.

A detailed mathematical model of miRNA mediated regulation was developed to study RNA mediated regulation at post-transcriptional level. With the known biological details of mechanism of regulation, the unintuitive observations of activating effect of miRNA were explained using the developed mathematical model such that these observations are no longer unexpected. In addition to the analysis of miRNA effect on the steady state of target protein, a simple method was developed to incorporate dynamic effects of intronic miRNA into existing mathematical models of cellular processes. Using the developed method a mathematical model of cell cycle was modified to include regulation by one intronic miRNA that resulted in improved model performance.

Contributions of intrinsic and extrinsic sources of noise to the observed protein distribution were analysed in the present study. A detailed model of gene expression was developed to investigate sources of gene-expression extrinsic noise that contribute to the observed noise floor at high protein abundance. It was observed that time scale of

fluctuations of a parameter compared to that of the protein of interest determine the observed contribution of extrinsic noise to the observed noise floor. Global sensitivity analysis was performed to estimate the relative contribution of major steps in gene expression to the intrinsic noise at steady state protein level. Interestingly, different measures of noise were observed to be sensitive to different parameters. Thus the study suggested that the observed sensitivity should be attributed to the specific measure and not to the generic 'noise' or variability.

In the present study the mathematical model of TBP was developed and analyzed using the deterministic approach. Analysis using stochastic approach would be important as the low copy number of TBP gene can result in significant fluctuations in TBP level. The mathematical model of miRNA mediated regulation can be extended to study the effect of multiple miRNA on multiple mRNA targets, as observed in cellular systems. The separate models of transcriptional and post-transcriptional regulation can be combined to study the properties of cross talk between regulatory layers. The model of gene expression though detailed, contained many assumptions and simplifications. Certain details were not included due to lack of experimental details. The model can be improved to incorporate biological details once experimental data is available. As the developed model of gene expression is modular, the modules of sub-processes can be used separately. Being a detailed model regulation at multiple levels can be incorporated in the single model.

In summary, the present study contributed to advance the understanding of certain aspects of eukaryotic gene expression. It demonstrated the use of mathematical modeling to gain deeper insight into gene expression process. The study showed the utility of dimensionless numbers for predictions of outcome of biological processes. The sensitivity analysis highlighted the fact that a mathematical measure or representation is a proxy for a concept and the interpretation may not always be generalized. In this study certain falsifiable hypotheses were generated that can be experimentally tested to get improved understanding of the underlying biological process in eukaryotic gene expression.

References

Aboobaker, A. A., P. Tomancak, et al. (2005). "Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development." Proceedings of the National Academy of Sciences of the United States of America **102**(50): 18017.

Aguda, B. D., Y. Kim, et al. (2008). "MicroRNA regulation of a cancer network: consequences of the feedback loops involving miR-17-92, E2F, and Myc." Proc Natl Acad Sci U S A **105**(50): 19678-83.

Alberts, B., D. Bray, et al. (2002). Molecular biology ofthe cell.

Alon, U. (2007). An introduction to systems biology: design principles of biological circuits, CRC Press.

Alvarez-Garcia, I. and E. A. Miska (2005). "MicroRNA functions in animal development and human disease." Development **132**(21): 4653-4662.

Angeli, D., J. E. Ferrell, et al. (2004). "Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems." Proceedings of the National Academy of Sciences of the United States of America **101**(7): 1822-1827.

Ardehali, M. B. and J. T. Lis (2009). "Tracking rates of transcription and splicing in vivo." Nature Structural & Molecular Biology **16**(11): 1123-1124.

Arigo, J. T., K. L. Carroll, et al. (2006). "Regulation of yeast NRD1 expression by premature transcription termination." Molecular cell **21**(5): 641-651.

Arnold, S., M. Siemann, et al. (2001). "Kinetic modeling and simulation of in vitro transcription by phage T 7 RNA polymerase." Biotechnology and bioengineering **72**(5): 548-561.

Audibert, A., D. Weil, et al. (2002). "In vivo kinetics of mRNA splicing and transport in mammalian cells." Molecular and Cellular Biology **22**(19): 6706.

Auld, K. L., C. R. Brown, et al. (2006). "Genomic association of the proteasome demonstrates overlapping gene regulatory activity with transcription factor substrates." Molecular cell **21**(6): 861-871.

Bai, L., A. Shundrovsky, et al. (2004). "Sequence-dependent kinetic model for transcription elongation by RNA polymerase." Journal of molecular biology **344**(2): 335-349.

Balagopal, V., L. Fluch, et al. (2012). "Ways and means of eukaryotic mRNA decay." Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms **1819**(6): 593-603.

Bar-Even, A., J. Paulsson, et al. (2006). "Noise in protein expression scales with natural protein abundance." nature genetics **38**(6): 636-643.

Bar N.S. (2009). "Analysis of protein synthesis dynamic model in eukaryotic cells: Input control." Mathematical Biosciences **210**: 84–91.

Barik, S. (2008). "An intronic microRNA silences genes that are functionally antagonistic to its host gene." Nucleic Acids Research **36**(16): 5232.

Barrandon, C., B. Spiluttini, et al. (2008). "Non-coding RNAs regulating the transcriptional machinery." Biology of the Cell **100**: 83-95.

Bartel, D. P. (2004). "MicroRNAs genomics, biogenesis, mechanism, and function." Cell **116**(2): 281-297.

Baskerville, S. and D. P. Bartel (2005). "Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes." RNA **11**(3): 241.

Baumeister, W., J. Walz, et al. (1998). "The Proteasome: Paradigm Review of a Self-Compartmentalizing Protease." Cell **92**: 367-380.

Becskei, A., B. Seraphin, et al. (2001). "Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion." The EMBO journal **20**(10): 2528-2535.

Beelman, C. A. and R. Parker (1995). "Degradation of mRNA in eukaryotes." Cell **81**(2): 179-183.
Ben-Asouli, Y., Y. Banai, et al. (2002). "Human interferon-[gamma] mRNA autoregulates its translation through a pseudoknot that activates the interferon-inducible protein kinase PKR." Cell **108**(2): 221-232.
Bergmann, J. E. and H. F. Lodish (1979). "A kinetic model of protein synthesis. Application to hemoglobin synthesis and translational control." Journal of Biological Chemistry **254**(23): 11927.
Bernard S., Cajavec B., et al. (2006). "Modelling transcriptional feedback loops: the role of Gro/TLE1 in Hes1 oscillations." Phil. Trans. R. Soc. A **364**: 1155–1170.
Bhattacharyya, S. N., R. Habermacher, et al. (2006). "Relief of microRNA-mediated translational repression in human cells subjected to stress." Cell **125**(6): 1111-1124.
Biran, A. and E. Meshorer (2012). "Concise Review: Chromatin and Genome Organization in Reprogramming." Stem Cells **30**(9): 1793-1799.
Blake, W. J., M. Kærn, et al. (2003). "Noise in eukaryotic gene expression." Nature **422**(6932): 633-637.
Blossey, R. and H. Schiessel (2008). "Kinetic proofreading of gene activation by chromatin remodeling."
Boeger, H., J. Griesenbeck, et al. (2008). "Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription." Cell **133**(4): 716-726.
Borchert, G. M., W. Lanier, et al. (2006). "RNA polymerase III transcribes human microRNAs." Nature Structural & Molecular Biology **13**(12): 1097-1101.
Borggrefe, T., R. Davis, et al. (2001). "Quantitation of the RNA polymerase II transcription machinery in yeast." Journal of Biological Chemistry **276**(50): 47150-47153.
Brantl, S. and E. G. H. Wagner (2002). "An antisense RNA-mediated transcriptional attenuation mechanism functions in Escherichia coli." Journal of bacteriology **184**(10): 2740.
Cao, D. and R. Parker (2001). "Computational modeling of eukaryotic mRNA turnover." Rna **7**(9): 1192.
Carè, A., D. Catalucci, et al. (2007). "MicroRNA-133 controls cardiac hypertrophy." Nature medicine **13**(5): 613-618.
Carninci, P., T. Kasukawa, et al. (2005). "The transcriptional landscape of the mammalian genome." Science **309**(5740): 1559-1563.
Carrier, T. A. and J. D. Keasling (1997). "Mechanistic modeling of prokaryotic mRNA decay." Journal of theoretical biology **189**(2): 195-209.
Carrington, J. C. and V. Ambros (2003). "Role of microRNAs in plant and animal development." Science **301**(5631): 336.
Carthew, R. W. (2006). "Gene regulation by microRNAs." Current opinion in genetics & development **16**(2): 203-208.
Casolari, J. M., C. R. Brown, et al. (2004). "Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization." Cell **117**(4): 427-439.
Chalut, C., Y. Gallois, et al. (1995). "Genomic structure of the human TATA-box-binding protein (TBP)." Gene **161**(2): 277-82.
Chang, T. C., D. Yu, et al. (2007). "Widespread microRNA repression by Myc contributes to tumorigenesis." Nature genetics **40**(1): 43-50.
Chang, Y.-F., J. S. Imam, et al. (2007). "The nonsense-mediated decay RNA surveillance pathway." Annu. Rev. Biochem. **76**: 51-74.
Cheng C., Yaffe M.B., et al. (2006). "A positive feedback loop couples Ras activation and CD44 alternative splicing." Genes & Development **20**: 1715–1720.

Cherry, J. L. and F. R. Adler (2000). "How to make a biological switch." Journal of theoretical biology **203**(2): 117-133.
Cho, K. H., S. Y. Shin, et al. (2003). "Experimental Design in Systems Biology, Based on Parameter Sensitivity Analysis Using a Monte Carlo Method: A Case Study for the TNFÎ±-Mediated NF-Îº B Signal Transduction Pathway." Simulation **79**(12): 726-739.
Cho, R. J., M. Huang, et al. (2001). "Transcriptional regulation and function during the human cell cycle." nature genetics **27**(1): 48-54.
Chou, T. (2007). "Peeling and sliding in nucleosome repositioning." Physical review letters **99**(5): 58105.
Ciocchetta, F., J. Hillston, et al. (2008). "Modelling co-transcriptional cleavage in the synthesis of yeast pre-rRNA." Theoretical Computer Science **408**(1): 41-54.
Clague, M. J. and S. Urbe (2010). "Ubiquitin: same molecule, different degradation pathways." Cell **143**(5): 682-685.
Coleman, R. A. and B. F. Pugh (1997). "Slow dimer dissociation of the TATA binding protein dictates the kinetics of DNA binding." Proc Natl Acad Sci U S A **94**(14): 7221-6.
Coleman, R. A., A. K. Taggart, et al. (1995). "Dimerization of the TATA binding protein." J Biol Chem **270**(23): 13842-9.
Cormack, B. P. and K. Struhl (1992). "The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells." Cell **69**(4): 685-696.
Costa, F. c. F. (2005). "Non-coding RNAs: new players in eukaryotic biology." Gene **357**(2): 83-94.
Cranz, S., C. Berger, et al. (2004). "Monomeric and dimeric bZIP transcription factor GCN4 bind at the same rate to their target DNA site." Biochemistry **43**(3): 718-727.
Crick, F. (1970). "Central dogma of molecular biology." Nature **227**(5258): 561-563.
Crick, F. H. C. (1958). "The Biological Replication of Macromolecules." Symp. Soc. Exp. Biol **XII**: 138.
Cuccato, G., A. Polynikis, et al. (2011). "Modeling RNA interference in mammalian cells." BMC Systems Biology **5**(1): 19.
Darzacq, X., Y. Shav-Tal, et al. (2007). "In vivo dynamics of RNA polymerase II transcription." Nature Structural & Molecular Biology **14**(9): 796-806.
De Silvaa E., Krishnana J. , et al. (2010). "A mathematical modelling framework for elucidating the role of feedback control in translation termination " Journal of Theoretical Biology **264**(3): 808-821
Degasperi, A. and S. Gilmore (2008). "Sensitivity analysis of stochastic models of bistable biochemical reactions." Formal Methods for Computational Systems Biology: 1-20.
Delbruck, M. (1940). "Statistical fluctuations in autocatalytic reactions." The Journal of Chemical Physics **8**: 120.
Denissov, S., M. van Driel, et al. (2007). "Identification of novel functional TBP-binding sites and general factor repertoires." The EMBO Journal **26**: 944-954.
Deribe, Y. L., T. Pawson, et al. (2010). "Post-translational modifications in signal integration." Nature Structural & Molecular Biology **17**(6): 666-672.
Diaz, R., J. Silva, et al. (2008). "Deregulated expression of miR-106a predicts survival in human colon cancer patients." Genes, Chromosomes and Cancer **47**(9).
Dimelow, R. J. and S. J. Wilkinson (2009). "Control of translation initiation: a model-based analysis from limited experimental data." Journal of The Royal Society Interface **6**(30): 51.
Drew, D. A. (2001). "A mathematical model for prokaryotic protein synthesis." Bulletin of mathematical biology **63**(2): 329-351.

Edelmann, L., L. Zheng, et al. (1998). "The TATA binding protein in the sea urchin embryo is maternally derived." Developmental Biology **204**(1): 293-304.

Elowitz, M. B. and S. Leibler (2000). "A synthetic oscillatory network of transcriptional regulators." Nature **403**(6767): 335-338.

Farley, B. M. and S. P. Ryder (2008). "Regulation of maternal mRNAs in early development." Critical Reviews in Biochemistry and Molecular Biology **43**(2): 135-162.

Ferrell, J. E. and E. M. Machleder (1998). "The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes." Science **280**(5365): 895-898.

Ferrell Jr, J. E., T. Y. C. Tsai, et al. (2011). "Modeling the cell cycle: why do certain circuits oscillate?" Cell **144**(6): 874-885.

Fire, A., S. Q. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-811.

Fong, Y. W., C. Cattoglio, et al. (2012). "Transcriptional regulation by coactivators in embryonic stem cells." Trends in Cell Biology.

Fraser, H. B., A. E. Hirsh, et al. (2004). "Noise minimization in eukaryotic gene expression." PLoS biology **2**(6): e137.

Friedman, M. J., C. E. Wang, et al. (2008). "Polyglutamine expansion reduces the association of TATA-binding protein with DNA and induces DNA binding-independent neurotoxicity." J Biol Chem **283**(13): 8283-90.

Fry, C. J. and C. L. Peterson (2001). "Chromatin remodeling enzymes: who's on first?" Current Biology **11**(5): R185-R197.

Gadgil, C., C. H. Lee, et al. (2005). "A stochastic analysis of first-order reaction networks." Bulletin of mathematical biology **67**(5): 901-946.

Gardner, T. S., C. R. Cantor, et al. (2000). "Construction of a genetic toggle switch in Escherichia coli." Nature **403**(6767): 339-342.

Gebauer, F. and M. W. Hentze (2004). "Molecular mechanisms of translational control." Nature Reviews Molecular Cell Biology **5**(10): 827-835.

Ghaemmaghami, S., W.-K. Huh, et al. (2003). "Global analysis of protein expression in yeast." Nature **425**(6959): 737-741.

Ghildiyal, M. and P. D. Zamore (2009). "Small silencing RNAs: an expanding universe." Nature Reviews Genetics **10**(2): 94-108.

Ghosh, T., K. Soni, et al. (2008). "MicroRNA-mediated up-regulation of an alternatively polyadenylated variant of the mouse cytoplasmic {beta}-actin gene." Nucleic Acids Research.

Gilchrist, M. A. and A. Wagner (2006). "A model of protein translation including codon bias, nonsense errors, and ribosome recycling." Journal of theoretical biology **239**(4): 417-434.

Gillespie, D. T. (1976). "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions." Journal of computational physics **22**(4): 403-434.

Glisovic, T., J. L. Bachorik, et al. (2008). "RNA-binding proteins and post-transcriptional gene regulation." FEBS letters **582**(14): 1977-1986.

Gokhale, S., D. Nyayanit, et al. (2011). "A systems view of the protein expression process." Systems and Synthetic Biology: 1-12.

Golding, I., J. Paulsson, et al. (2005). "Real-time kinetics of gene activity in individual bacteria." Cell **123**(6): 1025-1036.

Gonzalez, S., D. G. Pisano, et al. (2008). "Mechanistic principles of chromatin remodeling guided by siRNAs and miRNAs." Cell cycle **7**(16): 2601-2608.

Groppo, R. and J. D. Richter (2009). "Translational control from head to tail." Current Opinion in Cell Biology **21**(3): 444-451.

Guajardo, R. and R. Sousa (1997). "A model for the mechanism of polymerase translocation1." Journal of molecular biology **265**(1): 8-19.

Guhaniyogi, J. and G. Brewer (2001). "Regulation of mRNA stability in mammalian cells." Gene **265**(1-2): 11-23.

Gunawan, R., Y. Cao, et al. (2005). "Sensitivity analysis of discrete stochastic systems." Biophysical journal **88**(4): 2530.

Halvorsen, Y. C., K. Nandabalan, et al. (1990). "LAC9 DNA-binding domain coordinates two zinc atoms per monomer and contacts DNA as a dimer." Journal of Biological Chemistry **265**(22): 13283-13289.

Hamada, H., Y. Tashima, et al. (2009). "Sophisticated Framework between Cell Cycle Arrest and Apoptosis Induction Based on p53 Dynamics." PLoS ONE **4**(3).

Hastings, M. L. and A. R. Krainer (2001). "Pre-mRNA splicing in the new millennium." Current Opinion in Cell Biology **13**(3): 302-309.

Hasty, J., D. McMillen, et al. (2001). "Computational studies of gene regulatory networks: in numero molecular biology." Nature Reviews Genetics **2**(4): 268-279.

Heinrich, R., S. M. Rapoport, et al. (1977). "Metabolic regulation and mathematical models." Prog. Biophys. Mol. Biol **32**(1): 1–82.

Heinrich, R. and T. A. Rapoport (1980). "Mathematical modelling of translation of mRNA in eucaryotes; steady states, time-dependent processes and application to reticulocytest* 1." Journal of theoretical biology **86**(2): 279-313.

Hernandez, N. (1993). "TBP, a universal eukaryotic transcription factor?" Genes Dev **7**(7B): 1291-308.

Heyd, A. and D. A. Drew (2003). "A mathematical model for elongation of a peptide chain." Bulletin of mathematical biology **65**(6): 1095-1109.

Hieronymus, H., C. Y. Michael, et al. (2004). "Genome-wide mRNA surveillance is coupled to mRNA export." Genes & development **18**(21): 2652-2662.

Hirose, S. (1998). "Chromatin remodeling and transcription." JOURNAL OF BIOCHEMISTRY-TOKYO- **124**: 1060-1064.

Höfer, T. and R. J. Malte (2005). On the kinetic design of transcription. Genome Informatics.

Holstege, F. C. P., E. G. Jennings, et al. (1998). "Dissecting the regulatory circuitry of a eukaryotic genome." Cell **95**(5): 717-728.

Holzhütter, H. G. and P. M. Kloetzel (2000). "A kinetic model of vertebrate 20S proteasome accounting for the generation of major proteolytic fragments from oligomeric peptide substrates." Biophysical journal **79**(3): 1196-1205.

Hornberger, G. M. and R. C. Spear (1981). "Approach to the preliminary analysis of environmental systems." J. Environ. Manage.;(United States) **12**(1).

Houseley, J. and D. Tollervey (2009). "The many pathways of RNA degradation." Cell **136**(4): 763-776.

Hutti J.E., Turk B. E., et al. (2007). "IκB Kinase Phosphorylates the K63 Deubiquitinase A20 To Cause Feedback Inhibition of the NF-κB Pathway." Molecular And Cellular Biology **27**(21): 7451–7461.

Hutvagner, G. and P. D. Zamore (2002). A microRNA in a multiple-turnover RNAi enzyme complex. **297:** 2056-2060.

Isaacs, F. J., J. Hasty, et al. (2003). "Prediction and measurement of an autoregulatory genetic module." Proceedings of the National Academy of Sciences **100**(13): 7714-7719.

Isik, M., H. C. Korswagen, et al. (2010). "Expression patterns of intronic microRNAs in Caenorhabditis elegans." Silence **1**(1): 5.

Ivanova I. P., Loughrana G., et al. (2010). "Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1)." PNAS **107**(42): 18056–18060.

Ivanovska, I., A. S. Ball, et al. (2008). "MicroRNAs in the miR-106b Family Regulate p21/CDKN1A and Promote Cell Cycle Progression?†." Molecular and Cellular Biology **28**(7): 2167-2174.

Jackson-Fisher, A. J., C. Chitikila, et al. (1999). "A role for TBP dimerization in preventing unregulated gene expression." Mol Cell **3**(6): 717-27.

Jangra, R. K., M. K. Yi, et al. (2010). "miR-122 regulation of hepatitis C virus translation and infectious virus production." Journal of Virology: JVI. 00417-10v1.

Jessica E. Hutti, B. E. T., John M. Asara, Averil Ma,Lewis C. Cantley, and Derek W. Abbott (2007). "IκB Kinase Phosphorylates the K63 Deubiquitinase A20 To Cause Feedback Inhibition of the NF-κB Pathway." Molecular And Cellular Biology **27**(21): 7451–7461.

Jin, Y., X. Peng, et al. (2008). "Uniform design-based sensitivity analysis of circadian rhythm model in *Neurospora*." Computers & Chemical Engineering **32**(8): 1956-1962.

Johnson, S. M., H. Grosshans, et al. (2005). "RAS is regulated by the let-7 microRNA family." Cell **120**(5): 635-647.

Johnson, T. L. and J. Vilardell (2012). "Regulated pre-mRNA splicing: The ghostwriter of the eukaryotic genome." Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms **1819**(6): 538-545.

Jopling, C. L., M. K. Yi, et al. (2005). "Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA." Science **309**(5740): 1577.

Kapranov, P., S. E. Cawley, et al. (2002). "Large-scale transcriptional activity in chromosomes 21 and 22." Science **296**(5569): 916-919.

Kelen, K. V. D., R. Beyaert, et al. (2009). "Translational control of eukaryotic gene expression." Critical reviews in biochemistry and molecular biology **44**(4): 143-168.

Keller, A. D. (1995). "Model genetic circuits encoding autoregulatory transcription factors." Journal of theoretical biology **172**(2): 169-185.

Kerppola, T. K. and C. M. Kane (1991). "RNA polymerase: regulation of transcript elongation and termination." The FASEB Journal **5**(13): 2833.

Khanin, R. and D. J. Higham (2007). A minimal mathematical model of post-transcriptional gene regulation by microRNAs, University of Strathclyde.

Khanin, R. and V. Vinciotti (2008). "Computational modeling of post-transcriptional gene regulation by microRNAs." J Comput Biol **15**(3): 305-16.

Kierzek, A. M., J. Zaim, et al. (2001). "The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression." Journal of Biological Chemistry **276**(11): 8165.

Kim, D., B. J. Debusschere, et al. (2007). "Spectral methods for parametric sensitivity in stochastic dynamical systems." Biophysical journal **92**(2): 379.

Kim H. D and O'Shea E.K (2008). "A quantitative model of transcription factor–activated gene expression." Nat Struct Mol Biol. **15**(11): 1192–1198.

Kim, K. H. and H. M. Sauro (2010). "Sensitivity summation theorems for stochastic biochemical reaction systems." Mathematical Biosciences **226**(2): 109.

Koff, A., A. Giordano, et al. (1992). "Formation and activation of a cyclin E-cdk2 complex during the G1 phase of the human cell cycle." Science **257**(5077): 1689-94.

Komili, S. and P. A. Silver (2008). "Coupling and coordination in gene expression processes: a systems biology view." Nature Reviews Genetics **9**(1): 38-48.

Komorowski, M., M. J. Costa, et al. (2011). "Sensitivity, robustness, and identifiability in stochastic chemical kinetics models." Proceedings of the National Academy of Sciences **108**(21): 8645-8650.

Komorowski, M., J. Mi kisz, et al. (2009). "Translational repression contributes greater noise to gene expression than transcriptional repression." Biophysical Journal **96**(2): 372-384.

Kondo, T., T. Mori, et al. (1997). "Circadian rhythms in rapidly dividing cyanobacteria." Science **275**(5297): 224-227.

Konishi, T. (2005). "A thermodynamic model of transcriptome formation." Nucleic acids research **33**(20): 6587.

Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.

Kugel, J. F. and J. A. Goodrich (2000). "A kinetic model for the early steps of RNA synthesis by human RNA polymerase II." Journal of Biological Chemistry **275**(51): 40483-40491.

Kuli, I. M. and H. Schiessel (2003). "Chromatin dynamics: nucleosomes go mobile through twist defects." Physical review letters **91**(14): 148103.

Kussell, E. and S. Leibler (2005). "Phenotypic diversity, population growth, and information in fluctuating environments." Science **309**(5743): 2075-2078.

Kwek, K. Y., S. Murphy, et al. (2002). "U1 snRNA associates with TFIIH and regulates transcriptional initiation." Nature Structural & Molecular Biology **9**(11): 800-805.

Larson, D. R., R. H. Singer, et al. (2009). "A single molecule view of gene expression." Trends in Cell Biology **19**(11): 630-637.

Lee J., Choi K., et al. (2010). "Delineating role of ubiquitination on nuclear factor-kappa B pathway by a computational modeling approach." Biochemical and Biophysical Research Communications 391 **391**: 33–37.

Lee, R. C., L. F. Rhonda, et al. (1993). "The C. elegans Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to &lin-14." Cell **75**: 12.

Lee, T. I., N. J. Rinaldi, et al. (2002). "Transcriptional regulatory networks in Saccharomyces cerevisiae." Science Signaling **298**(5594): 799.

Lee, T. I. and R. A. Young (2013). "Transcriptional Regulation and Its Misregulation in Disease." Cell **152**(6): 1237-1251.

Lee, Y., M. Kim, et al. (2004). "MicroRNA genes are transcribed by RNA polymerase II." The EMBO journal **23**(20): 4051-4060.

Levine, E., E. Ben Jacob, et al. (2007). "Target-specific and global effectors in gene regulation by MicroRNA." Biophys J **93**(11): L52-4.

Levine, M. and R. Tjian (2003). "Transcription regulation and animal diversity." Nature **424**(6945): 147-151.

Lewin, B., J. E. Krebs, et al. (2009). Essential genes, Jones & Bartlett Publishers.

Li, B., J. Vilardell, et al. (1996). "An RNA structure involved in feedback regulation of splicing and of translation is critical for biological fitness." Proceedings of the National Academy of Sciences of the United States of America **93**(4): 1596.

Lim, L. P., M. E. Glasner, et al. (2003). "Vertebrate microRNA genes." Science **299**(5612): 1540.

Liu, T., T. Papagiannakopoulos, et al. (2007). "Detection of a microRNA signal in an in vivo expression set of mRNAs." PLoS One **2**(8).

Lu, H., R. J. Buchan, et al. (2010). "MicroRNA-223 regulates Glut4 expression and cardiomyocyte glucose metabolism." Cardiovascular research **86**(3): 410.

Luca Mariani, E. G. S., Maria H Lexberg, Caroline Helmstetter, Andreas Radbruch, Max Lohning and Thomas Hofer (2010). "Short-term memory in gene induction reveals the regulatory principle behind stochastic IL-4 expression." Molecular Systems Biology **6**(359).

Luciani, F., C. Kesmir, et al. (2005). "A mathematical model of protein degradation by the proteasome." Biophysical journal **88**(4): 2422-2432.
Ma, F., X. Liu, et al. (2010). "MicroRNA-466l upregulates IL-10 expression in TLR-triggered macrophages by antagonizing RNA-binding protein tristetraprolin-mediated IL-10 MRNA degradation." The Journal of Immunology **184**(11): 6053.
Marathe, R., V. Bierbaum, et al. (2012). "Deterministic and stochastic descriptions of gene expression dynamics." Journal of Statistical Physics **148**(4): 607-626.
Marathe, R., D. Gomez, et al. (2012). "Sources of stochasticity in constitutive and autoregulated gene expression." Physica Scripta **2012**(T151): 014068.
Marguerat, S., A. Schmidt, et al. (2012). "Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells." Cell **151**(3): 671-683.
Mariani L., Schulz E.G, et al. (2010). "Short-term memory in gene induction reveals the regulatory principle behind stochastic IL-4 expression." Molecular Systems Biology **6**(359).
Martinez, N. J. and A. J. M. Walhout (2009). "The interplay between transcription factors and microRNAs in genome scale regulatory networks." Bioessays **31**(4): 435-445.
Mattick, J. S. and I. V. Makunin (2006). "Non-coding RNA." Human molecular genetics **15**(suppl 1): R17.
McCracken, S., N. Fong, et al. (1997). "5 -Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II." Genes & development **11**(24): 3306.
Milo, R., P. Jorgensen, et al. (2010). "BioNumbers- ”the database of key numbers in molecular and cell biology." Nucleic acids research **38**(suppl 1): D750-D753.
Misirli, Z., E. T. Oner, et al. (2007). "Real imaging and size values of Saccharomyces cerevisiae cells with comparable contrast tuning to two environmental scanning electron microscopy modes." Scanning **29**(1): 11-19.
Mitarai, N., A. Andersson, et al. (2007). "Efficient degradation and expression prioritization with small RNAs." Physical Biology **4**: 164.
Monnier A., Belle R., et al. (2001). "Evidence for regulation of protein synthesis at the elongation step by CDK/cyclinB phosphorylation." Nucleic Acids Research **29**(7): 1453-1457.
Mor, A. and Y. Shav-Tal (2010). "Dynamics and kinetics of nucleo-cytoplasmic mRNA export." Wiley Interdisciplinary Reviews: RNA **1**(3): 388-401.
Morris, M. D. (1991). "Factorial sampling plans for preliminary computational experiments." Technometrics **33**(2): 161-174.
Mortensen, R. D., M. Serra, et al. (2011). "Posttranscriptional activation of gene expression in Xenopus laevis oocytes by microRNAâ€“protein complexes (microRNPs)." Proceedings of the National Academy of Sciences **108**(20): 8281.
Nandi, A., C. Vaz, et al. (2009). "miRNA-regulated dynamics in circadian oscillator models." BMC Systems Biology **3**(1): 45.
Narula J., Smith A. M., et al. (2010). "Modeling Reveals Bistability and Low-Pass Filtering in the Network Module Determining Blood Stem Cell Fate." PLoS Computational Biology **6**(5).
Nayak S. , Siddiqui J.K., et al. (2011). "Modelling and analysis of an ensemble of eukaryotic translation initiation models." IET Syst. Biol. **5**(1): 2–14.
Newman, J. R. S., S. Ghaemmaghami, et al. (2006). "Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise." Nature **441**(7095): 840-846.
Nikolov, D. B., H. Chen, et al. (1996). "Crystal structure of a human TATA box-binding protein/TATA element complex." Proceedings of the National Academy of Sciences **93**(10): 4862-4867.

Noble, K. N. and S. R. Wente (2010). "Nuclear mRNA on the move." Nature cell biology **12**(6): 525-527.

Norbury, C. and P. Nurse (1992). "Animal cell cycles and their control." Annu Rev Biochem **61**: 441-70.

Noula Shembade1, A. M. a. E. W. H. (2010). "Inhibition of NF-κB Signaling by A20 Through Disruption of Ubiquitin Enzyme Complexes." Science **327** (5969): 1135-1139

Novak, B. and J. J. Tyson (1997). Modeling the control of DNA replication in fission yeast, National Acad Sciences. **94:** 9147-9152.

Novak, B. and J. J. Tyson (2004). "A model for restriction point control of the mammalian cell cycle." J Theor Biol **230**(4): 563-79.

Novick, A. and M. Weiner (1957). "Enzyme induction as an all-or-none phenomenon." Proceedings of the National Academy of Sciences of the United States of America **43**(7): 553.

Obeyesekere, M. N., E. S. Knudsen, et al. (1997). "A mathematical model of the regulation of the G1 phase of Rb+/+ and Rb-/-mouse embryonic fibroblasts and an osteosarcoma cell line." Cell Proliferation **30**(3): 171-194.

Oeffinger, M. and D. Zenklusen (2012). "To the pore and through the pore: a story of mRNA export kinetics." Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms **1819**(6): 494-506.

Ohbayashi, T., E. E. Schmidt, et al. (1996). "Promoter structure of the mouse TATA-binding protein (TBP) gene." Biochem Biophys Res Commun **225**(1): 275-80.

Onouchi H., Nagami Y., et al. (2005). "Nascent peptide-mediated translation elongation arrest coupled with mRNA degradation in the CGS1 gene of Arabidopsis." Genes & Development **19**: 1799–1810.

Orom, U. A., F. C. Nielsen, et al. (2008). "MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation." Molecular cell **30**(4): 460-471.

Ozbudak, E. M., M. Thattai, et al. (2002). "Regulation of noise in the expression of a single gene." nature genetics **31**(1): 69-73.

Padgett, R. A., P. J. Grabowski, et al. (1986). "Splicing of messenger RNA precursors." Annual review of biochemistry **55**(1): 1119-1150.

Pagano, M., R. Pepperkok, et al. (1992). "Cyclin A is required at two points in the human cell cycle." EMBO J **11**(3): 961-71.

Pain, V. M. (1996). "Initiation of protein synthesis in eukaryotic cells." European Journal of Biochemistry **236**(3): 747-771.

Park, C., T. M. Kim, et al. (2013). "Transcriptional Regulation of Endothelial Cell and Vascular Development." Circulation research **112**(10): 1380-1400.

Paulsson, J. (2004). "Summing up the noise in gene networks." Nature **427**(6973): 415-418.

Paulsson, J. (2005). "Models of stochastic gene expression." Physics of life reviews **2**(2): 157-175.

Payer, B., M. Saitou, et al. (2003). "Stella is a maternal effect gene required for normal early development in mice." Current Biology **13**(23): 2110-2117.

Peters, B., K. Janek, et al. (2002). "Assessment of proteasomal cleavage probabilities from kinetic analysis of time-dependent product formation." Journal of molecular biology **318**(3): 847-862.

Petersen-Mahrt, S. K., C. Estmer, et al. (1999). "The splicing factor-associated protein, p32, regulates RNA splicing by inhibiting ASF/SF2 RNA binding and phosphorylation." The EMBO journal **18**(4): 1014-1024.

Peterson, C. L. (2002). "Chromatin remodeling enzymes: taming the machines." EMBO reports **3**(4): 319-322.

Petrocca, F., R. Visone, et al. (2008). "E2F1-regulated microRNAs impair TGFbeta-dependent cell-cycle arrest and apoptosis in gastric cancer." Cancer Cell **13**(3): 272-86.

Pillai, R. S., S. N. Bhattacharyya, et al. (2007). "Repression of protein synthesis by miRNAs: how many mechanisms?" Trends Cell Biol **17**(3): 118-26.

Proudfoot, N. J., A. Furger, et al. (2002). "Integrating mRNA processing with transcription." Cell **108**(4): 501-512.

Pugh, B. F. and R. Tjian (1991). "Transcription from a TATA-less promoter requires a multisubunit TFIID complex." Genes Dev **5**(11): 1935-45.

Raj, A., C. S. Peskin, et al. (2006). "Stochastic mRNA synthesis in mammalian cells." PLoS Biol **4**(10): e309.

Raj, A. and A. van Oudenaarden (2008). "Nature, nurture, or chance: stochastic gene expression and its consequences." Cell **135**(2): 216-226.

Rajala T., Hakkinen A., et al. (2010). "Effects of Transcriptional Pausing on Gene Expression Dynamics." PLoS Computational Biology **6**(3).

Raney A., Law G. L., et al. (2002). "Regulated Translation Termination at the Upstream Open Reading Frame in S-Adenosylmethionine Decarboxylase mRNA." The Journal Of Biological Chemistry. **277**(8): 5988–5994.

Raser, J. M. and E. K. O'Shea (2004). "Control of stochasticity in eukaryotic gene expression." Science **304**(5678): 1811.

Raser, J. M. and E. K. O'Shea (2005). "Noise in gene expression: origins, consequences, and control." Science **309**(5743): 2010.

Ravid, T. and M. Hochstrasser (2008). "Diversity of degradation signals in the ubiquitin–proteasome system." Nature Reviews Molecular Cell Biology **9**(9): 679-689.

Reppert, S. M. and D. R. Weaver (2001). "Molecular analysis of mammalian circadian rhythms." Annual Review of Physiology **63**(1): 647-676.

Richardson, J. P. and J. W. Roberts (1993). "Transcription termination." Critical reviews in biochemistry and molecular biology **28**(1): 1-30.

Richter, J. D. (2008). "Think you know how miRNAs work? Think again." Nature Structural & Molecular Biology **15**(4): 334-336.

Rocco, A. (2009). "Stochastic control of metabolic pathways." Physical Biology **6**(1): 016002.

Rodriguez, A., S. Griffiths-Jones, et al. (2004). "Identification of mammalian microRNA host genes and transcription units." Genome research **14**(10a): 1902.

Ron Milo, P. J. a. M. S. (May 2007). BioNumbers.

Rosenfeld, N., M. B. Elowitz, et al. (2002). "Negative autoregulation speeds the response times of transcription networks." Journal of molecular biology **323**(5): 785-793.

Rosenfeld, N., J. W. Young, et al. (2005). "Gene regulation at the single-cell level." Science Signalling **307**(5717): 1962.

Roussel, M. R. and R. Zhu (2006). "Stochastic kinetics description of a simple transcription model." Bulletin of mathematical biology **68**(7): 1681-1713.

Rozenblatt-Rosen, O., T. Nagaike, et al. (2009). "The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3 mRNA processing factors." Proceedings of the National Academy of Sciences **106**(3): 755.

Sans M.D., Xie Q., et al. (2004). "Regulation of translation elongation and phosphorylation of eEF2 in rat pancreatic acini." Biochemical and Biophysical Research Communications **319**: 144–151.

Saunders, A., L. J. Core, et al. (2006). "Breaking barriers to transcription elongation." Nature Reviews Molecular Cell Biology **7**(8): 557-567.

Schmidt, E. E. and U. Schibler (1995). "High accumulation of components of the RNA polymerase II transcription machinery in rodent spermatids." Development **121**(8): 2373-83.

Schwanhausser, B., D. Busse, et al. (2011). "Global quantification of mammalian gene expression control." Nature **473**(7347): 337-342.

Schwanhausser, B., D. Busse, et al. (2013). "Corrigendum: Global quantification of mammalian gene expression control." Nature **495**(7439): 126-127.

Sedighi M. and Sengupta A. M (2008). "Epigenetic chromatin silencing: bistability and front propagation." Phys Biol. **4**(4): 246–255.

Seo, J. and K.-J. Lee (2004). "Post-translational modifications and their biological functions: proteomic analysis and systematic approaches." Journal of biochemistry and molecular biology **37**(1): 35-44.

Sevignani, C., G. A. Calin, et al. (2006). "Mammalian microRNAs: a small world for fine-tuning gene expression." Mammalian Genome **17**(3): 189-202.

Shahrezaei, V., J. F. Ollivier, et al. (2008). "Colored extrinsic fluctuations and stochastic gene expression." Molecular systems biology **4**(1).

Shahrezaei, V. and P. S. Swain (2008). "Analytical distributions for stochastic gene expression." Proceedings of the National Academy of Sciences **105**(45): 17256-17261.

Sheppard, P. W., M. Rathinam, et al. (2012). "A pathwise derivative approach to the computation of parameter sensitivities in discrete stochastic chemical systems." The Journal of Chemical Physics **136**: 034115.

Shimoni, Y., G. Friedlander, et al. (2007). "Regulation of gene expression by small non-coding RNAs: a quantitative view." Mol Syst Biol **3**: 138.

Shis, D. L. and M. R. Bennett (2013). "Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants." Proceedings of the National Academy of Sciences.

Sigal, A., R. Milo, et al. (2006). "Variability and memory of protein levels in human cells." Nature **444**(7119): 643-646.

Singh, J. and R. A. Padgett (2009). "Rates of in situ transcription and splicing in large human genes." Nature Structural & Molecular Biology **16**(11): 1128-1133.

Singh, S., H. Y. O. Yang, et al. (2007). "A kinetic-dynamic model for regulatory RNA processing." Journal of biotechnology **127**(3): 488-495.

Skjondal-Bar, N. and D. R. Morris (2007). "Dynamic model of the process of protein synthesis in Eukaryotic cells." Bulletin of mathematical biology **69**(1): 361-393.

Smolen, P., D. A. Baxter, et al. (2000). "Modeling transcriptional control in gene networks—methods, recent results, and future directions." Bulletin of mathematical biology **62**(2): 247-292.

Sonenberg, N. and A. G. Hinnebusch (2009). "Regulation of translation initiation in eukaryotes: mechanisms and biological targets." Cell **136**(4): 731-745.

Spudich, J. L. and D. E. Koshland Jr (1976). "Non-genetic individuality: chance in the single cell." Nature **262**(5568): 467.

Stark, A., J. Brennecke, et al. (2005). "Animal MicroRNAs confer robustness to gene expression and have a significant impact on 30UTR evolution." Cell **123**(6): 1133-1146.

Stewart, M. (2007). "Ratcheting mRNA out of the nucleus." Molecular cell **25**(3): 327-330.

Storz, G., S. Altuvia, et al. (2005). "An abundance of RNA regulators*." Annu. Rev. Biochem. **74**: 199-217.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**(6765): 41-45.

Strubin, M. and K. Struhl (1992). "Yeast and human TFIID with altered DNA-binding specificity for TATA elements." Cell **68**(4): 721-730.

Suter, D. M., N. Molina, et al. (2011). "Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics." Science **332**(6028): 472.

Swain, P. S., M. B. Elowitz, et al. (2002). "Intrinsic and extrinsic contributions to stochasticity in gene expression." Proceedings of the National Academy of Sciences of the United States of America **99**(20): 12795.

Taniguchi, Y., P. J. Choi, et al. (2010). "Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells." Science **329**(5991): 533.

Thattai, M. and A. Van Oudenaarden (2001). "Intrinsic noise in gene regulatory networks." Proceedings of the National Academy of Sciences of the United States of America **98**(15): 8614.

Thattai, M. and A. Van Oudenaarden (2004). "Stochastic gene expression in fluctuating environments." Genetics **167**(1): 523-530.

Thomas, M. C. and C. M. Chiang (2006). "The general transcription machinery and general cofactors." Critical reviews in biochemistry and molecular biology **41**(3): 105-178.

Tjian, R. (1996). "The biochemistry of transcription in eukaryotes: a paradigm for multisubunit regulatory complexes." Philosophical Transactions: Biological Sciences: 491-499.

To, T.-L. and N. Maheshri (2010). "Noise can induce bimodality in positive transcriptional feedback loops without bistability." Science **327**(5969): 1142-1145.

Tripathi, T. and D. Chowdhury (2008). "Interacting RNA polymerase motors on a DNA track: Effects of traffic congestion and intrinsic noise on RNA synthesis." Physical Review E **77**(1): 11921.

Tripathi, T. and D. Chowdhury (2008). "Transcriptional bursts: A unified model of machines and mechanisms." EPL (Europhysics Letters) **84**: 68004.

Tsang, J., J. Zhu, et al. (2007). "MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals." Molecular cell **26**(5): 753-767.

van Roon-Mom, W. M. C., S. J. Reid, et al. (2005). "TATA-binding protein in neurodegenerative disease." Neuroscience **133**(4): 863-872.

Vasisht, R. T. (2006). "Thermodynamic and kinetic modeling of transcriptional pausing." Proceedings of the National Academy of Sciences of the United States of America **103**(12): 4439.

Vasudevan, S., Y. Tong, et al. (2007). "Switching from repression to activation: microRNAs can up-regulate translation." Science **318**(5858): 1931.

Vasudevan, S., Y. Tong, et al. (2008). "Cell-cycle control of microRNA-mediated translation regulation." Cell Cycle **7**(11): 1545-9.

Velculescu, V. E., L. Zhang, et al. (1997). "Characterization of the Yeast Transcriptome." Cell **88**(2): 243-251.

Venters, B. J. and B. F. Pugh (2009). "How eukaryotic genes are transcribed." Critical reviews in biochemistry and molecular biology **44**(2-3): 117-141.

Verma, M., S. Rawool, et al. (2006). "Biological significance of autoregulation through steady state analysis of genetic networks." Biosystems **84**(1): 39-48.

Vervoorts J., Luscher-Firzlaff J., et al. (2006). "The Ins and Outs of MYC Regulation by Posttranslational Mechanisms." The Journal Of Biological Chemistry **281**(46): 34725–34729.

Vohradsky, J., J. Panek, et al. (2010). "Numerical modelling of microRNA-mediated mRNA decay identifies novel mechanism of microRNA controlled mRNA downregulation." Nucleic acids research **38**(14): 4579.

Volfson, D., J. Marciniak, et al. (2005). "Origins of extrinsic variability in eukaryotic gene expression." Nature **439**(7078): 861-864.

Voliotis, M., N. Cohen, et al. (2008). "Fluctuations, pauses, and backtracking in DNA transcription." Biophysical journal **94**(2): 334-348.

Volpe, T. A., C. Kidner, et al. (2002). "Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi." Science **297**(5588): 1833.

Von Der Haar, T. (2008). "A quantitative estimation of the global translational activity in logarithmically growing yeast cells." BMC Systems Biology **2**(1): 87.

von Heijne, G., L. Nilsson, et al. (1978). "Models for mRNA translation: theory versus experiment." Eur. J. Biochem **92**: 397-402.

von Hippel, P. H. (1998). "An integrated model of the transcription complex in elongation, termination, and editing." Science **281**(5377): 660.

von Hippel, P. H. and T. D. Yager (1991). "Transcript elongation and termination are competitive kinetic processes." Proceedings of the National Academy of Sciences **88**(6): 2307.

von Hippel, P. H. and T. D. Yager (1992). "The elongation-termination decision in transcription." Science **255**(5046): 809.

von Roretz, C. and I. E. Gallouzi (2008). "Decoding ARE-mediated decay: is microRNA part of the equation?" Journal of Cell Biology **181**(2): 189.

Wang, H., A. Iacoangeli, et al. (2002). "Dendritic BC1 RNA: functional role in regulation of translation initiation." Journal of Neuroscience **22**(23): 10232.

Wang X. and Proud C.G. (2008). "A Novel Mechanism for the Control of Translation Initiation by Amino Acids, Mediated by Phosphorylation of Eukaryotic Initiation Factor 2B." Molecular and Cellular Biology **28**(5): 1429–1442.

Wei, Z., L. Yan, et al. (2011). "MiR-206-mediated dynamic mechanism of the mammalian circadian clock." BMC Systems Biology **5**: 141.

Weideman, C. A., R. C. Netter, et al. (1997). "Dynamic interplay of TFIIA, TBP and TATA DNA." J Mol Biol **271**(1): 61-75.

White, R. J., S. P. Jackson, et al. (1992). "A role for the TATA-box-binding protein component of the transcription factor IID complex as a general RNA polymerase III transcription factor." Proceedings of the National Academy of Sciences **89**(5): 1949-1953.

Winter, J., S. Jung, et al. (2009). "Many roads to maturity: microRNA biogenesis pathways and their regulation." Nat Cell Biol **11**(3): 228-34.

Xie, Z., K. D. Kasschau, et al. (2003). "Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation." Current Biology **13**(9): 784-789.

Xie, Z. R., H. T. Yang, et al. (2007). "The role of microRNA in the delayed negative feedback regulation of gene expression." Biochem Biophys Res Commun **358**(3): 722-6.

Yamada, Y. R. and C. S. Peskin (2009). "A look-ahead model for the elongation dynamics of transcription." Biophysical journal **96**(8): 3015-3031.

Yang, H., W. Kong, et al. (2008). "MicroRNA expression profiling in human ovarian cancer: miR-214 induces cell survival and cisplatin resistance by targeting PTEN." Cancer research **68**(2): 425.

Yang X.O., Angkasekwinai P., et al. (2009). "Requirement for the basic helix-loop-helix transcription factor Dec2 in initial TH2 lineage commitment." Nature Immunology **10**: 1260 - 1266.

Yazgan, O. and J. E. Krebs (2007). "Noncoding but nonexpendable: transcriptional regulation by large noncoding RNA in eukaryotes." Biochemistry and Cell Biology **85**(4): 484-496.

Ying, S.-Y. and S.-L. Lin (2005). "MicroRNA: fine-tunes the function of genes in zebrafish." Biochemical and biophysical research communications **335**(1): 1-4.

Ying, S. Y. and S. L. Lin (2004). "Intron-derived microRNAs—fine tuning of gene functions." Gene **342**(1): 25-28.

You, T., G. M. Coghill, et al. (2010). "A quantitative model for mRNA translation in Saccharomyces cerevisiae." Yeast.

Young, J. S., W. F. Ramirez, et al. (1997). "Modeling and optimization of a batch process for in vitro RNA production." Biotechnology and bioengineering **56**(2): 210-220.

Zenklusen, D., D. R. Larson, et al. (2008). "Single-RNA counting reveals alternative modes of gene expression in yeast." Nature Structural & Molecular Biology **15**(12): 1263-1271.

Zhdanov, V. (2008). "Switches in gene expression including microRNA and a large number of distinct mRNAs." JETP Letters **88**(7): 466-469.

Zhdanov, V. P. (2009). "Conditions of appreciable influence of microRNA on a large number of target mRNAs." Mol. BioSyst. **5**(6): 638-643.

Zi, Z. (2011). "Sensitivity analysis approaches applied to systems biology models." Systems Biology, IET **5**(6): 336-346.

Zi, Z., K. H. Cho, et al. (2005). "In silico identification of the key components and steps in IFN-Î³ induced JAK-STAT signaling pathway." FEBS letters **579**(5): 1101-1108.

Zouridis, H. and V. Hatzimanikatis (2007). "A model for protein translation: polysome self-organization leads to maximum protein synthesis rates." Biophysical journal **92**(3): 717-730.

Appendices

Appendix Ia – Concentration range for total TBP and total TBP binding sites

	Yeast Cell Radius: 1 µm (Misirli, Oner et al. 2007) and references therein Volume: $4.17*10^{-15}$ lit	**Mammalian Cell** Radius: 7.5 µm (Ron Milo May 2007) (http://bionumbers.hms.harvard.edu) Volume: $1.76*10^{-12}$ lit	**Sea urchin egg Cell** Radius: 50 µm (Alberts, Bray et al. 2002) Volume: $5.22*10^{-10}$ lit
TBP Concentration	4800 molecules = $1.91*10^{-6}$M 20,000 molecules = $7.96*10^{-6}$M $2*10^{9}$ molecules = 0.7M	4800 molecules = $4.5*10^{-9}$M 20,000 molecules = $1.88*10^{-8}$M $2*10^{9}$ molecules = $1.9*10^{-3}$M	4800 molecules = $1.5*10^{-11}$M 20,000 molecules = $6.3*10^{-11}$M $2*10^{9}$ molecules = $6.3*10^{-6}$M
TBP binding site concentration	3000 sites = $1.19*10^{-6}$M 10000 sites = $3.98*10^{-6}$M 80000 sites = $3.185*10^{-5}$M	3000 sites = $2.8*10^{-9}$M 10000 sites = $9.4*10^{-9}$M 80000 sites = $7.5*10^{-8}$M	3000 sites = $9.5*10^{-12}$M 10000 sites = $3.2*10^{-11}$M 80000 sites = $2.5*10^{-10}$M

Appendix Ib – Dimensionless parameter for mathematical model of TBP

Parameter with dimension	**Dimensionless parameter**
$k_1\ (M^{-1}s^{-1})$	$\kappa_1 = k_1 \frac{D_0}{k_8}$
$k_2\ (s^{-1})$	$\kappa_2 = \frac{k_2}{k_8}$
$k_3\ (M^{-1}s^{-1})$	$\kappa_3 = k_3 \frac{D_0}{k_8}$
$k_4\ (s^{-1})$	$\kappa_4 = \frac{k_4}{k_8}$
$k_5\ (Ms^{-1})$	$\kappa_5 = \frac{k_5}{D_0 k_8}$

$k_7\ (M)$	$\kappa_7 = \frac{k_7}{D_0}$
$k_0\ (Ms^{-1})$	$\kappa_0 = \frac{k_0}{D_0 k_8}$

Appendix Ic – Sensitivity for k_3 and k_5 for wide range of parameters

Table states ranges of D_0 concentration, k_5, k_7 and corresponding total TBP concentration, sensitivity to single parameter change, response time ratio, and the ratio of dimer concentration to TBP-DNA complex concentration.

Sr. No.	D_0 (M)	k_5 (M/s)	k_7 (M)	Total TBP (M)	Decrease in high-TBP state concentration		Ratio of response time by computational method	$[T_2]/[TD]$
					Sensitivity to k_3	Sensitivity to k_6		
1	2.50E-05	5.00E-11	1.25E-08	0.000117	0.08%	0.08%	2.62E-04	1.83E+00
2	2.50E-05	5.00E-12	1.25E-08	2.53E-05	2.69%	Not detectable	6.57E-01	1.88E-02
3	2.50E-05	5.00E-13	1.25E-05	1.72E-05	43.00%	6.53%	9.99E-01	1.14E-04
4	2.50E-05	5.00E-13	1.25E-06	1.93E-05	18.00%	1.45%	9.98E-01	2.35E-04
5	2.50E-05	5.00E-13	1.25E-07	1.93E-05	18.00%	0.10%	9.98E-01	2.37E-04
6	2.50E-05	5.00E-13	1.25E-08	1.93E-05	18.00%	0.05%	9.98E-01	2.37E-04
7	2.50E-05	5.00E-13	1.25E-09	1.93E-05	19.00%	Not detectable	9.98E-01	2.37E-04
8	2.50E-05	5.00E-14	1.25E-08	6.31E-06	42.00%	0.14%	1.00E+00	7.23E-06
9	2.50E-06	5.00E-12	1.25E-08	3.41E-06	1.99%	0.29%	2.14E-03	1.88E-01
10	2.50E-06	5.00E-13	1.25E-07	1.94E-06	18.00%	1.03%	9.76E-01	2.35E-03
11	2.50E-06	5.00E-13	1.25E-08	1.94E-06	18.00%	0.15%	9.76E-01	2.37E-03
12	2.50E-06	5.00E-14	1.25E-08	6.32E-07	42.00%	1.45%	9.99E-01	7.23E-05
13	2.50E-07	5.00E-12	1.25E-08	1.22E-06	0.50%	7.06%	2.98E-03	1.87E+00
14	2.50E-07	5.00E-13	1.25E-07	1.81E-07	43.93%	5.20%	8.37E-01	1.14E-02
15	2.50E-07	5.00E-13	1.25E-08	2.09E-07	17.41%	1.96%	1.58E-02	2.35E-02
16	2.50E-07	5.00E-13	1.25E-09	2.09E-07	17.24%	0.23%	1.65E-02	2.37E-02
17	2.50E-07	5.00E-14	1.25E-08	6.20E-08	47.40%	12.11%	9.93E-01	6.87E-04
18	2.50E-08	5.00E-12	1.25E-08	6.44E-07	3.00%	28.00%	4.07E-03	1.18E+01
19	2.50E-08	5.00E-13	1.25E-08	2.56E-08	47.44%	9.15%	1.08E+00	1.14E-01
20	2.50E-08	5.00E-13	1.25E-09	3.51E-08	10.41%	4.76%	1.19E+00	2.35E-01
21	2.50E-09	5.00E-11	1.25E-08	1.21E-07	33.70%	21 times increase	6.57E-02	2.11E+01
22	2.50E-09	5.00E-13	1.25E-09	1.01E-08	52.02%	22.70%	1.43E+00	1.14E+00
23	2.50E-09	5.00E-13	1.25E-10	1.77E-08	2.25%	10.15%	1.51E+00	2.35E+00

Appendix Id – Perturbation Study for TBP system by analytical method

In the neighbourhood of a steady state, for the three variable system of TBP (equations2.8 to 2.10, chapter 1), the rate of change of perturbation (σ_i), where perturbation for variable *i* is defined as [*i*]-[$i_{\_steady\ state}$], can be given as (Heinrich, Rapoport et al. 1977),

$$\frac{d\sigma_T}{dt} = \frac{\partial f1}{\partial [T]}\sigma_T + \frac{\partial f1}{\partial [T_2]}\sigma_{T_2} + \frac{\partial f1}{\partial [TD]}\sigma_{TD} = a_{11}\times\sigma_T + a_{12}\times\sigma_{T_2} + a_{13}\times\sigma_{TD}$$

$$\frac{d\sigma_{T_2}}{dt} = \frac{\partial f2}{\partial [T]}\sigma_T + \frac{\partial f2}{\partial [T_2]}\sigma_{T_2} + \frac{\partial f2}{\partial [TD]}\sigma_{TD} = a_{21}\times\sigma_T + a_{22}\times\sigma_{T_2} + a_{23}\times\sigma_{TD}$$

$$\frac{d\sigma_{TD}}{dt} = \frac{\partial f3}{\partial [T]}\sigma_T + \frac{\partial f3}{\partial [T_2]}\sigma_{T_2} + \frac{\partial f3}{\partial [TD]}\sigma_{TD} = a_{31}\times\sigma_T + a_{32}\times\sigma_{T_2} + a_{33}\times\sigma_{TD}$$

Where, a_{ij} are elements of Jacobian matrix J,

$$\begin{pmatrix} -4\times k_1\times[T] - k_3\times[D_0] + k_3\times[TD] - k_8 & 2\times k_2 & k_3\times[T] + k_4 + \dfrac{k_5\times k_6\times {k_7}^{k_6}\times[TD]^{k_6-1}}{\left({k_7}^{k_6} + [TD]^{k_6}\right)^2} \\ 2\times k_1\times[T] & -k_2 & 0 \\ k_3\times[D_0] - k_3\times[TD] & 0 & -k_3\times[T] - k_4 \end{pmatrix}$$

The general solution for the above system is given by,

$$\sigma(t) = c1\times e^{\lambda_1 t}\times v_1 + c2\times e^{\lambda_2 t}\times v_2 + c3\times e^{\lambda_3 t}\times v_3$$

Where λ_1, λ_2 and λ_3 are three eigenvalues and v_1, v_2 and v_3 are corresponding eigenvectors of J. Analytical expression for eigenvectors was obtained in terms of eigenvalues and reaction rate constants.

The eigenvectors are,

$$v_1 = \begin{pmatrix} \dfrac{a_{12} + a_{13}\left(\dfrac{(\lambda_1 - a_{11})(\lambda_1 - a_{22}) - a_{12} \times a_{21}}{a_{13} \times a_{21} + a_{23}(\lambda_1 - a_{11})}\right)}{(\lambda_1 - a_{11})} \\ 1 \\ \dfrac{(\lambda_1 - a_{11})(\lambda_1 - a_{22}) - a_{12} \times a_{21}}{a_{13} \times a_{21} + a_{23}(\lambda_1 - a_{11})} \end{pmatrix}$$

$$v_2 = \begin{pmatrix} \dfrac{a_{12} + a_{13}\left(\dfrac{(\lambda_2 - a_{11})(\lambda_2 - a_{22}) - a_{12} \times a_{21}}{a_{13} \times a_{21} + a_{23}(\lambda_2 - a_{11})}\right)}{(\lambda_2 - a_{11})} \\ 1 \\ \dfrac{(\lambda_2 - a_{11})(\lambda_2 - a_{22}) - a_{12} \times a_{21}}{a_{13} \times a_{21} + a_{23}(\lambda_2 - a_{11})} \end{pmatrix}$$

$$v_3 = \begin{pmatrix} \dfrac{a_{12} + a_{13}\left(\dfrac{(\lambda_3 - a_{11})(\lambda_3 - a_{22}) - a_{12} \times a_{21}}{a_{13} \times a_{21} + a_{23}(\lambda_3 - a_{11})}\right)}{(\lambda_3 - a_{11})} \\ 1 \\ \left(\dfrac{(\lambda_3 - a_{11})(\lambda_3 - a_{22}) - a_{12} \times a_{21}}{a_{13} \times a_{21} + a_{23}(\lambda_3 - a_{11})}\right) \end{pmatrix}$$

In the perturbation study, the initial perturbation for TBP is 10% of steady state free TBP concentration and zero for the other two species. From these initial conditions, analytical expressions for c1, c2 and c3 were obtained.

From, the general solution and the above expression for eigenvectors,

$$\sigma_T(t) = B_{11} \times e^{\lambda_1 t} + B_{12} \times e^{\lambda_2 t} + B_{13} \times e^{\lambda_3 t}$$

Where, $B_{11} = \dfrac{-(x(k_2 + \lambda_1)(k_2 + [D_0] \times k_3 + k_8 + 4 \times k_1 \times [T] - k_3 \times [TD] + \lambda_2 + \lambda_3))}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)}$,

$B_{12} = \dfrac{x(k_2 + \lambda_2)(k_2 + [D_0] \times k_3 + k_8 + 4 \times k_1 \times [T] - k_3 \times [TD] + \lambda_1 + \lambda_3)}{(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)}$ and

$B_{13} = \dfrac{-(x(k_2 + \lambda_3)(k_2 + [D_0] \times k_3 + k_8 + 4 \times k_1 \times [T] - k_3 \times [TD] + \lambda_1 + \lambda_2))}{(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)}$

Here, *x* is initial perturbation in [T].

In this case the response time is the time required to recover 90% of the perturbation. In this case, therefore, $\sigma_T(t)$ is 10% of the initial perturbation.

We compared the response time obtained for the two systems. The response time obtained with this method was found to be almost same to that obtained by computational method (Figure A1). The correlation coefficient between the results obtained by the two methods was 0.99.

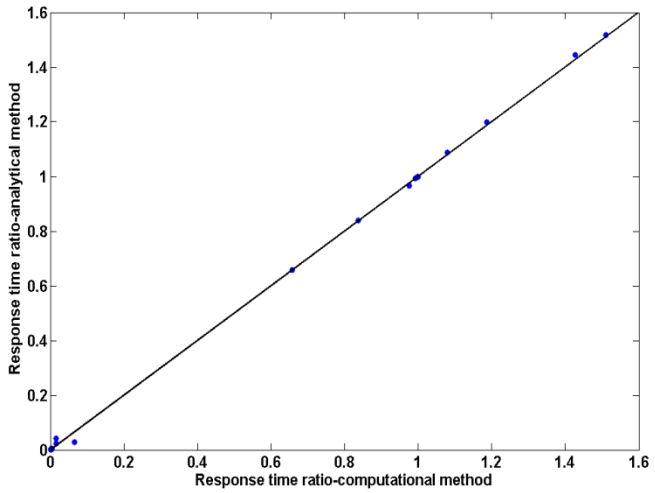


Figure A1 – Graph of ratio of response time by analytical method vs. ratio of response time by computational method.

Appendix IIa – Representative parameter scheme for calculation of relative noise

For reported values of reaction rate parameter, the values of four dimensionless numbers *a*, *b*, *c*, and *d* are 4, 0.6, 363.64, and 0.5 respectively.

For these values of dimensionless numbers following scheme of reaction rate parameters was used

a		***c***				***d***	
k_1	k_2	k_1	k_3	k_1	k_8	k_5	k_6
0.0002	0.001	0.0002	100	0.0002	0.000025	0.002	0.0002
0.002	0.01	0.002	10	0.002	0.00025	0.0025	0.00025
0.004	0.02	0.004	5	0.004	0.0005	0.003333	0.000333
0.02	0.1	0.02	1	0.02	0.0025	0.005	0.0005
0.1	0.5	0.1	0.2	0.1	0.0125	0.01	0.001
0.2	1	0.2	0.1	0.2	0.025	0.02	0.002
2	10	2	0.01	2	0.25	0.03	0.003
						0.04	0.004
						0.05	0.005

Dimensionless number *b* was changed by changing parameter q_2.

Using such scheme for varying reaction rate parameters 23 sets of parameters were generated ensuring that all the dimensionless numbers are varied for each ratio value.

	a	***b***	***c***	***d***	***r***
set2	4.00	0.60	363.64	0.50	0.1257
set8	4.00	0.40	363.64	0.50	0.1257
set14	4.00	0.60	4.02095	0.25	0.1257
set17	3.20	0.60	480.296	0.40	0.1257
set22	4.00	0.60	363.64	1.00	0.2506
set23	2.40	0.60	3.16339	0.30	0.2506
set24	4.00	0.40	363.64	1.00	0.2506
set11	4.00	0.60	363.64	2.00	0.5004
set13	4.00	0.6	0.554175	1	0.5004
set15	4.00	0.80	390.234	2.00	0.5004
set18	3.20	0.60	0.465752	0.40	0.5004
set3	4.00	0.60	363.64	5.00	1.2498
set6	4.00	0.80	363.64	5.00	1.2498
set9	4.00	0.40	363.64	5.00	1.2498
set16	4.00	0.60	0.269821	6.00	1.2498

set19	3.20	0.60	480.296	4.00	1.2498
set4	4.00	0.60	363.64	10.00	2.4988
set20	3.20	0.60	480.296	8.00	2.4988
set21	4.00	0.40	346.914	10.00	2.4988
set10	4.00	0.40	363.64	10.00	2.4989
set5	4.00	0.80	363.64	0.50	0.1258
set12	4.00	0.60	363.64	7.00	1.7494
set7	4.00	0.80	363.64	10.00	2.4987

Appendix IIb – Incorporation of constant *a*verage inhibition

The ratio of average expression in the presence of miRNA mediated regulation and that in the absence of miRNA regulation was considered as average inhibition. To incorporate average inhibition the original rate expression was modified as,

$$cdc14 = (\text{Rate expression for cdc14 used in Tyson model}) \times average\_inhibition$$

$$TFE = (\text{Rate expression for TFE used in Tyson model}) \times average\_inhibition$$

The target protein synthesis rates in basic model were reduced by the average inhibition by miRNA dependent reduction for the respective targets. The average inhibition for cdc14 for all three km_1 values was up to 10%. Repressing cdc14 to higher values results in perturbation in cell cycle. On the other hand, average inhibition observed for TFE for the two km_2 values was up to 60%. Figure A1 shows the time course profiles of cyclin A, B and E for average inhibition.

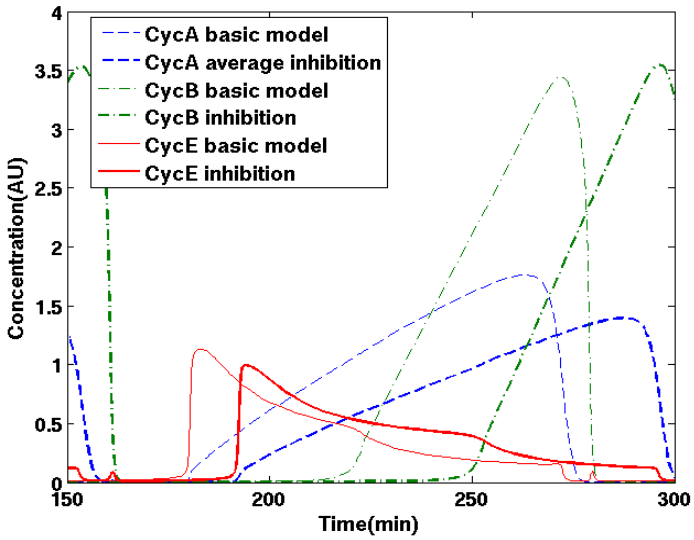
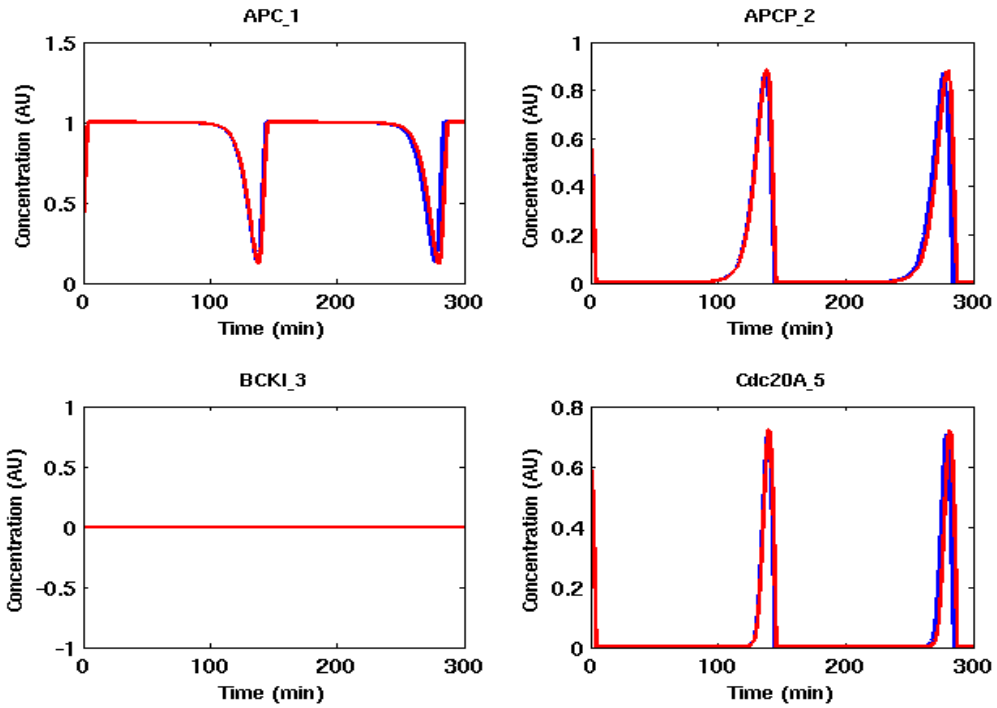


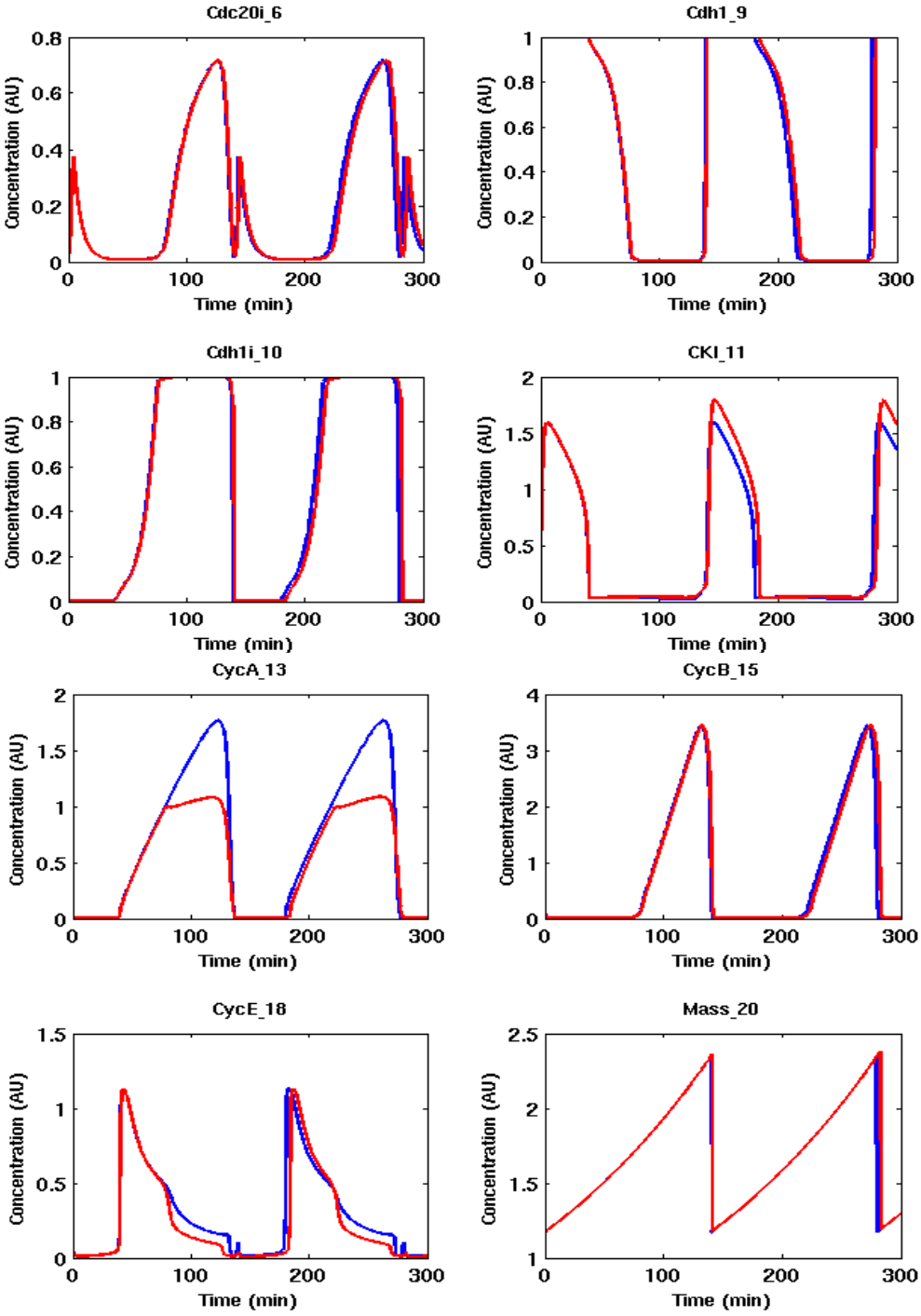
Figure A1: Comparison of time course simulation profiles of cyclin proteins for cell cycle model and cell cycle with average inhibition

Average inhibition of cdc14: cdc14 = cdc20A*0.9214

Average inhibition of TFE: TFE = GK(Vatf,Vitf,Jatf,Jitf)*0.6779)

Appendix IIc – Time course profiles for 16 species in cell cycle model




Cdc20i_6
Concentration (AU)
Time (min)
Cdh1_9
Concentration (AU)
Time (min)
Cdh1i_10
Concentration (AU)
Time (min)
CKI_11
Concentration (AU)
Time (min)
CycA_13
Concentration (AU)
Time (min)
CycB_15
Concentration (AU)
Time (min)
CycE_18
Concentration (AU)
Time (min)
Mass_20
Concentration (AU)
Time (min)

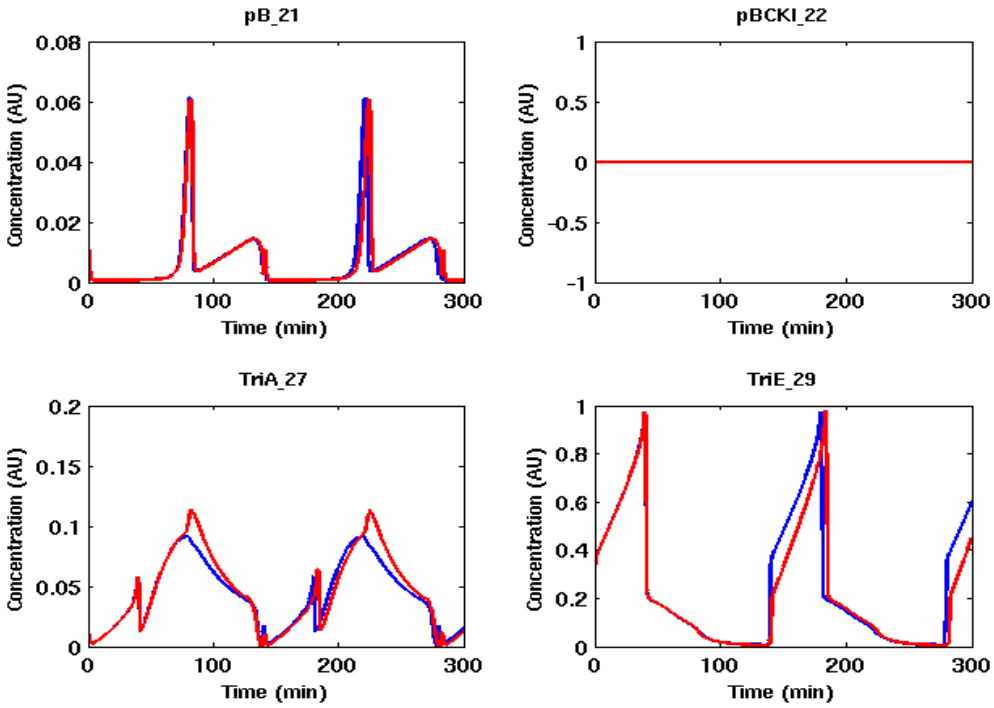


Figure A2 -Time course profiles of unmodified model and model with hsa-miR-25 regulation for all 16 species

Blue line indicates profile of speceies in unmodified model; red line indicates profile of species in model with hsa-miR-25 regulation.

Appendix IId – Time course profile of target protein with intronic miRNA regulation and with host protein mediated inhibition when the inhibitory effect is removed

Time course profile of target protein with intronic miRNA regulation and with host protein mediated inhibition when the inhibitory effect is removed by setting the specific rate of synthesis (k_2 and k_5) to zero. Figure A3 shows the time course profiles of target protein for different initial conditions.

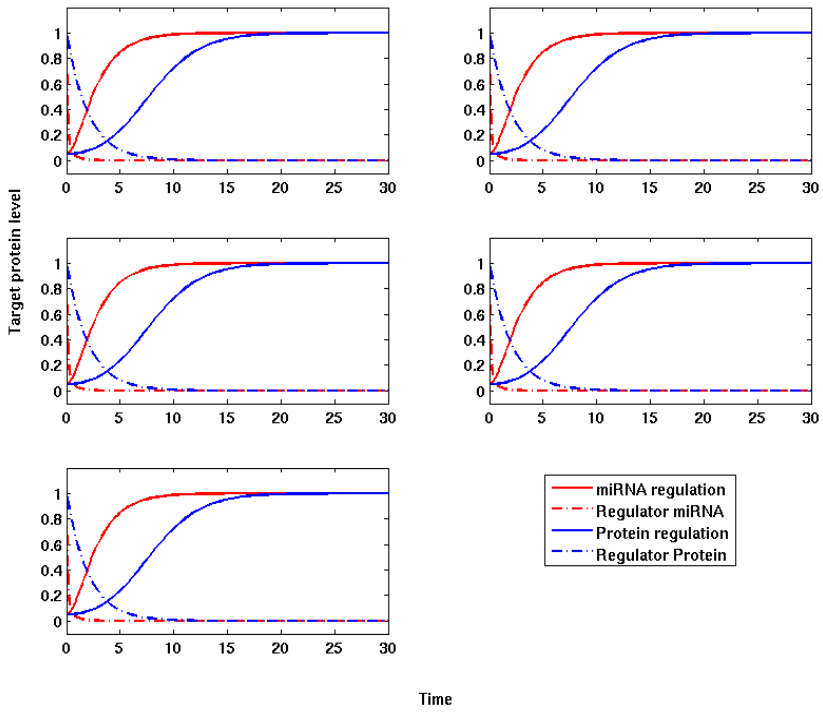


Figure A3: Time course profile of target protein with intronic miRNA regulation and with host protein mediated inhibition when the inhibitory effect is removed

Initial conditions – (a) all components at zero level. (b) Target protein at unregulated steady state, all other components at zero level. (c) Target mRNA at unregulated steady state level, all other components at zero level. (d) Target mRNA and protein at unregulated steady state, all other components at zero level. (e) Regulator molecule (miRNA and host protein) at steady state, all other components at zero level.

It was observed that the target protein profiles were qualitatively similar. The observed quantitative difference can be attributed to different half lives of miRNA and regulator protein. Therefore, the effect on the time course profiles was analysed by changing the degradation constant of miRNA and protein. Figure A4 shows time course profile of target protein with intronic miRNA regulation and with host protein mediated inhibition when the inhibitory effect is removed by blocking the synthesis of miRNA and regulator protein. The degradation reaction rate constant for regulator protein (k_{10}) is set to the same value as the miRNA degradation reaction rate constant (k_s).

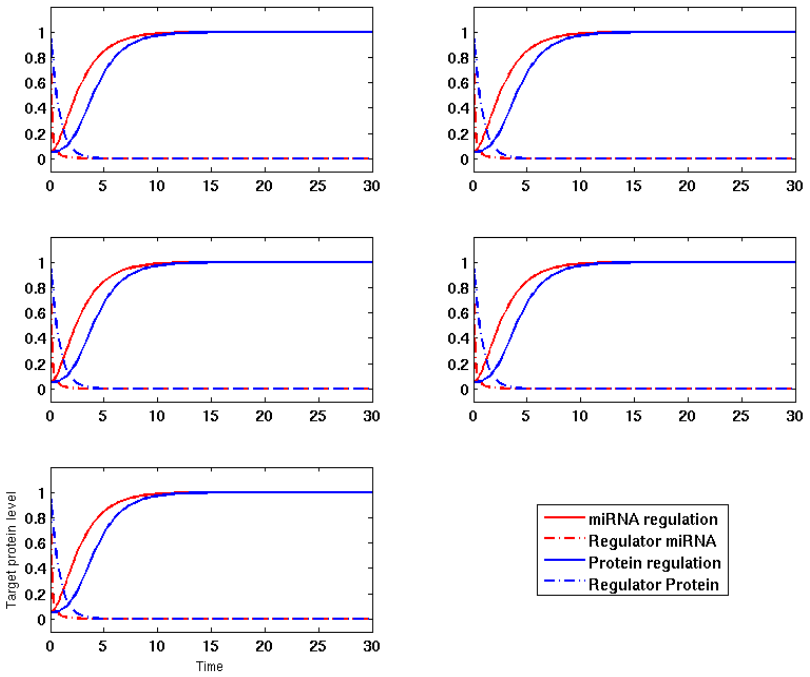


Figure A4: Time course profile of target protein with intronic miRNA regulation and with host protein mediated inhibition when the inhibitory effect is removed for changed specific rate of degradation.

Initial conditions are same as in FigureA3.

Appendix IIIa – Detailed list of mathematical models of gene expression

Table contains some details of mathematical models of gene expression. Modeling approach states whether the model is deterministic or stochastic, discrete or continuous, and analyzed at steady state or the kinetics is studied. The output column lists the major finding of the study.

Process	Model	Modeling approach	Processes considered	Output
Chromatin remodeling and histone modification	(Blossey and Schiessel 2008)	Deterministic/ Continuous/ Steady state	Histone tail modification	A quantitative model to test the role of histone tail modification in kinetic proofreading of gene activation
	(Kuli and Schiessel 2003)	Deterministic/ Discrete/ Equilibrium	Nucleosome sliding	Estimation of mobility of nucleosomes as function of thermal motion of DNA and underlying base pair sequence
	(Chou 2007)	Stochastic/ Discrete/ Kinetic	Histone sliding and unwrapping	Calculation of mean histone detachment time and mean detachment distance as a function of remodeler motor speed
	(Boeger, Griesenbeck et al. 2008)	Stochastic/ Discrete/ Kinetic	Nucleosome sliding	Identification of nucleosome disassembly to be a rate limiting step, particularly for PHO5 promoter
	(Raj, Peskin et al. 2006)	Stochastic/ Discrete/ Kinetic	Gene activation and inactivation	Statistical properties of transcriptional burst and effect of gene activation kinetics of burst size and frequency
	(Kim H. D and O'Shea E.K 2008)	Deterministic/ Continuous/ Kinetic	Nucleosome association, dissociation and gene activation PHO5 gene	Relationship between transcription factor affinity for DNA and gene expression
	(Mariani L., Schulz E.G et al. 2010)	Stochastic/ Discrete/ Kinetic	Gene opening and closing for il4 gene	Study of cell to cell variability as a result of stochastic gene opening and closing events
	(Narula J., Smith A. M. et al. 2010)	Deterministic/ Continuous/ Kinetic	Enhancement in gene expression particularly for Scl-Gata2-Fli1 triad network module	Study of steady state and transient of components in the triad, effect of mutation in the regulatory region on the components of this network module
	(Sedighi M. and Sengupta A. M 2008)	Stochastic/ Continuous/ Steady state	Positive feedback by chromatin remodeling proteins	Study of effect of change in chromatin silencing on stability of chromatin state
Transcription	(Young, Ramirez et al. 1997)	Deterministic/ Continuous/ Kinetic	mRNA synthesis	Prediction of optimal conditions for in vitro mRNA synthesis, in terms of NTP and Mg ion concentration
	(Arnold, Siemann et al. 2001)	Deterministic/ Continuous/ Kinetic	Initiation and elongation	Study of effect of RNA polymerase, promoter and nucleotide concentration on mRNA synthesis rate
	(Bai, Shundrovsky et al. 2004)	Deterministic/ Discrete/ Equilibrium	Elongation	Prediction of back-tracking and pauses during transcription depending upon the sequence information
	(Höfer and Malte 2005)	Stochastic/ Discrete/ Steady state	Initiation	Study of effect of kinetics of PIC assembly on mRNA noise
	(Yamada and Peskin 2009)	Deterministic/ Discrete/ Kinetic	Elongation with look-ahead feature of polymerase	Estimation of window size for RNA polymerase
	(Roussel and Zhu 2006)	Stochastic/ Discrete/ Kinetic	Elongation	Identification of probability distribution of transcriptional delay and elongation rate
	(Voliotis, Cohen et al. 2008)	Stochastic/ Discrete/ Steady state	Initiation and elongation	Analytical expression for transcription time
	(Tripathi and Chowdhury	Stochastic/ Discrete/	Elongation with RNA polymerase traffic	Identification of mRNA synthesis rate and RNA polymerase average density

	2008; Tripathi and Chowdhury 2008)	Kinetic		
	(von Hippel and Yager 1991)	Deterministic/ Continuous/ Equilibrium	Elongation and termination	Prediction of elongation and termination phases during transcription depending upon sequence information
	(von Hippel and Yager 1992)	Deterministic/ Continuous/ Equilibrium	Elongation and termination	Prediction of elongation, back-tracking and termination efficiency
	(von Hippel 1998)	Deterministic/ Continuous/ Equilibrium	Elongation, termination and editing	Prediction of control at any template site and error correction
	(Guajardo and Sousa 1997)	Deterministic/ Continuous/ Equilibrium	Elongation	Identification of force generation by polymerase translocation as a function of energy available from NTP binding
	(Vasisht 2006)	Stochastic/ Continuous/ Equilibrium	Elongation	Prediction of polymerase pause and back-tracking
	(Konishi 2005)	Deterministic/ Continuous/ Equilibrium	mRNA synthesis and degradation	General relation between transcript level produced by sequence specific interaction between DNA and protein factors
	(Kugel and Goodrich 2000)	Deterministic/ Continuous/ Kinetic	Initiation specific at AdMLP	Identification of promoter escape to be a rate limiting step in transcription initiation
	(Bernard S., Cajavec B. et al. 2006)	Deterministic/ Continuous/ Kinetic	Transcription regulation with feedback for Hes1 transcription factor	Mathematical model transcriptional regulation by feedback loop, which shows oscillations in Hes1 expression
	(Rajala T., Hakkinen A. et al. 2010)	Stochastic/ Discrete/ Kinetic	Initiation, elongation with pausing, editing, termination	Relation between RNA polymerase pausing and transcriptional dynamics in terms of interval between successive mRNA production
RNA processing and degradation	(Singh, Yang et al. 2007)	Deterministic/ Continuous/ Kinetic	Transcription, pre-mRNA splicing, pre-mRNA turnover and mRNA degradation	Quantitative model with identification of reaction rate parameters for steps in RNA processing
	(Ciocchetta, Hillston et al. 2008)	Stochastic/ Discrete/ Kinetic	Co-transcriptional cleavage and alternate splicing	Determination of relative frequency of alternate splicing pathways
	(Cao and Parker 2001)	Deterministic/ Continuous/ Kinetic	mRNA synthesis and degradation	Study of effect of degradation reaction rate parameters on mRNA level
	(Carrier and Keasling 1997)	Stochastic/ Discrete/ Kinetic	Degradation by endonuclease and protection by ribosomes	Study of effect of 3 different mechanisms of degradation in relation to ribosome loading and translation rate
	(Khanin and Higham 2007)	Deterministic/ Continuous/ Steady state	miRNA mediated post transcriptional regulation	Study of effect of presence of miRNA on target mRNA level
	(Levine, Ben Jacob et al. 2007)	Deterministic/ Continuous/ Steady state	sRNA mediated post transcriptional regulation	Study of effect of presence of sRNA on target protein level
	(Aguda, Kim et al. 2008)	Deterministic/ Continuous/ Kinetic	miRNA mediated post transcriptional regulation in feedback	Occurrence of oscillations in target protein level due to presence of negative feedback
Translation	(von Heijne, Nilsson et al. 1978)	Deterministic/ Continuous/ Kinetic	Initiation, Elongation and Termination	Study of effect of the ribosome movement and RNA secondary structure on the elongation rate.
	(Bergmann and Lodish 1979)	Deterministic/ Continuous/ Steady state	Initiation, Elongation and Termination	Relationship between polysome size and protein synthesis.
	(Heinrich and Rapoport 1980)	Deterministic/ Continuous/ Kinetic	Initiation, Elongation, Termination	Identification of elongation to be an important regulatory step in translation; effect of ribosome concentration on mRNA
	(Dimelow and Wilkinson 2009)	Deterministic/ Continuous/ Steady state	Initiation	Distribution of the rate parameters in initiation.
	(You, Coghill et al. 2010)	Deterministic and stochastic/ Continuous/ Steady state	Initiation	Effect on initiation factor concentration on protein synthesis rate and variation of protein synthesis rate in population

	(Skjondal-Bar and Morris 2007)	Deterministic/ Continuous/ Kinetic	Initiation and elongation	Relationship between amino acids, tRNA and other factors. Ribosomal density on RNA.
	(Gilchrist and Wagner 2006)	Stochastic/ Discrete and Continuous/ Kinetic	Initiation, Elongation, Termination	Study of effect of nonsense errors at any codon on capability of ribosome recycling
	(Drew 2001)	Stochastic/ Discrete/ Steady state	Initiation, elongation and termination	Study of effect of binding of regulator to DNA on protein synthesis
	(Heyd and Drew 2003)	Stochastic/ Discrete/ Steady state	Elongation	Study of effect of concentration of amino acids and elongation factors on elongation rate
	(Nayak S. , Siddiqui J.K. et al. 2011)	Deterministic/ Continuous/ Kinetic	Initiation	Detailed model of translation initiation; effect of regulation at translation initiation on protein output
	(Zouridis and Hatzimanikatis 2007)	Deterministic/ Continuous/ Steady state	Initiation ,Elongation(in detail),Termination	Study of effect of polysome size and distribution of ribosome along mRNA on protein synthesis rate
	(Bar N.S. 2009)	Deterministic/ Continuous/ Dynamic	Initiation	Study of effect of regulation by eIF2 on translation initiation rate
	(De Silvaa E., Krishnana J. et al. 2010)	Deterministic/ Continuous/ Steady state	Termination	Mathematical model of feedback regulation of translation termination by termination factors; the effect of premature stop codon on termination regulation
Protein degradation	(Holzhütter and Kloetzel 2000)	Deterministic/ Continuous/ Kinetic	Protein degradation	Protein fragment pattern by proteasomal cleavage
	(Peters, Janek et al. 2002)	Deterministic/ Continuous/ Kinetic	Proteasome mediated degradation	Dynamics of protein fragment generation
	(Luciani, Kesmir et al. 2005)	Deterministic/ Continuous/ Kinetic	Proteasome mediated degradation	Study of effect of proteasomal gate size on protein degradation.
	(Lee J., Choi K. et al. 2010)	Deterministic/ Continuous/ Kinetic	Ubiquitination of NF-κB	Study of effect of ubiquitination of NF-kB activation; effect of mutant NF-kB on ubiquitination

Appendix IIIb – Global stochastic sensitivity analysis using Morris method

To verify the observed differential sensitivity of CV and Fano factor to parameters using MPSA, global sensitivity was performed using Morris method (Morris 1991) for 4-reaction model of gene expression. The method is based on calculation of Elementary Effect (EE) for each parameter. The elementary effect was obtained by using a predefined sampling strategy in order to minimize the sample size. The elementary effect is represented as,

$$EE_i = \frac{f(\mathrm{k}_1, \mathrm{k}_2, ..., \mathrm{k}_i + \Delta, \mathrm{k}_{i+1}, ..., \mathrm{k}_n) - f(\mathrm{k})}{\Delta}$$

where, the output *f(k)* is a deterministic function of parameters.

To calculate elementary effect for each parameter, a design matrix was formed using the design strategy as given in (Morris 1991; Jin, Peng et al. 2008). The design matrix for *n*

parameters contained n rows and $(n+1)$ columns. The matrix was designed such that value of only one parameter was changed between any two consecutive rows. Thus one elementary effect can be obtained by comparing two rows of the matrix. To form a design matrix, it was assumed that each k_i scaled in [0.1] interval. To determine the values of each parameter, a p-level grid was formed. Therefore, each parameter can take values **{**0, 1/(p-1), 2/(p-1), …, 1**}**. The design matrix was therefore an n-dimensional p-level grid. Δ is a predetermined multiple of 1/(p-1). It is suggested that the value of p should be even. In this study the value of p was set to 10 and Δ was equal to p/[2(p-1)] as used in previous other studies. In the original design matrix the input parameters were defined to be within an interval of [0, 1]. For the gene expression system where the parameters are out of the specified range, Δ was rescaled as, $\Delta_{scaled} = \Delta \times (\mathrm{UB} - \mathrm{LB})$ where, UB and LB are upper bound and lower bound respectively. Therefore, the range of parameter values would be **{**0, 1/(p-1), 2/(p-1), …, 1-Δ**}***(UB_i- LB_i)+LB_i, for each parameter k_i.

In case of 4-reaction model of gene expression 100 design matrices were used to obtain 100 elementary effects for each parameter. The average elementary effect was used as a measure of sensitivity of steady state CV and Fano factor to the parameters.

The sensitivity results obtained using Morris method, were observed to be in agreement with MPSA. Figure A1 shows the bar graph of average EE for each parameter.

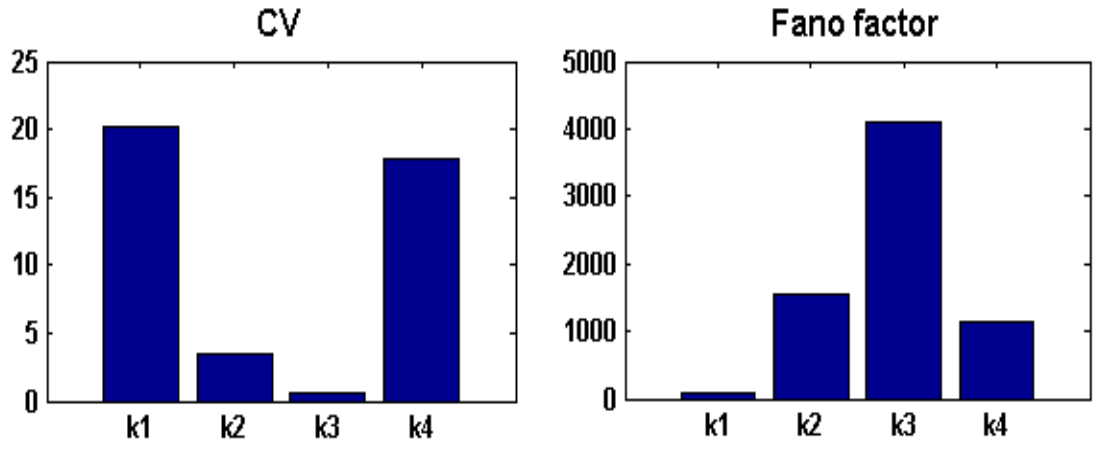


Figure A1: Bar plots showing average EE for %CV and noise strength for transcription (k1), mRNA degradation (k2), translation (k3) and protein degradation (k4)

The average values of elementary effect are summarized in Table A1.

Table A1: Average Elementary Effect values for 4-reaction model

Parameter	EE for %CV	EE for noise strength
Transcription (min^{-1})	20.18	96.06
mRNA degradation (min^{-1})	3.39	1555.28
Translation (min^{-1})	0.58	4110.62
Protein degradation (min^{-1})	17.79	1139.99

From the bar plots and the average EE values, it was observed that CV was most sensitive to transcription and least sensitive to translation. The converse was observed in case of Fano factor. It was observed to be most sensitive to translation. These results are in agreement with the qualitative and numerical sensitivity obtained using MPSA.

Appendix IV – Papers published

A kinetic model of TBP auto-regulation exhibits bistability

Sucheta A Gokhale, Reema Roshan, Vivek Khetan, Beena Pillai*, Chetan J Gadgil*

Abstract

Background: TATA Binding Protein (TBP) is required for transcription initiation by all three eukaryotic RNA polymerases. It participates in transcriptional initiation at the majority of eukaryotic gene promoters, either by direct association to the TATA box upstream of the transcription start site or by indirectly localizing to the promoter through other proteins. TBP exists in solution in a dimeric form but binds to DNA as a monomer. Here, we present the first mathematical model for auto-catalytic TBP expression and use it to study the role of dimerization in maintaining the steady state TBP level.

Results: We show that the autogenous regulation of TBP results in a system that is capable of exhibiting three steady states: an unstable low TBP state, one stable state corresponding to a physiological TBP concentration, and another stable steady state corresponding to unviable cells where no TBP is expressed. Our model predicts that a basal level of TBP is required to establish the transcription of the TBP gene, and hence for cell viability. It also predicts that, for the condition corresponding to a typical mammalian cell, the high-TBP state and cell viability is sensitive to variation in DNA binding strength. We use the model to explore the effect of the dimer in buffering the response to changes in TBP levels, and show that for some physiological conditions the dimer is not important in buffering against perturbations.

Conclusions: Results on the necessity of a minimum basal TBP level support the in vivo observations that TBP is maternally inherited, providing the small amount of TBP is required to establish its ubiquitous expression. The model shows that the system is sensitive to variations in parameters indicating that it is vulnerable to mutations in TBP. A reduction in TBP-DNA binding constant can lead the system to a regime where the unviable state is the only steady state. Contrary to the current hypotheses, we show that under some physiological conditions the dimer is not very important in restoring the system to steady state. This model demonstrates the use of mathematical modeling to investigate system behaviour and generate hypotheses governing the dynamics of such nonlinear biological systems.

Biology Direct, 2010, 5(1): 50.

A systems view of the protein expression process

Sucheta Gokhale, Dimpal Nyayanit, Chetan Gadgil*

Abstract

Many biological processes are regulated by changing the concentration and activity of proteins. The presence of a protein at a given sub-cellular location at a given time with a certain conformation is the result of an apparently sequential process. The rate of protein formation is influenced by chromatin state, and the rates of transcription, translation, and degradation. There is an exquisite control system where each stage of the process is controlled both by seemingly unregulated proteins as well as through feedbacks mediated by RNA and protein products. Here we review the biological facts and mathematical models for each stage of the protein production process. We conclude that advances in experimental techniques leading to a detailed description of the process have not been matched by mathematical models that represent the details of the process and facilitate analysis. Such an exercise is the first step towards development of a framework for a systems biology analysis of the protein production process.

Systems and Synthetic Biology, 2011, 5.3-4: 139:150

Analysis of miRNA regulation suggests an explanation for 'unexpected' increase in target protein levels

Sucheta A. Gokhale and Chetan J. Gadgil*

Abstract

MicroRNA (miRNA) has been mostly associated with decrease in target protein expression levels. Recently, 'unexpected' observations of increase in target protein expression attributed to microRNA regulation have been reported. We formulate a comprehensive model for regulation by miRNA that includes both reversible mRNA–miRNA binding and selective return of RNA. We use this mathematical model incorporating multiple individual steps in the regulation process to study the simultaneous effects of these steps on the target protein level. We show that four dimensionless numbers obtained from 12 rate constants are sufficient to define the relative change in steady state target protein levels. We quantify the range of these numbers for which such pleiotropic increase in protein levels is possible, and interpret the experimental findings in the framework of our model such that the results are no longer unexpected. Finally, we show through stochastic simulation that the nature of the target protein distribution remains unchanged and the relative steady state noise levels are also completely defined by the values of these dimensionless numbers, irrespective of the individual reaction rate constants.

Mol. BioSyst., 2012, 8(3): 760 - 765.

A simple method for incorporating dynamic effects of intronic miRNA mediated regulation

Sucheta Gokhale, Manoj Hariharan, Samir K. Brahmachari* and Chetan Gadgil*

Abstract

The importance of microRNA (miRNA) in modulating gene expression at the posttranscriptional level is well known. Such regulation has been shown to influence the dynamics of several regulatory networks including the cell cycle. In this study we incorporated regulatory effects of intronic miRNA into an existing mathematical model of the cell cycle through the use of an existing 'proxy' protein – the host protein. It was observed that the incorporation of intronic miRNA mediated regulation improved the performance of the model resulting in a closer match to experimental results. To test the universality of this approach we compared the effects of intronic miRNA mediated regulation and host protein mediated regulation. Further, we compared miRNA mediated and protein mediated positive and negative feedback regulations of the target protein. We found that the target protein profiles were predominantly similar. These observations show the applicability of our method for incorporating intronic miRNA mediated dynamic effects in models for regulation of gene expression.

Mol. BioSyst., 2012, 8(8): 2145-2152.