



The Microsoft Research - University of Trento
Centre for Computational
and Systems Biology

Technical Report CoSBI 03/2008

Rule based modeling of gene regulation and biosynthesis of tryptophan in *E. coli*

Valerio Passini

University of Camerino

*Department of Biology MCA & Department of
Mathematics and Informatics*

`valerio.passini@unicam.it`

Cédric Lhoussaine

Université des Sciences et Technologies de Lille

`lhousai@lifl.fr`

Mirabelle Nebut

Université des Sciences et Technologies de Lille

`nebut@lifl.fr`

Céline Kuttler

*The Microsoft Research - University of Trento
Centre for Computational and Systems Biology*

`kuttler@cosbi.eu`

UNIVERSITÀ DEGLI STUDI DI TRENTO

THE MICROSOFT RESEARCH - UNIVERSITY OF TRENTO
CENTRE FOR COMPUTATIONAL AND SYSTEMS BIOLOGY

SECOND LEVEL INTERNATIONAL MASTER
IN COMPUTATIONAL AND SYSTEMS BIOLOGY

MASTER'S THESIS

**Rule based modeling of gene regulation and
biosynthesis of tryptophan in *E. coli***

Student

Valerio Passini

Advisors

Céline Kuttler

Cédric Lhoussaine

Mirabelle Nebut

Defended March 2008

Abstract

The genetic regulation of the Trp operon in the bacterium *E. coli* relies on a sophisticated control mechanism. It tightly couples the advance of transcribing RNA polymerase to the efficiency of the contemporaneous translation of the nascent transcript by a ribosome. The concurrent control of this process involves interdependencies between multiple molecular actors. Within process algebra based modeling languages focused on pairwise interaction, its representation required sophisticated coding tricks. In this work, we abstract the mechanism of transcriptional attenuation within a novel rule base modeling language. It allows non-trivial concurrent control by representing molecules as parametrized terms.

Acknowledgements

I would like to thank the following persons: Céline Kuttler, whose precious help allowed to accomplish this Master's thesis; Cédric Lhoussaine and Mirabelle Nebut at INRIA for their help in science, friendship and a wonderful time in Lille; Joachim Nieheren, for everything. I am also grateful to the Università di Camerino and the project for LITBIO (Laboratory for Interdisciplinary Technologies in Bioinformatics) for economic support. Last but not least, I thank both the administrative as the teaching staff at the Microsoft Research - University of Trento Centre for Computational and Systems Biology for everything during my time in their Master's programme.

Contents

Contents	iv
List of Tables	v
List of Figures	vi
1 Introduction	1
2 Genetic regulation at the Trp operon	3
2.1 Genetic organization of the tryptophan operon	3
2.1.1 Transcription repression	6
2.1.2 Attenuation of transcription	6
2.2 Biosynthesis	11
2.2.1 Metabolic pathway: from chorismate to Trp	13
2.2.2 Feedback inhibition	13
3 Rule based modeling of Trp regulation and biosynthesis	15
3.1 Our modeling language.	15
3.2 Genetic expression and regulation	18
3.2.1 Transcription	18
3.2.2 Translation.	20
3.2.3 Attenuation	22
3.3 Biosynthesis	25
3.3.1 Metabolic pathway	25
3.3.2 Feedback inhibition	26
3.4 Previous models of Trp regulation	26
4 Conclusion	28
Bibliography	30
Appendix	33

List of Tables

2.1	Genes and enzymes	11
2.2	Intermediates and enzymes	13
3.1	Species in our abstractions of DNA and mRNA	17
3.2	Complex species in our model	22

List of Figures

2.1	DNA and mRNA of Trp	4
2.2	Repression	7
2.3	Alternative mRNA loops	8
2.4	Attenuation	9
2.5	Super attenuation and basal level read-through	10
2.6	Metabolic pathway: from chorismate to Trp	12
3.1	Visualization of rules 4 to 7 (transcription)	20
3.2	Visualization of rules 20-23 and 29-32 (mRNA folding and attenuation)	23
3.3	Visualization of rules 43 to 45 (biosynthesis).	25

Chapter 1

Introduction

The genetic regulation and the biosynthetic pathway of tryptophan in the bacterium *E. coli* is one of the best studied systems in molecular biology. Over 50 years of research have established detailed knowledge, both quantitative and qualitative. Although it remains a relatively small system, it is already so complex that no single experiment can investigate all key aspects at the same time.

Models allow to integrate all the available knowledge, produced by various experiments in the wet lab. Thus regulation of tryptophan operon and Trp biosynthesis have attracted wide attention in theoretical and computational biology. The Tryptophan system has been used to explore the flexibility of several modelling tools, and to clarify their applicability and limits, with respect to representing the system on various levels of abstraction.

On the other hand, building a model based on the currently available knowledge helps to structure ideas and to find unexplained issues. As an example, the *in silico* study of Elf et al. (2003) could shed light on generic attenuation and codon preferences usage in control loops. The authors have been able to demonstrate the relationship between charging of isoacceptor tRNA and sensitivity to aminoacid starvation.¹

In this work, we model the Tryptophan system within a novel rule based modeling language, that integrates aspects of concurrent control. Tryptophan is a good test case for the expressivity of this modeling language, allowing comparison with previous work, based on mathematical or computational abstractions within other frameworks. To investigate our formalism in detail exceeds the purpose of this work. We apply it here to tryptophan regulation, but it can be used for other biological cases.

¹ This interesting discovery does not apply to Trp operon. The Trp aminoacid unlike almost all other aminoacids is encoded by only one codon, and thus there is not any possible choice. See section 2.1.2

Our purpose is to show that our formalism simplifies the modeling of complex concurrent control in Tryptophan regulation by its unique features. In future work, we hope to quantify our model, and to obtain interesting results by simulation.

Thesis organization. This work is organized as follows. In chapter 2 we present the biology and regulatory mechanisms at *E. coli*'s tryptophan operon. Its aim is to depict a biological process that requires a good background in molecular biology in a way that eases to understand what are the main aspects to model. It claims neither to be complete from a biological point of view nor to be formally exact, but to let the reader understand the process and the following core chapter. Chapter 3 is the core part of this thesis. It introduces our modeling language, and discusses the most relevant parts of our model of the Trp operon. We discuss the previous models from other authors in the last section of this chapter. Chapter 4 concludes. The full model, comprising 60 reaction rules, is available in the appendix.

Chapter 2

Genetic regulation at the Trp operon

Tryptophan (Trp) is one of the 20 fundamental amino acids used in the translation of the proteins. While in many higher organisms, as Man, the only way to obtain Trp amino acid is through the diet that comprises it among the composition of digested proteins (Lodish et al., 2007), other simple organisms, as bacteria, are capable of synthesizing it if the environment cannot supply it in the necessary amount.

The enzymes needed in tryptophan biosynthesis are encoded in the tryptophan operon in *Escherichia coli*. The biosynthetic pathway for tryptophan in *E. coli*, is controlled at three different levels:

1. Repression of the transcription of the genes
2. Attenuation of the transcription of mRNA
3. Feedback inhibition mechanism acting on TrpE (Anthranilate Synthase)

These regulative mechanisms apply respectively to the gene, to the mRNA and to the protein.

2.1 Genetic organization of the tryptophan operon

By definition, an operon is an organized entity in the bacterial genome to promote the transcription of several genes in once. This is a strategy for having all the genes needed in a peculiar biochemical pathway expressed

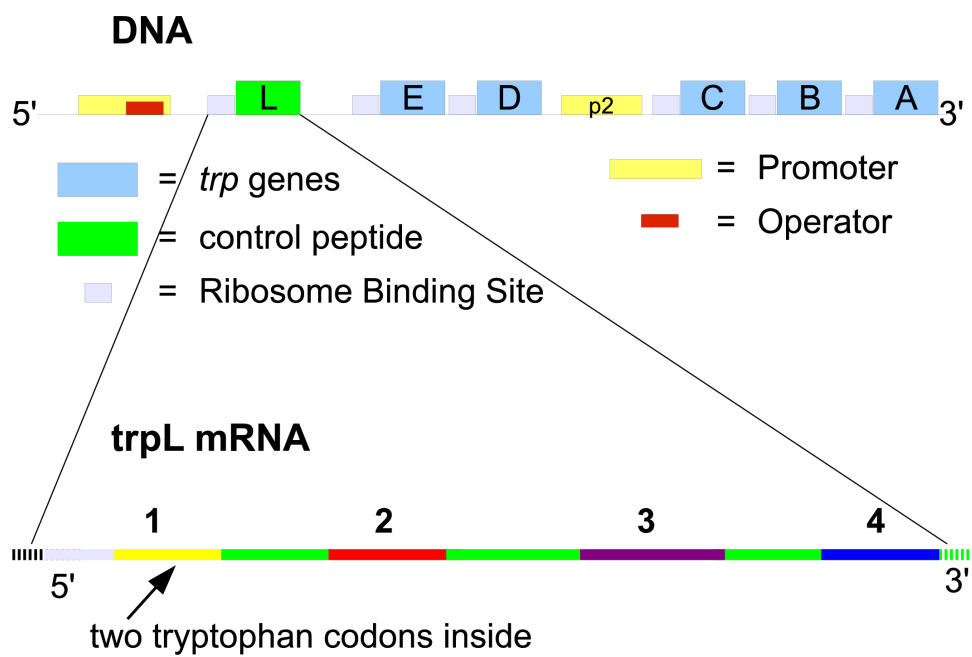


Figure 2.1: Organization of Trp operon on DNA, and transcript of the leader segment (mRNA)

when it is needed as a functional unity, even if they are devised in many single enzymes. The tryptophan operon in *Escherichia coli* is organized in this way and, as an historical remark, it was the first repressible operon discovered by Jacob and Monod in 1953. The tryptophan operon in *E. coli* bears five genes encoding the enzymes that perform the seven biochemical reactions needed for the synthesis of tryptophan starting from chorismate. Other organisms, like *B. subtilis*, may have the seven catalytic domains split in more enzymes, but the number of reactions needed doesn't change (Yanofsky, 2004). The five genes in the tryptophan operon are listed starting from the 5' terminus to the 3' terminus, they are also expressed in a way that follows the steps in the various reactions: *trpE*, *trpD*, *trpC*, *trpB* and *trpA*.¹

Tryptophan operon (Fig.2.1) has a promoter region for binding the RNA polymerase, within this region there is another binding site, called operator, to which a repressor, named TrpR, can be bound repressing the transcription (Yanofsky and Horn, 1994; Yanofsky, 2003; Khodursky et al., 2000). This mechanism relies on steric hindrance, that is a prohibition to bind to RNA polymerase because the binding site is already occupied. The transcription of the polycistronic mRNA starts at the beginning of *trpL* that encodes a peptide useful in attenuation of transcription: all the attenuation mechanism resides in *trpL* and its small peptide called TrpL. This peptide is not an enzyme and has not any other function than controlling transcription "sensing" the presence of tryptophan. The other genes are in the order told previously. Between *trpD* and *trpC* there is a second promoter, called *p2*, for binding RNA polymerase (Horowitz and Platt, 1982). This second promoter helps *E. coli* to express at a basal level some of the enzymes required by the tryptophan synthetic pathway, to have a quick activation of this pathway in case of lack of this amino acid. All the genes have a Ribosome Binding Site (RBS), also called Shine Dalgarno sequence (Lewin, 2004), this means that the translation of the genes can be parallel and independent from the others, when the mRNA has been completely synthesized. The transcription termination happens in a ρ dependent way (Yanofsky et al., 1981; Nudler and Gottesman, 2002) in case the transcription proceeds to *trpA* (Yanofsky, 2001).

¹Typographical conventions. In biology there are some rules in writing genes and proteins in texts so that it is easy to distinguish them, here we apply them as follows: a gene is written in lowercase and in italic like *trpE* or *hsp70*. The product of the gene that is the protein is written with the first letter in capital, like TrpE or Hsp70. The names of species are always written in italic: *Escherichia coli*, *Saccharomyces cerevisiae*, *Homo sapiens*. For shortcut they could be written with only the first letter of the genus but the full name of species: *E. coli*, *S. cerevisiae*, *H. sapiens*.

2.1.1 Transcription repression

The first mechanism by which the transcription is regulated is the repression (Yanofsky and Horn, 1994; Yanofsky, 2003). An operator (Fig. 2.2, part A), called *trpO*, overlaps the promoter, called *trpP*. When TrpR, a protein that is able to bind to the operator, is bound to DNA, the RNA polymerase can not access the promoter so the transcription of the gene is impossible. For this role in impeding the transcription, TrpR is a so called repressor. TrpR is composed of identical half-molecules and senses the presence of tryptophan, because each has a tryptophan binding site. When it is not complexed with tryptophan, the dimeric form, termed the TrpR aporepressor, cannot bind tightly to operator *trpO*. When two tryptophan molecules bind to their respective binding sites, the TrpR aporepressor is converted to the functional repressor (Fig.2.2B). In absence or with very scarce tryptophan, the repressor is converted to the aporepressor and can unbind the operator site leaving the promoter region, see Fig. 2.2C (Yanofsky, 2004). RNA polymerase can now initiate the transcription of the DNA sequence.

2.1.2 Attenuation of transcription

A decisive aspect to understand the attenuation mechanism is the tight *coupling of transcription and translation* in bacteria. Other than in eukaryotes, no intermediary mRNA processing occurs. As soon as the length of a nascent transcript allows, a ribosome may bind to its specific RBS site, and start the synthesis of a new protein (Oppenheim and Yanofsky, 1980). The contemporaneity of transcription and translation is used in attenuation.

The tryptophan polycistronic mRNA starts with the *leader sequence trpL*, containing four self complementary regions, and two codons for tryptophan (UGG). The tryptophan codons are crucial for the advance of translating ribosome. The complementary regions are able to pair, which affects transcribing RNA polymerase. The pairing rules illustrated by Figure 2.3 are:

- region 1 can only pair with 2,
- region 2 can pair with 3 *or* 1,
- region 3 can pair with 2 *or* 4, and
- region 4 can only pair with 3.

When the mRNA regions pair, they form 3-dimensional structures called stem loops, similar in shape to an hairpin. As soon as the RNA polymerase has synthesized the region 2, a first stem loop forms between regions 1 and

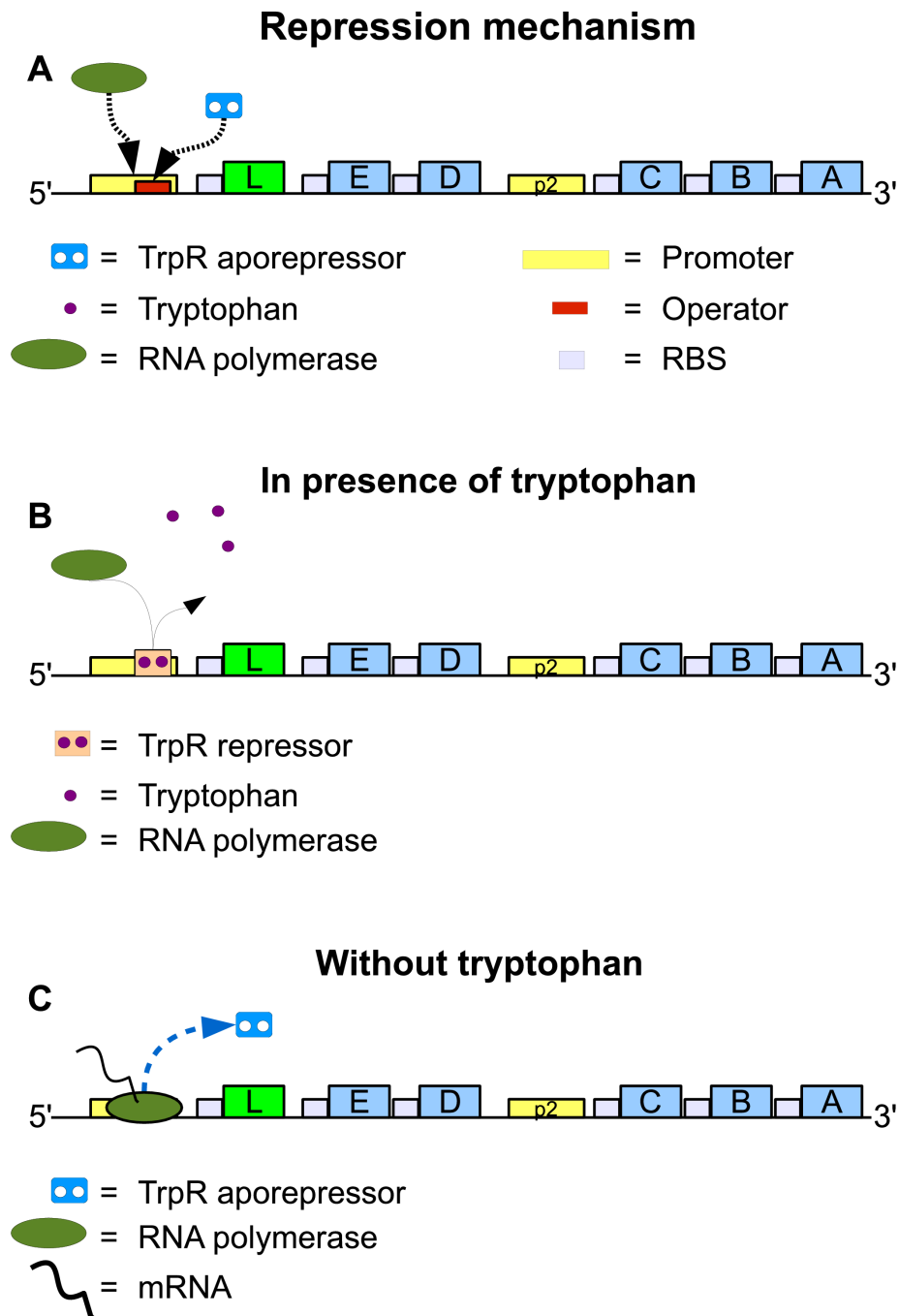


Figure 2.2: Repression of transcription initiation at the promoter: the repressor is only active in presence of Trp.

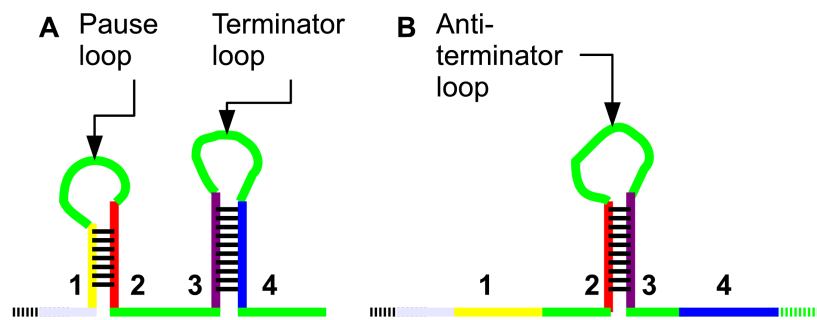


Figure 2.3: Alternative stem loops forming in the leader sequence of Trp mRNA.

2. It is called the *pause loop*, and obliges the RNA polymerase to wait until translation has started. The ribosome, moving along the mRNA unwinds the pause loop. RNA polymerase resumes transcription, and is now *synchronized* with the ribosome. Figure 2.4B illustrates. While RNA polymerase continues transcription, the ribosome encounters the two tryptophan codons encoded in region 1 of *trpL*. What happens at this point determines if the transcription continues until the end of the operon or not.

If the intracellular level is high, translation can proceed. While the ribosome is moving over region 2, RNA polymerase synthesizes the region 3 and 4. Region 3 couples with region 4 and forms a so called *terminator loop*. It leads to transcription termination, see Fig. 2.4C.

If the tryptophan *level is low*, the ribosome stalls over region 1. While the ribosome is stalling, the RNA polymerase continues polymerizing the region 3 of *trpL*. Region 2 is free to couple region 3, forming the so called *anti-terminator loop*. This prohibits the formation of the terminator loop. RNA polymerase continues to transcribe the whole operon, as illustrated by Fig. 2.4D.

Less usual behaviors: super-attenuation and basal level read-through The situation described above is what happens most frequently, but alternative behaviors are possible.

Super-attenuation (Fig. 2.5) happens when no ribosome binds to the RBS site in the newly synthesizing mRNA after the formation of the pause loop. The RNA polymerase waits a little, eventually continues transcription, and as soon as regions 3 and 4 of *trpL* are synthesized, the terminator loop forms and the transcription halts (Yanofsky and Horn, 1994; Yanofsky, 2000; Elf and Ehrenberg, 2005).

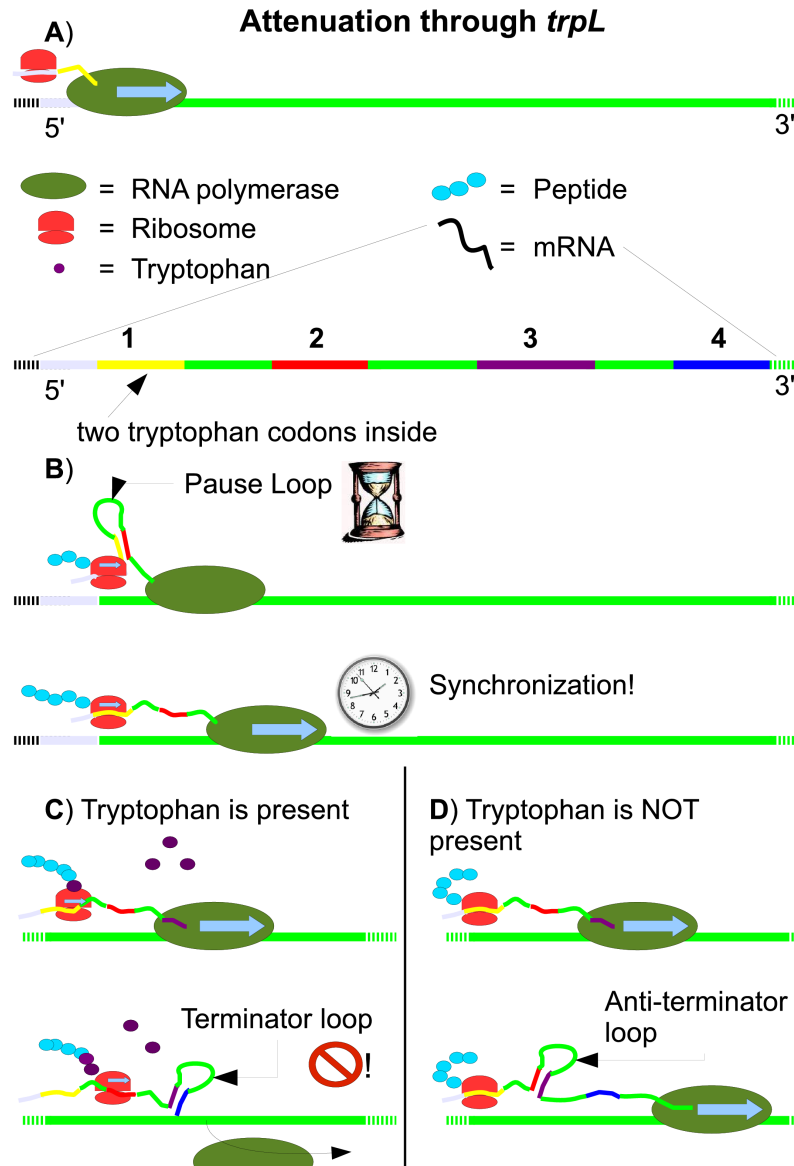


Figure 2.4: Attenuation

Super-attenuation and Basal Level Readthrough

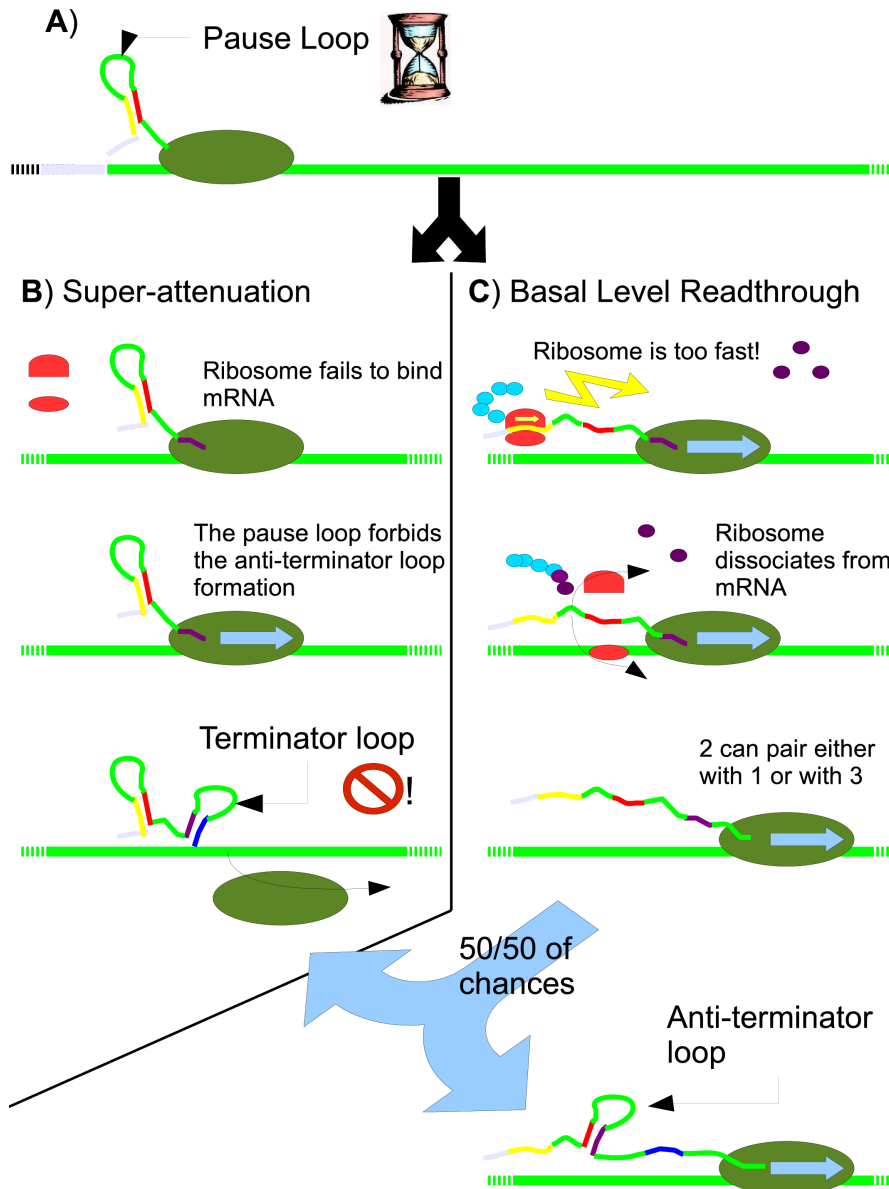


Figure 2.5: Super attenuation and basal level read-through

Gene name	Enzyme name	Active domains	Complexed with
<i>trpE</i>	TrpE	TrpE	TrpD
<i>trpD</i>	TrpD	TrpD/TrpG	TrpE
<i>trpC</i>	TrpC	TrpC/TrpF	
<i>trpB</i>	TrpB	TrpB	TrpA
<i>trpA</i>	TrpA	TrpA	TrpB

Table 2.1: Genes and enzymes

In *basal level read-through*, if a ribosome binds to the *trpL* RBS, but translates too fast TrpL, then two things are likely going to happen. Because the affinity of the segment 2 in *trpL* is the same for the segments 3 and 1, if 2 and 1 pair, then 3 and 4 are free to form the terminator loop, ending the transcription. If 2 and 3 regions pair, then the anti-terminator loop forms and the transcription keeps going. This effect is responsible for the basal expression (Fig. 2.5) that gives to the cell a minimal amount of the enzymes in all conditions. In fact, this phenomenon happens only on a small number of mRNA transcribed in conditions that should halt transcription: Yanofsky (2000) reports that only about 5-10% of mRNA in ongoing transcription binds to a ribosome that is too fast and leaves the messenger too early. Only the half of this amount of mRNA is completed.

2.2 Biosynthesis

Actually the biosynthesis of tryptophan involves a group of several other amino acids, enzymes and intermediates, that are partially common to other biosynthesis. Because of that, it is possible to distinguish the synthesis of tryptophan from other biochemical pathways starting from a certain catalytic step: the conversion of the chorismate into anthranilate (see Fig.2.6). Chorismate also participates in other biosynthetic pathways. The conversion in anthranilate ensures that the biosynthesis of Trp has just begun (Khodursky et al., 2000). Depending on the organism, the number of enzymes contributing to Trp biosynthesis might vary, merge in functional units or work inside functional complexes,² but the number and kind of reactions are the same, and involve the same intermediates and collateral metabolites.

Table 2.1 summarizes the short names of the genes, the names as enzymes, the active domains, and the last column tells if the enzyme is part of a complex, and with which other subunit.

²*B. subtilis* has 7 enzymes encoded by 7 genes (Yanofsky, 2004)).

Tryptophan Biosynthetic Pathway

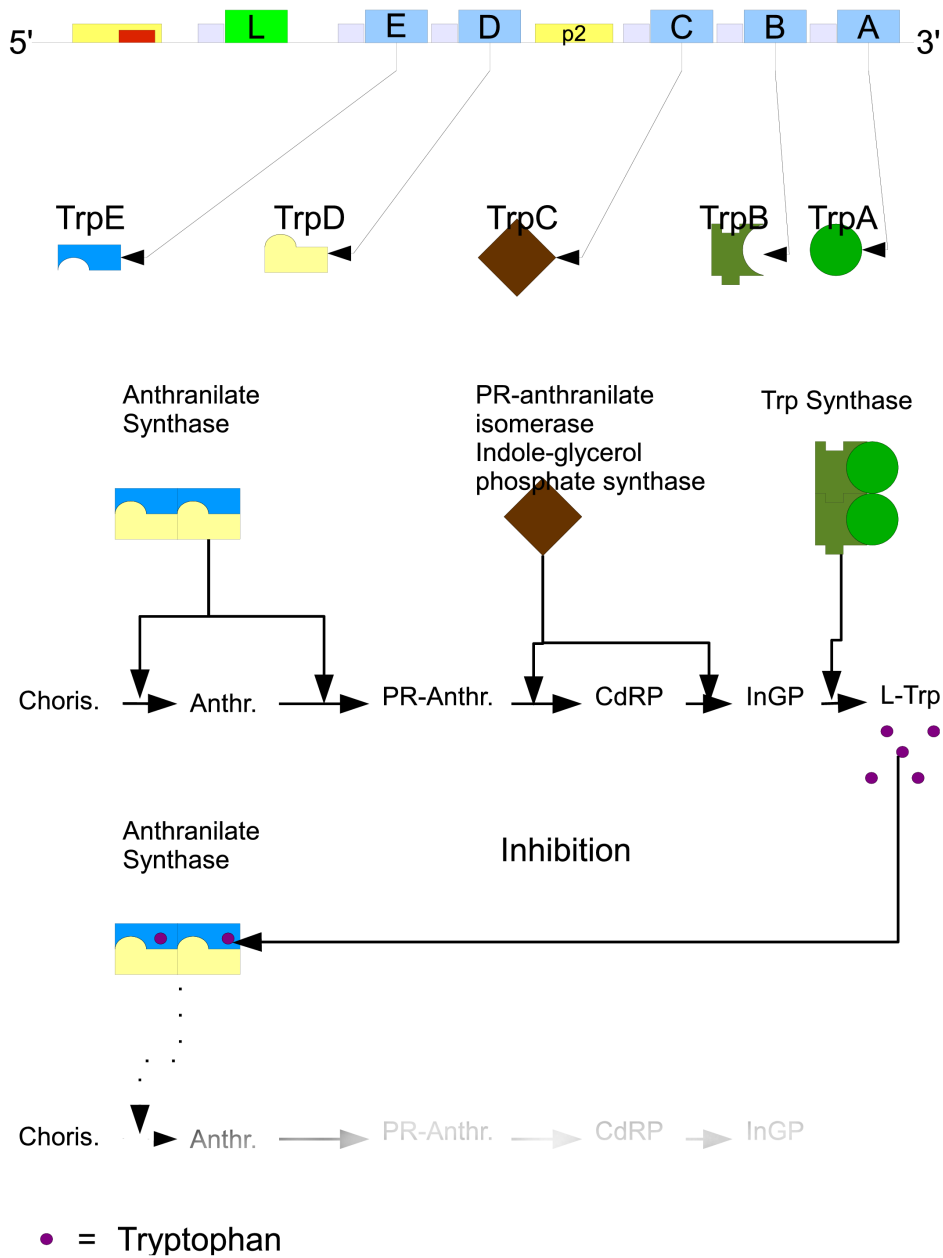


Figure 2.6: Metabolic pathway: from chorismate to Trp

Intermediate	Short name	Enzyme
Anthranilate	Anthr.	Anthranilate synthase
N-5'-phosphoribosyl-anthranilate	PR-Anthr.	Anthranilate synthase
1-(o-carboxyphenylamino-1-deoxyribulose 5-phosphate)	CdRP	PR-anthranilate isomerase-indole glycerol phosphate synthase
indole3-glycerol phosphate	InGP	PR-anthranilate isomerase-indole glycerol phosphate synthase
L-Tryptophan	L-Trp	Tryptophan synthase

Table 2.2: Intermediates and enzymes

2.2.1 Metabolic pathway: from chorismate to Trp

The biosynthesis starts with anthranilate synthase (see Fig.2.6), an heterotrimer that is an enzyme complex³, composed by two TrpE and two TrpD subunits. The biosynthesis continues with TrpC, also called PR-anthranilate isomerase-indole glycerol phosphate synthase, that has two active sites. The pathway ends with another heterotetramer constituted by two TrpB and two TrpA subunits, called tryptophan synthase (Yanofsky, 2003; Xie et al., 2003; Oppenheim and Yanofsky, 1980). The intermediates in the reactions and all the additional substrates are less interesting. They are not used in our model and are not useful to the purpose of understanding this pathway and its regulation. Table 2.2, lists them all, and the abbreviations used in Fig. 2.6.

2.2.2 Feedback inhibition

The pathway of tryptophan biosynthesis is regulated by feedback inhibition (Yanofsky and Horn, 1994; Yanofsky et al., 1971; Xie et al., 2003). Tryptophan is the effector molecule for the enzymatic complex called anthranilate synthase, which is the first enzyme to catalyze a reaction in the catalytic pathway, synthesizing anthranilate from chorismate. This feedback inhibition is achieved when every TrpE subunit in anthranilate synthase is individually bound by a tryptophan molecule (see Fig.2.6). When this event happens, anthranilate synthase conformation changes and the enzyme stops

³It is not possible to explain here how complexes form, but is worth to underline that the subunits alone can have only slight enzymatic activity or no activity at all.

its catalytic activity. The feed back inhibition is correlated to Trp concentration in the cell. Thus, an excess of intracellular tryptophan inactivates most of the anthranilate synthase proteins and avoids the production of more tryptophan.

Chapter 3

Rule based modeling of Trp regulation and biosynthesis

Rule-based Modeling (RBM) is a modeling style that resembles chemical reactions, as biologists and chemists are familiar with. It is easy to understand, yet more powerful to represent subtle aspects of complex phenomena as transcription/translation, biosynthesis.

We introduce our modeling language in Sec. 3.1, and sketch key differences between it and related frameworks. The core part of this thesis follows in Sec. 3.2: the application of our language to genetic expression and regulation at the Trp operon.

3.1 Our modeling language.

Our rule based modeling language is based on *species* and *reactions*. We were inspired by BioCham (Chabrier-Rivier et al., 2005), that however suffers from limitations with respect to concurrent control. Our language's additional features that overcome these limitations are:

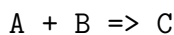
- while in BioCham, molecular species are *atomic*, ours may be parametrized *terms*.
- a **new** operator, inspired by the π -calculus, allows to create fresh *names*. We use these to parametrize instances of molecular species. With this, we can *identify* individual molecules belonging to the same species, or *group* molecules belonging to the same macromolecular complex.

To the best of our knowledge, these features are not supported by other rule based modeling frameworks as BioNetGen (Faeder et al., 2005).

When modeling transcriptional attenuation we need to express complex side conditions. Based on rules with several ($n > 2$) reactants, we can easily cover the core parts of our model. The κ -calculus (Danos and Laneve, 2004) that is closely related to BioNetGen only allows *binary* reaction rules, as does the π -calculus (Milner, 1999). In process algebra based modeling languages, reactions involving multiple reactants would have required sophisticated encoding techniques (Ciocchetta and Priami, 2006; Cleaveland et al., 2001). Our rule base modeling language provides an understandable alternative, that is accessible to non-experts modelers.

Language elements and conventions. Modeling any biological system requires a certain level of *abstractions* and *key assumptions* to be made. We introduce some in this section, they serve as illustrations of our rule based modeling language and its notational conventions. For sake of clarity, our detailed assumptions will follow in the technical discussion of the model in the other sections of this chapter.

Our *rules* are similar to reactions, as generally used by biologists or chemists. The left side of a rule lists one or more reactants. An arrow indicates that these reactants can be transformed into one or more products, listed on the right side of the rule. A simplistic example is:



In a model ready for simulation, the kinetics of every reaction would complete the rules.

Our *notational conventions* are similar to those of BioCham. *Species names* are alphanumeric strings, with underscore `_` to structure species names. For instance, `RNA_L1` represents the mRNA obtained by the transcription of `Leader1` DNA sequence. White spaces are disallowed, because the parser would read it like two or more compounds instead of one.

Examples from our model. The names used in modeling are supposed to be understandable at a first glance, i.e. to clearly recall the names used in the biology chapter. Table 3.1 introduces central species names of our model. Their choice reflects the organization of Trp Operon. It has its own logic that underlies our *modeling assumptions*, and our *choice of abstraction levels*, the reason why will be clearer in the next sections.

We distinguish elements of DNA and mRNA. The first DNA segment is the **Promoter**. There is no equivalent in the abstraction of the mRNA strand, because it is the place where the RNA polymerase binds, and transcription only starts *after* it. **RBS** and **RNA_RBS** are the *Ribosome Binding Site* on DNA

segment	Promoter	RBS	1	2	3
DNA	Promoter	RBS	Leader1	Leader2	Leader3
mRNA		RNA_RBS(n)	RNA_L1(n)	RNA_L2(n)	RNA_L3(n)

Table 3.1: Species in our abstractions of DNA and mRNA

and RNA respectively. The DNA species `Leader1`, `Leader2`, etc represent the various portions of the tryptophan leader sequence *trpL*. They are transcribed into *one* long strand of mRNA, that is composed from different components.

For species related to DNA we use simple names (without underscores), while the abstractions of mRNA places a prefix `RNA_` in front of these names, that are possibly shortened. Thus the products of transcription are named `RNA_L1`, `RNA_L2`, etc.

New operator, parametric species, and instances. Parameters allow to distinguish different instances of species, identified by a *name*. Here a minimal example in which an instance of `C` is identified by a unique name `n`:

$$A + B \Rightarrow \text{new } n.C(n)$$

Table 3.1 gave examples related to mRNA, that we now discuss. At the beginning of each session of transcription we create a unique *identifier* with the **new** operator. The identifier is passed as parameter to mRNA components created in the subsequent round of transcription. The usefulness to modeling is great. Without it we could not distinguish between the mRNA produced in two rounds of transcription. But now, `RNA_L1(a)` is clearly distinct from `RNA_L1(b)`. Only components of the *same* mRNA molecule representation may interact in loop formation:

$$\text{RNA_L1}(n) + \text{RNA_L2}(n) \Rightarrow \dots$$

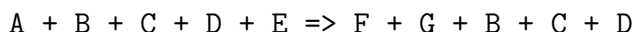
It is important to emphasize that the reaction can only be executed if the two parameters `n` on the left hand side of the rule *match*.

Complex formation. By convention we use a dash `-` for complex formation between two or more reactants, see:

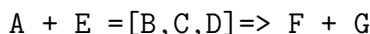
$$A + B \Rightarrow A-B$$

By this convention we avoid to introduce new species names for complexes. In our model `Promoter-RNAP` denotes the complex of promoter and RNA polymerase. To make the model simpler to write and understand, we will allow a few well-justified *exceptions*. For example, in the formation of the pause loop by `RNA_L1(n)` and `RNA_L2(n)` we rewrite the complex as `PauseLoop1_2(n)`. Table 3.2 on page 22 lists all these cases. Another detail is that a complex may bear a *parameter*, that none of its constituents had.

Convention: reactions with side conditions. Another useful convention is to place within square brackets [] compounds appearing on both sides of a reaction, when they are not consumed. In this manner we simplify the lengthy reaction

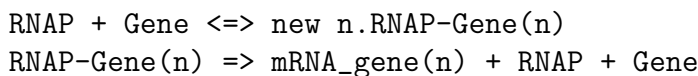


into the following:



This is useful when we require conditions to trigger events, called *side conditions*. Examples are in the model of mRNA loop formation (see section 3.2.3). Note that for us, this is just a notational shortcut. BioCham makes use of a similar convention, but it is limited to only one compound, motivated by enzymes that catalyze a reaction, and enforces the application of Michaelis Menten kinetics.

Rule application. The following example summarizes most of the conventions, and a simple use of parameters. Notice how a reversible complex `RNAP-Gene` is created, and how it carries the the newly created name `n` as *optional* parameter:



While this minimal rule set represents a basic model for gene transcription, it remains too poor to deal with the attenuation mechanism.

3.2 Genetic expression and regulation

3.2.1 Transcription

Assumptions and abstractions The two rules of our previous example were a high-level abstraction of transcription. In reality, RNA polymerase is

constituted by many subunits, needs σ factors to bind the promoter, moves from one nucleotide to the next on the DNA strand, etc. To simplify our model, the σ factor is not considered. Its inclusion would make sense for other cases of transcription, as modeling shock responses.

In our model, the RNA polymerase is a whole atomic unit, and instead of a physically realistic base pair stepping, it *jumps* from one large segment of nucleotides to the next. The segments as introduced in the previous section are chosen to represent the *essential properties we are interested in*, i.e. the key regulatory events reported experimentally: regulation of the expression, mRNA loop formation, and transcription of the genes.

Transcription model. Table 3.1 lists our species names for the representation of DNA and mRNA of the Trp operon. In Figure 3.1 our model for transcription is represented as a graph. The RNA polymerase representation RNAP must bind the promoter, and form a complex Promoter-RNAP before it may start transcription.

This step is reversible: i.e. the RNAP can release the promoter without starting transcription. Both reactions are modeled in rule 4 (see Appendix 4). If the transcription proceeds, then the RNAP binds to the first RBS of the tryptophan operon, forming the RBS-RNAP complex, and leaves the promoter (rule 5). From now on RNAP cannot move backward and will move along the other regions. Then RNAP moves, or better “jumps”, from RBS to Leader1 and produce a new complex: RNA-RBS. This is the first complex species appearing in transcription. The next two rules (6, 7) follow the same rationale, this time the new species produced is RNA_L1 as RNAP advances on to the DNA sequence Leader2.

Reactions. The species used are those collected in Table 3.1 as RNA_RBS(n), generated using parameters. Transcription begins when RNA polymerase binds to the promoter in the DNA (Fig. 2.2C), as explained in section 2.4. This happens in reaction rule 4.

```
rule 4: Promoter + RNAP <=> Promoter-RNAP
rule 5: Promoter-RNAP + RBS => new rna.RBS-RNAP(rna) + Promoter
rule 6: RBS-RNAP(rna) + Leader1
      => RBS + Leader1-RNAP(rna) + RNA_RBS(rna)
rule 7: Leader1-RNAP(rna) + Leader2
      => Leader1 + Leader2-RNAP(rna) + RNA_L1(rna)
```

In reaction 5, the RNA polymerase moves from the promoter to the RBS sequence on DNA. An instance of the complex RBS-RNAP is created. It takes

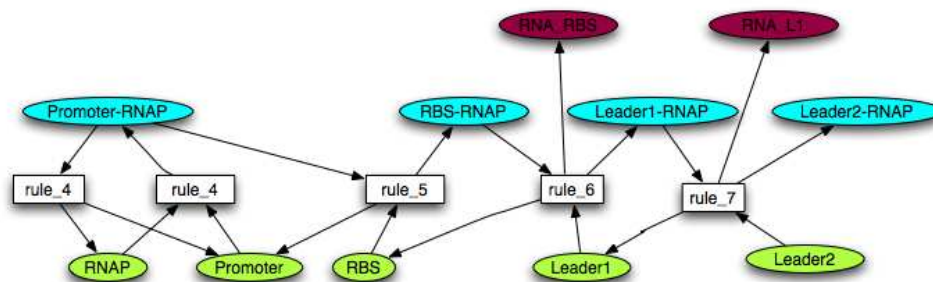


Figure 3.1: Visualization of rules 4 to 7, that model transcription. Color codes: green for molecules in initial configuration, blue for intermediate complexes, red for molecules dynamically created during model execution. Figure automatically generated from our rule set with BioCham (Chabrier-Rivier et al., 2004).

the newly created name `rna` as optional parameter. This name will be passed as parameter to all segments of the transcript, showing they are part of the same macro-molecule's abstraction. Segments of mRNA created in other rounds of transcription would be grouped together by distinct names (see 3.2.3). In reaction 6, the first piece of mRNA is produced and it is the `RNA_RBS(rna)`, where a ribosome can bind to start translation. At the same time the DNA sequence `RBS` is released while `Leader1` sequence is read by RNA polymerase, forming the complex `RNAP-Leader1(rna)`. In reaction 7 the above scheme is repeated: the final product is the first segment of the Leader sequence, `RNA_L1(rna)`, and the RNA polymerase keeps moving along the sequence.

3.2.2 Translation.

Assumptions and abstractions. In section 2.1.2, we already made many simplifications regarding translation. The reason is to focus on *crucial points* of the process, and to discard aspects that do not matter for our purposes. The need for details in our modeling of translation, are related to attenuation. We thus make the following simplifying assumptions:

1. while a ribosome translates over the leader segment `trpL`, it advances by transitions from one segment to the next, instead of base pair stepping, or stepping from codon to codon;

2. we don't distinguish between Trp and tRNA charged with Trp ¹;
3. we do not explicitly represent the growing peptide of amino acids, assembled while a ribosome translates the leader mRNA. We do however check for the presence of Trp while translating over the corresponding codons;
4. we don't explicitly render the internal RBS of the polycistronic mRNA, but only the one within the leader segment.
5. translation of the protein coding genes is simplified into an atomic step in rules 37-41, abstracting away from the required amino acids and transfer RNA, the availability of which is assumed *not* to be rate-limiting.
6. we do not represent a stop codon while translating *trpL*, we consider the presence of the ribosome on Leader2 as a side condition to stop translation of TrpL, this is not physically realistic but functional;
7. translation only starts after pause loop formation in rule 29; this is important for synchronization between RNAP and ribosome.

Translation model. In our model of translation ribosome binds the first segment of the mRNA, represented by `RNA_RBS`, when `PauseLoop1_2` has already been created from `RNA_L1` and `RNA_L2`. This action creates the complex `Ribosome-RNA_RBS`. The ribosome then moves on from `RNA_RBS` to `RNA_L1`, breaking the `PauseLoop1_2`. This has the consequence of freeing RNA polymerase (see section 3.2.3, regarding attenuation).

When ribosome is over `RNA_L1` it waits for two `Trp` molecules, this is the attenuation core, the effects of which are refined further in the appropriate section. Translation of the mRNA segments is modeled very simply, as:

```
rule 37: Ribosome + RNA_trpE(n) => TrpE + Ribosome + RNA_trpE(n)
```

Each encoded enzyme is translated independently, reflecting the existence of independent RBS on the polycistronic mRNA. It is important to notice that we do *not* parametrize newly created instances on enzyme species, as `TrpE`. Thus, any enzyme subunit remain able to form complexes with any products of other mRNAs. The details in formation of enzymatic complexes will follow in Section 3.3.1.

¹Actually ribosomes use only charged tRNA, but modeling tRNA charging is outside our purposes, unlike in other work (Elf and Ehrenberg, 2005)

Complex species	rule	impact
PauseLoop1_2(n)	20	Forces RNAP to pause and wait for the ribosome to unfold the pause loop.
TerminatorLoop3_4(n)	21	Forces the RNAP to abort transcription.
AntiTerminatorLoop2_3(n)	22	Allows RNAP to continue transcription into the enzyme coding regions.

Table 3.2: Complex species in our attenuation model, the rules in which they are instantiated, and their regulatory impact on RNA polymerase.

3.2.3 Attenuation

Attenuation is the most difficult process to model: it requires that translation and transcription keep a certain pace to continue transcription, as introduced in Section 2.4. The dynamics in the formation of these loops, and the interplay between the molecular actors in attenuation is highly non-trivial. Figure 3.2 gives a first impression of *parts* of our attenuation model and its complexity. Anyhow, the loop formation is the core concept among the rules discussed in the following sections. Here we have preferred to allow new species for complexes (see Table 3.2), instead of explicitly representing complexes in terms of their constituents (e.g. RNA_L1-RNA_L2). This choice keeps our model clean and easy to read. It tells that the segments are no longer available, because they are caught in loops.

Pause loop formation and unwinding. Rule 20 abstracts the pause loop formation (see Figure 2.3).

```
rule 20: RNA_L1(n) + RNA_L2(n) => PauseLoop1_2(n)
```

Our following model refines the general scheme of translation. We assume that only after the pause loop has formed, the ribosome can bind to RBS and unwind the loop, and continue translation towards segment 1.

```
rule29: Ribosome + RNA_RBS(n)
      = [PauseLoop1_2(n), Leader3-RNAP(n)] =>
      Ribosome-RNA_RBS(n)
```

Note that RNAP must be bound to the third DNA leader segment. The next rule breaks the pause loop, and represents how the ribosome advances to leader segment 1.


```
rule30: Ribosome-RNA_RBS(n) + PauseLoop1_2(n)
      =[Leader3-RNAP(n)]=>
      RNA_RBS(n) + Ribosome-RNA_L1(n) + RNA_L2(n)
```

From now on, Ribosome and RNAP are synchronized, and continue translation and transcription respectively in a race.

Ribosomal stalling. Rules 31 and 32 refine the general scheme of translation for the transition from mRNA segment one to two.

```
rule 31: Ribosome-RNA_L1(n) + Trp => Trp-Ribosome-RNA_L1(n)
```

```
rule 32: Trp-Ribosome-RNA_L1(n) + Trp + RNA_L2(n)
=> Ribosome-RNA_L2(n) + RNA_L1(n)
```

If there is enough Trp, the Ribosome moves until it is over the Leader2 portion of mRNA, forming the intermediary complex `RibosomeRNA_L2(rna)`. Here, we remember that according to our assumptions `RNA_L2` is the place where ribosome stops translation.

Terminator loop formation. If there was enough Trp, and the Ribosome continued to segment 2, the terminator loop can form.

In reaction 21 this complex is used as a condition to form the terminator loop using `RNA_L3(rna)` and `RNA_L4(rna)`.

```
rule21: RNA_L3(n) + RNA_L4(n)
      =[Ribosome-RNA_L2(n)]=> TerminatorLoop3_4(n)
```

Anti-terminator loop formation. On the other hand, if there is not enough Trp, Ribosome stalls over the Leader1 portion of mRNA. The two possible intermediate complexes produced by reactions 30 and 31 are `Ribosome-RNA_L1(rna)` and `RibosomeTrpRNA_L1(rna)`. These are used as conditions for the anti-terminator loop formation in the rules 22 and 23 respectively.

```
rule 22: RNA_L2(n) + RNA_L3(n)
      =[Ribosome-RNA_L1(n), Leader4-RNAP(n)]=>
      AntiTerminatorLoop2_3(n)
```

```
rule 23: RNA_L2(n) + RNA_L3(n)
      =[Trp-Ribosome-RNA_L1(n), Leader4-RNAP(n)]=>
      AntiTerminatorLoop2_3(n)
```

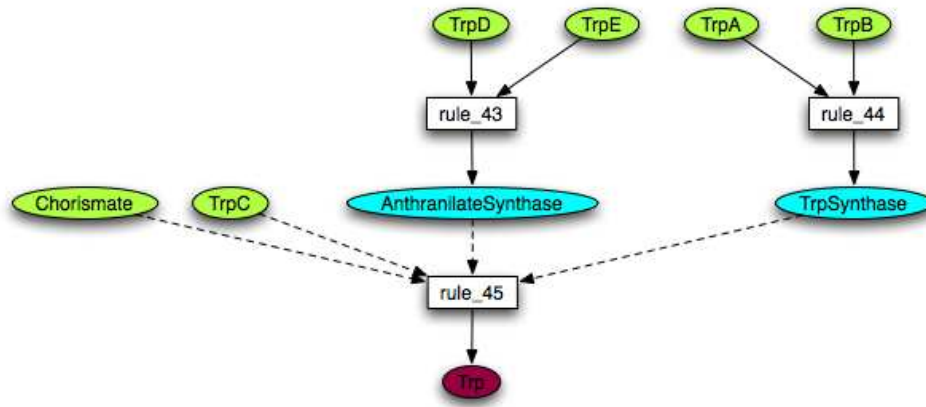


Figure 3.3: Visualization of rules 43 to 45 (biosynthesis).

Anti-terminator loop is formed by coupling the mRNA Leader2 (RNA_L2) sequence with mRNA Leader3 (RNA_L3), both created in the same round of transcription.

3.3 Biosynthesis

In this section we present our model of the metabolic pathway of Trp biosynthesis, and its feedback inhibition.

3.3.1 Metabolic pathway

We assume that modeling the biosynthesis of tryptophan does not require to specify every single intermediate produced in the various steps (see Fig.2.6). Given that there is no interest in a collateral pathway that uses one of the intermediates (i.e: anthranilate) as a substrate because its effects are negligible, thus for sake of simplicity we assume here that to synthesize tryptophan we only need chorismate and all the complexes/enzymes in place.

For this reason, we payed more attention to model the formation of the complexes, that are required for the various steps, than to the synthesis itself:

rule 43: $\text{TrpE} + \text{TrpE} + \text{TrpD} + \text{TrpD} \Rightarrow \text{AnthranilateSynthase}$

rule 44: $\text{TrpB} + \text{TrpB} + \text{TrpA} + \text{TrpA} \Rightarrow \text{TrpSynthase}$

rule 45:

$\text{Chorismate} + \text{AnthranilateSynthase} + \text{TrpC} + \text{TrpSynthase} \Rightarrow$
 $\text{AnthranilateSynthase} + \text{TrpC} + \text{TrpSynthase} + \text{Trp} + \text{Chorismate}$

In reactions 43 and 44 it is possible to see the two complexes formation: both are heterotetramers and requires two pairs of subunits. Once the complexes are ready the biosynthesis of tryptophan can start. In reaction 45, chorismate is never consumed because it appears on the right side of the reaction too. Of course this is not realistic, but because we did not want to model the synthesis of chorismate and we said that this substrate is considered to be always present, we have had two choices: supplying an infinite amount of it or to never consume it. We think the second choice is a fair solution

3.3.2 Feedback inhibition

Feedback inhibition (see section 2.2.2 on page 13), is the process which negatively controls the biosynthesis acting on the first enzymatic complex (Anthranilate Synthase).

```
rule 46: AnthranilateSynthase + Trp <=> AnthranilateSynthase-Trp
rule 47: AnthranilateSynthase-Trp + Trp
      <=> AnthranilateSynthase-Trp-Trp
```

Starting from reaction 46, Anthranilate Synthase is subtracted from the biosynthetic pathway (as seen previously in reaction 45) because it changes in a complex with Trp. The more Trp is present, the more Anthranilate Synthase complexes are inhibited. The reaction is reversible, thus even the opposite is possible: when Trp concentration decreases, the pathway is back again.

3.4 Previous models of Trp regulation

Previous modeling studies about tryptophan biosynthesis and regulation applied various modeling frameworks. We briefly present three recent and significant examples.

The goal of Simao et al. (2005) is to establish a method to translate logical regulatory networks models into Petri nets (PN). It is an approach that solves the issues related to systems of ordinary differential equations: First, the increasing complexity when adding new regulatory processes which prevents proper analysis. Second, difficult dynamical insights are only enabled by numerical computation. The main advantage in translating logical regulatory networks into PN, is the possibility to use the tools already available to analyze Petri nets on the model. The authors decided to focus on repression and feedback inhibition, while for the sake of simplicity they omit attenuation. Their model is further simplified because they consider that

TrpR and TrpE are inhibited by a single molecule of Trp. Their approach is able to show inside a model the role of specific circuits and to establish its importance in regulation. Anyhow we believe that attenuation is the most interesting regulatory element inside Trp regulation in *E. coli* and should be not underestimated as it is accountable for modulating sensitivity to amino acids starvation.

The model of Elf and Ehrenberg (2005) has been designed to answer general questions related to attenuation, and the preference in codon usage related to charged isoacceptors. Their model is general purpose, that is, it can be applied to different attenuators. It has been developed using ODE, and is probabilistic. It considers the step by step movements of ribosome and RNA polymerase as a Poisson Process. The authors focused on Trp and His (histidine) operons to understand the relationship between on one side length of the attenuator region, the number of codons inside the attenuator encoding the amino acid produced by the regulated pathway and bias in codons type usage, and on the other side the hyper-sensitivity to amino acid starvation and basal level expression. As said before, Trp is encoded by only one triple of nucleotides (UGG) and thus there are not multiple isoacceptors. The focus of *our* work is not on comparing different attenuators. However Elf and Ehrenberg's work is a great example for the meaningful results achievable with a good model.

The work of Santillán and Zeron (2004) covers the three main regulatory processes in Trp biosynthesis: using differential equations, they take care of repression, feedback inhibition and attenuation. What is really interesting in their approach is the attempt to differentiate the effects of every regulation process on the level of tryptophan inside the cell. As an example, disabling enzyme inhibition has the most important effect on production of Trp, causing an excess of synthesis of Trp at the beginning. This excess is then counterbalanced by the repression activity which reduces the amount of total enzymes translated. The initial excess can be explained by the fact that without inhibition, the excess of Trp is sensed by a longer circuit and it is regulated only by repression after a while. On the contrary, if repression is absent, feedback inhibition completely preserves the standard behavior in Trp synthesis. These results are difficult to carry out in laboratory and they well represent one of the advantages in modeling and simulation. We believe that our model should be exploited to perform similar experiments adding a stochastic behavior to the system.

Chapter 4

Conclusion

Our goal was to test our novel rule based modelling formalism and its expressivity. As example, we selected the Trp biosynthesis and regulation. They constitute a well known biological process which has been modelled in several ways, leading to interesting results even recently. This allows us to compare our model and the results we can retrieve from it with other similar works. We have been able to demonstrate the usefulness of an element available in our modeling language, *parametrization*, that is not supported by other rule based formalisms.

At the example of transcriptional attenuation (see section 3.2.3) we showed how parametrization allows to distinguish mRNA produced at different time points, i.e. to identify the sub-parts of a polycistronic mRNA, created in the same round of transcription. Without parametrization, after a first iteration of the model it would not be possible for the system to distinguish clearly between two different mRNAs. This would easily lead the system to inconsistent simulation results. But parametrization prohibits these in a simple manner. Let's suppose that a RNA polymerase is transcribing the Leader4 while a Ribosome is blocked over RNA_L1 and thus there are RNA_L2 and RNA_L3 available. We expect that an anti-terminator loop is going to form, but at the same time another RNA polymerase has started transcription, is over Leader2 and has transcribed another RNA_L1. In this situation, without parameters, a pause loop could form between RNA_L2 synthesized by the first RNA polymerase and RNA_L1 synthesized by the second RNA polymerase.

To the best of our knowledge, the control in transcriptional attenuation could not well be rendered in related concurrent modeling languages. The reaction rules of BioCham (Chabrier-Rivier et al., 2005) rewrite *atoms*, not parametrized terms as does our language. Process algebra based languages take an object centered perspective, and focus on interactions between *pairs*

of molecular actors (Milner, 1999; Regev and Shapiro, 2002). Our case requires to consider *multiple* actors, and their synchronization required to carry out one reaction step. Encoding such dependencies in the object centered paradigm of process calculi requires sophisticated tricks (Ciocchetta and Priami, 2006; Cleaveland et al., 2001), that are hardly applicable or understandable for non experts. It remains to clarify if variants of Petri nets might capture the necessary control. For a finite number of transcription rounds, colored Petri nets seem useful.

Summarizing, in this thesis we demonstrated that our rule based modeling language allows to cover the concurrent control of transcriptional attenuation. We believe it is accessible to a wider community of modelers than previous concurrent frameworks, and that it can be applied to many other biological systems.

Future work We validated the interactions incorporated in our model by inspection of reachability and interaction graphs generated by BioCham (see Fig.3.1 and Fig.3.2). We discarded an initial ambition of a more detailed qualitative validation of our model, via model checking in BioCham. The problem was that, as mentioned above, parameters are not provided by BioCham, such that key properties of our model could actually not be investigated. The next step before quantitative prediction would be to determine the necessary reaction rates and kinetics, to complete our model. Much data is already included in previous publications (Santillán and Zeron, 2004; Elf and Ehrenberg, 2005) regarding the same topic and books on molecular biology (Lodish et al., 2007; Lewin, 2004), but we have found some inconsistency among different authors that we could address. This would require the implementation of a dedicated simulation engine for our language, which for reasons explained above, should include parametrization. After, we will be ready for simulation and for gathering the results. If we obtained encouraging results, then we could try to apply our language to other biological cases, and check if it fits well to new needs. We think that modelling many different biological processes will force us to face the limits of our language. In fact it was our purpose to model Trp operon to show us the benefits given by parametrization.

Bibliography

- Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter (2007, November). *Molecular Biology of the Cell* (5 ed.). Garland Science.
- Bhartiya, S., N. Chaudhary, K. V. Venkatesh, and F. J. I. Doyle (2006). Multiple feedback loop design in the tryptophan regulatory network of *Escherichia coli* suggests a paradigm for robust regulation of processes in series. *Journal of the Royal Society Interface* 3(8), 383–391.
- Chabrier-Rivier, N., M. Chiaverini, V. Danos, F. Fages, , and V. Schächter (2004). Modeling and querying biomolecular interaction networks. *Theoretical Computer Science* 325(1), 24–44.
- Chabrier-Rivier, N., F. Fages, and S. Soliman (2005). The biochemical abstract machine BioCham. In *Proceedings of CMSB 2004*, Volume 3082 of *Lecture Notes in Bioinformatics*, pp. 172–191.
- Ciocchetta, F. and C. Priami (2006). Biological transactions for quantitative models. In *Proceedings of MeCBIC 2006*, Electronic Notes in Theoretical Computer Science. Elsevier. to appear.
- Cleaveland, R., G. Lüttgen, and V. Natarajan (2001). *Handbook of Process Algebra*, Chapter Priority in Process Algebras, pp. 711–765. Elsevier.
- Danos, V. and C. Laneve (2004). Formal molecular biology. *Theoretical Computer Science* 325(1), 69–110.
- Elf, J. and M. Ehrenberg (2005, June). What makes ribosome-mediated transcriptional attenuation sensitive to amino acid limitation? *PLoS computational biology* 1, e2. PMID: 16103903.
- Elf, J., D. Nilsson, T. Tenson, and M. Ehrenberg (2003). Selective charging of tRNA isoacceptors explains patterns of codon usage. *Science* 300, 1718–1722.

- Faeder, J. R., M. L. Blinov, B. Goldstein, and W. S. Hlavacek (2005). Rule-based modeling of biochemical networks. *Complexity* 10(4), 22–41.
- Horowitz, H. and T. Platt (1982). Identification of trp-p 2, an internal promoter in the tryptophan operon of *Escherichia coli*. *Journal of Molecular Biology* 156, 257–267.
- Khodursky, A. B., B. J. Peter, N. R. Cozzarelli, D. Botstein, P. O. Brown, and C. Yanofsky (2000). DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 97(22), 220414297.
- Lewin, B. (2004). *Genes VIII*. Pearson Prentice Hall Upper Saddle River, NJ.
- Lodish, H., A. Berk, C. A. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh, and P. T. Matsudaira (2007, August). *Molecular Cell Biology* (6Rev Ed ed.). W.H.Freeman & Co Ltd.
- Milner, R. (1999). *Communicating and Mobile Systems: the π -calculus*. Cambridge University Press.
- Nudler, E. and M. E. Gottesman (2002). Transcription termination and anti-termination in *E. coli*. *Genes to Cells* 7, 755–768.
- Oppenheim, D. S. and C. Yanofsky (1980). Translational coupling during expression of the tryptophan operon of *Escherichia coli*. *Genetics* 95, 785–795.
- Regev, A. and E. Shapiro (2002). Cells as computation. *Nature* 419, 343.
- Santillán, M. and E. S. Zeron (2004). Dynamic influence of feedback enzyme inhibition and transcription attenuation on the tryptophan operon response to nutritional shifts. *Journal of Theoretical Biology* 231, 287–298.
- Simao, E., E. Remy, D. Thieffry, and C. Chaouiya (2005). Qualitative modelling of regulated metabolic pathways: application to the tryptophan biosynthesis in *E. Coli*. *Bioinformatics* 21, 190–196.
- Trun, N. J. and J. E. Trempy (2004). *Fundamental Bacterial Genetics*. Blackwell Publishing.

- Xie, G., N. O. Keyhani, C. A. Bonner, and R. A. Jensen (2003). Ancient origin of the tryptophan operon and the dynamics of evolutionary change. *Microbiology and Molecular Biology Reviews* 67, 303.
- Yanofsky, C. (2000). Transcription attenuation: Once viewed as a novel regulatory strategy. *Journal of Bacteriology* 182(1), 1–8.
- Yanofsky, C. (2001). Advancing our knowledge in biochemistry, genetics, and microbiology through studies on Tryptophan metabolism. *Annual Review of Biochemistry* 70, 1–37.
- Yanofsky, C. (2003). *Using Studies on Tryptophan Metabolism to Answer Basic Biological Questions*, Volume 278. ASBMB.
- Yanofsky, C. (2004, August). The different roles of tryptophan transfer RNA in regulating trp operon expression in *E. Coli* versus *B. subtilis*. *Trends in Genetics* 20, 367–374.
- Yanofsky, C. (2007). RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. *RNA* 13, 1141.
- Yanofsky, C. and V. Horn (1994). Role of regulatory features of the trp operon of *Escherichia coli* in mediating a response to a nutritional shift. *Journal of Bacteriology* 176, 6245.
- Yanofsky, C., V. Horn, M. Bonner, and S. Stasiowski (1971). Polarity and enzyme functions in mutants of the first three genes of the tryptophan operon of *Escherichia coli*. *Genetics* 69, 409–433.
- Yanofsky, C., T. Platt, I. P. Crawford, B. P. Nichols, G. E. Christie, H. Horowitz, M. V. Cleemput, and A. M. Wu (1981). The complete nucleotide sequence of the tryptophan operon of *Escherichia coli*. *Nucleic Acids Research* 9, 6647–6668.

Appendix

Complete rule set

```
=====
#Components of the system:
#One gene Tryptophan Operon divided in:
#Promoter + RBS + Leader1 + Leader2 + Leader3 + Leader4 +
# Leader5 + trpE + trpD + trpC + trpB + trpA + RhoTerm
#Ribosomes
#RNA Polymerases = RNAP
#Chorismate
#Tryptophan = Trp
#Repressor
#Protease
#RNase

###Repression
1 Trp + Repressor <=> Repressor-Trp
2 Trp + Repressor-Trp <=> Repressor-Trp-Trp
3 Promoter + Repressor-Trp-Trp <=> Repressor-Trp-Trp-Promoter

###Transcription
4 Promoter + RNAP <=> Promoter-RNAP
5 Promoter-RNAP + RBS => new n.RBS-RNAP(n) + Promoter
6 RBS-RNAP(n) + Leader1 => RBS + Leader1-RNAP(n) + RNA_RBS(n)
7 Leader1-RNAP(n) + Leader2 => Leader1 + Leader2-RNAP(n) + RNA_L1(n)
8 Leader2-RNAP(n) + Leader3 => Leader2 + Leader3-RNAP(n) + RNA_L2(n)
9 Leader3-RNAP(n) + Leader4 =[RNA_L2(n)]=>
Leader3 + Leader4-RNAP(n) + RNA_L3(n)
10 Leader4-RNAP(n) + Leader5 => Leader4 + Leader5-RNAP(n) + RNA_L4(n)
11 Leader5-RNAP(n) + trpE =[AntiTerminatorLoop2_3(n)]=>
Leader5 + RNA_L5(n) + trpE-RNAP(n)
```

```

12 Leader5-RNAP(n) =[TerminatorLoop3_4(n)]=> Leader5 + RNAP
13 Leader5-RNAP(n) =[PauseLoop1_2(n), TerminatorLoop3_4(n)]=>
Leader5 + RNAP

###Transcription of all the Operon
14 trpE-RNAP(n) + trpD => trpE + trpD-RNAP(n) + mRNA_trpE(n)
15 trpD-RNAP(n) + trpC => trpD + trpC-RNAP(n) + mRNA_trpD(n)
16 trpC-RNAP(n)+ trpB => trpC + trpB-RNAP(n) + mRNA_trpC(n)
17 trpB-RNAP(n) + trpA => trpB + trpA-RNAP(n) + mRNA_trpB(n)
18 trpA-RNAP(n) + RhoTerm => trpA + RhoTerm-RNAP(n) + mRNA_trpA(n)
19 RhoTerm-RNAP(n) => RNAP +RhoTerm

###Loop formation for attenuation
20 RNA_L1(n) + RNA_L2(n) => PauseLoop1_2(n)
21 RNA_L3(n) + RNA_L4(n) =[Ribosome-RNA_L2(n)]=> TerminatorLoop3_4(n)
22 RNA_L2(n) + RNA_L3(n) =[Ribosome-RNA_L1(n), Leader4-RNAP(n)]=>
AntiTerminatorLoop2_3(n)
23 RNA_L2(n) + RNA_L3(n) =[Ribosome-RNA_L1_trp(n), Leader4-RNAP(n)]=>
AntiTerminatorLoop2_3(n)

##Loop formation in read through
24 RNA_L2(n) + RNA_L3(n) =[RNA_L1(n), RNA_L4(n), Leader5-RNAP(n)]=>
AntiTerminatorLoop2_3(n)
25 RNA_L1(n) + RNA_L2(n) + RNA_L3(n) + RNA_L4(n) =[Leader5-RNAP(n)]=>
PauseLoop1_2(n) + TerminatorLoop3_4(n)

###Superattenuation
26 Leader3-RNAP(n) + Leader4 =[PauseLoop1-2(n)]=>
Leader3 + Leader4-RNAP(n) + RNA_L3(n)
27 Leader4-RNAP(n) + Leader5 =[PauseLoop1_2(n), RNA_L3(n)]=>
Leader5-RNAP(n) + RNA_L4(n)
28 RNA_L3(n) + RNA_L4(n) =[Leader5-RNAP(n), PauseLoop1_2(n)]=>
TerminatorLoop3_4(n)

###Translation
29 Ribosome + RNA_RBS(n) =[PauseLoop1_2(n), Leader3-RNAP(n)]=>
Ribosome-RNA_RBS(n)
30 Ribosome-RNA_RBS(n) + PauseLoop1_2(n) =[Leader3-RNAP(n)]=>
RNA_RBS(n) + Ribosome-RNA_L1(n) + RNA_L2(n)
31 Ribosome-RNA_L1(n) + Trp => Ribosome-RNA_L1_trp(n)
32 Ribosome-RNA_L1_trp(n) + Trp + RNA_L2(n) => Ribosome-RNA_L2(n) + RNA_L1(n)

```

```
33 Ribosome-RNA_L2(n) => RNA_L2(n) + Ribosome + TrpL

##freeing ribosome from stalled position
34 Ribosome-RNA_L1(n) + Trp
=[AntiTerminatorLoop2_3(n), RNA_L4(n), RNA_RBS(n), RNA_L5(n)]=>
Ribosome-RNA_L1_trp(n)
35 Ribosome-RNA_L1_trp(n) + Trp + AntiTerminatorLoop2_3(n)
=[RNA_L4(n) + RNA_L5(n) + RNA_RBS(n)]=>
RNA_L1(n) + Ribosome-RNA_L2(n) + RNA_L3(n) +
36 RNA_RBS(n) + RNA_L1(n) + RNA_L2(n) + RNA_L3(n) +
RNA_L4(n) + RNA_L5(n) => mRNA_trpL(n)

###Translation of all the enzymes
37 Ribosome + mRNA_trpE(n) => TrpE + Ribosome + mRNA_trpE(n)
38 Ribosome + mRNA_trpD(n) => TrpD + Ribosome + mRNA_trpD(n)
39 Ribosome + mRNA_trpC(n) => TrpC + Ribosome + mRNA_trpC(n)
40 Ribosome + mRNA_trpB(n) => TrpB + Ribosome + mRNA_trpB(n)
41 Ribosome + mRNA_trpA(n) => TrpA + Ribosome + mRNA_trpA(n)
42 Ribosome + mRNA_trpL(n) => TrpL + Ribosome + mRNA_trpL(n)

###Biosynthesis of tryptophan
43 TrpE + TrpE + TrpD + TrpD => AnthranilateSynthase
44 TrpB + TrpB + TrpA + TrpA => TrpSynthase
45 Chorismate + AnthranilateSynthase + TrpC + TrpSynthase =>
AnthranilateSynthase + TrpC + TrpSynthase + Trp + Chorismate

###Inhibition of Biosynthesis of tryptophan
46 AnthranilateSynthase + Trp <=> AnthranilateSynthase-Trp
47 AnthranilateSynthase-Trp + Trp <=> AnthranilateSynthase-Trp-Trp

###Degradation of mRNA
48 mRNA_trpL(n) + RNase => mRNA_trpL-RNase(n)
49 mRNA_trpE(n)+mRNA_trpL-RNase(n) => mRNA_trpE-RNase(n)
50 mRNA_trpD(n)+mRNA_trpE-RNase(n) => mRNA_trpD-RNase(n)
51 mRNA_trpC(n)+mRNA_trpD-RNase(n) => mRNA_trpC-RNase(n)
52 mRNA_trpB(n)+mRNA_trpC-RNase(n) => mRNA_trpB-RNase(n)
53 mRNA_trpA(n)+mRNA_trpB-RNase(n) => mRNA_trpA-RNase(n)
54 mRNA_trpA-RNase(n) => RNase

###Degradation of enzymes
```

55 AnthranilateSynthase + Protease => Protease
56 AnthranilateSynthase-Trp + Protease => Protease
57 AnthranilateSynthase-Trp-Trp + Protease => Protease
58 TrpC + Protease => Protease
59 TrpSynthase + Protease => Protease
60 TrpL + Protease => Protease