AUTONOMOUS MOBILE ROBOTS IN NATURE:

TRANSFER RNAs IN *Escherichia coli* BACTERIA

A THESIS IN MECHATRONICS

Presented to the faculty of the American University of Sharjah
College of Engineering
in partial fulfillment of
the requirements for the degree

MASTER OF SCIENCE

by
SÁNDOR PIROS
M.S.E.E. 1984

Sharjah, UAE
May 2008

We approve the thesis of Piros Sandor — Date of signature

Ghaleb Husseini — May 18, 2008

Dr. Ghaleb A. Husseini
Associate Professor, Chemical Engineering
Thesis Advisor

H. Nashash — May 21, 08

Dr. Hassan Al-Nashash
Professor, Electrical Engineering
Graduate Committee

Imad Abukhalaf — May 21, 2008

Dr. Imad Abukhalaf
DuBiotech Director of Science Affairs
Graduate Committee

R. Dhaouadi — May 21, 2008

Dr. Rached Dhaouadi
Associate Professor
Coordinator, Mechatronics Graduate Program

Jamal A. Abdalla — 21/5/2008

Professor Jamal A. Abdalla
Director, Graduate Programs

Yousef Al-Assaf — 21/5/08

Professor Yousef Al-Assaf
Dean, College of Engineering

K Mitchell — 21/5/2008

Mr. Kevin Lewis Mitchell
Director, Graduate & Undergraduate Programs

AN ABSTRACT IN AN AMERICAN
UNIVERSITY OF SHARJAH THESIS:

AUTONOMOUS MOBILE ROBOTS IN NATURE: TRANSFER RNAs IN *Escherichia coli* BACTERIA

Sándor Piros, Candidate for the Master of Science Degree
American University of Sharjah, 2008

ABSTRACT

Transfer RNAs (tRNAs) are the smallest autonomous robots on earth. They can recognize a specific amino acid from the possible pool of 20 different amino acids. They are able to transport these protein building blocks to the ribosome, the site of amino acid assembly into protein chains. Accurate and rapid selection of tRNAs by the ribosome is critical for cell viability. There is no description in literature about their movement in the cytoplasm, but there is extensive research about tRNA recognition, selection and their relative movements into or inside the ribosome.

The aim of this master thesis is to develop a model of tRNA molecular movement in bacterial cytoplasm. One of the main criteria for protein synthesis is the availability of the necessary amino acid in the vicinity of the ribosome, according to the sequence coded by a gene.

The theme of this master thesis is the spatial movement/placement of aminoacylated tRNAs in the cytoplasm - viewed from the perspective of that particular aminoacylated tRNA (charged with a specific amino acid). A kinetic model of the messenger RNA – ribosome – tRNA system was developed and a computerized simulation was built to visualize different scenarios. The purpose of the simulation is to show the conditions necessary for the tRNA to deliver the particular amino acid to the ribosome within a similar timeframe to what happens *in vivo*.

Simulations results show that it is very unlikely that tRNAs are able to get to the ribosome by random movement, and there should be a certain mechanism to preselect the cognate tRNAs. Three hypotheses were developed to give possible explanations to this process, and provide a useful guide to future research and simulations in biology and microbiology.

CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

LIST OF APPENDICES

Appendix Page

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Ghaleb Al-Husseini for his guidance and his assistance during my thesis and throughout my study period.

Special thanks to Professor Mohammed Ameen Al-Jarrah for his helpful comments and generous support of my thesis, and to Dr. Khaled Assaleh who made me an expert in Matlab programming, to Dr. Mohamed Gadalla in mechanical studies.

Finally I have to mention Dr. Erika Zilahi, Dr Rita Kövesdi, Dr. Ida Miklós and Dr. Zoltán Papp; they were my first supporters, when I started my self-study in molecular biology and genetics.

I would also like to thank my family for their understanding and efforts in arranging the supporting atmosphere to focus on my thesis and to my wife for providing actual help in the literature search.

Thanks to Ms. Lujein Osama Farhat, AUS student for helping me with the editing process.

VITA

Sándor Piros was born 27 Oct 1960, in Debrecen, Hungary. He completed his secondary school studies in 1979 in Fazekas Mihály Secondary School on Special Mathematics Specialization.

He received his M.Sc. in electrical engineering in 1984 from the Technical University of Budapest, Hungary. He was an active member of the university's Scientific Student's Circle. In the academic year 1980/81 student competition, which was organized by the Faculty of Electrical Engineering, he was awarded 3rd prize,.

Since 1996 he has been enthusiastically studying molecular and developmental biology besides engineering.

He started his Master Degree Program in Mechatronics at the American University of Sharjah in 2005.

ABBREVIATIONS

Abbreviation	Term
aa	aminoacyl
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
tRNA	transfer RNA
ctRNA	cognate tRNA
mRNA	messenger RNA
rRNA	ribosomal RNA
A	adenine
C	cytosine
G	guanine
U	uracil
T	thymine
nt	nucleotide
bp	base pair
kD	kilo Dalton
S	Svedberg

GLOSSARY

Gene – logical unit of inheritance; they are physically stored in the chromosomes as strand of double helix DNA.

Amino acid – an organic compound containing at least one amino group and one carboxyl group. In the 20 different amino acids that compose proteins, an amino group and carboxyl group are linked to a central carbon atom, to which a variable side chain is bound.

Aminoacyl-tRNA – activated form of an amino acid, used in protein synthesis, consisting of an amino acid linked via a high energy ester bond to the 3`-hydroxyl group of a tRNA molecule.

Anticodon – sequence of three nucleotides in a tRNA that is complementary to a codon in an mRNA. During protein synthesis, base pairing between a codon and anticodon aligns the tRNA carrying the corresponding amino acid for addition to the growing peptide chain.

Base pair – association of two complementary nucleotides in a DNA or RNA molecule stabilized by hydrogen bonding between their base components. Adenine pairs with thymine or uracil, and guanine pairs with cytosine.

Chromosome – in eukaryotes, the structural unit of the genetic material consisting of a single, linear double-stranded DNA molecule and associated proteins. In prokaryotes, a single, circular double-stranded DNA molecule constitutes the bulk of the genetic material.

Codon – sequence of three nucleotides in DNA or mRNA that specifies a particular amino acid during protein synthesis; also called triplet. Of the 64 possible codons, three are stop codons, which do not specify amino acids.

DNA – long linear polymer, composed of four kinds of deoxyribose nucleotides, that is the carrier of genetic information. In its native state, DNA is a double helix of two antiparallel strands held together by hydrogen bonds between complementary purine and pyrimidine bases.

Elongation factor – one of a group of nonribosomal proteins required for continued translation of mRNA following initiation.

Enzyme – a biological macromolecule that acts as a catalyst. Most enzymes are proteins.

Eukaryotes – class of organisms, composed of one or more cells containing a membrane-enclosed nucleus and organelles, that constitutes one of the three distinct evolutionary lineages of modern-day organisms. Includes all organisms except viruses and prokaryotes.

Gene – physical and functional unit of heredity, which carries information from one generation to the next. In molecular terms, it is the entire DNA sequence – including exons, introns, and noncoding transcription-control regions – necessary for production of a functional protein or RNA.

Genome – total genetic information carried by a cell or organism.

Hydrogen bond – a noncovalent bond between an electronegative atom and a hydrogen atom covalently bonded to another electronegative atom. Particularly important in stabilizing the three-dimensional structure of proteins and formation of base pairs in nucleic acids.

Macromolecule – any large, usually polymeric molecule (e.g. a protein, nucleic acid, polysaccharide) with a molecular mass greater than a few thousand daltons.

mRNA – the RNA that specifies the order of amino acids in a protein.

Nucleotide – a nucleoside with one or more phosphate groups linked via an ester bond to the sugar moiety. DNA and RNA are polymers of nucleotides.

Nucleus – large membrane-bounded organelle in eukaryotic cells that contains DNA organized into chromosomes; synthesis and processing of RNA and ribosome assembly occur in the nucleus.

Peptide bond – covalent bond that links adjacent amino acid residues in proteins; formed by a condensation reaction between the amino group of one amino acid and the carboxyl group of another with release of a water molecule.

Primary structure – in proteins, the linear arrangement (sequence) of amino acids and the location of covalent bonds within a polypeptide chain.

Prokaryotes – class of organisms, including the eubacteria and archaea, that lack a true membrane-limited nucleus and other organelles.

Protein – a linear polymer of amino acids linked together in a specific sequence and usually containing more than 50 residues. Proteins form the key structural elements in cells and participate in nearly all cellular functions.

Reading frame – the sequence of nucleotide triplets that runs from a specific translation start codon in a mRNA to a stop codon. Some mRNAs can be translated into different polypeptides by reading them in two different reading frames.

Ribosome – a large complex comprising several different rRNA molecules and more than 50 proteins, organized into a large subunit and small subunits; the site of protein synthesis.

RNA – linear, single-stranded polymer, composed of ribose nucleotides, that is synthesized by transcription of DNA or by copying of RNA. Three types of cellular RNA play roles in protein synthesis.

rRNA – any one of several large RNA molecules that are structural and functional components of ribosomes.

Secondary structure – in proteins, local folding of a polypeptide chain into regular structures including the α helix, β sheet, and U-shaped turns and loops. tRNA molecules also have secondary structure by local folding of nucleotides forming a double-stranded chain.

Tertiary structure – in proteins, overall three-dimensional form of a polypeptide chain, which is stabilized by multiple noncovalent interactions between side chains. tRNA molecules also have tertiary structure.

Transcription – process whereby one strand of a DNA molecule is used as a template for synthesis of a complementary RNA by RNA polymerase.

Translation – the ribosome-mediated production of a polypeptide whose amino acid sequence is specified by the nucleotide sequence in an mRNA.

tRNA – a group of small RNA molecules that functions as amino acid donors during protein synthesis. Each tRNA becomes covalently linked to a particular amino acid, forming an aminoacyl-tRNA.

The definitions of the terms were taken from [1].

PREFACE

I began the “self-research” process a long time ago, but recently acquired knowledge through Mechatronics (at the American University of Sharjah) especially in the field of modeling and simulation of dynamic systems, provided me with more powerful tools in my investigation in order to find possible explanations.

Work on this master thesis was preceded by years of preliminary studies in Biology. Most valuable sources of knowledge were books before the internet boom. Now, of course, the net is a major source of information.

There is a need to give a glimpse into an engineer’s mind or the way he/she usually approaches such a problem, taken from a seemingly non-engineering discipline. Some unanswered question need to be investigated. This effort is for everyone to appreciate and build upon.

CHAPTER 1.

INTRODUCTION

Biology as an independent science was developed in the 19th century, when scientists discovered that organisms shared basic characteristics. Molecular biologists made important findings and learned to isolate, characterize, and manipulate the molecular components of cells and organisms in the last five decades. The structures of macromolecules such as deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and proteins are already well known. Molecular biology also studies the complex interactions between the systems of the biological molecules, including the interrelationship of DNA, RNA, and protein biosynthesis, in addition to learning how these interactions are regulated [2], [3].

Proteins are macromolecules that are the building-blocks of cells. Proteins are assembled from 20 different amino acids; usually a few hundred of them are connected into a chain by peptide bonds. The amino acid sequence of the protein is coded by the messenger RNA (mRNA). The factory where the assembling takes place is the ribosome. The ribosome is a conglomerate of RNA molecules and proteins. The template of these RNA molecules is stored in the chromosome – similar to a library – in the form of double helix DNA chains. DNA contains the genetic information used in the development of all living organisms. DNA plays a central role in the long-term storage of genetic information [4], [5].

1.1. Problem Statement

The manufacturing of proteins in the ribosome is carried out by the help of tiny industrial robots: the transfer RNA (tRNA) molecules. They are the smallest autonomous robots on earth. They can recognize a specific amino acid from the possible pool of 20 different amino acids. They are able to transport these blocks to the ribosome – the site of amino acid assembling into protein chains. A tRNA molecule acts like a mobile robot, as it delivers the spare parts – amino acids – to the assembly-line (the ribosome). The structure of tRNA molecules is also well known (Figure 1). There is a wide range of research on tRNA interaction with the ribosome,

but all focus on their biochemical reactions. There are other factors of great significance that play an integral part of this assembly. The ribosome is capable of distinguishing the cognate tRNA (ctRNA) from all other tRNA molecules, but it accomplishes this task by trial and error, which is extremely time consuming. Probably there is a method in nature to select the ctRNA even before it enters the ribosome A site. There is no description in literature on their movement in the cytoplasm, but there is extensive research on tRNA recognition, tRNA selection and relative movements into or inside the ribosome [17], [22].

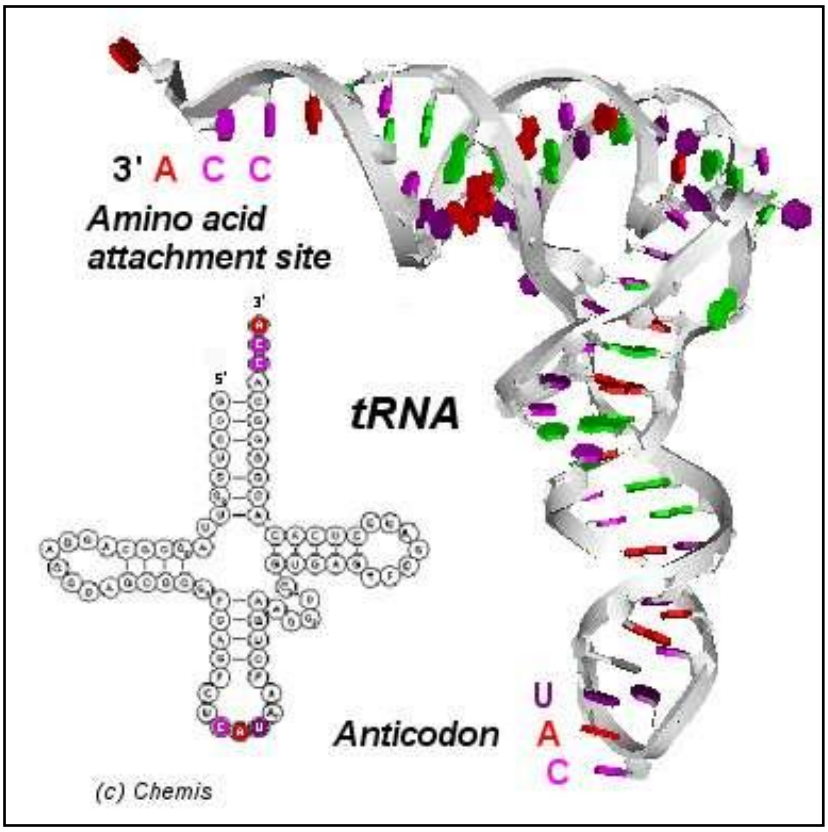


Figure 1. Schematic picture of tRNA [6]

1.2. Aim of the Thesis

The aim of this master thesis is to develop a model of tRNA molecular movement in the bacterial cytoplasm and run simulations according to different hypotheses. The main criterion of protein synthesis is the availability of the necessary amino acid in the vicinity of the ribosome according to their order, which is governed by the sequence of that particular gene.

The theme of this master thesis is the spatial movement or placement of aminoacyl-tRNA (aa-tRNA) molecules in the cytoplasm – viewed from the perspective of that particular aa-tRNA (charged with a specific amino acid). The aim is to set up a kinetic model of the mRNA – ribosome – tRNA system and build a computerized simulation to visualize different scenarios. The purpose of the simulation is to show the conditions necessary for the tRNA to be able to deliver the particular amino acid to the ribosome within the same timeframe as it happens *in vivo*.

Since the protein biosynthesis is similar in prokaryotes and eukaryotes, *Escherichia coli* (*E. coli*), a prokaryotic microorganism, has been chosen as a model organism in this study.

1.3. Significance of the Research

Life sciences are very important in science and medicine, because their topics are related to the existence of humankind and their results directly affect our well-being. The building-blocks of our body are proteins. Proteins are assembled from 20 different amino acids; usually few hundreds of them are connected into a chain by peptide bonds (the carboxyl group of one acid is linked to the amino group of the other). The amino acid sequence of the protein is coded by the mRNA. Three constitutive ribonucleic acids determine one code. We can consider an amino acid as a one digit code with 20 letters, and nucleic acids as three digit code with four letter alphabets. The venue where the assembly takes place is called a ribosome; the ribosome is a conglomerate of RNA molecules and proteins. The templates of these RNA molecules are stored in the chromosomes like a library – in the form of double helix DNA chains (the material of inheritance).

The manufacturing of proteins in the ribosome is carried out by the help of tiny industrial robots; the tRNAs. The significance of this research is that it attempts to give a clearer picture and understanding on tRNA's "modus operandi" in assembling amino acids into a protein chain. The more accurate our knowledge is, the more benefit we could harvest for humankind, and the more efficient antibiotics can be synthesized.

CHAPTER 2.

REVIEW OF THE LITERATURE

Cells consist of different components, including DNA, RNA, and proteins. These macromolecules are the most interesting and characteristic molecules of living systems. DNA stores the genetic information of a cell which consists of thousands of genes. Each gene serves as a template on how to build a protein molecule. Proteins are among the most essential macromolecules, which perform vital tasks for the cell functions and serve as building blocks in human tissues. The orientation of the genetic information defines the protein composition and their functions for each cell. For a more complete understanding of this system, a summary of the structure, the importance, and role of these macromolecules is given below.

2.1. DNA Molecule

The principal role of DNA molecules is the long-term storage of genetic information. The unit of DNA that carries this genetic information is called a gene. DNA segments have functional purposes. There are other DNA sequences with structural tasks. They are also involved in regulating the usage of the genetic information. DNA is arranged into structures called chromosomes and the complete set of chromosomes make up a genome within a cell. Eukaryotic organisms such as fungi, plants, and animals contain their DNA inside the cell nucleus, while in prokaryotes such as bacteria (e.g. *E. coli*), DNA is found in the cytoplasm [7], [8].

2.1.1. DNA Structure

DNA is a long polymer, made of monomers called nucleotides. Nucleotides consist of a pentose carbon sugar, a phosphate group and one of four different nitrogenous bases: guanine (G), adenine (A), cytosine (C), and thymine (T). Sugars and phosphate groups are joined by ester bonds (Figure 2).

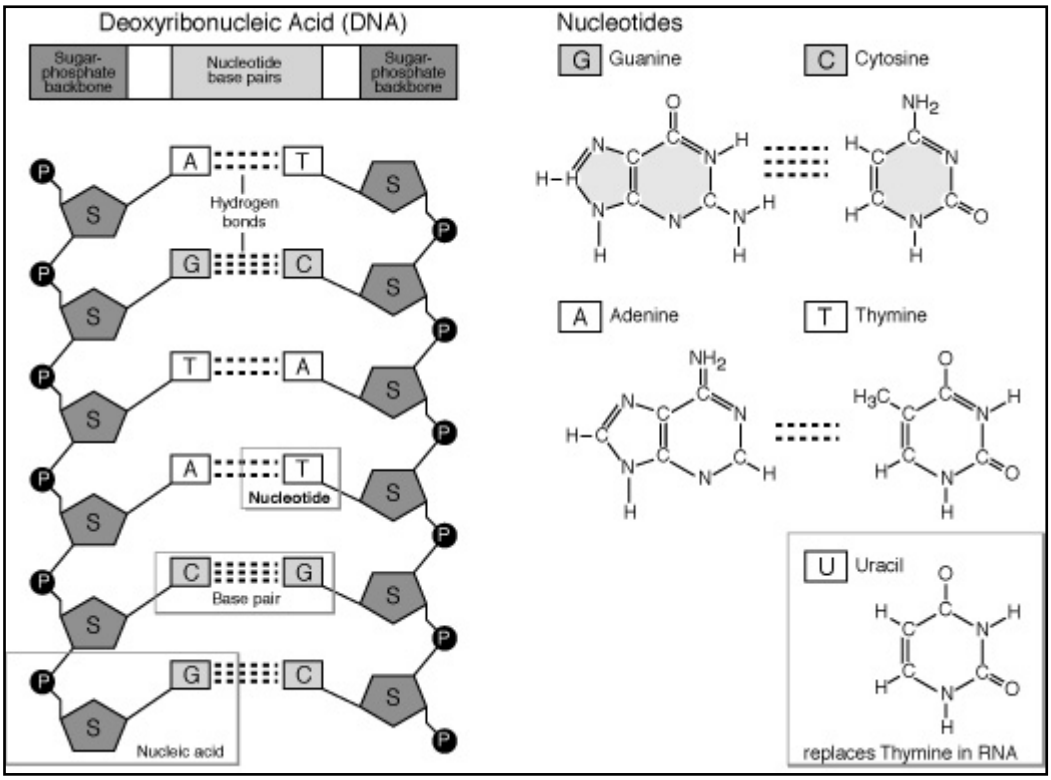


Figure 2. Elements of DNA molecule [9]

In biochemical terminology, a nitrogenous base (G, A, C, or T) linked to a sugar is called nucleoside but a base linked to a sugar and one or more phosphate groups is called a nucleotide (nt). Each individual unit is very small, but DNA polymers can be enormous molecules joined by millions of nucleotides by thevpolymerization of nucleotides. The *Escherichia coli (E. coli)* genome is made of one single chromosome containing around 4640000 nt [9]. DNA are polymers of nucleotides; a phosphodiester bond is formed between the 3` hydroxyl group of one nucleotide and 5` phosphate group of another. These asymmetric bonds imply that a polynucleotide chain has a direction, one end terminating in a 5` phosphate group and the other in a 3` hydroxyl group. In a double helix, the direction of the nucleotide in one strand is opposite to their direction in the reverse strand. This arrangement of DNA strands is called antiparallel. The DNA double helix is stabilized internally by hydrogen bonds between complementary bases (Figure 2 and Figure 3).

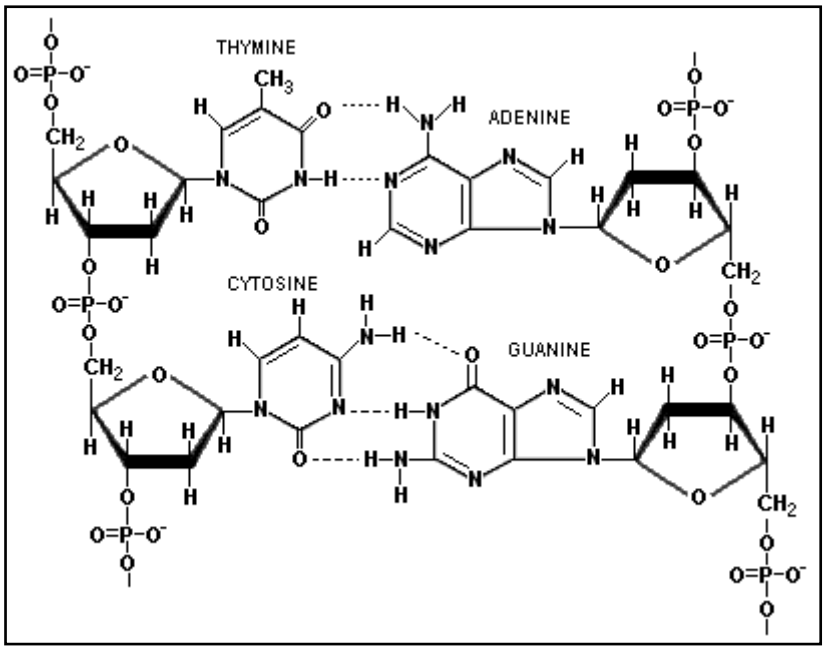


Figure 3. Structure of DNA [10]

The complementary bases are inside the DNA double helix. The two chains are joined by hydrogen bonds between complementary base pairs: C bonding only to G, and A bonding to T. This is called complementary pairing (Figure 1). Bases form hydrogen bonds which can be broken and rejoined easily. The complementary pairing results in the information on the double-stranded sequence of a DNA helix being duplicated on each strand, which is essential in DNA replication and critical to conserve all the functions of DNA in living organisms.

In a genome, genetic information is held within genes. A gene is a unit of heredity and is a region of the DNA that determines a particular characteristic in an organism. *E. coli* has 4441 genes in its genome [9].

2.2. RNA Molecules

RNA is a nucleic acid polymer. It consists of nucleotide monomers. RNA has three distinct roles in protein synthesis, when the genetic information is translated from the DNA into protein products. RNA acts as a messenger between the DNA and the protein synthesis (mRNA), forms an essential part of the ribosome (ribosomal

RNA), and plays an essential role as a carrier molecule for the amino acids used in protein synthesis (tRNA).

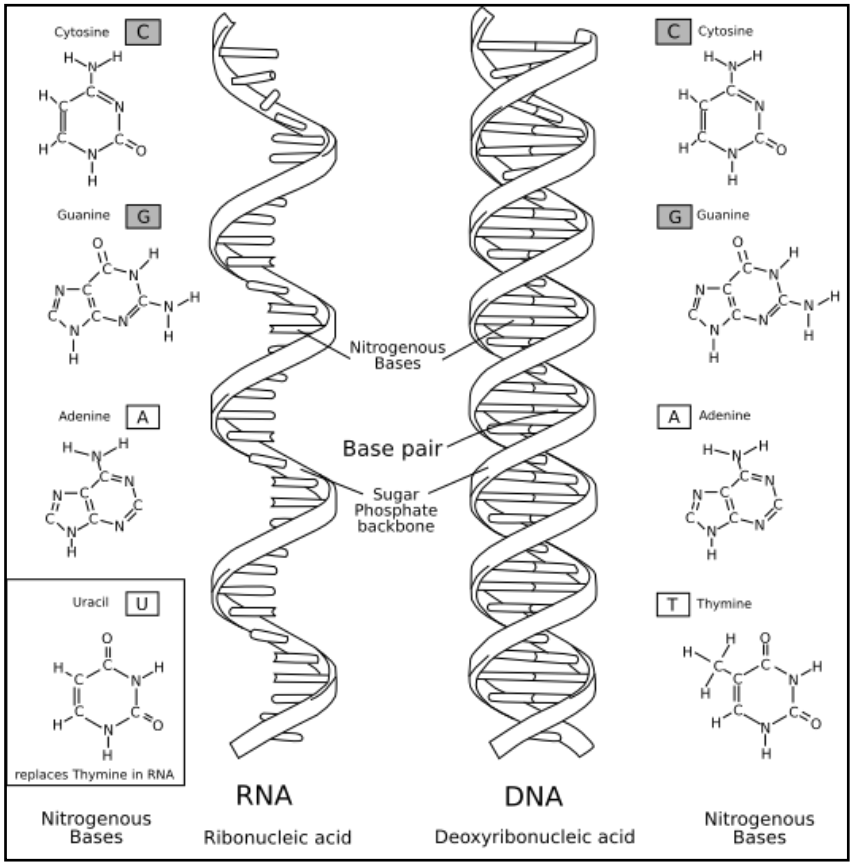


Figure 4. Structure of RNA; Differences between RNA and DNA [11]

2.2.1. RNA Structure

RNA is a polymer which is composed of a ribose and phosphate backbone and four different nitrogenous bases: adenine (A), guanine (G), cytosine (C), and uracil (U). A, G, and C are commonly found in DNA, but T is replaced by U in RNA as the base complementary to adenine (Figure 4).

2.2.2. Types of RNA Molecules

There are three kinds of RNA molecules: mRNA, tRNA, and ribosomal RNA (rRNA). They have diverse but related functions in protein synthesis.

2.2.2.1. Messenger RNA (mRNA)

Messenger RNA carries information from the DNA to the ribosome in cells during protein synthesis, it has been transcribed from DNA. In eukaryotic cells it happens before being transported from the nucleus into the cytoplasm. In the cytoplasm, mRNA is bound to the ribosome and translated into protein with the contribution of tRNAs and other enzymes. It is a different process in prokaryotes (cells without nucleus). In prokaryotic cells, mRNA can bind to the ribosome while it is being transcribed.

2.2.2.2. Transfer RNA (tRNA)

The first yeast tRNA ‘Ala’ was sequenced in 1965 [1]. Since then sequences of more than 4000 tRNAs from more than 30 organisms have been stored in the tRNA database [2] and these numbers increase due to the many genome projects being conducted.

Transfer RNA is a small RNA chain that carries a specific amino acid to the ribosomal site of protein synthesis during translation. It has a 3` terminal site for amino acid attachment (Figure 5). The amino acid attachment is catalyzed by an enzyme called aminoacyl tRNA synthetase. tRNA has an anticodon region that can base pair to the corresponding codon region on mRNA (Figure 5). It also contains a three base region called the anticodon that can pair the base to the corresponding three base codon region on mRNA (Figure 5). One tRNA molecule can be attached to only one type of amino acids. The genetic code contains multiple codons that determine the same amino acid, and different tRNA molecules having different anticodons may also transfer the same amino acid.

The following sections provide more details on tRNAs.

2.2.2.2.1. Transfer RNA Structure

Transfer RNA is a small RNA chain. All have between 74-95 nt. They carry a specific amino acid to the ribosomal site of protein synthesis during translation to the growing polypeptide. The tRNA structure contains at least two specific regions: a site for amino acid attachment, and an anticodon region for codon recognition.

There are more than 20 different tRNA molecules in an organism. tRNAs have a similar structure in all organisms. There are four arms and three loops: the acceptor, the anticodon and the TΨC loops. Sometimes tRNA molecules have an extra or variable loop (Figure 5).

Transfer RNA consists of two parts: the body and the acceptor stem. The body of the tRNA is derived from a tRNA gene. The acceptor stem has conserved its structures as it is the same for all tRNA molecules. It is the main part of the tRNA, and joining after the body is synthesized. The acceptor stem can be replaced during the lifetime of a tRNA molecule and is the site for amino acid attachment by aminoacyl tRNA synthase. The anticodon reads the codon region on mRNA, and base pair to the corresponding codon (Figure 5).

tRNA has a primary, secondary and tertiary structures. All the tRNAs have similar L-shaped 3D structure. It allows the tRNA to fit into the P, A, and E sites of the ribosome. P, A, and E sites are tRNA-binding sites of the ribosome during protein synthesis. All tRNAs end with the "CCA" sequence at the 3`-end. Amino acids are attached to the 3`-ends of tRNAs to create aminoacyl-tRNAs and are referred to as charged tRNAs.

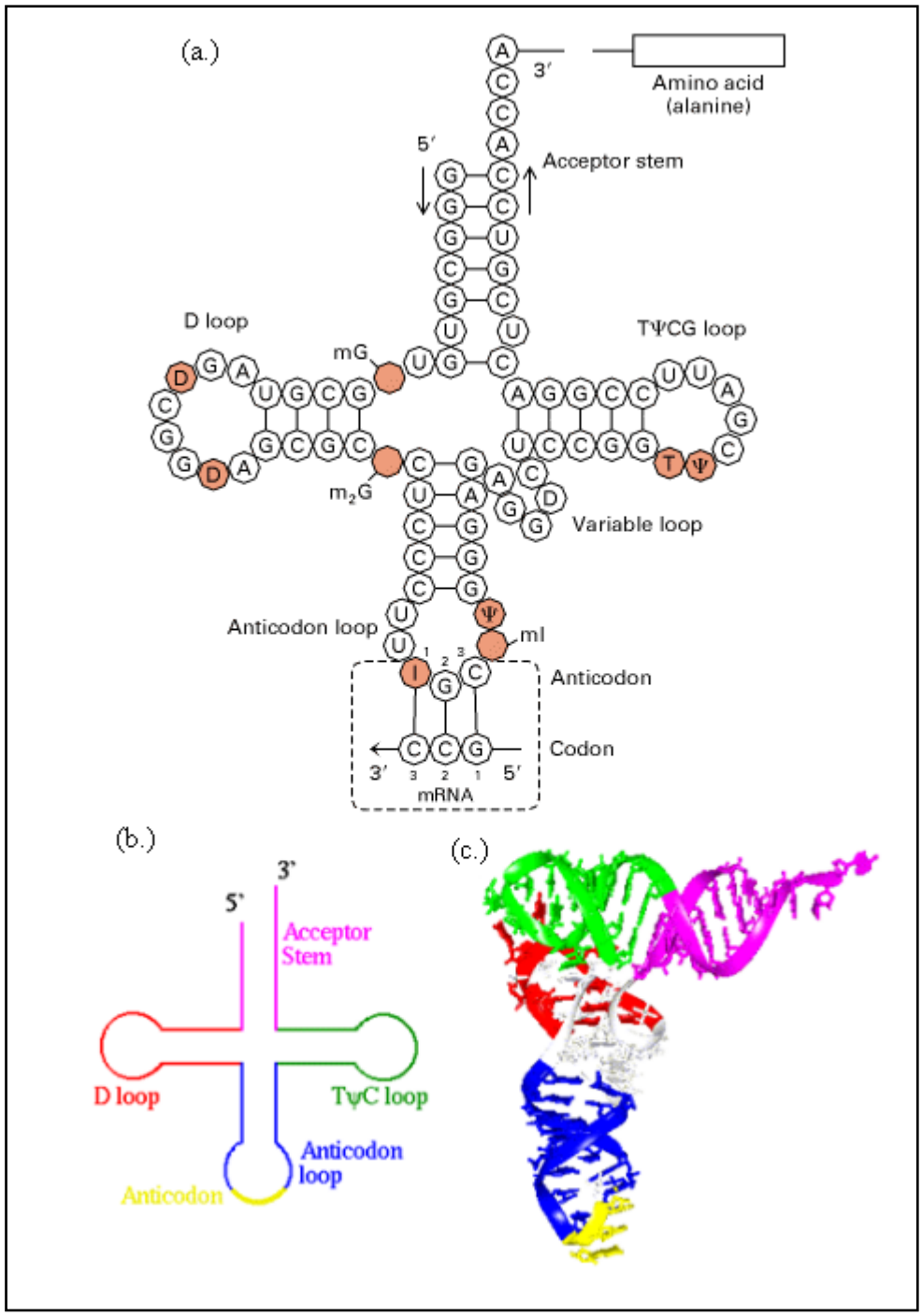


Figure 5. Structure of tRNAs; Secondary and Tertiary structures of tRNA [3]

2.2.2.2.2. Transfer RNA Function

Transfer RNAs play fundamental roles in protein synthesis. They are linked to a particular amino acid and they base pair with a codon on mRNA resulting in the growing of the peptide chain. Each tRNA molecule is recognized by one of the 20

aminoacyl tRNA synthetases. The enzymes bind one of the 20 amino acids to a particular tRNA forming an aminoacyl-tRNA. Attachment of amino acids to tRNAs requires energy which is provided by an adenosine triphosphate (ATP) molecule. The whole process consists of two steps. In the first step, the amino acid is joining to adenosine monophosphate deaminase (AMP). This step results in an aminoacyl AMP intermediate molecule. In the second step, the amino acid is transferred to the 3` CCA terminus of the acceptor region of tRNA. After it is completed the AMP is released. Both steps are catalyzed by aminoacyl tRNA synthetases [4]. After the suitable amino acid is attached, the tRNA recognizes a codon in mRNA, supplying its amino acid to the growing polypeptide chain.

2.2.2.2.3. Number of tRNAs in the Cells

Theoretically**,** one tRNA molecule and one codon specify an amino acid. Sixty one tRNA molecules would be required per cell in these circumstances. Many cells contain less than 61 types of tRNAs because the wobble base is capable of binding to more codons that specify a particular amino acid.

Organisms vary in the number of tRNA genes in their genome. There are 86 tRNAs in *E. coli*. Accordingly, many amino acids have more tRNAs, many tRNAs can attach to more than one codon. Most amino acids are encoded by more than one codon which means that some tRNAs can recognize more than one codon. For example, the amino acid glycine is coded by the codon sequences GGU, GGC, GGA, and GGG (Table 1).

2.2.2.3. Ribosomal RNA (rRNA)

Ribosomal RNA is the catalytic component of ribosomes. Ribosomes are the sites of protein synthesis in prokaryotic and eukaryotic cells. A prokaryotic ribosome contains 23S and 16S rRNA. Eukaryotic ribosome contains four different rRNA molecules: 18S, 5.8S, 28S, and 5S rRNAs. rRNA molecules are designated in units of

svedbergs (S), S the sedimentation rate of suspended particles centrifuged under standard conditions [5].

2.2.2.4. Ribosome

Ribosomes are found in the cytoplasm. The ribosome has two subunits: a large and a small subunit containing characteristic proteins and rRNAs. These molecules come together to form a decoding machine. Cells typically contain many ribosomes, for example there are about 20,000 ribosomes in an *E. coli* cell. The number of ribosomes can reach 10 million in a mammalian cell [5]. These numbers show the importance of ribosomes in the protein synthesis. Figure 6 shows the schematic model of the ribosome which consists of the small and the large subunit with the A (Aminoacyl-tRNA binding site), P (Peptidyl-tRNA binding site), and E (Exit site) sites. A, P, and E sites are tRNA-binding sites of the ribosome and they have important roles during protein synthesis. Several ribosomes may be attached to a single mRNA at any time.

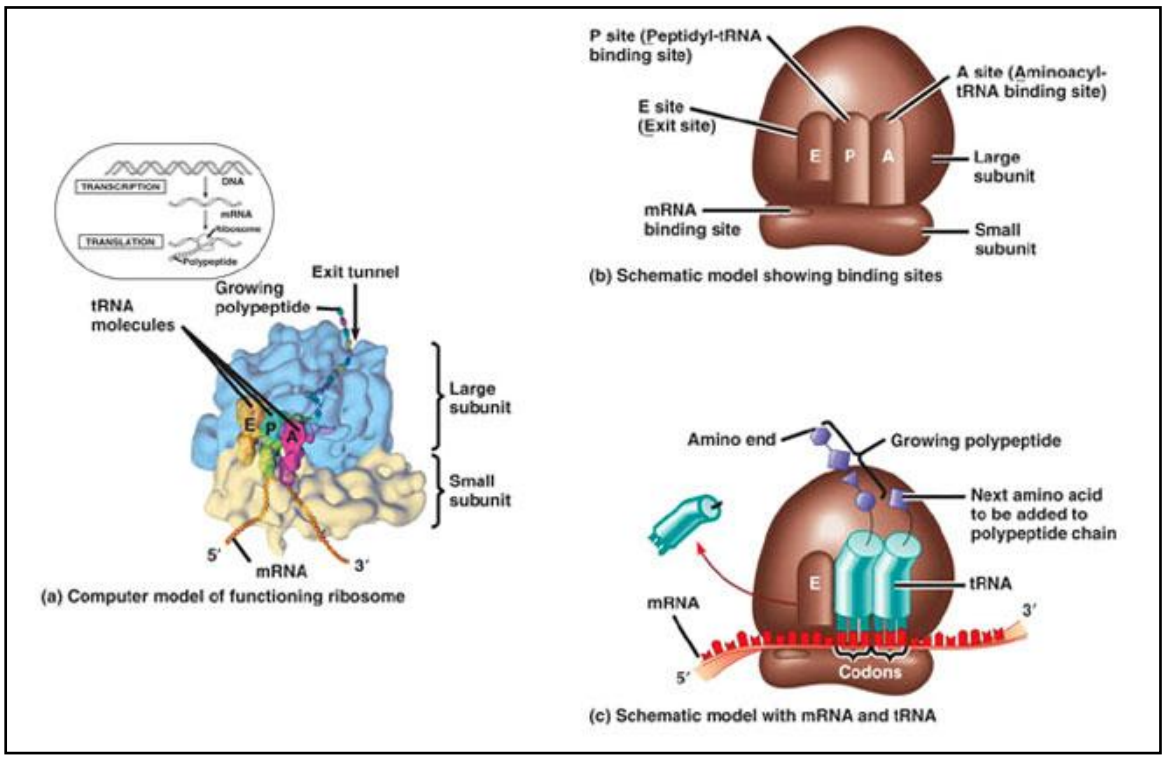


Figure 6 (a-c). Schematic models of the ribosome [6]

2.3. Proteins

Proteins are macromolecules, linear chains of amino acids. Proteins are polymers of 20 different amino acids. In a polypeptide chain the amino acids are joined together by peptide bonds between the carboxyl and amino groups of the adjacent amino acid. The sequence of amino acids in a protein is determined by a gene. A typical protein contains 200-300 amino acids but some are much smaller, and others much larger.

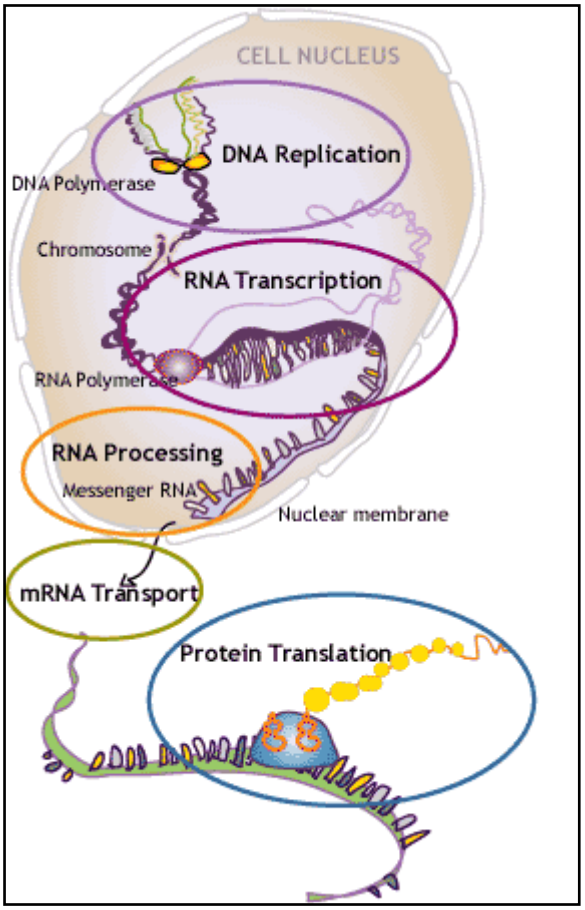


Figure 7. The route from DNA to protein [7]

2.4. The Route from DNA to Protein

DNA is arranged into chromosomes, and is located either in the nucleus in the eukaryotic cells or in the cytoplasm in prokaryotic cells. Every cell has genetic information which is duplicated before a cell divides. This process is called

replication (Figure 7). The building-blocks of the cell are proteins. When they are needed for the cell, the corresponding genes are activated, and transcribed into RNA during transcription. In eukaryotic cells non-coding parts of the mRNA are removed (processed) and is then transported out of the nucleus by an mRNA transport. Bacterial mRNA does not undergo RNA processing. In the following steps, the proteins are built based upon the code in the mRNA. This step is called translation and it takes place in the cytoplasm of the cell (Figure 7). The following section details the above mentioned processes.

2.4.1. Transcription

Transcription is the process when one strand of a DNA molecule is used as a template for synthesis of a complementary RNA by the enzyme RNA polymerase. Transcription is followed by translation when the nucleotide sequence of an mRNA is used to order and join amino acids in a protein. mRNA carries the genetic information translated from the DNA in the form of a three-base code. Each three-base code stands for a particular amino acid. Genetic code is used by cells during translation. The genetic code is a triplet code, which means that every three nucleotides are recognized from a starting point in the mRNA. The starting point is called a “start” codon.

The genetic code (Table 1) was first translated for *E.coli* [8]. The genetic code is universal, i.e. the codes are applicable for cells of all species.

The genetic code contains 64 codons, 61 define individual amino acids, and three are stop codons. Synthesis of all proteins begins with an amino acid called methionine present in the cells. The sequence of codons between a start codon to a terminating stop codon is called a reading frame. The genetic code is an overlapping triplet code so that one mRNA theoretically might be translated in three different reading frames. After the code is transcribed the mRNA is further processed; special head and tail regions are added and some parts are spliced out. mRNA leaves the nucleus in eukaryotes and carries the code into the cytoplasm, then associates with a ribosome. mRNA processing is simplified in prokaryotic cells, where there is no splicing.

Table 1. The genetic code

Codon	tRNA anticodon	Amino acid	Codon	tRNA anticodon	Amino acid	Codon	tRNA anticodon	Amino acid	Codon	tRNA anticodon	Amino acid
UUU		Phe	CUU	GAAgu	Leu	AUU	UAAgu	Ile	GUU	CAAgu	Val
UUC	AAGa	Phe	CUC	GAGa	Leu	AUC	UAGau	Ile	GUC	CAGa	Val
UUA	AAUc	Leu	CUA	GAUc	Leu	AUA	UAU	Ile	GUA	CAUc	Val
UUG	AAC	Leu	CUG	GAC	Leu	AUG	UAC	Met	GUG	CAC	Val
UCU	AGAgu	Ser	CCU	GGAgu	Pro	ACU	UGAgu	Thr	GCU	CGAgu	Ala
UCC	AGGa	Ser	CCC	GGGa	Pro	ACC	UGGa	Thr	GCC	CGGa	Ala
UCA	AGUc	Ser	CCA	GGUc	Pro	ACA	UGUc	Thr	GCA	CGUc	Ala
UCG	AGC	Ser	CCG	GGC	Pro	ACG	UGC	Thr	GCG	CGC	Ala
UAU		Tyr	CAU		His	AAU		Asn	GAU		Asp
UAC	AUGa	Tyr	CAC	GUGa	His	AAC	UUGa	Asn	GAC	CUGa	Asp
UAA		Stop	CAA	GUUc	Gln	AAA	UUUc	Lys	GAA	CUUc	Glu
UAG		Stop	CAG	GUC	Gln	AAG	UUC	Lys	GAG	CUC	Glu
UGU		Cys	CGU	GCAgu	Arg	AGU		Ser	GGU	CCAgu	Gly
UGC	AGCa	Cys	CGC	GCGa	Arg	AGC	UCGa	Ser	GGC	CCGa	Gly
UGA		Stop	CGA	GCUc	Arg	AGA	UCUc	Arg	GGA	CCUc	Gly
UGG	ACC	Trp	CGG	GCC	Arg	AGG	UCC	Arg	GGG	CCC	Gly

Data was compiled from different sources, based on [8].

2.4.2. Translation

Translation is the second stage of protein synthesis, also referred to as decoding. Translation takes place in the cytoplasm where the ribosomes are located. During translation, mRNA is decoded to produce a specific polypeptide chain according to the genetic code. Translation has four phases: activation, initiation, elongation, and termination.

In activation, the correct amino acid is attached to the correct tRNA by covalent bonding. An ester bond exists between the carboxyl group of the amino acid and the 3` OH of the tRNA (Figure 6). This is a two step ATP-requiring reaction catalyzed by a specific aminoacyl-tRNA synthase as an enzyme. Each of the 20 different synthetases recognizes one amino acid. At the end, the amino acid is linked

to the tRNA by a high-energy bond. Now the tRNA is called charged or activated. In initiation, the initiation factors bind to the small subunit of the ribosome. This preinitiation complex, together with Met-tRNA then binds to the mRNA. The assembly is completed after the large ribosomal subunit, with the tRNA-binding sites (A, P, and E sites), has joined the preinitiation complex. Elongation starts on the initiation complex with Met-tRNA in the P-site. Then the correct aminoacyl-tRNA decodes the second codon and enters the ribosome A site with the aid of elongation factors and guanosine 5`-triphosphate (GTP). Termination of the polypeptide happens when the A site of the ribosome faces a stop codon. At this point a releasing factor recognizes "nonsense" codons and releases the polypeptide chain.

According to the above mentioned aminoacyl-tRNA synthetase, tRNA and ribosomes play a major role in the translation of mRNA accompanied with the aid of enzymes. Decoding is a key element during protein synthesis that enables information transfer from RNA to protein. This step is critical for the survival of all organisms.

2.4.2.1. Transfer RNA During Translation

Selection of cognate aa-tRNAs is carried out by the translation machinery with high accuracy and speed. It involves kinetic proofreading and induced-fit mechanisms. During decoding, the ribosome must choose the tRNA whose anticodon corresponds to the codon in the mRNA so as to incorporate the correct amino acid into the growing polypeptide chain. Simulations have demonstrated that the flexibility of the acceptor stem of the tRNA is essential for correct tRNA selection. The recognition of the tRNA as an active player in translation will change the traditional view of tRNA as only an adaptor, and would determine more specific tRNA roles. The contribution of tRNA to decoding was studied for more than four decades [9], but the movement of tRNAs in the cytoplasm is poorly explained up to this day. How the tRNAs approach the ribosome remains a question.

2.5. *Escherichia coli* (*E. coli*)

E. coli is frequently studied in micro- and molecular biology. It is a prokaryotic organism. Its structure is clear, and it makes an excellent target for investigation and experimentations. *E. coli* plays an important role in modern biological engineering because of its long history in experimentation and laboratory work. Bacteria can also be grown easily and its genetic components are comparatively simple and easy to manipulate, making it one of the best studied prokaryotic model organisms in biotechnology.

The sequence of *E.coli* reveals 4441 open reading frames corresponding to 4322 proteins and 122 rRNAs and tRNAs. The entire sequence of the *E. coli* genome has already been identified. Data and statistics are available for the calculations conducted in this research. The model organism *E. coli* is 2-4 µm long with a diameter of 500-800 nm. The cytoplasm of the whole cell thus contains about 40 million molecules in addition to water molecules or, if we omit ions and small organic molecules, about half a million macromolecules [10].

Data collected on *E.coli* was derived or measured using different sources [10], [18]. The data were needed to provide sufficient and necessary details to help in cellular simulation. The complete set of data is shown in Appendix D (below).

Table 2. Quantities of molecule types in *E. coli*. [10]

Quantity	Molecule Types
225,000	Proteins
15,000	Ribosomes
170,000	tRNA-Molecules
15,000,000	Small organic molecules
25,000,000	Ions
70%	Water

2.5.1. The Bacterial Cytoplasm

Bacteria are prokaryotic organisms. Their nucleus is not separated from the cytoplasm by a cell membrane. The conditions inside a bacterial cytoplasm are often assumed to be very close to those in a test tube. The cube of size 100 nm of the *E. coli*

cytoplasm contains about 450 proteins, 30 ribosomes, 340 tRNA molecules, other RNA molecules, 30000 small organic molecules, 50000 ions and about 70% of the volume is water. The cytoplasm of the whole cell thus contains about 40 million molecules besides water or, if we omit ions and small organic molecules, about half a million large organic molecules [10].

Little is known about the movement of macromolecules in bacterial cells. In early experiments, it has been shown that the motion of macromolecules was consistent with simple diffusion on a time scale of < 1 sec and bacterium length of ~ 1 µm [11]. Deich et al. [12] investigated individual fluorescently labeled proteins in the membrane of *C. crescentus* and characterized their two-dimensional motion as diffusive. Golding et al. have found that cytoplasmic motion is subdiffusive on a time scale of seconds to minutes [13].

2.6. Simulation of Biochemical Cellular Processes

Many approaches exist for the simulation of biochemical cellular processes using deterministic and stochastic modeling approaches. Three types of cell models are generally discussed: macroscopic, mesoscopic and microscopic models [10].

2.6.1. Macroscopic Models

Macroscopic models deal with molecular concentrations determined by stochastic differential equations. Forces between or inside the molecules are ignored. Usually an infinite reaction volume is assumed.

2.6.2. Mesoscopic Models

Mesoscopic models deal with individual molecular dynamics of biochemical reactions. Generally physical forces between or inside the molecules are not considered.

2.6.3. Microscopic Models

Microscopic models are the only models that deal with physical forces within or between molecules. This type of modeling is fine-graded and are not suitable for whole cell simulation because of computational restrictions, and the nature of the interactions considered. A microscopic model was used for describing molecular folding - e.g. secondary structure of proteins [14], [15].

Let summarize what are the known facts and what is not know about the tRNA, the subject of this thesis.

The functions of the tRNA and how they are carried out is known, also well known and documents is how tRNAs are assembled and how they get charged by amino acids. Information could not be found, however, on the existence of any signaling between the tRNAs and the ribosomes. Possible forms of communication could be mechanical vibrations through the cytoplasm as a liquid, or by electromagnetic waves. If tRNAs are able to receive any signal emitted by the ribosome, they should have some special receptors for this purpose. tRNAs tertiary structures contain four loops (D, TΨC, anticodon and variable loops) and their shape and diameter are specific to each type of tRNA. They could have different resonant frequencies if they are really suitable to serve as receptors and not only for their believed purpose. During activation, the tRNA type is recognized by the aminoacyl-tRNA synthetase enzyme according to the geometry of these loops [5].

Where exactly in the bacterial cytoplasm is the tRNAs getting charged is still a question. Either it is located in a specific place or it happens generally throughout the bacteria's cytoplasm.

The relatively small quantity of tRNAs compared to the number of ribosomes per bacterial cells is another important fact. According to statistics there are approximately ten times more tRNAs present in *E.coli* than in the ribosomes [18]. If the quantity of each amino acid specific tRNA is about 2%, then there is only one tRNA molecule for five ribosomes. It means that the tRNAs are quite busy and well utilized in providing amino acid molecules for the ribosomes.

After collecting sufficient details about our subject it is ripe to start to find suitable methods and means of modeling and consequently to begin conducting simulations.

CHAPTER 3.

SUMMARY OF APPLIED RESEARCH METHOLODOGY

Large molecules have physical properties besides chemical. Most of the research methods present in literature focus exclusively on the chemical properties of these giant molecules. Sometimes engineering methods are more fruitful or efficient in describing or explaining any reasoning behind an observed phenomenon.

Most cell simulations concentrate on biochemical interactions, but in this thesis we extend our research into the physical domain. The model sought does not fit into any of the above mentioned categories. Instead of two dimensions, a 3 dimensional model can be used even if it is not a whole cell model. The planned model involved the spherical partition of the cytoplasm. The mRNA is depicted on the 'z' axis, and the proportionally sized ribosome should be at the origin of the coordinate system. Since several ribosomes act on the same mRNA molecule, the boundary of our proposed model is placed according to the average space between ribosomes. (In eukaryotic cells ribosomes do not move freely, so their distribution could not be assumed even, but the subject of our investigation is a type of bacteria, and until now it is widely believed that their cytoplasm is unstructured.) The volume of interest and the number of considered particles is assumed to be proportional to this 'L' distance (length of the cylinder = 2L, the radius = L). Particles are considered to be following the rules of Brownian Motion. From other researcher's experiments, the speed of differently sized particles can be approximated. The statistical average time interval (i.e. the delivery time of a particular amino acid to the ribosome) when the same tRNA interacts with the same ribosome, is included [9], [16], [17].

Most of the data for calculations were taken from Project CyberCell *E.coli* Statistics, Appendix D [18]. The average rate of amino acid assembly is around ~20 ms, so if our simulation results would fit into that time frame, it is possible to ascertain that the selection process is purely statistical or random. If our simulations show otherwise, then some other explanation about how these selections happen in reality must be found [11].

3.1. Software Employed in this Research

3.1.1. MATLAB

MATLAB® is a high level computational language, the software was developed by MathWorks for numerical computation, matrix manipulation, data analysis and visualization. Its main incentive is to provide a faster and easier method for programming when compared to other traditional programming languages.

By the help of optional toolboxes, it is suitable for modeling, simulation, control design, image and signal processing for many applications. Its ability for visualizing 3D graphics proved to be very useful in this thesis.

3.1.2. Paint.net

Paint.NET is an image and photo editing software which was used in this thesis to create explanatory figures. Different layers could be overlaid, which rendered the program useful for producing animation like graphics.

CHAPTER 4.

MODELING

4.1. Particle Movement in a Force Field

Preparation has been undertaken to study particle movement in a force field, and MATLAB code has been prepared to calculate this. An initial idea was that there should be a tRNA specific force exerted by the ribosome. Even after a thorough hunt for a clue in literature, no hint of any force could be found. Only van der Waals interactions exist in this context and they are weak forces that exist between the molecules, acting in the close proximity of the ribosome.

4.2. Particle Collision

In the proposed model, molecular particles move in 3 dimensional space and collide with each other. The most important part of the simulation program is to detect which particle collides with which object, when and where, as well as be able to calculate their velocity and direction of movement after the collision.

There are different types of collisions identified in physics. First we need to decide whether collisions between atomic particles are elastic or inelastic. Since molecules could be considered as rigid bodies and their collision energy does not cause any chemical interaction, collisions were considered fully elastic, in this thesis.

Available literature in this regards provides descriptions for 1D and 2D cases [19].

4.3. Elastic Collision of Particles

In elastic collisions, kinetic energy and momentum are conserved, i.e. there is no energy loss in the form of heat, etc.
Kinetic energy (KE) of a moving object is:

$$\mathbf{KE} = \frac{1}{2}\mathbf{mv^2} \quad \textbf{(1)}$$

And its momentum (P) is:

$$\mathbf{P} = \mathbf{mv} \quad \textbf{(2)}$$

Where m is the mass and v is the velocity. Velocity and momentum are vectorial quantities.

If we have two objects colliding centrally, then

$$\sum_i \mathrm{KE}_i = \sum_i \mathrm{KE}_i' \quad \textbf{(3)}$$

(Apostrophe indicates the value after collision)

$$\sum_i p_i = \sum_i p_i' \quad \textbf{(4)}$$

Conservation of kinetic energy gives the following equation:

$$\frac{1}{2}m_1v_1^2+\frac{1}{2}m_2v_2^2=\frac{1}{2}m_1v_1'^2+\frac{1}{2}m_2v_2'^2 \quad \textbf{(5)}$$

and the conservation of momentum provides the equation below

$$m_1v_1+m_2v_2=m_1v_1'+m_2v_2' \quad \textbf{(6)}$$

4.3.1. Elastic Collision in 1 Dimension (1D)

Object1 and object2 are moving at v_1 and v_2 velocities, shown in Figure 8.

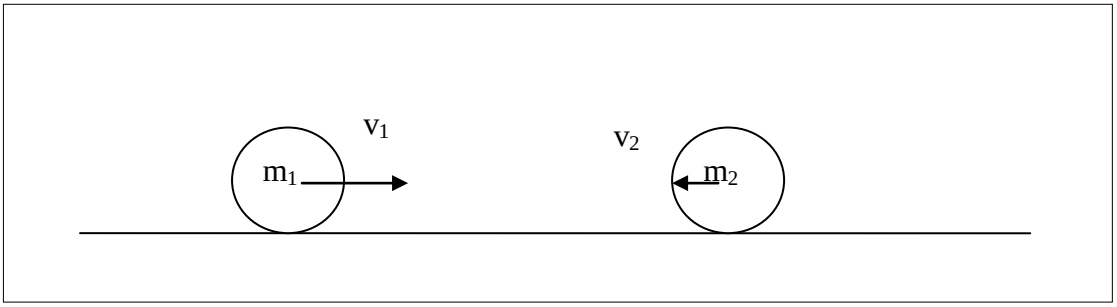


Figure 8. Before 1D collision

From eq. 5 and 6 we are able to calculate the objects velocities after collision:

$$v_1' = \frac{(m_1 - m_2)v_1 + 2m_2v_2}{m_1 + m_2} \quad \textbf{(7)}$$

and

$$v_2' = \frac{(m_2 - m_1)v_2 + 2m_1v_1}{m_1 + m_2} \quad \textbf{(8)}$$

respectively [19].

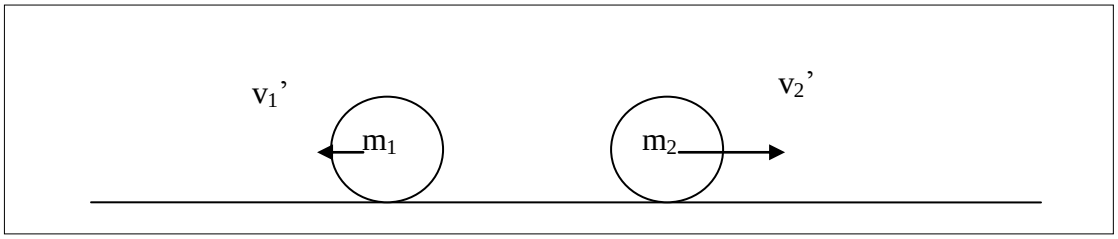


Figure 9. After 1D collision

4.3.2. Collision in 2 Dimensions (2D)

The unit normal vector of collision is:

$$\overrightarrow{u_n} = \frac{\vec{n}}{|\vec{n}|} \quad \textbf{(9)}$$

where: $\vec{n} = (x_2 - x_1, y_2 - y_1)$

and the unit tangential vector is:

$$\overrightarrow{u_t} = (-u_{ny}, u_{nx}) \quad \textbf{(10)}$$

(x_1,y_1) and (x_2,y_2) are the coordinates of the objects centers at the moment of collision. u_n and u_t are shown in Figure 12.

Before collision:

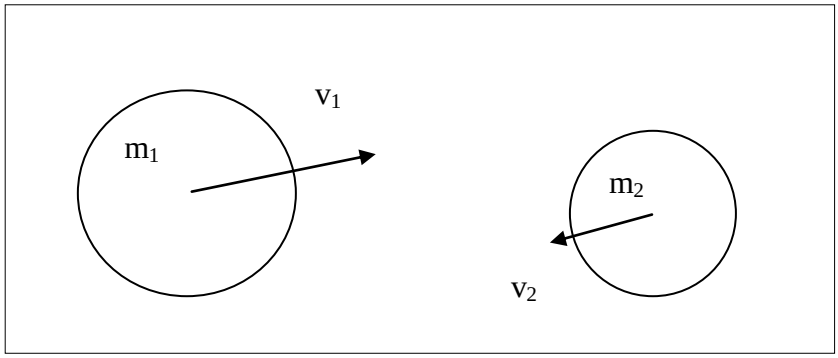


Figure 10. Before 2 D collision

Collision:

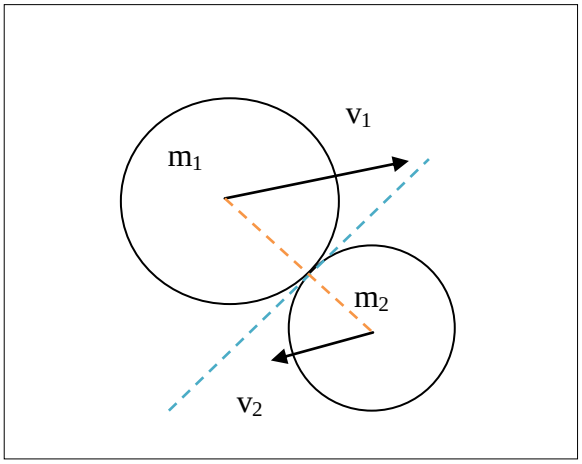


Figure 11. Collision (2 D)

After the collision, tangential velocity components are unchanged.

v_{1n} is the scalar velocity in the normal direction,

v_{1t} is the tangential component.

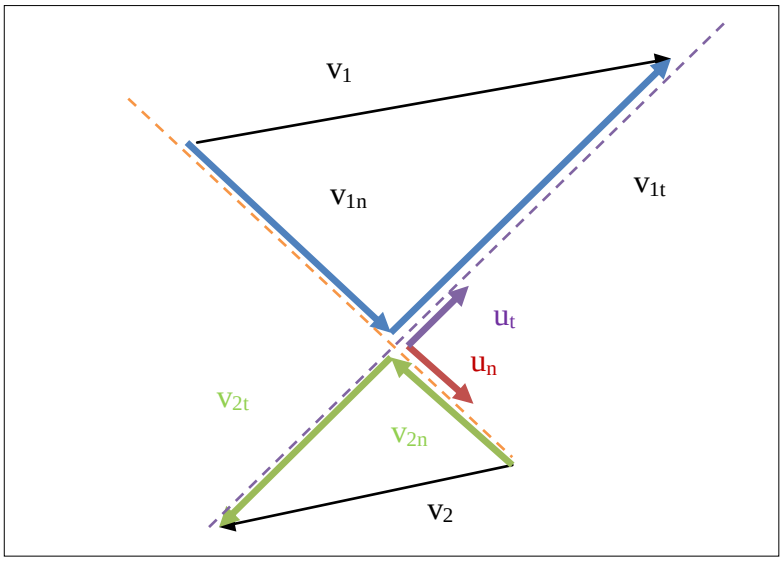


Figure 12. Velocity vector components before collision

Using dot products we can calculate the velocity components:

$$\mathbf{v}_{1n} = \vec{\mathbf{u}}_n \bullet \vec{\mathbf{v}}_1 \quad \textbf{(11)}$$

$$\mathbf{v}_{1t} = \vec{\mathbf{u}}_t \bullet \vec{\mathbf{v}}_1 \quad \textbf{(12)}$$

and

$$\overrightarrow{v_{2n}} = \overrightarrow{u_n} \bullet \overrightarrow{v_2} \quad \textbf{(13)}$$

$$\overrightarrow{v_{2t}} = \overrightarrow{u_t} \bullet \overrightarrow{v_2} \quad \textbf{(14)}$$

Because tangential components do not change during collision, they will be the same after the collision. Thus

$v_{1t}'=v_{1t}$ and $v_{2t}'=v_{2t}$

To get the normal components of the velocity vector after the collision, a formula similar to the 1D case could be used:

$$v_{1n}' = \frac{(m_1 - m_2)v_{1n} + 2m_2 v_{2n}}{m_1 + m_2} \quad \textbf{(15)}$$

and

$$v_{2n}' = \frac{(m_2 - m_1)v_{2n} + 2m_1 v_{1n}}{m_1 + m_2} \quad \textbf{(16)}$$

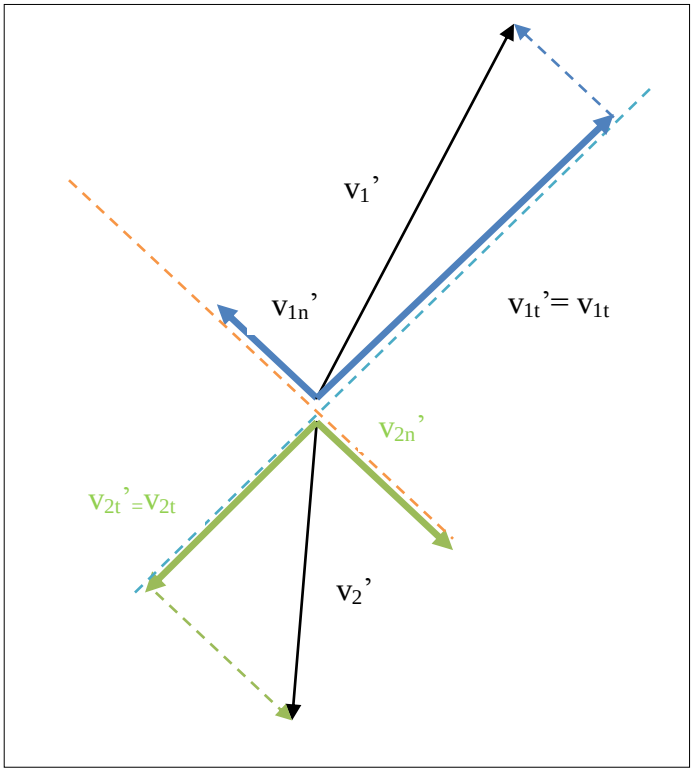


Figure 13. Velocity vector components after collision

To get back the vectorial values of velocity after collision, normal vectors should be multiplied by the corresponding scalar velocity components:

$$\overrightarrow{v_{1n}'} = v_{1n}' \cdot \overrightarrow{u_n} \quad \textbf{(17)}$$

$$\overrightarrow{v_{1t}'} = v_{1t} \cdot \vec{u}_t \quad \textbf{(18)}$$

and

$$\overrightarrow{v_{2n}'} = v_{2n}' \cdot \vec{u}_n \quad \textbf{(19)}$$

$$\overrightarrow{v_{2t}'} = v_{2t}' \cdot \vec{u}_t \quad \textbf{(20)}$$

Finally to get each object's velocity after collision, normal and tangential vector components are added together:

$$\overrightarrow{v_1'} = \overrightarrow{v_{1n}'} + \overrightarrow{v_{1t}'} \quad \textbf{(21)}$$

$$\overrightarrow{v_2'} = \overrightarrow{v_{2n}'} + \overrightarrow{v_{2t}'} \quad \textbf{(22)}$$

4.3.3. Collision in 3 Dimensions (3D)

Instead of circles as is the case with 2D, particles are approximated as spheres. The 3 Dimensional case is even more complicated. In the literature there is no complete calculation method for 3D collisions, the only one that could be found was reported by Thomas Smid for a special case, when one of the particles is considered stationary [20]. Appendix C contains a sample of his calculations.

- His study was formulated in a coordinate system where the initial velocity of sphere 2 is zero
- The relative position vector coincides with the z-axis
- Sphere 1 is at the origin moving with the same relative velocity as both balls
- Sphere 2 rests on the z-axis, at a point corresponding to the distance between the two balls
- Two successive rotations about the y and z axis of the coordinate system were used

Applying the above equations and reversing the coordinate transformations gives the velocities after the collision in the original coordinate system. The final answer comes after coordinate transformation of the two objects system. Thus it is better to develop an alternative method for calculating 3D collision behavior.

Similarly to the 2D case the normal unit vector of the collision was found, but instead of having one tangential component, there are 2 orthogonal components due to the fact that we have a plane rather than a line of collision.

To find the normal vector of the plane of collision

$$\vec{u_n} = \frac{\vec{n}}{|\vec{n}|} \tag{23}$$

where: $\vec{n} = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$ **(24)**

An example of a pair of tangential unit vectors is given below:

$$\vec{t_1} = (u_{ny} - u_{nz}, u_{nx} - u_{nz}, u_{nx} - u_{ny}) \tag{25}$$

$$\vec{t_2} = (u_{ny}(u_{nx} - u_{ny}) - u_{nz}(u_{nx} - u_{nz}), u_{nz}(u_{ny} - u_{nz}) - \\ - u_{nx}(u_{nx} - u_{ny}), u_{nx}(u_{nx} - u_{nz}) - u_{ny}(u_{ny} - u_{nz})) \tag{26}$$

u_{nx}, u_{ny}, u_{nz} are the x, y and z coordinate's of the collision normal vector, t_1 and t_2 are the tangential vectors. Their unit vectors with x, y and z components are given below:

$$\overrightarrow{u_{t1}} = \frac{\vec{t_1}}{|\vec{t_1}|} \tag{27}$$

$$\overrightarrow{u_{t2}} = \frac{\vec{t_2}}{|\vec{t_2}|} \tag{28}$$

Similarly to the 2D case equations (11) and (14) we can calculate velocity components using the dot product.

Normal component of v_1 velocity is:

$$v_{1n} = \overrightarrow{u_n} \bullet \overrightarrow{v_1} \tag{29}$$

and v_{1t1} v_{1t2} are its tangential components:

$$v_{1t1} = \overrightarrow{u_{t1}} \bullet \overrightarrow{v_1} \tag{30}$$

$$v_{1t2} = \overrightarrow{u_{t2}} \bullet \overrightarrow{v_1} \tag{31}$$

To obtain the second object's velocity components v_{2n}, v_{2t1} and v_{2t2}:

$$v_{2n} = \overrightarrow{u_n} \bullet \overrightarrow{v_2} \tag{32}$$

$$v_{2t1} = \vec{u}_{t1} \bullet \vec{v}_2 \quad \textbf{(33)}$$

$$v_{2t2} = \vec{u}_{t2} \bullet \vec{v}_2 \quad \textbf{(34)}$$

Similar to the 2D case the tangential components are the same after collision as they were before. The normal velocity components are obtained by the help of the same equations as in the 1D and 2D cases (equations 15 and 16).

$$v_{1n}' = \frac{(m_1 - m_2)v_{1n} + 2m_2 v_{2n}}{m_1 + m_2} \quad \textbf{(35)}$$

$$v_{2n}' = \frac{(m_2 - m_1)v_{2n} + 2m_1 v_{1n}}{m_1 + m_2} \quad \textbf{(36)}$$

So the velocity vectors after collision become:

$$\vec{v}_1' = \vec{v}_{1n}' + \vec{v}_{1t1} + \vec{v}_{1t2} \quad \textbf{(37)}$$

$$\vec{v}_2' = \vec{v}_{2n}' + \vec{v}_{2t1} + \vec{v}_{2t2} \quad \textbf{(38)}$$

More simplification can be made. Notice that $v_{1t1}'=v_{1t1}$, $v_{1t2}'=v_{1t2}$ and $v_{2t1}'=v_{2t1}$, $v_{2t2}'=v_{2t1}$, so

$$\vec{v}_1 = \vec{v}_{1n} + \vec{v}_{1t1} + \vec{v}_{1t2} \quad \textbf{(39)}$$

furthermore,

$$\vec{v}_{1t1} + \vec{v}_{1t2} = \vec{v}_1 - \vec{v}_{1n} \quad \textbf{(40)}$$

Equation (37) above becomes:

$$\vec{v}_1' = \vec{v}_{1n}' + \vec{v}_1 - \vec{v}_{1n} \quad \textbf{(41)}$$

similarly equation (38) above becomes:

$$\vec{v}_2' = \vec{v}_{2n}' + \vec{v}_2 - \vec{v}_{2n} \quad \textbf{(42)}$$

4.4. Collision Detection

Assume that we have two objects approaching each other. Object1 is moving with a velocity of v_1, object2 with a velocity of v_2. C_1 and C_2 are the objects' centers of mass. If object1 and object2 are to collide, the line segment connecting C_1 and C_2 should be parallel to itself until the objects collide.

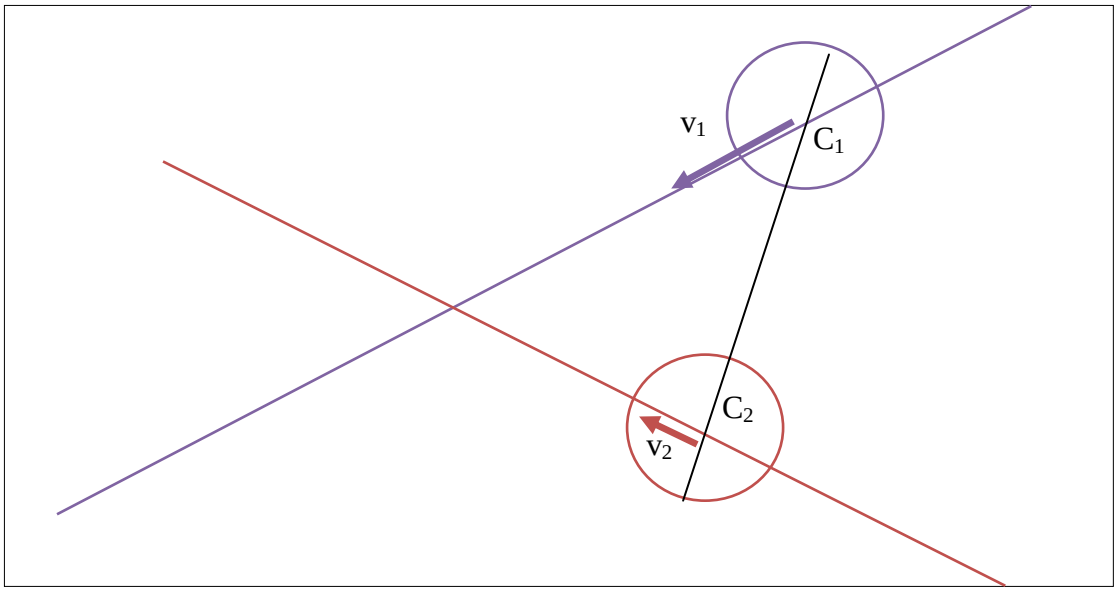


Figure 14. Two objects before collision

Because the model is based on discrete time intervals and not continuous ones, the exact moment of the collision cannot be caught.

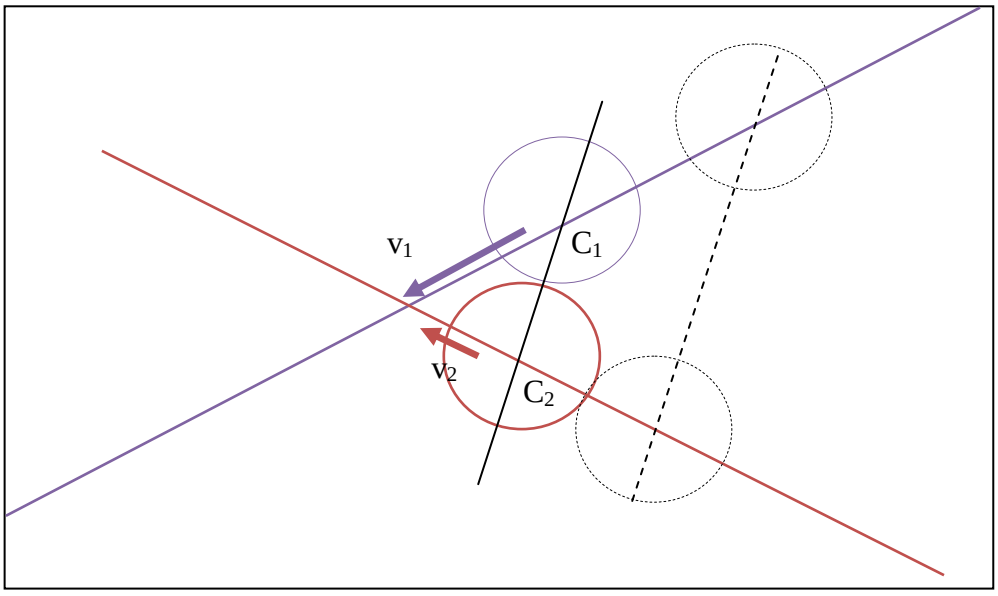


Figure 15. Two objects at the moment of collision

Consider that any two particles have collided when the distance between C_1 and C_2 is smaller than the sum of the two radii. The delta_t sampling time interval should be small enough to be able to detect all of the collisions. Since the line segment between C_1 and C_2 is always heading in the same direction, i.e. parallel to its previous state, it is not necessary to determine the exact moment of collision to get an

approximation of the normal vector of the collision. The direction shown by the C_1 and C_2 line segment is

$$\vec{n} = (x_2 - x_1, y_2 - y_1, z_2 - z_1) \quad \textbf{(43)}$$

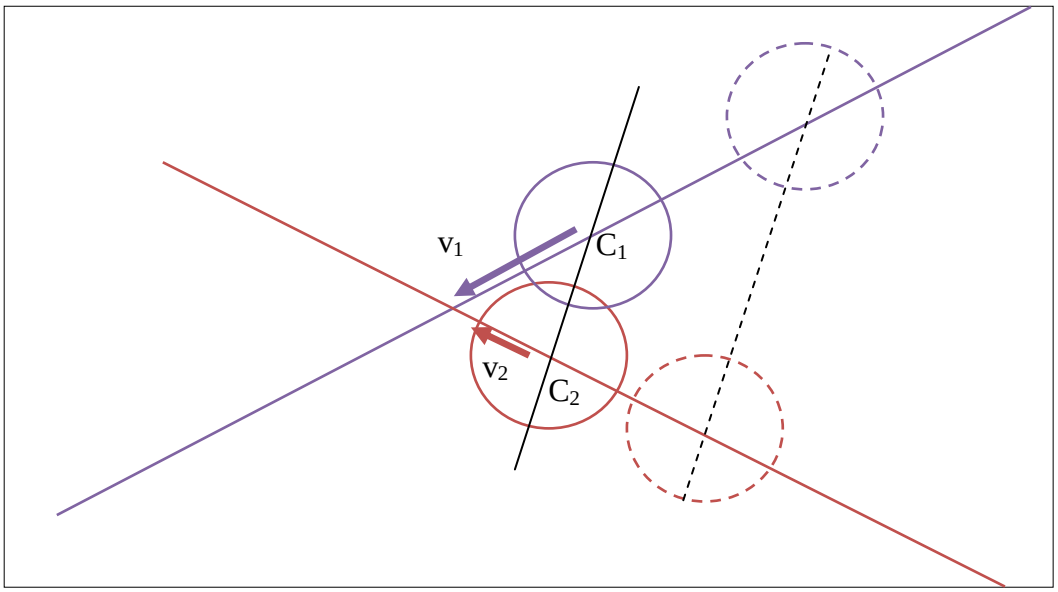


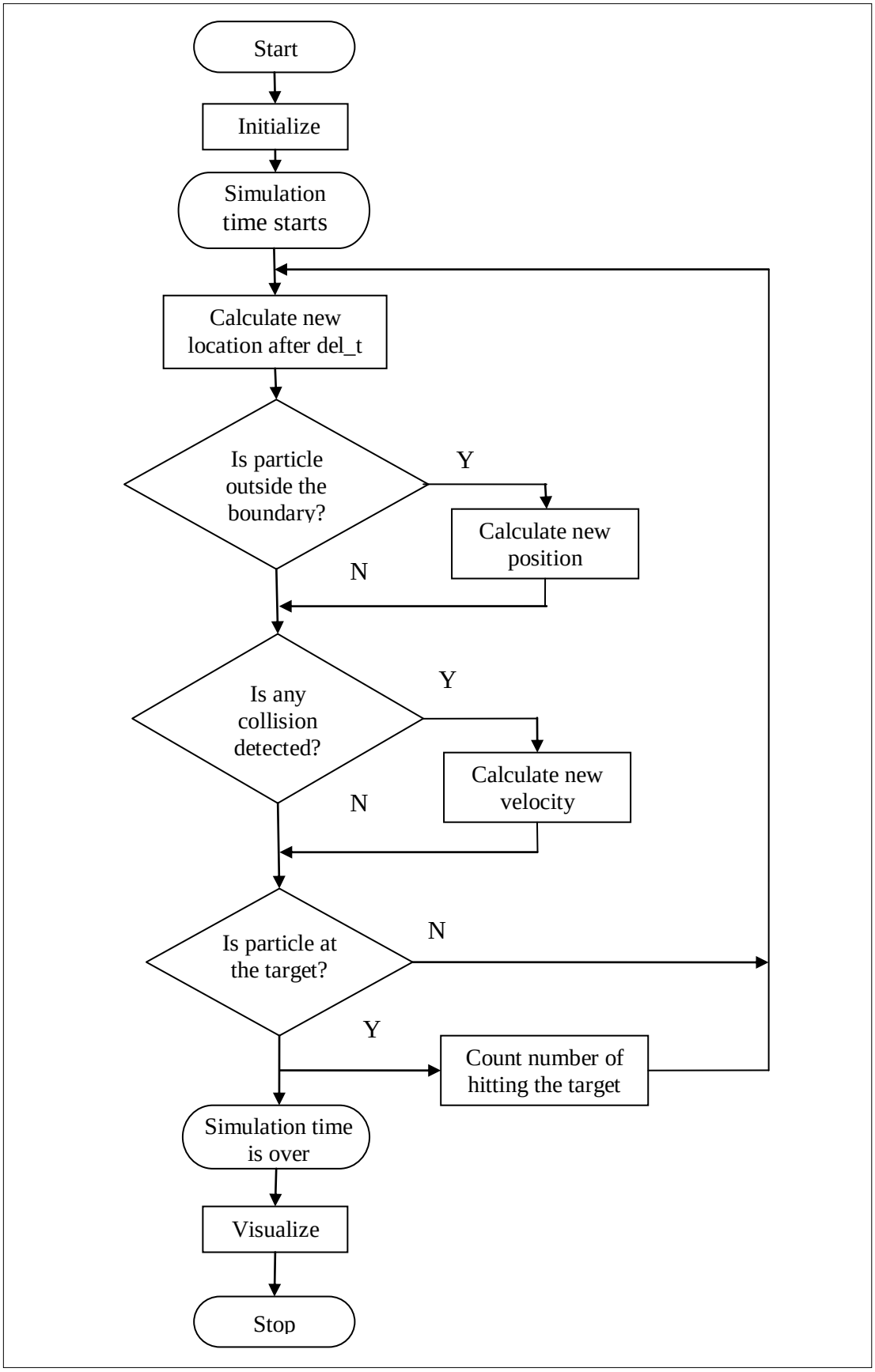
Figure 16. Two objects in overlapping

4.5. Flowchart

The MATLAB software was used for computation. A flowchart was developed (Figure 17) to organize the different subroutines. Computational input data, representing the initial conditions, was based on biological measurements and they are summarized in Table 3. During initialization we create a simulation space, place particles inside that space, randomly assign each a mass, radius, initial velocity and location. Placing objects one by one into the simulation space and checking for any overlapping guarantees that no object is contained inside another. Ribosome's velocities are assumed to be zero, to simulate them as fixed objects in the cytoplasm.

The simulation calculates new location for each particle every delta_t time interval and checks whether any collision occurs or any particle is located outside the simulation space. If there is a collision detected, then new velocities are calculated to both particles involved. When any particle crosses the boundary, it would change direction.

Figure 17. Flowchart



After each step the program checks whether any cognate tRNA reached the target, the ribosome, located in the centre of the simulation space. Figure 17 shows the flowchart used to code the MATLAB program.

Table 3. *E.coli* cell dimensions [18]

Cell length	2 μm or 2x10-6 m
Cell diameter	0.8 μm or 0.8x10-6 m
Cell total volume	1x10-15 L or 1x10-18 m3 (other est. at 0.88x10-15 L)
Average size of protein	360 residues
Average diameter of ave. protein	5 nm
Average MW of protein	40 kD
Average size of mRNA	1100 bases
Average length of mRNA	370 nm
Mean Velocity of 70 kD protein (cytoplasm)	3 nm/ms = 3x10-6 m/s
Mean Velocity of 40 kD protein (cytoplasm)	5 nm/ms = 5x10-6 m/s
Mean Velocity of 30 kD protein (cytoplasm)	7 nm/ms = 7x10-6 m/s
Mean Velocity of 14 kD protein (cytoplasm)	10 nm/ms = 10x10-6 m/s
Mean Velocity of small molecules (cytoplasm)	50 nm/ms = 5x10-5 m/s
Volume occupied by water	70%
Volume occupied by protein	17%
Volume occupied by all RNA	6%
Volume occupied by rRNA	5%
Volume occupied by tRNA	0.8%
Volume occupied by mRNA	0.2%
Volume occupied by DNA	1%
Volume occupied by ribosomes	8%
Translation rate	40 aa/sec
Number of mRNA/cell	4000
Number of tRNA/cell	200,000
Number of ribosomes/cell	18,000
MW of ribosome	2700 kD
Diameter of ribosome	20 nm
Volume of ribosome	4.2 x 10-24 m3

In Chapter 3 literature suggested to use the spherical coordinate system in modeling this biological system, but since we were able to propose a more suitable calculation method, by the help of vector geometry with Cartesian coordinates, we went ahead and used the new mathematical technique.

To calculate the velocity of particles after collision, the following equations from the previous section were used in the MATLAB codes:

Collision normal vector: $$\vec{n} = (x_2 - x_1, y_2 - y_1, z_2 - z_1) \quad \textbf{(44)}$$

Normal unit vector: $$\overrightarrow{u_n} = \frac{\vec{n}}{|\vec{n}|} \quad \textbf{(45)}$$

$Object_1$ velocity normal component before collision:

$$v_{1n} = \overrightarrow{u_n} \bullet \overrightarrow{v_1} \quad \textbf{(46)}$$

$Object_2$ velocity normal component before collision:

$$v_{2n} = \overrightarrow{u_n} \bullet \overrightarrow{v_2} \quad \textbf{(47)}$$

Velocity values after the collision are:

$Object_1$ normal component: $$v_{1n}' = \frac{(m_1 - m_2)v_{1n} + 2m_2 v_{2n}}{m_1 + m_2} \quad \textbf{(48)}$$

$$\overrightarrow{v_{1n}'} = v_{1n}' \cdot \overrightarrow{u_n} \quad \textbf{(49)}$$

$Object_2$ normal component: $$v_{2n}' = \frac{(m_2 - m_1)v_{2n} + 2m_1 v_{1n}}{m_1 + m_2} \quad \textbf{(50)}$$

$$\overrightarrow{v_{2n}'} = v_{2n}' \cdot \overrightarrow{u_n} \quad \textbf{(51)}$$

$Object_1$ velocity: $$\overrightarrow{v_1'} = \overrightarrow{v_{1n}'} + \overrightarrow{v_1} - \overrightarrow{v_{1n}} \quad \textbf{(52)}$$

$Object_2$ velocity: $$\overrightarrow{v_2'} = \overrightarrow{v_{2n}'} + \overrightarrow{v_2} - \overrightarrow{v_{2n}} \quad \textbf{(53)}$$

See Appendix A. for MATLAB Codes.

CHAPTER 5.

SIMULATION

After developing the computer program to simulate the tRNA movement, simulations were run using different settings.

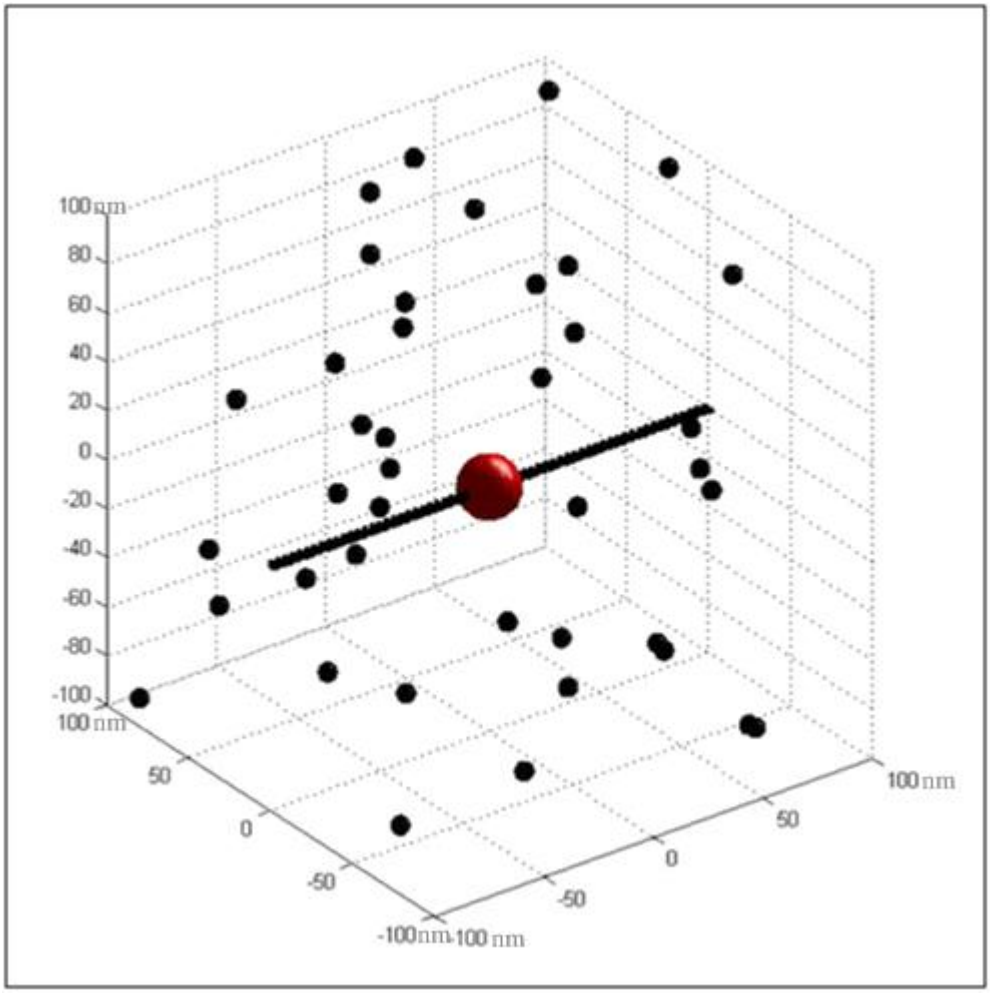


Figure 18. MATLAB simulation according chapter 5.1.1.

For visualization purposes an ‘avi’ movie file was created using the particles location matrix. Each frame comprises its 3D MATLAB graph taken every delta_t time interval. In this simulation (Figure 18). we can see the central ribosome with the mRNA and the ctRNAs moving in the simulation space. For simplicity objects other than ctRNA are not shown; but in the background, there are other objects involved in

the simulation, including other ribosomes (Figure 19) and other noncognate tRNAs (Figure 20).

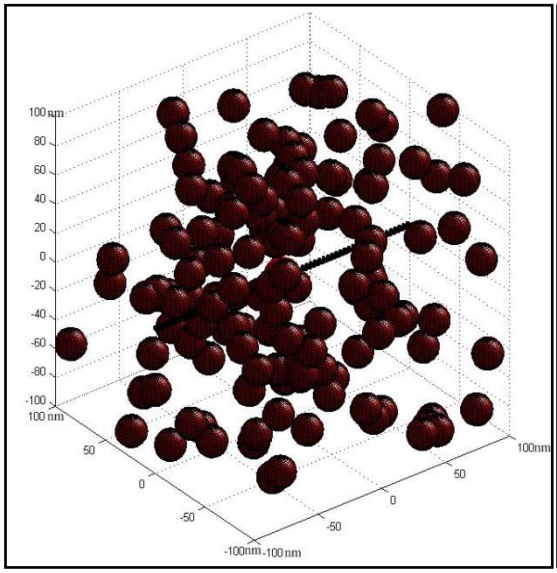


Figure 19. Messenger RNA and 144 ribosomes

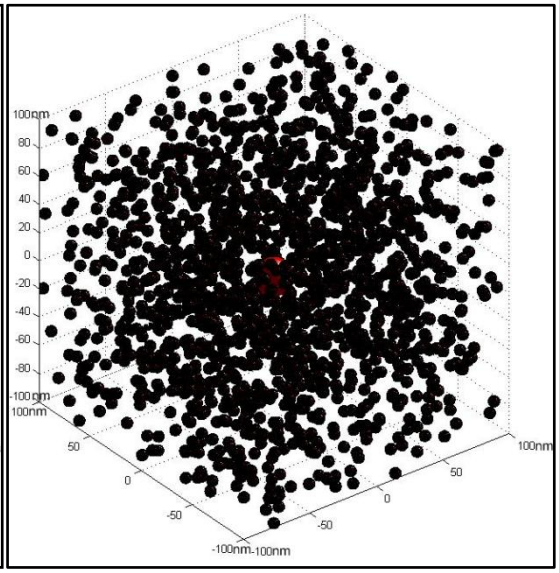


Figure 20. 1600 tRNAs

There are other particles also present, including different proteins and other macromolecules which are able to divert moving tRNAs in the cytoplasm. Figure 21 shows the simulation with 1,000 other objects present, it fills about 5% of the space. Figure 22 shows 10,000 other particles[1] addition to 144 ribosomes, and more than 1600 tRNAs. All these molecules occupy 32% of the space.

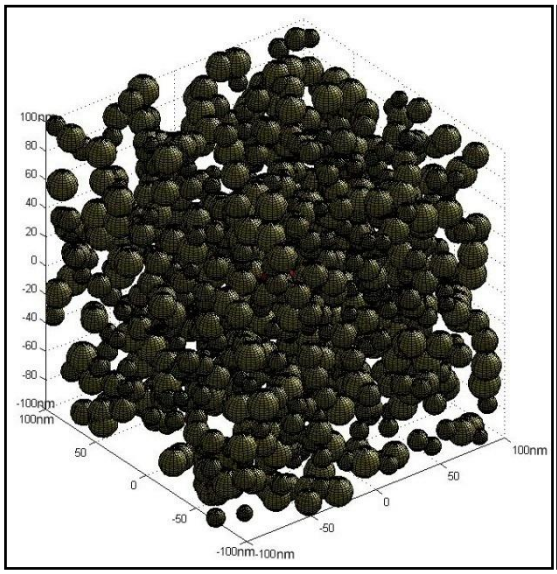


Figure 21. 1,000 other particles

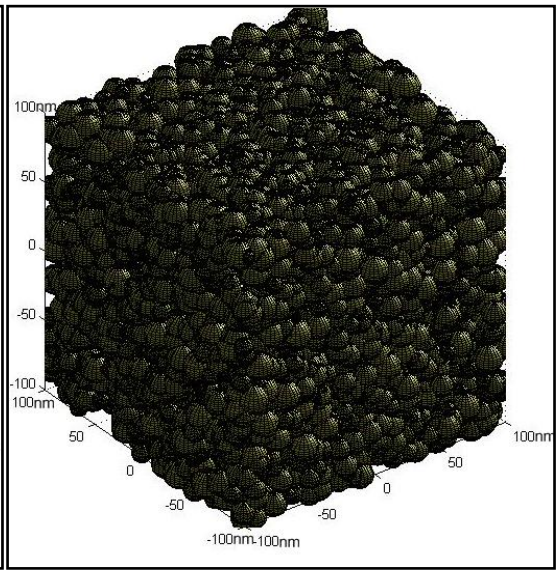


Figure 22. 10,000 other particles

[1] "other particles" in this document refers to other than tRNA, mRNA or ribosome

5.1. First Set of Simulations

The model volume is 200 nm x 200 nm x 200 nm cube. The particles are spaced randomly into this cube. The ribosome radius is set to 10 nm, tRNA radius to 3.5 nm, the size of other macromolecules that are capable of affecting the movement of tRNA under investigation are randomly selected between 5 and 10 nm in radius. The initial velocities of zero for ribosomes are assumed, because these structures could be considered stationary in the cytoplasm. The initial velocities of the tRNA and other particles are taken to be (10 nm/ms). The initial velocities are assumed to be omnidirectional.

To have a basis for comparison, initially 40 ctRNAs and the central ribosome (with the mRNA) are placed in the model space and 10 simulations are run.

Table 4. Test runs to validate the model

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
1	**1**	**0**	**2**	**1**	**0**	**0**	**0**	**0**	**0**	**0.5**

1. Normal cell

   The following particles are placed into the simulation space to simulate conditions in an average bacterial cytoplasm:

   Number of tRNA=1600

   Number of ctRNA=40

   Number of ribosomes=144

   Number of other particles=5 000

Length of the simulation time is set to 20 ms, to follow the normal amino acid assembling rate. The simulation time step is selected to be = 0.1 ms, to be able to detect all the collisions among these particles, according the average particle velocity and dimensions. Each set of conditions was simulated 10 times. Because of the enormous computational burden, some of the simulations lasts several hours. Details of the simulations are found in Appendix B.a.

5.1.1. Results

Table 5. Results of the first set of simulations

Simulation sets (Omnidirectional initial velocity)		Results (Average of 10 runs)
Serial number	Special conditions	
0	Validation test	**0.5**
1	Normal cell	**0.3**
2	Cell is dry[2]	**0.1**
3	Diluted cytoplasm[3]	**0.3**
4	Extremely dry cell[4]	**0.3**
5	Dry cell+ 5x longer time frame[5]	**0.9**
6	ctRNA density doubled[6]	**0.2**
7	ctRNA density tripled[7]	**0.4**
8	Increased metabolism (x2)[8]	**0.2**
9	ctRNAs velocity agitated (x10)[9]	**2**
10	Sampling rate increased (x2)[10]	**0.1**

*Result 1 means that any ctRNA reached the ribosome in the given timeframe, but it does not mean that it is inserted in tothe "A site" of the ribosome, since the surface area of the ribosome is 33 times larger than the "A site" itself.

[2] **Simulated by doubling the number of other particles.**

[3] **Number of other particles is reduced by 80%.**

[4] **Number of other particles is increased by 300%.**

[5] **Initial conditions are selected similar to set#2, but simulation time is increased by 400%, (simulating slower metabolism).**

[6] **Number of ctRNAs are increased, simulating tRNAs with concentration higher in the genome.**

[7] **Similar to set#6.**

[8] **The initial velocity of all moving objects in the simulaton space is doubled.**

[9] **Only ctRNAs initial velocity is increased 10 fold, to simulate a case when ctRNAs are able to move faster than other noncognate tRNAs.**

[10] **These simulations are meant to check whether our estimation for sampling rate are correct and no any collision is overlooked.**

5.2. Second Set of Simulations

Because of the unsatisfactorily low number of target hits in the first set of simulations, we attempted to provide some orderliness into particle movements. The bacteria, as a prokaryote, lack any internal structure, so its cytoplasm is considered homogeneous. To imitate what could happen if there is some organized fluid motion inside the bacteria, all the previous simulations were repeated. Instead of omnidirectional initial velocities, each particle's initial velocity had positive random components only. (Velocity vectors occupy only the first octant of the omnidirectional space.)

First, the initial simulation was repeated with the 40 ctRNAs and one ribosome only.

Table 6. Test runs to validate the model

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
0	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**2**	**1**	**0.4**

This result is similar to the reference test.

1. Normal cell

   The following particles are placed into the simulation space to simulate conditions in an average bacterial cytoplasm:

   Number of tRNA=1600

   Number of ctRNA=40

   Number of ribosomes=144

   Number of other particles=5 000

Length of the simulation time is set to 20 ms, to follow the normal amino acid assembling rate. The simulation time step is selected to be = 0.1 ms, to be able to detect all the collisions among these particles, according the average particle velocity and dimensions. Each set of conditions was simulated 10 times. Because of the enormous computational burden, some of the simulations lasts several hours. Details of the simulations are found in Appendix B.b.

5.2.1. Results

Table 7. Results of the second set of simulations

Simulation sets (Non-omnidirectional initial velocity)		Results (Average of 10 runs)
Serial number	Special conditions	
0	Validation test	**0.4**
1	Normal cell	**0.8**
2	Cell is dry[11]	**0.6**
3	Diluted cytoplasm[12]	**0.6**
4	Extremely dry cell[13]	**0.6**
5	Dry cell+ 5x longer time frame[14]	**1.6**
6	ctRNA density doubled[15]	**0.8**
7	ctRNA density tripled[16]	**1.2**
8	Increased metabolism (x2)[17]	**0.6**
9	ctRNAs velocity agitated (x10)[18]	**2.2**
10	Sampling rate increased (x2)[19]	**0.4**

[11] **Simulated by doubling the number of other particles.**

[12] **Number of other particles is reduced by 80%.**

[13] **Number of other particles is increased by 300%.**

[14] **Initial conditions are selected similar to set#2, but simulation time is increased by 400%, (simulating slower metabolism).**

[15] **Number of ctRNAs are increased, simulating tRNAs with concentration higher in the genome.**

[16] **Similar to set#6.**

[17] **The initial velocity of all moving objects in the simulaton space is doubled.**

[18] **Only ctRNAs initial velocity is increased 10 fold, to simulate a case when ctRNAs are able to move faster than other noncognate tRNAs.**

[19] **These simulations are meant to check whether our estimation for sampling rate are correct and no any collision is overlooked.**

5.3. Summary of the Simulation Results

Table 8. Summary of the simulations

Simulation sets		Average of 10 runs	
Serial number	Special conditions	First set of simulations	Second set of simulations
0	Validation test	**0.5**	**0.4**
1	Normal cell	**0.3**	**0.8**
2	Cell is dry[20]	**0.1**	**0.6**
3	Diluted cytoplasm[21]	**0.3**	**0.6**
4	Extremely dry cell[22]	**0.3**	**0.6**
5	Dry cell+ 5x longer time frame[23]	**0.9**	**1.6**
6	ctRNA density doubled[24]	**0.2**	**0.8**
7	ctRNA density tripled[25]	**0.4**	**1.2**
8	Increased metabolism (x2)[26]	**0.2**	**0.6**
9	ctRNAs velocity agitated (x10)[27]	**2**	**2.2**
10	Sampling rate increased (x2)[28]	**0.1**	**0.4**

So it is obvious, that a prospective consecutive ctRNA has no realistic chance of reaching a ribosome at a realistic rate in normal circumstances. The simulation gives more than one tRNA to hit the target, when the number of ctRNA-s or the time interval is exaggerated. It is to be emphasized, that even getting at least one suitable

[20] **Simulated by doubling the number of other particles.**

[21] **Number of other particles is reduced by 80%.**

[22] **Number of other particles is increased by 300%.**

[23] **Initial conditions are selected similar to set#2, but simulation time is increased by 400%, (simulating slower metabolism).**

[24] **Number of ctRNAs are increased, simulating tRNAs with concentration higher in the genome.**

[25] **Similar to set#6.**

[26] **The initial velocity of all moving objects in the simulaton space is doubled.**

[27] **Only ctRNAs initial velocity is increased 10 fold, to simulate a case when ctRNAs are able to move faster than other noncognate tRNAs.**

[28] **These simulations are meant to check whether our estimation for sampling rate are correct and no any collision is overlooked.**

tRNA hitting the target ribosome does not necessarily mean an actual success, because it should precisely reach to the A site (aminoacyl-tRNA binding site), not only any side of the ribosome.

The surface area of the ribosome could be approximated as the surface area of a sphere: $D^2*\pi=20^2*\pi$ $nm^2=1256nm^2$. The area of A site could be approximated as the area of a circle similar in size to a tRNA cross-section: $R^2*\pi=3.5^2*\pi$ $nm^2=38.5nm^2$.

The ratio of the areas is about 1:33, which means any tRNA that hits the ribosome has a small chance to reach the A site of the ribosome.

CHAPTER 6.

CONCLUSIONS

Since the already presented model proved that it is virtually impossible for the tRNA to reach the A site of the ribosome by random motion, we have to find possible explanations for this biological phenomenon.

Three different scenarios can be suggested;

1. the ribosome stores tRNAs and preselects them,
2. signaling between the ribosome and the cognate tRNA exists,
3. the tRNAs reach the ribosome in a preselected manner.

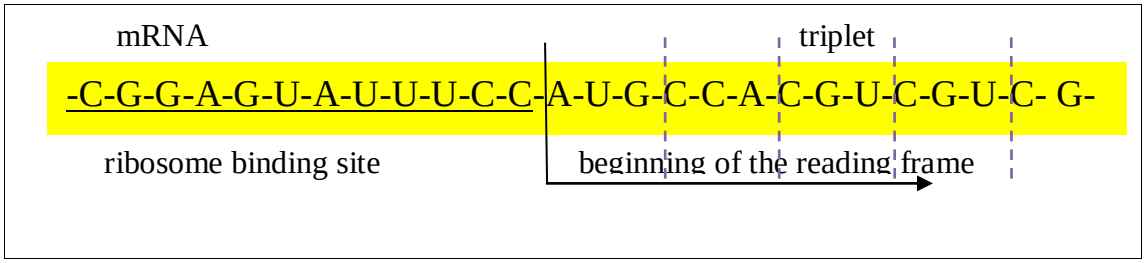


Figure 23. Reading frame

This is a portion of a gene of an mRNA: *E. coli* |EG10906|rpsG: 540 bp - 30S ribosomal subunit's S7 protein [21].

Table 9. Ribosomal 30S subunit protein S7 [21]

```
      gagttttggacaatcctgaattaacaacggagtatttcc
    1 - atg cca cgt cgt cgc gtc att ggt cag cgt
   31 - aaa att ctg ccg gat ccg aag ttc gga tca
   61 - gaa ctg ctg gct aaa ttt gta aat atc ctg
        “
  451 - ttc gca cac tac cgt tgg tta tcc ctt cgg
  481 - agt ttt agt cac cag gcg ggc gct tcc agt
  511 - aag cag ccc gct ttg ggc tac tta aat tga
```

Messenger RNA contains the information in the form of nucleotide base triplets. The difficulty in decoding this message is that if the beginning of the code is not known, depending on where the deciphering starts, three different meaning, senses could be interpreted. This is called the reading frame. Once the ribosome is attached to the mRNA, the beginning of the reading frame is fixed. During translation the ribosome moves along triplet by triplet until it reaches the stop codon.

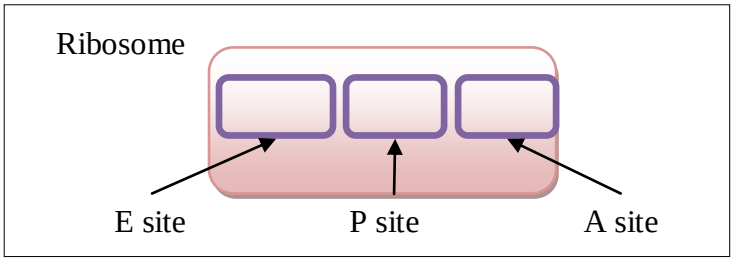


Figure 24. Three sites of the ribosome

*A site: aminoacyl-tRNA binding site

P site: peptidyl-tRNA binding site

E site: Exit site

6.1. Transfer RNAs are Stored in or at the Ribosome

The results of the simulations show that statistically, the number of delivered consecutive amino acids is not sufficient for protein synthesis, in the given time frame.

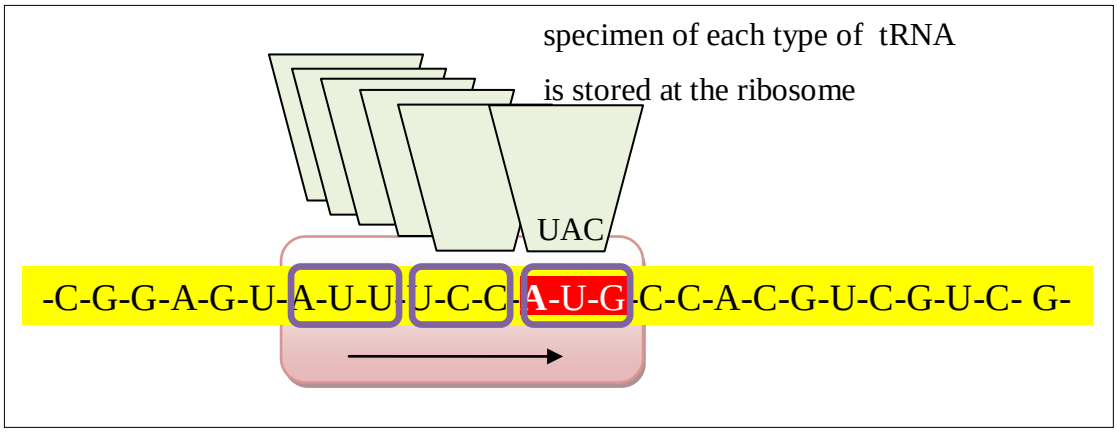


Figure 25. First hypothesis

Suppose the ribosome collects available tRNAs in its neighborhood and stores them for future amino acid assembly. This could happen only, when all type of tRNAs are available at hand to provide any amino-acid molecule for protein synthesis. The problem is that, there is no evidence in literature that ribosomes are able to store tRNAs. Another difficulty with this hypothesis is that there is no time for trial and error "fitting probe" for all types of tRNAs. The duration of one try is about 3 ms, so the ribosome can make an average of 6 or 7 "fitting probe" within the 20 ms time interval. The number of possible tRNA types is 42 (the actual number of different tRNAs in *E.coli* bacteria is more than 80, but some of them are redundant). We should have enough time for 21 trials, or 63 ms, but we only have 20 ms. Therefore, this hypothesis seems to be less likely (Figure 25).

6.2. Signaling between Ribosome and tRNA

i.e. tRNA as an autonomous agent

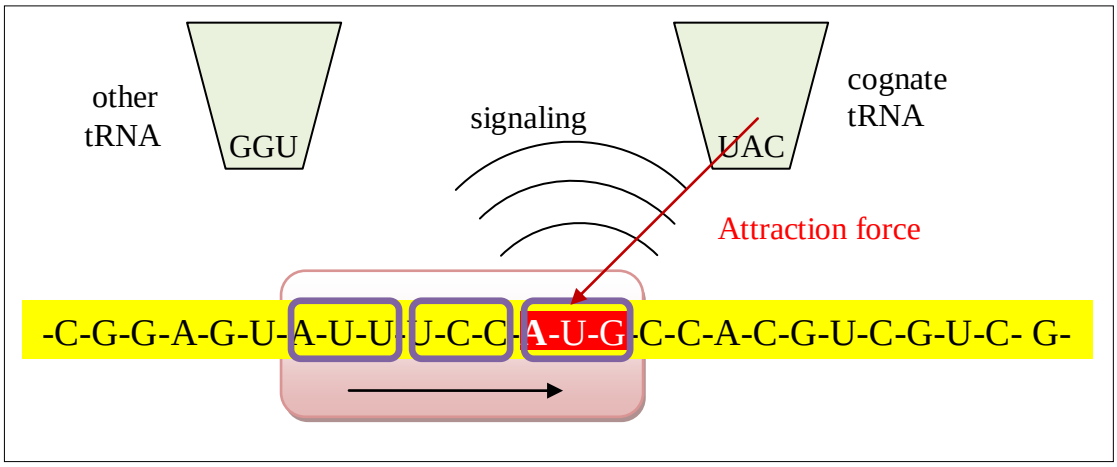


Figure 26. Second hypothesis

At the beginning of my research I had a preconception that ctRNAs act as autonomous agents, that are able to follow a beacon signal emitted by the ribosome and propel themselves right to the ribosome (Figure 26). Simulation results show however, that when the ctRNA's velocities are increased enormously, more ctRNAs can reach to the ribosome. The only criteria needed to justify this hypothesis is that ctRNA should move faster than other tRNAs. No data is available in literature on the

existence of any specific force acting between the ribosome and tRNAs when they are relatively far from each other. Cells are known to emit electromagnetic radiation, but there is no data on the existence of this kind of signaling between these two entities.

6.3. Transfer RNAs are Preselected Some Other Way

6.3.1. 'Reading Brackets'

i.e. tRNAs are preselected by the mRNA itself.

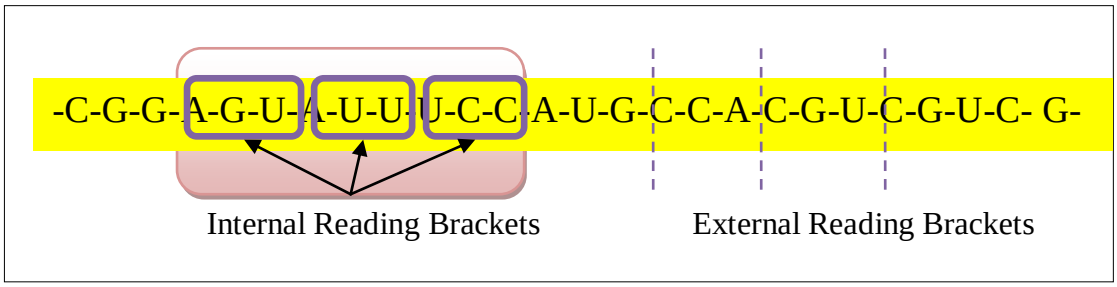


Figure 27. Reading brackets

RNA molecules and nucleotides tend to form pairs with their antisense counterpart. For example, ribosome's 16S rRNA component is the antisense to the Shine-Dalgarno sequence of the ribosome binding site and can bind to it. Once the ribosome subunits are assembled on the mRNA, the beginning of the actual reading frame is determined by its position. The ribosome has three sites to accommodate tRNA; they act as mechanical constraints. I call them 'Reading Brackets' (Figure 27).

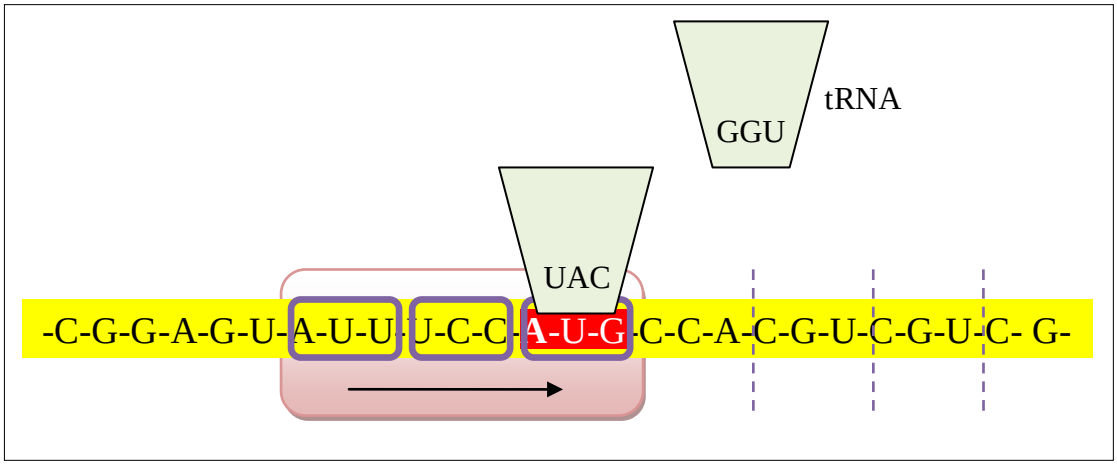


Figure 28. Third hypothesis; initiation

The ribosome steps onward, triplet by triplet, maintaining the correct translation of the entire reading frame from the start codon (AUG) to the terminating stop codon (UAA, UAG or UGA) (Table 1).

I propose, that the ribosome is able to fix a reading bracket externally too, beside its normal 'internal reading brackets' (Figure 28). The ribosome's body may be able to constrain tRNAs mechanically not only by its internal sites, but by its external shape (Figure 29).

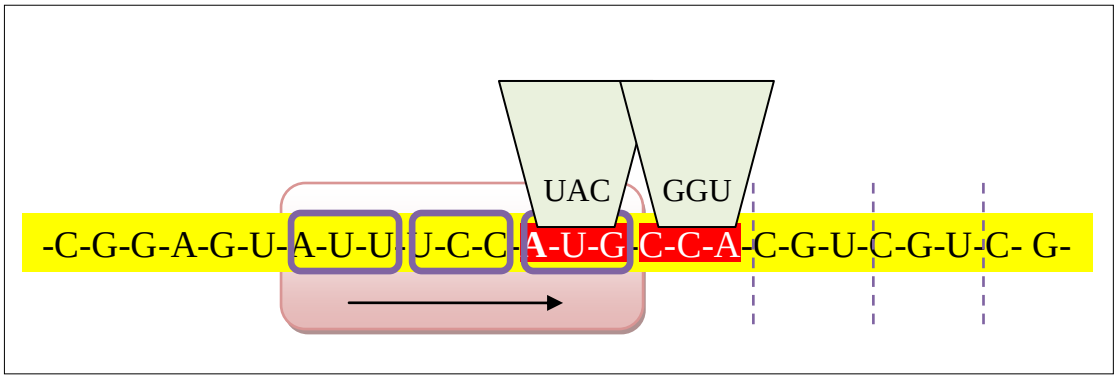


Figure 29. Third hypothesis; beginning of elongation

The free floating tRNAs have equal chance to bind to any part of the mRNA where the consecutive three bases are complementing its anticodon, regardless of the position of the reading frame. The ribosome is able to determine, or increase the possibility, that the three bases next to it will be occupied by a suitable charged tRNA.

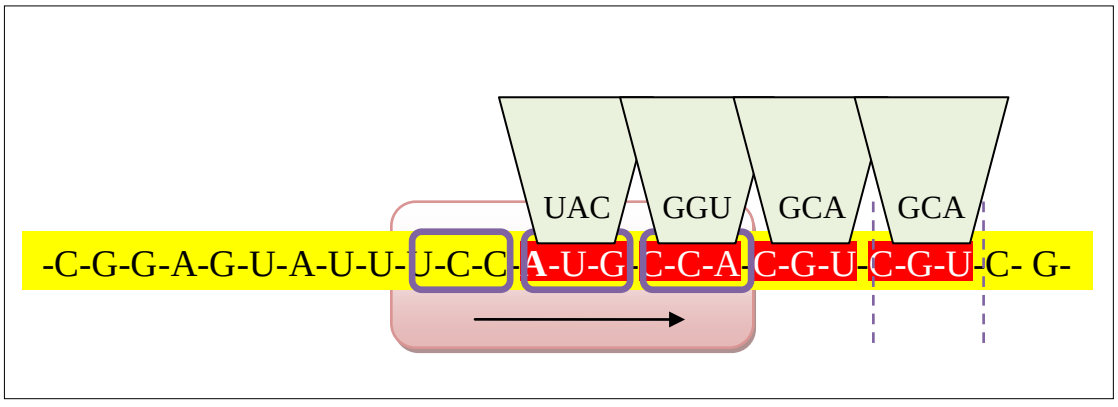


Figure 30. Third hypothesis; elongation in progress

This hypothesis (Figure 30) is proposed to account for the results of the simulation reported in Chapter 5. This hypothesis should be proved or disproved by conducting biological experiments or observations.

6.3.2. Transfer RNA Movement in the Bacterial Cytoplasm by Diffusion to the Ribosome

The interior of *E.coli* is without compartments, unlike eukaryotic cells. So the only mean of material transport viable is through diffusion [22], [11].

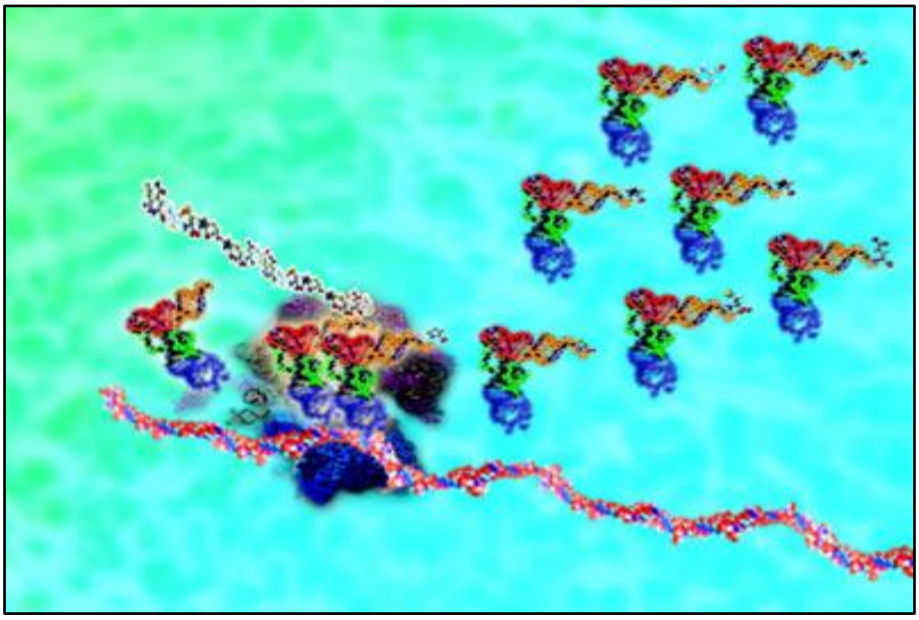


Figure 31. Original schematic view of protein synthesis from Berkeley Lab [23]

Despite this fact, in all the articles and animations explaining the translation process, tRNAs seem to be floating directly to the A site of the ribosome. See Figure 31. Moreover, the cognate always reaches the required site. In reality, at any given moment only one type of tRNA's anticodon is matching the actual codon out of more than 40 possible tRNA (see Table 1). The number of mismatches to a tRNA successful proofreading ratio is not yet known.

6.3.3. Transfer RNA Movement in the Bacterial Cytoplasm by Diffusion to the mRNA

The essence of my theoretical proposal is that the mRNA itself is capable of attracting tRNA and storing them. This process is proceeding in a parallel manner contrarily to what was imagined, that it took place at the ribosome itself in a serial manner. The timing is not so critical in this case. Moreover tRNAs are preselected automatically if they are able to align with each other without any gap before reaching the ribosome, or at least the chance is increased dramatically that there would only be three types of tRNA present next to the ribosome before it makes a step to receive the next tRNA.

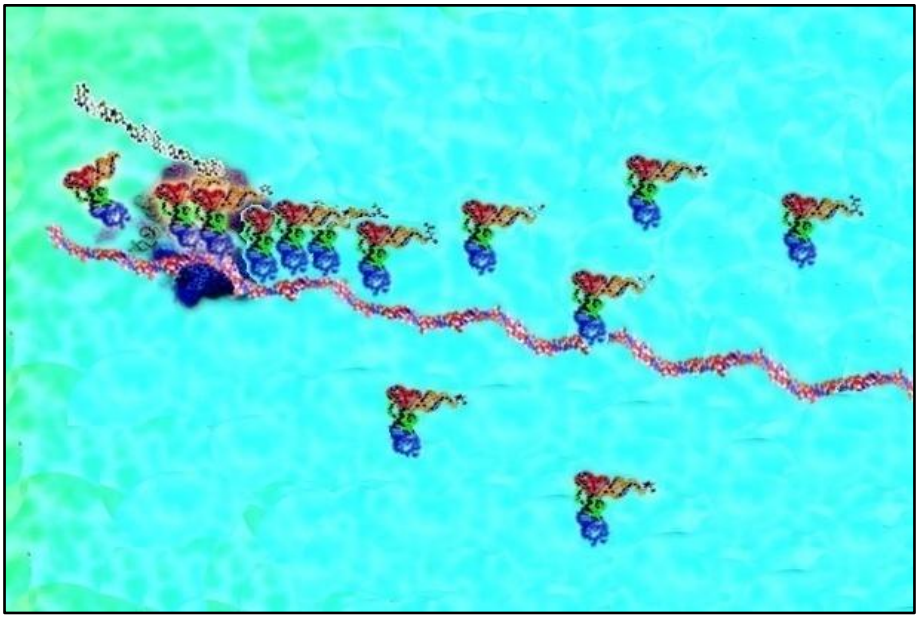


Figure 32. Protein assembly proposing that tRNA-s were preselected by the mRNA itself even before they enter into ribosome's "A site"

Figure 32 is a modified version of the previous citation, showing how the preselection process is proposed. The tRNAs are moving randomly in the cytoplasm (by diffusion) and are interacting with the mRNA. Once the ribosome is assembled and attaches itself to the mRNA, the "reading brackets" are established. As the ribosome moves on, these external reading brackets too move forward.

6.3.4. Proposed Explanation of Protein Assembling in Details

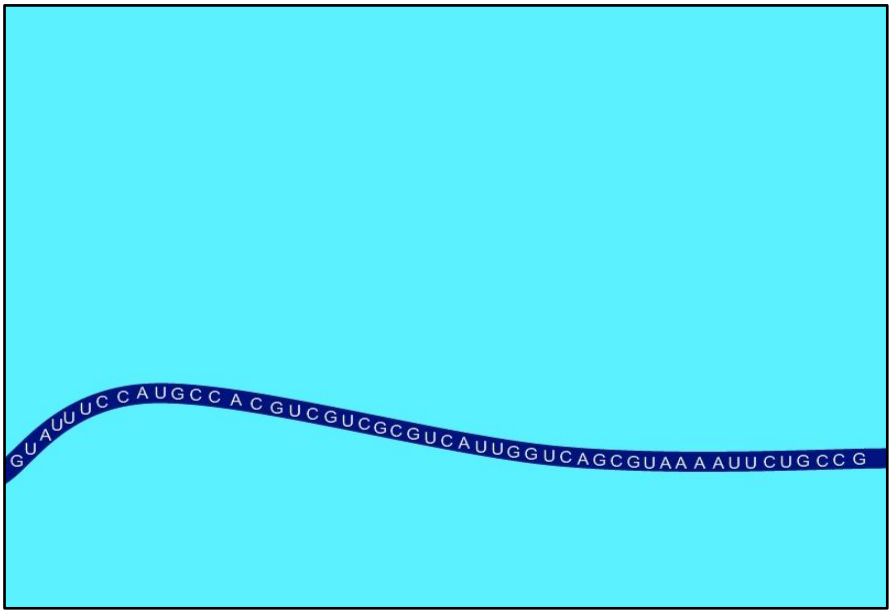


Figure 33. Growing messenger RNA chain

After synthesis mRNA is ready for translation. (Figure 33; It is not shown how tRNAs are able to stick to the growing mRNA chain randomly.)

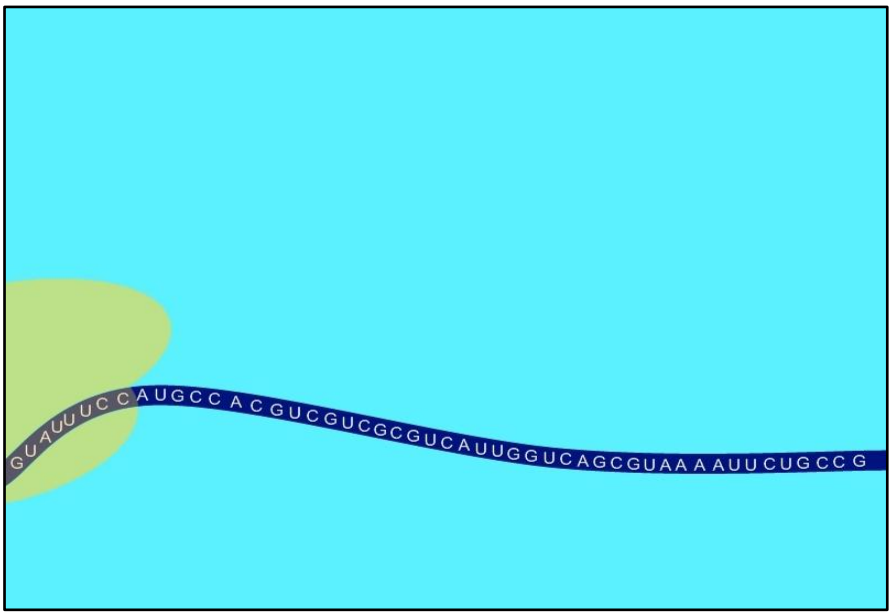


Figure 34. Ribosome is assembled

The first step of translation is when the ribosome is assembled from rRNAs and proteins. 16S rRNA attaches itself to the Shine-Dalgarno sequence, ready to start protein polymerization (Figure 34).The first amino-acid to start a protein chain is always methionine [5].

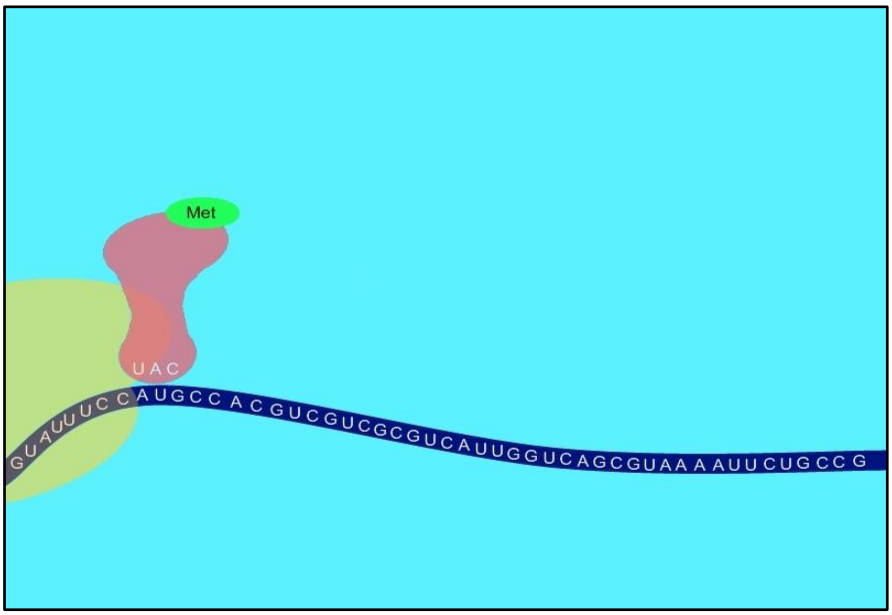


Figure 35. The ribosome is waiting for the first tRNA

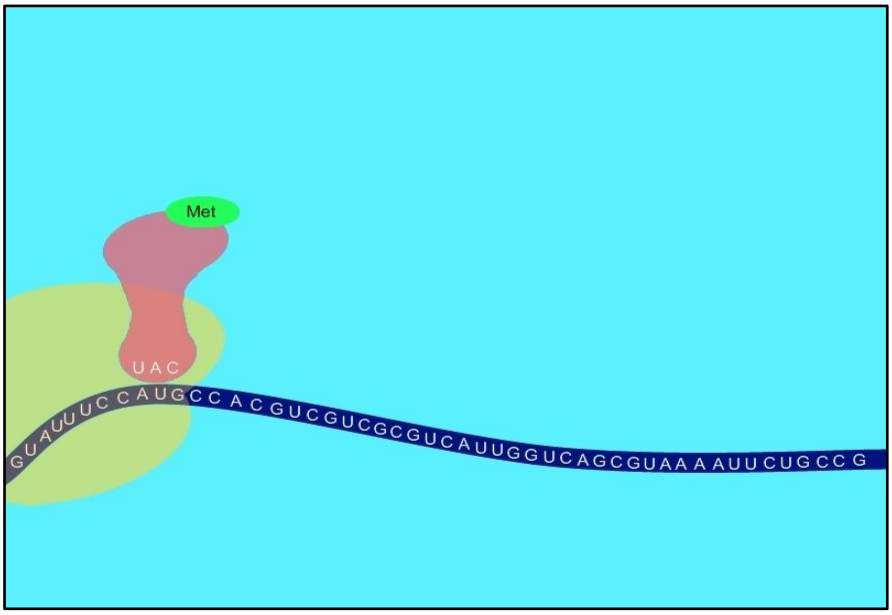


Figure 36. First tRNA entered to A site

The first tRNA, Met-tRNA, enters into the ribosome A site (Figure 35 and Figure 36), waiting for the arrival of the next amino-acid. The next codon is: C C A.

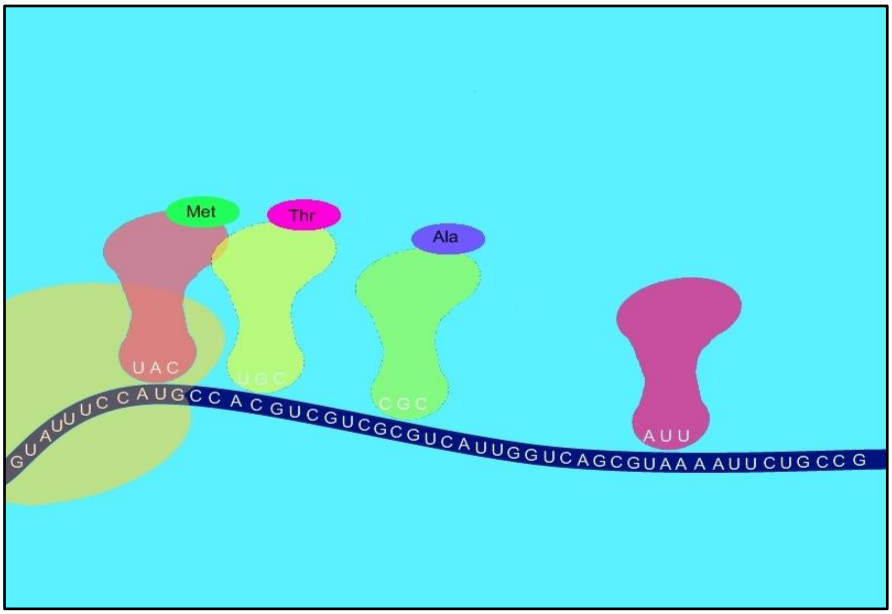


Figure 37. Preselection 1

Three type of tRNA have the highest chance to be present in the vicinity of the ribosome; two of them are out of the reading frame. Thr-tRNA matches with A C G (Figure 37) and His-tRNA matches with C A C base sequences (Figure 38), but not with the C C A. (Other tRNAs shown here are attached randomly to the mRNA.)

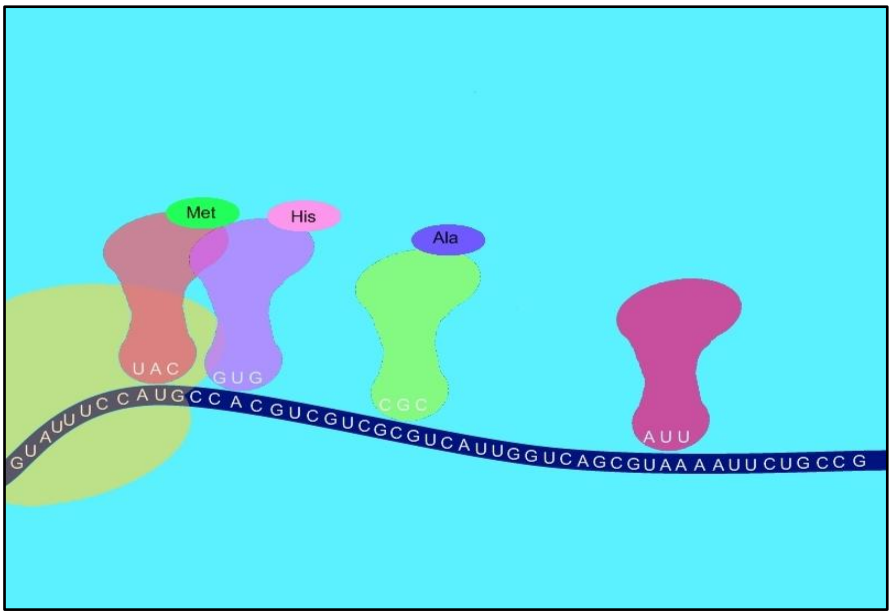


Figure 38. Preselection 2

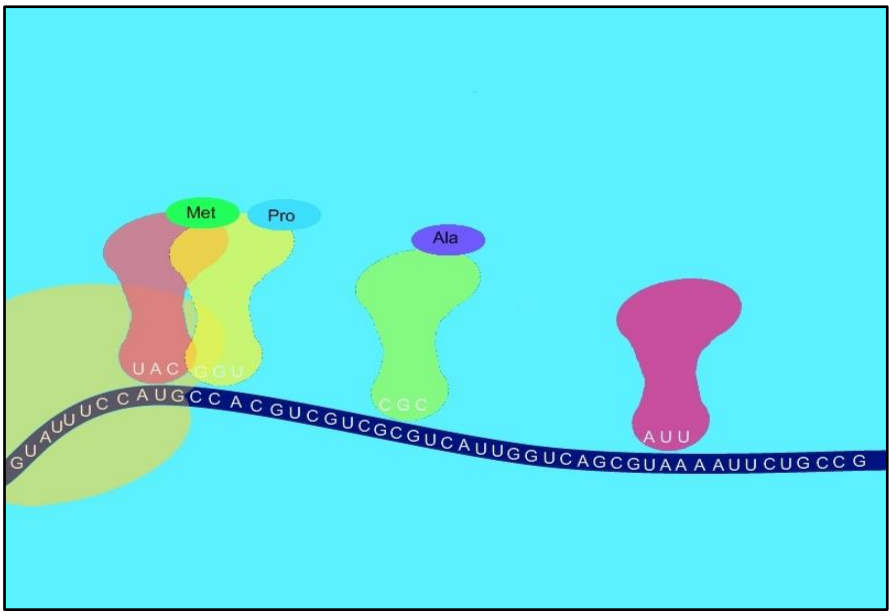


Figure 39. Preselection 3

Only Pro-tRNAs G G U anticodon matches with the C C A codon (Figure 39).

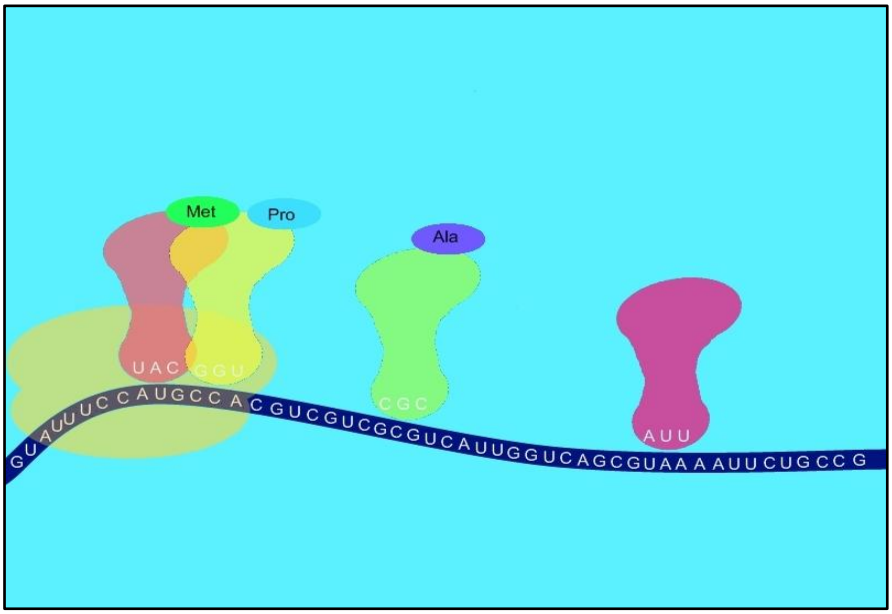


Figure 40. Proofreading

The preselection process ends when the ctRNA is found and accepted by proofreading at the ribosome A site.

6.3.5. Brainstorming

The following detailed explanation is purely speculative. If the intuition in the previous chapter is proved to be true, ones might think of the idea by mathematical induction.

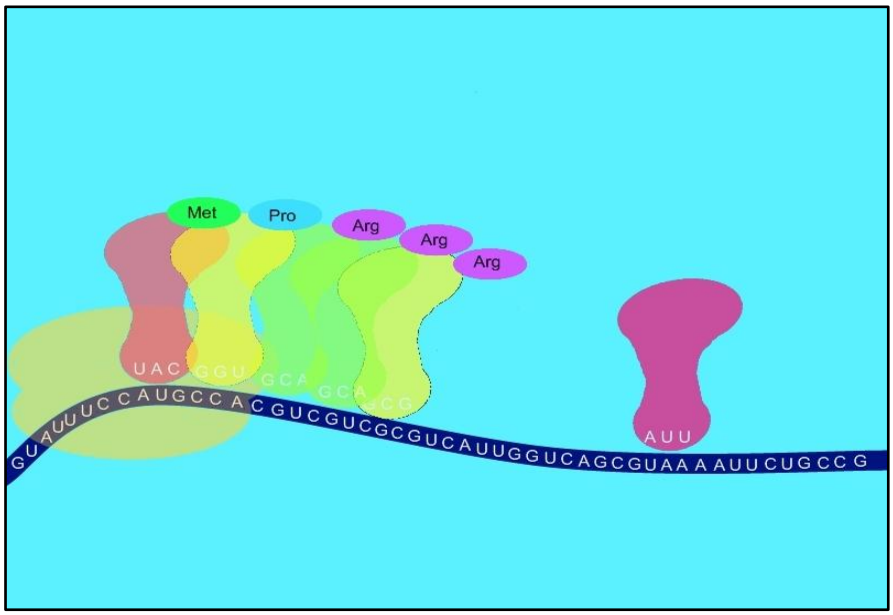


Figure 41. Elongation

The tRNA, according to their anticodon, line up outside the ribosome. The tRNA is leaning against the ribosome and bonds with the first two bases of the codon's triplet using hydrogen bonds. Therefore the interaction between the third position codon and anticodon becomes weak and this could explain the wobbling effect too.

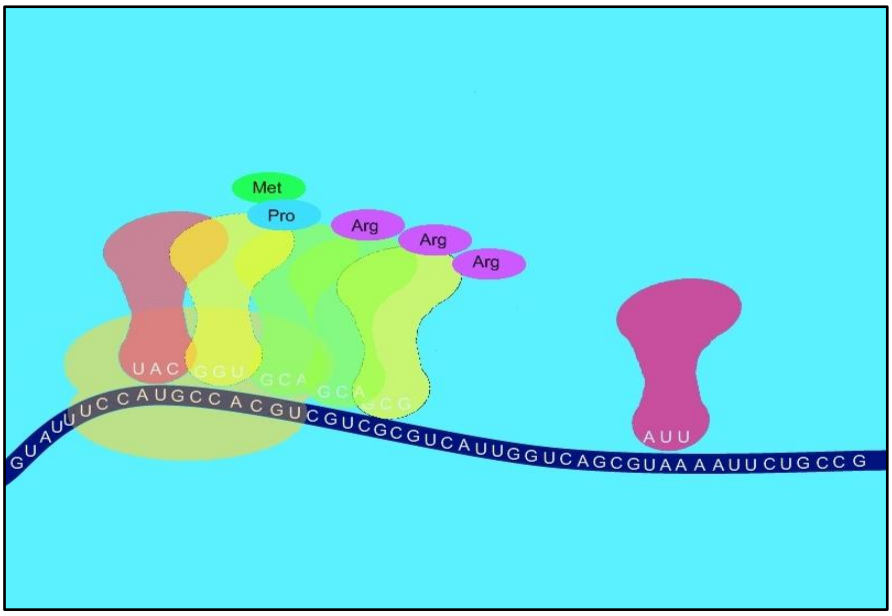


Figure 42. Peptide bonding

The ribosome steps forward and the polypeptide chain grows.

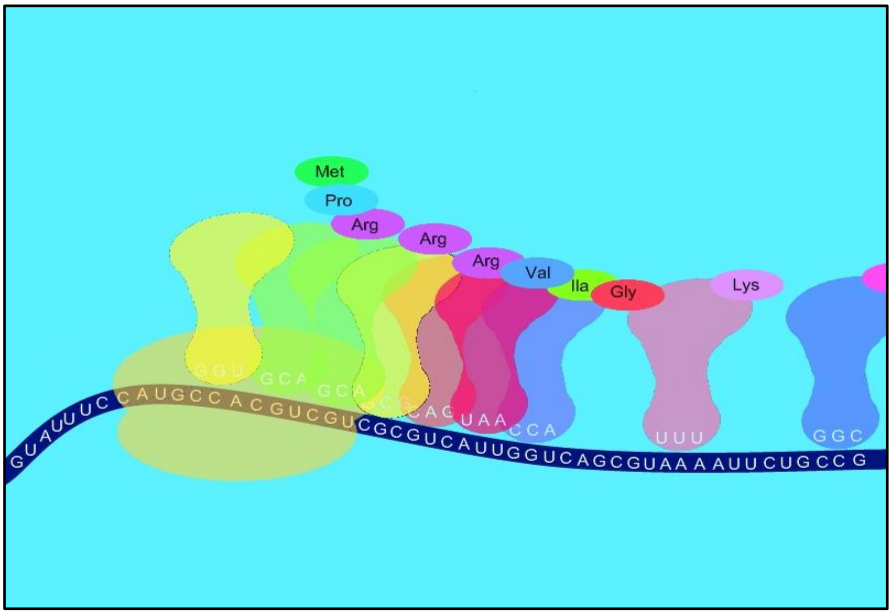


Figure 43. Translation in progress

So as the ribosome progresses along the mRNA, more and more tRNA is deposited, then it is selected, following the reading frame. Any other tRNA attached to the mRNA previously off the reading frame could be washed away by water molecules continuously bombarding these molecules. But tRNAs that are touching the ribosome side or leaning towards each other could withhold it and remain attached to

the mRNA. Additionally ribosome stepping forward during protein assembly can easily scrape these tRNAs.

The RNA secondary shape is helical, one full turn is made of 11 bases. So when the ribosome steps forward relative to the mRNA, it should turn 33° to right, giving free space to the tRNA which was leaning towards it previously, to enter the A site.

The most important elements of this scenario are the water molecules. Seventy percent of the cytoplasm is water; thus it could be considered the most important component of the whole cell.

CHAPTER 7.

SUMMARY

The aim of this master thesis is to set up a kinetic model of tRNA molecular movement in bacterial cytoplasm, develop a computational program, run simulations and visualize particle movement.

Three hypothesizes are based on simulation results and intuition are also discussed. The main features of these hypothesizes are summarized in Table 10.

Table 10. Comparison chart

Hypothesis	Pros	Cons	Opinion
1. The ribosome stores tRNAs and preselects them.		Timeframe is not sufficient.	Very unlikely.
2. Signaling exists between the ribosome and the cognate tRNA.	Logical explanation. No contradiction against it.	There is no observation on the existence of signaling or specific force between the tRNA and the ribosome. It cannot explain the wobbling effect.	Possible. Should be proved or disproved by conducting biological experiments or observations.
3. The tRNAs reach the ribosome in a preselected manner.	Logical explanation. RNA-RNA interactions are known phenomena. It can explain the wobbling effect.	There is not any observation.	Possible. Should be proved or disproved by conducting biological experiments or observations. New modeling and simulation is needed.

REFERENCES

1. Holley, R W, et al.*Nucleotide sequence in the yeast alanine transfer ribonucleic acid.* 1965, J.Biol.Chem, Vol. 240, pp. 2122-2128.

2. *uni-bayreuth.de.* [Online] http://www.uni-bayreuth.de.

3. *cgl.ucsf.edu.* [Online] www.cgl.ucsf.edu/.../tutorial/trna/trna.gif.

4. Cooper, Geofrey M. *The Cell.* 1997.

5. Lodish, Harvey, et al. *Molecular Cell Biology.* 2000.

6. *fig.cox.miami.edu.* [Online] http://fig.cox.miami.edu/~cmallery/150/gene/c7.17.16ribosome.jpg.

7. *nobelprize.org.* [Online] http://www.nobelprize.org/educational_games/medicine/dna/index.html.

8. Nirenberg, M. and Leder, P. *RNA codewords and protein synthesis.* 145, Sep 25, 1964, Science, pp. 1399-407.

9. Sanbonmatsu, K.Y., Joseph, S. and Tung, C.S. *Simulating movement of tRNA into the ribosome during decoding.* 2005, PNAS, pp. 15854-15859.

10. Schwehm, M. *Parallel Stochastic Simulation of Whole-Cell Models.* s.l. : ICSB, 2001.

11. Elowitz, M.B., et al. *Protein Mobility in the Cytoplasm of Escherichia coli.* 1999, Journal of Bacteriology, pp. 197-203.

12. Deich, J, et al. *Visualization of the movement of single histidine kinase molecules in live Caulobacter cells.* 45, Nov. 9, 2004, PNAS, Vol. 101, pp. 15921-15926.

13. Golding, I and Cox, Edward C. *Physical Nature of Bacterial Cytoplasm.* 2006, Physical Review Letters, pp. 96-99.

14. Broderick, G., et al. *A Life-Like Virtual Cell Membrane Using Discrete Automata.* 2005, In Silico Biology, pp. 163-178.

15. Plimpton, Steven J. and A., Slepoy. *Microbial Cell Modeling via Reacting Diffusive Particles.* 2005, Journal of Physics, pp. 305-09.

16. Sanbonmatsu, K.Y. *Energy landscape of the ribosomal decoding center.* 2006, Biochimie, pp. 1053-1059.

17. Sanbonmatsu, K.Y. *Alignment/misalignment hypothesis for tRNA selection by the ribosome.* 2006, Biochimie, pp. 1075-1089.

18. Institute for Biomolecular Design. *E.coli* Statistics ; Project CyberCell. [Online] http://redpoll.pharmacy.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi.

19. Berchek, Chad. http://www.geocities.com/vobarian/2dcollisions/. *www.geocities.com/vobarian/2dcollisions/2dcollisions.pdf* . [Online] www.geocities.com/vobarian/2dcollisions/2dcollisions.pdf .

20. www.plasmaphysics.org.uk/collision3d.htm. [Online]

21. Colibri. http://genolist.pasteur.fr/Colibri/genome.cgi. *http://genolist.pasteur.fr/Colibri/genome.cgi.* [Online] 1999.

22. Fraser, Christopher S and Hershey, John WB. *Movement in ribosome translocation.* 2, Journal of Biology, Vol. 4.

23. *lbl.gov.* [Online] http://www.lbl.gov/Science-Articles/Archive/sabl/2005/November/ribosome-groove.html.

24. Li, Wen-Hsiung. *Molecular Evolution.* 1997.

25. Podani, János. *Bevezetés a többváltozós biológiai adatfeltárás rejtelmeibe.* 1997.

26. Stansfield, William D. *Genetika Elmélet és gyakorlat.* 1997.

27. Gerhart, John and Kirschner, Marc. *Cells, Embryos and Evolution.* 1997.

28. Gilbert, Scott F. *Developmental Biology.* 1997.

29. Smith, John Maynard and Szathmáry, Eörs. *The origins of life.* 1999.

30. Schwartz, Jeffrey H. *Sudden origines Fossils, Genes and emergence of species.* 1999.

31. Csaba, György. *Sejtbiológia.* 1996.

32. Lagerkvist, Ulf. *DNA Pioneers and Their Legacy.* 1998.

33. Ptashne, Mark. *A Genetic Switch.*

34. Rédei, George P. *Genetics.*

35. Watson, James D. *Molecular Biology of the Gene.*

36. Cochella, Luisa and Green, Rachel.*An Active Role for tRNA in Decoding Beyond Codon-Anticodon Pairing.* 2005, Science, pp. 1178-1180.

37. Chechetkin, V.R. *Genetic code from tRNA point of view.* 2006, Journal of Theoretical Biology, pp. 922-934.

38. Pape, Tillmann, Wintermeyer, Wolfgang and Rodnina, Marina. *Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome.* 1999, The EMBO Journal, pp. 3800-3807.

39. Graversen, S.E. and Peskin, G. *Maximal inequalities for the ornstein-Uhlenbeck Process.* 2000, Proceedings of the American Mathematical Society, pp. 3035-3041.

40. Ninio, Jacques. *Multiple stages in codon-anticodon recognition:double-trigger mechanisms and geometric constraints.* 2006, Biochimie, pp. 963-992.

41. Dittmar, K.A., et al. *Selective Charging of tRNA isoacceptors induced by Amino-acid starvation.* 2004, EMBO Reports, pp. 151-157.

42. Daviter, T., Gromadski, K.B. and M.V., Rodnina. *The ribosome`s response to codon-anticodon mismatches.* 2006, Biochimie, pp. 1001-1011.

43. O`Connor, Michael. *tRNA hopping:effects of mutant tRNAs.* 2003, Biochimica et Biophysica Acta, pp. 41-46.

44. Valle, Mikel, et al. *Locking and Unlocking of Ribosomal Motions.* 1, Cell, Vol. 114.

45. chemis/Chemis-NucleicAcid/Graphics/tRNA.jpg. *www.geneticengineering.org.* [Online] http://www.geneticengineering.org/chemis/Chemis-NucleicAcid/Graphics/tRNA.jpg.

46. *www.ncbi.nlm.nih.gov.* [Online] http://www.ncbi.nlm.nih.gov/Class/MLACourse/Original8Hour/Genetics/nucleotide2.gif.

47. *scientificpsychic.com.* [Online] http://scientificpsychic.com/fitness/dna0.gif.

48. [Online] http://upload.wikimedia.org/wikipedia/commons/0/04/NA-comparedto-DNA_thymineAndUracilCorrected.png.

Appendices

Appendix A.

MATLAB Codes

```
function [Lcts,Trg,Stats,Vol]=main
%  Main Program
%Author:Piros Sandor
%AUS Mechatronics Master Student
%Al Ain April 2008
%
%Initialization
%
Lcts=0;
Trg=0;
Stats=0;
Trt=0;
%Read in initial data for simulation
[P0,Quant]=simdata;
% Units used: nm, ms, kD, nm/ms
% Nr: number of ribosomes
Nr=Quant(1);
% Nct:Number of cognate tRNAs
Nct=Quant(2);
% Nt: number of tRNAs
Nt=Quant(3);
% No: number of other particles
No=Quant(4);
% del_t: delta t, time step
del_t=Quant(5);
% dur: duration of simulation
dur=Quant(6);
```

```
%K: number of simulations
K=Quant(7);
% Vol: space
Volu=Quant(8);
Vol=[Volu/2;Volu/2;Volu/2];
% P0: list of particles, initial state [ty;r;m;v0[3,1];l0[3,1]]
%ty: type of particle: (1,ribosome; 2,cognate tRNA, 3 other tRNA, 4, other)
%r: radius of particle
%m: mass of particles
%v0:initial speed
%l0:initial location
Trgt=0; %times cognate tRNA hit the target
T=int16(dur/del_t);   %number of time steps
%Lct output matrix gives the locations of cognate tRNAs at each timestep
%Size(Lct)=(3,Nct,T)
%L: location matrix; location of each particle in 'delt' timesteps
%L0:location of each particle at time 0
N=Nr+Nct+Nt+No;
%P0=[Ty;R;M;V0;L0];
Ty=P0(1,:);
L0=P0(7:9,:);
V0=P0(4:6,:);
R=P0(2,:);    %radii
M=P0(3,:);
%Start of main procedure
Lcts=zeros(3*K,Nct,T);
for k=1:K
    Trt=0;
    Lct=zeros(3,Nct,T);
    tgt=0;
    Nr=int16(Nr);
    Nct=int16(Nct);
    if k==1
```

```
    else
        %redistribute particle before next simulation
        [L0]=distrib(R,Vol);
    end
    V=V0; %actual speed
    L=L0;   %actual location
    for t=1:T
    [L,V]=outb(L,Vol,V);                %out of boundary
    [L,V]=nloc(L,V,Nr,Nct,del_t,M);       %calculate new location
    [Trt]=trgt(R,L,Nr,Nct,Trt,t,tgt);       %target detction
    %[V]=cdet(R,V,L,M);            %collision detection
    %or
    [V]=cdets(R,V,L,Nr,Nct,Nt,M);   %collision detection simple
    [V]=colRNA(R,Vol,V,L,Nr,Nct,Nt,M);   %collision with mRNA
    Lct(:,:,t)=L(:,Nr+1:Nr+Nct);       %output data
    end
    sTrt=size(Trt,2);
    ka=int16(2*k);
    Trg((ka-1):ka,1:sTrt)=Trt;
    kb=int16(3*k);
    Lcts(kb-2:kb,:,:)=Lct;
    Stats(k)=sum(ne(Trt(1,:),0));
end
%Prepare statistic
Stats(K+1)=sum(Stats)/K;
%
%   Results:
%Lcts:
%   First dimension of matrix:     Coordinates of ctRNAs
%   Second dimension of matrix:    ctRNAs in sequence
%   Third dimension of matrix:     time intervals
%Trg:
%   Odd rows:   serial number of ctRNA reached to target
```

```
%   Even rows:  time stamp
%Stats:
%       first K elements are the results of simulations
%       (K+1)st element is the average of K simulations
```

```
function [P0,Quant]=simdata
%  Simulation data input
%
%Author:Piros Sandor
%AUS Mechatronics Master Student
%Al Ain April 2008
echo on
% Units used: nm, ms, kD, nm/ms
% Nr: number of ribosomes        (144)
Nr=143;
Nr=input('Enter the number of ribosomes (143) "Nr">');
Nr=Nr+1;
% Nct:Number of cognate tRNAs       (40)
Nct=40;
Nct=input('Enter the number of cognate tRNAs       (40) "Nct">');
% Nt: number of tRNAs               (1600)
Nt=1600;
Nt=input('Enter the number of tRNAs       (1600) "Nt">');
% No: number of other particles     (10000)
No=10000;
No=input('Enter the number of other particles     (10000) "No">');
% del_t: delta t, time step         (0.1 ms)
del_t=0.1;
del_t=input('Enter delta t, time step         (0.1 ms) "del_t">');
% t: duration of simulation         (20 ms)
dur=20;
dur=input('Enter duration of simulation         (20 ms) "dur">');
% K: number of simulations
K=int16(10);
K=input('Enter number of simulations         (10) "K">');
% Vol: space                      [100;100;100]
Volu=200;
Volu=input('Enter dimension of simulated space         (200 nm) "Vol">');
```

```
Vol=[Volu/2;Volu/2;Volu/2];
%Vat: average speed of tRNAs
Vat=10;
Vat=input('Enter average speed of tRNAs     (10nm/ms) "Vat">');
%Vao: average speed of other particle
Vao=10;
Vao=input('Enter average speed of other particle     (10nm/ms) "Vao">');
%Omni: distribution of initial velocities, omnidirectional or not?
Omni='Y';
Omni=input('Enter initial velocity distribution of particle, omnidirectional?("Y" or
"N") >');
% P0: list of particles, initial state [ty;r;m;v0[3,1];l0[3,1]]
%r: radius of particle
%m: mass of particles
%v0:initial speed
%l0:initial location
Trgt=0; %times cognate tRNA hit the target
T=single(dur/del_t);   %number of time steps
%Lct output matrix gives the locations of cognate tRNAs at each timestep
%Size(Lct)=(3,Nct,T)
%L: location matrix; location of each particle in 'delt' timesteps
%L0:location of each particle at time 0
%Initialization
N=uint16(Nr+Nct+Nt+No);
%Ty: type of particle: (1,ribosome; 2,cognate tRNA, 3 other tRNA, 4, other)
Ty=[ones(1,Nr),2*ones(1,Nct),3*ones(1,Nt),4*ones(1,No)];
Rr=10*ones(1,Nr);
Rct=3.5*ones(1,Nct);
Rt=3.5*ones(1,Nt);
Ro=5*(1+rand(1,No));
R=single([Rr,Rct,Rt,Ro]);  %give radius for particle
%volume of particles*100/V space
Volume=single((sum(R.^2)*4*pi/3)*100/(2^3*Vol(1)*Vol(2)*Vol(3)))
```

```
%distribute particles randomly in simulation space
L0=distrib(R,Vol);  %subroutine call
Mr=2700*ones(1,Nr);
Mct=25*ones(1,Nct);
Mt=25*ones(1,Nt);
Mo=1*Ro.^2;
M=single([Mr,Mct,Mt,Mo]); %give mass for particle
%Omnidirectional initial velocity distribution
if Omni=='Y';
   for i=1:3;
   V0(i,Nr+1:N)=[Vat*2*(-0.5+rand(1,Nct)),Vat*2*(-0.5+rand(1,Nt)),Vao*2*(-
0.5+rand(1,No))];
   end
% Non omnidirectional distribution
else
   for i=1:3;
   V0(i,Nr+1:N)=[Vat*rand(1,Nct),Vat*rand(1,Nt),Vao*rand(1,No)];
   end
end
%Calculate mean velocity of cognate tRNAs
for k=Nr+1:Nr+Nct;
Va(k-Nr)=norm(V0(:,k));
end
%Vamean: mean velocity of cognate tRNAs
Vamean=mean(Va)
%subroutine output values:
%P0:particle parameters
P0=[Ty;R;M;V0;L0];
%Quantities used:
Quant=[Nr,Nct,Nt,No,del_t,dur,K,Volu];
```

```
function [L0]=distrib(R,Vol)
%  Particle distribution in simulation space
%
%Author:Piros Sandor
%AUS Mechatronics Master Student
%Al Ain April 2008
%
N=length(R);
L0(:,1)=[0;0;0];    %location of target ribosome
for n=2:N
   for i=1:3;
   L0(i,n)=-Vol(i)+(2*Vol(i))*rand;    %distribute particles randomly
   end
rn=R(n);    %radius of particle n
lno=L0(:,n);    %location vector of particle 'n'
   for m=1:n-1;  %check whether this location is already occupied
      if norm(L0(:,m)-lno)<=rn+R(m);    %try new location
      n=n-1;
      end
   end
end
```

```
function [L,V]=outb(L,Vol,V)
%  OUTSIDE OF BOUNDARY DETECTION
%This function detects when any particle is outside of the boundary V space.
%New location will be the opposite side, with same speed.
%Author:Piros Sandor
%AUS Mechatronics Master Student
%Al Ain Oct 2007
%
for m=1:3
   W=Vol(m);
   for n=1:size(L,2)
   if L(m,n)>=W
%        L(m,n)=-W;
        V(m,n)=-V(m,n);
   elseif L(m,n)<=-W
%        L(m,n)=W;
       V(m,n)=-V(m,n);
   else
   end
   end
end
```

```
function [L,V]=nloc(L,V,Nr,Nct,del_t,M)
%  New location vector
%This function calculates new location of particles after del_t time interval.
%Author:Piros Sandor
%AUS Mechatronics Master Student
%Al Ain Oct 2007
%
cons=0;   % possible force coefficient
for nct=Nr+1:Nr+Nct %force applied on cognate tRNA
      F=cons*L(:,nct)/norm(L(:,nct));
       %Acting force:  F=cons*L/norm(L)
      V(:,nct)=V(:,nct)+F*del_t/M(nct);
end
L=L+V*del_t;
%if ribosome considered stationary:
%L(:,Nr+1:end)=L(:,Nr+1:end)+V(:,Nr+1:end)*del_t;
```

```
function [Trt]=trgt(R,L,Nr,Nct,Trt,t,tgt)
%  TARGET DETECTION
%This function detects when a cognate tRNA hits the target.(The distance between
the
%centers of tRNA and ribosome is smaller than the summ of the 2 radii.
%Author:Piros Sandor
%Al Ain Oct 2007
N=size(L,2);
for m=Nr+1:Nr+Nct
   rm=R(m);    %radius of particle m
   lm=L(:,m);    %location vector of particle 'm'
   if norm(L(:,m))<=R(m)+R(1);    %count hitting target
     ct=m-Nr;
     if   sum(eq(Trt(1,:),ct)) == 0
     tgt=tgt+1;
     Trt(1,tgt)=ct;           %serial No of ctRNA
     tm=t;
     Trt(2,tgt)=tm;           %time stamp
     else %tRNA reached to the target previously too
     end
   else
   end
end
```

```
function [V]=cdet(R,V,L,M)
%  COLLISION DETECTION
%Author:Piros Sandor
%AUS Mechatronics Master Student
%Al Ain Oct 2007
%
%This function detects when 2 particles collide. (The distance between the
%centers is smaller than the summ of the 2 radii.
%
N=size(L,2);
for m=1:N
   rm=R(m);    %radius of particle m
   lm=L(:,m);    %location vector of particle 'm'
   for n=m+1:N
   if norm(L(:,n)-lm)<=rm+R(n);    %call elcol3d or elcol3ds
%       [V(:,m),V(:,n)]=elcol3d(V(:,m),V(:,n),M(m),M(n),lm,L(:,n));
      [V(:,m),V(:,n)]=elcol3ds(V(:,m),V(:,n),M(m),M(n),lm,L(:,n));
   end
   end
end
```

```
function [V]=cdets(R,V,L,Nr,Nct,Nt,M)
%  COLLISION DETECTION (simplified, only for tRNA)
%Author:Piros Sandor
%Al Ain Oct 2007
%
%This function detects when 2 particles collide. (The distance between the
%centers is smaller than the summ of the 2 radii.
%
N=size(L,2);
for m=1:Nr+Nct+Nt;
   rm=R(m);    %radius of particle m
   lm=L(:,m);    %location vector of particle 'm'
   for n=m+1:N
   if norm(L(:,n)-lm)<=rm+R(n);    %call elcol3d or elcol3ds
%       [V(:,m),V(:,n)]=elcol3d(V(:,m),V(:,n),M(m),M(n),lm,L(:,n));
      [V(:,m),V(:,n)]=elcol3ds(V(:,m),V(:,n),M(m),M(n),lm,L(:,n));
   end
   end
end
```

```
function [v1_,v2_]=elcol3d(v1,v2,m1,m2,c1,c2)
%This function calculates velocity vectors after elastic collision
%Author:Piros Sandor
%AUS Mechatronics Master Student
%Al Ain June 2007
%
% v1 & v2: velocity vectors before collision
% c1 & c2: sphere center's coordinates at the moment of collision
% m1 & m2:masses
%v1_ & v2_: calculated velocities after collision
n=(c2-c1); n=n/norm(n); %normal unit vector of collision
t1=[n(2)-n(3);n(1)-n(3);n(1)-n(2)];
t1=t1/norm(t1); %tangent unit vectors of collision
t2=[n(2)*t1(3)-n(3)*t1(2);n(3)*t1(1)-n(1)*t1(3);n(1)*t1(2)-n(2)*t1(1)];
t2=t2/norm(t2);
v1n=v1'*n; v1t1=v1'*t1; v1t2=v1'*t2; %components of velocity vectors before
collision
v2n=v2'*n; v2t1=v2'*t1; v2t2=v2'*t2;
v1n_=(v1n*(m1-m2)+2*m2*v2n)/(m1+m2); %normal  components of velocity
vectors after collision
v2n_=(v2n*(m2-m1)+2*m1*v1n)/(m1+m2);
v1_=v1n_*n+v1t1*t1+v1t2*t2;
v2_=v2n_*n+v2t1*t1+v2t2*t2;  % velocity vectors after collision
```

```
function [v1_,v2_]=elcol3ds(v1,v2,m1,m2,c1,c2)
%This function calculates velocity vectors after elastic collision (simple way)
%Author:Piros Sandor
%AUS Mechatronics Master Student
%Al Ain Oct 2007
%
% v1 & v2: velocity vectors before collision
% c1 & c2: sphere center's coordinates at the moment of collision
% m1 & m2:masses
%v1_ & v2_: calculated velocities after collision
n=(c2-c1); n=n/norm(n); %normal unit vector of collision
v1n=v1'*n;  %components of velocity vectors before collision
v2n=v2'*n;
v1n_=(v1n*(m1-m2)+2*m2*v2n)/(m1+m2); %normal  components of velocity
vectors after collision
v2n_=(v2n*(m2-m1)+2*m1*v1n)/(m1+m2);
v1_=v1+(v1n_-v1n)*n;
v2_=v2+(v2n_-v2n)*n;  % velocity vectors after collision
```

```
function [V]=colRNA(R,Vol,V,L,Nr,Nct,Nt,M)
%  COLLISION with mRNA( only for tRNA)
%Author:Piros Sandor
%Al Ain April 2008
%
%This function detects when a tRNA collide with the mRNA.
%
N=size(L,2);
for m=Nr+1:Nr+Nct+Nt;
   rm=R(m);    %radius of particle m
   lm=L(:,m);    %location vector of particle 'm'
   xm=lm(1);   %x coordinate of particle 'm'
  X=[xm;1*sin((xm/0.3)*2*pi/11);1*cos((xm/0.3)*2*pi/11)];
   if norm(X-lm)<=rm+1;    %call elcol3d or elcol3ds
%       [V(:,m)]=elcol3d(V(:,m),[0;0;0],M(m),10*M(m),lm,X);
      [V(:,m)]=elcol3ds(V(:,m),[0;0;0],M(m),10*M(m),lm,X);
   end
   end
end
```

```
function [G]=cytgraphics(Lct,Vol,Trt)
%  Graphics
%This function visualize particles
%Author:Piros Sandor
%Al Ain Jan 2008
%
M=size(Lct,3);
R=3.5;  %radius of tRNA
for m=1:M
   set(gca,'NextPlot','replacechildren')
[xx yy zz] = sphere;
s = surf(10*xx,10*yy,10*zz);  %radius of ribosome=10
set(s, 'EdgeColor', 'r', 'FaceColor', 'none');
axis([-Vol(1) Vol(1) -Vol(2) Vol(2) -Vol(3) Vol(3)]);
axis on;
set(gca, 'DataAspectRatio' , [1  1 1]);
light;
set(s, 'LineWidth', 6)
hold on;
[xx yy zz] = sphere;
N=size(Lct,2);
for n=1:N
s(n)=surf(xx*R+Lct(1,n,m), yy*R+Lct(2,n,m), zz*R+Lct(3,n,m));
set(s(n), 'CData', rand(21), 'FaceColor', 'interp');
end
colormap(cool(100));
lighting phong;
set(gca, 'CameraViewAngle',  7);
set(gcf, 'color', [1 1 1]);
G(m)=getframe;
end
%movie(G)
```

```
function [G]=rnagraphics(Lct,Vol,Trt)
%  Graphics
%This function visualize particles and
%the mRNA with the ribosome
%Author:Piros Sandor
%Al Ain Jan 2008
%
M=size(Lct,3);
Rt=3.5;  %radius of tRNA
Rm=1;  %radius of mRNA helix
for m=1:M
   set(gca,'NextPlot','replacechildren')
[xx yy zz] = sphere;
s = surf(10*xx,10*yy,10*zz);  %radius of ribosome=10
set(s, 'EdgeColor', 'r', 'FaceColor', 'none');
axis([-Vol(1) Vol(1) -Vol(2) Vol(2) -Vol(3) Vol(3)]);
axis on;
set(gca, 'DataAspectRatio' , [1  1 1]);
light;
set(s, 'LineWidth', 6)
hold on;
[xx yy zz] = sphere;
X=int16(2*Vol(1)/0.3);
for x=1:X;
   p=double(x);
sm(x)=surf(xx*Rm-Vol(1)+0.3*p, yy*Rm+1*sin(p*2*pi/11),
zz*Rm+1*cos(p*2*pi/11));
set(sm(x), 'CData', rand(21), 'FaceColor', 'interp');
end
N=size(Lct,2);
for n=1:N
s(n)=surf(xx*Rt+Lct(1,n,m), yy*Rt+Lct(2,n,m), zz*Rt+Lct(3,n,m));
set(s(n), 'CData', rand(21), 'FaceColor', 'interp');
```

```
end
colormap(cool(100));
lighting phong;
set(gca, 'CameraViewAngle',  7);
set(gcf, 'color', [1 1 1]);
G(m)=getframe;
end
%movie(G)
```

Appendix B.

Simulation Results

B.a. Results of the Chapter 5.1.1. Simulations

5.1.1. Results

1. Initial conditions:
   Number of tRNA=1600
   Number of ctRNA=40
   Number of ribosomes=144
   Number of other particles=5 000
   time interval=20 ms
   time step=0.1 ms

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
0	0	1	1	0	0	1	0	0	0	0.3

2. Initial conditions:
   Number of tRNA=1600
   Number of ctRNA=40
   Number of ribosomes=144
   Number of other particles=10 000
   time interval=20 ms
   time step=0.1 ms

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
0	0	0	0	1	0	0	0	0	0	0.1

3. Initial conditions:

Number of tRNA=1600

Number of ctRNA=40

Number of ribosomes=144

Number of other particles=1 000

time interval=20 ms

time step=0.1 ms

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
0	0	1	0	1	0	1	0	0	0	0.3

4. Initial conditions:

Number of tRNA=1600

Number of ctRNA=40

Number of ribosomes=144

Number of other particles=20 000

time interval=20 ms

time step=0.1 ms

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
0	0	0	0	1	0	0	1	1	0	0.3

5. Initial conditions:

Number of tRNA=1600

Number of ctRNA=40

Number of ribosomes=144

Number of other particles=10 000

time interval=100 ms

time step=0.1 ms

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
1	0	1	1	2	1	1	0	1	1	0.9

6. Initial conditions:

Number of tRNA=1600

Number of ctRNA=80

Number of ribosomes=144

Number of other particles=1 000

time interval=20 ms

time step=0.1 ms

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
1	0	0	0	0	0	0	1	0	0	0.2

7. Initial conditions:

Number of tRNA=1600

Number of ctRNA=120

Number of ribosomes=144

Number of other particles=1 000

time interval=20 ms

time step=0.1 ms

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
1	1	0	0	1	0	0	0	1	0	0.4

8. Initial conditions: Velocity of particles was doubled.

Number of tRNA=1600

Number of ctRNA=40

Number of ribosomes=144

Number of other particles=1 000

time interval=20 ms

time step=0.1 ms

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
1	0	0	0	0	0	0	0	0	1	0.2

9. Initial conditions: 10*velocity for tRNA's only

Number of tRNA=1600

Number of ctRNA=40

Number of ribosomes=144

Number of other particles=1 000

time interval=20 ms

time step=0.1 ms

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
1	1	1	4	2	3	1	1	4	2	2.0

10. Initial conditions:

Number of tRNA=1600

Number of ctRNA=40

Number of ribosomes=144

Number of other particles=1 000

time interval=20 ms

time step=0.05 ms

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
0	0	0	0	0	0	0	0	1	0	0.1

B.b. Results of the Chapter 5.2.1. Simulations

5.2.1. Results

1. Initial conditions:
   Number of tRNA=1600
   Number of ctRNA=40
   Number of ribosomes=144
   Number of other particles=5 000
   time interval=20msec
   time step=0.1msec

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
0	2	0	1	1	1	2	1	0	0	0.8

2. Initial conditions:
   Number of tRNA=1600
   Number of ctRNA=40
   Number of ribosomes=144
   Number of other particles=10 000
   time interval=20msec
   time step=0.1msec

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
0	1	2	0	0	1	1	0	1	0	0.6

3. Initial conditions:

Number of tRNA=1600

Number of ctRNA=40

Number of ribosomes=144

Number of other particles=1 000

time interval=20msec

time step=0.1msec

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
0	2	2	0	0	1	0	0	0	1	0.6

4. Initial conditions:

Number of tRNA=1600

Number of ctRNA=40

Number of ribosomes=144

Number of other particles=20 000

time interval=20msec

time step=0.1msec

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
1	1	0	2	1	1	0	0	0	0	0.6

5. Initial conditions:

Number of tRNA=1600

Number of ctRNA=40

Number of ribosomes=144

Number of other particles=10 000

time interval=100msec

time step=0.1msec

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
3	5	2	1	1	1	1	1	1	0	1.6

6. Initial conditions:

Number of tRNA=1600

Number of ctRNA=80

Number of ribosomes=144

Number of other particles=1 000

time interval=20msec

time step=0.1msec

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
1	0	2	0	3	1	0	0	1	0	0.8

7. Initial conditions:

Number of tRNA=1600

Number of ctRNA=120

Number of ribosomes=144

Number of other particles=1 000

time interval=20msec

time step=0.1msec

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
2	2	3	1	0	1	1	0	1	1	1.2

8. Initial conditions: Velocity of particles was doubled.

Number of tRNA=1600

Number of ctRNA=40

Number of ribosomes=144

Number of other particles=1 000

time interval=20msec

time step=0.1msec

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
0	1	1	1	1	0	2	0	0	0	0.6

9. Initial conditions: 10*velocity for tRNA's only

Number of tRNA=1600

Number of ctRNA=40

Number of ribosomes=144

Number of other particles=1 000

time interval=20msec

time step=0.1msec

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
1	1	3	4	4	1	2	4	0	2	2.2

10. Initial conditions:

Number of tRNA=1600

Number of ctRNA=40

Number of ribosomes=144

Number of other particles=1 000

time interval=20msec

time step=0.05msec

Attempts										Average
1	2	3	4	5	6	7	8	9	10	
1	0	2	0	0	0	0	0	1	0	0.4

Appendix C.

3D COLLISION BY THOMAS SMID

“(9) $v_{z,2}' = v_{z,2} + \Delta v_{z,2}'$,

(10) $\mathbf{v_{x,2}'} = v_{x,2} + \tan(\theta)\cdot\cos(\varphi)\cdot\Delta v_{z,2}'$,

(11) $\mathbf{v_{y,2}'} = v_{y,2} + \tan(\theta)\cdot\sin(\varphi)\cdot\Delta v_{z,2}'$,

(12) $\mathbf{v_{z,}}_1' = v_{z,1} - m_2/m_1\cdot\Delta v_{z,2}'$,

(13) $\mathbf{v_{x,}}_1' = v_{x,1} - \tan(\theta)\cdot\cos(\varphi)\cdot m_2/m_1\cdot\Delta v_{z,2}'$,

(14) $\mathbf{v_{y,1}'} = v_{y,1} - \tan(\theta)\cdot\sin(\varphi)\cdot m_2/m_1\cdot\Delta v_{z,2}'$.

Note that in contrast to the corresponding 2D equations, Eqs.(8)-(14) are still formulated in a coordinate system where the initial velocity of ball 2 is zero. This is because the angles θ and φ still have to be specified. The problem here is that the angles for the position and velocity vectors can not be simply added as in the 2D case as the vectors do in general not lie in a coordinate plane. One can therefore not characterise the angles by the corresponding arctan function but has to choose a different approach: the easiest way to define the scattering geometry for the 3D case is to choose the coordinate system such that the relative position vector coincides with the z-axis, i.e. ball 1 is at the origin (moving with the relative velocity of both balls) and ball 2 rests on the z-axis at a point corresponding to the distance between both balls). This involves not only subtracting the velocity of ball 2 (as implicitly done already above in Eqs.(1)-(4)) and subtracting the position vector of ball 1, but in addition a corresponding rotation of the coordinate system (in fact two successive rotations about the y and z axis). With these transformations, the required angles are then directly given by the polar coordinates of the velocity vector of ball 1. Applying the above equations and reversing the coordinate transformations gives then the velocities after the collision in the original coordinate system. The positions can then also be updated analogously to the 2D-case. “ [20]

Appendix D.

E.coli STATISTICS

Table 11. *E.coli* Statistics

General Statistics		
1	Cell length	2 µm or $2x10^{-6}$ m
2	Cell diameter	0.8 µm or $0.8x10^{-6}$ m
3	Cell total volume	$1x10^{-15}$ L or $1x10^{-18}$ m^3 (other est. at $0.88x10^{-15}$ L)
4	Cell aqueous volume	7 x 10^{-16} L
5	Cell surface area	$6x10^{-12}$ m2
6	Cell wet weight	1x10-15 kg or $1x10^{-12}$ g
7	Cell dry weight	$3.0x10^{-16}$ kg or $3.0x10^{-13}$ g
8	Periplasm volume	$6.5x10^{-17}$ L
9	Cytoplasm volume	$6.7x10^{-16}$ L
10	Envelope volume	$1.6x10^{-16}$ L
11	Nuclear (DNA+protein) volume	$1.6x10^{-16}$ L
12	Inner Membrane thickness	$8x10^{-9}$ m
13	Outer Membrane thickness	$8x10^{-9}$ - $15x10^{-9}$ m
14	Periplasm thickness	$1x10^{-8}$ m
15	Average size of protein	360 residues
16	Average diameter of ave. protein	5 nm
17	Average MW of protein	40 kD
18	Average prot. oligomerization state	4 proteins/complex
19	Average MW of protein entity	160 kD

20	Average size of mRNA	1100 bases
21	Average length of mRNA	370 nm
22	Average MW of all RNAs	400 kD
23	Average MW of single DNA	$3.0x10^9$ D or $3.0x10^6$ kD
24	Average MW of all DNA	$7 x 10^6$ kD
25	Average length of DNA (chrom.)	1.55 mm
26	Diameter of chromosome	490 µm
27	Diameter of condensed chromosome	17 µm
28	Spacing between small organics	3.6 nm/molecule
29	Spacing between ions	2.1 nm/molecule
30	Ave. spacing between proteins	7 nm/molecule
31	Spacing between protein entities	9 nm/molecule
32	Mean Velocity of 70 kD protein (cytoplasm)	3 nm/ms = $3x10^{-6}$ m/s
33	Mean Velocity of 40 kD protein (cytoplasm)	5 nm/ms = $5x10^{-6}$ m/s
34	Mean Velocity of 30 kD protein (cytoplasm)	7 nm/ms = $7x10^{-6}$ m/s
35	Mean Velocity of 14 kD protein (cytoplasm)	10 nm/ms = $10x10^{-6}$ m/s
36	Mean Velocity of small molecules (cytoplasm)	50 nm/ms = $5x10^{-5}$ m/s
37	Mean Velocity of protein in H2O	27 nm/ms = $2.7x10^{-5}$ m/s
38	Mean Velocity of small molecules in H2O	87 nm/ms = $8.7x10^{-5}$ m/s
39	Concentration of protein in cell	200-320 mg/mL (5-8 mM)
40	Concentration of RNA in cell	75-120 mg/mL (0.5-0.8 mM)
41	Concentration of DNA in cell	11-18 mg/mL (5 nM)
42	Volume occupied by water	70%
43	Volume occupied by protein	17%
44	Volume occupied by all RNA	6%
45	Volume occupied by rRNA	5%

46	Volume occupied by tRNA	0.8%
47	Volume occupied by mRNA	0.2%
48	Volume occupied by DNA	1%
49	Volume occupied by ribosomes	8%
50	Volume occupied by lipid	3%
51	Volume occupied by LPS	1%
52	Volume occupied by murein	1%
53	Volume occupied by glycogen	1%
54	Volume occupied by ions	0.3%
55	Volume occupied by small organics	1%
56	Translation rate	40 aa/sec
57	RNA polymerase transcription rate	70 nt/sec

Large Molecule Copy Numbers		
1	Number of cell walls/cell	1
2	Number of membranes/cell	2
3	Number of chromosomes/cell	2.3 (at mid log phase)
4	Number of mRNA/cell	4000
5	Number of rRNA/cell	18,000
6	Number of tRNA/cell	200,000
7	Number of all RNA/cell	222,000
8	Number of polysaccharides/cell	39,000
9	Number of murein molecules/cell	240,000-700,000
10	Number of lipopolysaccharide/cell	600,000
11	Number of lipids/cell	25,000,000
12	Number of all lipids/cell	25,000,000
13	Number of phosphatidylethanolamine	18,500,000
14	Number of phosphatidylglycerol	5,000,000
15	Number of cardiolipin	1,200,000
16	Number of phosphatidylserine	500,000
17	Number of LPS (MW = 10kD)	600,000
18	Average SA of lipid molecule	25 Ang2
19	Fraction of lipid bilayer=lipid	40%
20	Fraction of lipid bilayer=protein	60%
21	Number of outer membrane proteins	300,000
22	Number of porins (subset of OM)	60,000
23	Number of lipoproteins (OM)	240,000

24	Number of inner membrane proteins	200,000
25	Number of nuclear proteins	100,000
26	Number of cytoplasmic proteins	1,000,000 (excluding ribo proteins)
27	Number of ribosomal proteins	900,000
28	Number of periplasmic proteins	80,000
29	Number of all proteins in cell	2,600,000
30	Number of external proteins (flag/pili)	1,000,000
31	Number of all proteins	3,600,000

Statistics on Larger Molecule Complexes		
1	Number of protein types to make flagella	42
2	Length of flagella	10-20 μm or ~15,000 nm
3	Diameter of flagella	25 nm
4	Number of protofilaments in flagellum	11
5	Diameter of each fliC monomer	5 nm
6	Number of fliC monomers in filament	3000x11=33,000
7	Number of flagella/cell	10
8	Number of fliC proteins	330,000
9	Speed at which *E. coli* move	50 um/sec = 18 $x10^{-5}$ km/h
10	Number of protein types to make pilus	1
11	Length of pili/fimbrae	200-2000 nm
12	Diamter of pili	6.5 nm
13	Number of papA/nm pilus	1.5
14	Number of papA monomers/pilus	3000-30,000
15	Number of pili/cell	100-300
16	Number of papA/cell	300,000-900,000
17	Number of ribosomes/cell	18,000
18	Number of protein types to make ribosome	55
19	Number rRNA types to make ribosome	3
20	Number of proteins in 30S subunit	21
21	Number of proteins in 50S subunit	34
22	Number of rRNA in 30S subunit	1
23	Number of rRNA in 50S subunit	2

24	Length of all rRNA	5520 nt
25	MW of ribosome	2700 kD
26	MW of RNA component	1700 kD
27	MW of protein component	1000 kD
28	Diameter of ribosome	20 nm
29	Volume of ribosome	4.2×10^{-24} m^3

Small Molecule Copy Numbers		
1	Number of water molecules/cell	2.34x1010 (23.4 billion)
2	Number of ions/cell	120,000,000 (300 mM)
3	Number of small organics/cell	18,000,000 (40-50 mM)
4	Number of K ions	90,000,000 (200-250 mM)
5	Number of Na ions	2,000,000 (5 mM)
6	Na (in): Na (out)	1:20 (in concentration)
7	Number of Ca ions	2,300,000 (6 mM)
8	Number of free Ca ions	40 (100 nM)
9	Number of Cl ions	2,400,000 (6 mM)
10	Number of Mg ions	4,000,000 (10 mM)
11	Number of Fe ions	7,000,000 (18 mM)
12	Number of Mn ions	1,700,000 (4 mM)
13	Number of Zn ions	1,700,000 (4 mM)
14	Number of Mo ions	1,700,000 (4 mM)
15	Number of Cu ions	1,700,000 (4 mM)
16	Number of PO4 ions	2,000,000 (5 mM)
17	Number of glucose/cell	200,000-400,000 (0.5-1 mM)
18	Number of PEP/cell	1,100,000 (2.8 mM)
19	Number of pyruvate/cell	370,000 (0.9 mM)
20	Number of gluc-6-PO4/cell	20,000 (0.05 mM)
21	Number of ATP/cell	500,000 – 3,000,000 (1.3-7.0 mM)
22	Number of ADP/cell	70,000 (0.17 mM)
23	Number of NADP/cell	240,000 (0.63 mM)

24	Number of NADPH/cell	220,000 (0.56 mM)
25	Number of all amino acids/cell	6,000,000 (1.5 mM)
26	Number of free Alanine/cell	350,000 (0.8 mM)
27	Number of free Cysteine/cell	80,000 (0.2 mM)
28	Number of free Aspartate/cell	530,000 (1.34 mM)
29	Number of free Glutamate/cell	200,000 (0.5 mM)
30	Number of free Phenylalanine/cell	170,000 (0.4 mM)
31	Number of free Glycine/cell	350,000 (0.8 mM)
32	Number of free Histidine/cell	80,000 (0.2 mM)
33	Number of free Isoleucine/cell	200,000 (0.5 mM)
34	Number of free Lysine/cell	190,000 (0.46 mM)
35	Number of free Leucine/cell	300,000 (0.7 mM)
36	Number of free Methionine/cell	40,000 (0.1 mM)
37	Number of free Asparagine/cell	200,000 (0.5 mM)
38	Number of free Proline/cell	200,000 (0.5 mM)
39	Number of free Glutamine/cell	200,000 (0.5 mM)
40	Number of free Arginine/cell	170,000 (0.4 mM)
41	Number of free Serine/cell	300,000 (0.7 mM)
42	Number of free Threonine/cell	1,400,000 (3.49 mM)
43	Number of free Valine/cell	240,000 (0.6 mM)
44	Number of free Tryptophan/cell	80,000 (0.2 mM)
45	Number of free Tyrosine/cell	300,000 (0.7 mM)
46	Osmotic pressure (pushing out)	75 lb/in2

E. coli Metabolism		
1	1 glucose generates (total)	36-38 ATP
2	glycolysis yields	6-8 ATP
3	oxidation of pyruvate yields	6 ATP
4	Krebs cycle/e- transport yields	24 ATP
5	Number ATP to make 1 DNA	72,289,000
6	Number ATP to make 1 protein (360 aa)	1500
7	Number ATP to make 1 lipid	7
8	Number ATP to make 1 polysaccharide	2000
9	Number ATP to make 1 RNA (1000 nt)	2000
10	Number ATP to make 1 cell	55 billion ATP
11	Number Glucose molecules consumed	1.4 billion molecules
12	Cell division rate	1 division/30 minutes

Pasted from [18].