İSTANBUL TECHNICAL UNIVERSITY

FACULTY OF SCIENCE AND LETTERS

GRADUATION PROJECT

QUANTUM ALGORITHMS AND GENETIC CODE

Onur Pusuluk

Department : Physical Engineering

Supervisors : Assist. Prof. Dr. Oğuzhan Gürlü, İ.T.U.

Assoc. Prof. Dr. Cemsinan Deliduman, M.S.G.S.U.

2009 SPRING SEMESTER

SUMMARY

The research we present in this thesis is devoted to the modelling and understanding of complex biological realities from a bioinformatic perspective in which simple physical concepts are used. In the mean time, some chemical realities such as catalytic activity, are included to the field of physics and knowledge about the nature of catalyzers is tried to be boradened beyond the ones obtained with perspective of modern chemistry.

In this sense, each step of biopolymer formation, like the ones in DNA replication or protein synthesis, is defined as an unsorted database search. Then, it is shown that Grover's Algorithm is optimum algorithm for this task and alphabet lengths of genetic and proteomic languages coincide with the distinguishable element numbers by one quantum query and three quantum queries respectively. Since these numbers are equal to the base pairing numbers required in replication and transcription, quantum nature of base pairing is studied. In the meantime, several scenarios are constructed on the basis of quantum search algorithm and possible roles of enzymes (biological catalyzers) in these scenarios are examined. It is suggested that active site of the enzymes can provide an environment which minimizes the decohorent fluctuations of cellular environment. Also, quantum mechanical scenario reveals that enzymes can create superpositions of chemically distinct molecules and enable tunnelling through energy barrier rather than lowering activation energy. Morever, an alternative scenario based on wave dynamics introduces that enzymes can make energy transfer from several interactions to one interaction.

Finally, proteomic language is considered as a structural language. Proper structure and maximum distinguishable number of building blocks is searched by lattice modelling. It is suggested that primitive proteomic alphabet could consist of ten amino acids which correspond to present class II amino acids and predecessor of the genetic code for these amino acids could be a doublet code. It is also suggested that present triplet code and proteomic alphabet containing common amino acids could evolve by duplication and divergence of this doublet code. This divergence could be directed in a way which increases the stability of proteins.

ACKNOWLEDGEMENTS

I would like to thank to my supervisors Assist. Prof. Dr. Oğuzhan Gürlü and Assoc. Prof. Dr. Cemsinan Deliduman for their many contributions to my life, both in the scientific sense and personally. They guided me toward a direction where physics become more interesting and itegrated into actual life. Also, they trusted me and they appreciated my workings. Working with them has been a great luck and pleasure for me.

I would also like to thank to Berkin Malkoç who is one of the most objective people I have seen. I think everyone needs such a friend with whom a healthy discussion about any subject is always available.

My beautiful wife has recently helped me alot in the emotional sense and made my life easier. I would also like to thank to my parents for those they have/haven't been doing and for always being with me. I am very lucky that I have such a beauiful family.

When you actually dedicate yourself to a research, age of the research is equal to your age and your whole life has an effect on its value. Therefore, I would also like to thank to everything in my life that contributes to my personality.

Onur Pusuluk

May 2009

CONTENT

ABBREVIATIONS

DNA	: Deoxyribonucleic acid
RNA	: Ribonucleic acid
mRNA	: Messenger ribonucleic acid
tRNA	: Transfer ribonucleic acid
rRNA	: Ribosomal ribonucleic acid
bp	: Base pair
aa	: Amino acid
aaRS	: Aminoacyl–tRNA synthetases
OH	: Hydroxyl group
COO^-	: Carboxyl group
H_3N^+	: Amino group

LIST OF SYMBOLS

A : Adenine base

G : Guanine base

C : Cytosine base

T : Thymine base

U : Uracile base

I : Inosine base

R : Purine bases, adenine or guanine

Y : Pyrimidine bases, cytosine, thymine or uracile

N : Anyone of the bases

: Nitrogen atom (contexual usage)

H : Hydrogen atom

O : Oxygen atom

C : Carbon atom

Mg : Magnesium atom

ψ : Dihedral angle around C_α–C bond in polpeptide backbone

ϕ : Dihedral angle around N–C_α bond in polpeptide backbone

θ : Any arbitrary angle

Å : Angstrom

$\hbar$: Planck constant

N : Element number of a database

Q : Minimum query number required to find a desired element

$Q_{1/2}$: Required query number to find desired element with probability ½

k : Any arbitrary variable

i : Desired variable

p_i : Probability to find desired element in the query i

Ω : Probability to find desired element in an independent query

$|s\rangle$: Symmetric superposition state

$|k\rangle$: Any arbitrary state

$|i\rangle$: Desired state

$|i_\perp\rangle$: Symmetric superposition of the eigenstates different from state $|i\rangle$

G : Operator for Grover's Algorithm

O : Oracle operator in Grover's Algorithm

D : Diffusion operator in Grover's Algorithm

$W^{\otimes n}$: Hadamard Matrix for n–qubits

R : Rotation matrix

I : Identity matrix

α : Probability amplitude

φ : Relative phase

H : Hamiltonian

E : Energy

U : Unitary time evolution operator

t : Time

1 INTRODUCTION

1.1 GENETIC CODE AS A COMPLEX BIOLOGICAL REALITY

All the living things have a hereditary material that is responsible for three different functions: replication, gene expression, and mutation. [1]

Replication is the genotypic function of the hereditary material. Hereditary material should contain genetic information which can be passed from the organisms to their offsprings with a minumum error. Gene expression is the phenotypic function of the hereditary material. Hereditary material should regulate the metabolism of the organism and should determine the phenotype of the organism during its all life. Mutation is the evolutionary function of the hereditary material. Hereditary material should be avaliable to change in order to increase the survivability of the organism by reconstruction of itself in different ways. [1]

Chromosomes which consist of both nucleic acids and proteins, are the hereditary biochemical materials. However, there is no direct relation between proteins and storage or change of hereditary information. Proteins help nucleic acids in only minimization of storage volume and regulation of gene expression. [1, 2]

1.1.1 Nucleotides as Building Blocks of Nucleic Acids

Fundamental subunits of nucleic acids are called as nucleotides which also have several important functions in the metabolism. All the nucleotides are made up from three basic parts: [1, 2]

- A phosphate group.

- A five-carbon sugar such as ribose in ribonucleic acids and deoxyribose in deoxyribonucleic acids.

- A nitrogenous base such as adenine (A), guanine (G), cytosine (C), thymine (T) and uracile (U). Ribobucleic acid (RNA) does not generally contain T, whereas deoxyribonucleic acid (DNA) does not generaly contain U. In addition to these major bases, some minor bases that are modificated forms of major basis, can be found in nucleic acids for the purpose of regulation or protection.

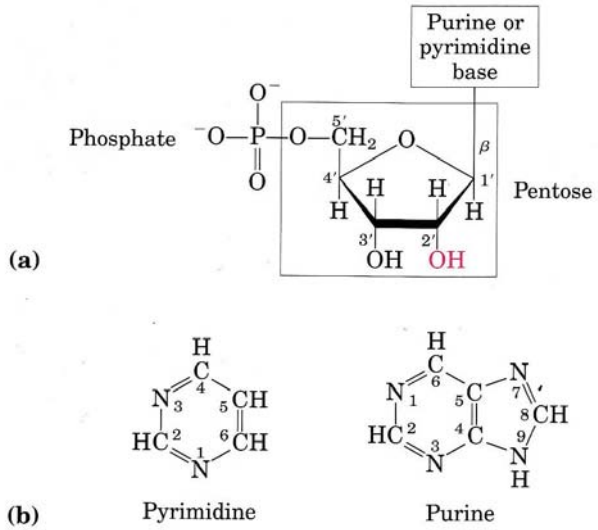

Figure 1.1 Structure of nucleotides (a) This figure shows the structure of a ribonucleotide. Deoxyribonucleotides contain a –H group on the 2′ carbon instead of a –OH group. (b) This figure shows the classifications of the bases. A an G are derivatives of purine (R) compound, whereas C, T and U are derivatives of pirimidine (Y) compound. [2]

Structure of nucleotides is not static and the evolutionary function of hereditary material depends on this fact. For example, a hydrogen atom of a nucleotide can change its position from an amino group to a ring nitrogen. Such random, reversible and infrequent conversions in the covalent structure of nucleotides are called as tautomeric shifts. [1]

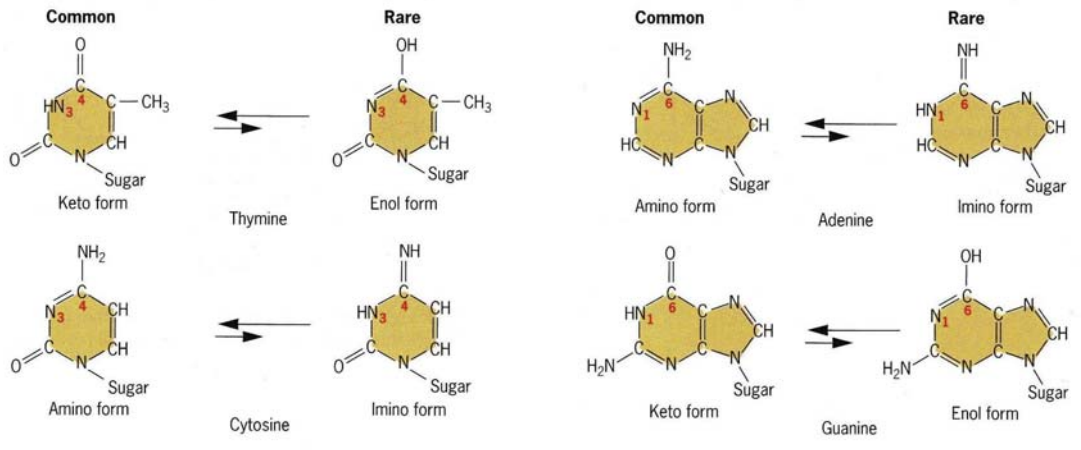

Figure 1.2 Tautomeric forms of nitrogenous bases G and T are the bases which are most stable in keto forms, whereas A and C occur commonly in amino form. [1]

In addition to this proton delocalization, nitrogenous bases can lose their exocylic amino groups via a random process called as deamination. For example, one of each 10^7 cytosine residue in DNA randomly turns into the uracil for every day and DNA has approximately 10^9 cytosine residues. These uracil residues are turned into cytosines by enzymes and if uracil was already contained by DNA, enzymes would not have the ability to distinguish pre-existing uraciles from others. In fact, this is the reason of uracil's exclusion from the DNA which is responsible for long-term storage of hereditary information. So, starting to use tymine as the fourth base instead of uracil is one of the most important turning points during evolution [2].

Methylation is another process which gives a dynamic structure to nucleotides. Normally, methylation of DNA is used for distinguishing organisms own DNA from the foreign ones. Also, it helps to determine which strand of the DNA is the template during protein synthesis [2].

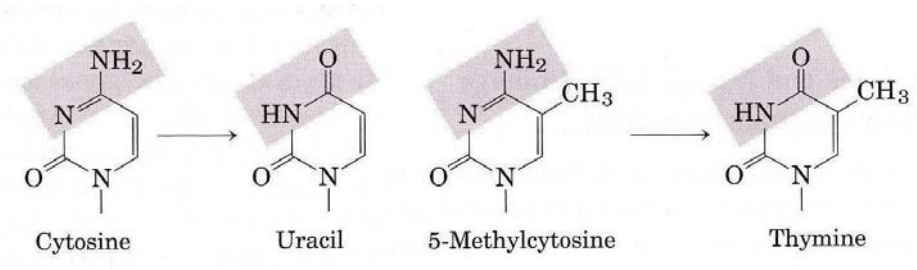

Figure 1.3 Deamination and methylation of nucleotide bases [2] C → U and C → T mutations by random deaminations of C and methylated C is 100–fold higher in single strand DNA of some viruses. [3]

All in all, it can be said that nucleotides have a dynamic covalent structure even when they form the parts of hereditary information.

1.1.2 Nucleic Acid Metabolism

Most common nucleic acid type which includes fundamental units of hereditary information, is Deoxyribonucleic Acid – DNA as indicated previous section. These fundamental units are called "genes", and other parts of DNA which does not encode information are responsible for structure and regulation [1, 2].

DNA is formed by two antiparallel polynucleotide strands and complementarity of these strands gives a right-handed double helix shape to DNA.

Then, negatively charged double helix is wrapped around positively charged proteins which are known as histones. This wrapping makes hereditary material more stable at biological pH by neutralization and gives a 10^4–fold condense tertiary structure. [1]

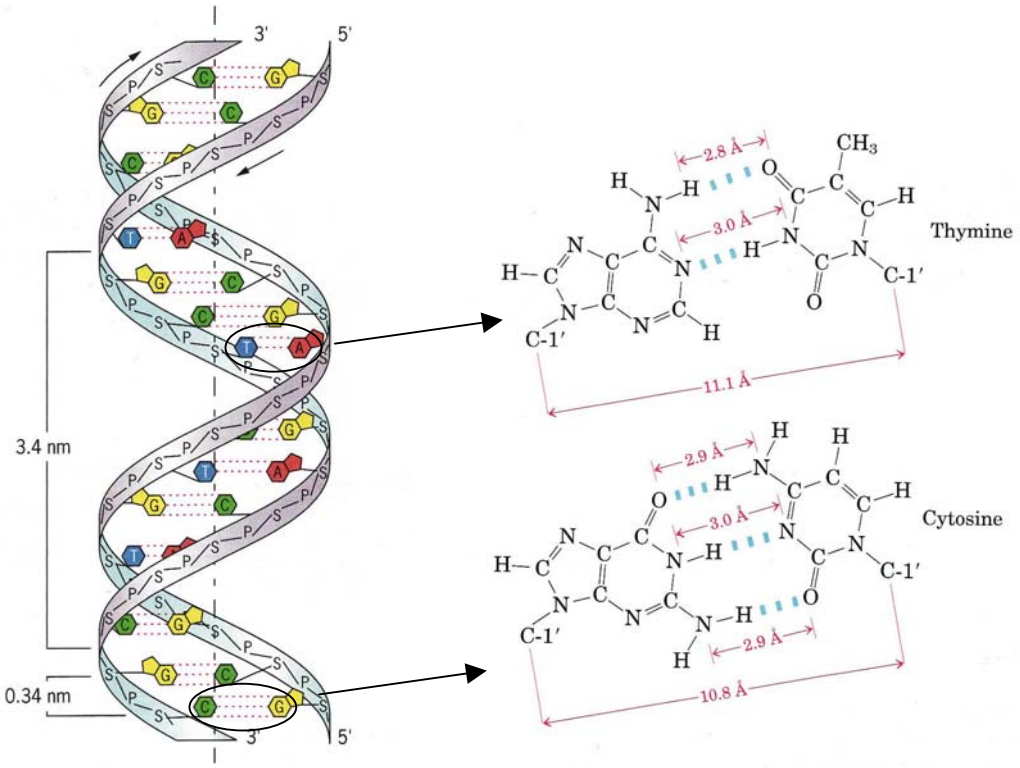

Figure 1.4 Secondary structure of DNA: Each polynucleotide is made up by phosphodiester linkage between nucleotides. Phosphodiester bonds are formed between 5′ phosphate group of one nucleotide and 3′ –OH group of the next one by the activity of DNA polymerase. Thus, covalent backbones of DNA strands elongate in opposite direcions and strands gain opposite polarities. Nitrogenous bases can be thought as side chains of the backbones and they interact each other via hydrogen-bonding. There are two different hydrogen bonds between A and T bases. Similiarly, there are three hydrogen bonds between G and C, but two of them have the same bond length. [1, 2]

The width of a DNA strand changes between 22 – 26 Å. Length of one nucleotide unit in this polynucleotide is 3.3 Å [4] and there are at most 220 million base pairs (bp) in one human chromosome [5]. Nitrogenous bases in the less stable tautomer forms can also join base pairings. Such base pairs cause transition and transversion mutations if they are formed during DNA replication. This is because sequence of the bases encodes the hereditary information and this information is passed from the organisms to their offsprings via DNA replication. [1]

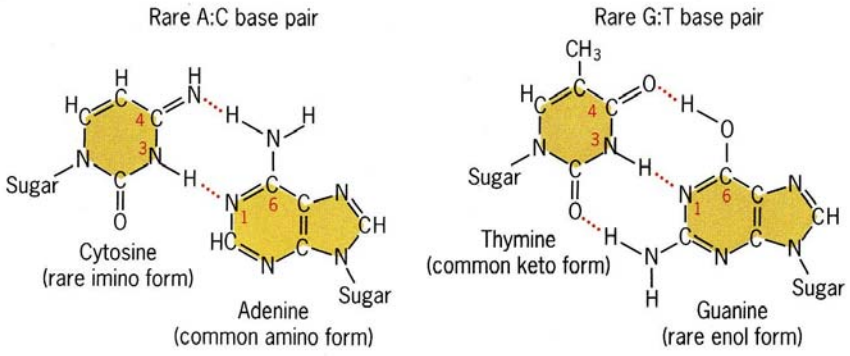

Figure 1.5 Rare base pairs caused by less stable tatutomer forms [1]

During the DNA replication, newly synthesized strands elongate with a rate 3,000 nucleotide per minute in humans. Morever, this rate increase to 30,000 nucleotide per minute in bacteria. [1] In addition to this elongation rate, replication occurs with only one error in each 10^9 to 10^{10} base addition in bacteria *E. coli*. Since bacteria has a chromosome that includes approximately 4.6 x 10^6 bp, it means that only one error occurs during 1,000 to 10,000 replications. [2] In order to optimize elongation rate and accuracy in such a way, each new DNA strand is synthesized by using one of the pre-existing DNA strands as a template. So, duplication of DNA is called as semiconservative replication. [1]

Each pre-existing strand is dublicated by $5' \rightarrow 3'$ activity of DNA polymerase on the complement strand with an error rate of 10^{-4} to 10^{-6} per nuclotide. During the replication, DNA polymerase decreases this rate 10^2- to 10^3-fold by the $3' \rightarrow 5'$ exonuclease activity. In addition to this proofreading mechanism in replication, DNA repair mechanisms continue decreasing error rate after the replication. [2]

There are at least five different DNA polymerase enzymes in bacteria *E. coli*. Activity of these enzymes can be different from each other and also each enzyme can have several active sites for different activities. In addition, DNA polymerase enzymes are not the only ones involved in the replication process. There are at least twenty different enzymes and proteins in the DNA replicase system and each element of this replisome have specific tasks. For example, helicases seperate double helix by breaking hydrogen bonds, whereas topoisomerases decrease the stress caused by the this seperation. Moreover, replications in humans are more complex. [2]

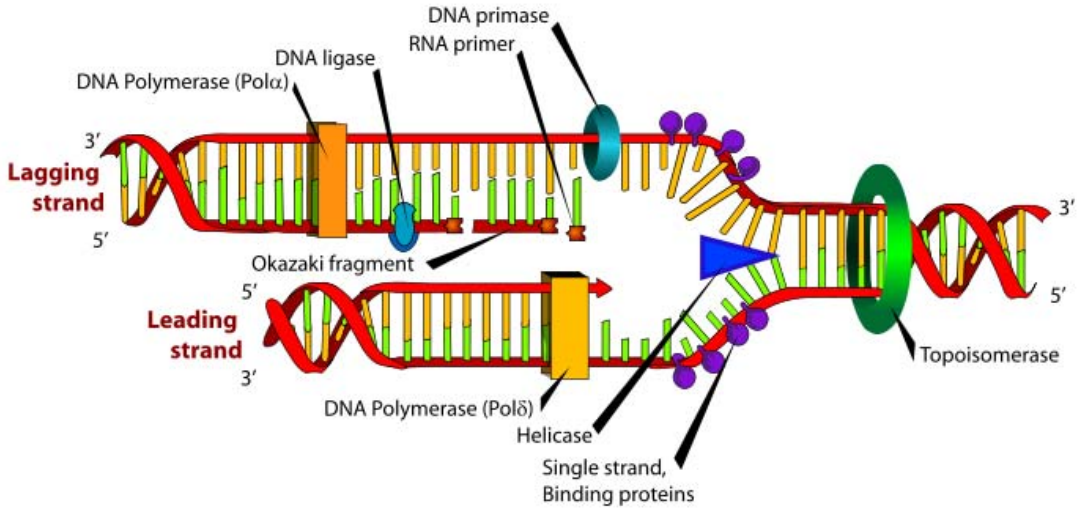

Figure 1.6 Replisome: [6] Elongation of polynucleotides are directional and replication starts at origin points. Thus, elongation of lagging strand can not be continuous as the ones of leading strand. In lagging strand, first primase synthesizes small RNA primers which are called as Okazaki fragments. Then, DNA polymerase replaces these RNA primers with DNA segments by converting ribose sugars into deoxyribose sugars. Finally, DNA ligase joins these segments together. [2]

Hereditary information stored in DNA is transmitted to messenger RNAs (mRNAs) by transcription. Then, this transmitted information is separated from mainly non-informative sequences by small nuclear RNAs (snRNAs) and finally informative sequences are used with the help of adoptor transfer RNAs (tRNAs) and ribosamal RNAs (rRNAs) by translation.

mRNA molecules are synthesized complementary to a template DNA strand by DNA-dependent enzyme RNA polymerase. This complex enzyme takes a free ribonucleoside $5'$-triphospate from the envirenment and adds it as a ribonucleotide to growing RNA in the $5' \rightarrow 3'$ direction with the help of Mg^{2+} molecules. [2]

During the transcription, newly synthesized ribopolynucleotide elongates with a rate 3,000 to 5,400 ribonucleotide per minute in bacteria *E. coli*. This rate is much smaller than the ones in DNA replication. However, accuracy of transcription is also smaller than accuracy of replication. RNAs are synthesized by $5' \rightarrow 3'$ activity of RNA polymerase with an error rate of 10^{-4} to 10^{-5} per ribonuclotide. Absence of a proofreading mechanism based on $3' \rightarrow 5'$ exanuclease activity has an important influnce on this difference in error rates, but this difference makes no sense since mRNA molecule has an half-life less than five minutes. [1, 2]

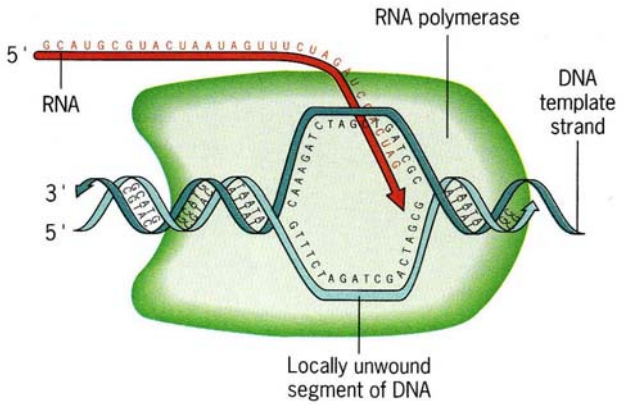

Figure 1.7 DNA-dependent RNA synthesis: In addition to elongate mRNA, RNA polymerase binds to a prometer region of DNA, unwinds double helix about 17 bp at a given time and rewind it after copying DNA sequence in the $3' \rightarrow 5'$ direction. Other types of RNA are also synthesized in the same way. [1, 2]

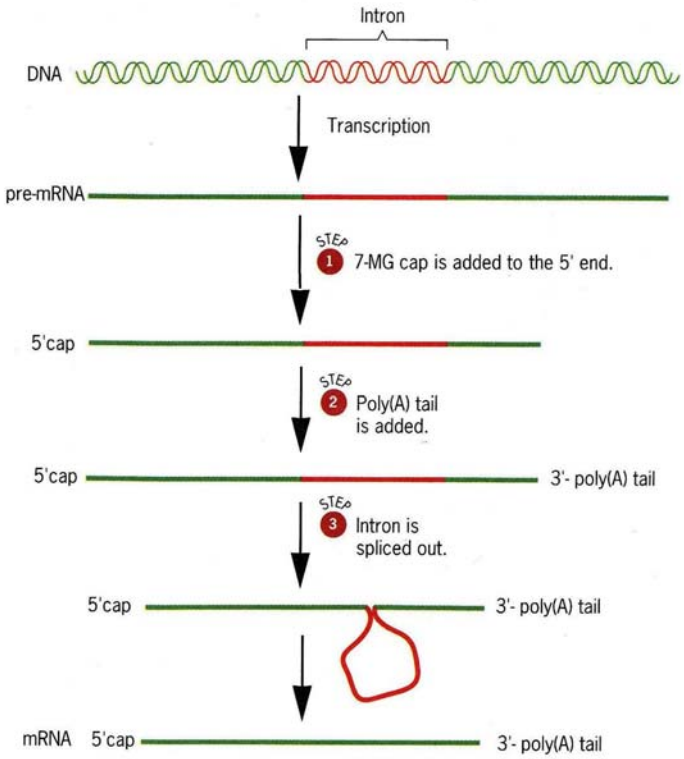

Figure 1.8 Posttranscriptional modifications of mRNA: In eukaryotes, half-life of the transcripts is increased up to five hours by postranscriptional modifications. These modifications are also required for transportation from nucleus to cytoplasm and for splicing mainly non-informative sequences. [1]

1.1.3 Triplet Genetic Code

Transcriped hereditary information is translated into 3–dimensional structural language by protein synthesis in respect to cellular and environmental conditions. During this translation, amino acid (aa) molecules are joined together in an order according to consecutive tree-base sequences in mRNA. These nonoverlapping ribonucleotide base triplets are called as codon and there is no structural comma between them. Also, ribonucleotide base triplets complementary to codons in tRNAs are called as anticodon and their deoxyribonucleotide versions in the template DNA strand are known as code. [1, 2]

There are 64 different codons since each nucleotide in a triplet can be one of the four ribonucleotides. However, three of these codons are stop codons and all the other ones are translated into only twenty amino acids. This degenerate triplet genetic code is present in every living organisms and amino acid meaning of a few codons shows a variety only in mitocondrial mRNAs. [1]

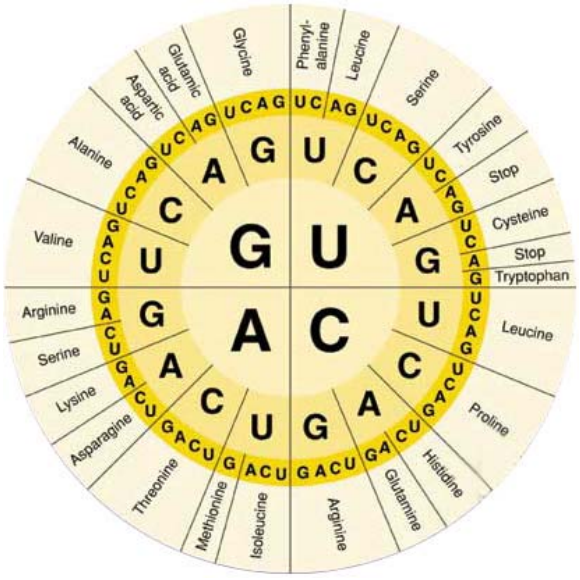

Figure 1.9 Standard Genetic Code: [7] Bases in the inner circle represent the first base of the codon. This figure shows that degeneracy in the genetic code can be classified into two major types with only two exceptions. First type is partial degeneracy in which the third codon can be one of two purines (R) or one of two pyrimidines (Y). Second type is complete degeneracy and the 3rd codon can be any of the four bases (N) since it can not change the amino acid meaning of the codon. [1]

Table 1.1 Known variant codon assigments in mitocondria [2]

	UGA	*AUA*	*AGR*	*CUN*	*CGG*
Standart genetic code	STOP	Isoleucine	Arginine	Leucine	Arginine
Animals					
Vertebrates	Trytophan	Methionine	STOP	+	+
Drosophila	Trytophan	Methionine	Serine	+	+
Yeasts					
Saccharomyces cerevisiae	Trytophan	Methionine	+	Threonine	+
Torulopsis glabrata	Trytophan	Methionine	+	Threonine	?
Schizosaccharomyces pombe	Trytophan	+	+	+	+
Filementous fungi	Trytophan	+	+	+	+
Trypanosomes	Trytophan	+	+	+	+
Higher plants	+	+	+	+	Trytophan
Chlamydomonas reinhardtii	?	+	+	+	?

"+", codon has same meaning with standard one; "?", codon is not observed

The third base of the codon has a binary effect on the amino acid meaning at best. Both of the two exceptions (see Figure 1.9) are removed in vertebrates' mitocondrial code by the alterations shown in Table 1.1. In addition, a non-degenerative one base change in the codon generally causes a new amino acid meaning with similar chemical properties. Therefore, one of the possible interpretations is that genetic code evolved in a way which maximized the error tolerance. [1] Morever, a recent mathematical model that employs this possibility as an assumption, gives a satisfactory explanation of the degeneracy of the genetic code as shown below. [8]

Table 1.2 Topological limit to the number of amino acid meanings [8]

Structure of genetic code	*Discernable base number for 1st base*	*Discernable base number for 2nd base*	*Discernable base number for 3rd base*	*Codon number*	*Maximum aa number*
Singlet	1	4	1	4	4
Doublet	4	4	1	16	11
Triplet	4	4	2	32	16
Triplet	4	4	3	48	20
Triplet	4	4	4	64	25

Results of this bioinformatic model are consistent with the wobble hypothesis which is a common consensus in biochemicstry. This hypothesis is based on the assumption that only the first two bases of the codons have an important effect on the

amino acid meaning. This is because, codon–amino acid connection is physically made by the adoptor tRNA molecules and hydrogen bonds between the first base of anticodon and the third base of codon are weakened by the steric effect. [1, 2]

According to the wobble hypothesis, the number of the codons that can be recognized by a tRNA in protein synthesis, is determined by the first base of the anticodon. For example, first base's being A or C makes hydrogen bonds strong and consequently, only one codon can be recognized. Hydrogen bonds are not so strong when the first base is U or G and this fact increases the recognized codon number to two. If the base inosine (I) is present at the first position, number of recognizable codons becomes three. [2]

Such a weakening effect on codon–anticodon interactions makes protein synthesis faster. Therefore, it can be said that genetic code increases both accuracy and elongation rate in protein synthesis. [2] A similiar optimization was found in DNA replication via semiconservative duplication in previous section.

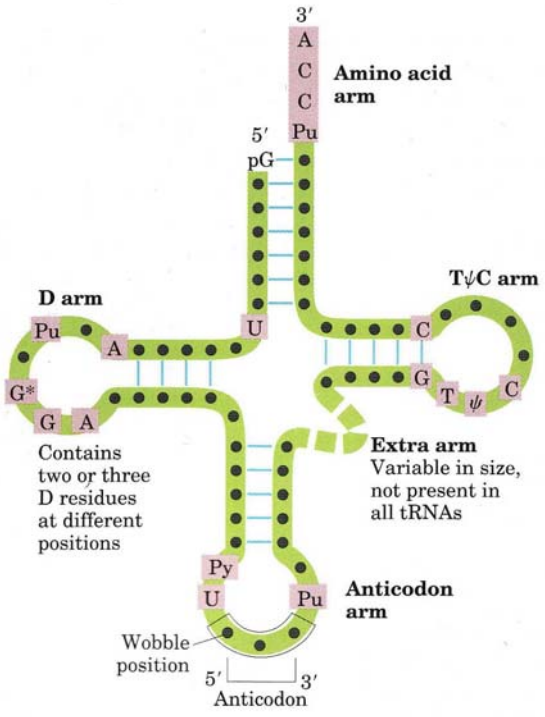

Figure 1.10 Structure of tRNA: There are 73 to 93 nucleotide in the mature structure of tRNA. tRNAs of the organelles mitochondria and chloroplast are special to them and they are smaller than the other ones. Each of 20 amino acids has at least one type of tRNA and there should be more than 32 tRNAs to recognize all codons. [2]

Physical connection of anticodons and amino acids is made via enzymes called as aminoacyl–tRNA synthetases (aaRS). Each of the 20 amino acids has one special aaRS in bacteria. In eukaryotes, cytoplasmic aaRSs and mitochondrial aaRSs are differnt from each other. These aaRS enzymes are classified into two families which have different structures and different activity mechanisms. [2, 9] Analyses show that classes evolved before the split of Bacteria, Archae, and Eucarya [9] and there is no known common ancestor of them [2].

There is no physical interaction between anticodon and enzyme in the recognition of tRNA for a specific amino acid. However, it was shown that the part of tRNA which includes base pairs at the positions 1:72, 2:71, 3:70 and unpaired 73th base, is responsible for successive interactions with right aaRS. This 3–dimensional structure is also called as operational code or second genetic code and it is a candidate for being the predecessor of genetic code. [9]

1.1.4 Protein Metabolism and Enzyme Activity

Although amino acid number is more than 700 in nature, there are 20 amino acids which are made of proteins and they share a common structure as shown in Figure 1.11. Tetrahedral bond structure of central carbon atom gives a chirality to these amino acids except glycine since it has one more hydrogen atom as a R group. As a result of chirality, there are two possible configurations for common amino acids and stereoisomers are named by D and L notational system. Generally, there is a biological selection of one stereoisomer when a biomolecule has a chirality and the dominant amino acids in living systems are L isomers. D isomers are present in some biological structures, but hereditary material does not code for them. [2, 10]

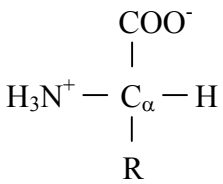

Figure 1.11 Structure of α-amino acids: Amino acid classification is based on the location of amino group on the carbon chain that has carboxylic acid function. All the common amino acids are α–amino acids and the functional group R makes them chemically different from each other. When a carboxyl group ($–COO^-$) of one amino acid and an amino group ($–H_3N^+$) of another one interact, a covalent bond known as peptide bond is formed between amino acids by dehydrolyses reaction. [2, 10]

Chemical properties of a polypeptide depend on the R groups of amino acid ingredients. This is because R groups become side chains of polypeptide backbone after each amino acid is added to the previous one by dehydrolyses and each R group has a distinct size, shape, electric charge, acid-base property, or tendencies to make ionic bonding, covalet bonding, hydrogen bonding, and so on. [2, 10]

Table 1.3 Common† Amino Acids [2]

Name of the Amino acid	*3-letter*	*1-letter*	*Property of the R group*	*Mol. Weight*	*Codon(s) in mRNA code*	*Codon number*	*AaRS class*
Glycine	Gly	G	Nonpolar	75	GGN	4	II
Alanine	Ala	A	Nonpolar	89	GCN	4	II
Proline	Pro	P	Nonpolar	115	CCN	4	II
Valine	Val	V	Nonpolar	117	GUN	4	I
Leucine	Leu	L	Nonpolar	131	UUR, CUN	6	I
Isoleuicine	Ile	I	Nonpolar	131	AUY	2	I
Serine	Ser	S	Polar	105	UCN, AGY	6	II
Threonine	Thr	T	Polar	119	ACN	4	II
Asparagine	Asn	N	Polar	132	AAY	2	II
Cysteine	Cys	C	Polar	121	UGY	2	I
Methionine	Met	M	Polar*	149	AUR	2	I
Glutamine	Gln	Q	Polar	146	CAR	2	I
Aspartate	Asp	D	Negative charge	133	GAY	2	II
Glutamate	Glu	E	Negative charge	147	GAR	2	I
Lysine	Lys	K	Positive charge	146	AAR	2	II
Arginine	Arg	R	Positive charge	174	CGN	4	I
Histidine	His	H	Ring/ Positive charge*	155	CAY	2	II
Phenylalanine	Phe	F	Aromatic Ring	165	UUY	2	II
Tyrosine	Tyr	Y	Aromatic Ring	181	UAY	2	I
Tryptophan	Trp	W	Aromatic Ring	204	UGR	2	I

"*", methionine is a polar structure because of its sulphur atom but standard biochemistry text books take it as nonpolar. Also, charge of the histidine is nearly neutral. [10]

"†", Other amino acids which found in protein structure are generaly derived from common amino acids by posttranslational modifications. However, recently it was shown that selenocysteine and prolysine can be also coded by hereditary material. [2, 10]

Nonpolar R groups generally make polypeptide molecule more stable in water due to their hydrophobic properpities. However, R group which makes polypeptide most flexible belongs to glycine because of its size. Also, nonpolar proline decreases flexibility by its compact R group. [2, 10]

Polar R groups increase solubility in water by their tendency to interact with water molecules. However, polar threonine has a low tendency to make hydrogen bonding, whereas similar amino acid tyrosine does not so. Morever, R groups of cysteine amino acids contain sulphur atoms and so, two of these groups can make a covalent disulfide bond which stabilize polypeptide in a higher level. [2, 10]

Charged R groups have a tendency to make ionic bonds. However, basic R group of histidine has a proton delocalization property like aromatic amino acids. [10]

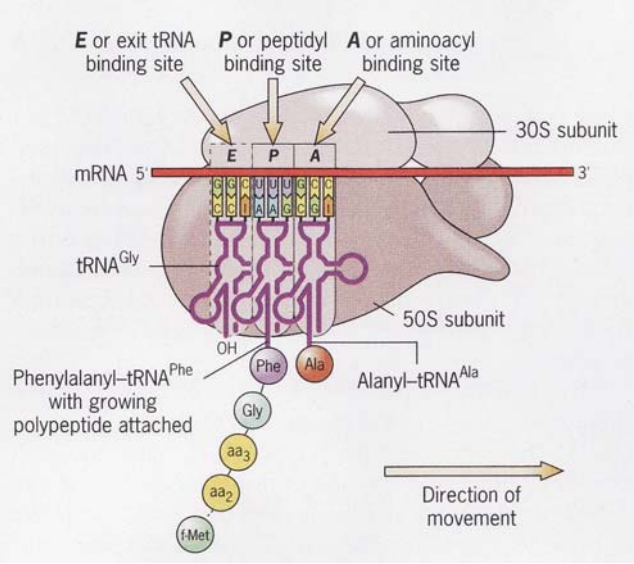

Figure 1.12 Protein synthesis: Proteins are polypeptides that contains thousands of amino acids and synthesized in ribosome. Protein synthesis starts by binding of certain one of two methionine's tRNAs. When an aminoacyl–tRNA goes from *P* site to *E* site of the ribosome by the movement of the ribosome, tRNA leaves from the organel but its cognate amino acid remains to be bound with the previous amino acid. Until a stop codon reaches to *P* site, elongation of peptide continue in this way. [1, 2]

Amino acid sequence of proteins is called as primary structure. After being synthesized, protein spatially rearranges itself by the interactions with in backbone. Most stable rearrangements are α–helix and β–sheet conformations in both of which peptide bonds have a planar geometry. [10]

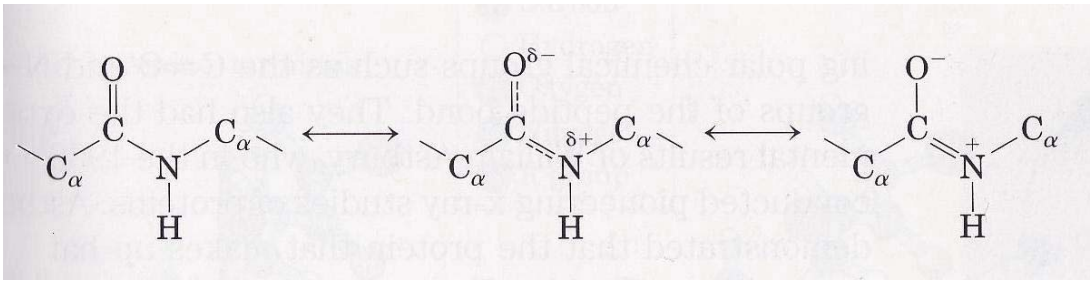

Figure 1.13 Reason of planar peptide bond geometry: Backbone of the protein includes repetitions of N–C_α–C arrangement. C–N bond of a peptide is smaller than the one in free amino acid. This is because these carbonyl oxygen and backbone nitrogen atoms partially share two electron pairs. As a result, C–N bond gains a double bond character that prevents rotations around this bond. Therefore, the six atoms shown in the figure becomes co-planar. [2]

The configuration in right side of Figure 1.13 is known as *trans* configuration. Although 99.95% of all 19 common amino acids are in this form, 6% of proline residues are present in *cis* form in which the plane of N–H and N–C_α bonds turns around nitrogen atom about 180^0. [2]

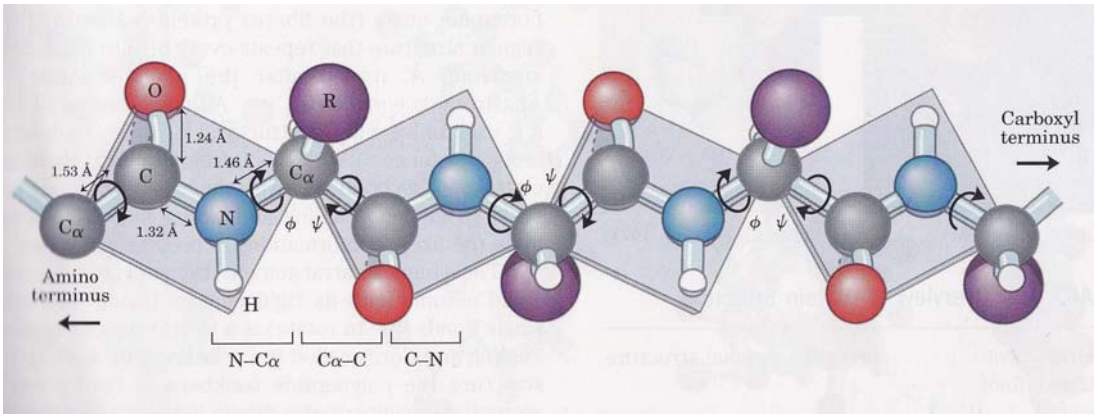

Figure 1.14 Planar geometry of peptide bonds: Dihedral angles around C_α–C and N–C_α bonds are named as ψ and ϕ by a common convention. Their values change between -180^0 and 180^0. Only their being 0^0 is forbidden because it gives an impossible conformation in which carbonyl oxygen of amino acid n and amino hydrogen of amino acid n+2 should overlap. It is assumed that each one of these angles is 180^0 in the fully extended conformation. [2]

In α–helix conformation, (ψ, ϕ) dehidral angels are distributed around the line $\phi = -\psi$ and corresponding rotations are coupled. Individual dehidral angels for each common amino acid except glycine and proline, are mainly focused at the point

(-41.1^0, -63.8^0) with a deviation (± 2, ± 2). Maximum of proline is (-36.5^0, -61.0^0), whereas this point for glycine is (-42.4^0, -59.1^0). [11]

These dehidral angles correspond to a helical structure in which backbone turns around an imaginary axis and side chains stick out from it. Each turn of helix around imaginary axis has 3.6 amino acid residues and a longitudinal length of 5.4 Å. In such a geometry, every partially negative backbone nitrogen atom of amino acid n+4 makes a hydrogen bond with the partially positive carbonyl oxygen atom of amino acid n. This optimal usage of internal hydrogen bonds makes α–helix conformation more stable. [2]

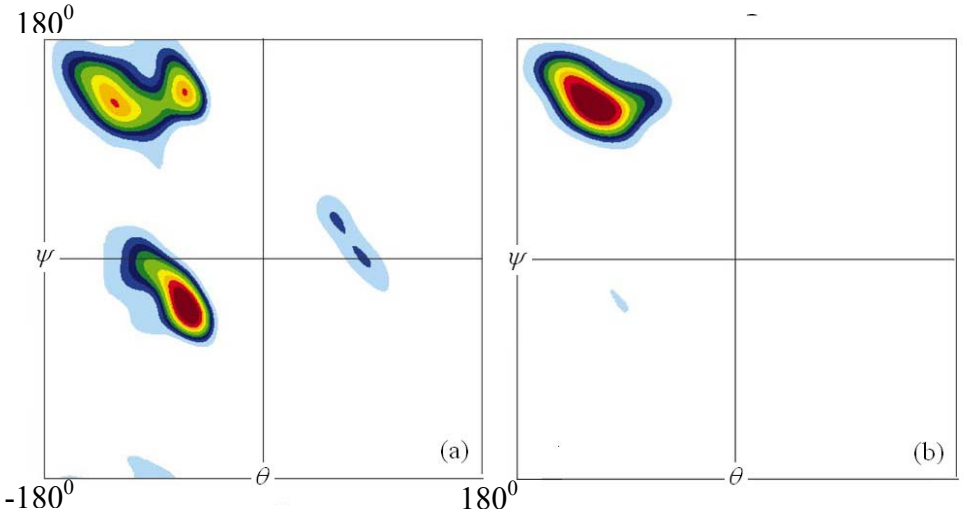

Figure 1.15 Individual Ramachandran Plots for all amino acids: From white to blue and from blue to dark red, amino acid density increases. (a) Ramachandran Plot for amino acids are inside α–helices. (b) Ramachandran Plot for amino acids are inside β–sheets. [11]

In β–sheet conformation, (ψ, ϕ) dehidral angles are distributed in the upper left corner of Ramachandran Plot. Individual dehidral angles for each common amino acid except proline and glycine, are mainly focused at the point (136^0, -123^0) if sheets are antiparallel and at the point (128^0, -116^0) if sheets are parallel. ϕ value of proline is less negative than other ones and focusing is very narrow, whereas there ise wide focus of glycine at (180^0, 180^0). [11]

These dehidral angles correspond to a sheet structure in which planes of peptide bonds make zigzag and side chains stick out from it. In such a geometry, every partially negative backbone nitrogen atom in one sheet makes a hydrogen bond

with the corresponding partially positive carbonyl oxygen atom in the adjacent sheet. In antiparallel sheets, each adjacent sheet has a length of 7 Å and their amino-to-carboxyl orientations are opposite. It means that this conformation is the one in which the closest pairs of the sheets make hydrogen bonding. Sheet length becomes 6.5 Å in parallel sheets and their amino-to-carboxyl orientations are the same. Thus, this conformation is less stable than the first one. [2]

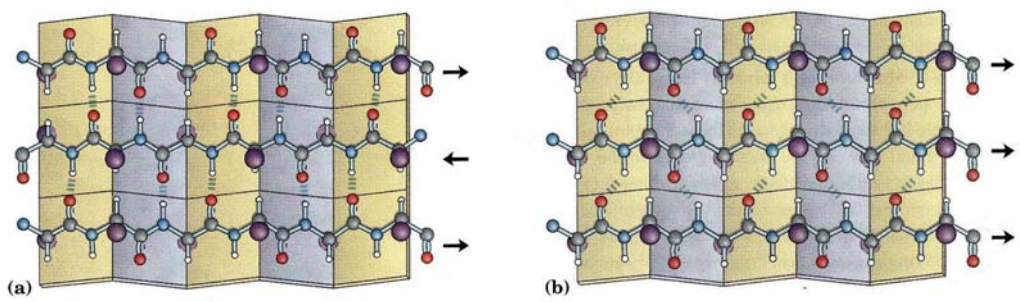

Figure 1.16 β–sheet conformations: (a) Antiparallel β–sheets, (b) Parallel β–sheets. [2]

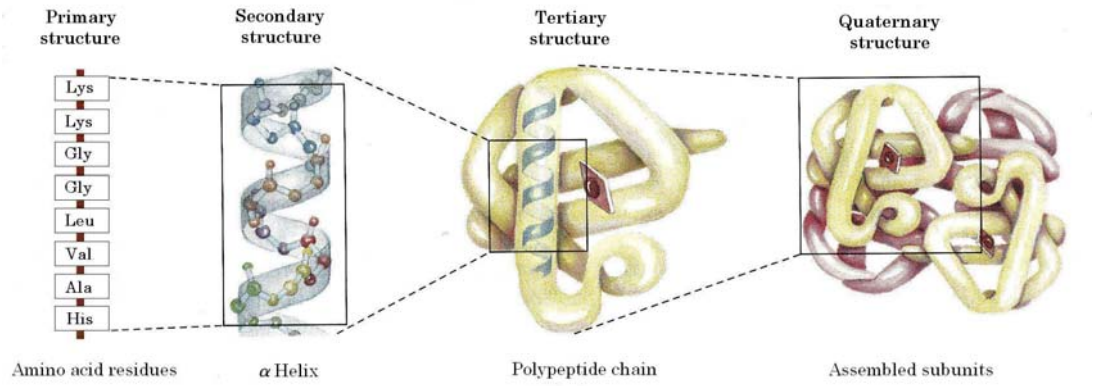

Figure 1.17 Protein folding: After a stable conformation is done, each protein starts spatially fold into a special 3–dimensional shape by long range interactions of side chains. [2]

These special 3–dimensional shapes can be classified into two major groups: Fibrous Shapes in which only one type of secondary conformation is present and Globular Proteins in which several types of secondary conformation are present. In globular proteins, glycine and proline residues cause β–turns between distinct succesive secondary structures. This is because, very small and flexible glycine does not join the side chain interactions and proline can be found in *cis* configuration. [2, 10]

Although insoluble fibrous proteins are mainly responsible for support, shape, and external protection in living systems, each globular protein has a special function such as regulation of certain biologic process like transportation or catalysis of a certain biochemical reaction as being an enzyme. [2]

Neutral pH, mild-temperature or aqueous enviroment of the cells is suitable for stabilization of most biomolecules and so, biochemical reactions can be slow in these conditions. Moreover, some reactants should be in unstable transition states or should collide with each other in certain orientations and cellular conditions does not allow them. Such reactions can occur only with in the specific environments of the active site of their enzymes. An active site includes amino acid residues which have suitable chemical properties for binding to substrate molecules and for transforming them into transition states. In some situations, active site completely seperates its substrate from the solution. [2]

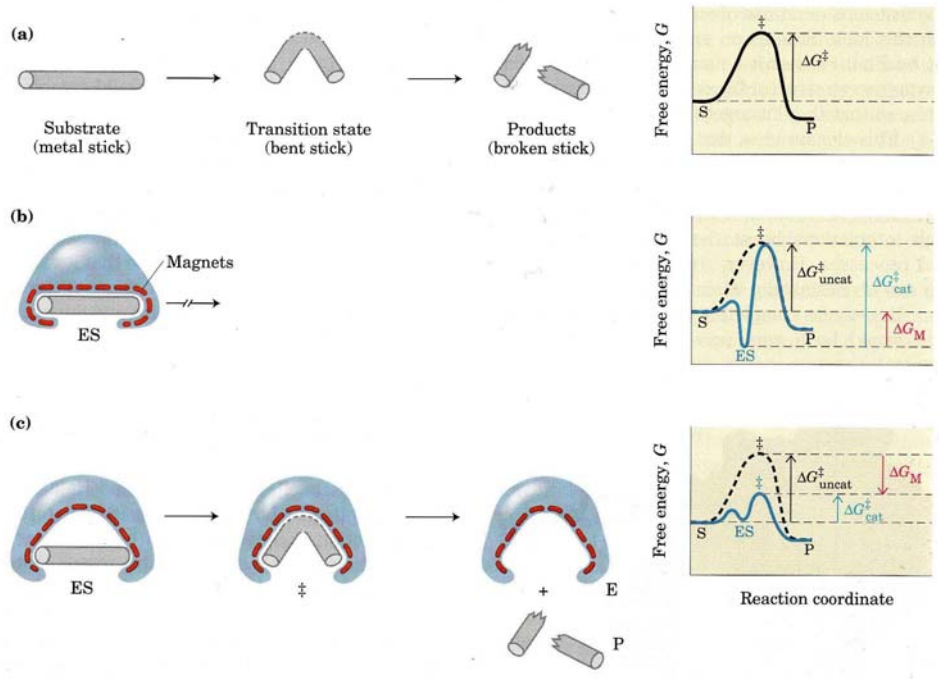

Figure 1.18 An imaginary model for enzyme activity: Enzymes can not change the equlibria of reactions, they make reaction rate bigger than the ones in reactions with no enzyme. (a) No enzyme situation: Substrate (S) and product (P) are called as ground states. (b) Situation of enzyme (E) complementary to S: Because of the steric complementarity, enzyme–substrate (ES) complex makes S more stable at first. So, this interaction increase the activation energy ΔG^{+}. (c) Situation of enzyme complementary to transition state: ES complex makes S less stable and this interaction decreases the activation energy. [2]

Enzymes can increase reaction rates up to 10^{17}. During reaction, covalent bond rearrangements decrease activation energy but main contribution comes from weak nonpolar interactions which are also responsible for the specifity of interactions. [2]

1.2 EVOLUTION AND INFORMATION THEORY

1.2.1 Systems and Complexity

Systems that show complex properties like genetic machinery can be classified into three major groups: [12]

- Simple systems that generate complex dynamics

- Complicated systems

- Complex systems

A system is called as simple system if it has not so much elements and its dynamics depend on well understood laws. For example, a pendulum system is a simple system since it has only one element and since its dynamics depend on Newton's laws. However, this simple system can generate chaotic properties if pendulum is periodically forced by gravity. Moreover, system remains chaotic when another pendulum is added to the present one. [12]

Complicated systems have a lot of elements whose roles are well defined and whose regulation is done by well undestood rules. For example, an airplane can contain 3×10^6 elements each of which has to work in a harmony for a successfull flight. It is a typical property of complicated systems. If one of the elements makes an error, whole system becomes unsuccessful. As a result, such systems can not reconstruct themselves according to long-range environmental changes. [12]

There are a lot of elements in also complex systems and they can be identical or different. However, roles of these elements are not so rigid as in complicated systems and can be updated by time. Interactions are nonlinear and do not bind elements permanently to each other. Morever, there is noise in interactions and dynamics of the system depends on some puzzling internal laws. Such a system is more than the sum of its elements and organization of it does not require any external principle. For example, although a bird flock does not have so many elements as an airplane, flock can adapt its flying to changing environmental conditions. [12]

Nature of the complex systems makes their studies quite difficult. Some useful tools are nonlinear dynamics, chaos, statistical physics, and network theory. Such tools can explain a lot of things about complex interactions between genes, proteins, and environment. [12] However, assumption of genetic code's being a complex system is made from a biochemical perpective. Genetic code includes complex biochemical molecules and their complex interactions, but eventually it is the evolutionary determinated way of storage and usage of hereditary information. Morever, non-informative elements are so many in this system and informative ones are only four nucleotides, twenty amino acids, and some operational codes. Perhaps, character of the complexity of genetic code changes and some other aspects of life become available to be explained, when system is studied with a bioinformatic perspective.

1.2.2 Information Theory

Information can be classified into three different types as syntactic information, semantic information and pragmatic information. Syntactic information is related to the form of the information. Considerations are based on the symbols used in a message and their arrangement. Meaning is exactly excluded. Information Theory (IT) based on this type information is known as communication theory or mathematical IT. Measurement of information, basic limits on tranmission and compression of the information, and the ways of building information processing systems that can reach these limits are the main subjects of this theory. [13]

On the contrary, both of the other two information types consider the information content in spite of the form of information. Semantic information is relative to meanings, whereas pragmatic information is contexual. IT based on these two information types is known as British tradition of IT which chooses its main research subjects on social sciences. [13]

According to mathematical IT, syntactic information is related to reduction of possibilities or uncertainty. [13] Thus, avarage amount of information of outcome x can be defined as the entropy of the ensemble X [14]

$$H(\mathrm{X}) = -\sum_{x \in A_{\mathrm{X}}} p(x) \log_b p(x) \tag{1.1}$$

This measure is known as Shannon's information measure and its unit is determined by *b*. Unit is bit for $b = 2$, nat for $b = e$, and digit for $b = 10$. [13]

According to equation (1.1), Shannon's information measure of a message M which is made up from *L* of *N* possible building blocks can be formulated as

$$H(M) = -\sum_{j=1}^{L}\sum_{i=1}^{N} p_j(x_i)\log_b p_j(x_i) \qquad (1.2)$$

at where $p_j(x_i)$ is the relative frequency of building block x_i at the position *j*.

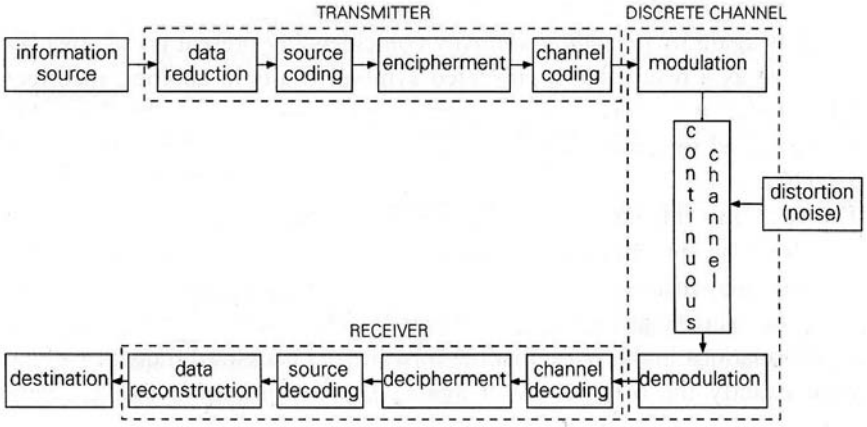

Figure 1.19 Communication model [13] A messages is transported over a noisy channel during communication and syntactic information is distorted by noise. Mathematical IT searches systematic solutions for perfect communications over noisy channels. As referred before, Tlusty shows that degeneracy in the genetic code can be such a systemamic solution [8].

Equation (1.2) means that maximization of information amount in a message can be achieved by increasing the randomness both in each block and in overall arrangement of blocks. Probabilities with which building blocks are chosen is of importance and not the building blocks themselves.

Generally, there is a maximization in the amount of syntactic information of biochemical messages, but it is not enough for living organisms. Bioinformatic perpective should not exactly exclude biochemistry. Information carried by biochemical messages are related to specific biochemical properties of chosen building blocks. Thus, biological information should also be considered as semantic information. [15]

In this sense, genes and proteins can be thought as molecular words which have operational and functional meanings. Language of genes has an alphabet that consists of four different nucleotides, whereas language of proteins has an alphabet that consists of twenty (or sometimes twenty one) amino acids. Consequently, triplet genetic code is a molecular dictionary which is used for the translation of genetic language to the proteomic language. [8, 15]

Stability of meanings in a practical language is related to the ability of coping with local disturbances which can cause errors in alphabetic level. In addition, practicability of a language depends on the usage of physical sources such as space, time, energy, and so on, with maximum efficiency. [15] So, a practical language can be obtained by constructing its alphabet from an optimal number of easily distinguishable building blocks. [16]

As mentioned in detail before, semiconservative replication with enzymes makes duplication of hereditary information accurate and fast. Similarly, protein synthesis gains both accuracy and speed by translation according to the triplet genetic code with enzymes. From this perspective, it seems that both of the genetic and proteomic languages are practical, and optimization of the building block number of their alphabets should be somehow related to the physical mechanisms underlying replication (or transciption) and translation.

In fact, each of the genetic machinery mentioned above can be defined as an unsorted database search in which each building block addition is a successful search. This is because, the nucleic acids and proteins are both synthesized by finding right building blocks from the surrounding environment of mixed building blocks and by arranging them according to a pre-existing template. Thus, the database of the replication (or tanscription) consists of four building blocks, whereas database of the translation consists of twenty (sometimes twenty one) building blocks and one stop signal. Moreover, evolution of the genetic and proteomic languages with their alphabets can be thought as an optimization of unsorted database search. [16]

2 QUANTUM ALGORITHMS AND GENETIC CODE

2.1 UNSORTED DATABASE SEARCH

Suppose that there are *N* different elements which are distributed randomly in a database. A query consists of an identifier yes/no question about a desired element and *Q* is the minimum query number which is needed to find this element.

2.1.1 Classical Database Search

i) If probability is uniformly distributed among the elements in each query and an element which has been already queried is put into database again, probability p_i to find desired element in the query *i* can be formulated as

$$p_i = \Omega_i \prod_{k=1}^{i-1}\left(1-\Omega_k\right) = \Omega\prod_{k=1}^{i-1}\left(1-\Omega\right) = \Omega\left(1-\Omega\right)^{i-1} \tag{2.1a}$$

where the term $\Omega = 1/N$ is the probability to find desired element in an independent query and the term $\left(1-\Omega\right)^{i-1}$ is the probability not to find desired element in *i* – 1 independent queries.

Then, the avarage number of *Q* for large values of *N* is

$$\left\langle Q\right\rangle = \sum_{k=1}^{N\to\infty} k\, p_k \tag{2.2}$$

$$= \sum_{k=1}^{N\to\infty} k\,\Omega\left(1-\Omega\right)^{k-1} = -\Omega\frac{\partial}{\partial\Omega}\sum_{k=1}^{N\to\infty}\left(1-\Omega\right)^{k}$$

$$= -\Omega\frac{\partial}{\partial\Omega}\left(\frac{1}{1-\left(1-\Omega\right)}\right) = -\Omega\left(-\frac{1}{\Omega^2}\right) = \frac{1}{\Omega} = N$$

Also, it can be shown that there should be at least *N*/2 queries to find a desired element with a probability of 1/2 in this situation.

$$\frac{1}{2} = \sum_{k=1}^{Q_{1/2}} p_k \tag{2.3}$$

$$= \sum_{k=1}^{Q_{1/2}} \Omega\left(1-\Omega\right)^{k-1} = \frac{\Omega}{1-\Omega}\sum_{k=1}^{Q_{1/2}}\left(1-\Omega\right)^{k}$$

$$= \frac{\Omega}{1-\Omega}\frac{\left(1-\Omega\right)\left(1-\left(1-\Omega\right)^{Q_{1/2}}\right)}{\left(1-\left(1-\Omega\right)\right)} = 1-\left(1-\Omega\right)^{Q_{1/2}} \approx 1-\left(1-Q_{1/2}\Omega\right) = Q_{1/2}\Omega = \frac{Q_{1/2}}{N}$$

$$\Rightarrow Q_{1/2} = \frac{N}{2} \tag{2.4}$$

ii) If an element which has already been queried is not put into database again, this value of $Q_{1/2}$ does not change in spite of the fact that element number of the database is decreasing in each query.

$$\Omega_k = \frac{\binom{1}{1}\binom{N_k - 1}{0}}{\binom{N_k}{1}} = \frac{1}{N_k} = \frac{1}{N-(k-1)} \tag{2.5}$$

$$p_i = \Omega_i \prod_{k=1}^{i-1} (1-\Omega_k) = \frac{1}{N-(i-1)} \frac{N-(i-1)}{N-(i-2)} \cdots\cdots \frac{N-2}{N-1} \frac{N-1}{N} = \frac{1}{N} \tag{2.6}$$

$$\frac{1}{2} = \sum_{k=1}^{Q_{1/2}} p_k = \sum_{k=1}^{Q_{1/2}} \frac{1}{N} = \frac{Q_{1/2}}{N} \Rightarrow Q_{1/2} = \frac{N}{2} \tag{2.7}$$

However, avarage number of Q is smaller than the ones in the first situation:

$$\langle Q \rangle = \sum_{k=1}^{N} k\, p_k = \frac{1}{N} \sum_{k=1}^{N} k = \frac{1}{N} \frac{N(N+1)}{2} = \frac{N+1}{2} \tag{2.8}$$

All in all, it can be said that $O(N)$ queries are required in order to obtain a desired element by classical algorithms and there should be at least $N/2$ queries to find this element with a probability of 1/2.

2.1.2 Quantum Database Search

Quantum mechanical systems are defined with the help of probability amplitudes rather than probabilities and evolve by the action of a transition matrix on the amplitude vector (/state vector). In such systems, qubit (quantum bit) is used instead of bit and each qubit can be any linear combination of the states 0 and 1. Although the number of the possible linear combinations is infinite, measurement of a qubit makes its state only 0 or 1 with a probability which depends on the amplitudes before measurement. This means that information amount of qubit decreases by measurement. This hidden information of the system exponentially increases with the number of qubits. Consequently, properties such as interference and superposition of the quantum states, give a simple parallelism which makes quantum algorithms faster than classical ones. [17, 18]

2.1.2.1 Original Version of Grover's Algorithm

In the Quantum Search Algorithm, which is also called as Grover's Algorithm, database can be defined as *N*–dimensional Hilbert Space whose eigenstates are the elements of the database. Algorithm has three basic operations: [18]

1) Initiation: Formation of the system with a superposition state $|s\rangle$ in which all *N* eigenstates have the same probabilty amplitude.

$$|s\rangle = 1/\sqrt{N}\,|111\cdots1\rangle \tag{2.9}$$

2) Query: Turning over the sign of the amplitude of desired eigenstate $|i\rangle$ by a conditional rotation.

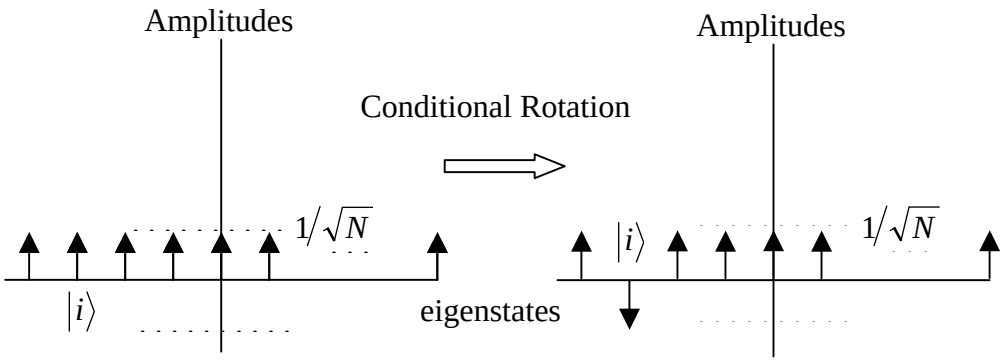

Figure 2.1 Conditional rotation in a query of Grover's Algorithm

3) Diffusion Transition [17] (/Walsh-Hadamard Transition [18]): Invertion of all amplitudes about avarage by Diffusion Matrix.

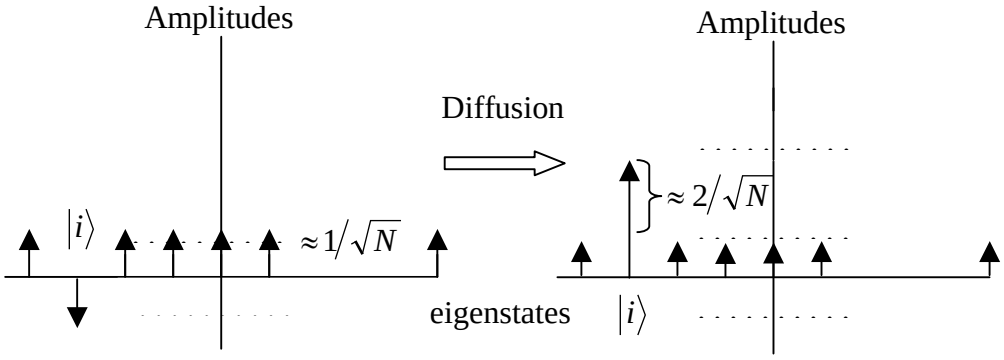

Figure 2.2 Inversion about avarage in a query of Grover's Algorithm

Grover's Algortims starts after obtaining the symmetric superposition state $|s\rangle$ and consists of repetition of operations (2) and (3) for *Q* times: [17, 18]

$$G^Q|s\rangle = |i\rangle \tag{2.10}$$

As shown in the Figure 2.2 above, when operator *G* acts on state vector for one time, amplitude of the desired state $|i\rangle$ increases approximately by $2/\sqrt{N}$ times, whereas other amplitudes decrease. As the step number of algorithm closes to $\sqrt{2N}/4$, amplitude of $|i\rangle$ goes to $1/\sqrt{2}$. After step number reaches to $\sqrt{N}/2$, all the other amplitudes vanish and state vector turns into the desired state. So, it can be said that value of *Q* is approximately $0.5\sqrt{N}$.

Consequently, Grover's Algorithm requires only $O(\sqrt{N})$ queries in order to find desired eigenstate $|i\rangle$ with a probability of 1/2 while a classic algorithm uses at least $O(N)$ queries for the same purpose.

2.1.2.1.1 Initiation: Symmetric Superposition State

Suppose that $N = 2^n$ and states can be described by *n* qubits since a system consisting of *n* qubits has 2^n possible eigenstates. Then, symmetric superposition state $|s\rangle = 1/\sqrt{N}\,|111\cdots1\rangle$ can be obtained by using Hadamard Matrix $W^{\otimes n}$. [18]

Fractal Hadamard Matrix is represented as

$$W^{\otimes k+1} = \frac{1}{\sqrt{2}}\begin{pmatrix} +W^{\otimes k} & +W^{\otimes k} \\ +W^{\otimes k} & -W^{\otimes k} \end{pmatrix} \tag{2.11a}$$

$$W^{\otimes 0} = 1 \tag{2.11b}$$

Let's look at the situation in which *n* is one. Action of $W^{\otimes 1}$ on the state $|0\rangle$ gives the symmetric superposition state, whereas action of the operator on the state $|1\rangle$ gives the antisymmetric superposition state:

$$W^{\otimes 1} = \frac{1}{\sqrt{2}}\begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix} \tag{2.11c}$$

$$W^{\otimes 1}|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) = |s\rangle \tag{2.12a}$$

$$W^{\otimes 1}|1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}+1 & +1\\ +1 & -1\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}+1\\-1\end{pmatrix} = \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right) = |a\rangle \qquad (2.12b)$$

In each case, both of the eigenstates have the same probabilty of 1/2, but probability amplitudes are equal only if $W^{\otimes 1}$ is applied on the state $|0\rangle$.

Now, let's look at the situations in which *n* is greater than one. $W^{\otimes k+1}$ can be also formulated as a tensor product like $W^{\otimes k} \otimes W^{\otimes k}$. Thus, action of $W^{\otimes n}$ on *n* qubits can be written as *n* independent action of $W^{\otimes 1}$ on each qubit as shown below.

$$W^{\otimes n}|\tilde{0}\rangle = W^{\otimes 1}|0\rangle \otimes W^{\otimes 1}|0\rangle \otimes \cdots \otimes W^{\otimes 1}|0\rangle \qquad (2.12c)$$

$$= \frac{1}{2^{n/2}}\left(|0\rangle + |1\rangle\right)^{\otimes n} = \frac{1}{2^{n/2}}\left(|00\cdots0\rangle + |01\cdots0\rangle + \cdots + |00\cdots1\rangle\right)$$

$$= \frac{1}{2^{n/2}}\left(|\tilde{1}\rangle + |\tilde{2}\rangle + \cdots + |\tilde{N}\rangle\right) = \frac{1}{\sqrt{N}}|11\cdots1\rangle = |\tilde{s}\rangle$$

Example for *n* equals to two,

$$W^{\otimes 2}|\tilde{0}\rangle = W^{\otimes 2}|00\rangle = W^{\otimes 1}|0\rangle \otimes W^{\otimes 1}|0\rangle \qquad (2.12d)$$

$$= \frac{1}{2}\left(|0\rangle + |1\rangle\right) \otimes \left(|0\rangle + |1\rangle\right) = \frac{1}{2}\left(|00\rangle + |01\rangle + |10\rangle + |11\rangle\right) = |\tilde{s}\rangle$$

Therefore, symmetric superposition state in which each eigenstate has a probability amplitude of $1/\sqrt{N}$ can be obtained by applying $W^{\otimes n}$ on the state $|\tilde{0}\rangle$. Since *n* equals $\log_2(N)$, this operation requires $O(\log_2(N))$ steps.

2.1.2.1.2 Construction of Operators for Grover's Algorithm

Query operator is also known as Quantum Oracle, because its action on desired state $|i\rangle$ is different from the ones on other states: It turns over the sign of the amplitude of desired state, whereas amplitudes of other states are not affected from it. This operation looks like a reflection about $|i_\perp\rangle$, symmetric superposition of the eigenstates different from state $|i\rangle$. [17, 18]

A reflection operator can be in the form of $P = 1 - 2|x\rangle\langle x|$ since *P* satisfies the reflection condition $P^2 = 1$. [16]

$$P^2 = (1-2|x\rangle\langle x|)(1-2|x\rangle\langle x|) \tag{2.13}$$

$$= 1-4|x\rangle\langle x| + 4|x\rangle\langle x|x\rangle\langle x| = 1-(4-4)|x\rangle\langle x| = 1$$

Therefore, oracle operator can be constructed as $O = 1-2|i\rangle\langle i|$. Operator O only flips the sign of the amplitude of state $|i\rangle$ as shown below.

$$O|i\rangle = (1-2|i\rangle\langle i|)|i\rangle = |i\rangle - 2\langle i|i\rangle|i\rangle = |i\rangle - 2|i\rangle = -|i\rangle \tag{2.14a}$$

$$O|i_\perp\rangle = (1-2|i\rangle\langle i|)|i_\perp\rangle = |i_\perp\rangle - 2\langle i|i_\perp\rangle|i\rangle = |i_\perp\rangle - 0 = |i_\perp\rangle \tag{2.14b}$$

In a similar way, diffusion transition operator can be constructed as $D = 2|s\rangle\langle s|-1$ because of the fact that invertion of all amplitudes about avarage is the same operation with reflection about symmetric superposition state $|s\rangle$. [18] In order to show this fact, let $\vec{\nu}$ be a *N*–dimensional amplitude vector with an avarage amplitude of α.

$$\vec{\nu} = \sum_{j=1}^{N} \alpha_j |j\rangle \tag{2.15}$$

$$D\vec{\nu} = (2|s\rangle\langle s|-1)\vec{\nu} = \left(\frac{2}{N}|11\cdots1\rangle\langle 11\cdots1|-1\right)\vec{\nu} = \frac{2}{N}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}_{NxN}\vec{\nu} - \vec{\nu} \tag{2.16a}$$

$$\left(D\vec{\nu}\right)_k = \sum_{j=1}^{N} D_{kj}\,\alpha_j = \sum_{j=1}^{N}\left(\frac{2}{N}-\delta_{jk}\right)\alpha_j = 2\alpha - \alpha_k = \alpha + (\alpha - \alpha_k) \tag{2.16b}$$

Diffusion Transition is also known as Walsh-Hadamard Transition since operator *D* can be expressed in terms of Hadamard Matrix. [18]

$$D = W^{\otimes n} R W^{\otimes n} \tag{2.17}$$

at where *R* is a rotation matrix as shown below.

$$R = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 \end{pmatrix}_{NxN} = \underbrace{\begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 \end{pmatrix}_{NxN}}_{R_1} + \underbrace{\begin{pmatrix} 2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}_{NxN}}_{R_2} \tag{2.18}$$

Then, put (2.18) into equation (2.17)

$$D = W^{\otimes n}\left(R_1 + R_2\right)W^{\otimes n} \tag{2.19}$$

$$= W^{\otimes n}R_1 W^{\otimes n} + W^{\otimes n}R_2 W^{\otimes n} = W^{\otimes n}\left(-I\right)W^{\otimes n} + W^{\otimes n}R_2 W^{\otimes n}$$

$$= \underbrace{-W^{\otimes n}W^{\otimes n}}_{D_1} + \underbrace{W^{\otimes n}R_2 W^{\otimes n}}_{D_2}$$

Let's calculate the parts of the equation (2.19) independently.

$$D_1 = -W^{\otimes n}W^{\otimes n} \tag{2.19a}$$

$$= -\frac{1}{\sqrt{2}}\begin{pmatrix}+1 & +1\\ +1 & -1\end{pmatrix}^{\otimes n}\frac{1}{\sqrt{2}}\begin{pmatrix}+1 & +1\\ +1 & -1\end{pmatrix}^{\otimes n} = -\left(\frac{1}{2}\begin{pmatrix}+1 & +1\\ +1 & -1\end{pmatrix}\begin{pmatrix}+1 & +1\\ +1 & -1\end{pmatrix}\right)^{\otimes n}$$

$$= -\left(\frac{1}{2}\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}\right)^{\otimes n} = -\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}^{\otimes n} = -I_{NxN} = -1$$

$$\left(D_2\right)_{jk} = \sum_{l,m}\left(W^{\otimes n}\right)_{jl}\left(R_2\right)_{lm}\left(W^{\otimes n}\right)_{mk} \tag{2.19b}$$

$$= \sum_{l,m}\left(W^{\otimes n}\right)_{jl}\left(2\delta_{l,0}\delta_{m,0}\right)\left(W^{\otimes n}\right)_{mk} = 2\left(W^{\otimes n}\right)_{j0}\left(W^{\otimes n}\right)_{0k}$$

$\left(W^{\otimes n}\right)_{jk}$ should be calculated in order to continue proving the equation (2.17). Let's look at the situations in which *n* is two and try to obtain a general formula for $\left(W^{\otimes n}\right)_{jk}$.

$$W^{\otimes 2} = \frac{1}{\sqrt{2}}\begin{pmatrix}+1 & +1\\ +1 & -1\end{pmatrix}\otimes\frac{1}{\sqrt{2}}\begin{pmatrix}+1 & +1\\ +1 & -1\end{pmatrix} \tag{2.20a}$$

$$= \frac{1}{2}\left(\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} + \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix} + \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix} - \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\right)^{\otimes 2}$$

$$= \frac{1}{2}\left(|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| - |1\rangle\langle 1|\right)^{\otimes 2}$$

$$= \frac{1}{2}\left((-1)^{0\cdot 0}|0\rangle\langle 0| + (-1)^{0\cdot 1}|0\rangle\langle 1| + (-1)^{0\cdot 0}|1\rangle\langle 0| + (-1)^{1\cdot 1}|1\rangle\langle 1|\right)^{\otimes 2}$$

$$= \frac{1}{2}\left((-1)^{\binom{0}{0}\cdot\binom{0}{0}} |00\rangle\langle 00| + (-1)^{\binom{0}{0}\cdot\binom{0}{1}} |00\rangle\langle 01| + \cdots + (-1)^{\binom{1}{1}\cdot\binom{1}{1}} |11\rangle\langle 11| \right)$$

$$= \frac{1}{2}\left((-1)^{\binom{0}{0}\cdot\binom{0}{0}} |\tilde{1}\rangle\langle\tilde{1}| + (-1)^{\binom{0}{0}\cdot\binom{0}{1}} |\tilde{1}\rangle\langle\tilde{2}| + \cdots + (-1)^{\binom{1}{1}\cdot\binom{1}{1}} |\tilde{4}\rangle\langle\tilde{4}| \right)$$

$$= 2^{-2/2} \sum_{j,k=1}^{2^2} (-1)^{\bar{j}\cdot\bar{k}} |j\rangle\langle k| \tag{2.20b}$$

In the equation (2.20b), $\bar{j}\cdot\bar{k}$ is the dot product of the 2–dimensional vector representations of the states. This equation can be generalized as shown below by using the fractal structure of Hadamard Matrix.

$$W^{\otimes n} = 2^{-n/2} \sum_{j,k=1}^{N} (-1)^{\bar{j}\cdot\bar{k}} |j\rangle\langle k| \tag{2.21a}$$

$$\left(W^{\otimes n}\right)_{jk} = 2^{-n/2} (-1)^{\bar{j}\cdot\bar{k}} \tag{2.21b}$$

Now, $\bar{j}\cdot\bar{k}$ is the dot product of the *N*–dimensional representations of the states. Since $\left(W^{\otimes n}\right)_{jk}$ is known now, let's put (2.21b) into (2.19b) and continue proving the equation (2.17).

$$\left(D_2\right)_{jk} = 2\left(W^{\otimes n}\right)_{j0}\left(W^{\otimes n}\right)_{0k} = 2/N\,(-1)^{\bar{j}\cdot\bar{0}+\bar{0}\cdot\bar{k}} = 2/N\,(-1)^0 = 2/N \tag{2.22a}$$

$$D_2 = \frac{2}{N}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}_{NxN} = \frac{2}{N}|11\cdots 1\rangle\langle 11\cdots 1| = 2|s\rangle\langle s| \tag{2.22b}$$

Now, put (2.19a) and (2.22b) into equation (2.19) and finish proving the equation (2.17).

$$D = W^{\otimes n} R W^{\otimes n} = 2|s\rangle\langle s| - 1$$

All in all, Grover's Algortihm can be formulated as

$$G^Q|s\rangle = (DO)^Q|s\rangle = \left(\left(2|s\rangle\langle s| - 1\right)\left(1 - 2|i\rangle\langle i|\right)\right)^Q|s\rangle = |i\rangle \tag{2.23}$$

2.1.2.1.3 Geometric solution for Grover's Algorithm

Let angle between the states $|s\rangle$ and $|i_\perp\rangle$ be $\theta/2$. Then, $|s\rangle$ can be written as

$$|s\rangle = \sin(\theta/2)|i\rangle + \cos(\theta/2)|i_\perp\rangle \tag{2.24}$$

The angle between state vector and $|i_\perp\rangle$ remains same [16] after the Oracle, whereas angle between state vector and $|i\rangle$ goes from $(\pi-\theta)/2$ to $(\pi+\theta)/2$.

$$O|s\rangle = (1-2|i\rangle\langle i|)|s\rangle = |s\rangle - 2\langle i|s\rangle|i\rangle = |s\rangle - 2\sin(\theta/2)|i\rangle \tag{2.25}$$

$$\langle i_\perp|O|s\rangle = \langle i_\perp|(|s\rangle - 2\sin(\theta/2)|i\rangle) = \langle i_\perp|s\rangle - 2\sin(\theta/2)\langle i_\perp|i\rangle = \cos(\theta/2) \tag{2.26}$$

$$\langle i|O|s\rangle = \langle i|(|s\rangle - 2\sin(\theta/2)|i\rangle) \tag{2.27}$$

$$= \langle i|s\rangle - 2\sin(\theta/2)\langle i|i\rangle = -\sin(\theta/2) = \cos((\theta+\pi)/2)$$

Therefore, Oracle reflects the state vector about $|i_\perp\rangle$ by an angle of $\theta/2$.

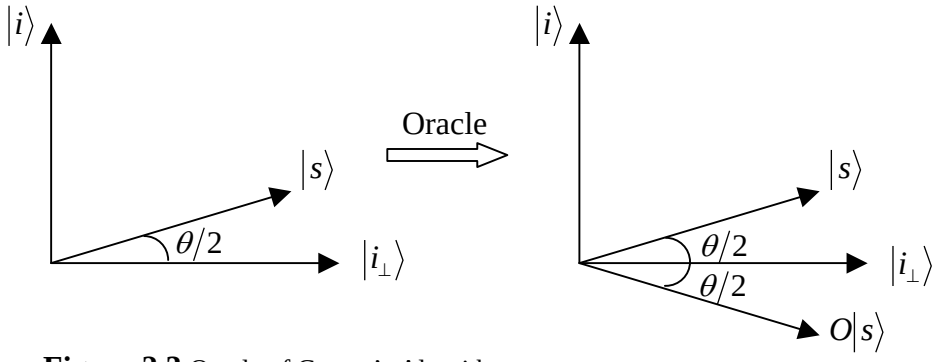

Figure 2.3 Oracle of Grover's Algorithm

Similiarly, the angle between the state vector and $|s\rangle$ remains same [16] after the Diffusion, whereas angle between the state vector and $|i_\perp\rangle$ goes from $\theta/2$ to $3\theta/2$.

$$DO|s\rangle = (2|s\rangle\langle s|-1)O|s\rangle \tag{2.28}$$

$$= (2|s\rangle\langle s|-1)(|s\rangle - 2\sin(\theta/2)|i\rangle)$$

$$= 2|s\rangle - 4\sin^2(\theta/2)|s\rangle - |s\rangle - 2\sin(\theta/2)|i\rangle$$

$$= (1-4\sin^2(\theta/2))|s\rangle - 2\sin(\theta/2)|i\rangle$$

$$\langle i_\perp | DO | s \rangle = \langle i_\perp | \left(\left(1 - 4\sin^2(\theta/2)\right) | s \rangle - 2\sin(\theta/2) | i \rangle \right) \tag{2.29}$$

$$= \left(1 - 4\sin^2(\theta/2)\right) \langle i_\perp | s \rangle - 2\sin(\theta/2) \langle i_\perp | i \rangle$$

$$= \left(1 - 4\sin^2(\theta/2)\right) \cos(\theta/2)$$

$$= \left(\cos^2(\theta/2) - 3\sin^2(\theta/2)\right) \cos(\theta/2)$$

$$= \left(\cos(\theta) - 2\sin^2(\theta/2)\right) \cos(\theta/2)$$

$$= \cos(\theta)\cos(\theta/2) - 2\sin(\theta/2)\cos(\theta/2)\sin(\theta/2)$$

$$= \cos(\theta)\cos(\theta/2) - \sin(\theta)\sin(\theta/2)$$

$$= \cos(3\theta/2)$$

$$\langle s | DO | s \rangle = \langle s | \left(\left(1 - 4\sin^2(\theta/2)\right) | s \rangle - 2\sin(\theta/2) | i \rangle \right) \tag{2.30}$$

$$= \left(1 - 4\sin^2(\theta/2)\right) \langle s | s \rangle - 2\sin(\theta/2) \langle s | i \rangle$$

$$= 1 - 4\sin^2(\theta/2) - 2\sin^2(\theta/2) = \cos^2(\theta/2) - \sin^2(\theta/2) \ = \cos(\theta)$$

Therefore, Diffusion reflects the state vector about $|s\rangle$ by an angle of θ.

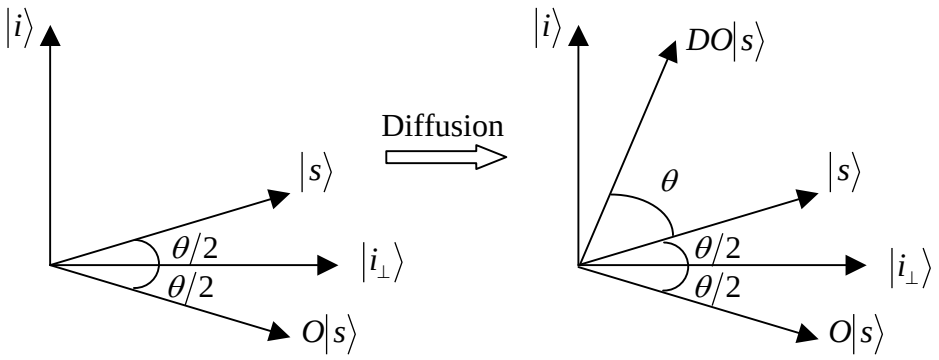

Figure 2.4 Diffusion Transition in Grover's Algorithm

All in all, the angle between the state vector and $|i_\perp\rangle$ goes from $\theta/2$ to $3\theta/2$ when Grover's Algorithm is applied on state vector for one time. Since two consequtive reflections equal to a rotation, it can be said that during each step, state vector is rotated by a constant angle θ in the plane including both of $|s\rangle$ and $|i\rangle$.

$$G^Q|s\rangle = \sin\left(\frac{\theta}{2}+Q\theta\right)|i\rangle + \cos\left(\frac{\theta}{2}+Q\theta\right)|i_\perp\rangle \tag{2.31}$$

L.H.S. of equation (2.31) also equals to desired state $|i\rangle$ according to equation (2.10). So, cosine term of the equation should vanish. It means that argument of the cosine function should be $\pi/2$ on modulo 2π . Then,

$$\frac{\theta}{2}+Q\theta = \frac{2Q+1}{2}\theta = \frac{\pi}{2} \tag{2.32a}$$

$$\Rightarrow \theta = \frac{\pi}{2Q+1} \tag{2.32b}$$

One more equation for angle θ can be derived from equation (2.24):

$$\sin(\theta/2) = \langle s|i\rangle = 1/\sqrt{N}\left(\langle i_\perp| + \langle i|\right)|i\rangle = 1/\sqrt{N} \tag{2.33a}$$

$$\Rightarrow \theta = 2\sin^{-1}\left(\frac{1}{\sqrt{N}}\right) = \frac{\pi}{2Q+1} \tag{2.33b}$$

Combination of equations (2.32b) and (2.33b) gives an equation which shows the relation between element number of database and minimum query number which is needed to find a desired element.

$$(2Q+1)\sin^{-1}\left(\frac{1}{\sqrt{N}}\right) = \frac{\pi}{2} \tag{2.34a}$$

Let $1/\sqrt{N}$ be sufficiently smaller than one. Consequently arcsine goes to its argument and *Q* becames sufficiently bigger than one. Then,

$$(2Q+1)\sin^{-1}\left(\frac{1}{\sqrt{N}}\right) \approx (2Q+1)\frac{1}{\sqrt{N}} \approx \frac{2Q}{\sqrt{N}} = \frac{\pi}{2} \Rightarrow Q \approx \frac{\pi\sqrt{N}}{4} \tag{2.35b}$$

According to this equation, *Q* is asymptotically converging to $\pi\sqrt{N}/4$ for large values of *N*. This value equals to $0.79\sqrt{N}$ and it is not so different from the ones obtained from Figure 2.2 before.

2.1.2.1.5 Analytic Solution for Grover's Algorithm

Let α_i be probability amplitude of desired state $|i\rangle$ and α_k be probability amplitude of any other state (k = 1, 2, ..., N and $k \neq i$). Suppose that α_i is a positive

real number c and each α_k equals to another positive number l such as c over l is smaller than $\sqrt{N}$. Since Grover's Algorithm starts with a superposition state $|s\rangle$ in which all N eigenstates have the same probabilty amplitude of $1/\sqrt{N}$ and $\sqrt{N}$ is bigger than one, these assumptions are not conflicted with initial condition. It will be shown that, these assumptions are also valid for any other step of the algorithm.

Let's firstly apply operator O on state vector:

$$\begin{aligned} O|s\rangle &= |s_O\rangle \\ &= \left(1-2\left|\cdots \overset{i^{th}element}{\overbrace{0\tilde{1}0}}\cdots\right\rangle\left\langle\cdots 010\cdots\right|\right)\left|\cdots l\,\overset{i^{th}element}{\overbrace{c\tilde{l}}}\cdots\right\rangle \\ &= \left|\cdots l\,c\,l\cdots\right\rangle - 2c\left|\cdots 010\cdots\right\rangle \\ &= \left|\cdots l\,\overset{i^{th}element}{\overbrace{(-c)}}\,l\cdots\right\rangle \end{aligned} \tag{2.36}$$

Then, apply operator D on the state vector

$$\begin{aligned} D|s_O\rangle &= |s'\rangle \\ &= \left(2/N\left|11\cdots1\right\rangle\left\langle11\cdots1\right| - 1\right)\left|\cdots l\,\overset{i^{th}element}{\overbrace{(-c)}}\,l\cdots\right\rangle \\ &= 2/N\left((N-1)l - c\right)\left|11\cdots1\right\rangle - \left|\cdots l\,(-c)\,l\cdots\right\rangle \\ &= \left|\alpha_1\cdots\alpha_{i-1}\,\alpha_i\,\alpha_{i+1}\cdots\alpha_N\right\rangle \end{aligned} \tag{2.37}$$

New values of amplitudes α_i and α_k after one step of the algorithm can be found from the equation (2.37) as

$$\alpha_i = 2/N\left((N-1)l - c\right) + c = \frac{2(N-1)}{N}l + \frac{N-2}{N}c \tag{2.38a}$$

$$\alpha_k = 2/N\left((N-1)l - c\right) - l = \frac{N-2}{N}l - \frac{2}{N}c \tag{2.38b}$$

If N is bigger than two, α_i remains positive according to equation (2.38a). In order to make α_k also positive, $\left((N-2)/N\right)l$ should be bigger than $(2/N)c$. According to assumptions, c over l should be smaller than $\sqrt{N}$. So,

$$c/l < (N-2)/2 \Rightarrow (N-2)/2 \geq \sqrt{N} \Rightarrow N - 2\sqrt{N} - 2 \geq 0 \tag{2.39}$$

Solution of equation (2.39) shows that N should be bigger than 7.46 to make α_k also positive. Therefore, it can be said that if N is bigger than 7.46 and c over l is smaller than $\sqrt{N}$, amplitudes remain positive after the action of operator G on state vector for one time.

Let's take the square of both of the equations (2.38a) and (2.38b) to try to find a relation between the increase in the amplitude of desired state and decrease in the other amplitudes.

$$\alpha_i^{\,2} = \left(4 - \frac{8}{N} + \frac{4}{N^2}\right) l^2 + \left(1 - \frac{4}{N} + \frac{4}{N^2}\right) c^2 + \left(4 - \frac{12}{N} + \frac{8}{N^2}\right) lc \tag{2.40a}$$

$$\alpha_k^{\,2} = \left(1 - \frac{4}{N} + \frac{4}{N^2}\right) l^2 + \left(\frac{4}{N^2}\right) c^2 + \left(-\frac{4}{N} + \frac{8}{N^2}\right) lc \tag{2.40b}$$

Multiply equation (2.40b) by $(N - 1)$ and add it to equation (2.40a)

$$\begin{aligned} \alpha_i^{\,2} + (N-1)\alpha_k^{\,2} = & \left(4 - \frac{8}{N} + \frac{4}{N^2} + N - 4 + \frac{4}{N} - 1 + \frac{4}{N} - \frac{4}{N^2}\right) l^2 \\ & + \left(1 - \frac{4}{N} + \frac{4}{N^2} + \frac{4}{N} - \frac{4}{N^2}\right) c^2 \\ & + \left(4 - \frac{12}{N} + \frac{8}{N^2} - 4 + \frac{8}{N} + \frac{4}{N} - \frac{8}{N^2}\right) lc \\ = & \; c^2 + (N-1) l^2 \end{aligned} \tag{2.40c}$$

$$\Rightarrow \alpha_i^{\,2} - c^2 = (1 - N)\left(\alpha_k^{\,2} - l^2\right) \tag{2.40d}$$

Generalize the equation (2.40d) by replacing c with $\alpha_i(j)$ and l with $\alpha_k(j)$. If query marker $j = 1, 2, \ldots, Q$

$$\alpha_i^{\,2}(j+1) - \alpha_i^{\,2}(j) = \underbrace{(1 - N)}_{<0}\left(\alpha_k^{\,2}(j+1) - \alpha_k^{\,2}(j)\right) \tag{2.40e}$$

Equation (2.40e) clearly shows the relation between the increase in the amplitude of desired state and decrease in the other amplitudes. Now, let's have a look at $\Delta\alpha_i$ if c is smaller than $1/\sqrt{2}$. Since c over l is smaller than $\sqrt{N}$, l is bigger than $1/\sqrt{2N}$. Then,

$$\Delta\alpha_i = \left(2/N\left((N-1)l - c\right) + c\right) - c = (2 - 2/N)l - (2/N)c \tag{2.41a}$$

$(2-2/N)l$ is bigger than $\sqrt{2}/\sqrt{N} - \sqrt{2}/N\sqrt{N}$ and $-(2/N)c$ is bigger than $-\sqrt{2}/N$. Consequently, there is a lower limit for $\Delta\alpha_i$.

$$\Delta\alpha_i > \frac{\sqrt{2}}{\sqrt{N}} - \frac{\sqrt{2}}{N\sqrt{N}} - \frac{\sqrt{2}}{N} \tag{2.41b}$$

$$> \frac{\sqrt{2}}{\sqrt{N}}\left(1 - \frac{1}{N} - \frac{1}{\sqrt{N}}\right) = \frac{\sqrt{2}}{\sqrt{N}}\left(1 - \frac{1}{\sqrt{N}}\left(1 + \frac{1}{\sqrt{N}}\right)\right)$$

For large values of N, $1/\sqrt{N}$ goes to zero and $\Delta\alpha_i$ becames bigger than $\sqrt{2}/\sqrt{N}$. It means that there should be $\sqrt{N}/\sqrt{2}$ steps to make the amplitude of desired element $|i\rangle$ equal to one.

All in all, order of Q is found as $\sqrt{N}$ again and all three ways used for this purpose is summarized in Table 2.1 below.

Table 2.1 Number of queries to find a desired element by Grover's Algorithm

Solution Way	*Asymptotic Value of Q*
Figurative Solution as in Figure 2.2	$\sqrt{N}/2 = 0.50\sqrt{N}$
Geometric Solution as in equation (2.35b)	$\pi\sqrt{N}/4 = 0.79\sqrt{N}$
Analytic Solution	$\sqrt{N}/\sqrt{2} = 0.71\sqrt{N}$

2.1.2.2 Analogue Version of Grover's Algorithm

It can be shown that Grover's algorithm is optimal for the task of unsorted database search. [17, 19] Classical algorithms based on Boolean logic are always inferior to quantum algorithms since they can not use superposition and interference of the states. However, analogue algorithms based on wave dynamics can also use both of superposition and interference features. [20]

Wave algorirthms can not be accurate as much as quantum algorithms since they do not discretize or digitize the information. [20] Discretization gives an ability to reach any desired accuracy easily and to maintain this accuracy by coping with the local disturbances. [16, 21] In spite of this fact, wave algorithms are more stable than

quantum algorithms. Because decoherence is weaker and there is no entanglement in wave algorithms. [20]

In the analogue version of Grover's Algorithm, database includes *N* different wave modes instead of *n* qubits. This version can be described by a classical coupled pendulum system. Symmetric superposition state is replaced by the center-of-mass mode in which there is a synchronisation between *N* identical harmonic oscillators that are coupled to a big, motionless oscillator. Then, reflection operator which is a discrete unitary Oracle in the original version, becomes a continuous time interaction Hamiltonian which acts on the desired state in a different way from the others. It can be achieved by elastically tapping of the desired oscillator which causes a phase shift of half a period. Diffusion operator should also act in the velocity space. If time interval of tapping is chosen as π, coupled oscillator can redistribute kinetic energy and this corresponds to the action of diffusion operator. Amplitude amplification increases the energy of desired oscillator instead of its probability. Since inelastic tapping does not contribute to the total energy of the system, it can be said that system makes energy transfer from other modes to the desired one. [20]

Most important property of the analogue version is that database search problem is solvable in both classical and quantum domains. Quantum system can be constructed in the same way mentioned above and this makes Grover's Algorithm useful in the analysis of the systems which have crossovers between classical and quantum behaviours. [20]

2.2 GENETİC CODE AND SIMPLE PHYSICAL CONCEPTS

There is no mechanism that can distinguish already queried molecules from the other ones in genetic machinery. Therefore, if a classical algorithm is used for database search in DNA replication (or transcription), there should be four queries in each nucleotide base addition according to equation (2.2). Similiarly, there should be approximately twenty one queries in each amino acid addition during protein synthesis in which classical search algorithm is used. These numbers are quite big when base pair numbers required for replication (or transcription) and translation are considered. Base pair number is 1 for replication (or transcription) and 3 for translation. Each base pairing includes two or three hydrogen bond formations as mentioned in the introduction section. However, required hyrogen bond numbers are also smaller than the numbers which are required by classical algorithms.

In contrast, if a quantum search algortihm is used for database search in DNA replication (or transcription), there is a one-to-one correspondence between the required base pair number and required query number which is the solution of the equation (2.34a). Also, the required query number in translation differs from the requied base pair number with a small non-integer part which can be considered as an indicator of intrinsic error.

$$(2Q+1)\sin^{-1}\left(\frac{1}{\sqrt{N}}\right)=\frac{\pi}{2}\Rightarrow\begin{cases}N=4 & \Rightarrow Q=1\\ N=20.2 & \Rightarrow Q=3\\ N=21 & \Rightarrow Q=3.07\end{cases} \tag{2.42}$$

First of all, such a matching between base pair numbers and quantum query numbers necessitates a direct link between base pairing and quantum mechanics. Although atoms are attracted to each other by electrostatic forces at the beginning of hydrogen bonding, hydrogen bonds have a partially quantum mechanical nature. Experiments show that two electrons are likely to be covalently shared between the hydrogen nucleus and hydrogen bond acceptor atom. [22] Also, some recent research is making good progress in elucidation of the quantum nature of proton transfer through a hydrogen bond. [23]

In addition to hydrogen bonding, electron and proton delocalizations are involved in bond formation between bases. As mentioned before, pairing genenerally occurs when bases are in stable tautomer forms. It can be shown that a coupling

between tautomeric shifts and alteration in elasticity of DNA makes proton tunelling possible in base pairing. [24, 25] Consequently, base pairing can be considered as a quantum mechanical process.

After base pairing, phosphodiester bonds are formed between nucleotides in replication (or transcription), whereas peptide bonds are formed between amino acids in translation. Both the phophodiester and peptide bonds have a covalent nature which can be explained only by quantum mechanics. Therefore, consideration of replication (or transcription) and translation as a quantum database search does not cause a contradiction with the present knowledge about these processes.

2.2.1 Language of the Genes

Consequtive steps of Grover's Algorithm look like a discretized evolution of the state vector. Hamiltonian *H* which is responsible for this evolution has two different terms which correspond to $|i\rangle\langle i|$ and $|s\rangle\langle s|$ components of operator *G* respectively. First term behaves as a potential energy which attracts the state towards $|i\rangle$. Therefore, conditional rotation due to Oracle should represent base pairing via hydrogen bonding. Second term behaves as a kinetic energy which diffuses state. It means that diffusion due to Walsh-Hadamard Transformation should be independent from bases and so, it should represent the processes that occur on sugar–phosphate backbones. [16]

Finally, base pairing should show some good quantum features which makes crossovers between quantum and classical domains possible in order to make transformations stable against decoherent fluctuations of cellular environment. [16]

2.2.1.1 Base Pairing as Quantum Oracle

Firstly, each nucleotide joining to usual base pairing in DNA should have a special qubit representation. One qubit is required for the information about the structure of the nucleotide. Also, a second qubit is required for the information about tautomeric form of the nucleotide. So, 2–qubit representation is enough to dintinguish four nucleotides as shown in Table 2.2. In addition, this representation gives a pairing condition which says that pairing occurs between the bases which are different states in each qubit.

Table 2.2 2–qubits representations of bases from informatic perspective

Bases $\left\|k\right\rangle$	*Structure*[*]	*Rare tautomeric form*[**]	*2–qubit representation*
$\mathrm{A}=\left\|1'\right\rangle$	$\mathrm{R}=\left\|0\right\rangle$	$\mathrm{Imino}=\left\|0\right\rangle$	$\left\|0\right\rangle\left\|0\right\rangle$
$\mathrm{T}=\left\|2'\right\rangle$	$\mathrm{Y}=\left\|1\right\rangle$	$\mathrm{Enol}=\left\|1\right\rangle$	$\left\|1\right\rangle\left\|1\right\rangle$
$\mathrm{G}=\left\|3'\right\rangle$	$\mathrm{R}=\left\|0\right\rangle$	$\mathrm{Enol}=\left\|1\right\rangle$	$\left\|0\right\rangle\left\|1\right\rangle$
$\mathrm{C}=\left\|4'\right\rangle$	$\mathrm{Y}=\left\|1\right\rangle$	$\mathrm{Imino}=\left\|0\right\rangle$	$\left\|1\right\rangle\left\|0\right\rangle$

"*", R is used for purine structure and Y is used for pyrimidine structure.

"**", Unrare amino is coded by $\left|1\right\rangle$ and unrare keto form is coded by $\left|0\right\rangle$.

Unpaired nucleotide $\left|k\right\rangle$ exists in excited state $\left|1\right\rangle$ which has higher energy than ground sate $\left|0\right\rangle$ of paired ones. This energy difference is defined as binding energy and it is likely to be released as an inelastic collision or phonon after bond formation. So, there should also be another 2–entangled state representation in order to examine bond formation during base pairing. One qubit is needed for the state of the nucleotide and another one is needed for the state of the energy quanta which will be released. [16]

In this sense, interaction Hamiltonian *H* should raise the state of nucleotide when it lowers the state of energy quanta and vice versa. Let's construct *H*

$$H_{\mathrm{int}}=\Delta E_H\left(a^{+}{}_{base}\otimes a_{\ quanta}+a_{\ base}\otimes a^{+}{}_{quanta}\right) \tag{2.43}$$

$$=\Delta E_H\left(\begin{pmatrix}0&1\\0&0\end{pmatrix}\otimes\begin{pmatrix}0&0\\1&0\end{pmatrix}+\begin{pmatrix}0&0\\1&0\end{pmatrix}\otimes\begin{pmatrix}0&1\\0&0\end{pmatrix}\right)$$

$$=\Delta E_H\left(\begin{pmatrix}0&0&0&0\\0&0&1&0\\0&0&0&0\\0&0&0&0\end{pmatrix}+\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&1&0&0\\0&0&0&0\end{pmatrix}\right)$$

$$=\Delta E_H\begin{pmatrix}0&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&0\end{pmatrix}$$

Eigenvalues and eigenstates of H can be easily calculated as shown below respectively.

$$H_{\text{int}} - \lambda I_{4x4} = \begin{pmatrix} -\lambda & 0 & 0 & 0 \\ 0 & -\lambda & \Delta E_H & 0 \\ 0 & \Delta E_H & -\lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{pmatrix} = 0 \tag{2.44}$$

$$\Rightarrow -\lambda^2\left(\lambda^2 - \Delta \mathrm{E_H}^2\right) = 0$$

$$\Rightarrow \lambda_{1,2} = \mp \Delta \mathrm{E_H}$$

$$H_{\text{int}}\vec{\upsilon}_{1,2} = \mp\lambda_{1,2}\vec{\upsilon}_{1,2} \tag{2.45}$$

$$\Rightarrow \Delta E_H \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ x \\ y \\ 0 \end{pmatrix} = \mp\Delta E_H \begin{pmatrix} 0 \\ x \\ y \\ 0 \end{pmatrix}$$

$$\Rightarrow x_{1,2} = \mp y_{1,2}$$

$$\Rightarrow \vec{\upsilon}_{1,2} = 1/\sqrt{2}\begin{pmatrix} 0 & 1 & \mp 1 & 0 \end{pmatrix}^T = 1/\sqrt{2}\left(|0\,1\rangle \mp |1\,0\rangle\right)$$

After that unitary time evolution operator can be constructed by spectral decomposition. ($\hbar = 1$)

$$U = e^{-iH_{\text{int}}t_b} \tag{2.46a}$$

$$= \frac{1}{2}\left(|01\rangle - |10\rangle\right)e^{i\Delta E_H t_b}\left(\langle 01| - \langle 10|\right) + \frac{1}{2}\left(|01\rangle + |10\rangle\right)e^{-i\Delta E_H t_b}\left(\langle 01| + \langle 10|\right) \tag{2.46b}$$

Since $\langle\psi'|\psi'\rangle = \left(\langle\psi|e^{i\Delta E_H t_b}\right)\left(e^{-i\Delta E_H t_b}|\psi\rangle\right) = \langle\psi|\psi\rangle$, global phase $e^{-i\Delta E_H t_b}$ has no effect on the final probabilities. However, relative phase between paired and unpaired bases affects propabilities. Let relative phase be φ. Then, evolution of an unpaired base can be formulated as shown below.

$$U|1\rangle|0\rangle = \varphi|0\rangle|1\rangle \tag{2.47a}$$

φ can be calculated if R.H.S. of the equation (2.46b) is put into (2.47) as U

$$U|1\rangle|0\rangle = \frac{1}{2}\left(|01\rangle - |10\rangle\right)e^{i\Delta E_H t_b}\left(\langle 01| - \langle 10|\right)|01\rangle + \frac{1}{2}\left(|01\rangle + |10\rangle\right)e^{-i\Delta E_H t_b}\left(\langle 01| + \langle 10|\right)|01\rangle$$

$$= -\frac{1}{2}\left(|01\rangle - |10\rangle\right)e^{i\Delta E_H t_b} + \frac{1}{2}\left(|01\rangle + |10\rangle\right)e^{-i\Delta E_H t_b}$$

$$= -\frac{1}{2}\left(e^{i\Delta E_H t_b} - e^{-i\Delta E_H t_b}\right)|01\rangle + \frac{1}{2}\left(e^{i\Delta E_H t_b} + e^{-i\Delta E_H t_b}\right)|10\rangle \qquad (2.47b)$$

In order to keep equality of (2.47a) and (2.47b), $e^{i\Delta E_H t_b} + e^{-i\Delta E_H t_b}$ should be zero, whereas φ is equal to $-e^{i\Delta E_H t_b} + e^{-i\Delta E_H t_b}$. First, use former identity and then use second one.

$$e^{i\Delta E_H t_b} + e^{-i\Delta E_H t_b} = 0 \qquad (2.47c)$$

$$\Rightarrow e^{-i\Delta E_H t_b} = -e^{i\Delta E_H t_b} = -1/e^{-i\Delta E_H t_b}$$

$$\Rightarrow e^{-i\Delta E_H t_b} = \sqrt{-1} \qquad (2.47d)$$

$$\Rightarrow \varphi = 1/2\left(2e^{-i\Delta E_H t_b}\right) = \sqrt{-1} \qquad (2.47e)$$

It means that φ is a geometric phase which makes crossovers between quantum and classical domains possible in order to make transformations stable against decoherent fluctuations of cellular environment.

Hydrogen bonds during base pairing is highly sensitive to distance because tunneling amplitude strongly depends on the shape of energy barrier. Since lenght of the hydrogen bonds in each base pairing have only two different values, it can be said that there are two effective consequtive hydrogen bondings in each base pairing. In other words, deexcitaion of a nucleotide should be a two-step process. This inference duplicates relative phase between unpaired and paired nucleotides. Since φ^2 is equal to minus one, conditional rotation due to Oracle can represent base pairing via hydrogen bonding.

To obtain this inference, it is assumed that there is no phase information in energy quanta and entanglement is likely to be broken as soon as the effect of the quanta's release on quantum coherence is minimum. [16]

2.2.1.2 Quantum Search Scenarios

2.2.1.2.1 Quantum Mechanical Search Scenario

Initiation of the replication requires symmetric superposition state $|s\rangle$ of nucleotides in this scenario. Special environment of the polymerase's active site and interaction with DNA can make transitions between chemically different nucleotides possible. In this sense, after an arbitrary nucleotide $|k\rangle$ is picked up from the surrounding environment, it should be relaxed to state $|s\rangle = 1/\sqrt{N}\sum_k |k\rangle$ in t_r seconds. In addition, the nucleotide in state $|s\rangle$ should be oscillate between different components of $|s\rangle$ with a period t_{osc} that can be determined by magnitudes of transition matrix elements. It is not likely to calculate these magnitudes, but t_{osc} should be quite smaller than t_r since this oscilation is a part of relaxation [16].

When oscillation brings nucleotide to the right one, orientation should be fixed by a sudden hydrogen bond formation via proton tunelling. So, oscillation period should be quite bigger than binding time t_b. [16]

$$t_b << t_{osc} << t_r \tag{2.48}$$

This binding should transform symmetric superposition state to the state $O|s\rangle|0\rangle = 1/\sqrt{N}\left(\sum_{k\neq i}|k\rangle|0\rangle - |i\rangle|2\rangle\right)$. However, nucleotide should still have a tendency to relax towards $|s\rangle$ in the presence of special environment of the polymerase's active site and interaction with DNA.

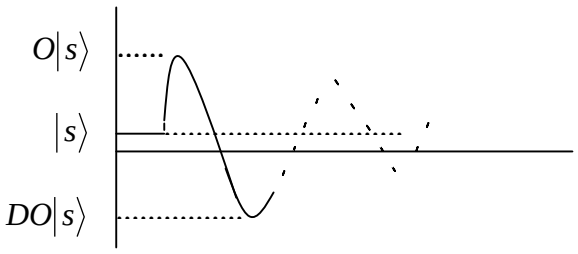

Figure 2.5 Relaxation towards symmetric superposition state [16]

When relaxation brings nucleotide to the state $DO|s\rangle|0\rangle = |i\rangle|2\rangle$, energy quanta should leave the system with minimum disturbance to quantum coherence.

Leaving of the quanta should make hydrogen bonding certain and this certainty should be supplier of the projective measurement. Impossibility of quanta's return should give an irreversibility to the process and replication should go along. [16]

Since binding energy of a single hydrogen bond is approximately $7k_BT$ [16], order of t_b can be determined by energy–time uncertainty principle.

$$\Delta E_H t_b \approx \hbar \tag{2.49a}$$

$$\Rightarrow t_b \approx 1.055\times10^{-34}\,J\,\text{sec}/\,7\times1.3806\times10^{-23}\,JK^{-1}\times300K = 3.63\times10^{-15}\,\text{sec}$$

Moreover, this femto second order can be supported by the equation (2.47d)

$$e^{-i\Delta E_H t_b/\hbar} = \sqrt{-1} \Rightarrow \Delta E_H t_b = (3\pi/2)\hbar \tag{2.49b}$$

$$\Rightarrow t_b = 3.63\times10^{-15}\,\text{sec}\times 3\pi/2 = 17.11\times10^{-15}\,\text{sec}$$

In additional to t_b, order of t_r can be also estimated in spite of the fact that there are a lot of variables that participate in it's determination. Elongation rate is 30,000 bp per minute in bacteria and this rate is mainly determined by t_r. Thus, order of t_r should be somewhat about 10^{-2} seconds. Consequently, oscillation period of symmetric superposition state should be somewhere between $O(10^{-12})$ and $O(10^{-5})$ according to equtaion (2.48).

2.2.1.2.2 Wave Mechanical Search Scenario

This scenario is based on the analogue version of the Grover's algorithm. Initiation of the base pairing requires center-of-mass mode in which there is a synchronisation between N identical harmonic oscillators that are coupled to a big, motionless oscillator. Interaction of each nucleotide with polymerase's active site can be imagined as an identical harmonic oscillator if sugar or phosphate parts of nucleotides mainly interact with enzyme's active site. If so, a back and forth bulk movement of enzyme–nucleotides complex in the direction that is perpendicular to the elongation direction can start base pairing when right nucleotide comes to the suitable orientation for proton tunelling. In this sense, proton tunelling between nucleotide and template DNA strand should be the cause of the tapping which shifts the phase of the right oscillator by half a period. Also, the period of this back and forth bulk movement should be π in order to redistribute kinetic energy in a way that increases the amplitude of right oscillator.

Amplitude amplification by this way should increase the energy of the right nucleotide by energy transferring from other oscillators. When this energy reachs an appropriate level to be bound to both template and growing strands, enzyme–nucleotides complex should be deconstructed and replication should go along.

2.2.2 Language of the Proteins

Appropriate scenarios can be constructed for translation in a same way with replication (or transcription) if quantum query number Q is chosen as 3 instead of 1. In fact, Q should be 3.07 to distinguish 21 different elements (20 building blocks and 1 stop signal) according to equation (2.42), but non-integer part can not be realized. So, algorithm has an intrinsic error when it is used in translation.

In addition, proton tunnelling should occur less efficiently in translation, since lengths of the hydrogen bonds in base pairing between thefirst base of anticodon and the third base of codon are changed by streric effects. This less efficient proton tunneling should be insufficient to conditional rotation of the amplitudes at the third query. So, amplitude amplification should not increase the probability (or energy) of the right nucleotide to the desired value. It means that the algorithm has an extrinsic error when it i used in translation.

All in all, it can be said that quantum database search can explain the relation between triplet genetic code and length of the alphabet of the protein language. It is also successful in the explanation of the accuracy difference between replication and protein synthesis. However, if genetic code was doublet, it could distinguish 11 different elements (10 building blocks and 1 stop signal) with an intrinsic error in algorithm. ($Q = 2 \Rightarrow N = 10.5$) So, why is genetic code triplet? Alternatively, why does the alphabet of the protein language include only 20 amino acids among more than 700 natural amino acids? If usage of quantum algorithm is actually related to an optimization during evolution, answers should be avaliable from this bioinformatic perspective.

2.2.2.1 Structural Information and Proteomic Language

Operational and functional meanings of the proteomic words depend on 3–dimensional shape and size of the proteins. This structural information is translated from a read–only memory (DNA) as 1–dimensional amino acid chains and amino acid sequences determine the rigid body transformation (folding) of proteins.

Rigid body transformations are directed by translation and rotation operators. Since biological languages require discretization of the information, the groups that are formed by these operations should be discretized. Such a discretization of space can be examined by lattices at molecular level. [21, 25]

Simplex is the simplest lattice discretization unit by which exact description of *d*–dimensional space is possible. This volume element consists of *d*+1 points which can represent *d*+1 equivalent and equidistant states. [25] Discretization of 3–dimensional translations is easy if each building block takes up equal space like component of amino acids in the backbone of the polypeptide chain. However, discretization of 3–dimensional rotations is not so easy and requires the largest set of equivalent and equidistant states which can be obtained by a 3–dimensional simplex, a tetrahedron. In this sense, building bocks of a language which encodes structural information should have a tetrahedral geometry. [21]

If quantum mechanics is used in the determination of building blocks, orthogonality is needed for the formation of Hilbert space. Tetrahedral quantum states satisfies this orthogonality condition and they can be formed by the sp^3 hybridization of the atomic orbitals. [21, 25] Both carbon and silicon atoms use sp^3 hybridization for covalent bonding, but silicon atoms do not have a tendency to form aperiodic chains. So, words that carry structural information should be written by buiding blocks which include a carbon atom with four covalent bonds. [21]

In addition, if these building blocks are amino acids, they should be α–amino acids. This is because of the fact that only the geometry of α–amino acid chains fit to a lattice which has a tetrahedral unit. Backbone of the α–amino acid chains consists of N–C_α–C repetitions and atoms that are covalently bound to central C_α atom can be placed into the corners of a tetrahedra. However, backbone of β–amino acids consists of N–C_β–C_α–C repetitions and atoms that are found in the backbone can not be placed into the corners of a tetrahedra. Backbones including amino acids that are in other forms than α and β have more than two carbon atoms between N and C atoms in each repetitive unit.

There is an isomorphism between tetrahedral group and permutation group P_4. Twelve elements of this group are rotation elements and they correspond to even permutations in P_4. The other twelve elements which correspond to odd permutations

in P_4 are reflection elements and they are not responsible for rotation operations, but for parity operations.

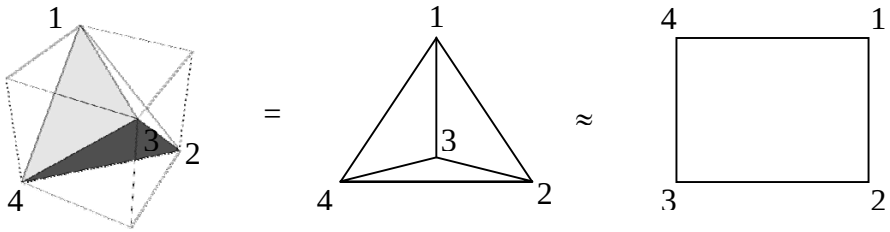

Figure 2.6 Tetrahedral Geometry Left hand side of the figure shows formation of a regular tetrahedron whose centre overlaps with the centre of cube. Whereas, right hand side of the figure represents the isomorphism between tetrahedral group and permutation group P_4.

Alternating vertices of a cube can form a regular tetrahedron whose centre overlaps with centre of the cube. Then, rotation elements can be classified into three groups: identity element, elements of the rotations around 2–fold axes and elements of the rotations around 3–fold axes. 2–fold axes are the axes which join the centre of opposite faces in cube. They also join the midpoints of the disconnected edges of tetrahedra. 180^0 rotations about these axes corresponds to the elements $(1\,2)(3\,4)$, $(1\,3)(2\,4)$ and $(1\,4)(2\,3)$. 3–fold axes are the space diagonals of the cube. They join the centre of tetrahedron with its corners. 120^0 rotations about these axes corresponds to the elements $(1\,2\,3)$, $(1\,3\,2)$, $(2\,3\,4)$, $(2\,4\,3)$, $(3\,4\,1)$, $(3\,1\,4)$, $(4\,1\,2)$ and $(4\,2\,1)$.

When a carbon atom that has sp^3 hybrid orbitals, is placed at the centre of tetrahedron, 2–fold axes are crossing in the middle of bond agles. If two of the bonds combine together to form a double bond, one of these axes becomes direction of the double bond. Then, 180^0 rotations about them convert configuration of the molecule from trans to cis or vice versa [21].

In the same sense, 3–fold axes are the bond axes and 120^0 rotations about them change the conformation of the molecule. Since there are 8 different rotations about 3–fold axes and 1 identity rotation, tetrahedral geometry allows 9 different conformations.

Diamond lattice is a fcc lattice with a two–point basis. It can be obtained by superimposition of two fcc lattices. Also, each fcc lattice is a simple cubic lattice with four–point basis. These four basis are located in the alternating corners of a cube. So, they can form a tetrahedra. In addition, one base of the other fcc lattice in

diamond structre, is located at the centre of this cube. So, diamond lattice has a tetrahedral unit which allows 9 different conformations.

Suppose that an α–amino acid chain elongates on a diamond lattice. This supposition requires two simplifications. First simplification is to take all the bonds in a same length. Second simplification is about bond angles. Peptide bond angles have an average about 120^0. [21] However, bond angles are 109.5^0 in tetrahedron as shown below.

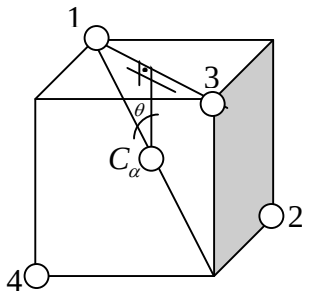

Figure 2.7 Bond formation in tetrahedral geometry Let edge of the cube be *a*. Then tangent of the angle θ should be equal to $a\sqrt{2}/a = \sqrt{2}$. This makes the bond angle 2θ be 109.5^0.

As mentioned in Figure 1.13, C_α–C–N–C_α atoms in the backbone of α–amino acid chains, are co-planar with O and H atoms because the rotation around C–N bond is not allowed. So, when a new amino acid is added to polypeptide chain, locations of the N, H and C_α atoms of this amino acid should be fixed. However, there are 9 possible orientations left for C and O atoms in this amino acid. Three locations are available for C atom when location of C_α atom is fixed. After location of C atom is fixed, there are also three possible locations for O atom in order to define a plane with C_α and C atoms. (see Figure 1.14) Therefore, there are 9 different orientations for each α–amino acid of a polypeptide chain which elongates on a diamond lattice.

Since this possible orientation number coincides with the allowed conformation number, it can be said that choice of the α–amino acids as the building blocks of proteins is related to information type which is encoded by proteomic language. However, why does proteomic alphabet include only 20 α–amino acids?

2.2.2.2 Scenario for the Evolution of Triplet Code

Each amino acid addition to polypeptide chain has an effect on the orientation of the previous amino acids. So, proteomic words are overlapping structural codes. It means that, in order to maximize diversity on 3–dimensional shape space by using all of the possible 9 orientations, proteomic alphabet should contain at least 9 different amino acids. [21] This alphabet length is close to the one that is required in doublet code. Both of the classical information theory (see Table 1.2) and quantum search algorithm limit the distinguishable element number to 11. Since one of these elements is stop signal, a doublet code can distinguish 10 amino acids at most. Therefore, predecessor of the triplet code can be a doublet code according to this optimization criterion since proteins evolved before genes. [21, 26] If so, evolution of the triplet code would require further optimization criteria.

If predecessor of the triplet code is really a doublet code and this doublet code evolved according to described criterion, it can also explain the evolutionary choice between stereoisomer types. In fact, 9 possible orientations are determined by considerations on the allowed values of dihedral angles (ψ, ϕ). Locations for C atom is determined by allowed values of dihedral angle ϕ, whereas locations for O atom is determined by allowed values of dihedral angle ψ. (see Figure 1.14) In additional to these orientations, there are two different location possibilities for R group and H atom that are bound to the same C_α with C and O atoms. Each of these possibilities corresponds to one isomer type and so, isomer type does not differ orientation number. However, stable α–helix and β–sheet conformations can not be constructed if mixed isomer types are used in protein synthesis. [21] Therefore, in order to use all of the possible 9 orientations effectively, proteomic building blocks could be chosen from only one stereoisomer type.

As mentioned in the introduction section, common amino acids are divided into two classes according to class of their cognate aaRS enyzme and each class includes 10 members. In addition, amino acids that belong to different classes, but have similiar chemical properties, differ only by the size of their R groups. It seems so that there can be relation between this classification and a doublet predecessor. Since a class lable is required when two independent doublet codes for two different amino acid classes are combined into one code, emergence of the third base of the

codon could be result of a such combination during evolution. Also, such an evolution of a binary valued label can explain the binary effect of third base on the amino acid meaning and error tolerance maximization in codon–amino acid matching. [21]

In fact, new genes can emerge in several ways. First, intragenic mutations can modify existing genes, but it is very dangerous by oneself. Gene duplication can make it more safe. When intragenic mutations modifies one of the copies, other copy can implement its normal tasks. Also, segment shuffling can cause emergence of new genes. Several segments of two or more genes can be combined into a new gene. For example, such duplication and divergence mechanisms are responsible for the emergence of 47% of the genes in bacteria *Bacillus subtilis*. [27]

In addition, horizontal transfer of functional genes from different species is possible. Such horizontal transfers are seen generally between procaryotes. However, after ancestral eucaryotic cell started to a symbiotic relationship with bacterial ancestors of mitocondria and choloroplasts, many of symbiont genes coding for proteins required inside mitocondria and choloroplasts were moved into eucaryotic genome while some of them were kept in symbiont genome inside corresponding organalles. [27]

Since duplication and horizontal transfer are the main mechanisms used in the evolution of the genetic code, they could be used in the transition from doublet code to triplet code. [26] Smaller class II amino acids are more likely to be primitive amino acids. If so, duplicated copy of doublet code could make the protein structures more stable. This is because of that, R groups of class I amino acids can fill up big cavities which can not be sufficiently filled up by small R groups of class II amino acids. [21]

As mentioned before, two of class II amino acids, proline and glycine can cause β–turns between distinct succesive secondary structures which are required in the formation of globular proteins. It shows that class II amino acids have the ability to form proteins which have special functions. This ability increases the possibility described above. In addition, this ability can explain the increase in amino acid number from 9 to 10. Proline residues at the locations of β–turns are found in cis configuration and cis configuration does not fit in the fcc diamond lattice. So, after the optimization that allows 9 distinguishable building blocks, there could be a

further optimization that allows β–turns. Since a doublet code can distinguish 10 amino acids at most, proline that can be present in both of trans or cis configurations could be added to the primitive proteomic alphabet.

Finally, this scenario can be supported with the fact that operational code in tRNAs is the candidate for being the predecessor of genetic code. [26] Since operational code consists of the base pairs at the positions 1:72, 2:71, 3:70 and unpaired 73^{th} base of tRNA, base pairs at the positions 1:72 and 2:71 can be related to first base of the doublet code, whereas base pair at the position 3:70 and unpaired 73^{th} base can be related to second base of the doublet code or vice versa.

2.2.3 Role of the Enzymes

Previously described scenarios for both of quantum search and evolution of triplet code hypothesize alternative roles for enzymes. First of all, enzymes in quantum search scenarios provide an environment which minimizes the decohorent fluctuations of cellular environment. Quantum mechanical proton transfer through a hydrogen bond requires such a protective environment. If so, this passive role of the enzymes can explain the fact that base pairing does not occur randomly in the cellular environment. Also, it can explain the proofreading mechanisms in replication and transcription. As long as the enzymes keep quantum coherence, these processes can be reversible. [16]

According to quantum mechanical search scenario, enzymes can create superpositions of chemically distinct molecules. It is likely to be when dynamic covalent structure of nucleotides (see introduction) and cut–and–paste jobs done by enzymes are considered. If so, enzymes should enable tunnelling through energy barrier rather than lowering activation energy. This ability to control transition matrix elements can also explain absence of nucleotide U in DNA and absence of nucleotide T in RNA. [16]

In addition, stereoisomer selection way of biological processes can be understood in this sense. Chirality flips are allowed quantum transformations which are directed by parity operations mentioned section 2.2.2.1. However, molecular size of amino acids make chirality flips more diffucult and racemase enzymes convert D–isomers into L–isomers in living cells [21]. If enzymes can provide an environment which minimizes decohorent fluctuations of cellular environment and control

transition matrix elemets between components of superposition states, they can make such chirality flips more easy.

However, wave mechanical search scenario hypothesizes a third mechanism for enzyme activity in reactions. According to this scenario enzymes can interact with several identical substrate molecules at same time and then they can make energy transfer from all these interactions to one interaction between one of the substrates and its cognate molecule.

Consequently, standard model for enzyme activity mechanism which includes complex hydrogen bond patterns based on complex molecular shapes, can be replaced with well defined simple physical models that can generate complex dynamics, like coupled harmonic oscillator systems. By this way, both similarities and differences between different enzyme activities can be understood more precisely.

3 DISCUSSIONS AND FUTURE DIRECTIONS

There are a lot of molecules each of which has one or more special functions in both of the storage and usage of the genetic information. When processes are tried to be defined with a basis of biochemical interactions, genetic machinery seems to be complex system. However, this research has been shown that main processes involved in genetic machinery can be defined as unsorted database search in which a quantum search algorithm is used. Then, a bioinformatic approach can be used in the understanding of the determination of chosen building block structures and distinguishable building block numbers in the storage and usage of genetic information. In the mean time, complex biochemical interactions can be considered as simple and well defined physical systems that generate complex dynamics. Therefore, it can be said that quantum bioinformatic approaches can be useful in the modelling and understanding of complex biological realities such as DNA replication and protein synthesis.

In this research, some assumptions have been done to make a relation between quantum search algorithm and base pairing mechanism and these assumptions should be checked. First of all, alternative roles that have been given to enzymes should be experimentally verified. Then, absence of phase information in energy quanta should be based on some theoretical knowledge. Finally, effects of the release of entangled quanta on quantum coherence should be clarified by both theoretically and experimentally. If these assumptions can be supported by further researches, knowledge about the nature of biochemical systems can be boradened beyond the ones obtained with the perspectives of modern molecular biology and biochemistry.

There is a common consensus in computational biology like that stability of protein structure depends on its energy configuration. However, it has been assumed that locally stable configurations can be enough to make proteins structurally strong. This assumption should be supported by computational models. In addition, effects of the aqueous environment has not been considered in lattice modelling of potein folding. Interactions between water molecules and R groups of amino acids should be also included in computational models. Morever, it should be checked whether there is a relation between R group properties of common amino acids and 9 possible orientations based on the rotations around 3–fold axes.

In addition, some assumptions have been done to undertsand evolution of the triplet code. First, class II amino acids have been assumed as the primitive proteomic building blocks. In order to support this assumption, it should be found that primitive proteins had consisted of mainly class II amino acids. Alternatively, synthesis of functional proteins by using only class II amino acids can support this assumption. Secondly, duplication and divergence of the doublet code have been related to optimization of the protein structure stability. This relation should be also clarified.

Some gaps in this research are filled by the results of other researches based on classical information theory and communication theory. Since these theories can be expanded to the quantum ones, communication between genetic and proteomic languages can be examined by quantum information theory and quantum communication theory. Perhaps, such researches can be related to this research and consequently, quantum aspects of life can be explained in a way that is not possible with the tools of classical biology.

REFERENCES:

[1] **Snustad, D. P. and Simmons, M. J.** (2003). "*Principals of Genetics*" (Third Edition). John Wiley & Sons, Inc, USA.

[2] **Nelson, D. L. and Cox, M. M.** (2005). "*Lehninger Principals of Biochemistry*" (Fourth Edition). W. H. Freeman and Company, New York.

[3] **Xia, X.** (2007). "*Bioinformatics and the Cell: Modern Computational Approaches in Genomics, Proteomics, and Tripcriptomics*" (First Edition). Springer Scince+Business Media, New York.

[4] **Mandelkern, M., Elias, J., Eden, D., Crothers, D.** (1981). "The dimensions of DNA in solution". *Journal of Molecular Biology* **152** (1): 153–61.

[5] **Gregory, S., *et al.*** (2006). "The DNA sequence and biological annotation of human chromosome 1". *Nature* **441** (7091): 315–21.

[6] Wiki files {http://en.wikipedia.org/wiki/File:DNA_replication.svg}

[7] Internet Files

{http://www.biology.lsu.edu/heydrjay/1201/Chapter17/SCI_Amino_Acid_CIRCLE.jpg}

[8] **Tlusty, T.** (2007). "A model for the emergence of the genetic code as a transition in a noisy information channel". *Journal of Theoritical Biology* **249 :** 331-347.

[9] **Ribas de Pouplana, L. and Schimmel, P.** (2001). "Operational RNA Code for Amino Acids in Relation to Genetic Code in Evolution". *Journal of Bioloical Chemistry* **276** (10): 6881–6884.

[10] **Carey, F. A.** (2008). "*Organic Chemistry*" (Seventh Edition). Mc-Graw Hill, New York.

[11] **Hovmöller, S., Zhou, T., and Ohlson, T.** (2002). "*Conformations of amino acids in proteins*". *Acta Crystallographica D* **58** : 768-776.

[12] **Amaral, L. A. N. and Ottino, J. M.** (2004). "*Complex networks: Augmenting the framework for the study of complex systems*". *The European Physical Journal B* **38** : 147-162.

[13] **van der Lubbe, J. C. A.** (1997). "*Information Theory*" (First Edition). Cambridge University Press, U.K.

[14] **MacKay, D. J. C.** (2003). "*Information Theory, Inference, and Lerarning Algortihms*" (--- Edition). Cambridge University Press, U.K.

[15] **Patel, A.** (2008). "*Towards Understanding the Origin of Genetic Languages*" in "*Quantum Aspects of Life*" (Editors: Abbott, D., Davies, P. C. W., Pati, A. K.), Imperior College Press, UK.

[16] **Patel, A.** (2000). "*Quantum Algorithms and the Genetic Code*" in *Proceedings of the Winter Institute on Foundations of Quantum Theory and Quantum Optics* (Calcutta), *Paramana* **56** : 367-381 {\tt quant-ph/0002037}

[17] **Nielsen, M. A. and Chuang, I. L.** (2000). "*Quantum Computation and Quantum Information*" (First Edition). Cambridge University Press, U.K.

[18] **Grover, L. K.** (1996). "*A fast quantum mechanical algorithm for database search*" in *Proceedings* of *28th Annual ACM Symposium on the theory of Computing* (STOC), 212-219. {\tt quant-ph/9605043}

[19] **Zalka, C.** (1999). "*Grover's quantum searching algorithm is optimal*" *Physical Reviews A,* **60** : 2746-2751. {\tt quant-ph/9711070v2}

[20] **Patel, A.** (2006). "*Wave Algorithms: Optimal Database Search and Catalysis*" in *Proceedings of American Institute of Physics,* 261-272, {\tt quant-ph/0609042}.

[21] **Patel, A.** (2002). "Carbon---The First Frontier of Information Processing". *Journal of Biosciences* **27** : 207-218

[22] **Isaacs, E. D., *et al*.** (1999) "*Covalency of the Hydrogen Bond in Ice: A Direct X-Ray Measurement*". *Physical Review Letters,* **82** (3) : 600-603.

[23] **Tuckerman, M. E., *et al*.** (2009) "*On the Quantum Nature of the Shared Proton in Hydrogen Bonds*" *Science* **275** : 817-820.

[24] **Golo, V. L., Kats, E. I., and Peyrard, M.** (2001). "*Concerted motion of protons in hydrogen bonds of DNA-type molecules*" *JETP Letters* **73**(4): 202-206 {\tt cont-mat/0101372v2 (2008)}; **Golo, V. L. and Volkov, Yu. S.** (2001) **"***Tautomeric transitions in DNA***"**{\tt cont-mat/0110599v1}

[25] **Patel, A.** (2003). "*Mathematical Physics and Life*" in "*Computing and Information Sciences: Recent Trends*" (Editors: Misra, J. C., Goswami, A. and Kumar, P.), Narosa Publising House, India : 271-294 {\tt quant-ph/0202022}

[26] **Patel, A.** (2005). "*The Triplet Genetic Code had a Doublet Predecessor*" *Journal of Theoritical Biology*, *233* : 527-532, {\tt q-bio.GN/0403036}

[27] **Alberts, B., *et al*.** (2002). "*Moleculr Biology of the Cell*" (Fourth Edition) Garland Science, New York.