Genetics Basics: Exploring DNA & Heredity

CHAPTER 5: The Genetic Code & Translation

5.1. Introduction

Within the expansive fabric of life on our planet, an elaborate molecular communication system known as the genetic code intricately regulates the fundamental nature of being. The genetic code, which is present in the DNA of all living organisms, functions as a fundamental plan that coordinates the production of proteins, which are essential components of life. The mechanism through which this code is decrypted and converted into operational proteins is a remarkable feat of biological engineering, elucidating the mysteries surrounding the emergence of life's variety and intricacy from a comparatively basic set of principles. This chapter delves into the captivating realm of the genetic code and translation, revealing the fascinating interaction between genotype and phenotype, the relationship between protein structure and function, the universal nature and redundancy of the genetic code, the complexities of the translation process, the significance of RNA-RNA interactions, and the application of these concepts in understanding genetic disorders such as Lesch-Nyhan Syndrome (Koonin & Novozhilov, 2009).

The fundamental essence of all living organisms resides in a central storage unit of genetic data, which is contained within the molecules of DNA (deoxyribonucleic acid). The data presented in the form of nucleotide sequences contain the encoded instructions necessary for constructing and sustaining the complex mechanisms that comprise the fundamental essence of life (Ambrogelly et al., 2007).

The chapter delves into the fundamental concept of the intriguing correlation between genotype and phenotype. The term "genotype" pertains to an organism's precise genetic constitution, denoted by the arrangement of nucleotides in its DNA sequence. In contrast, the concept of phenotype encompasses the range of observable characteristics and traits displayed by an organism, encompassing both physical appearances and physiological functions. The comprehension of the interaction between phenotype and genotype is crucial for unraveling the source of an organism's traits and their inheritability (Neumann, 2012).

The Lesch-Nyhan Syndrome, an uncommon hereditary condition, exemplifies the intricate correlation between genotype and phenotype. The etiology of this syndrome can be attributed to a genetic mutation occurring in the HPRT1 gene, resulting in a deficiency of the hypoxanthine-guanine phosphoribosyltransferase (HPRT) enzyme. The manifestation of major neurological symptoms and behavioral disruptions, including self-injurious behaviors, occurs due to the biochemical imbalance in purine metabolism. Examining this disorder yields significant insights into the significance of particular genetic mutations and their influence on an individual's phenotype, thereby providing a glimpse into the intricacies of genetic disorders (Bossi & Roth, 1980).

Proteins, which serve as the primary functional components within cells, are intricately associated with the genetic code through the biological process known as translation. Proteins, which consist of lengthy sequences of amino acids, perform a diverse range of tasks, including facilitating biochemical reactions and offering structural reinforcement (Rodin et al., 2011).

The genetic code functions via a trinucleotide-based system, where groups of three nucleotides, referred to as codons, are associated with distinct amino acids or act as initiation or termination signals for the process of protein synthesis. The genetic code consists of a total of 64 codons, which are formed by the combination of four nucleotide bases: adenine, cytosine, guanine, and thymine. Among the total of 64 codons, 61 are responsible for encoding amino acids, whereas the remaining three serve as stop codons, indicating the termination of protein synthesis. The present codon system exemplifies the remarkable efficiency and accuracy of the genetic code in converting a wide range of nucleotide sequences into a limited set of amino acids (Yarian et al., 2002).

One notable characteristic of the genetic code is its nearly universal presence among all extant organisms. The genetic code exhibits conservation across various organisms, ranging from bacteria to animals and plants, thereby highlighting the shared evolutionary origins of life on our planet. Moreover, the genetic code exhibits a certain level of depravity, whereby multiple codons have the ability to encode for the identical amino acid. The presence of redundancy in protein synthesis provides a mechanism for enhancing robustness, serving as a protective measure against mutations that could potentially have negative consequences on the observable characteristics of an organism (Novozhilov et al., 2007).

The process of translation is a crucial component of cellular biology, whereby the genetic information contained within the DNA molecule is translated into messenger RNA (mRNA) and then transformed into proteins. The process of protein synthesis occurs within the ribosomes, which are sophisticated molecular structures that play a crucial role in coordinating the incorporation of amino acids to form fully functional proteins. The translation process consists of three primary phases: initiation, elongation, and termination. A distinct set of factors and regulatory elements controls each stage (Hoffmann, 1974).

The genetic code serves as the fundamental basis for the translation process. Still, it is vital to note that various RNA-RNA interactions play a significant role in refining this process and regulating the expression of genes. RNA-binding proteins, microRNAs, and tiny regulatory RNAs are prominent constituents of this regulatory network. Through the modulation of translation rate and efficiency, these interactions facilitate cellular responses to environmental stimuli and exert precise regulation over the expression of particular genes. Consequently, they play a pivotal role in the organism's adaptive capacity to its surroundings (Wong et al., 2016).

In summary, the chapter pertaining to "The Genetic Code and Translation" explores the intricate molecular mechanisms that serve as the foundation for the remarkable phenomena observed in living organisms. This chapter elucidates the complexities of the genetic code, its relationship with the

structure and function of proteins, and the remarkable phenomenon of translation. It explores the biological language governing all living organisms' fundamental design. In addition, an examination of the genotype-phenotype correlation, as demonstrated by genetic disorders such as Lesch-Nyhan Syndrome, underscores the significant influence of genetic data on the intricacy and variety of living organisms. The examination of these fundamental biological processes enhances our comprehension of the interconnectedness of living organisms and provides invaluable perspectives into the marvels of the natural realm (Vetsigian et al., 2006).

5.1.1. History

The historical trajectory of the genetic code is a compelling narrative characterized by rigorous scientific investigation, encompassing numerous years of pioneering breakthroughs and cooperative endeavors among esteemed scholars in the disciplines of genetics, biochemistry, and molecular biology. The comprehension of life's fundamental processes has been significantly influenced by the elucidation of the crucial molecular language that regulates the conversion of genetic information into proteins. Let us undertake a scholarly exploration of the significant milestones in the process of unraveling the genetic code throughout history (Shandell et al., 2021).

The origins of the endeavor to unravel the genetic code can be attributed to the groundbreaking contributions of Gregor Mendel, widely regarded as the progenitor of contemporary genetics. During the mid-19th century, Gregor Mendel conducted a series of experiments involving pea plants, which established the fundamental principles underlying the concept of heredity. These groundbreaking investigations catalyzed subsequent generations of scientists, motivating them to delve into the genetic underpinnings that contribute to the remarkable diversity observed in living organisms (Knight et al., 1999).

In the 20th century, there was a notable convergence between genetics and biochemistry, which laid the foundation for significant breakthroughs and transformative advancements. During the 1930s, Friedrich Miescher, a biochemist, made a significant discovery by identifying DNA as a distinct substance present within the cell nucleus. This seminal finding served as the foundation for subsequent investigations into the nature and properties of nucleic acids.

During the 1950s, the utilization of X-ray crystallography by Rosalind Franklin and Maurice Wilkins initiated the invention of the double-helical configuration of DNA (Fournier et al., 2010).

The investigation into the genetic code experienced a surge in progress during the initial years of the 1960s, as researchers specializing in genetics and biochemistry directed their focus towards unraveling the mechanisms by which DNA's genetic data is converted into proteins. The primary focus of this study revolved around the inquiry: What is the mechanism by which the four nucleotides (adenine, cytosine,

guanine, and thymine) present in DNA are transformed into the twenty amino acids that comprise proteins (Wolf & Koonin, 2007)?

A significant advancement occurred in 1961 when Marshall Nirenberg and Johann Matthaei made a crucial breakthrough in their research. The researchers performed experiments utilizing synthetic RNA composed solely of uracil (poly-U) and determined that this specific RNA sequence encoded the amino acid phenylalanine. This discovery represents the initial instance in which scientists successfully decoded a particular codon within the genetic code, establishing a foundation for subsequent investigations (Kato, 2019).

Expanding upon the advancements above, Nirenberg, in collaboration with Philip Leder and Har Gobind Khorana, initiated a sequence of experiments during the mid-1960s to unravel additional codons and their respective amino acids. By employing an assortment of in vitro translation systems and synthetic RNA molecules, the researchers successfully elucidated a greater number of codons, thereby establishing the universal nature of the genetic code across diverse organisms (Osawa et al., 1992).

The Nobel Prize in Physiology was awarded in 1968 to Nirenberg, Khorana, and Robert Holley for their significant contributions to unraveling the genetic code. The acknowledgment served to emphasize the importance of their findings in furthering the comprehension of molecular biology (Sun & Caetano-Anollés, 2008).

As the investigation into the genetic code progressed, it became evident that the genetic code exhibits degeneracy. The identification was made that there exists a phenomenon wherein multiple codons can encode for a single amino acid, thereby facilitating redundancy and enhancing the resilience of the process of protein synthesis. This observation served to enhance scientists' comprehension of the efficacy and adaptability of the genetic code (Wiltschi & Budisa, 2007).

During the late 1960s and early 1970s, significant progress was made in decoding the genetic code, leading to a comprehensive understanding of its structure. Consequently, researchers redirected their efforts toward investigating the mechanisms underlying translation and the crucial involvement of transfer RNA (tRNA) in facilitating this process. The identification of transfer RNA (tRNA) molecules, which possess the capacity to recognize codons and convey the corresponding amino acids selectively, established a pivotal connection between the genetic code and the process of protein synthesis (Di Giulio, 1989).

The historical narrative surrounding the genetic code serves as a testament to the formidable impact of scientific cooperation and unwavering determination. The elucidation of the genetic code, starting with Mendel's experiments on pea plants and culminating in the seminal contributions of Nirenberg, Khorana, and other scientists, has had a profound impact on our comprehension of genetics. This breakthrough has not only transformed our understanding of the field but has also laid the groundwork

for significant progress in biotechnology, medicine, and molecular biology. As researchers persist in deciphering the complexities of the genetic code, this captivating endeavor persists as a constantly evolving narrative that contributes to our understanding of the intricate molecular intricacies of life (Freeland & Hurst, 1998).

5.2. Lesch-Nyhan Syndrome & The Relation Between Phenotype & Genotype

In the year 1962, an investigation was conducted by Dr. William Nyhan and Michael Lesch on a male child who was afflicted with a perplexing array of symptoms indicative of a severe illness. The male child exhibited hematuria, elevated levels of uric acid in his bloodstream, and unmanageable muscular contractions affecting his upper and lower extremities. The individual exhibited cognitive impairments and engaged in self-injurious behaviors by biting his fingers and lips. Upon thoroughly examining the male individual, Nyhan and Lesch determined that he was suffering from an unspecified ailment. Subsequently, additional individuals presenting comparable symptoms were documented, leading to the designation of the condition as the Lesch-Nyhan syndrome (Fu et al., 2014).

One of the initial manifestations of Lesch-Nyhan syndrome is the emergence of orange-colored particulate matter, resembling "sand," which is composed of uric acid crystals, observed in diapers within a few weeks following the infant's delivery. Within a span of approximately one year, the child initiates the manifestation of twisting movements in the extremities, specifically the hands and feet, alongside the occurrence of involuntary spasms. Approximately 50% of the children experience seizures. Following a period of approximately two to three years, certain individuals within the child population demonstrate a tendency towards compulsive self-inflicted harm (Genao & Torres, 2014).

Lesch-Nyhan syndrome predominantly manifests in males and is passed on as an X-linked recessive disorder. The disease is attributed to the presence of a defective gene on the sole X chromosome of affected males. In the year 1967, a group of scientists affiliated with the National Institutes of Health conducted a study that concluded that the disease in question arises due to a flawed version of the gene responsible for encoding the enzyme hypoxanthine-guanine phosphoribosyl transferase (HGPRT). In the cellular environment, the degradation of DNA and RNA results in the release of purines. These purines are salvaged by the enzyme HGPRT, which utilizes them to synthesize new nucleotides for RNA and DNA (Sampat et al., 2011). Individuals diagnosed with Lesch-Nyhan syndrome exhibit a genetic mutation within the HGPRT gene, resulting in an alteration of the enzyme's amino acid sequence, thereby causing a loss of functionality. As a consequence, the recycling process of purines is impeded, leading to their accumulation and subsequent conversion into uric acid. The manifestation of disease symptoms is attributed to elevated concentrations of uric acid (Ceballos-Picot et al., 2015).

The Lesch-Nyhan syndrome serves as an exemplification of the connection between genotype and phenotype, wherein a genetic mutation impacts the production of a protein, subsequently giving rise to the manifestation of disease symptoms. The preceding chapters have elucidated the process by which genetic information is encoded within DNA and subsequently transferred from DNA to RNA. This chapter focuses on the analysis of translation. This fundamental biological process involves converting the nucleotide sequence present in mRNA into a protein's corresponding amino acid sequence (Nyhan et al., 1970).

5.3. The Molecular Relation Between Genotype & Phenotype

Archibald Garrod was the individual who initially proposed the presence of a correlation between genotype and proteins. In 1908, Garrod put forth the accurate proposition that genes can encode enzymes. Regrettably, his theory failed to impact his peers during that time significantly (Andresen et al., 1997).

Did you know?

In the past, non-coding regions of DNA were commonly regarded as non-functional and insignificant. However, recent scientific investigations have revealed that these regions actually serve crucial regulatory functions in the process of gene expression.

5.3.1. The One Gene, One Enzyme Hypothesis

Beadle and Tatum employed the filamentous fungus Neurospora as a model organism to investigate the biochemical consequences of genetic mutations. The utilization of Neurospora as a model organism is advantageous due to its ease of cultivation in laboratory settings and the haploid nature of its primary vegetative component. The haploid state facilitates the convenient observation of recessive mutations, as depicted in Figure 5.1 (Horowitz & Leupold, 1951).

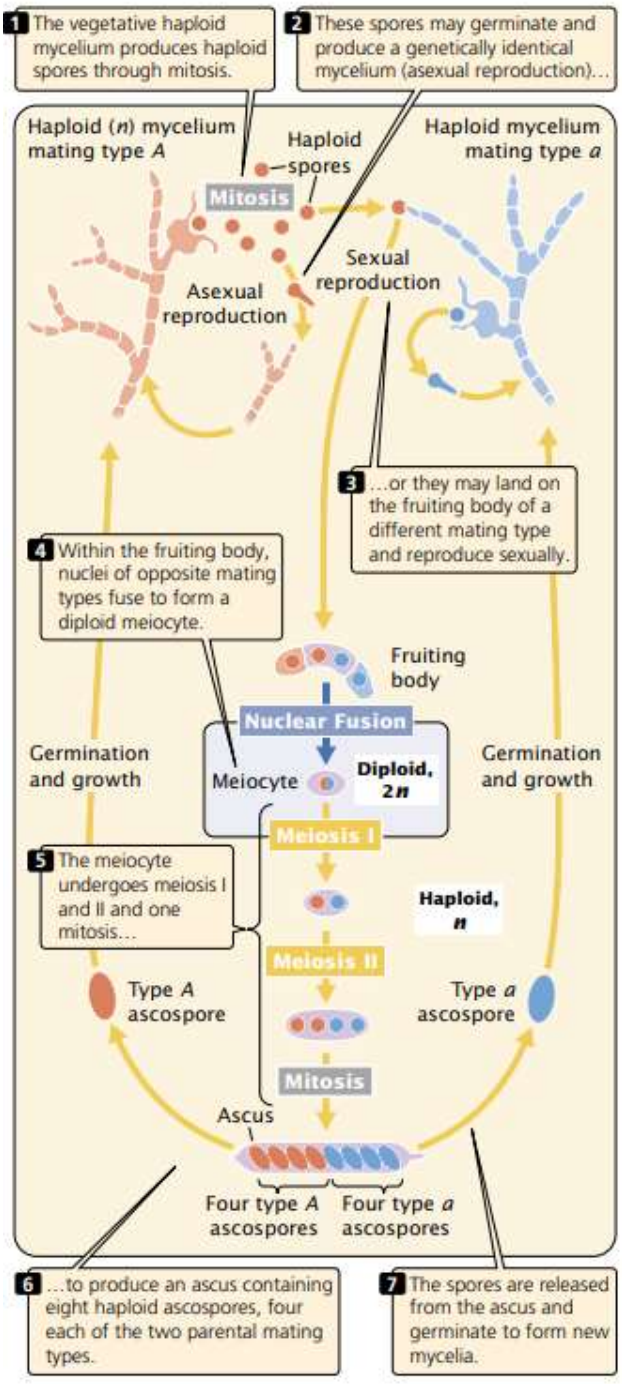


Figure 5-1. Beadle and Tatum employed Neurospora, a fungus with a complicated life cycle, to study gene-protein relationships. (Source: Kevin, Creative Commons License)

Wild-type Neurospora exhibits growth on a minimal medium comprising solely of inorganic salts, nitrogen, a carbon source such as sucrose, and the essential vitamin biotin. The fungus has the capacity to biosynthesize all necessary biological molecules from fundamental compounds. Nevertheless, it is

possible for mutations to occur that impede fungal growth by compromising the fungus's capacity to produce one or more vital biological compounds (Hickman & Cairns, 2003).

The auxotrophs, which are mutants lacking essential nutrients, exhibit an inability to thrive in a minimal medium. However, their growth can be sustained in a medium supplemented with the specific substance they are no more capable of synthesizing. The initial step undertaken by Beadle and Tatum involved subjecting spores of Neurospora to irradiation in order to induce mutations. Following irradiation, individual spores were transferred into separate culture tubes containing a complete medium, a nutrient-rich medium that provides all the necessary biological substances for growth. Subsequently, the spores were transferred from each culture to tubes that were filled with a minimal medium. Fungi harboring auxotrophic mutations exhibited growth on a nutrient-rich medium yet failed to increase on a nutritionally limited medium. This observation facilitated the identification of cultures by Beadle and Tatum that possessed such mutations (Rashevsky, 1962).

Following identifying an auxotrophic mutation within a specific culture, Beadle and Tatum embarked on a quest to ascertain the precise impact of this mutation. The spores of each mutant strain were transferred from a medium containing all necessary nutrients to a set of tubes. Each tube contained a minimal medium supplemented with a specific essential biological molecule, such as an amino acid. Upon observing the growth of spores within a tube, Beadle and Tatum successfully discerned the substance introduced as the biological molecule whose synthesis had been impacted by the mutation. An instance of an auxotrophic mutant, which exhibits growth solely on a minimal medium supplemented with arginine, can be attributed to a mutation that interferes with the biosynthesis of arginine. The utilization of this procedure by patients facilitated the genetic dissection of complex biochemical pathways involving multiple steps (Yanofsky, 2005).

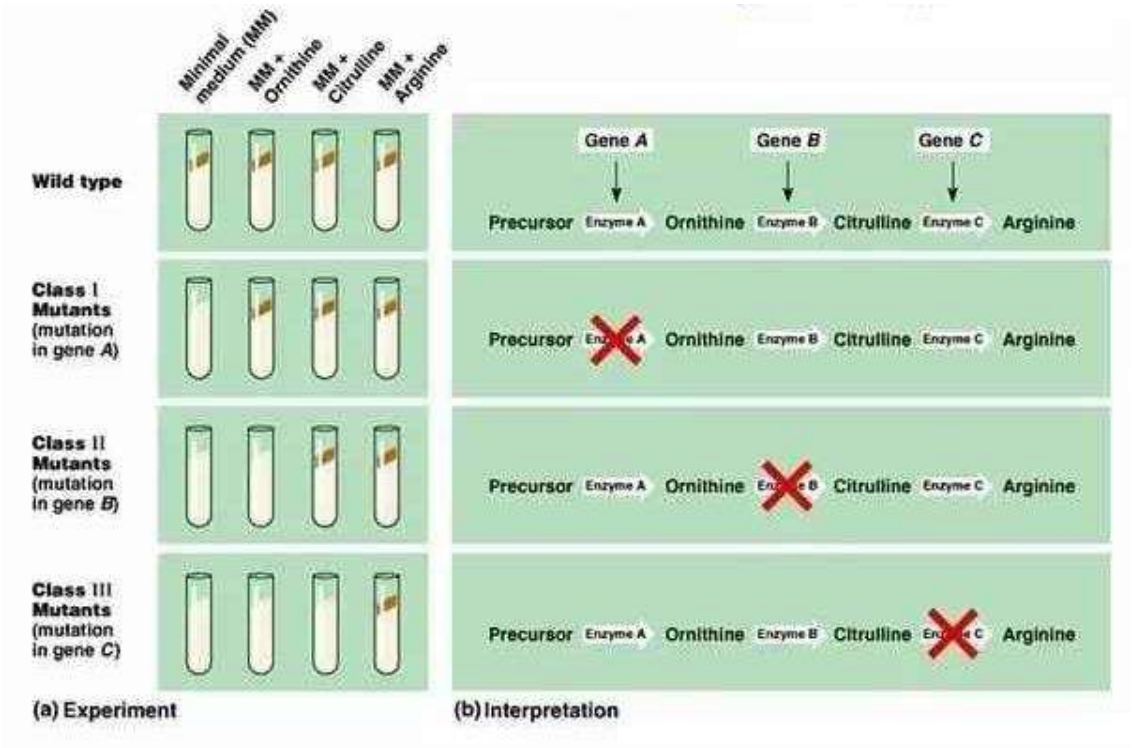


Figure 5-2. Beadle and Tatum found that enzymes influence each process step. (Source: Horace, Creative Commons License)

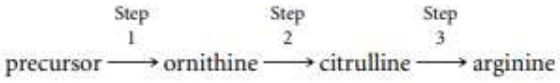
A series of auxotrophic mutants, whose growth was dependent on the presence of arginine, were initially isolated (Smith, 1994). Subsequently, the mutants were subjected to assessment regarding their growth potential on a minimal medium that was supplemented with three specific compounds, namely ornithine, citrulline, and arginine. Based on the obtained results, the mutants were categorized into three distinct groups (as presented in Table 5.1) according to their ability to exhibit growth in the presence of specific substances. The mutants belonging to Group I exhibited growth when cultivated on a minimal medium that was supplemented with ornithine, citrulline, or arginine. The mutants belonging to Group II exhibited growth on a minimal medium that was supplemented with either arginine or citrulline. However, these mutants did not display any growth when the medium was solely supplemented with ornithine (Robinson, 1974).

Table 5.1. Arginine auxotrophic mutant growth on minimum medium with nutrients (Source: Marshal, Creative Commons License)

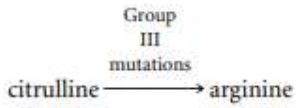
Mutant Strain Number	**Ornithine**	**Citrulline**	**Arginine**

Group I	+	+	+
Group II	-	+	+
Group III	-	-	+

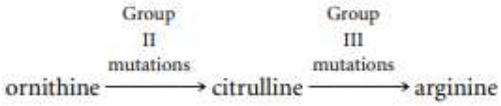
Ultimately, the group III mutants exclusively exhibited growth when the medium was supplemented with arginine. Srb and Horowitz suggested that the biochemical pathway responsible for the synthesis of the amino acid arginine consists of a minimum of three sequential enzymatic steps (Stark, 1977):



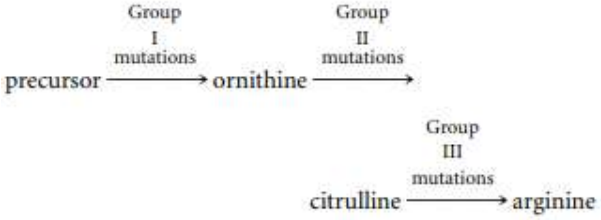
The researchers reached a conclusion that the mutations observed in group I had an impact on the initial step of this pathway. In contrast, the mutations in group II affected the second step, and the mutations in group III affected the third step. However, it remains unclear how the researchers determined the accuracy of the sequence of compounds in the biochemical pathway. It is worth noting that in the event of a mutation obstructing step 1, the introduction of either ornithine or citrulline facilitates growth, as these substances can still undergo conversion into arginine. Similarly, in the event that step 2 encounters an obstruction, the introduction of citrulline facilitates growth, whereas the introduction of ornithine does not yield any discernible impact. In the event that step 3 encounters an obstruction, the proliferation of spores will solely occur upon the introduction of arginine into the medium. The fundamental tenet posits that an auxotrophic mutant is incapable of biosynthesizing any molecule beyond the point at which a mutation hinders the metabolic pathway. By applying this line of reasoning in conjunction with the data presented in Table 5.1, it becomes evident that the inclusion of arginine in the growth medium facilitates the growth of all three categories of mutants. Hence, the biochemical processes influenced by all the mutant variations occur prior to the step leading to the production of arginine. The inclusion of citrulline in the growth medium enables the proliferation of group I and group II mutants, while group III mutants are unable to grow under these conditions. Consequently, it can be inferred that group III mutations impact a biochemical process occurring subsequent to citrulline synthesis but preceding arginine synthesis (Wang et al., 2004).



Group I mutants can grow with the addition of ornithine, but group II and group III mutants cannot; this suggests that the mutations in these two groups influence stages beyond the generation of ornithine. We know that group II mutations interfere with citrulline formation. Hence we know that group II mutations prevent ornithine from being converted into citrulline (Lacks & Hotchkiss, 1960).



Given that group I mutations impact a specific stage preceding the synthesis of ornithine, it can be inferred that these mutations disrupt the process of converting a precursor substance into ornithine. The biochemical pathway leading to the synthesis of ornithine, citrulline, and arginine can now be delineated (Burn et al., 2002).



It is essential to acknowledge that this particular procedure does not inherently identify all stages within a given pathway. Instead, it solely identifies the steps that yield the compounds under examination. By employing mutations and employing this particular mode of reasoning, Beadle and Tatum successfully discerned the genes responsible for regulating multiple biosynthetic pathways within the organism Neurospora. It has been determined that individual enzymes regulate each step within a given pathway, as illustrated in Figure 5.2, which specifically pertains to the arginine pathway. The findings from genetic crosses and mapping studies have provided evidence that mutations impacting a specific step within a pathway consistently correspond to the same chromosomal location. Beadle and Tatum postulated that mutations impacting a specific biochemical pathway were associated with a solitary genetic locus responsible for encoding a specific enzyme (Awai et al., 2006).

5.3.2. The Structure & Function of Proteins

Proteins play a pivotal role in all biological processes (Figure 5.3). Numerous proteins exhibit enzymatic properties, serving as pivotal catalysts that facilitate the chemical reactions within cellular systems. Additionally, certain proteins fulfill structural roles by offering a framework and reinforcement for various biological structures such as membranes, filaments, bone, and hair. Certain proteins play a crucial role in facilitating the transportation of substances, while others serve as regulators, facilitators of communication, or defenders within biological systems (Johnson & Barford, 1993).

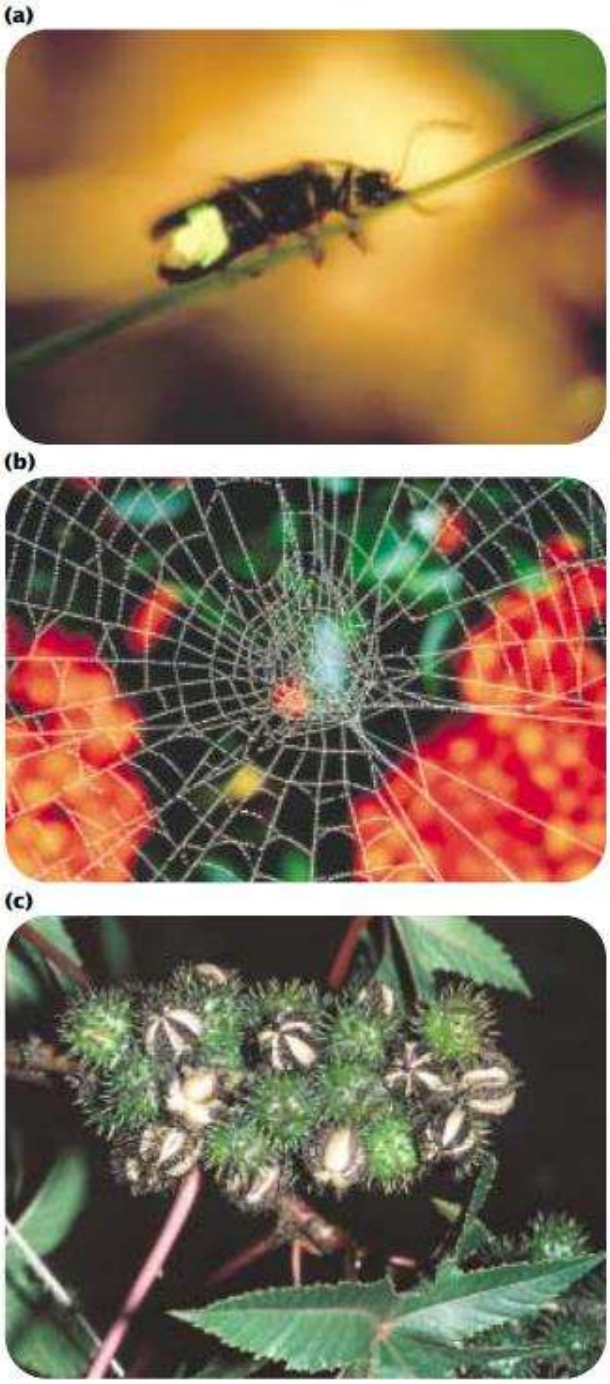


Figure 5-3. Proteins are essential to all cellular activities and serve a wide variety of biological purposes. (Source: Shay, Creative Commons License)

Proteins are comprised of amino acids that are connected in a linear fashion. Proteins consist of a set of 20 frequently occurring amino acids, as illustrated in Figure 5.4, which displays their respective three-letter and one-letter abbreviations. Modified forms of the common amino acids are occasionally

observed in proteins. The 20 amino acids that are commonly found share a similar structural framework, with the only distinguishing factor being the variations in the structures of their R (radical) groups (Ellis & Jones, 1998).

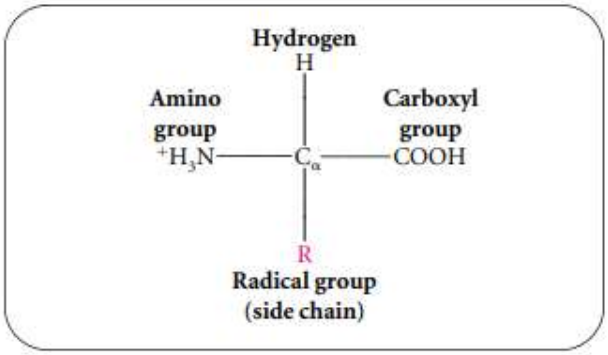
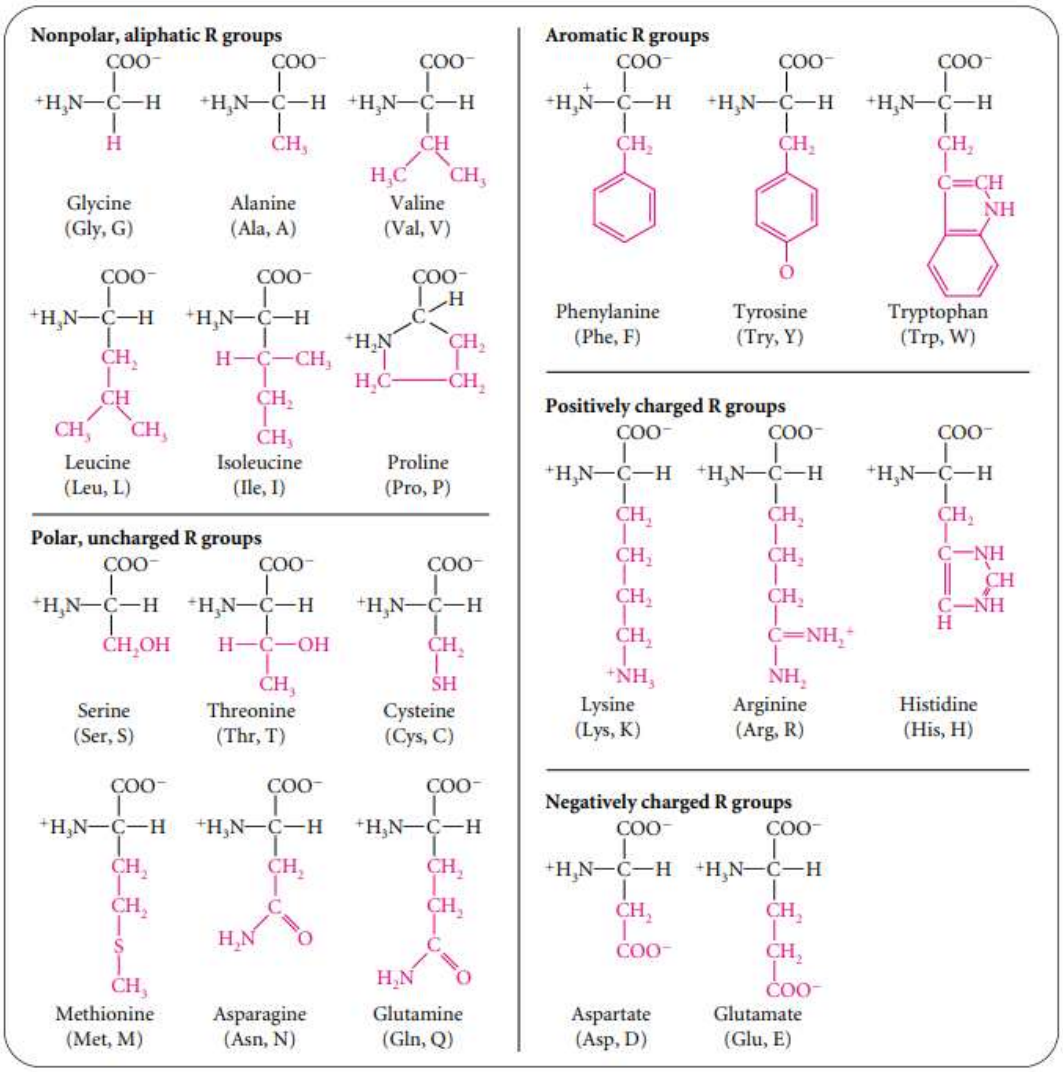


Figure 5-4. Typical amino acids share structural similarities. There are four types of groups that can bind to an amino acid's core carbon atom (Source: Philip, Creative Commons License)



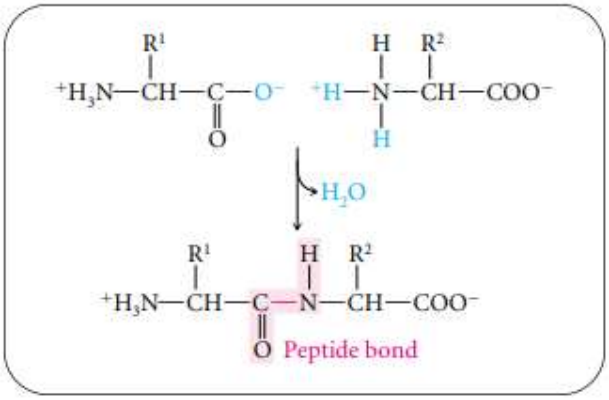


Figure 5-5. Peptide bonds connect amino acids together. (Source: Philip, Creative Commons License)

The linkage between amino acids in proteins occurs through peptide bonds (as depicted in Figure 5.4), resulting in the formation of polypeptide chains (Schumann et al., 1990). A protein is composed of one or multiple polypeptide chains. Similar to nucleic acids, polypeptides exhibit polarity, characterized by the presence of a free amino group (NH2) at one end and a free carboxyl group (CO2H) at the other end. Certain proteins are composed of a limited number of amino acids, while others possess a significantly larger quantity, potentially reaching into the thousands. Similar to nucleic acids, proteins exhibit multiple levels of molecular organization in their structure. The fundamental configuration of a protein is determined by the linear arrangement of its constituent amino acids (Figure 5.6a). The folding and twisting of a polypeptide chain into a secondary structure, as depicted in Figure 5.6b, is facilitated by the interactions among adjacent amino acids. The beta-pleated sheet and the alpha helix are two frequently observed secondary structures in proteins. The secondary structures of proteins engage in interactions and subsequently undergo folding processes to give rise to a tertiary structure (as depicted in Figure 5.6c). The tertiary structure represents the complete, three-dimensional conformation of the protein (Helmreich & Hofmann, 1996).

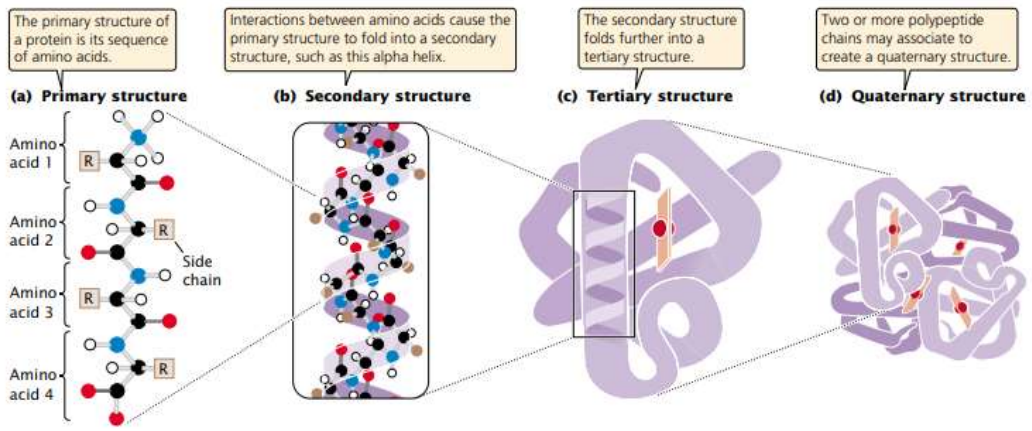


Figure 5-6. Proteins are hierarchically structured. (Source: Marshal, Creative Commons License)

Did you know?

In recent years, significant progress has been made in the field of genetic engineering, enabling researchers to manipulate the genetic material of various organisms. This breakthrough holds promise for addressing a range of issues, including the treatment of diseases and the mitigation of environmental challenges.

5.4. The Genetic Code

In the year 1953, the structure of DNA was successfully explained by Watson and Crick. Nevertheless, the mechanism by which the DNA base sequence dictated the arrangement of amino acids in proteins, known as the genetic code, was not readily apparent and continued to elude scientists for an additional decade. One of the initial inquiries regarding the genetic code pertaining to the requisite number of nucleotides for the specification of a singular amino acid (Clary & Wolstenholme, 1985). The fundamental component of the genetic code, responsible for encoding a solitary amino acid, is referred to as a codon. It was acknowledged by numerous early researchers that codons necessitate a minimum of three nucleotides. The mRNA molecule consists of nucleotide positions that can be filled with any of the four bases: adenine (A), guanine (G), cytosine (C), or uracil (U). If a codon were composed of a solitary nucleotide, the limited number of nucleotide options (A, G, C, and U) would be insufficient to encode the 20 distinct amino acids that are typically present in proteins. If codons were composed of two nucleotides, such as GU or AC, the total number of possible codons would be 16 (4 x 4). However, this number would still be insufficient to encode all 20 amino acids. The number of possible codons can be determined by considering that each codon consists of three nucleotides. With four possible nucleotides (A, C, G, and T), the total number of codons can be calculated as 4 x 4 x 4, resulting in 64 possible codons. This number is more than sufficient to encode the 20 different amino acids. Hence, the utilization of a triplet code, which necessitates the inclusion of three nucleotides per codon, emerges as the optimal approach for encoding the complete set of 20 amino acids. In 1961, Francis Crick and his colleagues conducted experiments involving mutations in bacteriophage, which provided confirmation that the genetic code operates as a triplet code (Sánchez et al., 2006).

5.4.1. Breaking The Genetic Code

Once it had been definitively proven that the genetic code is composed of codons comprising three nucleotides, the subsequent objective was to ascertain the specific grouping of three nucleotides that correspond to each amino acid. The accomplishment of this objective necessitated the creation of a cell-free system dedicated to protein synthesis. This system would enable the investigation of the translation process pertaining to a specific mRNA molecule (Kwon et al., 2003).

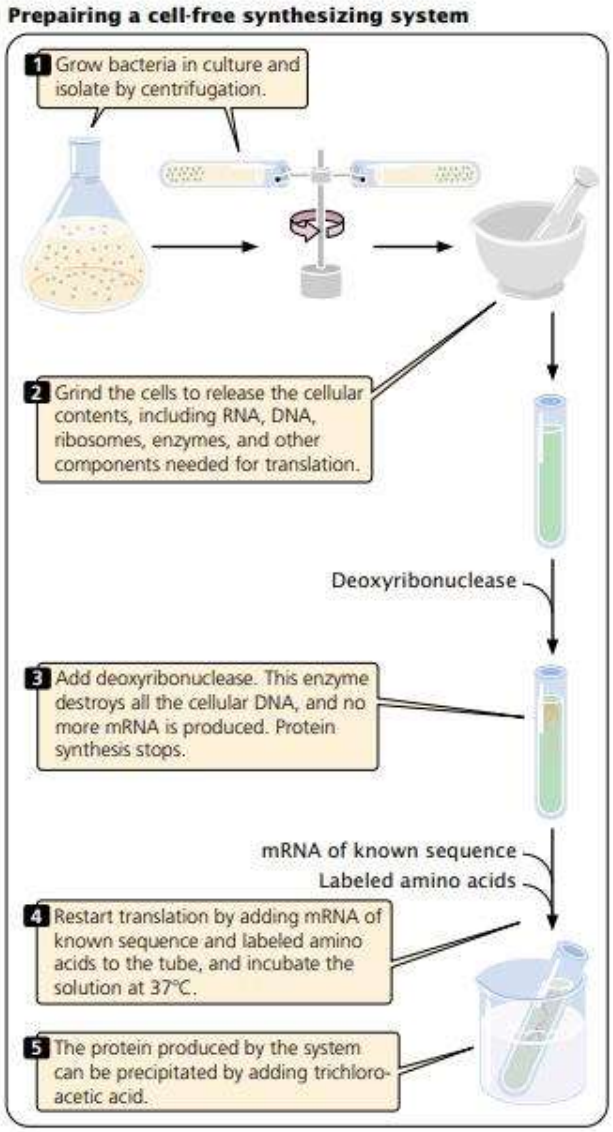


Figure 5-7. A cell-free protein synthesis mechanism was essential for deciphering the genetic code. (Source: Franklin, Creative Commons License)

From a logical standpoint, the most straightforward approach to deciphering the code would involve ascertaining the fundamental sequence of a specific RNA fragment, subsequently introducing it into a protein-synthesizing system devoid of cells, and facilitating the synthesis of a protein under the guidance of said RNA fragment (Hornos et al., 1999). The amino acid sequence of the recently synthesized protein can subsequently be ascertained, and its sequence can be juxtaposed with that of the RNA. Regrettably, during that period, the determination of the nucleotide sequence of an RNA fragment was not feasible. Consequently, alternative approaches were required to decipher the genetic code. The initial indications regarding the genetic code were discovered in 1961 through the research

conducted by Marshall Nirenberg and Johann Heinrich Matthaei. The researchers generated artificial ribonucleic acids (RNAs) through the utilization of polynucleotide phosphorylase, an enzyme. In contrast to RNA polymerase, polynucleotide phosphorylase does not necessitate a template for its activity. It possesses the capability to non-specifically connect RNA nucleotides that are accessible in a random manner (Nirenberg, 2004).

The initial synthetic messenger RNAs (mRNAs) employed by Nirenberg and Matthaei were composed of homopolymers, which are RNA molecules comprised solely of a single nucleotide type. As an illustration, the introduction of polynucleotide phosphorylase to a solution comprising uracil nucleotides resulted in the production of RNA molecules exclusively composed of uracil nucleotides, consequently comprising solely UUU codons. The poly(U) RNAs were subsequently introduced into 20 separate tubes, each of which contained a cell-free protein synthesis system along with a mixture of 20 distinct amino acids, with one of them being radioactively labeled. The process of translation occurred in all twenty tubes. However, the presence of radioactive protein was observed solely in one of the tubes, specifically the tube that contained phenylalanine labeled with a radioactive marker (refer to Figure 5.8). The findings of this study indicate that the codon UUU is responsible for encoding the amino acid phenylalanine. The outcomes of comparable experiments involving poly(C) and poly(A) RNA have indicated that the codon CCC corresponds to the amino acid proline, while the codon AAA corresponds to the amino acid lysine. However, due to technical limitations, the results obtained from poly(G) were inconclusive and could not be reliably interpreted. In order to acquire further knowledge regarding extra codons, Nirenberg and his colleagues conducted an experiment involving the synthesis of RNA molecules that incorporated two or three distinct nucleotide bases. The RNAs synthesized by polynucleotide phosphorylase exhibit random incorporation of nucleotides, resulting in the presence of mixed bases (Lenstra, 2014).

Consequently, these RNAs are referred to as random copolymers. In the context of this experiment, it was observed that the combination of adenine and cytosine nucleotides in the presence of polynucleotide phosphorylase resulted in the generation of RNA molecules exhibiting a total of eight distinct codons. These codons were specifically identified as AAA, AAC, ACC, ACA, CAA, CCA, CAC, and CCC. The utilization of cell-free protein-synthesizing systems resulted in the production of proteins comprising six distinct amino acids, namely asparagine, glutamine, histidine, lysine, proline, and threonine, through the involvement of poly (AC) RNAs (Tejedor & Valcárcel, 2010).

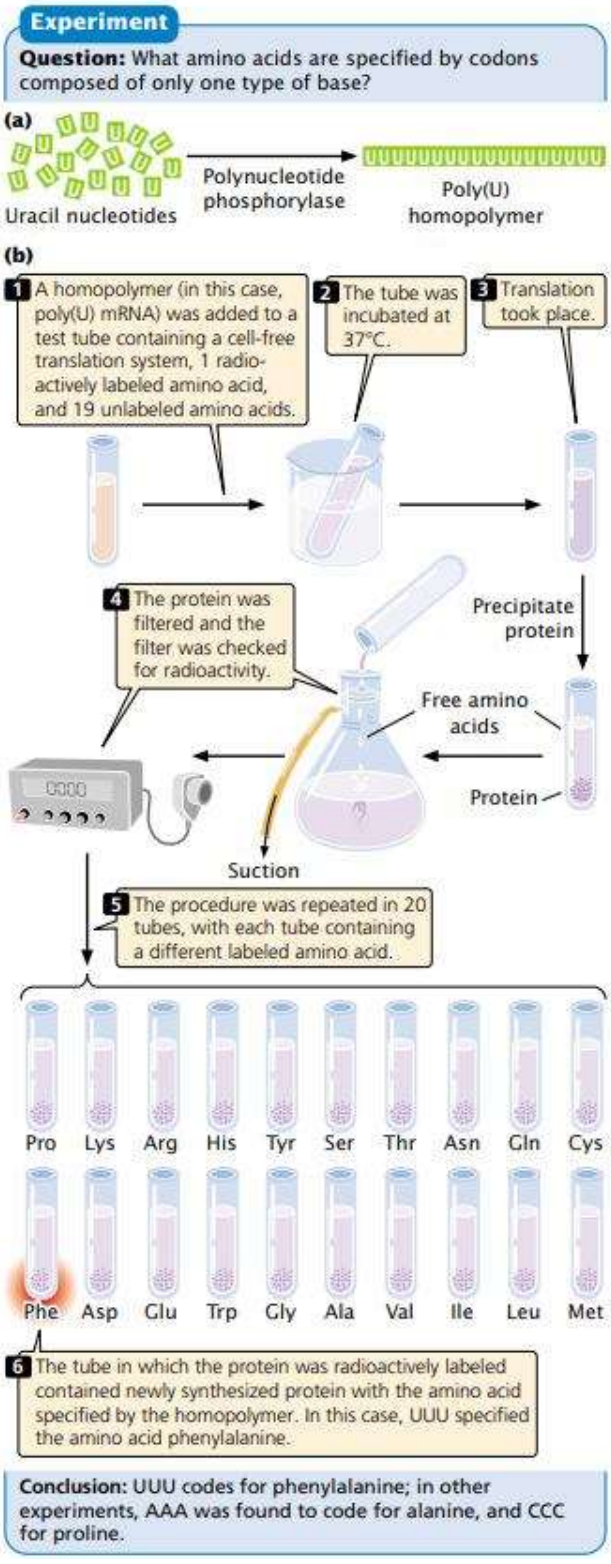


Figure 5-8. Nirenberg and Matthaei devised a homopolymer-specific amino acid identification approach. (Source: Kevin, Creative Commons License)

The distribution of amino acids within proteins was contingent upon the relative abundance of the two nucleotides employed during the synthesis of artificial mRNA. Furthermore, the likelihood of encountering a specific codon could be theoretically determined by evaluating the ratios of the constituent bases. If a ratio of 4:1 of C to A was employed in the synthesis of RNA, the likelihood of C appearing at any specific position within a codon is determined to be, while the probability of A being present within the codon is calculated to be 1/5 (Hornos & Hornos, 1993). When bases are incorporated randomly, the probability of observing a codon with two Cs and one A can be calculated as follows: 4/5 x 4/5 x 1/5 = 16/125= 0.13, or 13%. Similarly, the probability of observing a codon with two As and one C (AAC, ACA, or CAA) can be calculated as 1/5 x 1/5 x 4/5 = 4/125= 0.032 or about 3%. Hence, it can be inferred that the prevalence of an amino acid encoded by two cytosines (C) and one adenine (A) would be higher compared to an amino acid encoded by two adenines (A) and one cytosine (C). The base composition of codons was derived by comparing the percentages of amino acids in proteins produced by random copolymers with the expected theoretical frequencies for the codons. The experiments conducted did not yield any significant findings regarding the codon base sequence. It was observed that histidine was encoded by a codon consisting of two Cs and one A, as depicted in Figure 5.9. However, the specific identity of this codon, whether it was ACC, CAC, or CCA, remained undetermined (Antoneli & Forger, 2011).

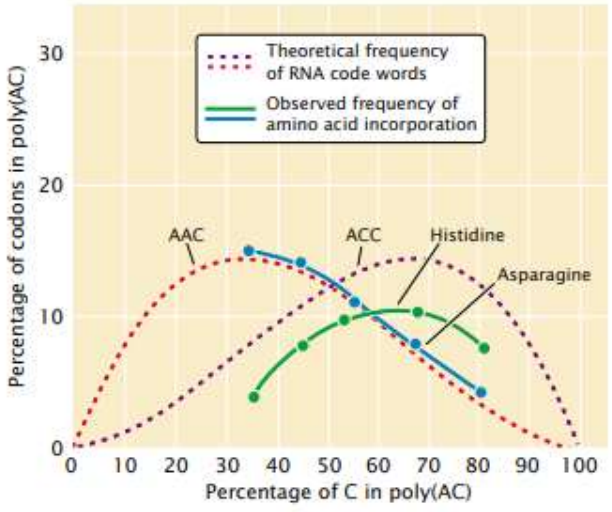


Figure 5-9. Nirenberg and Matthaei's random copolymers revealed the genetic code. (Source: Kevin, Creative Commons License)

In 1964, Nirenberg and Philip Leder devised an alternative method utilizing ribosome-bound transfer RNAs (tRNAs) to address the constraints associated with random copolymers. It was discovered that a concise mRNA sequence, even comprising only a solitary codon, can bind with a ribosome (Mat et al., 2010). The codon on the shorter messenger RNA molecule would subsequently form base pairs with the corresponding anticodon located on a tRNA molecule, which transports the amino acid stated by the codon (refer to Figure 5.10). The mRNA that was associated with ribosomes was combined with transfer RNAs (tRNAs) and amino acids and subsequently subjected to filtration using a nitrocellulose filter. The transfer RNAs (tRNAs) that formed pairs with the mRNA molecules bound to the ribosome were observed to adhere to the filter, while tRNAs that were not bound remained able to pass through. One notable benefit of this system lies in its compatibility with concise synthetic mRNA molecules, which can be readily synthesized with a predetermined sequence. Nirenberg and Leder synthesized more than 50 short messenger RNAs (mRNAs) that contained codons with known sequences. These synthesized mRNAs were subsequently introduced individually into a mixture consisting of ribosomes and transfer RNAs (tRNAs). Subsequently, the researchers proceeded to isolate the tRNAs bound to the ribosomes and ascertained the specific amino acids associated with these bound tRNAs. As an illustration, the presence of a tRNA molecule carrying valine was observed in synthetic RNA sequences containing the codon GUU, while no such tRNA association was observed in RNA sequences containing the codons UGU and UUG. By employing this methodology, Nirenberg and his colleagues successfully ascertained the specific amino acids that are encoded by over 50 codons (Martin, 1984).

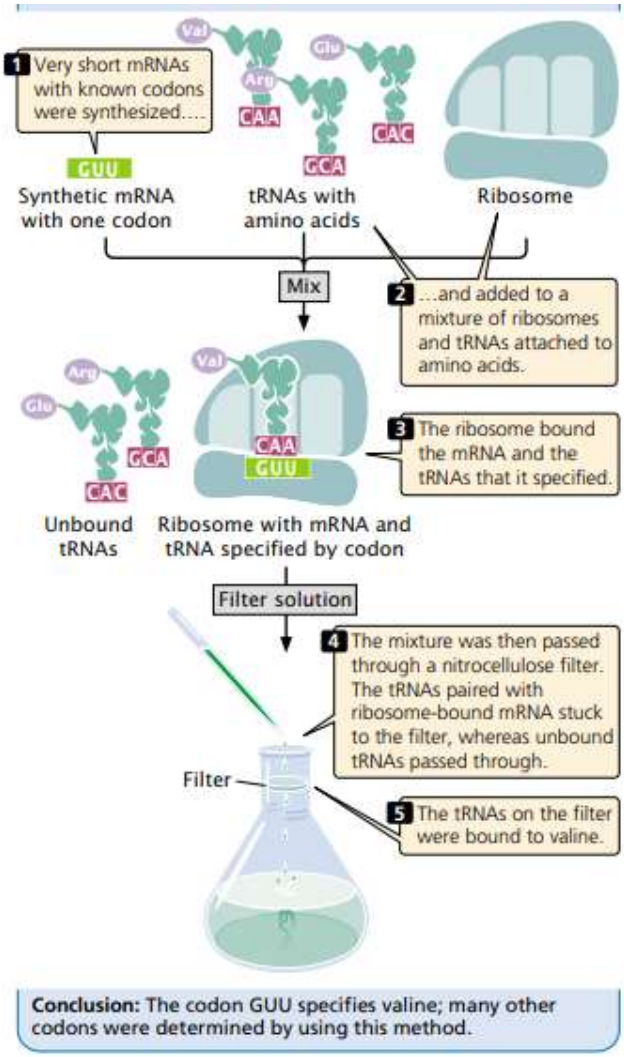


Figure 5-10. Nirenberg and Leder used ribosome-bound tRNAs to deliver genetic code information. (Source: Kevin, Creative Commons License)

An additional method was employed to provide supplementary information regarding the genetic code. Gobind Khorana and his coworkers employed chemical methodologies to artificially produce RNA molecules comprising established repetitive sequences (Greenough, 2023). The researchers postulated that if an mRNA sequence consisted of alternating uracil and guanine nucleotides (UGUG UGUG), it would be interpreted during the process of translation as two alternating codons, UGU GUG UGU GUG. Consequently, this would result in the synthesis of a protein consisting of two alternating amino acids. Upon introduction of the synthetic mRNA into a cell-free protein synthesis system, Khorana and his colleagues observed the generation of a protein composed of cysteine and valine residues arranged in an alternating pattern. The aforementioned technique was unable to ascertain the specific assignment