

Oligonucleotide Transcription, Translation and Replication





Oligonucleotide Transcription, Translation and Replication

Authors

Prof. Tom Brown Sr.

Dr. Tom Brown Jr.

DNA, RNA and Protein Synthesis

The genetic material is stored in the form of DNA in most organisms. In humans, the nucleus of each cell contains 3×10^9 base pairs of DNA distributed over 23 pairs of chromosomes, and each cell has two copies of the genetic material. This is known collectively as the human genome. The human genome contains around 30 000 genes, each of which codes for one protein.

Large stretches of DNA in the human genome are transcribed but do not code for proteins. These regions are called introns and make up around 95% of the genome. The nucleotide sequence of the human genome is now known to a reasonable degree

of accuracy but we do not yet understand why so much of it is non-coding. Some of this non-coding DNA controls gene expression but the purpose of much of it is not yet understood. This is a fascinating subject that is certain to advance rapidly over the next few years.

The Central Dogma of Molecular Biology states that DNA makes RNA makes proteins (Figure 1).

The process by which DNA is copied to RNA is called transcription, and that by which RNA is used to produce proteins is called translation.

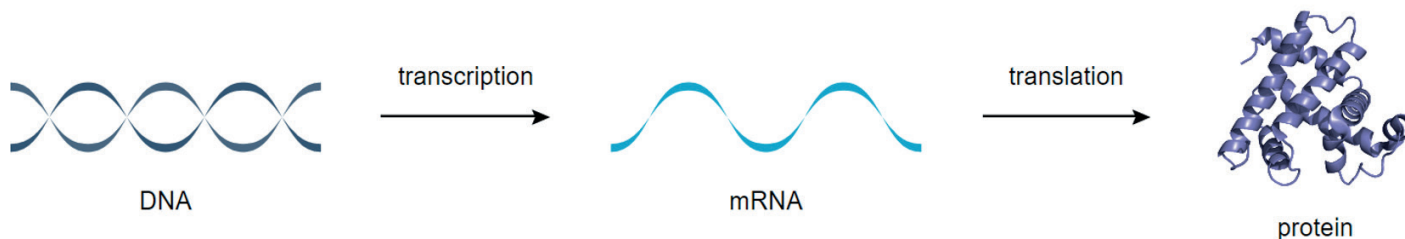


Figure 1. The Central Dogma of Molecular Biology. DNA makes RNA makes proteins.

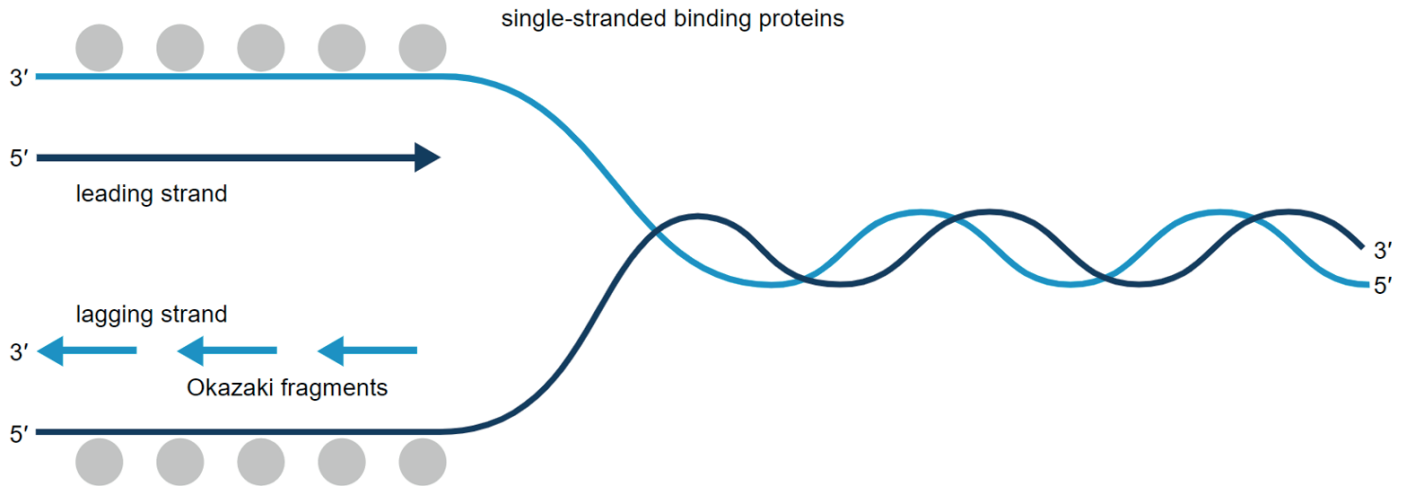


Figure 2. DNA replication in bacteria. Simplified representation of DNA replication in bacteria..

DNA Replication

Each time a cell divides, each of its double strands of DNA splits into two single strands. Each of these single strands acts as a template for a new strand of complementary DNA. As a result, each new cell has its own complete genome. This process is known as DNA replication. Replication is controlled by the Watson-Crick pairing of the bases in the template strand with incoming deoxynucleoside triphosphates, and is directed by DNA polymerase enzymes. It is a complex process, particularly in eukaryotes, involving an array of enzymes. A simplified version of bacterial DNA replication is described in Figure 2.

DNA biosynthesis proceeds in the 5'- to 3'-direction. This makes it impossible for DNA polymerases to synthesize both strands simultaneously. A portion of the double helix must first unwind, and this is mediated by helicase enzymes.

The leading strand is synthesized continuously but the opposite strand is copied in short bursts of about 1000 bases, as the lagging strand template becomes available. The resulting short strands are called Okazaki fragments

(after their discoverers, Reiji and Tsuneko Okazaki). Bacteria have at least three distinct DNA polymerases: Pol I, Pol II and Pol III; it is Pol III that is largely involved in chain elongation. Strangely, DNA polymerases cannot initiate DNA synthesis de novo, but require a short primer with a free 3'-hydroxyl group. This is produced in the lagging strand by an RNA polymerase (called DNA primase) that is able to use the DNA template and synthesize a short piece of RNA around 20 bases in length. Pol III can then take over, but it eventually encounters one of the previously synthesized short RNA fragments in its path. At this point Pol I takes over, using its 5'- to 3'-exonuclease activity to digest the RNA and fill the gap with DNA until it reaches a continuous stretch of DNA. This leaves a gap between the 3'-end of the newly synthesized DNA and the 5'-end of the DNA previously synthesized by Pol III. The gap is filled by DNA ligase, an enzyme that makes a covalent bond between a 5'-phosphate and a 3'-hydroxyl group (Figure 3). The initiation of DNA replication at the leading strand is more complex and is discussed in detail in more specialized texts.

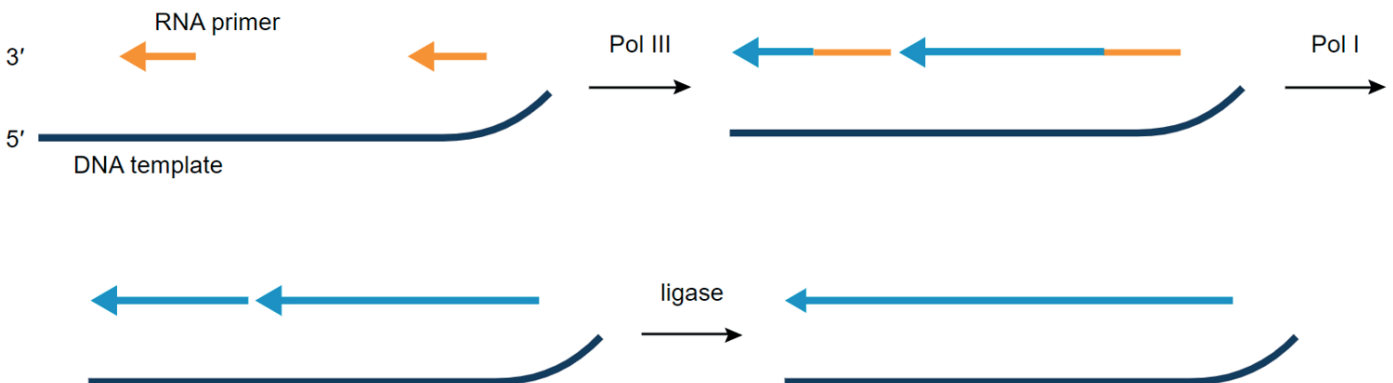


Figure 3. DNA polymerases in DNA replication. Simplified representation of the action of DNA polymerases in DNA replication in bacteria.

Mistakes in DNA Replication

DNA replication is not perfect. Errors occur in DNA replication, when the incorrect base is incorporated into the growing DNA strand. This leads to mismatched base pairs, or mispairs. DNA polymerases have proofreading activity, and a DNA repair enzymes have evolved to correct these mistakes. Occasionally, mispairs survive and are incorporated into the genome in the next round of replication. These mutations may have no consequence, they may result in the death of the organism, they may result in a genetic disease or cancer; or they may give the organism a competitive advantage over its neighbours, which leads to evolution by natural selection.

Transcription

Transcription is the process by which DNA is copied (transcribed) to mRNA, which carries the information needed for protein synthesis. Transcription takes place in two broad steps. First, pre-messenger RNA is formed, with the involvement of RNA polymerase enzymes. The process relies on Watson-Crick base pairing, and the resultant single strand of RNA is the reverse-complement of the original DNA sequence. The pre-messenger RNA is then “edited” to produce the desired mRNA molecule in a process called RNA splicing.

Formation of Pre-messenger RNA

The mechanism of transcription has parallels in that of DNA replication. As with DNA replication, partial unwinding of the double helix must occur before transcription can take place, and it is the RNA polymerase enzymes that catalyze this process.

Unlike DNA replication, in which both strands are copied, only one strand is transcribed. The strand that contains the gene is called the sense strand, while the complementary strand is the antisense strand. The mRNA produced in transcription is a copy of the sense strand, but it is the antisense strand that is transcribed.

Ribonucleoside triphosphates (NTPs) align along the antisense DNA strand, with Watson-Crick base pairing (A pairs with U). RNA polymerase joins the ribonucleotides together to form a pre-messenger RNA molecule that is complementary to a region of the antisense DNA strand. Transcription ends when the RNA polymerase enzyme reaches a triplet of bases that is read as a “stop” signal. The DNA molecule re-winds to re-form the double helix.

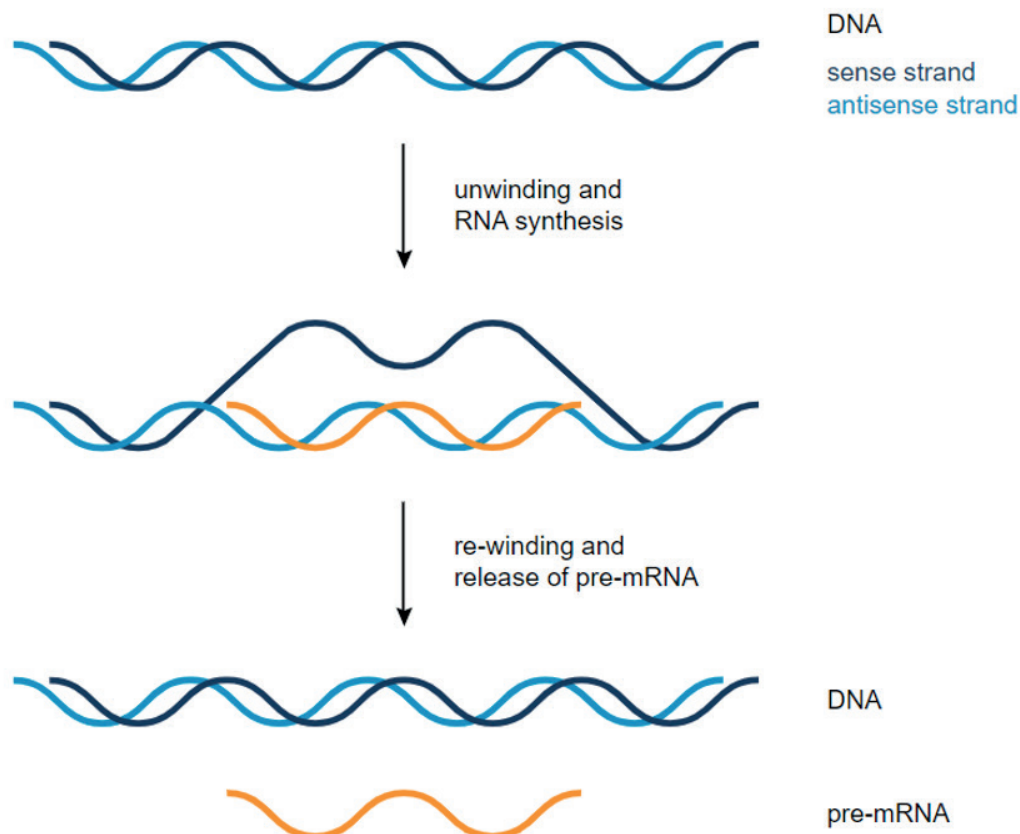


Figure 4. Transcription. Simplified representation of the formation of pre-messenger RNA (orange) from double-stranded DNA (blue) in transcription.

RNA Splicing

The pre-messenger RNA thus formed contains introns which are not required for protein synthesis. The pre-messenger RNA is chopped up to remove the introns and create messenger RNA (mRNA) in a process called RNA splicing (Figure 5).

Alternative Splicing

In alternative splicing, individual exons are either spliced or included, giving rise to several different possible mRNA products. Each mRNA product codes for a different protein isoform; these protein isoforms differ in their peptide sequence and therefore their biological activity. It is estimated that up to 60% of human gene products undergo alternative splicing. Several different mechanisms of alternative splicing are known, two of which are illustrated in Figure 6.

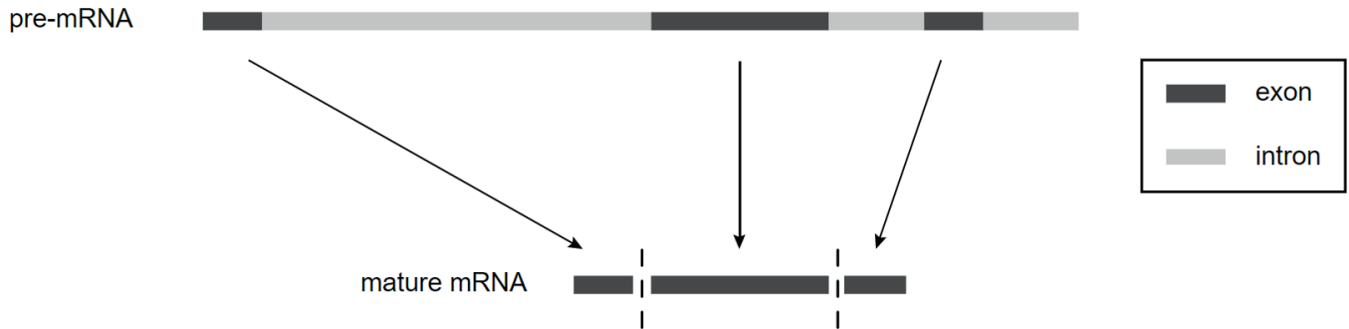


Figure 5. RNA splicing. Introns are spliced from the pre-messenger RNA to give messenger RNA (mRNA).

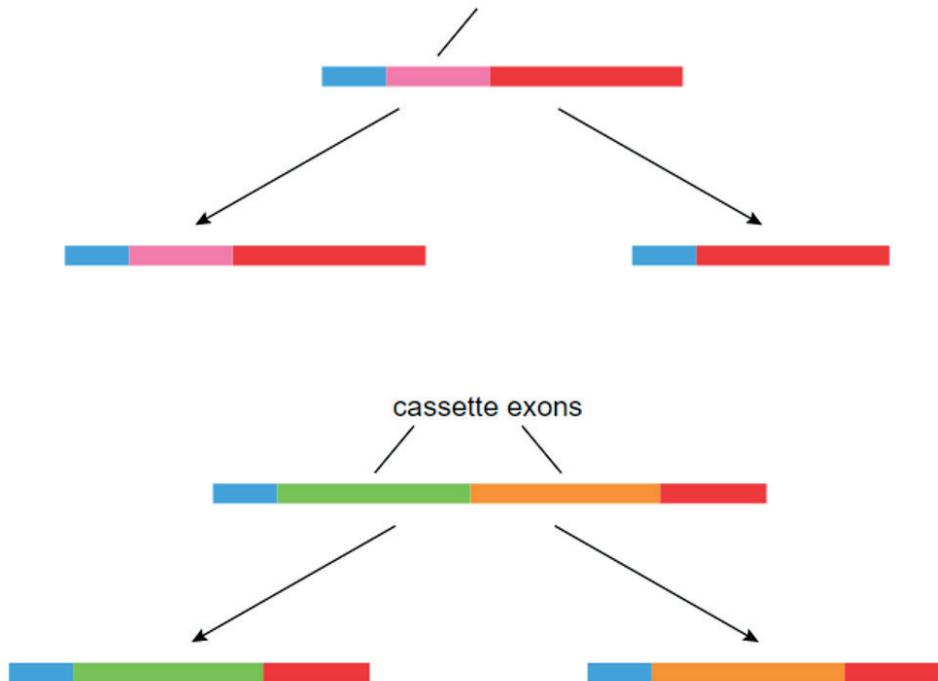


Figure 6. Alternative splicing. Several different mechanisms of alternative splicing exist - a cassette exon can be either included in or excluded from the final RNA (top), or two cassette exons may be mutually exclusive (bottom).

Alternative splicing contributes to protein diversity - a single gene transcript (RNA) can have thousands of different splicing patterns, and will therefore code for thousands of different proteins: a diverse proteome is generated from a relatively limited genome. Splicing is important in genetic regulation (alteration of the splicing pattern in response to cellular conditions changes protein expression). Perhaps not surprisingly, abnormal splicing patterns can lead to disease states including cancer.

Reverse Transcription

In reverse transcription, RNA is “reverse transcribed” into DNA. This process, catalyzed by reverse transcriptase enzymes, allows retroviruses, including the human immunodeficiency virus (HIV), to use RNA as their genetic material. Reverse transcriptase enzymes have also found applications in biotechnology, allowing scientists to convert RNA to DNA for techniques such as PCR.

Translation

The mRNA formed in transcription is transported out of the nucleus, into the cytoplasm, to the ribosome (the cell’s protein synthesis factory). Here, it directs protein synthesis. Messenger RNA is not directly involved in protein synthesis - transfer RNA (tRNA) is required for this. The process by which mRNA directs protein synthesis with the assistance of tRNA is called translation.

The ribosome is a very large complex of RNA and protein molecules. Each three-base stretch of mRNA (triplet) is known as a codon, and one codon contains the information for a specific amino acid. As the mRNA passes through the ribosome, each codon interacts with the anticodon of a specific transfer RNA (tRNA) molecule by Watson-Crick base pairing. This tRNA molecule carries an amino acid at its 3’-terminus, which is incorporated into the growing protein chain. The tRNA is then expelled from the ribosome. Figure 7 shows the steps involved in protein synthesis.

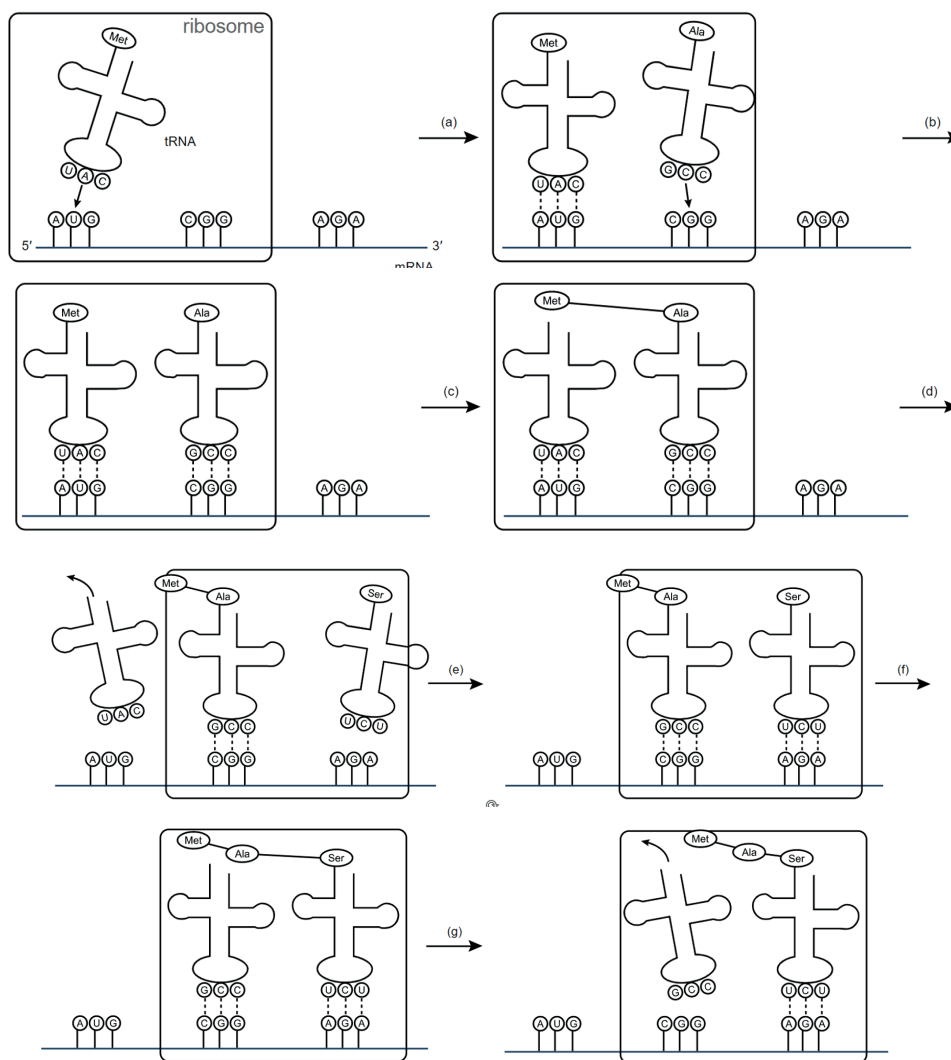


Figure 7. Translation(a) and (b) tRNA molecules bind to the two binding sites of the ribosome, and by hydrogen bonding to the mRNA; (c) a peptide bond forms between the two amino acids to make a dipeptide, while the tRNA molecule is left uncharged; (d) the uncharged tRNA molecule leaves the ribosome, while the ribosome moves one codon to the right (the dipeptide is translocated from one binding site to the other); (e) another tRNA molecule binds; (f) a peptide bond forms between the two amino acids to make a tripeptide; (g) the uncharged tRNA molecule leaves the ribosome.

Transfer RNA

Transfer RNA adopts a well defined tertiary structure which is normally represented in two dimensions as a cloverleaf shape, as in Figure 7. The structure of tRNA is shown in more detail in Figure 8.

Each amino acid has its own special tRNA (or set of tRNAs). For example, the tRNA for phenylalanine (tRNA^{Phe}) is different from that for histidine (tRNA^{His}). Each amino acid is attached to its tRNA through the 3'-OH group to form an ester which reacts with the -amino group of the terminal amino-acid of the growing protein chain to form a new amide bond (peptide bond) during protein synthesis (Figure 9). The reaction of esters with amines is generally favourable but the rate of reaction is increased greatly in the ribosome.

Each transfer RNA molecule has a well defined tertiary structure that is recognized by the enzyme aminoacyl tRNA synthetase, which adds the correct amino acid to the 3'-end of the uncharged tRNA. The presence of modified nucleosides is important in stabilizing the tRNA structure. Some of these modifications are shown in Figure 10.

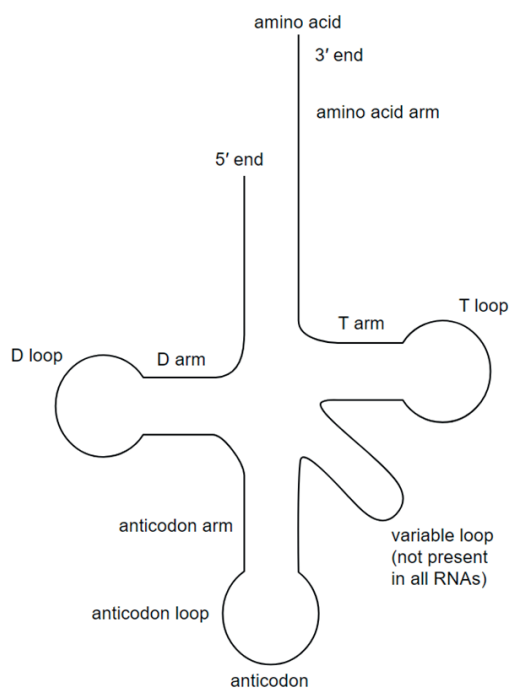


Figure 8. Two-dimensional structures of tRNA (transfer RNA). In some tRNAs the DHU arm has only three base pairs.

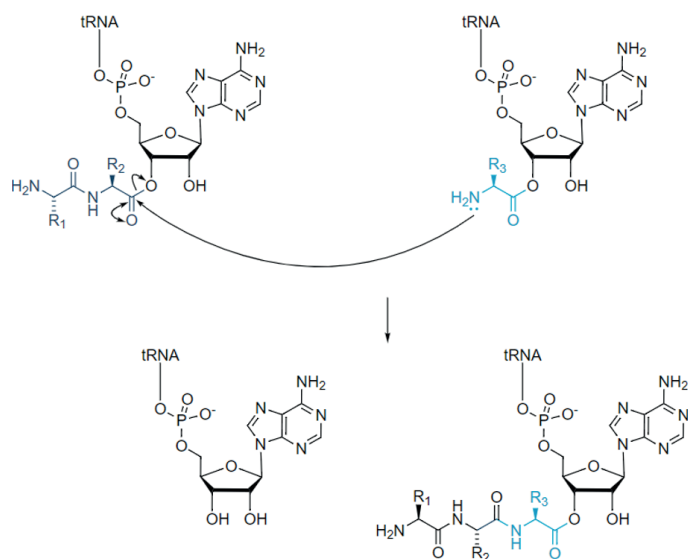


Figure 9. Protein synthesis. Reaction of the growing polypeptide chain with the 3'-end of the charged tRNA. The amino acid is transferred from the tRNA molecule to the protein.

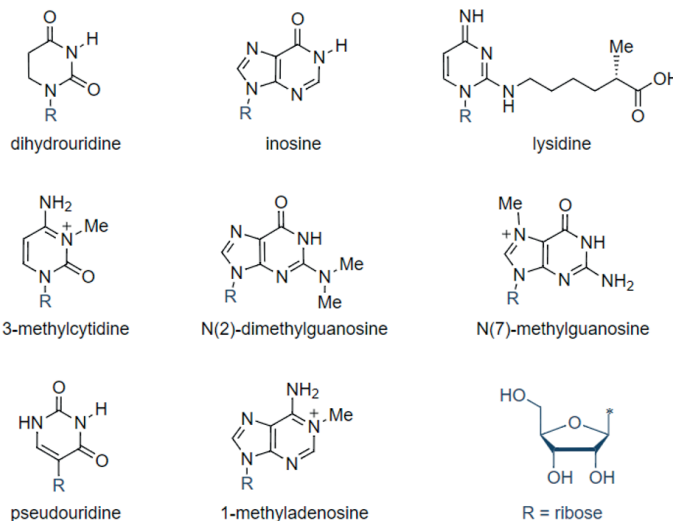


Figure 10. Modified bases in tRNA. Structures of some of the modified bases found in tRNA.

The Genetic Code

The genetic code is almost universal. It is the basis of the transmission of hereditary information by nucleic acids in all organisms. There are four bases in RNA (A,G,C and U), so there are 64 possible triplet codes ($4^3 = 64$). In theory only 22 codes are required: one for each of the 20 naturally occurring amino acids, with the addition of a start codon and a stop codon (to indicate the beginning and end of a protein sequence). Many amino acids have several codes (degeneracy), so that all 64 possible triplet codes are used. For example Arg and Ser each have 6 codons whereas Trp and Met have only one. No two amino acids have the same code but amino acids whose side-chains have similar physical or chemical properties tend to have similar codon sequences, e.g. the side-chains of Phe, Leu, Ile, Val are all hydrophobic, and Asp and Glu are both carboxylic acids (see genetic code). This means that if the incorrect tRNA is selected during translation (owing to mispairing of a single base at the codon-anticodon interface) the misincorporated amino acid will probably have similar properties to the intended tRNA molecule. Although the resultant protein will have one incorrect amino acid it stands a high probability of being functional. Organisms show “codon bias” and use certain codons for a particular amino acid more than others. For example, the codon usage in humans is different from that in bacteria; it can sometimes be difficult to express a human protein in bacteria because the relevant tRNA might be present at too low a concentration.

An Exercise in the use of the Genetic Code

One strand of genomic DNA (strand A, coding strand) contains the following sequence reading from 5’ to 3’:

TCGTCGACGATGATCATCGGCTACTCGA

This strand will form the duplex

5’-TCGTCGACGATGATCATCGGCTACTCGA-3’
3’-AGCAGCTGCTACTAGTAGCCGATGAGCT-5’

The sequence of bases in the other strand of DNA (strand B) written 5’ to 3’ is therefore

TCGAGTAGCCGATGATCATCGTTCGACGA

In the mRNA transcribed from strand A of DNA, the sequence of bases written 5’ to 3’ is

UCGAGUAGCCGAUGAUGAUCGUCGACGA

resulting in an amino acid sequence

Ser-Ser-Ser-Arg-STOP

However, if DNA strand B is the coding strand the mRNA sequence will be

UCGUCGACGAUGAUGAUCGGCUACUCGA

and the amino-acid sequence will be

Ser-Ser-Thr-Met-Ile-Ile-Gly-Tyr-Ser-

| First base (5'-end) | Middle Base | Third Base ('3-end) | | | |
|---------------------|-------------|---------------------|-----|------|------|
| | | U | C | A | G |
| U | U | Phe | Phe | Leu | Leu |
| | C | Ser | Ser | Ser | Ser |
| | A | Tyr | Tyr | Stop | Stop |
| | G | Cys | Cys | Stop | Trp |
| C | U | Leu | Leu | Leu | Leu |
| | C | Pro | Pro | Pro | Pro |
| | A | His | His | Gln | Gln |
| | G | Arg | Arg | Arg | Arg |
| A | U | Ile | Ile | Ile | Met |
| | C | Thr | Thr | Thr | Thr |
| | A | Asn | Asn | Lys | Lys |
| | G | Ser | Ser | Arg | Arg |
| G | U | Val | Val | Val | Val |
| | C | Ala | Ala | Ala | Ala |
| | A | Asp | Asp | Glu | Glu |
| | G | Gly | Gly | Gly | Gly |

The Wobble Hypothesis

Close inspection of all of the available codons for a particular amino acid reveals that the variation is greatest in the third position (for example, the codons for alanine are GCU, GCC, GCA and GCG). Crick and Brenner proposed that a single tRNA molecule can recognize codons with different bases at the 3'-end owing to non-Watson-Crick base pair formation with the third base in the codon-anticodon interaction. These non-standard base pairs are different in shape from A·U and G·C and the term wobble hypothesis indicates that a certain degree of flexibility or “wobbling” is allowed at this position in the ribosome. Not all combinations are possible; examples of “allowed” pairings are shown in Figure 11.

The ability of DNA bases to form wobble base pairs as well as Watson-Crick base pairs can result in mismatches occurring during DNA replication. If not repaired by DNA repair enzymes, these mismatches can lead to genetic diseases and cancer.

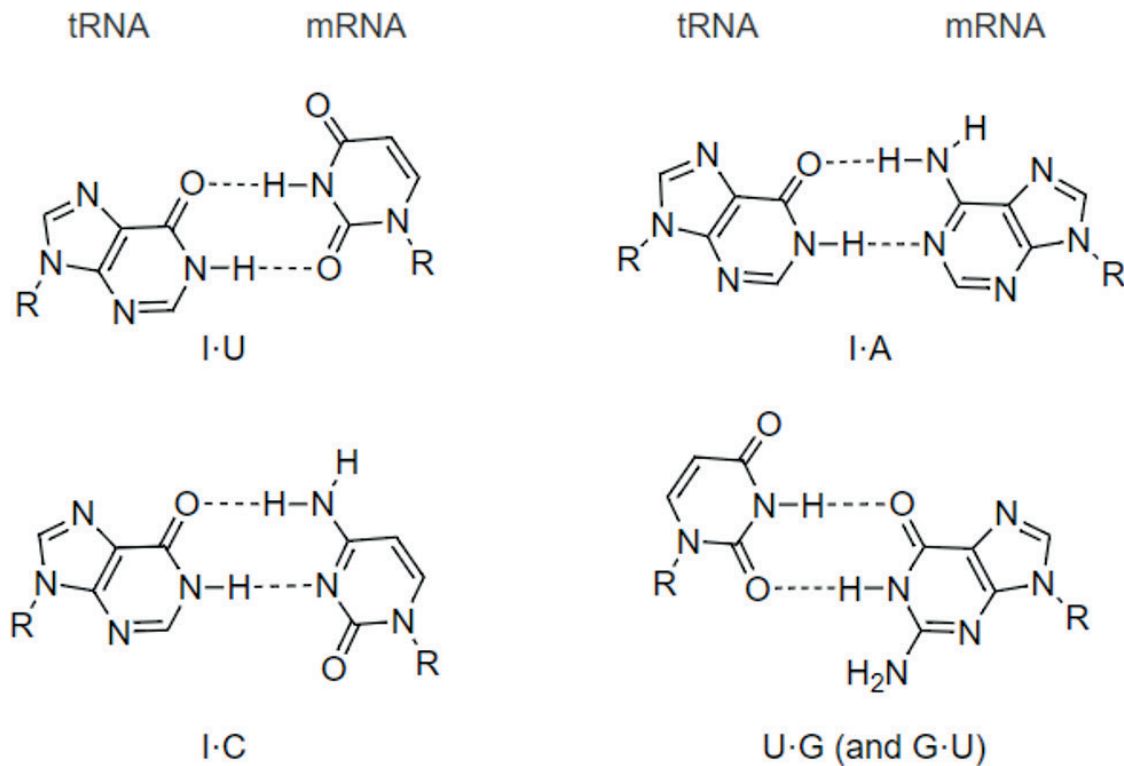


Figure 11. Structures of wobble base pairs found in RNA.

ATDBio, now a part of Biotage, offer Complex Bespoke DNA/RNA Oligonucleotides for your Diagnostic and Therapeutic Needs

Oligonucleotides are the backbone of precision medicine and the future of personalized therapy. We synthesize and deliver the best DNA/RNA oligos, aptamers and triphosphates for your precise needs.

The use of oligonucleotides in therapeutic and diagnostic research and development is rising exponentially, as new solutions are sought for the diagnosis and treatment of illnesses. Cutting edge research often requires non-standard oligonucleotides which can be extremely challenging to synthesize with any consistency of quality, yield and purity. Biotage have the in-house expertise to tackle these complex requirements and can partner with you to develop bespoke oligonucleotides for your precise needs.

The Custom Synthesis of Oligonucleotides

- » Synthesis scales from 0.2–30 μmol , larger scales available on request
- » Over 400 modifications — from dyes and quenchers to modified bases and backbones
- » Purification using HPLC or PAGE
- » Analysis using CE, HPLC, or Mass Spec
- » Lyophilised or adjusted to a specific concentration
- » Lengths of up to 170 bases, including complex modifications — longer modified oligos possible using patent-protected click chemistry technology



now part of





Your Complete Partner for Effective Chemistry

Biotage is a worldwide supplier of instruments and accessories designed to facilitate the work of scientists in life sciences. With our deep knowledge of the industry, academic contacts and in-house R&D teams, we can deliver the best solutions to your challenges. We take great pride in our flexibility and ability to meet our customer's individual needs. With strong foundations in analytical, organic, process, and biomolecule chemistry, we can offer the widest range of solutions available on the market.

EUROPE

Main Office: +46 18 565900
Fax: +46 18 591922
Order Tel: +46 18 565710
Order Fax: +46 18 565705
order@biotage.com
Support Tel: +46 18 56 59 11
Support Fax: + 46 18 56 57 11
eu-1-pointsupport@biotage.com

NORTH & LATIN AMERICA

Main Office: +1 704 654 4900
Toll Free: +1 800 446 4752
Fax: +1 704 654 4917
Order Tel: +1 800 446 4752
Order Fax: +1 704 654 4917
ordermailbox@biotage.com
Support Tel: +1 800 446 4752
us-1-pointsupport@biotage.com

JAPAN

Tel: +81 3 5627 3123
Fax: +81 3 5627 3121
jp_order@biotage.com
jp-1-pointsupport@biotage.com

CHINA

Tel: +86 21 68162810
Fax: +86 21 68162829
cn_order@biotage.com
cn-1-pointsupport@biotage.com

KOREA

Tel: +82 31 706 8500
Fax: +82 31 706 8510
korea_info@biotage.com
kr-1-pointsupport@biotage.com

INDIA

Tel: +91 11 45653772
india@biotage.com

Distributors in other regions
are listed on www.biotage.com

Literature Number: PPS702

© 2023 Biotage. All rights reserved. No material may be reproduced or published without the written permission of Biotage. Information in this document is subject to change without notice and does not represent any commitment from Biotage. E&OE. A list of all trademarks owned by Biotage AB is available at www.biotage.com/legal. Other product and company names mentioned herein may be trademarks or registered trademarks and/or service marks of their respective owners, and are used only for explanation and to the owners' benefit, without intent to infringe.

