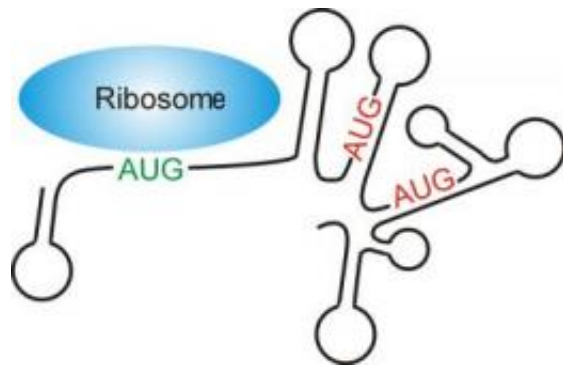# A lack of structure facilitates protein synthesis

June 28 2011

This is the mechanism for the recognition of an AUG start codon by the ribosome (blue) in the absence of a Shine-Dalgarno sequence. The correct AUG start codon (green) differs from all other AUG triplets (red) through its position in a single-strand (unstructured) region of the messenger RNA (black). Credit: MPI of Molecular Plant Physiology

Having an easily accessible starting point on messenger RNA increases protein formation, scientists from the Max Planck Institute of Molecular Plant Physiology in Potsdam have discovered.

Texts without spaces are not very legible, as they make it very difficult for the reader to identify where a word begins and where it ends. When genetic information in our cells is read and translated into proteins, the enzymes responsible for this task face a similar challenge. They must find the correct starting point for protein synthesis. Therefore, in

organisms with no real nucleus, a point exists shortly before the start codon, to which the enzymes can bind particularly well. This helps them find the starting point itself. However, genes that do not have this sequence are also reliably translated into proteins. Scientists from the Max Planck Institute of Molecular Plant Physiology in Potsdam have discovered that the structure of the messenger RNA probably plays a crucial role in this process.

The DNA of all organisms consists of the four bases adenine, cytosine, thymine and guanine, which are abbreviated using the letters A, C, T and G. In RNA, thymine is replaced by uracil (U). The bases are connected with each other by their sugar-phosphate backbone. They are comparable to the letters in our alphabet, which can be put together to form words. In the language of DNA, groups of three bases, known as triplets, code for the 20 amino acids, from which all proteins are made. Because there are no empty spaces between the triplets in the DNA, it is difficult to recognise the three bases that belong to one triplet and, particularly, to identify the starting point for protein synthesis on the nucleic acid strand.

Before proteins can be manufactured, the DNA is transcribed into its transport form, the messenger RNA (mRNA), and introduced into the cell plasma. Small protein factories, the ribosomes, bind to the mRNA here and commence with their work. They "read" the series of bases and translate them into amino acids. They begin this task neither directly at the beginning of the mRNA nor at a random point, but always at an AUG base triplet, the start codon. This triplet codes for the amino acid methionine which thus constitutes the first amino acid in every protein. However, methionine can also appear at other locations in the protein. Therefore, the question arises as to how the ribosomes know whether an AUG codon is a start signal or not.

To this effect, the Shine-Dalgarno sequence (SD sequence) usually

comes to the aid of prokaryotes, unicellular organisms without a true nucleus. The SD sequence is an mRNA base sequence that has remained virtually unchanged over the course of evolution and is found near the start codon. The ribosomes have an anti-Shine-Dalgarno sequence that can form a strong bond with the SD sequence. If a ribosome wanders along the mRNA in search of the start codon, it is detained by the SD sequence and consequently recognises the correct starting point for protein synthesis. However, mRNAs exist that do not have a Shine-Dalgarno sequence; their ribosomes nevertheless succeed in tracking down the correct AUG triplet. The mechanism that enables the correct identification of the start signal has been completely unclear until now.

According to the latest findings, the structure – or to be more precise the lack of structure – of the mRNA appears to be the factor at play here. Lars Scharff and Liam Childs from the Max Planck Institute of Molecular Plant Physiology in Potsdam examined tens of thousands of genes from different prokaryotes and cell organelles for the presence of a Shine-Dalgarno sequence. They discovered that, depending on the organism, between 15 and 50 percent of all genes do not have an SD sequence. The fact that the ribosomes also recognise the start codon on these mRNAs is probably because it is particularly easy to access. The mRNA is not usually present as a long thread but forms loops and hairpin structures. However, a ribosome can only bind to unstructured areas of the mRNA and this, it would appear, is where the secret lies: "Unlike in genes with a Shine-Dalgarno sequence, the mRNA around the start codon in genes with no Shine-Delgarno sequence has hardly any folded structures," Scharff explains.

In an experiment, the researchers introduced mutations to destroy the SD sequence, and the rate at which the mRNA was translated into proteins declined drastically. "As soon as we inserted a second mutation that simultaneously unfolded the structure of the mRNA at the start codon, this effect was reduced and protein building increased again," explains

Childs. Despite the missing SD sequence, the AUG codon is identified by the ribosome, as it is easier to access and not concealed in loops and convolutions.

Based on structural analyses of mRNAs, these results will facilitate the prediction of protein synthesis rates in future. In addition, it may become possible to influence the amount of proteins formed by modifying the mRNA structure in one direction or another.

More information: Lars B. Scharff, Liam Childs, Dirk Walther, Ralph Bock, Local Absence of Secondary Structure Permits Translation of mRNAs that Lack Ribosome-binding Sites, *PLoS Genet* 7(6): e1002155, doi:10.1371/journal.pgen.1002155

Provided by Max-Planck-Gesellschaft