8.3: Making a polypeptide in a bacterial cell

A bacterial cell synthesizes thousands of different polypeptides. The sequence of these polypeptides (the exact amino acids from N- to C-terminal) is encoded within the DNA of the organism. The genome of most bacteria is a double-stranded circular DNA molecule that is millions of base pairs in length. Each polypeptide is encoded by a specific region of this DNA molecule. So, our questions are how are specific regions in the DNA recognized and how is the information present in nucleic acid-sequence translated into polypeptide sequence.

To address the first question let us think back to the structure of DNA. It was immediately obvious that the one-dimensional sequence of a polypeptide could be encoded in the one-dimensional sequence of the polynucleotide chains in a DNA molecule. The real question was how to translate the language of nucleic acids, which consists of sequences of four different nucleotide bases, into the language of polypeptides, which consists of sequences of the 20 (or 22) different amino acids. As pointed out by the physicist George Gamow (1904-1968) the minimum set of nucleotides needed to encode all 20 amino acids is three; a sequence of one nucleotide (4) could encode at most four different amino acids, a sequence two nucleotides in length could encode (4) or 16 different amino acids (not enough), while a sequence of three nucleotide (4) could encode 64 different amino acids (more than enough). Although the actual coding scheme that Gamow proposed was wrong, his thinking about the coding capacity of DNA influenced those who set out to experimentally determine the actual rules of the "genetic code".

The genetic code is not the information itself, but the algorithm by which nucleotide sequences are "read" to determine polypeptide sequences. A polypeptide is encoded by the sequence of nucleotides. This nucleotide sequence is read in groups of three nucleotides, known as a codon. The codons are read in a non-overlapping manner, with no spaces (that is, non-coding nucleotides) between them. Since there are 64 possible codons but only 20 (or 22 - see above) different amino acids used in organisms, the code is redundant, that is, certain amino acids are encoded for by more than one codon. In addition there are three codons, UAA, UAG and UGA, that do not encode any amino acid but are used to mark
the end of a polypeptide, they encode "stops" or periods.

The region of the nucleic acid that encodes a polypeptide begins with what is known as the "start" codon and continues until one of the three stop codons is reached. A sequence defined by in-frame start and stop codons (with some number of codons between them) is known as an open reading frame or an ORF. At this point it is important to point out explicitly, while the information encoding a polypeptide is present in the DNA, this information is not used directly to specify the polypeptide sequence. Rather, the process is indirect. The information in the DNA is first copies into an RNA molecule (known as a messenger RNA) and it is this RNA molecule that directs polypeptide synthesis. The process of using information within DNA to direct the synthesis of an RNA molecule is known as transcription because both DNA and RNA use the same language, nucleotide sequences. In contrast polypeptides are written in a different language, amino acid sequences. For this reason the process of RNA-directed polypeptide synthesis is known as translation.

References

231 Nature of the genetic code finally revealed!: http://www.nature.com/nrmicro/journal...micro2707.html

232 when he was a professor at UC Boulder

233 The Big Bang and the genetic code: Gamow, a prankster and physicist, thought of them first: http://www.nature.com/nature/journal.../404437a0.html

Contributors

• Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.