Math, the Genetic Code and Living Organisms

J. T. Trevors

Laboratory of Microbial Technology
Department of Environmental Biology
University of Guelph
Guelph, Ontario
Canada, N1G 2W1
Tel: (519) 824-4120 ext. 3367
Fax: 519-837-0442
Email: jtrevors@uoguelph.ca

Abstract: It is known that the secrets of life are more complex than DNA and the genetic code. One secret of life is the self-assembly of the first cell with a genetic blueprint that allowed it to grow and divide. Another secret of life may be the mathematical control of life as we know it and the logical organization of the genetic code and the use of math in understanding life.

Key words: codons, logic, life, math

The Genetic Code

Math has a fundamental role in understanding the complexities of living organisms. For example, the genetic code-triplets of three bases in messenger ribonucleic acid (mRNA) that encode for specific amino acids during the translation process (synthesis of proteins using the genetic code in mRNA as the template) have some interesting and mathematical logic in their organization (Cullman and Labouygues, 1984). An examination of this logical organization may allow us to better understand the logical assembly of the genetic code and life.

The genetic code in mRNA is composed of U for uracil, C for cytosine, A for adenine and G for guanine. One noticeable feature of the genetic code is that some amino acids are encoded by several different but related base codons or triplets. There are 64 triplets or codons. Three triplets (UAA, UAG, and UGA) are nonsense codons- no amino acid corresponds to their code. The remaining 61 codons represent 20 different amino acids.

The genetic code is encoded in combinations of the four nucleotides found in DNA and then RNA. There are 16 possible combinations ($4^2$) of the four nucleotides of nucleotide pairs. This would not be sufficient to code for 20 amino acids (Prescott et al., 1993). The solution is mathematically simple. During the self-assembly and evolution of life, a code word (codon or triplet) evolved that provides for 64 or $4^3$ possible combinations. This simple code determines all the proteins necessary for life.

The genetic code is also degenerate. For example, up to 6 different codons are available for some amino acid. Another noteworthy aspect of biological messages is that minimal information is necessary to
encode the messages (Peusner 1974) and the messages can be encoded and decoded and put to work in amazingly short periods of time. A bacterial *Escherichia coli* cell can grow and divide in 1/2 an hour depending on the growth conditions. Mathematically, it could not be more simple. The genetic code is as follows.

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino acid</th>
<th>Codon</th>
<th>Amino acid</th>
<th>Codon</th>
<th>Amino acid</th>
<th>Codon</th>
<th>Amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU</td>
<td>phenylalanine</td>
<td>UCU</td>
<td>serine</td>
<td>UAU</td>
<td>tyrosine</td>
<td>UGU</td>
<td>cysteine</td>
</tr>
<tr>
<td>UUC</td>
<td>phenylalanine</td>
<td>UCC</td>
<td>serine</td>
<td>UAC</td>
<td>tyrosine</td>
<td>UGC</td>
<td>cysteine</td>
</tr>
<tr>
<td>UUA</td>
<td>leucine</td>
<td>UCA</td>
<td>serine</td>
<td>UAA</td>
<td>stop signal</td>
<td>UGA</td>
<td>none (stop signal)</td>
</tr>
<tr>
<td>UUG</td>
<td>leucine</td>
<td>UCG</td>
<td>serine</td>
<td>UAG</td>
<td>stop signal</td>
<td>UGG</td>
<td>tryptophan</td>
</tr>
<tr>
<td>CUU</td>
<td>leucine</td>
<td>CCU</td>
<td>proline</td>
<td>CAU</td>
<td>histidine</td>
<td>CGU</td>
<td>arginine</td>
</tr>
<tr>
<td>CUC</td>
<td>leucine</td>
<td>CCC</td>
<td>proline</td>
<td>CAC</td>
<td>histidine</td>
<td>CGC</td>
<td>arginine</td>
</tr>
<tr>
<td>CUA</td>
<td>leucine</td>
<td>CCA</td>
<td>proline</td>
<td>CAA</td>
<td>glutamine</td>
<td>CGA</td>
<td>arginine</td>
</tr>
<tr>
<td>CUG</td>
<td>leucine</td>
<td>CCG</td>
<td>proline</td>
<td>CAG</td>
<td>glutamine</td>
<td>CCG</td>
<td>arginine</td>
</tr>
<tr>
<td>AUU</td>
<td>isoleucine</td>
<td>ACU</td>
<td>threonine</td>
<td>AAU</td>
<td>asparagine</td>
<td>AGU</td>
<td>serine</td>
</tr>
<tr>
<td>AUC</td>
<td>isoleucine</td>
<td>ACC</td>
<td>threonine</td>
<td>AAC</td>
<td>asparagine</td>
<td>AGC</td>
<td>serine</td>
</tr>
<tr>
<td>AUA</td>
<td>isoleucine</td>
<td>ACA</td>
<td>threonine</td>
<td>AAA</td>
<td>lysine</td>
<td>AGA</td>
<td>arginine</td>
</tr>
<tr>
<td>AUG</td>
<td>methionine</td>
<td>ACG</td>
<td>threonine</td>
<td>AAG</td>
<td>lysine</td>
<td>AGG</td>
<td>arginine</td>
</tr>
<tr>
<td>GUU</td>
<td>valine</td>
<td>GCU</td>
<td>alanine</td>
<td>GAU</td>
<td>aspartic acid</td>
<td>GGU</td>
<td>glycine</td>
</tr>
<tr>
<td>GUC</td>
<td>valine</td>
<td>GCC</td>
<td>alanine</td>
<td>GAC</td>
<td>aspartic acid</td>
<td>GGC</td>
<td>glycine</td>
</tr>
<tr>
<td>GUA</td>
<td>valine</td>
<td>GCA</td>
<td>alanine</td>
<td>GAA</td>
<td>glutamic acid</td>
<td>GGA</td>
<td>glycine</td>
</tr>
<tr>
<td>GUG</td>
<td>valine</td>
<td>GCG</td>
<td>alanine</td>
<td>GAG</td>
<td>glutamic acid</td>
<td>GGG</td>
<td>glycine</td>
</tr>
</tbody>
</table>

- 16 codons have U as the first base
- 16 codons have C as the first base
- 16 codons have A as the first base
- 16 codons have G as the first base
- 16 codons have U as the second base
- 16 codons have C as the second base
- 16 codons have A as the second base
- 16 codons have G as the second base
- 16 codons have U as the third base
- 16 codons have C as the third base
- 16 codons have A as the third base
- 16 codons have G as the third base

- 6 codons for leucine
- 6 codons for arginine
- 6 codons for serine
- 5 codons - There are no sets of five codons for any amino acids.
- 4 codons for valine
- 4 codons for proline
- 4 codons for threonine
- 4 codons for alanine
- 4 codons for glycine
- 2 codons for phenylalanine
- 2 codons for tyrosine
- 2 codons for histidine
- 2 codons for asparagine
- 2 codons for lysine
- 2 codons for aspartic acid
- 2 codons for glutamic acid
- 2 codons for cysteine
- 2 codons for glutamine
- 1 codon for histidine
- 1 codon for glutamic acid
- 1 codon for glutamine
- 1 codon for tryptophan
- 1 codon (AUG) for methionine (start codon for translation)

Selenocysteine (21st amino acid encoded by the genetic code) codon is UGA, normally a stop codon. Selenocysteine is a derivative of cysteine in which the sulfur atom is replaced by a selenium atom that is an essential atom in a small number of proteins, notably glutathione peroxidase. These proteins are found in prokaryotes and eukaryotes ranging from *Escherichia coli* to humans. The selenocysteine is incorporated into proteins during translation in response to the UGA codon. This amino acid is readily oxidized by oxygen. Enzymes containing this amino acid must be protected from oxygen. Possibly it is one of the first amino acids that was present before oxygen was present on the Earth. As the oxygen concentration increased, the
selenocysteine may have been gradually replaced by cysteine with the codons UGU and UGC (Madigan et al. 1997).

The 3 base code sometimes only differs in the third base position. For example, the genetic code for glycine is GGU, GCC, GGA or GGG. Only the third base is variable. A similar third base change pattern exists for the amino acids lysine, asparagine, proline, leucine and phenylalanine. These relationships are not random. For example, UUU codes for the same amino acid (phenylalanine) as UUC. In some codons the third base determines the amino acid. The second base is also important. For example, when the second base is C, the amino acid specified comes from a family of four codons for one amino acid except for valine.

Biological expression is in the form of coded messages. These messages are the information on structure (shapes of biomolecules) and function (biochemical reactions necessary for life). The coded message determines the protein which folds into a shape that requires the minimal amount of energy. Therefore, the total energy of attraction and repulsion between atoms is minimal.

How did this genetic code come to be the code of life as we know it? Nature had abundant time (billions of years) to experiment with different coding schemes and eventually adopted the genetic code we have today. It is simple in terms of mathematics. It is also conserved but can be mutated at the DNA level and also repaired. The code is thermodynamically possible and consistent with the origin, evolution and diversity of life.

**Life and math**

Math as applied to understanding biology has countless uses. It is used to elucidate trends, patterns, connections, and relationships in a quantitative manner that can lead to important discoveries in biology. How can math be used to understand living organisms? One way to explore this relationship is to use examples from the bacterial world. The reader is also referred to an excellent text by Stewart (1998) that illustrates how math can be used to elucidate a fuller understanding of the natural world. For example, the exponential growth of bacterial cells (1 cell → 2 cells → 4 cells → 8 cells → 16 cells and so on) is essential information that is one of the foundations of microbiology research. Exponential growth over known periods of time is essential in the understanding of bacterial growth in countless areas of research. The ability to use math to describe growth per units of time is an excellent example of the interrelationship between math and the capability to understand this aspect of life.

For example, the basic unit of life is the cell, an entity of one. Bacteria also multiply by dividing. Remember that life is composed of matter and matter is composed of atoms and that atoms especially in solids are arranged in an efficient manner into molecules that minimize the energy needed to take on specific configurations. Often these arrangements or configurations are repeating units of monomers that make up polymers. Stewart (1998) described it very well in his excellent book when he posed the question. "What could be more mathematical than DNA?" The ability of DNA to replicate itself exactly and at the same time change ever so slightly allows evolutionary changes to occur. The mathematical sequence of 4 different bases (adenine, thymine, guanine and cytosine) in DNA are the blueprint of life. Again, the order of the 4 bases determines the mRNA (messenger ribonucleic acid) sequence and then the protein that is synthesized.

DNA in a cell is also capable of precisely replicating itself in a cell. The replicated DNA can then partition into each new cell when one cell divides and becomes 2 cells. The DNA can only replicate with the assistance of enzymes that unwind the DNA and allows the DNA strands to act as templates for the synthesis of the second strand. The ability of a cell to unwind its DNA, replicate or copy new strands and then partition them between 2 new cells has a mathematical basis. The four bases are paired in a specific manner-A (adenine) with T (thymine), C (cytosine) with G (guanine) on the opposite strands along a sugar phosphate backbone. Each strand can contain all 4 bases in any order. However, A must bond with T and C with G on opposite strands. This precise mathematical pairing must be obeyed.

Living organisms also have amazing mathematical order and symmetry (arrangement of even distribution of elements, as in a design). The repeating units of fatty acids, glycerol and phosphate that make up a phospholipid membrane bilayer are one example.

An excellent example of mathematical symmetry is the S-layer in many Archaea bacterial (prokaryotes consisting of methanogens, most extreme halophiles and hyperthermophiles and Thermoplasma) cell walls that exhibit a hexagonal configuration. A cell that can assemble the same repeating units countless times is efficient and reduces the numbers of errors incorporated into the assembly. This is exactly the characteristic that is needed for a living cell to grow and divide. Yet, a little bit of change can occur over time.

Biochemical reactions in cells are accompanied by gains or losses in energy during the reactions. Some of the energy is lost as heat and is not available to do work. In humans, heat is used to maintain a normal body temperature. The energy available to the cell is expressed as free energy and can be expressed as kJ/mole. Without the use of math and units of measurement, it would be impossible to describe energy metabolism in cells. Nor would we be able to
describe the rates of enzyme reactions necessary for the self-assembly and functioning of life.

Without units of temperature, we would not be able to describe the lower, upper and optimum growth temperatures of specific organisms microorganisms. The pH ranges for bacterial growth and the optimum pH values for enzyme reactions would be unknown without math to describe the values. Water availability values and oxygen concentrations would not be able to be described for growth of specific organisms. The examples are numerous. Without the use of math and scientific units to express values, our understanding of life would be minimal and biology would not have made the great advances that it has in the past decades.

One central characteristic of living organisms is reproduction. From nutrients in their environment, they can self-assemble new cells in virtually exact copies. Secondly, living organisms are interdependent on each other and their activities. The Earth's biosphere with its abundance of oxygen and living organisms was self-assembled by living organisms. From a chaotic lifeless environment on the early Earth, life self-assembled with the cell as the basic unit, with mathematically precise order, symmetry and base pairing in DNA as the genetic blueprint and with triplet codons as the genetic code for protein synthesis.

There is more to life than the genomic blueprint of each organism. Life functions within the natural laws that we know and the ones that we do not know. Math can be used to understand life from the molecular to the biosphere level. For example, this includes the origin and evolution of organisms, the nature of the genomic blueprints and the universal genetic code as well as ecological relationships. Math as applied to life, helps us understand how monomers become polymers necessary for the assembly of cells. Math helps us look for trends, patterns and relationships that may or may not be obvious to scientists. Math allows us to describe the dimensions of genes, sizes of organelles, cells, organs and whole organisms. Without this knowledge, a paucity of information would still exist on many aspects of life.

Acknowledgements
This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) operating grants program.

REFERENCES

Call For Resolutions
The Steering Committee of ACUBE requests that the membership submit resolutions for consideration at the 2000 Annual meeting to the Chair of the Resolutions Committee. The proposed resolutions will be published in the August 2000 issue of Bioscene. Submit proposed resolutions to:

Dr. Richard Wilson
Dept. of Biology
Rockhurst University
1100 Rockhurst Rd
Kansas City, MO 64110
Phone (816) 501-4048
wilson@vax1.rockhurst.edu