Using DNA for Data Storage Including Quantum Dot Identifiers and Blockchain

Austin, Tamarron (School: Navajo Preparatory School)

My research provides the theoretical storage of data in DNA, expressed using the four natural “A,” “C,” “G,” and “T” DNA nucleobases. The DNA-stored information may be actual biological genomes. Alternately, the DNA-stored information may be man-made for data-storage, such as data generated by a computer [1]. My research used Hexadecimal (Base16), which counts from zero to fifteen, and Base64, which counts from zero to sixty-three. Base16 encodes a binary byte, such as 11111111 into two Hexadecimal numbers, such as FF. Base64 encodes three binary bytes, such as 11111111 11111111 11111111, or six hexadecimal numbers, such as FF FF FF, into four Base64 symbols, such as 111111 111111 111111 111111, which is represented by //// in Base64. Base64 lends itself to the DNA molecule because the nucleobases A,C,G, and T form a Base4 (0,1,2,3) system. A triplet of three nucleobases forms a DNA Codon, and there are sixty-four DNA Codons in all. This number of DNA Codons, sixty-four, is arrived at by there being four choices for the first nucleobase in the codon triplet, four choices for the second nucleobase, and four choices for the third nucleobase, and 4*4*4 gives sixty-four DNA Codons in all. A doubles of two nucleobases gives 4*4 or 16 unique nucleobase pairs which maps into Base16 for data storage. Base64 is preferred if the required number of data-bearing nucleobases is divisible by three and thus the data-bearing nucleobases can be organized into Codons. However, Base16 is preferred if the required number of data-bearing nucleobases is divisible by 2. The data-bearing strands of DNA can be labeled or identified by quantum-dots at the beginning and end of the strand. Blockchain can ensure the integrity of the stored data.