General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

CTS - B - 70 - 3

A MECHANISM FOR THE EVOLUTION OF THE GENETIC CODE

MICHAEL CONRAD

APRIL, 1970

Center For Theoretical Studies
University of Miami / Coral Gables, Florida 33124

A MECHANISM FOR THE EVOLUTION OF THE GENETIC CODE*

Michael Conrad
Center for Theoretical Studies
University of Miami, Coral Gables, Florida 33124

ABSTRACT

Multiple coding is proposed as a mechanism facilitating the evolution of the genetic code. Multiple coding can occur when several information storing molecules share the same cytoplasm. These molecules may code for different construction machinery--that is, for different coding systems. Initially this will reduce the efficiency of the system, but will not be lethal if good proteins are still produced. An alternative coding system is retained if it can lead to the production of one useful protein. Under certain conditions the information in the genetic molecules associated with the alternate coding system will be rectified, and this coding system will become predominant.

April, 1970

*Research supported by the National Aeronautics and Space Administration Grant Number NGL 10-007-010.

Descriptions of the historical development of nature are often not as easily verifiable or verifiable in the same way as descriptions of phenomena accessible to direct observation or experimentation. This is especially so apropos the evolution of the genetic code since there is no historical record. However, some question has arisen as to whether the code could have evolved by natural selection, since slight alterations are almost always lethal in present day organisms. It therefore seems worthwhile to exhibit a mechanism by which the modern code could evolve from a primitive form in small steps.

Hereditary Processes

As is well known, biological systems store information in nucleic acids. The sequence of bases in DNA describes the sequence of amino acids in structural and enzymatic proteins, and also the sequence of bases in some RNA molecules which are not translated into proteins. DNA does not describe itself, since this would raise logical difficulties; rather it is replicated. The replication process must be under some control if it is to occur at the appropriate time. The mapping between nucleic acids and proteins is determined by construction machinery. In modern cells this includes messenger and transfer RNA molecules, ribosomes, and protein enzymes. In particular, the transfer RNA molecules are adaptors which bind specifically to a code word in the nucleic acid alphabet and to an amino acid of the protein alphabet.

The essence of this process is the specific cross catalysis of nucleic acids and proteins. At present, the specificity for replication of nucleic acids is apparently in the proteins (Commoner, 1964). Accordingly, processes of transcription and translation are both determined by template nucleic acid and specific catalyst, although the

sequencing can be predicted from the former alone. This fact is based on the high degree of correlation between information store and construction machinery--on the fact that the information store describes the construction machinery which reads it out. The simultaneous appearance of such compatible sequences is unlikely and could not be expected to arise from any straightforward chemical reaction or cataclysmic mutation. Certainly the code evolved by natural selection once a primitive coding system established itself.

Evolutionary Processes

Evolution through natural selection requires a set of heritable traits from which a subset may be selected. Ordinarily such a set or distribution characterizes a population of organisms. This is possible if the capacity of the system to survive is not an all or none function of the traits.

In the present case the traits are the coding systems, and the question naturally arises as to whether a distribution of coding systems is possible. The coding system is described by nucleic acid in the information store, and novel coding systems can arise through mutations. However, alterations in the code are coordinated to alterations in the enzymes. For example, suppose that the code word associated with amino acid A becomes associated with amino acid B. Then there will be massive replacement of amino acid A by amino acid B. In a present day system this would be lethal, even if the new coding system were more fit in some sense. Biological systems may have been simpler when the code assumed its final form. However, a coding system, just insofar as it is a dictionary of associations between code words and amino acids, must subsume considerable complexity and specificity.

A distribution of coding systems can be achieved if this distribution characterizes what is in some sense the organism rather than the population. This condition of double or multiple coding might be achieved in a number of ways. For example, imagine that several nucleic acid descriptions share what is essentially one cytoplasm, and that one of these descriptions codes for different construction machinery than the others. This will not necessarily be lethal since most of the enzymes synthesized will be synthesized according to the original code. Initially there will be waste syntheses since many of the enzymes produced through the new coding system will not be fully active. However, the new coding system will be retained if it results in the production of one or two useful proteins, so that the overall capabilities of the system are improved. In the course of evolution the information encoded in the description associated with the new coding system could be corrected or rectified. For example, suppose that in code I there are two words for amino acid A and that in code II one of these words is assigned to amino acid B. Then the rectification process would involve a sufficient number of point mutations to restore the original situation, or perhaps a smaller number if some of the replacements are advantageous. Such a process is not very improbable since each correction will increase the fitness of the system. Of course, rectification is not the only possible eventuality, but in certain instances it would be favored. This would be true, for example, if amino acid B were newly taken into the system and therefore opened up many possibilities for improvement. In any case, those systems in which a superior code became predominant would have preferential survival curves. In this sense a transfer of function has taken place,

since the original function of the new code was to provide a useful protein, but the final function is to provide the dominant form of coding.

Essentially, double coding is possible because of a synergic relationship between two coding systems. Under these circumstances it is possible that there will be cross reactions. That is, coding system I, associated with description I, might interact with description II, and conversely. Extensive cross reactions could cancel the advantages of such a relationship, however, and it must be supposed that such reactions are limited in some way. For example, descriptions and construction machinery may have formed units, in the sense of being localized in distinct areas, while substances such as high energy compounds and enzymes would be more diffusible. Alternatively, cross reactions would not be too severe if a large number of descriptions shared the same cytoplasm, so that descriptions coding for the unusual construction machinery formed a small minority; initially such descriptions might be maintained because they cannot be eliminated in a short amount of time, but later they could acquire a function.

The forms of organization capable of supporting multiple coding are almost inevitable, and in fact might be favored. The duplication of the nucleic acid description and cytoplasm is coordinated by a control device. This control may be set so that the duplication processes are in step, or may be set otherwise. In any event the control in early systems could not have been too precise, and furthermore the control is subject to mutation. Therefore, situations must have arisen in which several descriptions in the nucleic acid alphabet shared the same cytoplasm or its primitive equivalent. These descriptions also

code for the construction machinery and therefore for the code itself. It follows that the system could exhibit several coding processes at once, or multiple coding, since the description of the construction machinery is also mutable. Furthermore, if the duplication of the nucleic acid description were complete, but the duplication of the cytoplasm only partially complete, the conditions for limiting cross reactions would be achieved. Such a form of organization has advantages even in the absence of multiple coding. For example, it allows for some diversity in the system's enzymes, but this diversity is more consistent with enzymes specifically designed for various external conditions than it would be if it arose from unreliable syntheses. Furthermore, recombination processes are more easily achieved within this framework. Both of these mechanisms are homeostatic. They may have been important for primitive systems which could not support more sophisticated mechanisms for homeostasis--mechanisms for which a sophisticated code is a precondition.

Regardless of these advantages, multiple coding per se will persist only so long as it is of some survival value to the system--that is, if it results in the production of a useful protein. This protein need not be an immediate consequence of the appearance of the new coding system, for there will be a period of time over which it can develop. This is so since it will take a certain amount of time to eliminate a system with slightly decreased efficiency. In particular this will be true if competition is not severe, or in case of geographic isolation.

Examples of Multiple Coding

Some hypothetical examples might be worthwhile. The modern code exhibits redundancies. For example, a change in a single nucleotide

will often, but not always, result in a word which codes for the same or a homologous amino acid. The origin of these redundancies has been explained in terms of their survival value in attenuating the effects of mutation (Sonneborn, 1965) and mistranslation (Woese, 1965), or in terms of the way new amino acids were brought into the system (Crick, 1967). Both of these mechanisms are readily consistent with multiple coding, and, in fact, in this context they are not necessarily exclusive of one another. For example, a set of triplets coding for a single amino acid in one code might be partitioned to code for two amino acids in a derivative code. Multiple coding would generally be necessary since it is otherwise unlikely that the resulting substitutions in all of the system's enzymes would be nonlethal. The redundancies may have come about largely on the basis of evolutionary factors, although steric influences are not inconsistent with multiple coding. This would be true since homologous substitutions are more likely to be satisfactory, with the consequence that fewer rectifications would be necessary. Likewise, initial inefficiency prior to the rectification process would be less significant if a homologous amino acid were brought in. Furthermore, the overall fitness of such systems would not be unaffected by the results of mutation or the accuracy of translation. Thus, in many instances one of a number of competing modifications of the code would have a better chance of establishing itself if it were associated with a redundancy. It might be noted, however, that the elimination of nonsense mutations is one process for which multiple coding is not necessary and for which it might not be as efficient, since nonsense mutations would not be so deleterious in multiply coded systems.

As a second example, consider a change in word length in the

nucleotide language--this is undoubtedly the most difficult situation. Suppose that at some stage the primitive cytoplasm contained some construction machinery which reads in doublets and a small amount which reads wholly or partially in triplets. This might arise, for example, through changes in the structure of the adaptors or by a modification of the construction machinery which distorts the adaptors. As usual the proteins synthesized through the triplet code are initially inactive, but this will not be lethal. The triplet code is clearly superior since it can specifically bring more amino acids into the system. However, the original value of triplet coding may have been associated with the possibility of greater accuracy or redundancy. Evidently the rectification process in this case is more difficult than for the previous example, since much more information must be lost. However, since this type of transition would have taken place early in the history of life, the information stored in the nucleic acid molecules would not be so large.

Of course, it is altogether possible that no such transition took place, although there are some arguments in its favor (Jukes, 1966). The point is that one could construct a large number of pathways for such a process, or for any other process having to do with the development of the code.

Early Evolution

It is important that the mutated description, coding for the new construction machinery, must initially be read by the original construction machinery. For example, the mutated description will produce coding system B under the influence of coding system A (the initial coding system); the same description will produce coding system C under

the influence of coding system B, and so forth. There is no guarantee that the process will converge, and in general a sequence of coding systems will be generated. Furthermore, a sequence of descriptions will also be generated if the description codes for the replication device, i.e., if the transcription is not a passive one dimensional crystal growth process.

These problems do not arise if the construction machinery consists of nucleic acids, assuming that these do not affect the processes of transcription. However, if alteration of the code involves a protein the mutated description must be rectified with respect to this protein also. The description must be rectified in such a way that it describes the coding machinery which reads it out. This is necessary if offspring traits are to be correlated to parental traits.

Darwinian evolution is possible only to the extent that traits of selected systems are correlated to traits of their offspring. However, it must also be possible for the offspring to be more complicated than the parent--otherwise the code could not evolve. Biological systems satisfy these two conditions by using an information store to tailor reaction constraints. This is why specific cross cataysis is the essence of heredity. (See Conrad, 1969.)

Early biological molecules may have been selected or fractionated by physical processes of the environment (Pattee, 1966; Banda, 1968). The initial interactions among such molecules may have arisen on this basis. However, Darwinian evolution could only begin when these interactions allowed for some parent-offspring correlations, with the possibility for more complicated offspring. Under these conditions we have the possibility for a transfer of function, in which nucleic

acid acquires the function of an information store, the processes of transcription and translation are isolated, and the possibilities for evolution are increased. If correlation between parent and offspring is only partial, either because of a blurred genotype-phenotype distinction or because of incomplete rectification, we might expect an ensemble of coding systems, the members of which are maintained by a synergic relationship.

Concluding Remarks

It is not necessary that the evolution of the code proceeded by multiple coding only. However, the frequency with which some general mechanism is utilized in the course of evolution must be proportional to the probability that changes favored by the current selective forces will arise through this mechanism. On this basis multiple coding should be important since the distribution of codes arises from the fact that some members of the set of nucleic acids would be maintained in a virtual state by other members of the set. Thus, the old code serves as a correction device which maintains the new code despite its initial maladaption to the external environment. The only fortuitous event has to do with the appearance of a useful protein before the system is driven to extinction by a decrease in efficiency. In the absence of multiple coding, on the other hand, it must be supposed not only that such a protein appears, but that the alterations brought about by the recoding are not lethal, and that the chances for survival are increased, despite these alterations. Furthermore, if any rectification of the encoded information is necessary it would be equally necessary in the presence or absence of multiple coding.

It is altogether possible that the code was frozen out in its

final form before it became completely optimized (Crick, 1968). Thus, as the code became more sophisticated and as enzymes became more elaborate, multiple coding may have become much more expensive in terms of initially inactive proteins. Transfer of function, both as regards the appearance of a useful function and the accumulation of information in terms of the new code, would become more improbable. If this freeze process occurred it must have done so early in the history of life since the code has a universal character. However, it might be remarked that the picture required for multiple coding is reminiscent of coenecytic organisms--where several nuclei share a common cytoplasm. A similar situation may arise in bacteria when the genetic molecules duplicate more rapidly than the cytoplasm, and it is therefore not an implausible form of organization. It is possible that if there are any deviations from the code they will be found most conspicuously in such forms, assuming that the evolution of the code continued for a few more steps, but that the opportunity never arose to exploit any slight increase in efficiency.

Acknowledgements

I am indebted to Professor Behram Kursunoglu for hospitality at the Center for Theoretical Studies. I would also like to acknowledge stimulating conversations in the general area with Dr. H.H. Pattee.

References

Banda, P.W. (1968). Thesis, Biophysics Program, Stanford Univer[illegible]ty.

Commoner, B. (1964). Nature, 202, 960.

Conrad, M. (1969). Thesis, Biophysics Program, Stanford University.

Crick, F.H. (1967). Nature, 213, 119.

Crick, F.H. (1968). J. Mol. Biol., 38, 367.

Jukes, T.H. (1966). "Molecules and Evolution", New York: Columbia University Press.

Pattee, H.H. (1966). In "Natural Automata and Useful Simulations", p. 73 (Pattee, Edelsack, Fein, and Callahan, eds.) Washington, D.C.: Spartan Books, Inc.

Sonneborn, T.M. (1965). In "Evolving Genes and Proteins", p. 377 (V. Bryson and H.J. Vogel, eds.) New York: Academic Press.

Woese, C.R. (1965). Proc. U.S. Nat. Acad. Sci., 54, 1546.