

EVALUATION OF THE INTEREST TO PHYLOGENY OF ESTABLISHING THE AMINO-ACID SEQUENCE OF ACTIVE SITES OF ENZYMES.

*Ernie Zuckerkandl*

~~BASES AND PROTEIN MOLECULES WITH THAT ARE OF THE SAME GENERAL TYPE~~  
~~AND ARE OF THE SAME TYPE~~

With respect to problems of phylogeny, studies of sequences around the active sites of enzymes, ~~may~~ such as performed by Tuppy in cytochrome c, may be expected to be as a rule the least telling studies. Indeed <sup>(a)</sup> the changes in these regions may be expected to be very much smaller than in many other regions of the polypeptide chains. <sup>(b)</sup> Moreover what changes are tolerated by natural selection may be expected to be more limited in kind. The lys preceding the first <sup>and chicken</sup> cys in horse, beef, pig, ~~and~~ salmon/cytochrome c is changed to arg in the silkworm and in yeast, but it appears doubtful from this record that an amino-acid other than one of the ~~two~~ basic ones will be found at this site in any cytochrome c. <sup>(c)</sup> Finally, the changes that are observed around the active site more probably reflect peculiar functional requirements than phylogenetic relationships. Thus, in cytochrome c, the residue following the first cys is ala in horse, beef, pig, salmon and ~~chick~~ the silkworm, and ser in chicken. Yet, evidently, chicken is more closely related to horse, pig and beef than ~~is~~ are the salmon and the silkworm.

As a corollary to (b), it may be stated that whatever changes are observed in the ~~functionally~~ functionally most critical region of the enzyme are most likely one step -substitutions, since the chances for an intermediary step to be tolerated by natural selection are low. In accordance with this consideration the switches between ala and ser and ~~ala~~ between ala and glu (mentioned in (c) above) involve one step-mutations according to the genetic code as proposed by JUKES (1962). The genetic code proposed by SMITH (1962) is in agree-

ment inasmuch as the ala-ser change is concerned, but an ala - glu change appears here to involve at least two ~~steps~~ mutational steps. According to both codes a ser-glu change (or vice-versa) could be only indirect.

The glu -residue, at this site, is found in yeast only. If the genetic code in its final version should confirm that the change from glu to ala or to ser cannot be a one step mutation, then this fact could be used as an argument in favor of the evolutionary independence of the yeast cytochrome c and the animal cytochromes c.

Considerations of this kind will make possible statements about the likelihood of a phylogenetic relationship between two functionally analogous proteins belonging to groups that are very far removed from each other. By phylogenetic ~~relationship~~ <sup>kinship</sup> is meant here the existence of a common ancestor who already performed the function of the descendant ~~one~~ molecules. (A wider definition of phylogenetic kinship is not useful here, since the possibility cannot be excluded that all proteins, animal and vegetal, have a common ancestor).

Functionally analogous proteins belonging to vastly different forms of life may share a few characters of sequence and differ in all other characters of sequence. The question then arises whether the similarities have arisen by convergence from evolutionarily unrelated polypeptides, or whether a common polypeptide ancestor endowed with a similar function has existed, but the similarities in sequence have ~~been wiped out~~ mostly been wiped out by divergent evolutionary change. If changes are observed at the active sites of enzymes that cannot be due to one-step mutations, then the first hypothesis can be stated to be the more likely. Its likelihood will be the greater the more such differences are found between the two forms of the enzyme that are compared. The study of the amino-

acid sequence in closely or moderately related forms is, as stated, of little interest to phylogeny, but on the other hand, if forms are examined that are extremely far removed from each other on the evolutionary scale, these critical regions of the molecule are probably the most informative.

One may put it this way : If a common ancestry of two polypeptide chains is already ascertained by a sufficient number of structural homologies, the indirectness of an amino-acid substitution will indicate a degree of functional tolerance to change at a given molecular site; if the common ancestry is not established, then the indirectness of any amino-acid substitution at a molecular site presumed to be very intolerant to change represents some evidence in favor of the contention that the two polypeptide chains do not have a common ancestor (in the sense defined above).

Regarding extra-terrestrial life : If life on Mars should prove very similar to that on earth; if it is based on the same genetic code and on the activity of proteinaceous enzymes that carry out the same fundamental functions; if, furthermore, in a great number of these enzymes, whose function has been ascertained and found identical with that of corresponding telluric enzymes, amino-acid substitutions are noted at molecular sites that are found invariable on earth; if, finally, these substitutions, according to the genetic code, cannot have been one step mutations; then we shall have grounds for believing that living matter on earth and on Mars, although very similar, has not had a common origin, but must have arisen independently.

These considerations possibly furnish the best criterion to date for making a tentative statement about the independent origin of nearly identical forms of life.

## QUALIFICATION OF NOTE 1 OF JULY 3rd

An association of chemical units that is completely random must not be expected to be found in macromolecules of either chemogenic or biogenic origin :

## 1) Macromolecules of chemogenic origin.

a) The proportions of the different chemical units that make up such macromolecules cannot be expected to be equal in the original reaction mixture, hence the chemogenic product will be non-random in the sense of the predominance of some units over others.

b) The condensation of the chemical units that make up the macromolecule should perhaps not be expected to lead usually to products whose composition is very close to that of the original reaction mixture (cf. experimental results of Fox and Harada, 1960)§.

c) (Corollary to (b)): Certain associations between units, ~~xxx~~ i.e. certain neighbours in the macromolecule, may not occur, or occur to a small extent, for chemical reasons.

## 2) Macromolecules of biogenic origin.

a) On earth, the proportions of the different amino-acids in proteins, considering average figures of a number of them in different groups, are not random (cf. Vegotsky and Fox, 1962).

b) Mutations, except for the probable inherent instability of some of them, probably tend to introduce complete randomness, but this trend is counteracted by natural selection.

c) (= (c) of paragraph (1) ). Even under conditions of enzymatic synthesis certain sequences may be excluded for chemical reasons, such as perhaps the ~~xxxxxxxx~~ association of glu and pro (cf. Vegotsky and Fox).

The question arises whether, between certain limits, a more highly randomized sequence, given a certain number of distinct basic chemical units, may not be more indicative of the biogenic origin of a macromolecule than a less randomized sequence (because of the trend toward greater randomization through mutation in "primary" macromolecules, as compared to chemogenic condensation).