J. Mol. Evol. 3, 103—108 (1974)

Evolutionary Divergence in the Two Kinds of Subunits of Ribulose Diphosphate Carboxylase Isolated from Different Species of *Nicotiana*

Shiu Yuen Kwok* and S. G. Wildman

Department of Biology, Molecular Biology Institute, University of California, Los Angeles, California 90024

Received September 17, 1973

Abstract. A χ^2 analysis was made of previously reported values for the amino acid composition of the large and small subunits of ribulose diphosphate carboxylase (E.C. 4, 1, 1, 39) isolated from five species of *Nicotiana*. The distributions of these values were then compared with published values for hemoglobin α chains and cytochrome *c*'s of diverse origins. It was concluded that evolutionary diversity in the large and small subunits of RuDP carboxylase occurs even within the limited taxonomic category of the genus *Nicotiana*. The large subunit was as stable towards mutation during evolution as the α chains and cytochromes, whereas the small subunit was much less stable. The hypothesis is discussed that the large subunit played a more significant role than the small subunit in the enzyme function and/or structural integrity of the oligomeric protein. It was also speculated that the addition of small subunits to the molecule, which may have been a recent evolutionary event, enables the fixation of many evolutionarily favorable mutations. Thus, survival of the enzymatic activity in changing environments is favored.

Key words: Statistical Analysis — Amino Acid Composition — Evolution of Subunits in Ribulose Diphosphate Carboxylase.

Introduction

Ribulose diphosphate carboxylase (RuDPCase, E.C. 4, 1, 1, 39) is the enzyme contained within chloroplasts which catalyzes the primary fixation of CO_2 during photosynthesis and is probably the most abundant protein in the world. It has a molecular weight of about 5.5×10^5 daltons in spinach leaves (Kawashima and Wildman, 1970) and consists of two distinct kinds of subunits (Rutner and Lane, 1967). We assume that a similar organization probably prevails for enzyme found in other genera and families of plants. Kawashima (1969) noted a marked similarity in the amino acid composition of the large subunits of enzyme obtained from *Nicotiana tabacum* and *Spinacea oleracea* but found pronounced differences in composition of the

* Present address: The Rockefeller University, New York, N. Y., USA.

small subunits signifying a difference in the evolutionary patterns for the two subunits. In the case of enzyme from *Nicotiana* species, it has recently been shown that chloroplast DNA contains the code for the primary structure of the large subunit (Chan and Wildman, 1972) whereas nuclear genes code for the small subunit (Kawashima and Wildman, 1972). In the present study, we have estimated the degree of evolutionary divergence of the two kinds of subunits by a statistical analysis of the amino acid compositions of enzyme isolated from five species representing a random sample of the genus *Nicotiana*.

Materials and Methods

1. Plants

Five species of *Nicotiana* plants were used: *N. tabacum*, var. Turkish Samsun; *N. glutinosa*; *N. sylvestris*; *N. glauca* and *N. rustica*. The genus *Nicotiana* is classified into 3 subgenera, *viz. Rustica*, *Tabacum* and *Petunioides* (Goodspeed, 1954). Of the 5 species used, *N. glauca* and *N. rustica* belong to the subgenus *Rustica*; *N. glutinosa* and *N. tabacum* to *Tabacum*, while *N. sylvestris* belongs to *Petunioides*. Thus these 5 species constitute a random sampling of the genus *Nicotiana*.

2. Statistical Analysis of the Amino Acid Compositions of the Subunits

In his analysis of the α-chains of 11 different hemoglobins and 25 cytochrome *c*'s in order to ascertain the degree of relatedness between the proteins in evolutionary terms, Shapiro (1971) developed a parameter of relatedness called the Y value, which is essentially the χ^2 value of the amino acid compositions of the two proteins or two groups of homologous proteins. The Y parameter was proposed to be more sensitive than previous methods (Metzger *et al.*, 1968). Using Shapiro's procedure, the data were first transformed into number percentage of amino acid residues, *i. e.* number of an amino acid residue per 100 total residues. Contingency tables (2×16) were constructed for the amino acid residues in the large and small subunits of any two of the five species of *Nicotiana* RuDPCase comparing the frequencies of 16 different amino acids analyzed. From these, χ^2 values were computed. For example, to calculate the χ^2 value for the homologous subunits from 2 species of *Nicotiana* i and j, as compositions of r amino acid residues, $A_1 \ldots A_r \ldots A_{16}$, are known,

$$\chi^2_{ij} = \sum_{r=1}^{16} \left[\frac{(\text{observed no. of } A_{ir} - \text{expected no. of } A_{ir})^2}{\text{expected no. of } A_{ir}} + \frac{(\text{observed no. of } A_{jr} - \text{expected no. of } A_{jr})^2}{\text{expected no. of } A_{jr}} \right].$$

The expected number of A_{ir} can be obtained by dividing the product of total number of A_r residues in both subunits and the total number of all amino acid residues in subunit i by the total number of all amino acids in both subunits. Similarly A_{jr} can be obtained.

Experimental Results

1. The Amino Acid Composition of Subunits

The average result of duplicate amino acid analyses of the large and small subunits of RuDPCase derived from 5 species of *Nicotiana* have been

Table 1. Means and standard deviations of percentage amino acid composition for the two subunits of RuDPCase isolated from five species of *Nicotiana* (numbers calculated from original data contained in Tables 1 and 2 of Kawashima *et al.*, 1971, where probable integer values for each amino acid are also found)

Amino acid	Large subunit			Small subunit		
	Mean	Standard deviation	Standard deviation / Mean	Mean	Standard deviation	Standard deviation / Mean
Lysine	4.70	0.31	0.065	7.55	0.92	0.122
Histidine	2.84	0.11	0.039	0.66	0.23	0.348
Arginine	6.67	0.32	0.048	4.33	0.79	0.182
Aspartic acid	8.88	0.46	0.052	7.24	0.41	0.057
Threonine	5.38	0.37	0.067	4.17	0.43	0.103
Serine	3.12	0.11	0.035	3.63	0.52	0.143
Glutamic acid	10.24	0.42	0.041	15.76	0.54	0.034
Proline	4.15	0.12	0.029	6.82	0.42	0.062
Glycine	10.28	0.66	0.064	7.07	0.77	0.109
Alanine	8.44	0.41	0.048	5.33	0.30	0.056
Valine	7.35	0.22	0.003	6.72	0.71	0.106
Methionine	1.60	0.18	0.115	1.76	0.37	0.210
Isoleucine	4.37	0.11	0.024	4.46	0.34	0.076
Leucine	8.95	0.44	0.049	8.11	0.23	0.028
Tyrosine	3.68	0.19	0.051	7.22	0.64	0.089
Phenylalanine	4.50	0.19	0.042	4.38	0.56	0.128

Table 2. χ^2 values of the amino acid compositions of large and small subunits of RuDPCase isolated from five species of *Nicotiana*

	N. tabacum	*N. glutinosa*	*N. rustica*	*N. sylvestris*
		Large subunit		
N. glutinosa	3.84			
N. rustica	3.15	4.37		
N. sylvestris	3.30	7.97	4.90	
N. glauca	7.89	4.13	8.53	7.61
		Small subunit		
N. glutinosa	10.22			
N. rustica	11.18	16.69		
N. sylvestris	8.23	10.81	15.29	
N. glauca	23.55	26.12	38.42	38.01

presented in Tables 1 and 2 of a recent publication (Kawashima *et al.*, 1971). The probable numbers of each amino acid in the large and small subunit peptides were also calculated and included in the tables. The means and standard deviation of these data have been calculated and are presented in Table 1 of this paper. Among the different enzymes isolated from these closely related *Nicotiana* plants, the large subunits of RuDPCase are found

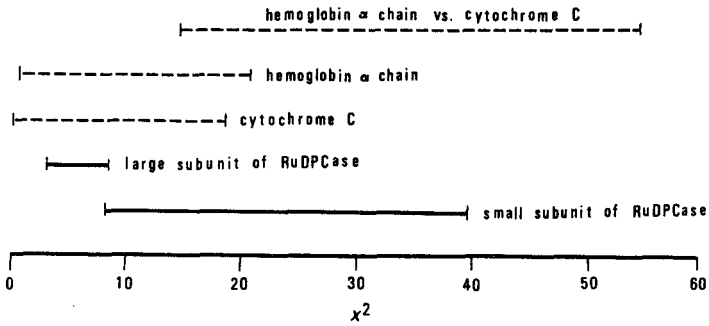

Fig. 1. Comparison of χ^2 values calculated for the amino acid composition of the large and small subunits of RuDPCase isolated from five species of *Nicotiana* to χ^2 values (Shapiro, 1971) for cytochrome *c* and hemoglobin α chain

to have very similar overall amino acid compositions, whereas the small subunits show more variations. The peptide chain of the small subunit of the *Nicotiana* RuDPCases is about the same size as the 104 residues in cytochrome *c* and 2/3 the size of the 140 residues in the hemoglobin *α* chain, whereas the peptide chains of the large subunits are about twice as large. Since the χ^2 (or *Y* values) computed by Shapiro (1971) for cytochromes and hemoglobin *α* chains appeared to be a measure of the degree of protein relatedness, the same kind of computation for the amino acids of the large and small subunits of the different RuDPCases was performed to provide a basis for estimating the degree of relatedness of the two kinds of subunits. As shown in Table 2, the χ^2 values for the amino acids of the large subunits range from 3.15 (*N. tabacum vs. N. rustica*) to 8.53 (*N. glauca vs. N. rustica*). For the small subunit, they range from 8.23 (*N. tabacum vs. N. sylvestris*) to 38.42 (*N. glauca vs. N. rustica*).

Fig. 1 depicts the distributions of the χ^2 values of the amino acids of the large and small subunits of the RuDPCases compared to those for hemoglobins and cytochromes. It is obvious that the χ^2 values obtained for the large subunits span a smaller range than the χ^2 values calculated for either the hemoglobins or cytochromes. The smaller variation is easily understood because the plants tested all belong to a single genus, *Nicotiana*, while the protein standards were derived from different classes (carp hemoglobin *vs.* pig hemoglobin), or even different kingdoms (*Rhodospirillum rubrum* cytochrome *vs.* horse cytochrome). In striking contrast, the range of χ^2 values for the small subunits exceeds those of the two protein standards at the upper limits but is still lower than the upper limit of cytochromes *vs.* hemoglobins. Whereas χ^2 (or *Y*) values were used by Shapiro (1971) to test the degree of equality of amino acid compositions and hence the relatedness of proteins, we have used his concept as an indicator of evolutionary divergency and regard the χ^2 values as providing a qualitative or even a semi-quantitative measurement of the rate of evolution. It should be noted,

however, that our χ^2 values have 15 (16–1) degrees of freedom, while Shapiro's Y values have 19 degrees of freedom. Therefore, the comparison of our data to those of Shapiro is made only in a qualitative sense, since no correction was attempted concerning the additional 4 amino acids found in the proteins analyzed by Shapiro. Our analysis suggest that while the large subunits may have evolved at a rate no greater and probably slower than hemoglobin α chains and cytochrome c's, the small subunits have evolved at a rate some order of magnitude higher.

Discussion

Differences in the rate of evolution for various proteins have been suggested to be the consequence of differences in the constraints of natural selection (Dickerson, 1971). This argument may also be applied to the difference in rate of evolution of the two subunits of RuDPCase. If the two subunits appeared at the same time to build the first enzyme, and, if mutation of the genes coding for the subunits is at random, the larger subunits should mutate at a faster rate than the small subunits. In theory, only about one-half as many coding triplets are required in the genes coding for the small subunit as the number coding for the large subunit. However, should the large subunits play a more important role in the enzymatic function or structural stability of the enzyme, the enzyme would remain functional and the plant itself could survive only when alterations in the primary structure of the large subunits were not deleterious. Akazawa (1970) has evidence that the large subunits carry the active sites of the enzyme and we therefore suspect that most mutations in the large subunit are lethal because carbon dioxide fixation during photosynthesis is prevented and the plant containing the faulty enzyme cannot survive. Since the precise function of the small subunit is still unknown, we may assume that more drastic changes in its primary structure may not greatly affect the enzyme activity, and thus the changes can become incorporated into the protein structure without greatly reducing the viability of the plants containing the mutant. An analogy is the fibrinopeptides in humans and related animals which have been demonstrated to be rapidly evolving peptides that do not have important physiological functions (Doolittle, Wooding, Liu and Riley, 1971). Here, the reader is reminded that natural selection is acting on the phenotypes of organisms (Mayr, 1964). Therefore, natural selection may operate to conserve the amino acid sequences of the large subunit while permitting wider viability of mutations in the small subunit of RuDPCase. Alternatively, it is possible that the unusually high rate of evolution of the small subunit provides an exceptionally favorable survival value. The two subunits may have appeared at different times in evolution, the large subunit being the first to appear as a member of the primordial photosynthetic system. As the organisms which contained the system

evolved and developed into more complicated forms, the intracellular and extra-cellular environments for the enzyme might also have changed. It was at this point in evolution that small subunits somehow became associated with the primitive enzyme. Not being of crucial importance to enzyme activity, these small subunits could undergo mutations which could provide for essential changes in protein conformation required for survival of enzymatic activity in the changing environments. Therefore the small subunits and their higher rate of evolution in their primary structures may be an unusual example of "useful" mutations in the game of evolution. Further study of the primary structures of RuDPCase enzymes from a wide spectrum of photosynthetic organisms could shed more light on this hypothesis. Of particular interest would be further information on the non-photosynthetic, *Hydrogenomas* bacteria since their RuDPCase is reported to be composed of only one kind of subunit (Kuehn and McFadden, 1969).

Acknowledgements. This work was supported by a Research Grant AI00536 from the National Institutes of Health, US Public Health Service, and Contract AT (11-1)-34, Project 8, from the Division of Biology and Medicine, US Atomic Energy Commission.

The authors are indebted to Professors Paul D. Boyer and John Fessler for making the amino acid analyses reported here possible, and to Dr. Karen Yuen for allowing us to use a computer program to calculate the χ^2 values.

We are indebted to a reviewer of this paper for having pointed out to us the possible "... theoretical difference in the nature of natural selection with regard to chloroplast and nuclear genes." In his view, mutations in chloroplast proteins could be selected at the level of competition between chloroplasts in the cell as well as at the level of competition between plants. "The greater intensity of selection between chloroplasts might make the adoption of nearly neutral mutations less likely."

References

Akazawa, T.: Progress in Phytochemistry **2**, 107 (1970)
Chan, P. H., Wildman, S. G.: Biochim. Biophys. Acta **277**, 677 (1972)
Dickerson, R. E.: J. Mol. Evol. **1**, 26 (1971)
Doolittle, R. F., Wooding, F. L., Liu, Y., Riley, M.: J. Mol. Evol. **1**, 74 (1971)
Goodspeed, T. H.: The genus *Nicotiana*. Waltham, Mass.: Chronica Botanica 1954
Kawashima, N.: Plant and Cell Physiol. **10**, 31 (1969)
Kawashima, N., Kwok, S. Y., Wildman, S. G.: Biochim. Biophys. Acta **236**, 578 (1971)
Kawashima, N., Wildman, S. G.: Ann. Rev. Plant Physiol. **21**, 325 (1970)
Kawashima, N., Wildman, S. G.: Biochim. Biophys. Acta **262**, 42 (1972)
Kuehn, G. D., McFadden, B. A.: Biochem. **8**, 2403 (1969)
Mayr, E.: Fed. Proc. **23**, 1231 (1964)
Metzger, H., Shapiro, M. B., Mosimann, J. E., Vinton, J. E.: Nature **219**, 1166 (1968)
Rutner, A. C., Lane, M. D.: Biochem. Biophys. Res. Commun. **28**, 531 (1967)
Shapiro, H. M.: Biochim. Biophys. Acta **236**, 725 (1971)

Dr. S. G. Wildman
Department of Biology
Molecular Biology Institute
University of California
Los Angeles, California 90024, USA