Journée du 3 juillet 1969

Première séance

MUTATION ET PROCESSUS DE L'ÉVOLUTION

PRÉSIDENT D. GLASER

Ch. YANOFSKY

Protein structure and evolution
Discussions

E. MARGOLIASH and W.M. FITCH

The evolutionary information
content of protein amino acid sequences

J. MAYNARD SMITH

Population genetics and molecular evolution
Discussions

S.E. BRESLER

Physical and Chemical processes leading to a mutation

De la Physique théorique à la Biologie, C.N.R.S., 1971.

INTRODUCTION

D. GLASER

My name is Glaser and I have been asked to be chairman this morning because Professor Dulbecco has not been able to come to the meeting. Also Professor Bresler of Leningrad will not be here, but we are fortunate in having a contribution from Professor Eigen, who will speak to us after the interval. All the talks before this interval will be devoted to the main topics of this mornings meeting.

For the sake of the physicists it might be useful to make a few general remarks about developments in molecular biology that have made possible a really new confrontation of the classical theory of evolution. You will learn from the 3 talks that we will hear first this morning, that it is beginning to be possible to make a really quantitative examination of the theory of evolution, because it is possible to define evolutionary events at the molecular level in a way that makes one of them strictly comparable with another.

This provides a sound basis for building a quantitative theory in which the rate of evolution would be predicted by the rate of mutation together with specification of selection pressures and such properties of the population as migration and mating patterns.

I could make one remark that the number 10^{11} agrees roughly with the results of measurements of phenotypic mutation rates in bacteria which are in the neighbourhood of 10^{-8}. That's the probability of finding a phenotypic mutant which is an auxotroph, or has acquired drug resistance, per generation per bacterium. But only a small portion of all the base changes will be seen phenotypically. That's based on the fact that a typical cistron has about 1000 nucleotides in it, let's say. Now if you say that the phenotypically detectable mutations constitute only 1 % all the base changes there's a discrepancy of a factor 100 between the estimate you quoted and the final one. Another remark is that I think the mutation rate probably is not dominated by thermal effects, but more likely by errors in the action of polymerases. These inaccuracies are the result of selection for the structure of the polymerase and one can make a qualitative argument that the mutation rate is optimized, and that it may not be the object of evolution to produce polymerases which are the most accurate possible within the limitations of quantum machanics and of kT, but rather to pick one which is a compromise between accurency and a mutation rate which allows evolution. I don't know how to estimate what the accuracy limit of

a polymerase could be. The theoretical chemists have to do that for us some day.

A difficulty arises when one can't define the importance or calculate the probability of a particular step in evolution. Clearly the development of an eye is a much larger event than the change of skin pigment, for example.

The assignment of a quantitative measure to the size of an evolutionary step in gross biology is very difficult. When one can speak of a single base change at the nucleic acid level and can make the chemical statement that a large number of base changes are equally likely, perhaps all base changes under some conditions, then a single base change can be taken as a unit of evolutionary change and the number of such changes per century can be taken as an input to a quantitative theory. The papers that we will hear this morning will contain descriptions of measurements of rates of evolution defined at the molecular level together with explorations of mechanisms which can account for these evolutionary steps at the DNA level. These basic events will be correlated with phenotypic results of evolutionary events at the protein level and to some extent at the organismic level.

With that brief introduction to the physicists describing the significance of these developments in molecular kinetics and their application to evolution, I'd like to call on the first speaker, Professor Yanofsky of Stanford, who will speak on the " Protein structure and evolution ".

De la Physique théorique à la Biologie, C.N.R.S., 1971.

PROTEIN STRUCTURE AND EVOLUTION

CHARLES YANOFSKY

Department of Biological Sciences, Stanford University,
Stanford, California 94305

Present day genetic and biochemical techniques provide the means by which we can attempt to answer fundamental questions on the molecular evolution of functional proteins. The considerable knowledge acquired in recent years in studies of gene structure-protein structure relationships serves as the basis for the design of experiments which may reveal why a protein in a particular organism has a unique primary structure, and how that structure changes when the organism is subjected to the forces of evolution. In this article I would like to describe mutational studies we have performed which provide some insight into structure-function relationships in a specific protein. I will also discuss experiments which are directed towards achieving the ' evolution ' of a functional protein.

The tryptophan operon of *E. coli*

The gene cluster we have studied in our analyses of gene structure-protein structure relationships is the tryptophan operon of *Escherichia coli*. This operon

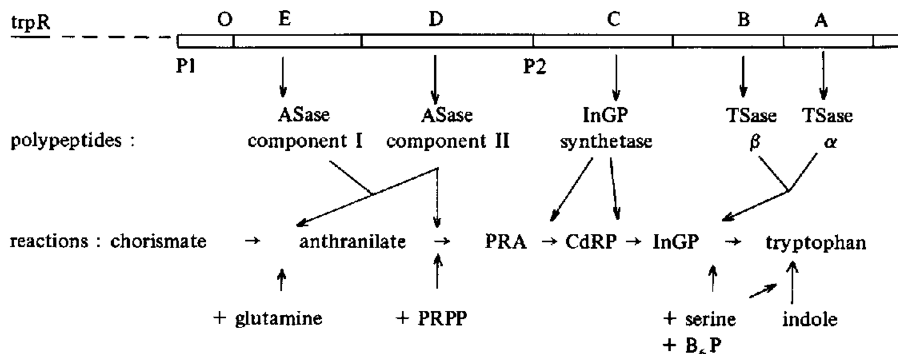

Fig. 1. The tryptophan operon of *E. coli*. The operon consists of 5 structural genes and adjacent controlling elements. TrpR is an unlinked gene specifying a protein repressor of the operon. O is the operator region and P1 and P2 are promoter regions. The various reactions in the pathway and the enzymes and enzyme complexes that serve as catalysts are indicated.

consists of five structural genes, each specifying a polypeptide which by itself or as a component of an enzyme complex catalyzes one or more of the terminal reactions in the biosynthesis of tryptophan (Fig. 1). Extensive mutational studies performed with this operon suggest that no segment of it is concerned with other essential bacterial functions. One implication of the existence of gene clusters of this type is that the component genes were derived from a common ancestral gene. To establish this point is of course one major objective of modern biology.

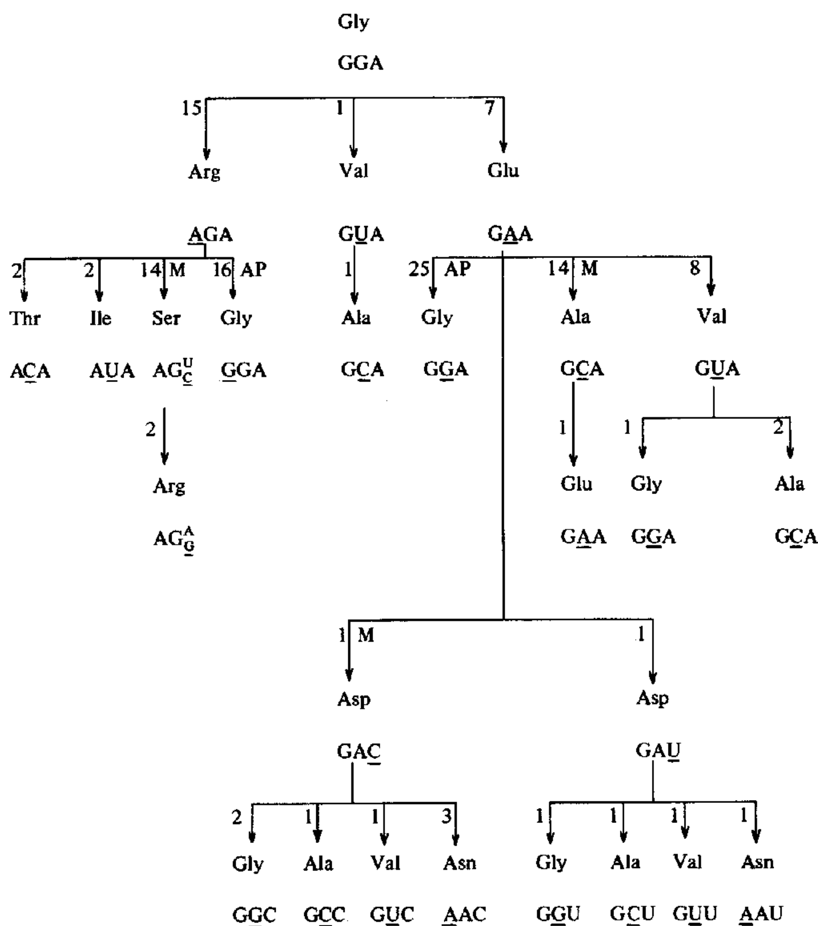

Fig. 2. Amino acid changes that have been observed at position 210 in the A protein and the probable corresponding codons [2, 3]. The number alongside each arrow indicates how many times the change was demonstrated by protein structure analyses. AP and M indicate changes favored by 2-aminopurine and the Treffers mutator gene, respectively. A bar under a codon letter identifies the nucleotide that is presumed to be introduced by the mutational change.

Tryptophan synthetase A protein alterations

Most of our gene structure-protein structure studies have been performed with the tryptophan synthetase A gene and A protein. The A protein is a single polypeptide chain 267 amino acid residues in length [1]; it has been shown to correspond linearly with the genetic map of the A gene [1]. Mutational changes in the A gene often lead to the production of altered A proteins which have single amino acid differences from the wild-type protein. At several positions in the A protein multiple amino acid changes have been detected. At position 210, for example, ten different amino acids have been inserted [2] (Fig. 2). Each of the observed amino acid substitutions is consistent with the interpretation that single mutational events involve single base-pair changes [2, 3]. Multiple amino acid substitutions have also been observed at positions 182 and 233 [3] (Fig. 3). It is clear from these cases and from comparable ones with other gene-protein systems that different amino acids can occupy a given position in a protein and permit function. We also know from amino acid sequence comparisons that enzymes isolated from related or unrelated species may have many sequence differences and nevertheless exhibit comparable enzymic activity. These observations focus on an important question : To what extent is the amino acid residue at each position in a protein essential for maximum effectiveness of that protein in its respective organism ? The same question phrased in terms familiar to the evolutionary biologists is : Are neutral mutational changes preserved during evolution ? In order to attack this problem experimentally we sought some means of rigorously assessing the

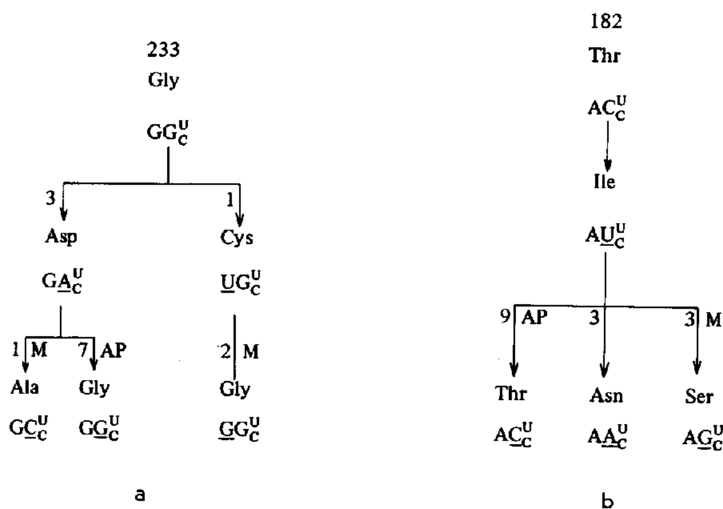

Fig. 3. Amino acid changes at position 182 (Fig. 3 A) and 233 (Fig. 3 B) and the probable corresponding codon changes [3]. See legend to Fig. 2 for other explanatory information.

functional capability of active A proteins with different amino acids at the same position. Many of the properties of the isolated altered A proteins have been examined but, since it is *in vivo* activity which is of concern to the organism, only tests performed in the growing cell could be considered relevant.

Perhaps the most sensitive means of examining the catalytic capability of an altered A protein in an organism such as *E. coli* is to determine the (tryptophan operon) enzyme levels attained when a culture is growing logarithmically in a minimal synthetic medium lacking tryptophan. Since under these conditions the organism must synthesize tryptophan to sustain its growth, a decrease in the catalytic capability of an enzyme should result in the production of elevated levels of all the enzymes specified by the operon; i.e., *E. coli* normally responds to a tryptophan deficiency by derepressing or turning on the synthesis of the enzymes of the pathway. Thus a very sensitive measure of true *in vivo* synthetic capacity is the effect on the levels of the biosynthetic enzymes. As can be seen in Table 1, when experiments of this type are per-

TABLE 1

Tryptophan synthetase B protein specific activities and generation times of strains with different amino acids at position 210 in the A protein.

Amino acid at position 210	Tryptophan synthetase B protein specific activity	Generation time (min)
Gly (wild type)	3.0 3.2 3.3	60, 60
Ala	3.2 3.3 3.5	60, 60
Ser	3.7 3.9 3.9	59, 60
Thr	7.7 8.0 8.2	58, 59
Val	30.1 31.3 31.9	69, 77
Ile	27.9 24.8 27.7	74, 74
Asn	50.9 51.4 55.1	104, 106

To ensure a constant genetic background each A gene was introduced by transduction into the same strain, a mutant with the A gene deleted. Several non-lysogenic colonies from each transduction were isolated and purified. Specific activities were determined with independent cultures harvested during log phase growth on minimal medium. Generation times were determined at 37 °C with cultures growing in minimal medium with glucose as carbon source. Estimates are based on cell population increases from 2×10^8 to 8×10^8/ml.

formed some of the proteins appear to be as active as the wild-type protein while others are probably less efficient and therefore signal the production of increased amounts of the biosynthetic enzymes. The same relationship is evident from the data presented in Table 2 for strains with altered proteins with amino acid changes at other positions in the A protein. Thus it is clear from this test that some amino acids are equally as effective as the wild-type amino acid. However, other amino acids at the same protein positions limit *in vivo* enzyme activity.

TABLE 2

Tryptophan synthetase B protein specific activities and generation times of strains with different amino acids at positions 182 and 233 in the A protein.

Amino acid at	Tryptophan synthetase B protein specific activity	Generation time (min)
position 182		
Thr (wild type)	2.6	58, 59
Ser	2.2	60, 60
Asn	9.1 9.3 10.8	58, 61
position 233		
Gly (wild type)	3.0 3.2 3.3	60, 60
Ala	11.3 12.9 13.1	61, 61

See legend to Table 1 for experimental conditions.

In view of these findings we might ask a related question: When an organism produces elevated enzyme levels does it do so at the expense of its ability to perform other metabolic reactions? For example, in the wild-type strain growing in minimal medium the tryptophan biosynthetic enzymes constitute 0.4 % of the soluble protein. If the organism were forced to increase this level to ca. 4 % to provide suffiicent tryptophan for maximal growth rates would it do so at the expense of its ability to perform other metabolic reactions, thereby limiting its growth rate? It is evident from the data in Tables 1 and 2 that significant increases in enzyme levels can be tolerated without any noticeable effect on the generation time. Thus, as can be seen in Table 1, when either serine or threonine occupies position 210 in the A protein the generation time is unaffected (Table 1). However, when the enzyme levels are in-

creased 10-fold (valine and isoleucine proteins) a significant lengthening of the generation time is evident. When still higher levels of enzyme are produced (asparagine protein) even longer generation times are observed. On the basis of the latter finding it seems likely that the enzyme levels and generation times in the valine and isoleucine strains represent the consequence of a *balance* between the rate of tryptophan synthesis and the effect of the formation of large amounts of these proteins on the growth rate of the organism. In Table 2 we also see that significant increases in specific activity are not correlated with appreciable changes in generation time. We might have expected to see such increases when enzyme levels are increased 3- to 4-fold. It should be pointed out, however, that studies performed in the manner described in Tables 1 and 2 are incapable of detecting minor changes in generation time. Despite this, we may tentatively conclude that different amino acids are equally acceptable at certain positions in the A protein and that moderate increases in enzyme levels can be tolerated without exerting a noticeable effect on the growth rate. Thus neutral or near-neutral mutational changes probably can occur-whether they are preserved is a much more difficult question to answer.

Compensating amino acid changes

In many missense mutants reversion events occur at second sites within the A gene as well as in the codon affected by the primary mutation [4, 5]. Several cases of second-site reversion have been analyzed and the findings obtained have revealed structural relationships within the folded protein molecule. For example, the change from glycine to glutamic acid at position 210 in the A protein is reversed by a change from tyrosine to cysteine at position 174 (Fig. 4). Interestingly, only the latter change reverses the effect of the presence of glutamic acid at position 210, i.e., mutational changes in other A gene codons cannot restore functional activity and only the change from tyrosine to cysteine at position 174 is effective. Similarly, mutant A187, an auxotroph with two amino acid differences from the wild-type protein, valines instead of glycines at positions 210 and 212, reverts at three positions,

Strain	Amino acids at corresponding positions			Activity of protein	Locations of genetic changes
	174-175-176		210-211-212		
wild type	-Tyr-Leu-Leu-	33 residues	-Gly-Phe-Gly	active	
A46	-Tyr-Leu-Leu-		-Glu-Phe-Gly	inactive	
A46PR8	-Cys-Leu-Leu-		-Glu-Phe-Gly	active	

Fig. 4. Second-site reversion of mutant A46 [4]. As indicated, a Tyr → Cys change at position 174 activates the protein with Glu at position 210.

Strain	Amino acids at corresponding positions		Activity of protein	Locations of genetic changes
	174-175-176	210-211-212		
wild type	Tyr-Leu-Leu- 33 residues	-Gly-Phe-Gly	active	
A46	Tyr-Leu-Leu-	-Glu-Phe-Gly	inactive	
A46PR9	Tyr-Leu-Leu-	-Val-Phe-Val	active	
A187	Tyr-Leu-Leu-	-Val-Phe-Val	inactive	
A187SPR4	Tyr-Leu-Leu-	-Val-Phe-Gly	active	
A187SPR3	Tyr-Leu-Leu-	-Val-Phe-Ala	active	
A187SPR5	Tyr-Leu-Leu-	-Gly-Phe-Val	active	
A187SPR2	Tyr-Leu-Leu-	-Ala-Phe-Val	active	
A187SPR1	Tyr-Leu-Arg-	-Val-Phe-Val	active	

Fig. 5. Second-site reversion of mutant A 187 [5]. The A 187 protein has two changes; the Gly residues at positions 210 and 212 are replaced by Val residues. When either Val is replaced by Gly or Ala, the protein is functional. Both valines are retained in a functional protein in which the Leu residue at position 176 is replaced by Arg.

210, 212 and 176. At positions 210 and 212 the replacement of valine by either glycine or alanine restores activity, demonstrating that the A 187 protein is inactive only because both valines are present. One further point of interest is that the position of the distal reversion change, at 176, is two residues from the position of the second-site reversion in mutant A 46. These observations suggest that the two regions of the polypeptide chain indicated in Figs. 4 and 5 interact in the native molecule. We may conclude from these studies that because of the spatial relationships in the folded molecule the effects of an amino acid change in one region of the molecule can only be overcome by distal changes by specific alterations in an interacting region.

These observations raise the possibility that a neutral mutational change at one site may permit a subsequent change to confer a selective advantage. This gain in functional acceptability would then preserve what originally was a neutral event.

Attemps to "evolve" a functional A protein in strains lacking a segment of the A gene

I would like to known whether it is possible to produce a functional A protein by mutationally altering a protein fragment lacking the 20 or so amino acid residues at the carboxyl end of the molecule. To determine this, deletion mutants lacking the end of the A gene were subjected to mutagenic treatments and the treated populations were added to a medium which would only sustain the growth of cells with a functional A protein. The deletion

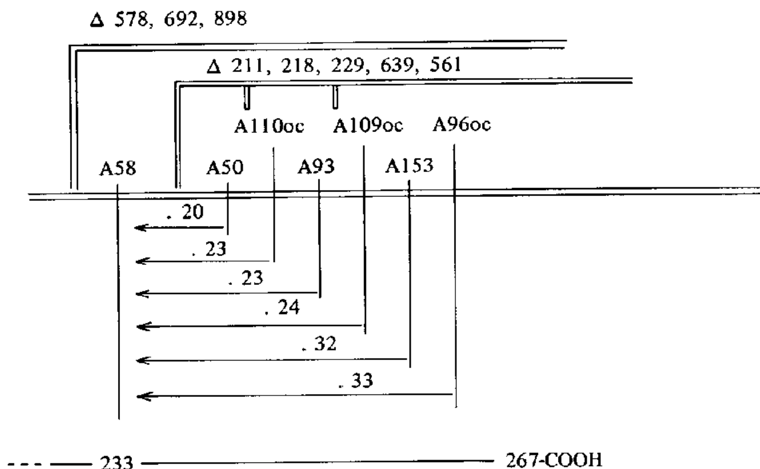

Fig. 6. Mutant sites and deletion termini in the region of the A gene specifying the carboxyl end of the A protein. Map distances are indicated above the arrows. Three of the point mutants (A 110, A 109, A 96) are ochre nonsense mutants. The precise terminus of each deletion is not known but it ends in the region indicated.

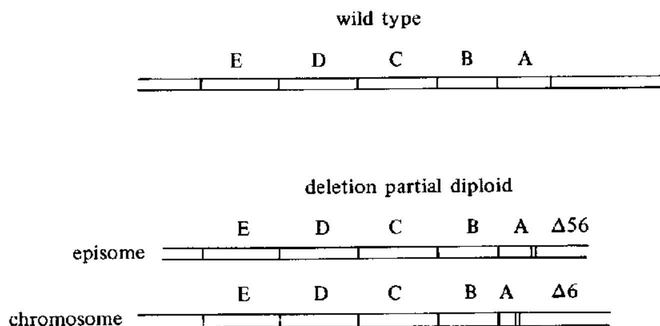

Fig. 7. A diploid strain employed in a prototrophy selection experiment. Deletion 6 removes most of the A gene while deletion 56 removes that region specifying the last thirty or so residues.

mutants examined are listed in Figs. 6 and 7. They were heavily mutagenized with nitrosoguanidine and ICR 191, powerful mutagens which produce base substitutions and base deletions and additions, respectively. Selective conditions were employed which would have permitted the outgrowth of bacteria with a functional A protein. In none of the deletions listed was a functional A protein detected — this despite the fact that large populations of bacteria were employed and the bacteria were permitted to divide several times before the

tryptophan supplement in the growth medium was depleted. These negative results suggest that the carboxyl-terminal end of the A protein is essential for enzyme activity and that alterations in the remaining portion of the molecule cannot compensate for the loss. This conclusion is supported by the fact that nonsense codons in the terminal region of the A gene result in enzyme inactivity (see Fig. 6). Diploid strains with terminal deletions were also examined in these studies to eliminate the possibility that inactivating mutations occurred in other genes of the operon concomitantly with mutations in the A gene. Furthermore, episomes with A gene terminal deletions were transferred out of a heavily mutagenized population into haploid cells lacking only the A gene. In every case except one, to be described below, we did not detect a functional A protein. Despite these negative results other findings to be described subsequently suggest that the amino acid sequence at the carboxyl terminus of the wild-type A protein is not the only sequence that will permit this protein to be catalytically active. In the diploid strain described in Fig. 7, in which A gene deletions were present on both chromosome and episome, an active A protein was formed as a consequence of mutational changes in the A gene segment. The prototrophic strain obtained grows very poorly without tryptophan, however, suggesting that the functional A protein that is produced is at best inefficient. To eliminate the possibility that in the diploid strain one of the other genes of the tryptophan operon was assuming the function of the A gene, the *trp* operon of the episome was introduced by transduction into a haploid strain, replacing the operon of the recipient; i.e., the transductants had only one copy of each of the genes of the operon. These transductants were slow-growing prototrophs, suggesting that the mutation or mutations responsible for A protein activity were in or near the A gene.

In related studies we attempted to modify mutationally the E, D, C or B gene in a diploid strain so that the altered protein it produced could function as an A protein. The partial diploid prepared for the experiment had 90 % of the A gene deleted on both chromosome and episome. To date, these experiments have also given negative results, suggesting that each of the operon

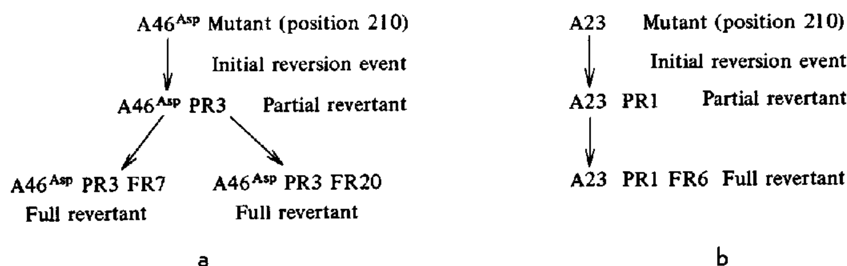

Fig. 8. Sequential reversion events [2] starting with mutant A46^{Asp} (aspartic acid at position 210), A, and mutant A23 (arginine at position 210), B.

proteins is considerably different from its ancestral protein and thus many amino acid changes would be required for it to acquire A protein activity.

Frameshift mutations and protein function

Mutant A46^{Asp} and A23 both yield slow-growing partial revertant strains in which prototrophy is due to second-site mutations [2]. In order to analyze the effect of these mutations on the structure of the A protein, faster-growing full revertants were selected from the partial revertant strains (Fig. 8). This extra step was necessary because the partial revertant A proteins were extremely labile and could not be isolated. When the full revertant A proteins were analyzed [2] we were surprised to find that several contiguous amino acids had changed in each (Fig. 9). The amino acid differences in each revertant could be explained by assuming that the primary mutational event resulted in a single base addition and the second mutation involved a single base deletion. The greater activity of the full revertant A proteins compared to the partial revertant proteins is readily understandable since, with the exception of the residues in the vicinity of position 210, the amino acid sequences would be unaltered. The activity of the partial revertant A proteins is surprising, particularly in view of the conclusions reached in the previous section. We would expect that in these strains (A46^{Asp} PR3 and A23 PR1) the entire sequence of the terminal portion of the A protein would be altered as a result of the single

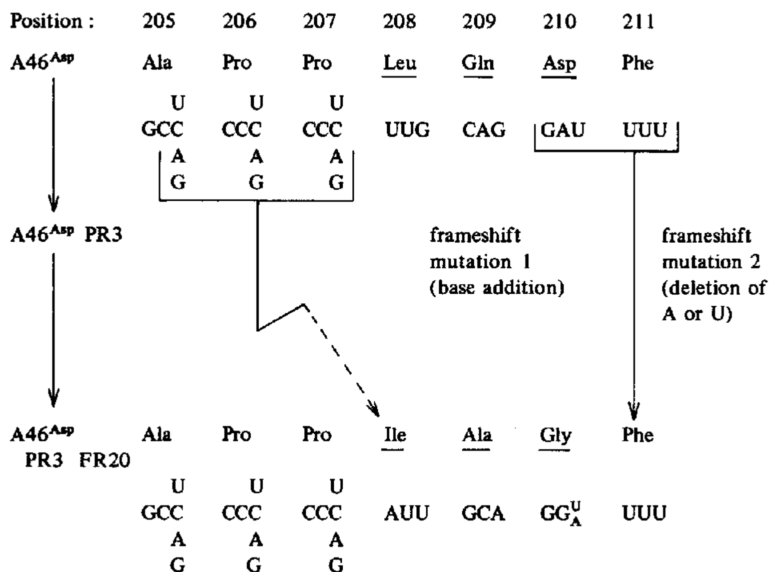

Fig. 9 a

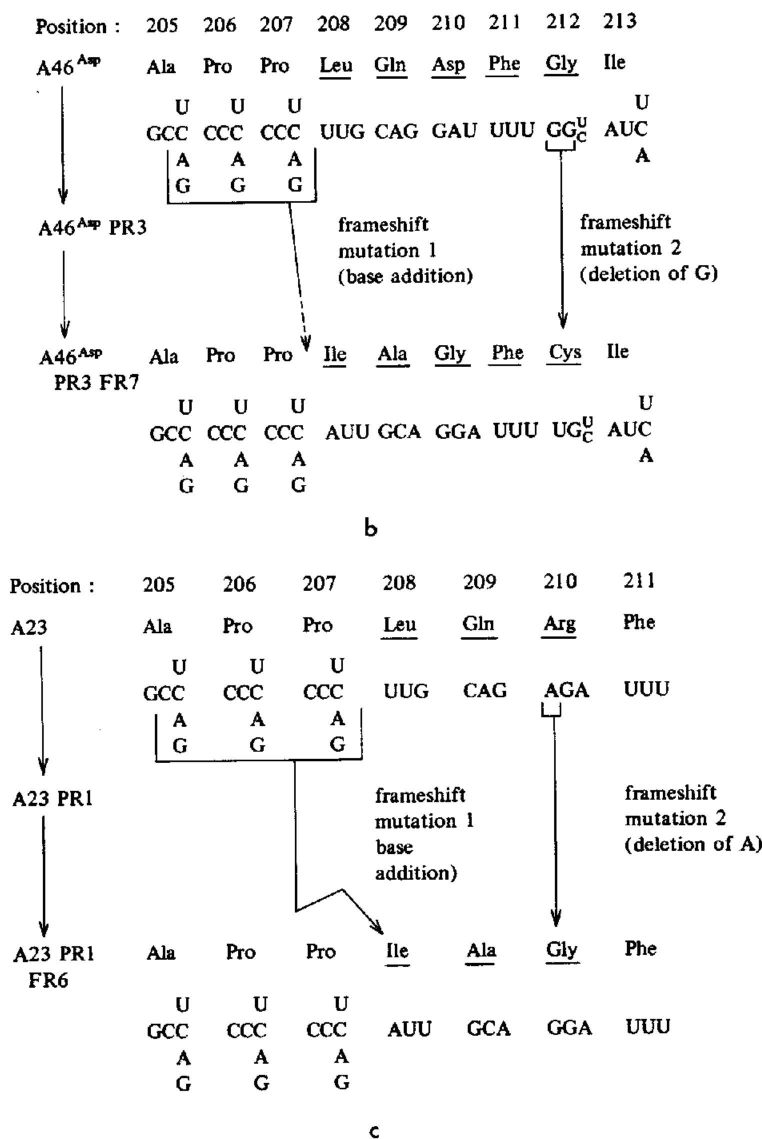

Fig. 9. Amino acid changes and probable corresponding nucleotide sequences in frameshift revertants A46^{Asp} PR3 FR20 (A), A46^{Asp} PR3 FR7 (B) and A23 PR1 FR6 (C) [2].

base addition. The new sequences that would be generated in the vicinity of position 210 are shown in Fig. 10. In strain A46^{Asp} PR3 a glycine residue would presumably replace the aspartic acid residue which is present at position 210 in A46^{Asp} and is responsible for enzyme *inactivity*. In A23 PR1, however, the arginine residue at position 210 would be replaced by a charged

	207	208	209	210	211	212
wild type sequence	Pro –	Leu –	Gln –	Gly –	Phe –	Gly
A46^{Asp}	Pro –	Leu –	Gln –	Asp –	Phe –	Gly
	CCA	UUG	CAG	GAU	UUU	GGU
	+ 1					
A46^{Asp} PR3	CC·	AUU	GCA	GGA	UUU	UGG
	Pro –	Ile –	Ala –	Gly –	Phe –	Trp
A23	Pro –	Leu –	Gln –	Arg –	Phe –	Gly
	CCA	UUG	CAG	AGA	UUU	GGU
	+ 1					
A23 PR1	CC·	AUU	GCA	GAG	AUU	UGC
	Pro –	Ile –	Ala –	Glu –	Ile –	Trp

Fig. 10. Hypothetical amino acid and nucleotide sequences in partial revertants A46^{Asp} PR3 and A23 PR1. It is assumed that the reading frame would remain shifted and the polypeptide would terminate at the first " new phase " terminator.

amino acid, glutamic acid. Thus it is not obvious from these hypothetical sequences why the partial revertant proteins are active, especially since we would expect that the entire carboxyl sequence starting at position 208 would be different from that of the wild-type protein. We do not know the length of the partial revertant proteins; as many as 14 terminator codons could be introduced as a result of the frameshift in these strains [2]. Since this number is large it seems likely that the protein is less than 267 residues in length in the partial revertant strains. The most reasonable explanation for these findings is that the new sequence that is generated as a consequence of the frameshift can perform the function peculiar to the carboxyl end of the normal protein.

If we compare the hypothetical wild-type amino acid sequence which would result from the initial base addition with the presumed sequences in PR3 and PR1 (Fig. 11), we see that the sequences are quite similar — in fact, the only difference between the PR3 and wild-type sequences is at posi-

	207	208	209	210	211	212
wild type	Pro –	Leu –	Gln –	Gly –	Phe –	Gly
	CCA	UUG	CAG	GGA	UUU	GGU
	+ 1					
	CC·	AUU	GCA	GGG	AUU	UGG
hypothetical sequence	Pro –	Ile –	Ala –	Gly –	Ile –	Trp
A46^{Asp} PR3 sequence	Pro –	Ile –	Ala –	Gly –	Phe –	Trp
A23 PR1	Pro –	Ile –	Ala –	Glu –	Ile –	Trp

Fig. 11. Hypothetical amino acid and nucleotide sequences if the frameshift occurred in the wild-type strain.

tion 211. Here, different hydrophobic amino acids are present — isoleucine and phenylalanine. In view of these comparisons it is perhaps surprising that functional A proteins were not detected in the mutagenesis studies with the deletion mutants. This may indicate that in A46^{Asp} PR3 and A23 PR1 the polypeptide chain has a near-normal length.

Comparative studies with the A proteins of *Escherichia coli, Salmonella typhimurium* and *Aerobacter aerogenes*

We are presently determining the amino acid sequences of the tryptophan synthetase A proteins from *Salmonella typhimurium* and *Aerobacter aerogenes* so that they may be compared with the sequence from *E. coli*. Our principal reason for performing these studies is based on the different GC contents of the DNA's of these organisms. *E. coli* DNA has approximately 50 % GC base pairs, *S. typhimurium* DNA about 51 % GC base pairs, while *A. aerogenes* DNA contains approximately 56-57 % GC base pairs [6]. At the present time about two-thirds of the *S. typhimurium* sequence and one-half of the *A. aerogenes* sequence are known. We estimate that there are ca. 10-15 % amino acid differences when we compare either sequence with that of *E. coli* [7]. The differences seem to be randomly distributed throughout the proteins and the *Salmonella* differences and *Aerobacter* differences are not at identical positions [8]. Thus it is not possible on the basis of these data to establish the evolutionary order of the three bacterial species. Most of the amino acid differences can be explained by a single base change per codon; thus the present structures probably differ from an ancestral molecule by no more than a single base change per codon. When we deduce the probable base change responsible for each amino acid change we find that the evolution of an *E. coli*-type protein to a *S. typhimurium*-type protein involved seven A or T → G or C changes and eight G or C → A or T changes. Of the seven identified AT changes, three were AT → GC and four were AT → CG. If we consider the differences proceeding from an *E. coli* protein to an *A. aerogenes* protein, there are nine A or T → G or C changes and only five G or C → A or T changes. This distribution is, of course, consistent with the higher GC content of *A. aerogenes* DNA. Furthermore, of the nine AT changes, two involved AT → GC and seven AT → CG. It appears, therefore, that the higher GC content of *A. aerogenes* DNA may be due to a preferential increase in the proportion of mutations from AT → CG. This conclusion, if substantiated by further studies, would be particularly interesting in view of our findings with the mutator gene of *E. coli* discovered by Treffers [9]. This mutator gene preferentially increases the base-pair change AT → CG [10, 11]. The presence of such a mutator gene at some period in the evolution of *A. aerogenes* would explain its high GC content and the apparent increase in the proportion of

AT → CG changes. In this connection, some recent work of Drake [12] should be cited. Drake has shown that certain mutations in the gene of phage T4 which specifies its DNA polymerase result in a reduction of the spontaneous mutation frequency. This finding suggests that at least in this organism DNA polymerase mistakes are responsible for a significant portion of spontaneous mutations. A bacterium with a highly active mutator gene may be being subjected to an exaggeration of a specific mistake-making mechanism.

The randomness of the positions of amino acid differences in the various A proteins examined and the fact that most of the amino acid differences can be explained by a single base change per codon, in my opinion support the conclusion that neutral mutations do occur and are preserved during evolution. More extensive and convincing data on this point have been discussed by King and Jukes [13] and by Margoliash (see article in this volume).

Acknowledgment

The author is indebted to Miriam Bonner, Virginia Horn and Susan Stasiowski for their excellent assistance with the studies described in this paper. These investigations were supported by grants from the National Science Foundation and the United States Public Health Service.

References

[1] C. Yanofsky, G.R. Drapeau, J.R. Guest and B.C. Carlton, The complete amino acid sequence of the tryptophan synthetase A protein (α subunit) and its colinear relationship with the genetic map of the A gene. *Proc. Natl. Acad. Sci. U.S.*, **57**, 296-298 (1967).

[2] H. Berger, W.J. Brammar and C. Yanofsky, Analysis of amino acid repacements resulting from frameshift and missense mutations in the tryptophan synthetase A gene of *Escherichia coli. J. Mol. iBol.*, **34**, 219-238 (1968).

[3] C. Yanofsky, H. Berger and W.J. Brammar, *In vivo* studies on the genetic code. *Proceedings of the XII International Congress of Genetics*, **3**, 155-165 (1969).

[4] D.R. Helinski and C. Yanofsky, A genetic and biochemical analysis of second site reversion. *J. Biol. Chem.*, **238**, 1043-1048 (1963).

[5] C. Yanofsky, V. Horn and D. Thorpe, Protein structure relationships revealed by mutational analysis. *Science*, **146**, 1593-1594 (1964).

[6] N. Sueoka. Compositional correlation between deoxyribonucleic acid and protein. *Cold Spring Harbor Symposium on Quantitative Biology*, **26**, 35-43 (1961).

[7] T.E. Creighton, G. Johnson and C. Yanofsky, unpublished observations.

[8] T.E. Creighton, D.R. Helinski, R.L. Somerville and C. Yanofsky, Comparison of the tryptophan synthetase α subunits of several species of Enterobacteriaceae *J. Bact.*, **91**, 1819-1826 (1966).