MUTATIONS IN PROTEINS AND BASE CHANGES IN CODONS

by

Thomas H. Jukes

Space Sciences Laboratory
University of California, Berkeley, Calif. 94720

Received June 30,1975

SUMMARY

About 300 single-amino-acid mutations in proteins were compared with the corresponding changes in the amino acid code. Transitional base changes are more common than transversions. Among transitions AG interchanges are more common than CU interchanges and AU interchanges are the most infrequent of transversions. Of the 17 undiscovered single-amino-acid mutations, 15 are transversions that do not produce a change in charge.

Many mutant proteins have been found to differ from their normal or "wild-type" counterparts by replacement of a single amino acid. The first of these to be discovered was the replacement of glutamic acid by valine in the β chain of hemoglobin S (Ingram, 1957) and, with only two exceptions, all such replacements are attributable to the change of a single base in the messenger RNA codon for the altered amino acid (Smith, 1962; Jukes, 1962). The substitution of the base takes place in DNA and is transcribed into messenger RNA which, in turn, is translated into the mutant protein. About 300 such mutations have been identified and it is worth examining them in terms of the amino acid code.

The base changes A ⇄ G and U ⇄ C, are termed transitions and A ⇄ C, A ⇄ U, G ⇄ C and G ⇄ U are termed transversions. (Freese, 1959). There are 75 single-base changes that produce amino acid changes in the amino acid code of 61 codons. (Table 1). Each such replacement may take place in either of two directions. Most of the changes in Table 1 are

---

Abbreviations for bases: Y = U or C; R = A or G; N = U, C, A, or G; for amino acids F = Phe, L = Leu, I = Ile; M = Met; V = Val; S = Ser; P = Pro; T = Thr; A = Ala; Y = Tyr; CT = chain termination; H = His; Q = Gln; N = Asn; K = Lys; D = Asp; E = Glu; C = Cys; W = Trp; R = Arg; G = Gly.

Table 1: Summary of amino acid changes in mutant proteins

Base Interchange in mRNA	Effect on Charge of Protein	Amino Acid Interchanges and kumbers Reported
A G	Change	DG-16(5); DN-13(1); EG-4(4); EK-24; RG-9(6)*; RH-9; RQ-3(1).
	No Change	AT-(2); CY-1(2); GS-(1); IV-(4); MV-3(1); NS-(4); IM(1)*; RK-(1).
U C	Change	RW-2(1)*; RC-0.
	No Change	LP-8(2); HY-6; FS-3(2); AV-1(5); IT-1(9); PS(1); FL-2(2)*; LS-(3); MT-(1).
A C	Change	AD-10(2); AE-7(2); KT-6(1); KQ-3.
	No Change	HP-1; NT-1; PT-(1); HN-0; PQ-0; IL-0; SY-0.
G U	Change	LR-7(1); DY-5; IR-(1); RM-(1).
	No Change	CG-2(2); IM-(1)*; GV-3(1); FV-2; IS-(1); GW-1; LV-0; AS-0; LW-0.
A U	Change	EV-7(2); DV-2(1); RW-2(1)*; KM-0.
	No Change	FY-2; LQ-1; IF-0; IK-0; IL-0; HL-0; IN-0; NY-0.
G C	Change	RG-0(6)*; DH-8; EQ-7; RP-6; RT-(1).
	No Change	AP-3; AG-1; CS-1; SW-1; IM-(1)*.
Ambiguous	Change	NK-14(2); RS-4(3).
Transversions	No Change	HQ-4(1); DE-(2); FL-2(2)*; CW-(1); LM-(1); ST-0.

Parentheses = other than globins
* = more than one base change can produce the same amino acid interchange

	Kinds	Numbers
Charge changes:	Possible 29 Found 27	Total 202 7.2 times per base change
No Charge changes:	Possible 51 Found 37	Total 98 2.6 times per base change

attributable to unambiguous base changes. However, there are some exceptions to this because of the degeneracy of the code. These are as follows:

(a) Interchanges between the following amino-acid pairs are caused by third-base changes of either of the two pyrimidines (U and C) to either

of the two purines (A and G): His ⇄ Gln, Asn ⇄ Lys, Asp ⇄ Glu, Ser ⇄ Arg.

(b) Interchanges between Cys and Trp are produced by U or C ⇄ G; between Ser and Thr by C ⇄ A or G ⇄ C; between Leu and Met by C or U ⇄ A. We have therefore listed (a) and (b) as "ambiguous transversions".

(c) Mutational interchanges between Phe and Leu may be produced by UUY ⇄ UUR or UUY ⇄ CUY.

(d) Mutational interchanges between Ile and Met may be produced by U, C or A ⇄ G; between Arg and Trp by C or A ⇄ U and between Arg and Gly by C or A ⇄ G. We have therefore listed the changes in (c) and (d) in more than one category, since they may be transitions or transversions (Table 1).

Most of the mutant proteins are globins, and the information on them was obtained from summaries by Lehmann (1975) and Lehmann and Huntsman (1974), who have compiled 211 single-amino-acid hemoglobin variants and four myoglobin variants. Other mutant proteins were compiled by Jukes (Jukes and Gatlin, 1971). The first point examined was whether base changes occur with equal frequency in both directions in the identified mutants. This is summarized in Table 2, which shows that this is usually the case, except for mutants induced by nitrous acid. The results also show that the GA change predominates over AG in globins. This was previously noted for a much smaller number of hemoglobin mutants and for evolutionary changes in cytochromes c, by Fitch (1967). Evolutionary changes may include intermediate steps that are undetectable. Lehmann and Carrell (1969) found about equal numbers of transitions and transversions had given rise to 95 hemoglobin variants, including 28 GA mutants. They suggested that a special mechanism might be involved in G to A transitions in man. In our summary of mutations in proteins other than globins, AG predominates over GA, and CU over UC. This imbalance is attributable to the effects of the deaminative chemical mutagen nitrous acid in the reported findings in tobacco mosaic virus coat protein. Deamination changes adenine to hypoxan-

Table 2: DNA base-pair changes as related to single-amino-acid replacements.

DNA Base-Pair Change	Base Change in mRNA	Number found in Globins	Number found in Other Proteins	Total
A·T ⇄ G·C	AG	26	21(10)	47(36)
	GA	37	5(3)	42(39)
	UC	15	5	20
	CU	8	18(9)	26(17)
A·T ⇄ C·G	AC	14	3	17
	CA	14	3	17
	UG	8	3	11
	GU	12	4	16
G·C ⇄ C·G	GC	14	1	15
	CG	13	-	13
A·T ⇄ T·A	AU	7	2	9
	UA	7	1	8
Purine → Pyrimidine		8	5	13
Pyrimidine → Purine		14	5	19

Transitions, 135; transversions, 138; uncharacterizable: Gly → Arg 10; Arg → Gly 5; Phe → Leu 3; Leu → Phe 1; Trp → Arg 1; Arg → Trp 1; Ile → Met 1.

Figures in parentheses show totals obtained after subtracting nitrous-acid-induced mutations in tobacco mosaic virus coat protein.

---

thine (which like guanine, pairs with cytosine), and changes cytosine to uracil. The mutations produced by nitrous acid in tobacco mosaic virus coat protein are as follows (AG changes result presumably from adenine to hypoxanthine):

Change	Locations in sequence of coat protein	Corresponding base change
Arg to Gly	46	AG
Arg to Lys	46	GA
Asn to Ser	25,33,73,126	AG
Asp to Gly	66	AG

Gln to Arg	99	AG
Glu to Gly	97	AG
Ile to Val	21,24,126,129	AG
Pro to Leu	20,156	CU
Pro to Ser	63	CU
Ser to Leu	55	CU
Ser to Phe	138,148	CU
Thr to Ile	5,59	CU
Thr to Met	107	CU
Thr to Ala	81	AG
Val to Met	11	GA

The next finding observable in Table 2 is that transitions predominate. These can take place either by A·T ⇌ G·T ⇌ G·C; or A·T ⇌ A·C ⇌ G·C. The pair G·T (or G·U) occurs with very little disruption of the double-helical structure of DNA (or RNA). This may be the explanation for the preponderance of transitions, for it is conceivable that "error-correcting mechanisms" in DNA would be less likely to remove intermediate G·T pairs than less compatible pairs such as A·A and T·T. This question has been extensively discussed by Fresco and Kopal (In preparation, 1975). In support of this, we have noted that pairing between U and G far exceeds all other "mis-pairing" on the helical regions of tRNA molecules (Holmquist, *et. al.*, 1973). A more recent summary of such mispairing in the helices of 54 tRNA molecules was made with the base pair adjoining the dihydrouridine loop omitted, because this pair does not consistently contribute to the helicity of this stem, i.e., it is not a "regular" hydrogen-bonded pair. The helical regions of the 54 molecules contained 52 pairs between G and U, 4 between U and U, 3 between C and A and no other mis-pairs. This indicates that the intermediate pair G·U is more probable than A·C in terms of minimizing the disruption of helicity. A ⇌ G interchanges far exceed C ⇌ U interchanges in the globins. Both types of interchanges are produced by the same change in a base pair. Since mutations can originate in either of the two DNA strands, it is not possible to attribute the difference to any preferential effect without further information. The amino acid composition of hemoglobins is such as to favor A + G slightly (about 11%) over C + T in the first two codon positions. This is not

enough to account for the disparity; and perhaps the predominance of A ⇄ G interchanges in hemoglobin variants is attributable to the ease with which Lys ⇄ Glu, Asp ⇄ Gly and Asp ⇄ Asn replacements are detected.

The disparity between the various classes of transversions is indeed striking. If all six possible interchanges of bases in DNA were equiprobable, one would expect the changes produced by A·T ⇄ C·G interconversions (total 61) to be equal to G·C ⇄ C·G (total 28) and A·T ⇄ T·A (total 17).

Most of the mutant hemoglobins and many of the other mutants were detected by electrophoresis near pH 8, and such mutations result from a change in electrostatic charge, produced by replacements of the four amino acidsArg, Asp, Glu and Lys. The commonest interchange found in globins is between Glu and Lys, which produces a "double change" in charge, between an acidic and a basic amino acid. This has twice as much electrophoretic effect as a change between a neutral amino acid and a basic or acidic amino acid. We therefore grouped the base substitutions with respect to their effects on charge, in Tables 1 and 3. Most of the replacements that produce a change in charge have been discovered. However, nearly all the undiscovered mutational replacements are transversions, no doubt due in part to the fact that six transitions have been discovered only with the aid of nitrous acid. The interchange between Ser and Thr, which can be produced by single-base changes between any of the possible 24 pairs of codons for Ser and Thr, has not been detected in a mutant protein. The summaries at the foot of Table 1 are adjusted for alternative possibilities. For example, there are two different base substitutions (C ⇄ G and A ⇄ G) that produce interchanges between Arg and Gly, and these 15 observed interchanges between Arg and Gly are counted as 15, rather than 30, in the totals of all mutations. The effect of charge-change on the number of mutations reported is quite marked, more than twice as many such changes have been reported as changes that produce no charge-change.

Table 3: Single-base mutations in the codons for amino acids.

	Discovered	Undiscovered
Transitions		
Changing charge	10	1
Not changing charge	17	0
Totals	27	1
Transversions		
Changing charge	16	1
Not changing charge	20	15
Totals	36	16

The relation, if any, of our findings to evolutionary changes in proteins is not evident. The mutant proteins we have listed were discovered in the majority of cases by electrophoretic comparisons with the wild types. In our compilation, transitions were more frequent than transversions but in a comparison of evolutionary changes between globins, 293 transitions and 548 transversions were counted (Jukes, 1965). This might indicate that transitions tend to be discarded more readily than transversions in the evolution of hemoglobin.

ACKNOWLEDGEMENT

I thank Richard Holmquist for useful suggestions, and Professors Jacques Fresco and Hermann Lehmann for communicating unpublished results.

Support by NASA Grant NGR 05-003-460 to the University of California, Berkeley, is acknowledged.

REFERENCES

Fitch, W.M. (1967) J. Mol. Biol, 26:499-507.
Freese, E. (1959) Proc. Nat. Acad. Sci., 45:622-633.
Holmquist, W. R., Jukes, T. H. and Pangburn, S. (1973) J. Mol. Biol. 78:91-116.
Ingram, V. (1957) Nature 180:326-327.
Jukes, T. H. (1962) Proc. Nat. Acad. Sci., 48:1809-1815.

Jukes, T. H. (1965) American Scientist, 53:477-487.
Jukes, T. H. and Gatlin, L (1971) Progress in Nucleic Acid Research, 11:303-350.
Lehmann, H. (1975) List of known haemoglobin substitutions and deletions (manuscript).
Lehmann, H. and Carrell, R. W. (1969) British Med. Bull., 25:14-23.
Lehmann, H. and Huntsmann, R. G. (1974) Man's Haemoglobins, pp. 188-204, North-Holland, Amsterdam.
Smith, E. L. (1962) Proc. Nat. Acad. Sci., 48:859-864.