Table 1 Replacement Data for Sixty-eight Arginine Residues in Vertebrate Proteins

Protein	Substituted amino acid His	Gln	Pro	Lys	Thr	Met	Ser	Ala	Asp	Glu	Pres
Cytochrome *c*											2
Haemoglobin α		1	1	1							0
Haemoglobin β	2	1		2				1			2
Myoglobin				2							2
Fibrinopeptides											2
Chymotrypsinogen				6	1					1	0
Pepsin		1							1		2
Carboxypeptidase		2		1							5
GPDH	1			1		1					8
Ribonuclease	1	2		1			1				2
Lysozyme											4
Carbonic anhydrase				1							3
Acidic trypsin inhibitor				1							2
Total	4	7	1	16	1	1	1	1	1	1	34
Assignment	CGX			AGR			?				

The families of proteins and the alignments used to identify substitutions are taken from ref. 8, and the data have been restricted to include only reasonably closely related sequences, to minimize the possibility of multiple base changes between codons for corresponding residues. The table notes the number of cases in which a given residue substitutes for arginine in each protein. The right-hand column records the number of arginine residues in each protein which are "preserved" (when no substitution has occurred in any member of the homology family).

In spite of these uncertainties there are sufficient data to say that the proportion of arginine codons in vertebrate mRNA which contain CpG is about one-half. If this is so, and a value of 4% is taken for the average frequency of arginine in vertebrate proteins[4], then the contribution of CGX arginine codons to the CpG frequency in polypeptide-specifying sequences in vertebrate DNA should be approximately 0.7%. Experimentally determined values for the CpG frequency in vertebrate DNAs[1-3] range from 0.8%–1.7%, and there are several reasons for supposing that this frequency is also typical of the polypeptide-specifying parts of these DNAs[6], the most striking of which is the finding of almost identical doublet frequencies to those of vertebrates in the nucleic acids of small viruses[3,11]. The evidence thus suggests that the bulk of CpG occurrence is in the CGX codon position, and that the doublet is rare in the XCG or XXC.GXX positions. As the expected CpG frequency contributed by the XCG position, assuming equal usage of all synonymous codons, is about 1.9%, it is clear that this position must contain substantially less CpG than expected even if (as is indicated by the dipeptide frequency data[6]) there is little or no CpG in the XXC.GXX position.

Differential selection pressure is the most plausible mechanism for maintaining the CpG shortage in vertebrate DNA because, in the absence of very specific differential mutations, a rapid upward drift in the frequency of this doublet would otherwise have been expected. This would be inconsistent with the maintenance of uniformly low levels of CpG in vertebrate classes of distant common ancestry. In these circumstances, why should the CGX arginine codons have apparently largely escaped the depletion in frequency characteristic of XCG and XXC.GXX codons? The answer may lie in the fact that only in the case of arginine do CpG-containing codons constitute a major proportion of all the codons for a particular amino acid. Perhaps they cannot all be removed without a drastic effect on protein function. It is also the case that CGY (Y = pyrimidine) codons are unique in that there is no single-step mutation from them to synonymous codons not containing CpG. It is surprising that they do not occur with noticeably higher frequency than CGR codons, which have the opportunity to lose their CpG by the mutation to AGR without an amino acid change. Unambiguous identification of a larger proportion of the arginine codons in known proteins is required before this question can be resolved satisfactorily.

I thank Professor J. H. Subak-Sharpe for helpful discussions.

R. A. ELTON

MRC Virology Unit,
Institute of Virology,
University of Glasgow,
Glasgow G11 5JR

Received January 11; revised April 13, 1973.

[1] Josse, J., Kaiser, A. D., and Kornberg, A., *J. Biol. Chem.*, **236**, 864 (1961).
[2] Swartz, M. N., Trautner, T. A., and Kornberg, A., *J. Biol. Chem.*, **237**, 1961 (1962).
[3] Subak-Sharpe, H., Bürk, R. R., Crawford, L. V., Morrison, J. M., Hay, J., and Keir, H. M., *Cold Spring Harbor Symp. Quant. Biol.*, **31**, 737 (1966).
[4] King, J. L., and Jukes, T. H., *Science*, **164**, 788 (1969).
[5] Jukes, T. H., *Curr. Top. Microbiol. Immunol.*, **49**, 178 (1970).
[6] Bullock, E., and Elton, R. A., *J. Mol. Evol.*, **1**, 315 (1972).
[7] Lanks, K. W., *Nature*, **225**, 1424 (1970).
[8] Dayhoff, M. O., *Atlas of Protein Sequence and Structure 1969* (National Biomedical Research Foundation, Silver Spring, Maryland, 1969).
[9] Fitch, W. M., *J. Mol. Biol.*, **26**, 499 (1967).
[10] Sneath, P. H. A., *J. Theor. Biol.*, **12**, 157 (1966).
[11] Subak-Sharpe, J. H., *Handbook of Molecular Cytology* (edit. by Lima-de-Faria, A.), 67 (North-Holland, Amsterdam, 1969).

Erratum

IN the article by Bromberg *et al.* (*Nature New Biology*, **243**, 177; 1973) the title should read "Haemoglobin Little Rock (β^{143} His→Gln; H21): A High Oxygen Affinity Haemoglobin Variant with Unique Properties".

Editorial, Advertising and Publishing Offices of NATURE

MACMILLAN JOURNALS LIMITED
4 LITTLE ESSEX STREET, LONDON WC2R 3LF
Telephone Number: 01-836 6633. Telegrams: Phusis London WC2R 3LF
Telex 262024

MACMILLAN JOURNALS LIMITED
711 NATIONAL PRESS BUILDING
WASHINGTON DC 20004
Telephone Number: 202-737 2355. Telex 64280

Advertisement Department
MACMILLAN JOURNALS LIMITED
4 LITTLE ESSEX STREET, LONDON WC2R 3LF
Telephone Numbers: U.K. 01-836 6633
U.S.A. 202-737 2355

Subscription Department
MACMILLAN JOURNALS LIMITED
BRUNEL ROAD, BASINGSTOKE, HANTS RG21 2XS
Telephone Number: Basingstoke 29242

Classified advertisements
T. G. SCOTT & SON, LIMITED
1 CLEMENT'S INN, LONDON WC2A 2ED
Telephone Number: 01-242 6264/01-405 4743
Telegrams: Textualist London WC2A 2ED

Registered as a newspaper at the Post Office