Biochem. J. (1972) **128**, 1353–1356
Printed in Great Britain

Short Communications

The Frequency of Errors in Protein Biosynthesis

By ROBERT B. LOFTFIELD* and DOROTHY VANDERJAGT
University of New Mexico School of Medicine, Albuquerque, N.Mex. 87106, U.S.A.

(*Received* 15 *May* 1972)

Every protein molecule consists basically of a longer or shorter chain of some 20 different amino acids linked head-to-tail by peptide bonds. The selection of a particular amino acid at a particular position in the chain is determined by a three-letter code-word of nucleotides in mRNA, and ultimately by a similar code-word in the genetic DNA. There are at least two places in the translation of the message in the mRNA where errors are possible. First, an amino acid, which may differ very slightly from other amino acids, must be 'activated' and attached to a specific tRNA under conditions where other amino acids are rarely attached to this specific tRNA. Secondly, the aminoacyl-tRNA must associate with a specific three-base sequence of mRNA with infrequent interference from other aminoacyl-tRNA species.

From the biological evidence of genetic continuity and fidelity and from the chemical success in establishing unique amino acid sequences for many proteins, we must conclude that the processes of transcription (DNA→DNA and DNA→RNA) and translation (RNA→protein) are remarkably precise, although it is impossible on these grounds to make quantitative assessments of the degree of precision. On the other hand physicochemical calculations and chemical intuition argue that there must be a limit to the capacity of any system to discriminate between similar structures or surfaces and that there must be a finite incidence of error in transcription and translation. Accurate information on the frequency of such errors could exclude some schemes for the activation and polymerization of amino acids because the schemes would be unable to account for the observed fidelity. Alternatively, good experimental results may require the revision of physicochemical theories if a particular reaction mechanism is established in spite of its apparent inability to meet calculated physicochemical requirements.

Several years ago an effort was made to determine the frequency with which isoleucine, valine and leucine substitute for each other in the oviduct synthesis of ovalbumin (Loftfield, 1963). Experimentally the approach consisted of incubating minced oviduct tissue under separate but identical conditions with highly radioactive isoleucine, valine and leucine. The three ovalbumins were isolated separately by techniques that, it was assumed, would not discriminate between ovalbumin molecules differing by a single amino acid substitution. Then the three ovalbumins were hydrolysed with chymotrypsin, and a number of chymotryptic peptides were isolated, again with techniques that should not have distinguished between closely related peptides. Finally, the amount of radioactivity found in valine in a peptide containing isoleucine but not valine was determined and compared with the specific radioactivity of valine-containing peptides. Thus the peptide (Cys,Ile,Pro) contained 47500c.p.m./μmol in isoleucine (0.96mol/mol), 24 c.p.m./μmol in valine (0.006mol/mol) and 7c.p.m./μmol in leucine (0.0001 mol/mol). From such results it was concluded that the frequency of mistranslation in this ill-defined system (the heterogeneous tissue, oviduct, and the chemically undefined protein, ovalbumin) was about 1 part in 3000.

Regrettably that estimate, the only reported attempt to determine chemically the frequency of translation errors, had to be qualified from the first because of numerous sources of uncertainty. Every effort had been made to isolate pure ovalbumin and to isolate pure peptides, but there were no adequate criteria of purity for either. In fact, a few peptides that appeared to contain very large amounts of 'errors' were found and were judged to be either impure or to be determined by genetic factors. Theoretically the techniques used for peptide purification should not have distinguished between homologous (i.e. correct and slightly erroneous) peptides, but no experimental evidence existed to support the theory. The estimate of 1 error in 3000 was based only on the results of the comparative studies of the five clearest experiments. Because of the amounts of radioactivity and other technical limitations, the value given was described as a probable maximum error level rather than as the actual error level. Since the frequency of non-genetic errors has important consequences both for theories of protein biosynthesis and for the physical chemistry of molecular interactions, it is desirable to improve on the techniques for this determination, to minimize

*Temporary address: Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, U.K.

the sources of error and to seek true values for the frequency of non-genetic errors instead of mere maxima.

We now report the first results of a new study in which many of the sources of earlier doubt have been minimized. The biological system is better defined, being a suspension of rabbit reticulocytes, which synthesize 90% of their protein as haemoglobin. The protein studied, rabbit haemoglobin, has been thoroughly characterized and its amino acid sequence determined. As a result of genetic variations haemoglobin molecules show differences in amino acid sequence in some places. However, none is to be found in the sites that we have investigated. Human haemoglobin, closely homologous with rabbit haemoglobin except in sites where we seek non-genetic errors, provides a carrier for incorrectly synthesized rabbit peptides. For instance, the tryptic peptide T-2 from human haemoglobin is identical with the rabbit T-2 peptide when the latter has a valine residue in place of isoleucine in position 10 (Table 1). Under these conditions the human T-2 peptide will serve as a carrier of the incorrect rabbit peptide, and it is thus possible to monitor whether purification techniques result in the enrichment of one or the other species. Unlike the work reported earlier (Loftfield, 1963), the present higher specific radioactivities and improved purification procedures have led to four almost identical estimates of error frequency, thus encouraging us to believe that we are observing a true error frequency rather than a mere maximum probable value.

Although there is still much to do in refining the technique and applying it to many more positions in haemoglobin, to many more proteins and to many more tissues (such as juvenile versus senescent), publication of these early results not only provides more accurate data than heretofore, but demonstrates that this approach to the determination of error frequency is feasible and could well be applied now in other laboratories.

The structures of several haemoglobins are known and the relevant parts of the human and rabbit α-chains are shown in Table 1. The peptides T-2, obtained by tryptic digestion of purified rabbit and human α(1→23)-chains, are entirely homologous except for the substitution of isoleucine for valine. The peptides T-4-C-1, obtained by tryptic digestion followed by chymotryptic digestion, are also very similar, differing only by three conservative substitutions, including isoleucine for valine. The rationale of our approach is to incubate rabbit reticulocytes (in rabbit blood) with highly radioactive valine. The haemoglobin formed under these physiological intact-cell conditions is then isolated and purified by chromatography on CM-cellulose at pH 5.0 with a 0–1.0 M-NaCl gradient. The α-chains are separated and purified with pyridine–formic acid chromatography on CM-cellulose (Hammel & Bessman, 1964). The α-chain is then cleaved at methionine-32 with CNBr (Gross, 1967). The α(1→32)-peptide is prepared and purified from fresh human blood in the same way. The two α(1→32)-peptides are then mixed in known proportions. Since the two peptides differ in amino acid composition at several positions, one can use amino acid analysis to determine whether there is enrichment of peptides from either source in subsequent hydrolyses and purifications. The mixed peptides (H+R)α(1→32) are then hydrolysed with trypsin and chymotrypsin to give the peptides shown in Table 1. These peptides are purified by a succession of preparative electrophoreses at pH 9.0, 6.5 and 3.5 (Loftfield, 1963). Finally, the peptides are hydrolysed to establish that the observed radioactivity is, in fact, present in valine, an important precaution since commercial preparations of radioactive valine may contain several tenths of a per cent of isoleucine or leucine.

Table 1. *Structures of the N-terminal sequence* (1→32) *of human and rabbit haemoglobin α-chains*

Incorporation of radioactive valine in positions identified by **bold** type was used as a measure of error frequency. Incorporation of radioactive valine in positions identified by *italic* type was determined as a measure of the extent of correct translation.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Rabbit	*Val*	Leu	Ser	Pro	Ala	Asp	Lys	Thr	Asn	**Ile**	Lys	Thr	Ala	Trp	Glu	Lys
Human	Val	Leu	Ser	Pro	Ala	Asp	Lys	Thr	Asn	**Val**	Lys	Ala	Ala	Trp	Gly	Lys
				Peptide T-3						Peptide T-2				Peptide T-1		

	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Rabbit	**Ile**	Gly	Ser	His	Gly	Gly	Glu	Tyr	Gly	Ala	Glu	Ala	*Val*	Glu	Arg	Met
Human	**Val**	Gly	Ala	His	Ala	Gly	Glu	Tyr	Gly	Ala	Glu	Ala	Leu	Glu	Arg	Met
					Peptide T-4-C-1							Peptide T-4-C-2				

Table 2. *Composition, analysis and radioassay of human (H), rabbit (R) and mixed (H+R) peptides from haemoglobin*

The peptides were not analysed for proline and the basic amino acids. The analyses are not corrected for loss of serine, threonine or tyrosine or for incomplete hydrolysis of isoleucine- and valine-containing peptides.

	α(1→23)				T-2, α(8→11)				T-4-C-1, α(17→24)				T-3, α(1→7)				T-4-C-2, α(25→31)			
	H	R	3R+H (calc.)	3R+H (found)	H	R	3R+H (calc.)	3R+H (found)	H	R	3R+H (calc.)	3R+H (found)	H	R	3R+H (calc.)	3R+H (found)	H	R	3R+H (calc.)	3R+H (found)
Asp	2	2	2.0	2.15	1	1	1.0	1.06					1	1	1.0	1.01				
Thr	1	2	1.75	1.64	1	1	1.0	0.93												
Ser	1	2	1.75	1.67						1	0.75	0.60	1	1	1.0	0.88				
Glu	3	4	3.75	3.69					1	1	1.0	1.05					2	2	2.0	2.03
Gly	4	4	4.0	3.96					2	3	2.75	2.78					1	1	1.0	0.97
Ala	7	4	4.75	4.71					2		0.5	0.60	1	1	1.0	1.05	2	2	2.0	1.99
Val	3	2	2.25	2.27	1		0.25	0.27	1		0.25	0.26	1	1	1.0	0.99		1	0.75	0.86
Ile	0	2	1.5	1.20		1	0.75	0.73		1	0.75	0.56								
Leu	2	1	1.25	1.76									1	1	1.0	0.95	1		0.25	0.15
Tyr	1	1	1.0	0.61					1	1	1.0	0.42								
Radioactivity in valine (c.p.m./μmol of rabbit peptide)*				4.8×10^6				705				750				2.2×10^6				2.1×10^6
Amount of valine (mol/mol of rabbit isoleucine site)								0.0003				0.0004								
Radioactivity in valine (c.p.m./μmol of rabbit peptide)†				1.6×10^5				10				40				6.7×10^4				5.9×10^4
Amount of valine (mol/mol of rabbit isoleucine site)								0.0002				0.0006								

* Radioassay from an experiment with 5.0mCi of [^{3}H]valine and α(1→32)-peptides in the ratio 3R:1H.
† Radioassay from a similar experiment with 100μCi of [^{14}C]valine and α(1→32)-peptides in the ratio 1R:1H.

On the basis of the amino acid analyses (Table 2) it is clear that the electrophoretic and enzymic preparation of these peptides has not led to a substantial enrichment of peptides from either source. Beginning with 10ml of rabbit blood and 5mCi of [^{3}H]valine or 100μCi of [^{14}C]valine, 2½h incubation yields incorporation of the order listed in Table 2. Usually about 4μmol of haemoglobin is isolated and the final yield of purified peptide is between 0.4 and 1.0μmol. Representative results appear in Table 2.

Each of the two isoleucine-containing peptides from rabbit haemoglobin shows a frequency of valine substitution for isoleucine that is in the range 2–6 parts per 10000. (Two other isoleucine sites in rabbit haemoglobin cannot be examined by this approach because of the proximity of valine sites.) It is important to note that we have not excluded the possibility that erroneously incorporated valine has substituted for amino acids other than isoleucine. Some substitutions (as for histidine or glutamic acid) would lead to peptides that would surely be lost in purification, but others might contribute to our estimate of error frequency. Intuitively it seems most likely that much of the erroneous incorporation is due to a valine-for-isoleucine substitution. Not only are the two amino acids very closely related chemically, but their codons are identical except for the first letter (GUN and AUN), and this first letter is only an exchange of one purine for another.

Our observations strongly support an estimate for an error frequency of about 3 parts per 10000 for valine in these peptides. It remains to be established whether the substitution is largely valine for isoleucine, whether other pairs of amino acids with similar structures or similar codons have similar frequencies of mistranslation and whether, in view of the multiplicity of codons and tRNA species, the error frequency is the same for a particular amino acid at different loci.

These findings affirm that the precision of the two sequential steps of peptide assembly is very great, far greater than can be deduced from the study of non-biological chemical reactions (Pauling, 1957). Perhaps the cell possesses a scavenging mechanism for eliminating errors. More probably Nature has found means for accentuating the small differences among amino acids and the similarly small differences among trinucleotide code-words.

This work was supported by U.S. Public Health Service Grant CA-08000.

Gross, E. (1967) *Methods Enzymol.* **11**, 238
Hammel, C. L. & Bessman, S. P. (1964) *J. Biol. Chem.* **239**, 2228
Loftfield, R. B. (1963) *Biochem. J.* **89**, 82
Pauling, L. (1957) *Festschr. Prof. Dr. Arthur Stoll Siebzigsten Geburtstag* 597