View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by NASA Technical Reports Server

REPETITIONS IN THE POLYPEPTIDE SEQUENCE OF CYTOCHROMES

Charles R. Cantor and Thomas H. Jukes

Department of Chemistry and Space Sciences Laboratory
University of California,
Berkeley, California*

Received

It seems likely that many proteins have evolved from comparatively short primordial peptides by processes of duplication, deletions, and amino acid substitutions due to point mutations (Eck and Dayhoff, 1966; Jukes, 1966a). In most cases, the evidence for this may have vanished due to the masking effect of evolutionary changes in the amino acid sequences. In a few cases, however, remnants of the early peptides are still discernible in the forms of partially repetitive sequences. Proteins from "primitive" organisms may be the best source in which to search for such manifestations.

The phenomena of gene duplications and deletions are perceptible as translations of the genetic message in the polypeptide sequences of certain proteins. Presumably this indicates an evolutionary origin and significance for the occurrence of these phenomena in these instances. Examples are to be found in the α and β chains of hemoglobin A, the genes for which occupy separate chromosomal loci. Their separation is thought to result from duplication and translocation (Ingram, 1963). The two chains also show the occurrence of deletions, indicated by the presence of gaps in the homology of the polypeptides, for instance -

*Supported by a National Science Foundation Cooperative Graduate Fellowship and by grant NsG 479 by the National Aeronautics and Space Administration to the University of California.

N66-23521
(ACCESSION NUMBER)
9
(PAGES)
CR-74433
(NASA CR OR TMX OR AD NUMBER)
(THRU)
1
(CODE)
04
(CATEGORY)
FACILITY FORM 602

GPO PRICE $
CFSTI PRICE(S) $
Hard copy (HC) 1.00
Microfiche (MF) .50

Residue No.		12													25
	α	ala	ala	trp	gly	lys	val	gly	ala	his	ala	gly	glu	tyr	gly
		13													24
	β	ala	leu	trp	gly	lys	val	-	-	asn	val	asp	glu	val	gly
Base Changes		0	2	0	0	0	0			1	1	1	0	2	0

The presence of a gap corresponding to the deletion of 6 consecutive base pairs in the gene is inferred from the comparison of these sequences. The comparison also shows that 7 point mutations have taken place in the gene to produce changes at 5 homologous pairs of loci corresponding to individual amino acid residues indicated by base changes. Such amino-acid changes, when tolerated, may become so numerous as to obscure the original homology, thus erasing the record of the primordial duplication. An example of this is given below for human cytochrome c. The comparison of any two polypeptide chains is aided by enumerating the minimum number of base changes in the coding triplets (Brimacombe, et al, 1965) corresponding to each pair of amino acid residues at two homologous sites. In the comparison above, these are listed underneath the sequences. In this way two amino acids that are related by a single-base change in the code are noted as being more closely related than two which are separated by a two-base or three-base change. Three-base changes at a locus are, of course, difficult to detect because of the extensive synonymity in the third positions of the coding triplets (Brimacombe, et al, 1965).

Partial genetic duplication is evident in the haptoglobins (Smithies, 1964) and in Clostridium pasteurianum ferredoxin (Jukes, 1966b).

Fitch (1966) has proposed a method for examining polypeptide chains for regions of partial internal gene duplication, depending upon determining the minimum number of nucleotides which must be altered to permit the conversion of one sequence into the other. The test is carried out by means of a computer. We have devised a similar test, details of which will be published later. Having detected a homology by means of this test, it is often possible to make it visually obvious by means of the procedure used above with the hemoglobin peptides.

An examination of the primary structure of the cytochromes c indicates a region of partial duplication in the cytochrome c of Neurospora crassa (Heller and Smith, 1965). Using throughout the numbering system employed for the vertebrate cytochromes c, residues 5 to 19 are homologous with residues 20 to 34 as follows:

Residue No.	5														19
	lys	gly	ala	asn	leu	phe	lys	thr	arg	cys	ala	glu	cys	his	gly
	20														34
	glu	gly	gly	asn	leu	thr	gln	lys	ile	gly	pro	ala	leu	his	gly
Base Changes	1	0	1	0	0	2	1	1	1	1	1	1	2	0	0

This internal homology is not perceptible in the cytochromes c of other species, including those of certain vertebrates (Margoliash and Smith, 1965), two yeasts (bakers yeast and Candida krusei) (Narita, et al, 1963; Narita and Titani, 1965) and the moth Samia cynthia (Chan and Margoliash, 1966). As an example, the corresponding sequences for human cytochrome c show no evidence of homology beyond what might be interpreted as being coincidental, as follows:

Residue No.	5														19
	lys	gly	lys	lys	ile	phe	ile	met	lys	cys	ser	gln	cys	his	thr
	20														34
	val	glu	lys	gly	gly	lys	his	lys	thr	gly	pro	asn	leu	his	gly
Base Changes	2	1	0	2	2	3	2	1	1	1	1	2	2	0	2

The only two identical pairs are the lysines at positions 7 and 22 and the histidines at 18 and 33. In other cytochromes c, this last vestige of a common origin of the two sequences in this region has disappeared; residues 22 and 33 are occupied by asparagine and tryptophan in tuna fish cytochrome c (Kreil, 1963) as follows:

Residue No.	5														19
	lys	gly	lys	lys	thr	phe	val	gln	lys	cys	ala	gln	cys	his	thr
	20														34
	val	glu	asn	gly	gly	lys	his	lys	val	gly	pro	asn	leu	trp	gly
Base Changes	2	1	1	2	2	3	2	1	2	1	1	2	2	3	2

Evolution can thus in some cases erode the traces of an incident in its early history. In other instances, such as in the often-cited example of the invariant sequence of 11 amino acid residues in loci 70 to 80 of the cytochromes c, an essential portion of the same molecule has remained unchanged during the evolution of a diversity of living species from a common origin (Margoliash and Smith, 1965). It would seen that evolutionary changes in the region of residues 5 to 34 have proceeded more slowly in Neurospora crassa cytochrome c than in the other cytochromes c whose primary structure has been described. However, another example of homology in the same region is evident when a typical vertebrate cytochrome c is compared with the diheme peptide from the variant heme protein RHP, of the photoanaerobe Chromatium (Dus, Bartsch and Kamen, 1962) as follows, assuming a single amino acid gap between residues 19 and 20 in the vertebrate cytochrome c:

i Tuna						10									19	
cytochrome c:						phe	val	gln	lys	cys	ala	gln	cys	his	thr	-
	20				24											
	val	glu	asn	gly	gly											
						1									10	
ii Chromatium						phe	ala	gly	lys	cys	ser	gln	cys	his	thr	leu
					16								20			
	val	ala	asp	glu	gly	-	-	-	-	ser	ala	lys	cys	his	thr	phe
						27										
	-	-	asp	glu	gly	ser										
Base changes, i vs. ii:	0	1	1	1	0	0	1	2	0	0	1	0	0	0	0	-
Base changes, internal in ii:	-	-	0	0	0	1	-	-	-	1	1	1	0	0	0	1

The primary structure of the remainder of the Chromatium heme-carrying protein is not available and would be of great interest as possibly revealing a second incident of internal duplication in the regions preceding and following the known peptide.

REFERENCES

Brimacombe, R., J. Trupin, M. Nirenberg, P. Leder, M. Bernfield, and T. Jaouni. 1965. Proc. Natl. Acad. Sci., U.S., 54:954.

Chan, S.K. and E. Margoliash. 1966. J. Biol. Chem. 241:335.

Dus, K., R. G. Bartsch, and M. D. Kamen. 1962. J. Biol. Chem. 237:3083.

Fitch, W. M. 1966. Fed. Proc. 25:775; J. Mol. Biol. 16:000.

Ingram, V.M. 1963. The Hemoglobins in Genetics and Evolution. Columbia University Press, N.Y.

Jukes, T.H. 1966a. Molecules and Evolution. Columbia University Press, N.Y. (in the press).

Jukes, T.H. 1966b. Science (in the press).

Kreil, G. 1963. Z. Physiol. Chem. 334:154.

Margoliash, E. and E. L. Smith. 1965. In V. Bryson and H. J. Vogel (Eds.), Evolving Genes and Proteins, p. 221. Academic Press, N.Y.

Narita, K., K. Titani, Y. Yaoi, and H. Murakami. 1963. Biochem. Biophys. Acta 77:688.

Narita, K. and K. Titani. 1965. Proc. Japan Acad. 41:831.

Smithies, O. 1964. Cold Spring Harbor Symp. Quant. Biol. 29:309.