Vol. 96, No. 2, 1980 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
September 30, 1980 Pages 648-655

INTERNALLY ELONGATED RODENT α-CRYSTALLIN A CHAIN: RESULTING FROM INCOMPLETE RNA SPLICING?

Wilfried W. de Jong, Louis H. Cohen, Jack A.M. Leunissen and Anneke Zweers

Department of Biochemistry, University of Nijmegen,
6500 HB Nijmegen, The Netherlands

Received July 9,1980

SUMMARY: αA^{Ins}, an elongated α-crystallin A chain previously observed in rat, was present beside the normal αA chain in mouse, gerbil and hamster, which places its origin at least 30 million years ago. Like in rat the sequences of golden hamster αA^{Ins} and αA were found to be identical, apart from the internal insertion of 22 residues in αA^{Ins}. The hamster chains only differed from the rat chains by a single substitution in the inserted sequence of αA^{Ins}. The origin of αA^{Ins}, by insertion of 22 residues in an otherwise unchanged αA chain, and its rigid evolutionary conservation are most easily explained by assuming the incomplete removal of a putative intervening sequence from the precursor mRNA of αA, leaving an intracistronic insert of 66 nucleotides in part of the eventually translated mRNA.

We previously described that the eye lens protein α-crystallin of the rat contains beside the normal αA and αB chains a small amount of an elongated αA-like chain (1, 2). This chain, designated as αA^{Ins}, turned out to be identical in sequence to the normal rat αA chain of 173 residues, apart from the presence of an inserted stretch of 22 residues between positions 63 and 64 of the normal αA chain (3). Both chains are encoded by the same 14S mRNA fraction. In order to elucidate the mutational mechanism by which the αA^{Ins} chain has arisen, the time of its origin, and the evolutionary changes it may have undergone, we now have studied the occurrence of the αA^{Ins} chain in individual rats and in other species, and established the sequences of αA and αA^{Ins} in the golden hamster.

MATERIALS AND METHODS

α-Crystallin was isolated by gel filtration over Ultrogel AcA-34 (3) of aqueous extracts of lenses from the following species: rat (*Rattus norvegicus*), mouse (*Mus musculus*), mongolian gerbil (*Meriones unguiculatus*), chinese hamster (*Cricetulus griseus*), golden hamster (*Mesocricetus auratus*), red squirrel (*Sciurus vulgaris*), guinea pig (*Cavia porcellus*), chinchilla (*Chinchilla laniger*), calf (*Bos taurus*), pig (*Sus scrofa*), rabbit (*Oryctolagus cuniculus*) and rhesus monkey (*Macaca mulatta*). The presence of αA^{Ins} chains was assessed by SDS polyacrylamide slab gel electrophoresis (4, 5).

0006-291X/80/180648-08$01.00/0

αA and αA^{Ins} chains (72 mg and 9 mg, respectively) of golden hamster were isolated from 270 mg of α-crystallin, obtained from 100 lenses, by ion-exchange chromatography over DEAE-cellulose (3). The aminoethylated chains were digested with trypsin, and peptides separated by fingerprinting and gel filtration, as described for the rat chains (3). Amino acid compositions of all peptides were determined, and differences with the corresponding rat chains were confirmed by dansyl-Edman degradation (6).

RESULTS AND DISCUSSION

The possibility that the elongated α-crystallin chain αA^{Ins}, previously studied and isolated from pooled rat lenses, might be the product of an allele of the normal αA gene was studied by analysis of α-crystallin from five individual adult Wistar rats, randomly taken from a large breeding colony. The animals all showed on SDS gel electrophoresis the presence of αA^{Ins} in their α-crystallin in the same proportion as seen in pooled rat lens α-crystallin. αA^{Ins} and αA are therefore not the products of allelic genes.

In order to assess whether the αA^{Ins} chain, being a rather bizarre structural variant which might well be a recently originated and evolutionary less fit phenomenon, is limited to inbred laboratory rats, we analysed the α-crystallin from other mammals. SDS gel electrophoresis revealed a chain with the electrophoretic mobility of αA^{Ins}, and present in approximately the same proportion as rat αA^{Ins}, in α-crystallin from mouse, gerbil and hamster (Fig. 1). The chain was not observed in other rodents: red squirrel, guinea pig and chinchilla, nor in any of the as yet investigated non-rodent mammalian species (7). The fact that αA^{Ins} is found in the rodent families Muridae (rat, mouse) and Cricetidae (hamster, gerbil), which separated from each other in the Miocene (8), indicates that this chain has originated at least some 30 million years ago.

To determine the changes accumulated in αA^{Ins} and αA during these 30 Myr of separation, we analysed the sequences of these chains in the golden hamster for comparison with the known rat sequences. The results are summarized in Fig. 2 and show that the hamster αA chain is identical to that of the rat. In the αA^{Ins} chain of the hamster only a single difference is found with the rat αA^{Ins} chain. Remarkably this substitution, 84 Gly-Ile, is present in the inserted

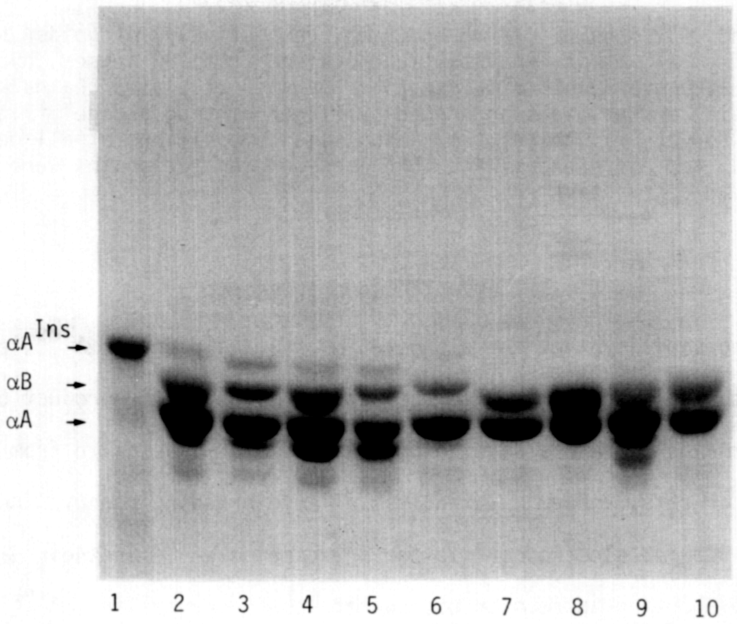

Fig. 1. The presence of the αA^{Ins} chain in different mammals. α-Crystallins were analysed by SDS-polyacrylamide slab-gel electrophoresis (4, 5). Samples are: 1. purified rat αA^{Ins}; α-crystallin from 2. rat; 3. mouse; 4. gerbil; 5. chinese hamster; 6. guinea pig; 7. calf; 8. pig; 9. rabbit, and 10. rhesus monkey. αA^{Ins} is seen to be present only in myomorph rodents (samples 2-5).

sequence, it requires two base substitutions, and is in contrast to most other substitutions observed in α-crystallin (7) a chemically "radical" replacement. It thus appears that both in rat and hamster the αA and αA^{Ins} chains are identical to each other, apart from the insertion in αA^{Ins}, and that these chains have remained unchanged, apart from the substitution at position 84 of αA^{Ins} in the rat or hamster lineage, for over 30 Myr.

Which molecular biological processes can account for the existence of the structurally closely related αA and αA^{Ins} chains? The origin of two non-allelic genes coding for these chains can reasonably be explained by gene duplication, a frequent evolutionary event (9). This duplication should then have taken place in the later stages of mammalian evolution, because both αA and αA^{Ins} of rat and hamster contain substitutions which occurred in earlier mammalian evolution (7)

Since the 173-residue αA chain occurs in all mammals, and therefore is the evolutionary oldest form, the gene duplication must have been followed by the

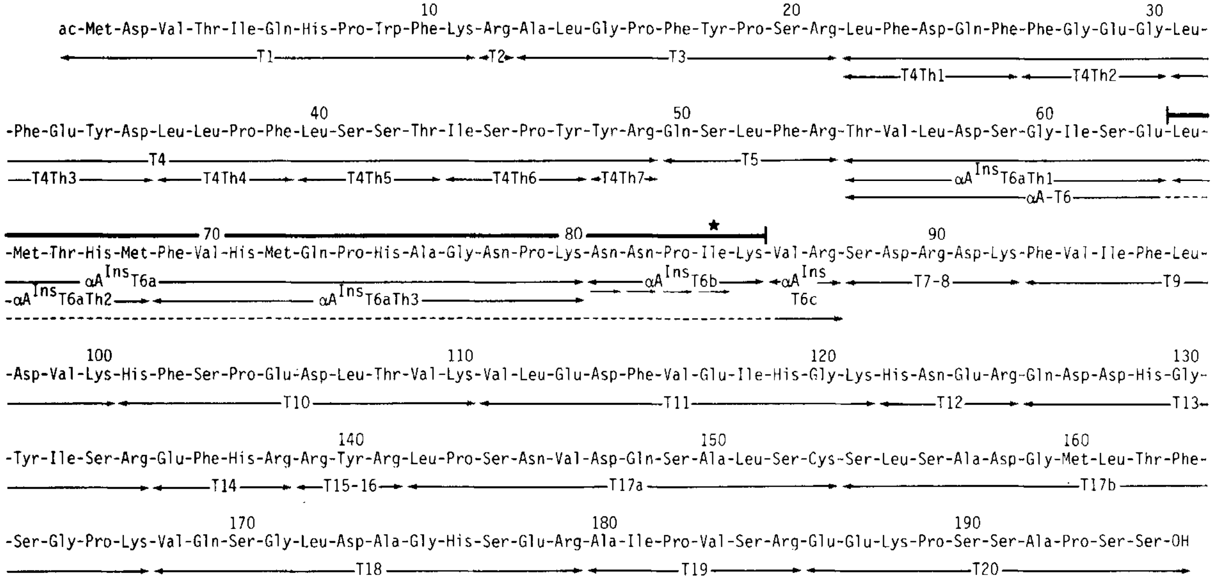

Fig. 2. Sequence of the αA and αA^{Ins} chains of the golden hamster. Both sequences are identical apart from the presence of an additional 22 residues in the αA^{Ins} chain, as indicated by the solid line. The asterisk marks the only difference with the corresponding rat chains (3): the replacement of residue 84-glycine by isoleucine. Amino acid analyses of all tryptic peptides (T) were found to be identical to those of the corresponding rat peptides (3), except αA^{Ins}-T6b which contained a residue of isoleucine instead of glycine as in the rat. The long tryptic peptides T4 and αA^{Ins}-T6a were further digested with thermolysin, and the resulting peptides (Th) isolated and analysed. The presence of isoleucine in position 84 was confirmed by four steps of dansyl-Edman degradation.

insertion of 66 base pairs coding for the additional residues in the αA^{Ins} chain. Such an intragenic insertion can be due to unequal crossing-over between homologous genes or to strand breakage followed by mispairing and repair, resulting in tandem-repeats in the products of the elongated genes (10). These mechanisms can not be responsible for the inserted sequence of the αA^{Ins} chain because this sequence shows no resemblance to neighbouring or other parts of the αA^{Ins} chain; also the corresponding inferred base-sequence, as deduced from the amino acid sequence, has no detectable homology with any other part of the mRNA. A second possibility for insertion of DNA-sequences into a gene could in principle be by means of the processes described for the different types of transposable genetic elements (11). Such elements usually involve at least several hundreds of base pairs, although a 66 bp mini-insert has been located in the yeast 21S rRNA gene (12). Moreover, such transposable genetic elements are usually characterized by the presence of inverted complementary sequences or direct repeats around the junctions of the insertion (11). We could not find indications for this type of regularities around the insertion in the base sequence of the αA^{Ins} mRNA, as deduced from the amino acid sequence of the αA^{Ins} chain, using a computer program designed to detect direct or complementary repeats in such inferred mRNA sequences (13).

According to these models no satisfactory explanation can be obtained for the origin of the inserted 66 nucleotides. Moreover, if the αA^{Ins} gene originated over 30 Myr ago, by gene duplication and insertion of 66 nucleotides, it is according to current opinions about protein evolution (9) highly unlikely that both gene products should not have accepted any additional substitutions. If αA and αA^{Ins} are functionally identical, one of the genes should have been relatively free to accumulate mutations. Even if αA^{Ins} fulfils some essential function, one would at least have expected to find some amino acid replacements to relieve the conformational strain which certainly must result from the insertion of 22 residues in an otherwise unchanged chain. Gene duplication followed by an insertion in one of the copies thus seems an unlikely explanation for the origin of αA^{Ins}.

The enigmatic features of the αA^{Ins} chain can, however, easily be explained in the light of the present knowledge about non-coding intervening sequences in eukaryotic genes (14). In fact changes in or near intervening sequences can be expected to lead to the insertion or deletion of amino acid sequences in otherwise normal polypeptide chains. A variant of human growth hormone appears to be the first example of the deletion of internal residues as a result of lengthening of an intervening sequence to include part of the adjacent coding sequence (15). The opposite process, failure to remove an intervening sequence or part thereof during processing of a primary RNA transcript, would lead to the appearance of inserted residues in a polypeptide chain. This phenomenon has not yet been reported, although read-through into intervening sequences possibly occurs in cytochrome b-deficient mutants of yeast mitochondrial genes (16).

Translation of an intervening sequence indeed provides the most plausible explanation for the existence of the αA^{Ins} chain. Although the structure of the α-crystallin genes is not yet known, they may be expected to contain intervening sequences like almost all other eukaryotic genes. The additional 66 nucleotides in the αA^{Ins} mRNA could correspond to either the 5' or 3' parts of an intervening sequence, or to a complete small intron; intervening sequences of comparable size have indeed been found in several genes (14, 17). The inferred base sequences around the insertion in the αA^{Ins} mRNA may well agree with the splice junction consensus sequences (18), apart that U or C have never been found as the 5' terminal base of intervening sequences (Fig. 3). One might speculate that this very deviation from the consensus sequence is responsable for the existence of αA^{Ins}, causing a "wobble" in the base-pairing to small nuclear ribonucleoproteins which have been proposed to be involved in RNA splicing (18).

It has been suggested that splicing of a primary RNA transcript need not be a hundred percent efficient (19), thus leading to the production by a single gene of two polypeptide chains differing in the presence of an inserted amino acid sequence in one of them. If this should be the case with the rat and hamster αA mRNA, it would immediately explain why αA and the αA-like part of

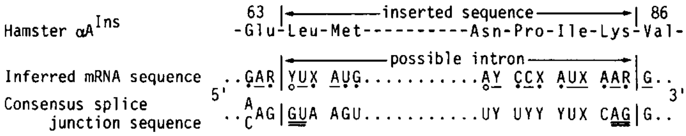

Fig. 3. Comparison of inferred base sequence in the αA^{Ins} mRNA, around the junctions of the inserted sequence, with the consensus splice junction sequence (18). The terminal intron consensus sequences GU and AG are the only ones which are constant in all hitherto reported splice junction sequences. The inferred αA^{Ins} mRNA bases are in agreement with the consensus sequence (—), or in possible agreement, c.q. have been found in the corresponding position in one or more other splice junctions (•) (18), apart from two bases (○). Y=U or C; R=A or G; X=U,C,A or G.

αA^{Ins} have remained identical for over 30 Myr. If such incomplete splicing were the result of a mutation in a single αA gene, for instance a base substitution in or near an intervening sequence, it is conceivable that such a situation can persist throughout 30 Myr of evolution. As long as the mutation allows the greatest part of the mRNA precursor to be processed and translated correctly, resulting in the production of the required amount of normal αA chains, natural selection has no means to remove the mutant gene. Even though the gene produces a small amount of functionally possibly less adequate αA^{Ins} chain it will, like other neutral or slightly deleterious mutants (20), have had a small but definite chance to maintain itself and reach fixation in an ancestral rodent population. The continuing selective pressure on the normal αA chain and the need for proper processing of the majority of the primary RNA transcript must have prevented further base substitutions, leading to amino acid replacements, in both the coding part and most of the normally removed intervening sequence. Although intron sequences change much faster in evolution than structural sequences, it may be that very small introns are more rigidly conserved (17). Obviously, restriction analysis and sequence determination of the rat αA gene(s) is required to conclusively establish the relation between the αA and αA^{Ins} chains.

ACKNOWLEDGEMENTS: We are grateful to Miss M. Versteeg for performing amino acid analyses. The work was supported in part by the Netherlands Foundation for Chemical Research (S.O.N.) and the Netherlands Organization for Pure Research (Z.W.O.).

REFERENCES

1. Cohen, L.H., Smits, D.P.E.M., and Bloemendal, H. (1976) Eur. J. Biochem. 67, 563-572.
2. Cohen, L.H., Westerhuis, L.W., Smits, D.P., and Bloemendal, H. (1978) Eur. J. Biochem. 89, 251-258.
3. Cohen, L.H., Westerhuis, L.W., De Jong, W.W., and Bloemendal, H. (1978) Eur. J. Biochem. 89, 259-266.
4. Laemli, U.K. (1970) Nature 227, 680-685.
5. De Jong, W.W., Zweers, A., and Cohen, L.H. (1978) Biochem. Biophys. Res. Commun. 82, 532-539.
6. Gray, W.R. (1972) Methods Enzymol. 25, 333-344.
7. De Jong, W.W., Gleaves, J.T., and Boulter, D. (1977) J. Mol. Evol. 10, 123-135.
8. Thenius, E. (1969) Phylogenie der Mammalia. Walter de Gruyter, Berlin.
9. Doolittle, R.F. (1979) in The Proteins, 3rd Ed., Vol. IV, Eds. Neurath, H., and Hill, R.L., Academic Press, New York, pp. 1-118.
10. Drake, J.W., and Baltz, R.H. (1976) Ann. Rev. Biochem. 45, 11-37.
11. Bukhari, A.I., Shapiro, J.A., and Adhya, S. (Eds.) (1977) DNA Insertion Elements, Plasmids and Episomes. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
12. Dujon, B. (1980) Cell 20, 185-197.
13. Wuilmart, C., Urbain, J., and Givol, D. (1977) Proc. Nat. Acad. Sci. USA 74, 2526-2530.
14. Crick, F. (1979) Science 204, 264-271.
15. Wallis, M. (1980) Nature 284, 512.
16. Van Ommen, G.J.B., Boer, P.H., Groot, G.S.P., De Haan, M., Roosendaal, E., Grivell, L.A., Haid, A., and Schweyen, R.J. (1980) Cell 20, 173-183.
17. Wahli, W., Dawid, I.B., Wyler, T., Weber, R., and Ryffel, G.U. (1980) Cell 20, 107-117.
18. Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L, and Steitz, J.A. (1980) Nature 283, 220-224.
19. Gilbert, W. (1978) Nature 271, 501.
20. Kimura, M., and Ohta, T. (1974) Proc. Nat. Acad. Sci. USA 71, 2848-2852.