Slow Rate of Evolution of αA Chain of α-Crystallin

The ever-increasing amount of comparative sequence data confirms that, in general, each protein evolves at a specific, more or less constant rate\(^1\)\(^{-3}\). The rate of acceptance of mutations in a particular gene is apparently influenced by the biological role and the size of the corresponding protein\(^2\). Knowledge of the rate and pattern of amino acid substitutions in more and functionally diverse proteins is, however, required to understand better the principles governing the evolution of proteins. It is moreover important to determine which residues in a given polypeptide chain are constant throughout evolution, because they can be considered to be of vital importance for the appropriate functioning of that protein.

Whereas most evolutionary studies have been performed on enzymes or other proteins with well known biological activities, we have started a comparative study of α-crystallin, a structural protein from the vertebrate eye lens which has no relationship to any other group of proteins for which sequence data are available. α-Crystallin offers some attractive properties for evolutionary studies: it is present in all vertebrates and can easily be isolated in considerable quantities.

In the bovine lens α-crystallin makes up 30% of the soluble protein. It is isolated as an aggregate with an average molecular weight of 800,000 (ref. 4) and is composed of two types of chains, each with a molecular weight of about 20,000, the acidic αA and basic αB chains, occurring in a ratio of 2 : 1. Both αA and αB chains are usually found in two electrophoretically distinguishable forms, designated αA\(_1\) and αA\(_2\), and αB\(_1\) and αB\(_2\), respectively, which are otherwise very similar\(^5\). αA\(_1\) has the same amino acid composition as αA\(_2\) and is thought to be derived from αA\(_2\) by in vivo deamidation\(^6\). Interest in α-crystallin has recently increased, since it seemed possible to isolate and translate the mRNA coding for the αA\(_2\) chain\(^7\). The complete amino acid sequence of the bovine αA\(_2\) chain has been established\(^8\). This provided the basis for a comparative study of αA chains from different mammals. The soluble lens proteins from ten mammals (cow, pig, horse, dog, rabbit, rat, mouse, guinea pig, rhesus monkey and human) were compared by acrylamide gel electrophoresis (Fig. 1). No difference in mobilities of the αA chains was observed among these species. This indicates that no substitutions involving charge differences exist between these chains, apart from the rather unlikely possibility that in a particular chain a substitution from a neutral to an acidic residue occurs together with a substitution from a neutral to a basic residue (or vice versa). More detailed studies were performed on the αA chains from cow, pig, dog and rabbit. α-Crystallin was prepared from aqueous lens extracts by isoelectric precipitation, followed by gel filtration on Sephadex G-200\(^9\). The αA chain was isolated by column chromatography on SE-Sephadex at pH 3.2 in 7 M urea\(^4\). The αA chain was reduced and aminoethylated\(^10\) and digested with 1% (by weight) trypsin in 0.1 M NH\(_4\)HCO\(_3\), pH 8.5, for 4 h at 37°C. The pH was lowered to 6.5 with 1 M HCl and the insoluble peptides removed by centrifugation. The soluble peptides were subjected to peptide mapping (Fig. 2). No significant differences were observed among the peptide maps of cow, pig and rabbit aminoethylated αA chains. The peptide map of canine αE-αA differed from the others in that peptide T11 had a considerably higher R\(_F\).

Peptides from fingerprints stained with 0.02% ninhydrin were eluted with 6 M HCl and hydrolysed for amino acid analysis. The results of these analyses, which will be published in full elsewhere, revealed only very few differences among the four species. Peptides which were identical in composition to the corresponding bovine ones were assumed to have the same sequence, which is reasonable considering the unlikeliness of reciprocal substitutions within a single short peptide. Differences in composition from the corresponding bovine peptides were observed in pig T3, T6, T18 and T20, in dog T3, T11 and T18, and in rabbit T1 and T18. These peptides were isolated by preparative paper electrophoresis and chromatography, using the systems described in the legend to Fig. 2. The unstained peptides were eluted from the paper with 10% acetic acid and sequenced according to the dansyl-Edman technique\(^11\). Dansyl amino acids were identified by polyamide thin-layer chromatography\(^12\). Assignment of amide groups was based on electrophoretic mobilities of peptides\(^13\).

The results are summarised in Table 1 and compared to the known bovine sequence\(^8\). All amino acid replacements can be accounted for by single base substitutions in the DNA. They are moreover of a chemically conservative character\(^14\), with the exception of the glutaminyl-leucyl interchange in dog T11.

Both dog and pig have alanine in position 13 where the other species have threonine. Since cow and pig are more closely related than pig and dog, at least two identical substitutions in different evolutionary lineages are needed to account for the occurrence of alanine or threonine in peptide T3.

Although this might seem to be a rare coincidence, it is known that in cytochrome c from 20% to 27% of all substitutions are in fact caused by parallel mutations\(^15\). The non-randomness of the evolutionary process is shown also by the presence of several substitutions in T1 and T18, whereas long stretches of the chain are unchanged. Comparative studies of other proteins have likewise demonstrated that although mutagenic events probably occur randomly over the entire gene, the distribution of accepted substitutions is non-random\(^16\)\(^{-18}\). Such departure from randomness reflects the constraints on the structure of a protein that are imposed by its function.

The analysed tryptic peptides of the four species account together for 137 residues of the total number of 173 residues in the αA chain. The remaining thirty-six residues are present in the insoluble 'core' peptides T4 and T9 which have not yet been analysed in pig, dog and rabbit. On the basis of our analyses we can make a first estimate of the rate of evolution of the mammalian αA chain. Among the investigated 137 residues we have observed the following number of substitutions (in parentheses the approximate time of divergence of the respective species\(^19\)\(^{-21}\)): between cow and pig four (50 m.y.);
The sequences which are identical in all four species are indicated by drawn lines. Except where the amino acid compositions of peptides differed from the corresponding bovine ones, the sequences for the other species were deduced from the bovine sequence. Peptides which differed in composition were sequentially degraded by the dansyl-Edman method (→). Peptides T4 and T9, which are insoluble at pH 6.5 and therefore not present on the peptide map, have not yet been investigated. Sites of cleavage with trypsin are indicated (↓). Because rabbit T1, like bovine T7, is blocked at the N-terminus, the relevant sequence was obtained by four degradation steps on the N-terminal cyanogen bromide fragment of the αA chain. Porcine T20 was split with thermolysin into two fragments, as indicated. Both thermolytic peptides were sequenced.
It seems that the rate of evolution of proteins slows down with increasing molecular weight, other factors being equal, because interior amino acid side chains are subjected to more stringent stereochemical requirements than exterior side chains. Although the αA chain itself has a rather low molecular weight (19,830), its slow rate of evolution apparently correlates better with the high molecular weight of the α-crystallin aggregate (average 800,000).

The slow rate of evolution of the αA chain agrees with the high degree of immunological cross-reactivity between α-crystallins from different mammalian species, although in fact no data are available on isolated αA chains.

If indeed the rate of evolution of proteins is limited by selective constraints on them, we must conclude that the αA chain is subjected to strong selective forces, hardly tolerating any change in the primary structure. Since α-crystallin is probably devoid of substitution-sensitive centres, like the substrate- or coenzyme-binding sites in enzymes, the selection must work solely at the level of the interactions within and between the polypeptide chains composing the molecule.

Apparently most amino acid substitutions in the αA chain would impair the integrity of the α-crystallin aggregate to such an extent that the appropriate functioning of the lens becomes seriously affected.

We thank M. Versteeg and G. Groenewoud for performing the amino acid analyses. Kiam Liem-The kindly provided rabbit α-crystallin. Lenses were obtained through the cooperation of the Central Animal Laboratory of the University of Nijmegen. This work was supported in part by the Netherlands Foundation for Chemical Research and by financial aid from the Netherlands Organization for the Advancement of Pure Research.

W. W. de Jong
J. M. van Amelsvoort
F. J. van der Ouderaa
H. Bloemendal

Department of Biochemistry,
University of Nijmegen,
Nijmegen

Received August 20; revised October 9, 1973.

3 McLaughlin, P. J., and Dayhoff, M. O., in Atlas of Protein Sequence and Structure (edit. by Dayhoff, M. O.), 5, 47 (National Biomedical Research Foundation, Silver Spring, Maryland, 1972).

Table 2 Sequence Comparison of Homologous Proteins and Rates of Evolution

<table>
<thead>
<tr>
<th>Protein</th>
<th>% Difference from cow</th>
<th>Pig</th>
<th>Dog</th>
<th>Rabbit</th>
<th>PAMs/100 m.y.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histone IV</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>α-Crystallin, αA chain</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Cytochrome c</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Insulin A and B chain</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Parathyroid hormone</td>
<td>12</td>
<td>20</td>
<td>18</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Haemoglobin a chain</td>
<td>16</td>
<td>19</td>
<td>21</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Haemoglobin b chain</td>
<td>16</td>
<td>19</td>
<td>21</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Pancreatic RNase</td>
<td>22</td>
<td>58</td>
<td>23</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Fibrinopeptides A and B</td>
<td>51</td>
<td>68</td>
<td>70</td>
<td>90</td>
<td></td>
</tr>
</tbody>
</table>