J Mol Evol (1995) 41:901–908

JOURNAL OF MOLECULAR EVOLUTION

A Reassessment of Mammalian αA-Crystallin Sequences Using DNA Sequencing: Implications for Anthropoid Affinities of Tarsier

Cynthia J. Jaworski

Section on Molecular Structure and Function, Laboratory of Molecular and Developmental Biology, National Eye Institute, Room 222, Building 6, National Institutes of Health, Bethesda, MD 28092, USA

Abstract. αA-crystallin, a major structural protein in the ocular lenses of all vertebrates, has been a valuable tool for molecular phylogenetic studies. This paper presents the complete sequence for human αA-crystallin derived from cDNA and genomic clones. The deduced amino acid sequence differs at two phylogenetically informative positions from that previously inferred from peptide composition. This led us to examine the same region of the αA-crystallin gene in 12 other mammalian species using direct sequencing of PCR-amplified genomic DNA. New sequences were added to the database, and corrections were made to all anthropoid sequences, defining clear synapomorphies for anthropoids as a clade distinct from prosimians. Within the anthropoids there are further synapomorphies delineating hominoids, Old World monkeys, and New World monkeys. Significantly, sequence revisions and the addition of new sequence for a prosimian, the sifaka, eliminate the previous support for the proposed anthropoid affinities of the tarsier inferred from αA-crystallin protein sequences. In addition, DNA sequences provide greater resolution of certain relationships. For example, although they are identical in protein sequence, comparison of DNA sequences clearly separates mouse and the common tree shrew, grouping the tree shrew closer to prosimians. These results show that adding DNA sequences to the existing αA-crystallin database can enhance its value in resolving phylogenetic relationships.

Key words: αA-crystallin — Tarsier — DNA sequencing — Anthropoid

Introduction

The transparency and refractive index of the vertebrate ocular lens depend upon families of structural proteins called crystallins (Bloemendal 1981; Berman 1991). These proteins are expressed at high levels in the lens and may survive without turnover throughout the life of the organism. Crystallins may be classified into two groups (Wistow 1993). The taxon-specific crystallins, which occur in phylogenetically restricted lineages, are enzymes that have acquired an additional role as structural proteins by a novel process of gene recruitment (Wistow 1993). In contrast, the ubiquitous crystallins, consisting of the α-crystallins and the β- and γ-crystallins, are represented in the lenses of all vertebrates.

Several features make αA-crystallin an excellent choice for phylogenetic comparisons. These are its broad distribution, high abundance, slow rate of evolutionary sequence change (de Jong 1981), and the fact that it is the product of a single-copy gene (King and Piatigorsky 1983; van den Heuvel et al. 1985). In the work of de Jong and co-workers (see de Jong and Goodman 1982; de Jong et al. 1984; Stapel et al. 1984; de Jong 1985), αA-crystallin was one of the first proteins to be used in the field of molecular systematics and constitutes one of the

U05569, Genbank accession number for human CDNA sequence; U24057, chimp; U24058, *Colobus;* U24059, capuchin; U24060, elephant; 024061, rhesus; U24062, sifaka; U24063, *Pteropus;* U24064, *Tupaia;* U24065, *Tonatia;* U24066, tarsier; U24067, tamarin; U24068, *Galago*

largest databases of sequences, suitable for constructing phylogenies encompassing all vertebrates (McKenna 1992; de Jong et al. 1993b). Such analyses have made several important contributions, among which was providing one of the key pieces of molecular evidence supporting anthropoid affinities of tarsier (de Jong and Goodman 1988).

Despite the increasingly widespread use of molecular data, the phylogenies inferred from molecular sequences are often problematic: there may be poor agreement with trees based on more traditional morphological methods, or the proposed relationships may contradict accepted biological concepts. Trees drawn from αA-crystallin protein data, for example, have been unable to separate all primates as a group from rodents and lagomorphs, and controversial positions are suggested for some species, such as the pangolin (de Jong et al. 1993b). Incongruities of molecular-based phylogenies often result from problems in the acquisition of data rather than problems inherent in the method. These include the presence of sequence errors in the database and the fact that sequence from a single individual is often the only representative of a diverse group of species. These factors can alter the topology of an inferred gene tree. In this work both areas are addressed with regard to αA-crystallin sequences.

First, the complete cDNA sequence of human αA-crystallin was determined. Unexpectedly, the deduced amino acid sequence reveals two discrepancies with the amino acid sequence originally reported (de Jong et al. 1975a). Both differences have been independently corroborated by partial protein sequencing (Takemoto and Emmons 1991) and more recently by partial cDNA sequence (Cooper et al. 1993). These discrepancies are, in part, the result of an apparent reciprocal substitution of amino acids which was not detected in compositional analysis of proteolytic digests. The revision of the human αA sequence implied that other related sequences might also need correction. In addition to the human cDNA sequence, the third exon of the αA-crystallin gene was sequenced from eight primates (one great ape: chimpanzee [*Pan troglodytes*]; two Old World monkeys: rhesus [*Macaca mulatta*] and colobus [*Colobus guereza*]; two New World monkeys: capuchin [*Cebus albifrons*] and golden lion tamarin [*Leontopithecus rosalia*]: three prosimians: sifaka [*Propithecus* sp., a lemur], galago [*Galago crassicaudatus*], and tarsier [*Tarsius syrichta*]); one mammal of uncertain affinity, the common tree shrew (*Tupaia glis*), and three nonprimate mammals (flying fox, a megachiropteran bat [*Pteropus hypomelanus*], round-eared bat, a microchiropteran bat [*Tonatia silvicola*] and African elephant [*Loxodonta africana*]). As predicted, the sequences determined for the anthropoids differ from those previously inferred from protein compositional data, and these differences are consistent with the revision of the human αA-crystallin sequence.

Phylogenetic analyses of the revised αA-crystallin sequences show that the anthropoid clade is defined by several synapomorphies. Within the anthropoid clade there are further synapomorphies including a rare deletion in Old World monkeys. However, in contrast to previous analyses, this expanded and revised dataset fails to support proposed anthropoid affinities for the tarsier (de Jong and Goodman 1988).

Materials and Methods

DNA Samples. Genomic DNA samples from capuchin (*Cebus albifrons*), tarsier (*Tarsius syrichta*), galago (*Galago crassicaudatus*), sifaka (Propithecus sp.), common tree shrew (*Tupaia glis*), flying fox (*Pteropus hypomelanus*), round-eared bat (*Tonatia silvicola*), and elephant (*Loxodonta africana*) were gifts from Dr. Morris Goodman and colleagues. Human and rhesus (*Macaca mulatta*) genomic DNA were purchased from Oncor (Gaithersburg, MD), and chimpanzee (*Pan troglodytes*) genomic DNA was purchased from Clontech (Palo Alto, CA). Colobus (*Colobus guereza*) and golden lion tamarin (*Leontopithecus rosalia*) DNA samples were prepared by standard methods (Davis et al. 1986) from postmortem eye tissues from the National Zoo, Washington, D.C. Oligonucleotides were synthesized on an Applied Biosystems 380B DNA synthesizer (Foster City, CA).

DNA Amplification Using the Polymerase Chain Reaction (PCR). The 3′ half of the cDNA for human αA-crystallin was amplified from single-stranded lens cDNA template which was made by reverse transcription (RT) of total RNA from 13.5-week fetal human lenses using an oligo(dT) primer. This lens template was subjected to PCR using the oligo(dT) primer in conjunction with primer #4206 (GGACGACTTTGTGGAGATCCACGG) which corresponds to amino acid residues 90–98 of the human αA-crystallin sequence, taken from previously determined genomic sequence (Jaworski and Piatigorsky 1989). Primers were used at final concentrations of 1 μM. PCR reactions (Mullis and Faloona 1987; Saiki et al. 1988) were performed using the AmpliTaq kit from Perkin-Elmer Cetus, according to the manufacturer's recommendations. The thermocycle profile consisted of 30 replications of 1 min at 94°C, 1 min at 55°C, followed by 1 min at 72°C.

For amplification of the third exon of αA-crystallin from genomic DNA the high conservation of αA-crystallin sequences (de Jong et al. 1984) facilitated the design of "generic" oligonucleotide primers, suitable for use when protein sequence itself is unknown. The 5′-oligonucleotide primer corresponds to the first 10 amino acids of exon 3 (positions 105–115) and the antisense primer corresponds to the last 10 amino acids and stop codon. The respective sequences are (5′): GACGACCACGGCTACATTTCCCGNGAGTTYCA and (3′): TTAGGACGAGGGAGCCGAGGWNGGCTTYTCYTC. These primers were used in PCR amplification of approximately 1.0 μg of genomic DNA using 30 cycles of the following profile: 1.5 min at 94°C, 2 min at 60°C, then 3 min at 72°C.

DNA Sequence Analysis. The PCR-amplified human cDNA was subcloned into the TA cloning vector (Invitrogen, San Diego, CA) in accordance with the manufacturer's instructions. Double-stranded sequencing of the cloned cDNA was performed with reagents and protocols from the Sequenase version 2.0 kit (USB, Cleveland, OH) and sequencing reactions were performed according to the USB manual, using αS^{35}-dATP (Amersham, Arlington Hts, IL). Samples were electrophoresed on 6% agarose-8M urea gels (BRL, Gaithersburg, MD) in TBE buffer at 70 V.

PCR products were also sequenced directly, as described elsewhere (*The Simple Fool's Guide to PCR*, version 2.0, Palumbi et al. 1991).

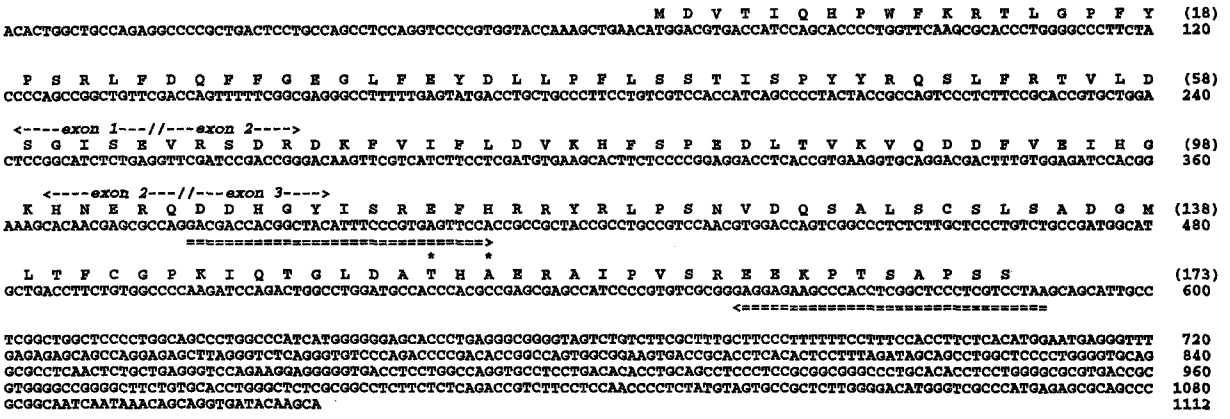

Fig. 1. Nucleotide and deduced amino acid sequence of human αA-crystallin cDNA. The sequence of the complete cDNA of human αA-crystallin was assembled from that of a reverse-transcribed PCR product from human lens RNA and previously published genomic sequences (Jaworski and Piatigorsky 1989; Jaworski et al. 1991). Exon boundaries are shown above the amino acid sequence. Revised amino acids sites, positions 153 and 155, are marked by *asterisks*. Positions of primers used for PCR of exon 3 are shown below the sequence with *arrows*.

PCR products were purified for direct sequencing by electrophoretic separation in 4% low melting agarose gels. After visualization with ethidium bromide staining, the band was excised and the DNA was purified using the Promega Magic (Wizard) PCR Preps kit (Promega, Madison, WI), with the entire product of each reaction in a final volume of 50 μl H_2O. The double-stranded PCR product (representing 10% of a 100-μl PCR reaction) was annealed to 10–60 pmol of primer in the presence of 2 μl of 5× Sequenase buffer and 1.0 μl of 5% NP-40 in a total volume of 11 μl. The sample was boiled for 2 min and then placed sequentially in dry ice/ethanol for 20 min, −20°C for 20 min, and finally equilibrated at room temperature. Each labeling reaction included 1.0 μl of 5% NP-40.

Computer Methods and Phylogenetic Analysis. Basic manipulations of DNA sequences were performed with the IDEAS package of programs (Kanehisa 1988) and the GCG package (Devereux et al. 1984) at the NCI Supercomputer Center, Frederick Cancer Research and Development Facility, Frederick, MD. Phylogenetic trees based on DNA sequences were produced with the Molecular Evolutionary Genetics Analysis (MEGA) program (Kumar et al. 1994), version 1.1 run on an IBM-compatible personal computer, using both the neighbor-joining (Saitou and Nei 1987) and the maximum parsimony options. All codon positions were weighted equally; gaps were treated by pairwise deletions, and genetic distances were calculated by the Tamura-Nei procedure (Tamura and Nei 1993). The MEGA program was similarly used for neighbor-joining trees drawn from amino acid sequences, with genetic distances calculated using the Poisson correction. Maximum parsimony analysis of amino acid sequences was performed using the branch and bound method of PAUP version 3.1 for the Macintosh (Swofford 1993), with the elephant as outgroup and gaps treated as a new character state.

Results and Discussion

Jaworski and Piatigorsky reported the cloning and DNA sequences of genomic fragments spanning the first two exons of human αA-crystallin including a novel pseudoexon (Jaworski and Piatigorsky 1989) and 1.8 kbp of the 5′-flanking sequences (Jaworski et al. 1991). To complete the nucleotide sequence of the expressed mRNA, a cDNA was obtained from human lens by RT-PCR. The PCR product was sequenced directly and confirmed by double-stranded sequencing of the cloned PCR product. The entire coding sequence and 5′- and 3′-untranslated regions, assembled from present and previous work, is presented in Fig. 1, with the deduced amino acid sequence shown above. Human αA-crystallin contains 173 amino acids and has a 3′-untranslated region of 522 bp, similar in length to that of mouse (532 bp) (King et al. 1982) but shorter than that of chicken (844 bp) (Thompson et al. 1987). The human and mouse 3′-UTRs are about 66% identical with the greatest similarity around the polyadenylation signal (not shown).

The deduced amino acid sequence confirms that originally reported by de Jong et al. (1975a), except for two residues in the C-terminal portion of the protein encoded by the third exon of the gene. The nucleotide sequence places a threonine at position 153 (153T) and an alanine at position 155 (155A), in contrast to the previous assignment of a deletion at position 153 (153Δ) and threonine at 155 (155T) in both human and rhesus αA-crystallin (de Jong et al. 1975b, 1984). The nucleotide sequence results have been independently confirmed by partial peptide sequence analysis of human αA-crystallin (Takemoto and Emmons 1991) and more recently by a partial cDNA sequence (Caspers et al. 1994).

Although DNA sequences have been determined for mouse, rat, cow, and chicken (de Jong et al. 1993a), in many of the species examined the αA-crystallin sequence was originally determined from the amino acid composition of proteolytic digests, with assignments made by homology with known sequences. This strategy facilitated the compilation of more than 50 mammalian αA-crystallin sequences, but these methods cannot detect reciprocal substitutions, something which had been recognized as a potential liability of the approach (de Jong 1982). The mis-assignments in the human αA-

```
                121
                 |
                 P  S  N  V  D  Q  S  A  L  S  C  S  L  S  A  D  G  M  L  T  F  C
HUMAN           CCGTCCAACGTGGACCAGTCTGCCCTCTCTTGCTCCCTGTCTGCCGATGGCATGCTGACCTTCTGT   (426)
CHIMP           CCGTCCAACGTGGACCAGTCGGCCCTCTCTTGCTCCCTGTCCGCCGATGGCATGCTGACCTTCTGT
RHESUS          NNNTCCAACGTGGACCAGTCGGCCCTCTCTTGCTCCCTGTCCGCCGACGGCATGCTGACCTTCTCT
COLOBUS         CCGTCCAACGTGGACCAGTCGGCCCTCTCTTGCTCCCTGTCCGCCGATGGCATGCTGACCTTCTCT
TAMARIN         CCGTCCAACGTGGACCAGTCGGCCCTGTCCTGCACCCTGTCCGCCGACGGCATGCTGACCTTCACT
CAPUCHIN        CCGTCCAACGTGGACCAGTCGGCCCTCTCTTGCTCCCTGTCCGCTGACGGCATGCTGACCTTCTCT
TARSIER         CCGTCCAACGTGGACCAGTCGGCCCTCTCTTGCTCCCTGTCCGCCGATGGCATGCTGACCTTCTCT
GALAGO          CCTTCCAACGTGGACCAGTCGGCCCTGTCCTGCTCCGTGTCCGCAGACGGCATGCTGACATTCTCT
SIFAKA          CCCTCCAACGTGGACCAGTCGGCGTTGTCCTGCTCCCTGTCCGCGGACGGCATGCTGACGTTCTCC
TUPAIA           CCTTCCAACGTGGACCAGTCGGCCCTCTCCTGCTCCCTGTCAGCAGACGGCATGCTGACCTTCTCG
PTEROPUS        CCCTCCAACGTGGACCAGTCGGCGCTGTCCTGCTCCCTGTCCGCCGACGGCATGCTGACCTTCTCC
TONATIA         CCCTCCAACGTGGACCAGTCGGCCCTCTCTTGCTCCCTGTCAGCCGACGGCATGCTGACCTTCTCT
ELEPHANT        CCCTCCAACGTGGACCAGTCTGCACTCTCTTGCTCCCTGTCTGCGGACGGCATGCTGACCTTCTGT

                143
                 |
                 G  P  K  I  Q  T  G  L  D  A  T  H  A  E  R  A  I  P  V  S  R
HUMAN           GGCCCCAAGATCCAGACTGGCCTGGATGCCACCCACGCCGAGCGAGCCATCCCCGTGTCGCGG     (489)
CHIMP           GGCCCCAAGATCCAGACTGGCCTGGATGCCACCCACGCCGAGCGAGCCATCCCCGTGTCGCGG
RHESUS          GGCCCCAAGATCCAGACTGGCCTAGACGCCACCCAT---GAGCGAGCCATTCCCGTGGCGCGG
COLOBUS         GGCCCCAAGATCCAGACTGGCCTAGATGCCACCCAC---GAGCGAGCCATTCCCGTGGCGCGG
TAMARIN         GGCCCCAAGATCCAGACTGGCCTGGACGCCACCCACGGCGAGCGAGCCATCCCCGTGTCGCGG
CAPUCHIN        GGCCCCAAGATCCAGACTGGCCTGGACACCACCCACGGCGAGCGAGCCATCCCCGTGTCGCGG
TARSIER         GGCCCCAAGATCCAGTCCGGCCTGGACGCCGGCCACAGCGAGCGAGCCATCCCCGTGTCACGG
GALAGO          GGCCCCAAGGTCCAGTCCGGCCTGGATGCCGGCCACAGCGAGCGAGCCATCCCCGTGTCGCGG
SIFAKA          GGCCCCAAGATCCCGTCTGGCCTGGACGCCGGCCACAGCGAGCGAGCCATCCCCGTGTCGCGG
TUPAIA          GGCCCCAAGGTCCAGTCGGGCCTGGACGCTGGCCACAGCGAGCGACCCATCCCCGTGTCCCGA
PTEROPUS        GGCCCCAAGGTCCCATCTGGCGTGGACGCCGGCCACAGCGAGCGAGCCATCCCCGTGTCGCGG
TONATIA         GGCCCCAAGATCCAGTCTGGCCTGGACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
ELEPHANT        GGCCCCAAGATCCAGTCTGGCATGGATGCCAGCCACAGTGAGAGAGCCATCCCCGTGTCCCGG
```

Fig. 2. Nucleotide sequences of αA-crystallin for 13 mammalian taxa. DNA sequences were obtained by direct sequencing of PCR products generated from genomic DNA samples using consensus oligonucleotide primers to exon 3. The region shown encompasses nucleotides 363–489 of αA-crystallin coding sequence; the corresponding amino acids of the human sequence, residues 121–163, are shown above in single-letter code. *N* denotes missing data; gaps are represented by *dashes.*

crystallin sequence were probably due to such reciprocal substitution, coupled with an underestimate of alanine content. This raised doubt about the assignments of residues at positions 153 and 155 in other primates. A possible consequence of corresponding revisions in other primate αA-crystallin sequences might be resolution of the primate clade and thus a more biologically acceptable topology for the αA-crystallin gene tree.

New and Revised αA-crystallin Sequences

To verify the amino acid assignments at positions 153 and 155, nucleotide sequences were determined for the third exon of αA-crystallin from 12 additional mammals, 8 primates (chimpanzee, rhesus, colobus, tamarin, capuchin, tarsier, galago, sifaka), the tree shrew (a species of uncertain primate affinities), 2 bats (1 megabat and 1 microbat), and African elephant. The critical region of the αA-crystallin gene was obtained by PCR amplification of genomic DNA samples, using primers corresponding to DNA sequences at the 5′ and 3′ termini of third exon coding sequences for αA-crystallin. PCR products were sequenced directly to avoid clonal selection of PCR-generated errors (Kunkel and Eckert 1989). This increased the size of the dataset for primate sequences and allowed for greater resolution in some species comparisons at the DNA level.

The resulting DNA sequences are shown in Fig. 2, compared to the corresponding sequence obtained from the human αA-crystallin cDNA. Figure 3 shows the deduced amino acid sequences with the mouse sequence as a reference. These results confirm the sequences and the 153G/155S pattern of galago and tarsier αA-crystallin and extend the list of prosimians by the addition of the sifaka, which also has the 153G/155S pattern.

All the anthropoid sequences contain distinct patterns consistent with the revised human sequence and different from those originally proposed. Chimpanzee αA-crystallin is identical to the revised human sequence at the amino acid level and contains only two differences at the nucleotide level (both in the third position of serine codons at positions 127 and 132). Like the hominoids, both the Old and New World monkeys have 153T. Thus the New World monkeys, tamarin and capuchin, have the 153T/155G pattern, the reverse of what had been reported for night monkey (*Aotus trivirgatus,* sometimes referred to as "owl monkey") (de Jong and Goodman 1988). The deletion at position 153 previously predicted for human and rhesus monkey is not present in human and chimpanzee sequences. However, there is a deletion in the rhesus sequence and in that of another Old World monkey, the colobus, but the deletion occurs at position 155 instead of 153. This deletion, a rare sequence change in αA-crystallins, occurs in a region of the C-terminal

	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142
MOUSE	P	S	N	V	D	Q	S	A	L	S	C	S	L	S	A	D	G	M	L	T	F	S
HUMAN	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
CHIMP	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
RHESUS	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
COLOBUS	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
TAMARIN	.	.	.	.	.	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T
CAPUCHIN	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
TARSIER	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GALAGO	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	.
SIFAKA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
TUPAIA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PTEROPUS	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
TONATIA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ELEPHANT	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
COW	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163
MOUSE	G	P	K	V	Q	S	G	L	D	A	G	H	S	E	R	A	I	P	V	S	R
HUMAN	.	.	.	I	.	T	.	.	.	.	T	.	A	.	.	.	.	.	.	.	.
CHIMP	.	.	.	I	.	T	.	.	.	.	T	.	A	.	.	.	.	.	.	.	.
RHESUS	.	.	.	I	.	T	.	.	.	.	T	.	Δ	.	.	.	.	.	.	A	
COLOBUS	.	.	.	I	.	T	.	.	.	.	T	.	Δ	.	.	.	.	.	.	A	.
TAMARIN	.	.	.	I	.	T	.	.	.	.	T	.	G	.	.	.	.	.	.	.	.
CAPUCHIN	.	.	.	I	.	T	.	.	.	T	T	.	G	.	.	.	.	.	.	.	.
TARSIER	.	.	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GALAGO	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SIFAKA	.	.	.	I	P	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
TUPAIA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PTEROPUS	.	.	.	.	P	.	.	V	.	.	.	.	.	.	.	.	.	.	.	.	.
TONATIA	.	.	.	I	.	.	.	.	.	–	–	–	–	–	–	–	–	–	–	–	–
ELEPHANT	.	.	.	I	.	.	.	M	.	.	S	.	.	.	.	.	.	.	.	.	.
COW	.	.	.	I	P	.	.	V	.	.	.	.	.	.	.	.	.	.	.	.	.

Fig. 3. Amino acid replacements within αA-crystallin residues 121–163. The amino acid sequences of residues 121–163, part of the third exon of αA-crystallin, were deduced from the DNA sequences of the listed species and compared to the amino acid sequence of the mouse (King et al. 1982). The bovine sequence is included as another mammalian αA-crystallin that has been verified by nucleotide sequence (Hay and Petrash 1987). *Single-letter* symbols for amino acids are used; *dots* indicate identity with the mouse; *triangles* represent a deletion, and *dashes* denote not determined. Previously published amino acid sequences for rhesus and human showed a deletion at position 153 and threonine at position 155; New World monkey residues at these positions were reported as 155-Gly and 155-Thr.

domain which seems to have undergone some unusual changes in Old World monkeys. In addition to the deletion, rhesus has substituted alanine for serine at position 162, thereby losing a site for O-GlcNAc modification conserved in αA-crystallins of mammals and birds (Roquemore et al. 1992).

Although all the anthropoid sequences are revised at positions 153 and 155, previously reported amino acid sequences for other groups are unchanged by the present data.

Increased Resolution of αA-crystallin Dataset Using NT Sequences

Trees drawn from the new and revised amino acid sequences presented in Fig. 3 did not produce significant changes in topology compared to inferences made from previous αA-crystallin data. Both distance methods (Fig. 4A) and maximum parsimony analysis (data not shown) show consistently that the anthropoids form a well-defined cluster but that the primates as a whole do not resolve as monophyletic.

Nucleotide sequences have proved useful for evolutionary studies of species with divergence times similar to those represented here, since more substitutions are present in nucleotide than in protein sequence, while "saturation" by multiple replacements has not yet been reached (Collins and Jukes 1994). Despite the limited length of the αA-crystallin exon 3, the increased resolution provided by nucleotide sequences nevertheless permits some refinements of the phylogeny.

The tree shrew is considered by some (see Martin 1990a) to be an early branch of the primate clade. On the basis of αA-crystallin protein sequence data (Fig. 3) the tree shrew is identical to mouse and neither can be phylogenetically separated from the prosimians. In contrast, mouse and tree shrew differ at 18 of 129 positions in αA-crystallin genomic sequence. In phylogenetic trees based on these data, mouse (and, as discussed below, tarsier) is separated from a branch which includes tree shrew, galago, and sifaka (Fig. 4B,C). The DNA data similarly improve the resolution among the anthropoids, and unlike the protein data, generate an inferred phylogeny consistent with the generally accepted branching order for New World monkeys, Old World monkeys, and hominoids.

A controversial issue in mammalian phylogeny recently addressed by molecular systematics concerns the monophyly of bats (Goodman 1991). Morphological considerations, including their shared adaptation for flight, have traditionally placed the megabats and microbats together. In contrast, neuroanatomical findings have shown that megabats, but not microbats, have a complex system of neural tracts that otherwise is unique to primates, suggesting that megabats may be the closest relatives of primates (Pettigrew 1986; Pettigrew et al. 1989), a view which has, however, been contradicted by others (Martin 1986). Analyses of DNA sequences from the globin locus (Bailey et al. 1992) and mitochondrial genes for cytochrome oxidase subunit I (Mindell et al. 1991) and subunit II (Adkins and Honeycutt 1991) support the monophyly of bats; however, the latter analysis suggests that the divergence between the suborders of bats is nearly as ancient as the chiropteran divergence from other mammalian orders. According to present and previous (de Jong et al. 1993b) αA-crystallin data, the two suborders of bats do not appear similar to each other. Thus αA-crystallin data are consistent with an early divergence of the chiropteran suborders, and as noted before, are uninformative with regard to the affinities of the chiroptera.

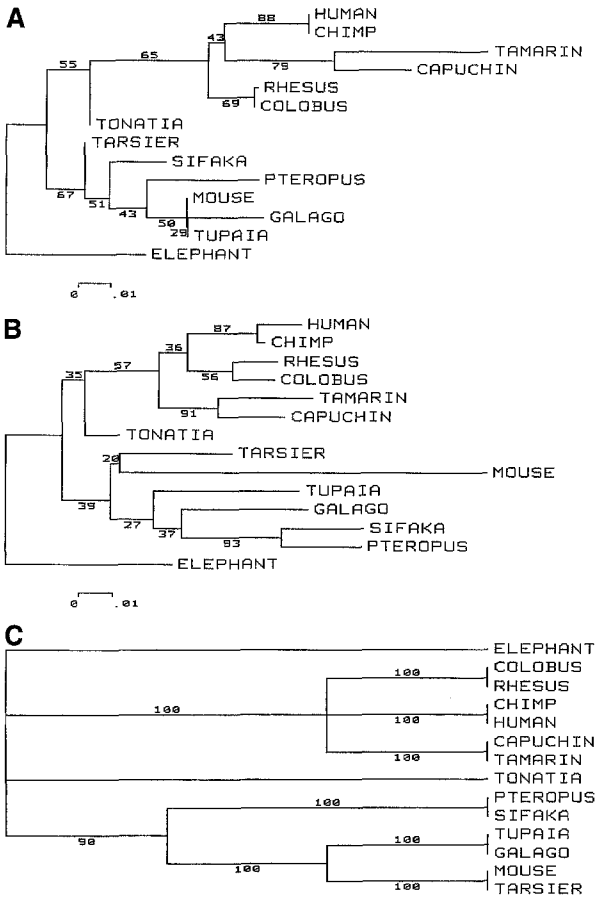

Fig. 4. Phylogenetic analysis of αA-crystallin exon 3 sequences was conducted with the MEGA program. **A** A phylogenetic tree inferred from deduced amino acid sequences of exon 3 (Fig. 3) was produced by the neighbor-joining method (Carver et al. 1993) with Poisson-corrected distances and elephant as outgroup. **B** Neighbor-joining tree based on nucleotide sequences shown in Fig. 2. Distances were calculated using the Tamure-Nei method, with 1,000 bootstrap replications. **C** Maximum parsimony tree was produced from the nucleotide sequences by the branch and bound method. The minimum length tree required 64 steps. The cladogram shown represents the consensus of ten trees, using the 50% majority rule.

Resolution of Anthropoid Clade

Using the revised αA-crystallin sequence data, the anthropoids form a distinct clade characterized by 13P (de Jong et al. 1993b), 148T, 153T, and lineage-specific substitutions at position 155; the latter two characters are revealed by the revisions presented here. The threonine at position 153 is unique among mammals, although 153T has also been found in swan and duck (Stapel et al. 1984). Most placental mammals (including all prosimians, rodents, lagomorphs, bats, and cattle) have 153G. At position 155, each anthropoid branch has characteristic substitutions: hominoids have alanine, New World monkeys have glycine, and Old World monkeys have a deletion. In contrast, all nonanthropoid mammals, with the single exception of the collared anteater, have 155S. Finally, among placental mammals 148T is found only in anthropoids although the marsupial opossum also contains 148T (de Jong et al. 1993b).

Reassessment of Tarsioid Affinities

The phylogenetic placement of the tarsier has been controversial (see reference Martin 1990b for overview). Some evidence has clustered tarsiers with prosimians, while other data support anthropoid affinities, with a division of primates into Strepsirhini (lemurs and lorises) and Haplorhini (tarsiers and anthropoids). An independent "third group" origin has also been proposed (see Culotta 1992). Molecular sequence data have been credited with supporting the Strepsirhini/Haplorhini division of primates (see Martin 1990b and Beard et al. 1991). However, this has been based on limited data consisting of the protein sequences of αA-crystallin (de Jong and Goodman 1988) and the hemoglobins (Czelusniak et al. 1990) and nucleotide sequences of β-globin (Koop et al. 1989). Thus, the αA-crystallin data set has had an important role in evaluating tarsier affinities.

The anthropoid cluster defined by the new and revised αA-crystallin sequences excludes the tarsier, contradicting earlier analyses based on αA-crystallin data (de Jong and Goodman 1988). Tarsier αA-crystallin, containing 13A (de Jong et al. 1993b), 148S, 153G, and 155S, has none of the anthropoid synapomorphies. Except for 13A, these diagnostic positions are the same as substitutions found in all prosimians.

The previous support from αA-crystallin protein sequences for anthropoid affinities of the tarsier had relied exclusively on the character at position 146: all prosimian sequences known at that time (potto, galago, lemur) contained 146V, while the tarsier and all anthropoids had isoleucine at this position. Inclusion of an additional prosimian (the sifaka) in the data provides an example of 146I in a prosimian species. With more primate taxa represented, it is evident that 146I, present in both anthropoids and prosimians, is not indicative of anthropoid affinities. Review of more distant mammalian αA-crystallin sequences shows that position 146 is homoplastic, with numerous parallel substitutions in different mammalian lineages. With an increased data density coupled with better resolution of the anthropoid clade, previous support from the αA-crystallin dataset for anthropoid affinities of the tarsier has been eliminated. Although a sister group cannot be identified from these data, αA-crystallin sequences suggest that the tarsiers are quite distant from the anthropoids.

This is consistent with other studies in which additional molecular sequences from the tarsier were examined. Studies of mitochondrial DNA from tarsiers by Hayasaka et al. (1988), later extended by Hasegawa et al. (1990), suggest that tarsiers are more closely related to lemurs then to anthropoids. In another study, the coding sequences from the IRBP gene (Stanhope et al. 1993) provided no resolution of the relative position of tarsiers.

Thus there are molecular data that suggest that if tarsiers belong to the anthropoid lineage, they must represent an ancient divergence. Divergence from prosimians or a third group origin are perhaps more likely.

The αA-crystallin data consist of unambiguously aligned coding regions of a single-copy nuclear gene. Expansion and revision of the data set, now including several representatives of anthropod and prosimian primates, have made the αA-crystallin sequences more informative for phylogenetic inferences, demonstrating the effect of an increase in the number of taxa on the topology of a gene tree (Wheeler 1992). Conceivably, the anthropoid synapomorphies in αA-crystallin sequences could be used in the classification of fossil primates of uncertain taxonomic status, such as adapid and omomyid species, if the fossils' "ancient DNA" proved to be suitably preserved.

Acknowledgments. I thank Dr. Morris Goodman for DNA samples and the Pathology Department of the National Zoo for tissue samples. I also thank Carl McIntosh for technical advice regarding direct sequencing of PCR products. I acknowledge the National Cancer Institute for allocation of computing time and staff support at the Advanced Scientific Computing Laboratory of the Frederick Cancer Research and Development Center. I appreciate the helpful comments and critical reading of the manuscript by Drs. Charles Sibley and Alfred Rosenberger and thank Dr. Joram Piatigorsky for support and encouragement during the completion of this work.

References

Adkins RM, Honeycutt RL (1991) Molecular phylogeny of the superorder Archonta. Proc Natl Acad Sci USA 88:10317–10321

Bailey WJ, Slightom JL, Goodman M (1992) Rejection of the "flying primate" hypothesis by phylogenetic evidence from the ϵ-globin gene. Science 256:86–89

Beard KC, Krishtalka L, Stucky RK (1991) First skulls of the early Eocene primate Shoshonius cooperi and the anthropoid-tarsier dichotomy. Nature 349:64–67

Berman ER (1991) Perspectives in vision research: biochemistry of the eye, 1st ed. Plenum Press, New York

Bloemendal H (ed) (1981) Molecular and cellular biology of the eye lens. Wiley, New York

Caspers GJ, Pennings J, de Jong WW (1994) A partial cDNA sequence corrects the human alpha A crystallin primary structure. Exp Eye Res 54:125–126

Collins DW, Jukes TH (1994) Rates of transition and transversion in coding sequences since the human–rodent divergence. Genomics 20:386–396

Culotta E (1992) A new take on anthropoid origins. Science 256:1216–1517

Czelusniak J, Goodman M, Koop BF, Tagle DA, Shoshani J, Braunitzer G, Kleinschmidt TK, de Jong WW, Matsuda G (1990) Perspectives from amino acid and nucleotide sequences on cladistic relationships among higher order taxa of Eutheria. In: Genoways HH (ed) Current mammalogy, vol 2. Plenum, New York, p 545

Davis LG, Dibner MD, Battey JF (eds) (1986) Basic methods in molecular biology. Elsevier, New York

de Jong WW (1981) Evolution of lens and crystallins. In: Bloemendal H (ed) Molecular and cellular biology of the eye lens. Wiley-Interscience, New York, p 221

de Jong WW (1982) Eye lens proteins and vertebrate phylogeny. In: Goodman M (ed) Macromolecular sequences in systematic and evolutionary biology. Plenum, New York, p 75

de Jong WW (1985) Superordinal affinities of Rodentia studied by sequence analysis of eye lens protein. In: Luckett WP, Hartenberger J-L (eds) Evolutionary relationships among rodents: a multidisciplinary analysis. Plenum Publishing, New York, p 211

de Jong WW, Goodman M (1982) Mammalian phylogeny studied by sequence analysis of the eye lens protein alpha-crystallin. Sonderdruck Z Saugetierkunde 47:257–276

de Jong WW, Goodman M (1988) Anthropoid affinities of Tarsius supported by lens αA-crystallin sequences. J Hum Evol 17:575–582

de Jong WW, Terwindt EC, Bloemendal H (1975a) The amino acid sequence of the A chain of human α-crystallin. FEBS Lett 58:310–313

de Jong WW, van der Ouderaa FJ, Versteeg M, Groenewoud G, van Amelsvoort JM, Bloemendal H (1975b) Primary structures of the α-Crystallin A chains of seven mammalian species. Eur J Biochem 53:237–242

de Jong WW, Zweers A, Versteeg M, Nuy-Terwindt EC (1984) Primary structures of the alpha-crystallin A chains of twenty-eight mammalian species, chicken and frog. Eur J Biochem 141:131–140

de Jong WW, Leunissen JAM, Voorter CEM (1993a) Evolution of the α-crystallin/small heat-shock protein family. Mol Biol Evol 10: 103–126

de Jong WW, Leunissen JAM, Wistow GJ (1993b) Eye lens crystallins and the phylogeny of placental orders: evidence for a macroscelidpaenungulate clade? In: Szalay FS, Novacek MJ, McKenna MC (eds) Mammal phylogeny: placentals. Springer Verlag, New York, p 5

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Goodman B (1991) Holy phylogeny! Did bats evolve twice? Science 253:36

Hasegawa M, Kishino H, Hayasaka K, Horai S (1990) Mitochondrial DNA evolution in primates: transition rate has been extremely low in the lemur. J Mol Evol 31:113–121

Hay RE, Petrash JM (1987) Nucleotide sequence of a bovine lens alpha A-crystallin cDNA. Biochem Biophys Res Commun 148:31–37

Hayasaka K, Gojobori T, Horai S (1988) Molecular phylogeny and evolution of primate mitochondrial DNA. Mol Biol Evol 5:626–644

van den Heuvel R, Hendriks W, Quax W, Bloemendal H (1985) Complete structure of the hamster αA-crystallin gene. Reflection of an evolutionary history by means of exon shuffling. J Mol Biol 185: 273–284

Jaworski CJ, Piatigorsky J (1989) A pseudo-exon in the functional human αA-crystallin gene. Nature 337:752–754

Jaworski CJ, Chepelinsky AB, Piatigorsky J (1991) The αA-crystallin gene: conserved features of the 5′-flanking regions in human, mouse, and chicken. J Mol Evol 33:495–505

Kanehisa M (1988) IDEAS-88: integrated database and extended analysis system for nucleic acids and proteins. Frederick Cancer Research Facility, Frederick, MD

King CR, Piatigorsky J (1983) Alternative RNA splicing of the murine alpha A-crystallin gene: protein-coding information within an intron. Cell 32:707–712

King CR, Shinohara T, Piatigorsky J (1982) αA-Crystallin messenger RNA of the mouse lens: more noncoding than coding sequences. Science 215:985–987

Koop BF, Tagle DA, Goodman M, Slightom JL (1989) A molecular view of primate phylogeny and important systematic and evolutionary questions. Mol Biol Evol 6:580–612

Kumar S, Tamura K, Nei M (1994) MEGA: molecular evolutionary genetics analysis software for microcomputers. Comput Appl Biosci 10:189–191

Kunkel TA, Eckert KA (1989) Fidelity of DNA polymerases used in

polymerase chain reactions. In: Erlich HA, Gibbs R, Kazazian HH Jr (eds) Polymerase chain reaction. Current communications in molecular biology. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p 5

Martin RD (1986) Are fruit bats primates? (news and views). Nature 320:482–483

Martin RD (1960a) Are tree-shrews primates? In: Primate origins and evolution. A phylogenetic reconstruction, 1st ed. Princeton University Press, Princeton, p 191

Martin RD (1990b) A provisional synthesis. In: Primate origins and evolution. A phylogenetic reconstruction. Princeton University Press, Princeton

McKenna MC (1992) The α crystallin A chain of the eye lens and mammalian phylogeny. Ann Zool Fennici 28:349–360

Mindell DP, Dick CW, Baker RJ (1991) Phylogenetic relationships among megabats, microbats, and primates. Proc Natl Acad Sci USA 88:10322–10326

Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In: Wu R (ed) Methods in enzymology, vol 155. Academic Press, San Diego, p 335

Palumbi S, Martin A, Romano S, McMillan WO, Stice L, Grabowski G (1991) The simple fool's guide to PCR, version 2.0. Department of Zoology and Kewalo Marine Laboratory, University of Hawaii, Honolulu

Pettigrew JD (1986) Flying primates? Megabats have the advanced pathway from eye to midbrain. Science 231:1304–1306

Pettigrew JD, Jamieson BGM, Robson SK, Hall LS, McAnally KI, Cooper HM (1989) Phylogenetic relations between microbats, megabats and primates (Mammalia: chiroptera and primates). Philos Trans R Soc Lond [Biol] 325:489–559

Roquemore EP, Dell A, Morris HR, Panico M, Reason AJ, Savoy L-A, Wistow GJ, Zigler JS Jr, Earles BJ, Hart GW (1992) Vertebrate lens α-crystallins are modified by O-linked N-acetylglucosamine. J Biol Chem 267:555–563

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis K, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487–491

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Stanhope MJ, Bailey WJ, Czelusniak JN, Goodman M, Si J-S, Nickerson J, Sgouros JG, Singer GA, Kleinschmidt TK (1993) A molecular view of primate supraordinal relationships from the analysis of both nucleotide and amino acid sequences. In: MacPhee RDE (ed) Primates and their relatives in phylogenetic perspective, chapter 8. Plenum Press, New York, p 251

Stapel SO, Leunissen JA, Versteeg M, Wattel J, de Jong WW (1984) Ratites as oldest offshoot of avian stem—evidence from α-crystallin A sequences. Nature 311:257–259

Swofford DL (1993) PAUP: phylogenetic analysis using parsimony, version 3.1. Computer program distributed by the Illinois Natural History Survey, Champaign

Takemoto L, Emmons T (1991) Truncation of αA-crystallin from the human lens. Exp Eye Res 53:811–813

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Thompson MA, Hawkins JW, Piatigorsky J (1987) Complete nucleotide sequence of the chicken αA-crystallin gene and its 5′ flanking region. Gene 56:173–184

Wheeler WC (1992) Extinction, sampling, and molecular phylogenetics. In: Novacek MJ, Wheeler QD (eds) Extinction and phylogeny, chapter 7. Columbia University Press, New York p 205

Wistow G (1993) Lens crystallins: gene recruitment and evolutionary dynamism. Trends Biochem Sci 18:301–306