Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 592–596, January 1989
Genetics

Structure of the rat argininosuccinate lyase gene: Close similarity to chicken δ-crystallin genes

(urea cycle/gene organization/sequence homology/lens protein/evolution)

TADASHI MATSUBASA*†, MASAKI TAKIGUCHI*, YOSHIHIRO AMAYA*, ICHIRO MATSUDA†, AND MASATAKA MORI*‡

*Institute for Medical Genetics, and †Department of Pediatrics, Kumamoto University Medical School, Kumamoto 862, Japan

Communicated by Motoo Kimura, October 6, 1988 (received for review August 10, 1988)

ABSTRACT **Argininosuccinate lyase (EC 4.3.2.1) is an enzyme of arginine biosynthesis and is also involved in the urea cycle in the liver of ureotelic animals. A comparison of cDNA-derived amino acid sequences revealed that argininosuccinate lyase is highly homologous with chicken δ-crystallin, a major structural protein of the eye lens. The gene for the rat argininosuccinate lyase was cloned and its structure was determined. This gene is a single-copy gene about 14 kilobases long and is split into 16 exons. A comparison with chicken δ-crystallin genes revealed that all introns interrupt the protein-coding regions at homologous positions. This close similarity in structural organization provides strong evidence for the view of Piatigorsky *et al.* [Piatigorsky, J., O'Brien, W. E., Norman, B. L., Kalmuck, K., Wistow, G., Borras, T., Nickerson, J. M. & Wawrousek, E. F. (1988) *Proc. Natl. Acad. Sci. USA* 85, 3479–3483] that chicken δ1- and δ2-crystallin genes evolved by recruitment and duplication of the preexisting argininosuccinate lyase gene and that δ2-crystallin is probably the direct homologue of argininosuccinate lyase.**

Argininosuccinate lyase (EC 4.3.2.1) is an enzyme involved in the urea cycle in the liver of ureotelic animals and catalyzes the conversion of argininosuccinate to arginine and fumarate. The enzyme, purified from steer liver (1) and from human liver (2, 3), consists of four identical subunits of about 50,000 Da. The steer enzyme has been extensively characterized by Ratner and colleagues (4). Argininosuccinate lyase is most active in the liver but is also present in all other tissues (4), where its primary role is presumably the biosynthesis of arginine. In ongoing work on the urea cycle, we isolated cDNA clones for rat argininosuccinate lyase (5) and determined the nucleotide sequence (6). The nucleotide sequence of the human enzyme cDNA is known (7, 8) and the gene was assigned to chromosome 7 (7).

δ-Crystallin is the dominant structural protein in lenses of birds and reptiles but apparently is not present in mammals. There are two known δ-crystallin genes (δ1 and δ2) in the chicken that are tandemly arranged and separated by about 4.2 kilobases (kb) of DNA (9, 10). The structures of δ1 (11, 12) and δ2 (13) genes were determined. The δ1 gene is actively expressed in the embryonic lens and encodes the major structural protein (δ1-crystallin), whereas the δ2 gene appears to be weakly expressed in various tissues. The nature of the δ2 gene product (δ2-crystallin) is unknown (13, 14). A comparison of the chicken δ1- and δ2-crystallins showed that they are highly homologous (51–69% identical) to human, rat, and yeast argininosuccinate lyase (6, 15). Piatigorsky *et al.* (16) have provided evidence suggesting that δ2-crystallin is a direct homologue of argininosuccinate lyase. We have now cloned the rat argininosuccinate lyase gene and determined its organization.§ The gene is about 14 kb long and contains 16 exons. The exon–intron organization of the gene is strikingly similar to that of the chicken δ-crystallin genes, thereby supporting the view that the two δ-crystallin genes were formed by duplication of an ancestral argininosuccinate lyase gene and that the chicken δ2-crystallin is argininosuccinate lyase.

MATERIALS AND METHODS

Isolation and Characterization of Genomic Clones. A rat genomic library, constructed by using bacteriophage λ Charon 4A as a vector, was provided by T. D. Sargent (National Institute of Child Health and Human Development). It was screened with nick-translated cDNA fragments excised from the rat argininosuccinate lyase cDNA plasmid pALr-3 (5). The cDNAs were nick-translated with [α-^{32}P]-dCTP (3000 Ci/mmol; 1 Ci = 37 GBq; Amersham) to a specific activity of $\approx 10^8$ cpm/μg of DNA. Phage DNAs of positive clones were characterized by restriction mapping and by Southern blot hybridization analysis. Subclones were constructed with the plasmid pUC18 as a vector, and the nucleotide sequences of the plasmid subclones were determined by the dideoxynucleotide chain-termination method (17), using alkali-denatured plasmid as a template (18) and synthetic oligonucleotide primers complementary to the vector or cDNA sequences. Oligonucleotide primers were synthesized by an Applied Biosystems model 381A DNA synthesizer.

RESULTS

Isolation and Characterization of Genomic Clones. A phage library constructed from rat liver DNA was screened with a rat argininosuccinate lyase cDNA as a probe. Six independent clones were isolated and analyzed by restriction enzyme digestion (Fig. 1). These clones overlapped and spanned about 29 kb. The cDNA hybridized with *Eco*RI restriction fragments of 6.5, 4.0, 3.0, and 2.8 kb. These fragments account for all fragments detected by Southern blot analysis of *Eco*RI digests of rat chromosomal DNA (data not shown). This observation and the results of the following sequence analysis indicate that the argininosuccinate lyase gene of the rat is a single-copy one, with no pseudogene. This is in contrast to the human gene, in which another hybridizing region was found on chromosome 22 in addition to the true gene on chromosome 7 (7).

To define positions and boundaries of the exon blocks, the restriction fragments identified by Southern hybridization were subcloned and the sequences were determined. A physical map of rat argininosuccinate lyase gene is shown in

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡To whom reprint requests should be addressed.

§The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04159).

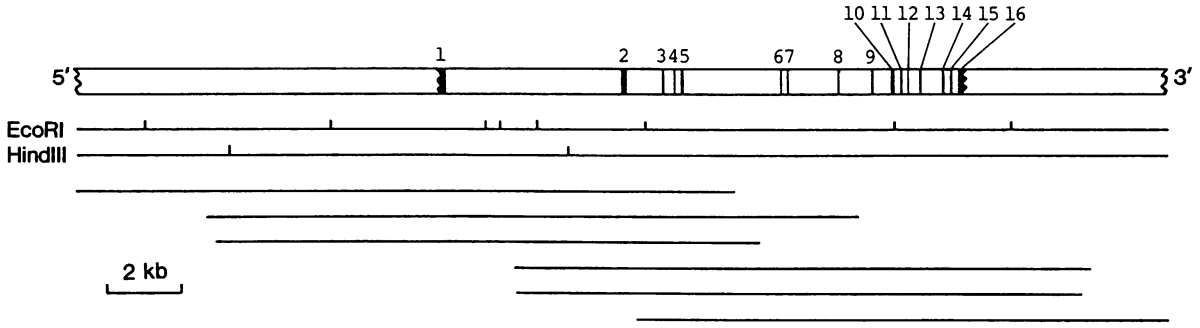

FIG. 1. Physical map of the rat argininosuccinate lyase gene. The structure of the gene is shown as the bar at the top of the diagram. Exons 1–16 are shown as vertical bars with exon numbers. *Eco*RI and *Hin*dIII sites are shown below the gene structure. The genomic DNA fragments contained in the phage clones are shown below the restriction maps.

Fig. 1. The gene is about 14 kb long and is divided into 16 exons. The 16 exons range from 53 bases (exon 8) to >175 bases (exon 16), and the 15 introns range from 104 bases (intron 14) to 4.5 kb (intron 1). The extent of exon 1 at the 5′ end and that of exon 16 at the 3′ end have not yet been determined; hence their exact sizes are unknown. Correspondingly, the existence of additional introns in the 5′ and 3′ untranslated regions is possible.

The sequences of the intron–exon junctions of the argininosuccinate lyase gene are compiled in Table 1 (see also Fig. 1). All of the splice donor and acceptor sites conform to the GT/AG rule for nucleotides immediately flanking exon borders, except intron 13, where the the donor/acceptor nucleotides are GC/AG. The introns with the unusual GC/AG sequence were found in several genes containing the mouse αA_2-crystallin gene (19). The binding site for argininosuccinate (20) is encoded by exon 2.

Comparison with Chicken δ-Crystallin Genes. A striking similarity of amino acid sequences of argininosuccinate lyase and δ-crystallins strongly suggests that these genes were evolved from a common ancestral gene. To shed light on the origins of the genes for rat argininosuccinate lyase and chicken δ1- and δ2-crystallins, the structural features of these genes were compared.

The position of introns in the argininosuccinate lyase and δ-crystallin genes relative to their coding sequences is shown

Table 1. Exon–intron organization of rat argininosuccinate lyase gene

Exon no.	Exon size, bp	Sequence at exon–intron junction		Intron size, bp	Amino acid interrupted
		5′ splice donor	3′ splice acceptor		
1	>169	GCA TCG GAG gtgtgt 12	tgtcctcttggccag AGC GGG AAG 13	≈4500	Glu-4/Ser-5
2	195	CTG GAC AAG gtagtc 207	ctgctacgtgttcag GTG GCT GAA 208	≈700	Lys-69/Val-70
3	84	CGC CTG AAG gtatga 291	gcctctccattccag GAA CTC ATT 292	212	Lys-97/Glu-98
4	57	AAT GAC CAG gtgtcg 348	ttctcctctccacag GTG GTC ACG 349	125	Gln-116/Val-117
5	98	GCA GAG GC gtgagc 446	cttcccctaccacag G GAG TGT 447	≈2500	Ala-149
6	78	ATC CAG AG gtaagg 524	acctatttcccacag T CAC GCC 525	113	Ser-175
7	78	CTG GGC AG gtaagg 602	cccttgtctctacag T GGG GCC 603	≈1200	Ser-201
8	53	TGT GCA G gtgagg 655	ctccctcacctccag AA CTG AAC 656	≈900	Glu-219
9	63	TTC GTG G gtaacg 718	ctgtgtcccttccag CT GAG TTC 719	447	Ala-240
10	115	GCC TAC AG gtaagg 833	acctctgtcccccag C ACC GGA 834	145	Ser-278
11	85	TTT GGA CGG gtgagc 918	tcactctcccctag TGC GCA GGA 919	114	Arg-306/Cys-307
12	60	GAC TTA CAG gtatga 978	ccacacctctctcag GAA GAC AAG 979	230	Gln-326/Glu-327
13	84	ACA CTG CAG gcaaga 1062	acctaacgcctccag ATT CAT CGT 1063	498	Gln-354/Ile-355
14	81	CGC AAA GGG gtgagt 1143	tgcctatgtacatag ATG CCA TTC 1144	104	Gly-381/Met-382
15	107	ACC GTC AG gtatgg 1250	cctctttctccccag T CCC CTG 1251	151	Ser-417
16	>175				

Exon sequences are shown in uppercase letters and intron sequences are in lowercase letters. Numbers immediately below the DNA sequences denote the nucleotide positions at which the introns interrupt the argininosuccinate lyase mRNA. The amino acids with numbers shown in the extreme right-hand column denote the positions of the indicated amino acids in the argininosuccinate lyase sequence interrupted by each intron. bp, Base pairs.

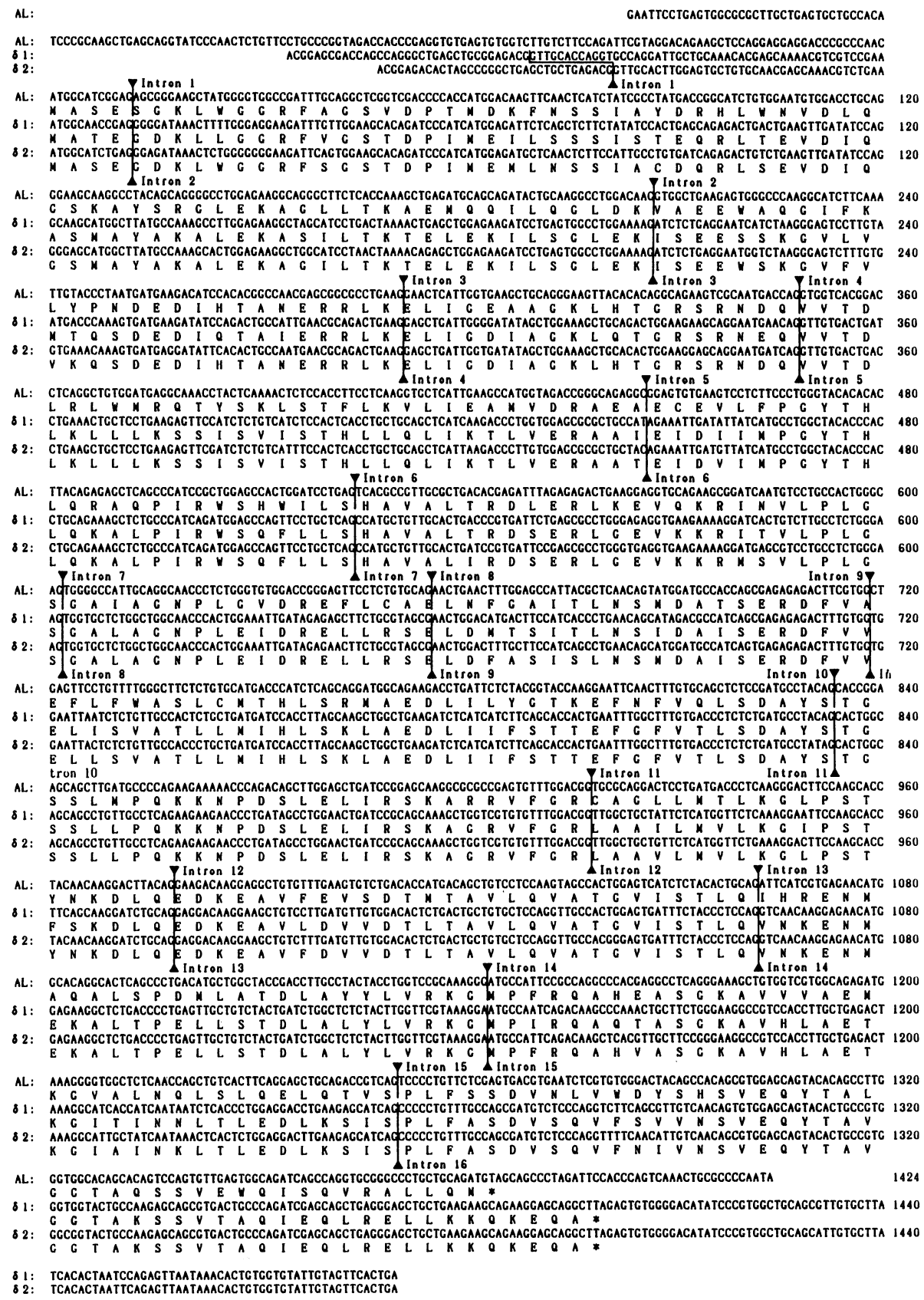

FIG. 2. Relative positions of introns in the rat argininosuccinate lyase gene (AL) and in the chicken δ1- and δ2-crystallin genes. Nucleotide and deduced amino acid sequences of argininosuccinate lyase are from Amaya *et al.* (6), those of δ1-crystallin are from Yasuda *et al.* (9), and those of δ2-crystallin are from Nickerson *et al.* (13). The nucleotide numbering is in the right margin. The positions of intronic interruptions are indicated by closed triangles with intron numbers. The intron positions of the argininosuccinate lyase gene are from Table 1, those of the δ1 gene are from Ohno *et al.* (11) and Nickerson *et al.* (12), and those of the δ2 gene are from Nickerson *et al.* (13). Lys-51 of argininosuccinate lyase is the binding site for argininosuccinate (20).

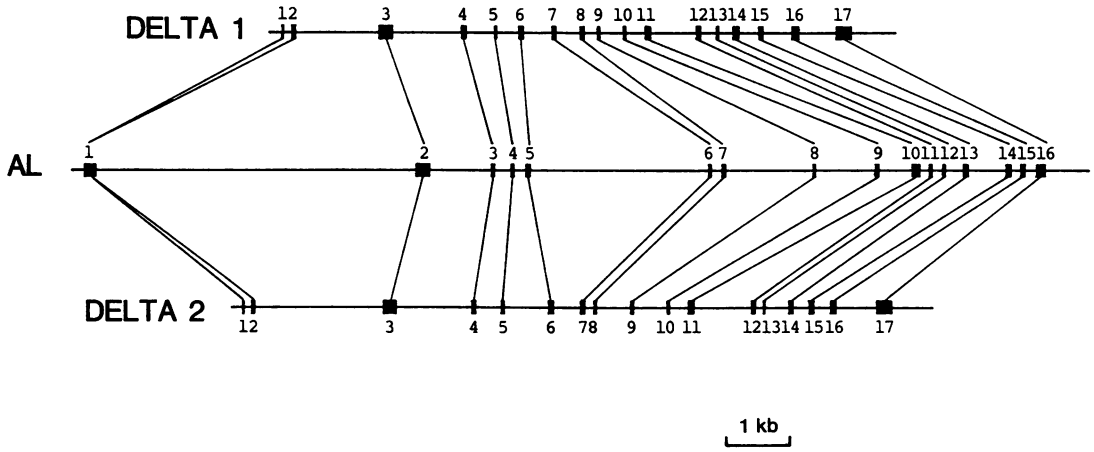

FIG. 3. Comparison of the structural organization of the rat argininosuccinate lyase gene (AL) and the chicken δ1- and δ2-crystallin genes. Exons are represented by vertical bars of various thickness with exon numbers, and the introns are represented by horizontal lines. The organization of the argininosuccinate lyase gene is from Fig. 1, that of the δ1 gene is from Ohno *et al.* (11) and Nickerson *et al.* (12), and that of the δ2 gene is from Nickerson *et al.* (13). Corresponding exons are linked by lines. Exon 1 of the argininosuccinate lyase gene corresponds to exons 1 and 2 of the δ-crystallin genes.

in Fig. 2. The intron location of the δ2-crystallin gene is that estimated based on remarkable similarity between the δ1 and δ2 genes (13). When the argininosuccinate lyase and δ-crystallin sequences are aligned for maximal identity, it is apparent that, with the exception of the first introns of δ-crystallin genes, the remaining 15 introns interrupt the protein-coding region at precisely the same positions. This striking similarity of intron positioning in the argininosuccinate lyase gene and δ-crystallin genes provides good evidence that these genes derived from a common ancestor.

A comparison of the overall architecture of the argininosuccinate lyase and δ-crystallin genes is shown in Fig. 3. In contrast to the conservation of the intron–exon junction positions, the length of the genes and of the corresponding introns varies among the three genes. The size difference between the argininosuccinate lyase gene and the δ-crystallin genes is more pronounced than that between the δ1- and δ2-crystallin genes. The argininosuccinate lyase gene (about 14 kb) is larger than the δ1 and δ2 genes (about 8 and 9 kb, respectively). Introns 1 and 5 of the argininosuccinate lyase gene are in particular much larger than the corresponding introns of the δ-crystallin genes.

DISCUSSION

Crystallins are the major proteins of the eye lens. Recently, unexpected relationships were seen to exist between crystallins and other proteins. The α-crystallins are related to the small heat shock proteins (21) and to p40, a schistosome egg antigen (22), and the β- and γ-crystallins are related to protein S, a calcium-binding bacterial spore coat protein (23). More interestingly, the major taxon-specific crystallins are either metabolic enzymes or closely related to enzymes (15). Duck ε-crystallin is identical to lactate dehydrogenase (LDH-B) (24), chicken δ-crystallin is closely related to argininosuccinate lyase (6, 15, 16), turtle τ-crystallin is related to enolase (15), squid S_{III}-crystallin is related to glutathione *S*-transferase (15), and frog ρ-crystallin is related to a family of NADPH-dependent reductases, including aldehyde and aldose reductase (25), and more closely related to prostaglandin F synthetase (26).

Two δ-crystallin genes of the chicken are tandemly arranged and separated by about 4.2 kb of DNA (9, 10). The δ1 gene encodes δ1-crystallin, the major structural protein of the chicken lens. On the other hand, the δ2 gene appears to be expressed at very low levels in various tissues and its product, δ2-crystallin, appears to be a non-lens-specific protein of low abundance. Sequence comparison showed that chicken δ-crystallins are highly homologous to human, rat, and yeast argininosuccinate lyase. The sequence homologies between the δ2-crystallin and human, rat, or yeast argininosuccinate lyase are 69%, 69%, and 55%, respectively. The homology between the δ2-crystallin and yeast argininosuccinate lyase (55% identity) is much the same as between the rat or human enzyme and the yeast enzyme (54% identity). These observations suggest that δ2-crystallin is argininosuccinate lyase. In fact, Piatigorsky *et al.* (16) most recently found that the two chicken δ-crystallin genes accounted for sequences in the chicken genome that cross-hybridize with human argininosuccinate lyase cDNA, with preferential hybridization to the δ2 gene. Our present study revealed that the structural organization of rat argininosuccinate lyase gene is close to that of the chicken δ-crystallin genes. All introns except intron 1 of the crystallin genes interrupt the coding regions at precisely the corresponding positions. All of these results provide strong evidence that the δ1 and δ2 genes were formed by duplication of an ancestral argininosuccinate lyase gene, that the δ2 gene has remained to encode argininosuccinate lyase, and that the δ1 gene was specialized for a lens structural protein. Relatively small amounts of argininosuccinate lyase activity found in various tissues of the chicken (15, 27) may reflect low-level expressions of the δ2 gene in these tissues. To determine whether or not δ2-crystallin is indeed argininosuccinate lyase, one may attempt to express the δ2 gene in cultured cells and then observe if the gene product has a reasonable argininosuccinate lyase activity.

Assuming that δ2-crystallin is argininosuccinate lyase, it is to be noted that the sequence homology between chicken δ1-crystallin and δ2-crystallin (argininosuccinate lyase) (91%) is much higher than interspecies homology between the chicken and mammalian argininosuccinate lyase (69%). The chicken δ1 and δ2 genes are also more similar in gene organization than are the δ2 gene and the rat argininosuccinate lyase gene (Fig. 3). Here, it may be that the gene duplication occurred within the period of bird evolution after birds had diverged from reptiles as a separate species. This possibility would be given support if reptiles have only one δ-crystallin or argininosuccinate lyase gene. On the other hand, if the reptiles have two δ-crystallin/argininosuccinate lyase genes, the gene duplication must have occurred before the divergence. It is also possible that two argininosuccinate lyase genes were present in the ancestors of mammals, birds, and reptiles and that one of the genes (perhaps equivalent to δ1 gene) was lost in mammals. In the latter two cases, a

strikingly high homology of the two δ-crystallin genes in birds might reflect "concerted evolution" resulting from "gene conversion."

We thank Dr. H. Kondoh of Nagoya University for valuable discussions and M. Ohara for comments. This work was supported in part by Grants-in-Aid 63618514 and 61870019 from the Ministry of Education, Science and Culture of Japan and by a grant from the Nissan Science Foundation.

1. Lusty, C. J. & Ratner, S. (1972) *J. Biol. Chem.* **247,** 7010–7022.
2. O'Brien, W. E. & Barr, R. H. (1981) *Biochemistry* **20,** 2056–2060.
3. Palekar, A. G. & Mantagos, S. (1981) *J. Biol. Chem.* **256,** 9192–9194.
4. Ratner, S. (1973) *Adv. Enzymol. Relat. Areas Mol. Biol.* **39,** 1–90.
5. Amaya, Y., Kawamoto, S., Oda, T., Kuzumi, T., Saheki, T., Kimura, S. & Mori, M. (1986) *Biochem. Int.* **13,** 433–438.
6. Amaya, Y., Matsubasa, T., Takiguchi, M., Kobayashi, K., Saheki, T., Kawamoto, S. & Mori, M. (1988) *J. Biochem. (Tokyo)* **103,** 177–181.
7. O'Brien, W. E., McInnes, R., Kalumuck, K. & Adcock, M. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7211–7215.
8. Matsuo, S., Tatsuno, M., Kobayashi, K., Saheki, T., Miyata, T., Iwanaga, S., Amaya, Y. & Mori, M. (1988) *FEBS Lett.* **234,** 395–399.
9. Yasuda, K., Kondoh, H., Okada, T. S., Nakajima, N. & Shimura, Y. (1982) *Nucleic Acids Res.* **10,** 2879–2891.
10. Hawkins, J. W., Nickerson, J. M., Sullivan, M. A. & Piatigorsky, J. (1984) *J. Biol. Chem.* **259,** 9821–9825.
11. Ohno, M., Sakamoto, H., Yasuda, K., Okada, T. S. & Shimura, Y. (1985) *Nucleic Acids Res.* **13,** 1593–1606.
12. Nickerson, J. M., Wawrousek, E. F., Hawkins, J. W., Wakil, A. S., Wistow, G., Thomas, G., Norman, B. L. & Piatigorsky, J. (1985) *J. Biol. Chem.* **260,** 9100–9105.
13. Nickerson, J. M., Wawrousek, E. F., Borras, T., Hawkins, J. W., Norman, B. L., Filpula, D. R., Nagle, J. W., Ally, A. H. & Piatigorsky, J. (1986) *J. Biol. Chem.* **261,** 552–557.
14. Borras, T., Nickerson, J. M., Cheperinsky, A. B. & Piatigorsky, J. (1985) *EMBO J.* **4,** 445–452.
15. Wistow, G. & Piatigorsky, J. (1987) *Science* **236,** 1554–1556.
16. Piatigorsky, J., O'Brien, W. E., Norman, B. L., Kalumuck, K., Wistow, G., Borras, T., Nickerson, J. M. & Wawrousek, E. F. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 3479–3483.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
18. Hattori, M. & Sakaki, Y. (1986) *Anal. Biochem.* **152,** 232–238.
19. King, C. R. & Piatigorsky, J. (1983) *Cell* **32,** 707–712.
20. Lusty, C. J. & Ratner, S. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3176–3180.
21. Ingolia, T. D. & Craig, E. A. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 2360–2364.
22. Nene, V., Dunne, D. W., Johnson, K. S., Taylor, D. W. & Cordingley, J. S. (1986) *Mol. Biochem. Parasitol.* **21,** 179–188.
23. Wistow, G., Summers, L. & Blundell, T. (1985) *Nature (London)* **315,** 771–773.
24. Wistow, G., Mulders, J. W. & de Jong, W. W. (1987) *Nature (London)* **326,** 622–624.
25. Carper, D., Nishimura, C., Shinohara, T., Dietzschold, B., Wistow, G., Craft, C., Kador, P. & Kinoshita, J. H. (1987) *FEBS Lett.* **220,** 209–213.
26. Watanabe, K., Fujii, Y., Nakayama, K., Ohkubo, H., Kuramitsu, S., Kagamiyama, H., Nakanishi, S. & Hayaishi, O. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 11–15.
27. Tamir, H. & Ratner, S. (1963) *Arch. Biochem. Biophys.* **102,** 249–258.