The Major Adult α-Globin Gene of Antarctic Teleosts and Its Remnants in the Hemoglobinless Icefishes

CALIBRATION OF THE MUTATIONAL CLOCK FOR NUCLEAR GENES*

(Received for publication, February 2, 1998, and in revised form, March 16, 1998)

Yuqiong Zhao‡‡, Manoja Ratnayake-Lecomwasam‡‡, Sandra K. Parker‡, Ennio Cocca‡, Laura Camardella‡, Guido di Prisco‡, and H. William Detrich III‡‡

From the ‡‡Department of Biology, Northeastern University, Boston, Massachusetts 02115 and the ¶Istituto di Biochimica delle Proteine ed Enzimologia, Consiglio Nazionale delle Ricerche, 80125 Naples, Italy

The icefishes of the Southern Ocean (family Channichthyidae, suborder Notothenioidei) are unique among vertebrates in their inability to synthesize hemoglobin. We have shown previously (Cocca, E., Ratnayake-Lecomwasam, M., Parker, S. K., Camardella, L., Caramella, M., di Prisco, G., and Detrich, H. W., III (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 1817–1821) that icefishes retain inactive genomic remnants of adult notothenioid α-globin genes but have lost the gene that encodes adult β-globin. Here we demonstrate that loss of expression of the major adult α-globin, α1, in two species of icefish (Chaenocephalus aceratus and Chionodraco rastrospinosus) results from truncation of the 5′ end of the notothenioid α1-globin gene. The wild-type, functional α1-globin gene of the Antarctic yellowbelly rockcod, Nototothenia coriiceps, contains three exons and two A + T-rich introns, and its expression may be controlled by two or three distinct promoters. Retained in both icefish genomes are a portion of intron 2, exon 3, and the 3′-untranslated region of the notothenioid α1-globin gene. The residual, nonfunctional α-globin gene, no longer under positive selection pressure for expression, has apparently undergone random mutational drift at an estimated rate of 0.12–0.33%/million years. We propose that abrogation of hemoglobin synthesis in icefishes most likely resulted from a single mutational event in the ancestral channichthid that deleted the entire β-globin gene and the 5′ end of the linked α1-globin gene.

* This work was supported by National Science Foundation Grants OPP-9120311 and OPP-9420712 (to H. W. D.) and by the Italian National Programme for Antarctic Research (to E. C., L. C., and G. d. P.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF049916, AF049914, and AF049915.

‡ These individuals contributed equally to the work described here and should be considered joint first authors.

‖To whom correspondence should be addressed: Dept. of Biology, Northeastern University, 414 Mugar Hall, 360 Huntington Ave., Boston, MA 02115. Tel: 617-373-4495; Fax: 617-373-3724; E-mail: iceman@neu.edu.

The icefishes of the Southern Ocean (family Channichthyidae, suborder Notothenioidei) are unique among vertebrates in their inability to synthesize hemoglobin. We have shown previously (Cocca, E., Ratnayake-Lecomwasam, M., Parker, S. K., Camardella, L., Caramella, M., di Prisco, G., and Detrich, H. W., III (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 1817–1821) that icefishes retain inactive genomic remnants of adult notothenioid α-globin genes but have lost the gene that encodes adult β-globin. Here we demonstrate that loss of expression of the major adult α-globin, α1, in two species of icefish (Chaenocephalus aceratus and Chionodraco rastrospinosus) results from truncation of the 5′ end of the notothenioid α1-globin gene. The wild-type, functional α1-globin gene of the Antarctic yellowbelly rockcod, Nototothenia coriiceps, contains three exons and two A + T-rich introns, and its expression may be controlled by two or three distinct promoters. Retained in both icefish genomes are a portion of intron 2, exon 3, and the 3′-untranslated region of the notothenioid α1-globin gene. The residual, nonfunctional α-globin gene, no longer under positive selection pressure for expression, has apparently undergone random mutational drift at an estimated rate of 0.12–0.33%/million years. We propose that abrogation of hemoglobin synthesis in icefishes most likely resulted from a single mutational event in the ancestral channichthid that deleted the entire β-globin gene and the 5′ end of the linked α1-globin gene.

Alone among vertebrate taxa, the 15 species of Antarctic icefishes (family Channichthyidae, suborder Notothenioidei) are unique in their failure to synthesize the respiratory oxygen transport protein hemoglobin (1). Although icefish blood contains “erythrocyte-like” cells in small numbers (2, 3), these cells are devoid of hemoglobin, and icefishes transport oxygen to their tissues solely in physical solution. In the cold (−1.86 to +1 °C), stable, and oxygen-rich environment experienced by these organisms, reduction of the hematocrit to near zero may have been selectively advantageous because it significantly diminishes the energetic cost associated with circulation of a highly viscous, corpuscular blood fluid (4–7). Indeed, hematocrit, mean cellular hemoglobin concentration, and hemoglobin chain multiplicity all decrease with increasing phylogenetic divergence among the red-blooded Antarctic notothenioid fishes (8), and the Bathymedusidae (the sister group to the channichthyids) approach the hematological extreme displayed by the white-blooded icefishes. Furthermore, di Prisco et al. (9) have shown that a red-blooded nototheniid fish, Pagotheria bernacchii, survives under resting conditions when its hemoglobin is converted to the carbon monoxide form in vivo. Apparently, red-blooded Antarctic fishes can sustain their basal metabolism using plasma-dissolved oxygen, and they draw on their hemoglobin stores primarily when respiratory demand increases. Nevertheless, the development in icefishes of compensatory physiological and circulatory adaptations that reduce tissue oxygen demand and enhance oxygen delivery (e.g. modest suppression of metabolic rates, enhanced gas exchange by large, well-perfused gills and through a scaleless skin, and large increases in cardiac output and blood volume) argues that loss of hemoglobin and erythrocytes was maladaptive under conditions of physiological stress. Thus, the most plausible evolutionary scenario is that the phylogenetic trend to reduced hematocrits and decreased hemoglobin synthesis in notothenioid fishes developed concurrently with enhancements to their respiratory and circulatory systems, leading ultimately to the acellular, hemoglobinless condition of the icefishes.

The channichthids diverged from other Antarctic notothenioids approximately 7–15 million years ago, but radiation of species within the icefish clade (i.e. lineage branch) appears to have been confined to the last one million years (10). Recently, we demonstrated that icefish species belonging to both primitive and advanced genera retain in their genomes inactive remnants of the major adult notothenioid α-globin gene but have lost the gene that encodes adult β-globin (11). Thus, the hemoglobinless phenotype appears to be a primitive channichthid character that was established by deletion or rapid mutation of the gene encoding β-globin before diversification of the clade. Our present objective is to determine the evolutionary fate of the channichthid α1-globin gene. In this report we describe the structures of the functional α1-globin gene of the red-blooded Antarctic rockcod Nototothenia coriiceps (family Nototheniidae) and of the α-globin gene remnants of two icefishes, Chaenocephalus aceratus and Chionodraco rastrospinosus. To our surprise, we find that the icefish α-globin gene is a truncated version of the α1-globin gene of red-blooded notothe-
a-Globin Gene Remnants in Antarctic Icefishes

noids. These remnants, which contain the 3′ portion of intron 2, all of exon 3, and the 3′-untranslated region of the a-globin gene, appear to be mutating randomly. Using transversion substitutions, we estimate that these nonfunctional nuclear gene fragments are diverging at the rate of 0.12–0.33%/million years. Because the notothenioid adult globin genes are tightly linked in 5′ to 5′ orientation, we now propose that the hemoglobinless phenotype was established by a single deletional event in the ancestral channichthyid that eliminated the entire β-globin gene and the 5′ half of the a-globin gene.

EXPERIMENTAL PROCEDURES

Collection of Animals and Storage of Tissues—Specimens of N. coriiceps, C. aceratus, and C. rastrospinosus were collected by bottom-trawling from the R/V Hero or from the R/V Polar Duke south of Low Island (Antarctic Treaty Protected Area System Marine Site of Special Scientific Interest (MSSSI) 35, Western Bransfield Strait) or west of Brabant Island (MSSSI 36, East Dallmann Bay). Fishes were transported alive to Palmer Station, Antarctica, where they were maintained in seawater aquaria at −1 to +1 °C. Blood from mature adults was collected into heparinized syringes by caudal venipuncture of unanesthetized specimens, and erythrocytes, isolated by differential centrifugation (12), were either used immediately or frozen and stored at −70 °C. Testes were dissected immediately after sacrifice of mature males, frozen in liquid nitrogen, and maintained at −70 °C until use.

Construction and Screening of Genomic Libraries—Genomic DNA of high molecular weight (>50 kilobase pairs) was purified from a single testis of each fish species by the method of Blin and Stafford (13). The N. coriiceps library was constructed in the λ vector Charon 35 (14). DNA was digested partially with MboI, and fragments of 15–20 kilobase pairs, obtained by sucrose gradient centrifugation, were ligated to the BamHI sites of the vector arms. Recombinant phage DNA was packaged in vitro (Packagene; Promega), and phage were amplified in Esherichia coli K802 to yield a stock at 8 × 10^10 plaque-forming units/ml. The C. aceratus library was constructed in LambdaGEM-11 (Promega). DNA was digested partially with MboI, fragments were ligated to phage arms containing XhoI half-sites, and recombinant phage DNA was packaged in vitro as described above. The latter library was propagated in E. coli KW251 to give an amplified stock with a titer of 2.4 × 10^7 plaque-forming units/ml.

Amplified C. aceratus and C. rastrospinosus genomic libraries (titer of 10^7–10^8) in Lambda Fix II (Stratagene) were generously provided by Drs. Michael Vayda and Bruce Sidell (University of Maine, Orono, ME). These libraries were propagated in E. coli XL1-Blue MRA (Stratagene).

Libraries were screened independently for clones containing a-globin gene sequences by hybridization of nylon (MagnaLift, MSI, Westboro, MA) replicates of bacteriophage plaque DNA to the N. coriiceps a-globin cDNA, NCbHb1-1 (11). 32P-labeled by random priming (15), approximately 300,000–400,000 phage were plated for each primary screen. Prehybridization and hybridization of the membranes were performed at moderate stringency (6 × SSC (1 × SSC = 0.15 M NaCl, 0.015 M trisodium citrate), 0.1% sodium pyrophosphate, 50 μg/ml heparin, 0.2% SDS, 62 °C) for 2 and 12–16 h, respectively. After hybridization, the membranes were washed twice in 3 × SSC, 0.1% SDS at 62 °C for 15 min each and then exposed to Fuji RX x-ray film at −70 °C for 16–34 h with intensification (DuPont Cronex Lightning Plus screens). Three positive N. coriiceps phage isolates, two from C. aceratus, and three from C. rastrospinosus were obtained after two additional rounds of plaque purification and screening, and single plaques were picked for clone stock preparation.

Subcloning and DNA Sequencing—Tertiary isolates were grown and DNA was extracted by the plate-lysis method (16). Genomic DNA inserts were excised from the recombinant phage by digestion with appropriate restriction endonucleases, the DNA digests were electrophoresed on 1% (w/v) agarose gels in 1× TBE (0.089 M Tris borate, 2 mM EDTA, pH 8.0), and DNA fragments containing a-globin gene sequences were transferred to nylon membranes (17) followed by hybridization of the gel replicates with 32P-labeled NCbHb1-1 (see “Construction and Screening of Genomic Libraries”). Positive fragments were excised from duplicate gels, DNA was recovered from low-melt agarose or by SPIN-X column (Costar Corp.) centrifugation, and the fragments were ligated into the multicloning sites of pUC19 or pBluescript (Stratagene). Recombinant plasmids were transformed into competent E. coli XL-1 Blue cells, the cells were grown on LB/vancomycin plates containing isopropyl-1-thio-β-D-galactopyranoside and 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside, and white colonies were picked for further analysis. Plasmid DNA was purified by alkaline-lysis plasmid mini-prep (16).

Parental clones and restriction-fragment subclones were sequenced manually on both strands by use of the dyeoxynucleotide chain termination method (18) and T4 DNA polymerase (Sequenase II; U. S. Biochemical Corp.). Both T3 and T7 universal primers and gene-specific primers were employed. Portions of the sequence (Fig. 1) were established by use of the PRISM Ready Reaction Dye Deoxy Termination Cycle Sequencing Kit (Applied Biosystems), and the products were resolved using an Applied Biosystems 373A automated DNA sequencer (University of Maine DNA Sequencing Facility).

DNA Sequence Analysis—Genomic a-globin sequences were aligned (k-tuple = 3, gap penalty = 3, largest gap = 50) by use of the algorithm of Wilbur and Lipman (19), and sequence similarities were calculated by the method of Dayhoff (20), as implemented by DNASTAR Align.

GenBank Accession Numbers—The Antarctic notothenioid a-globin gene sequences reported in this paper have been deposited in the GenBank data base under the following accession numbers: N. coriiceps, GenBank accession numbers AF049914; C. aceratus, GenBank accession numbers AF049915. These sequences have been scanned against the GenBank data base, and representative a-globin genes from other vertebrates are reported in Table I.

Erythocyte RNA Isolation—Total RNA from erythrocytes of adult N. coriiceps was purified from red blood cells by a modification (21) of the acid guanidinium isothiocyanate/phenol/chloroform method (22).

Mapping of a-Globin Transcription Start Sites by Primer Extension—Three antisense oligonucleotide primers derived from the 5′ end of the N. coriiceps a-globin gene (Fig. 2) were used to map potential transcription start sites. Primers were labeled at their 5′ termini by reaction with [γ-32P]ATP and T4 polynucleotide kinase (23). Labeled primers were annealed to total RNA (30 μg/reaction) from erythrocytes of N. coriiceps. Primer extension reactions were performed for 60 min using 25 units/μl avian myeloblastosis virus reverse transcriptase and 10 μM concentrations of the four dNTPs, and single-stranded RNA was degraded by treatment with RNase A. The products of each primer extension reaction were co-electrophoresed on a 9% polyacrylamide, 7 M urea DNA sequencing gel with sequencing ladders derived from the same primer.

RESULTS

Structure of the a-Globin Gene of N. coriiceps—To establish a framework for genetic comparison, we isolated and characterized the adult a-globin gene from the hemoglobin-expressing nototheniod N. coriiceps. After restriction digestion with EcoRI, three independent N. coriiceps genomic phage clones each gave a 5.2-kilobase pair a-globin-positive DNA fragment. This fragment was subcloned into pUC19 to yield the clone NCbHbG1. Fig. 1 presents the structural diagram (panel A) of and the strategy employed to sequence (panel B) the a-globin gene of NCbHbG1. Fig. 2 shows its nucleotide sequence and translation. The coding sequence and 5′- and 3′-untranslated regions of the NCbHbG1 a-globin gene exactly match the corresponding regions of the N. coriiceps a-globin cDNA (11), thus demonstrating that this gene encodes the major adult a-globin expressed by nototheniid fishes. The 432-bp 2 coding sequence is interrupted by two introns, the first (11) splitting codon 32 and the second (12) separating codons 101 and 102. These positions are conserved within a few nucleotides among all vertebrate a-globin genes studied to date (24). Both introns are substantially longer than their counterparts in the adult a-globin genes of other fishes, amphibians, or mammals (24–27) and are extremely rich in A + T residues. Intron 1 is 450 bp in length and 70% A + T, whereas intron 2 is 278 bp long and 64% A + T. Intron 1 also contains a 24-bp tandem repeat of the dinucleotide TA. The splice junctions of both introns conform to

1 A. Saeed, D. Lau, and H. W. Dietrich III, manuscript in preparation.

2 The abbreviation used is: bp, base pairs.
α-Globin Gene Remnants in Antarctic Icefishes

---

**Fig. 1. Sequencing strategy for the N. coriiceps α-globin gene.**

A, schematic representation of the notothenioid α-globin gene. Filled and open rectangles denote the three coding exons (E1-E3) and two introns (I1, I2), respectively. The 5′-upstream and 3′-downstream (3′-down) regions are indicated by vertically and diagonally hatched rectangles, respectively. B, sequence analysis by primer walking. Arrows indicate direction and extent of sequencing. The two long sense-strand reads were obtained by automated sequencing of a polymerase chain reaction-generated clone that links the adult α- and β-globin genes of *N. coriiceps* and overlaps the α clone described here.1

---

The GT/AG rule, containing GT at their 5′ donor junctions and CAG preceded by pyrimidine-rich tracts at their 3′ acceptor sites (28–30). The 5′-upstream and 3′-downstream sequences of the α-globin gene are also A + T-rich (58 and 61%, respectively), but the coding sequence is A + T-poor (44% A + T). Thus, the adult α-globin gene of *N. coriiceps* shares the intron/exon organization of other vertebrate α-globin genes, but its content of A + T nucleotides in noncoding regions (introns, 5′-upstream and 3′-downstream sequences) is high (see “Discussion”).

The *N. coriiceps* α-globin gene contains cis-acting regulatory elements comparable to those of globin genes in other vertebrates. Three potential transcription start (capping) sites were detected 38-bp (CAP 1; nucleotide designated +1), 380-bp (CAP 2), and 571-bp (CAP 3) upstream of the initiator codon (Fig. 2). Sites 1 and 2 begin with the AC dinucleotide common to other globin transcription initiation sites (24), whereas the most distal site 3 starts with AT. These putative cap sites are consistent with the size heterogeneity observed for *N. coriiceps* α-globin mRNAs (11). Immediately upstream of the cap sites are clusters of basal and hematopoietic promoter elements (Fig. 2), including probable TATA and CCAAT boxes, and GATA, CACCC, and NF-E2 motifs (31–34). Presumptive, but noncanonical, TATA boxes begin 18, 44, and 22 bp upstream of cap sites 1–3, respectively, and potential CCAAT boxes are located 72 and 165, 92, and 40 bp upstream of the three initiation sites. Four CACCC motifs, one GATA motif, and one NF-E2 element precede cap site 1, whereas two GATAs and one CACCC are proximal to site 2. Start site 3 is adjacent to two GATA elements but apparently lacks the CACCC motif. The functional analysis of both α- and β-globin promoters from *N. coriiceps* will be presented elsewhere.3 Finally, two consensus sequences for polyadenylation are present in the 3′-untranslated region, the second of which precedes the poly(A) tail of the α-globin cDNA by 14 nucleotides (11).

α-Globin Gene Remnants from Two Hemoglobinless Antarctic Icefishes

---

In contrast to the 2–3 α-globin-hybridizing DNA fragments observed in the genomes of red-blooded notothenioid fishes, the genomes of three icefishes, *C. aceratus*, *C. rastrospinosus*, and *Champsocephalus gunnari*, contain single DNA fragments that hybridize at reduced intensity to a notothenioid adult α-globin cDNA probe (11). One plausible interpretation of this observation is that the icefishes have lost, probably by deletion, a portion of the ancestral notothenioid adult α-globin gene. To investigate this possibility, we isolated and sequenced two independent α-globin-positive genomic clones from *C. aceratus* and one clone from *C. rastrospinosus*. Fig. 3 compares the structure of the icefish α-globin fragments to that of the *N. coriiceps* α-globin gene. The α-globin genes of the two icefishes contain a portion of intron 2, the entirety of exon 3, and all of the 3′-untranslated region. The apparent 5′ chromosomal breakpoint within intron 2 (between positions 931 and 932; see also Fig. 4) is identical in the two icefish genes, and the 5′-flanking sequences preceding the breakpoint bear no relation to any portion of the *N. coriiceps* α-globin gene. Beyond the second polyadenylation signal, the icefish and rockcod genes share sequence similarity for at least 180 bp (Fig. 4). These observations are consistent with deletional loss of 5′-upstream α-globin sequences (5′-untranslated region, exons 1 and 2, intron 1 and part of intron 2) before divergence of these two relatively advanced icefish species. Determination of the status of the α-globin gene in the ancestral channichthyid will require analysis of more primitive icefish species.

Fig. 4 shows the alignment of the α-globin gene remnants of *C. aceratus* and *C. rastrospinosus* with the corresponding region of the α-globin gene of *N. coriiceps*. In regions of overlap, the *C. aceratus* and *N. coriiceps* sequences are 96.4% similar, whereas those of *C. rastrospinosus* (the more advanced icefish) and *N. coriiceps* are 95.1% similar. The two icefish sequences in turn share 98.6% similarity. Although strongly related to the *N. coriiceps* sequence, the icefish α-globin remnants share a number of deletions, insertions, and point mutations with respect to the former. Particularly prominent in the icefishes is the 16-nucleotide deletion within the 3′-untranslated region at *N. coriiceps* position 1286 and the five nucleotide insertion after *N. coriiceps* position 1441, which occurs beyond the second polyadenylation signal. Furthermore, the sequences of the *C. aceratus* and *C. rastrospinosus* remnants contain 14 and 17 nucleotide substitutions, respectively, relative to the *N. coriiceps* α-globin sequence. Seven substitutions shared by the icefishes occur within exon 3, but no nonsense codons are introduced into the coding sequence. Transversions exceed transitions in both icefish species (11 transversions and 3 transitions for *C. aceratus*, 14 and 3 for *C. rastrospinosus*). The two icefish remnants have also diverged from each other, as shown by the two-nucleotide insertion and three nucleotide substitutions in the *C. rastrospinosus* sequence between positions 1010 and 1016 and the single nucleotide deletion after position 1398.

The inability to detect the 5′ segment of the α-globin gene in the icefish genomic clones might result from expansion of globin intron 2 in the icefish lineage. This appears unlikely given the failure to isolate clones containing the 5′ region in the large scale primary screens (>99% probability of detection for a single copy sequence; e.g. more than 3,800,000 total plaques were screened for *C. rastrospinosus*). Nevertheless, in an effort to recover the absent 5′ α-globin gene sequences, the *C. aceratus* and *C. rastrospinosus* genomic libraries were screened with a polymerase chain reaction-generated probe encompassing exons 1 and 2 of the *N. coriiceps* gene. Despite extensive screening (87–99% probability of detection for a single copy sequence; more than 200,000 plaques were screened for *C. rastrospinosus* and 480,000 for *C. aceratus*), no positive clones
were obtained from either library. Failure to recover the 5‘-end portion of the α1-globin gene from the icefish genomes is consistent with the proposed deletional loss of this region. Calibration of the Mutational Clock for Notothenioid Nuclear Genes in the Absence of Selective Pressure—The results presented above strongly suggest that failure of the icefishes to synthesize α1-globin is due to deletional loss of the 5‘-end of the α1-globin gene. The residual icefish α1-globin gene, no longer under positive selection pressure, subsequently experienced random mutational drift. Using the mitochondrial DNA-based age of radiation of the notothenioid fishes, 7–15 million years ago (10), we can estimate the rate of icefish nuclear gene divergence from the frequency of transversion substitutions in the α1-globin gene remnant (35, 36). C. acerai-

**FIG. 2.** Nucleotide sequence of the *N. coriiceps* α1-globin gene (*NcHbGα*) and deduced amino acid sequence (AA) of the encoded α1-globin. Coding sequences are indicated in uppercase roman text, and the encoded amino acid residues are presented below in the three-letter code. Introns are shown in lowercase with nucleotides at the intron/exon junctions in boldface. 5‘-Untranslated and -upstream sequences are given in lowercase italics, whereas 3‘-untranslated and -downstream sequences are given in lowercase. Three transcription initiation (CAP 1–3) sites, mapped by extension from primers 1, 2, and 3, are indicated in boldface text and by arrows. Potential TATA, CCAAT, GATA, CACCC, and NF-E2 promoter elements are shown in reversed text, boldface, shadowed text, double-underlined text, and single-underlined text, respectively. The Kozak sequence for translation initiation (48) is given in underlined boldface. The initiator methionine (encoded by the AUG codon at +39) is cleaved from the mature protein. Two potential polyadenylation signal sequences in the 3‘-untranslated region are underlined. Inton 1 contains a TA tandem repeat (underlined).
adapted notothenioids (4, 36).

Consistent with the low specific metabolic rates of the cold million years. This slow rate of nuclear gene divergence is the absence of selective pressure falls in the range 0.12–0.33%/11 and 14 transversions, respectively, in this fragment. There-remarkably similar to the

tus and the more advanced C. rastrosinosus have accumulated 11 and 14 transversions, respectively, in this fragment. Therefore, the estimated rate of icefish nuclear gene divergence in the absence of selective pressure falls in the range 0.12–0.33%/million years. This slow rate of nuclear gene divergence is consistent with the low specific metabolic rates of the cold-adapted notothenioids (4, 36).

**FIG. 3.** Comparison of the α-globin gene remnants of C. acetatus and C. rastrosinosus to the α-globin gene of N. coriiceps. Exons (E1–E3) and introns (I1, I2) are represented by filled and open rectangles, respectively, and the 5'-upstream and 3'-downstream regions are indicated by hatched rectangles. The 5'-flanking sequences of the icefish gene remnants are shown by stippled boxes. Box dimensions encompass the extent of sequencing of each gene. kb, kilobase pairs.

**DISCUSSION**

In this report we have demonstrated that the α-globin genes of two icefishes are partial 5'-truncated variants of the α1-globin gene of red-blooded notothenioid fishes. Furthermore, the residual icefish gene fragments have accumulated deletions, insertions, and nucleotide substitutions with respect to the functional globin gene. The striking similarity of the icefish remnants to each other, both in the apparent chromosomal breakpoint and in other mutations, strongly suggests that most of the changes evolved in the ancestral channichthyid approximatively 7–15 million years ago. The relatively minor differences between the gene remnants of C. acetatus and C. rastrosinosus probably evolved by mutational drift during the radiation of the icefishes over the past one million years. The apparent random mutation of the icefish α1 remnant should provide a useful tool for development of a molecular phylogeny of icefishes based on nuclear gene divergence.

**Structure of a Functional Notothenioid α-Globin Gene**—The structure of the functional α1-globin gene of N. coriiceps is remarkably similar to the α1-globin genes of other vertebrates. Intron positions have been maintained, and splice junctions conform to the GT/AG rule (28–30). Thus, the basic splicing mechanism is likely to have been conserved in the cold-adapted Antarctic fishes. The possibility that compensatory adaptations are required to permit efficient splicing in their cold thermal regime remains open.

Surprisingly, three potential transcription start sites were detected for the α1-globin gene. The two cap sites most proximal to the α1-coding sequence (CAP 1, CAP 2) are associated both with basal TATA and CCAAT boxes and with hematopoietic CACCC and GATA promoter elements. A single NF-E2 site precedes CAP 1. The third, most distal site (CAP 3) contains a putative TATA box but lacks the CACCC element. Like many other globin genes (24, 31), the sequences of the TATA boxes of the N. coriiceps promoters deviate from the vertebrate canon, TATAAA (33). The use of alternative promoters and multiple transcription initiation sites to regulate gene expression in eukaryotes is now recognized to be a common phenomenon (37). For example, Hu et al. (38) have shown that tissue-specific expression of prolactin receptors in the rat is controlled by three promoters, two of which contain noncanonical TATA elements. We propose that the 5'-upstream sequence of the N. coriiceps α1-globin gene has evolved multiple promoters to offset slow rates of pol II transcription at psychrophilic temperatures by simultaneous recruitment of multiple transcriptional complexes. Indeed, the adult α1/β-globin intergenic region of N. coriiceps stimulates strongly, in either orientation, the transcription of a luciferase reporter construct in differentiated MEL cells.

Other departures of the notothenioid α-globin gene from the norm in higher vertebrates entail the length and nucleotide composition of the two introns. Both are longer and very rich in A + T residues. The 5'- and 3'-noncoding sequences of the gene are similarly A + T-rich. Table I presents a comparison of the nucleotide compositions of these subregions in α-globin genes from cold-living, temperate, and warm-bodied vertebrates. It is immediately evident that the noncoding regions of the cold-bodied and temperate vertebrates are significantly richer in A + T residues than are the corresponding regions in warm-bodied organisms. By contrast, the coding sequences, which are presumably constrained by functional requirements of the encoded globin polypeptides, are relatively A + T-poor and quite similar in composition in all taxa. The introns of several other notothenioid genes (myoglobin (39), the dynein heavy chain (40), and three α-tubulins3 are also rich in A + T residues (61–71%). What functional significance can we attribute to the biased nucleotide compositions of genes from cold- and warm-bodied vertebrates?

Bernardi et al. (41–43) have demonstrated that the genomes of warm-bodied homeotherms (mammals, birds, etc.) are organized heterogeneously, with G + C- and gene-rich “isochores” interspersed with regions of lower G + C content. Among poikilotherms, warm-bodied (37–40 °C) Tilapia species also contain G + C-rich isochores, whereas their temperate congener lack them (42). High GC isochores may have evolved to maintain genome stability, to protect genes against DNA “breathing” and mutability (41), and to enhance mRNA and protein stability (42) at relatively high body temperatures. Conversely, the A + T-rich genomes of cold-living and temperate poikilotherms may facilitate DNA strand separation during transcription and replication in low temperature regimes (39). At the extreme temperatures experienced by Antarctic fishes, the latter advantage may be particularly important.

**Globin Gene Organization and Expression in Fishes**—In contrast to the distinct α- and β-globin gene clusters of higher vertebrates, the adult α- and β-globin genes of N. coriiceps, its temperate relative, N. angustata (the New Zealand black cod), and presumably other red-blooded notothenioids are tightly linked in 5’ to 5’ orientation.3 Head-to-head linkage of α/β gene pairs has also been observed in the Atlantic salmon (44), carp (45), and the zebrafish (27). Thus, this gene organization probably represents the ancestral condition of gnathostome fishes

---

3 S. K. Parker and H. W. Detrich III, unpublished results.
and is likely to have arisen by duplication, inversion, and divergence of the primordial globin gene of primitive jawless fishes (e.g. the lamprey) to give an α/β gene pair (27). One plausible advantage of head-to-head linkage is coordinate regulation of globin gene transcription mediated by shared promoter and/or enhancer elements located in the intergenic sequences (27).

In higher vertebrates, expression of the distinct α- and β-globin genes is regulated by separate promoters and enhancers located in the intergenic regions. This allows for the coordinated expression of these genes, which is important for the proper oxygen transport in the blood.

FIG. 4. Sequence alignment of the α1-globin gene remnants of C. aceratus and C. rastrospinosus to the α1-globin gene of N. coriceps. The icefish sequences (Ca-α1, Cr-α1) are numbered to correspond to the region of the N. coriceps α1-globin gene (Nc-α1) that begins at nucleotide 877, and the similarity of the three sequences commences at residue 932. Residues identical to the N. coriceps sequence (or to the C. aceratus sequence after position 1543) are indicated by periods. Nucleotides at positions that differ among the three fishes are shown in boldface, and those residues that differ from the consensus are underlined. Dashes indicate deletions. Exon 3 is enclosed by the large box.

### Table I

| Organism       | Gene  | GenBank accession No. | Tbody | Subregion of α-globin gene |
|----------------|-------|-----------------------|-------|---------------------------|
|                |       |                       |       | 5'-Upstream | Coding | Intron 1 | Intron 2 | 3'-Downstream |
| N. coriceps    | α1    | AF049916               | C     | 58 (678) | 44 (432) | 70 (450) | 64 (278) | 61 (357) |
| Carp           | α7    | AB004739               | C/T   | 63 (292) | 46 (432) | 68 (159) | 68 (92)  | 61 (157) |
| Salmon         | α1    | X97289                 | C/T   | 64 (640) | 46 (444) | 71 (225) | 62 (154) | 61 (536) |
| Zebrafish      | αA    | U50382                 | T     | 72 (704) | 46 (347) | 74 (150) | 61 (96)  | ND      |
| Xenopus        | αA    | X14259                 | T     | 53 (43)  | 54 (432) | 46 (162) | 70 (341) | 68 (235) |
| Chicken        | αA    | V00410                 | W     | 37 (338) | 41 (429) | 32 (133) | 29 (111) | 39 (213) |
| Human          | α2    | V00488                 | W     | 25 (154) | 35 (429) | 27 (117) | 28 (160) | 37 (318) |
| Mouse          | a     | V00714                 | W     | 47 (400) | 42 (435) | 51 (122) | 37 (140) | 47 (352) |

|                |       |                       |       | %A + T (length in bp) |
| N. coriceps    | α1    | AF049916               | C     | 68 (678) |
| Carp           | α7    | AB004739               | C/T   | 44 (432) |
| Salmon         | α1    | X97289                 | C/T   | 68 (159) |
| Zebrafish      | αA    | U50382                 | T     | 46 (347) |
| Xenopus        | αA    | X14259                 | T     | 54 (432) |
| Chicken        | αA    | V00410                 | W     | 41 (429) |
| Human          | α2    | V00488                 | W     | 35 (429) |
| Mouse          | a     | V00714                 | W     | 42 (435) |

*α* Body temperature: C, cold; T, temperate; W, warm.

*β* Coding and 3'-downstream (ND) sequences are incomplete.

and is likely to have arisen by duplication, inversion, and divergence of the primordial globin gene of primitive jawless fishes (e.g. the lamprey) to give an α/β gene pair (27). One plausible advantage of head-to-head linkage is coordinate regulation of globin gene transcription mediated by shared promoter and/or enhancer elements located in the intergenic sequences (27).

In higher vertebrates, expression of the distinct α- and β-glo-
bin gene complexes is regulated during development by locus control regions that govern the sequential, stage-specific expression of embryonic, fetal, and adult globin genes (32). Although locus control regions have not yet been described for fish globin genes, the developmental switching of globin genes in the phylogenetically ancient lamprey suggests that such elements exist throughout the piscine taxon (46). In zebrafish, paired embryonic globin genes are linked to the adult a/b pairs (27). Recently, we isolated an N. coriiceps genomic clone that contains two non-adult a/b-globin gene pairs, with each a/b pair linked 5' to 5' (We do not, as yet, know whether this presumptive embryonic/juvenile globin complex is contiguous to the adult a/b gene pair.) The common organization of embryonic, juvenile, and adult globin genes in several teleost fishes suggests that a single, primordial locus control region may have evolved to regulate developmental switching of pairs of globin genes before the development of distinct α- and β-globin gene complexes, each with its own locus control region, in higher vertebrates.

Mechanism of Globin Gene Loss in Antarctic Icefishes—With the demonstration that icefishes have undergone deletion of the adult β-globin gene (11) and 5' truncation of the major adult α-globin gene (this work), we can now propose a simple mechanism for globin gene loss. Fig. 5 shows that a single deletional event (scenario X) in the ancestral channichthyid, with chromosomal breakpoints located within intron 2 of the α-globin gene and downstream of the 3'-untranslated region of the β-globin gene, would abrogate expression of adult globin polypeptides. This mechanism is supported strongly by the observation that the residual globin loci of C. aceratus and C. rastrosinosus share common 5' breakpoints with respect to the N. coriiceps α-globin gene. Multiple deletions (Y1, Y2) occurring before diversification of the icefish clade that together yield the disrupted icefish globin locus are a formal but less likely possibility. The evolutionary fate of embryonic and juvenile globin genes in icefish genomes is unknown, but the linkage of embryonic and adult globin genes in the zebrafish (27) raises the possibility that the single deletional event postulated here (Fig. 5, scenario Z), or perhaps multiple events, may have removed almost the entire nototheniod globin gene complex.

Divergence of the Icefish α-Globin Gene Remnants and the Mutational Clock—Martin and Palumbi (36), summarizing nucleotide divergence rates among diverse taxonomic groups, have suggested that specific metabolic rate is the major parameter controlling the mutational clock. The rate of mutational change appears to be mediated by reactive oxygen species, generated metabolically, that can damage DNA either directly or indirectly. The data presented here indicate that rates of nuclear gene divergence in notothenioid fishes (0.12–0.33%/million years) in the absence of selective pressure are among the smallest observed in poikilotherms, in agreement with their low specific metabolic rates (4). Thus, the “nucleotide generation time” (average interval for a nucleotide to be copied through replication or repair) of the nuclear genes of Antarctic teleosts is likely to be long (36). The low mutational rates that we have estimated for the α-globin gene remnants of C. aceratus and C. rastrosinosus should be verified by analysis of these fragments in other icefish species and by examination of additional gene families in other notothenioid fishes. Nevertheless, our results suggest that the chronology of evolution of antifreeze glycoprotein genes from the notothenioid trypsinogen gene, estimated at 5–14 million years ago (47) based on mitochondrial divergence rates (0.5–0.9%/million years) of the salmon (36), may require reappraisal.

Acknowledgments—We gratefully acknowledge the excellent logistic support provided to our field research team at Palmer Station and on board the R/V Polar Duke by the personnel of Antarctic Support Associates, by the captains and crews of the R/V Polar Duke, and by the staff of the Office of Polar Programs of the National Science Foundation. We also acknowledge Patricia Singer (University of Maine DNA Sequencing Facility) for her excellent technical assistance in automated DNA sequencing.

REFERENCES
1. Ruud, J. T. (1954) Nature 173, 848–850
2. Hureau, J. C., Petit, D., Fine, J. M., and Marneux, M. (1977) in Adaptations Within Antarctic Ecosystems (Llano, G. A., ed) pp. 459–477, Smithsonian Institution, Washington, D. C.
3. Barber, D. L., Mills Westermann, J. E., and White, M. G. (1981) J. Fish Biol. 19, 11–28
4. Eastman, J. T. (1993) Antarctic Fish Biology: Evolution in a Unique Environment, Academic Press, San Diego, CA
5. D’Avinio, R., Caruso, C., Camardella, L., Schinia, M. E., Rutigliano, B., Romano, M., Carratore, V., Barra, D., and di Prisco, G. (1991) in Life Under Extreme Conditions. Biochemical Adaptation (di Prisco, G., ed) pp. 15–33, Springer-Verlag, Berlin
6. di Prisco, G., D’Avinio, R., Caruso, C., Tamburini, M., Camardella, L., Rutigliano, B., Carratore, V., and Romano, M. (1991) in Biology of Antarctic Fish (di Prisco, G., Maresca, B., and Tota, B., eds) pp. 263–281, Springer-Verlag, Berlin
7. Macdonald, J. A., Montgomery, J. C., and Wells, R. M. G. (1987) Adv. Mar. Biol. 24, 321–338
8. di Prisco, G. (1998) in Fishes of Antarctica. A Biological Overview (di Prisco, G., Pisano, E., and Clarke, A., eds) Springer-Verlag, Berlin, in press
9. di Prisco, G., Macdonald, J. A., and Brunori, M. (1992) Epitestula (Basel) 48, 473–475
10. Bargelloni, L., Ritchie, P. A., Patarnello, T., Battaglia, B., Lambert, D. M., and Meyer, A. (1994) Mol. Biol. Evol. 11, 854–863
11. Coca, E., Ratnayake-Lecamwasam, M., Parker, S. K., Camardella, L., Ciaramella, M., di Prisco, G., and Detrich, H. W., III (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 1817–1821
12. D’Avinio, R., and di Prisco, G. (1989) Eur. J. Biochem. 170, 699–705
13. Loenen, W. A. M., and Blattner, F. R. (1983) Nucleic Acids Res. 23, 503–517
14. Wilbur, W. J., and Lipman, D. J. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 726–739
15. Dayhoff, M. O. (1979) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington, D. C.
16. Puisant, C., and Houdebine, L. M. (1990) Biotechniques 8, 148–149
17. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156–159
18. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
19. Green, J. S., and Struhl, K., eds (1995) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
20. Southern, E. M. (1975) J. Mol. Biol. 98, 503–517
21. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
22. Wilbur, W. J., and Lipman, D. J. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 726–739
23. Dayhoff, M. O. (1979) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington, D. C.
24. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156–159
25. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
26. Wilbur, W. J., and Lipman, D. J. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 726–739
27. Dayhoff, M. O. (1979) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington, D. C.
28. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156–159
29. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
20. Wilbur, W. J., and Lipman, D. J. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 726–739
21. Dayhoff, M. O. (1979) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington, D. C.
22. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156–159
23. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
24. Wilbur, W. J., and Lipman, D. J. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 726–739
25. Patient, R. K., Elkington, J. A., Kay, R. M., and Williams, J. G. (1980) Cell 20, 565–573
26. Miyata, M., Hirono, I., and Aoki, T. (1993) Bull. Jpn. Soc. Sci. Fish. 59, 489–494.
α-Globin Gene Remnants in Antarctic Icefishes

27. Chan, F.-Y., Robinson, J., Brownlie, A., Shivdasani, R. A., Donovan, A., Brugnara, C., Kim, J., Lau, B.-C., Witkowska, E., and Zon, L. I. (1997) Blood 89, 688–700
28. Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383
29. Keller, E. B., and Noon, W. A. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 7417–7420
30. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., and Sharp, P. A. (1986) Annu. Rev. Biochem. 55, 1119–1150
31. Myers, R. M., Tilly, K., and Maniatis, T. (1986) Science 232, 613–618
32. Orkin, S. H. (1990) Cell 63, 665–672
33. Bucher, P. (1990) J. Mol. Biol. 212, 563–578
34. Strauss, E. C., and Orkin, S. H. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 5809–5813
35. Martin, A. P., Naylor, G. J. P., and Palumbi, S. R. (1992) Nature 357, 151–155
36. Martin, A. P., and Palumbi, S. R. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 4087–4091
37. Ayoubi, T. A. Y., and van de Ven, W. J. M. (1996) FASEB J. 10, 453–460
38. Hu, Z., Zhuang, L., and Dufau, M. L. (1996) J. Biol. Chem. 271, 10242–10246
39. Small, D. J., Vayda, M. E., and Sidell, B. D. (1998) J. Mol. Evol., in press
40. King, S. M., Marchese-Ragona, S. P., Parker, S. K., and Detrich, H. W., III (1997) Biochemistry 36, 1306–1314
41. Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M., and Rodier, F. (1985) Science 228, 953–958
42. Bernardi, G., and Bernardi, Gi. (1986) J. Mol. Evol. 24, 1–11
43. Bernardi, G., Moucheiroud, D., Gautier, C., and Bernardi, Gi. (1988) J. Mol. Evol. 28, 7–18
44. McMorrow, T., Wagner, A., Deryckere, F., and Gannon, F. (1996) DNA Cell Biol. 15, 407–414
45. Miyata, M., and Aoki, T. (1997) Biochim. Biophys. Acta 1354, 127–133
46. Lanfranchi, G., Pallavicina, A., Laveder, P., and Valle, G. (1994) Dev. Biol. 164, 402–408
47. Chen, L., DeVries, A. L., and Cheng, C.-H. C. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 3811–3816
48. Kozak, M. (1991) J. Cell Biol. 115, 887–903