The Journal of Experimental Biology 206, 131-139

doi:10.1242/jeb.00067

The myoglobin gene of the Antarctic icefish, *Chaenocephalus aceratus*, contains a duplicated TATAAAA sequence that interferes with transcription

Deena J. Small[1,2,*], Thomas Moylan[1,†], Michael E. Vayda[1,2] and Bruce D. Sidell[1,2,‡]

[1]*School of Marine Sciences* and [2]*Department of Biochemistry, Microbiology and Molecular Biology, University of Maine, Orono, ME 04469, USA*

*Present address: Maine Medical Center Research Institute, Center for Molecular Medicine, 81 Research Drive, Scarborough, ME 04074, USA

†Present address: Marine Education and Research Center, c/o Biological Sciences Department, California Polytechnic State University, San Luis Obispo, CA 93407, USA

‡Author for correspondence at address 1 (e-mail: bsidell@maine.edu)

Accepted 10 October 2002

Summary

Six of the 16 known species of Antarctic icefish (family Channichthyidae) have lost the ability to express cardiac myoglobin (Mb) *via* at least four independent events during radiation of these species. We report here that the lesion in *Chaenocephalus aceratus* Mb is a duplicated TATAAAA element that blocks transcription. This lesion is distinct from those of other icefish species that do not express cardiac Mb. The *C. aceratus* Mb gene is nearly identical to that of *Chionodraco rastrospinosus*, a closely related Mb-expressing icefish species, with one exception. A 15-bp segment is present in *C. aceratus* but absent from *C. rastrospinosus*; this insertion is located 648 bp upstream from the reference transcription start site of *C. rastrospinosus* and includes the sequence TATAAAA, which bound HeLa cell transcription factor IID (TFIID) and icefish nuclear proteins in gel-retardation assays. Reporter constructs containing the 'full-length' *C. aceratus* Mb promoter were not expressed in transient expression assays in oxidative skeletal muscle of live icefish. By contrast, constructs employing the nearly identical 'full-length' *C. rastrospinosus* Mb promoter were efficiently expressed in parallel assays in the same tissue. Truncated constructs of *C. aceratus* Mb that did not contain the 15-bp duplication were expressed at very low levels. These data confirm a third independent mechanism of Mb loss among channichthyid species, indicate that *C. aceratus* aerobic muscle is capable of expressing functional Mb genes and demonstrate that duplication of the muscle-specific TATAAAA sequence in an inappropriate context can result in loss of a gene's expression, resulting in significant physiological consequences.

Key words: myoglobin, Antarctic fish, heart muscle, gene expression, promoter regulation, oxygen transport, icefish, *Chaenocephalus aceratus*, Channichthyidae.

Introduction

Antarctic icefish (family Channichthyidae) are members of the Perciform suborder Notothenioidei that dominates the fish fauna of coastal Antarctic waters. This unique marine environment became isolated from the rest of the world's oceans approximately 25 million years ago, when the Drake Passage opened between South America and the Antarctic Peninsula, enabling subsequent development of strong circumpolar currents (Kennett, 1980; Eastman, 1993). Antarctic coastal waters are thermally stable and extremely cold, with temperatures that range only from –1.86°C to +0.3°C throughout the year (Littlepage, 1965; DeWitt, 1971). Divergence of mitochondrial DNA sequences suggests that the major radiation of Antarctic notothenioid species began 10–16 million years ago (Bargelloni et al., 1994; Bargelloni and LeCointre, 1998). Because channichthyids are the most derived notothenioid family, it is clear that this group has evolved much more recently than these dates.

Icefish are the only known vertebrates whose adult forms do not express the circulating oxygen-binding protein, hemoglobin (Ruud, 1954; Hemmingsen, 1991). All icefish exhibit dramatic cardiovascular adaptations to compensate for the lack of hemoglobin in the circulation (Hemmingsen, 1991; Johnston et al., 1993). The loss of hemoglobin expression, however, is not the only unusual feature of oxygen delivery systems in these fish. At least six species of icefish have lost the ability to express the intracellular oxygen-binding protein, myoglobin (Mb), within their cardiac muscle; oxidative skeletal muscle of all species in this family is devoid of Mb (Sidell et al., 1997; Vayda et al., 1997). Mb functions as an intracellular oxygen reservoir and facilitates delivery of oxygen to mitochondria in tissues that rely heavily on oxidative phosphorylation to meet their energetic needs (Wittenberg and Wittenberg, 1989).

Acierno et al. (1997) demonstrated that hearts of icefish that

do express Mb are capable of maintaining cardiac output at greater afterload challenges than are hearts of closely related species that do not express the protein. Enhanced performance of the Mb-containing hearts is erased in the presence of sodium nitrite, a specific inhibitor of Mb function, demonstrating that loss of Mb has significant physiological consequences. Ventricular tissues from icefish species that lack Mb also contain significantly higher densities of mitochondria and are spongier, with more spaces filled by luminal blood, than tissue from channichthyids that do express Mb (O'Brien and Sidell, 2000; O'Brien et al., 2000). Both of these features are thought to reduce the diffusion path length for oxygen and ensure its efficient delivery from blood to mitochondria.

In some species of icefish, the reason for loss of Mb expression is clear. For example, *Champsocephalus gunnari* expresses a non-functional Mb mRNA containing a five-nucleotide duplication that causes a frameshift resulting in premature termination of the Mb polypeptide (Vayda et al., 1997). We recently have found an identical 5-bp insertion in the Mb-coding sequence of congeneric *Champsocephalus esox* (T. J. Grove, J. Hendrickson and B. D. Sidell, unpublished observations). Mb mRNA is found in trace amounts in *Pagetopsis macropterus* cardiac muscle, its scarcity presumably due to an aberrant polyadenylation site (Vayda et al., 1997). The loss of Mb expression in *Chaenocephalus aceratus*, however, appears to be distinct because Mb mRNA cannot be detected in the heart ventricle of this species (Sidell et al., 1997) nor can Mb transcripts be detected by nuclear run-on assays (D. J. Small, M. E. Vayda and B. D. Sidell, unpublished results).

The *C. aceratus* Mb genomic sequence does not reveal aberrant splice junctions or other lesions in the transcriptional unit that can explain the absence of Mb expression. In fact, the *C. aceratus* Mb genomic sequence is 98% identical to the functional Mb genomic sequence of *Chionodraco rastrospinosus*, an Mb-expressing species (Small et al., 1998). The only differences between Mb genes of these icefish species are insertions/deletions within intron sequences, variations in the number of a simple sequence repeat ATCT located 141 bp upstream of transcription start relative to the *C. rastrospinosus* Mb gene sequence, and a 15-bp insertion in the *C. aceratus* Mb gene relative to that of *C. rastrospinosus*, which is located between residues –647 and –648 of the reference *C. rastrospinosus* sequence (Small et al., 1998). This insertion contains the sequence TATAAAA, which is identical to the muscle-specific transcription factor IID (TFIID) target sequence that is found 25 bp upstream of the *C. rastrospinosus* Mb transcription start site. This sequence is typical of the 'TATA' sequences of muscle-specific genes (Basel-Duby et al., 1993). The number of ATCT repeats varied between Mb promoters of *C. rastrospinosus* (13 repeats) and *C. gunnari* (25 repeats), both of which are transcribed, suggesting that repeat length of this microsatellite is unlikely to explain the failure of *C. aceratus* Mb (22 repeats) to be transcribed (Small et al., 1998). These observations prompted us to determine whether the duplicated TATAAAA sequence is responsible for the loss of Mb expression in *C. aceratus*. Results of our experiments show that the duplicated TATAAAA sequence binds TFIID and binds factors of icefish heart nuclear extracts and that constructs containing the duplicated TATAAAA sequence are not expressed. These results substantiate a third discrete molecular mechanism that has led to the loss of Mb expression within this unique family of fishes.

Materials and methods

Animals and tissues

Chaenocephalus aceratus Lönnberg used in this study were captured from depths of 80–190 m by 5.5 m Otter trawl deployed from *R/V Polar Duke* or *ARSV Laurence M. Gould*. Capture sites were either off the south shore of Low Island (63°25′S, 62°10′W) or the southwest shore of Brabant Island in the vicinity of Astrolabe Needle (64°10′S, 62°35′W), off the Antarctic Peninsula. Fish were transported live in shipboard aquaria to the US Antarctic research station, Palmer Station, where they were maintained in running seawater tanks until used in experiments. Animals were killed by a sharp blow to the head, followed by severing the spinal cord immediately posterior to the head.

Isolation of nuclear extracts

Intact nuclei were isolated from 4–10 g of heart ventricular tissue from freshly killed animals for each preparation. Tissue was rinsed with 50 ml ice-cold Ringer's solution (260 mmol l^{-1} NaCl, 5 mmol l^{-1} KCl, 2.5 mmol l^{-1} $MgCl_2$, 3 mmol l^{-1} $CaCl_2$, 2.5 mmol l^{-1} $NaHCO_3$ and 2 mmol l^{-1} NaH_2PO_4) and then diced and homogenized in 30 ml of homogenization buffer [300 mmol l^{-1} sucrose, 10 mmol l^{-1} Hepes pH 7.5, 5 mmol l^{-1} KCl, 0.75 mmol l^{-1} spermidine, 0.15 mmol l^{-1} spermine, 0.1 mmol l^{-1} EDTA, 0.1 mmol l^{-1} EGTA, 0.5 mmol l^{-1} phenylmethanesulfonyl fluoride (PMSF; Sigma, St Louis, MO, USA) and 2 μg ml^{-1} final concentration each of aprotinin, leupeptin and pepstatin] in a 40 ml dounce homogenizer by five strokes with pestle B (loose fitting) and three strokes with pestle A (tight fitting). The crude homogenate was then filtered through two layers of cheesecloth pre-wetted with the homogenization buffer and mixed with an equal volume of cushion buffer [9 volumes of 80% ultrapure sucrose solution mixed with 1 volume of 10× salts: 100 mmol l^{-1} Hepes pH 7.5, 50 mmol l^{-1} KCl, 7.5 mmol l^{-1} spermidine, 1.5 mmol l^{-1} spermine, 1 mmol l^{-1} EDTA, 1 mmol l^{-1} EGTA and 10 mmol l^{-1} dithiothreitol (DTT)]. The mixture was divided evenly and overlayed over 10 ml of cushion buffer in three Beckman SW25.1 ultracentrifuge tubes. Nuclei were collected by centrifugation at 59 750 ***g*** at 4°C for 60 min in a Beckman SW25.1 swinging bucket rotor. Nuclear pellets were resuspended in a 1 ml nuclei storage buffer (25% glycerol, 50 mmol l^{-1} Hepes pH 7.5, 3 mmol l^{-1} $MgCl_2$, 0.1 mmol l^{-1} EDTA, 1 mmol l^{-1} DTT, 0.1 mmol l^{-1} PMSF, 2 μg ml^{-1} each of aprotinin, leupeptin, pepstatin). 200 μl samples were placed in cryovials and flash-frozen in liquid nitrogen. Nuclei were stored in liquid nitrogen or at –70°C until use. Nuclear proteins

used in the gel mobility shift assays were extracted from frozen nuclei isolated from *C. rastrospinosus* heart ventricular tissue by incubating 150 µl of nuclei in nuclear extraction buffer (0.5 mmol l^{-1} PMSF, 2 µg ml^{-1} each of aprotinin, leupeptin and pepstatin, 0.5 mol l^{-1} NaCl, 0.7 mmol l^{-1} spermidine, 0.13 mmol l^{-1} spermine and 0.17 mmol l^{-1} EGTA) with constant rotation for 90 min at 4°C. Soluble nuclear proteins were separated from chromatin by centrifugation at 16 000 ***g*** in an Eppendorf microcentrifuge for 20 min. The supernatant containing the soluble nuclear proteins was removed, concentrated using a Micron 3 protein concentrator (Millipore, Bedford, MA, USA) and dialyzed against 1× GMS buffer [20 mmol l^{-1} Hepes pH 7.9, 25 mmol l^{-1} KCl, 2 mmol l^{-1} spermidine; 0.1 mmol l^{-1} EDTA, 10% glycerol, 0.5 mmol l^{-1} DTT and 100 µg ml^{-1} bovine serum albumin (BSA)].

Gel mobility shift assays

Sense and anti-sense oligonucleotides (purchased from Integrated DNA Technologies, Coralville, IA, USA) representing the DNA targets depicted in Fig. 1A used for the gel mobility shift assays were end-labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP as described in Sambrook et al. (1989). After labeling, complementary sense and antisense oligonucleotides were mixed together, heated to 96°C for 5 min and then slowly cooled to room temperature. The annealed oligonucleotides were then purified by electrophoresis through native 12% acrylamide gels cast in 50 mmol l^{-1} Tris-borate, 0.5 mmol l^{-1} EDTA, pH 8.3. After electrophoresis, the labeled double-stranded oligonucleotides were excised from the gel and eluted overnight in TE buffer. 2 pmol of the indicated oligonucleotide template was mixed with 5 µg of a nuclear extract or 2 ng of purified TFIID (Promega, Madison, WI, USA), 2 µg of poly(dI-dC) and 1× final concentration GMS buffer for a total volume of 25 µl. For competition assays, 100-fold excess (200 pmol) of unlabeled oligonucleotide was added to each sample. The reaction was incubated at room temperature for 30 min then mixed with 3 µl 10× gel-loading buffer (Promega), loaded onto 6% native acrylamide gels cast in 50 mmol l^{-1} Tris-borate, 0.5 mmol l^{-1} EDTA, pH 8.3 and subjected to electrophoresis at 25 V for 2 h. Gels were dried and exposed to film overnight.

Construction of reporter plasmids

Reporter plasmids were generated by PCR amplification of icefish Mb promoter sequences using DNA isolated from Mb genomic clones and PCR primers specific to conserved regions of the icefish promoter (Small et al., 1998). Each primer also contained short DNA sequences at the 5′ end that corresponded to either the *Sac*I (forward primer) or *Bgl*II (reverse primer) restriction site for cloning purposes. The forward primer used for constructs pR1L and pA1L was 5′-gccgagctcCTGCAGCC-CTCGAGTCGGTTTCTTC-3′ (position –1525 to –1501 of the *C. rastrospinosus* Mb sequence; –1575 to –1551 of the *C. aceratus* Mb sequence); the forward primer used for constructs pR2L and pA2L was 5′-gccgagctcGGTGTTTTCCGGGTG-TTGA-3′ (position –599 to –580 of the *C. rastrospinosus* Mb sequence; –634 to –403 of the *C. aceratus* Mb sequence); the forward primer used for constructs pR3L and pA3L was 5′-gccgagctcGGACAAGAAGAGGAAACATAGGATAGTG-3′ (position –398 to –369 of the *C. rastrospinosus* Mb sequence; –434 to –403 of the *C. aceratus* Mb sequence). The reverse primer used to generate all six constructs was 5′-gccagatctGATGTTGTACAAAATTCTTCTTGACCTGAC-3′ (complementary to positions +32 to +3 of both the *C. rastrospinosus* and *C. aceratus* Mb sequences). After PCR amplification, products were column purified (Qiagen, Valencia, CA, USA) and digested with *Sac*I and *Bgl*II. Digested PCR products were ligated to the pGL2 basic vector (Promega) at the *Sac*I and *Bgl*II sites. All constructs were sequenced at the University of Maine DNA sequencing core facility prior to use.

Transient expression assay in vivo

Specimens of *C. aceratus* and *C. rastrospinosus* were maintained in flowing seawater tanks at Palmer Station, Antarctica at an ambient temperature of –1°C. Individual fish were injected at multiple sites into their pectoral abductor muscle; each site was injected with a 1–5 µl solution containing 1 µg of a test plasmid promoter construct (pR1L, pA1L, pR2L, pA2L, pR3L, pA3L or pSV40Luc) and 0.5 µg of the *Renilla* luciferase internal reference construct driven by the cytomegalovirus (CMV) promoter (Promega). Each injection site was marked by a suture. Fish were returned to flowing seawater tanks and maintained until 0.3–0.6 g blocks of tissue surrounding each injection site were harvested 3–7 days post-injection (Schulte et al., 1998; Friedenreich and Schartl, 1990) and immediately frozen in liquid nitrogen. Luciferase assays were conducted using the dual-luciferase reporter assay system (Promega) according to the supplier. Light units emitted were measured using an MJ Research lumenometer (MJ Research, Waltham, MA, USA).

Results

The upstream insertion in the *C. aceratus* myoglobin promoter binds TFIID and icefish nuclear factors. Previous work has demonstrated that the –1575 bp sequence of the *C. aceratus* Mb promoter is 99% identical to that of the functional –1525 bp *C. rastrospinosus* Mb gene (Small et al., 1998). The difference in the length between the two sequences arises from differences in the length of the ATCT repeat element (52 bp for *C. rastrospinosus versus* 88 bp for *C. aceratus*) and a 15 bp sequence present in *C. aceratus* Mb but not present in *C. rastrospinosus* Mb. Fig. 1A illustrates the sequence of *C. aceratus* Mb containing this –695 bp upstream TATAAA sequence, –648 bp relative to the Mb-expressing icefish species *C. rastrospinosus*. This 15-bp insertion is proximal to a putative E2A site (Small et al., 1998) necessary for proper Mb developmental expression (Blackwell and Weintraub, 1990; Weintraub et al., 1991).

Oligonucleotides corresponding to this segment of the *C. aceratus* and *C. rastrospinosus* Mb promoters, and the

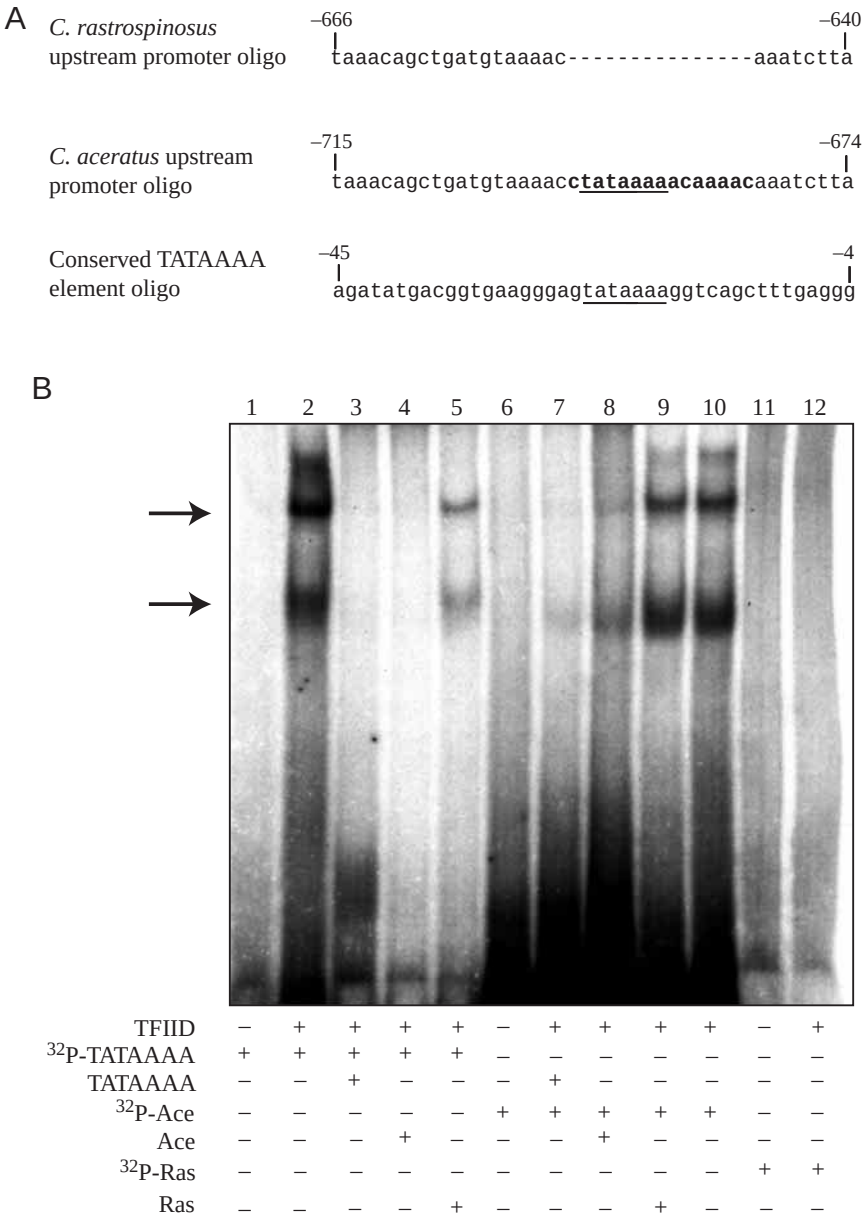

Fig. 1. Binding of HeLa cell transcription factor TFIID to icefish myoglobin (Mb) promoter sequences. (A) Oligonucleotides corresponding to both strands of the *Chaenocephalus aceratus* upstream promoter sequence (–715 to –674) that contains the TATAAA duplication (in bold) (*C. aceratus* upstream promoter oligo), the homologous region (–666 to –640) of the *C. rastrospinosus* promoter (*C. rastrospinosus* upstream promoter oligo) and the conserved promoter sequence (–45 to –4) containing the putative 'TATAAAA element' located within the core promoter region of both icefish (conserved TATAAAA element oligo) were synthesized based on the upstream promoter sequences obtained as described in Small et al., 1998. Numeric positions of sequences are relative to the transcription start site of each species. The TATAAAA sequence for both the *C. aceratus* upstream oligo and the conserved TATAAAA element oligo are underlined. Prior to use in the gel mobility shift assays, the complementary oligonucleotides were annealed and ^{32}P-end labeled using T_4 polynucleotide kinase and [γ-^{32}P]ATP prepared as described in the Materials and methods. (B) 2 pmol of the radiolabeled conserved TATAAAA element oligo (^{32}P-TATAAAA, lanes 1–5), radiolabeled *C. aceratus* upstream promoter oligo (^{32}P-Ace, lanes 6–10) or radiolabeled *C. rastrospinosus* upstream oligo (^{32}P-Ras, lanes 11 and 12) were incubated with (+) or without (–) 20 ng of purified human TFIID as indicated. Some samples also included 200 pmol of unlabeled conserved TATAAAA oligo (TATAAAA, lanes 3 and 7), *C. aceratus* upstream promoter oligo (Ace, lanes 4 and 8) or *C. rastrospinosus* upstream oligo (Ras, lanes 5 and 9) as competitor oligonucleotides. The DNA/TFIID complexes were then resolved by electrophoresis and autoradiography as described under Materials and methods. Arrows indicate the positions of shifted bands that presumably represent radiolabeled DNA/TFIID complexes.

authentic TATAAAA sequence common to both Mb genes (Small et al., 1998), were synthesized, labeled with ^{32}P using T4 polynucleotide kinase and used for gel-retardation assays (Fig. 1A). The *C. aceratus* upstream Mb sequence (–715/–674) bound purified human TFIID in gel-retardation assays (Fig. 1B, lane 10) as evident by the presence of two prominent retardation complexes. The authentic TATAAAA oligonucleotide (–45/–4) formed similar retardation complexes (Fig. 1B, lane 2). These two oligonucleotides competed for the same factor when one was present in excess, as indicated by a nearly complete elimination of labeled retardation complexes in Fig. 1B, lanes 3, 4, 7 and 8. The *C. rastrospinosus* upstream Mb sequence (–666/–640) did not form a retardation complex with TFIID (Fig. 1B, lane 12), nor was it effective in preventing the formation of retardation complexes with either of the TATAAAA-containing oligonucleotides (Fig. 1B, lanes 5 and 9). The *C. aceratus* upstream Mb sequence (–715/–674) binds to factors present in heart ventricle nuclear extracts from *C. rastrospinosus* to form several complexes (Fig. 2A, lanes 3 and 6). These complexes had similar mobility to those generated by incubation with HeLa TFIID (Fig. 2A, lane 1). Formation of labeled retardation complex was prevented by both an excess of the (–45/–4) TATAAAA oligonucleotide (Fig. 2A, lane 5) and by an excess of non-labeled *C. aceratus* (–715/–674) oligonucleotide (Fig. 2A, lane 4). The *C. aceratus* upstream (–715/–674) Mb oligonucleotide also prevented the formation of complexes between factors present in the icefish nuclear extract and labeled (–45/–4) TATAAAA oligonucleotide (shown by arrows in Fig. 2B, lane 3). These results suggest that the upstream insertion in the *C. aceratus* Mb promoter binds TFIID in an inappropriate context, which may explain the aberration of this promoter.

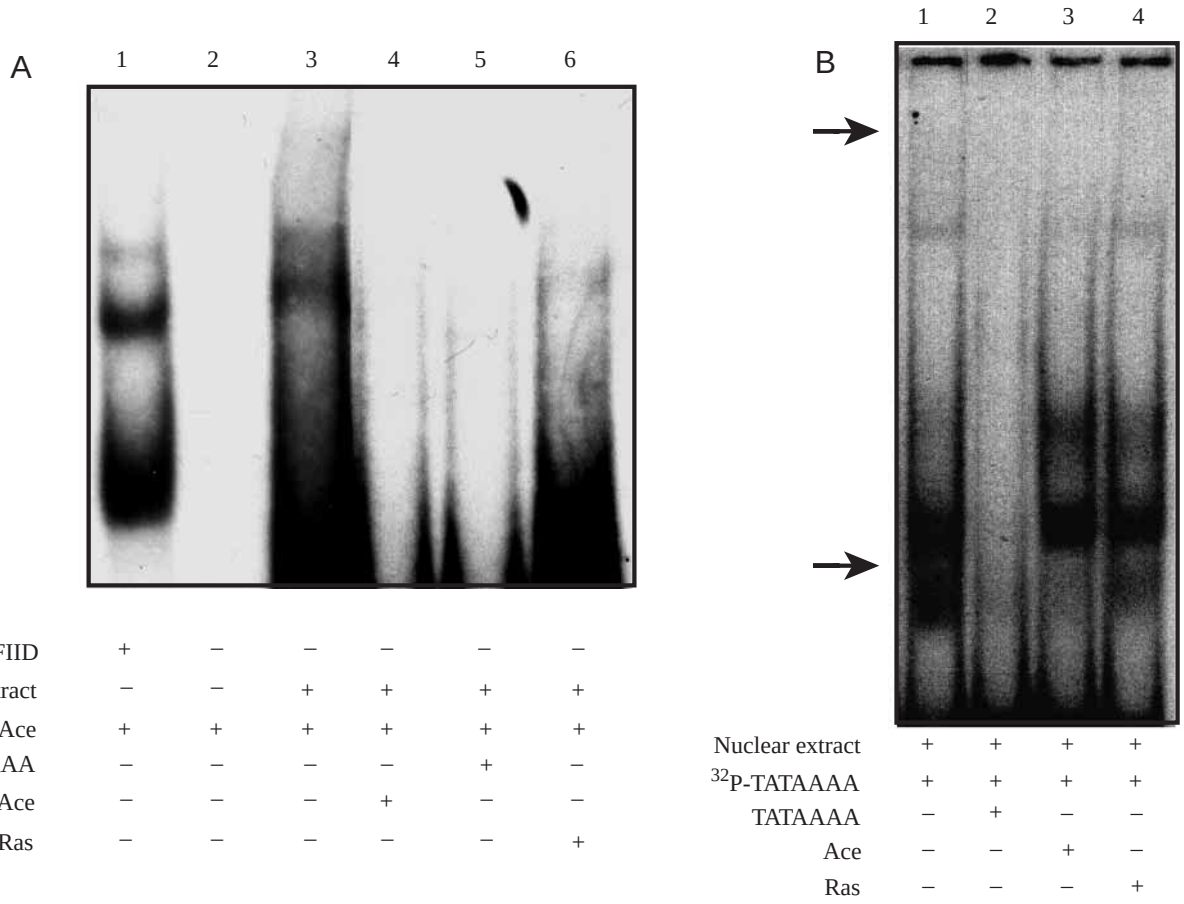

Fig. 2. Binding of factors present in *Chionodraco rastrospinosus* cardiac muscle nuclear extracts to myoglobin (Mb) promoter sequences. (A) 2 pmol of labeled *Chaenocephalus aceratus* upstream promoter sequence (–715 to –674) containing the TATAAA duplication shown in Fig. 1A were incubated with either 2 ng HeLa transcription factor IID (TFIID) (lane 1) or 5 μg of a *C. rastrospinosus* cardiac muscle nuclear extract (lanes 3–6) prepared as described under Materials and methods. No retardation bands were evident in the absence of added protein (lane 2). 200 pmol of unlabeled conserved TATAAAA oligo (TATAAAA, lane 5), *C. aceratus* upstream promoter oligo (Ace, lane 4) or *C. rastrospinosus* upstream oligo (Ras, lane 6) were used as competitor oligonucleotides as indicated. DNA/nuclear protein complexes were then resolved by electrophoresis and autoradiography as described under Materials and methods. (B) 2 pmol of labeled conserved TATAAAA element oligonucleotides (^{32}P-TATAAAA) shown in Fig. 1A were incubated with 5 μg of a *C. rastrospinosus* cardiac muscle nuclear extract prepared as described under Materials and methods. Retardation bands generated by incubation of the ^{32}P-labeled conserved TATAAAA sequence with nuclear extracts from *C. aceratus* are shown in lane 1. 200 pmol of unlabeled conserved TATAAAA oligo (TATAAAA, lane 2), *C. aceratus* upstream promoter oligonucleotide (Ace, lane 3) or *C. rastrospinosus* upstream oligonucleotide (Ras, lane 4) were used as competitor oligonucleotides as indicated. DNA/nuclear protein complexes were then resolved by electrophoresis and autoradiography as described under Materials and methods. Arrows denote DNA/protein complexes that are eliminated by 200 pmol of unlabeled *C. aceratus* upstream promoter oligonucleotide.

The presence of the –695 bp TATA sequences interferes with transient expression in *C. aceratus* muscle *in vivo*. In order to determine whether the –695 TATAAAA sequence inhibited transcription, segments of the *C. rastrospinosus* and *C. aceratus* Mb promoters were selectively amplified and cloned. The promoter segments were amplified using sense primers MP8 (–1525 of *C. rastrospinosus* Mb and –1575 of *C. aceratus* Mb), MP7F (–599 of *C. rastrospinosus* Mb and –634 of *C. aceratus* Mb) or MP6F (–398 of *C. rastrospinosus* Mb and –434 of *C. aceratus* Mb) and antisense primer Myo3B (+32) (Vayda et al., 1997; Small et al., 1998). Each of these amplified segments was linked to the firefly luciferase coding sequence reporter gene (Fig. 3A) and introduced into *C. aceratus* aerobic pectoral adductor muscle for transient expression assay. This tissue was competent to express the full-length (–1525) *C. rastrospinosus* Mb promoter construct (Fig. 3B). By contrast, the corresponding *C. aceratus* full-length (–1575) promoter construct containing the TATAAAA duplication was not expressed above the level of promoter-less controls. The *C. rastrospinosus* construct truncated at position –599 and the *C. aceratus* constructs truncated at position –634 both exhibited expression higher than the full-length *C. aceratus* construct containing the TATAAAA duplication (Fig. 3B). However, the low levels of expression provided by these truncated constructs indicate that an element upstream of –599 is essential for efficient transient expression *in vivo*. Further

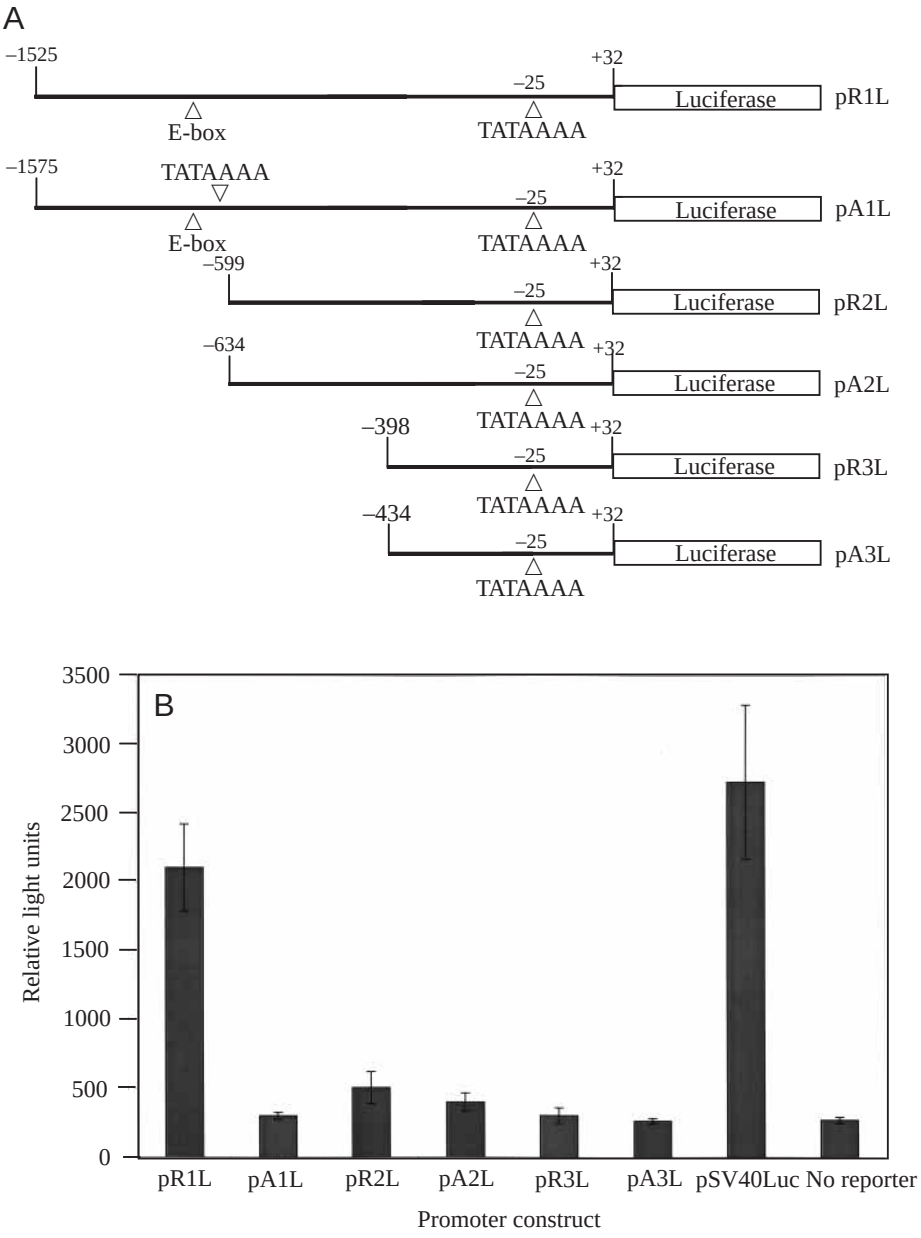

Fig. 3. Transient expression of *Chaenocephalus aceratus* and *Chionodraco rastrospinosus* promoter constructs in *C. aceratus* pectoral abductor muscle. (A) Schematic depiction of the upstream *C. rastrospinosus* myoglobin (Mb) promoter constructs (pR1L, pR2L and pR3L) and *C. aceratus* Mb promoter constructs (pA1L, pA2L and pA3L) generated as described in Materials and methods for use in the *in vivo* transient expression assay shown in (B). The upstream Mb promoter sequences inserted into the pGL2 basic luciferase plasmid are indicated numerically and are shown in Small et al. (1998). The position of the conserved TATAAAA element is indicated by a triangle at position –25 in all constructs. The putative E-box element found in the upstream promoter sequence of both species and the TATAAAA duplication found in *C. aceratus* are also indicated in constructs pR1L and pA1L. Promoter constructs were prepared as described in Materials and methods. (B) 1 μg of each test plasmid and 0.5 μg of the *Renilla* luciferase internal reference construct driven by the cytomegalovirus (CMV) promoter was injected into the pectoral adductor muscle of a *C. aceratus* individual. Activity of firefly luciferase in each tissue sample was normalized to that of *Renilla* luciferase. Values are means ± S.E.M. for N=7 trials. Expression of 'full-length' pR1L was significantly greater (P<0.001) than all other channichthyid Mb promoter constructs and was not significantly different from the positive control driven by the pSV40Luc promoter. Expression of 'full-length' pA1L was not significantly different from any truncated constructs from either species (pR2L, pR3L, pA2L or pA3L). 'No reporter' shows expression of the pGL2 luciferase plasmid lacking a promoter sequence.

truncation of either the *C. rastrospinosus* or *C. aceratus* Mb genes reduced transient expression to that of promoter-less controls (Fig. 3B). This striking difference between expressions of the full-length *C. rastrospinosus* Mb construct pR1L and the full-length *C. aceratus* Mb construct pA1L suggests that the presence of the duplicated TATAAAA element severely interferes with transcription.

Discussion

It is now clear that six of the 16 known species of Antarctic channichthyid icefish have lost the ability to express Mb in their heart ventricles; the remaining ten species of icefish do produce the protein in ventricular muscle. Furthermore, mapping the topology of Mb expression upon the best available cladistic phylogeny for this recently evolved group permits us to conclude that events leading to the loss of Mb expression in the icefish family have occurred independently at least four times during their radiation (Moylan and Sidell, 2000). Loss of expression of 'normally' physiologically important protein(s) in notothenioid fishes is not restricted to myoglobin. For example, Hofmann et al. (2000) have documented the absence of inducible production of heat-shock proteins in *Trematomus bernacchii*, and it has long been known that all members of the family Channicthyidae lack the circulating oxygen-binding protein, hemoglobin (Hb). Detrich and coworkers have established that loss of Hb expression in channichthyid icefish is due to an apparent complete deletion of the gene encoding β-globin and partial loss of the β-globin gene, rendering it impossible for the animals to produce a

functional Hb molecule (Cocca et al., 1995). The mechanisms accounting for the loss of Mb among species of icefish, however, are quite different.

In previous communications, we have described two of the specific mutations leading to the loss of Mb expression in other icefish species, *Champsocephalus gunnari* (5-nucleotide insertion leading to premature termination) and *Pagetopsis macropterus* (aberrant polyadenylation signal) (Vayda et al., 1997). In the present study, we report a third distinct mechanism by which Mb expression has been lost among channichthyid fishes. We can also conclude that the mutation we describe here must be very recent indeed. Compared with the gene from closely related Mb-expressing *Chionodraco rastrospinosus*, the now-silent Mb gene in *Chaenocephalus aceratus* does not exhibit any lesions within the normally transcribed sequence that would lead to inactivation of the gene product. Apparently, the relatively subtle event of duplicating a normal promoter element (the muscle-specific TATAAAA sequence) has resulted in the loss of Mb expression in this species. Numerous examples are known where removal of a negative regulatory element (thus eliminating binding of negative regulatory factor) results in the elevated expression of a promoter construct (e.g. Yan et al., 2001). However, we are unaware of any other examples where duplication of a TATAAAA element results in silencing of gene expression.

Lack of activity in transient expression assays of the 'full-length' promoter construct (pA1L) from *C. aceratus* indicates that this region of the gene contains a mutation that prevents its transcription. The only significant difference between this Mb promoter from *C. aceratus* and that from *C. rastrospinosus*, which is transcribed normally, is the 15-bp insertion containing a duplicated TATAAAA sequence at –695 bp. This insertion is located only 10 bp downstream from a putative 'E-box', a site that binds the basic helix-loop-helix (bHLH) class of transcription factors, such as MyoD, and the ubiquitously expressed E class factors, such as E12 (Small et al., 1998; Blackwell and Weintraub, 1990; Weintraub et al., 1991). Our gel mobility shift assays demonstrate that the transcription factor, TFIID, binds *in vitro* to the upstream insertion in *C. aceratus* and that this region also binds factors that are present in nuclear extracts from icefish. On the strength of these data, it is reasonable to conclude that the failure to observe Mb transcription *in vivo* is probably due to inappropriate recruitment of TFIID to the duplication at this site. TFIID bound to this duplicated TATAAAA sequence may sterically hinder the binding of MyoD or E-box class factors to putative binding sites at –575 and –663, respectively (Small et al., 1998). Binding of the TATA-binding protein (TBP) is known to induce a sharp 80° bend in DNA, distorting the backbone to force open the minor groove at the TATA sequence (Kim et al., 1993a,b). The introduction of such a conformational distortion between the potential MyoD-binding site at –575 or the E2A site at –663 could preclude binding of these skeletal muscle transcriptional activators. Although well-documented as skeletal muscle-specific transcriptional activators (Weintraub et al., 1991), E-box elements upstream of the start site are not essential for expression of Mb in mammalian cardiac muscle (Navankasattusas et al., 1992) or aerobic skeletal muscle (Yan et al., 2001). However, an intragenic E-box in close proximity to the transcriptional start site (+5 to +10) is known to act as a negative regulator element that probably contributes to myofiber type-specific Mb expression (Yan et al., 2001). We suggest that inappropriate positioning of the TATAAAA duplication near the E2A site may seed assembly of an inhibitory complex that silences Mb expression in *C. aceratus*.

Results from our *in vivo* transient expression assays with oxidative skeletal muscle of *C. aceratus* (Fig. 3) permit two additional conclusions. First, the minimal promoter required for *C. rastrospinosus* Mb expression is significantly longer than that required for mammalian Mb (Bassel-Duby et al., 1993) and includes at least one element distal to position –634. Second, despite the lack of endogenous Mb expression in aerobic skeletal muscle from any notothenioid fish that we have examined to date (Moylan and Sidell, 2000), this tissue apparently contains all the necessary transcription factors required for transient expression of the Mb promoter constructs. Thus, silencing of Mb expression in oxidative skeletal muscle, even in those species that normally express the protein in heart ventricle, must be the result of some type of developmental patterning, possibly rendering the gene inaccessible due to higher order DNA/chromosomal structure in the differentiated tissue.

Both the highly derived position of *C. aceratus* in channichthyid phylogeny and the punctuated loss of Mb expression prior to significant mutation in the coding sequence of the gene strongly indicate that the genetic lesion in this species has occurred quite recently. Our relatively extensive sequence analysis indicates that a single duplication event (duplication of the TATAAAA promoter element) apparently has led to the loss of Mb expression in *C. aceratus*. This observation is not unique among the channichthyid icefish. In fact, loss of Mb expression in *C. gunnari* and congeneric *C. esox* is also due to a single 5-bp duplication within the coding sequence of the gene (Vayda et al., 1997; T. J. Grove et al., unpublished data). We have identified a different mutation in the polyadenylation signal of the Mb gene in another icefish species, *P. macropterus*, which also has led to loss of Mb production (Vayda et al., 1997). Taken together, loss of myoglobin expression *via* both multiple events and multiple mechanisms during evolution of the icefish family seems to suggest that Mb function may not be of physiological relevance in the oxygen-rich and severely cold environment of the Southern Ocean. Compelling physiological evidence, however, suggests otherwise.

When isolated, perfused hearts from Mb-expressing and Mb-lacking icefish species are challenged with increasing afterload, hearts whose ventricles contain Mb are capable of greater pressure-work (Acierno et al., 1997). Furthermore, specific poisoning of Mb function in these preparations causes performance of Mb-expressing hearts to decrement below that of normally Mb-lacking hearts, indicating that the latter display mechanisms that partially compensate for the absence of this

oxygen-binding protein. In fact, one of our laboratories has documented several structural differences in Mb-lacking icefish hearts that probably contribute to this compensation. These include dense populations of mitochondria, which reduce intracellular diffusion distance for oxygen (O'Brien and Sidell, 2000), and a very spongy characteristic of the endocardium, which diminishes the diffusion path length for oxygen from blood-filled lumen of the heart to the interior of muscle cells (O'Brien et al., 2000). Some of these characteristics are analogous to those seen with Mb 'knockout' mouse strains. Murine Mb-knockouts were originally thought to show no significant functional disadvantage compared with Mb-expressing wild-type strains (Garry et al., 1998). However, more recent work has established that a large fraction of embryos from the Mb-deficient strain do not survive gestation. Furthermore, those Mb-deficient offspring that do survive display adaptive responses in their heart muscle that apparently compensate for lack of the oxygen-binding protein and permit normal function (Meeson et al., 2000). Notable among these is a significantly increased vascularization of the tissue, which reduces diffusion path length for oxygen, a response apparently mediated by induction of hypoxia-inducible transcription factors (HIF 1α and HIF 2). This observation opens up the intriguing possibility that a constitutive upregulation of the HIF transduction pathway might underlie the differences in mitochondrial densities and tissue architecture in Mb-deficient icefish hearts.

Results of experiments with isolated, perfused hearts from icefish that are described above, in combination with measurements of oxygen-binding kinetics of icefish Mb (Cashon et al., 1997), clearly indicate that Mbs from icefish are functional at cold body temperature and enhance cardiac performance. The conundrum presented by these observations is that modern population genetics theory dictates that loss of Mb expression, an apparently deleterious trait, should be subject to negative selective pressures and that such genetic mutation(s) ultimately should be eliminated from the population. Yet, Mb-lacking icefish species have arisen at least four times in the radiation of this family and continue to persist. We have argued previously that the uniquely cold and oxygen-rich waters of the Southern Ocean and relatively low niche competition in this marine habitat of relatively low species diversity during evolution of the icefish family have permitted icefish species lacking Mb to persist (Moylan and Sidell, 2000). A recent review by Montgomery and Clements (2000), in fact, cited the pattern of Mb loss in icefish as one example of a pattern of disaptation and recovery (i.e. detrimental genetic change followed by compensatory adaptation). As our knowledge of notothenioid phylogeny and the timing of events during the radiation of these fishes becomes more complete, we can anticipate that this group will offer even more lessons in evolutionary biology.

We gratefully acknowledge the excellent support provided by station personnel at the US Antarctic Program's Palmer Station, Antarctica, and the masters, crews and support personnel of *R/V Polar Duke* and *ARSV Laurence M. Gould*. The work was supported by US National Science Foundation (NSF) Grants OPP 94-21657, 99-09055 and 01-25890 to B.D.S.

References

Acierno, R., Agnisola, C., Tota, B. and Sidell, B. D. (1997). Myoglobin enhances cardiac performance in Antarctic fish species that express the protein. *Am. J. Physiol.* **273**, R100-R106.

Bargelloni, L. and Lecointre, G. (1998). Four years of notothenioid systematics: a molecular perspective. In *Fishes of Antarctica* (ed. G. di Prisco, E. Pisano and A. Clarke), pp. 259-273. Milan: Springer-Verlag Italia.

Bargelloni, L., Ritchie, P. S., Partarnello, T., Battagia, B., Lambert, D. M. and Meyer, A. (1994). Molecular evolution at sub-zero temperatures: mitochondrial and nuclear phylogenies of fishes from Antarctica (suborder Notothenoidei) and the evolution of antifreeze. *Mol. Biol. Evol.* **6**, 854-863.

Bassel-Duby, R., Grohe, C. M., Jessen, M. E., Parsons, W. J., Richardson, J. A., Chao, R., Grayson, J., Ring, W. S. and Williams, R. S. (1993). Sequence elements required for transcriptional activity of the human myoglobin promoter in intact myocardium. *Circ. Res.* **73**, 360-366.

Blackwell, T. K. and Weintraub, H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250**, 1104-1110.

Cashon, R. E., Vayda, M. E. and Sidell, B. D. (1997). Kinetic characterization of myoglobins from vertebrates with vastly different body temperatures. *Comp. Biochem. Physiol.* **117B**, 613-620.

Cocca, E., Ratnayake-Lecamwasam, M., Parker, S. K., Camardella, L., Ciaramella, M., diPrisco, G. and Detrich, H. W., III (1995). Genomic remnants of α-globin genes in the hemoglobinless Antarctic icefishes. *Proc. Natl. Acad. Sci. USA* **92**, 1817-1821.

DeWitt, H. H. (1971). Folio 15. In *Antarctic Map Folio Series* (ed. V. C. Bushnell), pp. 1-10. New York: American Geographical Society.

Eastman, J. T. (1993). *Antarctic Fish Biology: Evolution in a Unique Environment*. New York: Academic Press.

Friedenreich, H. and Schartl, M. (1990). Transient expression directed by homologous and heterologous promoter and enhancer sequences in fish cells. *Nucleic Acids Res.* **18**, 3299-3305.

Garry, D. J., Ordway, G. A., Lorenz, J. N., Radford, N. B., Chin, E. R., Grange, R. W., Bassel-Duby, R. and Williams, R. S. (1998). Mice without myoglobin. *Nature* **395**, 905-908.

Hemmingsen, E. A. (1991). Respiratory and cardiovascular adaptation in hemoglobin-free fish: resolved and unresolved problems. In *Biology of Antarctic Fish* (ed. G. di Prisco, B. Maresca and B. Tota), pp. 191-203. New York: Springer-Verlag.

Hofmann, G. E., Buckley, B. A., Airaksinen, S., Keen, J. E. and Somero, G. N. (2000). Heat-shock protein expression is absent in the Antarctic fish, *Trematomus bernacchii* (Family Nototheniidae). *J. Exp. Biol.* **203**, 2331-2339.

Johnston, I. A., Fitch, N., Zummo, G., Wood, R. E., Harrison, P. and Tota, B. (1993). Morphometric and ultrastructural features of the ventricular myocardium of the haemoglobinless icefish, *Chaenocephalus aceratus*. *Comp. Biochem. Physiol.* **76A**, 475-480.

Kennett, J. P. (1980). Paleoceanographic and biogeographic evolution of the Southern Ocean during the Cenozoic, and Cenozoic microfossil datums. *Paleogeogr. Paleoclimatol. Paleoecol.* **31**, 123-152.

Kim, J. L., Nikolov, D. B. and Burley, S. K. (1993a). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520-527.

Kim, Y., Geiger, J. H., Hahn, S. and Sigler, P. B. (1993b). Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-520.

Littlepage, J. L. (1965). Oceanographic investigation in McMurdo Sound, Antarctica. In *Biology of the Antarctic Seas*, vol. II (ed. G. A. Llano), pp. 1-37. Washington, DC: American Geophysical Union.

Meeson, A. P., Radford, N., Shelton, J. M., Mammen, P. P. A., DiMaio, J. M., Hutcheson, K., Kong, Y., Elterman, J., Williams, R. S. and Garry, D. J. (2001). Adaptive mechanisms that preserve cardiac function in mice without myoglobin. *Circ. Res.* **88**, 713-720.

Montgomery, J. and Clements, K. (2000). Disaptation and recovery in the evolution of Antarctic fishes. *Trends Ecol. Evol.* **15**, 267-271.

Moylan, T. J. and Sidell, B. D. (2000). Concentrations of myoglobin and

myoglobin mRNA in heart ventricles from Antarctic fishes. *J. Exp. Biol.* **203**, 1277-1286.

Navankasattusas, S., Zhu, H., Garcia, A. V., Evans, S. M. and Chien, K. R. (1992). A ubiquitous factor (HF-1a) and a distinct muscle factor (H1b/MEF-2) form an E-box-independent pathway for cardiac muscle gene expression. *Mol. Cell. Biol.* **12**, 1469-1479.

O'Brien, K. M. and Sidell, B. D. (2000). The interplay among cardiac ultrastructure, metabolism and the expression of oxygen-binding proteins in Antarctic fishes. *J. Exp. Biol.* **203**, 1287-1297.

O'Brien, K. M., Xue, H. and Sidell, B. D. (2000). Quantification of diffusion distance within the spongy myocardium of hearts from Antarctic fishes. *Resp. Physiol.* **122**, 71-80.

Ruud, J. T. (1954). Vertebrates without erythrocytes and blood pigment. *Nature* **173**, 848-850.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual*, 2nd edn, Cold Spring Harbor: Cold Spring Harbor Press.

Schulte, P. M., Powers, D. A. and Schartl, M. (1998). Efficient gene transfer into *Xipophorus* muscle and melanoma by injection of supercoiled plasmid DNA. *Mol. Mar. Biol. Biotechnol.* **7**, 241-247.

Sidell, B. D., Vayda, M. E., Small, D. J., Moylan, T. J., Londraville, R. L., Yan, M., Rodnick, K. J., Eppley, Z. A. and Costello, L. (1997). Variable expression of myoglobin among the hemoglobinless Antarctic icefishes. *Proc. Natl. Acad. Sci. USA* **94**, 3420-3424.

Small, D. J., Vayda, M. E. and Sidell, B. D. (1998). A novel vertebrate myoglobin gene containing three A+T-rich introns is conserved among Antarctic teleost species, which differ in myoglobin expression. *J. Mol. Evol.* **47**, 156-166.

Vayda, M. E., Small, D. J., Yuan, M., Costello, L. and Sidell, B. D. (1997). Conservation of the myoglobin gene among Antarctic notothenioid fishes. *Mol. Mar. Biol. Biotechnol.* **6**, 207-216.

Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T. K., Turner, D., Rupp, R., Hollenberg, S., Yuan, Z. and Lassar, A. (1991). The MyoD gene family: nodal point during specification of the muscle cell lineage. *Science* **251**, 761-766.

Wittenberg, B. A. and Wittenberg, J. B. (1989). Transport of oxygen in muscle. *Ann. Rev. Physiol.* **51**, 857-878.

Yan, Z., Serrano, A. L., Schiaffino, S., Bassel-Duby, R. and Williams, R. S. (2001). Regulatory elements governing transcription in specialized myofiber subtypes. *J. Biol. Chem.* **276**, 17361-17366.