BMC
Evolutionary Biology

RESEARCH ARTICLE   **Open Access**

The conserved Phe GH5 of importance for hemoglobin intersubunit contact is mutated in gadoid fish

Øivind Andersen[1,2*], Maria Cristina De Rosa[3], Prakash Yadav[2], Davide Pirolli[4], Jorge MO Fernandes[5], Paul R Berg[6], Sissel Jentoft[6] and Carl Andrè[7]

Abstract

Background: Functionality of the tetrameric hemoglobin molecule seems to be determined by a few amino acids located in key positions. Oxygen binding encompasses structural changes at the interfaces between the α1β2 and α2β1 dimers, but also subunit interactions are important for the oxygen binding affinity and stability. The latter packing contacts include the conserved Arg B12 interacting with Phe GH5, which is replaced by Leu and Tyr in the α^A and α^D chains, respectively, of birds and reptiles.

Results: Searching all known hemoglobins from a variety of gnathostome species (jawed vertebrates) revealed the almost invariant Arg B12 coded by the AGG triplet positioned at an exon-intron boundary. Rare substitutions of Arg B12 in the gnathostome β globins were found in pig, tree shrew and scaled reptiles. Phe GH5 is also highly conserved in the β globins, except for the Leu replacement in the β1 globin of five marine gadoid species, gilthead seabream and the Comoran coelacanth, while Cys and Ile were found in burbot and yellow croaker, respectively. Atlantic cod β1 globin showed a Leu/Met polymorphism at position GH5 dominated by the Met variant in northwest-Atlantic populations that was rarely found in northeast-Atlantic cod. Site-specific analyses identified six consensus codons under positive selection, including 122β(GH5), indicating that the amino acid changes identified at this position may offer an adaptive advantage. In fact, computational mutation analysis showed that the replacement of Phe GH5 with Leu or Cys decreased the number of van der Waals contacts essentially in the deoxy form that probably causes a slight increase in the oxygen binding affinity.

Conclusions: The almost invariant Arg B12 and the AGG codon seem to be important for the packing contacts and pre-mRNA processing, respectively, but the rare mutations identified might be beneficial. The Leu122β1(GH5) Met and Met55β1(D6)Val polymorphisms in Atlantic cod hemoglobin modify the intradimer contacts B12-GH5 and H2-D6, while amino acid replacements at these positions in avian hemoglobin seem to be evolutionary adaptive in air-breathing vertebrates. The results support the theory that adaptive changes in hemoglobin functions are caused by a few substitutions at key positions.

Keywords: Atlantic cod, Hemoglobin, Polymorphism, Coelacanth, Spotted gar, Positive selection

* Correspondence: oivind.andersen@nofima.no
[1]Nofima, P.O. Box 5010, Oslo N-1430 Ås, Norway
[2]Department of Animal and Aquaculture Sciences, Norwegian University of Life Sciences, P.O. Box 5003, Oslo N-1430 Ås, Norway
Full list of author information is available at the end of the article

Background

The hemoglobin molecule has evidently been optimized for oxygen binding under vastly different environmental and physiological conditions by the structural and functional divergence of the vertebrate globin chains [1-3]. The tetrameric hemoglobin consists of two α and two β subunits each containing eight alpha helices (A-H), and the amino acids are numbered either from the N-terminus (excluding the N-terminal Met) or according to helical positions. Whereas the amino acid sequences of both α and β subunits are highly variable with very few invariant positions, adaptive modifications of hemoglobin functions seems to be attributable to a very small number of amino acid substitutions at key positions [4]. These genetically based adaptations have evolved under the influence of natural selection and involve adjustments in heme-protein contacts, intersubunit interactions and binding sites for heterotropic ligands [1,5-7]. The cooperative oxygen binding results from the allosteric equilibrium between the low-affinity T (deoxy) state and the high-affinity R (oxy) state, and the α1β2 and α2β1 dimeric interfaces undergo the principal changes during the deoxy-to-oxy transition [8-10]. In addition to these sliding contacts, the oxygen binding also involves the α1β1 and α2β2 subunit contacts, which play a key role in stabilizing the bound oxygen [11,12]. Several studies of human hemoglobin mutations have documented that even small changes in these packing contacts may affect hemoglobin stability and oxygen binding affinity [13-16]. Further, allosteric effects of chloride ions at the intradimer interfaces cause significant changes in the rates of proton exchange upon ligand binding [17]. Intriguingly, the Leu55β(D6)- > Ser and Pro119α(H2)- > Ala replacements at the α1β1 interface in the bar-headed (*Anser indicus*) and Andean geese (*Chloephaga melanoptera*), respectively, were found to increase the hemoglobin oxygen affinity in these high-altitude species by the elimination of intersubunit contacts [18-21]. Correspondingly, replacing Met with the smaller Val residue in position 55 of the polymorphic β1 globin of Atlantic cod (*Gadus morhua*) was predicted to increase the intrinsic oxygen binding affinity as demonstrated in the human Met55β- > Ser mutant [20-22].

Human hemoglobin mutants have demonstrated the importance of the interaction between Arg B12 and Phe GH5 at the α1β1 and α2β2 interfaces for proper stability and function. Replacement of Arg with the smaller Lys residue containing only two N atoms caused slight anemia in Chinese Hb Kairouan (Arg31α- > Lys) mutants [23], while normal functional properties was found in the unstable Hb Prato (Arg31α- > Ser) mutant [24]. Corresponding Arg B12 mutations in the β globin are found in the unstable human Hb Tacoma (Arg30β- > Ser), but also in the rat zero β globin [25,26], while the mutant protein and transcript were undetectable in human Hb Monroe (Arg30β- > Thr) [27]. The importance of the interacting Phe GH5 was demonstrated by the polar Ser replacement in Hb Caruaru (Phe122β- > Ser) causing chronic haemolytic anaemia [28], but the stability and oxygen binding affinity of Hb Bushey (Phe122β- > Leu) were identical to those of HbA [29]. Phe GH5 seems to be highly conserved in the β chains of other tetrapods, but has been replaced by Leu and Tyr in the αA and αD chains, respectively, of birds and reptiles (Sauropsida) [30]. Phe GH5 mutations in human α globin include Hb Foggia (Phe117α- > Ser) exhibiting a phenotype typical of α thalassemia and was found to impair interactions with both β globin and the Alpha-Hemoglobin Stabilizing Protein (AHSP) [31,32].

A Leu122(GH5)Met polymorphism was recently reported in the Atlantic cod β1 globin, which also harbors the polymorphic positions Met55(D6)Val and Lys62(E6) Ala that modify the oxygen binding properties of the HbI-1 and HbI-2 isoforms [22,33]. This marine fish is widely distributed in temperate and Arctic waters in the North Atlantic, and the HbI-1 (Met55-Lys62) and HbI-2 (Val55-Ala62) variants dominate in southern and northern populations, respectively [22,34,35]. It is plausible that the diversification of Atlantic cod globin has been driven by positive selection, since these non-synonymous mutations alter hemoglobin function and likely confer an adaptive advantage. In protein coding genes, the ratio (ω) between non-synonymous (dN) and synonymous (dS) substitution rates is related to evolutionary constraints at the protein level [37]. A value of ω > 1 indicates positive Darwinian selection, whereas ω < 1 suggests negative selection. To gain insight into the evolutionary history and the functional implications of Phe GH5 mutations, we 1) searched for amino acid replacements at this position in gnathostome β globins, 2) modeled the structural variants of Atlantic cod hemoglobin, 3) examined the distribution of the Leu122β1Met polymorphism in trans-Atlantic cod populations, and 4) investigated genetic signatures of positive selection amongst gadiform β1 sequences.

Results

Invariant Arg31α and novel Phe122β mutations

We searched all known hemoglobin sequences from a broad range of gnathostome species for conservation of the interacting residues Arg B12 and Phe GH5, which correspond to positions 31α/30β and 117α/122β, respectively (Additional file 1: Figure S1). Arg B12 was invariantly found in all α globins examined and was also highly conserved in the β globins with a few exceptions. Arg was replaced by Asn and Ser in the pig β-like and in tree shrew β globin, respectively, while Gly, Cys, Asn and Lys substitutions were found in the β1 and β2 globins of scaled reptiles (Squamata) (Additional file 2: Table S1).

Whereas Phe GH5 has been replaced by Leu or Tyr in the sauropsid α globins [30], the position has been highly conserved in the β globins with a very few exceptions (Figure 1). We identified a Phe- > Leu substitution in the β1 globin of five marine gadoid species and gilthead seabream (*Sparus aurata*), but also the lobe-finned fish Comoran coelacanth (*Latimeria chalumnae*). Further, the freshwater gadiform species burbot (*Lota lota*) β1 globin possessed Cys at position GH5, while Ile was identified in the large yellow croaker (*Larimichthys crocea*). Gadoids possess up to five different β globins [22,33,38], but replacements of Phe GH5 were only found in the β1 chain. Similarly, the second β globin of gilthead seabream and Comoran coelacanth contains the conserved Phe GH5. Based on the nucleotide sequences available we predicted the sequential point mutations causing the identified Phe GH5 replacements (Figure 1).

Modelled 3D structure of mutated α1β1 contacts

Computational mutation analysis was performed to examine the structural changes at the α1β1 interface in the Atlantic cod hemoglobin caused by replacing Phe122β (GH5) with Leu or Met, but also with the Cys residue identified in burbot. The involvement of the amino acid replacement in the allosteric transition was addressed by generating computational models of both the T- and R-states of the Atlantic cod Hb3 tetramer comprising α1-α1-β1-β1 [38]. The resulting structures passed all stereochemical and geometric checks in PROCHECK and showed that Phe122β forms van der Waals contacts across

the interface with Arg31α, Val108α and Ile112α in both T- and R- states (Figure 2, Table 1). Both the Phe- > Leu and Phe- > Cys substitutions slightly decreased the number of intersubunit contacts in the T-state compared to the R-state, while the Phe- > Met change did not promote any difference in contacts at the α1β1 interface. We therefore predict a mild destabilization of the T-state as a consequence of the Leu or Cys replacement that might slightly increase the oxygen binding affinity of these variants. In contrast, replacing Phe122β with Leu in human HbA caused no difference in the number of van der Waals contacts with Arg31α and Val107α in the T-and R-state (Table 1). Consistently, the Hb-Bushey mutant exhibited identical stability and oxygen binding affinity as HbA [29].

Figure 2 Modeling of Phe122β1- > Leu replacement in Atlantic cod hemoglobin. Superimposition of the three-dimensional model structures of Phe122β (red) and Leu122β (blue) variants in **A)** deoxy- and **B)** oxy-states. Surface structure of the α1β1 dimer is displayed with the α- and β-chains highlighted in dark gray and light gray, respectively. The B, G helices of α-chain (yellow) and the G, H helices and GH corner of β-chain (orange) are shown in ribbon representation. The closest distances from 122β1 residue to the α-chain are shown in Å.

Figure 1 Mutations and amino acid replacements identified at position GH5 in gnathostome α and β globins. Possible sequential point mutations (underlined) are indicated, while ignoring any phylogenetic relationship between the given species.
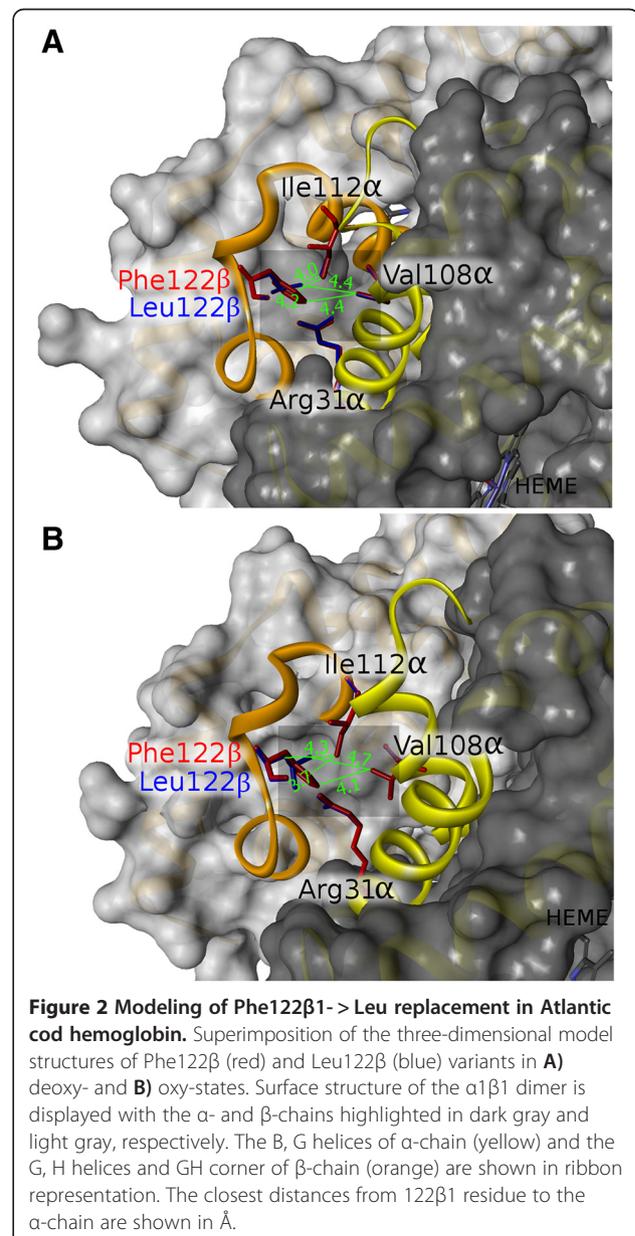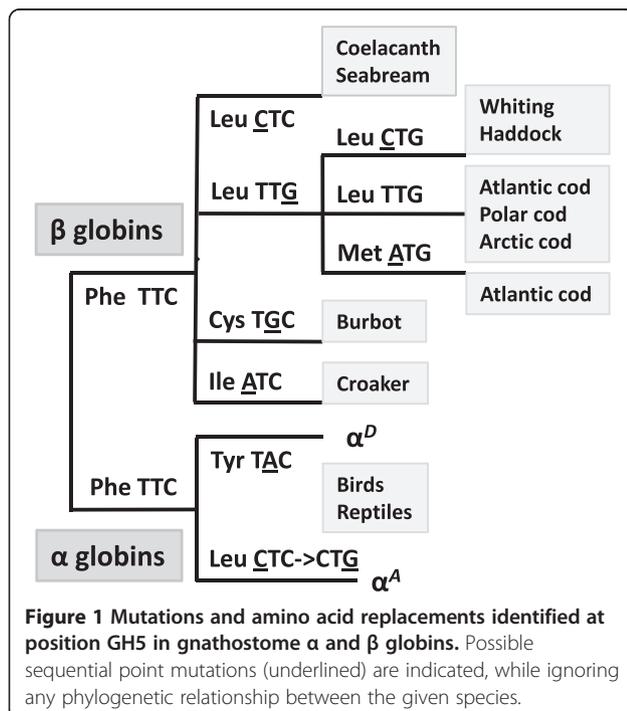
Table 1 Residues of α chain and number of atoms at a maximum distance of 4.5 Å from β GH5 position in Atlantic cod, burbot and human hemoglobin at the deoxy (T) and oxy (R) state

		Contacts	Number of atoms	
			Deoxy	Oxy
Atlantic cod	Phe	Arg31α	3	2
		Val108α	1	1
		Ile112 α	1	1
	Leu	Arg31α	3	2
		Val108α	-	1
		Ile112 α	1	1
	Met	Arg31α	3	2
		Val108α	1	1
		Ile112 α	1	1
Burbot	Cys	Arg31α	3	3
Human	Phe	Arg31α	3	3
		Val107α	1	1
	Leu	Arg31α	3	3
		Val107α	1	1
	Ser	Arg31α	3	3

Leu122β1Met polymorphism in trans-Atlantic cod populations

The distribution of the Leu122β1(GH5)Met polymorphism in trans-Atlantic populations of Atlantic cod was examined by SNP genotyping a total of 560 adult fish representing 15 populations (Additional file 3: Table S2). Population pairwise F_{ST} values demonstrated a clear genetic separation between northeastern and northwestern cod samples with intermediate Icelandic and Greenland populations (Additional file 4: Table S3). The Met allele dominated in the three Canadian populations (66.0-68.4%), and almost 50% of the Labrador and Newfoundland cod were homozygous for this allele (Figure 3). The allele frequency of Met122 decreased to 28.0 and 14.5%, respectively, in the Sisimiut and Nuuk populations, and only one Greenland fish was identified as homozygous Met. Even lower frequencies of this allele were found in Icelandic coastal and frontal populations (7.8%), Northeast Arctic cod (NEAC) population in Barents Sea (3%), Faroe bank and plateau populations (1.5%) and in spawning NEAC fish outside Lofoten islands (1%), while the Met allele was not identified in the North Sea, Kattegat and Baltic Sea populations (Additional file 5: Table S4). All samples conformed to Hardy-Weinberg expectations (Additional file 6: Table S5).

The polymorphic positions Met55Val and Lys62Ala of Atlantic cod β1 globin that discriminate between the HbI-1 (Met55-Lys62) and HbI-2 (Val55-Ala62) isoforms [22] were included in the genotyping to examine the distribution of the haplotypes. The Met122 variant was not identified in any HbI-1 fish and so exhibited the single haplotype Met55-Lys62-Leu122, while the Leu122Met polymorphism in the HbI-2 fish resulted in the two haplotypes Val55-Ala62-Leu122 and Val55-Ala62–Met122. The latter haplotype was rarely found in northeast-Atlantic populations, while the Val55-Lys62-Met122 recombination was identified in two individuals from Labrador and Georges Bay, and a Sisimiut recombinant exhibited the Val55-Lys62-Leu122 haplotype.

Positive selection

To detect genetic signatures of positive selection acting on the evolution of the β1 globin gene in Gadiformes, a site-specific likelihood analysis was performed using PAML and Datamonkey (Table 2). The one-ratio M0 model, which assumes the same ω ratio between non-synonymous (dN) and synonymous (dS) substitutions for all branches, had a log-likelihood value of −1092.88 with ω = 0.26. The log-likelihood value for the M3 model with three discrete site classes was −1064.41 with $\omega_0 = 0.08$, $\omega_1 = 0.08$ and $\omega_2 = 2.21$. Comparison of the two models using a likelihood ratio test revealed that model M3 provided a significantly better fit to our data set. When compared to a χ^2 distribution with 4 degrees of freedom (df), the difference in log-likelihood values (2ΔLnL) of 56.94 supports the rejection of the one-ratio model M0 (p = 0). Similarly, the other two evolution models that allowed for positive selection fitted the data better than those that did not (M2a versus M1a, p = 0.05; M8 versus M7, p = 0.01). Models M2a, and M3 identified positively selected sites at positions 6 (p = 0.995), 9 (p = 0.999), 13 (p = 0.997), 23 (p = 0.995), 55 (p = 0.995), 62 (p = 1), 122 (p = 1) and 123 (p = 0.993) with a ω ratio of 2.22 and 2.21, respectively. The same codons were identified in model M8 with ω = 2.22 and similar probability values greater than 0.99. The REL algorithm implemented in Datamonkey recognized positively selected sites with a Bayes factor greater than 50 at positions 6, 9, 13, 55, 62 and 122. The consensus from all methods indicated that codons 6 (Ala, Thr or Tyr), 9 (Arg, Ser, Thr or Lys), 13 (Ala, Thr or Gln), 55 (Met, Val or Leu), 62 (Lys, Ala, Asn, Gln or Arg) and 122 (Leu, Met or Cys) are under positive selection (Table 2; Additional file 7: Figure S2).

Discussion

The vast phylogenetic variation and adaptive modifications of the hemoglobin molecule have fascinated scientists since the pioneering work of Braunitzer [39] identifying only eight invariant positions in multiple vertebrate hemoglobins. The post-genomic era has later provided numerous hemoglobin protein and gene sequences from a variety of organisms and even from extinct species to investigate evolutionary conserved positions as well as adaptive changes in
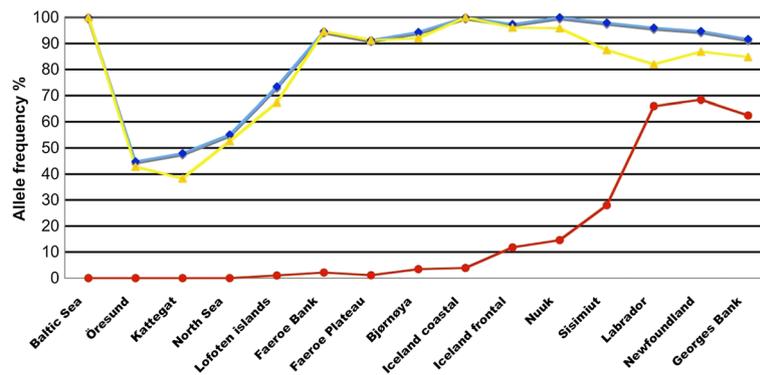
Figure 3 Allele frequencies of the three polymorphic sites in Atlantic cod β1 globin from samples collected across the North Atlantic. Met55Val (Val: blue line), Lys62Ala (Ala: yellow line) and Leu122Met (Met: red line). See Additional file 3: Table S2 for sampling locations and Additional file 5: Table S4 for SNP alleles frequencies.

specific lineages. Here we show that all α globins available from the gnathostomes exhibit the invariant Arg B12, which is predominantly coded by the AGG codon, whereas the six different Arg codons are found in other positions of the α globins in warm- and cold-blooded vertebrates [40]. Genomic sequences revealed that the Arg B12 codon spans the exon 1-exon 2 boundary of a phase 2 intron, and AG↓G is the most frequent signal for exon splicing [41]. Hence, the functional constraints of the invariant position are accompanied by independent requirements of exon splicing. Accordingly, the AGG- > ACG mutation in human Hb Monroe (Arg30β- > Thr) inhibited pre-mRNA splicing and no mutant protein and transcript was detected [27,42]. Further, the rare AGA codon is found in the β-like δ globin gene of primates, and higher primates produce only a small amount of Hb A_2 ($α_2δ_2$, <6% of total hemoglobin), while δ globin is a silent gene in Old World monkeys [43,44].

This contrasts with the gadoid *β1 globin*, which also has the AGA codon, but is highly expressed in the adult Atlantic cod at similar levels of β2 globin containing the conserved AGG codon [38,45]. The additional mutations identified at this position in β globins of pig, tree shrew and scaled reptiles raise the question about the importance of this codon for correct mRNA splicing. Possible effects of these amino acid replacements on hemoglobin function warrant further studies although the multiple changes identified at subunit contacts and heme contacts in cobra and sea snake hemoglobin appeared compatible with conserved overall functional properties [46,47]. The lizard β1 and β2 globins are probably products of a lizard-specific duplication event, and the phylogenetic positions of the β paralogs suggest that the common reptile ancestor may have possessed a fairly diverse repertoire of β-like globin genes [48].

Table 2 Identification of positively selected sites in gadoids and burbot β1 globin by maximum likelihood analysis using various models of evolution

Method	Model	Parameter estimates	Ln likelihood	Model comparison	Positively selected sites[a]
CODEML	M0: neutral	$ω = 0.26$	−1092.88		None
	M1a: nearly neutral	$ω_0 = 0.05$, $ω_1 = 1.00$	−1067.42		Not allowed
		$p_0 = 0.80$, $p_1 = 0.20$			
	M2a: positive selection	$ω_0 = 0.08$, $ω_1 = 1.00$, $ω_2 = 2.22$	−1064.41	M2 vs M1	6, 9, 13, 23, 55, 62, 122, 123
		$p_0 = 0.86$, $p_1 = 0$, $p_2 = 0.14$		$2ΔLnL = 6.02$, df = 2, p = 0.05	
	M3: discrete	$ω_0 = 0.08$, $ω_1 = 0.08$, $ω_2 = 2.21$	−1064.41	M3 vs M0	6, 9, 13, 23, 55, 62, 122, 123
		$p_0 = 0.40$, $p_1 = 0.46$, $p_2 = 0.14$		$2ΔLnL = 56.94$, df = 4, p = 0	
	M7: β	$p = 0.07$, $q = 0.21$	−1069.19		Not allowed
	M8: β + ωS > 1	$p = 9.24$, $q = 99$	−1064.45	M8 vs M7	6, 9, 13, 23, 55, 62, 122, 123
		$ω = 2.22$		$2ΔLnL = 9.48$, df = 2, p = 0.01	
		$p_0 = 0.86$, $p_1 = 0.14$			
HYPHY	REL	$ω = 2.72$			6, 9, 13, 55, 62, 122

[a]Only positively selected codons with Bayesian posterior probabilities equal or greater than 99% are indicated.

The highly conserved Arg B12 interacts with Phe GH5, which has been replaced by Leu and Tyr, respectively, in the sauropsid α^A and α^D chains forming the major HbA and minor HbD together with a common β chain [30,49]. The higher intrinsic oxygen affinity of avian HbD compared to HbA was proposed to involve substitutions at three positions in the α1β1 packing contacts, including the Leu117α(GH5)->Tyr change [30]. This mutation might represent an evolutionary adaptation to air breathing in reptiles and birds, while the reported Leu55β->Ser and Pro119α->Ala replacements have further increased the oxygen affinity in high-altitude birds [18,19]. We found Phe GH5 in the β globins of all sarcopterygians examined, except for the Phe->Leu replacement in a "living fossil"; the Comoran coelacanth. It should be noted that the Phe122β->Leu replacement is also found in the human Hb Bushey mutant, but the stability and binding affinity were shown to be identical to normal HbA [29]. On the other hand, multiple heme contacts and positions involved in sub-unit interface contacts have been replaced in the coelacanth hemoglobin, including the loss of an α1β2 contact that might be responsible for the easy dissociation of the tetrameric molecule [50]. The coelacanth genome harbors two α and two β globin genes, and the phylogenetic tree grouped the adult β1 and embryonic β2 globins together with amphibian embryonic β-chains in a clade that was lost in the amniote tetrapods [51,52]. Whereas the Phe122β->Leu mutation has disappeared in the amniote β globins, the same amino acid change has occurred in the sauropsid α^A chain. Our site-specific analyses identified six consensus codons under positive selection in gadiform β1 globins, and several amino acid substitutions observed at these positively selected sites produce significant changes in charge, size or hydrophobicity, which may affect hemoglobin function. Intriguingly, the identified sites under positive selection include positions 55, 62 and 122, which are polymorphic in Atlantic cod; amino acid substitutions at two of these positions in avian hemoglobin seem to offer a selective advantage in air-breathing vertebrates. Altogether, the results support the theory of Perutz [4] that adaptive changes in hemoglobin functions are caused by a few amino acid substitutions at key positions.

The adaptability of Atlantic cod to variable environmental conditions in Arctic and temperate North Atlantic waters seems to involve several genomic regions containing multiple polymorphic genes [22,53-56]. The selective advantage of possessing functionally different hemoglobin isoforms has been well documented in Atlantic cod, although contradicting results exist [22,57-61]. The HbI-1 and HbI-2 allelic variants differ in oxygen binding affinity and temperature sensitivity and are differentially distributed along a temperature gradient in northeast Atlantic populations [34-36]. The HbI-2 variant predominates in northwest-Atlantic waters, but the Met122β1Leu polymorphism in these populations might further increase the plasticity of this successful species to fluctuating environments. Paleoecological modelling and genetic studies of nuclear and mitochondrial markers suggest that cod populations have survived as least for 100, 000 years on both sides of the Atlantic [62-65]. The Leu122β1Met polymorphism likely originated in the Canadian populations and has expanded in these waters probably over a short historical time and driven by positive selection acting on this codon. Consistently, strong temporal shifts were observed in several gene-associated SNP loci in Canadian populations over an 80-year period indicating ongoing selection over short time-scales [66]. The intermediate frequencies of the Met122 allele in West Greenland populations and the rare distribution in Icelandic populations are supported by the occasional migration of adult cod from Canada to West-Greenland waters and the age-specific migration towards East Greenland and Iceland [67,68]. Consistently, postglacial gene flow was suggested by the spatiotemporal SNP analysis of trans-Atlantic cod populations demonstrating that samples from West Greenland offshore showed the greatest genetic affinity to Canada [69].

Conclusions

The importance of the α1β1 and α2β2 subunit contacts for hemoglobin stability and oxygen binding affinity is strongly supported by the conservation of the interacting residues Arg B12 and Phe GH5 in the gnathostomes. On the other hand, amino acid replacements identified at these positions in β globins of scaled reptiles and gadiform fishes might offer a selective advantage under certain conditions as indicated by the modeled interactions and genetic signatures of positive selection in the latter group. Intriguingly, the intradimer contacts B12-GH5 and H2-D6 are both polymorphic in Atlantic cod, while amino acid replacements at these positions in avian hemoglobin seem to be beneficial for air-breathing.

Methods

Vertebrate globin sequences

Conserved positions in the α and β globins available from gnathostome vertebrates were analysed by multiple sequence alignment of sequences available at http://www.ncbi.nlm.nih.gov and http://Ensembl.org using the BLAST tool on NCBI. Globin transcripts from whiting (*Merlangius merlangus*), haddock (*Melanogrammus aeglefinus*) and burbot were retrieved from the cod genome database at http://codgenome.no.

Protein modeling

Sequences of globin chains with known 3D structure were selected based upon similarity with Atlantic cod α1

and β1 globins using PSI-BLAST (blast.ncbi.nlm.nih.gov/ Blast). The homology model of T-state Atlantic cod hemoglobin was based on the structures of Antarctic rock cod (*Trematomus bernacchii*) (Protein Data Bank (PDB) 1HBH), the Dusky notothen (*Trematomus newnesi*) (PDB 2AA1), bluefin tuna (*Thunnus thynnus*) (PDB 1V4W), and rainbow trout (*Oncorhynchus mykiss*) (PDB 1OUT). For R-state Atlantic cod hemoglobin, the structures of the red-tailed Brycon (*Brycon cephalus*) (PDB 3BCQ), bluefin tuna (PDB 1V4U) and rainbow trout (PDB 1OUU) were used as templates. Sequence alignments were carried out using ClustalW (clustalw.genome.ad.jp/). Based on ClustalW alignments, three-dimensional models were generated by comparative protein modeling with MODELLER program [70] as implemented in Discovery Studio 3.5 (Accelrys Inc.). Twenty models, optimized by a short simulated annealing refinement protocol available in MODELLER, were generated for each globin chain. The geometrical consistency of the model was evaluated based on PDF violations provided by the program. The BUILD_MUTANT module of MODELLER was then used for computational mutagenesis experiments. One hundred alternative conformations of each hemoglobin mutant at position 122β position were generated by the program. After examination of the models with Discovery Studio (Accelrys Inc.), a representative model from each set was chosen that had few restraint violations and favourable stereochemical properties as determined using PROCHECK [71].

Population genotyping and analyses

A total of 560 adult Atlantic cod were collected during 2001−2011 from 15 trans-Atlantic populations (Additional file 2: Table S1). Genomic DNA was extracted using the DNeasy blood and tissue kit (Qiagen) from fin clips, spleen or muscle tissues stored in 95% ethanol or RNA-later (Qiagen). The β1 globin variants were genotyped by single nucleotide polymorphism (SNP) analysis. MassArrayTyper (Version 4 running Assay Editor 4.0.20.5) was used to design primers for multiplex PCR to amplify a region containing Met55Val and Lys62Ala, and a region containing the Met122Leu polymorphism. The sequences for the two primer pairs are: Sense, 5′-acgttg gatgtttggcgacctgagcaccga-3′, antisense, 5′-acgttggatgtg gtccagagccgtcctca-3′, and sense; 5′- acgttggatgtttcagctg ctgtgtgagtg-3′, antisense; 5′-acgttggatgacaggtacttctgcc acgc-3′. The SNPs were determined using downstream or upstream extension primers with the respective sequences: 5′- accgacgccgctatt-3′, 5′-ggccacgacgccgtgc-3′ and 5′-ctg catctccgggctca-3′. Samples were genotyped using a MassArray4 instrument, and the genotypes were assigned with MassArrayTyper (Version 4 running Typer Analyzer 4.0.22.67) and manually inspected using the MassArray Typer v. 3.3 software. SNP heterozygosities, linkage disequilibrium and population differentiation (F_{ST}), as well as statistical significances were calculated using Genepop 4.0.10 [72].

Analyses of positive selection

A 146-codon alignment spanning the full-length coding sequence of Atlantic cod β1 globin was obtained with MUSCLE (www.ebi.ac.uk/Tools/msa/muscle/) using a total of 13 sequences from six gadiformes, namely Atlantic cod, whiting, haddock, burbot, polar cod (*Boreogadus saida*) and Arctic cod (*Arctogadus glacialis*). This alignment was used for Bayesian inference of phylogeny, as previously reported [73]. The GTR evolutionary model with gamma-distributed rate variation across sites and a proportion of invariable sites was selected and 4 Markov chains were run for 500,000 generations sampling every 10th generation. A consensus tree was built after burning the first 5,000 trees. Individual sites under positive (diversifying) selection were identified using the maximum likelihood methods implemented in the CODEML program of PAML v4.7 [74], as detailed in [75]. We investigated the ratio (ω) between non-synonymous (dN) and synonymous (dS) substitutions using several branch-site models, which allow ω to vary among codons in the β1 globin protein. The models of positive selection used were M2a with three site classes (ω =1, $0 < \omega < 1$ and $\omega > 1$), M3 with three discrete site classes of different ω values and M8 with a β distribution of sites, including one class site with $\omega > 1$. These were then compared with the appropriate nested neutral or nearly neutral evolution models M0, M1a and M7 by likelihood ratio tests. Model M0 assumes the same ω ratio for all branches in the phylogeny and for all codons in the β1 globin gene, whereas model M1a allows for two site classes (ω = 1 and $0 < \omega < 1$) and model M7 uses a β distribution of class sites that does not allow for selection ($0 < \omega < 1$). Both naïve and Bayes empirical Bayes were used to determine Bayesian posterior probabilities (p) of positively selected sites. In addition, our data set was analyzed with the random effects likelihood model of molecular evolution (REL) implemented in Datamonkey to identify specific sites under positive selection [76].

Additional files

Additional file 1: Figure S1. Sequence alignment of human and Atlantic cod α and β globins. The A-H helices are indicated together with positions B12 and GH5 (arrow). The polymorphic sites Met55β1Val, Lys62β1Ala and Leu122β1Met in Atlantic cod are included. GeneBank accession numbers: Human: α; ABD95911, β; AAA16334. Atlantic cod: α1; ACV69830, β1; ACV69840.

Additional file 2: Table S1. Arg31β(B12) replacements identified in gnathostome species. Accession number is indicated for the protein, or mRNA when available.

Additional file 3: Table S2. Atlantic cod sample location and size. The Baltic cod were pooled into one sample in the analyses. The data storage tagged (DST) Icelandic cod have been previously described [77,78].

peptide and transcript but also interferes with the expression of wild allele. *Haematologica* 2007, **92**:1715–1716.

28. Bezerra MAC, Albuquerque DM, Santos MNN, Kimura EM, Jorge SEDC, Oliveira DM, Domingues BLTB, Peres JC, Araújo AS, Costa FF, Sonati MF: **Two new unstable haemoglobins leading to chronic haemolytic anaemia: Hb Caruaru [β122 (GH5) Phe → Ser], a probable case of germ line mutation, and Hb Olinda [β22 (B4) - 25 (B7)], a deletion of a 12 base-pair sequence.** *Eur J Haematol* 2009, **83**:378–382.

29. Wajcman H, Drupt F, Henthorn JS, Kister J, Prehu C, Riou J, Promé D, Galactéros F: **Two new variants with the same substitution at position β122: Hb Bushey [β122(GH5)Phe→Leu] and Hb Casablanca [β65(E9) lys→Met; β122(GH5)Phe→Leu].** *Hemoglobin* 2000, **24**:125–132.

30. Grispo MT, Natarajan C, Projecto-Garcia J, Moriyama H, Weber RE, Storz JF: **Gene duplication and the evolution of hemoglobin isoform differentiation in birds.** *J Biol Chem* 2012, **287**:37647–37658.

31. Lacerra G, Scarano C, Musollino G, Flagiello A, Pucci P, Carestia C: **Hb Foggia or α117(GH5)Phe - > Ser: a new α2 globin allele affecting the αHb-AHSP interaction.** *Haematologica* 2008, **93**:141–142.

32. Yu X, Mollan TL, Butler A, Gow AJ, Olson JS, Weiss MJ: **Analysis of human alpha globin gene mutations that impair binding to the α hemoglobin stabilizing protein.** *Blood* 2009, **113**:5961–5969.

33. Borza T, Stone C, Gamperl AK, Bowman S: **Atlantic cod (*Gadus morhua*) hemoglobin genes: multiplicity and polymorphism.** *BMC Genet* 2009, **10**:51.

34. Sick K: **Haemoglobin polymorphism of cod in North Sea and north Atlantic Ocean.** *Hereditas* 1965, **54**:49–69.

35. Frydenberg O, Møller D, Nævdal G, Sick K: **Haemoglobin polymorphism in Norwegian cod populations.** *Hereditas* 1965, **53**:257–271.

36. Wetten OF, Wilson RC, Andersen Ø: **High-resolution melting analysis of common and recombinant genotypes of the Atlantic cod (*Gadus morhua*) hemoglobin β1 gene in trans-Atlantic populations.** *Can J Fish Aquat Sci* 2012, **69**:525–531.

37. Yang Z, Nielsen R, Goldman N, Pedersen AM: **Codon-substitution models for heterogeneous selection pressure at amino acid sites.** *Genetics* 2000, **155**:431–449.

38. Verde C, Balestrieri M, de Pascale D, Pagnozzi D, Lecointre G, di Prisco G: **The oxygen transport system in three species of the boreal fish family gadidae - molecular phylogeny of hemoglobin.** *J Biol Chem* 2006, **281**:22073–22084.

39. Braunitzer G: **Phylogenetic variation in the primary structure of hemoglobins.** *J Cell Physiol Sup 1* 1966, **67**:1–20.

40. Hamada K, Horiike T, Kanaya S, Nakamura H, Ota H, Yatogo T, Okada K, Nakamura H, Shinozawa T: **Changes in body temperature pattern in vertebrates do not influence the codon usages of alpha-globin genes.** *Genes Genet Syst* 2002, **77**:197–207.

41. Ruvinsky A, Ward W: **Intron framing exonic nucleotides: a compromise between protein coding and splicing constraints.** *Open Evol* 2008, **2**:7–12.

42. Vidaud M, Gattoni R, Stevenin J, Vidaud D, Amselem S, Chibani J, Rosa J, Goossens M: **A 5′ splice-region G–––C mutation in exon 1 of the human beta-globin gene inhibits pre-mRNA splicing: a mechanism for beta + –thalassemia.** *Proc Natl Acad Sci USA* 1989, **86**:1041–1045.

43. Boyer SH, Crosby EF, Noyes AN, Fuller GF, Leslie SE, Donaldson LJ, Vrablik GR, Schaefer EV Jr, Thurmon EF: **Primate hemoglobins: some sequences and some proposals concerning the character of evolution and mutation.** *Biochem Genet* 1971, **5**:405–448.

44. Martin SL, Zimmer EA, Kan YW, Wilson AC: **Silent δ-globin gene I Old World monkeys.** *Proc Natl Acad Sci USA* 1980, **77**:3563–3566.

45. Wetten OF, Nederbragt AJ, Wilson RC, Jakobsen KS, Edvardsen RB, Andersen Ø: **Genomic organization and gene expression of the multiple globins in Atlantic cod: conservation of globin-flanking genes in chordates infers the origin of the vertebrate globin clusters.** *BMC Evol Biol* 2010, **10**:315.

46. Islam A, Persson B, Zaidi ZH, Jörnvall H: **Sea snake (*Microcephalopsis gracilis*) hemoglobin: primary structure and relationships to other forms.** *J Protein Chem* 1990, **8**:533–541.

47. Naqvi S, Abbasi A, Zaidi ZH: **Primary structure of hemoglobin from cobra *Naja naja naja*.** *J Protein Chem* 1994, **13**:669–679.

48. Hoffmann FG, Storz JF, Gorr TA, Opazo JC: **Lineage-specific patterns of functional diversification in the δ- and β- globin gene families of tetrapod vertebrates.** *Mol Biol Evol* 2010, **27**:1126–1138.

49. Moss G, Hamilton E: **Chicken definitive erythrocyte haemoglobins.** *Biochim Biophys Acta* 1974, **371**:379–391.

50. Amemiya CT, Alföldi J, Lee AP, Fan S, Philippe H, MacCallum I, Braasch I, Manousaki T, Schneider I, Rohner N, Organ C, Chalopin D, Smith JJ, Robinson M, Dorrington RA, Gerdol M, Aken B, Biscotti MA, Barucca M, Baurain D, Berlin AM, Blatch GL, Buonocore F, Burmester T, Campbell MS, Canapa A, Cannon JP, Christoffels A, De Moro G, Edkins AL, *et al*: **Comparative analysis of the genome of the African coelacanth, *Latimeria chalumnae*, sheds light on tetrapod evolution.** *Nature* 2013, **496**:311–316.

51. Schwarze K, Burmester T: **Conservation of globin genes in the "living fossil" *Latimeria chalumnae* and reconstruction of the evolution of the vertebrate globin family.** *Biochim Biophys Acta* 2013, **1834**:1801–1812.

52. Gorr T, Kleinschmidt T, Sgouros JG, Kasang L: **A "living fossil» sequence: primary structure of the coelacanth (*Latimeria chalumnae*) hemoglobin-evolutionary and functional aspects.** *Biol Chem Hoppe Seyler* 1991, **372**:599–612.

53. Andersen Ø, De Rosa MC, Pirolli D, Tooming-Klunderud A, Petersen PE, André C: **Polymorphism, selection and tandem duplication of transferrin genes in Atlantic cod (*Gadus morhua*) - Conserved synteny between fish monolobal and tetrapod bilobal transferrin loci.** *BMC Genet* 2011, **12**:51.

54. Bradbury IR, Hubert S, Higgins B, Bowman S, Borza T, Paterson IG, Snelgrove PVR, Morris CJ, Gregory RS, Hardie D, Hutchings JA, Ruzzante DE, Taggart CT, Bentzen P: **Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish.** *Evol Appl* 2013, **6**:450–461.

55. Hemmer-Hansen J, Nielsen ME, Therkildsen NO, Taylor MI, Ogden R, Geffen AJ, Bekkevold D, Helyar S, Pampoulie C, Johansen T, Carvalho GR: **A genomic island linked to ecotype divergence in Atlantic cod.** *Mol Ecol* 2013, **22**:2653–2667.

56. Karlsen BO, Klingan K, Emblem A, Jørgensen TE, Jueterbock A, Furmanek T, Hoarau G, Johansen SD, Nordeide JT, Moum T: **Genomic divergence between the migratory and stationary ecotypes of Atlantic cod.** *Mol Ecol* 2013. doi:10.1111/mec.12454.

57. Brix O, Forås E, Strand I: **Genetic variation and functional properties of Atlantic cod hemoglobins: introducing a modified tonometric method for studying fragile hemoglobins.** *Comp Biochem Physiol A Mol Integr Physiol* 1998, **119**:575–583.

58. Brix O, Thorkildsen S, Colosimo A: **Temperature acclimation modulates the oxygen binding properties of the Atlantic cod (*Gadus morhua* L.) genotypes-Hbl*1/1, Hbl*1/2, and Hbl*2/2-by changing the concentrations of their major hemoglobin components (results from growth studies at different temperatures).** *Comp Biochem Physiol A Mol Integr Physiol* 2004, **138**:241–251.

59. Petersen MF, Steffensen JF: **Preferred temperature of juvenile Atlantic cod *Gadus morhua* with different haemoglobin genotypes at normoxia and moderate hypoxia.** *J Exp Biol* 2003, **206**:359–364.

60. Gamperl AK, Busby CD, Hori TS, Afonso LO, Hall JR: **Hemoglobin genotype has minimal influence on the physiological response of juvenile Atlantic cod (*Gadus morhua*) to environmental challenges.** *Physiol Biochem Zool* 2009, **82**:483–494.

61. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, Wetten OF, Lanzén A, Winer R, Knight J, Vogel JH, Aken B, Andersen O, Lagesen K, Tooming-Klunderud A, Edvardsen RB, Tina KG, Espelund M, Nepal C, Previti C, Karlsen BO, Moum T, Skage M, Berg PR, Gjøen T, Kuhl H, *et al*: **The genome sequence of Atlantic cod reveals a unique immune system.** *Nature* 2011, **477**:207–210.

62. Arnason E: **Mitochondrial cytochrome B DNA variation in the high-fecundity Atlantic cod: trans-atlantic clines and shallow gene genealogy.** *Genetics* 2004, **166**:1871–1885.

63. Bigg GR, Cunningham CW, Ottersen G, Pogson GH, Wadley MR, Williamson P: **Ice-age survival of Atlantic cod: agreement between palaeoecology models and genetics.** *Proc R Soc B* 2008, **275**:163–172.

64. Marshall SM, Carr HG: **Intraspecific phylogeographic genomics from multiple complete mtDNA genomes in Atlantic cod (*Gadus morhua*): origins of the "codmother," transatlantic vicariance and midglacial population expansion.** *Genetics* 2008, **180**:381–389.

65. Bradbury IR, Hubert S, Higgins B, Borza T, Bowman S, Paterson IG, Snelgrove PVR, Morris CJ, Gregory RS, Hardie DC, Hutchings JA, Ruzzante DE, Taggart CT, Bentzen P: **Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature.** *Proc R Soc B* 2013, **277**:3725–3734.

66. Therkildsen NO, Hemmer-Hansen J, Hedeholm RB, Wisz MS, Pampoulie C, Meldrup D, Bonanomi S, Retzel A, Olsen SM, Nielsen EE: **Spatiotemporal**

SNP analysis reveals pronounced biocomplexity at the northern range margin of Atlantic cod *Gadus morhua*. *Evol Appl* 2013, **6**:690–705.

67. Hansen P: **Studies on the biology of the cod in Greenland waters.** *Rapports et Proces-Verbaux des Reunions du Conseil Int pour l'Exploration de la Mer* 1949, **123**:1–77.

68. Storr-Paulsen M, Wieland K, Hovgård H, Rätz HJ: **Stock structure of Atlantic cod (*Gadus morhua*) in West Greenland waters: implications of transport and migration.** *ICES J Mar Sci* 2004, **61**:972–982.

69. Therkildsen NO, Hemmer-Hansen J, Als TD, Swain DP, Morgan MJ, Trippel EA, Palumbi SR, Meldrup D, Nielsen EE: **Microevolution in time and space: SNP analysis of historical DNA reveals dynamic signatures of selection in Atlantic cod.** *Mol Ecol* 2013, **22**:2424–2440.

70. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints.** *J Mol Biol* 1993, **234**:779–815.

71. Laskowski RA, MacArthur MW, Moss DS, Thorntorn JM: **PROCHECK: a program to check the stereochemical quality of proteins structures.** *J Appl Crystallogr* 1993, **26**:283–291.

72. Rousset F: **Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux.** *Mol Ecol Resources* 2008, **8**:103–106.

73. Sundaram AY, Kiron V, Dopazo J, Fernandes JM: **Diversification of the expanded teleost-specific toll-like receptor family in Atlantic cod. Gadus morhua.** *BMC Evol Biol* 2012, **12**:256.

74. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24**:1586–1591.

75. Fernandes JM, Ruangsri J, Kiron V: **Atlantic cod piscidin and its diversification through positive selection.** *PLoS One* 2010, **5**:e9501.

76. Poon AF, Frost SD, Pond SL: **Detecting signatures of selection from DNA sequences using Datamonkey.** *Methods Mol Biol* 2009, **537**:163–183.

77. Pampoulie C, Jakobsdóttir KB, Marteinsdóttir G, Thorsteinsson V: **Are vertical behaviour patterns related to the Pantophysin locus in the Atlantic cod (*Gadus morhua* L.)?** *Behav Genet* 2008, **38**:76–81.

78. Thorsteinsson V, Pálsson ÓK, Jónsdóttir IG, Pampoulie C: **Consistency in the behaviour types of the Atlantic cod: repeatability, timing of migration and geo-location.** *Mar Ecol Prog Ser* 2012, **462**:251–260.