

Minireview

## Globin genes on the move

Ross C Hardison

Address: Center for Comparative Genomics and Bioinformatics, Huck Institutes of Life Sciences and Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA. Email: rch8@psu.edu

Published: 20 November 2008

*Journal of Biology* 2008, **7**:35 (doi:10.1186/jbiol92)

The electronic version of this article is the complete one and can be found online at <http://jbiol.com/content/7/9/35>

© 2008 BioMed Central Ltd

### Abstract

Recent data published in *BMC Biology* from the globin gene clusters in platypus, together with data from other species, show that  $\beta$ -globin genes transposed from one chromosomal location to another. This resolves some controversies about vertebrate globin gene evolution but ignites new ones.

### The importance of hemoglobin proteins and genes

In metazoans, hemoglobin carries oxygen from the lungs, gills or other respiratory organs to peripheral tissues that need the oxygen for efficient metabolism. Hemoglobin is composed of one or more globin polypeptide chains, which bind the iron-containing cofactor heme. Other organisms use the oxygen-binding and redox capacity of hemoglobin in a variety of other ways. Most multicellular species produce distinctive forms of hemoglobin at different developmental stages.

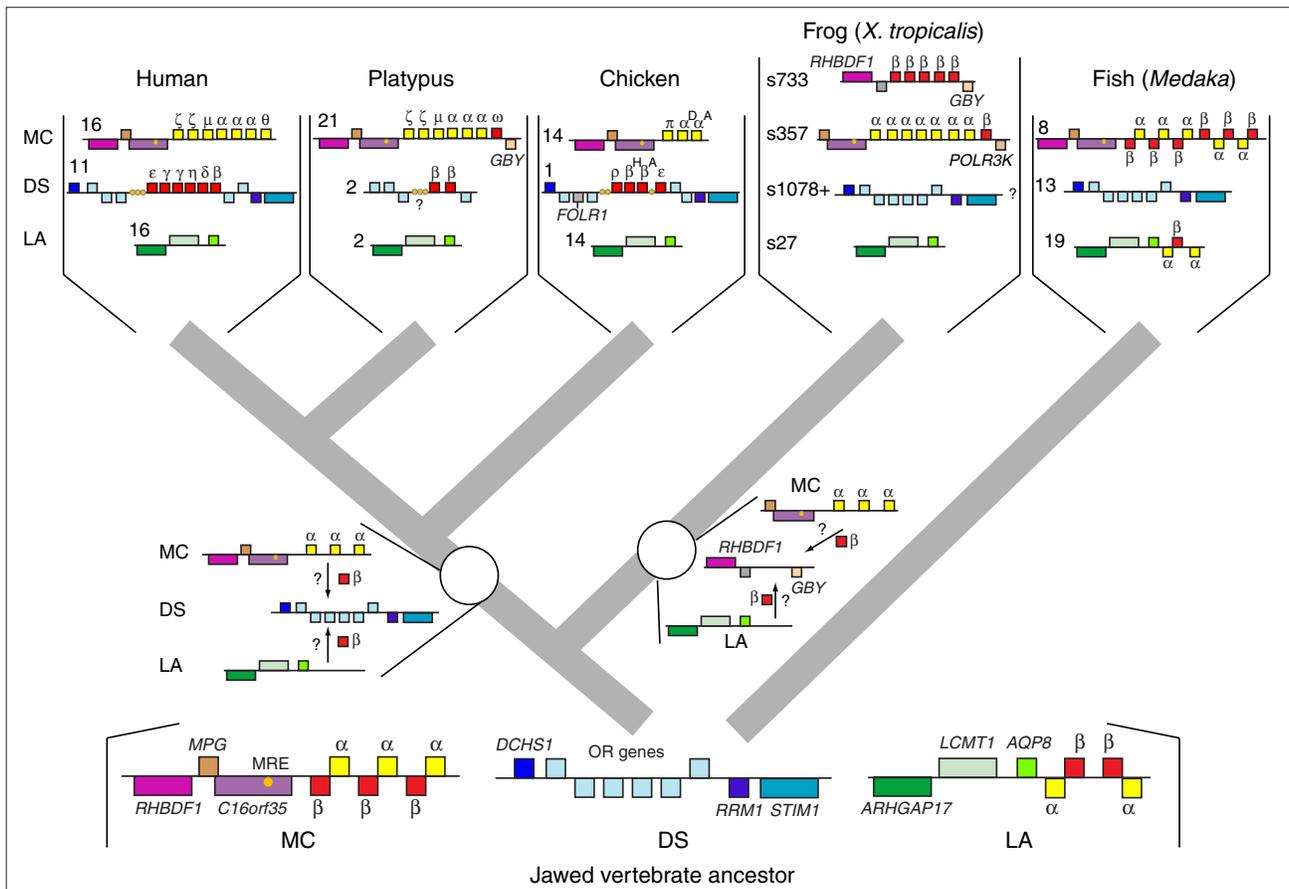
Studies of globin genes and their evolution are driven by a desire to understand how this diversity of protein structures and functions was generated. Current evolutionary ideas about gene duplication and divergence had their origin in studies of hemoglobin genes and other multigene families. In addition, interest in the hemoglobin genes is fueled by the realization that the most common genetic diseases in humans, such as sickle cell disease and inherited anemias, result from pathological variants in the hemoglobin genes. The quest to understand normal hemoglobin production and function and to learn what goes wrong in hemoglobinopathies should provide insights that could lead to effective therapeutic interventions.

The hemoglobins found in erythroid cells of jawed vertebrates (gnathostomes) include two  $\alpha$ -like globins and two  $\beta$ -like

globins, each bound by a heme molecule. Although the hemoglobins produced in embryonic or larval erythrocytes differ from those in adult erythrocytes, each has two copies of a polypeptide related to  $\alpha$ -globin, such as the embryonic  $\zeta$ -globin, and two copies of a polypeptide related to  $\beta$ -globin, such as the embryonic  $\epsilon$ -globin. The  $\alpha$ -like and  $\beta$ -like globin gene clusters are on different chromosomes in mammals and birds, but one or more pairs of  $\alpha$ -like and  $\beta$ -like globin genes are found in one locus in amphibians and fish (Figure 1). The arrangement of genes in these contemporary species strongly suggested that either  $\alpha$ -like or  $\beta$ -like globin genes moved to a new chromosomal location in the ancestor of birds and mammals [1]. How this occurred, whether by duplication and specific gene loss or by translocation of one type of gene, was far from clear. This mechanism and the effects such a movement had on the regulation of the genes are intriguing questions, and recent work, such as that of Patel *et al.* published recently in *BMC Biology* [2], has shed much light on them.

### Multiple globin gene clusters

One globin gene cluster is found in all gnathostomes examined, but frequently, perhaps invariably, it is accompanied by another globin gene cluster at a different genetic locus. The globin genes in the common cluster are flanked

**Figure 1**

One model for movement of globin genes during the evolution of jawed vertebrates. Top, gene clusters in contemporary species; bottom, inferred gene arrangements in the last common ancestor to jawed vertebrates; middle, some of the possible gene movements mapped onto an evolutionary tree (thick gray lines). Genes are indicated by boxes, with those above the line transcribed from left to right and those below the line from the right; red,  $\beta$ -like globin genes; yellow,  $\alpha$ -like globin genes; light blue, OR genes; other colors, other genes; small orange circles, major regulatory regions. All the known genes encoding hemoglobins expressed in erythroid cells are shown, as well as other genes that are most consistently diagnostic for the loci. Numbers to the left of each cluster specify the chromosome on which it is located; for the frog gene clusters, the scaffold number (preceded by s) is given. The Greek letter name is specified for hemoglobin genes in human, platypus and chicken, but generic ' $\alpha$ -globin' or ' $\beta$ -globin' is used for frog and fish because the genes are less well characterized. In the *X. tropicalis* genome assembly (version 4.1), scaffolds 733 and 357 are not connected. Maps of the gene clusters were derived from a combination of the assembled genomes and recent publications [2,4,15]. The gene maps are not complete or to scale; see [2,4] or the genome assemblies for more complete information.

by the genes *MPG* and *C16orf35* 'upstream' of the globin genes [3], and I will thus refer to this locus as 'MC', after these two genes (Figure 1). The major DNA region regulating expression of the globin genes (major regulatory element, MRE) is located in an intron of *C16orf35*. Frequently, the gene *RHBDF1* is adjacent to the *MPG* gene. In placental mammals and chickens, only  $\alpha$ -like globin genes (both embryonic and adult) are in this locus, whereas in amphibians [4] and fish [3], the  $\alpha$ -globin genes are paired with  $\beta$ -globin genes (Figure 1). The new analysis of platypus globin genes by Patel *et al.* [2] shows that a  $\beta$ -like globin gene, the  $\omega$ -globin gene, is closely linked to the  $\alpha$ -like globin

genes, just as it is in marsupials [5]. In addition, a platypus homolog to the gene for globin Y (*GBY*), a globin discovered in amphibians, is also in this locus [2]. Given its presence in all gnathostomes examined, we can infer with considerable confidence that the MC locus contained globin genes in the last common ancestor (LCA) of vertebrates (Figure 1).

A second locus contains  $\alpha$ - and  $\beta$ -globin genes in the pufferfish *Fugu rubripes* [6], and examination of the genome assemblies of zebrafish and *Medaka* shows a similar arrangement (Figure 1). The globin genes in this locus are flanked by the genes *LCMT1* and *AQP8*, and thus I will refer to it as the

LA locus. The gene *ARHGAP17* is also part of this locus in many species. The LA locus has a similar arrangement of non-globin genes in the genome assemblies of human, platypus, chicken and frog, but no globin genes. Given that the presence of globin genes in the LA locus appears to be restricted to fish, what is the ancestral arrangement? This cannot be answered with current data. The diagnostic features of the MC, DS (see below) and LA loci are widespread in fish, being present in *Medaka*, pufferfish, *Tetraodon* and zebrafish. Following the model of Gillemans *et al.* [6], Figure 1 shows gene arrangements in the gnathostome LCA that are similar to those in contemporary fish. This model emphasizes the presence of two different loci containing globin genes in many jawed vertebrates, and proposes that this feature is ancestral. However, it is also possible that the presence of globin genes at the LA locus is a derived feature in fish.

A third locus contains only  $\beta$ -like globin genes in placental mammals and chickens. The  $\beta$ -like globin genes are flanked by olfactory receptor (OR) genes [7]. In placental mammals, hundreds of OR genes are in this locus, along with additional multigene families such as *TRIM* genes. We therefore have to look several megabases away from the  $\beta$ -like globin genes to find single-copy genes that are distinctive for this locus, which are *DCHS1* on one side and *STIM1* on the other; I thus refer to the locus as DS (Figure 1). The *RRM1* gene is adjacent to *STIM1* in many species. The new data from platypus [2] reveal a pair of  $\beta$ -globin genes flanked by OR genes. Although the contigs are not sufficiently long to connect these to *DCHS1* and *STIM1*, it is likely that the  $\beta$ -globin genes in platypus are in the DS locus.

A fourth locus for hemoglobin genes is suggested by the gene arrangements in the *Xenopus tropicalis* genome assembly (version 4.1 released by the Joint Genomics Institute). Scaffold 733 is annotated with the gene *RHBDF1*, several  $\beta$ -like globin genes and *GBY* (Figure 1). Fuchs *et al.* [4] speculate that the globin gene clusters in scaffolds 733 and 357 are actually linked; however a more complete assembly and additional analyses are needed to resolve this issue. The genes associated with the DS locus are on short contigs in this assembly, and it is also possible that a future assembly will link the *RHBDF1*- $\beta$ -globin gene cluster with DS; however, it is notable that no OR genes flank the  $\beta$ -globin genes, whereas they do flank other genes in the DS cluster.

### Movement of globin genes

Globin genes are present in the MC locus in all gnathostomes examined, and it is therefore very likely that this was true in the LCA of jawed vertebrates (Figure 1). The DS locus has no globin genes in fish but it has  $\beta$ -like globin genes in amniotes. On the basis of these gene arrangements, Patel *et*

*al.* [2] argue that the  $\beta$ -like globin genes transposed from the MC locus and inserted into the DS locus. Jeffreys *et al.* [1] proposed 28 years ago that the separation of the  $\alpha$ -like and  $\beta$ -like globin genes involved either translocation or duplication followed by divergence. Identifying a likely source (MC) and target (DS) means that it is possible to argue strongly that the separation of these two loci involved a specific kind of translocation - transposition leading to insertion at a new site. This is because the proposed target locus (DS) has no globin genes in fish, but the diagnostic genes flanking the OR genes are present. If this were the ancestral arrangement, then it is simpler to invoke transposition of a  $\beta$ -globin gene into DS than to invoke a more complex series of chromosomal rearrangements.

Thus, the transposition model is strongly supported, but we should keep in mind alternative sources and targets. If the model in Figure 1 is correct in proposing that gene arrangements in the gnathostome LCA were similar to those in contemporary fish, then it seems equally likely that either the LA locus or the MC locus was the source of the transposed  $\beta$ -globin gene(s) in the DS locus. Of course, if the arrangements of globin genes changed between the gnathostome LCA and contemporary fish, then other possibilities arise. For example, the globin genes could have moved from the MC locus into the LA locus along this lineage. If that were the case, then one could propose that the gnathostome LCA had a single locus containing globin genes, the MC locus.

A different movement of  $\beta$ -globin genes may have occurred in the lineage to frogs. Fuchs *et al.* [4], using an early assembly of the *X. tropicalis* genome, propose that one large region contains all the hemoglobin genes and *GBY*. However, in the most recent assembly (version 4.1), scaffolds 733 and 357 are not connected. If indeed these are separate loci, then  $\beta$ -globin genes would have moved from either the MC or LA loci into a different locus flanked by *RHBDF1* and *GBY*.

### The consequences of moving

Despite the ambiguities that remain about several details, it is very likely that the  $\beta$ -like globin genes transposed, perhaps twice, during the evolution of jawed vertebrates. In the LCA of warm-blooded amniotes, the  $\beta$ -globin gene moved into a sea of OR genes that are expressed exclusively in nasal epithelium. How did this gene maintain its erythroid-specific expression? In the DS locus, high-level expression of  $\beta$ -like globin genes in erythroid cells requires an enhancer and also an insulator and boundary element [8,9]. We can expect that the regulatory sequences required *in cis* to the  $\beta$ -like globin genes accompanied them during the transposition; the only alternatives are that the regulatory regions were already present at the target locus (but why

would erythroid regulatory regions be in a locus expressed in nasal epithelium?) or they were formed as a result of the integration of the transposed DNA.

If the  $\beta$ -like globin genes came from the MC locus, then the obvious candidate for the source of the *cis*-regulatory regions is the MRE located in *C16orf35*. The other potential source, LA, is intriguing in this regard. Gillemans *et al.* [6] did not find evidence for a distal regulator in this locus in pufferfish; perhaps the transposed  $\beta$ -like globin gene carried sufficient proximal *cis*-regulatory elements to it to maintain erythroid expression. Tracing the course of evolution of *cis*-regulatory regions over such a long phylogenetic distance is a major challenge, because most are conserved, for example, only in placental mammals but no further [10]. Detailed studies of additional comparison species in each clade will give more power to this analysis.

Once the  $\beta$ -like globin genes entered the DS locus in the LCA of warm-blooded amniotes, their regulatory regions underwent considerable divergence and possibly duplication. They are controlled by multiple regulatory regions in both placental mammals and chickens, but none show clear DNA sequence matches in comparisons of entire loci between mammals and birds [8]. In contrast, the  $\alpha$ -globin MRE shows clear sequence alignments between placental and marsupial mammals [11], and the homologous regions from chickens and fish are active as enhancers [3]. Thus, it will be informative to search the platypus noncoding sequences for evidence of distal regulators and see whether they are related to distal regulatory regions in either placental mammals or birds - or even neither.

What is the fate of genes left at the source loci after transposition? Of course, our ability to detect the genes is much greater if they are still active, and most of the genes included in the diagrams in Figure 1 are still functional. Jeffreys *et al.* [1] speculated that a 'fossil'  $\beta$ -like globin gene would remain at the end of the  $\alpha$ -globin gene complex, and in fact the  $\beta$ -like  $\omega$ -globin gene is found in exactly that position in marsupials and monotremes (Figure 1). However, it is not a pseudogene, but rather it is transcribed and encodes a functional hemoglobin [2,5].

### Controversies resolved and ignited

The discovery of the  $\omega$ -globin gene in marsupials (and now monotremes) revealed another important component of the vertebrate hemoglobins, but accurately placing it in a phylogenetic tree has been challenging. Initial reports [5,12] concluded that  $\omega$ -globin is orthologous to avian  $\beta$ -like globins. If that were true, the  $\beta$ -like globin genes of placental mammals would not be orthologous to the avian  $\beta$ -like

globin genes, which in turn has an impact on the evolutionary analysis of coding and regulatory regions [13]. However, analyses of larger globin gene datasets [2,14,15] show that the  $\omega$ -globins form a distinct clade separate from all the other  $\beta$ -like globins of mammals and birds. Furthermore, the comparative gene mapping of Patel *et al.* [2] show convincingly that the  $\beta$ -like globin genes of birds and placental mammals are flanked by orthologous single-copy genes (DS locus in Figure 1), and thus these globin gene clusters are indeed orthologous.

Only two  $\beta$ -globin genes are present in the platypus DS locus [2,15], which seems similar to the embryonic  $\epsilon$ -globin and adult  $\beta$ -globin pair found in marsupials [16]. However, the platypus  $\beta$ -globin genes are more similar to each other than to any other  $\beta$ -like globin genes [2,15]. In the homologous gene clusters in every other amniote examined to date, the  $\beta$ -like globin gene located at the left end of the cluster (in the orientation shown in Figure 1) is expressed exclusively in embryonic erythroid cells. Thus, we would expect the gene in the comparable location in monotremes to have a similar expression pattern. Indeed, both groups describing this gene cluster call it  $\epsilon$ -globin, but evidence for grouping with therian (placental mammal and marsupial)  $\epsilon$ -globin is, at best, inconsistent [2]. It is called simply  $\beta$ -globin in Figure 1. Both  $\beta$ -globin genes are expressed in adult tissues, but no information is available on expression in embryos [2], so the range of developmental expression has not been determined. On the basis of extensive matches in the flanking regions, Opazo *et al.* [15] propose that the duplication to form the  $\beta$ -globin gene pair in monotremes occurred independently of the duplication that formed the ancestral therian  $\epsilon$ -globin and  $\beta$ -globin genes. In contrast, Patel *et al.* [2] argue that the gene pair had a common origin in mammals (including monotremes), but the  $\beta$ -like globin genes in platypus have become homogenized by gene conversions. As is the case for many evolutionary controversies, examination of homologous gene clusters in additional species, such as another monotreme, the echidna, could shed light on this debate.

The diversity of hemoglobins, their crucial functions, their exquisite regulation and the pathological consequences of some mutations make this a fascinating family of proteins and genes. Exploration of these genes in many different species continues to illuminate some and challenge other evolutionary models. The recent analysis of the globin gene clusters in platypus, combined with data mining in fish and frog genomes, has strongly supported previous ideas about the relationships of the globin gene clusters in vertebrates, with the  $\alpha$ - and  $\beta$ -globin genes in the MC locus deduced as ancestral.

Furthermore,  $\beta$ -globin genes moved into a cluster of OR genes at the DS locus in the LCA of birds and mammals,

and it will be informative to examine the homologous genes in reptiles. A draft genome sequence of the lizard *Anolis carolinensis* has been released by the Broad Institute, and it shows at least one  $\alpha$ -like globin gene in the MC locus, as expected. A full exploration of issues such as those discussed in the platypus globin gene papers [2,15] will probably require targeted exploration of these loci, with the genome assembly serving as a useful launching pad.

Likewise, deeper branchings should be resolved by analysis of globin gene loci in jawless vertebrates. A draft genome of the lamprey *Petromyzon marinus* has been released by the Washington University Genome Center. Again, the contigs are too short to find instant answers, but the genome assembly provides a great starting point for individual investigators. I predict that more clarity will emerge from such pursuits, as well as more surprises and controversies. Globin gene evolution will remain exciting for some time.

### Acknowledgements

This work was supported by NIH grants from NIDDK (DK65806, RH) and by Tobacco Settlement Funds from the Pennsylvania Department of Health. The genome assemblies of the frog *Xenopus tropicalis* and fish *Medaka* are public releases from, respectively, the Joint Genome Institute and a collaboration between the National Institute of Genetics and University of Tokyo, Japan.

### References

1. Jeffreys AJ, Wilson V, Wood D, Simons JP, Kay RM, Williams JG: **Linkage of adult alpha- and beta-globin genes in *X. laevis* and gene duplication by tetraploidization.** *Cell* 1980, **21**:555-564.
2. Patel VS, Cooper SJB, Deakin JE, Fulton B, Graves T, Warren WC, Wilson RK, Graves JAM: **Platypus globin genes and flanking loci suggest a new insertional model for beta-globin evolution in birds and mammals.** *BMC Biol* 2008, **6**:34.
3. Flint J, Tufarelli C, Peden J, Clark K, Daniels RJ, Hardison R, Miller W, Philipson S, Tan-Un KC, McMorrow T, Frampton J, Alter BP, Frischauf AM, Higgs DR: **Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster.** *Hum Mol Genet* 2001, **10**:371-382.
4. Fuchs C, Burmester T, Hankeln T: **The amphibian globin gene repertoire as revealed by the *Xenopus* genome.** *Cytogenet Genome Res* 2006, **112**:296-306.
5. Wheeler D, Hope RM, Cooper SJ, Gooley AA, Holland RA: **Linkage of the beta-like omega-globin gene to alpha-like globin genes in an Australian marsupial supports the chromosome duplication model for separation of globin gene clusters.** *J Mol Evol* 2004, **58**:642-652.
6. Gillemans N, McMorrow T, Tewari R, Wai AW, Burgdorf C, Drabek D, Ventress N, Langeveld A, Higgs D, Tan-Un K, Grosveld F, Philipson S: **Functional and comparative analysis of globin loci in pufferfish and humans.** *Blood* 2003, **101**:2842-2849.
7. Bulger M, Bender MA, von Doorninck JH, Wertman B, Farrell CM, Felsenfeld G, Groudine M, Hardison R: **Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse  $\beta$ -globin gene clusters.** *Proc Natl Acad Sci USA* 2000, **97**:14560-14565.
8. Hardison R: **Organization, evolution and regulation of the globin genes.** In *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*. Edited by Steinberg MH, Forget BG, Higgs DR, Nagel RL. Cambridge: Cambridge University Press; 2001: 95-116.
9. Gaszner M, Felsenfeld G: **Insulators: exploiting transcriptional and epigenetic mechanisms.** *Nat Rev Genet* 2006, **7**:703-713.
10. King DC, Taylor J, Zhang Y, Cheng Y, Lawson HA, Martin J, ENCODE groups for Transcriptional Regulation and Multispecies Sequence Analysis, Chiaromonte F, Miller W, Hardison RC: **Finding cis-regulatory elements using comparative genomics: some lessons from ENCODE data.** *Genome Res* 2007, **17**:775-786.
11. Hughes JR, Cheng JF, Ventress N, Prabhakar S, Clark K, Anguita E, De Gobbi M, de Jong P, Rubin E, Higgs DR: **Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences.** *Proc Natl Acad Sci USA* 2005, **102**:9830-9835.
12. Wheeler D, Hope R, Cooper SJB, Dolman G, Webb GC, Bottema CD, Gooley AA, Goodman M, Holland RAB: **An orphaned mammalian beta-globin gene of ancient evolutionary origin.** *Proc Natl Acad Sci USA* 2001, **98**:1101-1106.
13. Hardison RC: **New views of evolution and regulation of vertebrate beta-like globin gene clusters from an orphaned gene in marsupials.** *Proc Natl Acad Sci USA* 2001, **98**:1327-1329.
14. Aguileta G, Bielawski JP, Yang Z: **Evolutionary rate variation among vertebrate beta globin genes: implications for dating gene family duplication events.** *Gene* 2006, **380**:21-29.
15. Opazo JC, Hoffmann FG, Storz JF: **Genomic evidence for independent origins of beta-like globin genes in monotremes and therian mammals.** *Proc Natl Acad Sci USA* 2008, **105**:1590-1595.
16. Koop BF, Goodman M: **Evolutionary and developmental aspects of two hemoglobin beta-chain genes (epsilon M and beta M) of opossum.** *Proc Natl Acad Sci USA* 1988, **85**:3893-3897.