Cloning and expression of activation induced cytidine deaminase from *Bos taurus*

Authors **Subhash Verma[1,2], Tom Goldammer[3] and Robert Aitken[1*]**

Affiliations

1 Division of Infection and Immunity
Institute of Biomedical and Life Sciences
University of Glasgow, UK

2 Present address:
CSKHPKV Himachal Pradesh Agricultural University
Palampur, India

3 Research Institute for the Biology of Farm Animals (FBN)
Dummerstorf, Germany

* Corresponding author

Correspondence

Dr R. Aitken
Division of Infection and Immunity
Institute of Biomedical and Life Sciences
Glasgow Biomedical Research Centre
University of Glasgow
120 University Place
Glasgow G12 8TA
T +44 141 330 6659
F +44 141 330 4600
E r.aitken@bio.gla.ac.uk

Abbreviations

AICDA	Activation induced cytidine deaminase (gene)
AID	Activation induced cytidine deaminase (protein)
dNTP	an equimolar mixture of all four nucleoside triphosphates
DTT	dithiothreitol

Abstract

Activation induced cytidine deaminase is an enzyme crucial to somatic hypermutation and gene conversion, processes that are essential for the diversification of Ig V genes. The bovine Ig repertoire appears to be diversified by mechanisms that are significantly different to those that operate in humans, and mice. This study set out to test the hypothesis that differences in the organisation, coding sequence, expression or genomic location of the bovine AICDA gene enables the encoded enzyme to catalyse the unusual Ig diversification mechanism seen in cattle as well as conventional antigen-driven mutation. Characterisation of bovine AICDA excluded the first two possibilities. AICDA expression was detected in lymphoid tissues from neonatal and older cattle, but AICDA cDNA could not be detected in muscle tissue. The pattern of gene expression did not therefore differ from that in other vertebrates. The AICDA cDNA was cloned and expressed successfully in *Escherichia coli* generating a phenotype consistent with the mutating action of this deaminase. Using a whole genome radiation hybrid panel, bovine AICDA was mapped to a region of bovine chromosome 5 syntenic with the location of human AICDA on chromosome 12. We conclude that the unusual nature of Ig diversification in cattle is unlikely to be attributable to the structure, sequence, activity or genomic location of bovine AICDA.

Keywords

AID

1. Introduction

B lymphocytes that have undergone rearrangement at the Ig loci are able to carry out three further types of DNA modification reaction, depending on the host species and the environment in which the cell exists: class switch recombination (CSR) (Min and Selsing, 2005), somatic hypermutation (SHM) (Di Noia and Neuberger, 2007) and, notably in chickens, gene conversion (GC) (Ratcliffe, 2006).

While different parts of the Ig loci may be altered in these reactions, all three modification processes share a fundamental dependence upon the enzyme activation induced cytidine deaminase (AID). The gene encoding AID, AICDA, was first identified by subtractive cDNA hybridization studies of a murine lymphoma line under conditions designed to induce CSR (Muramatsu et al., 1999). The cDNA was recognised as coding for a member of the cytidine deaminase family which includes apolipoprotein B mRNA editing catalytic polypeptide 1 (APOBEC-1), APOBEC-2, and phorbolin. APOBEC-1 is a well-characterized RNA editing enzyme that recognizes the structure of the mRNA substrate through a cofactor, APOBEC-1 complementation factor. This binds to the mRNA and allows the deaminase activity of APOBEC-1 to act upon its target cytidine (Mehta et al., 2000). AID possesses a putative catalytic domain that is near-identical to that of APOBEC-1.

The importance of AID to the antibody response has been shown in a great many studies. For example, CSR is abolished in transgenic mice that carry mutations in AICDA and there is severe reduction of SHM (Muramatsu et al., 2000). All

human patients with hyper IgM syndrome type 2 carry mutations in AICDA (Revy et al., 2000), and all these mutated AICDA cDNAs are defective in CSR, as determined by *in vitro* assays (Ta et al., 2003). In cells that do not naturally express AICDA such as fibroblasts (Okazaki et al., 2002) and hybridomas (Martin and Scharff, 2002), activation of the enzyme was shown to induce isotype switching or mutation events on artificial constructs or the endogenous rearranged Ig loci. In the chicken B cell line DT40, AICDA disruption completely blocked Ig GC (Arakawa et al., 2002) and constitutive activation of the gene led to uncontrolled GC (Harris et al., 2002).

In cattle, rearrangement *per se* generates little Ig diversity owing to the limited number of Ig segments that can be recruited to the process and their high degrees of sequence similarity (Berens et al., 1997; Sinclair et al., 1997). Cattle are therefore reliant upon diversification processes that take place post-rearrangement. The mechanism and tissue location in which these processes occur remain elusive. It has been suggested that Ig diversification in cattle might parallel events in sheep, a species in which SHM in the ileal Peyer's patches is thought to diversify the Ig repertoire early in life (Jenne et al., 2006). SHM in this context must differ in its regulation or basic mechanism from antigen-driven events in the germinal centres of lymph nodes. Our study set out to test the hypothesis that the organisation, coding sequence, expression or genomic location of the AICDA gene in cattle differs with respect to other animals and that these differences contribute to diversification and antigen-driven SHM of Ig V genes in cattle.

2. Materials and Methods

2.1 Tissue samples

Samples of liver, spleen, mesenteric lymph node, thymus, and muscle were collected *post mortem* from a Holstein Friesian cross calf one day after birth and from three calves that were 13 weeks old. Animals were euthanized by barbiturate overdose by a qualified veterinary surgeon working within current UK legislation for animal welfare. At the time of collection, tissue samples were stabilised in RNAlater (Ambion) and stored at -80ºC. Samples were used for extraction of genomic DNA (DNeasy tissue kit, Qiagen) and RNA (RNeasy midi kit, Qiagen).

2.2 PCR amplification and partial cloning of AICDA gene

To recover of sections of the bovine AICDA gene, PCR primers were designed from the high degree of conservation of human, mouse, chicken, dog and zebrafish AICDA cDNA sequences. Sequences of AICDA cDNA from these animals were aligned and primers were designed using DNAStar Lasergene software. The sequences of oligonucleotides used in these experiments are given in Table 1. As described below, the majority of the AICDA gene was recovered in a reaction with primers AID F6 and AID R5 (Figure 1). A separate reaction (AID F1 and AID R4) recovered an overlapping section of the gene extending to the 3' region (Figure 1).

A touchdown PCR method was used to amplify sequences predicted to form the central region of the bovine AICDA gene (Figure 1). Bovine genomic DNA was used as template with primes AID F6 and AID R5 primers (Figure 1; Table 1).

After initial denaturation at 95°C for 5 min, product was amplified over 35 cycles. The initial five cycles used conditions of 95°C for 1 min, 64ºC for 1 min, and 72°C for 3 min. The annealing temperature was then reduced by 2ºC for each successive set of 5 cycles, the final 10 cycles of the reaction using an annealing temperature of 54ºC. Incubation times remained fixed throughout.

A conventional PCR protocol was used with primers AID F1 and AID R4 to recover the downstream region of the AICDA gene (Figure 1; Table 1). This reaction used 30 cycles of denaturation at 95°C for 40 sec, annealing at 58°C for 40 sec and extension at 72°C for 90 sec.

This strategy enabled recovery of the majority of the bovine AICDA gene but comparison with other species suggested that an intron of significant size separated exons 1 and 2. In addition, the first exon was expected to be very short preventing the design of a primer of sufficient length for recovery of the upstream section of the gene from genomic DNA.

2.3 Recovery of full-length AICDA cDNA

Total RNA was isolated from samples of bovine spleen using a commercial kit (RNeasy, Qiagen). Contaminating DNA was removed from RNA preparations by treating with DNase (TURBO DNase enzyme, Ambion) and the quality and integrity of the isolated RNA was established by spectrophotometric comparison of the yields of 18S and 28S RNA species using an Agilent 2100 Bioanalyser (Agilent Technologies). Final concentrations were measured by small-scale spectrophotometry with a NanoDrop-1000 instrument (NanoDrop Technologies).

First strand cDNA was synthesized by standard reverse transcription methods. A sample of 1µg of total RNA was added to a small 0.2 ml thin-walled PCR tube and the volume was made up to 11 µl with RNase-free water. To this, 1 µl of a 100 µM stock of an oligo(dT) primer (sequence: 5' GGCCAGTGAATTGTAATACGACTCACTATAGGGAGA-$(dT)_{24}$ 3') was added and the contents of the tube were mixed and incubated at 70ºC for 10 min. The tube was centrifuged briefly and the following reagents were then added in succession: 4 µl of 5x first strand buffer (Invitrogen), 2 µl of 0.1 M DTT, 1 µl of a 10 mM dNTP solution. The contents were mixed, briefly centrifuged and incubated at 42ºC, pausing after 2 min to add 1µl of Superscript II reverse transcriptase (200U/µl; Invitrogen). The incubation was continued for 1 hr, and then stopped by transfer to ice.

A GenBank entry from the bovine genome project (accession no. NW_001001418) allowed the design of primers against non-coding sequences upstream of exon 1 (AID exon 1F) and downstream of exon 5 (AID exon 5R). Primer sequences are shown in Table 1. Splenic cDNA was amplified with this primer pair using a conventional PCR protocol (initial denaturation at 94ºC for 2 min, and 30 cycles comprising denaturation at 94ºC for 1 min, annealing at 60ºC for 1 min and extension at 72ºC for 1min). The products were reamplified with a nested primer set, AICDA cDNA F and AICDA cDNA R (Table 1). By comparing sequence on the GenBank file NW_001001418 with the human AICDA gene, these primers were predicted to anneal at the termini of the AICDA reading frame. The nested reaction was carried out with conventional PCR cycling conditions (initial denaturation at 94ºC for 2 min, and 30 cycles of comprising denaturation at 94ºC for 40 sec, annealing at 56ºC for 40 sec and extension at 72ºC for 1min).

2.4 Sequencing

PCR products from genomic and cDNA amplifications were ligated into pCR-TOPO 2.1 and pCR-TOPO XL vectors (Invitrogen), transformed into *Escherichia coli*, and after identification of recombinant clones, samples of plasmid DNA from independent clones were sequenced at the Sir Henry Wellcome Functional Genomics Facility, University of Glasgow. The sequences were aligned, assembled and deposited at GenBank (accession no. DQ303466). The *Bos taurus* AICDA cDNA sequence was translated *in silico* and compared with the protein sequences from other species. Sequence editing, alignments and phylogenetic analysis were performed using DNAStar Lasergene software.

2.5 Expression of AICDA mRNA in bovine tissues

Nested amplifications were carried out with AID F6 and AID R5 as an outer primer set and AID F2 with AID R2 as internal pair (Table 1). cDNA samples prepared from spleen, lymph node, liver, muscle and thymus were used as templates. As an internal control for the quality of each cDNA sample, GAPDH (GenBank accession NM_001034034) was amplified using primers from Endogen.

2.6 Expression of AICDA in Escherichia coli

The contribution of AID to SHM is founded upon the mutagenic activity of the deaminase. In *E. coli,* nucleotide substitutions in *rpoB* can generate rifampicin resistance but the rate at which these mutations arise spontaneously is low (Miller, 1972). Petersen-Mahrt and colleagues showed that expression of human AICDA generated enhanced mutation frequency in *E.coli* leading to rifampicin

resistance (Petersen-Mahrt et al., 2002). The experiments described here tested if the expression of bovine AICDA would have similar effects.

In preparation for cloning into the pBAD/Myc-His vector (Invitrogen), cDNA was amplified with AID cDNA F (*Nco* I) and AID cDNA R (*Spe* I) (Table 1) to introduce *Nco* I and *Spe* I sites at the 5' and 3' termini of the amplicon. After initial denaturation at 94ºC for 5 min, reactions were cycled 30 times through 94ºC (1 min), 58ºC (30 sec) and 72ºC (30 sec) with a final incubation at 72ºC for 5 min. The PCR product was ligated into pCR-TOPO 2.1 and then subcloned into *Nco* I / *Xba* I cut pBAD/*Myc*-His-B using a Quick Ligation Kit (New England Biolabs). This placed expression of AICDA under the control of the *araBAD* promoter.

The recombinant plasmid was transformed into *E. coli* and expression was induced with 0.2% L-arabinose. The presence of recombinant protein was confirmed by SDS-PAGE and Western blotting using a mouse antibody to c-myc, a vector-encoded tag fused to the carboxy terminus of the bovine protein. Evidence for an elevation in mutation rate was sought by transforming *E. coli* with the pBAD vector or the recombinant construct carrying bovine AICDA. After selection on LB agar containing ampicillin, individual clones were cultured in LB medium and induced with 0.2% L-arabinose for 4 hours or overnight. Diluted aliquots from the cultures were then plated onto LB agar plates to determine the total number of bacteria, and to LB agar containing rifampicin (100 µg/ml). The number of rifampicin-resistant cells as a proportion of the total was then calculated. Data was compared between *E. coli* cultures carrying the pBAD vector and those with the AICDA expression construct and statistical significance assessed with the t test.

2.7 Physical mapping of bovine AICDA *using a radiation-hybrid (RH) panel*

The chromosomal location of AICDA was sought by analysis of a cattle-hamster somatic hybrid panel, and a whole genome radiation cell hybrid panel (Womack, 1997). The latter resource ($WGRH_{5000}$) contains fragments of the bovine genome, generated by exposure of bovine cells to an x-ray dose of 5000 rad and rescued into stable lines by fusion with hamster cells. Lines carrying the AICDA gene were identified by PCR using primers AID (RH) F and AID (RH) R. These spanned positions 8-30 and 111-130 of the AICDA cDNA sequence as derived in this study and were located on the gene's second exon. DNA from cells in the radiation hybrid panel was denatured at 95ºC for 15 min then amplified through 35 cycles of 94ºC for 30 sec, 58ºC for 30 sec and 72ºC for 45 sec. A final extension step was performed at 72ºC for 10 min. Once product had been confirmed from DNA samples from the cell hybrid panel, statistical analysis was carried out with the data (Chevalet and Corpet, 1986) to establish linkage with markers known to lie on particular chromosomes. Two-point linkage analysis for AICDA was done with markers on the cattle $WGRH_{5000}$ gene map (Band et al., 2000) using the software RHMAPPER 1.22 (Slonim et al., 1997). This enabled an estimate to be made of the proximity of AICDA to other markers on its assigned chromosome.

3. Results

3.1 Partial recovery of the Bos taurus *AICDA gene*

A PCR strategy was designed to recover the majority of the bovine AICDA gene. This approach was attempted because of the high degree of conservation of human, mouse, chicken, dog and zebrafish AICDA cDNA sequences (Zhao et al., 2005). AID F6 was designed to anneal to bases 6353 to 6373 in exon 2 of the human gene (GenBank accession no. AB040430) and AID R5 was targeted to bases 8468 to 8488 in exon 4. Thus, based upon their binding to the human gene, AID F6 and R5 (Figure 1) were predicted to amplify a PCR product of around 2.1 kb from bovine genomic DNA. In fact, PCR using genomic DNA from bovine liver as a template yielded a specific product of 3.4 kb (Figure 2). A second PCR reaction was performed to recover the downstream regions of the gene using primers AID F1 and AID R4 (Figure 1). AID F1 was predicted to anneal within exon 3 (bases 7814 to 7831 of the human gene) and AID R4 was designed from sequences in exon 5 (bases 8971 to 8994). A PCR product of 1.5 kb was obtained as opposed to the 1.2 kb predicted from the human AICDA sequence. The amplicons were cloned into TOPO vectors for sequencing. The F1 / R4 product could be completely sequenced using primers that annealed to vector sites flanking the insert. For the F6 / R5 product, a primer walking method was used to recover the complete sequence of the insert. Data were assembled into a contig. Since part of the gene from within exon 3 to part of exon 4 was present in both PCR products (Figure 1), a total of 4253 bases from the bovine AICDA gene were sequenced by this approach. A partial structure for the bovine gene could be derived by comparing the data with the sequence of the human AICDA gene (Figure 3). In summary, this revealed that the bovine gene shared with

multiple animal species a high degree of conservation of structure and protein coding sequences. Although intron sequences were divergent and two of the three introns were substantially longer in the bovine gene than in its human orthologue (Figure 3), the data refute the hypothesis that the bovine AICDA gene is fundamentally different to orthologues from other species.

Given the very small probable size of bovine exon 1 (8 nucleotides in human AICDA), the length of the intron separating exons 1 and 2 (5753 nucleotides in humans), and the divergence of intron sequences between bovine and human / mouse genes, it seemed unlikely that the same strategy would complete the characterisation of the bovine AICDA gene. Given the high similarity in gene structure and exon sequences of bovine AICDA and genes from human, chicken and mouse (Arakawa et al., 2002; Muto et al., 2000), it seemed likely that the lengths of bovine AICDA exons 1 to 5 were 8, 148, 274, 116 and 54 bp respectively (Figure 3).

3.2 Recovery of full length of AICDA cDNA

To further assess if small but potentially significant differences existed between the AID proteins of cattle and other animals, the full length bovine cDNA was recovered and sequenced. To do this, data from a BAC clone thought to carry the full AICDA sequence (GenBank accession no. NW_001001418) were analysed. Primers AID exon 1F and AID exon 5R were synthesised against sequences predicted to comprise untranslated regions upstream of exon 1 and downstream of exon 5, respectively. Nested primers (AID cDNA F and AID cDNA R) were based on homology of AICDA cDNAs from a range of species. PCR with these primers using splenic cDNA yielded a 600 bp product which after sequencing and

annotation was deposited in GenBank (accession no. NM_001038682). Analysis of the bovine AICDA cDNA sequence and comparison with genomic data revealed that the 600 nucleotides were derived from five exons as predicted (Figure 3), with exon 3 carrying one codon more than the human sequence. The predicted protein sequence comprised 199 amino acids and was 94.4%, 93.9%, 90.9% and 86.9% identical with dog, human, mouse and chicken sequences, respectively (Figure 4). However, several differences of potential significance were noted. The bovine AID was predicted to carry a pair of lysine residues (residues 6 and 7) positioned closer to the amino terminus of the protein than equivalent residues (arginine / lysine or lysine / lysine) in the other mammalian enzymes (Figure 4). Bovine AID possessed a proline residue at position 39 that might disrupt the secondary structure of the protein in this region; alanine was present at this position in the other vertebrate AIDs represented in Figure 4, although pufferfish AID is known to carry proline in this location (Ichikawa et al., 2006). Finally, the bovine AID contained an additional amino acid at position 118 (lysine; Figure 4). Other parts of the protein sequence were well-conserved, notably the cytidine deaminase motif and the carboxy-terminal leucine rich motif (Figure 4). Residues known to be critical for somatic hypermutation (Revy et al., 2000) were identical across the comparison (Figure 4) suggesting that the activity of the bovine AID protein was unlikely to differ substantially from that of enzymes found in the dog, mouse, human or chicken.

3.3 Expression of AICDA mRNA in bovine tissues

To assess for age dependence and tissue-specific expression of AICDA, cDNA was prepared from mesenteric lymph node, spleen, liver and muscle from day old and 13 week old cattle. This was used as template to assess whether AICDA

transcripts were present in the tissue samples. In those reactions that were successful, a product of the predicted 131 bp emerged. The analysis (Figure 5) revealed the presence of AICDA transcripts in lymph node, spleen and thymus samples from a neonatal calf and animals 13 weeks of age. There was no evidence of AICDA mRNA in muscle tissue (Figure 5). Although the reaction was not conducted on a quantitative basis, yields of the PCR product were substantially lower from liver than from lymphoid cDNA suggesting transcript abundance was lower. In contrast, the control reaction with primers against GAPDH was consistently successful and there was no visible difference in the amount of product (Figure 5).

3.4 Expression of AICDA in Escherichia coli

To assess if the predicted AID protein was stable and possessed the mutagenic properties that would be expected, the cDNA sequence was cloned into the *E. coli* expression vector pBAD/Myc-His. Expression of a novel protein of 26 kDa was detected by SDS-PAGE and Western blotting following induction with L-arabinose at 0.2% (data not shown). This was consistent with the size predicted for AID (24 kDa) taking into account the vector-encoded detection and purification tags.

E. coli TOP10 was transformed with pBAD/Myc-His and recombinant plasmid carrying AICDA cDNA. After induction with 0.2% L-arabinose for 4 hours and overnight, conversion to rifampicin resistance was detected with a frequency that was 2-6 times higher than that of controls (*ie E. coli* transformed with the pBAD vector lacking an insert). The data are shown in Table 2. Variation in bacterial count meant that mean frequencies of rifampicin resistance did not

achieve statistically significant differences between test and control cultures ($P < 0.1, > 0.05$) but the data suggest that bovine AID had some mutagenic effect in the bacterial host.

3.5 Chromosomal localisation of bovine AICDA

A cattle-hamster somatic hybrid panel consisting of 31 cell line was screened by PCR with AICDA specific primers. A statistically relevant concordance value of 0.97 was obtained with the reference markers ACO2 and BM2830 (mitochondrial aconitase and a non-coding sequence tagged site). Both of these markers are known to be located on the distal half of bovine chromosome 5 (BTA5). This allowed the syntenic assignment of AICDA to BTA 5. To refine the location of AICDA, PCR typing was used with the same primer sets using DNA from a cattle-hamster $WGRH_{5000}$ panel of 90 cell lines. The data from this analysis placed AICDA within 11.00 centriRay (cR_{5000}) of the marker MAGP2 with a LOD score greater than 10. One centriRay represents a separation that confers 1% probability that linkage between markers will break after exposure to 5000 rad of gamma radiation. MAGP2 is the gene encoding microfibril-associated glycoprotein 2. The predicted location of AICDA placed it close to M6PR, the gene for the mannose-6-phosphate receptor (Figure 6) although the proximity was not specifically estimated in these experiments. This assignment shows synteny with the location of human AICDA on chromosome 12.

4. Discussion

This study characterized activation induced cytidine deaminase from *B. taurus* testing the hypothesis that differences in the organisation, coding sequence, expression or genomic location of the bovine AICDA gene might account for the unusual Ig diversification mechanism seen in cattle as well as conventional antigen-driven mutation.

The cloning and sequencing of most of the gene was achieved through the design of homology primers and amplification from bovine genomic DNA. The organisation of the gene was similar to that in mice and humans and the degree of sequence identity in the exons implied a high level of stability during mammalian evolution and beyond. Although the chicken AICDA orthologue has a different gene structure (Figure 3), analysis of the cDNA has indicated that the coding sequence is remarkably similar to mammalian AICDA (Papavasiliou and Schatz, 2002). AICDA genes from *Xenopus* gene (Zhao et al., 2005), fugu, catfish and zebrafish (Ichikawa et al., 2006; Zhao et al., 2005) show more widespread change but despite this, the zebrafish enzyme retained the ability to catalyse CSR when transfected into mouse B cells (Barreto et al., 2005).

Recovery and analysis of the bovine AICDA cDNA revealed an open reading frame encoding 199 amino acid residues and a predicted protein sequence that was highly conserved and especially prominent within the cytidine deaminase motif which has been shown to be important for the activity of AID (Papavasiliou and Schatz, 2002). There were two differences of note. Many AID proteins including those distant from the mammalian enzymes (Ichikawa et al., 2006)

possess a cluster of basic residues at the amino terminus of the protein in a region predicted to form an alpha helix (Prochnow et al., 2007; Xie et al., 2004). This feature appeared to be shifted nearer to the amino terminus of the bovine enzyme. Secondly, the bovine AICDA gene carried an additional codon in exon 3 coding for lysine. This was located in a part of the protein sequence that is widely conserved across a wide range of vertebrates but AID from fugu, zebrafish and catfish also possess an additional amino acid at this position (Ichikawa et al., 2006; Zhao et al., 2005). There is no evidence that the additional amino acid present in this region in some species leads to a fundamental alteration in AID activity but the structural impact is presently unknown.

The contribution of different regions of the AID protein to its function has been investigated *via* mutational analysis. For example, it has been shown that deletion and mutation in the carboxy-terminal domain of AID uncoupled CSR from SHM and GC (Barreto et al., 2003; Ta et al., 2003). Likewise, studies of the amino-terminus of AID have been identified mutations that impact upon SHM activity but not CSR (Shinkura et al., 2004). Both regions of the bovine AID sequence are conserved when compared to other mammals. This refutes the hypothesis that the unusual nature of Ig diversification in cattle can be attributed to the properties of AID.

The expression of AICDA mRNA in a range of bovine tissues showed similarities to that reported for dogs and humans. In these species, AICDA mRNA has been shown to be present in both lymphoid and, more surprisingly, in some non-lymphoid tissues. For example, AICDA expression has been detected in canine

thymus, lungs, liver, small intestine and kidneys as well as in lymphoid tissues (Ohmori et al., 2004). In humans, strong expression was detected in lymph nodes and tonsils and at a lower level in thymus, kidney, pancreas and other tissues examined (Muto et al., 2000). Members of Honjo's laboratory (Muramatsu et al., 1999) have reported that AICDA transcripts were present in mouse bone marrow at a level almost comparable to that seen in spleen. In our study of *Bos taurus*, AICDA transcripts were readily detectable in lymphoid cDNA preparations, were recovered at much lower yields from bovine liver, and could not be detected in muscle tissues. Given that the assay was based upon a nested PCR reaction, it is possible that the low yield of AICDA amplicons from liver cDNA could have arisen from activated B cells in the samples, *en route* between lymphoid organs. Since high-sensitivity, nested PCR was used, another explanation for the apparent presence of AICDA transcripts in the liver and thymus would be that RT-PCR is detecting very low level ectopic expression of AICDA.

Radiation hybrid mapping placed the bovine AICDA gene on BTA5. In humans, AICDA is located on chromosome 12 (HSA12). Regions of synteny have been identified between BTA5 and HSA12 though other regions of the bovine chromosome bear genes located on HSA22 (Liu et al., 2003; Everts-van der Wind et al., 2004). Comparative mapping studies of some regions of the human and bovine genomes have revealed substantial similarity in the presence and order of markers on several chromosomes (Everts-van der Wind, 2005) but for BTA5 and HSA12, multiple internal rearrangements seem to have occurred during evolution. However, our studies do not support the notion that an unusual

genomic location impacts upon AICDA function in cattle, another potential trigger to Ig diversification.

The development of the bovine Ig system is incompletely understood, the most important deficiency being uncertainties of how, where and when Ig diversification takes place. By analogy with other vertebrates, SHM and GC are the most likely mechanisms for creation of Ig diversity (Zhao et al., 2006). Work with sheep (Jenne et al., 2006) and rabbits (Mage et al., 2006) has shown the importance of the gut associated lymphoid tissues - the ileal Peyer's patches (IPP) in sheep and the appendix in rabbits – in Ig diversification. Although there are many similarities between the Ig systems of sheep and cattle, important distinctions have also been noted (Zhao et al., 2006). Thus, it is far from clear whether the IPP serves as the site for Ig diversification in *Bos taurus*. Our studies have been unable to identify properties of bovine AICDA that might explain these differences, but our analysis may facilitate better assessment of events in the IPP of cattle. The characterisation of AICDA from the rabbit (Yang et al., 2005) illustrates the potential of this approach.

Achknowledgements

SV was recipient of a postgraduate scholarship from the Faculty of Biomedical and Life Sciences, University of Glasgow, and an Overseas Research Student award .

The authors express their thanks to Dr M. Dagleish (Moredun Research Institute, Edinburgh) for support in accessing biological materials for these studies.

References

Arakawa, H., Hauschild, J., Buerstedde, J.M., 2002. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. Science 295, 1301-1306.

Band, M.R., Larson, J.H., Rebeiz, M., Green, C.A., Heyen, D.W., Donovan, J., Windish, R., Steining, C., Mahyuddin, P., Womack, J.E., Lewin, H.A., 2000. An ordered comparative map of the cattle and human genomes. Genome Res. 10, 1359-1368.

Barreto, V., Reina-San-Martin, B., Ramiro, A.R., McBride, K.M., Nussenzweig, M.C., 2003. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. Molecular Cell 12, 501-508.

Barreto, V.M., Pan-Hammarstrom, Q., Zhao, Y., Hammarstrom, L., Misulovin, Z., Nussenzweig, M.C., 2005. AID from bony fish catalyzes class switch recombination. J. Exp. Med. 202, 733-738.

Berens, S.J., Wylie, D.E., Lopez, O.J., 1997. Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. Int. Immunol. 9, 189-199.

Chevalet, C., Corpet, F., 1986. Statistical decision rules concerning synteny or independence between markers. Cytogen. Cell Genetics 43, 132-139.

Di Noia, J., Neuberger, MS., 2007. Molecular mechanisms of antibody somatic hypermutation. Ann. Rev. Biochem. 76, 1-22.

Everts-van der Wind, A., Kata, S.R., Band, M.R., Rebeiz, M., Larkin, D.M., Everts, R.E., Green, C.A., Liu, L., Natarajan, S., Goldammer, T., Lee, J.H., McKay, S., Womack, J.E., Lewin, H.A., 2004. A 1463 gene cattle-human comparative map with anchor points defined by human genome sequence coordinates. Genome Res. 14, 1424-1437.

Everts-van der Wind, A., Larkin, D.M., Green, C.A., Elliott, J.S., Olmstead, C.A., Chiu, R., Schein, J.E., Marra, M.A., Womack, J.E., Lewin, H.A., 2005. A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution. Proc. Natl. Acad. Sci. (USA) 102, 18526-18531.

Harris, R.S., Sale, J.E., Petersen-Mahrt, S.K., Neuberger, M.S., 2002. AID is essential for immunoglobulin V gene conversion in a cultured B cell line. Current Biol. 12, 435-438.

Ichikawa, H., Sowden, M.P., Torelli, A.T., Bachl, J., Huang, P., Dance, G.S., Marr, S.H., Robert, J., Wedekind, J.E., Smith, H.C., Bottaro, A., 2006. Structural phylogenetic analysis of activation-induced deaminase function. J. Immunol. 177, 355-361.

Jenne, C.N., Kennedy, L.J., Reynolds, J.D., 2006. Antibody repertoire development in the sheep. Dev. Comp. Immunol. 30, 165-174.

Liu, Z., Hansen, M., Womack, J.E., Antoniou, E., 2003. A comparative map of interstitial bovine chromosome 5 with human chromosomes 12 and 22. Cytogen. Genome Res. 101, 147-154.

Mage, R.G., Lanning, D., Knight, K.L., 2006. B cell and antibody repertoire development in rabbits: the requirement of gut-associated lymphoid tissues. Dev. Comp. Immunol. 30, 137-153.

Martin, A., Scharff, M.D., 2002. Somatic hypermutation of the AID transgene in B and non-B cells. Proc. Natl. Acad. Sci. (USA) 99, 12304.

McBride, K.M., Barreto, V., Ramiro, A.R., Stavropoulos, P., Nussenzweig, M.C., 2004. Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. J. Exp. Med. 199, 1235-1244.

Mehta, A., Kinter, M.T., Sherman, N.E., Driscoll, D.M., 2000. Molecular cloning of Apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of Apolipoprotein B mRNA. Mol. Cell. Biol. 20, 1846-1854.

Miller, J.H., 1972. Experiments in molecular genetics. Cold Spring Harbor Press, Cold Spring Harbor.

Min, I.M., Selsing, E., 2005. Antibody class switch recombination: roles for switch sequences and mismatch repair proteins. Adv. Immunol. 87, 297-328.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., Honjo, T., 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102, 553-563.

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., Honjo, T., 1999. Specific expression of activation-induced cytidine

deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J. Biol. Chem. 274, 18470-18476.

Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K., Honjo, T., 2000. Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene. Genomics 68, 85.

Ohmori, K., Maeda, S., Okayama, T., Masuda, K., Ohno, K., Tsujimoto, H., 2004. Molecular cloning of canine activation-induced cytidine deaminase (AID) cDNA and its expression in normal tissues. J. Vet. Med. Sci. 66, 739-741.

Okazaki, I.M., Kinoshita, K., Muramatsu, M., Yoshikawa, K., Honjo, T., 2002. The AID enzyme induces class switch recombination in fibroblasts. Nature 416, 340.

Papavasiliou, F.N., Schatz, D.G., 2002. The activation-induced deaminase functions in a postcleavage step of the somatic hypermutation process. J. Exp. Med. 195, 1193-1198.

Petersen-Mahrt, S.K., Harris, R.S., Neuberger, M.S., 2002. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. Nature 418, 99.

Prochnow, C., Bransteitter, R., Klein, M.G., Goodman, M.F., Chen, X.S., 2007. The APOBEC-2 crystal structure and functional implications for the deaminase AID. Nature 445, 447-451.

1 Ratcliffe, M., 2006. Antibodies, immunoglobulin genes and the bursa of Fabricius in
2 chicken B cell development. Dev. Comp. Immunol. 30, 101-118.

3 Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N.,
4 Forveille, M., Dufourcq-Labelouse, R., Gennery, A., Tezcan, I., Ersoy, F., Kayserili,
5 H., Ugazio, A.G., Brousse, N., Muramatsu, M., Notarangelo, L.D., Kinoshita, K.,
6 Honjo, T., Fischer, A., Durandy, A., 2000. Activation-induced cytidine deaminase
7 (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome
8 (HIGM2). Cell 102, 565-575.

9 Shinkura, R., Ito, S., Begum, N.A., Nagaoka, H., Muramatsu, M., Kinoshita, K.,
10 Sakakibara, Y., Hijikata, H., Honjo, T., 2004. Separate domains of AID are required
11 for somatic hypermutation and class-switch recombination. Nature Immunol. 5,
12 707-712.

13 Sinclair, M.C., Gilchrist, J., Aitken, R., 1997. Bovine IgG repertoire is dominated by
14 a single diversified VH gene family. J. Immunol. 159, 3883-3889.

15 Slonim, D., Kruglyak, L., Stein, L., Lander, E., 1997. Building human genome maps
16 with radiation hybrids. J. Comp. Biol. 4, 487-504.

17 Ta, V.T., Nagaoka, H., Catalan, N., Durandy, A., Fischer, A., Imai, K., Nonoyama,
18 S., Tashiro, J., Ikegawa, M., Ito, S., Kinoshita, K., Muramatsu, M., Honjo, T., 2003.
19 AID mutant analyses indicate requirement for class-switch-specific cofactors.
20 Nature Immunol. 4, 843.

21 Womack, J.E., 1997. Mapping animal genomes. Adv. Vet. Med. 40, 157-189.

1 Xie, K., Sowden, M.P., Dance, G.S., Torelli, A.T., Smith, H.C., Wedekind, J.E.,
2 2004. The structure of a yeast RNA-editing deaminase provides insight into the fold
3 and function of activation-induced deaminase and APOBEC-1. Proc. Natl. Acad. Sci.
4 (USA) 101, 8114-8119.

5 Yang, G., Obiakor, H., Sinha, R.K., Newman, B.A., Hood, B.L., Conrads, T.P.,
6 Veenstra, T.D, Mage, R.G., 2005. Activation-induced deaminase cloning,
7 localization, and protein extraction from young V_H-mutant rabbit appendix. Proc.
8 Natl. Acad. Sci. (USA) 102, 17083-17088.

9 Zhao, Y., Jackson, S.M., Aitken, R., 2006. The bovine antibody repertoire. Dev.
10 Comp. Immunol. 30, 175-186.

11 Zhao, Y., Pan-Hammarstrom, Q., Zhao, Z., Hammarstrom, L., 2005. Identification
12 of the activation-induced cytidine deaminase gene from zebrafish: an evolutionary
13 analysis. Dev. Comp. Immunol. 29, 61-71.
14

Figure legends

Figure One. Strategy for partial recovery of bovine AICDA by PCR

A predicted structure of the bovine AICDA gene is shown, boxes indicating exons (roman numerals) and lines indicating introns. The model was based upon the known structure of AICDA orthologues from other vertebrates. Primers AID F6 and AID R5 were used for isolation of the central region of the bovine AICDA gene. AID F1 and AID R4 were used to recover 3' sequences. Primer sequences are shown in Table 1.

Figure Two. Amplification of the central region of the bovine AICDA gene with primers AID F6 and R5

PCR was carried as described in the main text and reaction products were analysed on a 1% agarose gel with TAE buffer. Lane 1 – PCR product of 3412 base pairs recovered from bovine genomic DNA using primers AID F6 and R5. Lane 2 - negative control from which genomic DNA was omitted. Lane 3 - 1 kb DNA ladder (Promega).

Figure Three. Experimentally-determined organisation of AICDA genes in a range of vertebrate species

The organisation of AICDA genes in *Bos taurus*, human, mouse and chicken. Exons are represented by boxes, introns are represented by lines. For bovine AICDA, numbers in the shaded boxes indicate exon size in bp, and Roman numerals above each box show the exon number for all species. Introns are represented by lines, and the length of each intron in bp is shown above each line for the bovine and

human genes to allow comparison. Grey toning is used to indicate that exons 2 and 5 of bovine AICDA could not be recovered completely by PCR. Exon 3 of bovine AICDA contained 3 extra bases as compared to the human gene but otherwise exon sequence was highly conserved. For bovine AICDA, exon 1 is unfilled and intron 1 is dotted to indicate that their sizes were not determined experimentally. The structure of the mouse AICDA gene was established from data at GenBank (accession number NT_039353). Genes are approximately centred upon the start of exon 3 in each case.

Figure Four. Alignment of AID protein sequences from cow, dog, mouse, human and chicken

The sequences are compared to bovine AID, dots indicating amino acid identity, a dash to indicate where a gap was introduced into the alignment. Where differences to bovine AID exist, amino acids are marked using the single letter code. The cytidine deaminase motif is shaded. Amino acids that are critical to AICDA mediated somatic hypermutation (Revy et al., 2000) are indicated with an arrowhead beneath the aligned sequences. The double underlined sequence is a putative Nuclear Exporter Sequence (NES) and overlaps with a conserved domain, important for CSR (McBride et al., 2004).

Figure Five. Expression of AICDA mRNA in bovine tissue

PCR reactions were carried out with cDNA from tissues (indicated) from calves of one day of age (panels 1 and 2) and 13 weeks (panels 3 and 4). Primers were used for amplification of AICDA transcripts (upper panels) and GAPDH (lower panels). Reactions were analysed by agarose gel electrophoresis and for simplicity, the relevant area of each area of the gel is presented.

Figure Six. Chromosomal assignment of bovine AICDA

Primers AID (RH) F and AID (RH) R were used with DNA from a whole genome radiation hybrid panel of cattle-hamster hybrids to determine the chromosomal location of bovine AICDA (Band et al., 2000). The vertical line represents bovine chromosome 5 (BTA5). Numbers on the left of the line represent distances along BTA5 in centriRays (cR_{5000}). One cR_{5000} represents a separation that confers 1% probability that linkage between markers will break after exposure to 5000 rad of gamma radiation. Abbreviations to the right of the line show framework markers for the chromosome. Many of the markers are non-coding and serve only as markers for BTA5 and reference points for the location of genes that can be assigned to this chromosome.

Figure 1

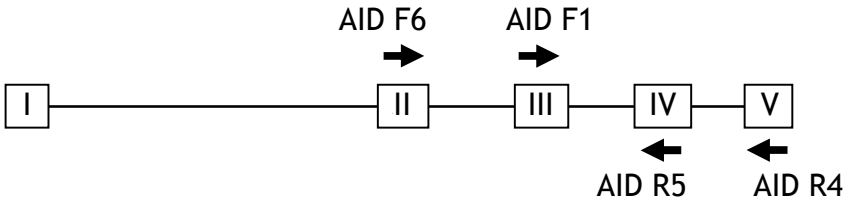


Figure 2

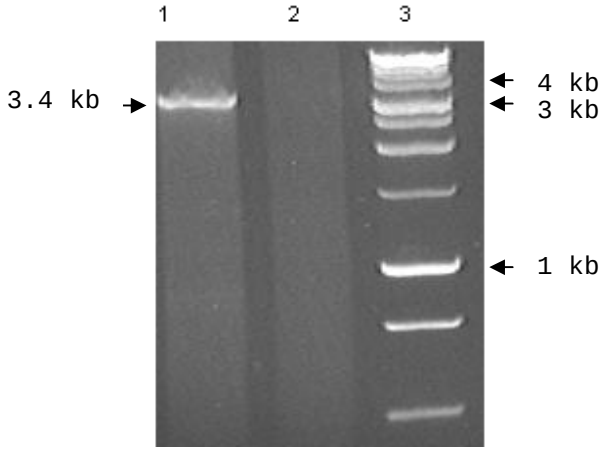


Figure 3

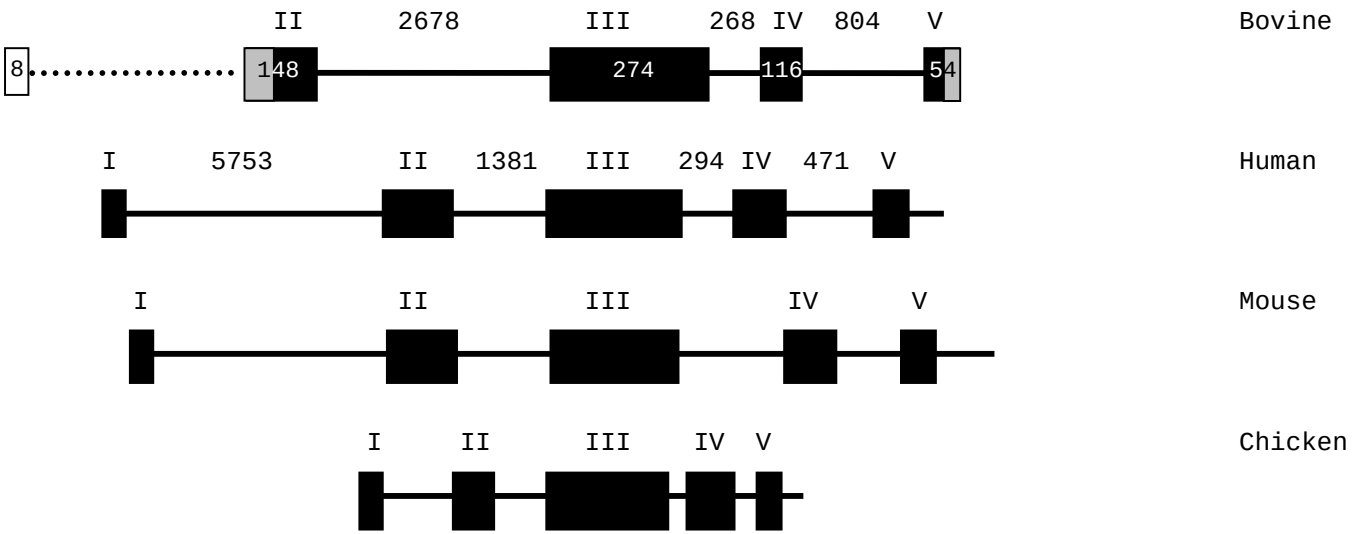

II
2678
III
268
IV
804
V
Bovine
8
148
274
116
54
I
5753
II
1381
III
294
IV
471
V
Human
I
II
III
IV
V
Mouse
I
II
III
IV
V
Chicken

Figure 4

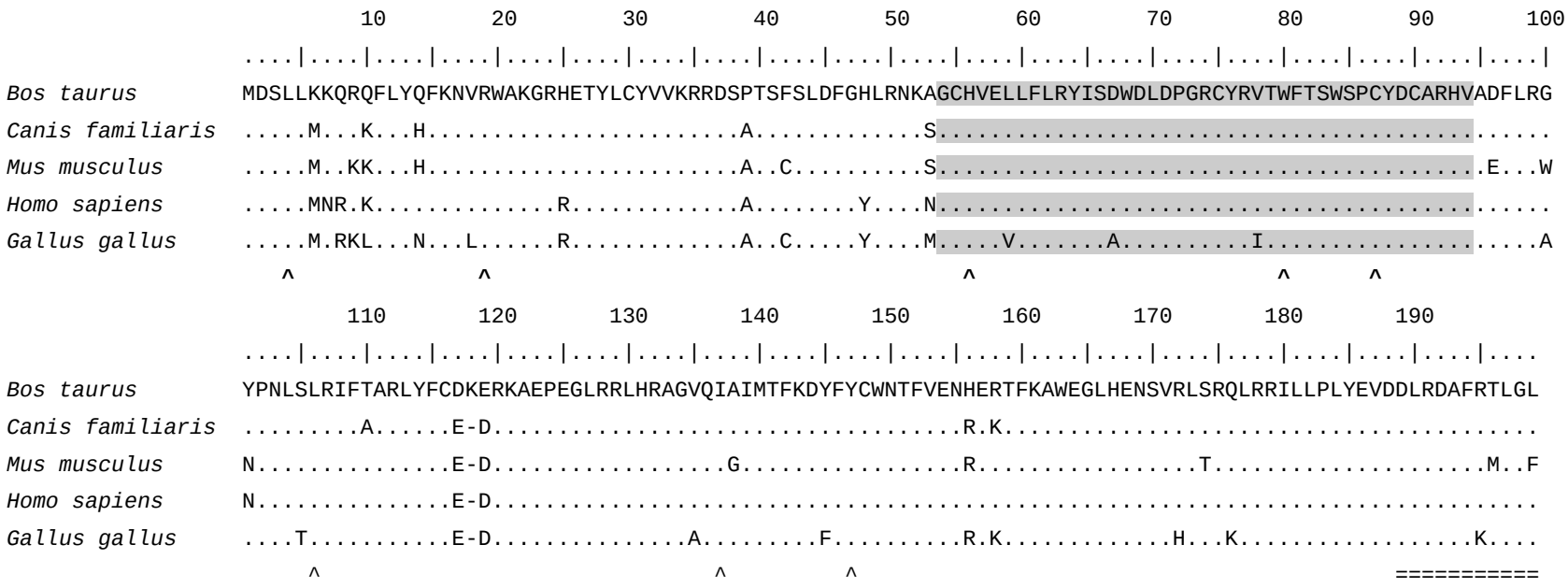

10 20 30 40 50 60 70 80 90 100
Bos taurus MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRG
Canis familiaris .....M...K...H.....A.....S.....
Mus musculus .....M..KK...H.....A..C.....S.....E...W
Homo sapiens .....MNR.K.....R.....A.....Y....N.....
Gallus gallus .....M.RKL...N...L.....R.....A..C.....Y....M.....V.....A.....I.....A
110 120 130 140 150 160 170 180 190
Bos taurus YPNLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL
Canis familiaris .....A.....E-D.....R.K.....
Mus musculus N.....E-D.....G.....R.....T.....M..F
Homo sapiens N.....E-D.....
Gallus gallus ....T.....E-D.....A.....F.....R.K.....H...K.....K....
==========

Figure 5

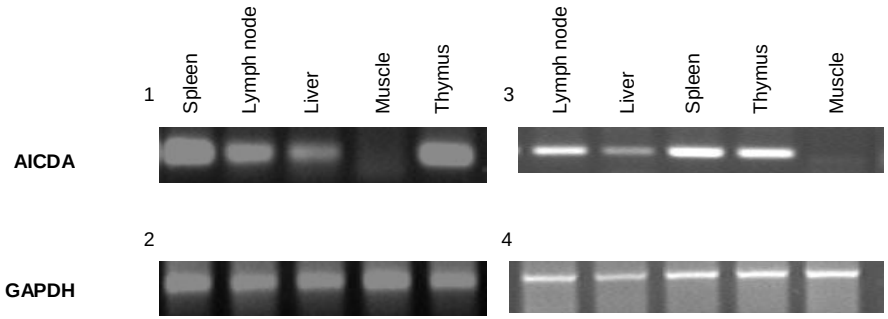
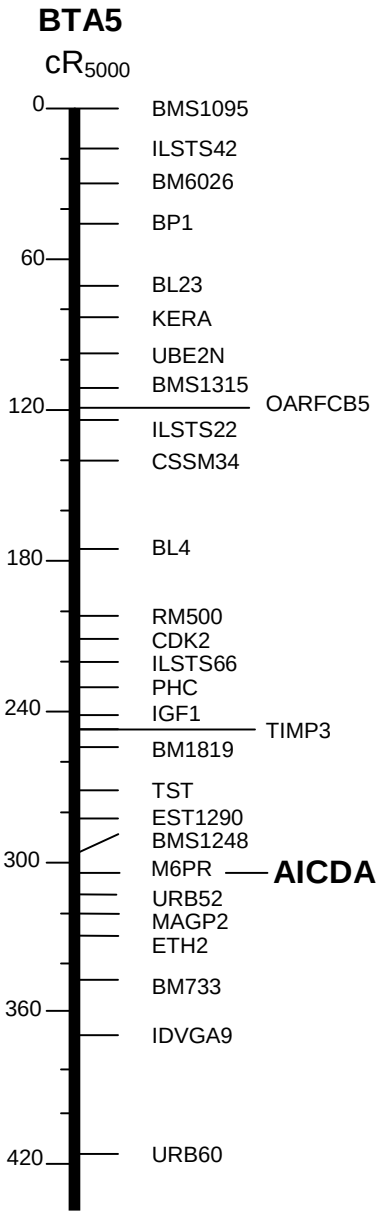

1
Spleen
Lymph node
Liver
Muscle
Thymus
3
Lymph node
Liver
Spleen
Thymus
Muscle
AICDA
2
4
GAPDH

Figure 6


BTA5
cR5000
0
60
120
180
240
300
360
420
BMS1095
ILSTS42
BM6026
BP1
BL23
KERA
UBE2N
BMS1315
OARFCB5
ILSTS22
CSSM34
BL4
RM500
CDK2
ILSTS66
PHC
IGF1
TIMP3
BM1819
TST
EST1290
BMS1248
M6PR
AICDA
URB52
MAGP2
ETH2
BM733
IDVGA9
URB60