Title

Biochemical Regulatory Features of Activation-Induced Cytidine Deaminase Remain Conserved from Lampreys to Humans

Journal

Molecular and Cellular Biology, 37(20)

ISSN

0270-7306

Authors

Quinlan, Emma M
King, Justin J
Amemiya, Chris T
et al.

Publication Date

2017-10-01

DOI

10.1128/mcb.00077-17

Peer reviewed

AMERICAN SOCIETY FOR MICROBIOLOGY | Molecular and Cellular Biology®

RESEARCH ARTICLE

Biochemical Regulatory Features of Activation-Induced Cytidine Deaminase Remain Conserved from Lampreys to Humans

Emma M. Quinlan,[a] Justin J. King,[a] Chris T. Amemiya,[b] Ellen Hsu,[c] Mani Larijani[a]

Department of Biomedical Sciences, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada[a]; Molecular Genetics Program, Benaroya Research Institute, Seattle, Washington, USA[b]; Department of Physiology and Pharmacology, SUNY Downstate Medical Center, New York, New York, USA[c]

ABSTRACT Activation-induced cytidine deaminase (AID) is a genome-mutating enzyme that initiates class switch recombination and somatic hypermutation of antibodies in jawed vertebrates. We previously described the biochemical properties of human AID and found that it is an unusual enzyme in that it exhibits binding affinities for its substrate DNA and catalytic rates several orders of magnitude higher and lower, respectively, than a typical enzyme. Recently, we solved the functional structure of AID and demonstrated that these properties are due to nonspecific DNA binding on its surface, along with a catalytic pocket that predominantly assumes a closed conformation. Here we investigated the biochemical properties of AID from a sea lamprey, nurse shark, tetraodon, and coelacanth: representative species chosen because their lineages diverged at the earliest critical junctures in evolution of adaptive immunity. We found that these earliest-diverged AID orthologs are active cytidine deaminases that exhibit unique substrate specificities and thermosensitivities. Significant amino acid sequence divergence among these AID orthologs is predicted to manifest as notable structural differences. However, despite major differences in sequence specificities, thermosensitivities, and structural features, all orthologs share the unusually high DNA binding affinities and low catalytic rates. This absolute conservation is evidence for biological significance of these unique biochemical properties.

KEYWORDS biochemistry, enzymes, evolution, evolutionary immunology

Received 22 February 2017 **Returned for modification** 28 March 2017 **Accepted** 10 July 2017

Accepted manuscript posted online 17 July 2017

Citation Quinlan EM, King JJ, Amemiya CT, Hsu E, Larijani M. 2017. Biochemical regulatory features of activation-induced cytidine deaminase remain conserved from lampreys to humans. Mol Cell Biol 37:e00077-17. https://doi.org/10.1128/MCB.00077-17.

Address correspondence to Mani Larijani, mlarijani@mun.ca.

Activation-induced cytidine deaminase (AID) is a DNA-mutating enzyme that is a member of the apolipoprotein B mRNA-editing catalytic component (APOBEC) family (1, 2). It is primarily expressed in mature B lymphocytes and initiates the antibody secondary diversification processes of somatic hypermutation (SHM) and class switch recombination (CSR) by mutating deoxycytidine (dC) to deoxyuridine (dU) in the variable and switch sequences of immunoglobulin (Ig) loci, respectively (3, 4). Since their discovery, murine and human AIDs (Hs-AIDs) have been subject to intense research, and AID has been shown to cause DNA damage resulting in tumorigenesis through targeting nonimmunoglobulin genes (5–11). Hs-AID preferentially deaminates dC to dU within WRC (W is A/T; R is A/G) motifs in single-stranded DNA (ssDNA) (12–17). We previously described certain unique biochemical properties of purified Hs-AID: it has a slow catalytic rate (one reaction every ~4 min) and exceptionally high affinity (nM range) for binding ssDNA, such that once bound to its ssDNA target, the bound AID:ssDNA complex lasts, on average, 5 to 8 min before the first dissociation event (18). These characteristics, which were subsequently confirmed by others (19), stand in sharp contrast to most typical human enzymes that have catalytic and substrate on/off rates

~1,000-fold faster (11, 20). We postulated that the enzymatic lethargy and high-affinity ssDNA binding have evolved to protect genomes from promiscuous AID activity by preventing rampant "jumping" around the genome (11, 18, 19, 21, 22). Direct proof of this concept was provided by the finding that AID "upmutants" that exhibited higher catalytic rates *in vitro* also inflicted more damage to the genomes of cells in which they were expressed (23).

Our recent solution of AID's functional and native structure provided insight into several features of the structure that are responsible for these unusual biochemical properties. We found that the surface of AID is home to two ssDNA binding grooves that pass over the catalytic pocket. AID's surface is extremely positively charged (net charge of +14 in neutral pH), explaining its unusually high affinity for binding negatively charged DNA; however, this also results in the vast majority of ssDNA binding nonspecifically and sporadically on the surface of AID such that the ssDNA does not pass over the catalytic pocket. Furthermore, we demonstrated that the majority (~75%) of AID conformations at any given time contain catalytic pockets that are closed and unable to carry out a mutation reaction. Taking these results together, we found that only ~1% of AID:ssDNA binding events result in mutation catalysis by AID, thus explaining the low catalytic rate (11, 21).

Aside from human and murine AIDs, AID orthologs in more distantly related vertebrates have been identified and some have been investigated for functionality. Due to evidence of SHM in the early diverging vertebrate fish lineages (24–28), Saunders et al. hypothesized that an AID ortholog could be found in bony fish and thus discovered AID transcripts in the channel catfish (Ip-AID) (29). Shortly thereafter, Zhao et al. determined that zebrafish also has a bona fide AID gene (the Dr-AID gene) and noted that it, along with the predicted AID genes from other ray-finned fish, encodes an additional 9 amino acids (aa) in the cytidine deaminase motif, along with a different N-terminal motif compared to tetrapod AIDs (30). Even though canonical CSR only occurs in tetrapods (31), multiple fish AID orthologs were able to initiate both SHM and CSR in *Escherichia coli*, *Saccharomyces cerevisiae*, and murine cells, albeit less effectively than mammalian AID (32–35), thus suggesting that CSR evolved due to the emergence of switch regions within immunoglobulin loci and not due to adaptations of the different AID orthologs.

The aforementioned studies examined the function of AID orthologs in cell-based reporter assays, but the biochemical and structural characteristics of these AIDs have not been studied. Since orthologous AIDs have been studied in the context of xenogeneic cell lines with the dampening effects of cellular DNA repair processes, it is difficult to form a direct comparison of their molecular properties. To address this issue, we previously compared the enzymatic properties of purified Dr-AID and Ip-AID with Hs-AID. We found that Dr-AID and Ip-AID were more active at a colder temperature (20 to 25°C) compared to Hs-AID (30 to 37°C). Furthermore, although both bony fish orthologs bound ssDNA with nanomolar range affinities akin to human AID, Dr-AID exhibited a faster catalytic rate than Hs-AID, while Ip-AID had a significantly slower catalytic rate (36, 37). Characterization of these differences using AID mutants and chimeric enzymes led to significant insights into structure-function relationships within AID, although it remained unclear how well conserved AID's low catalytic rate would be across evolution. Complicating interpretation was the fact that Dr-AID may be quite unique among orthologs in that it appears to play epigenetic roles outside the immune system. Dr-AID is required for neurogenesis in developing zebrafish embryos by turning on broad gene expression cascades through widespread demethylation of methylated promoter CpG motifs (38). We showed that Dr-AID is unique among all tested bony fish AID orthologs in that it is capable of deaminating 5-methylcytosine (mC) in CpG motifs, hence its ability to function in mC demethylation (37).

Here, we sought to study certain biochemical properties of AID orthologs that are the earliest diverged and cover the breadth of AID's evolution from its earliest emergence in jawless vertebrates in order to evaluate the conservation, or lack thereof, of AID's unusually lethargic activity and high binding affinity. To this end, we focused on AIDs from species representative of key divergence points in regard to evolution of

adaptive immune systems. These include the sea lamprey (Pm-CDA1), the nurse shark (Gc-AID), the "living fossil" fish coelacanth (Lc-AID), and a relatively more recently diverged bony fish, tetraodon (Tn-AID). We report that these earliest-diverged AID orthologs exhibit unique substrate specificities and optimal temperature tolerances but that AID's lethargic enzymatic rate and high affinity for ssDNA are remarkably conserved. These results shed light on the molecular evolution of AID in the context of diverging adaptive immune systems and highlight the importance of its unique biochemical properties.

RESULTS

Selection of AID orthologs covering the broadest window of AID's existence. Lamprey and hagfish are cyclostomes: the two remaining lineages of agnathans, the earliest-diverged extant vertebrates (Fig. 1A) (39). The jawless fish lamprey is of major interest for immunologists due to a unique adaptive immune system involving variable lymphocyte receptors (VLRs) rather than immunoglobulin (Ig) (40–50). The lamprey has three types of VLRs (A, B, and C), whose genes are proposed to be somatically rearranged by a type of gene conversion, the mechanisms of which are as yet unknown (40–44, 46–49, 51–54). While attempting to identify how VLRs were assembled when jawless vertebrates lack the recombination activating gene proteins (Rag 1 and 2) (55), Rogozin et al. identified cytidine deaminases 1 and 2 (CDA1/2), constituting the earliest-diverged AID orthologs identified to date (52). Both have a rather poor sequence similarity to other AID sequences, yet CDA1 was found to induce C to T mutations in *E. coli* and recombination in yeast (52). Two other groups identified the expression patterns of each CDA, determining that CDA1 was associated with VLRA and C cells which are T cell like, while CDA2 was expressed in B-cell-like VLRB cells (47, 53, 54). CDA1 exhibited mutagenic activity when expressed in both *E. coli* and yeast cells (52); however, it has not yet been purified in order to be directly studied. Therefore, we chose to purify CDA1 from *Petromyzon marinus* (Pm-CDA1) to biochemically compare to the well-characterized Hs-AID.

Another AID we chose to compare with Hs-AID comes from the cartilaginous fish nurse shark (Chondrichthyes) *Ginglymostoma cirratum* (Gc-AID), descendant of an early gnathostome lineage, representing the earliest-diverged Ig loci that is V(D)J-based (25, 56–60). Shark Ig loci do undergo SHM and recent studies suggest they may also undergo a type of rudimentary class-switching (26, 56, 59). To date, only a putative cytidine deaminase from elephant shark has been identified (61), and the functionality of AID from any shark species has not been evaluated. To this end, we isolated and identified an AID-encoding cDNA from the spleen of nurse shark and expressed and purified this putative enzyme.

Around the era the cartilaginous fish diverged from other early gnathostomes, the bony fish (Osteichthyes) class emerged, eventually giving rise to the ray-finned fish (Actinopterygii), who also evolved unique V(D)J immunoglobulin loci organization and isotypes (39). The AIDs of several earlier-diverged bony fish have been shown to be functional cytidine deaminases (29, 30, 32–34, 36–38, 62, 63). Here we chose a pufferfish belonging to a more recently derived lineage from the ray-finned fish clade, the pufferfish tetraodon (*Tetraodon nigroviridis*) to study its AID (Tn-AID).

Lastly, the closest fish ancestors to tetrapods are the lobe-finned fish (Sarcopterygii), of which coelacanth (*Latimeria chalumnae*) is an extant member whose genome has been recently sequenced, including its AID gene (the Lc-AID gene) (64). The other extant member of Sarcopterygii is the lungfish, whose genome apparently does not contain an AID ortholog, making coelacanth the only member of this clade to have an AID-encoding gene (55). Coelacanth Ig has also been shown to undergo V(D)J recombination (65, 66). Thus, as shown in Fig. 1A, the four fish AIDs chosen for this study are representative of key divergence points in the evolution of adaptive immunity, while covering the broadest evolutionary distance relative to Hs-AID and to each other.

Figure 1B shows the degree of amino acid identity and similarity among the orthologs, while Fig. 1C shows an alignment of the five AID orthologs' amino acid sequences. As expected, the ortholog from the earliest-diverged lineage, Pm-CDA1, is

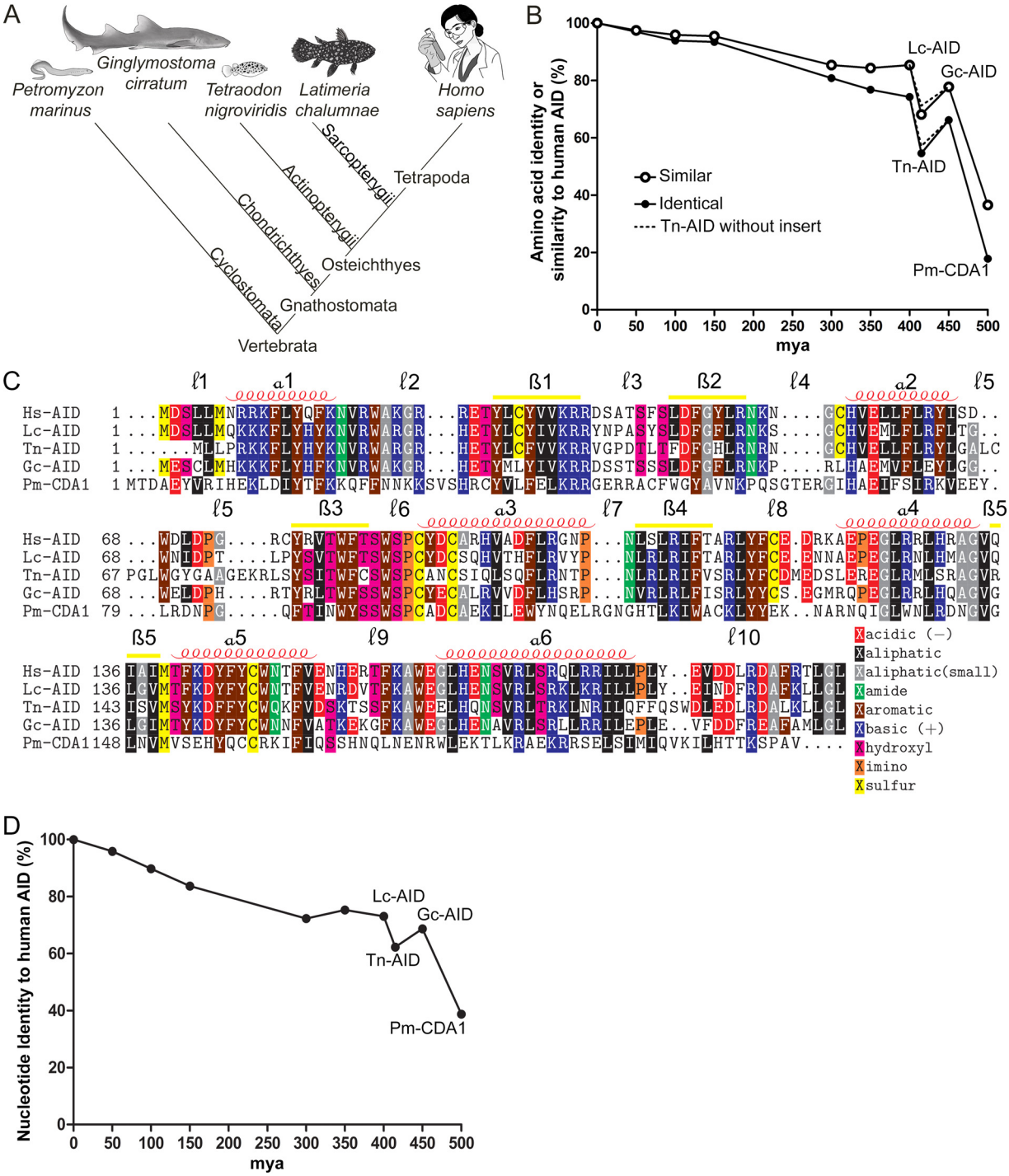

FIG 1 AID in the context of evolution. (A) Phylogenetic tree showing the evolutionary relationships between early-diverged fish species and humans: *Petromyzon marinus* (lamprey), *Ginglymostoma cirratum* (nurse shark), *Tetraodon nigrovirigis* (tetraodon), *Latimeria chalumnae* (coelacanth), and *Homo sapiens* (human). Images of the species were illustrated by Emma Quinlan. (B) Percent sequence identity and similarity were calculated by comparing identical and similar amino acids between the indicated AID orthologs and Hs-AID. To investigate the relative contribution of the ray-finned fish insert to sequence similarity, identical and similar homologies between Tn-AID and Hs-AID were calculated without this insert. The approximate period of species appearance (mya, million years ago) is shown on the *x* axis. Sequence identity and similarity to AID are based on published sequences of AID; some are predicted: 0 mya, *Homo sapiens*; 50 mya, *Callithrix jacchus*; 100 mya, *Pteropus vampyrus*; 150 mya, *Ornithorhynchus anatinus*; 300 mya, *Anolis carolinensis*; 350 mya, *Pleurodeles waltl*; 400 mya, *Latimeria chalumnae* and *Tetraodon nigroviridis*; 450 mya, *Ginglymostoma cirratum*; and 500 mya, *Petromyzon marinus*. (C) Alignment of AID protein sequences from human (Hs-AID), coelacanth (Lc-AID), tetraodon (Tn-AID), nurse shark (Gc-AID), and lamprey (Pm-CDA1). L, loop; α, helix; β, strand. (D) Percent sequence identity was calculated by comparing the nucleotide sequence between the AID-encoding cDNA sequences of fish AID orthologs and Hs-AID. The approximate period of species appearance is shown on the *x* axis. The cDNA sequences of the same nonfish species used to provide context for amino acid conservation were again used: 0 mya, *H. sapiens*; 50 mya, *C. jacchus*; 100 mya, *P. vampyrus*; 150 mya, *O. anatinus*; 300 mya, *A. carolinensis*; 350 mya, *P. waltl*; 400 mya, *L. chalumnae* and *T. nigroviridis*; 450 mya, *G. cirratum*; and 500 mya, *P. marinus*.

the least and quite poorly conserved with Hs-AID (17% identity, 37% similarity). At the other end of the spectrum, and also as expected, the most recently diverged AID ortholog relative to tetrapods, Lc-AID, exhibited the highest level of sequence similarity with Hs-AID (74% identity, 85% similarity). Tn-AID exhibits a modest level of conservation with Hs-AID between that of Pm-CDA1 and Lc-AID (55% identity and 68% similarity). Tn-AID has two inserts spanning positions 65 to 79; as of yet, this insert has only been found in ray-finned fish (Fig. 1C) (36, 37). However, Tn-AID's lower sequence similarity is not due to this insert: without it, Tn-AID's amino acid identity and similarity to Hs-AID only increases by 2 and 3%, respectively (indicated in Fig. 1B by a dotted line). Gc-AID exhibited higher conservation with Hs-AID (66% identity and 78% similarity) compared to Tn-AID. In addition to the amino acid sequences, we compared relative conservation among each AID's cDNA sequence (Fig. 1D). The nucleotide comparison shows a similar pattern of percent sequence identity between the AID orthologs as seen in the amino acid comparison. Lc-AID has the closest nucleotide conservation to Hs-AID (73%), whereas Pm-CDA1 is the least conserved (39%). Tn-AID is again modestly conserved with Hs-AID (62%), while Gc-AID is slightly more homologous (69%). We conclude that sequence similarity to Hs-AID generally corresponds with evolutionary distance; however, relative to the AIDs of land-dwelling tetrapods, all orthologs studied here have the lowest similarities to Hs-AID. This suggests that if the biochemical properties of AID are indeed divergent among species, this ought to be observable in our comparative study.

Fish AID orthologs are cold-adapted cytidine deaminases. In order to compare some of their biochemical properties, we generated expression constructs and expressed and purified AID orthologs from the aforementioned species as glutathione *S*-transferase (GST)-tagged fusion proteins, as previously described (18, 36, 37). First, we verified the quality and determined the concentrations of multiple independently purified preparations of each AID on denaturing SDS-PAGE gels (see Fig. S1A in the supplemental material). Once appropriate purification of each AID preparation was confirmed, we tested whether they are active cytidine deaminase enzymes. For initial activity tests, we utilized the standard alkaline cleavage assay for cytidine deamination which we and others have previously established to measure AID activity on ssDNA (15, 36, 37, 67–70). We incubated each AID at long intervals at three different temperatures (18, 25, and 37°C) with TGCbub7, which is a standard bubble-type substrate that contains the WRC motif TGC (Fig. 2A, top). The longer incubation at three different temperatures accounts for possible differences in absolute activity levels, as well as optimal temperatures among AID orthologs, and would allow for enzymatic activity, however faint, to reveal itself. As shown in Fig. 2A, when a representative preparation of each AID ortholog was incubated with TGCbub7 overnight, Pm-CDA1 and Hs-AID showed relatively similar levels of activity (54.7 and 93.0% deamination, respectively), whereas the other AID orthologs appeared to be less active (1.8, 5.1, and 16.3% for Gc-, Tn-, and Lc-AID, respectively).

To confirm that the observed activity was bona fide AID activity through cytidine deamination, we created catalytically "dead" AID mutants, wherein one of the zinc-coordinating amino acids and the proton-donating glutamic acid in the catalytic pocket, previously shown to be necessary for cytidine deamination (13, 21, 71), were mutated (H56Y and E58G for Hs-AID, Lc-AID, and Gc-AID; H53Y and E55G for Tn-AID; and H66Y and E68G for Pm-CDA1). We purified each mutant, confirming purification on SDS-PAGE (see Fig. S1B in the supplemental material). When a representative preparation of each catalytically dead AID mutant was incubated with TGCbub7 overnight and tested alongside the wild-type AID orthologs (Fig. 2A), no cleavage of the substrate was observed. As a second independent negative control of the purity and specificity of the AID purifications, the construct bearing only the GST tag with no AID insert was expressed, and the GST tag was purified in the same manner as the wild-type GST-tagged AID orthologs. The purity of this preparation was determined through visualization on an SDS-PAGE gel (see Fig. S1C in the supplemental material). When the GST

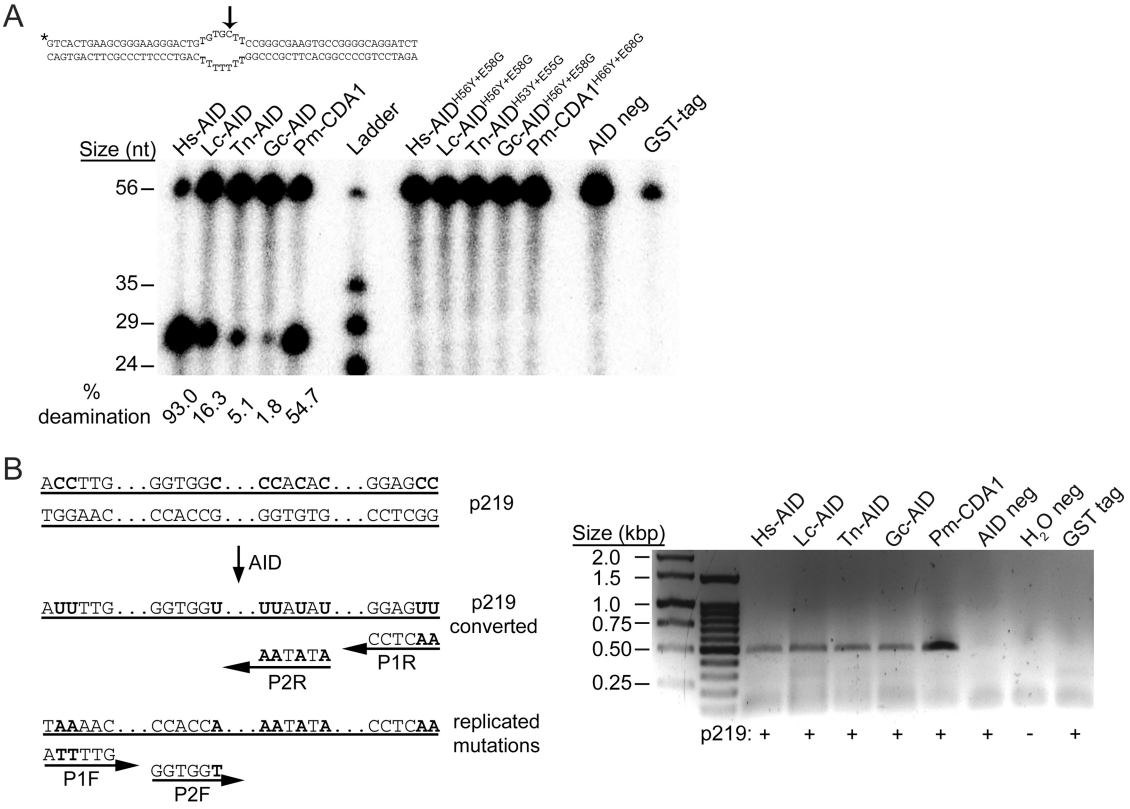

FIG 2 AID orthologs are active cytidine deaminase enzymes. (A, top) TGCbub7 substrate. *, ^{32}P label. The arrow shows the target cytidine in the bubble. (Bottom) Representative alkaline cleavage gel showing activity of 0.5 μg of Pm-CDA1, Gc-, Tn-, Lc-, and Hs-AID, and their corresponding catalytically dead mutants on TGCbub7, incubated at each enzyme's optimal temperature (14.5, 20.5, 20.0, 25, and 31°C, respectively). The substrate and product are 56 and 28 nucleotides in length, respectively. The negative control is the substrate TGCbub7 and activity buffer with no enzyme added. (B, left) Schematic of p219 showing the principle of the PCR assay for detecting AID activity on long stretches (several hundred nucleotides) of plasmid ssDNA. Outside primers: P1F (GGAAGGTATGAAAA TAGGAAAAGAAAATAAATAAATTTTG) and P1R (CCCCTAACTTTTATACCCAACCCTAACTCC). Nested primers: P2F (CCCCCCGATCCGTATT TTTGGATAGTTAGGTGGT) and P2R (CCCCCCGATCCAATTAACAACCCTAAAATATAA). Reverse primers are designed to preferentially anneal with deaminated cytidines; the forward primers anneal to the replicated mutations. (Right) PCR amplification of a section of p219 plasmid untreated (AID neg) or incubated with five AID orthologs or GST at their optimal temperature (25°C for GST) twice for 1 h each. For PCRs to amplify AID-mutated sequences, p219 plasmid incubated with Hs-AID, Pm-CDA1, and GST was annealed to deamination-specific primers at 52°C; Lc-, Tn-, and Gc-AID reactions were annealed at 51°C.

tag preparation was incubated overnight with TGCbub7 and visualized on an acrylamide gel alongside the wild-type and dead AID orthologs, the substrate was not cleaved (Fig. 2A). Thus, any activity seen in our alkaline cleavage assay can be attributed to cytidine deaminase activity of each purified AID ortholog.

To independently confirm that the fish AID proteins were active, we also used a second type of deamination assay. In this PCR-based assay that we previously established (14, 15), a plasmid substrate, rendered single stranded through restriction enzyme digestion and heat denaturation, is incubated with AID, followed by PCR using deamination-specific primers to selectively amplify AID-mutated DNA (Fig. 2B, left panel), which is then sequenced to verify AID-mediated mutations. Using this assay, we confirmed that all five AID orthologs catalyzed dC-to-dU mutations detectable by PCR (Fig. 2B, right panel). When the p219 plasmid was incubated with the GST tag alone, purified in parallel with the AID orthologs, no amplification using the deamination-specific primers was observed (Fig. 2B, right panel); therefore, any cytidine deaminase activity seen in this assay can be directly attributed to specific AID-mediated dC-to-dU conversion.

Previous studies using cell lines transfected with fish AID, as well as *in vitro* enzyme assays with purified fish AID, demonstrated that these orthologs mediate higher mutation rates at colder temperatures than 37°C, suggesting that AID orthologs could

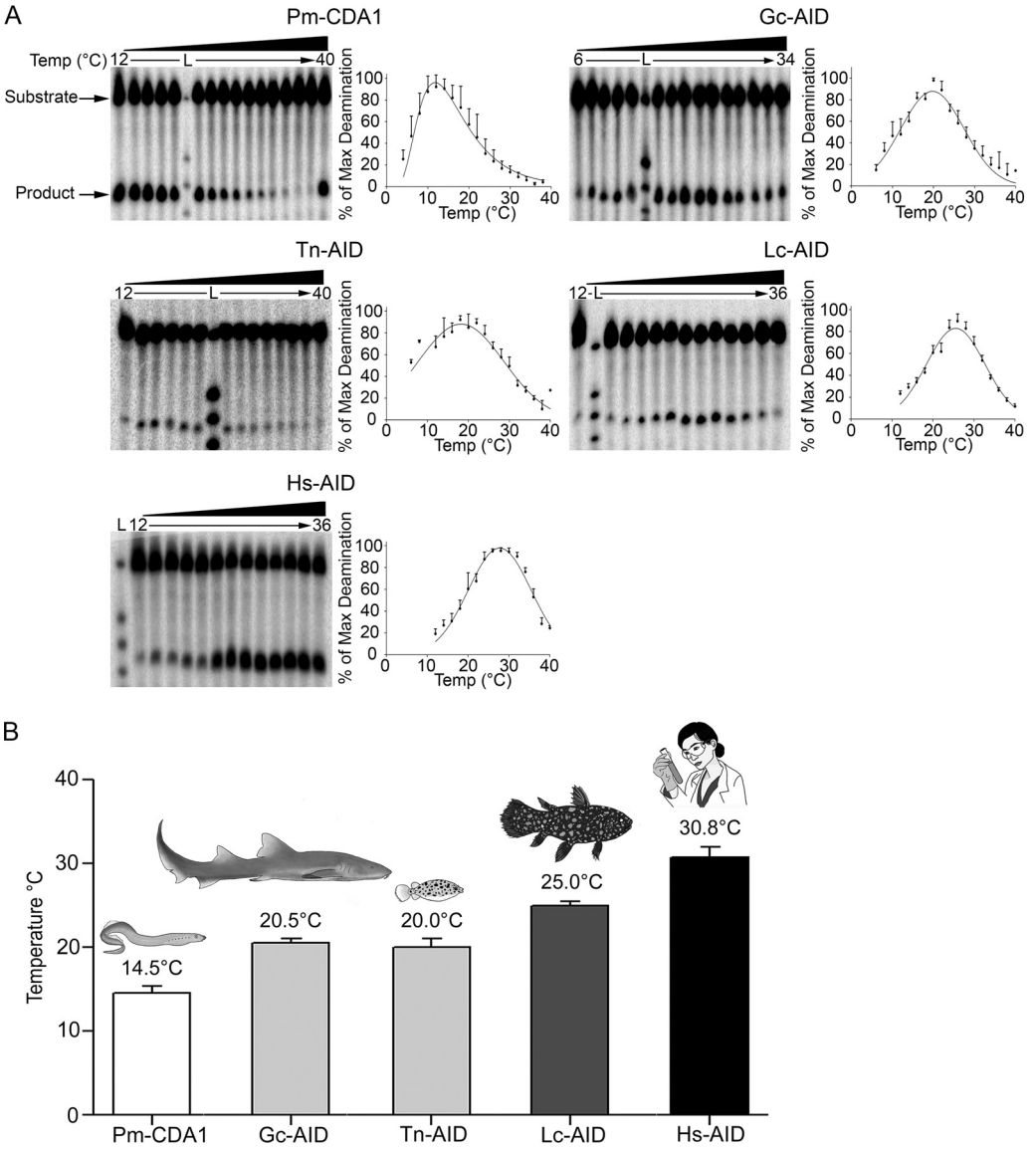

FIG 3 Fish AID orthologs are cold adapted compared to Hs-AID. (A) Representative AID activity alkaline cleavage gels and data from thermosensitivity assays. Gels show AID incubated with TGCbub7 for temperatures ranging from 4 to 40°C. Graphs present thermosensitivity curves, the peak of which indicates the AID ortholog's optimal temperature. The *y* axis shows the percentage of maximum deamination. Error bars represent standard deviations (SDs). (B) Graph of average optimal temperatures for each AID ortholog, indicated by name in the *x* axis, with each organism illustrated and the final average optimal temperature above each bar. Each average is determined from three to six individual experiments using two to four independently purified preparations expressed from two or three independently constructed expression vector clones for each AID ortholog, resulting in, from left to right, 7, 4, 5, 12, and 5 independent experiments. Error bars represent the standard errors of the mean (SEM).

have unique thermosensitivity preferences (2, 36). To evaluate this, we measured the activity of each AID on TGCbub7 at 18 temperature increments ranging from 4 to 40°C (Fig. 3A). In order from earliest to most recently diverged orthologs, Pm-CDA1 is the coldest adapted out of all five species tested, with an average optimal temperature of 14.5°C (Fig. 3B). Gc-AID and Tn-AID are both relatively cold adapted, with average optimal temperatures of 20.5 and 20.0°C, respectively. Lc-AID is the warmest adapted of the fish AIDs, preferring an optimal temperature of 25.0°C. Hs-AID is the warmest adapted of all the AID orthologs tested, with an optimal temperature of 30.8°C (Fig. 3B). Based on these data, we conclude that these fish AID orthologs are active cytidine deaminase enzymes that are colder adapted, one with as much difference in optimal temperature as ~15°C compared to Hs-AID.

AID orthologs have unique substrate sequence specificities. Hs-AID preferentially mutates dC in signature trinucleotide WRC motifs (14, 16, 17, 69, 72–77). To determine the sequence preference of the earlier-diverged AID orthologs, we tested them using the alkaline cleavage assay on six different substrates that are identical to TGCbub7 (Fig. 2A, top) except for the −2 and −1 position nucleotides immediately upstream of the target dC. Figure 4A demonstrates the WRC specificity of Hs-AID. As expected, Hs-AID is ~3-fold more active on the WRC substrates (on TGCbub7, TACbub7, and AGCbub7: 62.0, 47.0, and 42.7% deamination, respectively) than non-WRC substrates (on GACbub7, GTCbub7, and GGCbub7: 10.7, 18.0, and 27.1% deamination, respectively). We previously demonstrated that both Hs- and Dr-AID preferred the three tested WRC motifs at similar rates, whereas Ip-AID was 3-fold more active on TGCbub7 than on all other WRC and non-WRC motifs tested (36). Here, we found that Pm-CDA1 was 4-fold more active on TGCbub7 and TACbub7 than the non-WRC GGCbub7 and GACbub7; however, Pm-CDA1 also showed 2.5-fold activity on the non-WRC GTCbub7 compared to the WRC AGCbub7 (Fig. 4B). Both Gc-AID and Tn-AID exhibited a statistically significant preference for TGCbub7 over all other motifs tested (3- and 8-fold, respectively). Lc-AID also exhibited a slightly skewed WRC preference: it showed 5-fold more activity on the WRC motifs tested compared to GTCbub7 and GACbub7; however, it was also 5-fold more active on the non-WRC GGCbub7 compared to the other non-WRC substrates (Fig. 4B).

The alkaline cleavage assay thus revealed the relative preference of each AID on three WRC and three non-WRC-bearing oligonucleotide substrates (Fig. 4A and B); however, to examine the complete sequence specificity profile of each AID ortholog, we utilized the aforementioned deamination-specific PCR assay, since the plasmid substrate contains dC in all 16 possible trinucleotide NNC motifs. To this end, PCR amplicons representing plasmid DNA mutated by each of the five AID orthologs (Fig. 2B, right panel) were sequenced, and the mutations were mapped (Fig. 4C). Rates ranged from 0.1 to 6 mutations on average per 100 bp of the PCR amplicon, with all orthologs mediating either comparable or lower mutation loads than Hs-AID (Table 1). Pm-CDA1 mutated cytidines in an NYC motif (N, any nucleotide; Y, C or T) at higher than average rates compared to other motifs (Fig. 4D). Gc-AID mutated cytidines in the AAC and ACC motifs at 4- and 3-fold above average, respectively, while not mutating any C in a trinucleotide motif that begins with G. Tn-AID preferred the AAC motif, followed by AGC, CGC, and TGC, with the GNC motifs among the least mutated. Lc-AID mutated CCC motifs 2-fold above average, followed by AGC and CAC.

Complete agreement between the alkaline cleavage and PCR assays is not expected due to major differences in the nature of the substrates, one of which is an oligonucleotide with a single target dC and the other a plasmid with many dCs located in different regions. Nevertheless, when the substrate specificity data gleaned from the PCR deamination assay is focused on the same six motifs studied in the alkaline cleavage assay, the similarities between the two results become more apparent (Fig. 4E). The WRC versus non-WRC preferences of Gc-, Tn-, and Hs-AID are maintained, with all three AID orthologs preferring TGC, TAC, and AGC over GTC, GGC, and GAC. The unique preferences of Pm-CDA1 for GTC and Lc-AID for GGC (along with their preferred WRC motifs) are also conserved.

The corresponding sequence logo analysis of the PCR assay results depicted in Fig. 4F is in agreement with the initial sequence preference analysis. Pm-CDA1 is again unique among the other orthologs tested, since the −2 position in the XXC motif is not as relevant as the −1 position, showing equal weight for all four nucleotides in the −2 position while T and C in the −1 position are most preferred. In contrast, Gc-AID, Tn-AID, Lc-AID, and Hs-AID demonstrate a greater preference for certain nucleotides in the −2 position rather than the −1 position. Overall, we conclude that each AID ortholog has unique substrate sequence specificity patterns, with lowest preference for GNC trinucleotides being the only ubiquitous feature among three of the four jawed vertebrate orthologs.

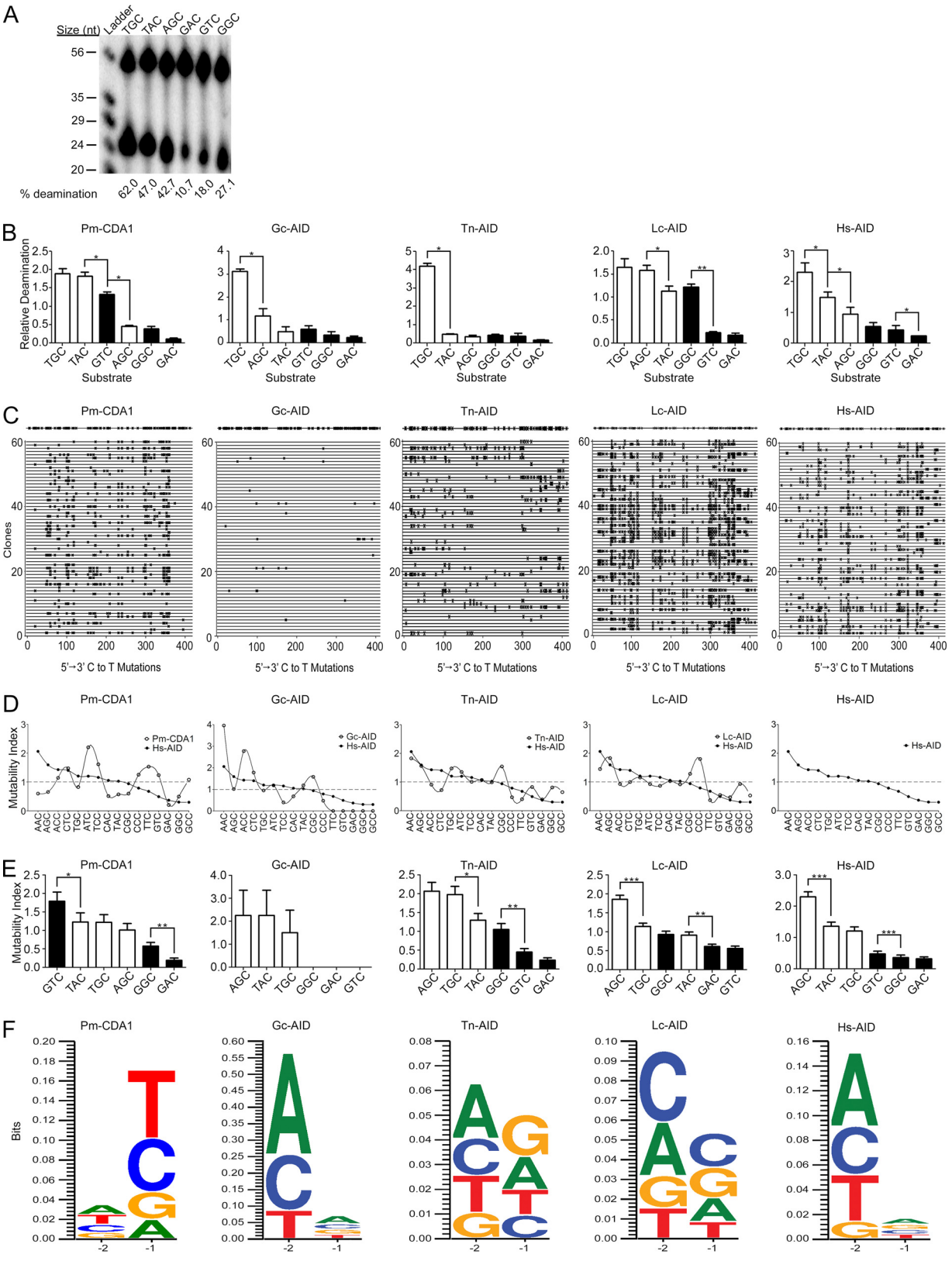

FIG 4 AID orthologs exhibit unique substrate specificity. (A) Representative alkaline cleavage gel of the substrate specificity assay, showing activity of Hs-AID on six bubble substrates (TGCbub7, TACbub7, AGCbub7, GACbub7, GTCbub7, and GGCbub7). (B) Bar graphs showing substrate specificity of each AID ortholog on six bubble substrates with the following target sequence: TGC, TAC, AGC, GGC, GTC, and GAC. Two or three purifications of one or two clones from each AID orthologs were tested two or three times for a total of three to five independent experiments for each species. The *y* axis shows relative deamination efficiency (preference relative to average of all six substrates) to enable comparison

(Continued on next page)

TABLE 1 Mutation rates of AID orthologs on denatured plasmid DNA[a]

AID ortholog	Avg dC mutations (per amplicon)	Mutation rate (per 100 nt)
Pm-CDA1	13	4
Gc-AID	0.4	0.1
Tn-AID	8	2
Lc-AID	22	6
Hs-AID	16	4

[a]nt, nucleotides.

All earlier-diverged AID orthologs have relatively low catalytic rates. We sought to assess the catalytic rates of AID orthologs by comparing initial Michaelis-Menten enzyme velocities, in the same manner previously used to demonstrate Hs-AID's lethargic catalytic rate (69). In order to determine the optimal duration of incubation with substrate for each AID ortholog to measure initial catalytic velocities, we first performed time course enzyme kinetics comparisons. Each AID ortholog was incubated with TGCbub7 at its optimal temperature for durations between 5 and 1,200 min (Fig. 5A). The substrate TGCbub7 was chosen because we found in this and previous works that it is the most preferred target of all AID orthologs in the alkaline cleavage assay, which is used to determine enzyme kinetics (36, 37). Following a similar pattern to the initial activity test seen in Fig. 2A, we found that Hs-AID was the most active, catalyzing the maximum deamination of the second most active AID ortholog, Pm-CDA1 (42 fmol/μg at 1,200 min), more than three times as fast. Tn-AID was as active as Pm-CDA1 until 360 min, after which Tn-AID catalyzed 20% less product than Pm-CDA1 at 1,200 min. Gc-AID activity plateaued at 360 min, after which it catalyzed $\sim$0.5 fmol/μg product every hour until it reached a low maximum of 17.5 fmol/μg at 1,200 min, which is 3-fold less than the activity of Hs-AID at this time. Lc-AID, catalyzed a maximum of 10 fmol/μg at 1,200 min, approximately one-fifth the product catalyzed by Hs-AID at this time point.

The optimal length of incubation for comparison of Michaelis-Menten kinetics between AID orthologs was determined based on the duration of the initial velocity phase of time course kinetics: Pm-CDA1, 240 to 420 min; Gc-, Tn-, and Lc-AID, 150 to 300 min; and Hs-AID, 120 min. We thus measured the catalytic kinetics of each AID ortholog at its optimal time and temperature on a range of TGCbub7 concentrations (Fig. 5B). Unsurprisingly, the order from highest to lowest initial catalytic rate followed the same trend as time course enzyme kinetics in Fig. 5A: Hs-AID exhibited the highest catalytic rate with an initial velocity of 0.75 fmol/μg/min. Pm-CDA1 is 33% less active, with an initial velocity of 0.50 fmol/μg/min. Tn-AID and Gc-AID had lower and similar initial velocities, between 0.18 and 0.22 fmol/μg/min, nearly a quarter the rate of Hs-AID. Similar to the time course kinetics, Gc-AID and Lc-AID were again the slowest, having initial velocities of 0.18 and 0.15 fmol/μg/min, respectively, one-fifth that of Hs-AID. Thus, based on time course and initial catalytic rate comparisons, we conclude

FIG 4 Legend (Continued)
between AID orthologs whose absolute activity levels on each substrate vary. Error bars represent the SEM. *, $P \leq 0.05$; **, $P < 0.005$. *P* values were determined by using a Mann-Whitney test. (C) Visualization of mutations. Each horizontal line is one individual sequenced amplicon of p219 that was mutated by the indicated AID ortholog and PCR amplified; each "X" is a C→T mutation. The *x* axis maps the section of p219 that is PCR amplified, minus the primers: 407 nucleotides from the first set of primers. The line at the top of each graph shows all possible dC mutations, indicated by an X. The numbers of amplicons analyzed for each AID ortholog were as follows: Pm-CDA1, 61; Gc-AID, 82; Tn-AID, 103; Lc-AID, 112; and Hs-AID, 114. (D) Comparison of substrate specificity of Hs-AID to early-diverged AID orthologs on p219. The *y* axis shows the mutability index, where 1 = average rate of mutation (dotted line) for all 16 NNC motifs. The relative preference for each individual NNC sequence was obtained by dividing its mutation rate by the average value for all 16 NNCs. The *x* axis shows XXC DNA motifs, ordered from most to least mutated by Hs-AID. Error bars represent the SD. (E) Bar graphs showing the substrate specificity of each AID ortholog on six NNC motifs—TGC, TAC, AGC, GGC, GTC, and GAC—in a PCR-based assay. The *y* axis shows the mutability index. The relative preferences for the six NNC sequences were obtained by dividing the mutation rate by the average value for the six NNC motifs identified. *, $P \leq 0.05$; **, $P < 0.005$, ***, $P < 0.0001$. *P* values were determined by using a Mann-Whitney test. Error bars represent the SEM. (F) Sequence logos showing the relative AID specificity of each nucleotide at the -2 and -1 nucleotide positions. The height of the stack shows the consensus of nucleotides at that position, and the height of each symbol within the stack indicates the relative frequency of each nucleotide at that position.

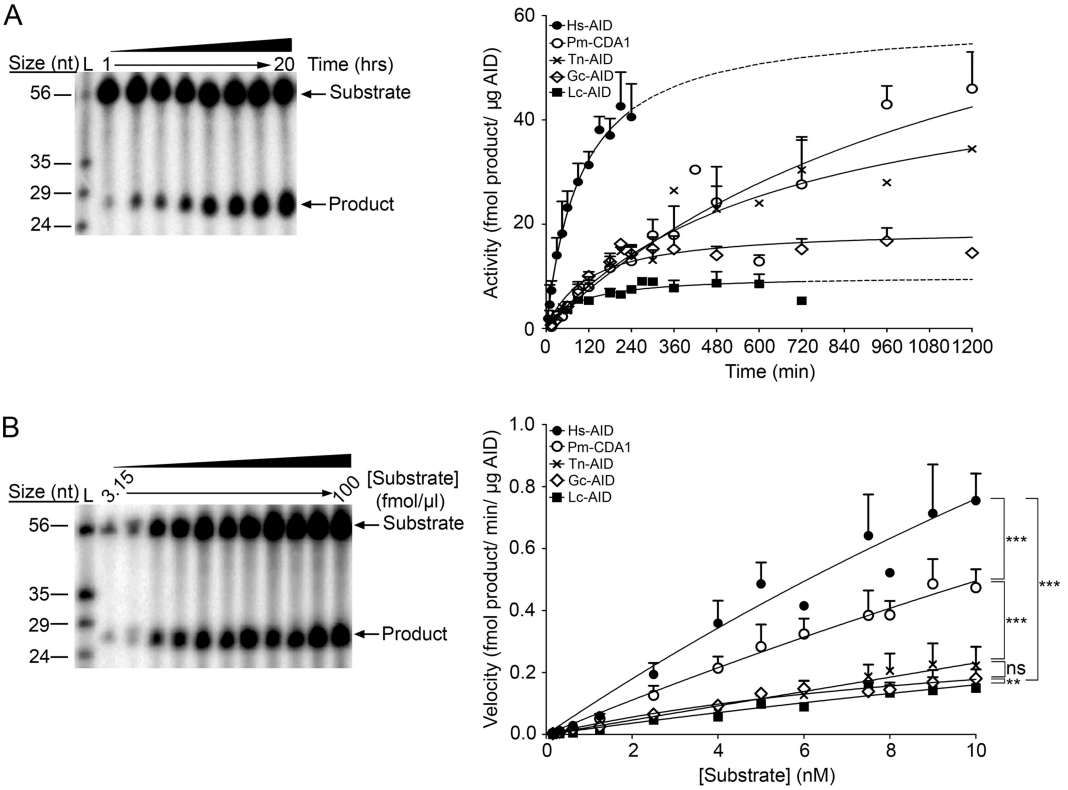

FIG 5 Early-diverged AID orthologs exhibit lower enzymatic efficiency compared to Hs AID. (A, left) Representative time course alkaline cleavage gel. Pm-CDA1 was incubated with TGCbub7 at optimal temperature (~14°C) for 1 to 20 h. (Right) Graph of combined time course data from two to seven experiments using three to six independently purified preparations of each AID ortholog; each point on the graph represents four to ten individual data points. Error bars represent the SD. (B, left) Representative Michaelis-Menten enzyme kinetics alkaline cleavage gel used to determine the initial enzymatic velocity of each AID ortholog. A 0.2-μg portion of Pm-CDA1 was incubated with concentrations of TGCbub7 ranging from 3.15 to 100 fmol/μl at the optimal temperature (~14°C) for 7 h. (Right) Graph of Michaelis-Menten kinetics showing initial velocities of the different AID orthologs. Each point represents four to six individual data points from three independent experiments on three to six independently purified preparations of each AID ortholog. Error bars represent the SD. **, $P < 0.005$; ***, $P = 0.0005$ (determined by two-tailed, nonparametric *t* test).

that all orthologs studies have a lower catalytic rate than Hs-AID. Interestingly, the earliest-diverged ortholog (Pm-CDA1) is the most active of the fish orthologs.

To ascertain whether the differences in catalytic activity between the AID orthologs are due to differences in affinity of binding to the substrate ssDNA, we performed electrophoretic mobility shift assays (EMSAs) previously optimized for measuring affinity of overall AID binding to DNA (Fig. 6A) (18, 36, 69). The average K_d values (in nM) were as follows: Pm-CDA1, 0.11; Gc-AID, 0.14; Tn-AID, 0.58; Lc-AID, 0.044; and Hs-AID, 0.14. The differences in binding affinities were not directly correlated with variation in catalytic activity (see Fig. S2 in the supplemental material); however, since all the ssDNA dissociation constants were extremely low (in the nM range), together they correspond to slower activity, compared to APOBEC3G, a sibling of AID that exhibits μM range ssDNA dissociation constants but a much faster catalytic rate (see Fig. S2 in the supplemental material) (78, 79). Thus, we conclude that all AID orthologs examined here are able to bind ssDNA efficiently with very high affinities in the nM range, similar to Hs-AID.

Predicted structures of AID orthologs provide insight into biochemical differences. In order to gain structure-function insights into the differences among orthologs, we modeled each AID's structure using several related solved APOBEC structures as the templates, similarly to the computational phase of our combined computational-biochemical approach used to solve the functional and native structure of Hs-AID (11, 21) (Fig. 7, first column). We examined the 25 lowest-energy predicted

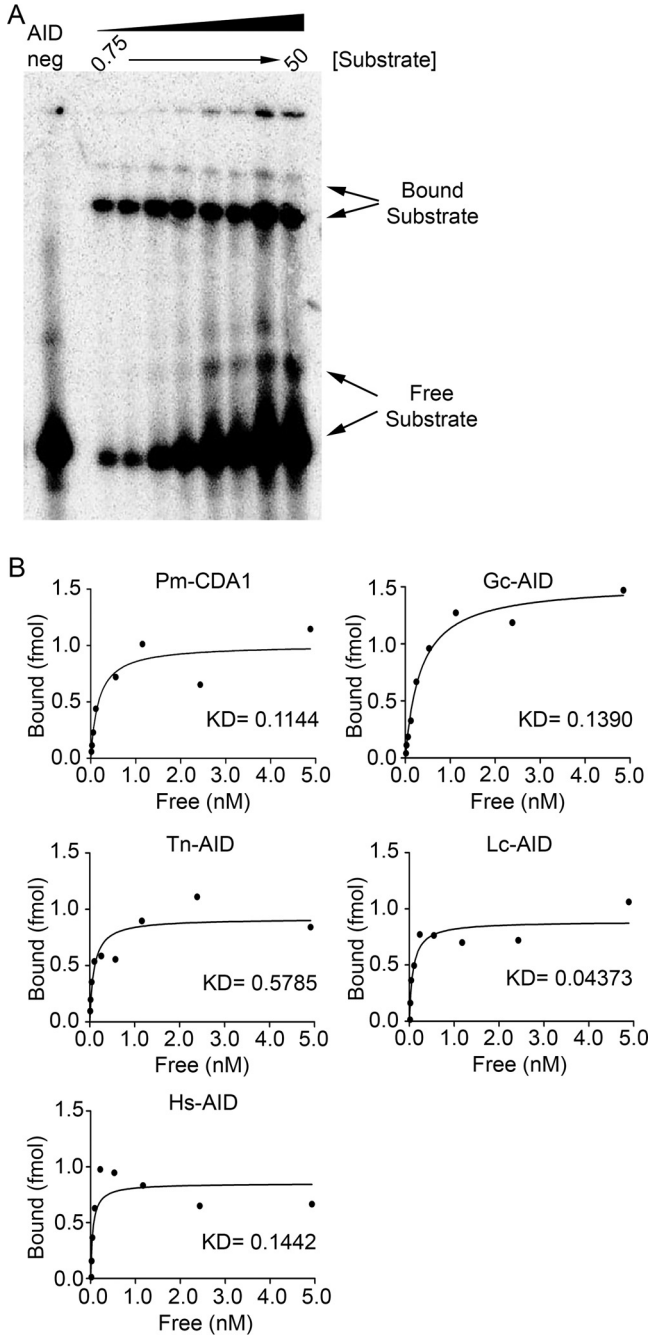

FIG 6 Differences in activity between AID orthologs are not due to ssDNA binding. (A) Representative EMSA gel of Hs-AID incubated with various concentrations of TGCbub7 at optimal temperature (31°C) for 1 h. AID in complex with substrate is found in the bound bands, while free substrate continues to migrate toward the bottom of the gel, as indicated. The negative control is TGCbub7 in binding buffer with no AID added. (B) In order to calculate K_d half-saturation binding affinities for each ortholog, the fraction of shifted substrate was quantitated, and a bound versus free plot was generated. The average K_d value for each AID ortholog is indicated within each graph, determined from three independently purified preparations of each AID ortholog.

conformations for each ortholog and found that all orthologs formed a core structure akin to the conserved core architecture of the AID/APOBEC family (21). Consistent with our finding that these orthologs are active cytidine deaminases, they all exhibited an enzymatically viable catalytic pocket at the center of which are the characteristic triad alpha-beta-alpha Zn^{2+}-coordinating motif of two cysteines and a histidine atop the

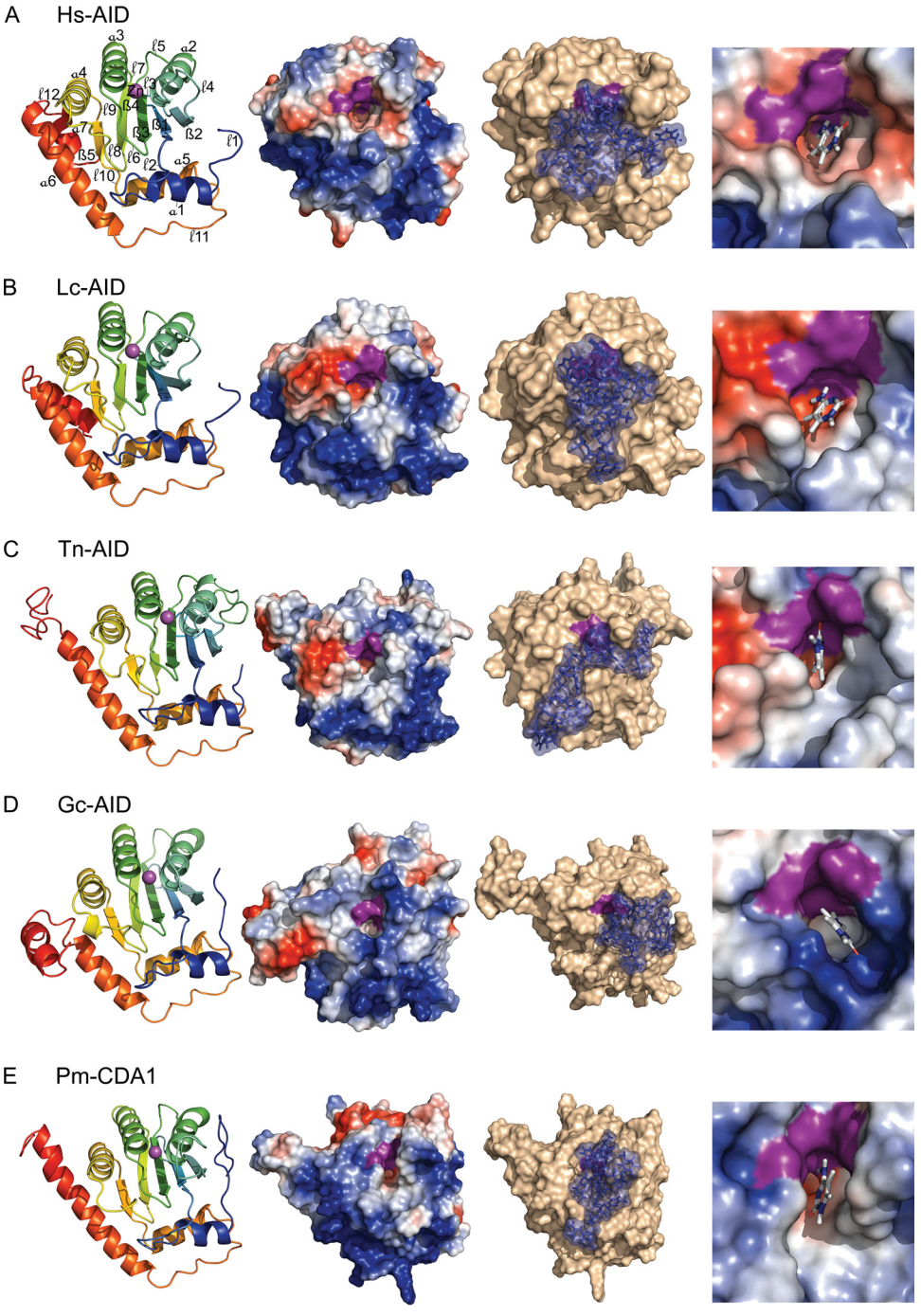

FIG 7 Predicted structures and substrate docking of AID orthologs. Shown are the cartoon (first column), surface topology (second column), surface model with single-stranded section of TGCbub7 in the DNA binding groove (third column), and catalytic pocket docked with dC (fourth column) of Hs-AID, Lc-AID, Tn-AID, Gc-AID and Pm-CDA1. Each structure is a representative of the 25 lowest-energy predicted conformations, using multiple related APOBEC X-ray and NMR structures as the templates. N-to-C terminus progression is shown from blue to red in the cartoon structures. The ribbon diagram shows Zn coordinated in the catalytic pocket (purple sphere): ℓ is loop, α is helix, and β is strand. All models illustrate the same global architecture with notable differences found in the N-terminal tails, connecting loops (ℓ2, ℓ4, ℓ5, ℓ6, and ℓ8), particularly in the loop 5 extension that is unique to bony fish, and in the structure/absence of the C-terminal α7. In the surface structures, positive and negative residues are blue and red, respectively. The catalytic pocket is seen as an indentation in the center, with the catalytic residues (Zn-coordinating triad of cysteines and histidine, catalytic glutamic acid) colored purple. A large proportion of the surface is positively charged, as reflected by the relatively high pI and net charge at neutral pH compared to other APOBEC enzymes, and similar to Hs-AID. The fourth column shows catalytically accessible conformations of each AID ortholog docked with dC in the catalytic pocket. All orthologs exhibited both open and closed catalytic pocket confirmations, similar to Hs-AID, and a representative open pocket confirmation was chosen for dC docking, since closed conformations are unable to dock dC in the pocket and thus represent an inactive conformation of AID.

TABLE 2 Percent substrate bound to AID near the catalytic pocket and correctly oriented within the DNA binding groove

AID ortholog	% substrate bound near the catalytic pocket		% catalytically viable states
	Not restricted	Restricted[a]	
Hs-AID	18 (59/320)	58 (466/800)	40 (10/25)
Lc-AID	39 (112/288)	74 (590/800)	36 (9/25)
Gc-AID	29 (56/192)	62 (498/800)	24 (6/25)
Tn-AID	42 (161/384)	63 (507/800)	48 (12/25)
Pm-CDA1	44 (99/224)	72 (572/800)	28 (7/25)

[a]That is, to within 30 by 30 by 30 Å of the catalytic pocket.

proton-donating catalytic glutamic acid (H66, E68, C97, and C100 for Pm-CDA1; H56, E58, C87, and C90 for Gc-AID; H53, E55, C93, and C96 for Tn-AID; H56, E58, C87, and C90 for Lc-AID; H56, E58, C87, and C90 for Hs-AID). We noted that each predicted AID structure exhibited a high concentration of surface-exposed positively charged residues distributed over the entire surface in locations proximal to the catalytic pocket, as we previously described for Hs-AID (21). Consequently, all AID orthologs exhibited similar isoelectric properties (pI values for Hs-, Lc-, Tn-, and Gc-AID and Pm-CDA1 of 10.0, 9.7, 10.1, 10.2, and 10.0, respectively) and charges at neutral pH (net charges for Hs-, Lc-, Tn-, and Gc-AID and Pm-CDA1 of +14.3, +11, +15.9, +11.7, and +12.7, respectively). This conservation is remarkable since there is significant divergence in isoelectric points and surface charges within the AID/APOBEC family (−2, −6, +0.5, −9, and −3.5 of human A3A, A3B-CTD, A3C, A3F-CTD, and A3G-CTD, respectively), thus making the net charge of approximately +14 being a unique and defining feature of AID (11). The conservation in high positive net charge is also consistent with our observation that all AID orthologs bind ssDNA with high-nM-range affinity (Fig. 6B).

Since we found that all AID orthologs examined had varying but consistently low catalytic robustness, we compared regions that govern the catalytic pocket and substrate specificity (Fig. 7, first and second columns). We previously described a set of secondary catalytic residues that support the aforementioned primary catalytic residues by stabilizing dC in the catalytic pocket and contributing to the physical architecture of the catalytic pocket (21). The secondary catalytic residues are contained in loops 2, 4, 6, and 8 in Hs-AID; relative breathing motion of these loops contributes to the fluidity of Hs-AID's catalytic pocket dynamics (11, 21). We previously showed that these loops contain much of the sequence divergence within the human AID/APOBEC family and that these differences mediate variable breathing dynamics of catalytic pockets (11). Here, we found that the AID orthologs also had significant differences in these loops. Pm-CDA1 had a three-residue insert and a replacement of two normally conserved secondary catalytic residues (R32 and C33) in loop 2. Nevertheless, the conformation of loop 2 remains similar to other AID and APOBECs. We noticed a large insert in loop 4 of Pm-CDA1 that compacts the residues neighboring the catalytic pocket. We also noted considerable sequence differences in loop 8 among all AID examined, with Pm-CDA1 having the shortest loop 8. In Tn-AID we observed a unique sequence in loop 5, which is conserved in AID from ray-finned fish. This sequence contains a high number of helix-breaking residues and was predicted to form an extension of loop 5. Furthermore, this extension contains a high number of hydrophobic residues and was found to mediate contact with $\alpha 2$ and the N terminus, which is likely important for dimerization.

We previously demonstrated that the catalytic pocket of AID transitions dynamically between accessible and occluded states and that only the former state can accommodate dC for deamination (21). Therefore, we sought to compare the efficiency of dC binding in the catalytic pocket of the AID orthologs. To this end, we docked a 7-nucleotide ssDNA containing the WRC motif TGC to the surface of AID and measured the proportions of ssDNA bound near the catalytic pocket and how many of those conformations were catalytically viable (Fig. 7, third and fourth columns; Table 2). As

previously demonstrated for Hs-AID (21), blind docking simulations showed sporadic binding of the ssDNA oligonucleotide on the surface, with a minor proportion of substrate bound proximal to the catalytic pocket (18% (59/320), 39% (112/288), 29% (56/192), 42% (161/384), and 44% (99/224) of DNA bound clusters for Hs-AID, Lc-AID, Gc-AID, Tn-AID, and Pm-CDA1, respectively). We then restricted the docking to the catalytic pocket and surrounding region and for Hs-AID found 58% of low-energy binding clusters resulted in DNA bound to the DNA binding grooves. For Tn-AID, Gc-AID, Lc-AID, and Pm-CDA1 we found 63% (507/800 clusters), 62% (498/800 clusters), 74% (590/800 clusters), and 72% (572/800 clusters) of low-energy ssDNA binding clusters bound to the DNA binding grooves located on the surface, respectively. For Tn- and Gc-AID, the proportions were similar to Hs-AID, but slightly higher in Lc- and Pm-CDA1. These higher proportions might be attributed to a slightly more favorably shaped DNA binding groove. We also observed that like Hs-AID, each ortholog had only a minority proportion of catalytic pockets in an open and accessible conformation able to dock dC (40% [10/25], 36% [9/25], 24% [6/25], 48% [12/25], and 28% [7/25]) for Hs-, Lc-, Tn-, and Gc-AID and Pm-CDA1, respectively (Table 2). This observation is consistent with our demonstration here that all early-diverged AID orthologs have a low catalytic rate and suggests that catalytic pocket occlusion is a crucial aspect of AID regulation that has been preserved during its divergence.

To address the structural basis of the variable substrate sequence specificities, we simulated TGC- and GTC-containing ssDNAs binding to Pm-CDA1 and Hs-AID. Our docking simulations reflected our *in vitro* results: we found that Hs-AID preferred TGC versus GTC (2.1% versus 0.9% of DNA complexes containing dC in the catalytic pocket), while Pm-CDA1 has a greater preference for GTC (2.1% versus 1.1% of DNA complexes containing dC in the catalytic pocket). These results not only explain our *in vitro* finding that Pm-CDA1 prefers GTC, the opposite preference to that of Hs-AID, but also serve as further validation of our AID:ssDNA structural prediction strategy.

DISCUSSION

Previous works have mainly focused on testing ray-finned fish AID in cell-based reporter assays, determining that these AID orthologs can initiate tetrapod-exclusive CSR in mammalian cells (32–35). This study is the first to compare purified versions of the earliest-diverging known AID orthologs. We found that all four early-diverged AID orthologs are active cytidine deaminases, supporting the theory that the downstream effects of AID mutation, like CSR and SHM, are likely more dependent on substrate availability, such as the concurrent exposure of single-stranded Ig switch-regions during transcription and transportation of AID into the nucleus of the activated B cell.

We also found that while each fish AID demonstrated various levels of activity, all AID orthologs share the exceptional biochemical properties of Hs-AID of relatively slow catalytic rates and high-nanomolar-range ssDNA binding affinities (19, 22, 69). The low catalytic rate and high binding affinity for ssDNA are consistent with predicted shared structural features of frequently occluded catalytic pockets and highly positively charged surfaces, respectively. We have shown that these two structural features of Hs-AID mediate sporadic ssDNA binding on its surface and infrequent positioning of dC in the catalytic pocket that, together, limit the activity of AID. Our finding here that even the most divergent orthologs studied share these biochemical and structural features points to the evolutionary importance of this safeguard (21).

In support of our finding that AID orthologs are lethargic *in vitro*, the relatively slow catalytic activity of wild-type AID has also been demonstrated *in vivo*: AID mutants that were more catalytically active than their wild-type Hs-AID counterpart were shown to increase antibody diversification and chromosomal translocations in B cells, suggesting that AID has evolved a suboptimal specific activity to avoid possibly tumorigenic DNA damage (23). To confirm our finding that AID's lethargy is consistent across species as a protective evolutionary trait against DNA damage, it will be critical to develop similar *in vivo* assays using AID upmutants of nonhuman AID orthologs.

The AID orthologs studied here show similarly high nanomolar range binding

affinities but also unusually low enzymatic activity. An enzyme requires a minimum binding strength to have any activity, but if the bound complex between product and enzyme is too strong, the product may remain bound too long and act like a competitor, decreasing activity. Indeed, when a related enzyme with a much higher specific activity, like APOBEC3G, is considered for comparison an inverse relationship between specific activities and binding affinities emerges (see Fig. S2 in the supplemental material). In this context, it is remarkable that even though the amino acid sequence of Pm-CDA1 is more different from Hs-AID than other human APOBECs are from Hs-AID, Pm-CDA1 has maintained the high binding affinity and low catalytic activity common to the other AID orthologs, further supporting the evolutionary importance of these biochemical properties.

Furthermore, as suggested in previous works (2, 36, 37), we found that purified fish AID are more cold adapted than their human counterpart, indicating that AID itself has adapted to changes in body temperature. Adult lamprey inhabit waters at ~20°C (80), nurse sharks prefer 25 to 30°C (81), tetraodon are found in tropical freshwater environments at ~26°C (82), coelacanths prefer deep water at 13 to 25°C (83), and humans keep an internal temperature of 37°C: these physiological temperatures are only slightly warmer than the optimal temperatures of the associated AID. The concurrence of the optimal temperature of an enzyme to the physiological temperature of the species is referred to as "corresponding states," where orthologous proteins from species living at different temperatures are in corresponding stable tertiary structures and exhibit similar thermodynamic stability at their respective optimal temperatures (84–86). Cold-adapted proteins are often more flexible than their warm-adapted counterparts so as to adopt a similar level of structural rigidity at lower temperatures (86, 87). In this context, we were surprised not to observe gross differences in apparent compaction of predicted AID ortholog structures. Rather, structure prediction suggests other reasons for variation in enzymatic activity. For example, in Tn-AID there is an extension of loop 5, which is unique to AID from ray-finned fish (not shown). This modified loop 5, enriched in proline/glycine residues, effectively bulges loop 5 away from the core toward the β2 face. Based upon the tetrameric structure of APOBEC2, AID has been proposed to dimerize at this interface, which involves interactions from β2 and loop 5 (88). Given the context of this insert, we suspect this loop 5 insert in Tn-AID may alter the dimer structure, which in turn may have ramifications on its other biochemical properties, such as activity. Furthermore, in Pm-CDA1 we noticed a distinct structure of the N-terminal tail in comparison to other AID structures examined. While the N-terminal in other AID structures forms a short extension which makes minor contact with β2, the N-terminal in Pm-CDA1 forms an extended conformation that maintains significantly more contact with β2 and partial contact with α2. This extended N-terminal tail effectively shields β2, analogous to the full-length structure of APOBEC2. Unlike the N-terminally truncated APOBEC2 which forms a tetramer, the full-length APOBEC2 structure exists solely as a monomer because the extended N-terminal tail, analogous to Pm-CDA1, effectively shields β2-β2 dimerization. This suggests that the N-terminal extension in Pm-CDA1 may result in shielding of the dimerization interface, possibly leading to prevalence of monomers, which would explain our observation of its relatively higher specific activity. The predicted models also reveal differences in the substrate specificity loops which explain the unique substrate preferences of these AID orthologs.

It has been hypothesized that Pm-CDA1 is responsible for initiating gene conversion between the placeholder untranslated region of VLRA/C and the leucine-rich repeats, which form the completed transcript (46, 47, 52). Unlike VLRB, which is predicted to perform the same actions in cyclostomes as Ig in gnathostomes, VLRA and VLRC are theorized to be more similar to T cell receptors. Therefore, it is possible that the unique preference for Pm-CDA1 for non-WRC motifs (NYC) is a reflection of coevolution with its target substrate (VLRA/C). Based on our work, we suggest differences present in loop 2 and 8 of Pm-CDA1, as well as the overall compacted state of the catalytic pocket and surrounding region, due to a four-residue insert (G60-R63) pushing loop 4 closer toward

the catalytic pocket, are responsible for its unique substrate preference. Concurrent with Ig loci appearing as the primary antigen recognition receptor in the shark family, AID seems to have evolved a preference for substrates with WRC motifs, its substrate specificity pattern becoming more like that of Hs-AID as the evolutionary distance between the species closes. Given the involvement of loop 8 in substrate recognition in Hs-AID and other APOBECs, it was not surprising to observe considerable differences of loop 8 between the AID orthologs.

This study is the first to carry out an *in vitro* comparison of specific biochemical properties in purified earliest-diverging orthologs of a tumorigenic human DNA-mutating enzyme. We discovered that the AID orthologs found in these fish species maintain the uniquely low enzymatic rate and high-affinity DNA binding of their human counterpart and that, despite structural differences that lead to various optimal temperatures and DNA substrate sequence preferences, the three defining regulatory aspects of the structure are remarkably conserved across species: an abundance of high positive surface charge, catalytic pocket inaccessibility, and frequently catalytically nonviable ssDNA binding. Together, these features inherent to AID are responsible for its lethargic catalytic activity, which we demonstrated is also conserved throughout evolution. Our results provide strong evidence that the inherent structural features that limit AID's "dangerous" mutagenic activities are biologically significant since they are remarkably conserved across evolution.

MATERIALS AND METHODS

AID expression and purification. AID-encoding open reading frames (ORFs) were synthesized in pBluescript (GenScript, USA). The ORF sequence of Pm-CDA1 was based on the published sequence (52). The Gc-AID ORF was based on a cDNA sequence obtained from nurse shark spleen cDNA library, amplified from the highly conserved central portion, which was then used as a probe for phage library. The ORF of Lc-AID was assembled from the published coelacanth genome (65). ORFs of these AID orthologs were cloned into PGEX-5X-3 (GE Healthcare, USA) to generate GST-AID expression constructs, as we have previously described for Hs-, Dr-, Ip-, and Tn-AID (36, 37). The catalytically dead AID mutants were created by site-directed mutagenesis, as previously described (36, 37). For each AID, a minimum of two independently cloned expression constructs of correct sequence in PGEX-5X-3 were selected for protein expression, and for each clone a minimum of two and up to four independent GST-AID preparations were purified, as described previously (36, 69). Purity and relative concentrations of each GST-AID preparation were assessed on Coomassie blue-stained SDS-PAGE, and protein concentrations were equalized using bovine serum albumin (BSA) standard curves, as previously described (36, 37, 69).

Preparation of substrates for enzyme assays. The partially single-stranded substrates containing a 7-nucleotide single-stranded bubble in the middle were prepared as previously described (18, 69). Six substrates were used, differing in the dinucleotide sequence immediately upstream of the target dC nucleotide; these substrates included three AID hot-spot (WRC, where W is T/A and R is G/A) and three cold-spot (non-WRC) motifs, where dC is the target for AID mutation. Polynucleotide kinase (NEB, USA) was used to 5′ label 2.5 pmol of the target strand with [γ-^{32}P]ATP, followed by purification using Mini-Quick spin DNA columns (Roche, USA) and annealing with a 3-fold excess of the complementary strand to generate bubble substrates, as previously described (36, 37, 69).

Alkaline cleavage deamination assay. To measure AID activity, we used the standard alkaline cleavage assays for deamination as previously described (11, 18, 36, 37, 69). Briefly, AID was incubated with radioactively labeled substrate in 100 mM phosphate buffer (pH 7.2; final volume, 10 μl) and then heat inactivated at 85°C for 25 min. After uracil excision by UDG, NaOH and heat were used to induce cleavage at the alkali-labile abasic site of the uracil and denaturation; loading dye was then added, and electrophoresis was performed on a 14% denaturing acrylamide gel (1× Tris-borate-EDTA [TBE], 25% formamide, 14% acrylamide-bisacrylamide, 45% urea). The gels were exposed to a Kodak Storage Phosphor Screen GP (Bio-Rad) and visualized using PhosphorImager Quantity One software (Bio-Rad, USA). To initially ascertain whether a preparation of GST-AID was enzymatically active, it was incubated with 50 fmol of TGCbub7, which has previously been demonstrated to be an optimal substrate for AID orthologs studied thus far (36, 37), at 18, 25, and 37°C overnight.

To determine thermosensitivity profiles, 50 fmol TGCbub7 was incubated with 0.5 to 1 μg of AID in 100 mM activity buffer (final volume 10 μl) for 0.5 to 5 h at temperatures ranging from 4 to 40°C. Reactions were then heat inactivated at 85°C for 25 min and subjected to UDG and NaOH treatment as described above. Two to six individual experiments were performed per unique preparation (2 to 4) of each purified AID ortholog. For each experiment, the percent deamination was plotted against temperature to generate a curve, with the peak revealing the optimal temperature of AID in each experiment. Optimal temperatures from 6 to 12 individual experiments were then averaged for each AID ortholog to arrive at its optimal temperature.

To determine substrate specificity, each AID ortholog was incubated with 50 fmol of ^{32}P-labeled bubble substrates with the following target sequence: TGC, TAC, AGC, GGC, GAC, or GTC at each AID ortholog's optimal temperature for 16 or 24 h. Three experiments were performed on each AID ortholog,

with two or three purifications of each AID ortholog being used. The relative percent deamination was used to normalize interpreparation variations in overall activity levels to enable comparison between different purifications of AID.

To determine time point kinetics, each AID ortholog was incubated with 25 to 50 fmol of TGCbub7 at each AID ortholog's optimal temperature for durations ranging from 5 min to 20 h. Three independent preparations of each AID ortholog were used, except for Lc-AID, for which six purifications were tested, all in two to seven separate experiments. The amount of deaminated substrate in fmol/μg of AID was plotted against time incubated. These experiments were used to determine time of initial exponential phase of AID activity in order to proceed to Michaelis-Menten kinetics.

To determine Michaelis-Menten kinetics for comparison of initial velocities, 0.2 μg of AID was incubated with 1.5 to 100 fmol of TGCbub7 at each AID ortholog's optimal temperature for 2 to 7 h (depending on the results of time point kinetics). Two or three independent preparations of each AID ortholog were tested in two or three individual experiments. The percent deamination was used to calculate the velocity (fmol of product/min of incubation/μg of AID), which was then plotted against the substrate concentration.

PCR-based deamination assay. We have previously described the deamination-specific PCR assay to detect activity of purified AID on an ~500-bp stretch of denatured plasmid (14, 15). Briefly, 150 ng of HindIII-digested p219 plasmid was denatured at 99°C for 10 min in phosphate buffer (pH 7.2), followed by snap-cooling in an ice bath. AID was added and the solution was incubated for 1 h at the optimal temperature for each AID ortholog as determined in this study. The plasmid substrate was again denatured and snap-cooled, followed by a second round of AID addition to a final volume of 40 μl, followed in turn by incubation for another hour. Then, 4 ng of plasmid (2 μl of the incubation solution) was PCR amplified using deamination-specific primers in a total volume of 25 μl, as previously described (15). The DNA was then gel purified (Qiagen) and TA-cloned (Invitrogen) and, for each AID ortholog, ~100 amplicons (each 407 nucleotides in size) were sequenced (Macrogen, South Korea). If mutations were not present, a second, more stringent nested deamination-specific PCR was carried out wherein 2 μl of the first PCR was added to the nested PCR, which had 7.5 pmol of a nested set of deamination-specific primers, resulting in amplicons of 329 nucleotides (15). In order to analyze substrate specificity patterns, annealing temperatures were kept at the minimum required to detect mutated sequences so as to avoid biasing the PCR products toward highly mutated amplicons. The resulting DNA was then TA cloned (Invitrogen) and sequenced (Macrogen). Amplicons with unique mutation patterns were then analyzed for the load and location of mutations induced by each AID ortholog.

EMSA. The use of an electrophoretic mobility shift assay (EMSA) to determine AID's binding affinity for ssDNA has previously been described (18, 36, 69). Briefly, 0.4 to 50 fmol of TGCbub7 was incubated with each AID ortholog in binding buffer (1 mM dithiothreitol, 2 μM $MgCl_2$, 50 μM NaCl, dH_2O) in a total volume of 10 μl at 25°C or its optimal temperature for 1 h, followed by UV-cross-linking on ice, as previously described (69). Then, 5 μl of loading dye was added to each sample, followed by electrophoresis on an 8% native acrylamide gel (0.5× TBE, 8% acrylamide-bisacrylamide, 6% glycerol, dH_2O) and electrophoresed at 4°C. The gels were exposed to a Kodak Storage Phosphor Screen GP (Bio-Rad) and visualized using a PhosphorImager (Bio-Rad, Hercules, CA) on Quantity One software (Bio-Rad). Two or three independent preparations of each AID ortholog were tested, each in two or three independent experiments. The amount of substrate bound (fmol) to AID was plotted against the concentration (nM) of free substrate, followed by nonsaturation binding kinetic analysis (Prism software; GraphPad, San Diego, CA) to obtain binding half-saturation K_d values.

Structure prediction of AID and AID-DNA complexes. We followed a similar approach as previously used in our combined computational-biochemical methodology for obtaining AID's native and functional structure (11, 21). Briefly, five resolved homologous APOBEC structures were chosen as the templates for homology modeling: mouse A2 NMR (PDB 2RPZ), A3A NMR (PDB 2M65), A3C (PDB 3VOW), A3F-CTD X-ray (PDB 4IOU), and A3G-CTD X-ray (PDB 3E1U). All APOBEC structures were obtained from the Protein Data Bank (http://www.rcsb.org) and visualized by using PyMOL v1.7.6 (http://www.pymol.org). Using the default parameters of I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/), full-length AID was modeled from the APOBEC templates to generate 25 models for each, totaling 125 models. For each model, the majority of the protein was homology-modeled with the exception of the nonhomologous 18 C-terminal amino acids, which were modeled *ab initio*. The catalytic pocket was defined by the indented space containing the Zn-coordinating and catalytic residues (H56, E58, C87, and C90 in Hs-AID). The protein charge and isoelectric point (pI) were calculated for each AID structure using the PARSE force field in PROPKA 3.0 (89–91).

DNA substrates were docked to each AID model using Swiss-Dock (http://www.swissdock.ch) (92, 93). Each substrate was constructed in Marvin Sketch v.5.11.5 (http://www.chemaxon.com/products/marvin/marvinsketch/), while surface topology and docking parameters were generated using Swiss-Param (http://swissparam.ch) (94). These output files served as the ligand file in Swiss Dock. 5′-TTTGCTT-3′ and 5′-TTGTCTT-3′ ssDNA substrates were chosen, since the former has been shown to be the preferred substrate of both human and bony fish AID (16–18, 69, 95), and the latter was among the preferred substrates of Pm-CDA1 in this study. Substrate docking simulations for each AID enzyme resulted in 5,000 to 15,000 binding modes, 8 or more of which were clustered based on root mean square values. The 32 lowest-energy clusters were selected, thus representing 256 of the lowest-energy individual binding events within 32 low-energy clusters for each AID. For blind docking, we simulated ssDNA binding with the entire surface of AID. For restricted docking, we simulated ssDNA binding within 30 by 30 by 30 Å (x, y, z) from the Zn-coordinating histidine in the catalytic pocket. In total, 7,008 low-energy clusters were analyzed for each protein, per substrate. In total, 5,600

low-energy docking clusters were analyzed for AID orthologs in UCSF Chimera v1.7 (https://www.cgl.ucsf.edu/chimera) (96). Deamination-conducive AID-DNA complexes were defined by the accessibility of the dC-NH_2 substrate group to the catalytic Zn-coordinating and the glutamic acid residues (H56, C87, C90, and E58 in wild-type Hs-AID).

Data quantitation and statistics. The amino acid alignment analysis was performed using STRAP (http://www.bioinformatics.org/strap/) version 2018-08-11. The nucleotide alignment analysis was performed using PROMALS3D (http://prodata.swmed.edu/promals3d/promals3d.php). The sequence logos for the PCR deamination assay were generated using WebLogo 3 software. The mutation frequencies of the XXC motifs generated by each AID ortholog were translated into sequence consensus data: the higher the frequency of mutation for a certain XXC motif, the higher the number of multiples of said motif in the consensus, and vice versa. These sequence consensus data were then entered into the WebLogo 3 algorithm.

For enzyme activity and binding assays, band densitometry was performed using Image Lab Analysis Software (Bio-Rad) and Quantity One as previously described (18, 21, 36, 37, 69). For each experiment, individual lanes of each gel were quantitated three times (variability was $<$5%). Values from individual experiments were then averaged and plotted as a single data point. For thermosensitivity and binding assays, the optimal temperatures and K_d values, respectively, were calculated for each AID ortholog in each experiment and were averaged to obtain final values. For substrate specificity, time course, and Michaelis-Menten kinetics experiments, data points gathered from all independent experiments were averaged for each AID ortholog. The data were graphed and statistically analyzed using GraphPad Prism software. Error bars represent standard errors of the mean (SEM) or standard deviations (SD) as indicated. *P* values for the substrate specificity assay were determined by the Mann-Whitney test, which compared the relative deamination of different substrates. The *P* values for the Michaelis-Menten kinetics assay were determined by using a two-tailed, nonparametric paired *t* test, which compared the initial velocities of the five AID orthologs.

Accession numbers. NCBI accession numbers were as follows: *Homo sapiens* activation-induced cytidine deaminase, AAS92920.1; *Callithrix jacchus* predicted single-stranded DNA cytosine deaminase isoform X1, XP_017832076.1; *Pteropus vampyrus* predicted single-stranded DNA cytosine deaminase, XP_011364076.1; *Ornithorhynchus anatinus* predicted single-stranded DNA cytosine deaminase, XP_001516174.2; *Anolis carolinensis* predicted single-stranded DNA cytosine deaminase isoform X1, XP_008102036.1; *Pleurodeles waltl* activation-induced cytidine deaminase, CBG76578.2; *Petromyzon marinus* cytosine deaminase, ABO15149.1. The UniProt (http://www.uniprot.org/) accession number was as follows: *Tetraodon nigroviridis* uncharacterized protein (CMP/dCMP-type deaminase), H3CRQ9.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/MCB.00077-17.

SUPPLEMENTAL FILE 1, PDF file, 0.5 MB.

ACKNOWLEDGMENTS

This study was supported by a Canadian Institutes of Health Research (CIHR) operating grant (MOP111132), a Canadian Cancer Society Research Institute (CCSRI) Innovation operating grant, and a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant (2015-047960) to M.L. E.M.Q. is supported by an NSERC Postgraduate Scholarship-Doctoral scholarship.

We have all completed the ICMJE uniform disclosure form and declare no support from any organization for the submitted work, no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years, and no other relationships or activities that could appear to have influenced the submitted work.

REFERENCES

1. Smith HC, Bennett RP, Kizilyer A, McDougall WM, KMP. 2012. Functions and regulation of the APOBEC family of proteins. Semin Cell Dev Biol 23:258–268. https://doi.org/10.1016/j.semcdb.2011.10.004.
2. Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS. 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol 22:367–377. https://doi.org/10.1093/molbev/msi026.
3. Keim C, Kazadi D, Rothschild G, Basu U. 2013. Regulation of AID, the B-cell genome mutator. Genes Dev 27:1–17. https://doi.org/10.1101/gad.200014.112.
4. 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. https://doi.org/10.1016/S0092-8674(00)00078-7.
5. Knisbacher BA, Gerber D, Levanon EY. 2016. DNA editing by APOBECs: a genomic preserver and transformer. Trends Genet 32:16–28. https://doi.org/10.1016/j.tig.2015.10.005.
6. Pettersen HS, Galashevskaya A, Doseth B, Sousa MM, Sarno A, Visnes T, Aas PA, Liabakk NB, Slupphaug G, Saetrom P, Kavli B, Krokan HE. 2015. AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature. DNA Repair (Amst) 25:60–71. https://doi.org/10.1016/j.dnarep.2014.11.006.
7. Robbiani DF, Nussenzweig MC. 2013. Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. Annu Rev Pathol 8:79–103. https://doi.org/10.1146/annurev-pathol-020712-164004.
8. Munoz DP, Lee EL, Takayama S, Coppe JP, Heo SJ, Boffelli D, Di Noia JM, Martin DI. 2013. Activation-induced cytidine deaminase (AID) is necessary for the epithelial-mesenchymal transition in mammary epithelial

cells. Proc Natl Acad Sci U S A 110:E2977–E2986. https://doi.org/10.1073/pnas.1301021110.
9. Gostissa M, Alt FW, Chiarle R. 2011. Mechanisms that promote and suppress chromosomal translocations in lymphocytes. Annu Rev Immunol 29:319–350. https://doi.org/10.1146/annurev-immunol-031210-101329.
10. Gazumyan A, Bothmer A, Klein IA, Nussenzweig MC, McBride KM. 2012. Activation-induced cytidine deaminase in antibody diversification and chromosome translocation. Adv Cancer Res 113:167–190. https://doi.org/10.1016/B978-0-12-394280-7.00005-1.
11. King JJ, Larijani M. 2017. A novel regulator of activation-induced cytidine deaminase/APOBECs in immunity and cancer: Schrodinger's CATalytic Pocket. Front Immunol 8:351. https://doi.org/10.3389/fimmu.2017.00351.
12. Bransteitter R, Pham P, Scharff MD, Goodman MF. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc Natl Acad Sci U S A 100:4102–4107. https://doi.org/10.1073/pnas.0730835100.
13. Dickerson SK, Market E, Besmer E, Papavasiliou FN. 2003. AID mediates hypermutation by deaminating single-stranded DNA. J Exp Med 197:1291–1296. https://doi.org/10.1084/jem.20030481.
14. Larijani M, Frieder D, Basit W, Martin A. 2005. The mutation spectrum of purified AID is similar to the mutability index in Ramos cells and in *ung*$^{-/-}$ *msh2*$^{-/-}$ mice. Immunogenetics 56:840–845. https://doi.org/10.1007/s00251-004-0748-0.
15. Larijani M, Frieder D, Sonbuchner TM, Bransteitter R, Goodman MF, Bouhassira EE, Scharff MD, Martin A. 2005. Methylation protects cytidines from AID-mediated deamination. Mol Immunol 42:599–604. https://doi.org/10.1016/j.molimm.2004.09.007.
16. Pham P, Bransteitter R, Petruska J, Goodman MF. 2003. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature 424:103–107. https://doi.org/10.1038/nature01760.
17. Sohail A, Klapacz J, Samaranayake M, Ullah A, Bhagwat AS. 2003. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. Nucleic Acids Res 31:2990–2994. https://doi.org/10.1093/nar/gkg464.
18. Larijani M, Petrov AP, Kolenchenko O, Berru M, Krylov SN, Martin A. 2007. AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. Mol Cell Biol 27:20–30. https://doi.org/10.1128/MCB.00824-06.
19. Pham P, Calabrese P, Park SJ, Goodman MF. 2011. Analysis of a single-stranded DNA-scanning process in which activation-induced deoxycytidine deaminase (AID) deaminates C to U haphazardly and inefficiently to ensure mutational diversity. J Biol Chem 286:24931–24942. https://doi.org/10.1074/jbc.M111.241208.
20. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R. 2011. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. Biochemistry 50:4402–4410. https://doi.org/10.1021/bi2002289.
21. King JJ, Manuel CA, Barrett CV, Raber S, Lucas H, Sutter P, Larijani M. 2015. Catalytic pocket inaccessibility of activation-induced cytidine deaminase is a safeguard against excessive mutagenic activity. Structure 23:615–627. https://doi.org/10.1016/j.str.2015.01.016.
22. Larijani M, Martin A. 2012. The biochemistry of activation-induced deaminase and its physiological functions. Semin Immunol 24:255–263. https://doi.org/10.1016/j.smim.2012.05.003.
23. Wang M, Yang Z, Rada C, Neuberger MS. 2009. AID upmutants isolated using a high-throughput screen highlight the immunity/cancer balance limiting DNA deaminase activity. Nat Struct Mol Biol 16:769–776. https://doi.org/10.1038/nsmb.1623.
24. Diaz M, Greenberg AS, Flajnik MF. 1998. Somatic hypermutation of the new antigen receptor gene (NAR) in the nurse shark does not generate the repertoire: possible role in antigen-driven reactions in the absence of germinal centers. Proc Natl Acad Sci U S A 95:14343–14348. https://doi.org/10.1073/pnas.95.24.14343.
25. Hinds-Frey KR, Nishikata H, Litman RT, Litman GW. 1993. Somatic variation precedes extensive diversification of germline sequences and combinatorial joining in the evolution of immunoglobulin heavy chain diversity. J Exp Med 178:815–824. https://doi.org/10.1084/jem.178.3.815.
26. Lee SS, Tranchina D, Ohta Y, Flajnik MF, Hsu E. 2002. Hypermutation in shark immunoglobulin light chain genes results in contiguous substitutions. Immunity 16:571–582. https://doi.org/10.1016/S1074-7613(02)00300-X.
27. Kaattari SL, Zhang HL, Khor IW, Kaattari IM, Shapiro DA. 2002. Affinity maturation in trout: clonal dominance of high-affinity antibodies late in the immune response. Dev Comp Immunol 26:191–200. https://doi.org/10.1016/S0145-305X(01)00064-7.
28. Yang F, Waldbieser GC, Lobb CJ. 2006. The nucleotide targets of somatic mutation and the role of selection in immunoglobulin heavy chains of a teleost fish. J Immunol 176:1655–1667. https://doi.org/10.4049/jimmunol.176.3.1655.
29. Saunders HL, Magor BG. 2004. Cloning and expression of the AID gene in the channel catfish. Dev Comp Immunol 28:657–663. https://doi.org/10.1016/j.dci.2004.01.002.
30. Zhao Y, Pan-Hammarström Q, Zhao Z, Hammarström L. 2005. Identification of the activation-induced cytidine deaminase gene from zebrafish: an evolutionary analysis. Dev Comp Immunol 29:61–71. https://doi.org/10.1016/j.dci.2004.05.005.
31. Zhu C, Lee V, Finn A, Senger K, Zarrin AA, Du Pasquier L, Hsu E. 2012. Origin of immunoglobulin isotype switching. Curr Biol 22:872–880. https://doi.org/10.1016/j.cub.2012.03.060.
32. Ichikawa HT, Sowden MP, Torelli AT, Bachl J, Huang P, Dance GS, Marr SH, Robert J, Wedekind JE, Smith HC, Bottaro A. 2006. Structural phylogenetic analysis of activation-induced deaminase function. J Immunol 177:355–361. https://doi.org/10.4049/jimmunol.177.1.355.
33. Barreto VM, Pan-Hammarstrom Q, Zhao Y, Hammarstrom L, Misulovin Z, Nussenzweig MC. 2005. AID from bony fish catalyzes class switch recombination. J Exp Med 202:733–738. https://doi.org/10.1084/jem.20051378.
34. Wakae K, Magor BG, Saunders H, Nagaoka H, Kawamura A, Kinoshita K, Honjo T, Muramatsu M. 2006. Evolution of class switch recombination function in fish activation-induced cytidine deaminase, AID. Int Immunol 18:41–47. https://doi.org/10.1093/intimm/dxh347.
35. Barreto VM, Magor BG. 2011. Activation-induced cytidine deaminase structure and functions: a species comparative view. Dev Comp Immunol 35:991–1007. https://doi.org/10.1016/j.dci.2011.02.005.
36. Dancyger AM, King JJ, Quinlan MJ, Fifield H, Tucker S, Saunders HL, Berru M, Magor BG, Martin A, Larijani M. 2012. Differences in the enzymatic efficiency of human and bony fish AID are mediated by a single residue in the C terminus modulating single-stranded DNA binding. FASEB J 26:1517–1525. https://doi.org/10.1096/fj.11-198135.
37. Abdouni H, King JJ, Suliman M, Quinlan M, Fifield H, Larijani M. 2013. Zebrafish AID is capable of deaminating methylated deoxycytidines. Nucleic Acids Res 41:5457–5468. https://doi.org/10.1093/nar/gkt212.
38. Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, Cairns BR. 2008. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell 135:1201–1212. https://doi.org/10.1016/j.cell.2008.11.042.
39. Laird DJ, De Tomaso AW, Cooper MD, Weissman IL. 2000. 50 million years of chordate evolution: seeking the origins of adaptive immunity. Proc Natl Acad Sci U S A 97:6924–6926. https://doi.org/10.1073/pnas.97.13.6924.
40. Pancer A, Saha Ratan N, Kasamatsu J, Suzuki T, Amemiya CT, Kasahara M, Cooper MD. 2005. Variable lymphocyte receptors in hagfish. Proc Natl Acad Sci U S A 102:9224–9229. https://doi.org/10.1073/pnas.0503792102.
41. Pancer Z, Amemiya CT, Ehrhardt GRA, Ceitlin J, Gartland GL, Cooper MD. 2004. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature 430:174–180. https://doi.org/10.1038/nature02740.
42. Deng L, Velikovsky CA, Xu G, Iyer LM, Tasumi S, Kerzic MC, Flajnik MF, Aravind L, Pancer Z, Mariuzza RA. 2010. A structural basis for antigen recognition by the T cell-like lymphocytes of sea lamprey. Proc Natl Acad Sci U S A 107:13408–13413. https://doi.org/10.1073/pnas.1005475107.
43. Tasumi S, Velikovsky CA, Xu G, Gai SA, Wittrup KD, Flajnik MF, Mariuzza RA, Pancer Z. 2009. High-affinity lamprey VLRA and VLRB monoclonal antibodies. Proc Natl Acad Sci U S A 106:12891–12896. https://doi.org/10.1073/pnas.0904443106.
44. Kasamatsu J, Sutoh Y, Fugo K, Otsuka N, Iwabuchi K, Kasahara M. 2010. Identification of a third variable lymphocyte receptor in the lamprey. Proc Natl Acad Sci U S A 107:14304–14308. https://doi.org/10.1073/pnas.1001910107.
45. Kirchdoerfer RN, Herrin BR, Han BW, Turnbough CLJ, Cooper MD, Wilson IA. 2012. Variable lymphocyte receptor recognition of the immunodominant glycoprotein of *Bacillus anthracis* spores. Structure 20:479–486. https://doi.org/10.1016/j.str.2012.01.009.
46. Das S, Hirano M, Aghaallaei N, Bajoghli B, Boehm T, Cooper MD. 2013. Organization of lamprey variable lymphocyte receptor C locus and

repertoire development. Proc Natl Acad Sci U S A 110:6043–6048. https://doi.org/10.1073/pnas.1302500110.
47. Hirano M, Guo P, McCurley N, Schorpp M, Das S, Boehm T, Cooper MD. 2013. Evolutionary implications of a third lymphocyte lineage in lampreys. Nature 501:435–438. https://doi.org/10.1038/nature12467.
48. Li J, Das S, Herrin BR, Hirano M, Cooper MD. 2013. Definition of a third VLR gene in hagfish. Proc Natl Acad Sci U S A 110:15013–15018. https://doi.org/10.1073/pnas.1314540110.
49. Holland SJ, Gao M, Hirano M, Iyer LM, Luo M, Schorpp M, Cooper MD, Aravind L, Mariuzza RA, Boehm T. 2014. Selection of the lamprey VLRC antigen receptor repertoire. Proc Natl Acad Sci U S A 111:14834–14839. https://doi.org/10.1073/pnas.1415655111.
50. Luo M, Velikovsky CA, Yang X, Siddiqui MA, Hong X, Barchi JJJ, Gildersleeve JC, Pancer Z, Mariuzza RA. 2013. Recognition of the Thomsen-Friedenreich pancarcinoma carbohydrate antigen by a lamprey variable lymphocyte receptor. J Biol Chem 288:23597–23606. https://doi.org/10.1074/jbc.M113.480467.
51. Mayer WE, Uinuk-ool T, Tichy H, Gartland LA, Klein J, Cooper MD. 2002. Isolation and characterization of lymphocyte-like cells from a lamprey. Proc Natl Acad Sci U S A 99:14350–14355. https://doi.org/10.1073/pnas.212527499.
52. Rogozin IB, Iyer LM, Liang L, Glazko GV, Liston VG, Pavlov YI, Aravind L, Pancer Z. 2007. Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat Immunol 8:647–656. https://doi.org/10.1038/ni1463.
53. Guo P, Hirano M, Herrin BR, Li J, Yu C, Sadlonova A, Cooper MD. 2009. Dual nature of the adaptive immune system in lampreys. Nature 459:796–801. https://doi.org/10.1038/nature08068.
54. Bajoghli B, Guo P, Aghaallaei N, Hirano M, Strohmeier C, McCurley N, Bockman DE, Schorpp M, Cooper MD, Boehm T. 2011. A thymus candidate in lampreys. Nature 470:90–94. https://doi.org/10.1038/nature09655.
55. Tacchi L, Larragoite ET, Munoz P, Amemiya CT, Salinas I. 2015. African lungfish reveal the evolutionary origins of organized mucosal lymphoid tissue in vertebrates. Curr Biol 25:2417–2424. https://doi.org/10.1016/j.cub.2015.07.066.
56. Malecek K, Brandman J, Brodsky JE, Ohta Y, Flajnik MF, Hsu E. 2005. Somatic hypermutation and junctional diversification at Ig heavy chain loci in the nurse shark. J Immunol 175:8105–8115. https://doi.org/10.4049/jimmunol.175.12.8105.
57. Hirano M. 2015. Evolution of vertebrate adaptive immunity: immune cells and tissues, and AID/APOBEC cytidine deaminases. Bioessays 37:877–887. https://doi.org/10.1002/bies.201400178.
58. Buonocore F, Gerdol M. 2016. Alternative adaptive immunity strategies: coelacanth, cod and shark immunity. Mol Immun 69:157–169. https://doi.org/10.1016/j.molimm.2015.09.003.
59. Hsu E. 2016. Assembly and expression of shark Ig genes. J Immunol 196:3517–3523. https://doi.org/10.4049/jimmunol.1600164.
60. Hinds KR, Litman GW. 1986. Major reorganization of immunoglobulin VH segmental elements during vertebrate evolution. Nature 320:546–549. https://doi.org/10.1038/320546a0.
61. Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, Hoon S, Gangu V, Roy SW, Irimia M, Korzh V, Kondrychyn I, Lim ZW, Tay BH, Tohari S, Kong KW, Ho S, Lorente-Galdos B, Quilez J, Marques-Bonet T, Raney BJ, Ingham PW, Tay A, Hillier LW, Minx P, Boehm T, Wilson RK, Brenner S, Warren WC. 2014. Elephant shark genome provides unique insights into gnathostome evolution. Nature 505:174–179. https://doi.org/10.1038/nature12826.
62. Saunders HL, Oko AL, Scott AN, Fan CW, Magor BG. 2010. The cellular context of AID expressing cells in fish lymphoid tissues. Dev Comp Immunol 34:669–676. https://doi.org/10.1016/j.dci.2010.01.013.
63. Magor BG. 2015. Antibody affinity maturation in fishes: our current understanding. Biology (Basel) 4:512–524.
64. Amemiya CT, Alföldi J, Lee A, Fan S, Philippe H, Maccallum I, Braasch I, Manousaki T, Schneider I, Rohner N, Organ C, Chalopin D, Smith J, Robinson M, Dorrington R, Gerdol M, Aken B, Biscotti M, Barucca M, Baurain D, Berlin A, Blatch G, Buonocore F, Burmester T, Campbell M, Canapa A, Cannon J, Christoffels A, De Moro G, Edkins A, Fan L, Fausto A, Feiner N, Forconi M, Gamieldien J, Gnerre S, Gnirke A, Goldstone J, Haerty W, Hahn M, Hesse U, Hoffmann S, Johnson J, Karchner S, Kuraku S, Lara M, Levin J, Litman G, Mauceli E, Miyake T, et al. 2013. The African coelacanth genome provides insights into tetrapod evolution. Nature 496:311–316. https://doi.org/10.1038/nature12027.
65. Amemiya CT, Dorrington R, Meyer A. 2014. The coelacanth and its genome. J Exp Zool B Mol Dev Evol 322:317–321. https://doi.org/10.1002/jez.b.22583.
66. Saha NR, Ota T, Litman GW, Hansen J, Parra Z, Hsu E, Buonocore F, Canapa A, Cheng JF, Amemiya CT. 2014. Genome complexity in the coelacanth is reflected in its adaptive immune system. J Exp Zool B Mol Dev Evol 322:438–463. https://doi.org/10.1002/jez.b.22558.
67. Beale RC, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS. 2004. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. J Mol Biol 337:585–596. https://doi.org/10.1016/j.jmb.2004.01.046.
68. Brar SS, Sacho EJ, Tessmer I, Croteau DL, Erie DA, Diaz M. 2008. Activation-induced deaminase, AID, is catalytically active as a monomer on single-stranded DNA. DNA Repair (Amst) 7:77–87. https://doi.org/10.1016/j.dnarep.2007.08.002.
69. Larijani M, Martin A. 2007. Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID. Mol Cell Biol 27:8038–8048. https://doi.org/10.1128/MCB.01046-07.
70. Petersen-Mahrt SK, Neuberger MS. 2003. In vitro deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). J Biol Chem 278:19583–19586. https://doi.org/10.1074/jbc.C300114200.
71. Mu Y, Prochnow C, Pham P, Chen XS, Goodman MF. 2012. A structural basis for the biochemical behavior of activation-induced deoxycytidine deaminase class-switch recombination-defective hyper-IgM-2 mutants. J Biol Chem 287:28007–28016. https://doi.org/10.1074/jbc.M112.370189.
72. Krause K, Marcu KB, Greeve J. 2006. The cytidine deaminases AID and APOBEC-1 exhibit distinct functional properties in a novel yeast selectable system. Mol Immun 43:295–307. https://doi.org/10.1016/j.molimm.2005.05.004.
73. Kohli RM, Abrams SR, Gajula KS, Maul RW, Gearhart PJ, Stivers JT. 2009. A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. J Biol Chem 284:22898–22904. https://doi.org/10.1074/jbc.M109.025536.
74. Carpenter MA, Rajagurubandara E, Wijesinghe P, Bhaqwat AS. 2010. Determinants of sequence-specificity within human AID and APOBEC3G. DNA Repair (Amst) 9:579–587. https://doi.org/10.1016/j.dnarep.2010.02.010.
75. Holtz CM, Sadler HA, Mansky LM. 2013. APOBEC3G cytosine deamination hotspots are defined by both sequence context and single-stranded DNA secondary structure. Nucleic Acids Res 41:6139–6148. https://doi.org/10.1093/nar/gkt246.
76. Rathore A, Carpenter MA, Demir Ö, Ikeda T, Li M, Shaban NM, Law EK, Anokhin D, Brown WL, Amaro RE, Harris RS. 2013. The local dinucleotide preference of APOBEC3G can be altered from 5′-CC to 5′-TC by a single amino acid substitution. J Mol Biol 425:4442–4454. https://doi.org/10.1016/j.jmb.2013.07.040.
77. Rogozin IB, Kolchanov NA. 1992. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim Biophys Acta 1171:11–18.
78. Chelico L, Prochnow C, Erie DA, Chen XS, Goodman MF. 2010. Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. J Biol Chem 285:16195–16205. https://doi.org/10.1074/jbc.M110.107987.
79. Chelico L, Sacho EJ, Erie DA, Goodman MF. 2008. A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV. J Biol Chem 283:13780–13791. https://doi.org/10.1074/jbc.M801004200.
80. Holmes JA, Lin P. 1994. Thermal niche of larval sea lamprey, *Petromyzon marinus*. Can J Fish Aquat Sci 51:253–262. https://doi.org/10.1139/f94-026.
81. Ferreira LC, Afonso AS, Castilho PC, Hazin FHV. 2013. Habitat use of the nurse shark, *Ginglymostoma cirratum*, off Recife, Northeast Brazil: a combined survey with longline and acoustic telemetry. Environ Biol Fishes 96:735–745. https://doi.org/10.1007/s10641-012-0067-5.
82. Dekkers WJ. 1975. Review of the Asiatic freshwater puffers of the genus Tetraodon Linnaeus, 1758 (Pisces, Tetraodontiformes, Tetraodontidae). Bijdrage Dierkunde 45:87–142.
83. Fricke H, Hissman K. 2000. Feeding ecology and evolutionary survival of the living coelacanth *Latimeria chalumnae*. Mar Biol 136:379–386. https://doi.org/10.1007/s002270050697.
84. Jaenicke R, Böhm G. 1998. The stability of proteins in extreme environments. Curr Opin Struct Biol, NY 8:738–748. https://doi.org/10.1016/S0959-440X(98)80094-8.

85. Wintrode P, Arnold F. 2001. Temperature adaption of enzymes: lessons from laboratory evolution. Adv Protein Chem 55:161–225. https://doi.org/10.1016/S0065-3233(01)55004-4.
86. Radestock S, Gohlke H. 2011. Protein rigidity and thermophilic adaptation. Proteins Struct Funct Bioinform 79:1089–1108. https://doi.org/10.1002/prot.22946.
87. Radestock S, Gohlke H. 2008. Exploiting the link between protein rigidity and thermostability for data-driven protein engineering. Eng Life Sci 8:507–522. https://doi.org/10.1002/elsc.200800043.
88. Prochnow C, Bransteitter R, Klein MG, Goodman MF, Chen XS. 2007. The APOBEC-2 crystal structure and functional implications for the deaminase AID. Nature 445:447–451. https://doi.org/10.1038/nature05492.
89. Bas DC, Rogers DM, Jensen JH. 2008. Very fast prediction and rationalization of pKa values for protein-ligand complexes. Proteins 73:765–783. https://doi.org/10.1002/prot.22102.
90. Li H, Robertson AD, Jensen JH. 2005. Very fast empirical prediction and rationalization of protein pKa values. Proteins 61:704–721. https://doi.org/10.1002/prot.20660.
91. Olsson MH, Sondergaard CR, Rostkowski M, Jensen JH. 2011. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput 7:525–537. https://doi.org/10.1021/ct100578z.
92. Grosdidier A, Zoete V, Michielin O. 2011. SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res 39:W270–W277. https://doi.org/10.1093/nar/gkr366.
93. Grosdidier A, Zoete V, Michielin O. 2011. Fast docking using the CHARMM force field with EADock DSS. J Comput Chem 32:2149–2159. https://doi.org/10.1002/jcc.21797.
94. Zoete V, Cuendet MA, Grosdidier A, Michielin O. 2011. SwissParam: a fast force field generation tool for small organic molecules. J Comput Chem 32:2359–2368. https://doi.org/10.1002/jcc.21816.
95. Xie K, Sowden MP, Dance GS, Torelli AT, Smith HC, Wedekind JE. 2004. The structure of a yeast RNA-editing deaminase provides insight into the fold and function of activation-induced deaminase and APOBEC-1. Proc Natl Acad Sci U S A 101:8114–8119. https://doi.org/10.1073/pnas.0400493101.
96. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera: a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20084.