The structure of a yeast RNA-editing deaminase provides insight into the fold and function of activation-induced deaminase and APOBEC-1

Kefang Xie, Mark P. Sowden, Geoffrey S. C. Dance†, Andrew T. Torelli, Harold C. Smith, and Joseph E. Wedekind‡

Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642

Activation-induced deaminase (AID) uses base deamination for class-switch recombination and somatic hypermutation and is related to the mammalian RNA-editing enzyme apolipoprotein B editing catalytic subunit 1 (APOBEC-1). CDD1 is a yeast ortholog of APOBEC-1 that exhibits cytidine deaminase and RNA-editing activity. Here, we present the crystal structure of CDD1 at 2.0-Å resolution and its use in comparative modeling of APOBEC-1 and AID. The models explain dimerization and the need for trans-acting loops that contribute to active site formation. Substrate selectivity appears to be regulated by a central active site ''flap'' whose size and flexibility accommodate large substrates in contrast to deaminases of pyrimidine metabolism that bind only small nucleosides or free bases. Most importantly, the results suggested both AID and APOBEC-1 are equally likely to bind single-stranded DNA or RNA, which has implications for the identification of natural AID targets.

Antibody diversity is generated during B cell development through class-switch recombination (CSR) and somatic hypermutation (SHM) (1), and in many vertebrates by gene conversion (2). Expression of activation-induced deaminase (AID) is critical for all three processes (3–5) and ectopic AID expression was sufficient to activate SHM in B cell lines, hybridomas, and fibroblasts (6–9), gene conversion in B cells (4, 5), as well as CSR in fibroblast cell lines (10). Mutations in the AID protein have been detected in humans with hyper-IgM type 2 syndrome (11). Expression of AID in *Escherichia coli* caused deoxycytidine-to-deoxyuridine deamination in actively transcribed host genes that was enhanced in strains deficient in the deoxyuridine repair enzyme DNA uracil *N*-glycosylase (12). DNA uracil *N*-glycosylase deficiency in mammals resulted in altered patterns of CSR and SHM (13, 14). *In vitro* studies revealed that AID catalyzed deamination on single-stranded DNA at WRCY hot spots, but not on double-stranded DNA, RNA–DNA hybrids, or single-stranded DNA (15, 16). Cytidine deamination on single-stranded DNA within transcription bubbles (17, 18) and the observation that transcription rates influenced SHM activity (17–19) suggested that the biological substrate for AID is DNA.

In contrast, the homology of AID to the mRNA-editing enzyme APOBEC-1 (3), suggested AID might edit RNA (3). This idea was intriguing because edited mRNAs either encode novel proteins or lose the ability to express proteins (20, 21). Consistent with expression of a novel protein from edited mRNA, *de novo* protein synthesis was required for CSR subsequent to AID activation (22).

APOBEC-1 and AID have proven difficult to purify at levels sufficient for structural studies. CDD1 from *Saccharomyces cerevisiae* (*Sc*CDD1) can be purified readily, which is relevant due to its orthology with APOBEC-1 at the level of both cytidine deaminase (CDA) sequence similarity (27%) and mRNA-editing activity on apolipoprotein B (*apoB*) substrates (23). We determined the crystal structure of *Sc*CDD1 to 2.0 Å resolution, which revealed that the fundamental CDA fold is necessary and sufficient for C-to-U deamination in pyrimidine metabolism, as well as RNA editing. Here, we present the CDD1 structure, its functional analysis, and comparative modeling of the AID and APOBEC-1 structures.

Materials and Methods

Protein Expression and Purification. His$_6$-tagged *Sc*CDD1 was expressed in *E. coli* BL21(DE3) CodonPlus (Stratagene) from vector pET-28a (Novagen) by induction with 1 mM isopropyl β-D-thiogalactoside at 30°C for 3 h. Cells were lysed in 50 mM Tris·HCl, pH 8.0/10 mM 2-mercaptoethanol/1 mg·ml^{-1} lysozyme/1 mM PMSF/2 mM benzamidine, and 5 μg·ml^{-1} each of aprotinin, leupeptin, and pepstatin A, and were sonicated and treated with nucleases (33 μg·ml^{-1} each of DNase I and RNaseA, 0.5% Triton X-100, 2 mM ATP, and 10 mM MgSO$_4$). *Sc*CDD1 was purified by using Ni-nitrilotriacetic acid agarose (Qiagen). Protein was eluted and dialyzed against 10 mM Tris·HCl, pH 8.0/0.15 M NaCl/1 mM DTT. The His$_6$ tag was cleaved with thrombin (Novagen). The flowthrough was exchanged with 20 mM Hepes, pH 8.0/0.25 M KCl/5% (vol/vol) glycerol/4 mM DTT, and was then concentrated to 6 mg·ml^{-1}.

Structure Determination. Crystals of native *Sc*CDD1 were grown at 20°C in hanging drops from solutions of 16.5% (wt/vol) polyethylene glycol monomethyl ether 5 K/0.45 M NH$_4$Cl/0.10 M Na-succinate, pH 5.5/10 mM DTT/1 mM NaN$_3$. Crystals were harvested after 3–4 weeks, were cryoprotected by adding 17.5% (vol/vol) polyethylene glycol monomethyl ether 550 to the mother liquor, and were flash-frozen. Diffraction data were collected at the Advanced Photon Source, Argonne National Laboratory, Argonne, IL. The native *Sc*CDD1 structure was solved by multiwavelength anomalous diffraction phasing (SBC BM19) using endogenous zinc as a phasing source (Table 1, which is published as supporting information on the PNAS web site). Data were reduced with HKL2000 (24) and four zinc atoms were located by use of SOLVE (25). Initial phases were density modified by using RESOLVE with 4-fold noncrystallographic symmetry averaging. These phases improved electron density maps and allowed skeletonization by use of O (26). The initial tetrameric asymmetric unit was built into maps by using phases derived from averaging with DM. Simulated annealing, positional, and individual *B* factor refinement were conducted in CNS (27). Electron density was continuous for regions A2–A142, B1–B139, C2–C141, and D3–D136. X-ray data collection and refinement statistics are in Table 1. Representative electron density for the refined structure is shown in Fig. 1*A*.
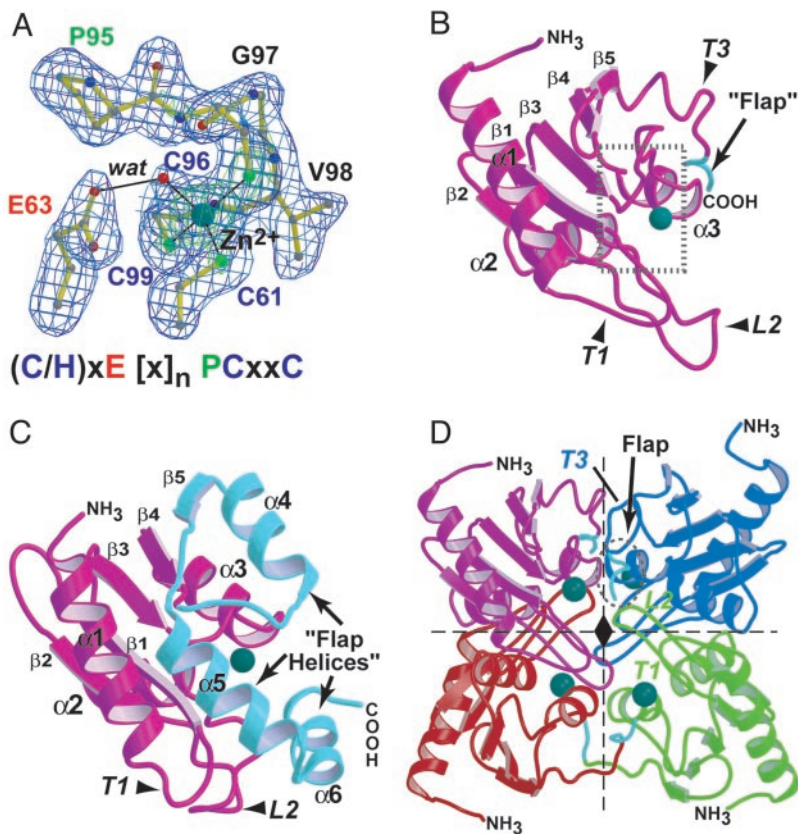
Fig. 1. Electron density map and ribbon diagrams of *Sc*CDD1 and *Sc*CD. (*A*) Portion of a 2.0-Å resolution simulated annealing ($F_o - F_c$) omit map of the CDD1 active site contoured at 3.5 (blue) and 8.5 σ (green). Ball-and-stick models are depicted for amino acids of the zinc-dependent deaminase signature motif (labels below). Black lines indicate ionic interactions to Zn^{2+} (dark green sphere) or hydrogen bonds to catalytic water (wat). (*B*) The CDD1 monomer depicting structural elements of the conserved catalytic β-triangle fold. The C-terminal flap and major secondary structure elements are labeled. (*Inset*) The position of signature amino acids from *A*. (*C*) Ribbon diagram of the *Sc*CD monomer in the same orientation as *B*. A helical flap is indicated at the C terminus (cyan). (*D*) Ribbon representation of the *Sc*CDD1 tetramer containing four active sites each with Zn^{2+}. Each colored monomer is a distinct polypeptide chain. Dashed lines and a diamond indicate twofold axes. The transoligomeric contributions that form the active site of the purple subunit are labeled: T1, L2 (green), flap (cyan with hatched circle), and T3 (blue). Figs. 1, 2 and 4 were drawn with BOBSCRIPT (54).

Modeling. An alignment template was constructed by use of a quadruple C^α superposition of CDAs including: *Sc*CDD1 (this study), *E. coli* (*Ec*)CDA (28), *Bacillus subtilis* (*Bs*) CDA (29), and the *Sc*CD monomer (30). Pairwise matches of tetrameric *Sc*CDD1 atoms with the tetrameric enzyme of *B. subtilis* (29) and the dimeric enzyme from *E. coli* (28) produced rms deviation values of 1.14 and 1.54 Å for 88% and 57% of spatially matched atoms within the CDA core. Differences in the subunit interface of the *Sc*CD dimer precluded superpositions with known CDAs. Equivalent C^α positions of the *Sc*CD β-triangle motif were superimposable at the monomeric level resulting in an rms deviation of 1.46 Å for 25% of matched positions. Amino acid alignments between *Homo sapiens* (*Hs*)APOBEC-1 (GenBank accession no. AAD00185) and *Hs*AID (GenBank accession no. NP_065712) with the four known structures were prepared before modeling. During calculation of the substrate position at the site of deamination, C or U nucleotides were restrained by distances derived from the coordinates of known CDA crystal structures bound to nucleoside substrate or analogs thereof (28, 29, 31). Details of template assembly and utilization were as described in the MODELLER manual. Other special methods are described in *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site.

Mutagenesis and Construction of Dimeric CDD1 Enzymes. *Sc*CDD1 plus 19 (Fig. 3*A*, construct 1) was PCR-amplified by using a 5′ CDD1-specific primer and a 3′ primer encoding the 19-aa "linker" of *Ec*CDA, and subcloned into pET28a. Fusion proteins of CDD1 were assembled as follows: the 5′ monomers containing the appropriate C-terminal *E. coli* linker or APOBEC-1 flap (Fig. 3*A*, construct 2 or 3) were PCR-amplified and subcloned as *Nde*I/ *Eco*RI fragments into pET28a. The N-terminally foreshortened 3′ monomer (Fig. 3*A*, construct 3) was PCR-amplified from an E63A mutant template of CDD1 that was ligated as an *Eco*RI/*Xho*I fragment to a 5′ monomer. The linker *Eco*RI site was mutagenized

(QuikChange, Stratagene) to restore the correct amino acid sequence. All CDD1 monomer and dimer cDNAs were amplified by using CDD1-specific primers and subcloned through *Eco*RI and *Xba*I sites into a modified pYES2.0 vector (23). *Ec*CDA was PCR-amplified from genomic DNA and subcloned into the modified pYES2.0 as described (23). The amino acid sequences of the flap (APOBEC-1) and linker (*E. coli*) are provided (Fig. 5, which is published as supporting information on the PNAS web site).

Editing Assay. Plasmids encoding CDD1 and reporter *apoB* RNA were cotransformed into yeast strain CL51 and CDD1 proteins expressed from individual transformants (23). Total yeast RNA was isolated as described (23), and editing activity was measured on *apoB* RNA by RT-PCR and poisoned primer extension (23, 32). Total protein was isolated from an equivalent number of cells and resolved by SDS/10.5% PAGE, transferred to nitrocellulose, and reacted with a hemagglutinin antibody (Babco, Richmond, CA; ref. 23).

CDA Assay. Deaminase activity on cytidine was measured by spectrophotometric assays (23). Specific activity for each purified recombinant protein was determined from at least three measurements, and were defined as nanomoles of product $min^{-1} \cdot mg^{-1}$ enzyme on cytidine (33).

Results and Discussion

The CDA Fold. The tertiary fold of the CDD1 monomer (Fig. 1*B*) exhibited high structural homology to the catalytic domains of known bacterial CDAs from *E. coli* and *B. subtilis* (29), as well as *Sc*CD (30), despite modest sequence identity; respective rms deviation values from superpositions were 1.32 Å (28% identity), 0.95 Å (43%), and 1.42 Å (16%). The tertiary fold of the deaminase catalytic domain was a triangular β-sheet comprising five core strands flanked on either broad face by three α-helices (Fig. 1*B*). A

conserved feature of this fold is the presence of the amino acid sequence (C/H)xEx$_n$PCxxC, identified as an essential signature motif that contained a Glu proton shuttle and Zn^{2+} ligands His or Cys (Fig. 1A). The subdomain structure $\alpha2\beta3\alpha3$ (Fig. 1B) and signature sequence have been dubbed a zinc-dependent deaminase motif characteristic of C-to-U deaminases of pyrimidine metabolism, as well as C-to-U and A-to-I RNA-editing enzymes (21, 34). Key structural differences exist between CDA and ScCD structures that affect substrate specificity. The ScCD structure exhibited a bulky polypeptide flap that folded over the active site entrance (Fig. 1C), excluding substrates larger than a single-nucleotide base. In contrast, the flap of CDD1 was a much shorter coil (Fig. 1B), which was consistent with its activity on cytidine or larger substrates (23). Selective pressure to maintain the overall β-triangle fold in the context of an active-site flap was suggested by the observation that strand $\beta5$ of ScCD was preserved, albeit in an orientation opposite to that of ScCDD1 (Fig. 1 B and C) whose fold typifies other CDA enzymes.

There has been a controversy regarding the structures of APOBEC-1 and AID, and it has been suggested that ScCD is a more appropriate modeling template (35) than the CDA enzymes (36). Our analysis indicated that a C-terminal flap like that of ScCD would preclude ribose binding by APOBEC-1 or AID, and that the ScCD sequence is too short to allow the assembly of a full-length template for modeling any APOBEC-1-related protein. A prerequisite template should, at a minimum, consist of structural features consistent with substrate binding because distinct modes of oligomerization by ScCD versus CDA enzymes impact active-site formation.

CDA Active-Site Formation. A comparison of the dimeric *E. coli* and tetrameric *B. subtilis* CDA structures revealed that their subunit interfaces are conserved, despite differences in quaternary structure (29). The CDD1 subunit interface adopted a similar structure whereby each tetramer could be generated from a single monomer (Fig. 1B) through application of three mutually perpendicular 2-fold axes (i.e., D_2 symmetry, Fig. 1D). The fact that the tetrameric CDD1 structure (Fig. 1D) could be superimposed on the dimeric EcCDA molecule (rms deviation of 1.54 Å) indicated homology within the N-terminal catalytic domain (NTCD; EcCDA residues 49–171, equivalent to Fig. 1B), the noncatalytic C-terminal domain (NCCTD; EcCDA residues 189–294), and the spatial arrangement between domains. Conservation of the subunit interfaces between tetramers and dimers implied that the latter arose through gene duplication of an ancestral monomer that adopted a tetrameric quaternary fold (21, 29). This observation accounts for the maintenance of the fundamental β-triangle fold in both N- and C-terminal domains of dimeric EcCDA (28). Hence, although the NCCTD of EcCDA exhibits a modest 11% sequence identity with its NTCD, the folds of these domains are nearly identical (rms deviation of 1.40 Å). This information provides a significant constraint in comparative modeling of any comparable NCCTD.

Preservation of both the CDA tertiary fold and the subunit interface can be attributed to the need for multiple polypeptides in formation of a single ribose-binding site (28). This requirement can be explained in the context of CDD1, wherein the NTCD (Fig. 1D, purple subunit) requires loop contributions T1 and L2 from the symmetry-related subunit (Fig. 1D, green), as well as T3 and the active-site flap of the blue subunit (Fig. 1D). In contrast, the C_2 subunit interface of dimeric Sc cytosine deaminase (CD) does not appear critical for enzymatic activity because each monomer possesses a self-contained active site (30). The simplicity of the base substrate may have allowed divergence in the ScCD domain that led to the acquisition of a large C-terminal flap (Fig. 1C). Such observations implied that D_2 oligomerization should be an essential aspect of the active sites of both APOBEC-1 and AID, whose polymeric substrates are more complex.
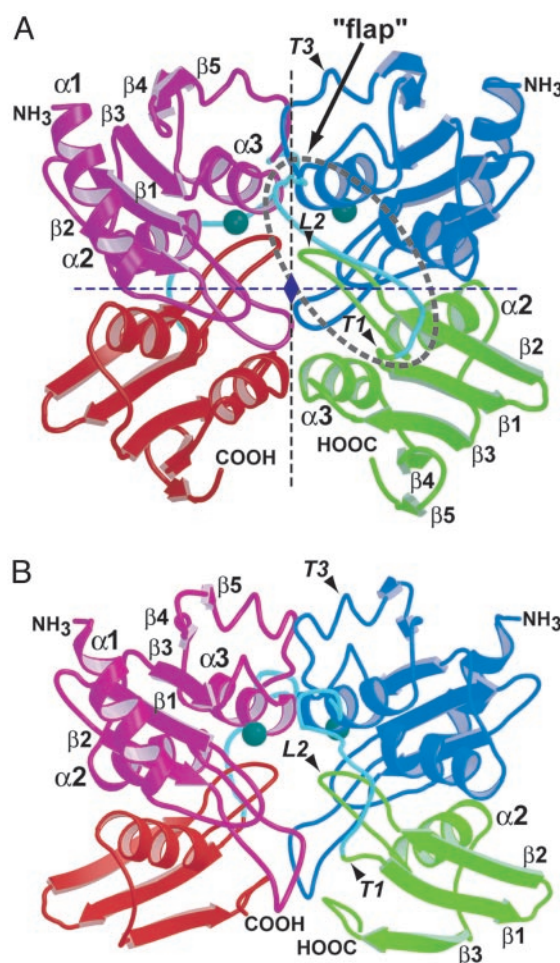

Fig. 2. Ribbon representations of *Hs*APOBEC-1 and *Hs*AID comparative models. (*A*) The dimeric APOBEC-1 model with polypeptide chains colored as follows: purple and red (NTCD and NCCTD) and blue and green (NTCD and NCCTD). A central flap (cyan with hatched oval) connects the NTCD to the NCCTD. Each NTCD coordinates Zn^{2+} (dark green sphere). Trans-acting structure elements that form the purple active site are: T1 and L2 (green NCCTD), flap (cyan with hatched circle), and T3 (blue NTCD). Symmetry axes are as described in Fig. 1D with the exception that blue axes represent improper (pseudo) two-fold rotations. (*B*) The dimeric AID structure. The subunit interface of AID obeys D_2 symmetry analogous to APOBEC-1, but axes were omitted for clarity.

Models of AID and APOBEC-1. To implement a modeling method relying on spatial restraints, a structural template was assembled for alignment of APOBEC-1 and AID sequences. The alignment (Fig. 5) resulted in sequence homologies of 51% and 41% in the NTCD, and 40% and 37% in the NCCTD for APOBEC-1 and AID, respectively. The latter approach was necessary because in cases where the degree of sequence identity is low (37), use of multiple structures improves the reliability of predictions (38). This method is further bolstered by the observation that active sites of distantly related enzymes of common function often exhibit homologous structures despite low sequence identity (39). Hence, our approach differs from prior efforts (21) to model APOBEC-1 and AID because it did not rely primarily on sequence comparisons, but instead made use of the highly conserved deaminase fold, and the knowledge that ScCDD1 acts in both pyrimidine metabolism and RNA editing.

The resulting comparative models of APOBEC-1 and AID featured bilobal domain organization on a single polypeptide chain (Fig. 2). The NTCD resembled the conserved catalytic domain of

*Sc*CDD1 (Fig. 1*B* versus Fig. 2, purple or blue), whereas the NCCTD exhibited a reduced topology in which structures lacked helix α1. The NCCTD of APOBEC-1 preserved the core five-stranded β-sheet (ββαβαββ) topology (Fig. 2*A*, red or green), but AID lacked the final αββ segment (Fig. 2*B*). Our models predicted that each dimer had two active sites on opposite faces of a head-to-head oligomer that functioned independently and that each could bind two nucleoside substrates or deaminate two sites within the same or different nucleic acid strands. These results differ from a previous model of APOBEC-1 derived from the *Ec*CDA structure in which topologies of both N- and C-terminal domains were altered to accommodate a double-stranded RNA-binding cleft that required cooperatively interacting active sites (40); our models do not exhibit double-stranded RNA- or double-stranded DNA-binding clefts.

The knowledge of several deaminase structures since the original *Ec*CDA-based model suggested that the cleft in the early APOBEC-1 study arose from an incorrect sequence alignment of the 229-aa APOBEC-1 sequence onto the 294-aa *Ec*CDA structure. Misalignment was likely because the N-terminal 48-aa helix bundle from *Ec*CDA does not appear in the structures of either *Sc*CDA, *B. subtilis* CDA, or *Sc*CD (29, 30), and therefore should not be considered part of the core deaminase fold. The need for new APOBEC-1-related protein models is underscored by the fact that the *Ec*CDA-based model has been adopted by other laboratories attempting to elucidate the structure and function of APOBEC-1-related protein family members (36, 41), but has been criticized subsequently for lack of structural rigor (21, 35).

Experimental evidence demonstrated the oligomeric states of AID and APOBEC-1 are dimeric (15, 36, 42), which was proven essential for APOBEC-1 RNA-editing activity (43). Inspection of the subunit interfaces of our models revealed that the sequences of APOBEC-1 and AID could be accommodated readily by a dimer with pseudo D_2 symmetry (Fig. 2). This symmetric domain arrangement of *Sc*CDD1 and *Ec*CDA assured that the NTCD (e.g., Fig. 2, purple subunit) was positioned to receive essential trans-acting loops from the NCCTD of the dyad-related molecules including T1 and L2 (Fig. 2, green subunit), T3 (Fig. 2, blue subunit), and the interdomain flap (Fig. 2, cyan coil between blue and green subunits). Although it has been suggested that APOBEC-1 could dimerize in a head-to-tail manner (44), our results indicated that ribose binding contributions from the respective NTCD and NCCTD used asymmetric interactions (i.e., a pseudo-D_2 subunit interface), as observed for *Ec*CDA (28). As such, the L4 loop of the NCCTD, predicted by our models to bind substrate in the APOBEC-1 and AID models, is significantly longer than that of the NTCD (Fig. 5). Hence, a head-to-tail dimer precludes positioning of the appropriately sized L4 loop. Furthermore, subunits in a head-to-tail orientation would not maximize the amount of buried surface area in the subunit interface because large and small domains would be paired; this result would be energetically unfavorable. Overall, the need for conserved subunit symmetry can be explained at the molecular level as the requirement to maintain an extensive interface (see Fig. 6, which is published as supporting information on the PNAS web site) that provides essential functional groups for substrate binding.

Corroboration of the APOBEC-1 Model. Details of the APOBEC-1 dimer, including the existence of a short interdomain flap, are key aspects of our work that differentiate the proposed models from prior studies (45). To evaluate features of the APOBEC-1 model in the context of enzymatic function a series of *Sc*CDD1 mutants were constructed. All native and mutant enzymes were assayed for deaminase activity on cytidine, as well as RNA-editing activity on reporter *apoB* mRNA (Fig. 3*A* and *B*). For editing assays, wild-type and mutant enzymes were coexpressed in yeast with a reporter *apoB* RNA (ref. 23 and Fig. 3C). *Sc*CDD1 plus 19 (Fig. 3*A*) addressed the importance of flap length by analogy to the *Ec*CDA
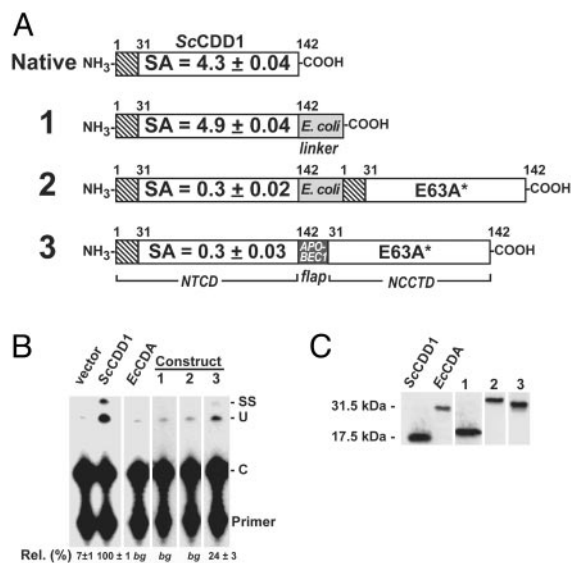

Fig. 3. Representative CDA and RNA-editing assays for *Sc*CDD1 and its mutants. (*A*) Schematic depictions of native CDD1 (open box) and chimeric constructs used in assays. Construct 1 (*Sc*CDD1 plus 19) is native *Sc*CDD1 with an additional 19 amino acids from *Ec*CDA (gray box). Construct 2 (*Sc*CDD1-*Ec*L19-CDD1*) has two *Sc*CDD1 domains joined by a central linker derived from *Ec*CDA. CDD1*, inactive deaminase domain E63A mutant. Construct 3 (*Sc*CDD1-*A*L-Δ(α1)-CDD1*) represents two *Sc*CDD1 domains joined by the central eight amino acid flap (black box) of APOBEC-1. Helix α1 (hatched box) of CDD1* was removed. Specific activities (SA) are shown in the NTCD. (*B*) Poisoned primer extension assays on reporter *apoB* RNA coexpressed in yeast with various CDA enzymes. U, Editing at C6666 of reporter *apoB* RNA; SS, "second stop" promiscuous editing of C6666, as observed for APOBEC-1 (32). Average C6666 editing by native *Sc*CDD1 was 7.8 + 0.8% (*n* = 4). Average editing values were scaled to native. bg, No significant editing above vector levels. (*C*) Western blots of hemagglutinin-tagged CDA proteins expressed in yeast.

and *Sc*CD enzymes (Fig. 1*C*), as well as hyper-IgM type 2 syndrome insertion mutants at the C terminus of AID (Fig. 5).

Constructs *Sc*CDD1-*Ec*L19-CDD1* and *Sc*CDD1-*A*L-Δ(α1)-CDD1* converted tetrameric *Sc*CDD1 into a dimer. In both constructs, the NCCTD harbored an E63A mutation, denoted CDD1*, which assured that only the NTCD was tested for function as reflected by the APOBEC-1 and AID models. *Sc*CDD1-*Ec*L19-CDD1* tested whether or not a long linker acting as a rigid flap, as modeled previously (45), was suited for RNA editing. Construct *Sc*CDD1-*A*L-Δ(α1T1)-CDD1* tested the feasibility of folding and activity for an *Sc*CDD1 dimer that lacked helix α1 of the NCCTD while using the putative central flap of APOBEC-1 (Fig. 2*A*).

CDA and RNA-editing assays were evaluated relative to native *Sc*CDD1 (23). Recombinant *Sc*CDD1 plus 19 (Fig. 3*A*) deaminated free cytidine at levels comparable with native (Fig. 3*A*), specific activity 4.9 ± 0.04 versus 4.3 ± 0.04 nmol·min^{-1}·mg^{-1}, but was incapable of editing reporter *apoB* RNA in yeast (Fig. 3*B*). These data supported a model for CDA activity in which the active-site flap restricted substrate access to the active site as observed for *Sc*CD (30). The observation that both the *Sc*CDD1-*Ec*L19-CDD1* and *Sc*CDD1-*A*L-Δ(α1)-CDD1* constructs deaminated cytidine *in vitro* demonstrated that the purified recombinant enzymes exhibited both folding and catalytic capabilities.

Deaminase activity on cytidine was 7-fold-diminished on a "per-active-site" basis for the dimeric constructs compared to native *Sc*CDD1. These results may be explained by the nonoptimal combination of amino acids at the active site resulting from mutagenesis. However, the level of activity on cytidine was significant for both dimers because neither *Ec*CDA nor *Sc*CDD1-*Ec*L19-CDD1*, a structural mimic of the previously proposed APOBEC-1
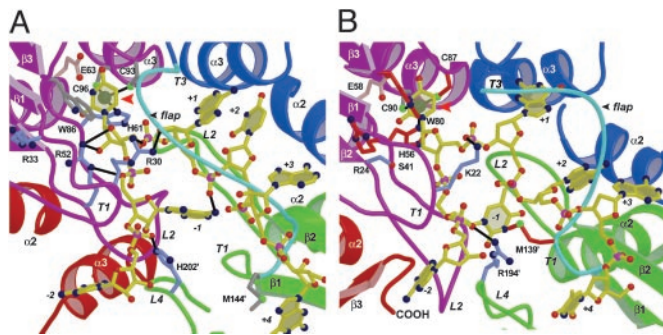
Fig. 4. Schematic diagrams of the *Hs*APOBEC-1 and *Hs*AID active-site models with bound substrates (ball-and-stick models, yellow). (*A*) The APOBEC-1 active site with bound *apoB* mRNA substrate 5′-GAUAU$_{6666}$AA-3′. (*B*) The AID active site with bound DNA substrate 5′-d(TAAGU*TA)-3′, described as an SHM hot spot (16). Residues mutated in hyper-IgM type 2 syndrome are red. In both diagrams, views represent expanded orientations of Fig. 2 *A* and *B*. Basic residues are light blue, acidic residues pale pink, and hydrophobic residues are gray. Predicted hydrogen bonds and ionic interactions with Zn^{2+} are depicted as black lines. The sites of base deamination are indicated by red arrowheads. Not all amino acids are shown.

linker structure (45), was competent to edit reporter *apoB* RNA (Fig. 3*B*). Strikingly, the *Sc*CDD1-*A*L-Δ(α1)-CDD1* construct, derived by analogy to our APOBEC-1 model (Fig. 2*A*), edited reporter RNA (Fig. 3*B*) at a level nearly half that of native *Sc*CDD1 (on a per-active-site basis). These results suggested that the modeled topology of APOBEC-1, and by analogy, AID, are structurally reasonable and that an active-site flap of the form described here makes important contributions in the selection of appropriately sized substrates. This hypothesis was supported by the structural analysis and comparisons of substrate specificity for *Sc*CD (30), *Ec*CDA, and *Sc*CDD1. A more detailed description of previous APOBEC-1 mutants mapped onto the APOBEC-1 model will be presented elsewhere.

Substrate Binding and the Active-Site Flap. Knowledge of binding to nucleoside substrates and analogs from the crystal structures of bacterial CDAs (28, 29) made it possible to model APOBEC-1 in the presence of the substrate *apoB* mRNA sequence 5′-GAUAU$_{6666}$AA-3′ (46) (Fig. 4*A*). Similarly, APOBEC-1 was modeled in the presence of the DNA sequence 5′-d(ATCTC*CG)-3′ (Fig. 7*B*, which is published as supporting information on the PNAS web site), which was a mutation hot spot in the *rpoB* gene when APOBEC-1 was overexpressed in *E. coli* (47). AID was modeled in the presence of a DNA SHM hotspot sequence 5′-d(TAAGU*TA)-3′ (Fig. 4*B*) that was targeted by the enzyme *in vitro* (16).

Modeling revealed many parallels between the mode of APOBEC-1 and AID substrate binding. In both structures a 7-mer was chosen because this was the minimal oligomer length necessary to pass from the protein surface into the active site at both 5′ and 3′ ends. Significantly, both models accommodated either RNA or DNA, but only in single-stranded form due to steric considerations (and the assumption that the overall CDA topology would be maintained). These results provided better agreement with the observation that affinity for substrate by APOBEC-1 is both modest ($K_d \approx 450 \mu M$; ref. 48) and nonspecific (40). Site-specific APOBEC-1-mediated *apoB* mRNA editing requires the protein APOBEC-1 complementation factor (ACF) whose three RNA-recognition motifs (49) bind RNA in a distinctly single-stranded manner (50); binding by ACF at a 3′ (mooring) sequence four nucleotides away from the edited C6666 has profound implications for the mode of RNA presentation to the APOBEC-1 active site. At present, it is unclear whether AID activity on either single-stranded DNA or RNA occurs within B cells in the context of complemen-

tation factors that regulate substrate specificity (36) by analogy to ACF (49).

Although neither APOBEC-1 nor AID sequences were aligned to each other during model construction, the amino acids that interacted with their respective substrates appeared homologous. Both APOBEC-1 and AID (in parentheses) exhibited basic residues such as R30 (K22) and H202 (R194) interacting with the phosphodiester backbone (Fig. 4). Conserved W86 (W80) was in van der Waals contact with the modified base (Fig. 4). Although ribose binding at the edited C of APOBEC-1 used the amphiphilic side chain of R52, which interacts with the 2′-OH group of RNA, and O4′ of RNA and DNA (Figs. 4*A* and 7*B*), the comparable position in AID was W20 (data not shown). However, AID possessed conserved S41 poised to hydrogen bond with a 2′-OH of ribose if it were present (Fig. 4*B*).

Based on the crystal structure of CDD1 and modeling of APOBEC-1 and AID, flap flexibility appeared to play an important role in substrate binding. Notably, all three enzymes exhibit a conserved Gly residue at the start of the flap sequence. Whereas CDD1 exhibits its C terminus, APOBEC-1 and AID maintain a second Gly residue equivalent to G138 of APOBEC-1 (see Fig. 5). Therefore, by analogy to the CDAs and *Sc*CD, the models predict that substrate selectivity based on size is achieved through specialized interdomain flaps that fold over each active site in a trans-manner, explaining the need for homodimerization in catalysis. Our functional data corroborated this hypothesis indicating a short flap hinged by two Gly residues (derived from APOBEC-1) was suited for RNA-editing activity, whereas longer sequences devoid of Gly, such as that of *Ec*CDA, were not. These results cannot be reconciled with a previous model of APOBEC-1 that called for a long *Ec*CDA-like interdomain linker (45).

Hyper-IgM Type 2 Mutants and the AID Model. To corroborate the model, hyper-IgM type 2 syndrome mutations (36) were mapped onto the predicted AID structure. Based on their locations, mutations could be assigned to one of three categories. The first set of mutations included H56Y and C87R, which mapped to the active site. The effects of these mutants could be predicted *a priori* because either substitution would impair Zn^{2+} coordination and nucleotide binding at the deaminated base (Fig. 4*B*). The second class of mutations includes W80R, R24W, M139, and the truncation mutation 190X. The effects of these mutations could not be predicted, although these also resided in the active site (Fig. 4*B*). The model predicted that the changes should adversely affect catalysis by interference with nucleotide binding (W80R), active-site folding (R24W), or substrate interactions (M139). Several frameshift and truncation mutations, including 190X are predicted to disrupt the C terminus at the L4 loop, which contacted the phosphate backbone of the substrate by means of R194 in our model (Fig. 4*B*). A discussion of the third mutation class, which maps to the surface of AID, is presented in *Supporting Materials and Methods*.

In summary, the structure of *Sc*CDD1 has been solved providing a tenable connection between known CDA and *Sc*CD enzymes, and the family of mammalian APOBEC-related proteins (21). Four structures were used in comparative modeling of *Hs*APOBEC-1 and *Hs*AID, and the preparation of a structural template revealed major differences exist in the modes of substrate binding by CDA and *Sc*CD enzymes. Comparative modeling of APOBEC-1-related proteins suggested that the domain interface between subunits influences the positioning of ribose binding loops as in CDAs and that a variably sized flap at the catalytic site regulates access based on substrate size, which was corroborated by our functional data. The recent discovery of the requirement for AID in CSR, gene conversion, and SHM (3, 5, 11), as well as APOBEC-3G (CEM15) in the suppression of HIV-1 infectivity (41), raised questions of how their substrates are targeted. Although much information is known for APOBEC-1, making it a model system, different paradigms have been proposed for AID in which either DNA or RNA

represents the primary substrate (22, 47, 51, 52). Both AID and APOBEC-1 targeted DNA in *E. coli* (12, 47). However, APOBEC-1 did not substitute for AID in CSR or SHM (52), and AID could not substitute for APOBEC-1 in *apoB* mRNA editing (51), suggesting each enzyme has an inherent specificity for its own substrate. The models of APOBEC-1 and AID presented here indicated either DNA or RNA substrates could be accommodated by their active sites, but only in single-stranded form. Given these observations it is important for investigators to consider both DNA-deamination and RNA-editing mechanisms as possible outcomes of AID biological activity.

Note. While this manuscript was under review, Ito *et al.* (53) demonstrated functional conservation between the NCCTD of AID and APOBEC-1, which were shown to serve as nuclear export signals. The models presented in this report predict homologous NCCTD structures and are consistent with the possibility of similar protein–protein interactions and function in this region.

1. Manis, J. P., Tian, M. & Alt, F. W. (2002) *Trends Immunol.* **23,** 31–39.
2. Flajnik, M. F. (2002) *Nat. Rev. Immunol.* **2,** 688–698.
3. Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y. & Honjo, T. (2000) *Cell* **102,** 553–563.
4. Harris, R. S., Sale, J. E., Petersen-Mahrt, S. K. & Neuberger, M. S. (2002) *Curr. Biol.* **12,** 435–438.
5. Arakawa, H., Hauschild, J. & Buerstedde, J. M. (2002) *Science* **295,** 1301–1306.
6. Rada, C., Jarvis, J. M. & Milstein, C. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 7003–7008.
7. Martin, A., Bardwell, P. D., Woo, C. J., Fan, M., Shulman, M. J. & Scharff, M. D. (2002) *Nature* **415,** 802–806.
8. Yoshikawa, K., Okazaki, I. M., Eto, T., Kinoshita, K., Muramatsu, M., Nagaoka, H. & Honjo, T. (2002) *Science* **296,** 2033–2036.
9. Faili, A., Aoufouchi, S., Gueranger, Q., Zober, C., Leon, A., Bertocci, B., Weill, J. C. & Reynaud, C. A. (2002) *Nat. Immunol.* **3,** 815–821.
10. Okazaki, I. M., Kinoshita, K., Muramatsu, M., Yoshikawa, K. & Honjo, T. (2002) *Nature* **416,** 340–345.
11. Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., *et al.* (2000) *Cell* **102,** 565–575.
12. Petersen-Mahrt, S. K., Harris, R. S. & Neuberger, M. S. (2002) *Nature* **418,** 99–103.
13. Imai, K., Slupphaug, G., Lee, W. I., Revy, P., Nonoyama, S., Catalan, N., Yel, L., Forveille, M., Kavli, B., Krokan, H. E., *et al.* (2003) *Nat. Immunol.* **4,** 1023–1028.
14. Rada, C., Williams, G. T., Nilsen, H., Barnes, D. E., Lindahl, T. & Neuberger, M. S. (2002) *Curr. Biol.* **12,** 1748–1755.
15. Dickerson, S. K., Market, E., Besmer, E. & Papavasiliou, F. N. (2003) *J. Exp. Med.* **197,** 1291–1296.
16. Bransteitter, R., Pham, P., Scharff, M. D. & Goodman, M. F. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 4102–4107.
17. Martin, A. & Scharff, M. D. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 12304–12308.
18. Ramiro, A. R., Stavropoulos, P., Jankovic, M. & Nussenzweig, M. C. (2003) *Nat. Immunol.* **4,** 452–456.
19. Sohail, A., Klapacz, J., Samaranayake, M., Ullah, A. & Bhagwat, A. S. (2003) *Nucleic Acids Res.* **31,** 2990–2994.
20. Bass, B. L. (2002) *Annu. Rev. Biochem.* **71,** 817–846.
21. Wedekind, J. E., Dance, G. S., Sowden, M. P. & Smith, H. C. (2003) *Trends Genet.* **19,** 207–216.
22. Doi, T., Kinoshita, K., Ikegawa, M., Muramatsu, M. & Honjo, T. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 2634–2638.
23. Dance, G. S., Beemiller, P., Yang, Y., Mater, D. V., Mian, I. S. & Smith, H. C. (2001) *Nuleic Acids Res.* **29,** 1772–1780.
24. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326.
25. Terwilliger, T. C. (1997) *Methods Enzymol.* **276,** 530–537.
26. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard. (1991) *Acta Crystallogr. A* **47,** 110–119.
27. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.
28. Betts, L., Xiang, S., Short, S. A., Wolfenden, R. & Carter, C. W., Jr. (1994) *J. Mol. Biol.* **235,** 635–656.
29. Johansson, E., Mejlhede, N., Neuhard, J. & Larsen, S. (2002) *Biochemistry* **41,** 2563–2570.
30. Ko, T. P., Lin, J. J., Hu, C. Y., Hsu, Y. H., Wang, A. H. & Liaw, S. H. (2003) *J. Biol. Chem.* **278,** 19111–19117.
31. Xiang, S., Short, S. A., Wolfenden, R. & Carter, C. W., Jr. (1996) *Biochemistry* **35,** 1335–1341.
32. Sowden, M., Hamm, J. K. & Smith, H. C. (1996) *J. Biol. Chem.* **271,** 3011–3017.
33. Kurtz, J. E., Exinger, F., Erbs, P. & Jund, R. (1999) *Curr. Genet.* **36,** 130–136.
34. Mian, I. S., Moser, M. J., Holley, W. R. & Chatterjee, A. (1998) *J. Comput. Biol.* **5,** 57–72.
35. Zaim, J. & Kierzek, A. M. (2003) *Nat. Immunol.* **4,** 1153–1154.
36. Ta, V. T., Nagaoka, H., Catalan, N., Durandy, A., Fischer, A., Imai, K., Nonoyama, S., Tashiro, J., Ikegawa, M., Ito, S., *et al.* (2003) *Nat. Immunol.* **4,** 843–848.
37. Rost, B. (1999) *Protein Eng.* **12,** 85–94.
38. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. & Sali, A. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29,** 291–325.
39. Chothia, C. & Lesk, A. M. (1986) *EMBO J.* **5,** 823–826.
40. Navaratnam, N., Bhattacharya, S., Fujino, T., Patel, D., Jarmuz, A. L. & Scott, J. (1995) *Cell* **81,** 187–195.
41. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. (2002) *Nature* **418,** 646–650.
42. Lau, P. P., Zhu, H. J., Baldini, A., Charnsangavej, C. & Chan, L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 8522–8526.
43. Oka, K., Kobayashi, K., Sullivan, M., Martinez, J., Teng, B. B., Ishimura-Oka, K. & Chan, L. (1997) *J. Biol. Chem.* **272,** 1456–1460.
44. Gerber, A. P. & Keller, W. (2001) *Trends Biochem. Sci.* **26,** 376–384.
45. Navaratnam, N., Fujino, T., Bayliss, J., Jarmuz, A., How, A., Richardson, N., Somasekaram, A., Bhattacharya, S., Carter, C. & Scott, J. (1998) *J. Mol. Biol.* **275,** 695–714.
46. Smith, H. C. (1993) in *Seminars in Cell Biology*, ed. Stuart, K. (Saunders Scientific, London), Vol. 4, pp. 267–278.
47. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. (2002) *Mol. Cell* **10,** 1247–1253.
48. Anant, S. & Davidson, N. O. (2000) *Mol. Cell. Biol.* **20,** 1982–1992.
49. Mehta, A., Kinter, M. T., Sherman, N. E. & Driscoll, D. M. (2000) *Mol. Cell. Biol.* **20,** 1846–1854.
50. Mehta, A. & Driscoll, D. M. (2002) *RNA* **8,** 69–82.
51. Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O. & Honjo, T. (1999) *J. Biol. Chem.* **274,** 18470–18476.
52. Eto, T., Kinoshita, K., Yoshikawa, K., Muramatsu, M. & Honjo, T. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 12895–12898.
53. Ito, S., Nagaoka, H., Shinkura, R., Begum, M., Muramatsu, M., Nakata, M. & Honjo, T. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 1975–1908.
54. Esnouf, R. M. (1999) *Acta Crystallogr. D* **55,** 938–940.

IMMUNOLOGY