The Journal of Immunology

A Founder Mutation in Artemis, an SNM1-Like Protein, Causes SCID in Athabascan-Speaking Native Americans

Lanying Li, Despina Moshous, Yungui Zhou, Junhua Wang, Gang Xie, Eduardo Salido, Diana Hu, Jean-Pierre de Villartay and Morton J. Cowan

J. Immunol. 2002;168;6323-6329
http://www.jimmunol.org/cgi/content/full/168/12/6323

The Journal of Immunology is published twice each month by The American Association of Immunologists, Inc., 9650 Rockville Pike, Bethesda, MD 20814-3994.

Print ISSN: 0022-1767 Online ISSN: 1550-6606.

A Founder Mutation in Artemis, an SNM1-Like Protein, Causes SCID in Athabascan-Speaking Native Americans[1]

Lanying Li,[2]* Despina Moshous,[2†] Yungui Zhou,* Junhua Wang,* Gang Xie,* Eduardo Salido,* Diana Hu,[‡] Jean-Pierre de Villartay,[†] and Morton J. Cowan[3]*

Athabascan SCID (SCIDA) is an autosomal recessive disorder found among Athabascan-speaking Native Americans and is manifested by the absence of both T and B cells ($T^-B^-NK^+$ SCID). We previously mapped the SCIDA gene to a 6.5-cM interval on chromosome 10p. SCIDA fibroblasts were found to have defective coding joint and reduced, but precise signal joint formation during V(D)J recombination. After excluding potential candidate genes, we conducted a combined positional candidate and positional cloning approach leading to the identification of nine novel transcripts in the refined SCIDA region. One of the transcripts showed significant homology with the mouse and yeast SNM1/PSO_2 and was recently reported (Artemis) to be responsible for another $T^-B^-NK^+$ SCID condition (radiation sensitive SCID) in 13 patients of primarily European origin. In our evaluation of this gene, we have identified a unique nonsense mutation in 21 SCIDA patients that is closely correlated to the founder haplotypes that we had previously identified. This nonsense founder mutation results in the truncation of the deduced protein product. The wild-type construct of the primary transcript can effectively complement the defective coding joint and reduced signal joint formation in SCIDA fibroblasts. The above results indicate that this SNM1-like gene (Artemis) is the gene responsible for SCIDA. We also discovered three additional alternative exons and detected at least six alternatively spliced SCIDA variants (SCIDA-V1, 2, 3, 4, 5, and 6) coexisting with the primary transcript in trace amounts. Finally, we found that the SCIDA primary transcript (Artemis) encodes a nuclear protein. ***The Journal of Immunology*****, 2002, 168: 6323–6329.**

Severe combined immunodeficiency (SCID) is a group of genetically and phenotypically heterogeneous disorders, all of which result in severe T cell and B cell immunodeficiency (1, 2). The most severe form of SCID has the $T^-B^-NK^+$ phenotype, accounting for ~20% of all cases in which patients present with a virtual absence of both circulating T and B cells, while maintaining a normal level and function of NK cells (1–3). This $T^-B^-NK^+$ condition is autosomal recessive. $T^-B^-NK^+$ SCID cases have been associated with mutations in recombinase-activating genes (RAG1[4] and RAG2) that initiate V(D)J recombination (4–6). A small group of patients with $T^-B^-NK^+$ SCID was also found to have defective V(D)J recombination in an extrachromosomal substrate assay and increased sensitivity to ionizing radiation (radiation sensitive (RS)-SCID) in bone marrow and fibroblast cells (7, 8).

Although it is very rare in the general population, $T^-B^-NK^+$ SCID has a high incidence (1:2000 live births) in Navajo and Apache Native Americans (Athabascan SCID (SCIDA), OMIM 602450) (9–11). The Navajo and Apache are believed to have originated from the Na-Dene subdivision of the Athabascan-speaking linguistic group that migrated into the southwestern U.S. from Alaska and western Canada between 700 and 1300 A.D., and diverged at the end of the 17th century (12, 13). The SCIDA condition found in these two groups appears quite unique and homogenous, characterized by the early onset of serious infections, failure to thrive, and severe oral and/or genital ulcers (14–16). It presents with a $T^-B^-NK^+$ immunophenotype, and affected children usually die from severe infections within 6 mo of age without a successful bone marrow transplant (16). SCIDA is inherited as an autosomal recessive trait with an estimated gene frequency of 2.1% in the Navajo population (10).

In a previous study, we established genetic linkage of the SCIDA gene to a region of 6.5 cM on chromosome 10p (11). We also presented data indicating a strong founder effect for this condition based on the results that in the SCIDA candidate region, certain rare alleles from the markers around the SCIDA locus were highly preserved on the SCIDA chromosomes, suggesting that these SCIDA chromosomes were descended from a common founder. A significant impairment in V(D)J coding joint formation, and reduced but precise signal joint formation were identified later in the fibroblast cells from a SCIDA patient that we included in our linkage study, using an extrachromosomal V(D)J recombination assay (17).

V(D)J recombination is the mechanism by which V, D, and J gene segments of Igs and TCRs are rearranged and assembled in the germline into the exon-encoding Ag recognition receptors in lymphoid cells (18). It has been well established that V(D)J recombination is initiated by the RAG1/2 complex, which recognizes the recombination signal sequences (RSSs) that flank the V,

*Department of Pediatrics, University of California, San Francisco, CA 94143; †Développement Normal et Pathologique du Système Immunitaire, Hôpital Necker Enfants Malades, Paris, France; and ‡Department of Pediatrics, Tuba City Indian Health Services, Tuba City, AZ 86045

Received for publication October 4, 2001. Accepted for publication April 19, 2002.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by grants from the National Institutes of Health (AI 28339), the March of Dimes (6-FY00-301), the Children's Health Initiative, Lucille Packard Foundation, and the Nigel Gough Memorial Fund.

[2] L.L. and D.M. are co-first authors of this manuscript.

[3] Address correspondence and reprint requests to Dr. Morton J. Cowan, Department of Pediatrics, University of California, 505 Parnassus Avenue, San Francisco, CA 94143-1278. E-mail address: mcowan@peds.ucsf.edu

[4] Abbreviations used in this paper: RAG, recombinase-activating gene; BAC/PAC, bacterial artificial chromosome/phage artificial chromosome; DSB, DNA double-strand break; HA, hemagglutinin; NLS, nuclear localization signal; PKcs, protein kinase catalytic subunit; RS-SCID, radiation-sensitive SCID; RSS, recombination signal sequence; SCIDA, Athabascan SCID.

0022-1767/02/$02.00

D, and J segments and introduce DNA double-strand breaks (DSBs) between the RSSs and the coding sequences (19, 20). The cleavage leaves blunt signal ends (with RSSs at the ends) that can be directly ligated, and covalently sealed hairpinned coding ends, which require further processing before the rejoining of different V, (D), and J segments into various exon-encoding Ag recognition sites. The functional significance of RAG proteins has been well demonstrated in RAG-deficient mice and also in $T^-B^-NK^+$ SCID cases found to have RAG mutations (4–6). The resolution of the coding and signal ends depends on several ubiquitously expressed proteins, including DNA-protein kinase catalytic subunit (PKcs), Ku 80/70 complex, XRCC4, and DNA ligase IV, which are essential factors for the general nonhomologous end joining machinery. To date, no human SCID cases have been associated with mutations in the above factors, which have also been excluded from causing RS-SCID or SCIDA (8, 11).

In an attempt to identify the SCIDA gene, we conducted a combined positional candidate and positional cloning approach. We developed a bacterial artificial chromosome/phage artifical chromosome (BAC/PAC) contig covering a ~2.5-cM SCIDA region defined by linkage disequilibrium analysis, from which we isolated over 30 microsatellite markers and further refined the region to a ~500-kb interval by haplotype analysis (21). We identified 28 transcript units from this interval, some of which were found to be derived from NMT2 and a pseudogene, while others were assembled into 9 novel transcripts (22). One of the transcripts showed significant homology with the mouse and yeast SNM1/PSO_2 gene, which has been shown to be involved in DNA cross-link damage repair, and the abbreviation SNM1 was used for "sensitive to nitrogen mustard." Recently, this SNM1-like gene has been reported to be responsible for RS-SCID and named Artemis (23). RS-SCID represents a similar $T^-B^-NK^+$ immunodeficiency condition found in patients of primarily European ethnic origin. Affected patients have been shown to have increased radiosensitivity in their bone marrow and fibroblast cells, and their fibroblasts also show impaired coding joint formation in an extrachromosomal V(D)J recombination assay (7, 8). The RS-SCID condition was previously linked with markers in the SCIDA candidate region (17). Eight various Artemis mutations have been found in 13 RS-SCID patients, including deletions, alternative splicing, and a nonsense mutation (23).

In this study, we report our results indicating that this SNM1-like gene, Artemis, is the SCIDA responsible gene, a founder mutation of which causes SCIDA in Navajo and Apache Native Americans. We also present the identification of several alternative exons and six alternative splicing products of SCIDA. Finally, we demonstrate the expected nuclear localization of the SCIDA/Artemis gene product by transient expression of epitope-tagged SCIDA cDNA.

Materials and Methods

Patients

In this study, we included 21 patients of Athabascan-speaking Native American ethnic origin, including 18 Navajo and 3 Apache, 18 of whom were diagnosed and treated at the University of California (San Francisco, CA) between 1984 and 2000 (16). All of the patients had the $T^-B^-NK^+$ immunophenotype. Patients 15-A1 and 25-A1 are first cousins, and 30-A1 is a paternal distant cousin of 14-A1. No other cross consanguinity could be traced among the families. We also included 30 normal, unrelated Navajo adult controls in this study. EDTA blood samples were obtained from patients, parents, siblings, and normal controls. Primary fibroblast cell lines for 19 of the 21 SCIDA patients were established from skin biopsies, and the cell line for one of the two remaining patients, 13-A2, was derived from amniocytes. The primary cell lines established from SCIDA patients 03-A2 and 07-A1 were also transformed and pseudoimmortalized with an ori^- SV40 construct pLAS-wt, provided by Dr. L. Daya-Grosjean (Institut Andre Lwoff, Villejuif, France) (24). DNA was extracted using standard methods from EDTA blood samples and fibroblast cells (25). Informed consent was obtained from all the participating subjects. The study was approved by the Committee on Human Research at the University of California, the Navajo Nation Health Board, and the Phoenix Area Indian Health Service IRB.

Genotyping and haplotype analysis

All of the microsatellite markers were genotyped, and the alleles for each marker were numbered using the same method that we described in our previous linkage study (11). Initially, we genotyped and analyzed the five microsatellite markers from our linkage study, D10S1664, D10S191, D10S1653, D10S674, and D10S1477 (11), in 36 Navajo SCIDA chromosomes and 60 normal control chromosomes of Navajo origin for the purpose of primary refinement of the SCIDA locus. After constructing the BAC/PAC contig covering the refined 2.5-cM SCIDA region from D10S191 to D10S1653, we included over 30 newly developed short tandem repeat polymorphism markers in our haplotype analysis of all 42 SCIDA chromosomes. These markers were identified from BAC/PAC clones by screening small subclones of BAC/PAC DNA using one vector primer and one di- or trinucleotide repeat primer (26). The positions of the new markers were determined by PCR amplification on the overlapped YAC and BAC/PAC clones (data not shown). We later included over 10 microsatellite markers from the human genome draft sequences of the correlated Sanger's BAC/PAC clones located within the refined ~500-kb SCIDA region. The full haplotypes of SCIDA chromosomes were unambiguously determined based on the genotypes of all family members except for patient 24-A1, whose haplotypes were determined only for those loci showing allele homozygosity because samples from family members were not available.

RT-PCR and 5′ RACE

RT-PCR analysis was performed to evaluate the SCIDA/Artemis gene. At first, we designed and used three pairs of primers to cover the SCIDA/Artemis coding region, exon1F/exon6R, exon4F/exon10R, and exon9F/exon14R. When we detected the multiple amplifications between exons 1 and 6, we sequenced the PCR products and also designed primers exon2F, exon2bF, and exon2cF to pair with exon3R and primers to pair exon3F with exon3bR, exon4R, and exon6R, respectively, in RT-PCR amplification. We also used two pairs, exon1F/exon4R and exon1F/exon6R, to evaluate the alternative exons in a multiple tissue cDNA panel of both fetal and adult brain, kidney, skeleton muscle, spleen, and thymus (Clontech Laboratories, Mountain View, CA). All the RT-PCR reagents were from Invitrogen (Carlsbad, CA) for total RNA extraction and first strand cDNA synthesis, and standard PCR conditions were used for all the evaluations. We also performed the 5′ RACE experiments to evaluate the 5′ region of the gene using the gene-specific primers (designed from exons 4 and 6) and the Smart RACE system from Clontech Laboratories, following the manufacturer's instructions.

Northern blot analysis

A Northern blot was prepared from total RNA extracted from the fibroblast cells of two SCIDA patients and two normal controls, and was probed with an RT-PCR product covering the coding sequences of the SCIDA gene.

Mutation screening

Mutation screening was performed first by RT-PCR amplification of first strand cDNA samples from fibroblast cells of patients and controls, using primers exon1F/6R, exon4F/10R, and exon9F/14R to cover the coding region, followed by sequencing the RT-PCR products. We also designed at least 14 pairs of primers from intron sequences to amplify and evaluate all the exons and exon-intron junctions using PCR amplification on genomic DNA, followed by sequencing analysis.

V(D)J recombination assay

The extrachromosomal V(D)J recombination assay and functional complementation were performed in SV40-transformed fibroblasts from patients 03-A2, 07-A1, and two normal controls by cotransfecting RAG1/2 expression constructs and the extrachromosomal substrate pGG49 (amp^r, for signal joint) or pGG51 (amp^r, for coding joint), using a previously described method (27–29). A full-length SCIDA primary transcript expression construct, pCMV-SCIDA (3.5 μg), was used to assess the complementation. pGG49 and pGG51 carry a *cam* gene that is interrupted from its promoter by a transcriptional terminator flanked by RSSs. Upon transfection, RAG proteins induce V(D)J recombination in the extrachromosomal plasmid substrates, resulting in the excision of the transcriptional terminator and the

activation of chloramphenicol resistance. The percentage of successful recombination is represented by the ratio of colonies grown on ampicillin/chloramphenicol (from recombined substrate only) vs ampicillin plates.

SCIDA subcellular localization

Both Cos7 and HK-293 cells were transfected with cDNA constructs containing the full length of the SCIDA primary transcript (Artemis) in frame with either the hemagglutinin (HA) or *myc* epitopes (Clontech Laboratories vector pCMV-HA or pCMV-Myc). Two days after lipofectamine-mediated transfection (Invitrogen), cells grown on glass slides were fixed with 2% paraformaldehyde, and the presence of HA- and *myc*-tagged proteins was detected by indirect immunofluorescence. Primary Abs (2 μg/ml) included mouse monoclonal anti-*myc* IgG and rabbit anti-HA affinity-purified IgG (Clontech Laboratories). FITC-labeled secondary Abs (Jackson ImmunoResearch Laboratories, West Grove, PA) were used at 5 μg/ml. Standard epifluorescence microscopy was performed using a Zeiss (Oberkochen, Germany) microscope, and confocal microscopy was performed on a Zeiss LSM 510 confocal microscope. Negative controls included untransfected cells and cells transfected in parallel, but not incubated with primary Abs.

We also used the PSORT II program to predict the subcellular localization site for the SCIDA protein. The program was developed by K. Nakai and P. Horton and was available from http://psort.nibb.ac.jp/psort/. PSORT II analyzes the input amino acid sequence by applying the stored rules for various sequence features of known protein-sorting signals, among which the subprogram NNCN was adopted from NNPSL (30), which makes the prediction by analyzing the amino acid composition according to the neural network. The reasoning system k-NN uses the k-nearest neighbor algorithm to assess the probability of localization at each candidate site (31).

Results

The genomic structure of the SCIDA gene and the alternative splicing events

The SCIDA primary transcript was detected by RT-PCR, containing exon 1–14, and is the same as the Artemis gene (23). When we used primer pairs of exon1F/exon6R to perform RT-PCR on different cDNA samples, we observed multiple weaker bands accompanying the major SCIDA primary transcript. We sequenced the related PCR products, and the results from analyzing the sequences suggested that the weaker bands resulted from alternative splicing events. Similar results were also observed in 5′ RACE and full-length amplification experiments. We detected three additional alternative exons, 2b, 2c, and 3b, in addition to the 14 exons in the SCIDA primary transcript (Artemis). By analyses of the SCIDA mRNA sequences and the corresponding DNA genomic draft sequences of the human genome, we detected the same exon-intron boundary consensus sequences (splice sites) for all 17 exons in the corresponding genomic sequences.

The organization of the SCIDA gene is presented in Fig. 1. Alternative splicing was observed with exons 2, 2b, 2c, 3b, and 5 in the evaluation of multiple normal tissues by RT-PCR analysis and 5′ RACE experiments, resulting in at least six alternatively spliced variants (GenBank accession AF395747-AF395752). Variant 1 (SCIDA-V1) includes all the exons except 2c and 3b; SCIDA-V2 contains all the exons except 2b and 3b; SCIDA-V3 has all the other exons except exons 2b, 2c, 3b, and 5; SCIDA-V4 has none of the alternative exons; SCIDA-V5 contains exons 1–14 plus 2b and 3b; and SCIDA-V6 has exons 1–14 plus 2c and 3b. The alternatively spliced variants, SCIDA-V1, V2, V5, and V6, result in a shorter open reading frame of 1689 bp, in which the first ATG appears at the position of the 121st aa of the deduced protein from the primary transcript (SCIDA/Artemis) and the second at the147th aa. SCIDA-V3 and SCIDA-V4 have a similar shorter open reading frame of 1734 bp with the first ATG appearing in a different frame in exon 4 and the second ATG also at the 147th aa. The stop codon of the six alternative spliced transcripts appears to be the same as the dominant one. The surrounding sequences of the ATG at the position of the 147th aa appear to best match the Kozack consensus sequence for translation initiation (32). We evaluated the expression level of SCIDA and its splice variants with the commercial panels of cDNAs (Clontech Laboratories) from both fetal and adult tissues of brain, kidney, skeleton muscle, spleen, and thymus. The alternatively spliced transcripts appeared to coexist in trace amounts with the primary transcript, and no obvious tissue-specific or developmental-specific distributions were observed (data not shown).

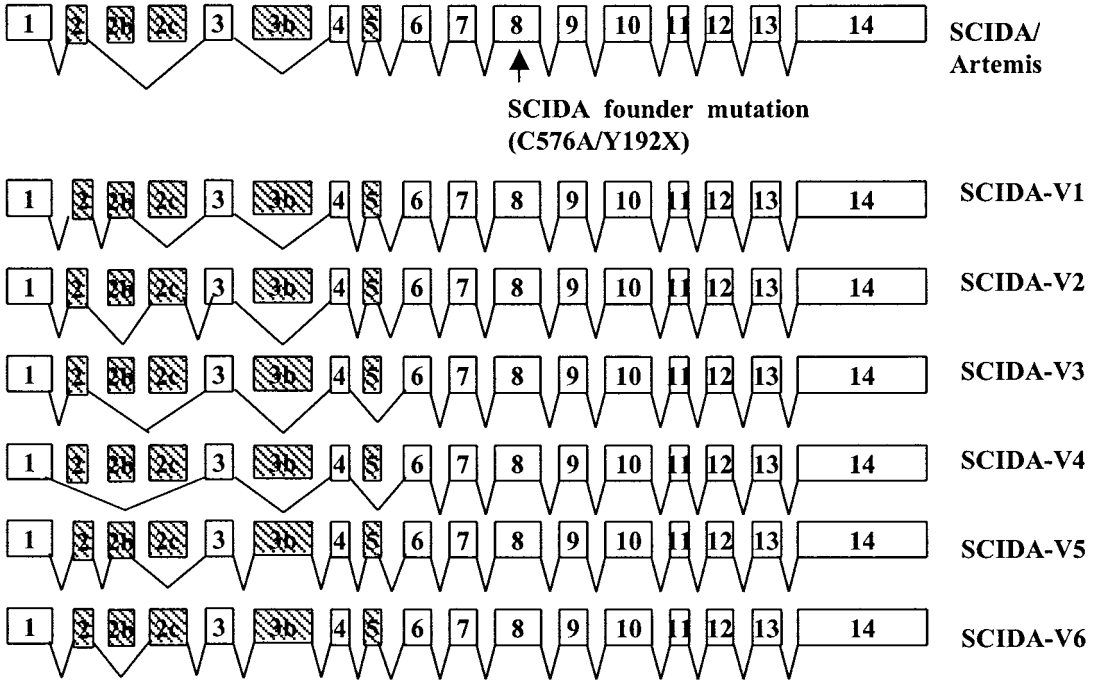

FIGURE 1. The organization of the human SCIDA gene and the six splice variants. The exon-intron organization is depicted. The constitutive exons are shown as □ and the alternative exons are shown as ▨.

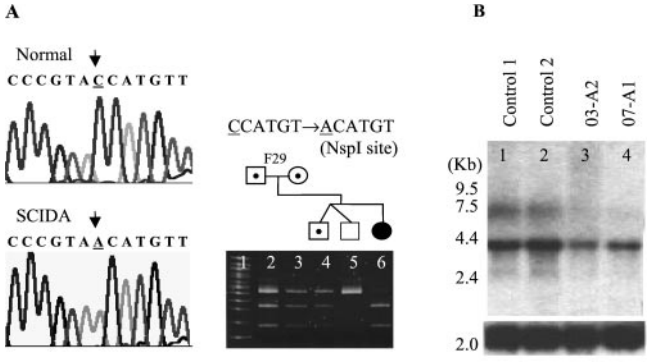

FIGURE 2. *A*, The nonsense homologous founder mutation C→A at exon 8 of the SCIDA gene in SCIDA patient 01-A1 and a normal control plus the *Nsp*I digestion of the SCIDA exon 8 PCR products in family 29. The affected child (*lane 6*) carried the homologous founder mutation that resulted in the total cleavage of the exon 8 PCR products into two bands. The parents and the carrier sibling (*lanes 2*, *3*, and *4*) present with about one-half of the products cleaved by *Nsp*I, and the normal sibling (*lane 5*) shows no cleavage. *B*, The low expression of the SCIDA transcript in SCIDA patient 03-A2 and 07-A1 fibroblast cells (*lanes 3* and *4*) compared with the normal expression of the SCIDA mRNA in control fibroblast cells (*lanes 1* and *2*).

The SCIDA founder mutation and the founder haplotypes

A homozygous single base change (TAC→TAA) in exon 8 was found in 17 of 18 Navajo SCIDA patients and all 3 Apache SCIDA patients (Fig. 2*A*). It was also present on the mRNA of the remaining Navajo SCIDA patient (30-A1), which appeared to result from the paternal allele that carries the mutation, and the absence of the maternal allele that carries no mutation at exon 8 in the genomic DNA. This absence of the maternal SCIDA allele in the mRNA expression was also confirmed by a single nucleic polymorphism in exon 9, in which the base that was carried homozygously on the mother's genomic DNA was also absent in the RT-PCR product of 30-A1 (data not shown). The nonsense mutation in exon 8 creates an *Nsp*I site (CCATGT→ACATGT), leading to a simple assay to confirm the mutation and the heterozygous status for patients and carriers (Fig. 2*A*). We evaluated this mutation in the parents and siblings and 60 normal chromosomes from Navajo controls, which eliminated its possibility of being a normal polymorphism. The nonsense mutation creates a premature stop codon in SCIDA, and its expression level was low in SCIDA patient fibroblast cells compared with normal control fibroblasts in Northern blot analysis (Fig. 2*B*), suggesting the involvement of nonsense-mediated mRNA decay (33). The deduced putative protein from the SCIDA primary transcript (Artemis) is composed of 686 aa, and the nonsense mutation in SCIDA patients resulted in a speculated truncation of SCIDA at the 192nd aa. The SCIDA truncation was seen in SCIDA patients in the Protein Truncation Test (data not shown).

The full haplotypes of 42 SCIDA chromosomes for 12 representative loci in the SCIDA region are presented in Table I. Numbers were used to represent the different alleles for each marker. The founder haplotype is bolded. Thirty-five of thirty-six Navajo SCIDA chromosomes presented in the first eight rows, and all six Apache SCIDA chromosomes listed in Table I carry the nonsense mutation, and they all have the haplotypes formed by the SCIDA-associated alleles in a rather large region. The maternal allele of patient 30-A1 (row 9) that carries an unidentified regulatory mutation appears to have a smaller region bearing the SCIDA-associated allele.

Table I. *Full haplotypes of 42 SCIDA chromosomes in the SCIDA region*[a]

No. of Chromosomes	Markers in SCIDA Candidate Region													Mutation(s)
	D10S1664	D10S191	85F9F1R1	22A2F1R1	1000B2F1R1	2K17F2R2	SCIDA gene	P853B18atFR	2O14-1FR	2O14-2FR	D10S1653	74L17F2R2	D10S674	
Navajo														
14	**4**	**8**	**7**	**2**	**2**	**2**		**2**	**1**[b]	**1**[b]	**2/4**[b]	**5/6**[b]	**4**	C576A/Y192X[c]
12	**4**	**8**	**7**	**2**	**2**	**2**		**1**	**3**[b]	**2**[b]	**2/4**	**5/6**	**4**	C576A/Y192X
2	**4**	**8**	**7**	**2**	**2**	**2**		**2**	**1**	**1**	**2/4**	**5/6**	**1**	C576A/Y192X
1	**4**	**8**	**7**	**2**	**2**	**2**		**1**	**3**	**2**	**2/4**	**5/6**	**1**	C576A/Y192X
2	**1**	**8**	**7**	**2**	**2**	**2**		**1**	**3**	**2**	**2/4**	**5/6**	**5**	C576A/Y192X
1	**6**	**8**	**7**	**2**	**2**	**2**		**2**	**1**	**1**	**2/4**	**5/6**	**5**	C576A/Y192X
1	**4**	**8**	**7**	**2**	**2**	**2**		**2**	**1**	**1**	6	7	**8**	C576A/Y192X
2	**7**	9	8	3	**2**	**2**		**2**	**1**	**1**	**2**	**6**	**4**	C576A/Y192X
1[d]	10	6	5	3	**2**	**2**		**2**	**1**	**1**	**2**	4	9	Unknown
Apache														
2	**4**	**8**	**7**	**2**	**2**	**2**		**1**	**3**	**2**	**4**	**5**	**4**	C576A/Y192X
2	**4**	**8**	**7**	**2**	**2**	**2**		**2**	**1**	**1**	**2**	**6**	**4**	C576A/Y192X
1	**4**	**8**	**7**	**2**	**2**	**2**		**2**	**1**	**1**	**2**	**6**	**4**	C576A/Y192X
1	6	2	1	3	**2**	**2**		**2**	**1**	**1**	**2**	**6**	**4**	C576A/Y192X

[a] The alleles at each locus are represented by numbers (in order) starting from 1 for the smallest allele seen in normal controls. SCIDA-associated alleles at each marker locus are boldface and the SCIDA locus is blank.
[b] Both alleles associated with SCIDA and rare in normal controls.
[c] The nonsense founder mutation.
[d] The haplotype from the 30-A1 maternal allele.

Table II. *Complementation of V(D)J recombination in SCIDA patient (13-A2 and 07-A1) fibroblasts transfected with a wild-type SCIDA construct (pCMV-SCIDA)*

Cell Line	pCMV-SCIDA	Coding (pGG51)			Signal (pGG49)		
		amp^r and *cam*^{r}	Total (*amp*^{r})	R%[a]	*amp*^{r} and *cam*^{r}	Total (*amp*^{r})	R%[b]
Control 1	–	376	37,880	9.9	10	980	1.02
Control 2	–	30	1,360	2.2	15	1,320	1.13
13-A2	–	2	81,920	0.02	532	85,080	0.62
	+	347	49,840	6.96	112	7,520	1.49
07-A1	–	0	59,280	<0.01	62	25,000	0.25
	+	1,040	153,280	6.78	71	4,480	1.58

[a] R (coding joints) = (*amp*^{r} and *cam*^{r})/(Total) × 1,000.
[b] R (signal joints) = (*amp*^{r} and *cam*^{r})/(Total) × 100.

Complementation of the defective V(D)J recombination activity by the SCIDA gene

In a previous study, it was demonstrated in one SCIDA cell line, AK5760 (03-A1), that coding joint formation was nearly absent while signal joint formation appeared precise, but with a lower frequency (17). In the current study, the same coding joint defect was detected in AK5760 and was effectively complemented when expressing the wild-type Artemis gene (data not shown). We obtained comparable results in fibroblast cells from SCIDA patients 03-A2 and 07-A1 using extrachromosomal substrates pGG49 (for signal joint) and pGG51 (for coding joint), plus pCMV-SCIDA for complementation (Table II). Furthermore, in addition to restoring coding joint function with pCMV-SCIDA, the apparently lower rate of signal end rejoining in the SCIDA fibroblast cells was also significantly increased to control levels.

SCIDA/Artemis gene nuclear localization

Cos7 and HK-293 cells transfected with either HA- or *myc*-tagged SCIDA/Artemis cDNA showed a characteristic nuclear pattern of expression by immunofluorescence. All positive cells (~10–20% of the cells on the slides) showed the tagged protein located essentially exclusively in the nucleus, with occasional cells also showing a punctate cytoplasmic staining, consistent with the endoplasmic reticulum (Fig. 3*A*). Since the anti-HA immunofluorescence experiments showed higher background staining, all the detailed analysis was performed with the anti-*myc* detection. The nuclear immunostaining clearly excluded the nucleoli and was as intense in the periphery as in the nucleoplasm interior (Fig. 3*B*). High resolution microscopy revealed spots of more intense staining distributed throughout the nucleoplasm. No staining was detected in untransfected cells or slides unexposed to the primary Ab.

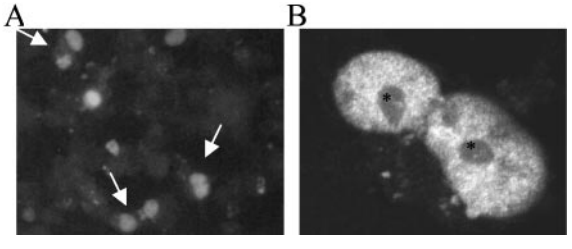

FIGURE 3. SCIDA nuclear localization from immunofluorescence experiment using *myc*-tagged SCIDA cDNA. *A*, The view from ×10 microscopy: the positive cells (arrows) show the tagged protein located exclusively in the nucleus, with occasional cells showing punctate cytoplasmic staining. *B*, The view from ×100 microscopy: the nuclear immunostaining excludes the nucleoli (*) and is as intense in the periphery as in the nucleoplasm interior.

The region of SCIDA/Artemis that is homologous to SNM1/PSO_2 does not include the region presenting the putative nuclear localization signal (NLS) (34). Although a bipartite NLS could not be identified in the SCIDA cDNAs, the basic residue cluster KKRK (aa 676–679 in the primary transcript product) was detected as a potential NLS. The NNCN prediction supports the nuclear localization for SCIDA/Artemis with 89% reliability. k-NN analysis gave a prediction of 69.6% possibility in the nucleus vs 13% in the cytoplasm, 8.7% in Golgi, 4.3% in the cytoskeleton, and 4.3% in the plasma membrane.

Discussion

Our results indicate that a unique nonsense mutation in an SNM1-like gene, Artemis causes SCIDA. Twenty of twenty-one SCIDA patients of Navajo and Apache origin in our study carried the homozygous nonsense mutation at exon 8 (TAC→TAA), which was not seen in any of the 30 normal controls, while it was carried by their parents. The same mutation was also presented on the paternal chromosome of the 21st patient and was pathogenic because the maternal SCIDA allele was not expressed in the mRNA sequence. This nonsense mutation creates a premature stop codon and results in a decreased level of SCIDA transcription (Fig. 2*B*) and the truncation of its protein product. The V(D)J recombination defect found in the SCIDA fibroblast cells can be effectively complemented by transfecting the wild-type SCIDA primary transcript into the fibroblast cells of SCIDA patients.

While RS-SCID represents a group of sporadic cases resulting from various mutations in Artemis (SCIDA), our results support a founder effect causing the high incidence of SCIDA in Navajo and Apache Native Americans. We previously proposed the founder effect for SCIDA based on the observations that the phenotype of all SCIDA cases is quite homogeneous and has relatively unique clinical aspects, in particular a high incidence of oral and genital ulcers, which appears very characteristic and has only been reported in SCIDA patients (14–16). In addition, we found that certain rare alleles from SCIDA-linked loci were highly preserved on the SCIDA chromosomes, suggesting a common origin(s). Finally, these two groups of Native Americans share a possible ancestry link. The Navajo and Apache are believed to have originated from the same Athabascan-speaking linguistic group that migrated into the southwestern U.S. between 700 and 1300 A.D. and diverged at the end of the 17th century (12, 13). In our studies, we found that almost all the Navajo and Apache disease chromosomes carry a significantly rare common haplotype formed by certain rare alleles from loci around the SCIDA locus ($p < 0.0001$) (Table I and data not shown). The nonsense mutation at exon 8 was detected in 35 of the 36 Navajo and all six Apache SCIDA chromosomes, all of which carried the

SCIDA common haplotypes. These results support that this nonsense mutation represents a mutation that occurred on the founder chromosome before the separation of the Navajo and Apache at the end of the 17th century. Our results support the view of common ancestors for Navajo and Apache populations.

The results of this study further support the belief that SCIDA (Artemis) is involved in V(D)J recombination/DNA repair (28). By using extrachromosomal substrates to demonstrate the V(D)J recombination, we show that SCIDA fibroblast cells carry a defect that severely impairs coding joint formation and apparently also reduces the rate of precise signal joint formation, similar to what has been previously reported (17). Furthermore, we show that both of these specific defects can be effectively complemented by the expression of a wild-type construct of the SCIDA primary transcript. Finally, the nuclear localization of the SCIDA (Artemis) protein that we found also correlates with its inferred functional involvement in V(D)J recombination, although its precise role remains unknown.

It has been previously reported that sequence comparison did not reveal informative orthologue or functional domains that might indicate the nature of the functional involvement of Artemis in V(D)J recombination (23). In the same report, it was noted that the C-terminal portion of 331 aa does not show obvious similarity with any known proteins, while the remaining portion appears to share significant sequence similarity with the murine and yeast SNM1/PSO_2 gene (23). SNM1/PSO_2 has been found to be involved in the repair of DNA damage caused by DNA interstrand cross-linking agents, but not ionizing radiation. Therefore, the C-terminal region of SCIDA/Artemis may be critical in directing itself to the specific DSBs (including the physiological coding and signal ends) repair pathway. We have recently found that DNA-PKcs coprecipitates with SCIDA in immunoprecipitation studies, that SCIDA can be phosphorylated by DNA-PK in vitro, and that the critical site(s) is located in the C-terminal region (data not shown). DNA-PKcs has been known to quickly appear, following Ku proteins, at DSB ends, and together they may function in recruiting other factors (XRCC4/ligase IV complex) and activating the specific DNA repair machinery (35–38). Therefore, although still preliminary, our results support the potential importance of the C-terminal region in directing SCIDA/Artemis to the V(D)J recombination/nonhomologous end joining repair pathway.

DNA-PKcs-, Ku-, XRCC4-, and ligase IV-deficient cells and animal models all show impaired lymphogenesis and increased radiosensitivity. The phenotype of DNA-PKcs deficiency appears relatively simple, and although the coding joint formation is severely impaired, the signal end rejoining appears relatively normal (39), which has been suggested to be the result of an alternative mechanism (40). In contrast, Ku-deficient mice show complete impairment of both coding and signal joint formation, plus signs of growth retardation and cellular phenotypes, suggesting additional functional roles (38, 41, 42). XRCC4 and ligase IV-deficient mice exhibit a different pleiotropic phenotype, including late embryonic lethality due to massive apoptosis of newly generated neurons, and various cellular defects (43, 44). Comparatively, SCIDA more closely resembles the DNA-PKcs deficiency phenotypes, but in SCIDA, the signal end rejoins accurately, although at a lower rate, while the coding joint formation is completely impaired. These results suggest that SCIDA may not be directly involved in the signal end rejoining reaction and that the defect in SCIDA may be related to coding end processing.

The mechanism for DSB end (including coding end) processing remains unclear. Based on the current knowledge, it is thought to involve a coordinated reaction of nucleases, polymerases, and other factors (45, 46). DNA-PKcs and Ku proteins are most likely involved based on the accumulation of hairpinned intermediates in targeted mice (41, 47), and it may be related to their potential roles in recruiting and/or activating factors that mediate the end processing. Therefore, the physiological and functional associations of SCIDA with DNA-PKcs and Ku are most likely significant and will be elucidated in further study.

With respect to the several SCIDA alternative splicing products that we identified, it is unclear as to their functional significance. They all exist in trace amounts without obvious tissue or developmental differences. However, alternative splicing has been recognized as a powerful and versatile regulatory mechanism that can affect the quantitative control of gene expression and accounts for a considerable proportion of proteomic complexity in higher eukaryotes (48, 49). A similar alternative splicing event was also observed in our analysis of the mouse *Scida* gene (data not shown), which appears to support the existence of putative functional alternative transcripts for SCIDA. The deduced proteins from the alternative splicing products lack the N-terminal domain containing the metallo-β-lactamase fold and retain the conserved region harboring sequences that may be related to nucleic acid metabolism (23) and the C-terminal region that appears related to the functional interaction with DNA-PKcs. Further studies will be required to determine whether or not these variants have any functional significance.

Finally, the identification of the specific SCIDA common mutation provides a direct way for prenatal diagnosis and carrier detection in Navajo and Apache populations. It is now feasible to establish a population-based screening program to identify the estimated 2.1% of individuals who carry the mutation, in addition to establishing a population-based newborn screening test considering its high incidence in this ethnic group. Pregnancies of diagnosed carriers can be monitored accurately by early DNA-based testing for family-specific mutations, which is considerably easier and more accurate than linkage analysis. The identification of the responsible gene and SCIDA mutation also opens the door for the prospect of the eventual cure of SCIDA by gene therapy.

Note added in proof. It has recently been reported that Artemis (SCIDA) associates with DNA-PKcs, and upon the formation of the complex and phosphorylation by DNA-Pkcs, it acquires endonuclease activity on 5′ and 3′ overhangs and hairpins (50).

Acknowledgments

We thank all the families for participating in the study. We thank Dr. Daya-Grosjean for providing us with the plasmid construct pLAS-wt, Dr. Michael Lieber for the substrates pGG49 and pGG51, Dr. Joanne Hesse for the RAG1 and RAG2 expression constructs, and Dr. Ann Feeney for her valuable comments on the manuscript revision. We thank Orla Clifford for her great administrative help.

References

1. Rosen, F. S., M. D. Cooper, and R. J. P. Wedgwood. 1995. The primary immunodeficiency. *N. Engl. J. Med. 333:431.*
2. Fisher, A., M. Cavazzana-Calvo, G. de Saint Basile, J. P. de Villartay, J. P. Di Santo, C. Hivroz, F. Rieux-Laucat, and F. Le Deist. 1997. Naturally occurring primary deficiencies of the immune system. *Annu. Rev. Immunol. 15:93.*
3. Buckley, R. H., R. I. Schiff, S. E. Schiff, M. L. Markert, L. W. Williams, T. O. Harville, J. L. Roberts, and J. M. Puck. 1997. Human severe combined immunodeficiency: genetic, phenotypic, and functional diversity in one hundred eight infants. *J. Pediatr. 130:378.*
4. Schwarz, K., G. H. Gauss, L. Ludwig, U. Pannicke, Z. Li, D. Lindner, W. Friedrich, R. A. Seger, T. E. Hansen-Hagge, S. Desiderio, et al. 1996. RAG mutations in human B cell-negative SCID. *Science 274:97.*
5. Villa, A., S. Santagata, L. Imberti, F. Bozzi, and L. D. Notarangelo. 1999. Omenn syndrome: a disorder of Rag1 and Rag2 genes. *J. Clin. Immunol. 19:87.*
6. Villa, A., C. Sobacchi, L. D. Notarangelo, F. Bozzi, M. Abinun, T. G. Abrahamsen, P. D. Arkwright, M. Baniyash, E. G. Brooks, M. E. Conley, et al. 2001. V(D)J recombination defects in lymphocytes due to RAG mutations: severe immunodeficiency with a spectrum of clinical presentations. *Blood 97:81.*

7. Cavazzana-Calvo, M., F. Le Deist, G. de Saint Basile, D. Papadopoulo, J. P. de Vellartay, and A. Fisher. 1993. Increased radiosensitivity of granulocyte macrophage colony-forming units and skin fibroblasts in human autosomal recessive severe combined immunodeficiency. *J. Clin. Invest. 91:1214.*
8. Nicolas, N., D. Moshous, D. Papadopoulo, M. Cavazzana-Calvo, R. de Chasseval, F. le Deist, A. Fisher, and J. P. de Villartay. 1998. A human SCID condition with increased sensitivity to ionizing radiations and impaired V(D)J rearrangements defines a new DNA recombination/repair deficiency. *J. Exp. Med. 188:627.*
9. Murphy, S., A. Hayward, G. Troup, E. Devor, and T. Coon. 1980. Gene enrichment in an American Indian population: an excess of severe combined immunodeficiency disease. *Lancet ii:502.*
10. Jones, J., C. Ritenbaugh, A. Spence, and A. Hayward. 1991. Severe combined immunodeficiency among the Navajo: characterization of phenotypes, epidemiology and population genetics. *Hum. Biol. 63:669.*
11. Li, L., D. Drayna, D. Hu, A. Hayward, S. Gahagan, H. Pabst, and M. J. Cowan. 1998. The gene for severe combined immunodeficiency disease in Athabascan-speaking Native Americans is located on chromosome 10p. *Am. J. Hum. Genet. 62:136.*
12. Krauss, M. E. 1979. *Native Languages of the Americas*. T. A. Sebeok, ed. Plenum Press, New York, p. 15.
13. Kehoe, A. B. 1981. *North American Indians: a Comprehensive Account.* Prentice-Hall, Upper Saddle River, pp. 449–530.
14. Rotbart, H. A., M. J. Levin, J. F. Jones, A. R. Hayward, J. Allan, M. F. McLane, and M. Essex. 1986. Noma in children with severe combined immunodeficiency. *J. Pediatr. 109:596.*
15. Kwong, P. C., A. O'Marcaigh, R. Howard, M. J. Cowan, and I. J. Frieden. 1999. Noma: a unique presentation of immunodeficiency in Athabacan-speaking Native American children with severe combined immune deficiency. *Arch. Dermatol. 135:927.*
16. O'Marcaigh, A., K. DeSantes, D. Hu, H. Pabst, B. Horn, L. Li, and M. J. Cowan. 2001. Bone marrow transplantation for T^-B^- severe combined immunodeficiency disease in Athabascan-speaking Native Americans. *Bone Marrow Transplant. 27:703.*
17. Moshous, D., L. Li, R. de Chasseval, N. Philippe, N. Jabado, M. J. Cowan, A. Fisher, and J. P. de Villartay. 2000. A new gene involved in DNA double-strand break repair and V(D)J recombination is located on human chromosome 10p. *Hum. Mol. Genet. 9:583.*
18. Grawunder, U., R. B. West, and M. R. Lieber. 1998. Antigen receptor gene rearrangement. *Curr. Opin. Immunol. 10:172.*
19. Oettinger, M. A., D. G. Schatz, C. Gorka, and D. Baltimore. 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science 248:1517.*
20. McBlane, J. F., D. C. van Gent, D. A. Ramsden, C. Romeo, C. A. Cuomo, M. Gellert, and M. A. Oettinger. 1995. Cleavage at the V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell 83:387.*
21. Zhou, Y., L. Li, and M. J. Cowan. 2000. Physical mapping and further refinement of Athabascan severe combined immunodeficiency disease region. *Am. J. Hum. Genet. 67:264.*
22. Wang, J., L. Li, Y. Zhou, and M. J. Cowan. 2000. Transcriptional mapping of SCIDA region. *Am. J. Hum. Genet. 67:264.*
23. Moushous, D., I. Callebaut, R. de Chasseval, B. Corneo, M. Cavazzana-Calvo, F. le Deist, I. Tezczn, O. Sanal, Y. Bertrand, N. Philippe, et al. 2001. Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell 105:177.*
24. Daya-Grosjean, L. 1987. An immortalized xeroderma pigmentosum, group C, cell line which replicates SV40 shuttle vectors. *Mutat. Res. 183:185.*
25. Ford, N., M. Nolan, M. Ferguson, and M. Ockler. 1989. Isolation of high-molecular-weight DNA from mammalian cells. In *Molecular Cloning: a Laboratory Manual*, 2nd Ed. J. Sambrook, ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 9.16–9.23.
26. Robic, A, C. Dubois, D. Milan, and J. Gellin. 1994. A rapid method to isolate microsatellite markers from cosmid clones. *Mamm. Genome 5:177.*
27. Gauss, G. H., and M. R. Lieber. 1993. Unequal signal and coding joint formation in human V(D)J recombination. *Mol. Cell. Biol. 13:3900.*
28. Gauss, G. H., and M. R. Lieber. 1996. Mechanistic constraints on diversity in human V(D)J recombination. *Mol. Cell. Biol. 16:258.*
29. Gauss, G. H., I. Domain, C.-L. Hsieh, and M. R. Lieber. 1998. V(D)J recombination activity in human hematopoietic cells: correlation with developmental stage and genome stability. *Eur. J. Immunol. 28:351.*
30. Reinhardt, A., and T. Hubbard. 1998. Using neural networks for prediction of the subcellular location of proteins. *Nucleic Acids Res. 26:2230.*
31. Horton, P., and K. Nakai. 1997. Better prediction of protein cellular localization sites with the k nearest neighbor classifier. *Proc. Int. Conf. Intell. Syst Mol. Biol. 5:147.*
32. Kozak, M. 1984. Compilation and analysis of sequences upstream from the translation start site in eukaryotic mRNAs. *Nucleic Acids Res. 12:857.*
33. Hilleren, P., and R. Parker. 1999. Mechanisms of mRNA surveillance in eukaryotes. *Annu. Rev. Genet. 33:229.*
34. Freidberg, E. C. 1992. Nuclear targeting sequences. *Trends Biochem. Sci. 17:347.*
35. Jackson, S. P., and P. A. Jeggo. 1995. DNA double strand break repair and V(D)J recombination: involvement of DNA-PK. *Trends Biochem. Sci. 20:412.*
36. Chen, L., K. Trujillo, P. Sung, and A. E. Tomkinson. 2000. Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. *J. Biol. Chem. 275:26196.*
37. McElhinny, S. A., C. M. Snowden, J. McCarville, and D. A. Ramsden. 2000. Ku recruits the XRCC4-ligaseIV complex to DNA ends. *Mol. Cell. Biol. 20:2996.*
38. Featherstone, C., and S. P. Jackson. 1999. Ku, a DNA repair protein with multiple cellular function? *Mutat. Res. 434:3.*
39. Taccioli, G. E., A. G. Amatucci, H. J. Beamish, D. Gell, X. H. Xiang, M. I. T. Arzayus, A. Priestley, S. P. Jackson, A. Marshak Rothstein, P. A. Jeggo, et al. 1998. Targeted disruption of the catalytic subunit of the DNA-PK gene in mice confers severe combined immunodeficiency and radiosensitivity. *Immunity 9:355.*
40. DiBiase, S. J., Z. C. Zeng, R. Chen, T. Hyslop, E. J. Curran, Jr., and G. Iliakis. 2000. DNA-dependent protein kinase stimulates an independently active, nonhomologous, end-joining apparatus. *Cancer Res. 60:1245.*
41. Zhu, C., M. A. Bogue, D. S. Lim, P. Hasty, and D. B. Roth. 1996. Ku86-deficient mice exhibit severe combined immunodeficiency and defective processing of V(D)J recombination intermediates. *Cell 86:379.*
42. Gu, Y., K. J. Seidl, G. A. Rathbun, C. Zhu, J. P. Manis, N. van der Stoep, L. Davidson, H. L. Cheng, J. M. Sekiguchi, K. Frank, et al. 1997. Growth retardation and leaky SCID phenotype of Ku70-deficient mice. *Immunity 7:653.*
43. Gao, Y., Y. Sun, K. M. Frank, P. Dikkes, Y. Fujiwara, K. J. Seidl, J. M. Sekiguchi, G. A. Rathbun, W. Swat, J. Wang, et al. 1998. A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell 95:891.*
44. Frank, K. M., J. M. Sekiguchi, K. J. Seidl, W. Swat, G. A. Rathbun, H. L. Cheng, L. Davidson, L. Kangaloo, and F. W. Alt. 1998. Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature 396:173.*
45. Lieber, M. R. 1999. The biochemistry and biological significance of nonhomologous DNA end joinning: an essential repair process in multicellular eukaryotes. *Genes Cells 4:77.*
46. Grawunder, U., and E. Harfst. 2001. How to make ends meet in V(D)J recombination. *Curr. Opin. Immunol. 13:186.*
47. Roth, D. B., J. P. Menetski, P. Nakajima, M. Bosma, and M. Gellert. 1992. V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in *scid* mouse thymocytes. *Cell 70:983.*
48. Graveley, B. R. 2001. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet. 17:100.*
49. Smith, C. W. J., and J. Valcárcel. 2000. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci. 25:381.*
50. Ma, Y., U. Pannicke, K. Schwarz, and M. R. Lieber. 2002. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell 108:781.*