Rch1, a protein that specifically interacts with the RAG-1 recombination-activating protein

CHRISTINA A. CUOMO*, SUSAN A. KIRCH††, JENO GYURIS*‡‡, ROGER BRENT*, AND MARJORIE A. OETTINGER†††

*Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114; †Department of Genetics, Harvard Medical School, Boston, MA 02115; and ‡Cell and Developmental Biology Program, Harvard University, Cambridge, MA 02138

Communicated by Keith R. Yamamoto, February 1, 1994

ABSTRACT RAG1 and RAG2 are lymphoid-specific genes that together induce V(D)J recombination activity in a variety of nonlymphoid cell types. While no other lymphoid-specific factors are required to induce recombination, other factors with more widespread expression patterns have been implicated in the reaction. However, none of these factors have been cloned, and their relationship to the RAG proteins is unclear. Using the yeast two-hybrid assay, we have identified RCH1, a gene encoding a protein of molecular weight 58,000 that interacts specifically with RAG-1. The predicted Rch1 protein sequence is 47% identical to yeast SRP1, a protein associated with the nuclear envelope. A truncated form of Rch1, which retains the ability to interact with RAG-1, reduces V(D)J recombination activity in HeLa cells.

Developing B and T cells assemble their antigen receptor molecules from component gene segments by a series of genomic rearrangement events collectively termed V(D)J recombination. Two genes involved in this assembly process, RAG1 and RAG2, were identified as being the only genes that need to be introduced into fibroblasts to induce recombination activity (1, 2). A targeted disruption of either RAG gene in transgenic mice results in the complete absence of V(D)J recombination, confirming the critical role of each gene in the reaction (3, 4).

Coexpression of RAG-1 and RAG-2 is restricted to immature cells of the lymphoid lineage (2), thereby explaining why V(D)J recombination is limited to pre-B and pre-T cells (1, 2). However, it is also clear that other factors must be involved in the recombination reaction. Such factors need not be specific for V(DJ) joining, but could instead play multiple roles in the cell and be recruited by the V(D)J recombination machinery. Examples of such factors include those defined by the mouse scid mutation (5) and by mutations in CHO (Chinese hamster ovary) cells which have defects both in V(D)J recombination and in some aspect of general DNA repair (6-10). None of these factors have been cloned, and their relationships to RAG-1 and RAG-2 are unclear.

To identify other factors that might play a role in V(DJ recombination, we used the yeast two-hybrid assay to find proteins that interact with RAG-1. In screening a cDNA library derived from HeLa cells, we were looking for widely expressed factors with which RAG-1 might interact to induce V(D)J recombination. In this paper we present one such factor, Rch1 (Rag cohort), which interacts specifically with RAG-1. A truncated form of Rch1 reduces V(D)J recombination in vivo.11

MATERIALS AND METHODS

Plasmids and Strains. The plasmids pJG4-5 and pL202Pl have been described previously (11-13). The RAG-1 acidic activation construct, pLAR1, contains the entire RAG1 structural gene and is the result of a three-part ligation of the Nco I–Sal I fragment from LG-6 [a derivative of M2SK (1)], an adaptor oligonucleotide, and EcoRl–Xho I-digested pJG4-5 vector. The LexA fusion construct, pLDR1, was prepared exactly as pLAR1 except substituting EcoRl–Sal I-digested vector pL202-Pl. The RAG-2 acidic activation and LexA DNA-binging fusion constructs pLAR2 and pLDL2 were created by subcloning the RAG2-containing EcoRI fragment from LG-4.3 [a derivative of R2RSK-1 (12)] in EcoRI-digested pJG4-5 and pL202-Pl vectors, respectively. The mammalian expression constructs for the Rch clones (HRCCD-1 and HRCCD-15) were derived by subcloning the Rch1 cDNAs in the CMV expression vector (14), using an adaptor oligonucleotide that supplied a Kozak sequence (15) and encoded the Myc epitope (16). Details of constructs are available on request.

All strains were derived from the parent Saccharomyces cerevisiae strain EGY48 [MATa trpl ura3 his3 LEU2::pLexAop6-LEU2 (13)], or EGY48 carrying LacZ reporter construct pJK103. An additional parent strain, EGY40 [MATa trpl ura3 his3 LEU2::pLexAop0-LEU2 (13)], was used to test the requirement for LexA operator sequences.

Library Screening. First, 5 × 106 independent tryptophan-synthesizing transformants of SKY34, a strain expressing pLDR1, were generated from a HeLa cDNA fusion library (13) that has an estimated complexity of 107. Then 5 × 107 cells were diluted into minimal media and plated on galactose plates lacking leucine. All resulting colonies (approximately 300) were picked after 4 days and streaked out on glucose plates lacking uracil, tryptophan, and histidine to select for the relevant plasmid markers under noninducing conditions. This master plate was replica-plated to test for transcription from the two reporter constructs—e.g., on plates containing 5-bromo-4-chloro-3-indolyl β-D-galactoside (X-Gal) and either glucose or galactose for the LacZ reporter, and on glucose or galactose plates lacking leucine. Twenty-five colonies passed this further screen. Plasmid DNA was recovered from each strain and used to transform the parent yeast strain SKY34 or an independently derived strain with identical genotype, SKY34B. Those plasmids that reproduced the interaction phenotypes were characterized further by digesting the plasmid DNA with EcoRI and Xho I to liberate the cDNA insert. Digested DNA was separated on a 0.8% gel, transferred to Zetabind membrane (AMF-Cuno), and hybridized sequentially with probes derived from independent library cDNAs.

Transient Assay for Reombinase Activity. The transient assays for recombinase activity were carried out as previously described (2). Briefly, V(D)J recombination events

1Present address: Mitotix Incorporated, 1 Kendall Square, Cambridge, MA 02139.
2To whom reprint requests should be addressed.
3The sequence reported in this paper has been deposited in the GenBank data base (accession no. U09559).

6156
were detected by using the reporter construct pJI200 (17), which constitutively expresses ampicillin resistance (Amp') while conferring chloramphenicol resistance (Cam') only after deletional rearrangement has been carried out. To test for possible effects of Rch1 on recombination proficiency, pJI200 was cotransfected with RAG-1 (M2CD7) and RAG-2 (R2RCD2) expression constructs in conjunction with CDM8 vector alone or Rch1 expression constructs HRCCD-1 or HRCCD-15 into HeLa cells by a standard calcium phosphate procedure. The amount of expression plasmid used in the transfections was typically 5 μg each of pJI200, CD8M, and the Rch1 expression plasmids HRCCD-1 or HRCCD-15. Experiments 1–5 used 6 μg of RAG1 and 4.8 μg of RAG2, while trials 6 and 7 used 1.2 μg of RAG1 and 1.0 μg of RAG2. Transfected cells were grown for 48 hr and plasmid DNA was recovered by a modified alkaline lysis procedure. Recovered plasmid DNA was digested with DpnI and introduced into MC1061 cells by electroshock and plated on the appropriate selective plates. The numbers of Amp' (representing total replicated plasmid) and Cam'/Amp' (representing recombined plasmid) colonies were counted after 15 hr of incubation at 37°C. Proper signal junction formation, indicative of a bona fide V(DJ) recombination event, was confirmed by hybridization with an oligonucleotide [SI2 (2)] specific for signal junction sequences.

RESULTS

The studies described in this paper rely on the yeast two-hybrid assay for detecting protein–protein interactions (12, 13, 18–20). Two expression plasmids are introduced into the appropriate yeast strain. One of these constitutively express a molecule in which the DNA-binding domain of LexA is fused to full-length RAG1 or RAG2. The other plasmid conditionally expresses an individual RAG protein or cDNA library yeast protein construct fused to the B42 acidic transcriptional activation domain (e.g., Activ-RAG1). Interaction between the protein fused to LexA and the protein fused to the acidic activation domain results in transcriptional activation from reporter constructs driven by promoters containing LexA operators. The selection system contains two different reporter constructs, each bearing a distinct promoter and marker gene (LEU2 or lacZ), thereby allowing two independent tests for protein–protein interaction.

RAG-1 and RAG-2 Fail to Interact in the Two-Hybrid Assay.

Given the requirement for both RAG-1 and RAG-2 in V(DJ)

Table 1. Rch1 interacts specifically with RAG-1

<table>
<thead>
<tr>
<th>Lex A fusion protein</th>
<th>Activation fusion protein</th>
<th>In glucose</th>
<th>In galactose</th>
<th>LacZ activity ratio, Gal/Glc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Leu⁺</td>
<td>LacZ activity</td>
<td>Leu⁺</td>
</tr>
<tr>
<td>RAG-1</td>
<td>Rch1-(33–529)</td>
<td>–</td>
<td>7</td>
<td>+</td>
</tr>
<tr>
<td>RAG-1</td>
<td>Rch1-(244–529)</td>
<td>–</td>
<td>7</td>
<td>+</td>
</tr>
<tr>
<td>RAG-1</td>
<td>Rch1-(83–529)</td>
<td>–</td>
<td>8</td>
<td>+</td>
</tr>
<tr>
<td>RAG-1</td>
<td>Rch1-(185–529)</td>
<td>–</td>
<td>8</td>
<td>+</td>
</tr>
<tr>
<td>RAG-1</td>
<td>pG4-5</td>
<td>–</td>
<td>8</td>
<td>+</td>
</tr>
<tr>
<td>10 proteins (see legend)</td>
<td>Rch1-(33–529)</td>
<td>–</td>
<td>16*</td>
<td>–</td>
</tr>
<tr>
<td>10 proteins</td>
<td>Rch1-(244–529)</td>
<td>–</td>
<td>15*</td>
<td>–</td>
</tr>
<tr>
<td>10 proteins</td>
<td>pG4-5</td>
<td>–</td>
<td>14*</td>
<td>–</td>
</tr>
</tbody>
</table>

Different Rch1 derivatives (amino acids present are indicated in parentheses) were tested for their ability to interact with the various LexA fusion proteins. An acidic activation construct not carrying a fusion partner (pG4-5 vector alone) is included in each set as a control. Strains carrying the indicated LexA or activation fusion proteins were tested in the presence of glucose or galactose for their ability to grow in the absence of leucine (Leu⁺). Growth was assessed after 4 days: + = growth, − = no growth. Additionally, β-galactosidase (LacZ) levels (average of three independent transformations, in standard units) were determined. The ratio of β-galactosidase activity under inducing (galactose) versus noninducing (glucose) conditions is indicated in the last column. The results with 10 different LexA fusion proteins (fusion partners: RAG-2, Myc, Max, Cdk2, Cdk3, Tr, Car, Fus3, Bcd, and Cdc2) were averaged, and this average value is indicated by . The highest ratio of LacZ expression in glucose compared with galactose, within this group of 10 proteins, was a value of 2.1 for LexA-Myc tested with Rch1-(244–529). Descriptions of the LexA fusion proteins used in addition to RAG-1 and RAG-2 can be found in ref. 12 and references therein, with the exception of Tr and Car, thymocyte receptor molecules which were gifts of Jee Lee (Ligand Pharmaceuticals) and David Moore (Massachusetts General Hospital, Boston).
found that RAG-2 failed to interact with Rch1. However, four independent Rch1 derivatives (representing different N-terminal deletions of Rch1) showed 10- to 20-fold increases in β-galactosidase expression in the presence of LexA-RAG-1. Finally, transcriptional activation indicative of the functional interaction of RAG-1 and Rch1 was dependent on the presence of LexA operators (data not shown).

**Rch1 Sequence.** The sequence of the longest human *RCH1* clone (Rch1-33-529) was determined (Fig. 1); it contains an open reading frame of 1491 bp capable of encoding a protein of molecular weight 55,000, a size consistent with the molecular weight observed by Western blot analysis with antibody directed against the influenza hemagglutinin epitope tag included in the library vector (data not shown). The sequence of three other *RCH1* derivatives was identical over the approximately 300 bp sequenced, suggesting that they all represented the same amino acid sequence as expected, from the DNA hybridization analysis. Because no obvious initial methionine was apparent in the sequence, it seemed likely that some N-terminal sequence was missing from the clone. The human *RCH1* cDNA was used as a probe to isolate a 2.2-kb cDNA from a mouse cDNA library derived from the pre-B cell 22D6 (1), and the N-terminal coding and 5' untranslated sequence was determined (Fig. 1). The first methionine in the sequence is embedded in a near consensus Kozak sequence and is downstream of the apparent translational initiation start. Because it is the identical authentic initial codon. Other portions of the mouse cDNA were sequenced and found to encode a protein so nearly identical to the predicted human Rch1 that the remaining sequence of the mouse cDNA was not determined. A composite sequence from the mouse and human data was used for sequence comparisons (see below). As was not surprising given the strong sequence similarity between Rch1 and yeast SRP1 (see below), *RCH1* hybridizing sequences were found in chicken and frog at a stringency comparable to that seen for *RAG1* (data not shown).

Sequence comparison with the BLAST program (21) revealed a striking degree of similarity between *RCH1* and a yeast gene, *SRP1*, that was identified as a suppressor of temperature-sensitive mutations in RNA polymerase I (22). These two protein sequences are 47% identical and, allowing for conservative substitutions, are 69% similar (Fig. 2). strikingly, both proteins contain eight nonidentical repeats of a 42-amino acid sequence (Fig. 2). In SRP1, two of eight repeats deviate from the 42-amino acid length, with one containing 43 amino acids and one 45. These same repeats in Rch1 are also longer, containing in both cases 43 amino acids, with the additional amino acids in the same position as in SRP1. One other repeat in Rch1 has an additional 5 amino acids. The amino acids showing the greatest degree of conservation between human Rch1 and yeast SRP1 are the same amino acids that are conserved between internal repeats. More importantly, individual repeat sequence nucleotides are numbered beginning with the first nucleotide of the mouse cDNA and continue through the overlapping region of the human cDNA. The predicted amino acid sequence is also indicated. The position of the first amino acid present in the encoded nucleotide sequence of the mouse cDNA was not determined.

**Fig. 1.** Nucleotide sequence of the *RCH1* cDNA and the predicted amino acid sequence of the encoded protein. The nucleotide sequence of the human Rch1 (33-529) cDNA used in the experiments presented in this paper is shown. In addition, the 5' untranslated sequence and first 96 nucleotides of the second transcriptional initiation start are shown. The first AUG indicated in the mouse sequence appears to be the initial methionine (see below). Nucleotides are numbered beginning with the first nucleotide of the mouse cDNA and continue through the overlapping region of the human cDNA. The predicted amino acid sequence is also indicated. The position of the first amino acid present in the encoded nucleotide sequence of the mouse cDNA was not determined.
glyceraldehyde-3-phosphate dehydrogenase

between repeats within each protein. Those amino acids not conserved between repeats are also less conserved between human and yeast. A consensus repeat based on the one proposed for SRPI (22), but including Rch1 sequence information, is shown in Fig. 2. In both SRPI and Rch1, the repeated sequences are flanked by sequences rich in charged residues, and the sequence similarity between the two proteins extends to these flanking regions as well. Taken together, these data strongly suggest that RCH1 is the human homologue of yeast SRPI.

Expression Pattern and Chromosomal Localization.

The expression of RCH1 in various cell types was determined by RNA hybridization. A single predominant RNA species with an approximate length of 2 kb was observed in all the cell lines examined, including the pre-B-cell lines 22D6 and 38B9, the mature B cells WEHI231 and MPCI1, the myeloma cell S194, the pre-T-cell 2017, the mature T cells rM11 and EL4, the fibroblastoid cell lines NIH-3T3 and 16B (an NIH-3T3 derivative expressing RAG-1 and RAG-2), and HeLa cervical carcinoma cells related to the cells from which the library was derived. The level of expression when normalized to human glyceraldehyde-3-phosphate dehydrogenase was comparable in all samples (representative data are shown in Fig. 3). Thus, as was expected given that RCH1 was isolated from a HeLa cell library, and unlike the RAG genes, RCH1 expression is not lymphoid-specific, nor is it confined to developmental stages where V(D)J recombination is occurring.

RCH1 was localized to human chromosome 17q23.1−23.3 (G. Duky, personal communication). No region of this chromosome is believed to be syntenic with mouse chromosome 16, where the scid mutation maps (24). Further, none of the CHO cell mutants described that affect both DNA repair and V(D)J recombination have defects mapping to human chromosome 17 (9). Thus, RCH1 does not appear to represent a gene previously implicated in V(D)J recombination.

A C-Terminal Fragment of Rch1 Inhibits VDJ Recombination in Vivo. We next asked whether Rch1 overexpression could affect VDJ recombinase activity induced in fibroblasts by transiently transfected RAG-1 and RAG-2 DNA. Expression of the nearly full-length Rch1 clone [Rch1-(33–529)] had little effect. Recombination frequencies were either slightly higher or the same as in controls (Fig. 4). This result might have been expected, given that HeLa cells already express Rch1 mRNA. However, we found that Rch1-(244–529), a truncated derivative that interacts with RAG-1, decreased the recombination frequency. Because the effect was not large, three different plasmid preparations of the truncated clone were used in seven independent transfections, each trial was done in duplicate or triplicate, and several different plasmid preparations of the full-length Rch1 clone and the CD8 control were also used. The decrease was repeatedly 2- to 3-fold compared with the vector-only control. The ratio of recombination frequencies for parallel transfections receiving Rch1-(33–529) versus Rch1-(244–529) (Fig. 4) indicates that overexpression of the truncated clone yields on average recombination frequencies ½ of those with the nearly full-length derivative.

DISCUSSION

We have identified a gene whose product interacts with the RAG-1 protein in the two-hybrid assay. The highly specific nature of the RAG-1–Rch1 interaction argues against an
FIG. 4. Expression of truncated Rch1 inhibits V(D)J recombination in HeLa cells. Hela cells were transiently transfected with RAG-1 and RAG-2 expression plasmids M2CD7 and R2IRC2D (1, 2), the pJH200 reporter construct (17), and either (i) CDM8 expression vector alone (hatched bars), (ii) Rch1(33–529) (empty bars), or (iii) Rch1(244–529) (filled bars) in expression vectors. Seven independent transfection assays are shown. For each assay, recombination frequencies are shown normalized to the CDM8 control and represent the average of duplicate or triplicate transfections. The ratio of recombination frequencies from cells transfected with Rch1(33–529) compared with those transfected with Rch1(244–529) is shown for each experiment in the line below the experiment number.

artefactual association of these two proteins. RCH clones were obtained at a very low frequency (<10^{-5}); only three genes were positive for interaction with RAG-1, and each of these genes was isolated repeatedly. Thus, RAG-1 protein does not simply associate nonspecifically with numerous human proteins. Further, the specificity tests indicate that Rch1 is likewise not promiscuous in its interactions.

The decreased recombination caused by a truncated derivative of Rch1 suggests that Rch1 is involved in V(D)J recombination. Although we cannot exclude the possibility that the truncated protein reduces recombination indirectly, it is important to note that Rch1 was identified by an entirely independent assay that involves its ability to specifically interact with RAG-1 in yeast. It seems unlikely that these two distinct properties of Rch1 are merely coincidental. Further, if Rch1 merely interacted with RAG-1 and prevented it from functioning normally, but played no role in V(D)J recombination, the full-length and truncated proteins should have had the same effect. Instead the nearly full-length and presumably functional protein does not inhibit the recombination activity, whereas the truncated protein acts as a dominant negative. Thus, the specific interaction between Rch1 and RAG-1 appears likely to be involved in the induction of V(D)J recombination.

Although the role of Rch1 in V(D)J recombination remains to be elucidated, the repeat structure of Rch1 and its homology with SRPI may provide useful clues. Biochemical fractionation indicates that SRPI is localized to the nucleus, and at least some of the protein appears to be associated with the insoluble nuclear matrix/lamina/pore fraction (22). Immunofluorescence microscopy of SRPI in yeast supports this conclusion, as a punctate staining pattern is seen near the nuclear envelope, perhaps at the nuclear pores (22). While no other known proteins have the same repeat structure as SRPI and Rch1, other proteins with multiple internal repeats (Schizosaccharomyces pombe nud2¹, S. cerevisiae Cdc23, and ankyrin and members of the ankyrin repeat family) are thought to form structures within the nucleus and to be associated with integral membrane proteins (discussed in ref. 22 and references therein). Although SRPI was isolated as a suppressor of temperature-sensitive mutations in RNA polymerase I, no biochemical evidence of a direct interaction between SRPI and RNA polymerase I was found, nor was SRPI found in the nucleolus (22). Thus, it was suggested that SRPI may play some structural role in the nucleus which then indirectly influences RNA polymerase I function. Interestingly, SRPI is an essential gene in yeast.

By analogy with SRPI, the Rch1 protein may be associated with matrix/membrane structures in the nucleus. The recent report that overexpressed RAG-1 gives a punctate staining pattern in tissue culture cells is interesting in this regard (26). It is appealing to speculate that Rch1 might serve to bring RAG-1 and other components of the recombinational apparatus together into an active complex associated with specific structures within the nucleus. The modification and rejoining of the cut ends during recombination could be facilitated if they were held in such a structure.

Note Added in Proof. It has recently been noted (27) that the repeated amino acid motif found in the armadillo protein of Drosophila is found in SRPI and a number of other proteins.

The authors thank F. Solomon, K. Struhl, and J. Kaplan for helpful discussions and L. Liebster and A. Mukherjee for excellent technical assistance. This work was supported by a postdoctoral fellowship from the Human Frontiers Science Program (J.G.), the Pew Scholar's program (M.A.O. and R.B.), and grants from the Human Frontiers Science Program (R.B.) and the National Institutes of Health (Grant GM48026; M.A.O.).