THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 272, No. 10, Issue of March 7, pp. 6416–6422, 1997
© 1997 by The American Society for Biochemistry and Molecular Biology, Inc.
Printed in U.S.A.

Molecular Definition of a Novel Human Galectin Which Is Immunogenic in Patients with Hodgkin's Disease*

(Received for publication, September 24, 1996, and in revised form, December 9, 1996)

Özlem Türeci, Holger Schmitt, Natalie Fadle, Michael Pfreundschuh‡, and Ugur Sahin

From the Medizinische Klinik und Poliklinik, Innere Medizin I, Universitätskliniken des Saarlandes, D-66421 Homburg, Federal Republic of Germany

Using autologous serum for the immunoscreening of a cDNA expression library derived from tissue involved by Hodgkin's disease, a new 36-kDa protein with the characteristics of galectins (S-type lectins) was detected. Sequence analysis of the cDNA clone HOM-HD-21 revealed two homologous motifs known as lectin domains with galactoside binding capacity. The two domains are linked by a stretch of about 30 amino acid residues and share a sequence homology of 39%. While the N-terminal lectin domain shows merely moderate homologies with known galectins, the C-terminal lectin domain is highly homologous to rat galectin-5 with an amino acid sequence identity of 70%. We ruled out mutations of the tumor-derived transcript by sequence comparison with the respective cDNA cloned from normal peripheral blood leukocytes. Recombinant protein expressed in Chinese hamster ovary cells was purified from lysates by lactose and galactose affinity chromatography, proving the galactoside binding capacity of this new galectin. Northern blot analysis revealed an expression spectrum restricted to peripheral blood leukocytes and lymphatic tissues. In accordance with the nomenclature of known galectins, we suggest to designate this novel galactoside binding protein galectin-9.

Galectins, formerly known as S-type lectins or S-Lac lectins (1, 2), are a growing family of soluble animal β-galactoside-binding proteins. Members of the galectin family are defined by two characteristic features: affinity to β-galactosides and a specific sequence motif called lectin domain. Based on the number of lectin domains two groups of galectins are distinguished. The majority of galectins, including galectin-1, -2, -3, -5, and -7 (3–9), have a single lectin domain. The second group of galectins, which includes rat galectin-4 (10), rat galectin-8 (11), and the 32-kDa galactoside-binding protein of *Caenorhabditis elegans* (12) is characterized by two tandemly arranged lectin domains connected by a linker peptide. Although none of the galectins contains a typical secretion signal, several galectins are externalized by nonclassical secretory mechanisms (13, 14) and play a role in modulating cell-to-cell or cell-to-matrix interactions. Galectins are involved in a number of different cellular events, including physiological (15, 16) and malignant cell adhesion (17, 18), activation and proliferation of immune cells (19, 20), as well as induction (21) and inhibition (22) of programmed cell death.

Recently, we established SEREX[1] (serological identification of antigens by recombinant expression cloning), a novel approach for the molecular definition of human tumor antigens using autologous serum from tumor patients. For SEREX, tumor-derived λ-phage expression libraries are screened for reactivity with high-titered IgG antibodies present in the autologous serum of the analyzed patient. By applying this method to several human neoplasms we identified numerous new tumor antigens (23, 24). In a cDNA library derived from the Hodgkin's disease-involved spleen of a 28-year-old female we detected four different antigens. The molecular and biochemical analysis of one of these antigens, which was encoded by the clone HOM-HD-21, revealed a novel human lectin, which shares the structural and functional properties of galectins. In conformity with the definition and the nomenclature of other known galectins (galectin 1–8) we propose to name this novel galactoside-binding protein galectin-9.

MATERIALS AND METHODS

Sera and Tissues—The study had been approved by the local ethical review board ("Ethikkommission der Ärztekammer des Saarlandes", Saarbrücken, Germany). Recombinant DNA work was done with the official permission and according to the rules of the state government of Saarland. Sera and tumor tissues were obtained during routine diagnostic or therapeutic procedures and stored at −80 °C until use. Normal tissues were collected from autopsies of tumor-free patients.

Construction of cDNA Expression Libraries—The construction of the Hodgkin's-derived cDNA expression library has been described elsewhere (23). In brief, a oligo(dT)-primed cDNA expression library resulting in 3.2×10^6 primary clones was established by directional cloning of the cDNA derived from the infiltrated spleen of a 28-year-old woman with a nodular sclerosis type of Hodgkin's disease into the *Eco*RI- and *Xho*I-digested λZAPII phage (Stratagene, Heidelberg, Germany).

Immunoscreening of Transfectants—Modifications of the previously described technique (23, 24) were implemented to circumvent the detection of false positive clones encoded by IgG-heavy chain transcripts derived from the numerous B-lymphocytes present in the spleen. After transfection for primary screening, plaques were transferred onto nitrocellulose membranes. After blocking with 5% (w/v) low-fat milk-TBS, nitrocellulose membranes were preincubated with an alkaline phosphatase-conjugated antibody specific for human IgG for 1 h. Reactive clones representing expressed IgG heavy chains were visualized by staining with alkaline phosphatase substrate and marked with a pencil. These prestained membranes were washed extensively with TBS, incubated with autologous patient's serum (1:1000), and the normal immunoscreening procedure was performed. Those plaques which appeared positive but had no pencil marks were considered as positives and subjected to retests and monoclonalization.

Sequence Analysis of Identified Antigens—Reactive clones were sub-

* 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.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) Z49107.

‡ To whom correspondence should be addressed: Medizinische Klinik I, Universität des Saarlandes, D-66421 Homburg, Federal Republic of Germany. Tel.: 49-6841-16-3002; Fax: 49-6841-16-3101; E-mail: ugursahin@aol.com.

[1] The abbreviations used are: SEREX, serological identification of antigens by recombinant expression cloning; CHO, Chinese hamster ovary; PBS, phosphate-buffered saline; TBS, Tris-buffered saline, H&RS, Hodgkin-Reed-Sternberg; Ni-NTA, nickel-nitrilotriacetic acid; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; MOPS, 4-morpholinepropanesulfonic acid; bp, base pair(s).

This paper is available on line at http://www-jbc.stanford.edu/jbc/

This is an Open Access article under the CC BY license.

cloned to monoclonality and submitted to *in vivo* excision (25) of pBluescript phagemids (26). Sequencing of cDNA inserts was carried out using a Sequenase 2.0 kit (U. S. Biochemical Corp., Bad Homburg, Germany) with vector-specific reverse and universal M13 primers according to the manufacturer's instructions. Specific internal oligonucleotides were designed as the sequencing progressed. Sequence alignments were performed with DNASIS (Pharmacia Biotech Inc.) and BLAST (27) softwares on EMBL (28) and GenBank™ (29) (Release 27.11.96) and PROSITE (30) data bases.

Northern and Southern Blot Analysis—Northern blots were performed with RNA extracted from tumors and normal tissues using guanidium thiocyanate as a chaotropic agent (31). RNA integrity was checked by electrophoresis in formalin/MOPS gels. Gels containing 10 μg of RNA/lane were blotted onto nylon membranes (Hybond N, Amersham Corp.). After prehybridization the membranes were incubated overnight at 42 °C in hybridization solution (50% formamide, 6 × SSC, 5 × Denhardt's, 0.2% SDS) with a ^{32}P-labeled specific full-length probe. The membranes were washed at progressively higher stringency, with the final wash in 1 × SSC and 0.2% SDS at 65 °C for 20 min. Autoradiography was conducted at −70 °C for 2 days using Kodak X-Omat-AR film and intensifying screen. After exposure the filters were stripped and rehybridized with glyceraldehyde 3-phosphate dehydrogenase to prove RNA integrity. Densitometry of autoradiographs was performed to compare the expression ratios of different tissues using a XRS scanner and whole band analyzer software (BioImage, Ann Arbor, MI). For Southern blot analysis 4 μg of human DNA digested with *Eco*RI, *Hin*dIII, or *Bam*HI, respectively, was blotted and hybridized with a probe specific for the C-terminal lectin domain.

Cloning of Human Galectin-9 cDNA from Normal Peripheral Blood Leukocytes—Total RNA was isolated from peripheral blood leukocytes obtained from buffy coats by density gradient centrifugation (Ficollpaque, Pharmacia, Freiburg, Germany). First strand cDNA was synthesized from 10 μg of total RNA with a $(dT)_{18}$-oligonucleotide and Superscript reverse transcriptase (Life Technologies, Inc., Eggenstein, Germany). For amplification of human galectin-9 cDNA, transcript-specific oligonucleotides comprising the entire open reading frame were used. PCR was performed for 35 cycles with an annealing temperature of 68 °C under standard conditions with *Tth* Polymerase (Goldstar, Eurogentec). The resulting 978-bp product was cloned into TA-cloning vector and sequenced.

Production of His-tagged Recombinant Protein for Immunization—Full-length human galectin-9 protein turned out to be toxic and not expressable in *Escherichia coli*. A 245-bp cDNA fragment coding for the first 81 N-terminal amino acids of galectin-9 was amplified using the identified HOM-HD-21 cDNA as template and the primers (5′-GCC TTC AGC GGT TCC CAG GCT CCC TAC-3′) and (5′-CCC AGC TTC CGT GCC TCG TGT TGG ACA-3′) as sense or antisense oligonucleotides, respectively. The PCR with *Pfu* polymerase (Stratagene) was conducted for 20 cycles with an annealing temperature of 60 °C. The blunt-end PCR product was gel-purified using the QiaEx II kit (Qiagen) and ligated in frame to *Sma*I-digested, dephosphorylated, and gel-purified pQE32 vector (Qiagen), allowing for the translation of a fusion protein bearing a 6-histidine tail at the N terminus. To express the His-tagged recombinant protein, the construct was transformed into *E. coli* SG13009 (pREP4) strain and selected on kanamycin/ampicillin-containing plates. Several colonies were picked, and the production of recombinant protein was induced in small scale by adding 2 mM isopropyl-β-D-thiogalactopyranoside to the culture medium in order to check for protein expression. Small scale purification over Ni-NTA-columns was performed for each clone. One clone coding for a protein of the expected length was selected and verified by sequence analysis. Subsequently, large scale induction of recombinant protein was performed. Cells were harvested 5 h after induction with 2 mM isopropyl-β-D-thiogalactopyranoside. Lysis was performed in buffer A (8 M urea, 100 mM Na_2PO_4, 10 mM Tris-HCl, pH 8, 0,01% Triton X-100) overnight. Debris was spun down, and supernatant was loaded onto preequilibrated Ni-NTA-resin. Washes were performed with 2 volumes of buffer A, pH 8.0, and at least 10 volumes of buffer A, pH 6.3. Elution was performed with 250 mM imidazol in buffer A. The yield of affinity-purified His-tagged protein ranged from 15 to 40 mg/liter of bacterial culture. Protein quantification was done by the Bradford procedure (32) following the recommendations of the manufacturer (Bio-Rad).

Production and Affinity Purification of Polyclonal Rabbit Antisera—Polyclonal rabbit antisera were obtained from a custom antibody service (Eurogentec, Brussels, Belgium). A total of four immunizations with 100 μg of purified His-tagged fusion protein per immunization were performed. Sera before and after immunization were tested by Western blot for reactivity with the fusion protein. Serum obtained after the final boost was submitted to affinity purification. Briefly, His-tagged protein in buffer A was loaded onto Ni-NTA-resin. Washes with sequentially decreasing molarities of urea in buffer A were performed to equilibrate the column with 100 mM Na_2PO_4, 10 mM Tris-HCl, pH 8, 0.01% Triton X-100, pH 7.4. After several washes polyclonal serum was eluted with 4 M $MgCl_2$, dialyzed against distilled H_20 for 1 h and against phosphate-buffered saline (PBS) overnight. Sensitivity and specificity of the affinity-purified rabbit serum was tested using purified His-tagged protein and CHO/HD21-lysate as positive controls and two unrelated His-tagged proteins (MAGE-1 and HOM-Mel 40) and CHO/pcDNA3 as negative controls. Affinity-purified serum was used for all experiments.

Western Blot Analysis—Rabbit antiserum obtained after immunization with the N-terminal fragment of human galectin-9 was used to detect protein expression in cell lysates. Samples of 2 μg of recombinant protein and 20 μg of cell lysates, respectively, were mixed with 2 × SDS buffer (0.1 M Tris-HCl, pH 6.8, 0.2 M dithiothreitol, 4% SDS, 0.2% bromphenol blue, and 20% glycerol), electrophoresed in 12% SDS-PAGE and then blotted onto nylon membranes (Schleicher & Schüll) by semi-dry transfer (Bio-Rad). After blocking unspecific binding with 5% low-fat milk in TBS for 1 h, the membranes were incubated with 1:100 diluted anti-galectin-9 rabbit serum. The blots were then incubated for 1 h with alkaline phosphatase-conjugated mouse anti-rabbit IgG (Dianova). The membranes were consecutively incubated for 30 min with rabbit anti-mouse Ig (Dianova) as bridging antibody, with anti-alkaline phosphatase and with 0.25 mg/ml alkaline phosphatase. After each incubation step the membranes were washed extensively in TBS and 0.01% Tween 20. Visualization of positive reactions was performed by staining with 5-bromo-4-chloro-3-indolyl phosphate and nitro blue tetrazolium according to the manufacturer's instructions (Bio-Rad).

Transfection and Expression of Human Galectin-9 in Chinese Hamster Ovary Cells—The HOM-HD-21 cDNA insert was excised from the pBluescript-SK phagemid using restriction digestion with *Eco*RI and *Xho*I and was ligated into the *Eco*RI and *Xho*I cut, gel-purified eucaryotic expression vector pcDNA3 (Invitrogen). The ligation product was transformed in *E. coli* TOP-10, and plasmids were purified on a silica gel matrix (Qiagen) after alkaline lysis of bacteria. 5 μg of pcDNA3 plasmid containing the ligated HOM-HD-21 fragment were used to transfect CHO cells. As a control CHO cells were transfected with pcDNA3 plasmid containing no insert. Liposome-mediated transfection was carried out in RPMI 1640 medium with 1% fetal calf serum using N-[1-(2,3-dioleoyloxy)propyl]-*N,N,N*-trimethylammonium methyl sulfate (DOTAP) (Boehringer Mannheim) according to the manufacturer's instructions. Transfected cells were designated CHO/HD21 or CHO/pcDNA, respectively, and were selected by a 4-week culture in RPMI 1640 medium containing 10% fetal calf serum and 250 μg/ml G418 (Sigma). Expression of human galectin-9 was analyzed by Western blot using rabbit serum.

Galactoside Affinity Purification—1×10^7 CHO/HD21 and CHO/pcDNA cells were lysed in a buffer containing 100 mM Tris-HCl, pH 8, 100 mM NaCl, 2 mM EDTA, 1% Triton X-100, 1 mM phenylmethylsulfonyl fluoride (Sigma). The lysate was loaded on a lactosyl-Sepharose column (Sigma). Unbound proteins were removed by extensively washing first with buffer B containing 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, and subsequently with PBS. Galactoside-binding proteins which had bound to the column were specifically eluted with 150 mM lactose-containing PBS. Samples of all fractions were analyzed by SDS-PAGE and subsequent silver staining according to the manufacturer's instructions (SilverPage, Biometra).

RESULTS

Identification of Human Galectin-9 and Sequence Analysis—Applying the modified immunoscreening as described under "Materials and Methods" to 1.0×10^6 recombinant plaques, 14 positive clones representing four different transcripts were identified. A group of six clones, designated HOM-HD-21, HOM-HD-202, HOM-HD-297, HOM-HD-303, HOM-HD-402, and HOM-HD-415 represented the same transcript. Restriction enzyme mapping and 5′ and 3′ sequencing revealed that these clones had inserts of different length ranging from 1586 to 1634 bp. The clone with the longest insert HOM-HD-21 (Fig. 1) had a 5′-untranslated region of 70 bp, an N-terminal methionine complying with the general features of a translation initiation site (33), followed by a 969-bp open reading frame coding for a putative protein with a predicted molecular mass

```
  49  GGT GTG AAA GGC AGC GGT GGC CAC AGA GGC GGC GGA GAG ATG GCC TTC    96
   1                                                       M   A   F     3

  97  AGC GGT TCC CAG GCT CCC TAC CTG AGT CCA GCT GTC CCC TTT TCT GGG   144
   4   S   G   S   Q   A   P   Y   L   S   P   A   V   P   F   S   G    19

 145  ACT ATT CAA GGA GGT CTC CAG GAC GGA CTT CAG ATC ACT GTC AAT GGG   192
  20   T   I   Q   G   G   L   Q   D   G   L   Q   I   T   V   N   G    35

 193  ACC GTT CTC AGC TCC AGT GGA ACC AGG TTT GCT GTG AAC TTT CAG ACT   240
  36   T   V   L   S   S   S   G   T   R   F   A   V   N   F   Q   T    51

 241  GGC TTC AGT GGA AAT GAC ATT GCC TTC CAC TTC AAC CCT CGG TTT GAA   288
  52   G   F   S   G   N   D   I   A   F   H   F   N   P   R   F   E    67

 289  GAT GGA GGG TAC GTG GTG TGC AAC ACG AGG CAG AAC GGA AGC TGG GGG   336
  68   D   G   G   Y   V   V   C   N   T   R   Q   N   G   S   W   G    83

 337  CCC GAG GAG AGG AGG ACA CAC ATG CCT TTC CAG AAG GGG ATG CCC TTT   384
  84   P   E   E   R   R   T   H   M   P   F   Q   K   G   M   P   F    99

 385  GAC CTC TGC TTC CTG GTG CAG AGC TCA GAT TTC AAG GTG ATG GTG AAC   432
 100   D   L   C   F   L   V   Q   S   S   D   F   K   V   M   V   N   115

 433  GGG ATC CTC TTC GTG CAG TAC TTC CAC CGC GTG CCC TTC CAC CGT GTG   480
 116   G   I   L   F   V   Q   Y   F   H   R   V   P   F   H   R   V   131

 481  GAC ACC ATC TTC GTC AAT GGC TCT GTG CAG CTG TCC TAC ATC AGC TTC   528
 132   D   T   I   F   V   N   G   S   V   Q   L   S   Y   I   S   F   147

 529  CAG CCT CCC GGC GTG TGG CCT GCC AAC CCG GCT CCC ATT ACC CAG ACA   576
 148   Q   P   P   G   V   W   P   A   N   P   A   P   I   T   Q   T   163

 577  GTC ATC CAC ACA GTG CAG AGC GCC CCT GGA CAG ATG TTC TCT ACT CCC   624
 164   V   I   H   T   V   Q   S   A   P   G   Q   M   F   S   T   P   179

 625  GCC ATC CCA CCT ATG ATG TAC CCC CAC CCC GCC TAT CCG ATG CCT TTC   672
 180   A   I   P   P   M   M   Y   P   H   P   A   Y   P   M   P   F   195

 673  ATC ACC ACC ATT CTG GGA GGG CTG TAC CCA TCC AAG TCC ATC CTC CTG   720
 196   I   T   T   I   L   G   G   L   Y   P   S   K   S   I   L   L   211

 721  TCA GGC ACT GTC CTG CCC AGT GCT CAG AGG TTC CAC ATC AAC CTG TGC   768
 212   S   G   T   V   L   P   S   A   Q   R   F   H   I   N   L   C   227

 769  TCT GGG AAC CAC ATC GCC TTC CAC CTG AAC CTC CGT TTT GAT GAG AAT   816
 228   S   G   N   H   I   A   F   H   L   N   L   R   F   D   E   N   243

 817  GCT GTG GTC CGC AAC ACC CAG ATC GAC AAC TCC TGG GGG TCT GAG GAG   864
 244   A   V   V   R   N   T   Q   I   D   N   S   W   G   S   E   E   259

 865  CGA AGT CTG CCC CGA AAA ATG CCC TTC GTC CGT GGC CAG AGC TTC TCA   912
 260   R   S   L   P   R   K   M   P   F   V   R   G   Q   S   F   S   275

 913  GTG TGG ATC TTG TGT GGA GCT CAC TGC CTC AAG GTG GCC GTG GAT GGT   960
 276   V   W   I   L   C   G   A   H   C   L   K   V   A   V   D   G   291

 961  CAG CAC CTG TTT GAA TAC TAC CAT CGC CTG AGG AAC CTG CCC ACC ATC  1008
 292   Q   H   L   F   E   Y   Y   H   R   L   R   N   L   P   T   I   307

1009  AAC AGA CTG GAA GTG GGG GGC GAC ATC CAG CTG ACC CAT GTG CAG ACA  1056
 308   N   R   L   E   V   G   G   D   I   Q   L   T   H   V   Q   T   323

1057  TAG GCG GCT TCC TGG CCC TGG GGC CGG GGG CTG GGG TGT GGG GCA GTC  1104
 324   *

1105  TGG GTC CTC TCA TCA TCC CCA CTT CCC AGG CCC AGC CTT TCC AAC CCT  1152

1153  GCC TGG GAT CTG GGC TTT AAT GCA GAG GCC ATG TCC TTG TCT GGT CCT  1200

1201  GCT TCT GGC TAC AGC CAC CCT GGA ACG GAG AAG GCA GCT GAC GGG GAT  1248
```

FIG. 1. **Nucleid acid and amino acid sequence of human galectin-9.** The predicted open reading frame in the nucleic acid sequence was translated into the amino acid sequence below. The stop codon is marked by an *asterisk*. The putative polyadenylation signal sequence is *underlined*.

of 35,918 Da and a long 3′-untranslated region of 595 bp. Sequences of all analyzed clones were identical. Length variation turned out to be due to incompleteness of the 5′ ends resulting from premature stops of the cDNA synthesis. The alignment of the predicted amino acid sequence with data bases demonstrated high homologies exclusively with the members of the galectin family and revealed two domains of 140–150 amino acids in head-to-tail orientation with a mutual homology of 39% linked by a stretch of 30 amino acids. Both domains contained sequence motifs that are conserved in the carbohydrate binding region of galectins (Fig. 2*a*; Ref. 2). While the N-terminal domain showed a moderate overall homology to other known galectins, the C-terminal domain had a 70% amino acid homology to rat galectin-5 (Fig. 2*b*). Considering

1249	TGC CTT CCT CAG CCG CAG CAG CAC CTG GGG CTC CAG CTG CTG GAA TCC	1296
1297	TAC CAT CCC AGG AGG CAG GCA CAG CCA GGG AGA GGG GAG GAG TGG GCA	1344
1345	GTG AAG ATG AAG CCC CAT GCT CAG TCC CCT CCC ATC CCC CAC GCA GCT	1392
1393	CCA CCC CAG TCC CAA GCC ACC AGC TGT CTG CTC CTG GTG GGA GGT GGC	1440
1441	CTC CTC AGC CCC TCC TCT CTG ACC TTT AAC CTC ACT CTC ACC TTG CAC	1488
1489	CGT GCA CCA ACC CTT CAC CCC TCC TGG AAA GCA GGC CTG ATG GCT TCC	1536
1537	CAC TGG CCT CCA CCA CCT GAC CAG AGT GTT CTC TTC AGA GGA CTG GCT	1584
1585	CCT TTC CCA GTG TCC TTA AAA TAA AGA AAT GAA AAT GCT TGT TGG CAA	1632
1633	AAA AAA AAA AAA AAA AAA A	1651

FIG. 1—*continued*

conserved amino acid exchanges the homology of this domain to rat galectin-5 raised to 93%. As described for other galectins, there was apparently no typical secretion signal in the peptide sequence of HOM-HD-21. Alignment of the linker peptide attaching both lectin domains revealed no significant homology to the linker peptides of other galectins. Three putative glycosylation sites were found at Asn^{34}, Asn^{79}, and Asn^{137}.

Production of Polyclonal Rabbit Antisera by Immunization with Recombinant Human Galectin-9 His-tagged Protein—The procaryotic expression of a 5′-fragment of HOM-HD-21 cDNA as His-tagged fusion protein yielded a product with an apparent molecular mass of 12 kDa, consistent with the expected molecular mass (Fig. 3). The product was purified by nickel-chelate affinity chromatography and used for the production of rabbit immune sera. Serum collected before and after immunization was tested in Western blot against the recombinant protein used for immunization and two unrelated recombinant His-tagged proteins. The immune serum had a specific reactivity with galectin-9 protein (data not shown). For further analysis the serum was affinity-purified using immobilized His-tagged antigen bound onto the nickel-chelate column.

Galactoside Binding Activity of Recombinant Human Galectin-9 Protein Expressed in CHO Cells—To confirm the galactoside binding activity of eucaryotically expressed galectin-9, CHO cells showing no detectable endogenous galectin-9 expression in Northern blot and Western blot were chosen. CHO/HD21 and CHO/pcDNA (negative control) cells were passaged in selection medium containing G418. After 3 weeks the expression of a specific 35-kDa protein was demonstrated in Western blots using rabbit antiserum raised against the N-terminal 81 amino acids of human galectin-9 (Fig. 4). The protein is comigrating with its putative counterpart derived from density gradient separated peripheral blood leukocytes (data not shown). No additional smaller bands were detected, implying that the two lectin domains of the molecule are not cleaved by endogenous proteases. Extracts from CHO/HD-21 and from CHO/pcDNA cells were loaded onto columns with immobilized galactose and lactose. After washing, competitive elution was performed with galactose and lactose, respectively. Silver staining of the initial cell lysate and the eluted fraction (Fig. 5) demonstrated binding of the 35-kDa protein to galactose and lactose. The specificity of the eluted protein was demonstrated with the rabbit antiserum raised against the N-terminal fragment of galectin-9 (data not shown). The binding to the galactose column was apparently weaker than to the lactose column, resulting in a higher portion of recombinant protein in the flow-through (data not shown). This is consistent with the nearly 100-fold higher affinity of galectins to lactose as compared with galactose (34). No binding of proteins was immunodetected using lysates from CHO/pcDNA cells transfected with vector lacking the human galectin-9 cDNA.

Human Galectin-9 from Hodgkin's-involved Tissue Is Not Mutated—To exclude mutations of the tumor-derived human galectin-9 transcript as a reason for immunogenicity in the autologous host, we cloned the respective cDNA from normal peripheral blood leukocytes by reverse transcription PCR. Sequencing of the entire open reading frame revealed no differences to the tumor-derived cDNA, ruling out mutations as the cause for the generation of the detected antibody response.

Expression Spectrum of Galectin-9—The expression of human mRNA was analyzed by Northern blot hybridization using 20 μg of total RNA blotted onto nylon membranes. We detected a moderate expression of a 1.7-kilobase transcript in peripheral blood leukocytes, lymph nodes, and tonsils. No expression was detected by Northern blot in several tissues, including breast, kidney, brain, skeletal muscle, skin, testis, and stomach. A very weak expression signal detected in some of the tissue samples derived from colon and lung was most likely caused by resident leukocytes. We detected a high expression of galectin-9 transcripts in the Hodgkin's-diseased spleen used for library construction and in two lymph node samples derived from other Hodgkin's patients. The densitometric determined expression level in each of the Hodgkin's disease involved tissues was at least 10-fold higher than the level of expression found in normal lymphatic tissues (Fig. 6).

Southern blot analysis with probes specific for the C-terminal lectin domain revealed hybridization of the probe with at least two distinct bands for each DNA restriction digestion (Fig. 7).

DISCUSSION

Hodgkin's disease is a complex lymphoproliferative malignant disorder. The histological diagnosis is based on the presence of Hodgkin-Reed-Sternberg (H&RS) cells surrounded by a cellular infiltrate composed of reactive lymphocytes, plasma cells, histiocytes, neutrophils, eosinophils, and stroma cells (35). The mononucleated Hodgkin and the multinucleated giant H&RS cells usually represent less than 1% of the cellular population of an involved tissue, although they are presumed to be the neoplastic cell population. Specific interactions of H&RS cells with the surrounding lymphocytes mediated by characteristic profiles of cytokines and growth factors have been reported. Hodgkin's disease is often associated with impaired immune functions (36). Aberrations of the humoral immunity include the presence of immune complexes in sera of patients with Hodgkin's disease. Previous investigations revealed that these immune complexes contain antigens that are present in the cytoplasm of H&RS cells (37). However, the nature of these antigens has remained undefined. Using the newly established approach of SEREX, we succeeded in identifying four different antigens in a Hodgkin's-derived cDNA expression library (23, 24). One of these 9 antigens, initially named HOM-HD-21, turned out to be a novel human galectin. It has two homologous lectin domains separated by a linker peptide. This structure is

a

```
HGalect9N      MAFSGSQAPYLSPAVPFSGTIQGGLQDGLQIT    32
hGalect9C      PAIPPMMYPHPAYPMPFITTILGGLYPSKSIL   210
rGalect5       MSSFSTQTPYPNLAVPFFTSIPNGLYPSKSIV    32
hGalect7               MSNVPHKSSLPEGIRPGTVLR       21
rGalect8N      LSLSNLQNIIYNPTIPYVSTITEQLKPGSLIV    33
rGalect4N      YVPAPGYQPTYNPTLPYKRPLPGGLSVGMSIY    35
rGalect3       PATGPFGAPTGPLTVPYDMPLPGGVMPRMLIT   145

hGalect9N      VNGTVLSSSGTRFAVNFQTGF---SGNDIAFH    61
hGalect9C      LSGTVL-PSAQRPHINL------CSGNHIAFH   235
rGalect5       ISGVVL-SDAKRFQINL------RCGGDIAFH    57
hGalect7       IRGLV-PPNASRFHVNLLCGEEQ--GSDAALH    50
rGalect8N      IRGHVPKDSE-RFQVDFQHGNSLKPRADVAFH    64
rGalect4N      IQG-IAKDNMRRFHVNFAVGQD--EGADIAFH    64
rGalect3       LIGTV-KPNANSITLNFK------KGNDIAFH   170

hGalect9N      FNPRF--EDGGYVVCNTRQNGSWGPEERRTH-    90
hGalect9C      LNLRF---DENAVVRNTQIDNSWGSEERSLPR   264
rGalect5       LNPRF---DENAVVRNTQINNSWGPEERSLPG    86
hGalect7       FNPRL---DTSEVVFNSKEQGSWGREER-GPG    78
rGalect8N      FNPRF--KRSNGIVCNTLTNEKWGWEEITHDM    94
rGalect4N      FNPRF--DGWDKVVFNTMQSGQWGKEEKKKS-    93
rGalect3       FNPRFNENNRRVIVCNTKQDNNWGREERQSA-   201

hGalect9N      -MPFQKGMPFDLCFLVQSSDFKVMVNGILFVQ   121
hGalect9C      KMPFVRGQSFSVWILCGAHQLKVAVDGQHLFE   296
rGalect5       SMPFSRGQRFSVWILCEGHCFKVAVDGQHICE   118
hGalect7       GVPFQRGQPFEVLIIASDDGFKAVVGDAQYHH   109
rGalect8N      -MPFRKEKSFELVIMVLKNKFHVAVNKHIILL   125
rGalect4N      -MPFQKGHHFELVFMVMSEHYKVVVNGTPFYE   124
rGalect3       -FPFESGKPFKLQVLVEADHFKVAVNDVHLLQ   232

hGalect9N      YFHRVPFHRVDTIFVNGSVQLSYISFQPP      150
hGalect9C      YYHRLRNLPTINRLEVGGDIQLTHVQT        323
rGalect5       YSHRLMNLPDINTLEVAGDIQLTHVET        145
hGalect7       FRHRLPLARVRLVEVGGDVQLDSVRIF        136
rGalect8N      YAHRLNPEKIDTLGIFGKVNIHSIGF--       152
rGalect4N      YGHRLPLQMVTHLQVDGDLELQSINFLGG      153
rGalect3       YNHRMKNLREISQLGIIGDITLTSASHAM      261
```

b

```
hGalect9C    FSTPAIPPMMYPHPAYPMPFITTILGGLYPSKSILLSGTVLP  217
rGalect5     MSSFSTQTP-YPNLAV--PFFTSIPNGLYPSKSIVISGVVLS   39

hGalect9C    SAQRFHINLCSGNHIAFHLNLRFDENAVVRNTQIDNSWGSEER  260
rGalect5     DAKRFQINLRCGGDIAFHLNPRFDENAVVRNTQINNSWGPEER   82

hGalect9C    SLPRKMPFVRGQSFSVWILCGAHQLKVAVDGQHLFEYYHRLRN  303
rGalect5     SLPGSMPFSRGQRFSVWILCEGHCFKVAVDGQHIGEYSHRLMN  125

hGalect9C    LPTINRLEVGGDIQLTHVQT 323
rGalect5     LPDINTLEVAGDIQLTHVET 145
```

FIG. 2. *a,* alignment of N-terminal domain of human galectin-9 (*hgalect9N*) with its C-terminal domain (*hgalect9C*) as well as with galectins from both human (*h*) and rat (*r*) tissues. These include rat galectin-3 (5), the N-terminal domain of rat galectin-4 (10), rat galectin-5 (8), human galectin-7 (9), and the N-terminal domain of rat galectin-8 (11). *Dashes* represent gaps to aid better alignment. Residues with shared identity or similarity (*italic letters*) are *boxed*. *b*, sequence alignment of the C-terminal domain of human galectin-9 (*hGalect9C*) and rat galectin-5. *Dashes* represent gaps to aid better alignment. *Boxed* residues are identical in human galectin-9 and rat galectin-5.

similar to the previously described members of the galectin family consisting of rat galectin-4 (10), rat galectin-8 (11), and *C. elegans* galectin (12).

Two lines of evidence demonstrate that HOM-HD-21 cDNA codes indeed for a novel galectin: first, its deduced amino acid sequence contains two domains with conserved motifs that are implicated in the carbohydrate binding of galectins; second, eucaryotically expressed recombinant protein is biologically active and possesses sugar binding activity. Based on the nomenclature (1) for the previously described galectins, we suggest to designate this new transcript galectin-9.

The two carbohydrate binding domains of galectin-9 share an amino acid sequence homology of 39% to each other. This is similar to the other previously defined two lectin domain ga-

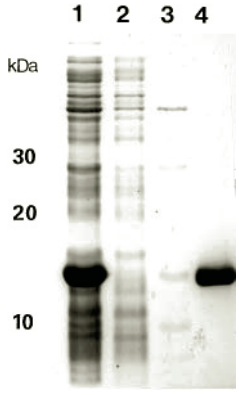

FIG. 3. **Procaryotic expression of a N-terminal fragment of human galectin-9.** A cDNA coding for a His-tagged 81-amino acid N-terminal fragment of galectin-9 was expressed in *E. coli* as described under "Material and Methods." Lysates of bacteria induced for expression of recombinant protein (*lane 1*) and uninduced control (*lane 2*) were transmitted to nickel-chelate affinity purification. Flow-through (*lane 3*) and specific eluate (*lane 4*) were analyzed by Coomassie staining after electrophoresis. In each lane with the exception of *lane 3*, 10 μg of protein were loaded.

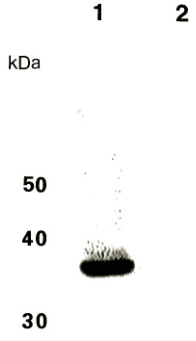

FIG. 4. **Western blot detection of human galectin-9 recombinantly expressed in trasnfected CHO-cells.** Lysates of CHO cells transfected with human galectin-9 cDNA cloned into pcDNA3 (*lane 1*) or pcDNA-vector alone (*lane 2*) were analyzed by immunoblot with the 1:100 diluted antiserum raised against an 81-amino acid N-terminal fragment of human galectin-9.

lectins, rat galectin-4 and rat galectin-8, which also have a limited interdomain homology of 35 and 33%, respectively. This limited homology may indicate that the two carbohydrate binding domains may recognize different ligands. The linker peptide sequence of human galectin-9 demonstrates no significant homology to the linker peptides of the other galectins. The same holds true for the previously described two-carbohydrate domain galectins. This may indicate that the structural sequence requirements for the functionality of such a linker peptide are limited.

The C-terminal lectin domain has a high homology to rat galectin-5. Although the published sequence of rat galectin-5 cDNA indicates only one lectin domain (8), splice variants with two lectin domains resident at the same gene locus cannot be excluded, since the genomic clone for rat galectin-5 has not been published to date. In addition, a human analogue of rat galectin-5 has not been identified yet. On this background the C-terminal lectin domain of human galectin-9 might be encoded by the gene for the human counterpart of rat galectin-5. To test this hypothesis we performed a Southern blot hybridization with genomic DNA obtained from human tissue using the C-terminal lectin domain of human galectin-9 as a probe. The finding that besides a strong unique hybridization signal at least one additional signal with weaker intensity was de-

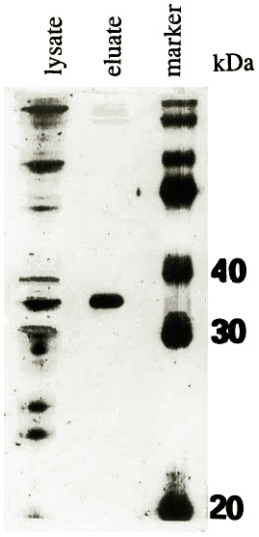

FIG. 5. **Specific binding of recombinant human galectin-9 to immobilized lactose.** Lysate of CHO cells transfected with human galectin-9 cDNA cloned into pcDNA3 was passed through a sephacryl column with immobilized lactose. The column was washed extensively with PBS. Carbohydrate-binding proteins were eluted specifically with buffer containing 150 mM lactose. Samples of the lysate and eluate were analyzed by SDS-PAGE and subsequent silver staining.

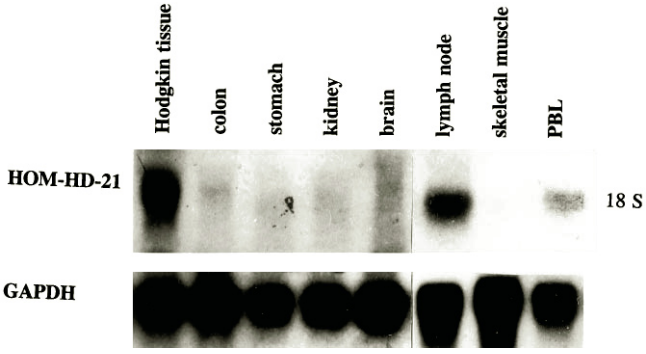

FIG. 6. **Northern blot analysis of human galectin-9 expression.** 20 μg of total RNA extracted from different human tissues were probed with radiolabeled HOM-HD-21 cDNA. Blots were stripped and retested after autoradiography with a radiolabeled glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) cDNA to check for integrity of RNA. The lane designated by *Hodgkin tissue* was loaded with RNA from the involved spleen of a patient with Hodgkin's disease.

tected (Fig. 7) does contradict this speculation. Furthermore, since the 70% homology of human galectin-9 protein to rat galectin-5 is significantly lower than the homology of other rat galectins to their human counterparts (90% for galectin-1, 81% for galectin-3), it appears more likely that the gene encoding human galectin-9 is not the human counterpart to rat galectin-5.

As reported recently, antibodies reactive with human galectin-9 were detected in about 50% of the sera derived from patients with Hodgkin's disease, but not in the sera of healthy individuals or patients suffering from other tumors (23), suggesting that the antibodies might have been generated as a tumor-specific response. Potential causes for antibody responses may be mutations, neoexpression of viral epitopes, overexpression, or reexpression of embryofetal proteins which are normally silenced in adult tissue. By sequencing we could rule out mutational alterations of the transcript derived from the Hodgkin's spleen cDNA library as initiators for the antibody response. Northern blot studies revealed a very limited expression pattern restricted to peripheral blood leukocytes and lymphatic tissues. Furthermore, a significantly higher ex-

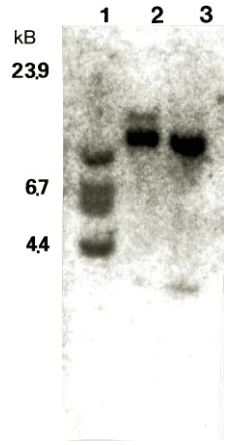

FIG. 7. **Southern blot hybridization of human genomic DNA with a radiolabeled cDNA from the C-terminal lectin domain of human galectin-9.** Human genomic DNA digested with *Eco*RI (*lane 1*), *Hin*dIII (*lane 2*), and *Bam*HI (*lane 3*) were hybridized with a radiolabeled cDNA probe. Autoradiography demonstrated one strong band and at least one additional band for each digestion.

pression of the galectin-9 mRNA in tissues involved by Hodgkin's disease was observed. This suggests that loss of tolerance and the development of a strong humoral immune response were initiated by overexpression of the galectin in the tumor tissue. The functional significance of the respective antibodies must be evaluated in future studies.

Although its function is not yet known with certainty, the restricted expression of the new galectin to lymphatic tissues also suggests that it might have an important role in the regulation of cellular interactions of the immune system. Such a role has also been demonstrated for other galectins, *e.g.* for galectin-1, which induces apoptosis in activated T-cells (21) and galectin-3, which is up-regulated in proliferating T-cells and can inhibit Fas-mediated apoptosis (22). As this new galectin is strongly overexpressed in Hodgkin's disease tissue, it is conceivable that it might participate in the interaction between the H&RS cells with their surrounding cells and might thus play a role in the pathogenesis of this elusive disease and/or its consistently associated immunodeficiency.

So far we have not been able to assign the cellular origin of the galectin-9 transcripts to defined subpopulations in the peripheral blood or in the Hodgkin's infiltrated tissues, since the polyclonal rabbit serum, which reacted with His-tagged protein in Western blot, did not work in immunocytology and immunohistology (data not shown). Preliminary studies using peripheral blood leukocytes separated by magnetic cell sorting indicate that the expression level of the transcript is similar in enriched populations of B-cells, T-cells, and macrophages (data not shown). The availability of monoclonal antibodies that function in immunocytology will enable us to define more precisely the subpopulation(s) with high human galectin-9 expression.

REFERENCES

1. Barondes, S. H., Castronovo, V., Cooper, D. N. W, Cummings, R. D., Drickamer, K., Feizi, T., Gitt, M. A., Hirabayashi, J., Hughes, R. C., Kasai, K. I., Leffler, H., Liu, F. T., Lotan, R., Mercurio, A. M., Monsigny, M., Pillai, S., Poirier, F., Raz, A., Rigby, P., Rini, J. M., and Wang, J. L. (1994) *Cell* **76,** 597–598
2. Barondes, S. H., Cooper, D. N. W., Gitt, M. A., and Leffler, H. (1994) *J. Biol. Chem.* **269,** 20807–20810
3. Clerch, L. B., Whitney, P., Hass, M., Rew, K., Miller, T., Werner, R., and Massoro, D. (1988) *Biochemistry* **27,** 692–699
4. Teichberg, V. I., Silman, I., Beitsch, D. D., and Resheff, G. (1975) *Proc. Natl. Acad. Sci., U. S. A.* **72,** 1383–1387
5. Gitt, M. A., Massa, S. M., Leffler, H., and Barondes, S. H. (1992) *J. Biol. Chem.* **267,** 10601–10606
6. Ho, M. K., and Springer, T. A. (1982) *J. Immunol.* **128,** 1221–1227
7. Cherayil, B., Chartovitz, S., Wong, C., and Pillai, S. (1990) *Proc. Natl. Acad.*

Sci. U. S. A. **87,** 7324–7328
8. Gitt, M. A., Wiser, M. F., Leffler, H., Herrmann, J., Xia, Y. R., Massa, S. M., Cooper, D. N. W., Lusis, A. J., and Barondes, S. H. (1995) *J. Biol. Chem.* **270,** 5032–5038
9. Madsen, P., Rasmussen, H. H., Flint, T., Gromov, P., Kruse, T. A., Honore, B., Vorum, H., and Celis, J. E. (1995) *J. Biol. Chem.* **270,** 5823–5829
10. Oda, Y., Herrmann, J., Gitt, M. A., Turck, C. W., Burlingame, A. L., Barondes, S. H., and Leffler, H. (1993) *J. Biol. Chem.* **268,** 5929–5939
11. Hadari, Y. R., Paz, K., Dekel, R., Mestrovic, T., Accili, D., and Zick, Y. (1995) *J. Biol. Chem.* **270,** 3447–3453
12. Hirabayashi, H., Satoh, M., and Kasai, K. (1992) *J. Biol. Chem.* **267,** 15485–15490
13. Cho, M., and Cummings, R. D. (1995) *J. Biol. Chem.* **270,** 5207–5212
14. Cleves, A. E., Cooper, D. N. W., Barondes, S. H., and Kelly, R. B. (1996) *J. Cell Biol.* **133,** 1017–1026
15. Baum, L. G., Pang, M., Perillo, N. L., Wu, T., Delegeane, A., Uittenbogaart, C. H, Fukuda, M., and Seilhamer, J. J. (1995) *J. Exp. Med.* **181,** 877–887
16. Inohara, H., and Raz, A. (1995) *Cancer Res.* **55,** 3267–3271
17. Ohannesian, D. W., Lotan, D., Thomas, P., Jessup, J. M., Fukuda, M., Gabius, H. J., and Lotan, R. (1995) *Cancer Res.* **55,** 2191–2199
18. Woynarowska, B., Skrincosky, D. M., Haag, A., Sharma, M., Matta, K., and Bernacki, R. J. (1994) *J. Biol. Chem.* **269,** 22797–22803
19. Jeng, K. C., Frigeri, L. G., and Liu, F. T.(1994) *Immunol. Lett.* **42,** 113–116
20. Yamaoka, A., Kuwabara, I., Frigeri, L. G., and Liu, F. T. (1995) *J. Immunol.* **154,** 3479–3487
21. Perillo, N. L., Pace, K. E., Seilhamer, J. J., and Baum, L. G. (1995) *Nature* **378,** 736–739
22. Yang, R. Y., Hsu, D. K., and Liu, F. T. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 6737–6742
23. Sahin, U., Türeci, Ö., Schmitt, H., Cochlovius, B., Johannes, T., Schmits, R., Stenner, F., Luo, G., Schobert, I., and Pfreundschuh, M. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 11810–11813
24. Türeci, Ö., Sahin, U., Schobert, I., Koslowski, M., Schmitt, H., Schild, H. J., Stenner, F., Seitz, G., Rammensee, H. G, and Pfreundschuh, M. (1996) *Cancer Res.* **56,** 4766–4772
25. Short, J. M., Fernandez, J. M. Sorge, J. A., and Huse, W. D. (1988) *Nucleic Acids Res.* **16,** 7583–7600
26. Alting-Mees, M. A., Sorge, J. A., and Short, J. M. (1992) *Methods Enzymol.* **216,** 483–495
27. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J.(1992) *J. Mol. Biol.* **215,** 403–410
28. Emmert, D. B., Stoehr, P. J., Stoesser, G., and Cameron, G. N. (1994) *Nucleic Acids Res.* **22,** 3445–3449
29. Benson, D. A., Boguski, M., Lipman, D. J., and Ostell, J. (1994) *Nucleic Acids Res.* **22,** 3441–3444
30. Bairoch, A., and Bucher, P.(1994) *Nucleic Acids Res.* **22,** 3583–3589
31. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159
32. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248–254
33. Kozak, M. (1991) *J. Cell Biol.* **108,** 229–241
34. Leffler, H., and Barondes, S. H. (1986) *J. Biol. Chem.* **261,** 10119–10126
35. Banks, P. M. (1990) *Semin. Oncol.* **17,** 683–695
36. Romagnani, S. P., Ferrini, P. L. R., and Ricci, M. (1985) *Semin. Hematol.* **22,** 41–48
37. Bepler, G., Zhen, Q., and Havemann, K. (1985) *J. Cancer Res. Clin. Oncol.* **110,** 56–64