Structural Basis for the Recognition of Carbohydrates by Human Galectin-7[†,‡]

Demetrios D. Leonidas,[§] Efstratia H. Vatzaki,[§] Henrik Vorum,[∥] Julio E. Celis,[∥] Peder Madsen,[∥] and K. Ravi Acharya*[,§]

Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, U.K., and Department of Medical Biochemistry and Danish Centre for Human Genome Research, University of Aarhus, DK-8000 Aarhus, Denmark

Received May 7, 1998; Revised Manuscript Received July 17, 1998

ABSTRACT: Knowledge about carbohydrate recognition domains of galectins, formerly known as S-type animal lectins, is important in understanding their role(s) in cell−cell interactions. Here we report the crystal structure of human galectin-7 (hGal-7), in free form and in the presence of galactose, galactosamine, lactose, and *N*-acetyl-lactosamine at high resolution. This is the first structure of a galectin determined in both free and carbohydrate-bound forms. The structure shows a fold similar to that of the prototype galectins -1 and -2, but has greater similarity to a related galectin molecule, Gal-10. Even though the carbohydrate-binding residues are conserved, there are significant changes in this pocket due to shortening of a loop structure. The monomeric hGal-7 molecule exists as a dimer in the crystals, but adopts a packing arrangement considerably different from that of Gal-1 and Gal-2, which has implications for carbohydrate recognition.

Molecular recognition is a key event in biological processes. Often, cell−cell or cell−matrix interactions are mediated by lectins through their interaction with cognate carbohydrates. The galectins belong to a family of β-galactoside-binding lectins which possess at least one carbohydrate recognition domain (CRD) and exhibit some sequence conservation (*1*, *2*). These lectins are soluble in the cytoplasm and are secreted, despite the lack of a classical signal peptide (*3*). Members of this family show a widespread distribution in a variety of species ranging from sponges to humans, and have been implicated in intra- and extracellular functions through glycoconjugate-mediated recognition. To date, ten mammalian galectins (galectins 1−10) and many from other species have been well-characterized (for a recent review see ref *4* and references therein). On the basis of their molecular architecture, in particular the CRD, galectins are classified into three groups: proto-, chimera-, and tandem repeat-type galectins (*5*).

One member of the galectin family, galectin-7 (hGal-7)[1] a 15 kDa protein originally found in the human epidermis, is a proto-type, monomeric galectin with a single CRD, and has been cloned and isolated in two different laboratories (*6*, *7*). This galectin has significant amino acid sequence identity to known galectins, in particular to human Gal-1 (38% identity) and human Gal-2 (27% identity). It was first identified by two-dimensional polyacrylamide gel electrophoresis of keratinocyte protein extracts (*6*), and differential in situ hybridizations indicated that it is specifically expressed in these cells (*7*). Human Gal-7 is expressed at all stages of epidermal differentiation (i.e., in basal and suprabasal layers) but is moderately repressed by retinoic acid, contrasting with the behavior of other keratinocyte markers sensitive to this agent, which, if basal, are induced, or if suprabasal, are repressed. The effect of retinoic acid on a keratinocyte cell marker such as hGal-7 resembles more of its metaplasiogenic effect in vivo than of its inhibitory effect on terminal epidermal differentiation in vivo (*7*).

Presently, all of the available data indicate that hGal-7 has a definitive role in cell−cell and cell−matrix interactions, in particular, in keratinocyte development, as it seems to be the case for other members of the galectin family (ref *1* and references therein). The protein localizes to areas of cell−cell contact, particularly in the upper layers of human epidermis, and it is strikingly down-regulated in SV40-transformed K14 keratinocytes which are anchorage-independent and unable to differentiate (*6*). Thus, the expression of hGal-7 may be required for the maintenance of the normal keratinocyte phenotype (*7*, *8*). Recently, hGal-7 has been identified as one of the p53 inducer genes and may be implicated in p53-mediated transcriptional activation resulting in apoptosis (*9*). Also, it has been reported that the expression of rat Gal-7 (a homologue of hGal-7) is elevated in chemically induced mammary carcinomas (*10*). Thus, different biological functions could be attributed to each galectin.

[†] This work was supported by a Medical Research Council (UK) Program Grant 9540039 and a Wellcome Trust (UK) Equipment Grant 043614/Z/95/Z to K.R.A., and partial support from the Danish Medical Research Council, the Danish Biotechnology and Development Program, the Danish Cancer Society, NOVO Nordisk research grants to P.M. and J.E.C., and the European Union through its support of the work at EMBL, Hamburg, through the HCMP Access to Large Installations Project (Contract Number CHGE-CT93-0040).

[‡] The atomic coordinates for the complexes of human Galectin-7 as well as for the free structure have been deposited with the Protein Data Bank (accession numbers 2GAL, 3GAL, 4GAL, 5GAL, and 1BKZ, respectively).

* Address correspondence to this author. Phone: +44-1225-826238. Fax: +44-1225-826779. E-mail: K.R.Acharya@bath.ac.uk.

[§] University of Bath.

[∥] University of Aarhus.

[1] Abbreviations: CRD, carbohydrate recognition domain; DMM, dimethyl mercury; FOM, figure of merit; Gal, galectin; hGal-7, human galectin-7; MIR, multiple-isomorphous replacement; PCR, polymerase chain reaction; rms, root-mean-square.

S0006-2960(98)01056-3 CCC: $15.00
Published on Web 09/15/1998

On the basis of structural studies, a considerable wealth of information on the molecular recognition of lectins by carbohydrates has been obtained over the past few years (for recent reviews see refs *11*−*14*). In the case of galectins, X-ray crystallographic studies on Gal-1 with biantennary saccharides [between 2.15 and 2.45 Å (*15*), *N*-acetyllactosamine at 1.9 Å (*16*), and Gal-2 in complex with lactose at 2.9 Å (*17*)] have been performed. Some preliminary structural information on Gal-3 and Gal-4 CRDs is also available (*18*). These structures have provided a description of the binding mode(s) of different carbohydrate moieties to galectin molecules (*11*, *18*). They have also provided some insights into the multivalent functionality of galectins in carbohydrate recognition. Interestingly, the crystal structure of Charcot-Leyden crystal protein from human eosinophils, a monomeric galectin, named as Gal-10, also has the CRD-containing "lectin fold" of Gal-1 and Gal-2 and possesses lysophospholipase activity (*19*). It has been shown that this protein has specific, *albeit* weak, carbohydrate-binding activity for *N*-acetyl-D-glucosamine and lactose (*19*, *20*), but the mode of carbohydrate-binding is yet to be established. This clearly illustrates the importance of elucidating the structural determinants of lectin-binding specificity for each galectin in order to understand the interaction of glycoconjugate ligands at physiological concentrations. Here, we report high-resolution structures for both native hGal-7 (free form) and hGal-7 in complex with galactose, galactosamine, lactose, and *N*-acetyl-lactosamine at better than 2.0 Å resolution. The knowledge of the detailed carbohydrate-binding interactions provides structural insights into the interactions of hGal-7 with other cellular constituents.

EXPERIMENTAL PROCEDURES

Overexpression of hGal-7. The coding part of hGal-7 (*6*), except for the first start methionine, was amplified with *Pfu* DNA polymerase (Stratagene). The PCR-amplified DNA segment was cut with the restriction enzymes *Bam*HI and *Hind*III and ligated into the hexa-His expression vector pT7-PL (*21*), a derivative of the plasmid pRK172 containing a T7 promoter. The plasmid was sequenced to ensure that no changes occurred in the encoded amino acid sequence during the amplification process. BL21(DE3) *Escherichia coli* cells were transformed with the recombinant plasmid and grown at 37 °C to an OD_{600} of 0.6−0.7, at which point expression of the recombinant protein was induced by adding IPTG to 0.5 mM (*22*). The cells were grown for an additional 4 h after induction, followed by centrifugation for 10 min (7000 rpm, Sorvall SS34) at 4 °C. Induction of the proteins was analyzed by SDS−PAGE.

Purification of hGal-7. After centrifugation, the bacterial pellet from 2 L of culture medium was resuspended in 50 mL of lysis buffer (300 mM NaCl, 20 mM imidazole, 1 mg/mL lysozyme, 1 mM PMSF, 50 mM sodium phosphate, pH 8.0), incubated for 30 min, and centrifuged for 1 h (11 000 rpm, Sorvall SS34) at 4 °C. The supernatant was then applied to a Ni^{2+}-NTA column which was pre-equilibrated with the same buffer. After intensive washing with wash buffer (300 mM NaCl, 20 mM imidazole, 1 mM PMSF, 50 mM sodium phosphate, pH 8.0), the protein was eluted from the column in a single step with elution buffer (300 mM NaCl, 250 mM imidazole, 50 mM sodium phosphate, pH 8.0). The purification was analyzed by SDS−PAGE.

Factor Xa Cleavage of hGal-7. The purified protein solution was diluted twice with distilled water to reach a final concentration suitable for Factor Xa cleavage (150 mM NaCl, 125 mM imidazole, 25 mM sodium phosphate, pH 8.0); NaN_3 was added to a final concentration of 0.1%. The protein was then digested for 15 h at 37 °C with the restriction protease (approximately 1/150 weight−weight) Factor Xa (Denzyme), a protease which cleaves C-terminal to the IEGR sequence (*23*). After cleavage, hGal-7 was separated from contaminating proteins by gel filtration on a Superdex 200 column (Pharmacia) pre-equilibrated with wash buffer (300 mM NaCl, 20 mM imidazole, 50 mM sodium phosphate, pH 8.0). Finally, the purified protein was passed through the Ni^{2+}-NTA column once again to remove uncleaved protein still containing the hexa-His tag. Each

Table 1: Statistics for Crystallographic Structure Determination[a]

crystal	native I	native II	10 mM $HgCl_2$	20 mM $HgCl_2$	DMM
internal scaling					
wavelength (Å)	0.87	0.8373	1.5418	1.5418	1.5418
resolution (Å)	40−2.0	40−1.9	40−2.5	40−2.6	40−2.5
no. of crystals	2	1	1	1	1
reflections measured	123 784	90 657	78 754	67 193	58 371
unique reflections	17 794	16 622	8867	8310	8933
completeness (%)	95.8	78.1	93.9	98.0	94.3
(outermost shell)	(93.7)	(82.1)	(63.2)	(99.5)	(60.7)
$I/\sigma I$	10.6	9.6	11.2	7.0	7.0
R_{symm}[b] (%)	10.6	5.1	9.5	11.9	12.2
R_{merge}[c] (%)	5.4				
overall completeness (%)	95.2				
derivative scaling					
resolution (Å)			20−2.5	20−2.6	20−2.5
mean fractional isomorphous difference (%)			28.1	32.1	18.6
heavy-atom refinement					
no. of sites			2	2	2
coordinating residue			Cys 38	Cys 38	Cys 38
phasing power[d] (centrics/acentrics)			1.3/1.6	2.4/3.2	2.3/3.3
R_{cullis}[e] (%) (centrics)			71.4	59.7	59.4
R_{cullis}[e] (%) (acentrics)			75.1	54.8	52.6
overall figure of merit	0.60				

[a] Abbreviations: $HgCl_2$, mercuric chloride; DMM, dimethyl mercury. [b] $R_{symm} = \sum_h \sum_i |I(h) - I_i(h)|/\sum_h \sum_i I_i(h)$, where $I_i(h)$ and $I(h)$ are the ith and the mean measurements of the intensity of reflection h. [c] R_{merge} is equal to R_{symm}, but it refers to merging data from different crystals and different sources. [d] Phasing power is the ratio of the rms calculated heavy-atom structure amplitude to the rms lack of closure. [e] R_{cullis}, cullis R factor: $\sum |||F_{PH}|_{obs} +/- |F_P|_{obs}| - |F_H|_{calc}|/\sum ||F_{PH}|_{obs} +/- |F_P|_{obs}|$, where F_{PH} and F_P are the observed structure factor amplitudes for the heavy-atom derivative and the native data sets.

Table 2: Data Collection Statistics for hGal-7−Carbohydrate Complexes

data set	hGal-7−galactose	hGal-7−galactosamine	hGal-7−lactose	hGal-7−*N*-Ac-lactosamine
wavelength (Å)	0.912	0.8373	0.98	0.8373
resolution (Å)	40.0−2.0	40.0−1.9	40.0−1.95	40.0−1.9
reflections measured	172 647	132 648	80 604	108 433
unique reflections	18 155	22 372	19 246	21 098
R_{symm}[a] (%)	7.7	5.8	4.7	6.3
completeness (%)	98.2	98.6	96.8	99.4
(outermost shell)	(97.4)	(98.0)	(92.4)	(98.3)
$I/\sigma I$	9.0	6.8	7.6	7.6

[a] $R_{symm} = \sum_h \sum_i |I(h) - I_i(h)|/\sum_h \sum_i I_i(h)$, where $I_i(h)$ and $I(h)$ are the ith and the mean measurements of the intensity of reflection h.

Table 3: Refinement Statistics

	native	galactose complex	galactosamine complex	lactose complex	*N*-Ac-lactosamine complex
resolution (Å)	20−1.9	20−2.0	20−1.9	20−1.95	20−2.0
R_{cryst}[a] (%)	19.4	18.7	21.1	20.0	21.3
R_{free}[b] (%)	26.0	24.5	25.8	25.5	29.6
no. of reflections	20 239	17 545	22 335	19 174	18 066
no. of protein atoms	2116	2081	2110	2116	2095
no. of solvent molecules	89	76	87	86	81
no. of carbohydrate atoms		24	24	23	26
deviations from ideality (rms)					
bond lengths (Å)	0.008	0.008	0.009	0.010	0.009
bond angles (deg)	1.7	1.6	1.7	1.7	1.6
dihedrals (deg)	30.1	30.0	30.1	30.1	30.0
impropers (deg)	0.8	0.8	0.8	0.8	0.8
average B factor (Å^2)					
main chain atoms (mol A/mol B)	27.4/38.2	18.8/28.3	21.3/39.7	28.5/44.5	28.5/39.1
side chain atoms (mol A/mol B)	32.0/42.0	23.5/32.4	27.0/43.1	32.4/47.3	32.9/41.8
all protein atoms (mol A/mol B)	29.7/40.0	21.1/30.3	24.1/41.4	30.4/45.8	30.6/40.4
solvent atoms	43.1	37.6	39.9	47.6	47.1
carbohydrate molecules		26.5/35.2	26.6/45.7	55.9	44.6

[a] $R_{cryst} = \sum_h |F_o - F_c|/\sum_h F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h, respectively. [b] R_{free} is equal to R_{cryst} for a randomly selected 5% subset of reflections not used in the refinement (*33*).

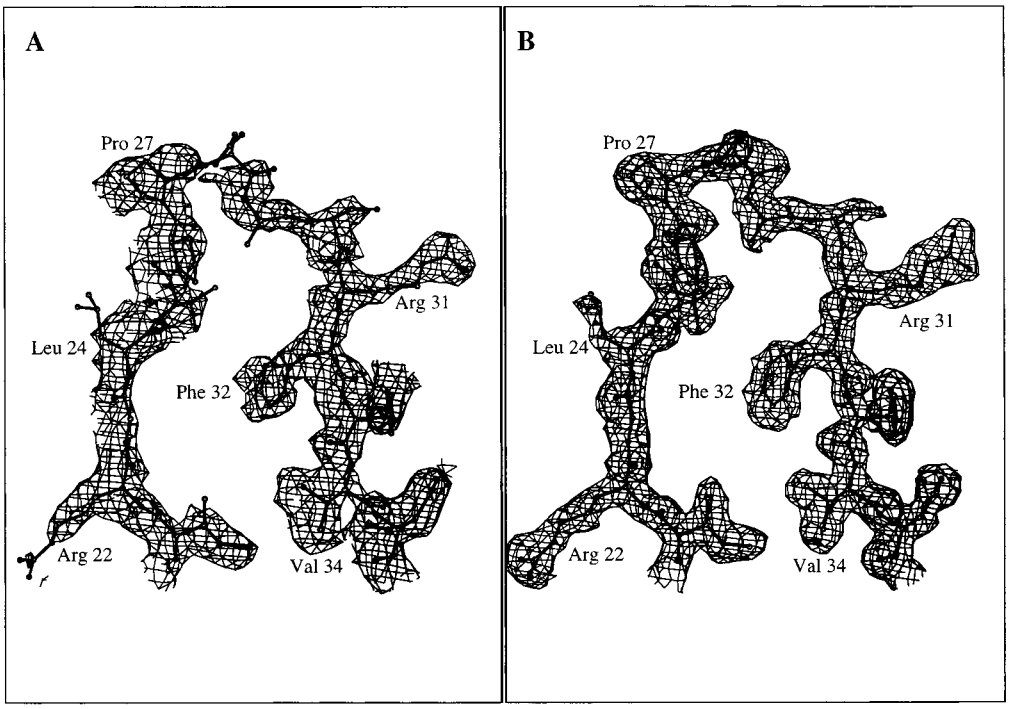

FIGURE 1: A view of a slice through the electron density map of the refined hGal-7 atomic model in the loop region comprising residues 21−35. The maps were calculated (A) using the observed structure factors and phases derived from SHARP after density modification (*27*) and (B) using phases from the final model (for details see Experimental Procedures). The map is contoured at 1σ and was calculated using data between 20 and 1.9 Å.

step during the purification was analyzed by SDS−PAGE. The final protein concentration was estimated by the Bradford procedure (*24*) using BSA as standard and was found to be approximately 10 mg/L of bacterial medium.

Crystallization. Crystals of recombinant hGal-7 were grown using the hanging drop vapor diffusion method from drops containing 9 mg/mL protein at pH 8.1 in 50 mM sodium phosphate buffer, 0.3 M sodium chloride, 20 mM imidazole, and 8.5% PEG 3350. Drops were equilibrated against reservoirs containing 50 mM sodium phosphate buffer, 0.3 M sodium chloride, 20 mM imidazole, and 17% PEG 3350 at pH 8.1. Single crystals appeared after 3−4 days at 16 °C. These crystals diffracted to a minimum Bragg spacing of 1.9 Å on a synchrotron radiation source. The systematic absences and symmetry were consistent with the space group $P2_12_12_1$, with unit cell dimensions $a = 54.23$ Å, $b = 65.34$ Å, and $c = 73.60$ Å. There are two hGal-7 molecules per crystallographic asymmetric unit, and approximately 50% of the crystal volume is occupied by solvent.

The hGal-7−lactose and hGal-7−*N*-acetyl-lactosamine complexes were obtained by soaking native hGal-7 crystals with 5 mM lactose or 50 mM *N*-acetyl-lactosamine for 2 h or 3 days, respectively. Cocrystals for the hGal-7−galactose and hGal-7−galactosamine complexes were grown using the native conditions but with the addition of 50 mM of the

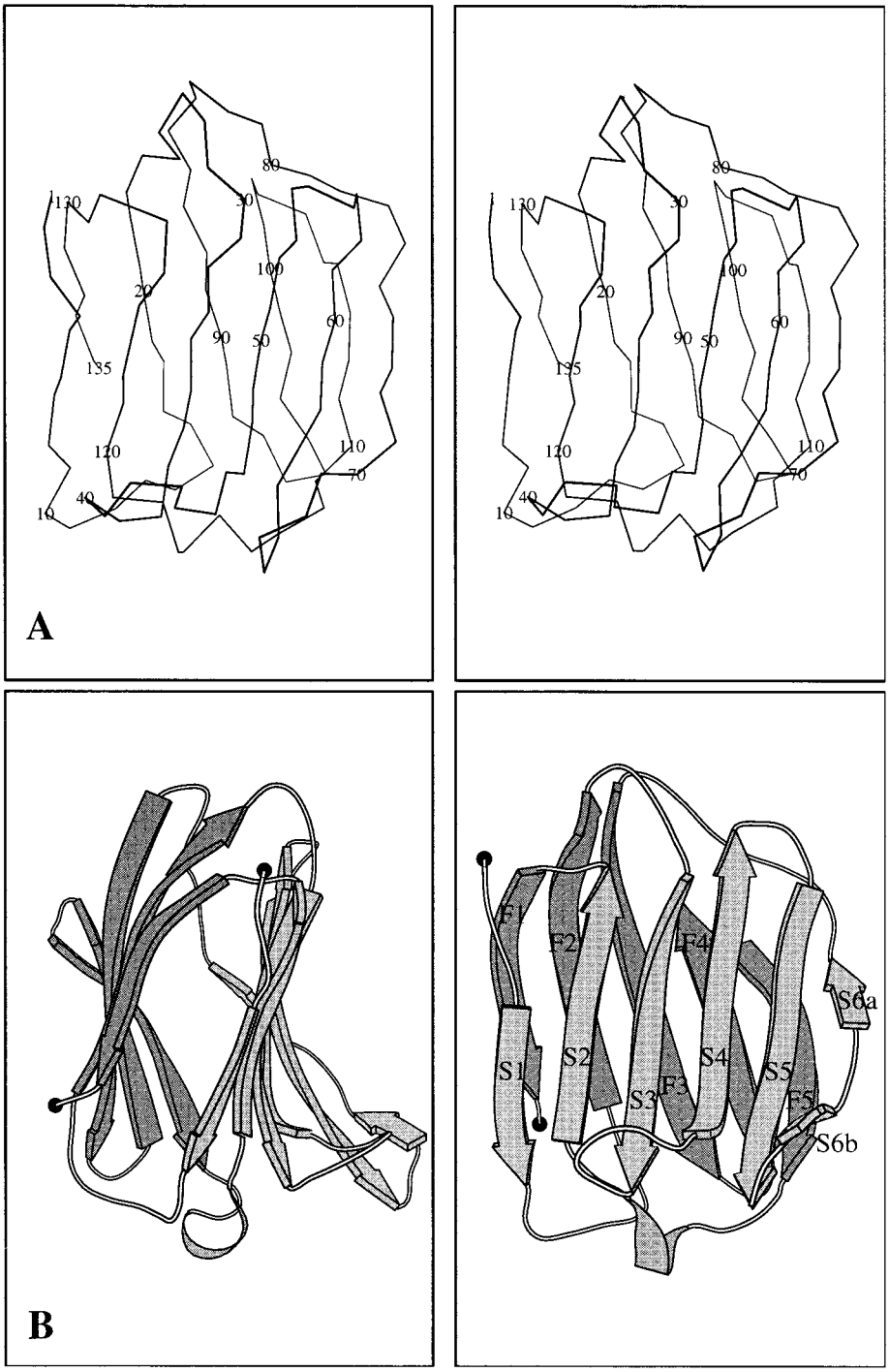

FIGURE 2: (A) Stereo picture of the C^{α} trace of hGal-7. (B) Orthogonal views of the hGal-7 molecule, drawn with the program MOLSCRIPT (*44*): The "β-sandwich" is formed by the packing of two β-sheets, S1-S6a/S6b and F1-F5.

respective carbohydrate in the crystallization buffer. The hGal-7−galactose complex crystals were orthorhombic and grew within a week at 16 °C (space group $P_12_12_1$ with unit cell dimensions $a = 54.47$ Å, $b = 64.52$ Å, and $c = 72.24$ Å). The hGal-7−galactosamine crystals belong to space group $C2$ with unit cell dimensions $a = 95.59$ Å, $b = 54.62$ Å, $c = 61.20$ Å, and $\beta = 115.06°$, with two molecules per asymmetric unit.

Data Processing and Reduction. Diffraction data were collected at 16 °C using the Synchrotron radiation source at Daresbury (U.K.) on station PX9.6 ($\lambda = 0.87$ Å) for the native to 2.0 Å resolution and on station PX9.5 ($\lambda = 0.98$ Å) for the hGal-7−lactose complex to 1.95 Å. Another native data set was collected to 1.9 Å resolution on station BW7B ($\lambda = 0.837$ Å) at EMBL, Hamburg, Germany. Data for the hGal-7−galactosamine and the hGal-7−*N*-acetyllactosamine complexes were also collected on station BW7B at Hamburg, while data for the hGal-7−galactose complex were collected on station X11 at Hamburg (for details see Tables 1 and 2). Heavy-atom derivative data sets for mercury chloride and dimethyl mercury (DMM) to 2.5 Å resolution were collected in-house on a 30 cm diameter MAR−Research image plate mounted on an Enraf-Nonius X-ray generator. The mercury derivatives were prepared by soaking native hGal-7 crystals for 24 and 48 h in solutions of the mother liquor containing 20 and 10 mM mercury

Table 4: Structural Superposition of Galectins[a]

	Gal-7 (A)	Gal-7 (B)	Gal-1 (A)	Gal-1 (B)	Gal-2 (A)	Gal-2 (B)
Gal-7 (B)	0.45					
	134/134					
Gal-1 (A)	1.64	1.57				
	43/128	43/129				
Gal-1 (B)	2.09	1.60	0.54			
	37/129	43/128	129/129			
Gal-2 (A)	1.77	1.67	0.75	0.88		
	35/128	35/128	55/128	55/128		
Gal-2 (B)	1.72	1.63	0.77	0.97	0.49	
	35/128	35/128	55/128	55/128	129/129	
Gal-10	1.40	2.06	1.86	1.97	2.01	2.03
	35/131	33/131	24/126	24/127	24/128	24/128

[a] Brookhaven Protein Data Bank accession numbers for the proteins used are the following: bovine spleen galectin-1 (Gal-1), 1SLT (*16*); human galectin-2 (Gal-2), 1HLC (*17*); human eosinophil galectin-10 (Gal-10), 1LCL (*19*). Galectins -1, -2, and -7 are dimers in the crystal; the monomers are designated as A and B. Gal-10 is a monomer. The first row corresponds to overall rms distance between the proteins, and the second row shows the number of identical amino acid residues divided by the number of structural equivalents between the two structures. Structural alignments were made with the SHP program (*37*).

chloride, respectively. The DMM derivative was prepared by vapor diffusion by pipetting a small amount of DMM into one end of the data collection capillary containing a crystal mounted in the standard way, and data were collected after approximately 24 h of equilibration at room temperature.

Raw data images were indexed, integrated, and corrected for Lorentz and polarization effects using the program DENZO (*25*). All data were scaled and merged using the program SCALEPACK (*25*). Intensities were truncated to amplitudes by the TRUNCATE program (*26*). The two native data sets were finally merged together resulting in a master native data set, which was 95.2% complete to 1.9 Å resolution. Details of data processing statistics are presented in Tables 1 and 2.

Phase Calculation. Structure determination using the method of "molecular replacement" with the known galectin structures (Gal-1, -2, and -10) as the starting model failed to give an unambiguous set of solutions for hGal-7. Hence the structure was determined by the method of "multiple isomorphous replacement" (MIR) using mercury chloride and DMM as derivatives (Table 1). The maximum-likelihood program SHARP (*27*) was used for heavy-atom refinement and phasing. The initial MIR phases were calculated using reflections from 50.0 to 2.5 Å resolution, which gave a final figure of merit (FOM) of 0.564. The protein phases were improved and extended to 2.0 Å resolution using density modification with the program SOLOMON (*28*) as implemented within the SHARP program assuming a solvent content of 0.5. The final FOM was 0.847. The resultant solvent-flattened electron density map was used to build a polyalanine model using the graphics program O (*29*).

The structure of the hGal-7−galactosamine complex (space group *C*2) was determined by the method of molecular replacement using the program AMoRe (*30*) with the refined coordinates of the native Gal-7 at 1.9 Å resolution as a search model.

Refinement. All crystallographic refinement was carried out using the program X-PLOR 3.851 (*31*). After one round of refinement using the polyalanine model and with tight noncrystallographic symmetry restraints, the crystallographic R factor (R_{cryst}) dropped to 0.417 (R_{free} 0.474; based on 5% of the reflections omitted from the refinement). Calculated phases obtained from the refined structure were combined with the original MIR phases and used to determine a new electron density map. Careful examination of this map and using the amino acid sequence of hGal-7 (*6*) allowed the building of two hGal-7 molecules. Alternating cycles of

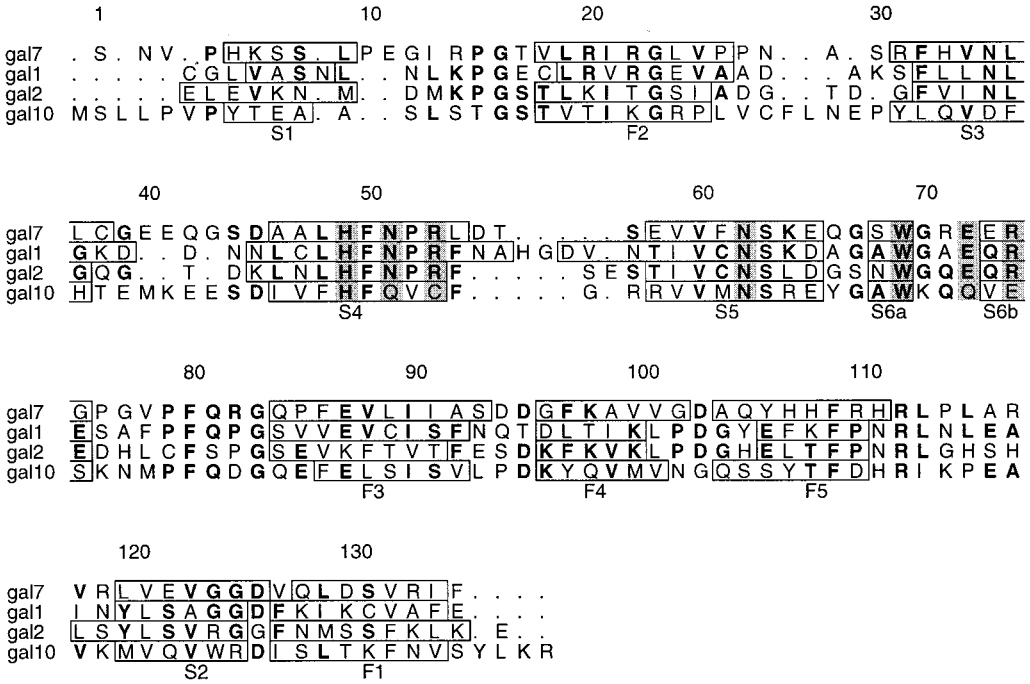

FIGURE 3: Structure-based sequence alignment of hGal-7, bovine spleen Gal-1 (1SLT), human Gal-2 (1HLC), and human eosinophil Gal-10 (1LCL) as determined using the program SHP (*37*). Every tenth residue in the hGal-7 sequence is numbered. Secondary structural elements (β-strands F1-F5 and S1-S6b), as determined by DSSP (*36*), are boxed. Residues that form part of the carbohydrate-binding site are shadowed, and residues which are identical between the galectins are highlighted in bold letters.

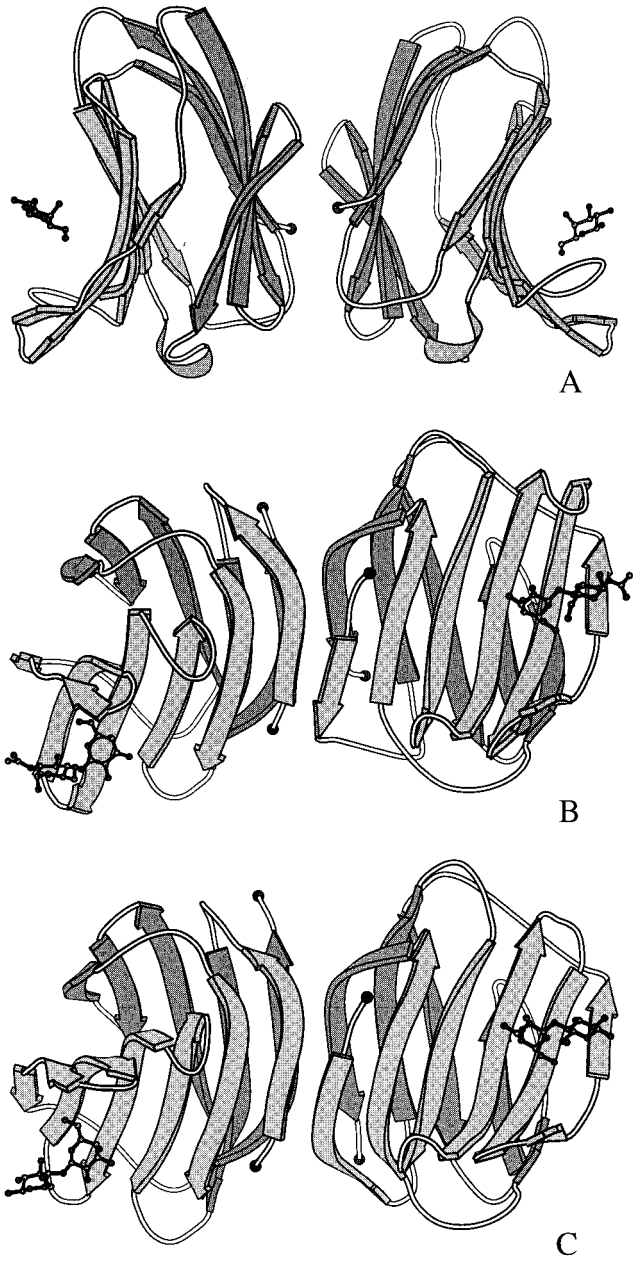

FIGURE 4: Dimer organization in hGal-7 (A), Gal-1 (*16*) (B), and Gal-2 (*17*) (C). The bound carbohydrates are shown in ball-and-stick representation.

manual building, conventional positional refinement, and the simulated annealing method as implemented in X-PLOR improved the quality of the model. Rebuilding was initially performed at 2.0 Å resolution, using electron density maps calculated from combined model/MIR phases. Extension of refinement from 2.0 to 1.9 Å was performed with the aid of SIGMAA-weighted $2|F_o| - |F_c|\phi_{calc}$ maps (*32*). The behavior of the R_{free} (*33*) value was monitored throughout refinement. Several rounds of refinement, model building, individual B-factor refinement, and solvent correction as implemented in X-PLOR 3.851 (*34*) (using all data from 20.0 to 1.9 Å resolution, a total of 17 687 reflections) gave a final model of native hGal-7. This model had an R_{cryst} of 0.194 and last recorded R_{free} of 0.26 (*33*). The rms deviation from ideality of bond lengths and bond angles were 0.008 Å and 1.73°, respectively. During the final stages of refinement, water molecules were inserted into the model only if there were peaks in the $|F_o| - |F_c|$ electron density maps with heights greater than 3σ and they were at hydrogen bond forming distances from appropriate atoms. $2|F_o| - |F_c|\phi_{calc}$ maps were also used to check the consistency in peaks. Water molecules with a temperature factor higher than 65 Å^2 were excluded from subsequent refinement steps.

For the complexes of hGal-7 with lactose, galactose, and *N*-acetyl-lactosamine, the refined native structure was used as a starting model for the refinement, while in the case of the hGal-7−galactosamine complex, the output model from AMoRe was used. In each case the carbohydrate molecule was included during the final stages of refinement using the respective coordinates from the Protein Data Bank. The details of refinement are given in Table 3.

The program PROCHECK (*35*) was used to assess the quality of the final structure. Analysis of the Ramachandran (φ-ψ) plot showed that all residues lie in the allowed regions. Secondary structural alignments and solvent accessibilities were calculated using the DSSP program (*36*). Structural superpositions were performed with the program SHP (*37*). Atomic coordinates for structural comparisons were obtained from the Protein Data Bank at the Brookhaven National Laboratory.

RESULTS AND DISCUSSION

Description of the Overall Structure. The three-dimensional crystal structures of hGal-7 in the native form and in complex with galactose, galactosamine, lactose, and *N*-acetyl-lactosamine were determined between 1.9 and 2.0 Å (Figure 1, Tables 1−3). The hGal-7 molecule crystallizes as a dimer, although in solution it is known to exist as a monomer unlike most other mammalian galectins, except Gal-10 (*6*). In molecule A, the first three residues at the N-terminus are disordered, while in molecule B residues 1−2 and 10−11 are disordered. The two monomers have nearly identical structures [excluding residues 1−3 at the N-terminus, the rms deviations are 0.45 Å (C^α atoms), 0.46 Å (main-chain), and 0.96 Å (all atoms)]. In the native structure, the dimer contains 89 bound water molecules.

The topology of the hGal-7 molecule consists of the "jelly roll" motif and is closely similar to that of Gal-1 (*16*), Gal-2 (*17*), and Gal-10 (*19*) (Figure 2). As found in Gal-10, a short 3_{10} helix is located between strands F5 and S2. A least-squares superposition of hGal-7 with Gal-1, Gal-2, and Gal-10 (Table 4) indicates that the hGal-7 structure is more similar to Gal-10 than to Gal-1 or Gal-2, although this is not reflected in pair-wise amino acid sequence comparisons. A structure-based sequence alignment of hGal-7 with the known galectin structures (Figure 3) reveals a few loop regions adopting significantly different positions. Specifically, in hGal-7, (i) there appears to be an insertion of two residues, Pro 10 and Glu 11, between strands S1 and F2 (Gal-1, -2, and -10 have similar structure in this region); (ii) the loop region encompassing residues 39−45 (between strands S3 and S4) has a conformation which differs from the corresponding loop in Gal-1 and -2, but which is similar to Gal-10; (iii) the short loop between residues 55 and 57 (connecting strands S4 and S5) adopts a very similar conformation to that in Gal-2 (in Gal-1, this loop has an extended structure due to the insertion of several residues, and in Gal-10, the corresponding loop has a different

Table 5: Intermolecular Contacts at the Dimer Interface

polar (distance < 3.3 Å)		distance	van der Waals interactions[a]		no. of
molecule A	molecule B	(Å)	molecule A	molecule B	contacts
Arg 14 Nη2	Asp 94 Oδ1	2.5	Arg 14	Asp 94, Asp 95	7
Arg 14 Nη2	Asp 95 Oδ1	3.1	Pro 15	Pro 15, Ser 93, Asp 94	4
Gly 16 O	Lys 98 Nζ	2.8	Gly 16	Gly 16, Ile 91, Ala 92, Lys 98	6
Arg 20 Nη1	Asp 103 Oδ1	3.0	Val 18	Val 18	1
Lys 98 Nζ	Gly 16 O	3.0	Arg 20	Asp 103	4
Lys 98 Nζ	Phe 138 OT	2.9	Ile 91	Gly 16	1
Asp 103 Oδ1	Arg 20 Nη1	2.6	Ala 92	Gly 16	1
Asp 103 Oδ2	Arg 20 Nη2	2.6	Ser 93	Pro 15	1
Asp 103 O	Arg 133 Nη1	3.2	Lys 98	Gly 16, Phe 135	5
Asp 103 O	Arg 133 Nη2	3.0	Val 100	Phe 135	2
			Asp 103	Arg 20, Arg 133, Phe 135	8
			Phe 135	Ile 91, Lys 98, Val 100, Asp 103	10

[a] van der Waals distances are the maximum allowed values of C−C, 4.1 Å; C−N, 3.8 Å; C−O, 3.7 Å; O−O, 3.3 Å; O−N, 3.4 Å; and N−N, 3.4 Å. OT, carboxy terminus.

Table 6: Hydrogen Bond Interactions between hGal-7 and β-Galactosides[a]

		galactose		galactosamine		lactose	*N*-acetyl-lactosamine
donor	acceptor	Mol A	Mol B	Mol A	Mol B	Mol B	Mol B
water	O1	3.2					
water	O1	2.7					
O2	Pro 10 O*					3.1	3.2
water	O2	3.0					
water	O3	2.5		2.7			
water	O3	3.1		3.1			
His 49 Nϵ2	O4	2.7	2.9	2.8	2.9	2.8	2.9
O4	Asn 51 Oδ1	3.2	3.2	3.1		2.9	2.9
Arg 53 Nη2	O4	3.2	2.9	2.9		2.6	2.6
water	O4	3.2	3.3	3.1			
Arg 53 Nη2	O5		3.1	3.1	3.0	2.9	2.8
O5	Glu 72 Oϵ1				3.0		
Asn 62 Nδ2	O6	2.6	2.9	2.8	2.9	2.4	2.6
O6	Glu 72 Oϵ1	2.9	2.5	2.7	2.6	3.1	2.8
O3′	Glu 72 Oϵ1					2.5	2.6
Arg 53 Nη1	O3′						3.1
O7′	water*						2.7

[a] Hydrogen bond interactions were calculated with the program HBPLUS (*43*). Values shown are distances in Å. Atoms from the glucose moiety of lactose and *N*-acetyl-lactosamine are shown with a prime, while atoms from a symmetry molecule are shown with a *.

structure due to the presence of Cys 57); and (iv) the loop containing residues 112−119 (between strands F5 and S2) matches closely the conformation of the corresponding loop in Gal-10 rather than that of Gal-1 and -2.

The hGal-7 molecule contains a single cysteine residue at position 38. It appears to be active and binds mercury atoms. It is situated at the end of strand S3 and is buried in both molecules of the dimer in the crystal. The thiol group does not appear to contribute to carbohydrate-binding (located some 15 Å away from the carbohydrate) and is not involved in any crystal-packing contacts.

Crystal Packing. In Gal-1 and Gal-2 (dimeric galectins with a single CRD), the monomers of the dimer pack approximately at 90° to the plane of the β-sheets, related by a 2-fold rotation axis. The narrow apolar interface is formed by four backbone hydrogen bonds. This structural organization gives rise to "subunit multivalency" for carbohydrate recognition (*11*, *18*). On the other hand, in hGal-7 the dimer arrangement in the crystal (noncrystallographic) is considerably different because this galectin recognizes carbohydrate(s) in its monomeric form (Figure 4) and does not possess multivalency. The dimer interface involves the association of the β-strands, F1−F5, from the two monomers (Figure 4A). These are held together by hydrogen bonding interactions involving 5 residues from molecule A, 8 residues from molecule B, and an extensive set of van der Waals interactions (Table 5). Thus, the dimer interface is relatively large in hGal-7, covering an area of 1484 Å^2, in contrast to areas of 1093 and 1179 Å^2 in Gal-1 and Gal-2, respectively (as calculated with the program NACCESS (*38*)). In general, molecule B has higher overall thermal parameter (Table 3) and makes fewer crystal-packing contacts than molecule A. Comparison of individual monomers of hGal-7, Gal-1, Gal-2, and Gal-10 reveals that each has a very similar topology.

The Carbohydrate Recognition Site. Both in the native and in the carbohydrate complex structures of hGal-7, each monomer adopts significantly different crystal-packing contacts. This has implications for carbohydrate recognition (discussed below). Nevertheless, the native structure does not undergo any significant conformational change upon carbohydrate-binding. In all four complex structures, the carbohydrate moiety has well-defined electron density. The rms deviations in atomic coordinates for the native and carbohydrate complexes range from 0.35 to 0.52 Å. In the case of the galactose and galactosamine complex structures (cocrystallization experiments), both monomers exhibit almost identical carbohydrate-binding modes (Table 6, Figure 5B,C). In binding galactose and galactosamine, residue Glu

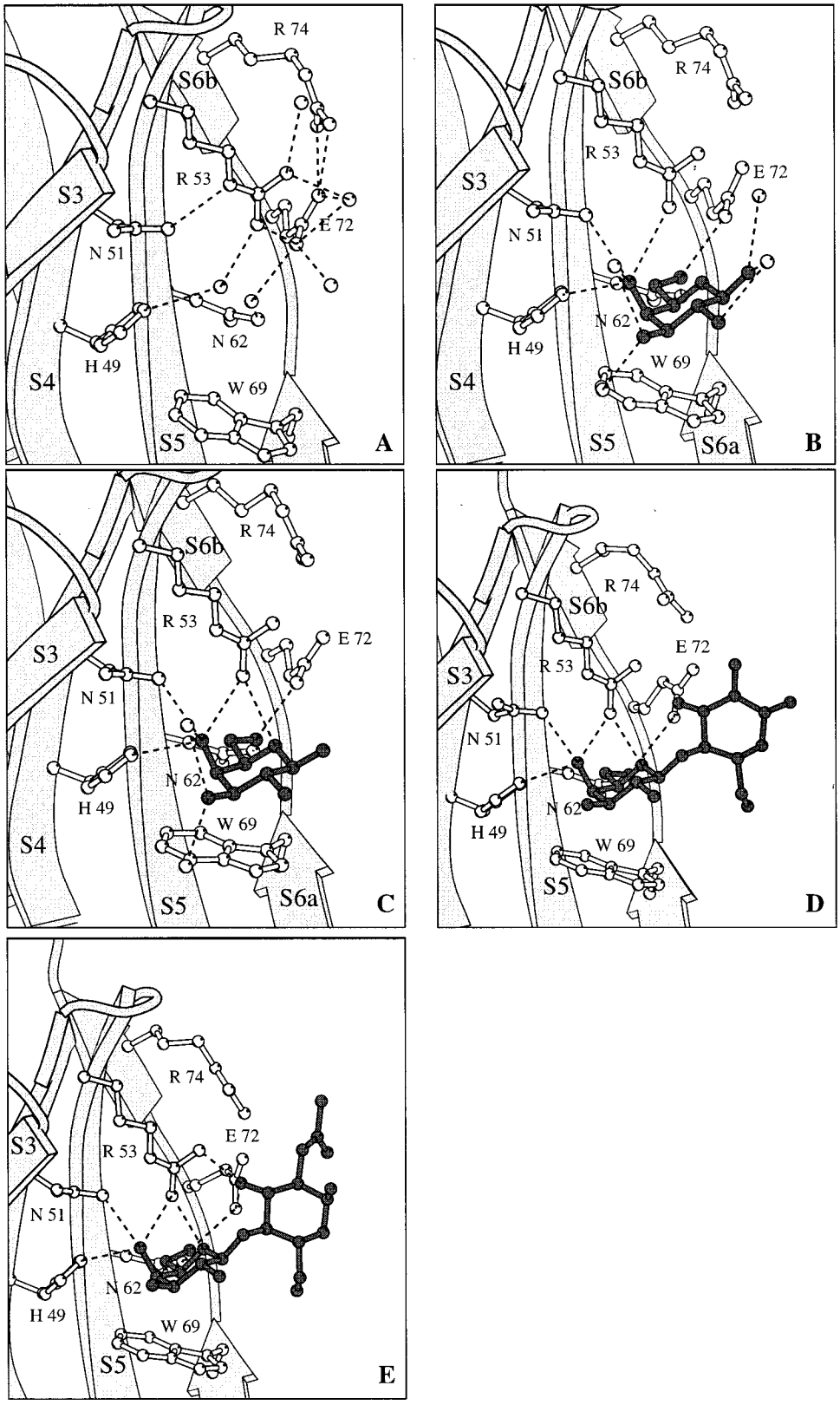

FIGURE 5: The carbohydrate recognition site of hGal-7: native (A), with bound galactose (B), with bound galactosamine (C), with bound lactose (D), and with bound *N*-acetyl-lactosamine (E). The side chains of residues comprising the hGal-7 carbohydrate-binding site are shown in ball-and-stick models. Bound waters are shown as spheres. Hydrogen bond interactions are represented in dashed lines. Figure prepared with the program MOLSCRIPT (*44*).

72 moves toward the carbohydrate moiety to form a hydrogen bond and three water molecules observed in the free structure are displaced by the carbohydrate (Table 6). Also, extensive water-mediated interactions with the carbohydrate moiety are identified (Figure 5). However, in the lactose and *N*-acetyl-lactosamine complex structures (soaking experiments; co-crystallization experiments failed to produce crystals for these complexes), carbohydrate-binding was observed only in

Table 7: Comparison of Polar Interactions between Galectins with Various Saccharides[a]

sugar atom	human Gal-7	bovine Gal-1	human Gal-2
Gal O6	Glu 72 (Oϵ2) 2.8	Glu 71 (Oϵ2) 2.8/3.1	Glu 68 (Oϵ1) 2.8/3.3
Gal O6	Asn 62 (Nδ2) 2.6	Asn 61 (Nδ2) 2.7/3.0	Asn 58 (Nδ2) 2.7/3.0
Gal O5	Arg 53 (Nη2) 2.8	Arg 48 (Nη2) 2.9/3.1	
Gal O4	Arg 53 (Nη2) 2.6	Arg 48 (Nη2) 3.0/2.8	
Gal O4	Asn 51 (Oδ1) 2.9		Asn 47 (Oδ1) 3.1/−
Gal O4	His 49 (Nϵ2) 2.9	His 44 (Nϵ2) 2.8/2.8	His 47 (Nϵ2) 2.7/2.9
Glc O2			Arg 70 (Nη2) 3.3/3.1
Glc O4			Arg 49 (Nη2) 3.3/−
Nag/Glc O3		Glu 71 (Oϵ2) 2.4/2.5	Glu 68 (Oϵ1) 2.8/2.7
Nag/Glc O3	Arg 53 (Nη1) 3.1	Arg 48 (Nη1) 3.2/3.3	Arg 49 (Nη2) 2.6/2.5
Nag/Glc O3		Arg 48 (Nη1) 2.8/2.5	
Nag/Glc O3		Arg 73 (Nη2) −/3.3	Arg 70 (Nη2) −/3.1

[a] Values given are distances in Å. Gal, Glc, and Nag are galactose, glucose, and *N*-acetyl-glucosamine moieties, respectively. The structures used are the following: Human Gal-7, *N*-acetyl-lactosamine complex (present study); bovine Gal-1, *N*-acetyl-lactosamine complex (*16*); human Gal-2, lactose complex (*17*).

molecule B (Table 6, Figure 5D,E). A close examination of the corresponding site in molecule A in these two complex structures reveals that residues 22−26 and 128−131 (from a symmetry-related molecule) block the disaccharide entry to this site. This observation is understandable since hGal-7 is a "monomeric galectin", while in the case of "dimeric galectins" (Gal-1 and Gal-2), interactions with lactose (*17*) and *N*-acetyl-lactosamine (*16*) are observed in both monomers. As described earlier, molecule B appears to be involved in fewer crystal-packing contacts and could accommodate lactose and *N*-acetyl-lactosamine in the carbohydrate-binding pocket during soaking experiments.

The carbohydrate recognition residues of Gal-1 and Gal-2 have been determined either by site-directed mutagenesis (*39*) or by X-ray crystallographic studies (*15*−*17*). These residues are conserved in hGal-7 (His 49, Asn 51, Arg 53, Asn 62, Trp 69, Glu 72, Arg 74; Figure 3) and are located in the concave face of the S4−S5−S6 β-sheet (Figure 4). The high degree of structural conservation in the carbohydrate-binding groove supports the suggestion that this region represents a "carbohydrate-binding cassette" present in all the members of the galectin family (*40*). In the native hGal-7 structure, this site is occupied by five water molecules in molecule A (none in molecule B) and contains a network of hydrogen bonds (Figure 5A). For example, His 49 forms hydrogen bonds with Asn 35 and Asn 62 and is involved in a water-mediated interaction with Arg 53. Asparagine 51 makes hydrogen bonds with Phe 32, Arg 53, and Val 60, while Arg 53 makes hydrogen bonds with Glu 58 and Glu 72. Asparagine 62 makes polar contacts with His 49, Ser 63, Gly 70, Arg 71, and Arg 112. Glutamic acid 72 makes hydrogen bonds with Arg 53 and Arg 74, which in turn makes a hydrogen bond to Glu 58.

Detailed analysis of hGal-7−carbohydrate complex structures show that His 49, Asn 51, Arg 53, Asn 62, and Glu 72 are the key residues involved in carbohydrate recognition through hydrogen bond interactions (Tables 6 and 7). The highly conserved residues His 49, Asn 51, and Arg 53 make hydrogen bonds with the galactose O4 in all four complexes (Figure 5). The galactose O5 makes two hydrogen bonds with Arg 53 and Glu 72, while O6 makes interactions with Asn 62 and Glu 72. Tryptophan 69 is involved in stacking interactions with the galactose moiety in a manner analogous to that seen in Gal-1 and Gal-2 structures (*16*, *17*). Residues Arg 53, Thr 56, Glu 58, Glu 72, and Arg 74 form a network of ionic interactions and may be important for the optimal orientation of carbohydrates in the pocket. Among these, Arg 53 and Glu 72 make direct interactions with the carbohydrates (Table 6). In the galactose and galactosamine complex structures, the O1, O2, and O3 atoms of the carbohydrate are involved in water-mediated interactions and contribute to the strength of carbohydrate-binding. In the galactose and galactosamine complex structures, the O1 atom is involved in hydrogen bond formation with Pro 85 and Ser 8 from a symmetry-related molecule, while in the lactose and *N*-acetyl-lactosamine complexes, the O2 and C3 atoms are involved in van der Waals crystal-packing contacts. Surface accessibility calculations using the program X-PLOR indicate that the Glc and GlcNAc moieties of lactose and *N*-acetyl-lactosamine are less buried than the galactose moiety; 40% and 46% of the lactose and *N*-acetyl-lactosamine accessible surfaces are buried in the hGal-7 CRD; only a few contacts are made with the galactose, mediated mainly through the O3 hydroxyl group's interaction with Arg 53. The glycosidic torsion angles [defined for the Gal (β1−4) GlcNAc linkage as φ = O5−C1−O4′−C4′ and ψ = C1−O4′−C4′−C5′] are −59° and −115° in the lactose and −65° and −115° in the *N*-acetyl-lactosamine complex structures, respectively.

Comparison of the carbohydrate-binding pockets in Gal-1 and Gal-2 with that of hGal-7 reveals significantly different architecture (Figure 6). Even though all of the carbohydrate-binding residues are conserved, the loop connecting the strands S4 and S5 is longer in Gal-1 than in Gal-2 or Gal-7, which are similarly sized in this region (Figures 3 and 6E). On the other hand, the corresponding region in Gal-10 is significantly different from these three galectins. The shortening of the loop structure appears to have some structural implications for carbohydrate-binding specificity in hGal-7 and has resulted in a significantly different set of hydrogen bond interactions in this site. Superposition of the carbohydrate-binding site reveals that Arg 31 in hGal-7 occupies the position of His 52 in Gal-1, which is located about 3.1 Å away from the carbohydrate moiety and hence this arginine residue could form part of the carbohydrate-binding region (Figure 6E).

Concluding Remarks. Even though the precise roles of the galectins have not been established, a number of studies have suggested that these proteins may be involved in cell−cell and cell−matrix interactions. Their expression is

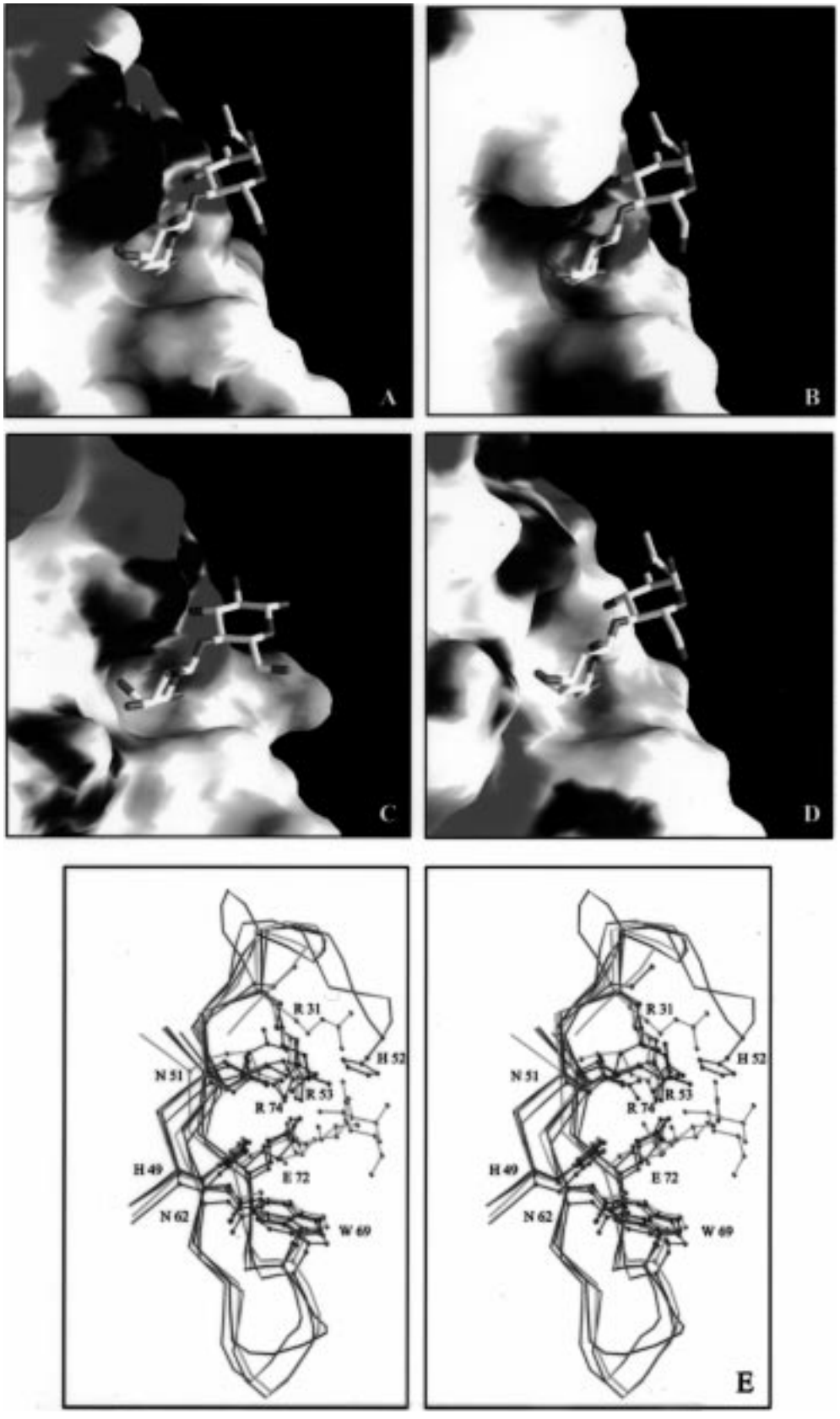

FIGURE 6: (A−D) Electrostatic surface potential map of the carbohydrate recognition pocket in hGal-7 with bound *N*-acetyl-lactosamine (A), Gal-1 with bound *N*-acetyl-lactosamine (B), Gal-2 with bound lactose (C), and Gal-10 with *N*-acetyl-lactosamine (based on modeling study) (D). Electrostatic potential was calculated with the program GRASP and is color coded on surface from blue (∼5 kt/e) to red (∼−4 kt/e) (*45*). (E) Comparison of carbohydrate-binding sites of Gal-1 (red), Gal-2 (green), hGal-7 (gray), and Gal-10 (cyan), shown in stereoview.

developmentally regulated and temporally associated with the appearance of specific carbohydrate-containing structures in tissues and cells (*1*, *41*). hGal-7 is no exception. All of the experimental data to date suggests that this molecule could make interactions with the polylactosamine oligosaccharides of extracellular matrix substrates. Our study on the

free and the carbohydrate-bound forms of hGal-7 now provides a structural basis for carbohydrate recognition. The results of this study may be useful in identifying naturally occurring glycoconjugates, especially glycoproteins, that interact with hGal-7.

The available structural and functional data on different galectins show that each galectin has a different specificity for carbohydrates. Using the structural data, it is now possible to identify the mechanisms governing galectin−carbohydrate specificities, and this should prove useful in designing more efficient ligand analogues to be used in functional assays. For example, in the case of Gal-3, different subsites involved in sugar recognition have been predicted (*42*). Preliminary structural data on Gal-3 and Gal-4 CRDs (*18*) have already indicated that these galectins do not obey the packing arrangements of Gal-1 and Gal-2, and this has implications for carbohydrate recognition. Thus, as new galectin structures are elucidated, new clues into the extracellular function of these molecules through carbohydrate recognition may become available.

NOTE ADDED IN PROOF

After this paper was submitted, the crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1 Å resolution was reported (*46*). The overall fold is considerably similar to hGal-7. The conformation of the CRD also appears to be conserved in Gal-3 and hGal-7 with the exception of residue Arg-74 (hGal-7).

ACKNOWLEDGMENT

We thank the staff at the Synchrotron Radiation Source, Daresbury, U.K., the staff at the EMBL Outstation, Hamburg, Germany for their excellent support, and Dr. Tassos Papageorgiou for the help during X-ray data collection. We also thank Dr. Daniel Holloway for the constructive criticisms of the manuscript.

REFERENCES

1. Barondes, S. H., Castronovo, V., Cooper, D. N. W., Cummings, R. D., Drickamer, K., Feizi, T., Gitt, M. A., Hirabayashi, J., Hughes, C., Kasai, K., Leffler, H., Liu, F. T., Lotan, R., Mercurio, A. M., Monsigny, M., Pillai, S., Poirer, F., Raz, A., Rigby, P. W. J., Rini, J. M., and Wang, J. L. (1994) *Cell 76*, 597−598.
2. Barondes, S. H., Cooper, D. N., Gitt, M. A., and Leffler, H. (1994) *J. Biol. Chem. 269*, 20807−20810.
3. Drickamer, K. (1988) *J. Biol. Chem. 263*, 9557−9560.
4. Leffler, H. (1997) *Trends Glycosci. Glycotechnol. 45*, 9−19.
5. Hirabayashi, J., and Kasai, K. (1993) *Glycobiology 3*, 297−304.
6. 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.
7. Magnaldo, T., Bernerd, F., and Darmon, M. (1995) *Dev. Biol. 168*, 259−271.
8. Magnaldo, T., and Darmon, M. (1997) *Trends Glycosci. Glycotechnol. 9*, 95−102.
9. Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W., and Vogelstein, B. (1997) *Nature 389*, 300−305.
10. Lu, J., Pei, H., Kaeck, M., and Thompson, H. J. (1997) *Mol. Carcinog. 20*, 204−215.
11. Rini, J. M. (1995) *Curr. Opin. Struct. Biol. 5*, 617−621.
12. Weis, W. I., and Drickamer, K. (1996) *Annu. Rev. Biochem. 65*, 441−473.
13. Gabius, H.-J. (1997) *Eur. J. Biol. 243*, 543−576.
14. Lis, H., and Sharon, N. (1998) *Chem. Rev. 98*, 637−674.
15. Bourne, Y., Bolgiano, B., Liao, D. L., Strecker, G., Cantau, P., Herzberg, O., Feizi, T., and Cambillau, C. (1994) *Nat. Struct. Biol. 1*, 863−870.
16. Liao, D. I., Kapadia, G., Ahmed, H., Vasta, G. R., and Herzberg, O. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1428−1432.
17. Lobsanov, Y. D., Gitt, M. A., Leffler, H., Barondes, S. H., and Rini, J. M. (1993) *J. Biol. Chem. 268*, 27034−27038.
18. Lobsanov, Y. D., and Rini, J. M. (1997) *Trends Glycosci. Glycotechnol. 9*, 145−154.
19. Leonidas, D. D., Elbert, B. L., Zhou, Z., Leffler, H., Ackerman, S. J., and Acharya, K. R. (1995) *Structure 3*, 1379−1393.
20. Dyer, K. D., and Rosenberg, H. F. (1996) *Life Sci. 58*, 2073−2082.
21. Christensen, J. H., Hansen, P. K., Lillelund, O., and Thogersen, H. C. (1991) *FEBS Lett. 281*, 181−184.
22. Studier, F. W., and Moffatt, B. A. (1986) *J. Mol. Biol. 189*, 113−130.
23. Nagai, K., and Thorgensen, H. C. (1987) *Methods Enzymol. 153*, 461−481.
24. Bradford, M. (1976) *Anal. Biochem. 72*, 248−254.
25. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
26. French, S., and Wilson, K. S. (1978) *Acta Crystallogr., Sect. A 34*, 517−525.
27. LaFortelle, E. D., and Bricogne, G. (1997) *Methods Enzymol. 276*, 472−494.
28. Abrahams, J. P. (1997) *Acta Crystallogr., Sect. D 53*, 371−376.
29. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr., Sect. A 47*, 110−119.
30. Navaza, J. (1994) *Acta Crystallogr., Sect. A 50*, 157−163.
31. Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science 235*, 458−460.
32. Read, J. (1986) *Acta Crystallogr., Sect. A 42*, 140−149.
33. Brünger, A. T. (1992) *Nature 355*, 472−475.
34. Jiang, J.-S., and Brünger, A. T. (1994) *J. Mol. Biol. 243*, 100−115.
35. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
36. Kabsch, W., and Sanders, C. (1983) *Biopolymers 22*, 2577−2637.
37. Stuart, D. I., Levine, M., Muirhead, H., and Stammers, D. K. (1979) *J. Mol. Biol. 134*, 109−142.
38. Hubbard, S. J., Campbell, S. F., and Thornton, J. M. (1991) *J. Mol. Biol. 220*, 507−530.
39. Abbott, W. M., and Feizi, T. (1991) *J. Biol. Chem. 266*, 5552−5557.
40. Gitt, M. A., Massa, S. M., Leffler, H., and Barondes, S. H. (1992) *J. Biol. Chem. 267*, 10601−10606.
41. Hirabayashi, J., and Kasai, K. (1994) *Glycoconjugate J. 11*, 437−442.
42. Henrick, K., Bawumia, S., Barboni, E. A. M., Mehul, B., and Hughes, R. C. (1998) *Glycobiology 8*, 45−57.
43. McDonald, I. K., and Thornton, J. M. (1994) *J. Mol. Biol. 238*, 777−793.
44. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
45. Nicholls, A., Bharadwaj, R., and Honig, B. (1991) *Biophys. J. 239*, 423−433.
46. Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S. H., and Rini, J. M. (1998) *J. Biol. Chem. 273*, 13047−13052.

BI981056X