J. Mol. Model. **1997**, *3*, 325 - 331

A Combined Molecular Modelling and CIDNP Study of Similarities in the Pattern of Ligand Binding in Mammalian and Avian Galectins[$]

Emadeddin Tajkhorshid[1,*], Hans-Christian Siebert[2,3], Maria Burchert[2], Herbert Kaltner[2], Gian Kayser[1], Claus-Wilhelm von der Lieth[1], Robert Kaptein[3], Johannes F. G. Vliegenthart[3], and Hans-Joachim Gabius[2]

[1]Zentrale Spektroskopie, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany; Tel: +49 6221 422339; Fax: +49 6221 422333 (E.Tajkhorshid@DKFZ-Heidelberg.de)

[2]Institut für Physiologische Chemie, Tierärztliche Fakultät, Ludwig-Maximilians-Universität, Veterinärstr. 13, D-80539 München, Germany

[3]Bijvoet Center for Biomolecular Research, Utrecht University, P.O. Box 80.075, NL-3508 TB Utrecht, The Netherlands

Received: 20 June 1997 / Accepted: 24 July 1997 / Published: 11 August 1997

Abstract

Galectins (Galactose binding lectins) from bacteria, plants and animals have been shown to possess tyrosine or tryptophan residues that form hydrophobic contacts with their ligands in the binding sites. At the present time, the X-ray structures of only two galectins from human and bovine tissues are known. In the present study we applied X-ray data of bovine heart galectin-1 as a template for homology modelling of a number of galectins from mammalian and avian tissues. The conservation of one tryptophan and at least one histidine in binding pocket can be observed from the comparison of the model structures.

We also show that it is possible to obtain information of the architecture of the binding pocket of several galectins in solution using CIDNP (Chemically Induced Dynamic Nuclear Polarisation) techniques. The CIDNP approach offers a possibility to analyse these lectins in solution thereby providing supplementary information to the available X-ray data. All studied galectins show comparable alterations when they are recorded by CIDNP-technique in the absence and in the presence of their specific carbohydrate ligands.

Keywords: Surface accessibility, Molecular Dynamics, Homology modelling, Carbohydrate binding

[$] *Presented at the 11. Molecular Modeling Workshop, 6 -7 May 1997, Darmstadt, Germany*

* *To whom correspondence should be addressed*

```
      .    :    .    :    .    :    .    :    .    :    .    :   60
1:  ACGLVASNLNLKPGECLRVRGEVAADAKSFLLNLGKDDNNLCLHFNPRFNAHGDVNTIV   59
2:  ACGLVASNLNLKPGECLRVRGEVAPDAKSFVLNLGKDSNNLCLHFNPRFNAHGDANTIV   59
3: SCQGPVCTNLGLKPGQRLTVKGIIAPNAKSFVMNLGKDSTHLGLHFNPRFDAHGDVNLIV   60
4: MEQGLVVTQLDVQPGECVKVKGKILSDAKGFSVNVGKDSSTLMLHFNPRFDCHGDVNTVV   60
5:                  ITIMGTVKPNANRIVLDF-RRGNDVAFHFNPRFNEN-NRRVIV  185

      .    :    .    :    .    :    .    :    .    :    .    : 120
1: CNSKDAGAWGAEQRESAFPFQPGSVVEVCISFNQTDLTIKLPDGYEFKFPNRLN-LEAIN  118
2: CNSKDGGAWGTEQREAVFPFQPGSVAEVCITFDQANLTVKLPDGYEFKFPNRLN-LEAIN  118
3: CNSKKMEEWGTEQRETVFPFQKGAPIEITFSINPSDLTVHLP-GHQFSFPNRLG-LSVFD  118
4: CNSKEDGTWGEEDRKADFPFQQGDKVEICISFDAAEVKVKVPEV-EFEFPNRLG-MEKIQ  118
5: CNTKQDNNWGKEERQSAFPFESGKPFKIQVLVEADHFKVAVNDAHLLQYNHRMKNLREIS  245

      .    :    .    :    .    :    .    :    .    :    .    : 180
1: YLSAGGDFKIKCVAFE                                             134
2: YMAADGDFKIKCVAFD                                             134
3: YFDTHGDFTLRSVSWE                                             134
4: YLAVEGDFKVKAIKFS                                             134
5: QLGISGDITLTSANHAMI                                           263
```

Figure 1. *Sequence Alignment of four modelled galectins and the bovine heart galectin which was used as the template for modelling. CIDNP sensitive aromatic amino acids are presented in bold face.*
1: galectin-1 (bovine)
2: galectin-1 (human)
3: CL-14 galectin from chicken intestine
4: CL-16 galectin from chicken liver
5: murine galectin-3, C-terminal part

Introduction

Galactoside specific lectins from bacteria, plants and animals have been shown to possess tyrosine or tryptophan residues which can form hydrophobic contacts with the ligands in the binding sites [1-3]. However, because only a few number of three dimensional structures of galectins-carbohydrate complexes have been experimentally solved [4-6], the molecular details of their interaction with different sugar molecules have yet to be understood. The crystal structures of galectin-1 from bovine heart and spleen emphasise the importance of Trp69 for this purpose [4-6]. Among the selected animal galectins, this tryptophan moiety is conserved, implying an indispensable functional significance (Figure 1).

X-ray crystallography of proteins provides detailed information on the three-dimensional structure. However, the precision of the derived spatial parameters of the protein in the crystal, should not be considered as proof for the solution structure. Consequently, it is desirable to employ a sensitive spectroscopic method which in combination with molecular modelling can verify distinct properties determined by crystallographic analysis for the solution structure. Laser photo CIDNP (chemically induced dynamic nuclear polarisation) is an NMR-based technique [7-9]. The generation of signals in this method depends on the surface accessibility of the side chains of Tyr, Trp and His residues as schematically shown in Figure 2. Therefore, this technique is a potentially valuable tool to investigate surface epitopes involving Tyr, Trp and His residues whose surface exposure can well be of notable physiological importance. Signals of binding-involved Tyr, Trp and His residues can be suppressed by the addition of specific ligands. Besides testing of the structural aspects predicted for a protein in solution CIDNP results can be taken into account to judge the reliability of knowledge-based homology modelling approaches [10].

Molecular dynamics simulations can provide useful information about the dynamic behaviour of the molecular systems under study. In the present work, we have studied the involvement of aromatic amino acids in the binding site of a number of animal and human galectins using CIDNP measurements. The experimental data have been verified by calculated surface accessibility of the amino acids after different sets of molecular dynamics simulations. Herein, we docu-

Table 1. *Calculated surface accessibility of aromatic amino acids of chicken intestine galectin. The calculated surface areas (Å²) are given as the average of the lowest potential energy 10 snapshots of the restrained (M1) or relax (M2) dynamics simulations.*

	M1	SD	M2	SD
His 41	53.2	13.9	52.9	8.8
His 45	25.3	11.5	31.6	10.8
His 53	76.8	7.7	86.7	11.9
His 100	53.6	9.4	75.3	9.1
His 104	54.1	11.1	40.0	10.0
His 123	67.5	10.4	85.6	9.2
Trp 69	89.9	5.6	90.3	8.4
Trp 133	85.3	20.2	28.7	11.5
Tyr 119	36.7	9.7	51.0	8.2

Table 2. *Calculated surface accessibility of aromatic amino acids of chicken liver galectin. The calculated surface areas (Å²) are given as the average of the lowest potential energy 10 snapshots of the restrained (M1) or relax (M2) dynamics simulations.*

	M1	SD	M2	SD
His 45	29.7	7.1	20.0	6.7
His 53	79.1	8.5	38.7	8.5
Trp 69	84.4	7.9	97.4	10.5
Tyr 119	34.6	8.5	43.0	10.8

ment that the CIDNP-results are readily reconcilable with the calculated surface properties for bovine galectin-1, as obtained by crystallography, and for four related galectins, as obtained by computer-assisted homology-modelled structures.

Materials and Methods

Knowledge Based Homology Modelling

The starting structures of the galectins were either taken from the Brookhaven Protein Data Bank or generated using knowledge-based homology modelling. Both **FastA** and **BLAST** algorithms report for high scores of similarity between the sequence of studied galectins and the sequence of bovine heart galectin (entries 1sla, 1slb and 1slc in the Brookhaven Protein Data Bank). Therefore we started to build a three dimensional model for the galectins based on the known structure of bovine heart galectin. The sequence alignment which was used for further framework construction is shown in Figure 1. The alignment conserves particular amino acids which are located in binding pocket of the galectins. The next step of homology modelling computations were performed in the GLAXO institute for molecular biology SA, using the Swiss-Model Automated Protein Modelling service which makes use of ProMod (Protein Modelling tool) [11, 12]. The generation of the H atoms, and automatic assignment of partial charges for each atom of the molecule was accomplished using the consistent valence force field (cvff) in INSIGHTII software.

To gain insight into the conformational flexibility and in order to simulate the dynamic behaviour of the protein structures, each molecular ensemble was subjected to restrained and/or relaxed molecular dynamics simulations. In the case of restrained molecular dynamics, position constraints were imposed to the α-carbon atoms of the amino acids using a force constant of 100 kcal/Å². The molecular dynamics calculations were performed using DISCOVER program running on an IBM-SP2 parallel machine. All simulations were carried out using the cvff at a temperature of 300 K and an integration step of 1 fs. A cut-off distance of 10 Å and a dielectric constant of ε = 4.0 were used in all calculations. 10 conformers with the lowest potential energy were selected from the complete 100 ps of production time excluding the first 10 ps of equilibration. These conformations were used to calculate average values of surface accessibilities. In all cases the selected conformations were equally spread out over

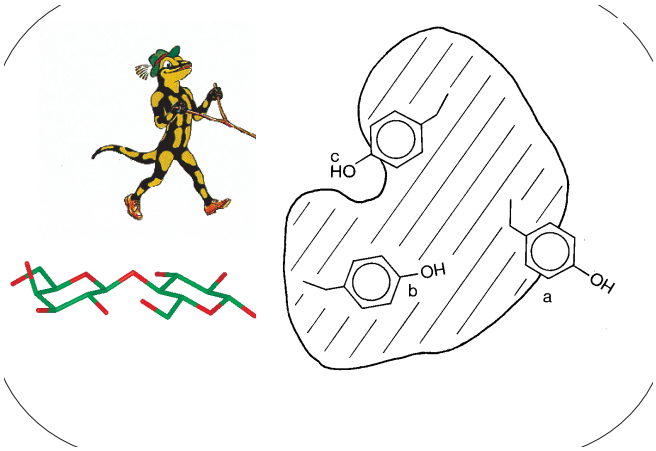

Figure 2. *The generation of signals in laser photo CIDNP method depends on the surface accessibility of the side chains of tyrosine, tryptophan and histidine.*
a. Completely exposed (strong CIDNP signal)
b. buried (No CIDNP signal)
c. Exposed in binding site (CIDNP signal can be suppressed by presence of ligand)

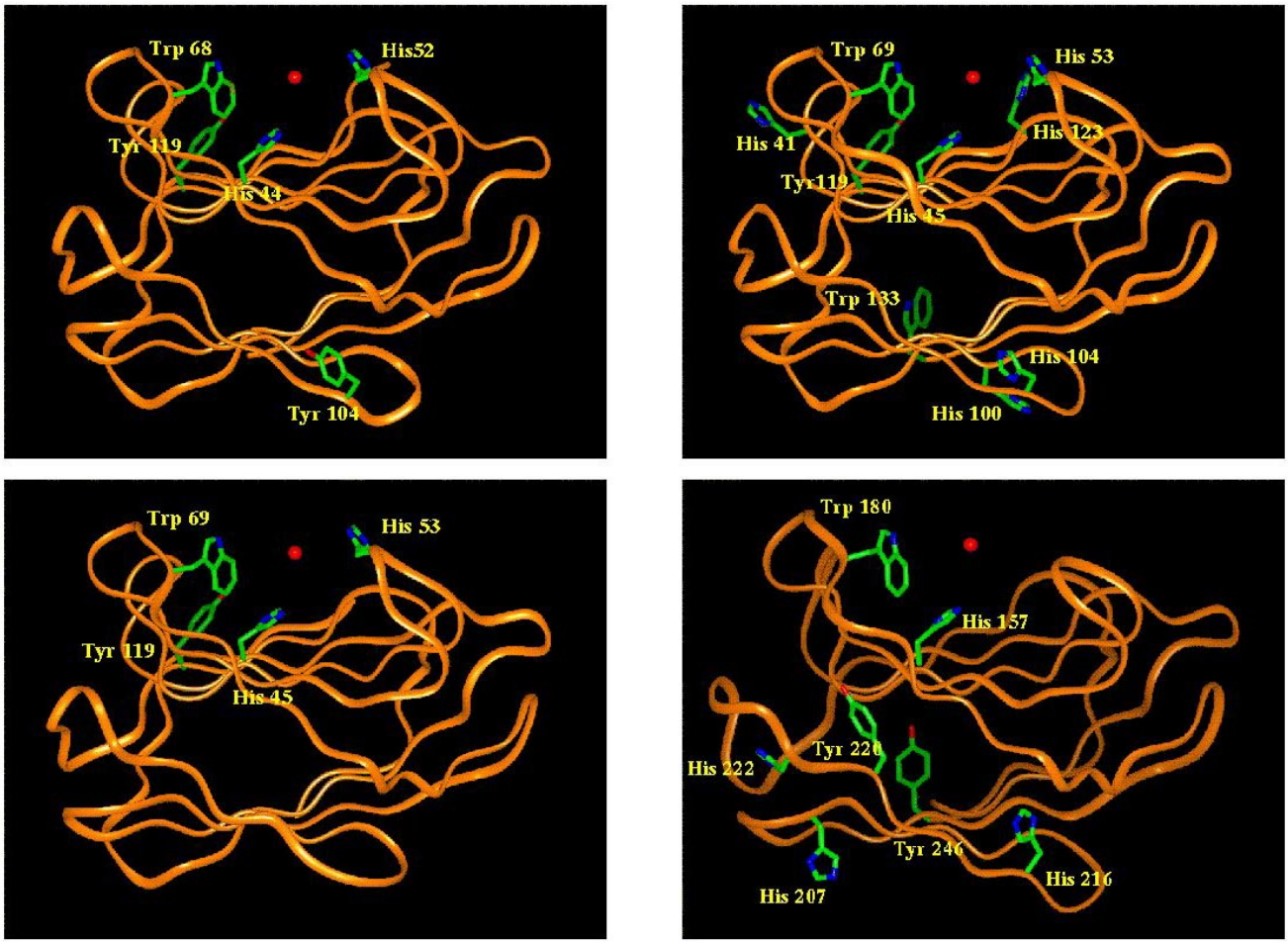

Figure 3. *CIDNP sensitive aromatic amino acids (Tyr, Trp & His) and their relative position to the binding site in the model structure of studied galectins. The position of the glycosidic oxygen in the disaccharide ligand is represented by the red sphere.*
top/left: Human galectin 1
top/ritght: Chicken intestine galectin
bottom/left: Chicken liver galectin
bottom/right: Murine galectin 3

the complete duration of the simulation and did not show cluster within a specific time period.

The surface accessibility of the principally CIDNP-reactive side chains were calculated with the help of the Connolly program. The calculated values are equivalent to the accessible surface of the side chain atoms of the aomatic amino acids obtained by smoothening the van der Waals surface with a test sphere that has the average radius of the solvent water (1.5 Å).

CIDNP-Method

The principle of the method has been described elsewhere [13, 14]. ^{1}H CIDNP NMR measurements were carried out at 360 MHz on a Bruker AM-360 NMR spectrometer. A typical experimental setting consisted of a presaturation pulse of 1 s, a light pulse of 0.5 s (5 W), a resonance frequency pulse (90º flip angle) of 5 µs, an acquisition time of 1 s and a delay of 5 s. In general, an adequate signal-to-noise ratio was obtained after 16 or 32 light scans. The resulting light spectrum was subtracted from the dark spectrum, thereby establishing the CIDNP difference spectrum, containing only signals of polarised residues. The experiments were performed at pH=8.2 (pD=8.6).

Results and Discussion

The structural characteristic under investigation is the accessibility of the ring systems of Tyr, Trp and His for a laser light-excitable dye, which leads to the generation of characteristic laser photo CIDNP-signals [7-9]. Beside the general relevance of a surface structure element, which may display the freedom for mobility in solution, the focus on aromatic side chains is especially interesting for galactoside-binding proteins. In their binding pocket animal lectins of this specificity can harbour such a moiety for exerting affinity-enhancing hydrophobic interaction [1-3]. Thus, ligand presence is expected to influence the respective signal intensity,

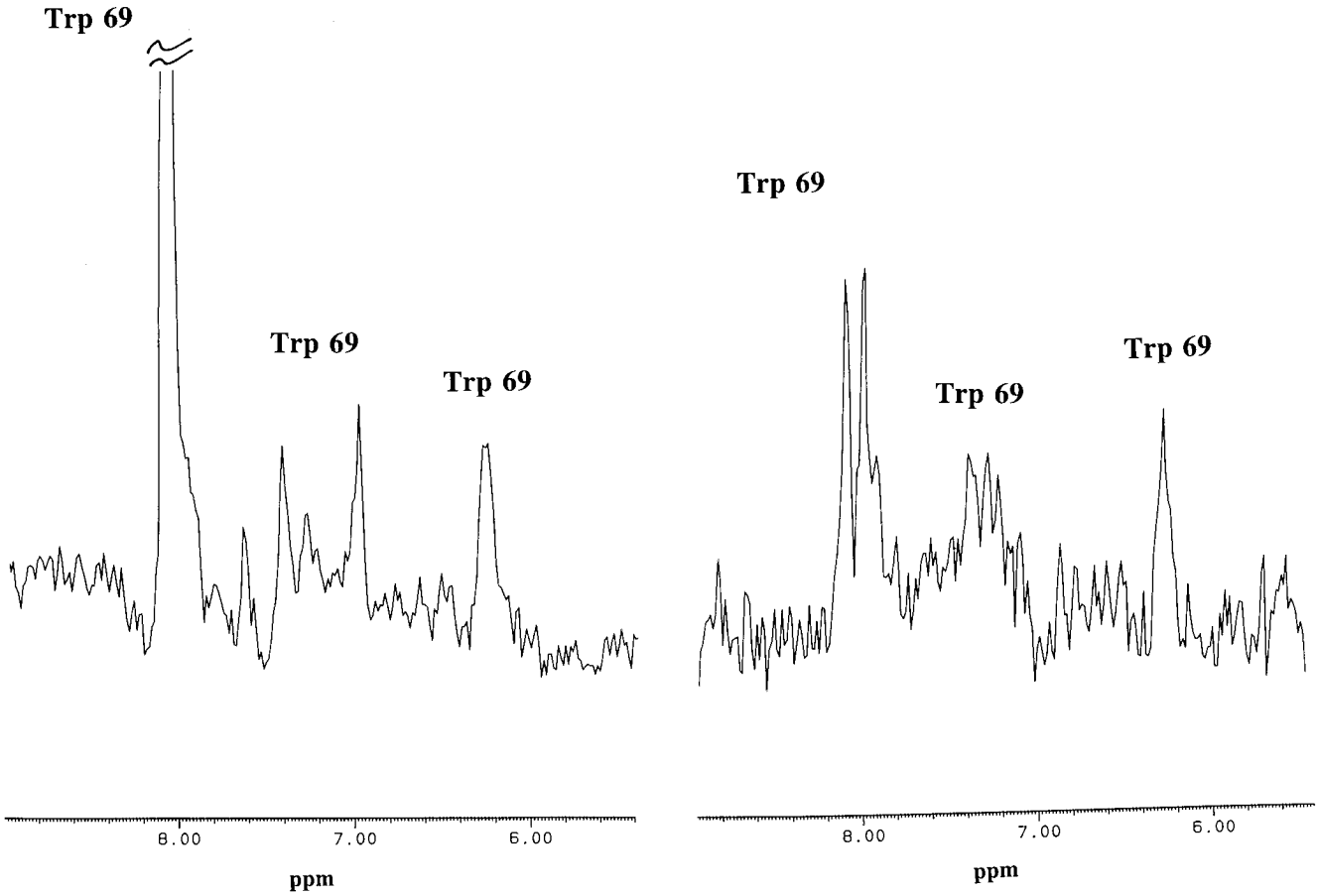

Figure 4. *Laser photo CIDNP difference spectra of chicken liver galectin in the absence (left) and presence (right) of lactose obtained at pH=8.2. Since His residues generate also very pronounced CIDNP signals at this pH, a contribution of the His residues to the highly intense CIDNP signal cannot be excluded.*

a further measure of the sensitivity and specificity of the technique.

The sequence of bovine galectin (Figure 1) whose 3D structure has been determined by X-ray crystallography contains 5 CIDNP sensitive residues: one tryptophan, two tyrosines and two histidines. The tryptophan and at least one of the two histidine residues are among the conserved part in the sequence of the galectins (Figure 1) and are located in the binding pocket of the proteins. The availability of two experimentally determined galectins enables to calculate the surface accessibility of all 5 residue and to compare these data directly with those obtained by CIDNP-measurements [15].

Our previous results from the CIDNP measurements and theoretically calculated surface accessibilities [15] indicate that the solution structure of bovine galectins show surface features which are in accordance with the crystallographic data. When lactose was added, the CIDNP signal for Trp69 was significantly weakened, concurrent with the complex-stabilising role of this residue in the crystal structure [15].

The sequence of human galectin-1 is very similar to that of the bovine homologue (Figure 1). To derive a data set for comparison to the experimental results, knowledge-based homology modelling was applied to human galectin-1. This close structural similarity between bovine and human galectin-1 translated into comparable spectra of the two lectins in the absence as well as in the presence of lactose. The relative positions of principally responsive side chains to the ligand-binding site illustrate the consistency of the results from the experiment and the modelling (Figure 3).

To monitor to what extent theoretical predictions will be valid for more distant members of the galectin family, two proteins of avian origin were examined. The sequence identity between mammalian and avian galectins is slightly above 50%, compared to more than 80% for inter-mammal galectins (Figure 1). Relevant conservation of amino acids comprises the placement of the tryptophan moiety, of two histidine residues (His45 and His53) as well as one tyrosine unit (Tyr119), when the galectins from chicken liver and intestine are considered (Figure 1).

Molecular dynamics simulations were performed in order to simulate the dynamic behaviour of the proteins which give a closer estimation of the molecular characteristics within the experimental conditions. Application of completely re-

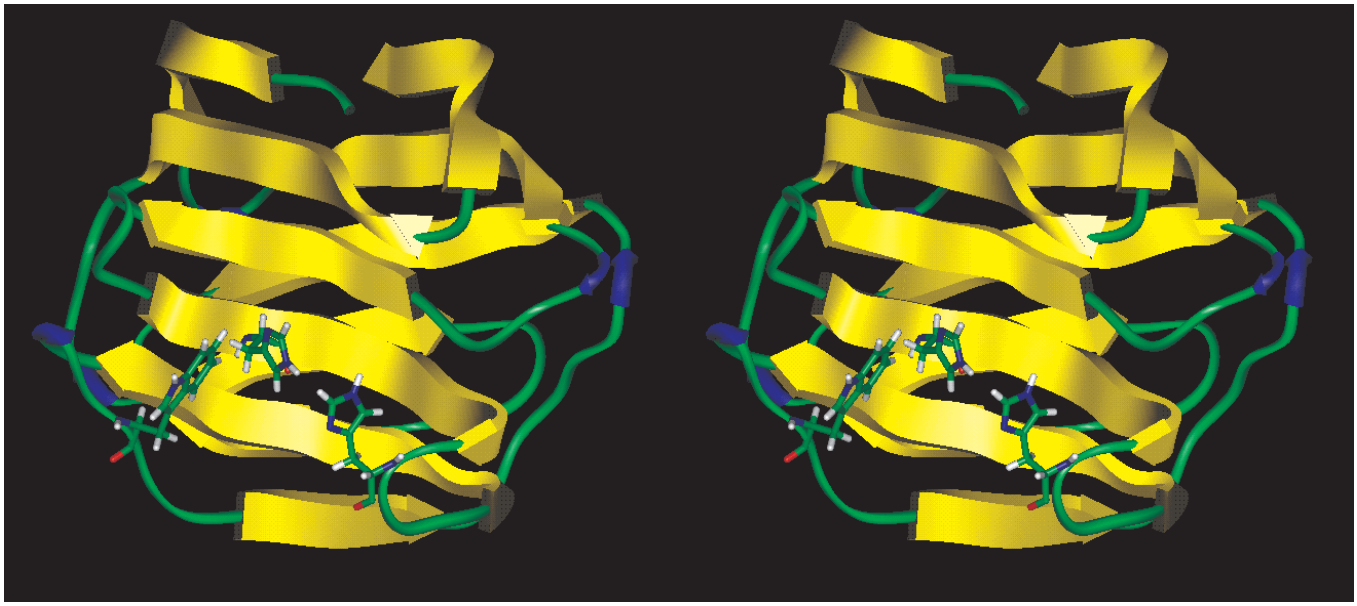

Figure 5. *Presentation of secondary structure and binding site of chicken intestine (left) and liver (right) galectins. Only CIDNP sensitive aromatic amino acids in the binding pocket are shown.*

laxed molecular dynamics simulation is preferred whenever the backbone structures of the studied proteins are stable enough. Examination of trajectories derived from relaxed molecular dynamics simulations revealed that the backbone structures of the studied models are stable within the applied simulation conditions and retain their main folding patterns. Application of relaxed molecular dynamics simulation resulted in slightly different values for the calculated average surface accessibility of the studied amino acids, compared to the restrained molecular dynamics simulations. However in most of the cases the differences were not significant to influence the conclusions

The surface accessibilities were calculated on the basis of the homology-modelled structures for the CIDNP-reactive side chains for the two chicken galectins (Figure 3, Table 1, 2). The pronounced surface presentation of Trp69 (Figures 4, 5, Table 1, 2) apparently is maintained. The respective signal intensity was sensitive to the presence of lactose for both avian galectins (Figure 4). Ligand presence has only at most a very minor influence of the dye-binding capacity of the single tyrosine unit in position 119, graphically displayed in the computer models (Figure 3). The absence of any significant signal (Figure 4) for tyrosine in the CIDNP spectra of chicken galectin indicates that the Tyr119 is not sufficiently exposed to the surface of the protein to produce CIDNP signals. This conclusion can also be confirmed by the calculated average values for the surface accessibility of this residue during molecular dynamics simulations (Table 1, 2).

The conservation of Trp68/69 in animal galectins, the deleterious influence of mutagenesis at this site on galectin-1 and the indication for stacking with the hydrophobic B-face of galactose in the crystal structures of galectin-1/2 imply a general role of this residue in complex stabilisation [5, 6, 16-19]. This evidence is clearly corroborated by the presented results, extending the concept of functional conservation from the proto-type mammalian galectins to related avian proteins. In case not too many CIDNP-reactive residues occur in the protein's sequence, this method has the advantage to enable correlations between the crystal structure/homology modelling parameters and the spectra for residues not shielded by the ligand. The structural parameters detected by the CIDNP methods can be correlated with detailed computational structure predictions such as the shielding of the surface exposed tryptophan after ligand addition in all studied galectins and buried tyrosine residue in the case of chicken liver galectin. As documented for the galectins, the two respective data sets are reconcilable. An indication for a ligand-induced conformational change, as e.g. noted for two animal C-type galactoside-binding lectins which show either increased surface hydrophobicity or trypsin susceptibility in the presence of lactose [20, 21], was only in case of chicken intestine galectin detectable (Figures 3, 5).

References

1. Sharon, N. *Trends Biochem. Sci.* **1993**, 18, 221.
2. Rini, J. M. *Ann. Rev. Biophys. Biomol. Struct.* **1995**, 24, 551.
3. Gabius, H. -J. *Eur. J. Biochem* **1997**. 243, 543.
4. Shaanan, B.; Lis, H.; Sharon, N. *Science* **1991,** 254, 862.
5. Liao, D.; Kapadia, G.; Ahmed, H.; Vasta, G. R.; Herzberg, O. *Proc. Natl. Acad. Sci. USA* **1994**, 91, 1428.

6. Bourne, Y.; Bolgiano, B.; Liao, D.; Strecker, G.; Cantau, P.; Herzberg, O.; Feizi, T.; Cambillau, C. *Nature Struct. Biol.* **1994**, 1, 863.
7. Kaptein, R.; Dijkstra, K.; Nicolay, K. *Nature* **1978**, 274, 293.
8. Kaptein, R. In *Biological Magnetic Resonance*; Berliner, L. J., Ed.; Plenum Press: New York, 1982; pp 145-191.
9. Hore, P. J.; Broadhurst, R. W. *Progr. NMR Spectroscopy* **1993**, 25, 345.
10. Johnson, M. S.; Srinivasan, N.; Sowdhamini, R.; Blundell, T. L. *Crit. Rev. Biochem. Mol. Biol.* **1994**, 29, 1.
11. Peitsch, M. C.; Jongeneel, C. V. *Int. Immunol.* **1993**, 5, 233.
12. Peitsch, M. C. *Bio/Technology* **1993**, 13, 658.
13 . Siebert, H.-C.; Tajkhorshid, E.; von der Lieth, C. -W.; Kleineidam, R.; Kruse, S.; Schauer R.; Kaptein, R.; Gabius, H. -J.; Vliegenthart, J. F. G. *J. Mol. Model.* **1996**, 2, 446.
14. Siebert, H.-C.; von der Lieth, C. -W.; Kaptein, R.; Beintema, J. J.; Dijkstra, K.; Nuland N.; Soedjanaatmadja, U. M. S.; Rice, A.; Vliegenthart, J. F. G.; Wright, C. S.; Gabius, H. -J. *Proteins.* **1997**, 28, 268.
15. Siebert, H. -C.; Adar, R.; Arango, R.; Burchert, M.; Kaltner, H.; Kayser, G.; Tajkhorshid, E.; von der Lieth, C. -W.; Kaptein, R.; Sharon, N.; Vliegenthart, J. F. G.; Gabius, H. -J. (submitted)
16. Paroutaud, P.; Levi, G.; Teichberg, V. I.; Strosberg, A. D. *Proc. Natl. Acad. Sci. USA* **1987**, 84, 6345.
17. Abbott, W. M.; Feizi, T. *J. Biol. Chem.* **1991**, 266, 5552.
18. Lobsanov, Y. D.; Gitt, M. A.; Leffler, H.; Barondes, S. H.; Rini, J. M. *J. Biol. Chem.* 1993, 268, 27034.
19. Hirabayashi, J. In *Glycosciences: Status and Perspectives*; Gabius, H. -J., Gabius S., Eds.; Chapman & Hall: London-Weinheim, 1997; pp 355-368.
20. Komano, H.; Kurama, T.; Nagasawa, Y.; Natori, S. *Biochem. J.* **1992**, 284, 227
21. Hatakeyama, T.; Furukawa, M.; Nagatomo, H.; Yamasaki, N.; Mori, T. *J. Biol. Chem.* **1996**, 271, 16915.