ELSEVIER

A molecular model of the carbohydrate recognition domain of a rat macrophage lectin and analysis of its binding site

Jürgen Bajorath

Bristol-Myers Squibb Pharmaceutical Research Institute, Seattle, Washington, and Department of Biological Structure, University of Washington, Seattle, Washington

A three-dimensional model of the carbohydrate recognition domain of a rat macrophage C-type lectin has been constructed by comparative modeling and assessed by inverse folding analysis. Comparative modeling in the presence of low sequence similarity was based on information provided by comparison of X-ray structures and sequence–structure alignments. The sequence–structure compatibility of the model was sound. Its binding site was analyzed in comparison to the X-ray structure of a galactose-specific mutant of the mannose-binding protein. The specificity of the macrophage lectin was discussed in light of mutagenesis data on asialoglycoprotein receptors. © 1996 by Elsevier Science Inc.

Keywords: protein structure prediction, sequence similarity, structural similarity, comparative modeling, model assessment, protein superfamily, computer graphics analysis, macrophage lectin, carbohydrate binding

INTRODUCTION

Members of protein superfamilies are thought to adopt similar global folds, despite sharing only limited sequence similarity. Protein superfamilies appear attractive as targets for comparative modeling,[1] if the structure of at least one member has been determined. It may then be possible to generate approximate three-dimensional models for other members of the family. However, the low level of sequence similarity shared by protein superfamily members, often 30% or less, makes it difficult to generate topologically meaningful alignments relative to structural template(s),[1] and hence the accuracy of such models may be limited and insufficient for more detailed analysis. This is particularly problematic in cases where information from structure comparison is not (yet) available to complement sequence alignments.

Calcium-dependent (C-type) lectins form a protein superfamily that includes a variety of mammalian carbohydrate-binding proteins.[2] C-type lectin domains specifically bind mono- or oligosaccharides in a calcium-dependent fashion and function as carbohydrate recognition modules.[3] The C-type lectin domain of the rat mannose-binding protein (MBP) was the first structure of a C-type lectin domain determined[4] and revealed a previously unobserved protein fold, consisting to ~50% of loops and other extended regions of unusual secondary structure. Structures of trimetric homologs of rat MBP[5,6] and of a mutant form of rat MBP, which is primarily specific for galactose (Gal),[7,8] have also been determined. In addition to these closely related molecules, the structure of the C-type lectin domain of E-selectin, a cell adhesion molecule, became available,[9] and the MBP and E-selectin structures have been compared in detail.[9,10] Thus, information from structure comparison is available to improve the accuracy of multiple sequence alignments of C-type lectins[10] and hence the ability to construct three-dimensional C-type lectin models by comparative model building.

In this study, it was attempted to build an accurate model of the extracellular carbohydrate recognition domain of the rat macrophage lectin (ML).[11] Macrophage lectin is a type II transmembrane glycoprotein receptor and includes a carboxy-terminal C-type lectin domain with specificity for Gal and *N*-acetylgalactosamine (GalNAc).[11] Macrophage lectin is implicated in carbohydrate-dependent endocytosis by macrophages[11] and closely related to the rat hepatic lectin/asialoglycoprotein receptor (RHL),[12] which also binds Gal,

Color Plates for this article are on pages 283 and 284.

Address reprint requests to: Jürgen Bajorath, Bristol-Myers Squibb Pharmaceutical Research Institute, 3005 First Avenue, Seattle, Washington 98121.

Received 19 September 1996; accepted 21 October 1996.

655 Avenue of the Americas, New York, NY 10010

0263-7855/97/$15.00
SSDI S0263-7855(96)00075-6

but preferentially GalNAc.[13] The ML model was generated on the basis of E-selectin/MBP structure comparison and using MBP as structural template. The sequence–structure fitness of the finalized model was assessed by energy profile analysis. The carbohydrate-binding site of the ML model was analyzed in comparison to a galactose-specific mutant of the mannose-binding protein and considering mutagenesis data.

METHODS

Topological sequence comparison

X-ray structures of E-selectin at 2.0-Å resolution[9] and of MBP at 1.7-Å resolution[14] were compared by backbone superposition as described.[10] Briefly, structurally conserved regions were identified by sequential least-squares superposition of backbone segments of increasing length followed by root mean square deviation (rmsd) comparison. Backbone segments that superimposed with an rmsd of less than 1 Å were determined, and a structure-based sequence alignment was generated. This structure-oriented alignment was complemented by sequence comparison with ML and provided the basis for comparative prediction of its structure.

Comparative model building and loop modeling

Structural manipulations and computer graphics analysis were carried out using InsightII (MSI, San Diego, CA). Backbone regions structurally conserved in MBP and E-selectin provided the framework for the ML model and were copied from MBP. With the exception of segment 257–265 (see below), the conformations of regions including insertions and deletions and other structurally variable (loop) regions in MBP and E-selectin were approximated by conformational search with CONGEN.[15] In these calculations, main-chain conformational space was uniformly searched in 30° torsion angle increments, and side-chain conformations were modeled using an iterative search procedure.[15] Conformations with acceptable interatom interactions and potential energy were sampled and, for each modeled loop, the conformation with lowest solvent-accessible surface within 2 kcal of the energy minimum conformation was selected and included in the model. Segment 257–265 was modeled using the corresponding loop conformation in the X-ray structure[8] of a galactose-binding MBP mutant (gbMBP).[7] This loop was included in the model using Insight's Loop Splice routine. Side-chain replacements were modeled via computer graphics in conformations as similar as possible to the original conformation or, for structurally unconstrained positions, using a rotamer search procedure[16] implemented in InsightII. In these rotamer search calculations, an 8-Å cutoff distance was used for nonbonded interactions. Color figures were generated using InsightII (version 95.0) on an SGI Indigo Impact and processed as RGB files.

Model refinement and assessment

The initially assembled model was refined by energy minimization with Discover (MSI, San Diego, CA) until the rms derivative of the energy function was ~1 kcal/Å. In these calculations, a distance-dependent dielectric constant ($1r$) and a 15-Å cutoff distance for the treatment of nonbonded interactions were used. During the energy minimization calculations, residues predicted to participate in the formation of calcium-binding sites were constrained to their original positions. The stereochemical quality of the refined model was confirmed and its sequence–structure compatibility was assessed by energy profile analysis using PROSAII (version 3.0).[17] Pairwise residues interaction energies were calculated using β-carbon interactions and, for graphical representation and analysis of the profiles, a 50-residue window was used for energy averaging at each residue position.[17]

Binding site modeling

The position of the *N*-acetylgalactosamine (GalNAc) ligand in the ML-binding site was modeled by superposition of structurally conserved regions of the gbMBP–GalNAc structure[8] and the ML model, followed by transfer of the ligand in its crystallographic conformation. Mutations in the binding site region were modeled using Insight's Replace command, followed by a search for the lowest energy side-chain rotamer conformation.[16]

RESULTS AND DISCUSSION

Structure-oriented sequence comparison

The sequences of the carbohydrate recognition domains of ML, MBP, and E-selectin are less than 30% identical. Comparison of the MBP and E-selectin X-ray structures reveals significant differences in some regions, although both proteins adopt the same fold. This information was incorporated in a sequence alignment of MBP and E-selectin, which thus reflects the spatial equivalence of residues (Figure 1). The sequence of ML was aligned against this template by matching core residues, structurally constrained positions, and consensus residues in structurally conserved regions. Using these criteria, regions structurally conserved in MBP, E-selectin, and ML could be assigned with confidence. Likewise, it was possible to identify the regions including insertions and deletions, all of which mapped to structurally variable regions (Figure 1). A schematic representation of the C-type lectin fold is shown in Color Plate 1 which highlights the activated secondary structure elements and extended regions of unusual secondary structure.

Model building and assessment

Backbone segments outside structurally conserved regions were considered variable, and their conformations were approximated using the loop modeling techniques described in Methods. The sequence–structure compatibility of the completed ML model was assessed by energy profile analysis,[17] a structure assessment technique belonging to the inverse folding approach.[18] These methods do not detect incorrectly modeled loop or side-chain conformations, which limit prediction accuracy. However, inverse folding techniques have made it possible to analyze the sequence–structure compatibility of a model and thus to assess its overall reliability. As shown in Figure 2, the average residue interaction energy is negative at each position in the model. This profile is consistent with an overall correctly folded

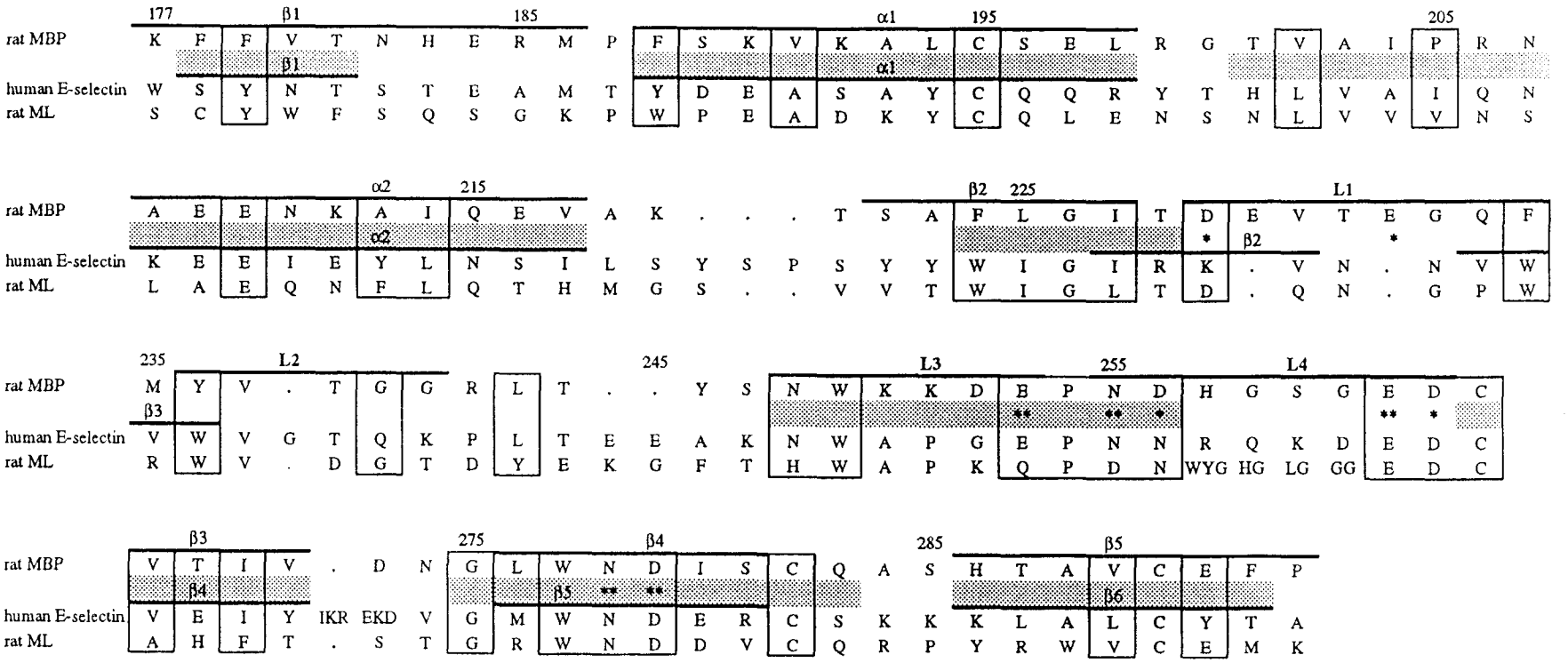

Figure 1. Topological alignment of the C-type lectin domains of E-selectin and MBP, on the basis of structural comparison. The sequence of rat ML was aligned against this template. The major secondary structure elements in MBP and E-selectin are labeled. L1–L4 are extended regions of nonclassic secondary structure in MBP. Residues that participate in the formation of two calcium-binding sites in MBP are labeled with one and two asterisks, respectively. Two asterisks indicate the functional calcium-binding site. Residues thought to be determinants of the C-type lectin fold are boxed. Horizontal shading shows the structurally conserved regions in MBP and E-selectin. Residue numbers are given for ML.

structure with no substantial errors in core regions.[17] Thus, the energy profile analysis carried out here indicated that the ML model was sound and sufficiently accurate to analyze some details.

In a previous report, model building of E-selectin and comparison with its later determined X-ray structure was described.[19] The approach was similar to the one presented herein in that the selectin model was built on the basis of the MBP X-ray structure and structure-oriented sequence comparison. The model was shown to be in good agreement with its X-ray structure, and good agreement was observed in the carbohydrate-binding region.[19] The major difference between the previous and present study is that no information from C-type lectin structure comparison was available at the time of selectin modeling. This information was used to aid in the model building of ML and is thought to further increase modeling accuracy.

The macrophage lectin molecular model and its carbohydrate-binding site

Color Plate 2 shows the ML model superposed on the MBP and E-selectin X-ray structures. The major differences in these structures occur in extended loop regions. The functional calcium-binding site, which is shared by these proteins, is directly involved in carbohydrate binding to MBP.[14] Compared to MBP, the functional calcium coordination sphere in ML contains two residue replacements (ML/MBP: Q253/E and D255/N) that favor the coordination of galactose over mannose.[20] Residues that participate in the formation of the functional calcium-binding site in ML and residues that surround this site are shown in Color Plate 3. The representation outlines the predicted carbohydrate-binding site. Residue W257, part of the glycine-rich loop 257–265, is critical for Gal binding.[7] This region in ML displays a five-residue insertion relative to MBP and E-selectin and correctly positions W257 relative to the calcium coordination sphere. To study the ML binding site in more detail, the position of the GalNAc ligand was modeled on the basis of its orientation and conformation observed in the gbMBP–GalNAc complex.

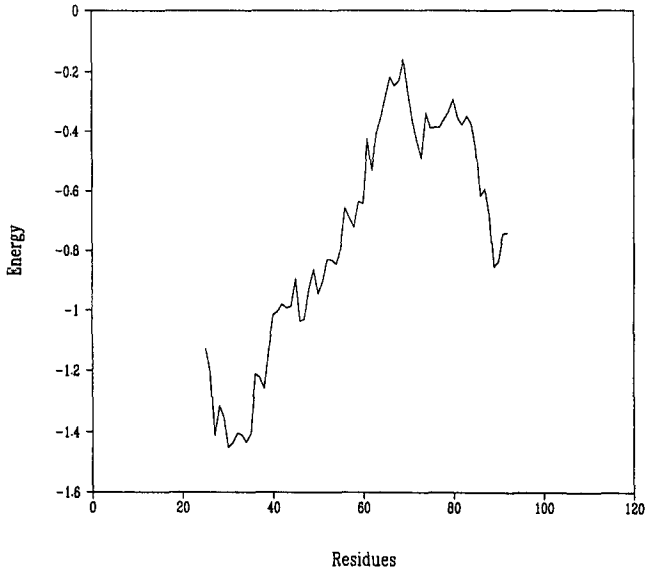

Figure 2. Energy profile of the ML model. Pairwise average residue interaction energy is given in units of E/kT (E, interaction energy [in kcal/mol]; k, *Boltzmann constant;* T, *temperature in degrees Kelvin) and plotted against residue positions. The interaction energy was calculated at each residue position, using a 50-residue window for energy averaging.*

No unfavorable contacts were detected in the model after transfer of the GalNAc ligand from the gbMBP–GalNAc X-ray structure (see Methods). The pentagonal–bipyramidal geometry of the calcium-binding site is conserved. The Gal O-3 and O-4 participate in the formation of the calcium-binding site and share an apical position. The contacts within the calcium coordination sphere in the gbMBP–GalNAc X-ray structure and the ML-GalNAc model are similar. In the model, the O-3–calcium and the O-4–calcium distance are both 2.50 Å, whereas the corresponding crystallographic contact distances are 2.52 and 2.56 Å, respectively. In the X-ray structure, the calcium–ligand distances range from 2.27 to 2.56 Å (average distance, 2.46 Å), and in the model the corresponding distances range from 2.30 to 2.58 Å (average, 2.45 Å). Interactions between the Gal moiety, the calcium coordination sphere, and residue W257 anchored the ligand in the calcium-binding site and determined its orientation. Only the *N*-acetyl position was slightly adjusted. Color Plate 4 shows a detailed view of this model. However, the modeled complex can be considered only as a first approximation and must be interpreted with caution.

Structure–function aspects

On the basis of the model, binding of the GalNAc *N*-acetyl group is thought to involve a subsite in ML formed by residues H270, R284, and Y286, which are, on average, at a distance of ~8 Å from the functional calcium. The location of this region approximately corresponds to the carbohydrate-binding site in E- and P-selectin.[9,21,22] Mutagenesis studies on MBP[7,20] and asialoglycoprotein receptors[13] have identified some residues important for Gal and GalNAc binding. In the model, residue H270, conserved in ML and RHL and critical for GalNAc binding,[13] is within van der Waals contact distance of the *N*-acetyl group. Residue H270 is predicted to correspond spatially to a glutamic acid in the selectins, which is important for carbohydrate binding.[21] Rat hepatic lectin has a greater preference for binding GalNAc than ML, which is mainly a consequence of four residue changes (ML/RHL: V222/N, A250/R, K252/G, S273/T).[13] The model suggested that only two of these mutations, A250/R and V222/N, affect ligand binding directly. However, two other residues, Y286 and R284, both conserved in ML and RHL, are within contact distance of each other and the *N*-acetyl group (Color Plate 4). Asparagine instead of valine at position 222 brings the asparagine side chain to within contact distance of both the *N*-acetyl group of GalNAc and the side chains of Y286 and H270. This network of interactions provides an explanation for the observed differences in GalNAc binding. The side chain of a serine at position 222, as seen in gbMBP, which does not show RHL-like selectivity for GalNAc,[8] is too short to form comparable interactions.

SUMMARY

The ML molecular model was constructed in the presence of low sequence identity with other C-type lectins. This was possible by combining sequence and structure comparison. Inverse folding analysis suggested that the accuracy of the model was sufficient to analyze some of its details. The predicted ML binding site was inspected, and residues important for carbohydrate binding were mapped. The majority of residues in the binding site are part of conserved regions of the C-type lectin fold. Thus, residues at these positions may be of more general importance to modulate carbohydrate specificities of members of the C-type lectin superfamily.

REFERENCES

1 Bajorath, J., Stenkamp, R., and Aruffo, A. *Protein Sci.* 1993, **2,** 1798
2 Drickamer, K. *J. Biol. Chem.* 1988, **263,** 9557
3 Drickamer, K. *Curr. Opin. Struct. Biol.* 1993, **3,** 393
4 Weis, W.I., Kahn, R., Fourme, R, Drickamer, K., and Hendrickson, W.A. *Science* 1991, **254,** 1608
5 Sheriff, S., Chang, C.Y., and Ezekowitz, R.A.B. *Nature Struct. Biol.* 1994, **1,** 789
6 Weis, W.I. and Drickamer, K. *Structure* 1994, **2,** 1227
7 Iobst, S.T. and Drickamer, K. *J. Biol. Chem.* 1994, **269,** 15512
8 Kolatkar, A.R. and Weis, W.I. *J. Biol. Chem.* 1996, **271,** 6679
9 Graves, B.J., Crowther, R.L., Chandran, C., Rumberger, J.M., Li, S., Huang, K.-S., Presky, D.H., Familetti, P.C., Wolitzky, B.A., and Burns, D.K. *Nature (London)* 1994, **367,** 532
10 Bajorath, J. and Aruffo, A. *Protein Sci.* 1996, **5,** 240
11 Ii, M., Kurata, H., Itoh, N., Yamashina, I., and Kawasaki, T. *J. Biol. Chem.* 1990, **265,** 11295
12 Drickamer, K., Mamon, J., Binns, G., and Leung, J. *J. Biol. Chem.* 1984, **257,** 939
13 Iobst, S.T. and Drickamer, K. *J. Biol. Chem.* 1996, **271,** 6686
14 Weis, W.I., Drickamer, K., and Hendrickson, W.A. (1992) *Nature (London)* **360,** 127
15 Bruccoleri, R.E., Haber, E., and Novotny, J. *Nature (London)* 1989, **355,** 483
16 Bajorath, J. and Fine, R.M. *Immunomethods* 1992, **1,** 137
16a Richardson, J.S. *Methods Enzymol.* 1985, **115,** 359
16b Kabsch, W. and Sander, C. *Biopolymers* 1983, **22,** 2577
17 Sippl, M.J. *Proteins* 1993, **17,** 355
18 Wodak, S.J. and Rooman, M.J. *Curr. Opin. Struct. Biol.* 1993, **3,** 247
19 Bajorath, J., Stenkamp, R. and Aruffo, A. *Bioconjug. Chem.* 1995, **6,** 3
20 Drickamer, K. *Nature (London)* 1992, **360,** 4810
21 Bajorath, J., Hollenbaugh, D., King, G., Harte, W., Jr., Eustice, D.C., Darveau, R.P. and Aruffo, A. *Biochemistry* 1994, **33,** 1332
22 Erbe, D.V., Watson, S.R., Presta, L.G., Wolitzky, B.A., Foxall, C., Brandley, B.K. and Lasky, L.A. *J. Cell. Biol.* 1993, **120,** 1227