Research Article

Structure of the Calcium-Dependent Lectin Domain from a Rat Mannose-Binding Protein Determined by MAD Phasing

WILLIAM I. WEIS, RICHARD KAHN, ROGER FOURME, KURT DRICKAMER, WAYNE A. HENDRICKSON

Calcium-dependent (C-type) animal lectins participate in many cell surface recognition events mediated by protein-carbohydrate interactions. The C-type lectin family includes cell adhesion molecules, endocytic receptors, and extracellular matrix proteins. Mammalian mannose-binding proteins are C-type lectins that function in antibody-independent host defense against pathogens. The crystal structure of the carbohydrate-recognition domain of a rat mannose-binding protein, determined as the holmium-substituted complex by multiwavelength anomalous dispersion (MAD) phasing, reveals an unusual fold consisting of two distinct regions, one of which contains extensive nonregular secondary structure stabilized by two holmium ions. The structure explains the conservation of 32 residues in all C-type carbohydrate-recognition domains, suggesting that the fold seen here is common to these domains. The strong anomalous scattering observed at the Ho L_{III} edge demonstrates that traditional heavy atom complexes will be generally amenable to the MAD phasing method.

MAMMALIAN MANNOSE-BINDING PROTEINS (MBP's) found in serum and liver mediate immunoglobulin-independent defensive reactions against pathogens. These proteins function by binding to specific cell-surface high-mannose oligosaccharides of various bacteria and fungi, and kill these organisms by complement-mediated cell lysis (*1*) or opsonization (*2*). The presence of stress-response promoter regions in the MBP gene (*3*) and increased levels of MBP mRNA during the acute phase response (*2*) suggest that concentrations of MBP rise upon exposure to pathogens. Patients with reduced levels of MBP, due to a point mutation in the MBP gene, show a propensity for repeated, severe bacterial infections (*4*). It has been suggested that MBP's confer a primitive form of immunological self versus non-self discrimination by recognizing carbohydrate structures absent from mammalian cell surfaces (*5*). MBP's also act as inhibitors of human immunodeficiency virus and influenza virus (*6*) infection, presumably by binding to the high-mannose carbohydrates of the viral envelope glycoproteins and blocking attachment to the host cell.

MBP's contain a COOH-terminal carbohydrate-recognition domain (CRD) common to the family of calcium-dependent carbohydrate-binding proteins known as C-type animal lectins (*7*). The family includes endocytic receptors such as the hepatic lectins (asialoglycoprotein receptors), which clear serum glycoproteins from circulation (*8*); selectins (lectin-containing cell adhesion molecules), which target leukocytes to lymphoid tissues and sites of inflammation (*9, 10*); macrophage receptors involved in the phagocytosis of pathogens (*11*); proteoglycan core proteins (*12*); and soluble proteins including pulmonary surfactant apoproteins (*13*). Each of these proteins contains a C-type CRD attached to domains responsible for the physiological functions of the molecule. In the MBP's, the CRD is attached to a collagenous domain presumed to be responsible for the ability to fix complement, as a similar collagenous domain is found in complement protein C1q. The C-type CRD was originally defined as a proteolytic fragment retaining Ca^{2+}-dependent carbohydrate-binding activity, and is now recognized by a sequence motif of approximately 30 conserved amino acids spread over about 120 amino acids (*7*). Although this motif defines a Ca^{2+}-dependent carbohydrate-binding domain, various C-type lectins display distinct carbohydrate-recognition specificities.

We have undertaken structural analysis of MBP's and other C-type lectins in an attempt to understand the basis of their specific, Ca^{2+}-dependent carbohydrate binding, as well as to define the three-dimensional basis of their shared sequence motif. We now describe the three-dimensional structure of the CRD of rat MBP-A with holmium ions substituted for calcium ions. The structure, which consists of a hitherto unknown fold, reveals the roles of the conserved residues that define the C-type CRD motif. The holmium ions form an integral part of the structure, and their ligation suggests a mechanism for the loss of ligand binding activity in the endosome by the endocytic receptor group of C-type lectins. The structure was solved by the multiwavelength anomalous dispersion (MAD) phasing method (*14*), with the use of the anomalous scattering at the Ho L_{III} edge. We have extended the MAD methodology by developing a procedure to refine the anomalous scattering factors f' and f'' against the diffraction data.

Structure determination. The bacterial expression, biochemical characterization, and crystallization of the COOH-terminal portion of rat MBP-A have been described (*15*). The expressed portion containing the neck and CRD did not yield crystals suitable for structure determination, whereas a proteolytic fragment of the expressed protein (designated MBP-A-F2) containing the minimal

W. I. Weis, K. Drickamer, and W. A. Hendrickson are with the Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032; W. I. Weis and W. A. Hendrickson are also with the Howard Hughes Medical Institute. R. Kahn and R. Fourme are with the Laboratoire pour l'Utilisation du Rayonnement Electromagnétique, CNRS and Université Paris-Sud, 91405 Orsay Cedex, France.

Table 1. Data statistics for MBP-A-F2 at 2.5 Å.

Nominal* wavelength (Å)	Reflections† (N)	Completeness (%)	Signal [% > $3\sigma(I)$]	R_{sym}‡
1.4400 (remote)	13,812	93.7	92.5	0.031
1.5356 (peak)	13,461	91.3	91.5	0.043
1.5363 (edge)	13,421	91.0	88.6	0.043

*Because the optical configuration of the beamline precluded measurement of fluorescence spectra from the sample crystal, the monochromator positions defining the three wavelengths were determined from measurement of a unique 3° swath of data containing a mirror plane, at seven equally spaced increments of the monochromator crystal that spanned the L_{III} absorption edge of a $HoCl_3$ powder. Five strong reflections common to the seven data sets were selected, and the average absolute difference between Bijvoet-related intensities of the five reflections was plotted against monochromator angle. The absorption edge and peak positions for the Ho^{3+}-MBP-A-F2 crystal were estimated from a curve of the form observed for the $HoCl_3$ transmission sketched through the seven points. †Unique reflections defined by point group symmetry *222* (not *mmm*) to distinguish acentric Bijvoet mates. Bijvoet-paired measurements were recorded an average of 1.2 times. ‡$R_{sym} = \Sigma_{\mathbf{h}} \Sigma_i |I_i(\mathbf{h}) - \langle I(\mathbf{h})\rangle| / \Sigma_{\mathbf{h}} \Sigma_i I_i(\mathbf{h})$, where $\mathbf{I}_i(\mathbf{h})$ is the i^{th} measurement and $\langle \mathbf{I}(\mathbf{h})\rangle$ is the weighted mean of all measurements of $\mathbf{I}(\mathbf{h})$.

CRD (*16*) (residues 107 to 221 of MBP-A) yielded high-quality crystals. In the presence of Ca^{2+}, crystals of MBP-A-F2 grew only as a complex with a high-mannose glycopeptide. In an attempt to prepare heavy atom derivatives for crystallographic phase determination, we found that various trivalent lanthanide ions (Ln^{3+}) can substitute for Ca^{2+} functionally, but that crystals grown from Ln^{3+}-substituted MBP-A-F2 were in different lattices from that of the native (Ca^{2+}-containing) form, and did not incorporate the glycopeptide required for growth of the native crystals. We also demonstrated that this fragment behaves as a dimer in solution, and that the dimer forms the asymmetric unit of both the monoclinic native and the orthorhombic Ln^{3+}-substituted MBP-A-F2 crystal forms (*15*). Each MBP-A-F2 protomer binds 2 Ca^{2+} or Ln^{3+} (*15*).

Although the Ln^{3+}-containing MBP-A-F2 crystals cannot be used for structure determination by the isomorphous replacement method, they are ideally suited for the MAD phasing method, in which phase information is obtained from the variation in anomalous scattering from a single crystalline species as a function of x-ray wavelength (*14*). MAD phasing requires the presence of a sufficiently strong anomalous scatterer to produce measurable variation in the diffracted amplitudes, and the lanthanides display extremely large anomalous dispersion effects at their L_{III} edges (*17*). In the MAD method, measurements of the Bijvoet differences at several wavelengths ($|^{\lambda i}F(+\mathbf{h})| - |^{\lambda i}F(-\mathbf{h})|$), and the dispersive differences at pairs of wavelengths ($|^{\lambda i}F(\mathbf{h})| - |^{\lambda j}F(\mathbf{h})|$), where λi represents the ith wavelength, yield phase angles by a linear least squares procedure (MADLSQ) that provides solutions for three wavelength-invariant quantities: $|^0F_T(\mathbf{h})|$, the structure factor magnitude due to normal scattering from all atoms in the structure; $|^0F_A(\mathbf{h})|$, the structure factor magnitude due to normal scattering from the anomalous scatterers only; and $\Delta\phi(\mathbf{h}) = {}^0\phi_T(\mathbf{h}) - {}^0\phi_A(\mathbf{h})$, the phase difference between $^0F_T(\mathbf{h})$ and $^0F_A(\mathbf{h})$. The $|^0F_A|$ values are used in a Patterson synthesis or in direct methods to determine the positions of the anomalous scatterers, from which the phases $^0\phi_A$ are computed. The electron density distribution for the entire structure can then be calculated with the use of Fourier coefficients $|^0F_T|\exp(-i^0\phi_T)$, where the relation $^0\phi_T = \Delta\phi + {}^0\phi_A$ gives the required phases. In addition to the diffraction data, the MADLSQ procedure requires the values of the anomalous scattering factors f' and f'' at each wavelength, which enter into the least-squares equations as known coefficients. The $f'(\lambda)$ and $f''(\lambda)$ spectra can be calculated accurately away from the absorption edge (*18*), but the theory breaks down near the edge. In most previous MAD experiments, f' and f'' have been extracted from x-ray fluorescence spectra of the sample crystal (*19*).

Diffraction data from an orthorhombic Ho^{3+}-MBP-A-F2 crystal were collected at three wavelengths near the Ho L_{III} absorption edge chosen from limited diffraction measurements (Table 1) to produce the largest spread of observations for optimal least-squares fitting: the wavelength of peak absorption to maximize f'' and thus the Bijvoet differences, and the absorption edge and a point remote from the edge to maximize the difference in f', which determines the dispersive differences. In order to minimize systematic errors from primary beam intensity changes, sample decay, and differential absorption, all measurements of a particular reflection entering into the determination of its phase were performed close in time and from one sample crystal. Specifically, ten blocks of data, each consisting of the same region of reciprocal space measured at the three wavelengths, were recorded (*20*) (Table 1). When computed without treating Bijvoet mates separately (that is, in point group *mmm*), the R value that measures agreement within and between scaled batches of data (Table 1) varied greatly among the ten blocks of data at the peak and edge wavelengths; within a block, the R value was fairly uniform. Since this R value reflects the magnitude of the Bijvoet differences, the variation indicated that the monochromator had not returned to exactly the same positions for each new block of data so that, because of the rapid change of the anomalous scattering factors at the L_{III} edge, f'' differed from block-to-block. The

Table 2. MAD phasing statistics at 2.5 Å. Values for the observed anomalous diffraction ratios represent $\langle\Delta|F|^2\rangle^{1/2}/\langle|F|^2\rangle^{1/2}$, where $\Delta|F|$ is the absolute value of the Bijvoet difference at one wavelength (diagonal elements) or the dispersive difference at the two wavelengths intersecting at an off-diagonal element. The differences between Bijvoet mates at each wavelength for centric reflections, which would be 0.00 for perfect data, are shown in parentheses and serve as an estimate of the noise in the anomalous signals. A linear scale factor was applied to the peak and edge data of each block to make their $\langle|F|^2\rangle^{1/2}$ equal to that of the remote, to place all data on approximately the same scale with the remote wavelength data as an absolute reference point. Wavelength-dependent linear scale factors relating the data at the three wavelengths were then determined from the values of f' and f'' and applied to each block. These scale factors were recomputed from the refined f' and f'' values, and the entire refinement procedure repeated to produce a final set of $|^0F_T|$, $|^0F_A|$, $\Delta\phi$, and f', f'' for each block of data. Initial estimates of (f', f'') for scattering factor refinement were (-14.1 *e*, 19.9 *e*) and (-28.0 *e*, 10.2 *e*) (*37*) at the peak and edge positions, respectively; at the remote wavelength, f' and f'' were held fixed at their theoretical values (*38*) shown below to provide an absolute reference scale.

	Observed diffraction ratios			Scattering factors (*e*)	
	Remote	Peak	Edge	f'	f''
Block 1					
Remote	0.098 (0.019)	0.101	0.121	−9.4	9.3
Peak		0.266 (0.021)	0.091	−20.4	25.7
Edge			0.102 (0.017)	−26.4	8.3
Block 2					
Remote	0.115	0.073	0.123	−9.4	9.3
Peak		0.238	0.119	−12.0	21.7
Edge			0.134	−29.3	10.2

Agreement of MADLSQ-derived quantities. The statistics are computed from 1689 redundant determinations (26 percent of phased reflections).

$R^*(|^0F_T|) = .065$† $\langle\Delta(\Delta\phi)\rangle$‡ = 22.8°†

$R^*(|^0F_A|) = .311$† $\langle\sigma(\Delta\phi)\rangle$ = 9.0°†

$\langle m\rangle$§ = 0.89

*$R = \Sigma_{\mathbf{h}} \Sigma_i ||F_i(\mathbf{h})| - \langle|F(\mathbf{h})|\rangle| / \Sigma_{\mathbf{h}} \langle|F(\mathbf{h})|\rangle$. †Does not include reflections with m = 0. ‡$\Delta(\Delta\phi)$ is the difference between two independent determinations of $\Delta\phi(\mathbf{h})$. §$\langle m\rangle$ is the mean figure of merit, and here includes 464 reflections (6.4 percent) with m = 0.

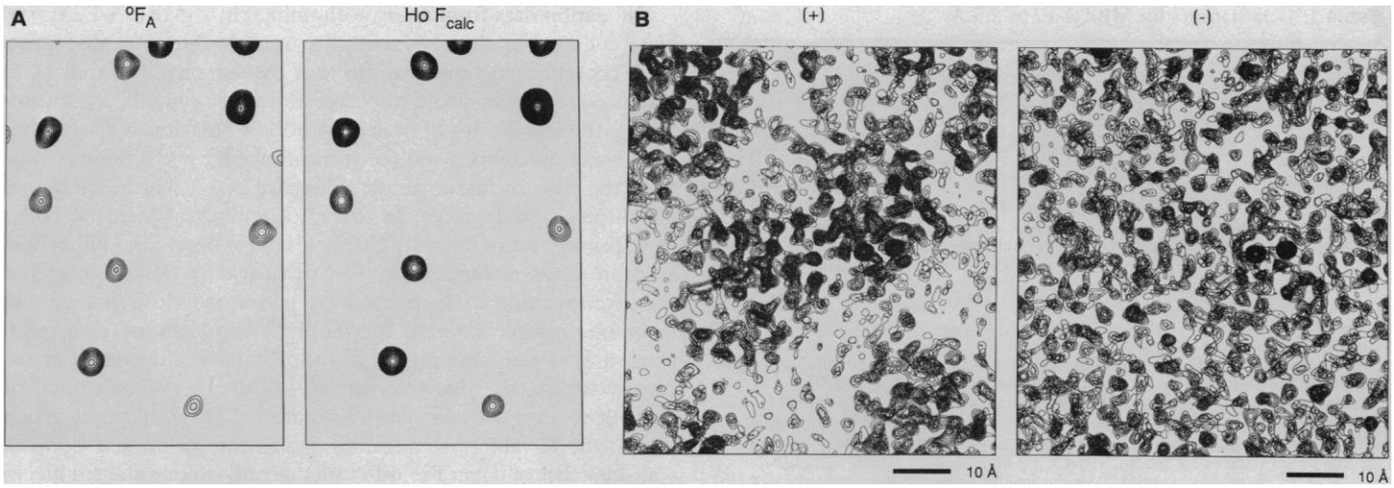

Fig. 1. Determination of Ho positions and choice of hand. (**A**) (Left) The *u* 1/2 *w* Harker section of the $|^{0}F_{A}|$ Patterson map, contoured in increments of 0.5 standard deviations (σ), starting at 2.5 σ. In addition to four self-vectors, there are numerous cross-vectors on this section because three of the four sites have *y* coordinates near 0 or 1/2. The map was made with 6190, or 76.8 percent of the possible $|^{0}F_{A}|$ values (86.0 percent of the reflections phased by MADLSQ) to Bragg spacings of 2.5 Å, meeting the criteria $|^{0}F_{A}| \leq 400$ (maximum $|^{0}F_{A}|$ for four Ho atoms is 1072), $|^{0}F| \geq 2\sigma(|^{0}F_{A}|)$, $\sigma(\Delta\phi) < 50$. Refinement of the four-site Ho model against this set of $|^{0}F_{A}|$ gave an *R* value of 0.260. A difference Fourier synthesis with coefficients $|^{0}F_{A}|_{obs} - |^{0}F_{A}|_{calc}$ and phases calculated from the refined Ho model showed no other sites. (Right) The same section calculated from the refined Ho model. (**B**) Comparison of electron density maps computed from the two enantiomorphic choices of Ho structure to maximum Bragg spacings of 2.5 Å. The Ho^{3+} appear as black circles. Each panel shows the equivalent 10 Å slab along the *c* axis of one unit cell. The correct enantiomer is shown on the left. Both maps are contoured in 0.5 σ increments, starting at 1.0 σ.

monochromator instability was not apparent at the remote point, where the anomalous scattering factors vary slowly. The larger overall merging *R* values for the peak and edge data sets compared to the remote wavelength data (Table 1) also suggested variation in *f'*, which will cause differences among redundant measurements of $|F(+\mathbf{h})|$ or $|F(-\mathbf{h})|$, although larger absorption effects at these longer wavelengths may also have contributed to the poorer agreement within these data sets. The anomalous diffraction ratios (Table 2) observed from each block confirmed the variability in the peak and edge scattering factors; for example, at the edge point the average Bijvoet signal varied from 0.087 to 0.257. Thus, we effectively had data from 21 different wavelengths (1 remote + 10 peak + 10 edge) instead of 3, and no precise knowledge of *f'* and *f"* at the edge and peak wavelengths of a block.

To overcome the problem of unknown anomalous scattering factors in the MADLSQ procedure, we implemented a procedure to refine their values (*21*). For each block of data, the linear least-squares determination of the local variables $|^{0}F_{T}|$, $|^{0}F_{A}|$, and Δφ was performed for each reflection with the use of estimates (Table 2) for the anomalous scattering factors. Next, a well-determined subset of $|^{0}F_{T}|$, $|^{0}F_{A}|$, and Δφ were chosen and treated as knowns in a nonlinear least-squares refinement of the global variables *f'* and *f"* at the edge and peak wavelengths. The updated estimates of *f'* and *f"* were then used in a redetermination of the local variables. This procedure was iterated until the shifts in *f'* and *f"* were on the order of their computed standard errors. Table 2 shows the observed anomalous differences and the refined scattering factor values for two of the ten blocks of data. The scattering factor refinement appears to have a large radius of convergence, as we observed changes as large as 9.8 *e* from the initial estimates of *f'* and *f"*. The set of refined values follow the form of the Ho *f'* and *f"* spectra, albeit with some scatter in the fit. The refinement of the anomalous scattering factors against the diffraction data can be expected for most cases to offer an advantageous alternative to extraction of these quantities from x-ray fluorescence spectra (*19*), since the few scattering factor variables are vastly overdetermined. The x-ray absorption spectrum remains, however, the most convenient basis for choice of wavelengths.

Redundant determinations of $|^{0}F_{T}(\mathbf{h})|$, $|^{0}F_{A}(\mathbf{h})|$, and $\Delta\phi(\mathbf{h})$ were combined to produce a unique set of 7195 reflections with maximum Bragg spacings of 2.5 Å (89.3 percent of those possible). The excellent agreement among independent determinations of these quantities (Table 2) indicates that the scattering factor refinement produced meaningful results, as almost all of the redundancies came from blocks of data with different values of *f'* and *f"*, and also emphasizes the quality of the MAD phases as a result of the accuracy of the data and the large anomalous scattering signal at the Ho L_{III} edge. The $|^{0}F_{A}|$ Patterson map was extremely clear (Fig. 1A), and

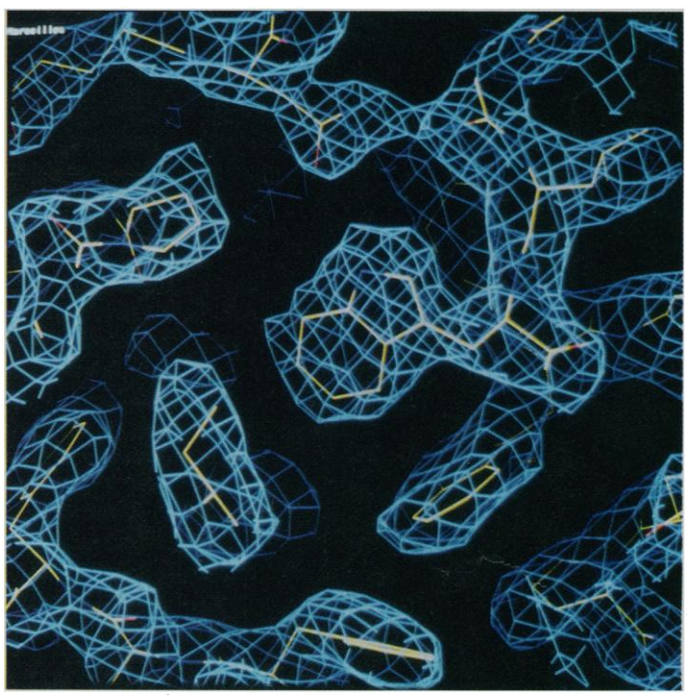

Fig. 2. A portion of the MAD-phased electron density at 2.5 Å Bragg spacings with the initial model. The 1.0 σ contours are shown.

agreed well both with a Bijvoet difference Patterson computed from the remote wavelength data (*15*) and a Patterson computed from the refined Ho model (Fig. 1A), providing independent evidence for the success of the scattering factor refinement. The choice of hand of the Ho structure, which cannot be distinguished from the centrosymmetric Patterson function, was determined by comparison of figure-of-merit weighted electron density maps made with ${}^{0}\phi_{T}$ derived from ${}^{0}\phi_{A}$ computed with the Ho atoms in either hand. The choice was obvious (Fig. 1B): one map had clearly defined solvent channels and a helix, while the other had unrecognizable features.

The polypeptide chain tracing, sequence registration and α-carbon coordinates were determined for one protomer by inspection of a minimap. After a model of this protomer (protomer 1) was built with the computer graphics program FRODO (*22*), protomer 2 was located in the minimap by identifying a distinctive cluster of aromatic residues in uninterpreted density. A second copy of the protomer 1 model was moved as a rigid body into the appropriate density by means of FRODO, and then refined as a rigid body with the program X-PLOR (*23*), while protomer 1 was kept fixed. Adjustments were then made to protomer 2, mainly to account for differences in lattice contacts made by the two protomers in the asymmetric unit. The quality of the MAD-phased electron density map (Fig. 2) allowed us to construct a molecular model with no significant deviations from ideal geometry (Table 3); in particular, the 2.5 Å Bragg spacing limit allowed unambiguous assignment of almost all main-chain carbonyl oxygen positions, so that only two residues of this model fell in the "disallowed" region of the Ramachandran diagram, and they remain as the only such residues after refinement. The model was refined by conjugate gradient minimization with X-PLOR, first at 2.5 Å and then at 2.3 Å resolution (Table 3); extension to higher resolution awaits measurement of a complete 1.5 Å data set. The first two residues of protomer 1, and the first three and last two of protomer 2, are poorly defined in the refined model. The current model has been deposited in the Protein Data Bank (*24*) (entry 1MSB).

Overall structure. The MBP-A CRD adopts a previously undescribed fold of approximate dimensions 40 by 25 by 25 $Å^3$ that is best appreciated by following the path of the polypeptide backbone through the structure (Fig. 3, A and B). The structure begins as a β strand from residues 110 through 119 (β1) that is interrupted by a bulge at residues 115 to 116. The chain then makes a turn into a 10-residue α helix, followed by 10 residues in an extended conformation, and a second α helix of 11 residues. The chain continues as a β strand (β2) that curves sharply at Gly^{158}. This is followed by several loops, each containing one classical tight turn (*25*) in addition to stretches of nonregular secondary structure. Loop 1

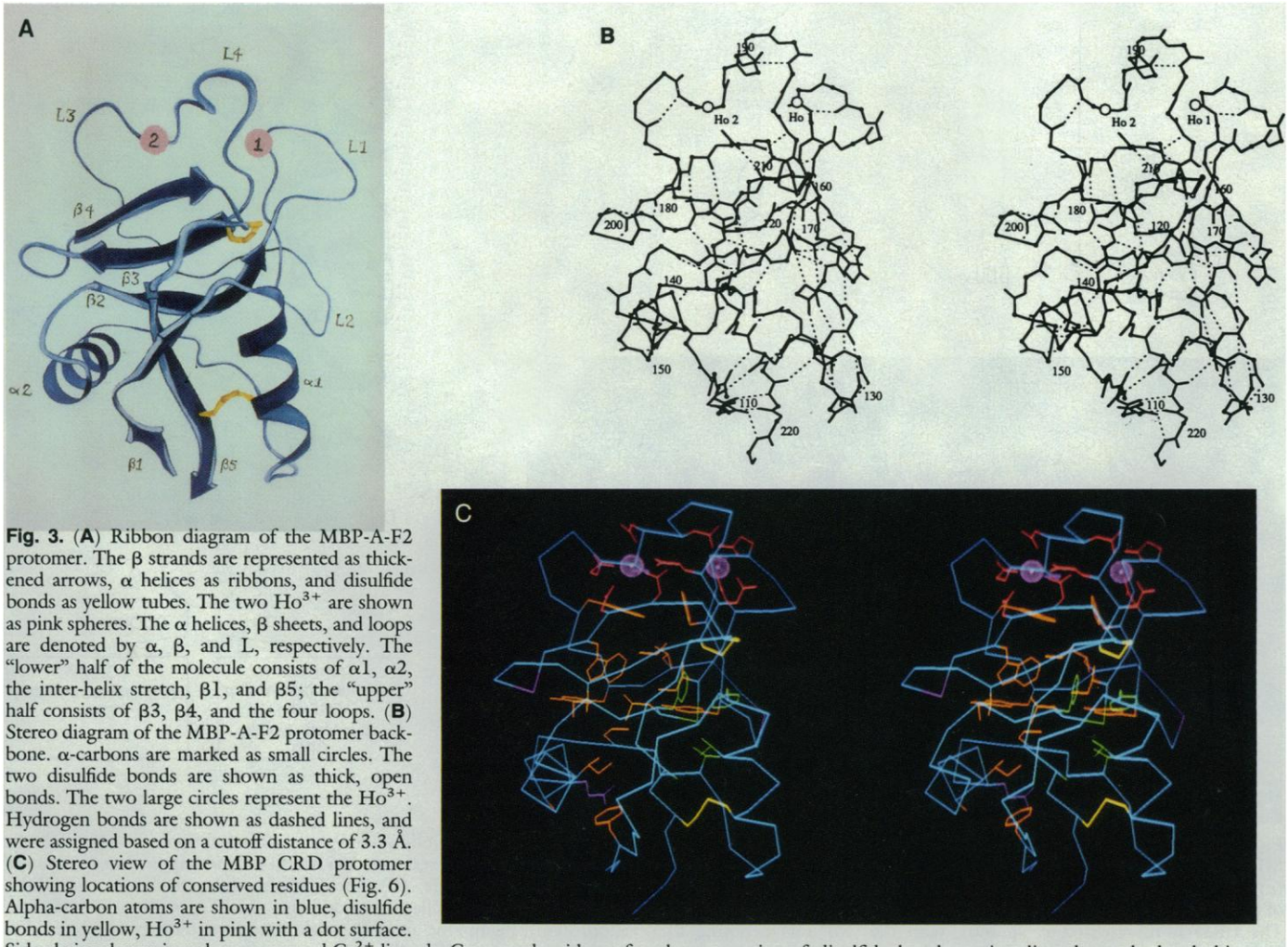

Fig. 3. (**A**) Ribbon diagram of the MBP-A-F2 protomer. The β strands are represented as thickened arrows, α helices as ribbons, and disulfide bonds as yellow tubes. The two Ho^{3+} are shown as pink spheres. The α helices, β sheets, and loops are denoted by α, β, and L, respectively. The "lower" half of the molecule consists of α1, α2, the inter-helix stretch, β1, and β5; the "upper" half consists of β3, β4, and the four loops. (**B**) Stereo diagram of the MBP-A-F2 protomer backbone. α-carbons are marked as small circles. The two disulfide bonds are shown as thick, open bonds. The two large circles represent the Ho^{3+}. Hydrogen bonds are shown as dashed lines, and were assigned based on a cutoff distance of 3.3 Å. (**C**) Stereo view of the MBP CRD protomer showing locations of conserved residues (Fig. 6). Alpha-carbon atoms are shown in blue, disulfide bonds in yellow, Ho^{3+} in pink with a dot surface. Side chains shown in red are conserved Ca^{2+} ligands. Conserved residues of the smaller hydrophobic core (F121, V124, V135, F156, G158, A215) are shown in green. Conserved residues of the larger hydrophobic core (F112, P138, A146, I147, L157, I159, F168, Y170, L176, Y178, W181, V196, I198, W204) are shown in orange. Conserved residues that do not fall into the categories of disulfide bonds, cation ligands, or hydrophobic core (E143, G173, P186, G202) are shown in purple. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

(residues 161 to 168) is a fairly open structure, while loop 2 (residues 170 to 174) is essentially a tight turn bounded on either side by one residue in β conformation. An extended stretch at the "back" of the molecule links loop 2 to loop 3 (180 to 185), which ends in a cis-proline (186) that turns the chain into loop 4 (187 to 194). The backbone continues as paired, antiparallel β strands (β3 and β4) that end in a tight turn linked by a disulfide bond to the beginning of β3. A β strand (β5), linked to the first α helix by another disulfide bond, ends the structure near its start point by antiparallel pairing with β1.

Loops and extended regions account for more than half of the MBP-A CRD structure: 29 percent of the residues are contained in the five β strands and 18 percent in the two helices, while 21 percent are found in classical tight turns and 32 percent in nonregular structure. Moreover, while the five β strands make up more than one-quarter of the structure, there is a remarkable lack of extended β-sheet structure (that is, containing three or more strands). In several places, two paired β strands form hydrogen bonds with a third stretch of residues in β-sheet patterns. In each case, however, the third "strand" runs orthogonally with respect to the other two, precluding formation of more than three hydrogen bonds (Fig. 3B). For example, in addition to pairing with β4, strand β3 pairs in a parallel fashion with residues 180 to 182 of loop 4. Similarly, residues 134 to 136, in the region between the helices, pair in an antiparallel manner with the middle of the β5 strand. Strand β2 highlights the irregular secondary structure of the CRD. The strand starts by pairing residues 154 to 156 in a parallel β structure with the beginning of β5. Residues 155 to 157 pair with β3 to form a pseudo-antiparallel sheet structure, while the main-chain amide of Gly^{158} forms a hydrogen bond with residue 136 of the stretch between the helices. After the turn at 158, residues 158 to 160 pair with loop 2 to form a structure akin to a three-strand antiparallel β sheet. Finally, three main-chain carbonyl oxygen atoms in β2 form hydrogen bonds with bound water molecules rather than with amino acids.

Strand β2 divides the structure into two halves (upper and lower in Fig. 3, A and B): the lower half contains the paired NH_2- and COOH-terminal β strands, the two α helices, and the extended inter-helix region, while the upper half contains the four loops, the inter-loop region, and the paired β strands 3 and 4. The β2 strand

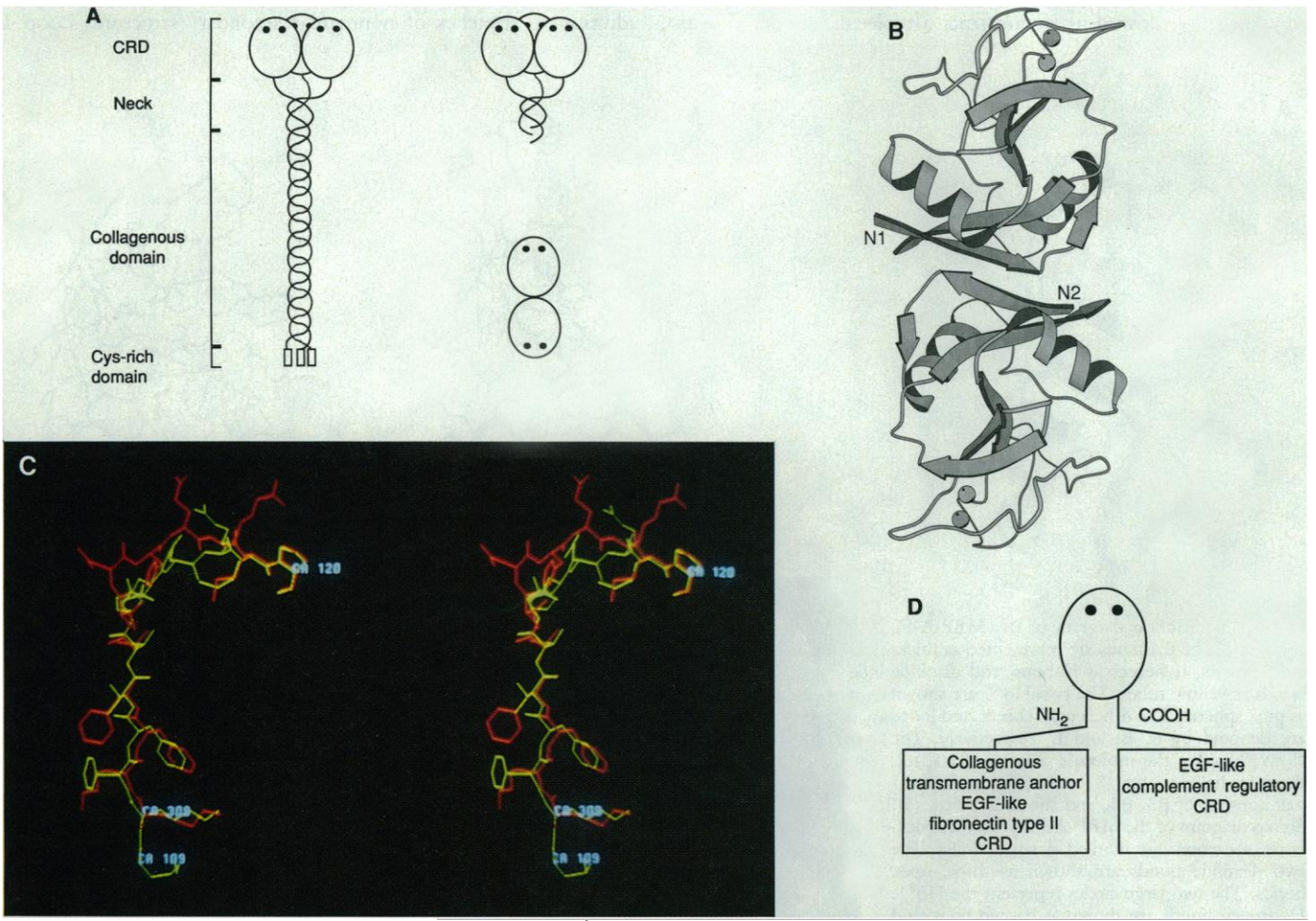

Fig. 4. Oligomeric structure. (**A**) Schematic representation of the intact MBP trimer (left), the trimeric fragment containing the neck and CRD (upper right), and the dimer studied here (lower right). The organization of the CRD's in the intact MBP and in the trimeric fragment is not known. The CRD is represented as an oval, with the two dots denoting the Ca^{2+}-binding end. (**B**) Ribbon diagram (*39*) of the MBP-A-F2 dimer. Structural elements are represented as in Fig. 3A. (**C**) Stereo view of the NH_2-terminal region of protomers 1 (green) and 2 (red) that comprise the crystallographic asymmetric unit. Residues 109 to 120 of each protomer are shown after applying a transformation that superimposes (*40*) residues 119 to 221. Note that residues 119 and 120 superimpose closely, while the main chain atoms of residues 109 to 114 superimpose but are out of register by one residue. (**D**) The CRD module attaches as a COOH-terminal domain to collagenous domains (MBP's, pulmonary surfactant apoprotein), membrane anchors (asialoglycoprotein receptors) or epidermal growth factor (EGF)-like domains (proteoglycan core proteins), and as an NH_2-terminal domain to EGF-like domains (selectins) or complement regulatory domains (proteoglycan core). The macrophage mannose receptor contains tandemly repeated CRD's preceded by a fibronectin type II repeat.

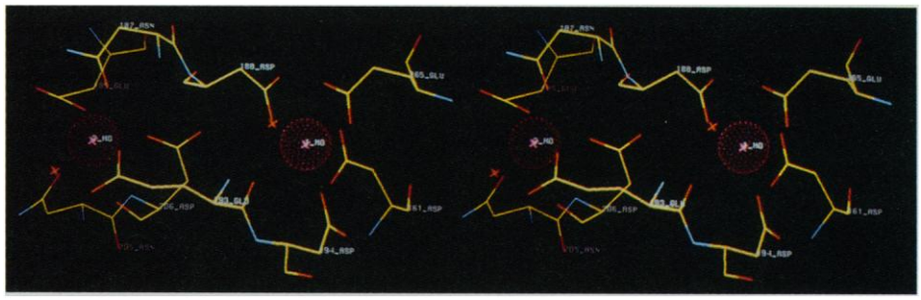

Fig. 5. Stereo view of Ho^{3+} ligands. Oxygen atoms are shown in red, nitrogen in blue, carbon in yellow. Water molecules are represented by red crosses. Ho^{3+} are shown as pink crosses surrounded by a dot surface. The two Ho^{3+} are separated by 8.5 Å.

also demarcates two distinct hydrophobic cores. One core, confined to the lower part of the molecule (bottom front right in Fig. 3, B and C), is composed of residues from β1, helix 1, part of the interhelix stretch, β2 and part of β5. The other, larger core extends through both halves of the domain (left and back of Fig. 3, B and C) and is composed of residues from the interhelix region, helix 2, β2, loops 1 and 2, the region between loops 2 and 3, and strands 3, 4, and 5. Residues 156 to 159, conserved in all C-type CRD's (see below), play a key role in linking the upper and lower halves of the CRD by alternately packing in the two hydrophobic cores. The two aliphatic side chains on either side of the Gly^{158} turn pack in the large core, while the aromatic side chain at position 156 and the α carbon of Gly^{158} pack in the small core.

Oligomeric structure. The MBP-A-F2 dimer (Fig. 4A) forms by antiparallel pairing of the two NH_2-terminal strands (Fig. 4B), which buries hydrophobic residues from parts of these strands, and the ends of the β5 strands and the two helices in an interface of approximately 650 $Å^2$ per protomer. Only two residues from each NH_2-terminal strand form hydrogen bonds between the protomers. A feature of the dimer structure is that the first 12 residues of each protomer violate the dimer symmetry. The transformation obtained by superposition of residues 119 to 221 [which agrees with the diad axis found in a self-rotation function (*15*)] superimposes the main-chain atoms of residues 110 to 118, except for the bulge, but they are out of register by one residue (Fig. 4C). The difference appears to be due to a lattice contact in the bulge region between the two parts of the β1 strand of the protomer 2 (residues 115 to 117), which "pulls" the chain one residue along the paired β5 strand.

The variability of the structure in the NH_2-terminal region suggests that the proteolytic digestion used to make MBP-A-F2 produces an artifactual dimer interface. In the intact molecule, the polypeptide preceding the MBP-A-F2 portion may interact with the β1 strand in a manner akin to the interaction between protomers, with hydrophobic contacts similar to those that form the dimer interface contained within the protomer. Native MBP's are oligomers of a trimeric building block that is stabilized by the collagenous triple helix (Fig. 4A), and it seems likely that the collagenous stalk imposes a quasi threefold symmetry relation among the COOH-terminal CRD's. If any inter-CRD contact surfaces exist in the intact MBP, they cannot be ascertained from the present dimeric structure. However, a fragment of MBP-A containing the CRD plus the 40 amino acid neck that connects the CRD to the collagenous domain is a trimer (Fig. 4A) (*15*). Completion of the crystal structure of this fragment should provide insight into the structure of this region and the relative disposition of the CRD's in the intact MBP.

Cation and carbohydrate binding sites. The two holmium ions (labeled 1 and 2 in Figs. 3, 5, and 6) form an integral part of the structure by pinning together several loop regions. The structure around the cation sites bears no resemblance to known Ca^{2+}-binding motifs such as the E-F hand (*26*), which is not surprising given the much weaker Ca^{2+} affinity of C-type lectins (~0.1 mM) (*15, 27*) commensurate with their presence in the extracellular environment; nor is it similar to the metal-binding sites in plant lectins such as concanavilin A (*28*). Loops 1, 3, and 4 contain most of the protein-derived ligands in both sites, with the remainder located in the β4 strand. Each Ho^{3+} appears to be ligated by one main-chain carbonyl oxygen atom, one water molecule, and five oxygen atoms from amino acid side chains (Fig. 5). The limited resolution of the present structure precludes definitive analysis of the coordination, as ambiguity remains as to whether certain carboxylate ligands are uni- or bidentate, but the sevenfold coordination appears to be similar to that often seen in protein-Ca^{2+} complexes (*29*). In site 1, all of the ligating side chain oxygen atoms come from carboxylates of aspartic and glutamic acid, while in the other site a mixture of three carboxylate and two carbonyl oxygen atoms, the latter from asparagine residues, serve as ligands. Two residues found in all C-type CRD's link the two sites: Glu^{193} binds to site 2 via its carboxylate side chain and to site 1 with its main-chain carbonyl oxygen atom, and Asp^{206} binds to site 2 with one of its carboxylate oxygen atoms and its main-chain carbonyl oxygen atom, while its other carboxylate oxygen atom forms a hydrogen bond with a water molecule that serves as a ligand in site 1.

The presence of carboxylate groups as cation ligands suggests that they contribute to the reversible loss of ligand binding at mildly acid pH seen in several different C-type CRD's, including the MBP's. This phenomenon is central to the function of the endocytic receptor members of the C-type lectin family, since these receptors bind to their ligands at the cell surface, dissociate from them in the

Table 3. Refinement statistics for MBP-A-F2. Noncrystallographic symmetry restraints were not applied. The root-mean-square (rms) discrepancy between isotropic temperature factors of main-chain atoms is 0.9 $Å^2$ and 1.3 $Å^2$ for covalent bond (1–2) and angle (1–3) related atoms, respectively, and 1.5 $Å^2$ and 1.8 $Å^2$ for 1–2 and 1–3 related side-chain atoms.

Item	Initial model	Current model
Data	${}^0F_T$ all data phased	Remote wavelength* all data observed
Resolution range (Å)	5.0–2.5	5.0–2.3
Atoms (No.)	1782	1839 (includes 57 water molecules)
R	0.359	0.176
Model geometry		
Bond lengths (Å)	0.022†	0.010
Bond angles (°)	2.8†	2.6
Dihedral angles (°)	25.0†	25.5
Improper torsion (°)	1.1†	1.1
Peptide ω angles (°)	1.4	3.7

*The merging *R* value on intensities (Table 1) for the 2.3 Å remote wavelength data set is .031. The shell between 2.5 and 2.3 Å is 44 percent systematically incomplete because of the data collection geometry; inclusion of these data had little effect on the coordinates, but gave more realistic temperature factors. Bijvoet mates were treated as separate observations, and the anomalous scattering factor corrections $f' = -9.4$, $f'' = 9.3$ *e* were applied. †Does not include the eight prolines, which are distorted by the model regularizer in FRODO (*22*).

low pH environment of the endosome, and then return to the cell surface for another round of binding (*30*). The release of ligand at low pH results from decreased Ca^{2+} affinity at the endosomal pH of 5.0 to 5.5 because of a conformational change in the CRD (*27*). We suggest that one effect of low pH is to shift the equilibrium toward the ligand nonbinding state by protonating the calcium-ligating carboxylate groups at the two sites, so that they no longer bind Ca^{2+}. The midpoint of the transition between sugar binding and nonbinding states is pH 6.2. Therefore, this mechanism requires that the ligating acids have aberrantly high pK_a values, which seems reasonable since the proximity of several negatively charged acidic groups should stabilize their protonated forms. Despite the presence of Ca^{2+}-ligating carboxylate groups, not all C-type CRD's release their carbohydrate ligands at low pH. Thus, while protonation of Ca^{2+} ligands may be necessary for carbohydrate ligand release, it is not sufficient. Several MBP-A mutants with reduced Ca^{2+} affinity, but normal ligand binding activity in the presence of sufficient Ca^{2+}, map to the central part of the molecule in the hydrophobic cores, away from the Ca^{2+} sites (*31*). The change to a nonbinding state thus appears to involve residues beyond the immediate vicinity of the calcium ions. It is conceivable that titratable groups other than those involved in cation ligation also contribute to the pH-induced triggering of the transition. If this is so, different C-type lectins may take distinct paths to the low pH, ligand nonbinding state, as there are no titratable residues other than Ca^{2+} ligands common to all CRD's that release ligand at low pH. Ultimately, the structure of the low pH, Ca^{2+}-free form of a CRD will be needed to understand the nature of this reversible change.

The location of the carbohydrate binding site is not obvious in the structure described. Inspection of the molecular surface reveals a number of relatively shallow depressions that might serve as binding sites, as found in other lectins with known three-dimensional structures. Results of random mutagenesis of MBP-A (*31*) show that amino acid substitutions that give a wild-type binding phenotype map to surfaces in the lower part of the domain and to loops 1 and 2, but the absence of mutations in the remaining surfaces precludes localization of a binding site. Since the actual number of sugars recognized by the MBP protomer is unknown, it is possible that the binding site will extend over a substantial portion of the surface. The structure of the Ca^{2+}-MBP-A-F2 oligosaccharide complex (*15*) remains to be determined.

Homology to other C-type lectins. Several features of the MBP-A-CRD structure suggest that it can be considered a prototype C-type CRD fold. Alignment of 22 CRD sequences reveals 14 absolutely conserved amino acids, with an additional 18 conserved in character (Fig. 6). Most of these residues fall into three classes: cysteines involved in disulfide bond formation, residues packed in the hydrophobic cores, and Ca^{2+} ligands (Fig. 3C). The four cysteines in the two disulfide bonds are conserved and, although the MBP-A disulfide bond pattern has never been determined chemically, the nested disulfide pattern seen in the three-dimensional structure (Cys^{128}-Cys^{217} and Cys^{195}-Cys^{209}) agrees with that determined by amino acid sequencing of several other C-type lectins (*32*). Many other conserved residues pack together elements of secondary structure in the hydrophobic cores. The conserved "WIGL" sequence (aromatic-aliphatic-glycine-aliphatic) at residues 156 to 159 in β2 indicates that the hydrophobic packing at this turn is common to all C-type CRD's. Residue Gly^{158} appears to be conserved because of its tight packing, which sterically excludes amino acids with larger side chains; the main-chain torsion angles at this position do not preclude other amino acids. In contrast, Pro^{186}, which packs into the larger core, appears to be conserved to allow the turn characteristic of the cis-peptide bond between loops 3 and 4.

Positions of insertions and deletions needed to align the groups of C-type CRD's (Fig. 6) map to the surface of the molecule: in the turn between α2 and β2, in loops 2 and 4, in the region between loops 2 and 3, and in the turn between β3 and β4. Therefore, it appears that the known C-type CRD sequences can be accommodated on the structure presented here without disturbing the basic fold. Nonetheless, the sequence alignments reveal interesting variations. For example, in the selectins, two bidentate carboxylate ligands at Ca^{2+} site 1 are missing (Fig. 6), and no oxygen-containing side chains are present in the vicinity to replace them. This suggests that site 1 may be absent in these proteins, although it is possible that main-chain carbonyl oxygens or water molecules (or both) serve as Ca^{2+} ligands in these molecules. (Whereas the selectins have been shown to require Ca^{2+} for function, the binding stoichiometry is not known.)

The proximity of the NH_2- and COOH-termini suggests that, despite the variable locations of the CRD's in the primary structures of C-type lectins, the CRD's can combine with their respective effector domains in spatially similar ways. For example, CRD's lie on the COOH-terminal side of the transmembrane domains of the asialoglycoprotein receptors (*33*), while in the selectins they are

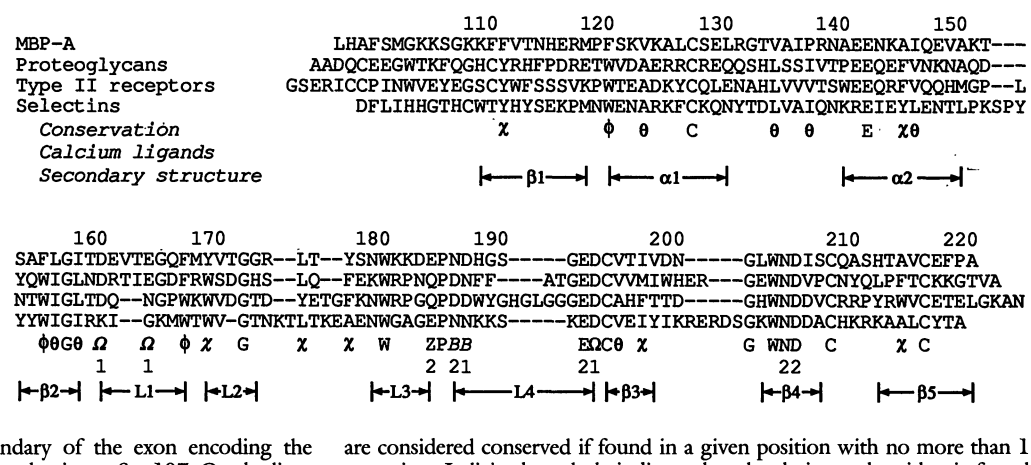

Fig. 6. Alignment of C-type CRD sequences. Twenty-two sequences were aligned with a multidimensional dynamic programming algorithm (*41*). One representative from each of four subgroups of C-type lectins (*42*) is shown. The four sequences shown are: rat MBP-A (*34*), representing group III (C-type lectins with collagenous domains); rat proteoglycan core protein (*12*) from group I (proteoglycans); major form of the rat asialoglycoprotein receptor (rat hepatic lectin I) (*43*) from group II (type II receptors); mouse lymphocyte homing receptor (MEL-14 antigen) (*10*) from group IV (selectins). The numbering of the intact rat MBP-A is shown at the top. The sequences shown start at the boundary of the exon encoding the designated CRD. The MBP-A-F2 fragment begins at Ser 107. On the lines marked "conservation," invariant amino acids are shown in single-letter code; positions conserved in character are shown with the following code: θ = aliphatic; ϕ = aromatic; χ = aliphatic or aromatic; Z = E or Q; B = D or N; Ω = side chain with carbonyl oxygen atom (D, N, E or Q). Residues are considered conserved if found in a given position with no more than 1 exception. Italicized symbols indicate that the designated residue is found with 2 to 4 exceptions. For example, position 161 is always D, N or E except in the 4 selectin sequences used in the alignments. The cation site (1 or 2) to which a side chain contributes oxygen ligands is shown below that residue. α helices, β sheets, and loops denoted by α, β, and L, respectively.

NH_2-terminal to epidermal growth factor (EGF) and complement-regulatory domains, as well as their membrane anchors (*9, 10*). It is easy to envision that the CRD's in these two types of membrane proteins would be similarly disposed with respect to the cell membrane, probably with the "upper" half (Fig. 3A) pointing away from the cell surface. The COOH-terminal orientation of the CRD's in both MBP's and the asialoglycoprotein receptors suggests that their attachment to collagenous domains (MBP's) or transmembrane domains (receptors) through short neck peptides is similar. This notion is supported by the observation that the CRD's of these proteins can be removed readily by proteolytic digestion in the neck region (*34, 35*). The macrophage mannose receptor (*11*) presents a more complicated example, as it contains eight tandemly repeated CRD's. It is clear, however, that the CRD is a tightly folded, modular unit that can be placed in many orientations within the structure of C-type lectins (Fig. 4D).

Prospects for MAD phasing. The accuracy of the phases obtained for the Ho^{3+}-substituted MBP-A CRD emphasizes the advantages of using L_{III}-edge anomalous scattering for MAD structure determination. K-edge scattering effects, which are substantially weaker and therefore make more stringent demands on data quality, have been used in most previous MAD experiments (*36*). Atoms commonly used for heavy atom derivatives in isomorphous replacement, such as Pt, Au, Hg, and Pb, should be generally applicable vehicles for MAD phasing, because their L_{III} edges fall at especially convenient wavelengths (~1.0 Å), although the variation of the anomalous scattering factors at the L_{III} edge is smaller for these atoms than for lanthanides. MAD offers the distinct advantage of not requiring isomorphism, which often limits the attainment of reliable phases past moderate resolution limits, and which would allow preparation of heavy atom derivatives prior to crystallization to ensure complete substitution. Finally, the direct refinement of the anomalous scattering factors, as described above, offers a convenient and potentially more accurate way to obtain these essential parameters.

REFERENCES AND NOTES

1. K. Ikeda, T. Sannoh, N. Kawasaki, T. Kawasaki, I. Yamashina, *J. Biol. Chem.* **262**, 7451 (1987).
2. R. A. B. Ezekowitz, L. E. Day, G. A. Herman, *J. Exp. Med.* **167**, 1034 (1988).
3. M. E. Taylor, P. M. Brickell, R. K. Craig, J. A. Summerfield, *Biochem. J.* **262**, 763 (1989).
4. M. Super, S. Thiel, J. Lu, R. J. Levinsky, M. W. Turner, *Lancet* **335**, 1236 (1989); M. Sumiya *et al.*, *Lancet* **337**, 1569 (1991).
5. R. A. B. Ezekowitz, *Curr. Biol.* **1**, 60 (1991).
6. _____, M. Kuhlman, J. Groopman, R. Byrn, *J. Exp. Med.* **169**, 185 (1989); E. M. Anders, C. A. Hartley, D. C. Jackson, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4485 (1990).
7. K. Drickamer, *J. Biol. Chem.* **263**, 9557 (1988).
8. G. Ashwell and J. Harford, *Annu. Rev. Biochem.* **51**, 531 (1982); D. F. Halberg *et al.*, *J. Biol. Chem.* **262**, 9828 (1987); M. Spiess, *Biochemistry* **29**, 10009 (1990).
9. M. P. Bevilacqua, S. Stengelin, M. A. J. Gimbrone, B. Seed, *Science* **243**, 1160 (1989); G. I. Johnston, R. G. Cook, R. P. McEver, *Cell* **56**, 1033 (1989); M. Bevilacqua *et al.*, *ibid.* **67**, 233 (1991).
10. L. A. Lasky *et al.*, *Cell* **56**, 1045 (1989); M. H. Siegelman, M. van de Rijn, I. L. Weissman, *Science* **243**, 1165 (1989).
11. M. E. Taylor, J. T. Conary, M. R. Lennartz, P. Stahl, K. Drickamer, *J. Biol. Chem.* **265**, 12156 (1990).
12. D. F. Halberg, G. Proulx, K. Doege, Y. Yamada, K. Drickamer, *ibid.* **263**, 9486 (1988).
13. R. T. White *et al.*, *Nature* **317**, 361 (1985).
14. W. A. Hendrickson, *Trans. Am. Crystallogr. Assoc.* **21**, 11 (1985).
15. W. I. Weis, G. V. Crichlow, H. M. K. Murthy, W. A. Hendrickson, K. Drickamer, *J. Biol. Chem.* **266**, 20678 (1991).
16. M. S. Quesenberry and K. Drickamer, *Glycobiology*, in press.
17. L. K. Templeton, D. H. Templeton, R. P. Phizackerley, K. O. Hodgson, *Acta Crystallogr.* **A38**, 74 (1982).
18. D. T. Cromer and D. Liberman, *J. Chem. Phys.* **53**, 1891 (1970).
19. W. A. Hendrickson, J. L. Smith, R. P. Phizackerley, E. A. Merritt, *Proteins* **4**, 77 (1988).
20. Diffraction data from a single Ho^{3+}–MBP-A-F2 crystal (space group $P2_12_12_1$; a = 66.39 Å, b = 73.12 Å, c = 45.29 Å) of dimensions 0.70 mm by 0.16 mm by (0.04 to 0.10) mm were measured at 4°C on the Mark II spherical drift multiwire proportional chamber detector on the D23 line at the LURE synchrotron (Orsay, France) [R. Kahn, R. Fourme, R. Bosshard, V. Saintagne, *Nucl. Instrum. Methods Phys. Res.* **A246**, 596 (1986)], which operated at 1.85 GeV and ring currents of approximately 210 to 310 mA. The detector was positioned at an angle of 20° with respect to the direct beam. Bijvoet-related reflections were measured across mirror planes by rotation about crystallographic diad axes. Data from a rotation through 106° about the **b*** axis were recorded; the crystal was then removed, remounted for rotation about the **a*** axis, and returned to the cooling stream for collection of an orthogonal 45° of data. The mosaic spread was larger after this procedure, but no other dimunution in the diffraction was observed. Exposure times per 0.10° frame were 30 seconds for the first 72° of data and 40 seconds for the remainder. Ten blocks of reciprocal space (5° to 24°) were recorded at each of the three wavelengths, in the order (remote, peak, edge). For the largest blocks, approximately 2.5 hours elapsed between measurement of the same region of reciprocal space at different wavelengths. Integrated intensities corrected for Lorentz-polarization were obtained with the MADNES software [A. Messerschmidt and J. W. Pflugrath, *J. Appl. Crystallogr.* **20**, 306 (1987)] in combination with a profile fitting algorithm [W. Kabsch, *ibid.* **21**, 916 (1988)], the latter being modified in order to process Bijvoet mates independently. Data from each wavelength were binned into 2° batches and scaled with the program ROTAVATA [CCP4, *The SERC (UK) Collaborative Computing Project No. 4, A Suite of Programs for Protein Crystallography* (Daresbury Laboratory, Warrington WA4 4AD, UK, 1979)]. Bijvoet mates were not combined, including centric reflections, which were used to estimate the noise in the anomalous signals (Table 2). Scale factors were applied and outlier measurements rejected (14 in total) by the CCP4 program AGROVATA, which was modified to reject outlying measurements only with respect to redundant copies of the same Friedel mate $I(+\mathbf{h})$ or $I(-\mathbf{h})$, rather than with respect to the mean of all $I(\pm\mathbf{h})$. Redundant measurements were not merged until after phase determination. The data were placed on a quasi-absolute scale by applying a linear scale factor to bring $\langle|F|^2\rangle^{1/2}$ at the average Bragg spacing (3.1 Å) to its theoretical value calculated from the chemical composition of the unit cell, including the anomalous scattering corrections at the wavelength under consideration.
21. Other procedures for refining f' and f'' have been reported in W. A. Hendrickson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2190 (1989); E. Fanchon and W. A. Hendrickson, *Acta Crystallogr.* **A46**, 809 (1990).
22. T. A. Jones, *Methods Enzymol.* **115**, 157 (1985).
23. A. T. Brünger, *X-PLOR Manual, Version 2.1* (Yale University, New Haven, CT, 1990).
24. F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977).
25. J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981).
26. R. H. Kretsinger, *CRC Crit. Rev. Biochem.* **8**, 119 (1980).
27. J. A. Loeb and K. Drickamer, *J. Biol. Chem.* **263**, 9752 (1988).
28. K. D. Hardman and C. F. Ainsworth, *Biochemistry* **11**, 4910 (1972); G. N. Reeke, J. W. Becker, G. M. Edelman, *J. Biol. Chem.* **250**, 1525 (1975); Z. Derewenda *et al.*, *EMBO J.* **8**, 2189 (1989).
29. A. L. Swain, R. H. Kretsinger, E. L. Amma, *J. Biol. Chem.* **264**, 16620 (1989).
30. M. DiPaola and F. R. Maxfield, *ibid.* **259**, 9163 (1984); P. P. Breitfeld, C. F. J. Simmons, G. J. A. Strous, H. J. Geuze, A. L. Schwartz, *Int. Rev. Cytol.* **97**, 47 (1985).
31. M. S. Quesenberry and K. Drickamer, in preparation.
32. J. Fuhlendorff, I. Clemmensen, S. Magnusson, *Biochemistry* **26**, 6757 (1987); J. D. Sandy, C. R. Flannery, R. E. Boynton, P. J. Neame, *J. Biol. Chem.* **265**, 21108 (1990).
33. K. Drickamer, J. F. Mamon, G. Binns, J. O. Leung, *J. Biol. Chem.* **259**, 770 (1984).
34. K. Drickamer, M. S. Dordal, L. Reynolds, *ibid.* **261**, 6878 (1986).
35. J. A. Loeb and K. Drickamer, *ibid.* **262**, 3022 (1987).
36. W. A. Hendrickson, *Science* **254**, 51 (1991).
37. R. Kahn *et al.*, *FEBS Lett.* **179**, 133 (1985).
38. D. T. Cromer, *J. Appl. Crystallogr.* **16**, 437 (1983).
39. P. Kraulis, *ibid.*, in press.
40. W. Kabsch, *Acta Crystallogr.* **A32**, 922 (1976).
41. D. J. Lipman, S. F. Altschul, J. D. Kececioglu, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4412 (1989).
42. K. Bezouska, G. V. Crichlow, J. M. Rose, M. E. Taylor, K. Drickamer, *J. Biol. Chem.* **266**, 11604 (1991).
43. J. O. Leung, E. C. Holland, K. Drickamer, *ibid.* **260**, 12523 (1985).

44. We thank C. Ogata and W. Yang for discussions, and R. Yarmolinsky for preparing Fig. 3A. Supported by NIH grants GM42628 and GM34102; by LURE; by a Life Sciences Research Foundation Fellowship from the Howard Hughes Medical Institute (W.I.W.); and an American Cancer Society Faculty Salary Award FRA-342 (K.D.).

24 September 1991; accepted 31 October 1991