

# Molecular Modeling of the Multidomain Structures of the Proteoglycan Binding Region and the Link Protein of Cartilage by Neutron and Synchrotron X-ray Scattering<sup>†</sup>

Stephen J. Perkins,<sup>\*,‡</sup> Adam S. Nealis,<sup>‡</sup> David G. Dunham,<sup>§</sup> Timothy E. Hardingham,<sup>§</sup> and I. Helen Muir<sup>§</sup>

*Department of Biochemistry and Chemistry, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, U.K., and Biochemistry Division, Kennedy Institute of Rheumatology, Bute Gardens, Hammersmith, London W6 7DW, U.K.*

*Received April 29, 1991; Revised Manuscript Received July 19, 1991*

**ABSTRACT:** The interaction of proteoglycan monomers with hyaluronate in cartilage is mediated by a globular binding region at the N-terminus of the proteoglycan monomer; this interaction is stabilized by link protein. Sequences show that both the binding region (27% carbohydrate) and the link protein (6% carbohydrate) contain an immunoglobulin (Ig) fold domain and two proteoglycan tandem repeat (PTR) domains. Both proteins were investigated by neutron and synchrotron X-ray solution scattering, in which nonspecific aggregate formation was reduced by the use of citraconylation to modify surface lysine residues. The neutron and X-ray radius of gyration  $R_G$  of native and citraconylated binding region is 5.1 nm, and the cross-sectional  $R_G$  ( $R_{XS}$ ) is 1.9–2.0 nm. No neutron contrast dependence of the  $R_G$  values was observed; however, a large contrast dependence was seen for the  $R_{XS}$  values which is attributed to the high carbohydrate content of the binding region. The neutron  $R_G$  for citraconylated link protein is 2.9 nm, its  $R_{XS}$  is 0.8 nm, and these data are also independent of the neutron contrast. The scattering curves of binding region and link protein were modeled using small spheres. Both protein structures were defined initially by the representation of one domain by a crystal structure for a variable Ig fold and a fixed volume for the two PTR domains calculated from sequence data. The final models showed that the different dimensions and neutron contrast properties of binding region compared to link protein could be attributed to an extended glycosylated C-terminal peptide with extended carbohydrate structures in the binding region. As a further control, these scattering curve models were used to define hydrodynamic sphere models, which were used to compute sedimentation coefficients of 3.4 and 2.6 S for the binding region and link protein, respectively, in good agreement with the experimental values. It is concluded that, under physiological conditions, the three globular domains of link protein and binding region are contained within dimensions of 10 nm × 2.5 nm × 3 nm and that the extended C-terminal peptide of the binding region has a length in the range of 8–17 nm.

**P**roteoglycans are specialized glycoproteins in which many long strongly anionic polysaccharide chains (glycosaminoglycans) are attached to a central extended protein core. In cartilage, proteoglycans occur at very high concentrations (50–100 mg/mL) and together with collagen are responsible for the important biomechanical properties of the tissue (Muir, 1979; Hardingham, 1986; Hassell et al., 1986). The proteoglycans are of high molecular weight and form supramolecular aggregates in which many proteoglycan monomers bind non-covalently to an extended hyaluronate polysaccharide chain. The interaction is mediated by ternary complex formation between a negatively charged globular G1 entity at the N-terminus of the proteoglycan protein core, a positively charged link protein, and hyaluronate.

Structural studies will elucidate the manner in which proteoglycans aggregate within cartilage. The sequences of G1 and link protein (Deak et al., 1986; Neame et al., 1986; Doege et al., 1987, 1991; Dudhia & Hardingham, 1989, 1990) show that both are constructed from a single copy of a variable immunoglobulin (Ig) fold domain and two copies of a proteoglycan tandem repeat (PTR) domain (Bonnet et al., 1986; Perkins et al., 1989). While link protein exists as a three-

domain structure, binding region is the N-terminal G1 domain of a proteoglycan monomer and is cleaved from it by mild proteolysis. The N-terminal region of proteoglycan monomers also contains a second globular entity G2, which is separated from G1 by an extended polypeptide chain E1. The binding region fragment released by proteolysis contains the three-domain G1 structures and a chemically undefined segment of E1. Since each PTR is encoded as a separate exon (Kiss et al., 1987), and since it has been found in single copies in two cell surface proteins, human CDw44 (alternative name: gp90<sup>Hermes</sup>) and mouse Pgp-1 (Stamenkovic et al., 1989; Goldstein et al., 1989; Wolffe et al., 1990), it can be inferred that a PTR domain most probably exists also as an independently folded structural entity. Tryptic digestion experiments show that the Ig folds in binding region and link protein interact with each other in the ternary complex, while the PTRs interact with hyaluronate (Neame et al., 1986; Périn et al., 1987; Goetinck et al., 1987). Structural studies are now required in order to elucidate the arrangement of these three domains in these two proteins.

Constrained X-ray and neutron solution scattering is a powerful low-resolution technique that will characterize the arrangement of domains in multidomain proteins, especially if known crystallographic structures are available for the domains (Perkins, 1988a,b; Perkins et al., 1991). The first investigation of binding region by neutron scattering showed that this has a globular, elongated structure with carbohydrate on its surface (Perkins et al., 1981). Sedimentation experiments

<sup>†</sup>Support from the Wellcome Trust, the Medical Research Council, and the Science and Engineering Research Council is gratefully acknowledged.

<sup>\*</sup>Author to whom correspondence should be addressed.

<sup>‡</sup>Royal Free Hospital School of Medicine.

<sup>§</sup>Kennedy Institute of Rheumatology.

have also shown that binding region and link protein possess elongated shapes (Bonnet et al., 1985). Electron microscopy has also indicated that both proteins have globular structures (Wiedemann et al., 1984; Paulsson et al., 1987; Mörgelin et al., 1988). These results were supported by consensus secondary structure analyses of the Ig fold and PTR sequences (Perkins et al., 1989). The original scattering and sedimentation studies were however affected by small amounts of nonspecific aggregated proteins. To inhibit their formation in the present scattering study, several of the surface lysine residues of both proteins were chemically modified by citraconylation to reduce the amounts of these aggregates. The resulting improvements in the scattering curves, together with the constrained analysis of the neutron and X-ray scattering curves on the basis of available crystal structures for the Ig fold and the fixed volume of the PTRs, lead to solution structures for the domains of binding region and link protein and an improved interpretation of the globular structures visualized by electron microscopy. The differences between the solution structures of the two proteins could thus be critically evaluated in relation to the similarity of their domain structures and their binding with hyaluronate.

#### MATERIALS AND METHODS

(a) *Preparations of Binding Region and Link Protein.* Porcine binding region and link protein were prepared following chondroitinase ABC (EC 4.2.2.4) and trypsin (EC 3.4.21.4) digestion of proteoglycan aggregates (Bonnet et al., 1985). The citraconylation of binding region and link protein and the keratanase digestion of binding region are described in Bonnet et al. (1985). Sample concentrations for the citraconylated samples were determined either prior to citraconylation assuming minimal losses in subsequent manipulations or by adding trace amounts of  $^{125}\text{I}$ -labeled binding region or link protein and determining the concentration of radioactivity prior to and after citraconylation. Before and after scattering experiments, samples were monitored by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (Bonnet et al., 1985). Radioimmunoassays for binding region and link protein (Ratcliffe & Hardingham, 1983) verified the antigenic integrity of the samples.

(b) *Neutron and X-ray Scattering Measurements.* Neutron data were obtained on the solution camera D17 at the ILL Grenoble. Guinier data were collected using a sample–detector distance (SD) of 3.46 m and wavelengths  $\lambda$  of 0.993, 1.087, 1.388 or 1.600 nm (wavelength spread of 10% full width at half-maximum). Data at larger scattering vectors  $Q$  ( $Q = 4\pi(\sin \theta)/\lambda$ , where  $2\theta$  is the scattering angle) were obtained with a SD of 1.40 m with  $\lambda$  of 0.993 or 1.000 nm, or a SD of 3.46 m with  $\lambda = 1.087$  nm and a detector rotation of  $4.85^\circ$  relative to the main beam. Other data were recorded on the solution camera D11 at a SD of 2.50 m and  $\lambda$  of 1.00 nm; the spread in  $\lambda$  is 9%. The  $Q$  range is therefore 0.06–1.61  $\text{nm}^{-1}$ . Samples were measured in phosphate-buffered saline at pH 7.2 (123 mM NaCl, 10.4 mM  $\text{Na}_2\text{HPO}_4$ , 7.7 mM  $\text{NaN}_3$ ), in buffers containing 0%, 10%, 70%, 80%, and 100%  $^2\text{H}_2\text{O}$  by volume. Samples were dialyzed for at least 36 h with four buffer changes; the last dialysate was used for the background buffer measurement. All data were recorded at 20  $^\circ\text{C}$ . Raw data were reduced using standard Grenoble software (Ghosh, 1981).

Synchrotron X-ray scattering was performed at the SRS Daresbury on the solution scattering camera at station 7.3 (Nave et al., 1985). The ring operated at 1.8 GeV and currents of 70–300 mA in four sessions, and 2.0 GeV and currents of 90–240 mA in two further sessions. The SD was 2.210 or

2.690 m, and  $\lambda$  was 0.1608 nm, giving  $Q$  ranges of 0.07–1.30 or 0.07–0.88  $\text{nm}^{-1}$ . A 500-channel linear detector of length 76 mm was used, calibrated using wet, slightly stretched rat tail collagen (diffraction spacing 67 nm) and dry white adductor muscle (diffraction spacing 14.4 nm) as standards. Exposure times were 10 min at 1.8 GeV and 6 min at 2.0 GeV. Samples were measured in phosphate-buffered saline at 11–13  $^\circ\text{C}$  after dialysis as above. Data reduction was based on the program SWANAL. Other details are given in Perkins et al. (1990a).

(c) *Scattering Data Analyses.* At low  $Q$ , the radius of gyration  $R_G$  and the forward scattering at zero angle  $I(0)$  are given by the Guinier relationship (Glatter & Kratky, 1982):

$$\ln I(Q) = \ln I(0) - R_G^2 Q^2 / 3$$

Molecular weights  $M_r$  are calculated from  $I(0)/c$  values ( $c$  = concentration) as absolute values from neutron data or relative values from X-ray data (Kratky, 1963; Jacrot & Zaccai, 1981). The X-ray  $M_r$  were determined relative to C-reactive protein and serum amyloid P component (Perkins & Pepys, 1986) and two mouse immunoglobulins IgG2a and IgG2b (S. J. Perkins and A. R. Rees, unpublished scattering data). The degree of elongation is obtained from the ratio  $R_G/R_0$ , where  $R_0$  is calculated from the hydrated (X-rays) or dry (neutrons) glycoprotein volume assuming a sphere of radius  $r$ , thus  $R_0 = (0.6r)^{1/2}$ . The cross-sectional radius of gyration  $R_{XS}$  and the cross-sectional intensity at zero angle  $[I(Q)Q]_{Q \rightarrow 0}$  are obtained:

$$\ln [I(Q)Q] = \ln [I(Q)Q]_{Q \rightarrow 0} - R_{XS}^2 Q^2 / 2$$

For an elongated structure similar to an elliptical cylinder, its longest axis  $L$  is calculated from

$$L^2 = 12(R_G^2 - R_{XS}^2)$$

Knowledge of  $I(0)$  and  $[I(Q)Q]_{Q \rightarrow 0}$  also gives  $L$  from

$$L = I(0)\pi/[I(Q)Q]_{Q \rightarrow 0}$$

The neutron contrast variation data in  $\text{H}_2\text{O}$ – $^2\text{H}_2\text{O}$  mixtures is analyzed in simplified Stuhrmann plots assuming that the internal structure is centrosymmetric (Ibel & Stuhrmann, 1975):

$$R_G^2 = R_{G-C}^2 + \alpha_G \Delta\rho^{-1}$$

$$R_{XS}^2 = R_{XS-C}^2 + \alpha_{XS} \Delta\rho^{-1}$$

where  $R_{G-C}$  and  $R_{XS-C}$  are the  $R_G$  of the macromolecule and its cross section at infinite contrast,  $\alpha_G$  and  $\alpha_{XS}$  correspond to the radial inhomogeneity of scattering density, and  $\Delta\rho^{-1}$  is the reciprocal solute–solvent contrast difference. Stuhrmann plots were analyzed using weighted two-parameter least-squares fits (Perkins & Weiss, 1983).

(d) *Hydrodynamic and Densitometric Analyses.* Frictional coefficients  $f$  are calculated from

$$f = M_r(1 - \bar{v}\rho_{20,w})/N_a s_{20,w}^0$$

where  $\bar{v}$  is the partial specific volume,  $\rho_{20,w}$  is the density of water at 20  $^\circ\text{C}$ ,  $N_a$  is Avogadro's constant, and  $s_{20,w}^0$  is the sedimentation coefficient. The degree of elongation is given from the frictional ratio  $f/f_0 = 6\pi\eta a$ , in which the glycoprotein is assumed to be a hydrated sphere of radius  $a$  (see next section). The  $\bar{v}$  values were calculated from compositions using the consensus volume data set of Perkins (1986).

(e) *Modeling of Binding Region and Link Protein.* Simulations of the scattering curves and the  $R_G$  and  $\alpha$  parameters used small Debye spheres to follow previously described procedures (Perkins & Weiss, 1983; Smith et al., 1990; Perkins

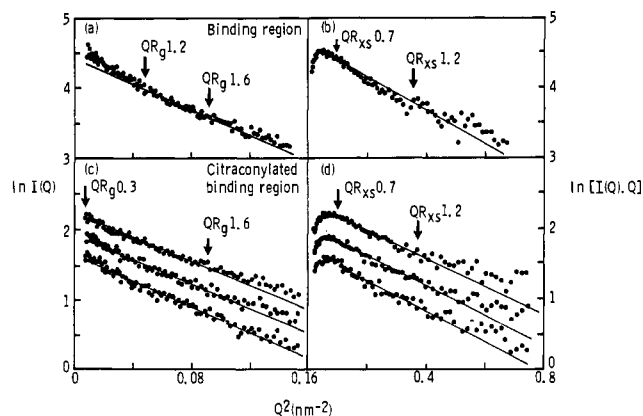


FIGURE 1: Synchrotron X-ray Guinier analyses for native and citraconylated binding region. The vertical scales are arbitrarily numbered. The  $QR_G$  and  $QR_{XS}$  ranges used to determine  $R_G$  and  $R_{XS}$  are arrowed, based on  $R_G$  fits out to  $Q = 0.3 \text{ nm}^{-1}$  and  $R_{XS}$  fits in the  $Q$  range  $0.3\text{--}0.6 \text{ nm}^{-1}$ .

et al., 1990b). Sequences were used to calculate dry volumes from the data of Chothia (1975) and Perkins (1986) as follows: chicken, pig, rat, and human link protein (Deak et al., 1986; Neame et al., 1986; Dudhia & Hardingham, 1989, 1990); rat and human G1 and G2 domains (Doege et al., 1987, 1991); human versican (Zimmerman & Ruoslahti, 1989); murine PGP-1 (Wolffe et al., 1990); human CD44/gp90Hermes (Goldstein et al., 1989; Stamenkovic et al., 1989). The dry structure was represented by cubes of side  $0.595 \text{ nm}$ , since this corresponds to the mean volume of a monosaccharide unit of hyaluronate and facilitates the comparison of the models with their physiological substrate. The overlapping spheres at the center of each cube are of diameter  $0.738 \text{ nm}$ , which is less than the nominal resolution of  $3.9 \text{ nm}$  of the scattering curve ( $2\pi/Q_{\text{max}}$ ) as required (Glatter & Kratky, 1982). For hydrodynamic simulations, the hydrated volume is calculated assuming a hydration of  $0.3 \text{ g of H}_2\text{O/g of glycoprotein}$  and a water molecule volume of  $0.0245 \text{ nm}^3$ . Hydrodynamic frictional coefficients were calculated using nonoverlapping spheres by the modified Oseen tensor procedure of Bloomfield (Garcia de la Torre & Bloomfield, 1977a,b) using the procedure described in Perkins (1989, 1990) and Smith et al. (1990).

## RESULTS AND DISCUSSION

**(a) X-ray Scattering of Citraconylated Binding Region.** Native binding region, even when aggregate-free by the criteria of ultracentrifugation (Bonnet et al., 1985) or gel filtration, generally exhibited minor aggregates as shown by the curvature in Guinier  $R_G$  plots at low  $Q$ , as exemplified in the synchrotron X-ray experiments of Figure 1a. In Figure 1a, traces of aggregates influence part of the  $R_G$  analysis of native binding region ( $6.5 \text{ mg/mL}$ ) at  $QR_G$  below  $1.2$ , but in Figure 1b these do not affect the  $R_{XS}$  analysis. It was inferred that these aggregates form soon after purification due to self-association, or possibly as a result of radiation damage effects. The Guinier analyses however led to satisfactory  $R_G$  and  $I(0)/c$  data by the use of a larger than usual  $QR_G$  range of  $1.2\text{--}1.6$ . To justify this procedure, the complete removal of aggregates was attempted. On the basis that aggregates resulted from the attraction of oppositely charged groups on different molecules, the charge on several lysine residues in the binding region were reversed by citraconylation (Materials and Methods; Bonnet et al., 1985; Bindels et al., 1985). In Figure 1c,d, a dilution series is shown for citraconylated binding region ( $10.5$ ,  $6.6$ , and  $5.0 \text{ mg/mL}$ ). The linear plots of Figure 1c in

Table I: Structural Properties of Binding Region and Link Protein

	binding region		link protein
	Compositional		
amino acid residues	(338)	446	326
$M_r$	(56 000)	69 500	39 100
sequence		67 000 <sup>a</sup>	42 000 <sup>a</sup>
biochemical		70 000 $\pm$ 4000	NM <sup>b</sup>
X-ray scattering		70 000 $\pm$ 5000	43 000 $\pm$ 3000
neutron scattering			
% wt carbohydrate	(33)	27	6
dry volume ( $\text{nm}^3$ )	(67.1)	84.5	49.7
hydrated volume ( $\text{nm}^3$ )		112.8	65.7
densitometric $\bar{v}$ ( $\text{mL/g}$ )	(0.695)	0.703	0.729
(calcd)			
$\Sigma b/M_r$ in $\text{H}_2\text{O}$ (fm)	(0.2279)	0.2289	0.2369
match point (% $\text{H}_2\text{O}$ )			
calcd	(43.0)	42.5	42.5
exptl		43.8, 45.1	43.2
$A_{280}$ (1%, 1 cm)	(9.0)	8.9	17.7
	Structural		
X-ray			
$R_G$ (nm)		$5.13 \pm 0.04$	NM
$R_{XS}$ (nm)		$2.04 \pm 0.02$	NM
Neutron			
$R_{G-c}$ (nm)		$5.1 \pm 0.2$	$2.9 \pm 0.2$
$R_{XS-c}$ (nm)		$1.86 \pm 0.04$	$0.8 \pm 0.2$
$\alpha_G$ ( $\times 10^{-5}$ )		0	0
$\alpha_{XS}$ ( $\times 10^{-5}$ )		$30 \pm 4$	0
$R_{G-c}/R_0$		2.4	1.6
$s_{20,w}^0$ (S)		$3.4 \pm 0.1^a$	$2.5 \pm 0.1^a$
$f$ ( $\times 10^{-8} \text{ g}\cdot\text{cm}^{-1}\cdot\text{s}^{-1}$ )		$9.7 \pm 0.3$	$7.6 \pm 0.3$
		$12 \pm 2^c$	
$f/f_0$		1.8	1.4
	Models		
$R_G$ (nm)		4.94	2.94
$s_{20,w}^0$ (S)		3.6	2.9

<sup>a</sup> Bonnet et al (1985). <sup>b</sup> NM, not measurable. <sup>c</sup> Perkins et al. (1981).

the  $QR_G$  range of  $0.3\text{--}1.6$  showed that this treatment was effective, as confirmed by molecular weight calculations (below). As a control, since the citraconylation procedure involved a protein denaturation step, radioimmune assays (Materials and Methods) were successfully used to confirm the integrity of the samples. It was however observed that, even for citraconylated preparations, high X-ray beam intensities occasionally caused the reappearance of aggregates. Two other approaches to reduce aggregates were unsuccessful, based either on the removal of about half the negatively charged keratan sulfate content by keratanase digestion (Materials and Methods) or on the use of  $0.1\%$  solutions of Nonidet P40 detergent to reduce hydrophobic interactions.

The sequence of rat proteoglycan (Doege et al., 1987) was used for data interpretation (Table I). Since the tryptic cleavage site which generates binding region from proteoglycans is unknown, two compositions of binding region were calculated assuming that, as limits, either this was composed only of G1 (i.e., an Ig fold and two PTR domains: 338 residues) or this included also the E1 extended polypeptide up to the start of G2 (446 residues). Both compositions were similar to that of pig binding region (Perkins et al., 1981). The carbohydrate analyses (Bonnet et al., 1985) were assumed to correspond to three keratan sulfate oligosaccharide ( $\text{Gal}_2\text{GalNAc}_3\text{GlcNAc}_{24}$  with 24 sulfate groups; Stuhlsatz et al., 1989) and three biantennary complex-type N-linked oligosaccharides. The calculated absorption coefficients  $A_{280}$  (Perkins, 1986) were  $9.0$  and  $8.9$  in that order; both were higher than that of  $6.46$  ( $1\%$ ,  $278 \text{ nm}$ ,  $1 \text{ cm}$ ) for porcine binding region (Perkins et al., 1981).

The X-ray Guinier analyses (Figure 1) showed no dependence of the  $R_G$ ,  $R_{XS}$ ,  $I(0)$ , and  $[I(Q)Q]_{Q \rightarrow 0}$  parameters on

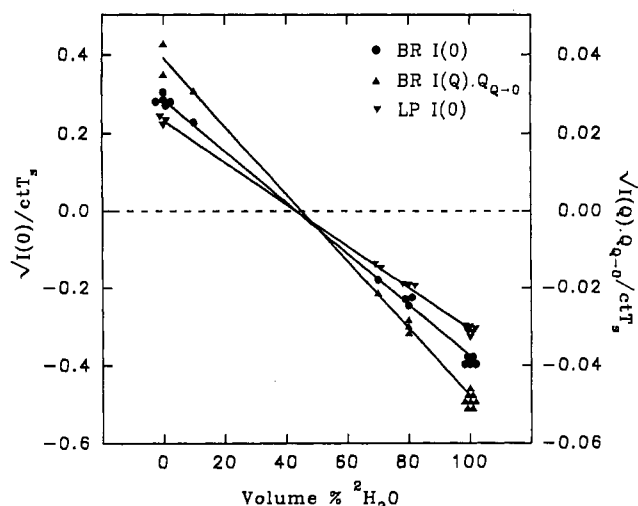


FIGURE 2: Match-point determinations for native and citraconylated binding region and citraconylated link protein by neutron scattering. Data for the  $I(0)$  (● and ▼) and  $I(Q)Q$  (▲) plots are shown ( $c$ : concentration;  $t$ : sample thickness;  $T_s$ : sample transmission).

concentrations between 5 and 10.5 mg/mL (data not shown). The  $R_G$  and  $R_{XS}$  values (Table I) were similar for native and citraconylated binding region. Three  $M_r$  calculations (Materials and Methods) gave 73 000, 70 000, and 68 000 for citraconylated binding region. These agreed with values of  $66\,500 \pm 5500$  from equilibrium ultracentrifugation and an estimate of ca. 70 000 by gel electrophoresis (Bonnet et al., 1985), and also with 69 500 calculated from the 446-residue sequence. For binding region, the sequence was thus taken to be closer to the 446-residue limit than the 338-residue limit. This limit was employed in subsequent analyses.

The mean X-ray  $R_G$  of native and citraconylated binding region was determined as  $5.13 \pm 0.04$  nm, and the mean  $R_{XS}$  as  $2.04 \pm 0.02$  nm. This led to an elongation ratio  $R_G/R_0$  ratio of 2.4, which was large compared to a typical value of 1.28 for globular proteins (Perkins, 1988b). This showed that the binding region has an elongated solution structure. The original neutron experiments on the binding region had also predicted an elongated structure (Perkins et al., 1981), but a lower  $R_G$  of 4.2 nm had been determined as the result of ambiguities caused by aggregation. Assuming that binding region can be represented by an elliptical cylinder, combination of the  $R_G$  and  $R_{XS}$  values (Materials and Methods) led to a length  $L$  of  $16 \pm 1$  nm. An alternative calculation based on the intensity parameters (Materials and Methods) gave a consistent value of  $L$  of  $19 \pm 2$  nm. If  $L$  is 16–19 nm, the hydrated volume is  $112.8$  nm<sup>3</sup>, and  $R_{XS}$  is 2.04 nm (Table I), the two short axes of the elliptical cylinder representing binding region were determined as 8 nm  $\times$  1 nm under the conditions of measurement in this high positive solute–solvent contrast difference.

(b) *Neutron Scattering of Native Binding Region.* Neutron contrast variation on native binding region in  $H_2O$ – $^2H_2O$  mixtures was used to confirm and extend the X-ray analysis. The  $R_G$  data were based on Guinier fits in a limited  $QR_G$  range between 0.7 and 1.5 ( $Q$  range of 0.13–0.30 nm<sup>-1</sup>). This is improved compared to the X-ray data and is probably the result of the absence of radiation effects. The  $R_{XS}$  data were obtained using the  $Q$  range of Figure 1. A  $M_r$  of  $70\,000 \pm 5000$  was calculated from  $I(0)/c$  in  $H_2O$  buffers, in good agreement with the other  $M_r$  values given above. The neutron match-point graphs based on the  $I(0)$  and the  $[I(Q)Q]_{Q \rightarrow 0}$  data were linear, with correlation coefficients of 0.999 or better. These gave match points of 43.8%  $^2H_2O$  and 45.1%  $^2H_2O$

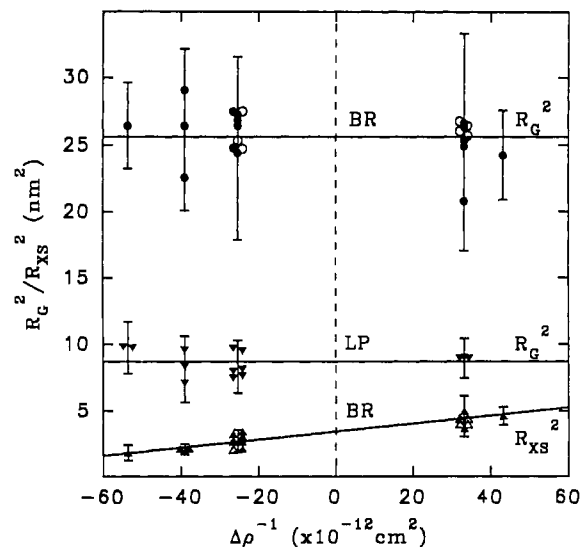


FIGURE 3: Stuhrmann plots of the neutron  $R_G^2$  and  $R_{XS}^2$  data as a function of the reciprocal solute–solvent contrast difference  $\Delta\rho^{-1}$  for binding region and link protein. For binding region,  $R_{G-C}$  is  $5.1 \pm 0.2$  nm (15 values) and  $\alpha_G$  is zero within error;  $R_{XS-C}$  and  $\alpha_{XS}$  are  $1.86 \pm 0.04$  nm and  $30 (\pm 4) \times 10^{-5}$ . Native binding region (●); citraconylated binding region in  $^2H_2O$  (○). X-ray data for citraconylated binding region (○ and ▲) are shown adjacent to the 0%  $^2H_2O$  neutron data points for comparison. For citraconylated trypsin-digested link protein (▼),  $R_{G-C}$  is  $2.9 \pm 0.2$  nm as the mean of the 14 values shown, and  $\alpha_G$  is zero within error.

(Figure 2), which were in satisfactory agreement to within 2.6%  $^2H_2O$  with a match point of 42.5%  $^2H_2O$  predicted from the composition. These agreements showed that molecular information could be obtained from the scattering data.

The Stuhrmann analyses (Figure 3) summarize the dependence of the  $R_G$  and  $R_{XS}$  data on the contrast. Even though the limited Guinier fit range led to large statistical errors in the  $R_G^2$  data of Figure 3, nonetheless  $R_{G-C}$  was found to be  $5.1 \pm 0.2$  nm, and this was in good agreement with the X-ray  $R_G$  of  $5.13 \pm 0.04$  nm. As a control, in 100%  $^2H_2O$ , Figure 3 showed that good agreement is seen for  $R_G$  and  $R_{XS}$  data measured for the citraconylated binding region. The length  $L$  was calculated as  $17 \pm 1$  nm from the  $R_G$  and  $R_{XS}$  data in all contrasts and  $20 \pm 2$  nm from the ratio of intensities, and these values agree well with  $L$  from X-ray data. Since the slope  $\alpha_G$  in Figure 3 was zero, this showed that there was no significant inhomogeneity in the distribution of hydrophobic and hydrophilic residues within the binding region. However, the neutron  $R_{XS}^2$  values showed a clear dependence on the solute–solvent contrast difference, where  $\alpha_{XS}$  was large at  $30 (\pm 4) \times 10^{-5}$ . The neutron  $R_{XS-C}$  of  $1.86 \pm 0.04$  nm was thus less than the X-ray  $R_{XS}$  of  $2.04 \pm 0.02$  nm. This large  $\alpha_{XS}$  may be explained in terms of an elongated glycoprotein structure composed of a central protein core (of lower scattering densities, as visualized by X-ray scattering or 0%  $^2H_2O$  buffers by neutron scattering), along the length of which are attached oligosaccharide chains (of higher scattering densities) in extended conformations [cf. C1 inhibitor of complement: Perkins et al. (1990b)]. The carbohydrate in binding region consists of about three sulfated O-linked and three N-linked complex-type oligosaccharides (Bonnet et al., 1985).

(c) *Neutron Scattering of Trypsin-Digested Link Protein.* Link protein occurs in two forms which differ in  $M_r$ . The larger form of  $M_r$ , 48 000 contains an additional glycopeptide that is not present in the smaller form of  $M_r$ , 42 000 (Baker & Caterson, 1979; Le Gledic et al., 1983). Pig laryngeal cartilage contains about 90% of the larger form. For this study, it was converted to a single form of  $M_r$ , 39 100 by mild

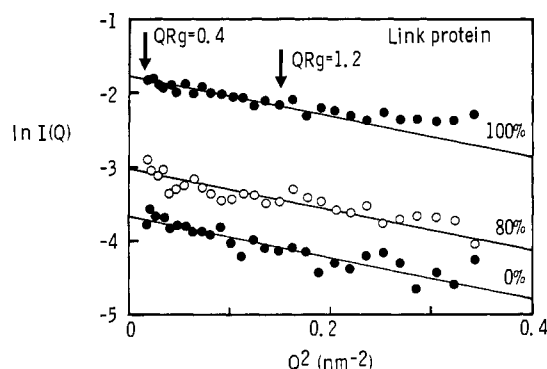


FIGURE 4: Neutron Guinier analyses for citraconylated trypsin-digested link protein. The vertical scale is referenced to the incoherent scattering of a 1-mm water sample. The  $QR_G$  range used to derive the  $R_G$  values using the  $Q$  range out to  $Q = 0.4 \text{ nm}^{-1}$  is arrowed.

tryptic digestion (Bonnet et al., 1985), in which it was assumed that the tryptic cleavage occurred between Arg28 and Val29 and that this contained a single biantennary complex-type N-linked oligosaccharide (Dudhia & Hardingham, 1989). Its predicted  $A_{280}$  is 17.7, which is higher than that of 13.8 reported for nontrypsinized bovine link protein (Tang et al., 1979). Both this and native link protein aggregated readily. In order to obtain samples suitable for scattering, citraconylation was used. These neutron data were successfully measured in 0%, 80%, and 100%  $^2\text{H}_2\text{O}$  buffers, with sample concentrations of 3.5, 2.8, and 2.6 mg/mL, in that order as shown in Figure 4. These samples however aggregated in synchrotron X-ray beams.

Neutron Guinier analyses (Figure 4) resulted in a  $M_r$  of  $43\,000 \pm 3000$  from  $I(Q)/c$ , in good agreement with a biochemical  $M_r$  of 42 000 from ultracentrifugation and a sequence  $M_r$  of 39 100. In Figure 2, regression of the 14 points for link protein gave a match point of 43.2%  $^2\text{H}_2\text{O}$ . The neutron match point of 43.2%  $^2\text{H}_2\text{O}$  (Figure 2) agreed with a prediction of 42.5%  $^2\text{H}_2\text{O}$  calculated from the sequence (Perkins, 1986). The Stuhmann plot (Figure 3) resulted in an  $R_{G-C}$  of  $2.9 \pm 0.2 \text{ nm}$  and an  $\alpha_G$  close to zero. The  $R_G/R_0$  ratio of link protein was thus 1.6. Comparison with that for binding region showed that link protein is less elongated than binding region (Table I). Cross-sectional analyses in a  $Q$  range of  $0.4\text{--}1.1 \text{ nm}^{-1}$  for link protein in  $^2\text{H}_2\text{O}$  buffers gave an  $R_{XS}$  of  $0.8 \pm 0.2 \text{ nm}$  (3 values). Assuming that link protein can be represented by an elliptical cylinder, the two length calculations gave  $12 \pm 1 \text{ nm}$  and  $10 \pm 1 \text{ nm}$  (Materials and Methods). From these lengths, the  $R_{XS}$ , and the volume, the two shorter axes of link protein were determined as 2.0 and 2.6 nm.

(d) *Scattering Models of Binding Region and Link Protein.* Initial modeling calculations approximated the structures of binding region and link protein by elliptical cylinders constructed from small spheres, whose length  $L$  and two shorter axes  $b$  and  $c$  are to be determined. Holding the volume constant in the model for binding region, lengths  $L$  of 17–18 nm gave  $R_G$  values close to 5.1 nm (Table I), irrespective of the lengths of  $b$  and  $c$ . A ratio of  $b$  and  $c$  of 1:14 gave the best account of the X-ray scattering curve (measured in positive solute-solvent contrast; not shown) in the  $Q$  range  $0.4\text{--}1.2 \text{ nm}^{-1}$ . A ratio of  $b:c$  of 1:8 gave the best fit for the neutron scattering curve in  $^2\text{H}_2\text{O}$  (measured in negative solute-solvent contrast; not shown). These different models reflect the large contrast dependence of the  $R_{XS}$  data in Figure 3. These models could be combined into a two-density oblate elliptical cylinder of  $30 \times 1 \times 14$  spheres (diameter 0.595 nm; overall dimensions  $18 \text{ nm} \times 1 \text{ nm} \times 8 \text{ nm}$ ). Here, good curve fits could be obtained by setting the scattering density of the 120 spheres

found in the outermost three rows on each long edge of this structure independently of the central eight rows of 254 spheres to allow for positive and negative solute contrast conditions. Similar simulations for link protein showed that an elliptical cylinder of  $17 \times 4 \times 5$  spheres (dimensions on  $10 \text{ nm} \times 2.5 \text{ nm} \times 3 \text{ nm}$ ) gave a good account of the neutron curves in  $\text{H}_2\text{O}$  and  $^2\text{H}_2\text{O}$  (not shown). In summary, without the use of sequence information, curve fitting alone led to noticeably different structures for these two proteins (Perkins et al., 1986).

Final scattering curve analyses for the binding region and link protein were based on full sequence data showing that both proteins have three similar domains (Figure 5). The variable Ig fold sequences in immunoglobulins differ the most from those in the proteoglycan sequences in that additional residues are found in the L1 and L2 hypervariable loops and at the N-terminal end of the Ig-fold in link protein and G1, as exemplified in Figure 5a. After removal of the N-terminal 13 residues immediately prior to the Ig fold of link protein in order to correspond to its trypsinized form, the dry volume (Chothia, 1975) of the Ig fold in both link protein and G1 were found to be the same at  $19.0 \pm 0.2 \text{ nm}^3$ . The crystal structure of the  $V_L$  domain of Fab McP603 (Satow et al., 1986) was assumed to resemble most closely the variable Ig fold of binding region and link protein on the grounds that this has the longest hypervariable loop L1 (Perkins et al., 1989). The dry volume of this  $V_L$  domain is  $15.0 \text{ nm}^3$ . Accordingly, the  $V_L$  coordinates were converted into 72 spheres, each of volume  $0.595^3 \text{ nm}^3$ , to which were added a further 18 spheres in the region close to the N-terminus, L1 and L2 (Figure 6a), in order to reach the required volume of  $19.0 \text{ nm}^3$ . N-Linked carbohydrate chains (11 spheres) were added in the region suggested by homology with the crystal structure (Figures 5a and 6a). Using the PTR sequences of Figure 5b, calculation of the dry volume showed that this was closely similar in 18 of the 20 sequences and resulted in a mean of  $14.3 \pm 0.3 \text{ nm}^3$ ; the exceptions were the PGP-1 and CD44 sequences, which had volumes of  $16.1 \text{ nm}^3$ . Each PTR in link protein and G1 was thus represented by 68 spheres. The final 239-sphere model for link protein in Figure 6b is of overall length 17 spheres, with an  $R_G$  of 2.94 nm, and readily led to highly satisfactory curve fits as shown in Figure 7a. If both PTRs are collinear with the Ig fold, this analysis suggests that each independently folded PTR domain is of approximate size  $3 \text{ nm} \times 3 \text{ nm} \times 2.5 \text{ nm}$  ( $5 \times 5 \times 4$  spheres). The overall dimensions of this model are similar to the simple elliptical cylinder model above for link protein, although here a slightly improved curve fit is obtained with the model of Figure 6b. Note however that the scattering analyses do not provide information on the location of the loops L1, L2, and L3 in the final model of Figure 6b.

Since the three globular domains of link protein are found also in the binding region, the large difference in the form of the scattering curves of these two glycoproteins and their contrast dependence (Figure 7, panel a vs panel b) is likely to be explained in terms of the additional C-terminal peptide and the increased glycosylation found in binding region. That the initial modeling of binding region led to an oblate ellipsoid structure is indicative that these components lie in an extended conformation. Trial-and-error modeling was performed with 22 distinct structures in which the C-terminal segment was represented either by an arbitrary oblate structure or by an explicit protein core with O-linked oligosaccharides attached to this (Figure 7b). Models with an  $R_G$  value close to the experimental value of 5.1 nm, in which the long axis of the Ig fold and PTR domains were collinear with the C-terminal

(a) Immunoglobulin fold domain

	←A→	←A'→	*←B→	←L1→	←C→	←C'→	←L2→
MCPC603	↓						
Chi LP	EPHPDNSSLEHERIIHQEENGPRLLVVAEQAKIFSQRGGNVTLPCFKYH			EHSTAGSGTHKI			RVKWTKLTS
Pig LP	DHLSNNYTLDDRVIIHQEENGPRLLVVAEQAKVFSHRGGNVTLPCFKYH			DPTAF GSGTHKI			RIKWTKLTS
Rat LP	DHLSDSYTPDQDRVIHQEENGPRLLVVAEQAKVFSHRGGNVTLPCFKYH			DPTAF GSGIHKI			RIKWTKLTS
Hum LP	DHLSDNVTLDDHRAIHQEENGPHLLVVAEQAKVFSHRGGNVTLPCFKYH			DPTAF GSGIHKI			RIKWTKLTS
Rat G1		EEVDPHDNLSVSIQPSPLKALLGTSITIPCFIDPMHPVTAPSTAPLT					KEVVLLVATEGVVRV
Hum G1		VETSDHDNLSVSIQPSPLRVLGTSITIPCFIDPMHPVTAPSTAPLA					KEVVLLVATEGVVRV
Hum Versican		LHKVKVGKSPVVRGSLGK VSLPCHFSTMTPLP					KNGKDLKETTVLVAQ

(b) Proteoglycan tandem repeat domain

	←C'→	*←D→	←E→	*←F→	←L3→	←G→
MCPC603	TRES	GV	PDR	FT	SGSGTDFTLTISVQAEDLAVYVCQNDHSYPLTFGAGTKL	
Chi LP	YGKY	QG	RV	FL	RESSENDASLIITNIMLEDYGRYKCEVIEGLE	DDTAVVALNLEGVV
Pig LP	YGKY	QG	RV	FL	KGGSNDASLVITDLETDYGRYKCEVIEGLE	DDTAVVALDLEGVV
Rat LP	YGKY	QG	RV	FL	KGGSNDASLIITDLETDYGRYKCEVIEGLE	DDTAVVALELQGVV
Hum LP	YGKY	QG	RV	FL	KGGSNDASLVITDLETDYGRYKCEVIEGLE	DDTAVVALDQGVV
Rat G1	NSIY	QD	KV	SLPNYPAIPSDATLEIQNLRSNDSGIYRCVMMHGE	DSEATLEVIVKGV	
Hum G1	NSAY	QD	KV	SLPNYPAIPSDATLEVQSLRSNDSGIYRCVMMHGE	DSEATLEVIVKGV	
Hum Versican	NGNIKIGQDYKGRVSVPTHPAEVGDASLTIVKLLASDAGLYRCDVMYIGIE				DTQDVTSLTVDGVV	

FIGURE 5: Sequence alignments for the Ig fold and PTR domains of binding region and link protein. (a) The full Ig fold in four link protein and two proteoglycan G1 sequences is compared with that of the  $V_L$  domain in McPC603 and that in versican. The positions of the three hypervariable loops L1, L2, and L3 are indicated above the sequences. The  $\beta$ -strands in the crystallographic model are denoted by the letters A–G (Chothia & Lesk, 1987). The locations of two putative N-linked oligosaccharide sites are asterisked. The cleavage site to yield trypsinised link protein is arrowed at Arg13. Further details are given in Figure 2 of Perkins et al. (1989). (b) Full PTR sequences in four link protein and two proteoglycan G1 sequences are shown together with PTR sequences found in versican and two cell surface proteins. The locations of two putative N-linked oligosaccharide sites in the first and second PTRs of G1 are asterisked; the first of these is found also in G2.

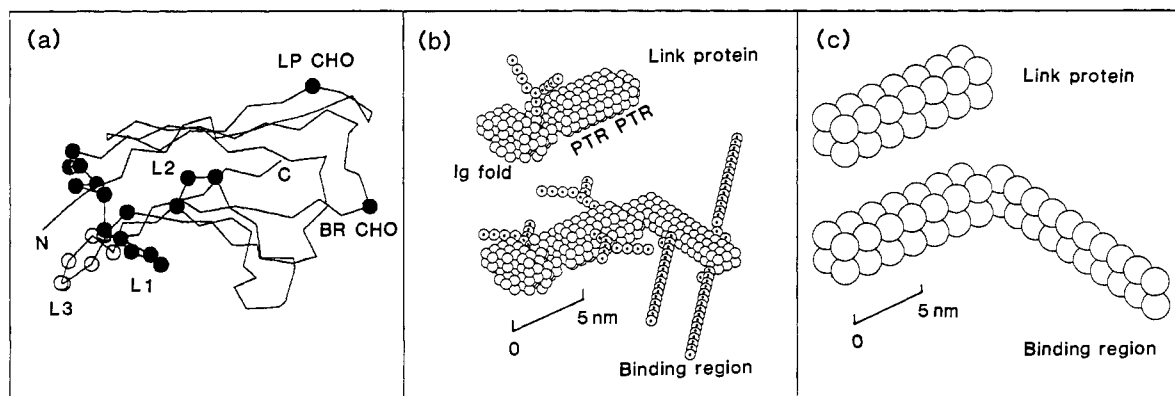


FIGURE 6: Molecular models for binding region and link protein. (a) An  $\alpha$ -carbon diagram of the  $V_L$  domain of Fab McPC603 used to initialize the Debye sphere modeling of the Ig fold in link protein and binding region. The locations of the L1, L2, and L3 hypervariable loops (13, 3, and 6 residues in length, respectively) are indicated in the foreground by circles. The  $\beta$ -strands D, E, B, and A are shown in the uppermost  $\beta$ -sheet, together with the putative oligosaccharide site of the Ig fold in link protein immediately preceding strand B (Arg18 in McPC603). The  $\beta$ -strands G, F, C, C', and C'' are shown on the lower  $\beta$ -sheet, together with the putative oligosaccharide site found in binding region between strands E and F (Glu87 in McPC603). (b) Debye scattering curve models for link protein and binding region (239 and 399 spheres, respectively, of diameter 0.595 nm). Both models are orientated so that the view is similar to that of the  $\alpha$ -carbon diagram of the  $V_L$  domain shown in (a). The locations of the Ig fold and PTR domains are indicated in the link protein model; they are similar in the binding region model. The oligosaccharides in link protein (one) and binding region (six) are represented by spotted spheres and are depicted in arbitrary extended conformations. (c) Hydrodynamic sphere models for link protein and binding region (28 and 48 spheres, respectively, of diameter 1.69 nm). The  $7 \times 2 \times 2$  sphere assembly corresponds to the G1 domain of proteoglycans, while the  $10 \times 2 \times 1$  assembly corresponds to part or all of the E1 domain.

segment, gave rise to wide-angle curves with intensities that were considerably higher than the experimental scattering curves (i.e., these models were insufficiently oblate in shape). A marked improvement was obtained by postulating that the two segments in the binding region were set at right angles (Figure 7b). The final model contained 399 spheres and had an  $R_G$  of 4.94 nm, which is within error of the experimental

value of  $5.1 \pm 0.2$  nm. In this, the C-terminal segment is of size  $2 \times 3 \times 14$  spheres (length 8 nm).

The contrast dependence of the scattering curves for the binding region was calculated by dividing this model into two sections. The scattering density of the 87 spheres corresponding to six oligosaccharide chains and 117 spheres corresponding to the outermost spheres along the length of the

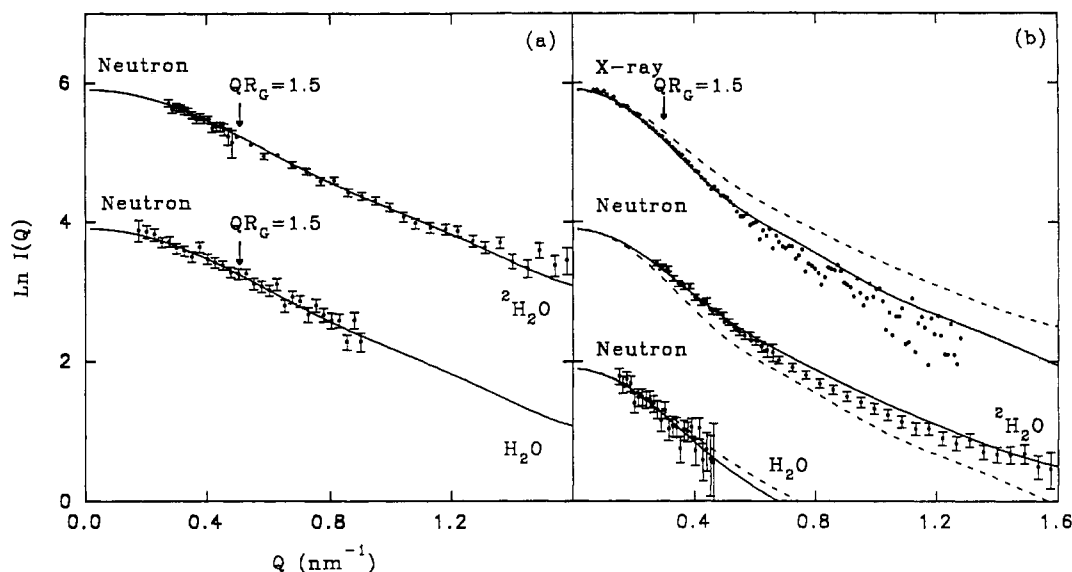


FIGURE 7: Final scattering curve simulations for link protein and binding region. (a) The simulated curve calculated from the link protein model shown in Figure 7 is compared with experimental neutron data collected on instrument D17 in  $^2\text{H}_2\text{O}$  (upper) and instrument D11 in  $\text{H}_2\text{O}$  (lower). The condition  $QR_G = 1.5$  is arrowed for comparison with Figure 4. Error bars correspond to the statistics of neutron data collection. Beam smearing corrections have been applied to the simulated curves (Materials and Methods). (b) The contrast-dependent simulated curves for the binding region model shown in Figure 7 are compared with the X-ray scattering curve of citraconylated binding region (upper), and the neutron scattering curves of native binding region measured on instrument D17 in  $^2\text{H}_2\text{O}$  (center) and on instrument D11 in  $\text{H}_2\text{O}$  (bottom) buffers. The contrast dependence of the curve fits is illustrated by the dashed curves; the top and bottom ones correspond to the simulated curve for neutrons in  $^2\text{H}_2\text{O}$  buffers, and the central one corresponds to that for X-rays.

protein were set to be distinct from that of the 195 remaining innermost protein spheres. From composition data, the electron density of the hydrophilic amino acid and carbohydrate residues was calculated as 468 electrons/ $\text{nm}^3$ , while that for the hydrophobic amino acid residues is 399 electrons/ $\text{nm}^3$  and that for water is 334 electrons/ $\text{nm}^3$ . The approximation was made that these hydrophilic and hydrophobic densities could be assigned to the two sections of binding region. In accordance with these calculated electron densities, the use of a 65:30 weighting for the outermost and innermost spheres in the binding region model resulted in a satisfactory X-ray curve fit (Figure 7b). The neutron match points were calculated as 50%  $^2\text{H}_2\text{O}$  and 34%  $^2\text{H}_2\text{O}$  for the hydrophilic and hydrophobic regions, respectively. In reasonable accord with this, it was found that a 50:60 weighting for the outermost and innermost spheres of the model for the curve obtained in 100%  $^2\text{H}_2\text{O}$  buffers gave the most satisfactory curve fit (Figure 7b). The reversed 50:40 weighting was successfully used to reproduce the limited neutron curve measured in 0%  $^2\text{H}_2\text{O}$  (Figure 7b). In conclusion, these results show that the large differences between the scattering curves for link protein and binding region are well described by models for binding region that contain an extended C-terminal segment. The modeling is however not able to define a precise structure or orientation for this C-terminal segment, and the model of Figure 7b is only indicative of a family of structures that leads to curve fits of similar quality.

(e) *Hydrodynamic Models of the Binding Region and Link Protein.* The sedimentation coefficients  $s_{20,w}^0$  of Bonnet et al. (1985) lead to frictional ratios  $f/f_0$  of 1.8 for native binding region and 1.4 for citraconylated link protein (Table I). This shows that both molecules are elongated. The experimental  $s_{20,w}^0$  values of binding region and link protein may be correlated with the scattering models by the use of hydrated nonoverlapping hydrodynamic spheres (Table I; Figure 6c). For link protein, the use of 28 spheres of diameter 1.688 nm in a  $7 \times 2 \times 2$  array resulted in a calculated  $s_{20,w}^0$  of 2.9 S, which is close to the experimental value of 2.5 S, given that the error in each of the calculated and experimental values

is about  $\pm 0.2$  S (Perkins, 1989). The dimensions of this model are  $12 \text{ nm} \times 3 \text{ nm} \times 3 \text{ nm}$ , which are in agreement with those of the scattering curve model of  $10 \text{ nm} \times 2.5 \text{ nm} \times 3 \text{ nm}$  once allowance is made for hydration. For binding region, it was necessary to add 20 further spheres to the link protein model in order to attain the required hydrated volume. No explicit account of oligosaccharide conformations was made since  $s_{20,w}^0$  data are only a one-parameter measure of macromolecular elongation. By adding an array of  $10 \times 2 \times 1$  spheres at one end of the link protein model, an  $s_{20,w}^0$  of 3.6 S was obtained, which is also close to the experimental  $s_{20,w}^0$  of 3.4 S. While the length of this segment is 17 nm, which is longer than the corresponding length of 8 nm in the scattering curve model, this difference in length is considered to reflect the ill-defined structure of the C-terminal peptide and its associated O-linked oligosaccharides.

## CONCLUSIONS

The principal aim of this study is to interpret the solution scattering results for both link protein and binding region in terms of their sequence similarities and their similar overall three-domain structures. It should however be noted that solution scattering data are limited in defining unique models. The data are spherically averaged because of the random orientations of the scattering particles, so without data from other sources it is not possible to prove any single model correct. Models incompatible with the data can be ruled out. The present analyses were constrained by the known volumes for both binding region and link protein from sequences (Table I), together with the crystal structure of the Ig fold (Figure 6a). It could thus be inferred that both structures contain a compact globular part with dimensions of  $10 \text{ nm} \times 2.5 \text{ nm} \times 3 \text{ nm}$ ; these dimensions define the overall size of the Ig fold and the two PTR domains within this structure. Crystal structures of a variable Ig fold can be used to model this domain; it was of interest to note that the two putative N-linked oligosaccharide sites occur at the opposite end of this domain to where the L1, L2, and L3 hypervariable loops are found. One hypothesis for the role of the L1, L2, and L3 loops



is that these provide the site of protein-protein interaction in the ternary complex formed between G1 domain, link protein, and hyaluronate [see Perkins et al. (1989) for a discussion]. Further insight into the assembly of these three domains awaits the determination of the molecular structures of the G1 and link protein Ig folds and a single PTR domain. The alignment of 20 full PTR sequences in Figure 5b shows that all PTR sequences determined to date can be readily aligned with remarkably few gaps; the exception is found between residues 77 and 78 in the G1 and G2 and versican sequences. This implies that molecular structures for other PTR domains can be obtained by the use of standard molecular graphics modeling techniques.

Binding region and link protein give rise to different elongated solution structures (Figure 6). These differences are explained by noting that the binding region contains part or all of the E1 polypeptide found between the G1 and G2 domains at the N-terminus of proteoglycan monomers. This E1 polypeptide was found to be extended and is heavily glycosylated by O-linked oligosaccharides. This is consistent with secondary structure predictions of the E1 polypeptide conformation, which had suggested that this has a high proportion of coil structure; this region has a high proline content which does not favor regular secondary structures (Perkins et al., 1989). By this solution study, the length of this E1 segment is not well-defined, but is expected to range between 8 and 17 nm.

Electron microscopy studies on binding region within a proteoglycan monomer (Weidemann et al., 1984; Paulsson et al., 1987; Mörgelin et al., 1988, 1989; Dennis et al., 1990) may suffer from artifacts caused by the nonphysiological conditions of measurement. Thus Paulsson et al. (1987) had shown that two circular domains G1 and G2 of diameters 8 and 6 nm, respectively, were visible, separated by  $21 \pm 4$  nm, and with a curved connection of length  $30 \pm 5$  nm between G1 and G2. If the G1 domain was indeed spherical as implied, its volume even allowing for 2 nm in diameter as a decoration artifact would be more than twice that calculated from its known sequence on the basis of standard residue volumes. A spherical model for binding region or link protein is entirely incompatible with the solution scattering results. However, the rotatory shadowing technique may lead to moderately elongated structures appearing more globular and spherical. The G1 domain corresponds to an Ig fold and two PTRs. It is of interest that its observed diameter of  $8 \pm 1.2$  nm from electron microscopy is close to the length of 10 nm determined by solution scattering. The G2 domain corresponds to two PTRs; its diameter of  $6 \pm 1.4$  nm is close to the expected length of two PTRs if each is of size  $3 \text{ nm} \times 3 \text{ nm} \times 2.5 \text{ nm}$  (see Results and Discussion, section d). In confirmation of these deductions, Mörgelin et al. (1989) report similar diameters for G1 and G2 in a survey of proteoglycans, which after correction for decoration effects are  $8\text{--}9 \pm 2$  nm and  $6\text{--}9 \pm 2$  nm, respectively. Furthermore, an estimate of a minimum distance of 12 nm between G1 domains bound to hyaluronate (Mörgelin et al., 1988) is also compatible with the solution scattering length of 10 nm. No comparisons can be made with electron micrographs of link protein since these have not yet been obtained. Turning to the C-terminal segment of binding region, the solution length of 8–17 nm for this by this study suggests that it forms just part of the 30-nm length estimated for the E1 region by electron microscopy, given that the site of cleavage is not known. The results of electron microscopy also suggested E1 to be stiffened, and when visible, it was also curved. These results are thus generally compatible with a

linear arrangement of three compact globular structures and an extended C-terminal tail at an angle to the main axis. This linear sequence of compact domain structures concurs with similar findings for many other multidomain proteins, for example, those found in the complex multidomain structures of components of the complement cascade of immune defense (Perkins et al., 1990c).

The binding region itself binds to a minimum length of hyaluronate  $\text{HA}_{10}$ . The interaction with hyaluronate will involve highly specific protein-carbohydrate contacts. The length of  $\text{HA}_{10}$  will be about 5 nm if fully extended; it is of interest that this is similar to the length of two PTR domains, since the PTR domains have been implicated in a functional role of binding to hyaluronate (Périn et al., 1987; Bonnet et al., 1985; Perkins et al., 1989). Biochemical evidence shows that a minimum length of  $\text{HA}_{22-24}$  (13–14 nm if extended) is required for effective ternary complex formation with link protein and hyaluronate (Kimura et al., 1979). This suggests that in the ternary complex the PTR domains of binding region and link protein are bound to hyaluronate in an extended arrangement. The dimensions of the models are thus compatible with the known stoichiometry of the binding of hyaluronate saccharides.

The present application of solution scattering to glycoprotein structures has shown again that this technique is useful to identify the extended or compact arrangement of oligosaccharide chains on the surface of a protein. This had previously been seen to be the case from modeling studies performed on heavily glycosylated structures such as  $\alpha_1$ -acid glycoprotein,  $\alpha_1$ -antitrypsin, and C1 inhibitor (Perkins et al., 1985, 1990b; Smith et al., 1990). Here, and previously for  $\alpha_1$ -antitrypsin, it has been shown that the relative electron densities of the hydrophilic and hydrophobic components of glycoproteins make X-rays a more sensitive monitor of the extended nature of oligosaccharide conformations than is the case with the scattering densities encountered in neutron studies. This advantage was well illustrated by the different solution structures seen for two otherwise similar multidomain proteins (Figure 6b). Thus the presence of extended oligosaccharide chains in a globular glycoprotein would be identified by high axial ratios by X-ray scattering (or neutron scattering in 0%  $^2\text{H}_2\text{O}$ ) and confirmed by scattering curve simulations. In this contrast, the carbohydrate moiety scatters more strongly than the protein moiety, and since the former is located on the surface of the protein core at the greatest distance from the center of scattering mass, the carbohydrate component can exert the maximal effect on the scattering curve. The presence of extended oligosaccharide structures would also be supported by neutron Stuhmann  $\alpha$  values that are higher than expected for the glycoprotein of interest.

Solution scattering as applied to glycoproteins of the extracellular matrix was handicapped initially by their tendency to aggregate. The citraconylation of some of the surface lysine groups to reverse their charges inhibits aggregate formation and is a mild procedure that is reversible, so that samples can be recovered if required. This technique had been successful in ultracentrifugation studies of link protein (Bonnet et al., 1985). Controls are however required to ensure that the chemical modification does not lead to a denatured protein structure, such as in the case of fully citraconylated  $\alpha$ -crystallin (Bindels et al., 1984), and this procedure may not therefore be generally applicable to all glycoproteins.

#### ACKNOWLEDGMENTS

We thank Dr. C. Nave, Dr. J. Torbet, and Dr. D. L. Worcester for generous instrumental support at the Insti-



tute-Laue-Langevin, Grenoble, and the Synchrotron Radiation Source, Daresbury.

## REFERENCES

- Baker, J. R., & Caterson, B. (1979) in *Glycoconjugate Research* (Gregory, J. D., & Jeanloz, R. W., Eds.) pp 329–340, Academic Press, New York.
- Bindels, J. G., Misdorn, L. W., & Hoenders, H. J. (1985) *Biochim. Biophys. Acta* 828, 255–260.
- Bonnet, F., Dunham, D. G., & Hardingham, T. E. (1985) *Biochem. J.* 228, 77–85.
- Bonnet, F., Périn, J.-P., Lorenzo, F., Jollès, J., & Jollès, P. (1986) *Biochim. Biophys. Acta* 873, 152–155.
- Chothia, C. (1975) *Nature (London)* 254, 304–308.
- Chothia, C., & Lesk, A. M. (1987) *J. Mol. Biol.* 196, 901–917.
- Deak, F., Kiss, I., Sparks, K. J., Argraves, W. S., Hampikian, G., & Goetinck, P. F. (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83, 3766–3770.
- Dennis, J. E., Carrino, D. A., Schwartz, N. B., & Caplan, A. I. (1990) *J. Biol. Chem.* 265, 12098–12103.
- Doege, K. J., Sasaki, M., Horigan, E., Hassell, J. R., & Yamada, Y. (1987) *J. Biol. Chem.* 262, 17757–17767.
- Doege, K. J., Sasaki, M., Kimura, T., & Yamada, Y. (1991) *J. Biol. Chem.* 266, 894–902.
- Dudhia, J., & Hardingham, T. E. (1989) *J. Mol. Biol.* 206, 749–753.
- Dudhia, J., & Hardingham, T. E. (1990) *Nucleic Acids Res.* 18, 1292.
- Garcia de la Torre, J., & Bloomfield, V. A. (1977a) *Biopolymers* 16, 1747–1761.
- Garcia de la Torre, J., & Bloomfield, V. A. (1977b) *Biopolymers* 16, 1779–1793.
- Ghosh, R. (1981) Institut Laue Langevin Internal Publication 81GH29T.
- Glatter, O., & Kratky, O., Eds. (1982) *Small angle X-ray scattering*, Academic Press, New York.
- Goetinck, P. F., Stirpe, N. S., Tsonis, P. A., & Carlone, D. (1987) *J. Cell Biol.* 105, 2403–2408.
- Goldstein, L. A., Zhou, D. F. H., Picker, L. J., Minty, C. N., Bargatze, R. F., Ding, J. F., & Butcher, E. C. (1989) *Cell* 56, 1063–1072.
- Hardingham, T. E. (1986) *Rheumatology* 10, 143–183.
- Hassell, J. R., Kimura, J. H., & Hascall, V. C. (1986) *Annu. Rev. Biochem.* 55, 539–567.
- Ibel, K., & Stuhmann, H. B. (1975) *J. Mol. Biol.* 93, 255–266.
- Jacrot, B., & Zaccari, G. (1981) *Biopolymers* 20, 2413–2426.
- Kimura, J. H., Hardingham, T. E., Hascall, V. C., & Solorsh, M. (1979) *J. Biol. Chem.* 254, 2600–2609.
- Kiss, I., Deak, F., Mestric, S., Delius, H., Soos, J., Dékány, K., Argaves, W. S., Sparks, K. J., & Goetinck, P. F. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 6399–6403.
- Kratky, O. (1963) *Prog. Biophys. Mol. Biol.* 13, 105–173.
- Le Glédic, S., Périn, J.-P., Bonnet, F., & Jollès, P. (1983) *J. Biol. Chem.* 258, 14759–14761.
- Mörgelin, M., Paulsson, M., Hardingham, T. E., Heinegård, D., & Engel, J. (1988) *Biochem. J.* 253, 175–185.
- Mörgelin, M., Paulsson, M., Malmström, A., & Heinegård, D. (1989) *J. Biol. Chem.* 264, 12080–12090.
- Muir, H. (1979) *Adult Articular Cartilage* (Freeman, M. A. R., Ed.) 2nd Ed., Pitman Medical, London.
- Nave, C., Helliwell, J. R., Moore, P. R., Thomson, A. W., Worgan, J. S., Greenall, R. J., Miller, A., Burley, S. K., Bradshaw, J., Pigram, W. J., Fuller, W., Siddons, D. P., Deutsch, M., & Tregear, R. T. (1985) *J. Appl. Crystallogr.* 18, 396–403.
- Neame, P. J., Christner, J. E., & Baker, J. R. (1986) *J. Biol. Chem.* 261, 3519–3535.
- Paulsson, M., Mörgelin, M., Wiedemann, H., Beardmore-Gray, M., Dunham, D. G., Hardingham, T. E., Heinegård, D., Timpl, R., & Engel, J. (1987) *Biochem. J.* 245, 763–772.
- Périn, J.-P., Bonnet, F., Thuriéau, C., & Jollès, P. (1987) *J. Biol. Chem.* 262, 13269–13272.
- Perkins, S. J. (1986) *Eur. J. Biochem.* 157, 169–180.
- Perkins, S. J. (1988a) *Biochem. J.* 254, 313–327.
- Perkins, S. J. (1988b) in *New Comprehensive Biochemistry* (Neuberger, A., & Van Deenen, L. L. M., Eds.) Vol. 18B, Part II, pp 143–264, Elsevier, Amsterdam.
- Perkins, S. J. (1989) in *Dynamic Properties of Biomolecular Assemblies* (Harding, S. E., & Rowe, A. J., Eds.) Chapter 15, pp 226–245, Royal Society of Chemistry, London.
- Perkins, S. J. (1990) *FEBS Lett.* 271, 89–92.
- Perkins, S. J., & Weiss, H. (1983) *J. Mol. Biol.* 168, 847–866.
- Perkins, S. J., & Pepys, M. B. (1986) *Protides Biol. Fluids* 34, 323–326.
- Perkins, S. J., Miller, A., Hardingham, T. E., & Muir, H. (1981) *J. Mol. Biol.* 150, 69–95.
- Perkins, S. J., Kerckaert, J. P., & Loucheux-Lefebvre, M. H. (1985) *Eur. J. Biochem.* 147, 525–531.
- Perkins, S. J., Dunham, D. G., Hardingham, T. E., & Muir, H. (1986) *Synchrotron radiation*, Appendix to the Daresbury Annual Report 1985/86, pp 178–178.
- Perkins, S. J., Nealis, A. S., Dudhia, J., & Hardingham, T. E. (1989) *J. Mol. Biol.* 206, 737–753.
- Perkins, S. J., Nealis, A. S., & Sim, R. B. (1990a) *Biochemistry* 29, 1167–1175.
- Perkins, S. J., Smith, K. F., Amatayakul, S., Ashford, D., Rademacher, T. W., Dwek, R. A., Lachmann, P. J., & Harrison, R. A. (1990b) *J. Mol. Biol.* 214, 751–763.
- Perkins, S. J., Smith, K. F., & Nealis, A. S. (1990c) *Biochem. Soc. Trans.* 18, 1151–1154.
- Perkins, S. J., Nealis, A. S., Sutton, B. J., & Feinstein, A. (1991) *J. Mol. Biol.* 221, 1345–1366.
- Ratcliffe, A., & Hardingham, T. E. (1983) *Biochem. J.* 213, 371–378.
- Satow, Y., Cohen, G. H., Padlan, E. A., & Davies, D. R. (1986) *J. Mol. Biol.* 190, 593–604.
- Smith, K. F., Harrison, R. A., & Perkins, S. J. (1990) *Biochem. J.* 267, 203–212.
- Stamenkovic, I., Amiot, M., Pesando, J. M., & Seed, B. (1989) *Cell* 56, 1057–1062.
- Stuhlsatz, H. W., Keller, R., Becker, G., Oeben, M., Lennartz, L., Fischer, D. C., & Greilung, H. (1989) in *Keratan sulphate: Chemistry, Biology, Chemical Pathology* (Greiling, H., & Scott, J. E., Eds.) pp 1–11, The Biochemical Society, London.
- Tang, L. H., Rosenberg, L., Reiner, A., & Poole, A. R. (1979) *J. Biol. Chem.* 254, 10523–10531.
- Weidemann, H., Paulsson, M., Timpl, R., Engel, J., & Heinegård, D. (1984) *Biochem. J.* 224, 331–333.
- Wolfe, E. J., Gause, W. C., Pelfrey, C. M., Holland, S. M., Steinberg, A. D., & August, J. T. (1990) *J. Biol. Chem.* 265, 341–347.
- Zimmermann, D. R., & Ruoslahti, E. (1989) *EMBO J.* 8, 2975–2981.