Crystal structure and carbohydrate-binding properties of the human cartilage glycoprotein-39

Fusetti, F; Pijning, T; Kalk, KH; Bos, E; Dijkstra, BW; Dijkstra, Bauke W.

Published in:
The Journal of Biological Chemistry

DOI:
10.1074/jbc.M303137200

Publication date:
2003

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 278, No. 39, Issue of September 26, pp. 37753–37760, 2003

Printed in U.S.A.

Crystal Structure and Carbohydrate-binding Properties of the Human Cartilage Glycoprotein-39*

Received for publication, March 26, 2003, and in revised form, June 12, 2003
Published, JBC Papers in Press, July 8, 2003, DOI 10.1074/jbc.M303137200

Fabrizia Fusetti‡, Tjaard Pijning‡, Kor H. Kalk‡, Ebo Bos§, and Bauke W. Dijkstra‡¶

From the ‡Laboratory of Biophysical Chemistry, Department of Chemistry, University of Groningen, Nijenborgh 4, 9747AG Groningen and §NV Organon, 5340BH Oss, The Netherlands

The human cartilage glycoprotein-39 (HCgp-39 or YKL40) is expressed by synovial cells and macrophages during inflammation. Its precise physiological role is unknown. However, it has been proposed that HCgp-39 acts as an autoantigen in rheumatoid arthritis, and high expression levels have been associated with cancer development. HCgp-39 shares high sequence homology with family 18 chitinases, and although it binds to chitin it lacks enzymatic activity. The crystal structure of HCgp-39 shows that the protein displays a $(\beta/\alpha)_8$-barrel fold with an insertion of an $\alpha + \beta$ domain. A 43-Å long carbohydrate-binding cleft is present at the C-terminal side of the β-strands in the $(\beta/\alpha)_8$ barrel. Binding of chitin fragments of different lengths identified nine sugar-binding subsites in the groove. Protein-carbohydrate interactions are mainly mediated by stacking of side chains of aromatic amino acid residues. Surprisingly, the specificity of chitin binding to HCgp-39 depends on the length of the oligosaccharide. Although chitin disaccharides tend to occupy the distal subsites, longer chains bind preferably to the central subsites in the groove. Despite the absence of enzymatic activity, long chitin fragments are distorted upon binding, with the GlcNAc at subsite −1 in a boat conformation, similar to what has been observed in chitinases. The presence of chitin in the human body has never been documented so far. However, the binding features observed in the complex structures suggest that either chitin or a closely related oligosaccharide could act as the physiological ligand for HCgp-39.

Rheumatoid arthritis (RA)[1] is a chronic, systemic, and extremely disabling disease that affects mainly women. It is characterized by inflammation of synovial tissues causing joint pain, swelling, and stiffness. At the moment, there is no cure for rheumatoid arthritis, and therefore current treatments focus on relieving the pain and inflammation and on stopping joint destruction. A better understanding of how rheumatoid arthritis affects the body may help to design and develop efficient treatments. Some of the most promising areas of research for the abatement of RA include oral tolerance therapy and the use of monoclonal antibodies and drugs to block the activity of interleukins, neutrophils, and T-cells (1).

The human cartilage glycoprotein-39 (HCgp-39 or YKL40) is a candidate autoantigen in RA having the capability to induce T-cell-mediated autoimmune response (2–4). Moreover, recent studies have shown that HCgp-39-derived peptides can induce immunologic tolerance in patients with persistent RA (5). HCgp-39 is secreted by articular chondrocytes and synovial cells (6), and it is expressed in peripheral blood-derived macrophages in association with their differentiation from monocytes to macrophages (7). Increased levels of HCgp-39 have been detected in the blood of RA patients (8, 9), in inflamed tissues (7, 10, 11), and recently also in spinal fluids following neural damage (12). Additionally, abnormal expression of HCgp-39 is related to the extent of liver fibrosis (13) and to the development of cancer (14–16). Based on these findings it has been suggested that HCgp-39 could be used as a clinical marker for early detection of disease activity and for testing the effectiveness of therapy (17).

The precise physiological function of HCgp-39 is unknown. However, it has been demonstrated recently that HCgp-39 stimulates cell proliferation of connective tissue acting synergistically with the insulin-like growth factor 1 (18), consistent with a proposed role in tissue remodeling and cellular differentiation (7, 19). A similar function has been suggested for the homologous imaginal disc growth factors that promote cell proliferation in *Drosophila* (20).

HCgp-39 is closely related to the mammalian proteins YM1 (21) and YKL39 (22), oviduct-specific glycoproteins (23), and to the human chitinolytic enzyme chitotriosidase (24). Based on amino acid sequence and structure relationships all of these proteins have been grouped in glycoside hydrolase family 18, which also includes bacterial and plant chitinases (25).[2] These mammalian chitinase-like proteins (CLPs), which lack catalytic activity, could function as lectins able to recognize specific glycan structures present in mammalian tissues. Crystal structures of CLPs have confirmed a high degree of structure similarity to the family 18 chitinases (27–29), but none of those crystal structures has clearly revealed how CLPs bind oligosaccharides.

Here we present the structure of HCgp-39 and for the first time for a mammalian CLP, the crystal structures of complexes with intact chitin fragments bound in the carbohydrate-binding groove. Surprisingly, the HCgp-39 interacts distinctly differently with short and long chitin fragments.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1NWR, 1NWS, 1NWT, and 1NWU) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ To whom correspondence should be addressed. Tel.: 31-50-3634381; Fax: 31-50-3634800; E-mail: bauke@chem.rug.nl.

[1] The abbreviations used are: RA, rheumatoid arthritis; HCgp-39, human cartilage glycoprotein-39; CLP, chitinase-like protein; BES, 2-[bis(2-hydroxyethyl)amino]ethanesulfonic acid; HA, hyaluronic acid;

[2] The reader is also referred to Coutinho, P. M., and Henrissat, B. (1999) Carbohydrate-Active EnZYmes server at afmb.cnrs-mrs.fr/~cazy/CAZY/index.html.

TABLE I
Structure determination of HCgp-39: Data processing and phasing
For the definition of standard crystallographic quantities the reader is referred to Ref. 26.

	Native	Hg (λ_1)[a]	Hg (λ_2)[a]	Pt[a]	U[a]
Cell parameters(Å)	$a = b = 127.3$ $c = 107.7$	$a = b = 127.9$ $c = 109.0$	$a = b = 127.9$ $c = 109.0$	$a = b = 128.2$ $c = 108.7$	$a = b = 128.7$ $c = 108.7$
Resolution range (Å)	50–2.6	50–2.7	50–2.7	35–3.5	35–3.5
Total reflections	205,355	105,799	102,143	59,507	53,652
Unique reflections	52,490	44,097	42,076	21,978	21,801
Completeness (%)‡[b]	99.6 (96.2)	91.0 (91.4)	86.5 (85)	97.3 (87.9)	96.8 (90.8)
R_{sym} (%)‡[b]	6.0 (25.9)	5.4 (18.7)	6.9 (18.5)	14.9 (38.3)	13.6 (39.4)
$I/\sigma(I)$[b]	18.2 (4.1)	12.5 (3.3)	10.4 (2.6)	8.0 (3.0)	8.5 (2.4)
Number of sites	ND[c]	4	4	18	4
Phasing power (iso/ano)		2.06/0.93	ND[c]/0.39	1.12/0.53	1.15
Overall figure of merit			0.40		

[a] Hg = $(C_2H_5Hg)_3PO_4$, Triethyl mercuri phosphate; Pt = *cis*-$Pt(NH_3)_2Cl_2$, *cis*-Platinum-di-amino-di-chloride; U = $U(NO_2)_4$, Uranium nitrite.
[b] The numbers in parentheses refer to the highest resolution bin.
[c] ND, not determined.

EXPERIMENTAL PROCEDURES

Crystallization and Heavy Atom Derivative Search—Recombinant HCgp-39, produced by a Chinese hamster ovary transfectant and purified to homogeneity via cation exchange chromatography, immunochromatography, and size exclusion chromatography, was kindly provided by NV Organon, Oss, The Netherlands.

HCgp-39 was crystallized using the hanging drop vapor-diffusion method. The protein solution contained 4 mg/ml HCgp-39 in 1 M NaCl, 10 mM BES buffer, pH 7.5. The precipitant consisted of 10% polyethylene glycol 8000, 0.5 M NaCl in 0.1 M sodium citrate buffer, pH 5.1. The crystals grew to a maximum size of $0.8 \times 0.15 \times 0.2$ mm^3. They belong to space group $P4_3$ with 4 molecules in the asymmetric unit and a solvent content of ~50%. HCgp-39 crystals, which were briefly soaked in a solution of 10% glycerol in mother liquor and then flash-frozen, diffracted to a resolution of 2.6 Å using synchrotron radiation. Subsequently, improved diffraction could be obtained by avoiding glycerol in the cryoprotectant solution. Instead, the HCgp-39 crystals were stabilized in 35% polyethylene glycol 8000, 0.2 M NaCl in 20 mM sodium citrate buffer, pH 5.5, or 20 mM BES buffer, pH 7.5, for a few hours and then flash-frozen in liquid nitrogen.

Heavy atom derivatives were obtained by soaking the crystals for 3 days in solutions of mother liquor containing 3 mM heavy atom compound. Among the over 20 different heavy atom compounds tested, 10 gave acceptable diffraction but only 4 could be used as isomorphous derivatives. The complex structures were obtained by soaking HCgp-39 crystals in the presence of 5–10 mM chitin oligosaccharides ($GlcNAc_2$, $GlcNAc_4$, $GlcNAc_5$) in 25% (w/v) polyethylene glycol 8000, 0.2 M NaCl, 20 mM BES-NaOH, pH 8.0, for about 15 h at room temperature.

Data Collection and Processing—Diffraction data for the native data set were collected at the ELETTRA Synchrotron, Trieste, Italy. To attempt structure determination by the MAD method, data on the $(C_2H_5)_3HgPO_4$ derivative was collected at beamline BM14 at the European Synchrotron Radiation Facility, Grenoble, France, with the wavelengths tuned to $\lambda_1 = 1.005$ Å and $\lambda_2 = 0.775$ Å (Table I). The other derivative data sets were collected at beamline X31 at the European Molecular Biology Laboratory outstation at Deutsches Elektronen-Synchrotron, Hamburg, Germany, with the wavelength tuned to $\lambda = 0.99$ Å, which, although not optimal, allowed the detection of the anomalous signal of the *cis*-$Pt(NH_3)_2Cl_2$ derivative (Table I). X-ray diffraction data on HCgp-39 crystals soaked with $(GlcNAc)_5$ were collected in-house on an ENRAF-Nonius CuKα rotating anode generator equipped with a MacScience DIP2030 image plate detector. Data on $(GlcNAc)_4$- and $(GlcNAc)_2$-complexed HCgp-39 were collected at the ESRF, Grenoble, at beamlines ID14–1 and ID14–2, respectively. All data were integrated and merged using the programs DENZO and SCALEPACK (30). Programs from the CCP4 suite (31) were used for various calculations. Details on data collections and processing statistics are given in Tables I and II.

Structure Determination and Refinement—Heavy atom position search, parameter refinement, and phases were calculated with programs from the PHASES package (32). The anomalous data were included in the calculations. The initial phases had a figure of merit of 0.40. Next, the non-crystallographic symmetry operators were determined by manual superposition of the heavy atom sites and refined with the program IMP of the O package (33). The program DM (34) was used to improve the initial phases by applying solvent flattening, histogram matching, and 4-fold non-crystallographic symmetry averaging. The final figure of merit was 0.70 for data to 3.0 Å. The obtained electron density map could be interpreted unambiguously, and all residues from positions 22 to 383 could be placed in electron density. The model was built with the program O (33). Refinement was performed with the program CNS (Crystallography & NMR System) (35). 10% of the reflections, selected in very thin resolution shells, was set aside for R_{free} factor calculations. Non-crystallographic symmetry restraints were imposed throughout the whole refinement procedure. Bulk solvent correction and overall anisotropic *B*-factor corrections were included from the beginning resulting in an R_{factor} of 23.5% and R_{free} of 25.8%. Incorporation of experimental phase information during refinement resulted in improved R_{factor} and R_{free} values of 21.6 and 24.0%, respectively, for data to 2.6 Å. The final model consists of 4×362 residues, 4×2 *N*-acetylglucosamine residues (*N*-linked glycosylation), and a total of 117 water molecules. The stereochemical quality was assessed with the program PROCHECK (36). Refinement statistics are given in Table II.

The structures of HCgp-39 in complex with chitin fragments were solved by standard methods starting from the coordinates of the native HCgp-39 structure determined at 2.6 Å resolution. To optimize the position of the 4 molecules in the asymmetric unit a three-dimensional rigid body refinement of the native HCgp-39 model was performed using data up to 3.5 Å. Subsequently the model was adjusted using a round of simulated annealing. At this stage the electron density corresponding to the soaked ligands was visible in the chitin-binding groove from both $2F_o - F_c$ and $F_o - F_c$ Fourier maps. The refinement, carried out with the program CNS (Crystallography & NMR System) (35), was followed by monitoring the behavior of the free R_{factor} calculated from 10% of the data. The $\beta(1, 4)$-linked *N*-acetylglucosamine residues were fitted in the electron density using the program O (33). After a few rounds of x-ray-restrained energy minimization followed by *B*-factor refinement, water molecules were added. Non-crystallographic symmetry restraints were kept for the protein core throughout the whole refinement procedure. Refinement statistics are reported in Table II. Coordinates and structure factors have been deposited in the protein data bank with entry codes 1NWR, 1NWS, 1NWT, and 1NWU.

RESULTS AND DISCUSSION

A recombinant glycosylated form of the human cartilage glycoprotein-39 was crystallized. The structure was determined using the multiple isomorphous replacement method supplemented with anomalous scattering from three heavy atom derivatives (Table II). The native structure of HCgp-39 was determined at 2.6 Å. The final model comprises 4 HCgp-39 monomers (residues 22–383) and 117 water molecules. The 4 crystallographically independent molecules are essentially identical, having a root mean square difference of 0.02 Å for 361 equivalent Cα atoms. In a Ramachandran plot 91.4% of the 1252 non-glycine and non-proline residues are in the most favored regions, 8.0% lie in the additional allowed sections, and the remaining 0.6% in the generously allowed parts.

In the crystal structure of HCgp-39 the first 21 N-terminal residues that form the leader sequence are not present. Disulfide bonds are formed between Cys^{26} and Cys^{51} and between Cys^{300} and Cys^{364}. Three *cis* peptide bonds are present (resi-

TABLE II
Structure determination of HCgp-39 complexes: Data processing and refinement statistics
For the definition of standard crystallographic quantities the reader is referred to Ref. 26.

	Native	$(GlcNAc)_2$	$(GlcNAc)_4$	$(GlcNAc)_5$
Cell parameters (Å)	$a = b = 127.3$ $c = 107.7$	$a = b = 128.2$ $c = 109.3$	$a = b = 128.7$ $c = 108.8$	$a = b = 128.3$ $c = 108.5$
Cryoprotectant	10% glycerol	35% PEG8K	35% PEG8K	35% PEG8K
Resolution range (Å)	50–2.6	30–2.7	50–2.2	30–2.5
X-ray source	ELETTRA	ESRF ID14–1	ESRF ID14–2	in-house
Total reflections	205,355	148,105	572,110	227,637
Unique reflections	52,490	48,088	107,198	59,975
Completeness (%),†[b]	99.6 (96.2)	96.0 (92.8)	98.8 (99.2)	97.9 (96.1)
$I/\sigma(I)$[b]	18.2 (4.1)	11.4 (2.5)	18.5 (2.8)	10.1 (2.0)
R_{sym} (%)‡[b]	6.0 (25.9)	8.7 (48.0)	7.8 (44.2)	11.4 (62.6)
R_{factor} (%)	21.6	20.9	22.5	21.7
R_{free} (%)[a]	24.0	25.2	24.5	25.3
Residues in the asymm. unit	4×362	4×362	4×362	4×362
Water molecules	117	122	439	502
Overall average β-factor (Å^2)	36.4	46.5	45.0	36.1
root mean square deviation from ideality				
Bond lengths (Å)	0.008	0.009	0.006	0.009
Bond angles (°)	1.3	1.4	1.3	1.4

[a] R_{free} corresponds to R_{factor} calculated using 10% of the total reflections selected randomly in thin shells and excluded during refinement.
[b] The numbers in parentheses refer to the highest resolution bin.

dues Ser^{57}-Phe^{58}, Leu^{140}-Tyr^{141}, and Trp^{352}-Ala^{353}). It is noteworthy that two of them are located in the sugar-binding groove and that all three *cis* peptide bonds have been found in corresponding positions in the structures of other family 18 glycosylhydrolases (37–39). HCgp-39 is *N*-glycosylated at Asn^{60} and 2 β(1, 4)-linked GlcNAc residues were clearly visible in the electron density. Inspection of the electron density map revealed a sequence mismatch at position 311 where an isoleucine residue is present instead of the threonine expected from the deposited amino acid sequence (Swiss-Prot P36222). The sequence mismatch has been identified in all four HCgp-39 molecules in the asymmetric unit, and it has been verified by mass spectrometry analysis (data not shown).

Protein Structure—HCgp-39 has approximate dimensions of $54 \times 40 \times 44$ Å, and it displays the typical fold of family 18 glycosylhydrolases (25). Its three-dimensional structure (Fig. 1) consists of an eight-stranded β/α-barrel domain with a second domain, composed of six antiparallel β-strands and one α-helix ($\alpha + \beta$ domain), inserted after strand β7. Superpositions of HCgp-39 with the bacterial chitinases ChiA and ChiB (37, 38) result in root mean square differences of 1.54 and 1.46 Å for 330 and 321 equivalent Cα pairs, respectively. The N-terminal fibronectin III-module domain and the C-terminal chitin-binding domain found in ChiA and ChiB, respectively, are absent in HCgp-39.

Carbohydrate Binding—A distinguishing feature of HCgp-39 is its ability to bind to insoluble chitin (40). By soaking HCgp-39 crystals in the presence of chitin fragments we could determine the structures of HCgp-39 in complex with $(GlcNAc)_2$, $(GlcNAc)_4$, and $(GlcNAc)_5$ (Table II, Fig. 2). These structures allowed the identification of 9 sugar-binding subsites in the 43 Å groove, across the C-terminal ends of the β-strands of the $(\beta/\alpha)_8$ barrel. They were numbered from -6 to $+3$ from the non-reducing end, as is the current convention (41), with the -1 and $+1$ subsites corresponding to the ones immediately upstream and downstream from the scissile bond in chitinases (Figs. 2 and 3). The carbohydrate-binding groove of HCgp-39 is lined with aromatic residues, which have hydrophobic stacking interactions with the hydrophobic sides of the bound sugar rings (Fig. 3). Among them, Trp^{99} and Trp^{352} additionally participate in ligand recognition by making hydrogen bonds with the C3-OH of the -1 GlcNAc and the C7-O of the -2 GlcNAc, respectively. The $\alpha + \beta$ domain also takes part in sugar binding, with Arg^{263} making hydrogen bonding interactions with the C6-OH of the -1 GlcNAc. Specific recognition of GlcNAc is achieved at subsites -5, -2, and $+2$. Binding specificity at subsite -5 is determined by a hydrogen-bonding interaction between the Glu^{70}-Oϵ and the nitrogen atom of the *N*-acetyl group of the bound GlcNAc. At subsite -2 the binding of GlcNAc is stabilized by a complex hydrogen-bonding network involving Trp^{352}, Asn^{100}, and water mediated interactions with Arg^{263}, Glu^{290}, and Gly^{97}. Finally, the *N*-acetyl group of $+2$ GlcNAc is hydrogen-bonded, via a water molecule, to the side chain hydroxyl groups of Thr^{184} and Ser^{179}. A more detailed description of protein-carbohydrate interactions is given in Fig. 3.

The presence of chitin fragments in the binding groove does not cause drastic conformational changes in the protein. The only exception is the rotation of the side chain of Trp^{99} to make a parallel stacking interaction with the $+1$ sugar ring (Figs. 3 and 4). This interaction is critical for stabilizing the structure of the chitin oligosaccharide, which becomes bent and twisted upon binding (Fig. 3). In family 18 chitinases, including the human chitotriosidase (42), the residue corresponding to Trp^{99} is already in this chitin-binding orientation in the uncomplexed enzyme. Because of the difference in orientation of Trp^{99}, the chitin-binding groove is wider in HCgp-39 than in human chitotriosidase. This could explain why the 39-kDa form of human chitotriosidase, unlike HCgp-39, cannot bind to insoluble chitin (40). Interestingly, in the plant chitinase hevamine, which acts as a defense endochitinase, the residue corresponding to Trp^{99} is a glycine, a substitution that ensures a completely open chitin-binding groove and full accessibility of the central subsites located near the catalytic residues.

HCgp-39 is a lectin, and bound carbohydrates are not hydrolyzed. Interestingly, the binding mode of chitin to HCgp-39 is similar to that observed in productive chitin-chitinase complexes. For instance, we see that the -1 *N*-acetylglucosamine residue has a boat conformation as occurs in chitin when bound to bacterial chitinases (43, 44). The distorted conformation, visible in the $(GlcNAc)_4$ and $(GlcNAc)_5$ complex structures, is stabilized by multiple interactions involving Trp^{99}, Trp^{352}, Arg^{263}, and Tyr^{141} (Fig. 3). Intriguingly, although 6 subsites, from -1 to -6, are available for chitin oligosaccharides to bind in a linear fashion this distorted conformation is preferred. Apparently, binding of GlcNAc to residues -2 to $+2$ is energetically more favorable than binding to sites from -1 to -4. Indeed, the protein-carbohydrate interactions at subsites -3

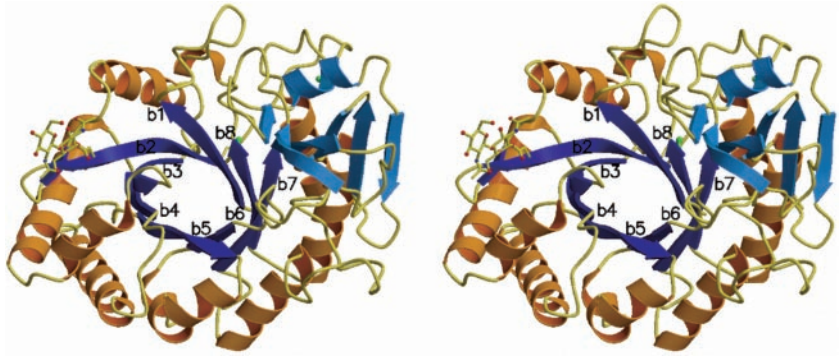

FIG. 1. **Stereo view of the HCgp-39 structure.** The ribbon representation was generated using the programs MOLSCRIPT (57) and RASTER-3D (58). The $(\beta/\alpha)_8$ barrel domain is colored *blue and yellow*. The $\alpha + \beta$ domain is represented in *light blue*. The *N*-glycosylation at residue Asn^{60} is shown as *ball-and-stick*. The β-strands of the $(\beta/\alpha)_8$ barrel are labeled *b1–b8*.

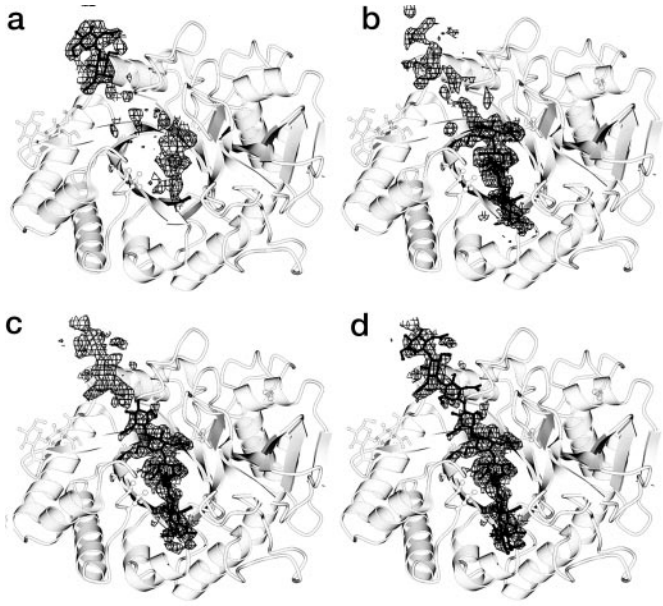

FIG. 2. **Structure of HCgp-39 in complex with chitin fragments.** Ribbon representation of the structure of HCgp-39 with overlaid initial F_o-F_c electron density calculated prior to modeling of the chitin ligand. *a*, $(GlcNAc)_2$; *b*, $(GlcNAc)_4$; *c*, $(GlcNAc)_5$. The modeled $GlcNAc_9$ is shown in *d* with electron density for $(GlcNAc)_5$. The figure was generated using the programs BOBSCRIPT (59) and RASTER-3D (58).

and −4 are not very strong (see below). This is in contrast to what has been observed in hen egg white lysozyme where a δ-lactone derived from $(GlcNAc)_4$ occupies subsites −1 to −4 (45). Furthermore, the *N*-acetyl carbonyl oxygen atom of the −1 GlcNAc bound to HCgp-39 is oriented away from the C1 atom, whereas in chitinase-substrate complexes the oxygen atom is close to the C1 carbon as expected for the substrate-assisted catalytic mechanism of family 18 chitinases (39, 46). This difference results from the E140L and D138A substitutions that in HCgp-39 eliminate the essential catalytic residues and create a large hydrophobic cavity where the *N*-acetyl group can be accommodated in extended conformation, causing total impairment of the chitinase activity.

The structure of HCgp-39 in complex with $(GlcNAc)_5$ allowed unambiguous identification of subsites −3 to +2. However, F_o-F_c and $2F_o-F_c$ Fourier maps, calculated with the refined model including 5 GlcNAc residues bound to HCgp-39 showed residual density features at both extremities of the built-in chitopentaose (Fig. 2). 1 *N*-acetylglucosamine residue was added to the oligosaccharide at the +3 site resulting in a slightly better R_{free} factor and a better appearance of the electron density map. Consequently the occupancy values were corrected to 0.5 for the residues in the −3 and +3 sites. These results suggest that in HCgp-39 a chitopentaose molecule binds in subsites −2 to +3 and subsites −3 to +2 with comparable frequency. Subsequent examination of the electron density maps revealed more uninterpreted features beyond the identified six sites, indicative of additional binding sites for the chitopentaose in HCgp-39 (Fig. 3). Similar results were obtained from soaking studies with $(GlcNAc)_4$, where the initial maps showed the strongest electron density in subsites −2 and −1. Nevertheless, residual electron density, although not clearly interpretable, was visible along the whole binding groove especially upstream subsite −3 (Fig. 3).

Compelling experimental evidence for occupation of subsites −5 and −6 came from the structure of HCgp-39 with chitobiose $((GlcNAc)_2)$ (Fig. 3). In these subsites Trp^{71}, Glu^{70}, and Tyr^{34} form a flat area on the protein surface on which the disaccharide can be accommodated. Initial electron density maps showed that binding of $(GlcNAc)_2$ also occurs in the middle part of the groove, although with lower occupancy (Fig. 2). Because of the poor quality of the electron density we have refrained from building any ligand in those subsites. Soaking of HCgp-39 crystals in the presence of $(GlcNAc)_3$ produced similar results (data not shown). Finally, HCgp-39 is glycosylated and although Asn^{60} is in the vicinity of the groove, the *N*-linked glycan, which was left intact during the crystallization, does not interfere with chitin binding in the crystal.

The described HCgp-39 crystallographic complexes clearly identify subsites −3 to +3, −5, and −6 in the carbohydrate-binding groove. Combining the experimental structures yielded a model in which an oligosaccharide made up of 9 GlcNAc residues is bound to HCgp-39 (Figs. 2 and 3). The model was obtained by inserting a β(1, 4)-linked GlcNAc at position −4. At subsite −4, no aromatic side chains are available to stabilize the binding, but the side chain of Glu^{70} could form a hydrogen bond interaction with the C6-OH of the GlcNAc residue.

Our results show that binding of either long or short oligosaccharides to HCgp-39 is possible. Remarkably, whereas chitin fragments of 4 or more GlcNAc residues tend to occupy the central part of the groove, shorter oligosaccharides bind preferentially at the more distant subsites on the protein surface. The existence of two distinct binding sites with selective affinity for long and short oligosaccharides has never been documented for chitinases or for chitinase-like proteins and could be unique for HCgp-39. This suggests the possibility that HCgp-39 like many other lectins could function by cross-linking two targets. Although, because the 2 binding sites are only about 10 Å apart, the simultaneous binding of 2 glycoproteins is questionable. Alternatively, ligand binding to the two sites could be either mutually exclusive and thus provoke different effects or synergic, with the surface site recruiting ligands that are subsequently fixed in the central part of the groove. However, the binding of a chito-oligosaccharide from subsites −6 to +3 may also occur, although the occurrence of such a long chito-oligosaccharide in mammals has not yet been demonstrated.

The observation that chitin oligosaccharides bind to

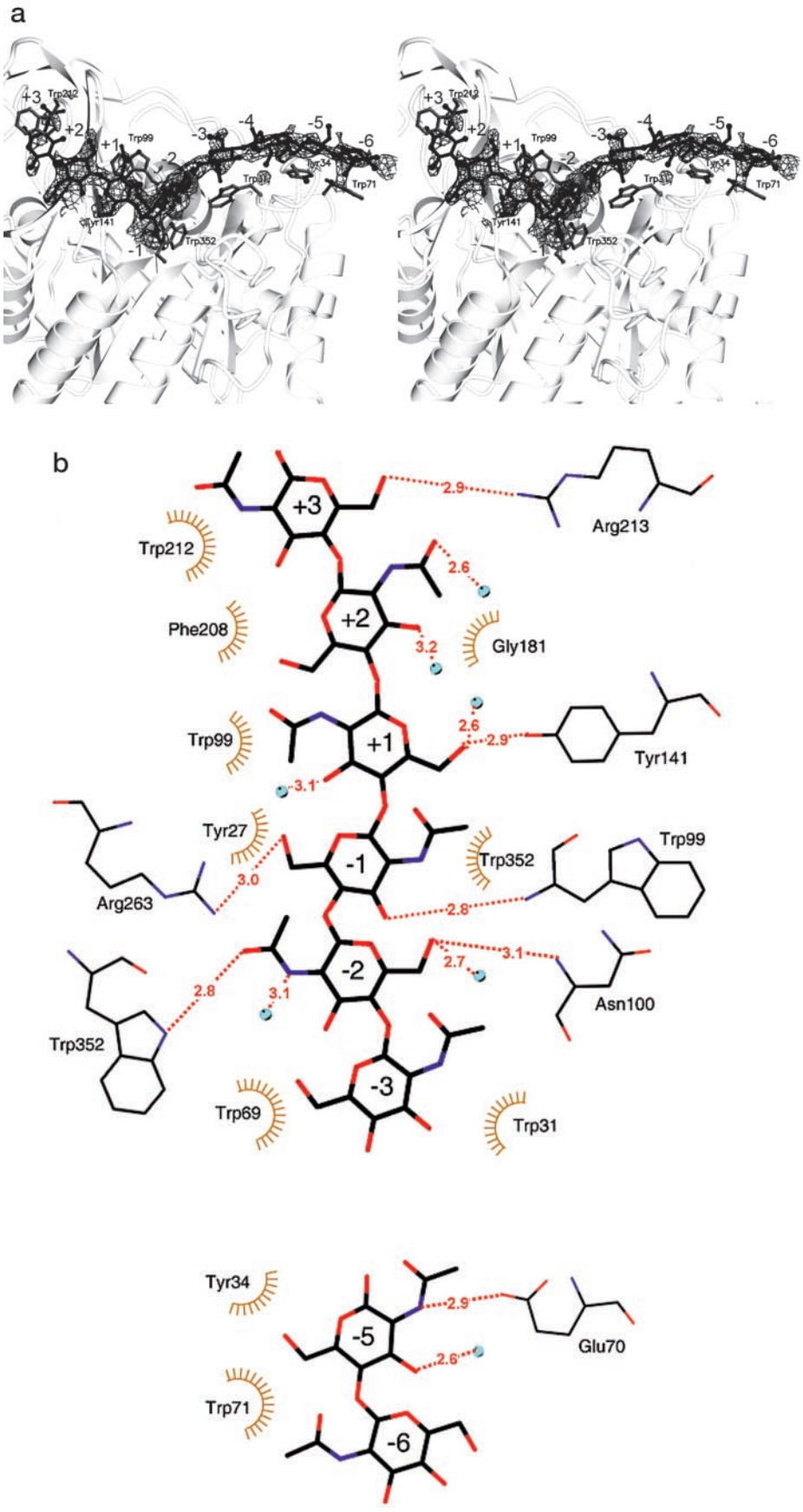

FIG. 3. **Model structure of HCgp-39 with a long chitin fragment.** *a*, stereo view of the binding site showing the structure of a bound chitin oligosaccharide and the lining of aromatic amino acids. The "boat" structure of GlcNAc is visible in the *−1* subsite. The $F_o - F_c$ electron density map was calculated with the HCgp-39-$(GlcNAc)_5$ crystallographic data prior to model building of the ligand. $(GlcNAc)_5$ is shown in *black* in *ball-and-stick* representation; the HCgp-39 structure is shown as a *ribbon* representation in *light gray*. The aromatic residues lining the carbohydrate-binding site are shown in *ball-and-stick* representation. The figure was generated using the programs BOBSCRIPT (59) and RASTER-3D (58). *b*, schematic representation of the chitin-binding site showing protein-carbohydrate interactions at subsites *−3* to *+3*, *−6*, and *−5*. The bound chitin and the amino acid side chains are shown in *thick* and *thin black lines*, respectively. The hydrogen-bonding interactions are shown as *red dashed lines* with atomic distances in Å. Amino acid residues having hydrophobic interactions with the bound chitin are schematically drawn as *brown half-circles*. Water molecules are shown as *blue spheres*. The figure was generated with the program LIGPLOT (60) followed by manual editing.

HCgp-39 suggests interesting speculations. It has been found in vertebrates that short chito-oligosaccharides are used as primers for the synthesis of hyaluronic acid (HA) and that they are kept at the reducing ends of the nascent glycans (47, 48). HA is composed of repeating disaccharide units of *N*-acetylglucosamine and D-glucuronic acid linked together by alternat-

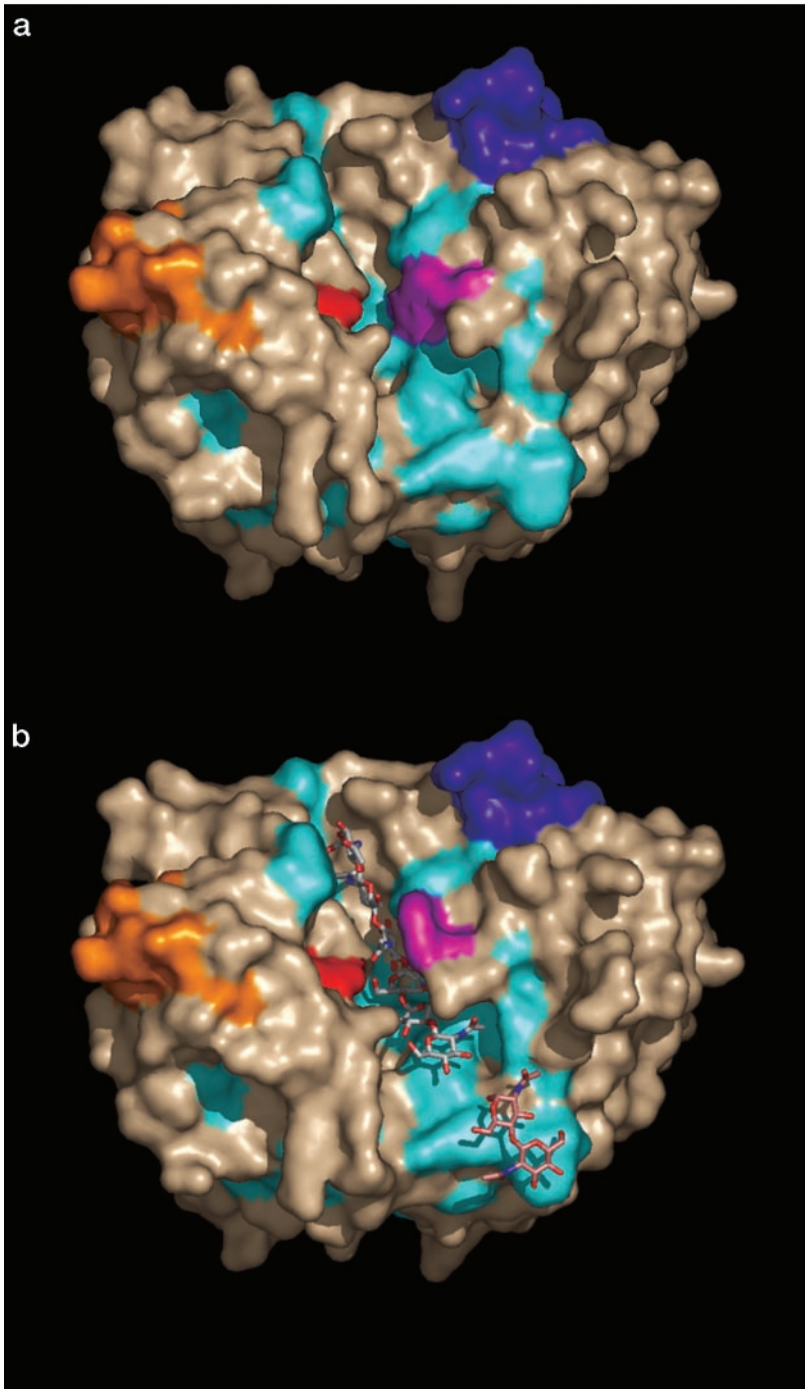

FIG. 4. **Surface representation of the structures of native HCgp-39 (*a*) and of HCgp-39 in complex with chito-oligosaccharides (*b*).** The protein surface is colored as follows: aromatic amino acids are shown in *cyan*, Trp^{99} is depicted in *violet*, the putative heparin-binding site GRRDKQH (residues 143–149) is colored *blue*, and the epitope 259–271 is colored in *red* (residues 259–265) and *orange* (residues 266–271). The bound $(GlcNAc)_5$ and $(GlcNAc)_2$ are shown in *white* and *pink sticks*, respectively. The figure was generated with the program PyMOL (61).

ing $\beta(1, 4)$ and $\beta(1, 3)$ glycosidic bonds. It is a structural component of hyaline cartilage, synovial joint fluid, and skin tissues. HA plays an important biological role in embryogenesis, cell proliferation, and tissue remodeling, and it is involved in acute and chronic inflammation processes (49). The expression of HCgp-39 *in vivo* has been related to similar events (7, 50, 51). Thus the function of HCgp-39 could be linked to the functions of HA. Because of its chitin-binding ability HCgp-39 could participate in specific signaling processes by perceiving the presence of newly synthesized HA chains that still contain the chito-oligosaccharide.

Furthermore it has been reported that both the bovine and porcine homologs of HCgp-39 (52, 53) are able to bind to heparin. HCgp-39 presents a cluster of basic amino acid residues, GRRDKQH, at position 143–149. Similar consensus sequences have been found in heparin-binding proteins (54). In HCgp-39 this putative heparin-binding site is located in a surface loop (Fig. 4). However, soaking of HCgp-39 crystals or co-crystallization in the presence of fully sulfated heparin fragments and heparin-like compounds did not result in conclusive evidence of binding at this site. Likewise, attempts to model the trisulfated disaccharide unit, $-IdoA(2\text{-}OSO_3)\text{-}GlcNSO_3(6\text{-}OSO_3)-$, abundant in heparin, into the HCgp-39 structure proved unsuccessful. Instead, the structurally related heparan sulfate, which bears less sulfate groups could be a more likely ligand of HCgp-39. Heparan sulfate proteoglycans are present at the cell surface and in the extracellular matrix where they play a role in growth factor/cytokine action, cell adhesion, organogenesis, and wound healing. Heparan sulfate is structurally heterogeneous depending on the producing cell type, and it contains GlcNAc and *N*-unsubstituted glucosamine units (55), the presence of which has also been implicated in the ability to bind to

L-selectins (56). Unlike heparin, heparan sulfate unsulfated fragments can be accommodated in the binding groove of HCgp-39. A similar function has been proposed for the related protein YM1, which showed specificity for glucosamine oligosaccharides (21).

Relationships with Other Chitinase-like Proteins—The crystal structures of HCgp-39 in complex with chitin also help to evaluate the carbohydrate-binding properties of other mammalian CLPs. A multiple sequence alignment indicates extensive sequence similarity among the mammalian CLPs. However, of all residues participating in chitin binding in HCgp-39 only Trp^{31} and Trp^{352} are highly conserved. All residues in HCgp-39 that are involved in making hydrogen-bonding interactions with the bound chitin are conserved in chitotriosidase (24), as well as in MGP40 (27), in bovine CLP1 (52), and in porcine gp38k (53) indicative of a similar ligand specificity for these proteins.

Comparison of the available three-dimensional structures of mammalian CLPs shows high structure similarity apart from small differences in the proximity of the binding cleft. In particular the crystal structure of MGP40 (27) can be superimposed onto HCgp-39 with a root mean square difference of 0.75 Å for 358 equivalent Cα pairs. All chitin-binding residues, including Trp^{99}, are conserved in both structures at the same position. Based on structure and sequence homology (83% identity) it can be concluded that MGP40 is the equivalent of HCgp-39 expressed by goat cells.

The structure of the imaginal disc growth factor-2 from *Drosophila* (29) (15.7% sequence identity) can be superimposed onto HCgp-39 with a root mean square difference of 1.57 Å for 327 equivalent Cα atoms. Although it has been proposed that both imaginal disc growth factor-2 and HCgp-39 could function by stimulating cell proliferation through a similar signaling pathway involving also the binding to an insulin-like receptor (18, 29), the sequence and structural features indicates that these proteins are likely to display different carbohydrate specificities. In fact, neither the tryptophan residues lining the carbohydrate-binding groove nor the polar residues determining chitin specificity in HCgp-39 and chitotriosidase are conserved in imaginal disc growth factor-2.

Conclusions—In the present study we determined the structure of HCgp-39 in complex with chitin fragments of different length by means of protein x-ray crystallography. Our structures show that HCgp-39 is able to bind to chitin in a similar fashion as seen in family 18 chitinases. Sugar binding occurs via van der Waals interactions with the side chains of aromatic amino acid residues and several hydrogen-bonding interactions involving the sugar-hydroxyl groups. Chitin oligosaccharides fit strikingly well in the binding groove of HCgp-39. However, the *N*-acetyl group, which is characteristic of GlcNAc, is only specifically recognized at subsites -2 and -5. Furthermore, we show that in HCgp-39 binding specificity also depends on the length of the chito-oligosaccharide. Although short chitin fragments bind at one end of the groove longer oligosaccharides bind in a distorted manner to the central subsites. The functional significance of this unique property needs further investigation.

HCgp-39 stimulates cell proliferation, and it is secreted during inflammation and in response to tissue injury. However the exact role of HCgp-39 and the nature of its physiological ligand remain unknown at present. Because of its chitin-binding ability HCgp-39 could function in host defense mechanisms against chitin-containing pathogens. Alternatively we propose that HCgp-39 could bind to chitin-like oligosaccharides present in nascent chains of hyaluronan, which is also involved in tissue-repairing events and inflammation. Modeling of alternative candidate ligands revealed that although binding of heparin in the HCgp-39 groove appears unlikely, fragments of unsulfated heparan sulfate can be modeled in place of chitin. The implications of carbohydrate binding for the biological function of HCgp-39 await further study.

Finally, HCgp-39 is a candidate autoantigen in RA. Although no evidence has yet been found of a direct involvement of anti-HCgp-39 antibodies in RA onset, it has been shown that immune response to the HCgp-39 epitopes 259–271 and 263–275 correlates with disease activity (2, 4, 5). The structure of HCgp-39 shows that although region 266–275 maps onto the protein surface, residues 259–265 are buried in the vicinity of the chitin-binding groove (Fig. 4).

REFERENCES

1. Firestein, G. S. (2003) *Nature* **423,** 356–361
2. Verheijden, G. F., Rijnders, A. W., Bos, E., Coenen de Roo, C. J., van Staveren, C. J., Miltenburg, A. M., Meijerink, J. H., Elewaut, D., de Keyser, F., Veys, E., and Boots, A. M. (1997) *Arthritis Rheum.* **40,** 1115–1125
3. Cope, A. P., Patel, S. D., Hall, F., Congia, M., Hubers, H. A., Verheijden, G. F., Boots, A. M., Menon, R., Trucco, M., Rijnders, A. W., and Sonderstrup, G. (1999) *Arthritis Rheum.* **42,** 1497–1507
4. Tsark, E. C., Wang, W., Teng, Y. C., Arkfeld, D., Dodge, G. R., and Kovats, S. (2002) *J. Immunol.* **169,** 6625–6633
5. Kavanaugh, A., Genovese, M., Baughman, J., Kivitz, A., Bulpitt, K., Olsen, N., Weisman, M., Matteson, E., Furst, D., Van-Vollenhoven, R., Anderson, J., Cohen, S., Wei, N., Meijerink, J., Jacobs, C., and Mocci, S. (2003) *J. Rheumatol.* **30,** 449–454
6. Hakala, B. E., White, C., and Recklies, A. D. (1993) *J. Biol. Chem.* **268,** 25803–25810
7. Kirkpatrick, R. B., Emery, J. G., Connor, J. R., Dodds, R., Lysko, P. G., and Rosenberg, M. (1997) *Exp. Cell Res.* **237,** 46–54
8. Johansen, J. S., Kirwan, J. R., Price, P. A., and Sharif, M. (2001) *Scand. J. Rheumatol.* **30,** 297–304
9. Vos, K., Steenbakkers, P., Miltenburg, A. M., Bos, E., van Den Heuvel, M. W., van Hogezand, R. A., de Vries, R. R., Breedveld, F. C., and Boots, A. M. (2000) *Ann. Rheum. Dis.* **59,** 544–548
10. Baeten, D., Boots, A. M., Steenbakkers, P. G., Elewaut, D., Bos, E., Verheijden, G. F., Berheijden, G., Miltenburg, A. M., Rijnders, A. W., Veys, E. M., and De Keyser, F. (2000) *Arthritis Rheum.* **43,** 1233–1243
11. Conrozier, T., Carlier, M. C., Mathieu, P., Colson, F., Debard, A. L., Richard, S., Favret, H., Bienvenu, J., and Vignon, E. (2000) *Ann. Rheum. Dis.* **59,** 828–831
12. Tsuji, T., Matsuyama, Y., Natsume, N., Hasegawa, Y., Kondo, S., Kawakami, H., Yoshihara, H., and Iwata, H. (2002) *Spine* **27,** 732–735
13. Johansen, J. S., Christoffersen, P., Moller, S., Price, P. A., Henriksen, J. H., Garbarsch, C., and Bendtsen, F. (2000) *J. Hepatol.* **32,** 911–920
14. Johansen, J. S., Cintin, C., Jorgensen, M., Kamby, C., and Price, P. A. (1995) *Eur. J. Cancer* **31A**, 1437–1442
15. Cintin, C., Johansen, J. S., Christensen, I. J., Price, P. A., Sorensen, S., and Nielsen, H. J. (1999) *Br. J. Cancer* **79,** 1494–1499
16. Mueller, A., O'Rourke, J., Grimm, J., Guillemin, K., Dixon, M. F., Lee, A., and Falkow, S. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 1292–1297
17. Peltomaa, R., Paimela, L., Harvey, S., Helve, T., and Leirisalo Repo, M. (2001) *Rheumatol. Int.* **20,** 192–196
18. Recklies, A. D., White, C., and Ling, H. (2002) *Biochem. J.* **365,** 119–126
19. Connor, J. R., Dodds, R. A., Emery, J. G., Kirkpatrick, R. B., Rosenberg, M., and Gowen, M. (2000) *Osteoarthr. Cartilage* **8,** 87–95
20. Kawamura, K., Shibata, T., Saget, O., Peel, D., and Bryant, P. J. (1999) *Development* **126,** 211–219
21. Chang, N.-C. A., Hung, S.-I., Hwa, K.-Y., Kato, I., Chen, J.-E., Liu, C.-H., and Chang, A. C. (2001) *J. Biol. Chem.* **276,** 17497–17506
22. Hu, B., Trinh, K., Figueira, W. F., and Price, P. A. (1996) *J. Biol. Chem.* **271,** 19415–19420
23. Buhi, W. C. (2002) *Reproduction* **123,** 355–362
24. Renkema, G. H., Boot, R. G., Muijsers, A. O., Donker-Koopman, W. E., and Aerts, J. M. F. G. (1995) *J. Biol. Chem.* **270,** 2198–2202
25. Henrissat, B., and Davies, G. (1997) *Curr. Opin. Struct. Biol.* **7,** 637–644
26. Drenth, J. (1999) *Principles of Protein X-ray Crystallography*, 2nd Ed., Springer-Verlag, Berlin
27. Mohanty, A. K., Singh, G., Paramasivam, K., Jabeen, T., Sharma, S., Yadav, S., Kaur, P., Kumar, P., Srinivasan, A., and Singh, T. P. (2002) *J. Biol. Chem.* **278,** 14451–14460
28. Sun, Y.-J., Chang, N.-C. A., Hung, S.-I., Chang, A. C., Chou, C.-C., and Hsiao, C.-D. (2001) *J. Biol. Chem.* **276,** 17507–17514
29. Varela, P. F., Llera, A. S., Mariuzza, R. A., and Tormo, J. (2002) *J. Biol. Chem.* **277,** 13229–13236
30. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
31. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. Sect. D* **50**, 760–763
32. Furey, W., and Swaminathan, S. (1995) *Methods Enzymol.* **277,** 590–620
33. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47**, 110–119
34. Cowtan, K., and Main, P. (1996) *Acta Crystallogr. Sect. D* **52**, 43–48
35. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D* **54**, 905–921

36. Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993) *J. Mol. Biol.* **231,** 1049–1067
37. van Aalten, D. M., Synstad, B., Brurberg, M. B., Hough, E., Riise, B. W., Eijsink, V. G., and Wierenga, R. K. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5842–5847
38. Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A. B., Chet, I., Wilson, K. S., and Vorgias, C. E. (1994) *Structure* **2,** 1169–1180
39. Terwisscha van Scheltinga, A. C., Kalk, K. H., Beintema, J. J., and Dijkstra, B. W. (1994) *Structure* **2,** 1181–1189
40. Renkema, G. H., Boot, R. G., Au, F. L., Donker-Koopman, W. E., Strijland, A., Muijsers, A. O., Hrebicek, M., and Aerts, J. M. (1998) *Eur. J. Biochem.* **251,** 504–509
41. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) *Biochem. J.* **321,** 557–559
42. Fusetti, F., von Moeller, H., Houston, D., Rozenboom, H. J., Dijkstra, B. W., Boot, R. G., Aerts, J. M. F. G., and van Aalten, D. M. F. (2002) *J. Biol. Chem.* **277**, 25537–25544
43. van Aalten, D. M. F., Komander, D., Synstad, B., Gaseidnes, S., Peter, M. G., and Eijsink, V. G. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 8979–8984
44. Papanikolau, Y., Prag, G., Tavlas, G., Vorgias, C. E., Oppenheim, A. B., and Petratos, K. (2001) *Biochemistry* **40,** 11338–11343
45. Ford, L. O., Johnson, L. N., Moachin, P. A., Phillips, D. C., and Tjian, R. (1974) *J. Mol. Biol.* **88,** 349–371
46. Tews, I., Terwisscha van Scheltinga, A. C., Perrakis, A., Wilson, K. S., and Dijkstra, B. W. (1997) *J. Am. Chem. Soc.* **119,** 7954–7959
47. Semino, C. E., Specht, C. A., Raimondi, A., and Robbins, P. W. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 4548–4553
48. Meyer, M. F., and Kreil, G. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 4543–4547
49. Lee, J. Y., and Spicer, A. P. (2000) *Curr. Opin. Cell Biol.* **12,** 581–586
50. Johansen, J. S., Jensen, H. S., and Price, P. A. (1993) *Br. J. Rheumatol.* **32,** 949–955
51. Johansen, J. S., Hvolris, J., Hansen, M., Backer, V., Lorenzen, I., and Price, P. A. (1996) *Br. J. Rheumatol.* **35,** 553–559
52. Nyirkos, P., and Golds, E. E. (1990) *Biochem. J.* **269,** 265–268
53. Shackelton, L. M., Mann, D. M., and Millis, A. J. (1995) *J. Biol. Chem.* **270,** 13076–13083
54. Capila, I., and Linhardt, R. (2002) *Angew. Chem. Int. Ed. Engl.* **41,** 390–412
55. Westling, C., and Lindahl, U. (2002) *J. Biol. Chem.* **277,** 49247–49255
56. Norgard-Sumnicht, K., and Varki, A. (1995) *J. Biol. Chem.* **270,** 12012–12024
57. Kraulis, P. (1991) *J. Appl. Crystallogr.* **24,** 946–950
58. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524
59. Esnouf, R. M. (1997) *J. Mol. Graphics Model.* **15,** 133–138
60. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein Eng.* **8,** 127–134
61. DeLano, W. L., The PyMOL Molecular Graphic System, DeLano Scientific LLC., San Carlos, CA