Accelerated Publications

Identification of Active-Site Residues of the Pro-Metastatic Endoglycosidase Heparanase†

Mark D. Hulett,‡ June R. Hornby,‡ Stephen J. Ohms,§ Johannes Zuegg,§ Craig Freeman,‡ Jill E. Gready,§ and Christopher R. Parish*,‡

Division of Immunology and Cell Biology and Division of Biochemistry and Molecular Biology, John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra ACT 2601, Australia

Received September 5, 2000; Revised Manuscript Received October 20, 2000

ABSTRACT: Heparanase is a β-D-endoglucuronidase that cleaves heparan sulfate (HS) and has been implicated in many important physiological and pathological processes, including tumor cell metastasis, angiogenesis, and leukocyte migration. We report herein the identification of active-site residues of human heparanase. Using PSI-BLAST and PHI-BLAST searches of sequence databases, similarities were identified between heparanase and members of several of the glycosyl hydrolase families (10, 39, and 51) from glycosyl hydrolase clan A (GH-A), including strong local identities to regions containing the critical active-site catalytic proton donor and nucleophile residues that are conserved in this clan of enzymes. Furthermore, secondary structure predictions suggested that heparanase is likely to contain an $(\alpha/\beta)_8$ TIM-barrel fold, which is common to the GH-A families. On the basis of sequence alignments with a number of glycosyl hydrolases from GH-A, Glu^{225} and Glu^{343} of human heparanase were identified as the likely proton donor and nucleophile residues, respectively. The substitution of these residues with alanine and the subsequent expression of the mutant heparanases in COS-7 cells demonstrated that the HS-degrading capacity of both was abolished. In contrast, the alanine substitution of two other glutamic acid residues (Glu^{378} and Glu^{396}), both predicted to be outside the active site, did not affect heparanase activity. These data suggest that heparanase is a member of the clan A glycosyl hydrolases and has a common catalytic mechanism that involves two conserved acidic residues, a putative proton donor at Glu^{225} and a nucleophile at Glu^{343}.

Heparan sulfate proteoglycans (HSPGs)[1] are a diverse group of molecules that comprise a protein core covalently linked to linear chains of heparan sulfate (HS), a complex sulfated glycosaminoglycan ≤400 modified sugar residues in length. HSPGs are a ubiquitous component of the cell membrane and the extracellular matrix (ECM) and have important functions as structural molecules and cell surface receptors (*1*−*4*). The HS moieties of HSPG interact with a wide range of molecules, including ECM components (e.g., collagen, laminin, and fibronectin) (*5*, *6*), growth factors and cytokines (e.g., basic fibroblast growth factor, platelet-derived growth factor, hepatocyte growth factor, and IL-2) (*7*−*9*), and enzymes (e.g., lipoprotein lipase) (*10*). By regulating the bioavailability and activity of these molecules, HS plays a major role in a number of cellular processes, including cell adhesion, migration, differentiation, and proliferation (*5*, *6*, *11*).

The enzymatic cleavage of HS chains by the endoglycosidase heparanase (*12*) is critical for the modulation of the biological function of HS-binding proteins. Heparanase activity is also essential in the disassembly of the ECM by invading cells (*13*), particularly metastatic tumor cells and leukocytes entering inflammatory sites. The levels of heparanase expression by these cells correlate with their ability to degrade the ECM and to traverse the vascular endothelium and underlying basement membrane (*13*,*14*). Indeed, heparanase inhibitors have been shown to reduce the incidence of tumor metastasis and inhibit T-cell-mediated delayed-type hypersensitivity and experimental autoimmune encephalo-

† This work was supported by Progen Industries Ltd. (Brisbane, Australia) and an Australian National University Institute of Advanced Studies Australian Government block grant.

* To whom correspondence should be addressed: Division of Immunology and Cell Biology, John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra ACT 2601, Australia. Phone: 61-2-6249-2604. Fax: 61-2-6249-2595. E-mail: Christopher.Parish@anu. edu.au.

‡ Division of Immunology and Cell Biology.

§ Division of Biochemistry and Molecular Biology.

[1] Abbreviations: HS, heparan sulfate; HSPG, heparan sulfate proteoglycan; ECM, extracellular matrix; GH-A, glycosyl hydrolase clan A; PSI, position specific iterated; PHI, pattern hit initiated; SOE PCR, splice overlap extension polymerase chain reaction; TIM, triosephosphate isomerase; NR, nonredundant; gb, Genbank; dbj, DNA database of Japan; pdb, Protein Structure Database; emb, EMBL; sp, Swissprot; dNTP, deoxynucleotide triphosphate; KLH, keyhole limpet hemocyanin; DME, Dulbecco's Modified Eagle's Medium; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate; DEAE, diethylaminoethyl; Tris, tris(hydroxymethyl)aminomethane; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis.

10.1021/bi002080p CCC: $19.00
Published on Web 11/29/2000

myelitis, in animal models of these diseases (*15*−*22*). The release by heparanase of HS-bound growth factors from depots in the ECM and on the cell surface is also believed to play an important role in regulating angiogenesis and wound healing (*23*). Clearly, the essential role heparanase plays in these physiological and pathological processes makes it a potentially important target for the development of anti-tumor and anti-inflammatory drugs.

A number of groups have recently reported the cDNA cloning and expression of an identical novel human heparanase (*24*−*29*). The cDNA was found to encode a unique protein of 543 amino acids that contained a potential signal peptide sequence (residues Met^{1}−Ala^{35}) and six putative N-linked glycosylation sites, but no other obvious structural features. The predicted 543-amino acid polypeptide was suggested to be a precursor form that is processed to a mature 50 kDa active form following removal of the 157 N-terminal amino acids (*24*). More recently, the active enzyme has been claimed to exist as a heterodimer, comprising the previously described 50 kDa polypeptide which is noncovalently linked to an 8 kDa peptide derived from the N-terminus of the heparanase precursor (residues Gln^{36}−Lys^{108}) (*28*). These data suggest that heparanase is expressed as a pre-pro form that is first processed into a 65 kDa pro form upon cleavage of the signal peptide and then further processed by removal of a 6 kDa bridging fragment (Glu^{109}−Gln^{156}) to give the active mature 50 and 8 kDa polypeptide components of the heterodimer. However, it should be noted that it is not yet clear if the heterodimeric form of heparanase represents the active enzyme.

Despite many early claims of the existence of several distinct heparanases, the recent cloning experiments together with biochemical studies suggest that mammalian cells express only a single dominant heparanase enzyme. The same human heparanase cDNA sequence was isolated using amino acid sequences of purified enzymes from a diverse range of tissues, including platelets, placenta, hepatoma cells, and embryonic fibroblasts (*24*−*29*). An identical cDNA sequence has also been isolated from a human T-cell line, and highly homologous sequences have been isolated from metastatic rat tumor cells and activated mouse splenic T-cells (*24*). In addition, physicochemical studies of purified heparanase from human platelets, rat tumors, and rat liver have suggested that these enzymes are structurally related (*30*, *31*). Therefore, the likely existence of one dominant heparanase expressed by all mammalian cells makes it an attractive target for new drug development.

While the cloning of mammalian heparanase was clearly a critical step toward understanding heparanase function, the molecular basis of how heparanase cleaves HS remained unclear. In the study presented here, we provide evidence to suggest that heparanase is related to members of the clan A glycosyl hydrolases (GH-A) and identify two residues crucial for heparanase function. The use of sensitive protein sequence alignment approaches in combination with secondary structure predictions suggested that heparanase contains sequences that are homologous to families 10, 39, and 51 of the GH-A, especially in the active-site regions. This clan of enzymes uses a general acid catalysis mechanism for the hydrolysis of glycosidic bonds that requires two critical residues, a proton donor and a nucleophile (*32*), both of which appear to be conserved in heparanase. Site-directed mutagenesis of these residues, but not others predicted to be outside the active site, completely abolished heparanase activity, suggesting that heparanase has a catalytic mechanism for cleaving HS chains similar to that used by the members of GH-A for their carbohydrate substrates. These data should prove to be useful for further characterization of the enzyme active site and the development of inhibitors of heparanase.

MATERIALS AND METHODS

Database Searches. The amino acid sequence analysis was performed using the Position Specific Iterated BLAST (PSI-BLAST) and Pattern Hit Initiated BLAST (PHI-BLAST) methods (*33*) at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/), and using the All Non-Redundant (NR) amino acid sequence database from April 2000, which includes SwissProt (sp), CDS translation of GenBank (gb), EMBL (emb), the DNA database of Japan (dbj), and the Protein Structure Database (pdb). Default amino acid replacement matrixes and gap penalties were used in all database searches. All secondary structure predictions were made using JPRED[2] (a consensus secondary structure prediction server) (*34*) (http://jura.ebi.ac.uk). Annotations for the glycosyl hydrolases were determined with PROSITE (http://expasy.proteome-.org.au/prosite).

Generation of Heparanase Mutant cDNAs and Expression Constructs. The mutant human heparanase cDNAs were constructed by splice overlap extension (SOE) PCR (*35*) using the wild-type human heparanase cDNA as a template. Briefly, two PCRs were performed to generate overlapping 5′ and 3′ fragments encoding each mutation, which were then spliced together in a third PCR. The 5′ fragments for the E225A, E343A, D367A, E378A, and E396A mutants were generated by amplification with oligonucleotides Hep1a and HepEA-1, -3, -5, -7, and -9, respectively, and the 3′ fragments with Hep1c and HepEA-2, -4, -6, -8, and -10, respectively. The reactions were performed on 100 ng of the human heparanase cDNA in the presence of 500 ng of each oligonucleotide primer, 1.25 mM dNTPs, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, and 1 unit of Pfu polymerase (Gibco-BRL, Gaithersburg, MD) for 25 amplification cycles. The 5′ and 3′ fragments (100 ng each) were spliced together and amplified to produce full-length cDNAs of each heparanase mutant, using oligonucleotides Hep1a and Hep1c under the PCR conditions described above. The sequences of the oligonucleotide primers used in the PCRs and their positions of hybridization on the human heparanase cDNA (*24*) are as follows: Hep1a, 5′-TTTGAATTCAGCCCAAGATGCTGCTGCG-3′ (nucleotide positions 44−62); Hep1c, 5′-TACTCTAGATCAGATGCAAGCAGCAACTTTGGC-3′ (positions 1660−1681); HepEA-1, 5′-CTGTTAGGTGCATTGCCTAG-3′ (positions 734−714); HepEA-2, 5′-CTAGGCAATGCACCTAACAG-3′ (positions 714−734); HepEA-3, 5′-CAGAGCTTGTTGCTCCTAAC-3′ (positions 1090−1071); HepEA-4, 5′-GTTAGGAGCAACAAGCTCTG-3′ (positions 1071−1090); HepEA-5, 5′-GCCCAATTTAGCCAGCCAC-3′ (positions 1161−1143); HepEA-6, 5′-GTGGCTGGCTAAATTGGGC-3′ (positions 1143−1161); HepEA-

[2] A dash indicates consecutive residues, and x(1,*n*) denotes 1−*n* residues of any type.

7, 5′-CATCACCACTGCTATTCCC-3′ (positions 1194−1176); HepEA-8, 5′-GGGAATAGCAGTGGTGATG-3′ (positions 1176−1194); HepEA-9, 5′-GGATCGAAGTTTG-CATCCAC-3′ (positions 1250−1231); and HepEA-10, 5′-GTGGATGCAAACTTCGATCC-3′ (positions 1231−1250). The mutant heparanase cDNA expression constructs were produced by subcloning the mutant cDNAs into the eukaryotic expression vector pCDNA3 (Invitrogen, San Diego, CA). Each cDNA was engineered in the PCRs to have an *Eco*RI site at its 5′ end (the 5′-flanking oligonucleotide primer Hep1a containing an *Eco*RI recognition site) and an *Xba*I site at its 3′ end (the 3′-flanking oligonucleotide primer Hep1c containing an *Xba*I recognition site), which enabled the cDNAs to be cloned into the *Eco*RI and *Xba*I sites of pcDNA3. The nucleotide sequence integrity of all clones was confirmed by automated sequencing using an Applied Biosystems 377 sequencer.

Transfection. Transient transfection of COS-7 cells was performed by the DEAE-dextran method as described previously (*36*), and cells were harvested 72 h post-transfection. COS-7 cells were maintained in Dulbecco's Modified Eagle's Medium (DME), containing 10% (v:v) Fetal Clone II (Hyclone Laboratories, Logan, UT).

Heparanase Activity Assay and Detection by Western Blot Analysis. Heparanase activity was assayed on 4×10^4 COS-7 cell transfectants as described previously (*37*). Western blot analysis of heparanase expression was performed as follows. Transfected COS-7 cells ($2-5 \times 10^6$) were lysed by three rounds of freezing and thawing in 0.5 M NaCl and 1% CHAPS, and cell debris was pelleted by centrifugation. Supernatants were added to 20 μL of concanavalin A−Sepharose beads (Pharmacia, Uppsala, Sweden) and incubated overnight at 4 °C. Beads were then washed twice in 0.5 M NaCl followed by once in normal saline and boiled in reducing SDS sample buffer for 5 min. Supernatants were then electrophoresed under reducing conditions by SDS−PAGE (10% acrylamide) and transferred to nitrocellulose, and Western blot analysis was performed using a purified rabbit anti-human heparanase polyclonal antibody as described previously (*30*). The activities of the mutant heparanase enzymes were normalized to take into account the variation in expression as determined by Western blot analysis. Densitometry was performed using a FujiFilm FLA-3000 Phosphoimager and ImageReader software to determine the intensities of the 50 kDa heparanase bands in the Western blots for each of the transfected cell lines. Activities were normalized on the basis of the expression levels using the formula normalized activity = (wild-type band intensity/mutant band intensity) × mutant activity.

Polyclonal Antibody Production. The anti-heparanase polyclonal antibody was produced by immunizing New Zealand white rabbits with a human heparanase peptide (KYLRLPYPFSNKQVDK, amino acids 462−477). This region in human heparanase was predicted to be solvent-accessible by the Kyte−Doolittle protein hydropathy algorithm (*38*) and differs from rat and mouse heparanase by eight and nine residues, respectively. The peptide was coupled to keyhole limpet hemocyanin (KLH) using Imject maleimide-activated KLH (Pierce, Rockford, IL). Rabbits were immunized subcutaneously with 100 μg of peptide KLH emulsified in complete Freund's adjuvant for the primary immunization, or incomplete Freund's adjuvant (Sigma, St. Louis, MO) for subsequent immunizations. The anti-peptide polyclonal antibody was affinity purified by passing serum from an immunized rabbit that had been diluted in PBS through a column comprising the peptide linked to Sulfolink gel support (Pierce) and eluting with 100 mM glycine-HCl (pH 2.3) followed by neutralization with 1 M Tris-HCl (pH 7.0).

RESULTS

Heparanase Is Related to the Glycosyl Hydrolase Clan GH-A. PSI-BLAST searches (*33*) were carried out with the complete amino acid sequences of human, mouse, and rat heparanases (Figure 1). All three sequences showed statistically significant matches to two putative proteins with unknown function from *Arabidopsis thaliana*, emb entry CAB62595 ($E = 4 \times 10^{-33}$, 31% identical to human heparanase, aligned region of 520 amino acids) and gb entry AAC62794 ($E = 6 \times 10^{-8}$, 29% identical to human heparanase, aligned region of 190 amino acids). Further weak similarities were found between mouse and rat heparanase and a number of glycosyl hydrolases, including an endoglucanase F precursor (gb entry AAC45377) from *Fibrobacter succinogenes* (which is identical to the cellulose-binding protein, dbj entry BAA10965.1), xylanase F1 from *Aspergillus oryzae* (dbj entry BAA75475.1), and a putative α-L-arabinofuranosidase from *Thermotoga maritima* (gb entry AAD35369). An additional PSI-BLAST search with the sequence of the endoglucanase F precursor showed that it was significantly similar to α-L-arabinofuranosidases [e.g., $E = 5 \times 10^{-23}$ and $E = 2 \times 10^{-13}$ for the matches to these enzymes from *T. maritima* (gb entry AAD35369) and *Bacillus subtilis* (sp entry ABFA_BACSU), respectively] (Figure 1). The α-L-arabinofuranosidases are members of glycosyl hydrolase family 51, which is contained within glycosyl hydrolase clan GH-A (*39*). Although the overall matches of the heparanase sequences to these glycosyl hydrolases were weak, they showed strong local similarity to part of the inferred active site (containing the proton donor residue) of the endoglucanase F precursor and the α-L-arabinofuranosidases. By inspecting the PROSITE annotations for glycosyl hydrolases (see Materials and Methods), we found that the heparanase sequences match the known (or inferred) proton donor sites for other families of GH-A (Figure 1). These included members of families 10 (eukaryotic and bacterial xylanases) and 39 (mammalian α-L-iduronidases and bacterial β-xylosidases), which, like members of family 51, are members of clan GH-A. Of these GH-A families, X-ray structures have been determined only for members of family 10, which has an $(\alpha/\beta)_8$ TIM-barrel fold (*32*, *39*, *40*). The inclusion of families 39 and 51 in clan GH-A is based on sequence similarities to other families in the clan (*32*, *39*). However, it is very probable that members of families 39 and 51 will have a three-dimensional structure similar to that of members of family 10.

Further inspection of sequence alignments of the heparanases and glycosyl hydrolase families 10, 39, and 51 revealed several other local similarities beyond the proton donor site. Although the initial PSI−BLAST search with the heparanase sequences found weak similarity only to members of family 51, the most striking local similarities (excluding the inferred proton donor site) were to xylanases from family 10. These local similarities include the regions D^{171}xLYSA (human

heparanase numbering), S[219]WELGNEP (in the proton donor site), L[242]GED, N[260]AKLY, G[376]IEV, and L[393]VDEN.

Secondary structure predictions for the three heparanases, using the JPRED[2] method (*34*), were carried out in parallel with the database searches and revealed an alternating α/β pattern (Figure 1), similar to the $(\alpha/\beta)_8$ TIM-barrel protein fold which is likely to be common to all the glycosyl hydrolase families within clan GH-A (*41*, *42*). In addition, the location of the predicted secondary structural elements of the heparanase sequences agreed with those found in the

X-ray structures of several xylanases (Figure 1) (*32*, *39*, *40*).

Prompted by the finding of similarity between regions of heparanase and the active sites of members of clan GH-A, we carried out PHI-BLAST searches with the sequence of rat heparanase and the pattern W-E-x-G-N-E[2] (based on the putative proton donor site in the endoglucanase F precursor). This search returned the endoglucanase F precursor as a significant match ($E = 2 \times 10^{-4}$), providing strong evidence that heparanase is indeed related to the endoglucanase F precursor and thus, indirectly, to α-L-arabinofuranosidases. Another PHI-BLAST search with the sequence of rat heparanase and the pattern K-K-x(1,3)-W-L-x-E (corresponding to the putative nucleophilic site in the endoglucanase F precursor and α-L-arabinofuranosidases) returned a match to endoglucanase F precursor with an E of 2×10^{-5}. In addition to this demonstration of homology of heparanase and glycosyl family 51, a final PHI-BLAST search with the sequence of rat heparanase and the patterns L-G-E-D-F-V and N-A-K-L-Y (corresponding to two regions of identity with xylanases) returned matches to xylanase F1 from *A. oryzae* (dbj entry BAA75475.1), with E values of 4×10^{-4} and 5×10^{-4}, respectively, providing strong evidence of homology between heparanase and glycosyl hydrolase family 10.

The catalytic mechanism of glycosyl hydrolases from clan GH-A involves two acidic residues, one acting as a proton donor and the other as a nucleophile (*32*). The proton donor site is predicted to be Glu[225] in human heparanase. In this region, the heparanase sequences are very similar to the sequences SWxxxN**E**, N**E**P, and WExGN**E** present in the xylanases, iduronidases, and α-L-arabinofuranosidases, respectively (Figure 1). In addition, the secondary structure prediction of heparanases shows this region has structural elements similar to those found in the crystal structures of the xylanases (*32*, *39*, *40*), namely, a loop between a β-strand and an α-helix. Loops between a β-strand and a α-helix are conserved locations not only for the proton donor and nucleophile in glycosyl hydrolases of family 10 but also generally for active-site residues in enzymes with the TIM-barrel protein fold. Secondary structure predictions of iduronidases and α-L-arabinofuranosidases also place the proton donor site in a loop between a β-strand and an α-helix, suggesting a similar protein fold for these enzymes. The identification of a possible nucleophilic residue in the heparanases is more equivocal because of the weak sequence similarity in the C-terminal region of the heparanase to these glycosyl hydrolases. However, on the basis of the pattern of alternating β-strands and α-helices and the fact that active-site residues are located in conserved loops between a β-strand and an α-helix, which is typical of TIM-barrel proteins, Glu[343] is predicted to be the putative site for the nucleophile. This site shares the WLx**E** motif with endoglucanase F (Figure 1) and a number of other putative glycosyl hydrolases from family 10, e.g., from *Ar. thaliana* (emb entry AL031394) or from *Zea mays* (emb entry AF063865).

In addition, the H[296]HY sequence in heparanases is also present in the endoglucanase F precursor or cellulose-binding protein and in α-L-arabinofuranosidases, from glycosyl hydrolase family 51. The region equivalent to this sequence in glycosyl hydrolase family 10 contains a strictly conserved histidine residue, His[210] in the xylanase from *Penicillium simplicissimum* (sw entry XYNA_PENSI and pdb entry 1BG4). In the crystal structure of XYNA_PENSI, this histidine is located adjacent to the putative active site nucleophile and is, therefore, part of the active site (*40*). The remaining regions with a high degree of similarity between the heparanases and xylanases involve areas more distant from the active site, such as the N[260]AKLY sequence of the heparanases, which in the xylanases is located on a loop connecting an α-helix to a β-strand at the opposite end of the TIM barrel to that containing the active-site residues.

It should be noted that following the completion of this study, the sequence of a novel β-endoglucuronidase from *Scutellaria baicalensis* Georgi (sGUS, dbj entry BAA97804.1) was reported to be homologous to heparanase (*43*). On the basis of this new information, we performed additional PSI-BLAST searches which demonstrated that sGUS provides a definitive link between the mammalian heparanases and members of GH-A (Figure 1). The initial PSI-BLAST searches with the complete sequence of human heparanase returned a significant match to sGUS ($E = 4 \times 10^{-26}$), with the next iteration of PSI-BLAST showing significant matches to the endoglucanase F precursor (gb entry AAC45377, $E = 4 \times 10^{-5}$). Convergence occurred on the fifth iteration of PSI-BLAST, and by this point, there were significant matches to numerous α-L-arabinofuranosidases (e.g., from *Bacteroides ovatus*, sw entry ABF2_BACOV, $E = 4 \times 10^{-87}$) and β-xylosidases (e.g., from *Thermoanaerobacterium* sp., sw entry XYNB_THESJ, $E = 5 \times 10^{-66}$) as well as to a xylanase from *Caldocellum saccharolyticum* (gb entry AAB97373) with an E of 1×10^{-60}.

Site-Directed Mutagenesis of Human Heparanase. On the basis of the above sequence alignments, Glu[225] and Glu[343] of human heparanase (*24*) were predicted to be the putative proton donor and nucleophilic residues, respectively, of the

FIGURE 1: Amino acid sequence alignment of heparanases (***Hep***, human, mouse, and rat) and selected members of the glycosyl hydrolase families 10 [***Xyna***, *Penicil.*, endo-1,4-β-xylanase from *P. simplicissimum* (sp entry XYNA_PENSI and pdb entry 1BG4); *Pseud.*, endo-1,4-β-xylanase A from *Pseudomonas fluorescens* (sp entry XYNA_PSEFL and pdb entry 1CLX); *Cellul.*, exo-glucanase/xylanase from *Cellulomonas fimi* (sp entry GUX_CELFI and pdb entry 1EXP)], 39 [***Idua***, α-L-iduronidases from human (sp entry IDUA_HUMAN) and mouse (sp entry IDUA_MOUSE)], and 51 [***Arab***, endoglucanase F from *Fibrobacter succinogenes* (gb entry AAC45377), α-L-arabinofuranosidases from *B. subtilis* (sp entry ABFA_BACU), and *T. maritima* (gb entry AAD35369.1)]. The alignment includes a putative protein from *Ar. thaliana* (emb entry CAB62595) with unknown function but with significant similarity to the heparanases, and the sequence of sGUS (dbj entry BAA97804.1), a β-endoglucuronidase from *S. baicalensis* (*43*). Predicted consensus secondary structures using JPRED[2] (*34*) (■ for α-helix and □ for β-sheet) are shown for the ***Hep***, ***Idua***, and ***Arab*** groups of sequences; the secondary structure for the three ***Xyna*** sequences is based on a consensus of their X-ray crystal structures. Boxes denote identical residues within each group. The shading indicates identical (dark gray shading and white letters) and similar (light gray shading and black letters) residues between the heparanase and other groups of sequences. Black shading with white letters indicates the known or inferred active-site Glu residues of the ***sGUS***, ***Xyna***, ***Idua***, and ***Arab*** sequences. Similar residues are defined as follows: D and E, acidic; A, G, I, L, and V, aliphatic; N and Q, side chain-containing amide group; F, W, and Y, aromatic; R, H, and K, basic; and S and T, side chain-containing hydroxyl group. Mutation sites for human heparanase are indicated for inactive mutants (▼), for active mutants (●), and for the mutant which was not expressed (◆).

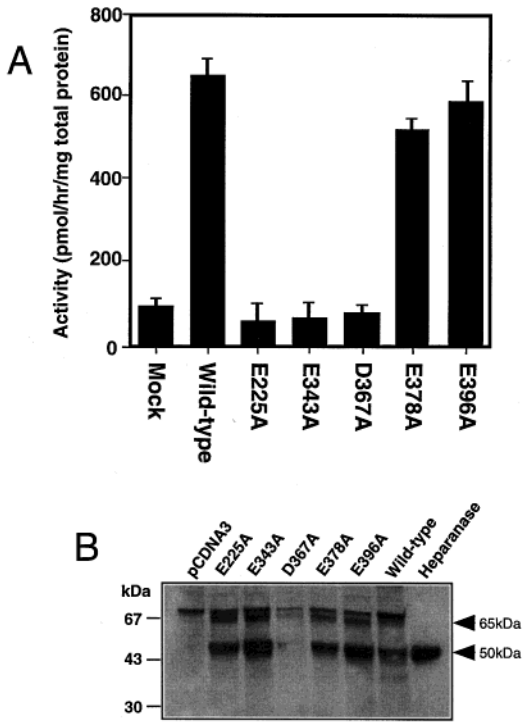

FIGURE 2: Activity and expression of human heparanase mutants. (A) COS-7 cells were transfected with wild-type human heparanase, mutant heparanase pCDNA3 expression constructs, or pCDNA vector alone. Heparanase activity was assayed on whole cell lysates, and the results have been normalized to take into account any variation in expression as determined by Western blot analysis (see Materials and Methods). The results are representative of five independent experiments. Error bars represent the standard error of the mean. (B) Lysates of COS-7 cells transfected with wild-type human heparanase, mutant heparanase pCDNA3 expression constructs, or pCDNA vector alone were precipitated with concanavalin A−Sepharose beads and electrophoresed under reducing conditions by SDS−PAGE on 10% acrylamide gels. Purified 50 kDa active human platelet heparanase (*30*) was also electrophoresed as a positive control. Samples were transferred to nitrocellulose, and Western blot analysis was performed with an anti-human heparanase polyclonal antibody. The positions of 67, 43, and 30 kDa molecular mass markers are shown. Arrows denote the 65 and 50 kDa forms of heparanase. An unknown 70 kDa cross-reacting molecule is present in all samples.

enzyme active site (Figure 1). These two residues, together with three additional acidic residues that were predicted to be outside the putative active site (Asp^{367}, Glu^{378}, and Glu^{396}), were targeted for site-directed mutagenesis. cDNAs encoding mutant heparanases were engineered so that each of the targeted residues was substituted with alanine, and mammalian cell expression constructs were generated by subcloning the mutant cDNAs into the vector pCDNA3. These constructs were transiently transfected into COS-7 cells, and the heparanase activity of lysates of these cells was assessed. COS-7 cells transfected with the E225A, E343A, and D367A mutants showed only low levels of heparanase activity [71.2 ± 79, 72 ± 51, and 81 ± 19 pmol h^{-1} (mg of protein)$^{-1}$, respectively] which were similar to that of the endogenous heparanase activity of mock transfected cells [107 ± 23 pmol h^{-1} (mg of protein)$^{-1}$] (Figure 2A). In contrast, cells transfected with the E378A and E396A mutants exhibited high levels of heparanase activity [578 ± 34 and 590 ± 71 pmol h^{-1} (mg of protein)$^{-1}$, respectively] that were comparable with that of wild-type heparanase-transfected cells [654 ± 59 pmol h^{-1} (mg of protein)$^{-1}$] (Figure 2A). The activities of the heparanase mutants have been normalized to take into account any variation in expression levels between the transfected cells (see below, and Materials and Methods).

To confirm that each of the mutant heparanases was appropriately expressed and processed in this system, Western blot analysis of transfected COS-7 cell lysates was performed using a rabbit anti-human heparanase polyclonal antibody. A 50 kDa protein, corresponding to that expected for active mature heparanase isolated from human platelets, was detected at similar levels in COS-7 cells transfected with pCDNA constructs encoding wild-type heparanase and the E225A, E343A, E378A, or E396A mutants, but not cells transfected with the D367A mutant or pCDNA3 alone (Figure 2B). A 65 kDa protein corresponding to the expected size for unprocessed heparanase was detected at lower levels in transfected cells also expressing the 50 kDa processed heparanase. These data demonstrate that each of the heparanase mutants was processed correctly and expressed at levels similar to that of wild-type heparanase, with the exception of the D367A mutant, which for unknown reasons was not expressed. On the basis of these results, it is clear that the substitution of Glu^{225} or Glu^{343} with alanine results in the complete loss of heparanase activity, which is consistent with the proposal that both of these residues contribute to the active site of heparanase. However, the possibility that these residues play an important structural role is also consistent with the results. The substitution of Glu^{378} or Glu^{396} with alanine did not reduce enzyme activity, which suggests that these residues do not contribute significantly to the activity or structural integrity of heparanase. The effect on heparanase activity of substituting Asp^{367} with alanine could not be determined due to the inability to express this mutant.

DISCUSSION

The definition of the structural basis of heparanase catalytic activity is clearly fundamental to the understanding of how heparanase functions in health and disease. The recent cloning of mammalian heparanase was an important advance, but failed to provide an immediately obvious clue as to the molecular basis of heparanase activity. The use of database screening to identify homologous protein sequences has proven to be an extremely powerful tool for aiding in the identification of functionally important domains and regions in novel proteins. In this study, we have used sensitive protein sequence alignment approaches in combination with secondary structure predictions to demonstrate that mammalian heparanase is related to members of glycosyl hydrolase families 10, 39, and 51 (grouped in the clan A glycosyl hydrolases) (*39*, *42*) and contains regions similar to the conserved active sites of these enzymes. Glycosyl hydrolases catalyze the hydrolytic cleavage of the glycosidic bond between two carbohydrate residues, or between carbohydrates and noncarbohydrate moieties (*32*). As the heparanase substrate is HS, a complex glycosaminoglycan, it is perhaps not surprising that it is homologous to glycosyl hydrolases. A sequence-based characterization system has categorized more than 60 families of glycosyl hydrolases which have been grouped into 10 clans (GH-A−J), and the structures of examples from at least 27 of these have been determined (*32*, *39*). GH-A comprises families 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 51, 53, and 72, which have been shown or predicted to contain $(\alpha/\beta)_8$ TIM-barrel folds. Our secondary

structure predictions also suggest that mammalian heparanase contains alternating α- and β-elements similar to a TIM-barrel fold, which is consistent with our reported sequence relationships to glycosyl hydrolase families 10, 39, and 51. On the basis of these observations, mammalian heparanase together with the homologous emb entry CAB62595 putative protein from *Ar. thaliana* and sGUS from *S. baicalensis* (*43*) (see below and Figure 1) is likely to form a new glycosyl hydrolase family in GH-A.

The identification of a significant degree of sequence identity between regions containing the active-site proton donor and nucleophile residues of specific family 10, 39, and 51 GH-A members (e.g., xylanases, iduronidases, and arabinofuranosidases) and heparanase suggested that heparanase was likely to have a similar catalytic mechanism. Indeed, the substitution of either the putative proton donor (Glu^{225}) or nucleophile (Glu^{343}) residues in heparanase completely abolished enzyme activity. The proposal that Glu^{225} of human heparanase is the proton donor is further supported by sequence analysis of the active site of more than 150 glycosyl hydrolases (*44*−*46*). The proton donor in each of these enzymes is immediately preceded by an invariant Asn and a conserved Trp five residues N-terminal, both of which are found in the human heparanase proton donor region (WELGN**E**). Interestingly, the WxxxN**E** motif is conserved in most confirmed retaining-type glycosyl hydrolases (*45*), which suggests that heparanase may well function via a retaining configuration at the hydrolysis site. Furthermore, the mapping of the heparanase putative Glu^{225} acid donor and the Glu^{343} nucleophile residues onto the crystal structure of endo-1,4-β-xylanase from *P. simplicissimum* (pdb entry 1BG4), based on the alignment shown in Figure 1, demonstrates that they are juxtaposed in the active-site cleft in a topology that is very similar to that known for other TIM-barrel glycosyl hydrolases of GH-A. The three other residues substituted with alanine, Asp^{367}, Glu^{378}, and Glu^{396}, are located in α-helix 7, β-strand 8, and α-helix 8, respectively, all of which are predicted to be outside the active site (Figure 3).

The alignment between the heparanases and xylanases, as shown in Figure 1, indicates that only six of the eight α/β units in the TIM-barrel fold can be assigned unequivocally in heparanase. Clear sequence homology between heparanase and the xylanases begins at the N-terminus of the 50 kDa mature heparanase polypeptide (at residue Lys^{158}), which corresponds to the start of the third α/β unit of the xylanases. This suggests that, in contrast to the xylanases, the first two α/β units of the heparanase TIM-barrel structure may be located elsewhere in the protein. One possibility is that the first two α/β units are shifted in the amino acid sequence to the C-terminal region (amino acids 410−543) but occupy the same spatial location as in the xylanases due to the circular nature of the TIM-barrel structure. Indeed, similar circularly permuted TIM-barrel structures have been described that have permutations in the same place as the heparanase/xylanase pair (*47*). The alignment of the first two α/β units of the xylanases to the C-terminal region of heparanase shows some weak local similarities, but no significant overall similarity. Another possible origin for the two missing heparanase TIM-barrel α/β units could be the 8 kDa polypeptide (residues 36−108). It has been proposed that this fragment forms a heterodimer with the 50 kDa

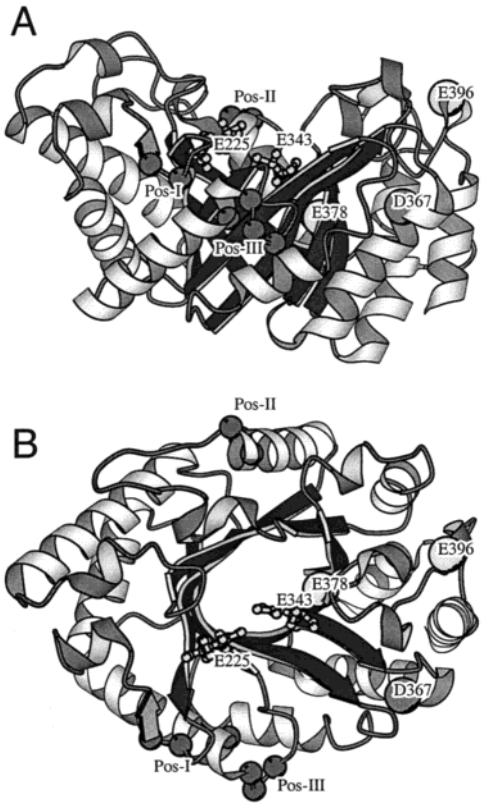

FIGURE 3: Mapping of the human heparanase mutants and basic amino acid clusters onto a xylanase structure. A ribbon diagram of the crystal structure of endo-1,4-β-xylanase from *P. simplicissimum* (pdb entry 1BG4) (*40*) is shown, (A) side view and (B) top view, with the core of the TIM barrel as black β-sheets and white α-helices and the additional structure in gray. The heparanase amino acid mutations and positively charged clusters have been mapped onto the xylanase structure using the alignment given in Figure 1. Balls and sticks represent the putative catalytic proton donor E^{225} and nucleophile E^{343} residues, the mutations of which produced inactive heparanases; large balls represent residues which when mutated had no effect on heparanase activity (E^{378} and E^{396}, light gray) or showed no expression (D^{367}, dark gray); and small balls represent basic amino acid residues in the three positively charged clusters close to the active site: Pos-I (K^{231} and K^{232}), Pos-II (K^{158}, K^{159}, and K^{161}), and Pos-III (R^{272}, R^{273}, K^{274}, and K^{277}). The diagrams were drawn using MOLSCRIPT (*54*).

polypeptide and may be crucial for enzyme activity (*28*); however, there is no sequence similarity with the two N-terminal α/β units of the xylanases. Alternate explanations are that heparanase does not contain a classic $(\alpha/\beta)_8$ TIM-barrel structure but has evolved a specialized structure comprising only six α/β units or that the 6 kDa sequence (residues 109−157) between the 8 and 50 kDa polypeptides originally contributes to the $(\alpha/\beta)_8$ barrel during folding of the pre-pro form of heparanase but is then excised during processing.

Despite the difficulty in defining the boundaries of the putative heparanase TIM barrel, it is clear that heparanase can be divided into three domains, the 157 N-terminal amino acids which appear to be crucial for expression and activity (contains a signal peptide and the 8 kDa fragment that forms a heterodimer with the 50 kDa fragment), the catalytic TIM-barrel domain (residues 158−410), and a C-terminal domain (residues 411−543) of unknown function. The suggestion that mature heparanase actually forms a heterodimer (which may be the active form) is interesting. As described above,

it is possible that the 8 kDa fragment contributes to the TIM barrel to produce an active $(\alpha/\beta)_8$ fold or, by another mechanism, is able to modify the catalytic activity of the active site. It is also possible that excision of the 6 kDa fragment activates the enzyme by revealing the active site, which is a common mechanism used by hydrolytic enzymes to regulate their activities. Clearly, further investigation of the molecular basis of the interaction between the 8 kDa polypeptide and the mature 50 kDa heparanase polypeptide is needed.

Several bacterial heparinases have been described that have no sequence homology to mammalian heparanase despite being able to cleave similar HS substrates. It has been proposed that these enzymes catalyze the cleavage of glycosidic bonds using a lyase mechanism (*48*−*50*), which is distinct from that of the GH-A members. However, positively charged residues in proximity to the active site have been suggested to contribute to the interaction of the bacterial heparinases with anionic HS substrates (*51*). Therefore, positively charged residues may also be important in the binding of HS by mammalian heparanase. A number of basic residues that are conserved in human, rat, and mouse heparanase are found in proximity to the proposed catalytic proton donor and nucleophile, e.g., KK (residues 231 and 232) near Glu^{225} and KK (residues 337 and 338) near Glu^{343}. Furthermore, two clusters of basic amino acids that conform to HS-binding protein consensus sequences (xBBBxxBx or xBBxBx) (*52*, *53*) are present in human heparanase, namely, PRRKTAKM (residues 271−278) and QKKFKN (residues 157−162). On the basis of the simple mapping of these residues onto the structure of endo-1,4-β-xylanase from *P. simplicissimum* (pdb entry 1BG4), three of these basic clusters (residues 231 and 232, 271−278, and 157−162) are predicted to be situated on the top of the TIM-barrel fold in proximity to the proposed active site (Figure 3) and may, therefore, participate in interactions with HS.

As described in the Results, the recently published amino acid sequence for a novel β-glucuronidase from *S. baicalensis* Georgi (sGUS) (*43*) is clearly homologous to heparanase, it being 25% identical over 487 aligned residues in the alignment generated by a PSI-BLAST search (Figure 1). The β-endoglucuronidase sGUS is also a clear homologue of α-L-arabinofuranosidases from glycosyl hydrolase family 51 (e.g., 21% identical to the endoglucanase F precursor over 356 aligned residues), thus linking heparanase to this family. Furthermore, sequence alignments suggest that the putative proton donor and nucleophile residues of heparanase are also conserved in sGUS (Figure 1). Indeed, Sasaki and co-workers (*43*) have performed site-directed mutagenesis studies targeting these putative proton donor and nucleophile residues in sGUS, the results of which are in agreement with those presented herein for heparanase. These data provide further strong supporting evidence that mammalian heparanase is related to members of GH-A, and has the active-site proton donor and nucleophile residues that are conserved with this family.

Also of interest is the identification of two putative *Ar. thaliana* proteins from our PSI-BLAST searches that show a high degree of identity with human, rat, and mouse heparanase. The two proteins, emb entry CAB62595 and gb entry AAC62794, are 31 and 29% identical with human heparanase and are encoded by distinct genes on chromosomes 5 and 3, respectively. The putative gb entry AAC62794 is predicted to consist of 196 amino acids and is homologous to the C-terminal end of heparanase (residues 371−542 of human heparanase). The putative emb entry CAB62595 is predicted to consist of 521 amino acids (homologous to residues 14−542 of human heparanase) and contains a region homologous to the putative TIM-barrel fold of heparanase, including conserved regions that encompass the proposed proton donor (WEFGN**E**DSG) and nucleophile (PWVG**E**SGAY) residues (Figure 1). Thus, emb entry CAB62595 is also likely to be related to the TIM-barrel glycosyl hydrolases and, as predicted for sGUS, is likely to be a novel plant β-glucuronidase.

In conclusion, we postulate that mammalian heparanase is a member of the clan A glycosyl hydrolases and that Glu^{225} and Glu^{343} are the putative acid/base catalytic residues. This work lays the foundation for further investigation of the molecular basis of how heparanase cleaves HS. Such information may facilitate the use of rational drug design approaches in developing compounds that block the heparanase active site or ligand-binding domain, thereby inhibiting heparanase cleavage of HS. Such drugs would be beneficial as therapeutic agents in the treatment of cancer and inflammatory diseases.

ACKNOWLEDGMENT

We thank Ms. Anna Browne for technical assistance with the rabbit immunizations.

REFERENCES

1. Dietrich, C. P., Nader, H. B., and Strauss, A. J. (1983) *Biochem. Biophys. Res. Commun. 111*, 865−871.
2. Kjellen, L., and Lindahl, U. (1991) *Annu. Rev. Biochem. 60*, 443−475.
3. Yanagishita, L., and Hascall, V. C. (1991) *J. Biol. Chem. 267*, 481−539.
4. David, G. (1993) *FASEB J. 7*, 1023−1030.
5. Jackson, R. L., Busch, S. J., and Cardin, A. L. (1991) *Physiol. Rev. 71*, 481−539.
6. Wight, T. N., Kinsella, M. G., and Qwarnstromn, E. E. (1992) *Curr. Opin. Cell Biol. 4*, 793−801.
7. Taipale, J., and Keski-Oji, J. (1997) *FASEB J. 11*, 51−59.
8. Guimond, S., Maccarana, M., Olwin, B. B., Lindahl, U., and Rapraeger, A. C. (1993) *J. Biol. Chem. 268*, 23906−23914.
9. Lyon, M., Deakin, J. A., Mizuno, K., Nakamura, T., and Gallagher, J. T. (1994) *J. Biol. Chem. 269*, 11216−11223.
10. Goldberg, I. J. (1996) *J. Lipid Res. 37*, 693−707.
11. Rapraeger, A. C. (1993) *Curr. Opin. Cell Biol. 5*, 844−853.
12. Nakajima, M., Iramura, T., Di Ferrante, N., and Nicholson, G. L. (1984) *J. Biol. Chem. 259*, 2283−2290.
13. Vlodavsky, I., Eldor, A., Haimovitz-Friedman, A., Matzner, Y., Ishai-Michaeli, R., Lider, O., Naparstek, Y., Cohen, I. R., and Fuks, Z. (1992) *Invasion Metastasis 12*, 112−127.
14. Nakajima, M., Irimura, T., Di Ferrante, D., Di Ferrante, N., and Nicolson, G. L. (1983) *Science 220*, 611−613.
15. Parish, C. R., Coombe, D. R., Jakobson, K. B., and Underwood, P. A. (1987) *Int. J. Cancer 40*, 511−515.
16. Parish, C. R., Hindmarsh, E. J., Bartlett, M. R., Staykova, M. A., Cowden, W. B., and Willenborg, D. O. (1998) *Immunol. Cell Biol. 76*, 104−113.
17. Parish, C. R., Freeman, C., Brown, K. J., Francis, D. J., and Cowden, W. B. (1999) *Cancer Res. 59*, 3422−3441.
18. Willenborg, D. O., and Parish, C. R. (1988) *J. Immunol. 140*, 3401−3405.
19. Nakajima, M., Iramura, T., and Nicholson, G. L. (1988) *J. Cell Biochem. 36*, 157−167.

20. Lider, O., Baharav, E., Mekori, Y. A., Miller, T., Naparstek, Y., Vlodavsky, I., and Cohen, I. R. (1989) *J. Clin. Invest. 83*, 752−756.
21. Hershkoviz, R., Mor, F., Miao, H. Q., Vlodavsky, I., and Lider, O. (1995) *J. Autoimmun. 8*, 741−750.
22. Vlodavsky, I., Mohsen, M., Lider, O., Svahn, C. M., Ekre, H. P., Vigoda, M., Ishai-Michaeli, R., and Peretz, T. (1995) *Invasion Metastasis 14*, 290−302.
23. Ishai-Michaeli, R., Eldor, A., and Vlodavsky, I. (1990) *Cell Regul. 1*, 833−842.
24. Hulett, M. D., Freeman, C., Hamdorf, B. J., Baker, R. T., Harris, M. J., and Parish, C. R. (1999) *Nat. Med. 5*, 803−809.
25. Vlodavsky, I., Freidmann, Y., Elkin, M., Aingorn, H., Atzmon, R., Isgai-Michaeli, R., Bitan, M., Pappo, O., Peretz, T., Michal, I., Spector, L., and Pecker, I. (1999) *Nat. Med. 5*, 793−802.
26. Toyoshima, M., and Nakajima, M. (1999) *J. Biol. Chem. 274*, 24153−24160.
27. Kussie, P. H., Hulmes, J. D., Ludwig, D. L., Patel, S., Navarro, E. C., Seddon, A. P., Giorgio, N. A., and Bohlen, P. (1999) *Biochem. Biophys. Res. Commun. 261*, 183−187.
28. Fairbanks, M. B., Mildner, A. M., Leone, J. W., Cavey, G. S., Mathews, W. R., Drong, R. F., Slighton, J. L., Bienkowski, M. J., Smith, C. W., Bannow, C. A., and Heinrikson, R. L. (1999) *J. Biol. Chem. 274*, 28587−29590.
29. Dempsey, L. A., Plummer, T. B., Coombs, S. L., and Platt, J. L. (2000) *Glycobiology 10*, 467−475.
30. Freeman, C., and Parish, C. R. (1998) *Biochem. J. 330*, 1341−1350.
31. Freeman, C., Browne, A., and Parish, C. R. (1999) *Biochem. J. 342*, 361−368.
32. Davies, G., and Henrissat, B. (1995) *Structure 3*, 853−859.
33. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389−3402.
34. Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., and Barton, G. J. (1998) *Bioinformatics 14*, 892−893.
35. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., and Pease, L. R. (1989) *Gene 77*, 61−68.
36. Seed, B., and Aruffo, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 3365−3369.
37. Freeman, C., and Parish, C. R. (1997) *Biochem. J. 325*, 229−237.
38. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105−132.
39. Hennrissat, B., and Davies, G. (1997) *Curr. Opin. Struct. Biol. 7*, 637−644.
40. Schmidt, A., Schlacher, A., Steiner, W., Schwab, H., and Kratky, C. (1998) *Protein Sci. 7*, 2081−2088.
41. Henrissat, B., and Bairoch, A. (1996) *Biochem. J. 316*, 695−696.
42. Coutinho, P. M., and Henrissat, B. (1999) Carbohydrate-Active Enzymes server at http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html.
43. Sasaki, K., Taura, F., Shoyama, Y., and Morimoto, S. (2000) *J. Biol. Chem. 275*, 27466−27472.
44. MacLeod, A. M., Lindhorst, T., Withers, S. G., and Warren, R. A. (1994) *Biochemistry 33*, 6371−6376.
45. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P., and Davies, G. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7090−7094.
46. Islam, M. R., Tomatsu, S., Shah, G. N., Grubb, J. H., Jain, S., and Sly, W. S. (1999) *J. Biol. Chem. 274*, 23451−23455.
47. MacGregor, E. A., Jespersen, H. M., and Svensson, B. (1996) *FEBS Lett. 378*, 263−266.
48. Sasisehkaran, R., Leckband, D., Godavarti, R., Ventataraman, G., Coony, C. L., and Langer, R. (1995) *Biochemistry 34*, 14441−14448.
49. Sasisehkaran, R., Ventataraman, G., Godavarti, R., Ernst, S. E., Coony, C. L., and Langer, R. (1996) *J. Biol. Chem. 271*, 3124−3131.
50. Godavarti, R., Cooney, C. L., Langer, R., and Sasisekharan, R. (1996) *Biochemistry 35*, 6846−6852.
51. Godavarti, R., and Sasisekharam, R. (1998) *J. Biol. Chem. 273*, 248−255.
52. Cardin, A. D., and Weintraub, H. J. R. (1989) *Arteriosclerosis 9*, 21−32.
53. Hileman, R. E., Fromm, J. R., Weiler, J. M., and Linhardt, R. J. (1998) *BioEssays 20*, 156−167.
54. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI002080P