Molecular Model of Human Heparanase with Proposed Binding Mode of a Heparan Sulfate Oligosaccharide and Catalytic Amino Acids

ARTICLE *in* BIOPOLYMERS · JANUARY 2012

Impact Factor: 2.39 · DOI: 10.1002/bip.21696 · Source: PubMed

4 AUTHORS, INCLUDING:

Nicolas Sapay
Bioaster Technology Research Institute
25 PUBLICATIONS **384** CITATIONS
SEE PROFILE

Anne Imberty
French National Centre for Scientific Research
361 PUBLICATIONS **10,360** CITATIONS
SEE PROFILE

Molecular Model of Human Heparanase with Proposed Binding Mode of a Heparan Sulfate Oligosaccharide and Catalytic Amino Acids

Nicolas Sapay,[1] Éric Cabannes,[2] Maurice Petitou,[2] Anne Imberty[1]

[1] *Centre de recherche sur les Macromolécules Végétales - CNRS, 601 rue de la chimie, BP 53, 38041 Grenoble cedex 9, France*

[2] *Endotis Pharma, Parc Biocitech, 102 avenue Gaston Roussel, 93230 Romainville, France*

Received 7 April 2011; revised 9 July 2011; accepted 11 July 2011
Published online 21 July 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.21696

ABSTRACT:

Heparan sulfate is abundantly present in the extracellular matrix. As other glycosaminoglycans, it is synthesized in the Golgi apparatus and then exposed on the cell surface. The glucuronidase activity of human heparanase plays a major role in the structural remodeling of the extracellular matrix, which underlies cell migration, hence tumor invasion. Heparanase is therefore a major target for anti-cancer treatment. Several inhibitors of its enzymatic activity have been synthesized. However, their design is limited by the absence of experimental structure of the protein. Homology modeling is proposed based on the structure of the endoxylanase from Penicillium simplicissimum co-crystallized with a series of xylan oligosaccharide. The new heparanase model is consistent with the few experimental data suited for the validation of such work. Furthermore, the presence of natural substrates in the template structure allowed us to propose a binding model for a hydrolyzed heparin sulfate pentasaccharide. Several lysine residues have been identified to play a key role in binding to the anionic polysaccharide substrate. In addition, two phenylalanine residues are also potentially important for the interaction with the substrate. The enzymatic mechanism investigated in the light of this new model allows for the proposal of several amino acids that can influence the protonation state of the nucleophile and the proton donor.

Keywords: glycosyl hydrolase family 79; glycosaminoglycan; homology modeling; molecular docking; cancer

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

INTRODUCTION

Heparan sulfate (HS) is a linear polysaccharide belonging to the group of glycosaminoglycans. It is ubiquitously present in the extracellular matrix where it is covalently attached to cell surface or matrix proteins, forming the proteoglycans.[1,2] The repeating unit of glycosaminoglycans is a disaccharide of hexuronic acid and hexosamine. In the case of heparan sulfate, the hexuronic acid is a β-D-glucuronic (GlcA) or an α-L-iduronic (IdoA) acid, with sulfation at position 2, and the hexosamine is a α-D-glucosamine (GlcN), with sulfation at position N, 3 or 6. Heparan sulfate is involved in numerous interactions with proteins from the extracellular matrix and the cell surface. In particular, HS interacts with cellular growth factors like the vascular endothelial growth factors (VEGF)[3,4] or the fibroblast growth factors (FGF).[5,6] It often serves for concentrating the diffusible growth factor at the cell surface, as coligand to the cell surface

Additional Supporting Information may be found in the online version of this article.
Correspondence to: Anne Imberty; e-mail: anne.imberty@cermav.cnrs.fr

receptor of those proteins, which makes HS a core gear of physiological processes such as wound healing,[7] angiogenesis,[8,9] or cancer development.[9]

The synthesis of heparan sulfate is initiated in the Golgi apparatus where it is synthesized as a repeat of GlcA 1–4 linked to an *N*-acetyl-glucosamine (GlcNAc).[10,11] In a second step, the GlcNAc is locally deacetylated, then N- and O-sulfated.[12] Some of the GlcA residues are epimerized at C-5, resulting into IdoA.[13–15] Finally, the polysaccharide 2-, 3-, and/or 6-O sulfated, these sulfations being limited to the initial N/O-sulfated regions.[10,11,16–18] Hence, the biosynthesis of HS leads to a heterogeneous polymer organized in distinct highly sulfated domains of variable size and composition (S-domains) separated by regions of low sulfation enriched in N-acetylated disaccharide units[15,19,20] (NA-domains). Ultimately, the HS chain is exported on the cell surface, covalently linked to cell surface proteins.

In contrast with the synthesis, only few enzymes are involved in the extracellular modification of heparan sulfate: the 6-*O*-endosulfatases Sulf-1 and Sulf-2[21] and the heparanase.[22–24] Sulf-1 and Sulf-2 are responsible for the 6-O desulfation of the highly sulfated S-domains. Heparanase is an *endo*-β-glucuronidase responsible for the hydrolysis of the HS chain into sulfated oligosaccharides of 5–10 kD, i.e., 10–20 sugar units.[25,26] Those three enzymes are tightly related to cancer development and metastasis; thus considered as priority targets for the anti-cancer drug development.[27–29] In particular, the glucuronidase activity of heparanase is required for the structural remodeling of the extracellular matrix underlying cell migration, hence tumor invasion.[26,30] Heparanase is therefore a target for anti-cancer treatment.[29–31] Several inhibitors of the heparanase activity have been synthesized, from sulfated oligosaccharides, like PI-88 (mixture of phosphomannopentaoses sulfate)[32] or PG545 (sulfated tetrasaccharide functionalized with a cholestanyl aglycon),[33] to nonsugar-related molecules, like 2,3-dihydro-1,3-dioxo-1*H*-isoindole-5-carboxylic acid[34] or furanyl-1,3-thiazol-2-yl and benzoxazol-5-yl acetic acid derivatives.[35]

Heparanase is encoded by the *HPSE1* gene.[22–24,36] The protein is initially expressed as a pre-pro-enzyme of 543 amino acids. The 65 kDa pro-enzyme is released after cleavage of the N-terminal signal peptide at the position A^{35} in the endoplasmic reticulum. The pro-enzyme is further processed by the cleavage of a 6 kDa polypeptide constituted by the amino acids S^{110} to $Q,^{157}$ resulting in mature heparanase. The maturation is operated by cathepsin L in lysosomes.[37,38] The active form of heparanase is a heterodimer comprising a small 8 kDa subunit (Q^{36}–E^{109}) noncovalently bound to a large 50 kDa subunit (K^{158}–I^{543}).[39–42]

Sequence analysis and secondary structure prediction have shown that heparanase is very likely organized into two structural domains[43,44]: a catalytic domain with a $(\alpha/\beta)_8$ TIM-barrel fold followed by a β-sandwich C-terminal domain. This latter domain has been shown to be responsible for the heparanase secretion and the activation of the Akt signaling pathway.[44] The catalytic domain presents *endo*-β-D-glucuronidase activity[43,45–48] and belongs to GH family 79 in the carbohydrate-active enzymes database (www.cazy.org).[49] Similar to other glycosyl hydrolases of GH clan A, which includes GH79, the heparanase catalytic reaction uses a retaining mechanism: an acidic amino acid operates a nucleophilic attack on the carbon C1, helped by a protonated acidic residue. For heparanase, a site-directed mutagenesis study has identified the nucleophile (E^{343}) and the proton donor (E^{225}).[43] The enzyme hydrolyzes the β1–4 linkage between a glucuronic acid and a sulfated glucosamine, thus producing oligosaccharides with reducing GlcA residue. Noteworthy, heparanase cleaves the HS chain at only few sites, resulting in HS fragments of large size ($\approx$10 kDa).[25,26] The exact nature of the HS sequence recognized by heparanase is still debated. Although the structure of the natural substrate of heparanase is still elusive, the enzymatic reaction is well known.

To date, no experimental structure of heparanase is available and no crystal structures have been solved for other enzymes of CAZY family 79. Sequence comparison demonstrated similarities between heparanase and GH families 10 (*endo*-1,3/4-β-xylanases), 39 (mammalian α-L-iduronidases and bacterial β-xylosidases), and 51 (bacterial α-L-arabinofuranosidases and endoglucanases).[43] These similarities were used to build three-dimensional models of the structure of the catalytic domain of heparanase.[50,51] The first models have been based on the *endo*-β-D-xylanase from *Penicillium simplicissimum* (GH family 10), for which a crystal structure is available.[52] However, none of them include the 8 kDa subunit, in spite of its requirement for the enzymatic activity of heparanase. This issue has been addressed by Fux and colleagues, who used protein structure prediction server for building a model of the whole mature form of heparanase,[44] based on the crystal structure of the α-L-arabinofuranosidase from *Geobacillus stearothermophilus* (GH family 51).[53] The α-L-arabinofuranosidase hydrolyzes the α1–3 linkage between a xylopyranose from the xylan main chain and an arabinofuranose substituent. Such activity is very different from the one of heparanase, which hydrolyses the β1–4 linkage between two hexopyranoses of the HS main chain. Hence, the choice of an arabinofuranosidase as a template may not be an optimal choice for modeling the binding site of heparanase.

The present study proposes a re-assessment of the template choice for the homology modeling of human heparanase. Sequence and structure analyses led us to build the model from the structure of the endoxylanase from *P. simplicissimum*

Table I Sequence Definition

Name	Definition	Sequence
HPSE	Mature form	Q^{36}–E^{109} (8 kDa) + K^{158}–I^{543} (50 kDa)
$HPSE_{catalytic}$	Catalytic domain of HPSE	L^{51}–E^{109} + K^{158}–T^{416}
$HPSE_{50\ kDa\ catalytic}$	Catalytic part of the 50 kDa domain	K^{158}–T^{416}
$HPSE_{50\ kDa\ noncatalytic}$	Noncatalytic part of the 50 kDa domain	K^{417}–I^{543}
GS3	Monomer analog of mature heparanase[40]	Q^{36}–E^{109} + GSGSGSQ + K^{158}–T^{416}
$GS3_{catalytic}$	Catalytic domain of GS3	L^{51}–E^{109} + GSGSGSQ + K^{158}–T^{416}

co-crystallized with a series of xylan oligosaccharide.[54] This new model corresponds to the GS3 construct, which is a monomeric analog of the mature heterodimeric heparanase, the 8 kDa subunit being linked to the 50 kDa subunit by the peptide GSGSGSQ. This construct has been demonstrated to be an active enzyme.[40] The heparanase model is consistent with the few experimental data suited for the validation of such work. Furthermore, the presence of natural substrates in the template structure allowed us to propose a binding model for a hydrolyzed HS pentasaccharide. It emphasizes the role of residues K^{231}, K^{277}, and R^{383} in the substrate binding. In addition to these three basic residues, aromatic residues F^{101} and F^{385} are also potentially important for the interaction with the substrate. For the first time, the enzymatic mechanism was also investigated. Residues R^{93}, K^{277}, H^{296}, and possibly Y^{298} may be directly involved in the protonation state of the nucleophile and the proton donor.

MATERIALS AND METHODS

Heparanase Sequence

The full-length human heparanase sequence was retrieved from the UniProtKB/Swiss-Prot[55] database (UniProtKB Id Q9Y251; www.uniprot.org). This sequence corresponds to the pre-proform of the enzyme. The different sequences of interest were defined (Table I), including the mature enzyme, the catalytic domain, and its fragments, as well as the chimeric GS3 construct.

Glycosyl Hydrolase Sequences

The glycosyl hydrolase sequences were retrieved from the PDB database (www.pdb.org)[55] by similarity search with the four heparanase sequences using BLAST and PSI-BLAST,[56] as available on the NCBI Web site (blast.ncbi.nlm.nih.gov). BLAST results were filtered by searching for the key words "glycosyl hydrolase" in the retrieved entries. Filtered results were sorted out according to the coverage of the heparanase sequences and the *E*-value.

Sequence Multiple Alignment

The sequences of the mature human heparanase and its GS3 analog were aligned with all the GH10 endoxylanases present in the PDB. The sequence alignment was performed using the Expresso method available on the T-coffee server (www.tcoffee.org).[57,58] The alignment was manually refined to optimized the overlap between the secondary structures of the $(\alpha/\beta)_8$ barrel between heparanase and the endoxylanases. For the heparanase sequences, the secondary structure was predicted with Jpred3[59] and the result was refined with DrawHCA, an implementation of the hydrophobic cluster analysis method.[60–63] Alignment and structure visualization were done with the MultiSeq extension[64] of VMD 1.8.7.[65]

To analyze the residue conservation among the heparanase-related enzymes, the Pfam multiple alignment corresponding to the GH79 family was recovered from the Pfam entry PF03662 (pfam.sanger.ac.uk).[66] It is important to note that the Pfam profile encompasses only the segment L^{150}–G^{370}. Noteworthy too, vertebrate heparanase II sequences were removed from the multiple alignment, as human heparanase II does not carry an enzymatic activity and is not processed by protease.[67] To complete the residue conservation analysis, all full-length heparanase sequences from vertebrates were retrieved from the UniProtKB database and aligned with T-coffee. Alignment visualization was done MultiSeq.

Homology Modeling

Homology modeling of heparanase was carried out with the Orchestrar extension of SybylX 1.2 (www.tripos.com). Structurally conserved regions were modeled using the CHORAL method[68] using a minimal score of -2. Structural gaps of three residues or less were fixed in the best model produce by CHORAL using the BRIDGE methods and a maximum anchor root mean square deviation (RMSD) of 0.8 Å. Structurally variable regions were modeled using the CODA methods[69] as implemented in Orchestrar. Missing segments were searched in the templates, the FREAD database[69] or the PETRA database,[70] this latter one being a database of *ab initio* models. PETRA was only employed for segments smaller than eight residues for which no hit was found in the templates or FREAD. Finally, side chains were added using the Adante method.[71] The disaccharide and trisaccharide initially present in the template structure were conserved all along the process.

In parallel, the catalytic domain or the mature heparanase sequences were submitted to the automatic molecular modeling servers Phyre[72] (www.sbg.bio.ic.ac.uk/~phyre/) and Swiss-Model[73] (swissmodel.expasy.org).

Model Refinement

The final model was energy minimized using NAMD 2.7b1.[74] The system topology was created with LEaP, included in the AMBER 11 package,[75] using the Amber99SB force field[76] for the protein part and the GLYCAM06 force field[77] for the sugar part. Partial charges

Table II The Six Glycosyl Hydrolases Retrieved in the PDB Using the Mature Heparanase Sequence. The Coverage is the Proportion of the Mature Heparanase Sequence Aligned with the Sequence of the Retrieved PDB Entry (the Chain of the PDB Entry is Specified in Parenthesis). The Local and Global Identities are, Respectively, the Percentage of Identity of the Covered Part and of the Whole HPSE Sequence with the PDB Entry. The Percentage of Similarity is Specified in Parenthesis

PDB Entry	Enzymatic Activity	GH Family[a]	Species	Coverage (%)	Local Identity (%)	Global Identity (%)
1B30 (A)[b]	*endo*-1,4-β-xylanase	GH10	*P. simplicissimum*	35	28 (43)	14 (40)
1XYZ (A)	*endo*-1,4-β-xylanase	GH10	*C. thermocellum*	30	22 (41)	12 (45)
1TA3 (B)	*ndo*-1,4-β-xylanase	GH10	*E. nidulans*	19	31 (50)	14 (40)
1TUX (A)	*endo*-1,4-β-xylanase	GH10	*T. aurantiacus*	14	36 (53)	10 (35)
1UMZ (A)	Xyloglucan endotransglycosylase	GH16	*P. tremula*	13	31 (47)	10 (31)
1ECE (A)	Endocellulase E1	GH5	*A. cellulolyticus*	7	34 (53)	12 (41)

[a] As inferred by CAZy (http://www.cazy.org).[49]
[b] The sequences of the PDB entries 1BG4, 1B30, 1B31, 1B3V, 1B3W, 1B3X, 1B3Y, and 1B3Z are strictly identical.

of sulfated sugar residues compatible with GLYCAM06 were previously described.[78] The energy minimization process was divided into four successive stages:

1. Geometry optimization of the hydrogen atoms;
2. Geometry optimization of the amino acid side chains (except amino acids E^{225} and E^{234}) and the O/N-sulfate, carboxyl, and hydroxyl moieties of the sugar residues;
3. Geometry optimization of amino acids not involved in the $(\alpha/\beta)_8$ TIM barrel;
4. Energy minimization of the whole system except the side chain of amino acids E^{225} and E^{234}, and the ring of the monosaccharide units $GlcA^{-1}$ and $GlcNS^{+1}$, which are directly upstream and downstream to the cleavage site;

The nonbonded interactions were calculated every step with no cut-off applied and a dielectric constant of 1. At each step, the force constant of the harmonic restraints was initially 100 kcal mol^{-1} $Å^{-2}$ until convergence, i.e., a change of the potential energy inferior to 1 kcal mol^{-1}. Thereafter, the force constraint on the atom freed at the next step was reduced to 5 kcal mol^{-1} $Å^{-2}$ over 10,000 minimization steps. Visualization and analysis of the minimization results were done with VMD 1.8.7, R 2.10[79] and the ggplot2 package.[80] The electrostatic potential on the model surface was analyzed with APBS 1.2.1.[81] Coordinates of the refined model are available from the authors upon request.

RESULTS AND DISCUSSION

Choice of the Template

Blast searching based on the sequence of mature human heparanase—or its monomeric analog GS3—retrieved six glycosyl hydrolases from the PDB (Table II). Four of them belong to the GH family 10, the best hit being the *endo*-β-D-xylanase from *P. simplicissimum*. For this latter result, the Blast alignment covers roughly the N-terminal third of the heparanase sequence, with a sequence identity of 35% and correct fit of the proton donors between the heparanase and the xylanase sequences.

The use of the full length or the truncated sequence of the catalytic domain ($HPSE_{catalytic}$ and $HPSE_{50\ kDa\ catalytic}$) leads to similar results: the GH10 xylanase from *P. simplicissimum* presents higher sequence similarities. More PDB entries are retrieved, corresponding mainly to other GH10 xylanases or to glucosaminidases and exoglucanases. In contrast, the sequence of the noncatalytic domain ($HPSE_{50\ kDa\ noncatalytic}$) retrieves only the endocellulase E1. It should be noted that the boundaries between the catalytic and noncatalytic domains were settled based on the secondary structures forming the $(\alpha/\beta)_8$ barrel. All the secondary structure predictions performed on the mammalian heparanases (human, bovine, mouse, rat) agree on the presence in the 8 kDa subunit of one conserved α-helix (P^{79}–L^{87} in the human sequence) and four conserved β-strands (D^{37}–F^{43}, S^{58}–D^{62}, Y^{91}–F^{94}, F^{101}–D^{105} in the human sequence). The first strand (D^{37}–F^{43}) belongs to the β-sandwich forming the noncatalytic domain. The catalytic domain starts between the first and the second strand (S^{58}–D^{62}), which initiates the α/β barrel and is followed by the first α-helix (P^{79}–L^{87}). Residue L^{51} was therefore chosen as the first amino acids of the catalytic domain. A similar reasoning was followed to define the end of the catalytic domain at position T^{416}.

When using the protein modeling servers Robetta, PHYRE, and SWISS-MODEL, members of other CAZY families are also retrieved (Table III). When the GS3 construct is used, a GH51 member is identified as the best template: the α-L-arabinofuranosidase from *G. stearothermophilus*. When only the catalytic domain of GS3 is used, PHYRE and SWISS-MODEL select a GH39 member as the best template: the β-D-xylosidase from *T. saccharolyticum*. Both templates possess a $(\alpha/\beta)_8$ barrel catalytic domain coupled to a

Table III Summary of the Results Obtained with the Automatic Molecular Modeling Servers

Server	Query Sequence	Template PDB Id	Enzymatic Activity	GH Family[a]	Sequence Identity (%)	*E*-value
Robetta[44]	GS3	1PZ3	α-L-arabinofuranosidase	GH51	—	—
PHYRE	GS3	1PZ2	α-L-arabinofuranosidase	GH51	12	$1.4e^{-40}$
PHYRE	$GS3_{catalytic}$	1PX8	α-D-xylosidase	GH39	13	$9.8e^{-20}$
Swiss-Model	GS3	1QW9	α-L-arabinofuranosidase	GH51	11	$1.7e^{-42}$
Swiss-Model	$GS3_{catalytic}$	1UHV	α-D-xylosidase	GH39	14	$2.3e^{-23}$

[a] As inferred by CAZy (http://www.cazy.org).[49]

β-sandwich domain of unknown function. The difference arises from their natural substrates. The GH51 arabinofuranosidase cleaves the β1–3 linkage between an arabinofuranose substituent and the xylan main chain. The GH39 xylosidase is an exoglucanase cleaving the β1–4 linkage at the nonreducing end of the xylan main chain.

The use of either a GH51 arabinofuranosidase or a GH39 xylosidase by the automated modeling servers as a template for the modeling of heparanase is questionable. None of these enzymes is an endoglucanase like heparanase. From this point of view, the GH10 *endo*-1,4-β-D-xylanase from *P. simplicissimum* is a better choice, although the substrate is a xylose polysaccharide. This hypothesis is supported by the comparison of the crystal structure of *G. stearothermophilus* arabinofuranosidase and *P. simplicissimum* endoxylanase. There is a good superimposition of the secondary structures forming the $(\alpha/\beta)_8$ barrel (Figure 1A). The pair of catalytic glutamate residues is consequently located almost at the same place in both enzymes. The shape of the binding site is very different, however. In the case of *G. stearothermophilus* arabinofuranosidase, the catalytic dyad is constituted by the amino acids E^{175} and E^{294}. They are both nested in a pocket situated below the end of a large groove on the protein surface (Figure 1B). In the case of *P. simplicissimum* endoxylanase, the catalytic dyad is constituted by the amino acids E^{132} and E^{238}. They are located at the bottom and in the middle of a shallow and narrow groove (Figure 1B). This catalytic pair is therefore directly accessible from the protein surface, in contrast with arabinofuranosidase.

As *P. simplicissimum* endoxylanase and heparanase are both endoglucanases, we hypothesize that the shape of the heparanase-binding site is similar to the endoxylanase one, i.e., a narrow groove located at the top of the $(\alpha/\beta)_8$ barrel with the catalytic site located in the middle. This kind of shape for the binding site is commonly found in endo-acting polysaccharidases, as it allows a random binding of several sugar units in polymeric substrates.[82] Conveniently, *P. simplicissimum* endoxylanase has been extensively studied and several crystal structures of the protein and its complexes with substrates are available.[52,54] More generally, the GH10 endoxylanases are well studied enzymes.[83–86] Since the *P. simplicissimum* xylanase is formed by a single domain, unlike heparanase, we focused the study on the modeling of the heparanase catalytic domain only. Noteworthy, there are GH10 xylanases organized like heparanase, i.e., a catalytic domain with in an $(\alpha/\beta)_8$ barrel folded tightly bound to a noncatalytic domain with a β-sandwich fold. However, the sequence similarity with human heparanase is lower than with the *P. simplicissimum* enzyme.

Structure Completion and Ligand Binding

The structure of the heparanase catalytic domain was therefore built by homology with the endoxylanase from *P. simplicissimum*. The identity and similarity ratios between the query and the model are 20 and 40%, respectively (see Figure 2). The quality of the alignment is the lowest in the loops connecting the secondary structure elements of the $(\alpha/\beta)_8$ barrel, in particular for the segments A^{63}–G^{76} and G^{95}–E^{109} in the 8 kDa subunit, and for the segments R^{302}–D^{314} and Q^{383}–P^{402} in the 50 kDa subunit. Noteworthy, the xylanase counterparts of two such segments are involved in the substrate binding. It can be hypothesized that in the case of heparanase, those loops may also be involved in the substrate binding.

Eight crystal structures of the endoxylanase from *P. simplicissimum* are available in the PDB (1BG4, 1B30, 1B31, 1B3V, 1B3W, 1B3X, 1B3Y, and 1B3Z). The 1BG4 entry corresponds to the free enzyme while the others are complexed with xylose, xylooligosaccharides, or xylose disaccharides and trisaccharides substituted by a β1–2 linked deoxyglucuronic acid—those two latter compounds are xylanase inhibitors. The structures of the free and bound forms of the enzyme are highly similar with an average RMSD below 0.11 Å for the backbone. Their resolution ranges from 1.75 Å (1BG4 and 1B31) to 2.60 Å (1B3W). All those crystal structures were used during the modeling process, especially for building the structurally conserved regions. The ligands present in the 1B3Z structure were retained during the modeling process. They consist of a xylotriose and a xylobiose, respectively, located upstream and downstream to the catalytic pair (Figure 1B) and corresponding to the end product of xylopentose hydrolysis. The loops and the amino acid side chains were modeled to avoid steric

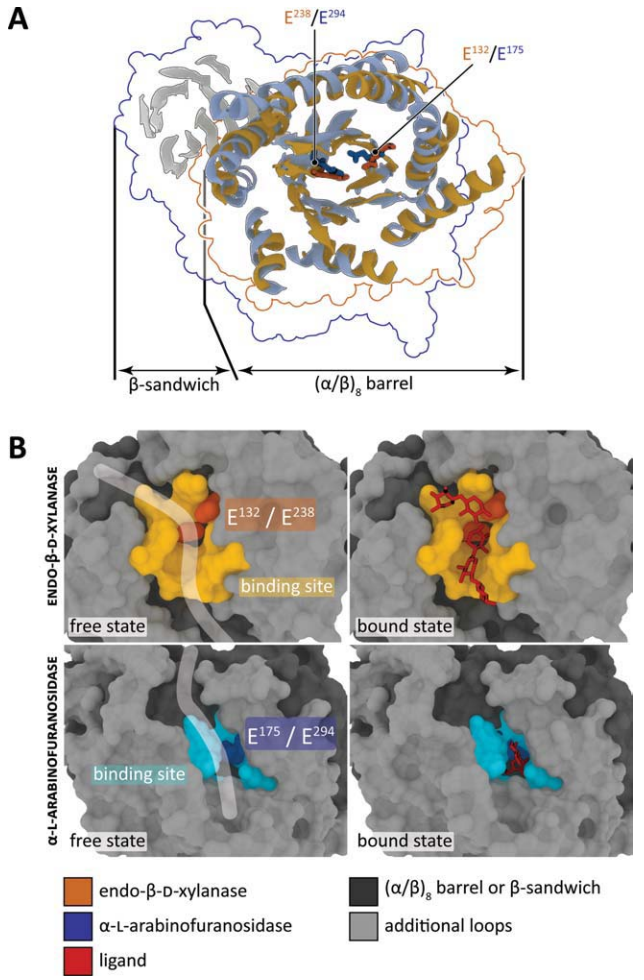

FIGURE 1 Comparison of the crystal structures of the α-L-arabinofuranosidase from *G. stearothermophilus* and the *endo*-1,4-β-D-xylanase from *P. simplicissimum*. A, Superimposition of the arabinofuranosidase (PDB entry 1P3Z) and the endoxylanase (PDB entry 1BG4) structures. The protein surface is outlined for each enzyme. The superimposition was done with DaliLite (www.ebi.ac.uk/Tools/dalilite) using the heavy atoms of the protein backbone. B, top panel—Catalytic site of the endoxylanase in free or in bound state (PDB entries 1BG4 and 1B3Z, respectively). The ligand is a hydrolyzed xylose pentasaccharide. The groove on the protein surface is outlined by a white transparent line. Both states correspond to the wild-type enzyme. B, bottom panel—Catalytic site of the arabinofuranosidase in free or in bound state (PDB entries 1PZ3 and 1QW8, respectively). The ligand is an arabinose-α(1–3)-xylose disaccharide. The groove in the protein surface is outlined by a white transparent line. For the bound state, the enzyme is the inactive mutant E175A.

clashed with ligands. A quality assessment of the model prior to minimization is available in Supporting Information Table IS. A superimposition of the model with the template is available in Supporting Information Figure 2S.

For the model refinement, the two xylooligosaccharides were replaced by two fragments of heparan sulfate: IdoA-(α1–4)-GlcNS-(β1–4)-GlcA and GlcNS-(α1–4)-IdoAS (see Figure 4). The choice of those fragments is inspired by the enzymatic assays realized by Okada et al.[45] They have shown that 2-N and 6-O sulfations of the GlcN directly upstream and downstream to the cleavage site are of primary importance; the upstream GlcNS residue is preferentially preceded by a nonsulfated IdoA, while the downstream GlcNS is preferentially followed by a 2-O sulfated IdoA. The replacement itself consists in extracting the coordinates of the oligosaccharide backbone, i.e., carbons and oxygens from the glycosidic rings and linkages, adding the appropriate substituents to the backbone structure, minimizing the position of those substituents and finally adding the new compound to the yet non-minimized heparanase structure.

General Organization of the Heparanase Structure and Validation

The geometry optimization protocol results in the creation of a narrow groove leading to the catalytic site with the hydrolyzed heparan sulfate pentasaccharide located in its middle (see Figure 3). The truncated 8 kDa subunit is interlocked in the $(\alpha/\beta)_8$ barrel (Figure 3A). This is in agreement with the biochemical data, which demonstrate that the 8 kDa subunit is necessary to the activity of the enzyme. The C-terminus of the 8 kDa subunit is close to the N-terminus of the 50 kDa subunit. Both are located on the same side than the active site, with the GS3 linker roughly pointing toward the groove entrance (Figure 3A). It can be hypothesized that a longer linker, like the 6 kDa peptide present in pro-heparanase, would block the access to the active site. Similarly, the N-terminus of the truncated 8 kDa subunit is close to the C-terminus of the truncated 50 kDa subunit. Both are located on the opposite face of the groove and points toward the same direction. This indicates that the noncatalytic domain is located far from the active site, as observed for the GH10, GH39, and GH51 members with a similar dual domain organization. Noteworthy, the $(\alpha/\beta)_8$ barrel face that would be in contact with the β-sandwich shows an electrostatic potential close to 0 kT e^{-1} (Figure 3B), which is compatible with the interior of a protein.

In spite of numerous experimental studies, data suited to the model validation are extremely scarce, in particular site-directed mutagenesis experiments. Five N-glycosylation sites,[87] one S-cysteinylation and two disulfide bridges[88] have been characterized (Supporting Information Figure 1S). The N-glycosylation sites are located on the model surface (Figure 5A), as expected. The two cysteine residues (C^{179} and C^{211}) contained in the catalytic domain are also exposed on the surface (Figure 5A). C^{179} has been shown by mass spectrometry to form a disulfide bridge with C^{127}, located in the

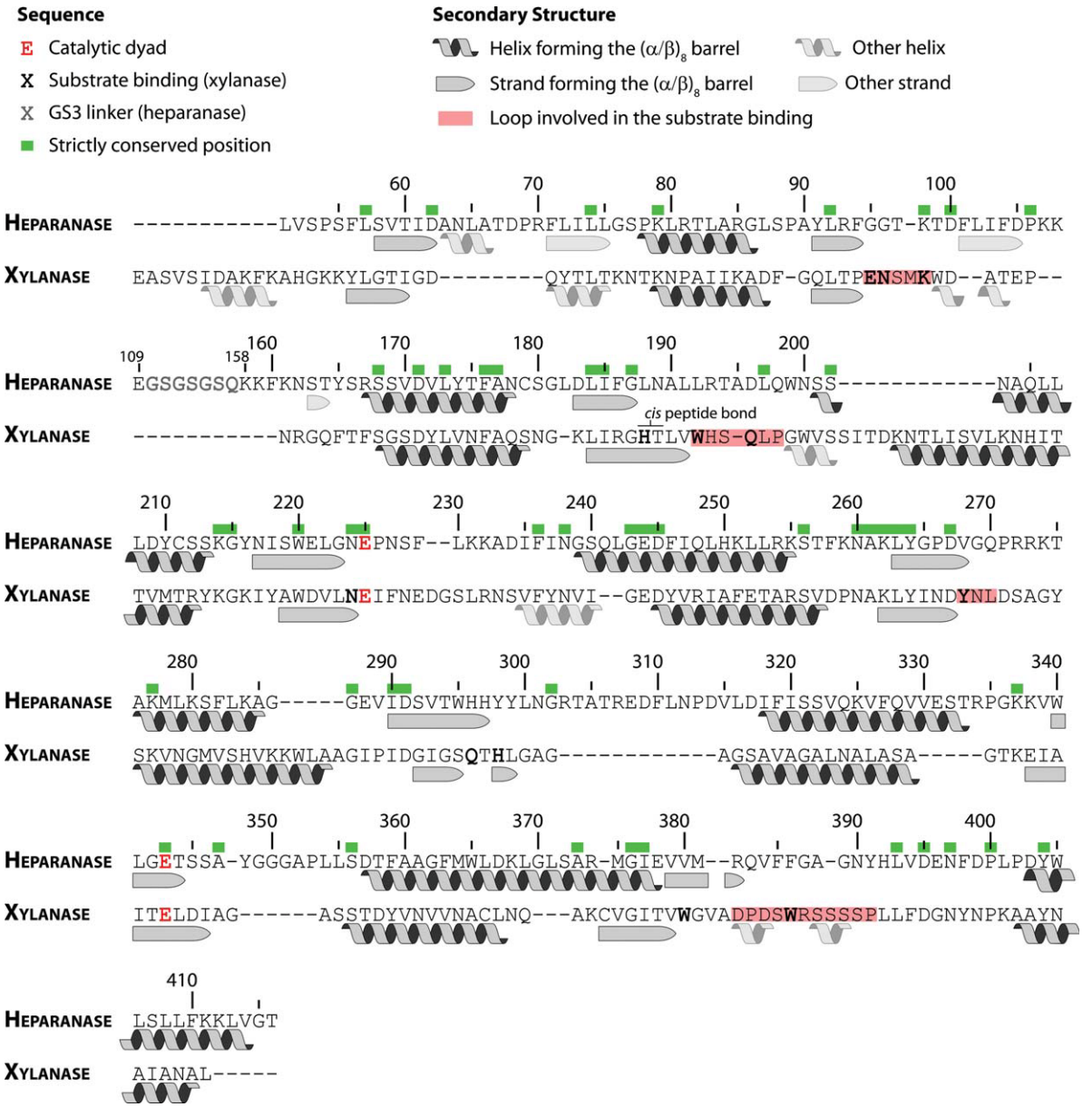

FIGURE 2 Refined T-coffee sequence alignment between human heparanase and *P. simplicissimum* endoxylanase (UniprotKB entries Q9Y251 and P56588, respectively). Only the catalytic domain of the heparanase sequence is considered here. The numbering is the same than for the heparanase UniprotKB entry. Consequently, the GS3 linker is outside of the numbering scheme.

6 kDa linker.[88] The formation of the C^{127}–C^{179} bridge is not necessary to the activation of heparanase or the enzymatic activity,[88] as expected since the 6 kDa linker is removed from the mature form of the enzyme. C^{211} has been shown to be S-cysteinylated.[88] However, its mutation into a serine does affect neither the heparanase activity nor the heparanase secretion. Noteworthy, in our model, both C^{179} and C^{211} are located on the face opposite to the heparanase-binding site and to the potential location of the β-sandwich (Figure 5A). This suggests that none of these cysteine residues play a major role in the ligand binding or the protein folding.

Structural Organization of the Substrate-Binding Site and Validation

The groove leading to the active site is narrow, due to the presence of several hydrophobic residues forming its walls (Figure 3C). Residues F^{385} and F^{386} on one side, and F^{101} and A^{195} on the other side form a lid over the groove entrance.

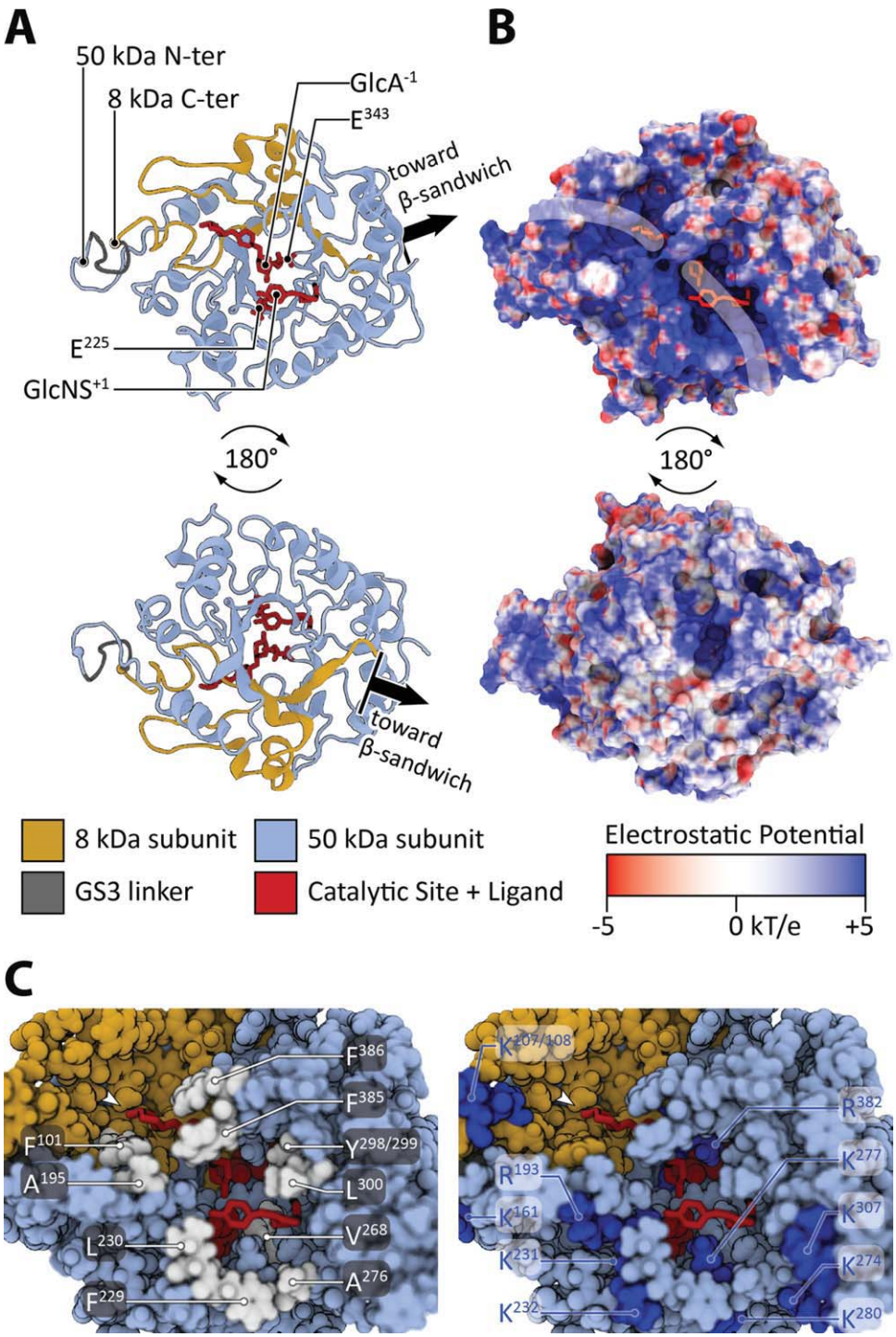

FIGURE 3 General organization of the heparanase molecular model. A, Overview of the heparanase structure bound to the disaccharide and trisaccharide. B, Overview of the electrostatic potential on the surface of heparanase. The groove leading to the catalytic site is highlighted by a white line. C, Overview of the groove forming the binding site. Left panel: hydrophobic amino acids forming the wall of the groove. Right panel: positively charged residues embedded in or at close range from the groove.

Table IV Details of the Heparanase–Oligosaccharide Interactions

Sugar Residue	Polar Interaction with the Backbone Amide	Polar Interaction with a Side Chain	Apolar Interaction with a Side Chain
$IdoA^{-3}$	L^{102}	N^{64}	F^{101}
$GlcNS^{-2}$	A^{195}, D^{196}, Q^{383}, F^{386}	Q^{383}, R^{382}	F^{385}
$GlcA^{-1}$	F^{385}	N^{224}, E^{343}, and R^{382}	—
$GlcNS^{+1}$	F^{229}, L^{230}	E^{225}, K^{231}, and K^{277}	V^{268}, Y^{298}, and L^{300}
$IdoAS^{+2}$	A^{276}	K^{277} and N^{301}	—

Aromatic residues F^{229}, Y^{298}, and Y^{299} and aliphatic residues L^{230}, V^{268}, A^{276}, and L^{300} are also present at the groove exit, near the reducing end of the disaccharide. This abundance of hydrophobic residues is also observed along the endoxylanase-binding site.[52,54]

The groove forming the binding site is highly electropositive because of the numerous positively basic amino acids forming its walls (Figure 3C), such as R^{93}, R^{193}, K^{231}, R^{272}, K^{277}, and R^{382}. More positively charged residues are found in the direct vicinity of the groove: K^{98}, K^{232}, R^{272}, R^{273}, K^{274}, K^{280}, and R^{307} (Figure 3B). To our knowledge, there is only one study attempting to experimentally identify the heparan sulfate-binding site.[89] The segments K^{158}–D^{171} and Q^{270}–K^{280} were found to be directly involved in this process. The deletion of the segments or the double mutation K158A/K161A abolished the ability of heparanase to bind heparan sulfate. In our model, K^{158} and K^{161} are located nearby the entrance of the basic groove, where it forms a basic cluster with K^{159} and with K^{107} and K^{108} from the 8 kDa subunit (Figure 3C, 5A, and Supporting Information Figure 1S). However, this basic cluster is not directly involved in the groove shaping or the interaction with the ligand. In contrast, the position of the Q^{270}–K^{280} segment in the model is in good agreement with the experimental data: K^{277} interacts directly with the ligand while R^{272}, R^{273}, K^{274}, and R^{307} form a large basic cluster at the exit of the groove, by symmetry to the one at the entrance (Figure 3C, 5A, and Supporting Information Figure 1S).

Overall, the structural organization of the binding site is in partial agreement with experiment. The absence of experimental constraints during the geometry optimization of the model is certainly an issue here. It should be noted that three additional sugar units at the nonreducing end of the trisaccharide would have been sufficient to reach the K^{158}–D^{171} segment. In addition, this part of the protein may potentially be a sub-site involved in the substrate recognition, while the Q^{270}–K^{280} segment would be directly involved in the interactions with the sugar units located at the cleavage site. More mutational data would certainly be useful to better assess the role of those two segments.

Protein–Ligand Interaction

The electrostatic interactions between the protein and the ligand seem to be concentrated on three charged residues: K^{231}, K^{277}, and R^{382} (Table IV) that are located within 4 Å range from the catalytic site (Figure 4A):

- K^{231} is bonded to oxygen O^4 from $GlcNS^{+1}$.
- K^{277} forms an electrostatic bridge between the *O*-sulfate from $GlcNS^{+1}$ and the carboxylate from $IdoAS^{+2}$.
- R^{382} forms an electrostatic bridge between the N-sulfate from $GlcNS^{-2}$ and the carboxylate from $GlcA^{-1}$.

To better estimate the importance of these residues, their conservation in the vertebrate heparanase sequences was inspected (Figure 5B, the multiple alignment is provided as Supporting Information). The catalytic pair $E^{225/343}$ is strictly conserved, as expected, as well as K^{231} and R^{382}. Residue K^{277} is conserved in mammalian and amphibian sequences, but substituted by histidine or aspartate in bird and fish sequences. To fulfill the analysis, the conservation in the GH family 79 was also inspected. The sequence alignment of all the GH79 members is available in the Pfam database (Pfam entry PF03662).The catalytic pair $E^{225/343}$ is still strictly conserved. However, K^{231} and K^{277} are not particularly well conserved or substituted by a residue of equivalent property (Figure 5B). It is important to note that GH79 family includes enzyme that act on different substrates and some of them do not hydrolyze heparan sulfate, or even a sulfated polysaccharide. In fact, the residue conservation in GH79 members more probably reflects the conservation of the hydrolysis mechanism rather than the binding of a particular substrate. At last, it should be noted that the Pfam profile only takes into account the segment L^{150}–G^{370} from human heparanase. Hence, the conservation of R^{382} could not be assessed.

In addition to basic residues, several hydrophobic amino acids seem to be important for the interaction with the ligand (Table IV and Figure 4):

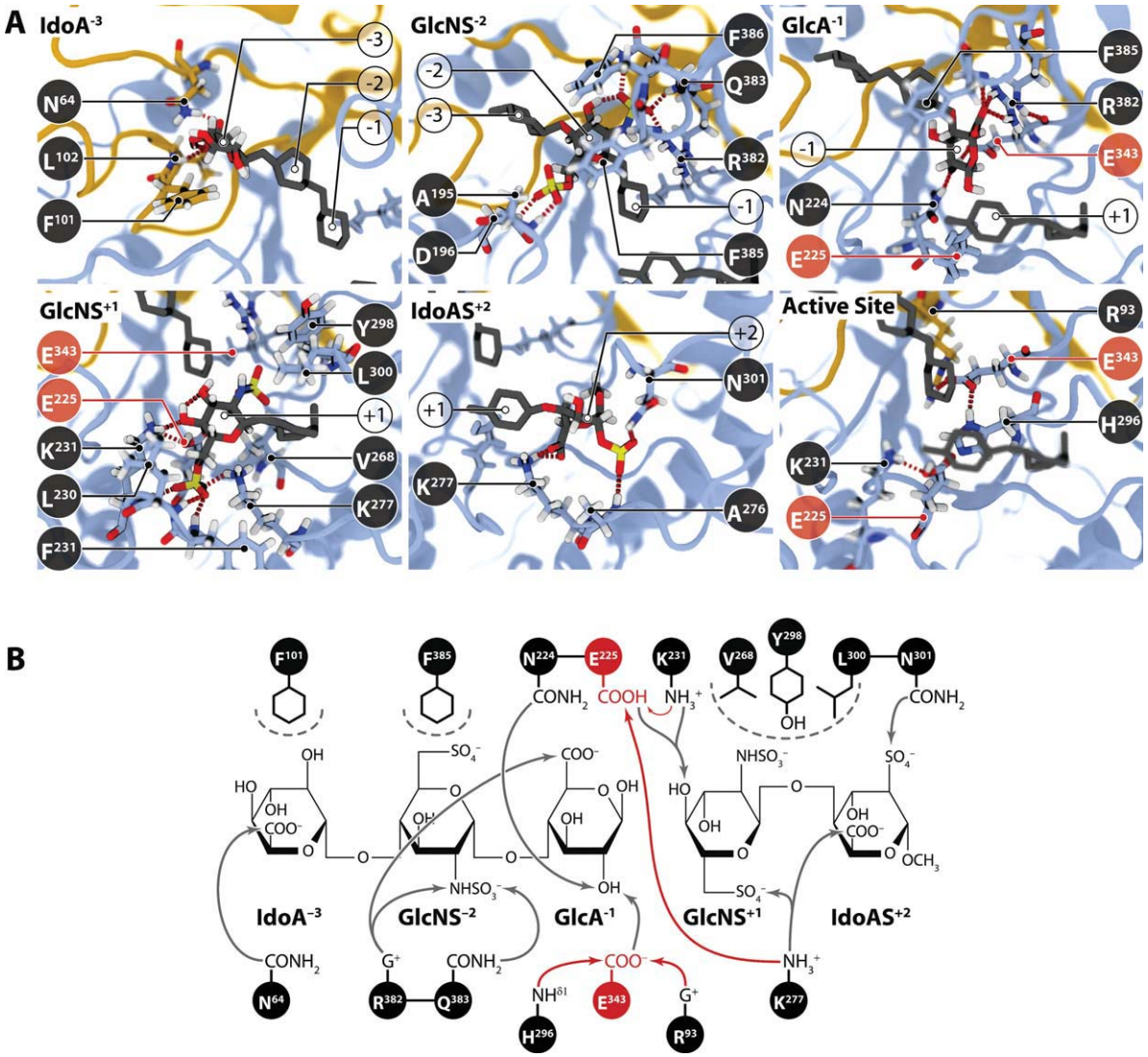

FIGURE 4 Details of the heparanase–oligosaccharides interaction. A, Details of the heparanase-binding site per monosaccharide unit. A detail of the amino acids interaction within the active site is also provided. B, Recapitulative scheme of the main interactions within the binding site. Polar interactions are represented by arrows. Apolar interactions are represented by dashed lines. Interactions with the catalytic site are depicted in red.

- F^{101} and F^{385} are stacked to the rings of $IdoA^{-3}$ and $GlcNS^{-2}$, respectively
- V^{268}, Y^{298}, and L^{300} form a hydrophobic pocket where the N-sulfate of $GlcNS^{+1}$ is located.

F^{101} is strictly conserved in vertebrate heparanases. F^{385} is almost strictly conserved in mammal heparanases and substituted by leucine or isoleucine in fish and frog sequences (Supporting Information and Figure 5B). Of particular interest, F^{385} and F^{386} seem to be the heparanase equivalent of W^{276} found in endoxylanase (Figure 3C). This tryptophan forms a lid over the groove when a substrate is present, but is very probably in an opened configuration in the free state enzyme.[52,54] The amino acids forming the hydrophobic pocket (V^{268}, Y^{298}, and L^{300}) are also conserved or substituted by amino acids of equivalent size and polarity in vertebrate heparanases and GH79 members. In fact, Y^{298} is almost strictly conserved in GH79 sequences. That probably indicates its importance in the enzymatic reaction specific to this family of hydrolases.

Overall, the number of amino acids important for ligand binding seems to be limited, in contrast with the length of the groove and the ligand size. Most of the identified residues

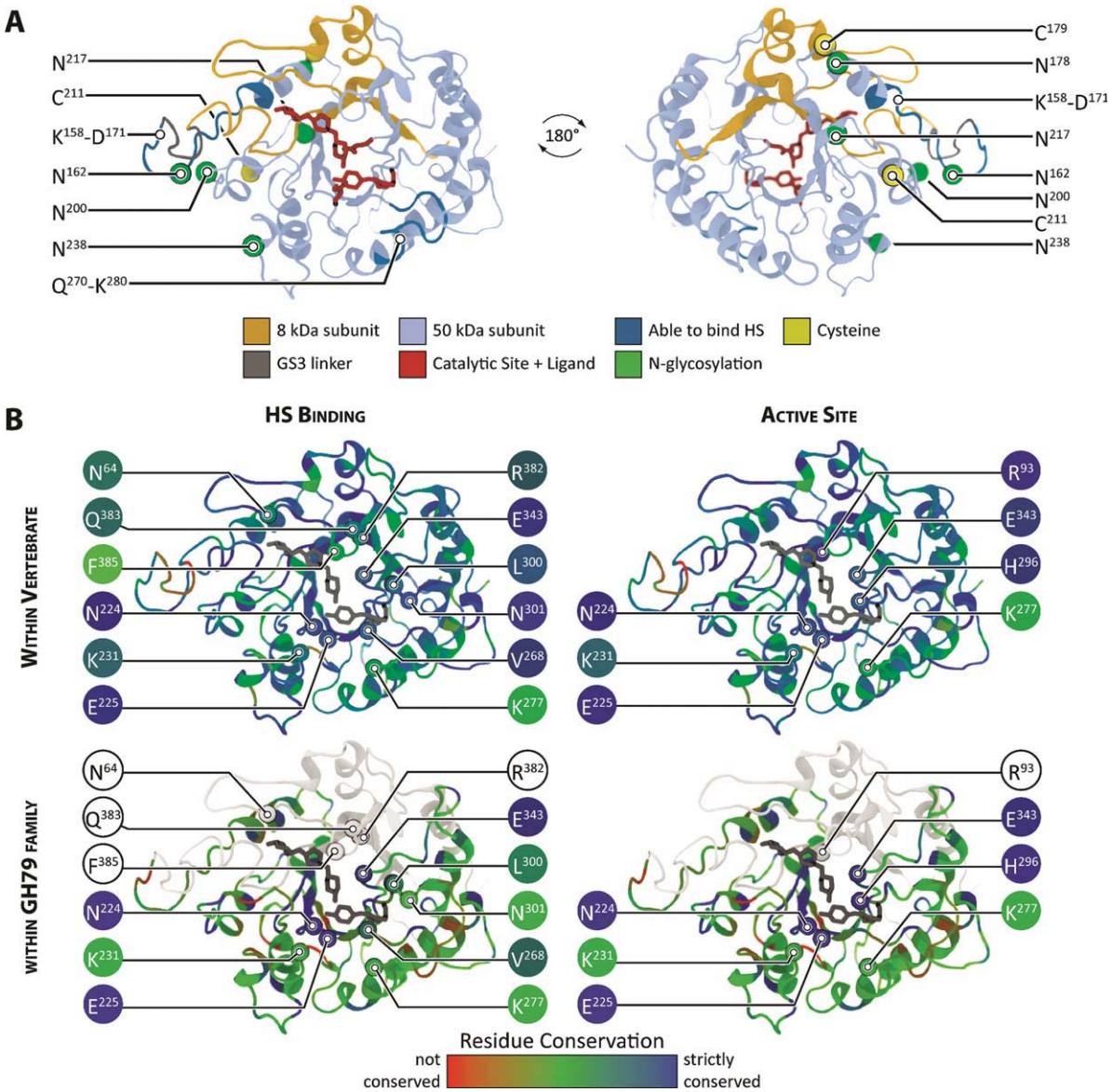

FIGURE 5 A, Depiction of the experimentally characterized parts of heparanase on the model. B, Amino acid conservation within vertebrate heparanase (top) or within the GH family 79 (bottom). Amino acids are colored as a function of the sequence similarity within the multiple sequence alignment, based on the BLOSUM60 matrix. The sequence parts not present in the GH79 alignment are depicted in white. It should be noted that Y^{298} is hidden by L^{300}. Y^{298} is conserved in vertebrate heparanases and GH family 79.

interact with sugar units directly upstream and downstream to the cleavage site. This is in agreement with the attempts to identify the structural feature of the natural heparanase ligand.[45,46] In this latter study, the carboxyl group of $GlcA^{-1}$, the O-sulfation of $GlcNS^{-2}$ and the N-sulfation of $GlcNS^{+1}$ have been shown to be of the uttermost importance for the substrate recognition. In our model, the carboxyl group of $GlcA^{-1}$ is bonded to the guanidyl group of R^{382} and the backbone amide of F^{385}. The epimerization of $GlcA^{-1}$ into $IdoA^{-1}$ would bring the carboxyl group within range of the nucleophile E^{343}. This would explain why IdoA is not favored at this particular position.[45,46,90,91] Noteworthy, R^{382} is also bounded to the N-sulfate of $GlcNS^{-2}$, which seems to have a promoting effect on the enzymatic activity.[45] The other important groups for the substrate recognition would be the 6-*O*-sulfate of $GlcNS^{-2}$, the 2-*N*-sulfate of $GlcNS^{+1}$, and probably the 2-*O*-sulfate of $IdoAS^{+2}$.[45,46] None of these groups interacts with the side chain of a charged amino acid. In fact, the 2-*N*-sulfate

of GlcNS^{+1} is event located in the hydrophobic pocket constituted by V^{268}, Y^{298}, and L^{300}. This very probably means that our model lacks of accuracy for a part of the binding site, due to the lack of constraint introduce during the minimization stage. Noteworthy, the carboxylate of IdoAS^{+2} and the 6-*O*-sulfate of GlcNS^{+1} are bridged by K^{277}. This is consistent with the fact that the 6-O-sulfation of GlcNS^{+1} promotes the substrate recognition.[45] Finally, it is worth to mention that the 3-O-sulfation of GlcNS^{+1} inhibits the enzyme. This is well consistent with our model, since a sulfate located at that place would be in front of the proton donor E^{225}.

Insights into Heparanase Catalytic Mechanism

The catalytic mechanism operating in heparanase is still elusive. So far, only the nucleophile and the proton donor have been identified.[43] In the present model, the nucleophile E^{343} interacts with the hydroxyl at position 2 of GlcA^{-1}, while the proton donor E^{225} interacts with the hydroxyl at position 4 of GlcNS^{+1}. In addition, E^{225} interacts with K^{231} and E^{343} interacts with R^{93} and H^{296} (see Figure 4).

In the xylanase template, the proton donor E^{132} is in proximity to N$^{\varepsilon 1}$ of W^{88} and the guanidyl group of R^{141}. Both are well conserved in GH10 endoxylanases. Analysis of the pH dependence of mutants suggested that R^{141} residue increases the pK_a of the protein donor glutamate in *Bacillus halodurans* and *Clostridium stercorarium*.[84] Due to the flexibility of lysine, both K^{231} and K^{232} could play the role of R^{141} in heparanase and K^{231} has been selected in the final model (see Figure 4). The effect of K^{231} on the pK_a of E^{225} is probably less important than R^{141} in xylanase, since the pK_a of lysine is lower than the arginine one. This could be compensated by the electrostatic environment of the active site, which is much electropositive in heparanase. Noticeably, N^{224}, K^{277}, H^{296} are within 3.5–4.0 Å from E^{225} and may influence its pK_a as well. Noteworthy, Both K^{231} and K^{232} are conserved in vertebrate heparanases, but not in GH family 79.

In the template, the nucleophile E^{238} is within 3 Å from H^{210}. This histidine is strictly conserved among the GH10 endoxylanases. NMR studies of the xylanase from *Cellulomonas fimi* have shown that this histidine pK_a is higher than 10.4 and that it is very likely protonated in the free enzyme at optimal pH.[92] Upon formation of noncovalent enzyme–inhibitor complexes or a long-lived covalent glycosyl–enzyme intermediate, the histidine remains positively charged.[92] This charged histidine tightly bridges the nucleophile glutamate and a strictly conserved aspartate residue (D^{240} in *P. simplicissimum* xylanase). The presence of these two acidic groups very probably participates to the protonation of histidine. In the case of heparanase, the residue equivalent to H^{210} would be H^{296} (see Figure 2). It is bonded to E^{343} in the model and is within 3.5 Å from N^{224}, but not to another acidic residue (Figure 4A). E^{343} also interacts with R^{93} and is within 4 Å from R^{382} (Figure 4A). It should be noted that H^{296} is protonated at position N$^{\delta 1}$ in the model, as a function of the favorable hydrogen bonds created in the structure before minimization. Noteworthy too, H^{296} and R^{93} are conserved in vertebrate heparanases, and H^{296} is also strictly conserved in the GH79 members. This supports the hypothesis that this histidine is directly involved in the heparanase enzymatic reaction rather than the substrate binding. In fact, the all segment H^{295}–Y^{298} seems to be conserved in GH family 79, H^{296} and Y^{298} being the most conserved position and H^{295} and Y^{297} being substituted respectively by arginine and isoleucine.

CONCLUSION

The molecular model of heparanase presented in this study is consistent with the structural features expected to be carried out by this enzyme: an $(\alpha/\beta)_8$ barrel fold of the catalytic domain, a basic binding site with a groove shape and with the catalytic site located in its middle, N-glycosylation and S-cysteinylation sites exposed to the proteins surface. However, the absence of experimental constraints hinders the accuracy of the model. Therefore, our model is mostly, but not totally in agreement with the experimental attempts to characterize the binding site or the natural substrate of heparanase.

The presence of a ligand in the template crystal structure allowed us to dock a heparan sulfate oligosaccharide to the heparanase active site. The modeled complex emphasized the potential role of three basic (K^{231}, K^{277}, and R^{382}) and two aromatic (F^{101} and F^{385}) amino acids in the substrate binding. We also identified amino acids potentially important for the enzymatic mechanism, like H^{296} and Y^{298}, in addition to the already characterized proton donor and nucleophile. The model proposed in this study is expected to be helpful to the molecular design of new heparanase inhibitors and anti-metastasis drugs.

The present study was funded by the "Pôle de Compétitivité MEDICEN" as part of the DOSCA project.

REFERENCES

1. Casu, B.; Lindahl, U. In Adv Carbohydr Chem Biochem; Academic Press: New York, 2001, pp 159–206.
2. Lindahl, U.; Li, J.; Kwang, W. J. In International Review of Cell and Molecular Biology; Academic Press: New York, 2009, p 105–159.

3. Stringer, S. E. Biochem Soc Trans 2006, 34, 451–453.
4. Stauffer, M. E.; Skelton, N. J.; Fairbrothe, W. J. J Biomol NMR 2002, 23, 57–61.
5. Raman, R.; Venkataraman, G.; Ernst, S.; Sasisekharan, V.; Sasisekharan, R. Proc Natl Acad Sci U S A 2003, 100, 2357–2362.
6. Schlessinger, J.; Plotnikov, A. N.; Ibrahimi, O. A.; Eliseenkova, A. V.; Yeh, B. K.; Yayon, A.; Linhardt, R. J.; Mohammadi, M. Mol Cell 2000, 6, 743–750.
7. Schultz, G. S.; Wysocki, A. Wound Repair Regen 2009, 17, 153–162.
8. Grünewald, F. S.; Prota, A. E.; Giese, A.; Ballmer-Hofer, K. Biochim Biophys Acta (BBA)-Proteins Proteom, 1804, 567–580.
9. Iozzo, R.; Zoeller, J.; Nyström, A. Molecules and Cells 2009, 27, 503–513.
10. Gorsi, B.; Stringer, S. E. Trends Cell Biol 2007, 17, 173–177.
11. Esko, J. D.; Selleck, S. B. Annu Rev Biochem 2002, 71, 435–471.
12. Aikawa, J.-i.; Grobe, K.; Tsujimoto, M.; Esko, J. D. J Biol Chem 2001, 276, 5876–5882.
13. Hagner-McWhirter, A.; Li, J. P.; Oscarson, S.; Lindahl, U. J Biol Chem 2004, 279, 14631–14638.
14. Hagner-Mcwhirter, A.; Lindahl, U.; Li, J. Biochem J 2000, 347 Pt 1, 69–75.
15. Carlsson, P.; Presto, J.; Spillmann, D.; Lindahl, U.; Kjellén, L. J Biol Chem 2008, 283, 20008–20014.
16. Rong, J.; Habuchi, H.; Kimata, K.; Lindahl, U.; Kusche-Gullberg, M. Biochemistry (Mosc) 2001, 40, 5548–5555.
17. Pinhal, M. A. S.; Smith, B.; Olson, S.; Aikawa, J. -i.; Kimata, K.; Esko, J. D. Proc Natl Acad Sci U S A 2001, 98, 12984–12989.
18. Habuchi, H.; Miyake, G.; Nogami, K.; Kuroiwa, A.; Matsuda, Y.; Kusche-Gullberg, M.; Habuchi, O.; Tanaka, M.; Kimata, K. Biochem J 2003, 371, 131–142.
19. Esko, J. D.; Lindahl, U. J Clin Invest 2001, 108, 169–173.
20. Murphy, K. J.; Merry, C. L. R.; Lyon, M.; Thompson, J. E.; Roberts, I. S.; Gallagher, J. T. J Biol Chem 2004, 279, 27239–27245.
21. Morimoto-Tomita, M.; Uchimura, K.; Werb, Z.; Hemmerich, S.; Rosen, S. D. J Biol Chem 2002, 277, 49175–49185.
22. Vlodavsky, I.; Friedmann, Y.; Elkin, M.; Aingorn, H.; Atzmon, R.; Ishai-Michaeli, R.; Bitan, M.; Pappo, O.; Peretz, T.; Michal, I.; Spector, L.; Pecker, I. Nat Med 1999, 5, 793–802.
23. Kussie, P. H.; Hulmes, J. D.; Ludwig, D. L.; Patel, S.; Navarro, E. C.; Seddon, A. P.; Giorgio, N. A.; Bohlen, P. Biochem Biophys Res Commun 1999, 261, 183–187.
24. Toyoshima, M.; Nakajima, M. J Biol Chem 1999, 274, 24153–24160.
25. Gong, F.; Jemth, P.; Galvis, M. L. E.; Vlodavsky, I.; Horner, A.; Lindahl, U.; Li, J. -p J Biol Chem 2003, 278, 35152–35158.
26. Vreys, V.; David, G. J Cell Mol Med 2007, 11, 427–452.
27. Rosen, S. D.; Lemjabbar-Alaoui, H. Expert Opin Therap Targets, 14, 935–949.
28. Barash, U.; Cohen-Kaplan, V.; Dowek, I.; Sanderson, R. D.; Ilan, N.; Vlodavsky, I. FEBS J, 277, 3890–3903.
29. McKenzie, E. A. Br J Pharmacol 2007, 151, 1–14.
30. Ilan, N.; Elkin, M.; Vlodavsky, I. Int J Biochem Cell Biol 2006, 38, 2018–2039.
31. Nadir, Y.; Brenner, B. Best Pract Res Clin Haematol 2009, 22, 85–92.
32. Parish, C. R.; Freeman, C.; Brown, K. J.; Francis, D. J.; Cowden, W. B. Cancer Res 1999, 59, 3433–3441.
33. Dredge, K.; Hammond, E.; Handley, P.; Gonda, T. J.; Smith, M. T.; Vincent, C.; Brandt, R.; Ferro, V.; Bytheway, I. Br J Cancer 2011, 104, 635–642.
34. Courtney, S. M.; Hay, P. A.; Buck, R. T.; Colville, C. S.; Porter, D. W.; Scopes, D. I.; Pollard, F. C.; Page, M. J.; Bennett, J. M.; Hircock, M. L.; McKenzie, E. A.; Stubberfield, C. R.; Turner, P. R. Bioorg Med Chem Lett 2004, 14, 3269–3273.
35. 5Courtney, S. M.; Hay, P. A.; Buck, R. T.; Colville, C. S.; Phillips, D. J.; Scopes, D. I.; Pollard, F. C.; Page, M. J.; Bennett, J. M.; Hircock, M. L.; McKenzie, E. A.; Bhaman, M.; Felix, R.; Stubberfield, C. R.; Turner, P. R. Bioorg Med Chem Lett 2005, 15, 2295–2299.
36. Hulett, M. D.; Freeman, C.; Hamdorf, B. J.; Baker, R. T.; Harris, M. J.; Parish, C. R. Nat Med 1999, 5, 803–809.
37. Zetser, A.; Levy-Adam, F.; Kaplan, V.; Gingis-Velitski, S.; Bashenko, Y.; Schubert, S.; Flugelman, M. Y.; Vlodavsky, I.; Ilan, N. J Cell Sci 2004, 117, 2249–2258.
38. Abboud-Jarrous, G.; Atzmon, R.; Peretz, T.; Palermo, C.; Gadea, B. B.; Joyce, J. A.; Vlodavsky, I. J Biol Chem 2008, 283, 18167–18176.
39. Levy-Adam, F.; Miao, H.-Q.; Heinrikson, R. L.; Vlodavsky, I.; Ilan, N. Biochem Biophys Res Commun 2003, 308, 885–891.
40. Nardella, C.; Lahm, A.; Pallaoro, M.; Brunetti, M.; Vannini, A.; Steinkühler, C. Biochemistry (Mosc) 2004, 43, 1862–1873.
41. Fairbanks, M. B.; Mildner, A. M.; Leone, J. W.; Cavey, G. S.; Mathews, W. R.; Drong, R. F.; Slightom, J. L.; Bienkowski, M. J.; Smith, C. W.; Bannow, C. A.; Heinrikson, R. L. J Biol Chem 1999, 274, 29587–29590.
42. McKenzie, E.; Young, K.; Hircock, M.; Bennett, J.; Bhaman, M.; Felix, R.; Turner, P.; Stamps, A.; McMillan, D.; Saville, G.; Ng, S.; Mason, S.; Snell, D.; Schofield, D.; Gong, H.; Townsend, R.; Gallagher, J.; Page, M.; Parekh, R.; Stubberfield, C. Biochem J 2003, 373, 423–435.
43. Hulett, M. D.; Hornby, J. R.; Ohms, S. J.; Zuegg, J.; Freeman, C.; Gready, J. E.; Parish, C. R. Biochemistry (Mosc) 2000, 39, 15659–15667.
44. Fux, L.; Feibish, N.; Cohen-Kaplan, V.; Gingis-Velitski, S.; Feld, S.; Geffen, C.; Vlodavsky, I.; Ilan, N. Cancer Res 2009, 69, 1758–1767.
45. Okada, Y.; Yamada, S.; Toyoshima, M.; Dong, J.; Nakajima, M.; Sugahara, K. J Biol Chem 2002, 277, 42488–42495.
46. Pikas, D. S.; Li, J. -p.; Vlodavsky, I.; Lindahl, U. J Biol Chem 1998, 273, 18770–18777.
47. Freeman, C.; Parish, C. R. Biochem J 1998, 330, 1341–1350.
48. Podyma-Inoue, K. A.; Yokote, H.; Sakaguchi, K.; Ikuta, M.; Yanagishita, M. J Biol Chem 2002, 277, 32459–32465.
49. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. Nucleic Acids Res 2009, 37, D233–D238.
50. Zhou, Z.; Bates, M.; Madura, J. D. Proteins 2006, 65, 580–592.
51. Ishida, K.; Hirai, G.; Murakami, K.; Teruya, T.; Simizu, S.; Sodeoka, M.; Osada, H. Mol Cancer Ther 2004, 3, 1069–1077.
52. Schmidt, A.; Kratky, C.; Schlacher, A.; Schwab, H.; Steiner, W. Protein Sci 1998, 7, 2081–2088.
53. Hovel, K.; Shallom, D.; Niefind, K.; Belakhov, V.; Shoham, G.; Baasov, T.; Shoham, Y.; Schomburg, D. EMBO J 2003, 22: 4922–4932.
54. Schmidt, A.; Gubitz, G. M.; Kratky, C. Biochemistry (Mosc) 1999, 38, 2403–2412.

55. Consortium, T. U. Nucleic Acids Res 2010, 38, D142–D148.
56. Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Nucleic Acids Res 1997, 25, 3389–3402.
57. Armougom, F.; Moretti, S.; Poirot, O.; Audic, S.; Dumas, P.; Schaeli, B.; Keduas, V.; Notredame, C. Nucleic Acids Res 2006, 34, W604–608.
58. Notredame, C.; Higgins, D. G.; Heringa, J. J Mol Biol 2000, 302, 205–217.
59. Cole, C.; Barber, J. D.; Barton, G. J. Nucleic Acids Res 2008, 36, W197–201.
60. Callebaut, I.; Labesse, G.; Durand, P.; Poupon, A.; Canard, L.; Chomilier, J.; Henrissat, B.; Mornon, J. P. Cell Mol Life Sci 1997, 53, 621–645.
61. Gaboriaud, C.; Bissery, V.; Benchetrit, T.; Mornon, J. P. FEBS Lett 1987, 224, 149–155.
62. Lemesle-Varloot, L.; Henrissat, B.; Gaboriaud, C.; Bissery, V.; Morgat, A.; Mornon, J. P. Biochimie 1990, 72, 555–574.
63. Woodcock, S.; Mornon, J.-P.; Henrissat, B. Protein Eng 1992, 5, 629–635.
64. Roberts, E.; Eargle, J.; Wright, D.; Luthey-Schulten, Z. BMC Bioinf 2006, 7, 382.
65. Humphrey, W.; Dalke, A.; Schulten, K. J Mol Graph 1996, 14, 33–38, 27–38.
66. Finn, R. D.; Mistry, J.; Tate, J.; Coggill, P.; Heger, A.; Pollington, J. E.; Gavin, O. L.; Gunasekaran, P.; Ceric, G.; Forslund, K.; Holm, L.; Sonnhammer, E. L. L.; Eddy, S. R.; Bateman, A. Nucleic Acids Res 2010, 38, D211–D222.
67. Levy-Adam, F.; Feld, S.; Cohen-Kaplan, V.; Shteingauz, A.; Gross, M.; Arvatz, G.; Naroditsky, I.; Ilan, N.; Doweck, I.; Vlodavsky, I. J Biol Chem 2010, 285, 28010–28019.
68. Montalvão, R. W.; Smith, R. E.; Lovell, S. C.; Blundell, T. L. Bioinformatics 2005, 21, 3719–3725.
69. Deane, C. M.; Blundell, T. L. Protein Sci 2001, 10, 599–612.
70. Deane, C. M.; Blundell, T. L. Proteins: Struct Funct Bioinf 2000, 40, 135–144.
71. Smith, R. E.; Lovell, S. C.; Burke, D. F.; Montalvao, R. W.; Blundell, T. L. Bioinformatics 2007, 23, 1099–1105.
72. Kelley, L. A.; Sternberg, M. J. E. Nat Protoc 2009, 4, 363–371.
73. Schwede, T.; Kopp, J. r.; Guex, N.; Peitsch, M. C. Nucleic Acids Res 2003, 31, 3381–3385.
74. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. J Comput Chem 2005, 26, 1781–1802.
75. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. J Comput Chem 2005, 26, 1668–1688.
76. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Proteins: Struct Funct Bioinf 2006, 65, 712–725.
77. Sapay, N.; Cabannes, E.; Petitou, M.; Imberty, A. Glycobiology 2011, doi: 10.1093/glycob/cwr052.
78. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. J Comput Chem 2008, 29, 622–655.
79. R-development-core-team.R: A language and environment for statistical computing: Vienna, 2005.
80. Wickham, H. ggplot2: Elegant Graphics for Data Analysis; Springer: New York, 2009.
81. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Proc Natl Acad Sci U S A 2001, 98, 10037–10041.
82. Davies, G.; Henrissat, B. Structure 1995, 3, 853–859.
83. Czjzek, M.; Bravman, T.; Henrissat, B.; Shoham, Y. Acta Crystallogr Sect D Biol Crystallogr 2004, 60, 583–585.
84. Lo Leggio, L.; Kalogiannis, S.; Eckert, K.; Teixeira, S. C.; Bhat, M. K.; Andrei, C.; Pickersgill, R. W.; Larsen, S. FEBS Lett 2001, 509, 303–308.
85. Nishimoto, M.; Kitaoka, M.; Fushinobu, S.; Hayashi, K. Biosci, Biotechnol, Biochem 2005, 69: 904–910.
86. Xie, H.; Flint, J.; Vardakou, M.; Lakey, J. H.; Lewis, R. J.; Gilbert, H. J.; Dumon, C. J Mol Biol 2006, 360, 157–167.
87. Simizu, S.; Ishida, K.; Wierzba, M. K.; Osada, H. J Biol Chem 2004, 279, 2697–2703.
88. Simizu, S.; Suzuki, T.; Muroi, M.; Lai, N. S.; Takagi, S.; Dohmae, N.; Osada, H. Cancer Res 2007, 67, 7841–7849.
89. Levy-Adam, F.; Abboud-Jarrous, G.; Guerrini, M.; Beccati, D.; Vlodavsky, I.; Ilan, N. J Biol Chem 2005, 280, 20457–20466.
90. Oldberg, A.; Heldin, C. H.; Wasteson, A.; Busch, C.; Hook, M. Biochemistry (Mosc) 1980, 19, 5755–5762.
91. Ogren, S.; Lindahl, U. J Biol Chem 1975, 250, 2690–2697.
92. Schubert, M.; Poon, D. K. Y.; Wicki, J.; Tarling, C. A.; Kwan, E. M.; Nielsen, J. E.; Withers, S. G.; McIntosh, L. P. Biochemistry (Mosc) 2007, 46, 7383–7395.

Reviewing Editor: J. Andrew McCammon