The EMBO Journal Vol. 20 No. 12 pp. 3008–3017, 2001

Structure of Golgi α-mannosidase II: a target for inhibition of growth and metastasis of cancer cells

Jean M.H.van den Elsen[1], Douglas A.Kuntz and David R.Rose[2]

Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, Ontario, Canada M5G 2M9

[1]Present address: Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK

[2]Corresponding author
e-mail: drose@uhnres.utoronto.ca

J.M.H.van den Elsen and D.A.Kuntz contributed equally to this work

Golgi α-mannosidase II, a key enzyme in *N*-glycan processing, is a target in the development of anti-cancer therapies. The crystal structure of *Drosophila* Golgi α-mannosidase II in the absence and presence of the anti-cancer agent swainsonine and the inhibitor deoxymannojirimycin reveals a novel protein fold with an active site zinc intricately involved both in the substrate specificity of the enzyme and directly in the catalytic mechanism. Identification of a putative GlcNAc binding pocket in the vicinity of the active site cavity provides a model for the binding of the $GlcNAcMan_5GlcNAc_2$ substrate and the consecutive hydrolysis of the α1,6- and α1,3-linked mannose residues. The enzyme–inhibitor interactions observed provide insight into the catalytic mechanism, opening the door to the design of novel inhibitors of α-mannosidase II.

Keywords: cancer therapy/drug design/
Golgi α-mannosidase II/*N*-glycosylation pathway/
X-ray crystallography

Introduction

In breast, colon and skin cancers, the unusual quantitative distributions of complex carbohydrate structures on the cell surface are associated with disease progression, metastasis and poor clinical outcome (Goss *et al*., 1995; Dennis *et al*., 1999a,b). This altered distribution is associated with abnormalities in the *N*-glycosylation pathway, and inhibition of key enzymes in this pathway has shown clinical potential in cancer treatment. Golgi α-mannosidase II (GMII; mannosyl oligosaccharide 1,3-1,6-α-mannosidase II; EC 3.2.1.114) plays a central role in the *N*-glycosylation pathway. In preliminary clinical trials with late-stage cancer patients, specific inhibition of GMII by the non-toxic, orally available compound swainsonine resulted in reduced tumor growth and metastasis, with minimal side-effects (Dennis and Laferte, 1985; Kiyohara *et al*., 1987; Goss *et al*., 1995, 1997). Swainsonine [(1*S*,2*R*,8*R*,8a*R*)-trihydroxy-indolizidine] is an indolizidine alkaloid naturally occurring in a number of Australian and North American plants. It interferes with the glycosylation pathway by binding to GMII with inhibition constants in the range of 20–50 nM, and leads to the increased accumulation of hybrid-type glycan structures, in relation to the complex structures.

The N-linked glycosylation pathway involves the assembly and maturation of sugar complexes attached to asparagine side chains of target proteins at a characteristic Asn-X-Ser/Thr sequence. In eukaryotic cells, this pathway is responsible for proper processing of proteins as they are synthesized in the endoplasmic reticulum (ER) and Golgi apparatus. It follows a well-defined series of steps and involves a considerable number of enzymes responsible for both the extension (glycosyl transferases) and trimming (glycosyl hydrolases) at various steps of the assembly process (recently reviewed by Herscovics, 1999). The pathway begins with the assembly and transfer of a $Glc_3Man_9GlcNAc_2$ (where Glc is glucose, Man is mannose and GlcNAc is *N*-acetyl-glucosamine) oligosaccharide precursor onto a newly formed polypeptide chain. Initial processing involves a series of trimming events by ER-resident glycosidases. As the pathway progresses through the Golgi, the trimming steps are followed by the action of a number of transferases and mannosidases, resulting in the mature glycosylation structure on the nascent protein destined for the cell surface, the lysosomes or secretion.

GMII belongs to glycosyl hydrolase family 38 (Henrissat, 1991; Coutinho and Henrissat, 1999) and is central to the Golgi processing pathway, as it specifically trims two mannose residues from the branched $GlcNAcMan_5GlcNAc_2$ mannose intermediate (Figure 1A) to form the core $GlcNAcMan_3GlcNAc_2$ glycosyl structure, an essential precursor for the further addition of *N*-acetyl-glucosamine units. GMII is a Type II transmembrane protein, ~125 kDa in size, composed of a short N-terminal cytoplasmic tail, a single-span transmembrane domain connected by a stalk segment to a large luminal C-terminal catalytic portion (Moremen and Touster, 1985, 1986). The enzyme is highly specific for the presence of the single GlcNAc attached in a β1,2 linkage to the Man α1,3-Man arm of the $GlcNAcMan_5GlcNAc_2$-Asn-X substrate (Harpaz and Schachter, 1980). It removes the di-mannose branch (M6, M7; Figure 1A) by hydrolysis of both glycosidic bonds with net retention of sugar anomeric configuration, resulting in the final tri-mannose $GlcNAcMan_3GlcNAc_2$ core. There is little or no experimental evidence to date addressing whether the two bonds are cleaved in separate binding sites or sequentially in the same binding site, nor whether or not the singly hydrolyzed product is released from the enzyme between the two cleavage events.

Understanding of the three-dimensional structure of human GMII and the atomic basis of its catalytic

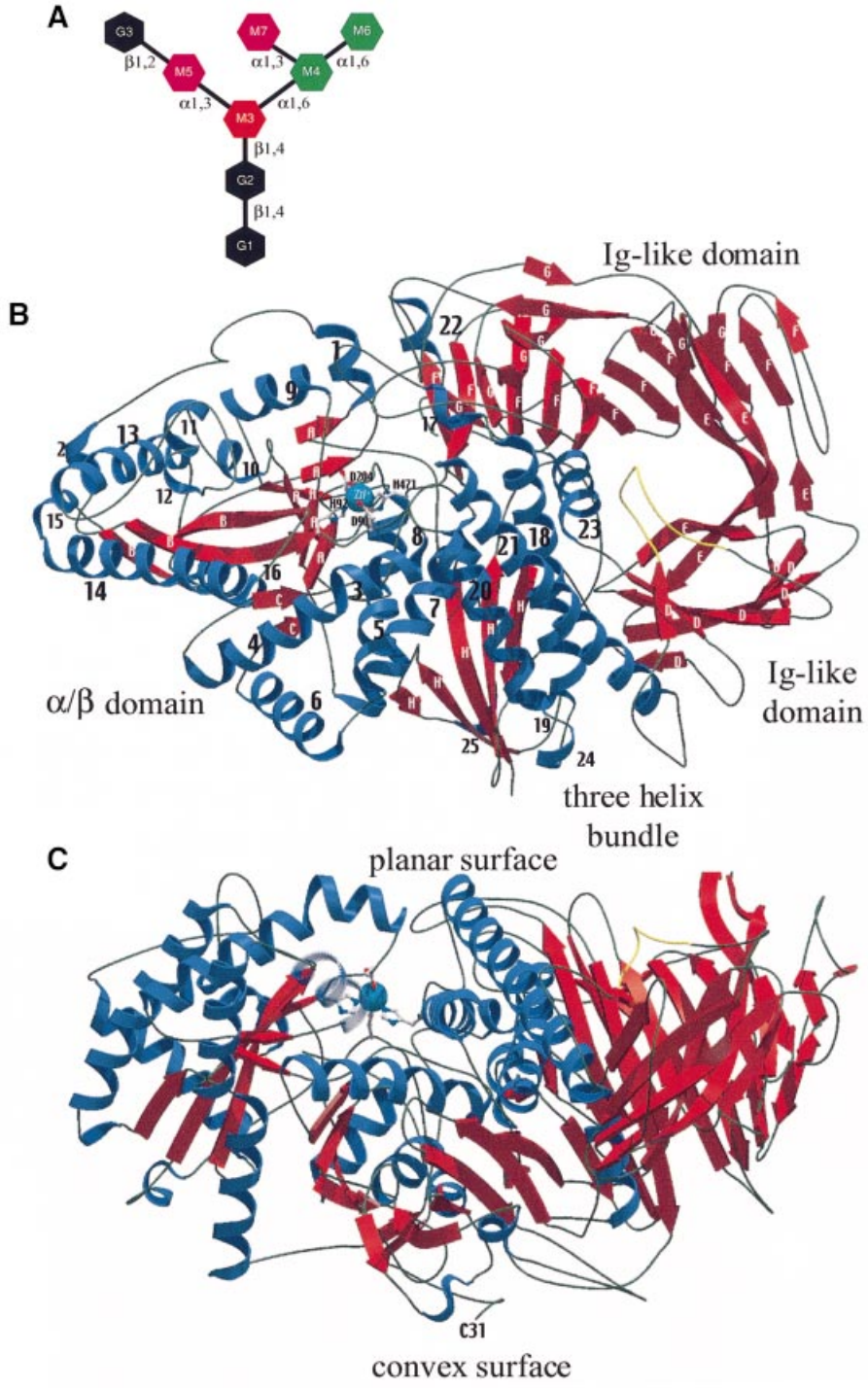

Fig. 1. (**A**) Schematic representation of the high mannose $GlcNAcMan_5GlcNAc_2$ substrate of dGMII. *N*-acetyl-glucosamine residues are shown as black hexagons, α1,6-linked mannose residues (M4 and M6) are colored in green, α1,3-linked mannose residues (M5 and M7) are colored in magenta and β1,4-linked mannose M3 is shown in red. (**B**) Ribbon representation of the dGMII structure, top view onto the planar surface; (**C**) side view. The planar and convex faces of the molecule are indicated. The loop formed by residues 527–540 is shown in yellow. All molecular images were prepared using MOLSCRIPT (Kraulis, 1991) and rendered using Raster3D (Merritt and Bacon, 1997).

mechanism and its inhibition by swainsonine would be a powerful approach to novel glycosylation inhibitors as anti-cancer agents. However, mammalian glycosidases in general, and GMII in particular, have been difficult to purify in quantities suitable for physical studies. Here we have used GMII from *Drosophila melanogaster*, which shows high sequence identity with human GMII (41% identity, 61% similarity), coupled with homologous expression in *Drosophila* cells as a model system for structural studies. We have shown (Rabouille *et al*., 1999) that *Drosophila* Golgi α-mannosidase II (dGMII) displays comparable kinetic properties and inhibitor sensitivity to mammalian GMII, as well as the same substrate specificity (K.W.Moremen, personal communication). In this paper

Table I. Data collection statistics

	MAD (Se-Met) dGMII			Native dGMII		
	Inflection	Peak	Remote	High resolution	DMNJ complex	Swainsonine complex
Wavelength (Å)	0.9790	0.9786	0.9770	1.0	1.0	1.54189
Resolution (Å)[a]	2.14	2.14	2.14	1.76	1.69	1.87
Highest resolution shell	2.31–2.14	2.31–2.14	2.31–2.14	1.90–1.76	1.75–1.69	1.91–1.87
Temperature (K)	100	100	100	100	100	100
No. of unique reflections						
overall	59 212	59 092	59 218	104 565	114 653	87 386
shell	11 288	11 297	11 296	19 882	10 722	5601
Completeness (%)						
overall	99.7	99.8	99.8	97.0	97.8	99.7
shell	96.2	96.3	96.2	94.9	92.6	96.9
R_{merge}[b]						
overall	0.050	0.054	0.057	0.056	0.086	0.078
shell	0.086	0.093	0.105	0.127	0.186	0.452
Correlation of anomalous differences	remote/peak 0.64	remote/edge 0.59	peak/edge 0.70			

[a]Resolution cut-off was determined by the highest resolution shell with completeness >90% and good redundancy. Diffraction data were observed beyond this point.
[b]$R_{merge} = \Sigma_h\Sigma_i|I_i - \langle I\rangle|/\Sigma_i I_i$, where $\langle I\rangle$ is the average of equivalent reflections and the sum is extended over all observations, *i*, for all unique reflections, *h*.

Table II. Refinement statistics

	dGMII	dGMII–swainsonine complex	dGMII–DMNJ complex
Resolution (Å)	500–2.14	500–1.87	500–1.69
R_{cryst} (%)	19.30	18.10	19.69
R_{free} (%)	21.05	20.90	21.56
Atoms (no.)	9194	9202	9199
Residues (no.)	1014	1014	1014
Water molecules (no.)	981	985	983
R.m.s.d. bonds (Å)	0.005	0.005	0.006
R.m.s.d. angles (°)	1.32	1.31	1.33
R.m.s.d. improper dihedrals (°)	0.81	0.78	0.80
Average *B*-factors (Å^2)	15.8	19.4	15.8
Average *B*-factors for inhibitor atoms (Å^2)	15.6	12.0	17.4
Crossvalidated σ_A coordinate error (Å)	0.10	0.14	0.11

$R_{cryst} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. For R_{free}, the sum is extended over a subset of reflections (~10%) excluded from all stages of refinement.

we present the first structure of a Golgi α-mannosidase. We also describe the structure in complexes with the potent inhibitor swainsonine and the mannose-like compound deoxymannojirimycin (DMNJ; 1,5-dideoxy-1,5-imino-D-manitol) bound in the active site of the enzyme. The structure exhibits a previously unobserved protein fold. It enables the proposal of a catalytic mechanism accounting for both the sequential cleavage of two glycosidic bonds in the same catalytic site, and the crucial dependency on the substrate possessing the single β1,2 GlcNAc substituent. These insights lead to new suggestions for GMII-specific inhibitors.

Results and discussion

Protein expression

The cDNA for dGMII is predicted to encode a protein of 1108 amino acids. For protein expression in *Drosophila* cells we eliminated the first 75 amino acids, consisting of the cytosolic and transmembrane domains and most of the stalk region. The remaining cDNA was cloned in-frame behind a secretion signal.

Numbering of our construct starts at the point where the expressed protein is expected to be cleaved, by signal peptidase, from the secretion signal. Three extra N-terminal residues, a His_6 tag, and a glycine, glutamine and phenylalanine were added in cloning. The first aspartate (D13) of the construct corresponds to Asp76 of the native protein. The first residue seen in the structure (C31) corresponds to C94, and the final residue S1044 to S1107 of the full-length sequence.

Structure determinations

The structure of dGMII has been determined by the multi-wavelength anomalous dispersion (MAD) phasing method using a data set collected from a crystal of selenomethionine (Se-Met)-derivatized enzyme (Table I). To our knowledge, it is the first reported structure of an Se-Met-substituted enzyme produced in a *Drosophila* overexpression system. The native dGMII structure has been refined

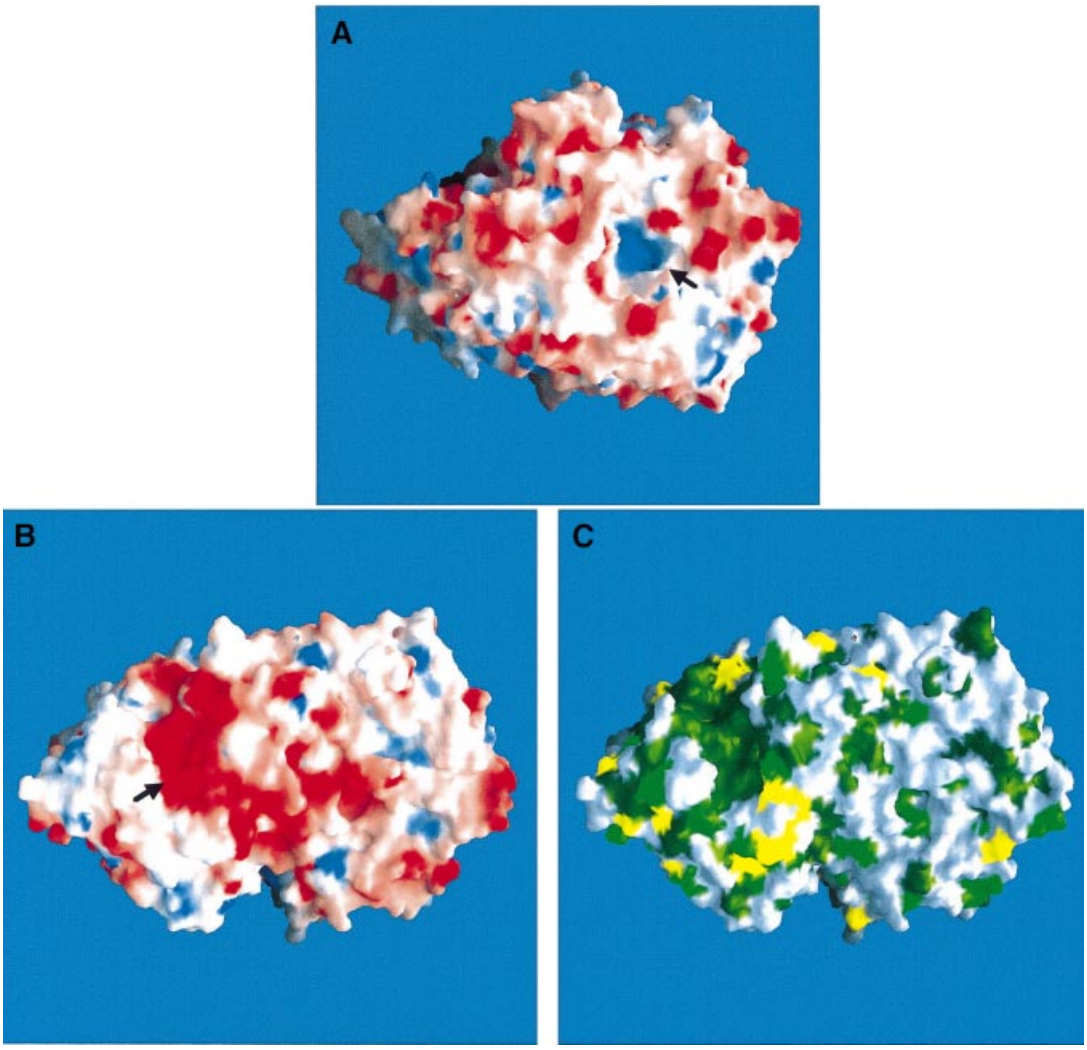

Fig. 2. Molecular surface representation of the convex (**A**) and the planar face (**B**) of the dGMII molecule. Molecular surface images are colored for electrostatic potential (red for negative, blue for positive). Arrows indicate the pore (A) and active site (B) regions. (**C**) Molecular surface representation of the planar face of dGMII, colored for homology with the sequence of human GMII (dark green for identical, light green for conserved, yellow for similar and white for different residues). Alignment of human and *Drosophila* GMII sequences (SwissProt accession Nos Q16706 and Q24451, respectively) was performed using the GAP program of the Wisconsin package (Version 10, Genetics Computer Group) using the default parameters without any manual intervention. The scores were used to color the molecular surface. All molecular surface images were produced using GRASP (Nicholls *et al*., 1991).

to a resolution of 2.14 Å (see refinement statistics presented in Table II). The model contains residues 31–1044 of the recombinant enzyme (numbered as described above), as well as a zinc ion, an *N*-glycan residue, a molecule of the cryo-protectant 2-methyl-2,4-pentanediol (MPD), and a 2-amino-2-hydroxy-methyl-propane-1,3-diol (Tris) molecule. The presence of the enzyme-bound zinc ion was confirmed by inductively coupled plasma atomic emission spectroscopy (ICP-AES) in a molar ratio of 1:1 (data not shown). The final structure of the dGMII–swainsonine complex has been refined to 1.87 Å resolution and the dGMII–DMNJ complex to 1.69 Å resolution.

Overall architecture of dGMII

The structure of dGMII reveals a previously unobserved protein fold consisting of an N-terminal α/β-domain, a three-helical bundle and an all-β C-terminal domain forming a single compact entity, connected by five internal disulfide bonds and stabilized by a zinc-binding site (Figure 1B). The oval-shaped molecule has two distinct faces (Figure 1C). The N-terminal face of the molecule is convex, whereas the opposing face of the enzyme has a planar surface. N-terminal residue Cys31 is the last residue of the so-called stalk region, the linkage between the catalytic domain and the transmembrane domain. Cys31 is located at the convex face of the molecule, indicating that this surface of the molecule presumably faces the inner side of the Golgi membrane, while the planar surface, containing the active site cavity (see below), faces the Golgi lumen.

The N-terminal α/β-domain is comprised of an inner core of three β-sheets (A, B and C; Figure 1B) consisting of 11, mostly parallel β-strands, surrounded by 16 α-helices. This domain contains a GlcNAc residue found in the electron density map at a consensus *N*-glycosylation

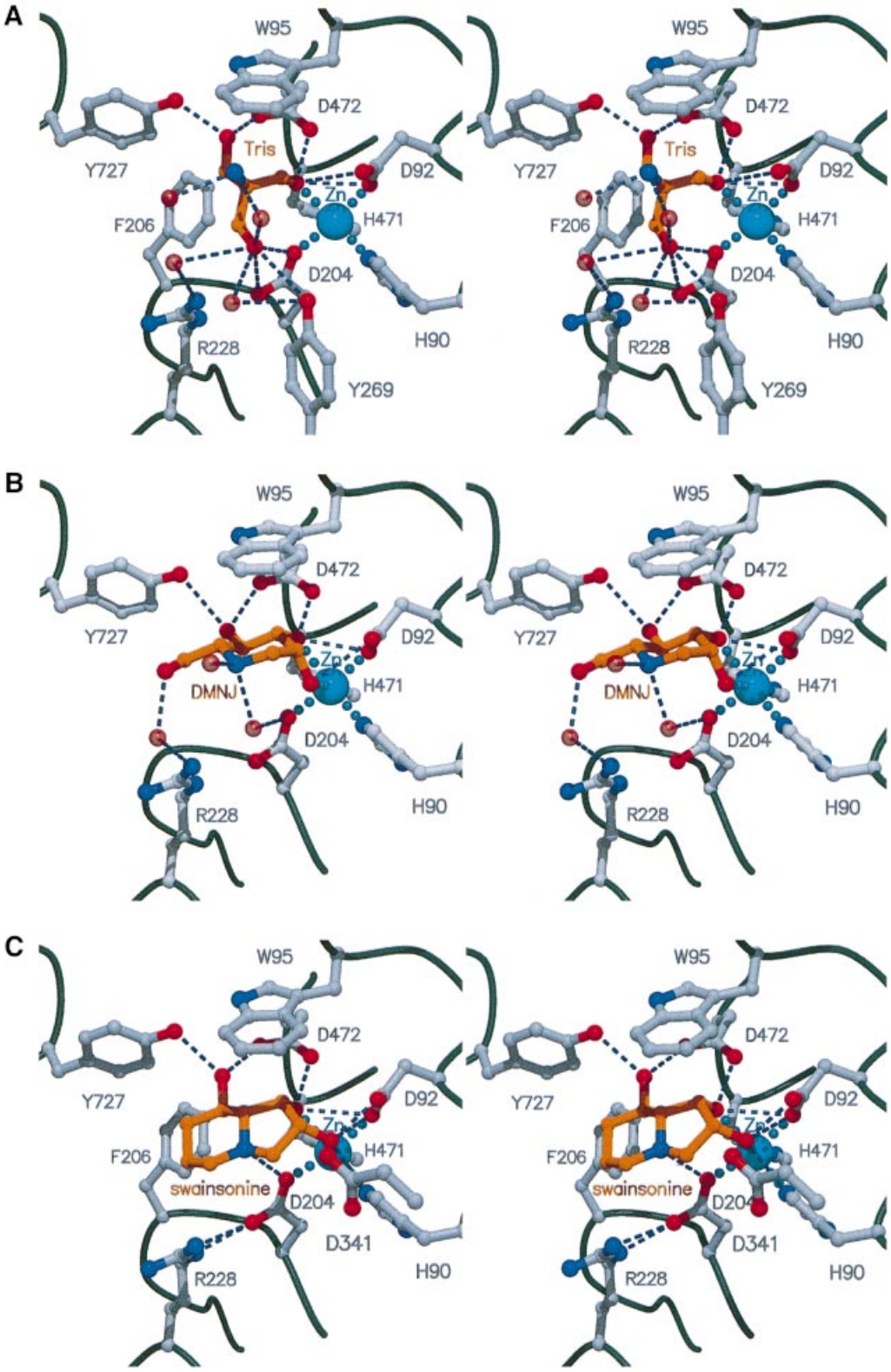

Fig. 3. Stereo views of the active site of dGMII with bound Tris (**A**), DMNJ (**B**) and swainsonine (**C**) molecules. The active site zinc ion is shown in turquoise, the bound inhibitor molecules are rendered in gold and water molecules are represented as transparent red spheres. Interatomic distances <3.2 Å are shown as blue dashed lines.

site (Asn194), located at the N-terminus of helix 7. The α/β-domain is stabilized by three disulfide bonds: between Cys31 and Cys1032 connecting the N- and C-terminal extremes of dGMII; Cys275 and Cys282 linking helices 10 and 11; and Cys283 and Cys297 linking helix 11 with a loop between helix 13 and the core of parallel β-sheets.

The cysteines forming the latter two disulfide bonds are conserved in the human GMII sequence.

The C-terminal half of the protein contains a three-helix bundle, comprised of helices 18, 20 and 21, and is connected to the N-terminal α/β-domain via a zinc-binding site. The zinc ion is coordinated in a rare pentavalent geometry (T_5-square-based pyramidal) involving residues Asp90, His92, Asp204 and His471. Furthermore, the C-terminal domain contains two immunoglobulin-like domains: a small β-sandwich consisting of 12 anti-parallel strands from β-sheets D and E, and a large 21-strand structure involving β-sheets F and G.

A barrel formed by the three-helix bundle and helix 23 together with the two β-sandwich structures results in a narrow pore in the center of the C-terminal domain. The pore is lined by six arginine residues (Arg540, 565, 617, 770, 777 and 893) contributing to the overall positive charge of the pore (Figure 2A, see arrow). A hairpin loop, connecting two strands of β-sheet D (Figure 1B and C, residues 527–540, shown in yellow) protrudes into the center of the barrel on the planar side of the molecule. Arg530, located at the tip of the type I β-turn in this loop, plugs the pore, preventing an open channel through the protein. The resulting crater-like cavity on the convex side of the molecule is 20 Å deep, with a diameter of 20 Å funneling to 8 Å at the bottom of the cavity. *B*-factor values of residues within the loop indicate a higher degree of flexibility compared with the rest of the structure (average *B*-factor values ~33 and ~15 $Å^2$, respectively). The function of this loop and of the C-terminal domain as a whole has yet to be identified.

Active site

The molecular surface representation of the planar face of dGMII reveals an extended pocket in the N-terminal α/β-domain, formed primarily by acidic residues (Figure 2B, see arrow). These same residues form the core of a large, contiguous, surface-exposed patch of highly conserved amino acids, in comparison with the human GMII sequence (Figure 2C). The active site of the enzyme is located in a small cavity in the side of this conserved, negatively charged region. The cavity is lined by aromatic residues Trp95, Phe206, Tyr269 and Tyr727, which are involved in hydrophobic and hydrogen bond interactions with a bound Tris molecule in the unliganded structure (Figure 3A). Tris is known to inhibit dGMII activity (IC_{50} 40 mM) (Rabouille *et al*., 1999). Additional hydrophobic and hydrogen bond interactions are observed with Asp92 and Asp204. At the open side of the cavity, the Tris molecule hydrogen bonds with Arg228, Tyr269 and Asp341 (not shown) via water molecules.

A key feature of the active site is the coordination of the zinc ion by the Tris hydroxyl group O2. In the enzyme–Tris complex, the zinc ion is bound in a T_5-square-based pyramidal geometry, coordinated by the OD1 oxygen moieties of Asp92 and Asp204, the NE2 nitrogens of His90 and 471, and the hydroxyl oxygen O2 of the bound Tris molecule, as represented in Figure 3A. The T_5 geometry is further stabilized by hydrogen bonds between the zinc coordinating atoms and by hydrogen bonds between the ND1 nitrogen atoms of His90 and His471 with the carbonyl oxygen of Se-Met 167 and a water molecule, respectively (not shown). Hydrogen bonding of oxygen-containing residues (or water) to zinc-bound histidines, the so-called 'elec-His-Zn motifs', is believed to increase the basicity and the ligand strength of the histidine and arrange it correctly for interaction with the metal (Alberts *et al*., 1998). In an uninhibited enzyme, Tris would probably be replaced by a coordinating water molecule. As discussed below, this arrangement has implications for substrate binding and transition state stabilization.

The occurrence of zinc in family 38 glycosyl hydrolases has been described by Snaith (1975) in Jack-bean α-mannosidase. A possible role for zinc in catalysis was indicated by inactivation of the enzyme by chelating agents and bivalent metal ions such as Cu^{2+}. Copper has also been shown to effectively inactivate *Drosophila* and mouse GMII (Rabouille *et al*., 1999).

Inhibitor binding

The structures of dGMII in complex with the inhibitors DMNJ (IC_{50} 400 μM) and swainsonine (IC_{50} 20 nM) show that both compounds bind to the same active site in a similar manner (Figure 3B and C). The binding of both inhibitors involves a large contribution of hydrophobic interactions involving aromatic residues Trp95, Phe206 and Tyr727, forming the walls of the cavity. The inhibitor ring structures are stacked against Trp95, a feature seen in several carbohydrate-binding and -hydrolyzing proteins (for review see Boraston *et al*., 2000), and stabilized by hydrogen bonds and interactions with the zinc ion. In the complexes of dGMII with either DMNJ or swainsonine, the T_5 geometry of the bound zinc ion, as seen in the Tris-bound enzyme, is transformed into T_6-octahedral coordination. In both the dGMII complexes, the inhibitor O2 hydroxyl oxygen replaces the O2 oxygen of Tris, and the O3 hydroxyl oxygen forms the apex of the second pyramid. In order to obey the restraints of the T_6 geometry, the plane of the swainsonine ring structure is tilted with respect to the saccharide-like ring of the bound DMNJ molecule. This enables the formation of a hydrogen bond between the zinc-coordinating OD1 oxygen of Asp204 and the N4 nitrogen at the fusion of the five- and six-membered rings of swainsonine. As in the Tris-bound enzyme, the zinc coordinating oxygen atoms of the inhibitors are involved in hydrogen bond interactions with the neighboring metal-binding residues of the enzyme.

The position of the DMNJ and swainsonine molecules is stabilized in the active site by hydrogen bonds between carboxylic oxygens OD1 and OD2 of Asp472 and hydroxyl oxygens O3 and O4 (O5 in swainsonine) of the inhibitors, analogous to the O1 and O2 interactions seen in the enzyme–Tris complex. As in the Tris-bound enzyme, DMNJ is involved in additional hydrogen bonds, via water molecules, with the NH2 nitrogen of Arg228, the hydroxyl oxygen of Tyr269, the backbone carbonyl oxygen of Arg876 (not shown) and the OD1 oxygen of Asp204.

The displacement of the Tris molecule by either of the inhibitors only slightly affects the zinc-binding site by weakening the internal hydrogen bonds between Asp204 and His90 and His471. No major conformational changes are observed between the Tris-bound and the inhibitor-bound mannosidase molecules, as their backbones are virtually superimposable, with root mean square

deviations (r.m.s.d.) between C_α atoms of 0.068 Å (dGMII–DMNJ complex) and 0.087 Å (dGMII–swainsonine complex).

Catalytic mechanism

GMII is a retaining mannosyl hydrolase, which cleaves the linkage between the C1 atom of M7 and M6 (Figure 1A) and, respectively, the O3 and O6 atom of the α1,6-linked mannosyl branch (M4) of $GlcNAcMan_5GlcNAc_2$. The catalytic mechanism is proposed to follow a very similar path to the corresponding retaining β-glycosidases (Braun *et al*., 1995; White and Rose, 1997; Numao *et al*., 2000). This is a two-stage reaction that usually involves two carboxylic acids, one acting as a nucleophile attacking the glycosidic bond, and the other as a general acid/base catalyst. Nucleophilic attack of one carboxylic acid results in glycosylation of the enzyme by forming a covalent intermediate followed by a second deglycosylation step, each step passing through an oxocarbonium ion-like transition state.

Based on the structure of the dGMII–inhibitor complexes we speculate that mannose residues M6 and M7 on the α1,6-linked mannosyl branch (M4) bind to the enzyme at the same site and in the same manner as mannose-like inhibitor DMNJ. Coordination of the zinc ion with the O2 and O3 hydroxyl oxygens thereby contributes to the enzyme's specificity for mannose. Four acidic amino acid residues, Asp92, Asp204, Asp341 and Asp472, are candidates for catalytic side chains based on their proximity to the active site (Figure 3C). Results from a recent study on the mechanism of catalysis in Jack-bean α-mannosidase, by Withers and co-workers, using reagents that trap the glycosyl–enzyme intermediate, identified an aspartate residue as the catalytic nucleophile in that enzyme (Howard *et al*., 1998). Comparison of the highly conserved sequence region surrounding this aspartate in Jack-bean α-mannosidase with the same sequence region in dGMII suggests that Asp204 in dGMII is the catalytic nucleophile that attacks the glycosidic linkage, as has been confirmed recently for the family 38 bovine lysosomal α-mannosidase (Numao *et al*., 2000). For this reaction it is required that Asp204 is close to the anomeric carbon of the mannose substrate. In the dGMII–DMNJ complex, however, the equivalent anomeric carbon is located 4.6 Å from the nucleophile. Binding of the C2 and C3 substituent hydroxyl oxygens of the flattened five-membered ring in swainsonine causes the inhibitor molecule to tilt, bringing its bridgehead nitrogen N4, in the analogous position to C1 in the substrate, significantly closer to the putative nucleophilic Asp204 (3.2 Å). This tilted binding mode, stabilized by a hydrogen bond between N4 and Asp204 and by van der Waals stacking interactions between the six-membered ring of swainsonine and Phe206, may resemble the mode of binding of the ring-flattened transition state mannosyl cation. Thus, Phe206 would stabilize the transition state by compensating for the loss of stacking interactions of the substrate with Trp95. The highly complementary shape of swainsonine with the active site of dGMII, and its structural analogy with the skewed boat transition state conformation, could therefore explain its 10 000 times higher binding affinity for the enzyme, compared with the substrate-mimic DMNJ.

The OD1 oxygen of Asp204, the putative nucleophile, directly coordinates the zinc ion, implicating a role for the zinc in positioning the nucleophile and in the stabilization of protonation states of the reacting partners. It is tempting to speculate that the change of zinc coordination from pentavalent (T_5) to the less favored hexavalent (T_6) state (Alberts *et al*., 1998) on substrate binding may also contribute to the mechanism. From the Tris and DMNJ structures, we would predict that the coordination would revert to T_5 on product release. If so, this transition may energetically facilitate the deglycosylation step. Such evidence of direct zinc involvement in the catalytic mechanism of a glycosyl hydrolase is unprecedented. Arg288 positions Asp204 for nucleophilic attack by virtue of hydrogen bond interactions between its NE and NH2 nitrogens and the OD2 oxygen of Asp204 (Figure 3C). Based on the expected distance between the two catalytic residues (~5.5 Å; Davies and Henrissat, 1995), likely candidates for the catalytic acid base are Asp341 and Asp472. Asp341 is the more likely candidate, based on geometric disposition. Furthermore, preliminary indications are that the D341N mutant is catalytically inactive (D.A.Kuntz, unpublished results). Recent data suggest that other residues, such as tyrosines, play a role in glycosidic bond cleavage. For example, in the family 11 xylanase structures, a Tyr stabilizes distortion of the bound covalent intermediate and, possibly, the transition state (Sabini *et al*., 1999; Sidhu *et al*., 1999; Zechel and Withers, 1999). In dGMII, Tyr269 and Tyr727 are positioned for a similar function.

Substrate binding and cleavage

The function of GMII is dependent on the presence of β1,2-GlcNAc (G3; Figure 1A), added to α1,3-linked mannose (M5) by GlcNAc transferase I (for reviews see Kornfeld and Kornfeld, 1985; Moremen *et al*., 1994). This β1,2-GlcNAc dependence suggests the presence of an additional saccharide-binding site in GMII. An indication for such a binding site is provided by the observation of an MPD molecule in the structure of dGMII, in the vicinity of the active site cavity. MPD was used as a cryo-protectant during the procedure of flash-freezing of the crystal, prior to data collection (see Materials and methods). The replacement of MPD by the alternative cryo-protectant glycerol resulted in the occupation of this same position by a glycerol molecule. Glycerol has been shown to mimic saccharide binding in structures of glycosyl hydrolases (Schmidt *et al*., 1998; Vallée *et al*., 2000).

The observation of the binding of MPD and glycerol near dGMII's active site (Figure 4A) enables a hypothesis regarding the binding and cleavage of α1,6- and α1,3-linked mannoses on the α1,6-linked mannose branch of the $GlcNAcMan_5GlcNAc_2$ oligosaccharide. In this hypothesis, the MPD-binding site is suggested to be the putative site of interaction for β1,2-GlcNAc (G3; Figure 1A), enabling anchoring of the oligosaccharide substrate in the conserved negatively charged pocket. In Figure 4B, a model is shown of a $GlcNAcMan_5GlcNAc_2$ structure with the β1,2-GlcNAc residue placed in the MPD-binding site and the α1,6-linked M6 mannose docked into the active site, with its hydroxyl oxygens O2 and O3 coordinating the zinc ion. The starting point of this model was an energy-minimized structure of

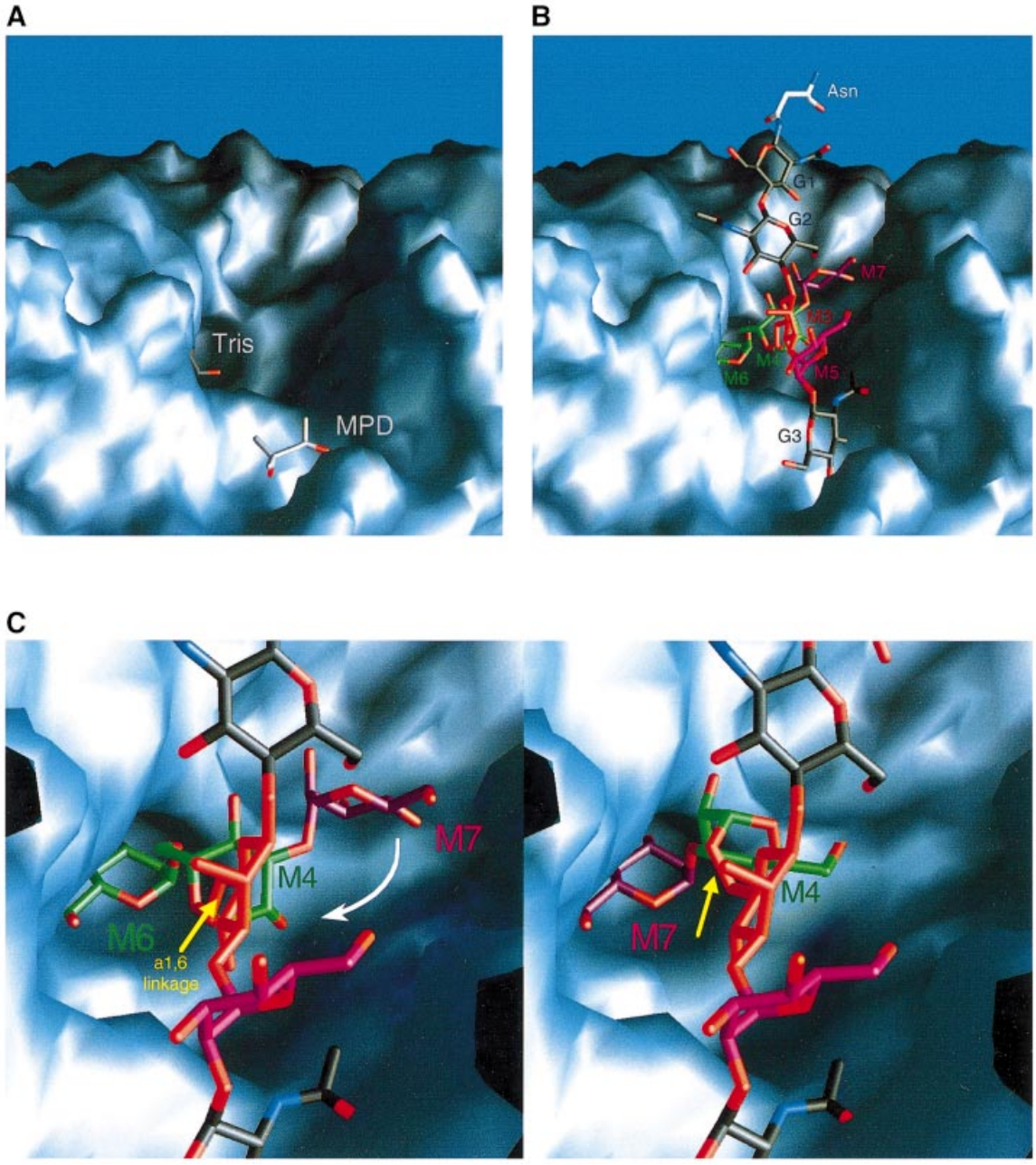

Fig. 4. (**A**) Molecular surface representation of dGMII showing the position of the active site-bound Tris molecule and the MPD-binding site. (**B**) Molecular surface representation of dGMII with the $GlcNAcMan_5GlcNAc_2$ substrate modeled into the binding pocket. The substrate molecule is positioned into the binding pocket with α1,6-linked mannose M6 (shown in green) docked into the active site and β1,2-GlcNAc residue G3 (shown in black) placed in the MPD-binding site. Individual mannose residues of the substrate are colored according to the coloring scheme used in Figure 1A. (**C**) Representation of the sequential trimming of the α1,6- (M6) and α1,3-linked (M7) mannose residues. All molecular surface images were produced using GRASP (Nicholls *et al*., 1991).

$GlcNAcMan_5GlcNAc_2$ (I.Tvaroska, personal communication) with minor adjustments of some of the flexible glycosidic linkages. No further minimization has been carried out. As required, the asparagine-linked β1,4-GlcNAc residues G1 and G2 extend away from the surface of the molecule (into the Golgi lumen). Both M4 and the second substrate α1,3-linked M7 mannose are located within the conserved negatively charged pocket, pointing away from the active site cavity. In this orientation, it can easily be visualized that after cleavage of the α1,6-linked M6, the α1,3-linked M7 can be brought into the active site cavity by an ~180° rotation, through the extended pocket, around the flexible α1,6-linkage of M4 (see Figure 4C). In addition to the dependence of GMII's action on the presence of the G3 β1,2-GlcNAc, this model provides a mechanism for the cleavage of both mannose residues without major conformational change of the enzyme, and more importantly, without release of the polypeptide–carbohydrate complex, anchored by the stationary GlcNAc, between the two cleavage events. Finally, this model suggests that the α1,6-linked M6 mannose is preferentially cleaved first, enabling the shorter

α1,3-linked M7 residue to rotate through the pocket with minimal steric hindrance; according to our model, the proposed 'swivel' mechanism would be slightly hampered should the M7 mannose be cleaved first. This is supported by data reported for α-mannosidase II from mung bean seedlings, *Xenopus* liver, rat liver Golgi and for enzyme activity in homogenates of insect cells, showing preferential hydrolytic activity on the M6 mannosyl residue (Kaushal *et al*., 1990; Altmann and Marz, 1995; Ren *et al*., 1997).

Conclusions

The structure of the catalytic domain of GMII provides the basis for its zinc ion-mediated specificity for mannose, as well as insight into its reaction mechanism. In addition, the result illustrates the structural basis for the mechanism of inhibition by the anti-cancer agent swainsonine, which we propose mimics aspects of the transition state binding. This understanding is critical for the rational design of swainsonine variants and/or novel mechanism-based compounds as specific α-mannosidase II inhibitors, for the treatment of several forms of cancer. A bound MPD molecule identifies a putative GlcNAc binding pocket, located near the active site, and enables a hypothesis explaining the enzyme's dependency on the single GlcNAc substitution of the $GlcNAcMan_5GlcNAc_2$ substrate for binding. Furthermore, it suggests a novel mechanism for successive hydrolysis of the α1,6- and α1,3-linked mannose residues, resulting in the tri-mannose core glycosyl structure. Finally, it opens the door to the design of novel highly specific inhibitors linking together functional sites in the enzyme.

Materials and methods

Protein overexpression and purification

Expression, purification and crystallization of the dGMII will be described in detail elsewhere. Briefly, the cDNA was inserted behind an inducible promoter, and used to stably transfect *Drosophila* S2 cells. Single-cell clones secreting high levels of dGMII were chosen and adapted to serum-free medium. Unlabeled dGMII was isolated from the supernatants of cells grown in Fernbach flasks, by batch binding to Blue–agarose (Sigma). The protein was eluted from the Blue–agarose using NaCl and further purified by Ni-NTA chromatography (Qiagen). EDTA (5 mM) was added to scavenge any free nickel. The protein was extensively dialyzed against 10 mM Tris pH 8 containing 100 mM NaCl, concentrated to 25 mg/ml, and stored in aliquots at –80°C.

For Se-Met labeling, we used a custom batch of Ex-Cell 420 (#006140E; JRH Biosciences, Lenexa, KS), which lacked any added methionine or copper. Cells were grown to high cell density in a spinner flask in standard medium, resuspended in the 'methionine-free' medium and allowed to starve for 4 h prior to the addition of 50 mg/l Se-Met (Sigma). After 70 h of induction, the protein was purified from the supernatant as outlined above except that 5 mM β-mercaptoethanol was present throughout the purification.

Crystallization and data collection

Crystals of dGMII and complexes of the enzyme with various inhibitors were grown at room temperature using vapor diffusion and micro-batch crystallization techniques. Crystals were obtained under a wide variety of conditions. Polyethylene glycol (PEG) was used as a precipitant (with sizes 4000, 6000, 8000, 10 000 and 20 000) at concentrations varying from 5 to 20%, in the presence of 5% MPD or 0–30% glycerol. Crystallization solutions were buffered at pH 7–7.5 using 100 mM buffer solutions of Tris, HEPES or Mes. The crystals belong to the orthorhombic space group $P2_12_12_1$ with cell dimensions a = 69 Å; b = 110 Å; c = 139 Å; α = β = γ = 90°. For the initial structure determination, Se-Met-derivatized mannosidase II crystals were grown in 8.5% PEG 6000, 5% MPD and 100 mM Tris pH 7.0, using micro seeds obtained from wild-type enzyme crystals. Data were collected from crystals that were frozen in liquid nitrogen after a stepwise increase of the MPD or glycerol concentration in the crystallization solution from 5 to 25%. Data collection was performed at the Advanced Photon Source facility at Argonne National Laboratories, Argonne, IL. Beam line BM14D was used for collection of MAD data and BM14C for collection of high-resolution data. Diffraction data were processed using DENZO and SCALEPACK (Otwinowski and Minor, 1997).

Structure determination

The structure of uncomplexed dGMII was determined by MAD phasing at the selenium absorption edge with data sets collected at an absorption peak wavelength of 0.9786 Å, inflection wavelength of 0.9790 Å and a remote wavelength of 0.9770 Å. Initial positions of 26 out of 28 selenium atoms were determined with the program Solve (Terwilliger *et al*., 1987) with an initial figure of merit of 0.67. The experimental map obtained after density modification, using the program DM of the CCP4 program package (Cowtan, 1994), showed continuous density of very high quality for the whole molecule. The structure was traced using the program O (Jones *et al*., 1991) using the density-modified experimental map. The model was refined using the program CNS (Brünger *et al*., 1998).

ICP-AES

The metal content in dGMII samples was analyzed by ICP-AES using the model 'Optima 3000 DV' (Dual View) from Perkin Elmer. The zinc content in the protein samples was determined relative to an equivalent amount of dGMII assay buffer.

Protein Data Bank accession Nos

The accession numbers for the coordinates of the structures reported in this paper are 1HTY, 1HWW and 1HXK.

Acknowledgements

We thank Tara Signorelli for her participation in this work. We are grateful to Dan Mathers (ANALEST, University of Toronto, Canada) for performing the ICP-AES analysis. dGMII cDNA was obtained from Dr David Roberts. We thank Drs Stephen Withers, Jeremy Carver and colleagues at GlycoDesign, and Harry Schachter and Kelley Moremen for their helpful, critical discussions. The staff of the BioCARS beamline at the Advanced Photon Source supported X-ray data collection. This research is funded by the Protein Engineering Network of Centres of Excellence and the Canadian Institutes of Health Research. J.M.H.v.d.E. was funded by a Medical Research Council of Canada Postdoctoral Fellowship.

References

Alberts,I.L., Nadassy,K. and Wodak,S.J. (1998) Analysis of zinc binding sites in protein crystal structures. *Protein Sci.*, **7**, 1700–1716.

Altmann,F. and Marz,L. (1995) Processing of asparagine-linked oligosaccharides in insect cells: evidence for α-mannosidase II. *Glycoconj. J.*, **12**, 150–155.

Boraston,A., Chiu,P., Warren,R.A.J. and Kilburn,D.G. (2000) Specificity and affinity of substrate binding by a family 17 carbohydrate-binding module from *Clostridium cellulovorans* cellulase 5A. *Biochemistry*, **39**, 11129–11136.

Braun,C., Brayer,G.D. and Withers,S.G. (1995) Mechanism-based inhibition of yeast α-glucosidase and human pancreatic α-amylase by a new class of inhibitors. 2-Deoxy-2,2-difluoro-α-glycosidases. *J. Biol. Chem.*, **270**, 26778–26781.

Brünger,A.T. *et al*. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.

Coutinho,P.M. and Henrissat,B. (1999) Carbohydrate-active enzymes server at URL: http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html

Cowtan,K. (1994) 'dm': an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsl. Protein Crystallogr.*, **31**, 34–38.

Davies,G. and Henrissat,B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure*, **3**, 853–859.

Dennis,J.W. and Laferte,S. (1985) Recognition of asparagine-linked oligosaccharides on murine tumor cells by natural killer cells. *Cancer Res.*, **45**, 6034–6040.

Dennis,J.W., Granovsky,M. and Warren,C.E. (1999a) Protein glycosylation in development and disease. *BioEssays*, **21**, 412–421.
Dennis,J.W., Granovsky,M. and Warren,C.E. (1999b) Glycoprotein glycosylation and cancer progression. *Biochim. Biophys. Acta*, **1473**, 21–34.
Goss,P.E., Baker,M.A., Carver,J.P. and Dennis,J.W. (1995) Inhibitors of carbohydrate processing: a new class of anticancer agents. *Clin. Cancer Res.*, **1**, 935–944.
Goss,P.E., Reid,C.L., Bailey,D. and Dennis,J.W. (1997) Phase IB clinical trial of the oligosaccharide processing inhibitor swainsonine in patients with advanced malignancies. *Clin. Cancer Res.*, **3**, 1077–1086.
Harpaz,N. and Schachter,H. (1980) Control of glycoprotein synthesis. V. Processing of asparagine-linked oligosaccharides by Golgi α-D-mannosidases dependent on the prior action of UDP-*N*-acetylglucosamine:α-D-mannoside β-2-*N*-acetylglucosaminyl-transferase I. *J. Biol. Chem.*, **255**, 4894–4902.
Henrissat,B. (1991) A classification of glycosyl hydrolases based on amino-acid sequence similarities. *Biochem. J.*, **280**, 309–316.
Herscovics,A. (1999) Importance of glycosidases in mammalian glycoprotein biosynthesis. *Biochim. Biophys. Acta*, **1473**, 96–107.
Howard,S., He,S. and Withers,S.G. (1998) Identification of the active site nucleophile in Jack-bean α-mannosidase using 5-fluoro-β-L-gulosyl fluoride. *J. Biol. Chem.*, **273**, 2067–2072.
Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
Kaushal,G.P., Szumilo,T., Pastuszak,I. and Elbein,A.D. (1990) Purification to homogeneity and properties of mannosidase II from mung bean seedlings. *Biochemistry*, **29**, 2168–2176.
Kiyohara,T., Dennis,J.W. and Roder,J.C. (1987) Double restriction in NK cell recognition is linked to transmethylation and can be triggered by asparagine-linked oligosaccharides on tumor cells. *Cell. Immunol.*, **106**, 223–233.
Kornfeld,R. and Kornfeld,S. (1985) Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.*, **54**, 631–664.
Kraulis,P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
Merritt,E.A. and Bacon,D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.
Moremen,K.W. and Touster,O. (1985) Biosynthesis and modification of Golgi mannosidase II in HeLa and 3T3 cells. *J. Biol. Chem.*, **260**, 6654–6662.
Moremen,K.W. and Touster,O. (1986) Topology of mannosidase II in rat liver Golgi membranes and release of the catalytic domain by selective proteolysis. *J. Biol. Chem.*, **261**, 10945–10951.
Moremen,K.W., Trimble,R.B. and Herscovics,A. (1994) Glycosidases of the asparagine-linked oligosaccharide processing pathway. *Glycobiology*, **4**, 113–125.
Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
Numao,S., He,S., Evjen,G., Howard,S., Tollersrud,O.K. and Withers,S.G. (2000) Identification of Asp197 as the catalytic nucleophile in the family 38 α-mannosidase from bovine kidney lysosomes. *FEBS Lett.*, **484**, 175–178.
Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscilation mode. *Methods Enzymol.*, **276**, 307–326.
Rabouille,C., Kuntz,D.A., Lockyer,A., Watson,R., Signorelli,T., Rose,D.R., van den Heuvel,M. and Roberts,D.B. (1999) The *Drosophila* GMII gene encodes a Golgi α-mannosidase II. *J. Cell Sci.*, **112**, 3319–3330.
Ren,J., Castellino,F.J. and Bretthauer,R.K. (1997) Purification and properties of α-mannosidase II from Golgi-like membranes of baculovirus-infected *Spodoptera frugiperda* (IPLB-SF-21AE) cells. *Biochem. J.*, **324**, 951–956.
Sabini,E., Sulzenbacher,G., Dauter,M., Dauter,Z., Jorgensen,P.L., Schulein,M., Dupont,C., Davies,G.J. and Wilson,K.S. (1999) Catalysis and specificity in enzymatic glycoside hydrolysis: a 2,5B conformation for the glycosyl–enzyme intermediate revealed by the structure of the *Bacillus agaradhaerens* family 11 xylanase. *Chem. Biol.*, **6**, 483–492.
Schmidt,A., Schlacher,A., Steiner,W., Schwab,H. and Kratky,C. (1998) Structure of xylanase from *Penicillium simplicissimum*. *Protein Sci.*, **7**, 2081–2088.
Sidhu,G., Withers,S.G., Nguyen,N.T., McIntosh,L.P., Ziser,L. and Brayer,G.D. (1999) Sugar ring distortion in the glycosyl–enzyme intermediate of a family G/11 xylanase. *Biochemistry*, **38**, 5346–5354.
Snaith,S.M. (1975) Characterization of jack-bean α-D-mannosidase as a zinc metalloenzyme. *Biochem. J.*, **147**, 83–90.
Terwilliger,T.C., Kim,S.H. and Eisenberg,J. (1987) Generalized method of determining heavy-atom positions using the difference Patterson function. *Acta Crystallogr. A*, **43**, 1–5.
Vallée,F., Lipari,F., Yip,P., Sleno,B., Herscovics,A. and Howell,P.L. (2000) Crystal structure of a class I α1,2-mannosidase involved in *N*-glycan processing and endoplasmic reticulum quality control. *EMBO J.*, **19**, 581–588.
White,A. and Rose,D.R. (1997) Mechanism of catalysis by retaining β-glycosyl hydrolases. *Curr. Opin. Struct. Biol.*, **7**, 645–651.
Zechel,D.L. and Withers,S.G. (1999) Glycoside mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.*, **33**, 11–18.

Received March 12, 2001; revised April 27, 2001; accepted April 30, 2001