



Cite this: *Chem. Soc. Rev.*, 2017,  
46, 5128

Received 11th December 2016

DOI: 10.1039/c6cs00897f

[rsc.li/chem-soc-rev](http://rsc.li/chem-soc-rev)

## The ENGases: versatile biocatalysts for the production of homogeneous *N*-linked glycopeptides and glycoproteins

Antony J. Fairbanks  <sup>ab</sup>

The *endo*- $\beta$ -*N*-acetylglucosaminidases (ENGases) are an enzyme class (EC 3.2.1.96) produced by a range of organisms, ranging from bacteria, through fungi, to higher order species, including humans, comprising two sub families of glycosidases which all cleave the chitobiose core of *N*-linked glycans. Synthetic applications of these enzymes, *i.e.* to catalyse the reverse of their natural hydrolytic mode of action, allow the attachment of *N*-glycans to a wide variety of substrates which contain an *N*-acetylglucosamine (GlcNAc) residue to act as an 'acceptor' handle. The use of *N*-glycan oxazolines, high energy intermediates on the hydrolytic pathway, as activated donors allows their high yielding attachment to almost any amino acid, peptide or protein that contains a GlcNAc residue as an acceptor. The synthetic effectiveness of these biocatalysts has been significantly increased by the production of mutant glycosynthases; enzymes which can still catalyse synthetic processes using oxazolines as donors, but which do not hydrolyse the reaction products. ENGase biocatalysts are now finding burgeoning application for the production of biologically active glycopeptides and glycoproteins, including therapeutic monoclonal antibodies (mAbs) for which the oligosaccharides have been remodelled to optimise effector functions.

### 1. Introduction

Protein glycosylation<sup>1</sup> is the most diverse form of post-translational modification, playing a key role in protein folding, and crucially affecting many important protein properties, such as conformation and stability, susceptibility to proteases, and circulatory lifetime. Protein glycosylation is central to many vital biological processes, such as cell-cell signalling, development, and immune response. In *N*-linked glycosylation,<sup>2</sup> the predominant form, the oligosaccharides/glycans are attached to the amide nitrogen of particular protein asparagine residues, whereas in the alternative *O*-linked glycosylation, glycans are most commonly attached *via* the side chains of serine or threonine residues.

Many important mammalian proteins of clinical and therapeutic significance are *N*-glycosylated. Amongst these immunoglobulin G (IgG) antibodies, the main antibodies found in blood and extracellular fluid, are of particular therapeutic and commercial importance. Indeed recombinant monoclonal antibodies (mAbs) are the most significant class of new therapeutic agents to be developed in recent times, 2016 marking the 30th anniversary of the first FDA approval of a mAb for human use.<sup>3</sup> Human IgGs contain two *N*-linked glycans attached to the two Asn-297 residues of the Fc region. Numerous studies have demonstrated that Fc-receptor mediated effector functions of IgGs (*e.g.* complement-dependent cytotoxicity [CDC] and antibody-dependent cellular

<sup>a</sup> Department of Chemistry, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand. E-mail: [antony.fairbanks@canterbury.ac.nz](mailto:antony.fairbanks@canterbury.ac.nz)  
<sup>b</sup> Biomolecular Interaction Centre, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand



Antony J. Fairbanks

Department at UC from 2010–2014 throughout the Christchurch earthquakes. He has long-standing interests in synthetic carbohydrate chemistry.



cytotoxicity, [ADCC]), are highly dependent on the structure of attached glycans.<sup>4</sup> Thus correct glycosylation of mAbs is important for therapeutic efficacy, and correspondingly significant efforts are being made to optimise the glycosylation patterns on mAbs.<sup>5</sup>

N-Linked glycosylation is also of significance for many mammalian enzymes; for example all lysosomal enzymes are *N*-glycosylated, and their targeting to the lysosome relies on the attachment of mannose-6-phosphate residues and transport mediated by the mannose-6-phosphate receptors (M6PR's).<sup>6</sup> Glycosylation therefore plays an important role in the efficacy of treatment of the lysosomal storage disorders (LSDs) using recombinant glycoproteins as enzyme replacement therapies.<sup>7</sup>

The biosynthesis of the glycan portions of *N*-linked glycoproteins is not under direct genetic control; rather a complex sequence of enzymatically catalysed events leads to the formation of diverse *N*-glycan structures. Firstly the enzyme oligosaccharyl transferase (OST) mediates the transfer of a 14-mer dolichol phosphate bound oligosaccharide ( $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ ) to particular

asparagine (Asn) residues (Asn-X-Ser/The)<sup>8</sup> as the protein is assembled. Then, sequential trimming of the oligosaccharide chain(s) occurs, and subsequently the action of several other glycosidases and glycosyl transferases leads to the production of glycoproteins bearing a variety of *N*-glycan structures, commonly classified into three groups; high mannose, hybrid, and complex (Fig. 1). Importantly all of these oligosaccharides contain the same core pentasaccharide unit ( $\text{Man}_3\text{GlcNAc}_2$ ) linked to Asn residues. Due to the action of numerous enzymes in competing processes, glycoproteins are invariably produced as inseparable heterogeneous mixtures of different 'glycoforms'; proteins in which different oligosaccharide structures are linked to the same peptide chain. Therefore all recombinant therapeutic glycoproteins, including mAbs, are currently manufactured and administered as mixtures of materials in which different oligosaccharide structures are linked to the same peptide chain.<sup>9</sup>

Access to pure single glycoforms of glycopeptides and glycoproteins has consequently become a major scientific objective.<sup>10</sup>

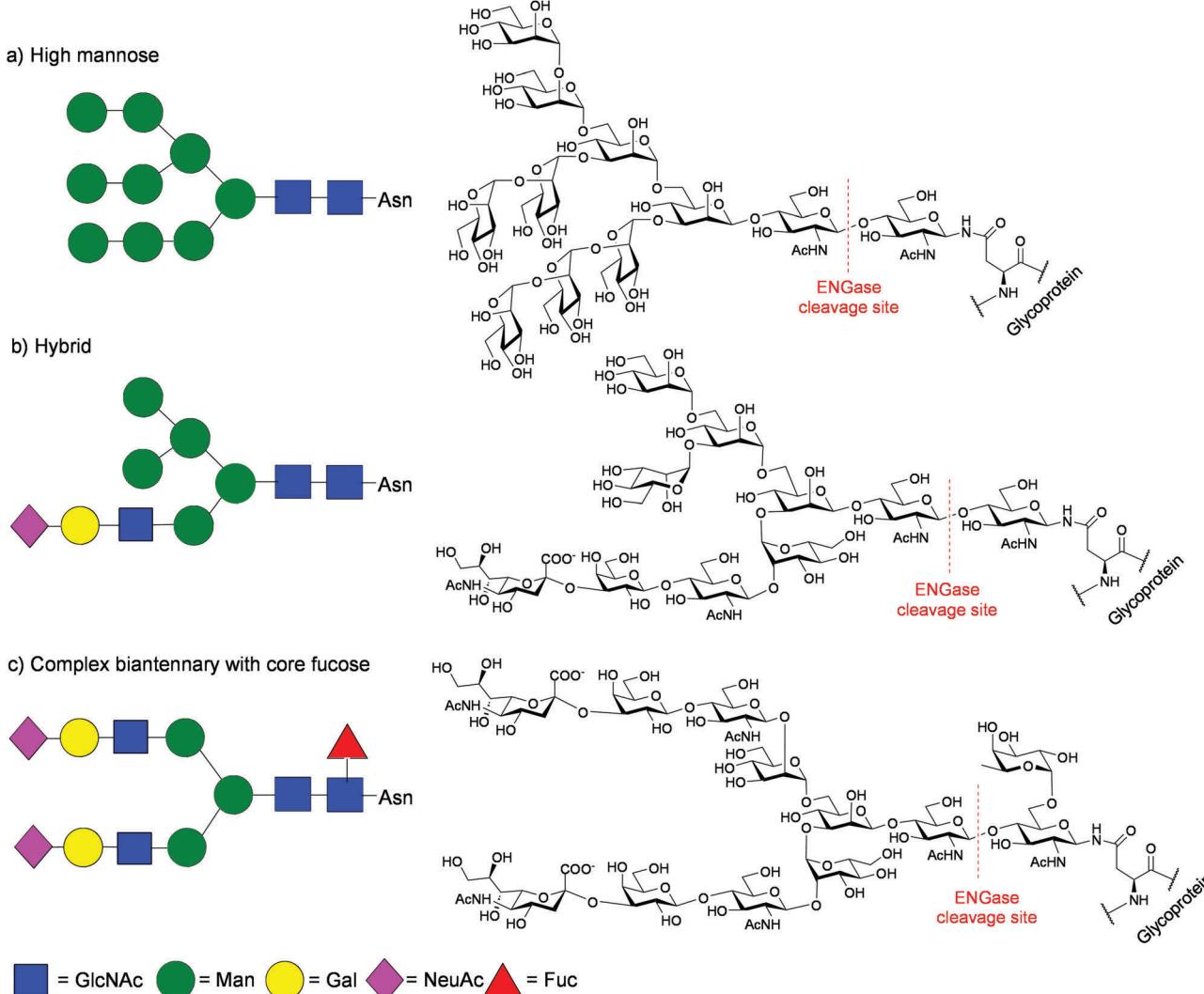


Fig. 1 Chemical structures and Glycobiology representations of typical *N*-glycans showing the ENGase cleavage site: (a) high mannose; (b) hybrid; (c) complex biantennary with core fucose.



Amongst the numerous approaches<sup>11</sup> that have been adopted to access glycopeptides and glycoproteins in homogeneous form, this review will focus on synthetic applications of *endo*- $\beta$ -N-acetyl-glucosaminidase (ENGase) enzymes.

## 2. The *endo*- $\beta$ -N-acetylglucosaminidases (ENGases)

### 2.1 Hydrolytic activity

The *endo*- $\beta$ -N-acetylglucosaminidases or ENGases (sometimes also referred to as endohexosaminidases, of which they are a sub-set) are an enzyme class (EC 3.2.1.96), first identified in the early 1970's, that are produced by a range of organisms, ranging from bacteria, through fungi, to higher order species, including humans. Interest in them initially arose because of their ability to act on glycoproteins, hydrolysing attached *N*-glycans. All of the ENGases so far identified have been classified by sequence analysis into one of two of the Carbohydrate Active enZYme database (CAZY)<sup>12</sup> families<sup>13</sup> of the glycoside hydrolases, either family GH18 or family GH85, and all have similar  $(\beta/\alpha)_8$ -TIM-barrel structures consisting of eight  $\alpha$ -helices and eight parallel  $\beta$ -sheets. Fig. 2 shows two views of an X-ray crystal structure<sup>14</sup> of Endo A, a family GH85 ENGase. The major domain comprises the eight  $\alpha$ -helices (red) and eight  $\beta$ -sheets (yellow) in a TIM-barrel fold; two other domains also comprise  $\beta$ -sheets.

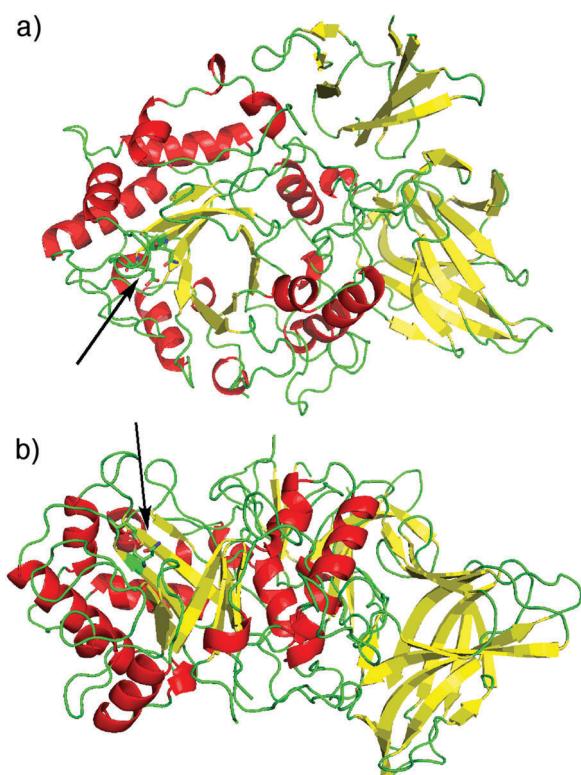


Fig. 2 Two different views of the X-ray crystal structure<sup>14</sup> of WT Endo A (PDB: 3FHA).  $\alpha$ -Helices are shown in red,  $\beta$ -sheets in yellow. Domain 1 (residues 2–350) displays a typical  $(\beta/\alpha)_8$ -TIM-barrel fold. An arrow shows the position of the catalytic site.

In their action as *endo*-glycosidases the ENGases specifically cleave the diacetylchitobiose core [GlcNAc  $\beta$ (1–4)GlcNAc] of *N*-linked glycans between the two *N*-acetylglucosamine (GlcNAc) residues.<sup>15</sup> When applied to an *N*-linked glycoprotein/glycopeptide the two products of hydrolysis are therefore truncated *N*-glycan oligosaccharides and a protein/peptide which possesses only a single GlcNAc residue at any *N*-glycosylation site. Their application as biocatalysts to achieve the reverse process, namely the attachment of *N*-glycans to GlcNAc residues, represents a highly effective convergent method to access homogenous glycopeptides and glycoproteins, which comprise *N*-linked glycans of optimal structure for desired biological activity.<sup>16</sup> Fig. 3 demonstrates the principles of glycoprotein remodelling and convergent glycopeptide synthesis using ENGases.

The two-step procedure for *in vitro* glycoprotein remodelling (Fig. 3a) firstly entails ENGase catalysed hydrolysis of the heterogeneous mixture of protein glycoforms that is invariably produced by expression, so that *N*-glycans are 'trimmed' back to a single GlcNAc residue at *N*-linked glycosylation sites. In a second step, a defined glycan is attached to this GlcNAc 'handle', again in a process catalysed by an ENGase, but one which is acting in the reverse sense. This synthetic process may be achieved either by a classical 'transglycosylation' reaction employing an oligosaccharide comprising two GlcNAc residues to act as the source of the glycan to be added ( $X = \text{GlcNAc}$ ), or alternatively by a 'glycosylation' reaction using an activated donor that possesses only a single GlcNAc residue which is activated at the reducing terminus ( $X = \text{leaving group}$ ). In a similar fashion ENGases can be applied to the glycosylation of synthetic peptides assembled by solid-phase synthesis that contain GlcNAc 'handles' attached to particular Asn residues (Fig. 3b). Thus, ENGase-catalysed attachment of an *N*-glycan, again either by 'transglycosylation' or 'glycosylation', represents an efficient convergent method to access defined glycopeptides.

Although all of the ENGases carry out the same hydrolytic reaction, they have different tolerances as to the precise structure of the *N*-glycans that they can hydrolyse. A summary of some of the most well studied ENGases to date is given in Table 1, including their hydrolytic specificities and their active site residues. The ENGases are all retaining glycosidases that hydrolyse substrates *via* a two-step mechanism involving general acid/base catalysis (Fig. 4). The family GH18 enzymes (Fig. 4a) all contain two carboxylic acid residues in the active site, a glutamic acid and an aspartic acid. In contrast the family GH85 enzymes (Fig. 4b) contain only one carboxylic acid residue, a glutamic acid; the other key catalytic residue is an asparagine. The pH/rate dependency for the hydrolytic activity of the ENGases is typically a bell-shaped curve centred around pH 5.0–8.0, the precise pH optimum depending on the enzyme.<sup>17–19</sup> The upper and lower limits of the curve generally correspond with the kinetic  $pK_a$  values of the two catalytic carboxylic acid residues in the enzyme active site; for example the pH optimum for Endo H is pH 5.5, and its hydrolytic activity drops rapidly as the pH is either increased from pH 6.0 to pH 8.0, or is decreased from pH 5.0 to pH 4.0.<sup>18</sup> Although the family GH85 ENGases actually only possess a single catalytic carboxylic acid residue in their active



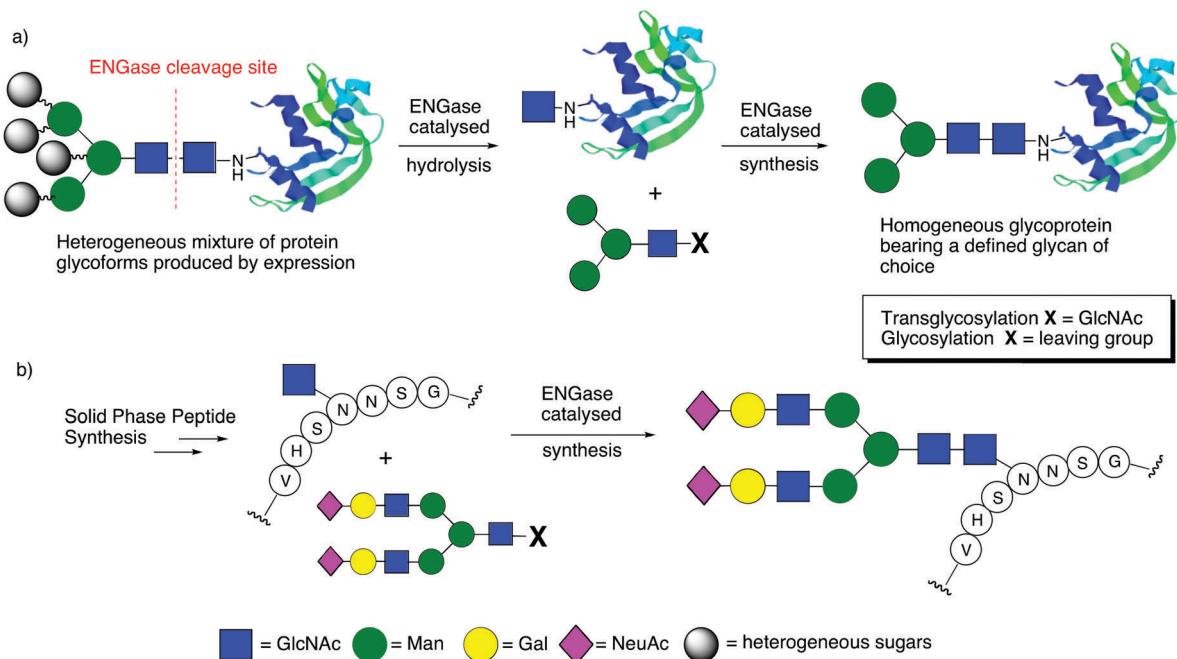


Fig. 3 The use of ENGases to catalyse: (a) glycoprotein remodelling; (b) convergent glycopeptide synthesis.

Table 1 A selection of the most well-studied ENGases

Enzyme	Source	GH family	Glycan structure hydrolysed	Glycans with core fucose hydrolysed?	Catalytic residues	Glycosynthase mutants
Endo-A	<i>Arthrobacter protophormiae</i>	85	High mannose/hybrid	No	E173 <sup>a</sup> N171 <sup>a</sup>	N171A <sup>a</sup> E173Q <sup>a</sup> E173H <sup>a</sup>
Endo-M	<i>Mucor hiemalis</i>	85	High mannose/hybrid and biantennary complex	No	E177 N175	N175A N175Q
Endo-D	<i>Streptococcus pneumoniae</i>	85	Truncated core structures only	Yes	E324 N322	N322A N322Q
Endo-CC1	<i>Coprinopsis cinerea</i>	85	High mannose/hybrid and biantennary complex	—	E182 N180	N180H N180Q
Endo-CC2	<i>Coprinopsis cinerea</i>	85	High mannose/hybrid and biantennary complex	—	E188 N186	—
Endo Om	<i>Ogataea minuta</i>	85	High mannose/hybrid and biantennary complex	No	E196 N194	—
Endo-CE	<i>Caenorhabditis elegans</i>	85	High mannose/hybrid	No	E154 N152	—
Endo H	<i>Streptomyces plicatus (griseus)</i>	18	High mannose/hybrid	Yes	E132 <sup>a</sup> D130 <sup>a</sup>	—
Endo F <sub>1</sub>	<i>Elizabethkingia meningoseptica</i>	18	High mannose/hybrid	No	E132 <sup>a</sup> D130 <sup>a</sup>	—
Endo F <sub>2</sub>	<i>Elizabethkingia meningoseptica</i>	18	Biantennary complex	Yes	E126 <sup>a</sup> D124 <sup>a</sup>	—
Endo F <sub>3</sub>	<i>Elizabethkingia meningoseptica</i>	18	Bi-/triantennary complex	Yes	E128 <sup>a</sup> D126 <sup>a</sup>	D126A <sup>a</sup> D126Q <sup>a</sup>
Endo S	<i>Streptococcus pyogenes</i>	18	Complex (IgG Fc specific)	Yes	E235 D233	D233A D233Q
Endo S2	<i>Streptococcus pyogenes (M49)</i>	18	High mannose/hybrid and biantennary complex	Yes	E186 D184	D184M D184Q

<sup>a</sup> Amino acid numbering refers to the peptide chain after cleavage of an N-terminal signal peptide.

sites, similar pH/rate profiles have generally been observed for their hydrolytic activity.<sup>17,19</sup>

Regardless of whether the active site contains one or two carboxylic acids, the hydrolytic mechanism<sup>20</sup> catalysed by the

ENGases involves neighbouring group participation of the 2-acetamide of the second GlcNAc residue. A glycosyl oxazoline (or oxazolinium ion) is therefore implicated as a high-energy intermediate in the ENGase mediated hydrolytic process. A key

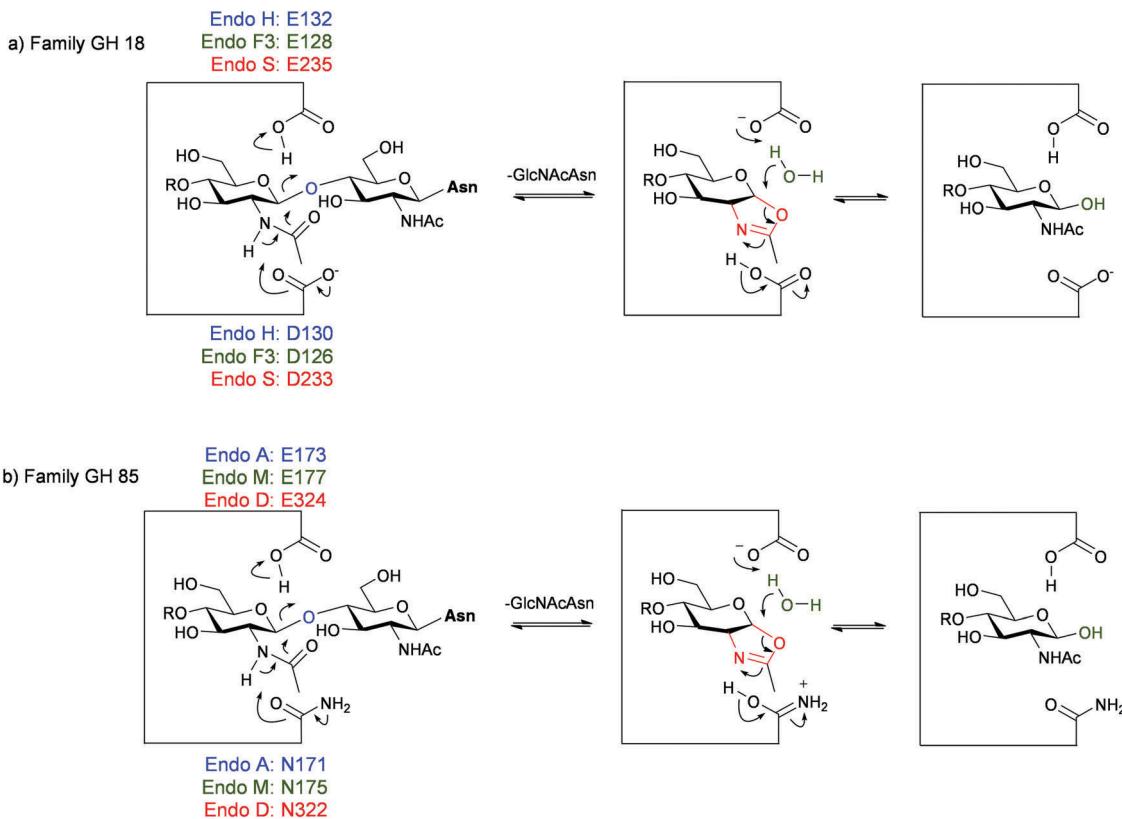
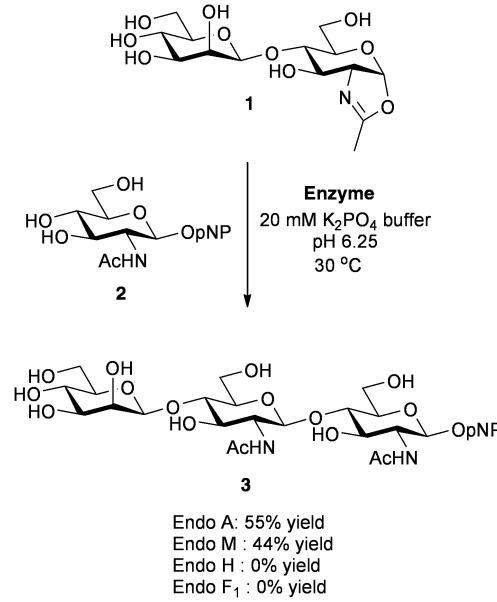


Fig. 4 Hydrolytic mechanism of the ENGases showing key catalytic residues for some enzymes from: (a) family GH18; (b) family GH85.

breakthrough in the field was the realisation that *N*-glycan oxazolines may act as substrates for these enzymes; in particular that they may be intercepted by a GlcNAc residue attached to a peptide or protein in a synthetic reaction that is formally the reverse of hydrolysis.

In 2001 Shoda and co-workers published their first seminal report in the field<sup>21</sup> in which they demonstrated that disaccharide oxazoline **1** could be used as an activated glycosyl donor substrate for some of the ENGases (Scheme 1). In particular Endo A and Endo M (both family GH85 enzymes) catalysed the glycosylation of GlcNAcOpNP **2** by the disaccharide oxazoline **1**, to produce the corresponding trisaccharide **3**. However neither Endo H nor Endo F<sub>1</sub> (both family GH18 enzymes), were found to catalyse this synthetic reaction. Besides paving the way for the use of more extended oxazolines as privileged donor substrates for ENGase catalysed glycosylation, these initial results also seemed to imply that perhaps only family GH85 enzymes would be useful synthetically; subsequently this led to an initial focus on the family GH85 enzymes. An additional advantage of the use of oxazolines as donors is that in many cases wild-type (WT) ENGases are able to catalyse synthetic reactions using either truncated or 'non-natural' oxazolines, and often the products of these reactions are not hydrolytic substrates for that enzyme. This was indeed the case in the first example by Shoda (Scheme 1), *i.e.* trisaccharide **3** was not a hydrolytic substrate for either Endo A or Endo M. Thus, in these cases WT ENGases can catalyse irreversible synthetic reactions using oxazoline donors.



Scheme 1 Shoda's demonstration of the first use of oxazolines as glycosyl donors for ENGase catalysed glycosylation.<sup>21</sup>

Since this first report oxazolines have become the donors of choice for the ENGase-catalysed synthesis.<sup>22</sup> Moreover, following further important developments made by Shoda,<sup>23</sup> they may now be readily synthesised directly at the reducing end of complex oligosaccharides without the need for any protecting groups.<sup>24</sup>

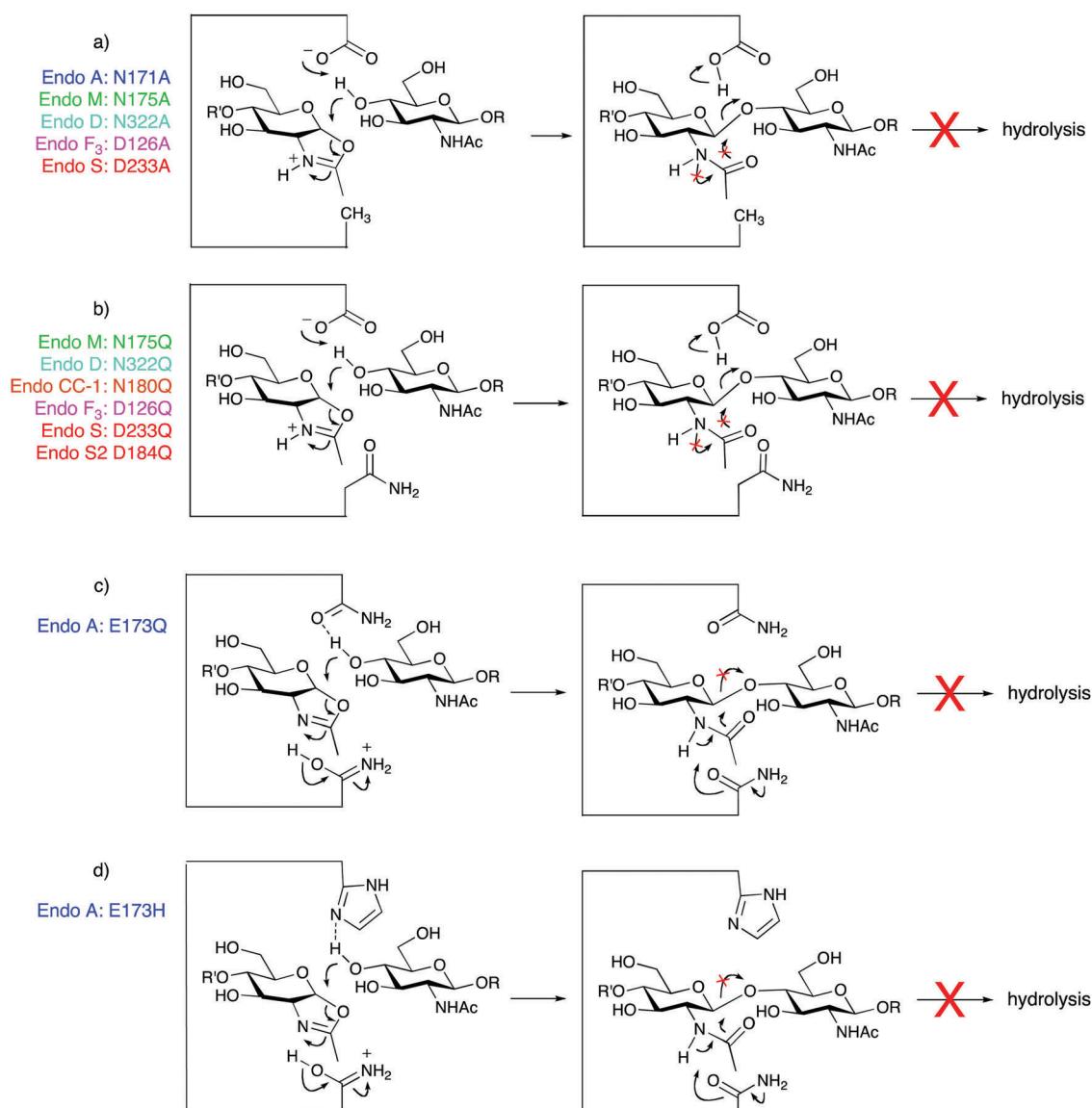


Besides the use of oxazolines, a variety of other factors can increase glycosylation efficiency, for example the use of variable amounts of organic co-solvents<sup>25,26</sup> and/or control of reaction pH;<sup>27</sup> modulation of both of these can decrease the rates of competitive hydrolytic reactions.

## 2.2 Mutant enzymes

The use of glycosidases for synthetic purposes, either in 'true' transglycosylation reactions, or when catalysing 'glycosylation' using activated glycosyl donors, such as oxazolines, is inevitably limited by the rate at which the enzyme hydrolyses the product. Although in certain cases when using truncated or 'non-natural' oxazolines as donors, the products of glycosylation are not hydrolysed by the ENGase, in general they are. Indeed when using full-length glycans the rate of product hydrolysis can

sometimes far exceed the rate of the synthetic reaction, meaning that synthesis is unachievable. A solution to this problem is to access mutant 'glycosynthase' enzymes,<sup>28</sup> as originally developed by Withers<sup>29</sup> and Planas,<sup>30</sup> by site-directed mutagenesis. In this original work, the catalytic nucleophile of retaining glycosidases was replaced with a non-participating residue, for example alanine; these mutant enzymes could process activated donors (such as glycosyl fluorides) of opposite anomeric configuration to the natural substrate, but were incapable of hydrolysing the reaction product. In the case of the ENGases, the catalytic nucleophile is the 2-acetamido group on the GlcNAc residue undergoing hydrolysis; therefore a direct analogy is not possible. Two approaches have been developed to accessing glycosynthase mutants of ENGases, namely mutation of one of the two key catalytic site residues. By far the most common approach has



**Fig. 5** Design of mutant glycosynthases of the ENGases: (a) mutations of the catalytic base (family GH18: D, family GH85: N) to alanine; (b) mutations of the catalytic base (family GH18: D, family GH85: N) to glutamine; (c) mutation of the catalytic acid (E) to glutamine; (d) mutation of the catalytic acid (E) to histidine.

been to mutate either the asparagine (family GH85) or aspartate (family GH18) residue that stabilises/deprotonates the oxazolinium intermediate during the hydrolytic reaction; *i.e.* the catalytic residue that acts as the base. The first glycosynthase mutant of any ENGase, reported by Wang and Yamamoto,<sup>31</sup> was a mutant of Endo M in which residue N175 was replaced by alanine (Fig. 5a), and this was a reasonably effective synthase enzyme. Further studies in which all other possible mutations were made at this site revealed that the asparagine to glutamine mutation was the most effective in terms of synthetic efficiency (Fig. 5b); this E175Q Endo M mutant<sup>32</sup> is now available commercially. Wang has subsequently performed the analogous mutations on a variety of other family GH85 enzymes including Endo A and Endo D, to produce a variety of synthetically useful synthase mutants. Moreover Wang extended this approach to some of the family GH18 enzymes, and has shown that replacement of the aspartic acid active site residue of both Endo F<sub>3</sub> and Endo S by either alanine (Fig. 4a) or glutamine (Fig. 5b), also produces extremely useful glycosynthase mutants. Whilst this N/D to A or Q mutation is the most reliable in terms of the production of an active glycosynthase enzyme, very recently two other mutations of the catalytic base residue have been reported (Endo CC1 N180H<sup>33</sup> and Endo S2 D184M<sup>34</sup>).

An alternative approach, developed by Fairbanks, is to mutate the catalytic glutamic acid residue (conserved for GH18 and GH85 enzymes) that acts as a general acid to protonate the anomeric oxygen during the hydrolytic process.<sup>35</sup> Adopting this approach Fairbanks made the first two glycosynthase mutants of Endo A, E173Q (Fig. 5c) and E173H (Fig. 5d); potential hydrogen

bond acceptors were chosen as the replacements for the glutamic acid as this residue also normally acts as a general base to direct the incoming nucleophile during the second step of the hydrolytic reaction. The E173Q and E173H Endo A mutants<sup>35</sup> were both capable of irreversibly catalysing glycosylation of GlcNAc acceptors, including de-glycosylated RNase B, with a tetrasaccharide oxazoline. However Yamamoto reported that the analogous Endo M mutants (E177Q and E177H) were not synthetically active.<sup>32</sup>

Further site directed mutagenesis studies have also revealed that mutations to other active residues can increase the synthetic efficiency of these enzymes, particularly by decreasing the rate of direct competitive oxazoline hydrolysis. For example the Y217F mutant<sup>32</sup> of Endo M and the Y205F mutant<sup>14</sup> of Endo A both showed increased rates of synthetic glycosylation; however such mutants are actually still capable of hydrolysing the product.

The total number of mutant ENGases that have now been made is increasing rapidly (*vide infra*), as more and more enzymes are cloned and expressed in a drive to expand synthetic capability. A summary of some of the most useful synthetic mutants produced so far is provided in Table 2. Finally a key point worth re-iterating is that these mutant enzymes will not catalyse transglycosylation using full-length *N*-glycans as substrates, because the first step of 'transglycosylation' necessarily involves hydrolysis of the full length *N*-glycan. In many previous papers the term transglycosylation has also (rather erroneously) been used to refer to ENGase-catalysed reactions using oxazolines as donors. The ability of mutant ENGases to catalyse one and not the other highlights the important, yet mostly overlooked, difference between the two.

Table 2 Summary of some of the most synthetically useful mutant ENGases

Mutant enzyme	Donor oxazolines structures processed	Fucosylation of the acceptor tolerated (or required)?	Protein and IgG applicability	Comments
Endo A N171A	High mannose/hybrid	No	Protein: yes IgG: Fc only	Will also transfer truncated glycans with bisecting GlcNAc Tolerates glycan phosphorylation
Endo A E173H	High mannose/hybrid	No	Protein: yes IgG: not known	High mannose only up to Man <sub>6</sub> structures
Endo M N175Q	High mannose/hybrid Biantennary complex	No	Protein: yes IgG: no	Commercially available Will not glycosylate the Fc region of IgGs
Endo M N175Q/W251N	High mannose/hybrid Biantennary complex	Yes	Protein: not known IgG: not known	Will act on IgG Fc-derived peptides
Endo D N322Q	Core structure only	Yes	Protein: yes IgG: yes	Only up to pentasaccharide
Endo F <sub>3</sub> D126A	Biantennary complex Triantennary complex	Required	Protein: not known IgG: yes	Will transfer triantennary structures
Endo S D233A Endo S D233Q	Core structure Biantennary complex	Yes	Protein: no IgG: yes	Endo S is highly IgG specific Will not transfer high mannose oxazolines
Endo S2 D184M	High mannose/hybrid Biantennary complex	Yes	Protein: not known IgG: yes	Transfers all three classes of glycan to IgGs, both with and without core fucose
Endo CC1 N180H	Biantennary complex	Not known	Protein: yes IgG: not known	Reportedly more thermostable than Endo M Not yet tested with high mannose oxazolines



Endo A	163	YGF DGF WFI NQ ETE GADE --- GTAEAMQAFLV - YLQE QK P -- EGM HIM WYDS M I - DT G	212
Endo M	167	YGF DGF WL FNI ECE FF PFP TNP KFK AELA KFLH - YFKE KLH NEI PGS QLW I WYDS M I - NEG	224
Endo D	314	YGY DGY F I NQ ETT GDLV --- KPL GEK M RQ FML - YSKEY AAK VNH PI KYS WYD AMT - YN Y	367
Endo CC-1	172	RGF DGF WL L NVE I GL QGG S E QAR GL --- AAW VALL Q L Q E V L K V G P H G L V I WYD S V I - VR G	226
Endo Om	186	YGF DGY LL N I ETT F SNT K --- I AAD LEP PFA E - QLK SG L H C L D S K N E L I WYD S Y V F P A N	239
Endo CE	144	FGF DGF WL L N I E N K I D A G --- K I D L L I Q F C S - S L T E K S K K M N E N S K V I WYD S V L - D N G	195
		* : * : * : * * : . . : . : *** :	
Endo A	213	AIA WQ N HLT DRN K M Y L Q - NG S T R V A D S M F L N F W W R D Q R Q S N E L - - - - -	254
Endo M	225	EI H WQ N Q L T W K N E L L F F K - - - - - N T D G I F L N Y W W K K E Y P E M A R - - - - - R	262
Endo D	368	G RY H Q D G L G E Y N Q F M Q P E G D K V P A D N F A N F N W D K A N G Y D T Y - - - - - A	411
Endo CC-1	227	D L W Q D R L N A F L L F F L - - - - - N S S G I F T N Y W W Y N D A P Q K Q I D F L S R V D P N L T Q T A E P	280
Endo Om	240	K V S Y T N G V T E S N Y N F F S - - - - - L S D A F F S N Y W W N I K N L Q E N - - - - - I	276
Endo CE	196	K L H W Q N E L N E M N R K F Y D - - - - - A C D A I Y L N Y N W K D K E L L - - - - - S	291
		* : * : * : * * : . . : . : *** :	

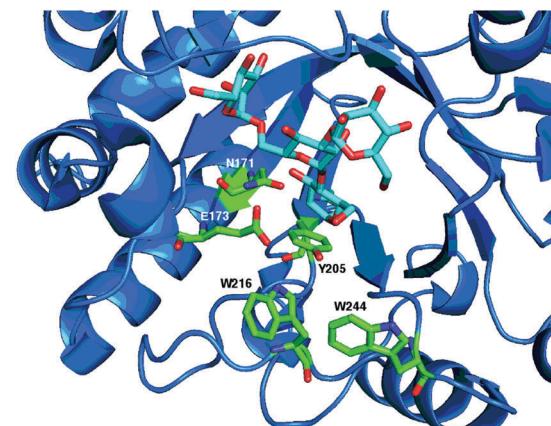
**Fig. 6** Amino acid alignment of the catalytic region of some family GH85 ENGases, showing conserved and key active site residues. In the case of Endo A the amino acid numbering refers to the peptide chain after signal peptide cleavage. Catalytic site residues are highlighted in pink; conserved residues across at least three enzymes are highlighted in grey. The symbol † shows the position of aromatic amino acids demonstrated as important for (trans)glycosylation activity for some enzymes. An asterisk indicates positions which have a single, fully conserved residue; a colon or a full stop indicates conservation between groups of strongly, or weakly, similar properties respectively.

### 2.3 Family GH85 enzymes

According to the CAZy database,<sup>12</sup> family GH85 currently contains 353 members, the vast majority of which come from bacterial sources. So far 11 enzymes have been characterised, and crystal structures for two of them have been published and deposited in the Protein Data Bank (PDB), confirming their  $(\beta/\alpha)_8$ -TIM-barrel structures; for Endo A (PDB accession codes WT: 3FHQ, and an E173Q mutant: 2VTF), and Endo D (PDB: 2W91 and 2W92). An amino acid alignment of the catalytic region of some of the most well-studied GH85 enzymes is shown in Fig. 6, showing conserved and key active site residues. Note there is some inconsistency in the amino acid numbering used in the literature; for GH85 enzymes uniquely for Endo A the numbering refers to the amino acid position after cleavage of an N-terminal signal peptide.

A crystal structure<sup>14</sup> showing the active site of Endo A in complex with a tetrassaccharide thiazoline, highlighting the key catalytic residues (N171 and E173) and aromatic residues (Y205, W216 and W244) important for (trans)glycosylation activity is shown in Fig. 7.

**Endo A.** Endo A, is a 621 amino acid (80 kDa) protein (plus a 24 amino acid signal peptide) produced by the bacterium *Arthrobacter protophormiae*.<sup>36</sup> The purified enzyme, which displays specific activity for high mannose and hybrid oligosaccharides, was used in a number of early reports of transglycosylation activity using full-length glycans (*i.e.* those possessing two GlcNAc residues) as donors,<sup>37</sup> including for the production of remodelled<sup>38</sup> and neo-glycoproteins.<sup>39</sup> The Endo A gene was first cloned and recombinant enzyme expressed by Takegeawa.<sup>40</sup> Besides identifying the key catalytic site residues as N171 and E173, early mutagenic studies also revealed the importance of W216 both for hydrolytic and transglycosylation activities.<sup>41</sup> Studies on transglycosylation to different acceptors revealed structural requirements of the enzyme; namely that the acceptor be in a  ${}^4\text{C}_1$  conformation and possess equatorial OH-groups at positions-3 and 4,<sup>42</sup> results that were supported by later work<sup>43</sup> revealing the inability of Endo A to glycosylate GlcNAc acceptors which possessed either fucose or a protecting group at position-3. In their key report Shoda and collaborators<sup>21</sup> revealed that Endo A would process a disaccharide oxazoline as a donor substrate. Wang then



**Fig. 7** X-Ray crystal structure<sup>14</sup> of WT Endo A (PDB: 3FHQ) in complex with a tetrasaccharide (Man<sub>3</sub>GlcNAc) thiazoline, showing catalytic active site residues (N171 and E173) and aromatic residues (Y205, W216 and W244) important for (trans)glycosylation activity.

reported the first application of Endo A (and simultaneously Endo M) with more extended oxazolines,<sup>44</sup> and applied this methodology for the synthesis of a variety of glycopeptides,<sup>45</sup> and later a glycoprotein.<sup>46</sup> Wang,<sup>47</sup> and Fairbanks<sup>48</sup> also probed the use of structurally modified *N*-glycan oxazolines as substrates for Endo A catalysed glycosylation. The PDB currently contains three Endo A structures: two for WT Endo (PDB: 3HFA [Fig. 2] and 3FHQ [Fig. 7])<sup>14</sup> and the E173Q mutant (PDB: 2VTF).<sup>49</sup> These structures provided confirmation of the identity and spatial arrangement of essential active site amino acids, and the proximity of hydrophobic residues Y205, W216, and W244, all identified as important for glycosylation/transglycosylation activity. The first glycosynthase mutants of Endo A (E173H and E173Q) were produced by Fairbanks;<sup>35</sup> these contrast with the other family GH85 and GH18 ENGase mutants reported to date in that it was the key catalytic glutamic acid that was replaced. Subsequently other Endo A mutants, most notably N171A, were produced by Wang, which have shown more wide-ranging synthetic capabilities.<sup>50</sup>

**Endo M.** The hydrolytic activity of the family GH85 ENGase Endo M, a 744 amino acid (85 kDa) protein (including signal

peptide), was first observed in a culture fluid from a soil sample containing the fungus *Mucor hiemalis*.<sup>51</sup> Purification,<sup>17</sup> and in depth study of substrate specificities<sup>52</sup> revealed that Endo M could cleave sialylated complex *N*-glycans in addition to high mannose and hybrid oligosaccharides, although hydrolysis activity on high-mannose oligosaccharides was significantly higher than on the complex and hybrid glycans. Early studies also revealed that Endo M was able to<sup>53</sup> catalyse transglycosylation of peptides bearing GlcNAc residues using full-length Asn-linked glycans, including sialylated complex bi-antennary glycans<sup>54</sup> albeit with low efficiencies (~20%). Yamamoto and collaborators applied this transglycosylation activity to the synthesis of a variety of glycopeptides<sup>55,56</sup> in which the GlcNAc peptide acceptor was assembled by solid phase synthesis, though the yields for transglycosylation were typically poor (e.g. ~8.5%).<sup>57</sup> Expression of Endo M<sup>58</sup> in *Candida boidinii* (a protease-deficient [pep4] strain) and kinetic characterisation of recombinant enzyme was accompanied by site directed mutagenesis experiments, which revealed residues E177 and W228 to be essential and important for hydrolytic and transglycosylation activities, respectively. Endo M then became the first ENGase to be commercially available, following which recombinant enzyme was employed<sup>59</sup> for the synthesis of glycopeptides by transglycosylation using full-length Asn-linked *N*-glycans.

As for Endo A, in their seminal paper Shoda and collaborators<sup>21</sup> showed that a disaccharide oxazoline was a good donor substrate for Endo M. Fairbanks<sup>60</sup> then went on to show that larger *N*-glycan oxazolines<sup>61,62</sup> were also donor substrates for Endo M, and additionally reported that non-natural oxazolines,<sup>63</sup> for example those containing glucose rather than mannose attached to the GlcNAc, could act as donors for Endo M, to give products that were not hydrolytic substrates for the enzyme. The issue of competitive product hydrolysis was completely overcome by the development of the first glycosynthase mutant of any ENGase, namely the Endo M N175A mutant by Wang and Yamamoto.<sup>31</sup> Additionally an Endo M Y217F mutant also showed enhanced glycosylation activity, though it was still capable of catalysing product hydrolysis. Subsequent studies in which all other possible mutations were made at position 175 revealed that the N175Q mutant was the most effective in terms of synthetic efficiency; a double N175Q/Y217F mutant was not more effective. This Endo M N175Q mutant<sup>32</sup> has so far proven to be the most widely applicable and useful glycosynthase of an ENGase, and is now commercially available from Tokyo Chemical Industry Co. Ltd (TCI). Its applications in synthesis will be discussed later.

Inazu mapped out the acceptor requirements<sup>64</sup> for transglycosylation catalysed by WT Endo M, confirming that free hydroxyls at positions-4 and 6 are needed on a GlcNAc acceptor (mannose and glucose also acted as acceptors), but that neither the ring oxygen nor the substituents at positions-1 or 2 were important. This minimum required structure corresponds to a 1,3-diol structure with secondary and primary hydroxyl groups; Inazu additionally showed that an acyclic diol would also act as an acceptor, but not glycerol, which possesses two primary hydroxyl groups.

This requirement for a free primary hydroxyl group at position-6 of GlcNAc means Endo M will not act, either

synthetically or hydrolytically, on glycans that have core fucose (*i.e.* with fucose attached to position-6 of the innermost GlcNAc), limiting the synthetic potential of its glycosynthases. Very recently Yamamoto<sup>65</sup> performed mutagenic studies to find a mutant enzyme that would cleave fucosylated glycans and discovered that the W251N mutation, asparagine being the amino acid at the corresponding position (N413) in Endo D, a family GH85 ENGase which can cleave fucosylated glycans, increased hydrolytic activity on fucosylated glycans. The incorporation of the double mutation N175Q/W251N<sup>65</sup> then produced a glycosynthase Endo M which could glycosylate (~20% yield) a Fuc $\alpha$ (1-6)GlcNAc-biotin conjugate using a complex biantennary oxazoline.

**Endo D.** The hydrolytic activity of Endo D (a 1646 amino acid, 182 kDa protein), first observed in the culture fluid of *Streptococcus pneumoniae* (*Diplococcus pneumoniae* subsequently being renamed *Streptococcus pneumoniae*) by Muramatsu in 1971,<sup>66</sup> was in fact the first ENGase activity to be disclosed. Endo D was subsequently isolated and purified from *S. pneumoniae*,<sup>67</sup> and its specific hydrolytic activity was investigated with respect to substrate structure. Much later it was cloned and expressed,<sup>68</sup> and truncation and mutational studies revealed the key catalytic active site residues.<sup>69</sup> Vocadlo and Boraston solved the first structure<sup>19</sup> of the catalytic domain of Endo D (SpGH85) (PDB: 2W91 and 2W92), which was the first structure of a GH85 enzyme to be published, providing structural insight into the catalytic site and oligosaccharide binding motifs. More recently a new structure has been deposited in the PBD (PDB: 3GDB).

Particular interest in the enzyme arose from its ability to hydrolyse substrates that are core fucosylated, a common modification of *N*-glycans attached to IgG's. The first demonstration of the ability of Endo D to process oxazolines as donors was reported by Fairbanks,<sup>70</sup> although significant limitations of utility of the wild type enzyme were revealed due to the high rates of direct oxazoline hydrolysis. Additionally the fact that Endo D only hydrolyses very truncated *N*-glycans precludes its application to larger structures.

Subsequently Wang produced two glycosynthase mutants of Endo D (N322A and N322Q),<sup>71</sup> which, although they displayed enhanced synthetic efficiency including the ability to transfer a tetrasaccharide oxazoline to an IgG-Fc fragment carrying a Fuc $\alpha$ (1,6)GlcNAc moiety, were unable to accept more extended oxazolines as substrates. Mutations of other residues (E324Q, Y360F and H371W) were also made by analogy with the Endo A/M sequences and corresponding mutations that had improved the glycosylation activity of those enzymes. However, for Endo D these were all found to be less effective than the N322A and N322Q mutants.

**Endo CC.** Recently two ENGases from the basidiomycete fungus *Coprinopsis cinerea*, termed Endo-CC1 and Endo-CC2, were reported,<sup>72</sup> that bear significant sequence similarities (46% and 40% respectively) to Endo M, and which are reportedly more thermostable and easier to express. Like Endo M, both enzymes displayed hydrolytic activity against high mannose and biantennary complex *N*-glycans, but neither sialic acid terminated tri- or tetra-antennary glycans, nor those with a bisecting GlcNAc (*i.e.* with a GlcNAc attached to position-4 of the central branching mannose) were cleaved. As Endo-CC1 (a 787 amino



acid protein, including a 74 amino acid N-terminal signal peptide) was easier to express it was examined in more detail. The WT enzyme was actually found to be incapable of catalysing the transglycosylation of de-glycosylated RNaseB using a full-length Asn-linked biantennary complex sialylglycopeptide (SGP) as the donor. However, site directed mutagenesis, in which the catalytic site asparagine N180 was replaced by all the other 19 amino acids, revealed that both N180H and N180Q mutants were capable of catalysing the transglycosylation of RNase B using SGP as the donor. It was stated that the latter enzyme also hydrolysed the product, which is logical since some hydrolytic activity is required for transglycosylation activity using full-length *N*-glycans donors (the two GlcNAcs must necessarily be cleaved in the first step). The glycosylation activity of the Endo CC1 N180H mutant using a complex biantennary glycan oxazoline and de-glycosylated RNase B was very recently reported.<sup>33</sup> Herein it was claimed that significantly reduced amounts of both oxazoline and enzyme were required to achieve the same yield as compared to when the same reaction was catalysed by the N175Q mutant of Endo M.

**Endo Om.** Database searches identified homologous sequences in four yeast strains, in which ENGase activity was then displayed in crude cell extracts. The gene from the yeast *Ogataea minuta* was amplified, sequenced, cloned, and expressed to produce a family GH85 ENGase called Endo Om.<sup>73</sup> Recombinant Endo Om, a 722 amino acid (87 kDa) protein (catalytic residues: E196, N194) was found to hydrolyze high mannose, hybrid, biantennary and (2,6)-branched triantennary oligosaccharides, but not tetra-antennary, (2,4)-branched triantennary, oligosaccharides with bisecting GlcNAc, or core-fucosylated biantennary glycans. Endo Om showed transglycosylation activity using an Asn-linked agalacto biantennary oligosaccharide as donor and pNP-glucose as acceptor.<sup>73</sup> Further investigations into synthetic applications of Endo Om, for example involving the use of oxazolines as donors, have not yet been reported.

**Endo CE.** Endo CE, a 294 amino acid (34 kDa) protein was identified by sequence analysis with Endo M (34% homology) and then cloned<sup>74</sup> from the nematode *Caenorhabditis elegans* by Yamamoto. Despite its homology with Endo M, Endo CE selectively cleaves high mannose glycans; complex biantennary oligosaccharides were found to be very poor substrates. Yamamoto also showed that Endo CE had transglycosylation activity using a Man<sub>6</sub>GlcNAc<sub>2</sub>-Asn linked donor and glucose as acceptor.

**Other enzymes.** Family GH85 enzymes have so far also been characterised from a variety of other species, including a human enzyme HsENGase,<sup>75</sup> an enzyme from *Bacillus halodurans* (Endo BH)<sup>76</sup> and two enzymes from *Aradopsis thaliana* (ENGaseAtA and ENGaseAtB).<sup>77,78</sup> Significant synthetic applications of these enzymes have so far not been reported.

## 2.4 Family GH18 enzymes

Family GH18 of the glycoside hydrolases, which currently contains more than 10 000 enzymes, also comprises many chitinases of classes II and IV (EC 3.2.1.14) in addition to numerous ENGase enzymes (EC 3.2.1.96). All of the GH18 enzymes have similar ( $\beta/\alpha$ )<sub>8</sub>-barrel folds, and the catalytic residues are located at the carboxyl end of  $\beta$ -strand 4. An amino acid alignment of the catalytic region of a selection of the family GH18 ENGases is shown in Fig. 8, highlighting the key catalytic acid (E) and base (D) residues. Again there is inconsistency in the amino acid numbering used in the literature, as all of the GH18 enzymes have N-terminal signal peptides; initially the numbering used referred to the amino acid position after signal peptide cleavage, but this has not been adopted for Endo S and Endo S2. Although crystallographic and mutagenic studies were performed on the GH18 enzymes before the GH85 enzymes, it is only more recently that the synthetic potential of this family of the ENGases has begun to be realised. An X-ray crystal structure<sup>79</sup> of WT Endo F<sub>3</sub> (PDB: 1EOM) in complex with a biantennary complex oligosaccharide is shown in Fig. 9. Catalytic site base (D126) and acid (E128) residues are shown.

Endo-H	8	GPTSVAYVEV-----NNNSMLNVGKYTLADGGGNAFDVAVI	FAANINYDTGKT	56
Endo-F1	9	NIKLFSTFTEV-----NDTNPLNNLNFTLKNSGKPLVDMVVLFSANINYDAANDK		57
Endo-F2	19	-----AGYRTWRDSATAS-----GNLPS-----MR-WLPDSLDMVMVFPDYPPE-----N	59	
Endo-F3	13	-----AYYIT-----DGRNPTFK-----LK-DIPDKVDMVILFGL-----41		
Endo-S	115	-----GGYFRTWHDKTSDFPT-----EKDKVNS-----MG-ELPKEVVLAFIFHDWTKDY-----S	158	
Endo-S2	68	-----AGYFRTWHDRASTGIDGKQQHPENT-----MA-EVPKEVVLIFVFPHDHTASD-----S	114	
	:	.	:	*
Endo-H	57	AYLHFNE-----NVQRVLDNAVTQIRPLQQQGKIVLVLGVNH-----QGAGFAN	101	
Endo-F1	58	VFVSNNP-----NVQHLLTNRAKYLKPLQDKGKIVLVSILGNH-----DRSGIAN	102	
Endo-F2	60	AYWNTLKTN-----XVPYLHKRGTKVITLGLDLSATT-----92		
Endo-F3	42	KYWSLQDTTKLPGGTGMMGFSFKSYKDLDTQIRSLQSRSRGIKVQNIDDDVSWQS-----94		
Endo-S	159	LFWKELATK-----HVPKLNKQGTRVIRTI	198	
Endo-S2	115	PFWSELKD-----YVHKLHQQGTALVQTIGVNELNGR-----TG	149	
	:		:	*
Endo-H	102	FPS-----QQAASAFAKQQLS-DAVAKYGLDGVDFDDEYAEGNN-----GTAQPNDS		147
Endo-F1	103	LST-----A-RAKAFQAQELK-NTCDLYNLDGVFFDDEYSAYQTPPPSGFVTPSNN		150
Endo-F2	93	---TGGQDSIGYSSWAKGIYDKWGEYNLDGIDIDIESPSGATL-----134		
Endo-F3	95	---SKPGGFASAAAGDAIKSIVIDKWKLDSIISLDIEHSGAKPNPIPTFGYAATGYNGW	151	
Endo-S	199	DTSKYPNTPEGNKALAKAIVDEYVYKYNLDGLDVVEHDSIPKVDK-----244		
Endo-S2	150	LSKDYPDTPEGNKALAAAIVKAFVTDRGVGVLIDIEHEFTNKRTP-----195		
	:	.	:	**:
			*	*

Fig. 8 Amino acid alignment of the catalytic region of some family GH18 ENGases, showing conserved and key active site residues. Amino acid numbering refers to the peptide chain after signal peptide cleavage, except in the cases of Endo S and Endo S2. Catalytic site residues are highlighted in pink. Conserved residues between at least three enzymes are highlighted in grey. An asterisk indicates positions which have a single, fully conserved residue; a colon or a full stop indicates conservation between groups of strongly, or weakly, similar properties respectively.



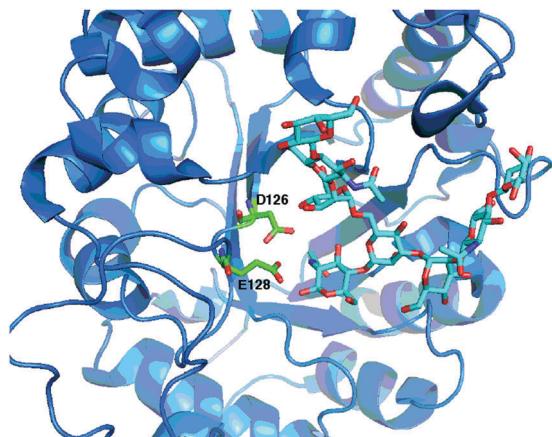


Fig. 9 X-Ray crystal structure<sup>79</sup> of WT Endo F<sub>3</sub> (PDB: 1EOM) in complex with a biantennary complex octasaccharide, showing catalytic active site residues (D126 and E128).

**Endo H.** Endo H, a 271 amino acid protein (plus a 42-amino acid N-terminal signal peptide) first purified from filtrates of *Streptomyces plicatus* (originally named *Streptomyces griseus*) by Tarentino and Maley<sup>80</sup> more than 40 years ago, rapidly became a standard Glycobiology tool for the removal<sup>81</sup> of heterogeneous high mannose glycans from glycoproteins.<sup>82</sup> It is specific for high mannose/hybrid glycans as an  $\alpha$ (1-3)-linked mannose residue (specifically  $\text{Man}\alpha(1-3)\text{Man}\alpha(1-6)\text{Man}\beta(1-4)\text{GlcNAc}\beta(1-4)\text{GlcNAc-Asn}$ ) is absolutely required for processing. Additionally it can hydrolyse hybrid N-glycans that contain core fucose. Following its cloning and expression in *E. coli*,<sup>83</sup> and identification of active site residues by mutagenesis,<sup>18</sup> the crystal structures<sup>84</sup> of WT Endo H (PDB: 1EDT) and then a variety of mutants were solved;<sup>85</sup> a total of 8 structures are now available (PDB: 1C3F, 1C8X, 1CBY, 1C90, 1C91, 1C92, 1C93). An early example of very low efficacy transglycosylation activity of Endo H (alongside Endo F<sub>1</sub>) was reported by Trimble<sup>86</sup> during the hydrolysis of Asn-linked high mannose N-glycans when transfer to glycerol was observed. However, Shoda later reported that Endo H was not capable of effecting glycosylation using a disaccharide oxazoline,<sup>21</sup> and larger N-glycan oxazolines were also found not to be donor substrates by both Fairbanks<sup>48</sup> and Wang.<sup>87</sup> Therefore further synthetic investigations of Endo H ceased. Indeed it was probably these early negative results with a family GH18 ENGase that caused synthetic attention to be focussed exclusively on the GH85 ENGases for a considerable time.

**Endo F<sub>1-3</sub>.** The ENGase activity of the bacterium *Elizabethkingia meningoseptica* (formerly named *Flavobacterium meningosepticum*), first observed by Elder and Alexander,<sup>88</sup> actually arises from three different ENGases, labelled as Endo F<sub>1</sub>,<sup>89</sup> F<sub>2</sub>, and F<sub>3</sub>.<sup>90</sup> Whilst Endo F<sub>1</sub> cleaves only high mannose and hybrid oligosaccharides, Endo F<sub>2</sub> preferentially processes biantennary complex glycans, and Endo F<sub>3</sub> is specific for bi- and triantennary complex oligosaccharides. A particular point of interest is that Endo F<sub>2</sub> and Endo F<sub>3</sub> can hydrolyse substrates possessing core fucose. Endo F<sub>1</sub> (289 amino acids, plus a 50 amino acid N-terminal signal peptide) was the first ENGase enzyme for which a crystal structure was solved

(PDB: 2EBN),<sup>91</sup> and is very closely related to Endo H, though a difference is that core fucosylation reduces the ability of Endo F<sub>1</sub> to process high mannose glycans by a factor of 50, whereas it has no effect on the hydrolytic activity of Endo H.<sup>92</sup> Endo F<sub>1</sub> was (simultaneously to Endo H), shown to have the ability to transglycosylate glycerol during the hydrolysis of high mannose Asn-linked N-glycans. Endo F<sub>2</sub>, (290 amino acids, plus a 45 amino acid N-terminal signal peptide) was also cloned and expressed as a fusion protein in *E. coli*.<sup>93</sup> The crystal structure of Endo F<sub>3</sub> (290 amino acids plus a 39-amino acid N-terminal signal peptide) was solved<sup>79</sup> both alone (PDB: 1EOK) and in complex with a biantennary complex octasaccharide (PDB: 1EOM). Interesting and contrasting glycosylation activity of the Endo F<sub>1-3</sub> enzymes using oxazolines as donors was reported by Wang.<sup>87</sup> In particular Endo F<sub>2</sub> and Endo F<sub>3</sub> were specifically able to use oxazolines to catalyse glycosylation of GlcNAc with a core fucose at position-6, but not GlcNAc without this core fucose. Contrastingly Endo F<sub>1</sub> displayed the converse activity; *i.e.* it was able to glycosylate non-fucosylated GlcNAc, but not GlcNAc with fucose at position-6. Glycosynthase mutants (D126A and D126Q) of Endo F<sub>3</sub> were recently reported by Wang (the numbering in this paper as D165 refers to the amino acid sequence before cleavage of the 39-mer signal peptide),<sup>94</sup> who revealed that they too were able to catalyse the glycosylation of fucosylated GlcNAc, but were also capable of transferring both bi- and tri-antennary complex glycans.

**Endo S.** Endo S<sup>95</sup> is a 995 amino acid (108 kDa) family GH18 ENGase produced by the bacterial pathogen *Streptococcus pyogenes*, which bears similarities to Endo F<sub>2</sub>. Like Endo H and Endo F<sub>1-3</sub> it comprises an N-terminal signal peptide (36 amino acids), but unlike the other GH18 enzymes the amino acid numbering in the literature usually refers to the full-length polypeptide. Additionally, Endo S has so far proven to be unique amongst the ENGases in that it specifically hydrolyzes the core glycans on human IgG antibodies (and not IgA or IgM), and is additionally selective for complex<sup>96</sup> as opposed to high mannose glycans. Remarkably Endo S is very specific for IgGs; for example it has enzymatic activity on natively folded IgG, but not on denatured IgG.<sup>97</sup> Following the first report by Davis<sup>98</sup> on the ability of Endo S to glycosylate IgGs using oxazolines as donors, the glycosynthase mutants D233A and D233Q were made by Wang<sup>99</sup> which were capable of catalysing the glycosylation of intact de-glycosylated antibodies (*i.e.* which had been trimmed back to  $\text{Fuc}\alpha(1-6)\text{GlcNAc}$  residues at Asn-297) using complex type N-glycan oxazolines and without product hydrolysis. The crystal structures of both WT Endo S (PDB: 4NUY) and the D233Q Endo S mutant (PDB: 4NUZ) were subsequently reported.<sup>100</sup> Most recently Wang has developed new mutants of a new family GH18 enzyme called Endo S2,<sup>101</sup> which has 37% identity with Endo S. Endo S2 is produced by a particular serotype (M49) of *S. pyogenes*, and Collin and co-workers have recently shown<sup>102</sup> that Endo S2 has broader hydrolytic activity on IgG's than Endo S; it was also able to cleave high mannose and hybrid glycans from a series of therapeutic mAbs in addition to complex oligosaccharides. Wang subsequently produced<sup>34</sup> two Endo S2 mutants, D184M and D184Q, which were found to have considerably relaxed substrate specificity and were capable of transferring high mannose, hybrid



and complex *N*-glycans to IgGs. Additionally they were also found to be considerably more active catalysts than the corresponding Endo S mutants.

**Other enzymes.** A number of other family GH18 ENGases have been identified, cloned, and in some cases crystallised, including: the high-mannose specific ENGase Endo T,<sup>103</sup> which is a 359 (32 kDa) amino acid protein produced<sup>104</sup> by the mesophilic fungus *Hypocrea jecorina* (PDB: 4AC1, catalytic residues E131, D129); Endo FV<sup>105</sup> an ENGase from the basidiomycete *Flammulina velutipes*, which displayed transglycosylation activity to glucose as an acceptor using full length Asn-linked high mannose and hybrid glycans; and Endo E,<sup>106</sup> an 88 kDa enzyme isolated from *Enterococcus faecalis*<sup>107</sup> by sequence similarity with Endo S, which possesses both a family GH18  $\alpha$ -domain, and a  $\beta$ -domain similar to the family GH 20 enzymes. Additionally coordinates have been deposited in the PDB for the ENGase Endo BT from *Bacteroides thetaiotaomicron* (PDB: 3POH) without an associated publication.

### 3. Synthetic applications of ENGase enzymes

ENGase enzymes may be used to transfer *N*-glycans onto a wide variety of other species providing that they possess a GlcNAc residue to act as an acceptor. An additional requirement is that other functionalization of this GlcNAc residue is tolerated by the enzyme in question; for example the presence of a fucose at position-6, a xylose at position-3, or protecting groups on any of the hydroxyl groups. Below are some recent highlights from three general areas of application of ENGase biocatalysis.

#### 3.1 Production of bioactive glycopeptides

Many early applications of ENGases to catalyse transglycosylation using full-length Asn-linked glycans as donors focussed on peptides with a GlcNAc residue attached to asparagine residue as the acceptor. For example Wang reported<sup>108</sup> the convergent ENGase (both Endo A and Endo M) mediated synthesis of CD52 glycopeptides, bearing both high mannose and complex biantennary glycans; CD52 is a 12-amino acid GPI-anchored cell-surface glycopeptide found on all human lymphocytes and sperm cells.

C34 is a 34-mer peptide derived from the C-terminal ecto-domain of HIV-1 envelope glycoprotein, gp41. Synthetic C34 peptides have been shown to possess potent anti-HIV activity (typical IC<sub>50</sub>'s  $\sim$  1.1 nM). Wang reported<sup>109</sup> the synthesis of a high mannose (Man<sub>9</sub>GlcNAc<sub>2</sub>) glycoform of C34 using Endo A to catalyse the transglycosylation of a synthetic 34-mer peptide using a full length Man<sub>9</sub>GlcNAc<sub>2</sub>Asn as donor in 11% yield. However, unfortunately glycosylation with the high mannose *N*-glycan actually decreased the anti-HIV activity of C34 by a factor of  $\sim$  7.

The very significant increase in efficiency achieved by using oxazolines as donors was clearly demonstrated by Wang in the Endo A and Endo M catalysed synthesis of glycosylated versions of both 34-mer HIV gp41 and 7-mer gp120 fragments using di- and tetrasaccharide oxazolines as donors (yields 73–82%).<sup>44</sup>

Similarly Wang also produced di-glycosylated 47-mer cyclic glycopeptides corresponding to the third variable domain (V3) of HIV gp120 domain, in 60% and 86% yields (Scheme 2a) respectively.<sup>45</sup>

With larger oligosaccharide oxazolines the rate of product hydrolysis frequently outcompetes the rate of the synthetic reaction. Therefore it was only by the use of glycosynthase mutants, for example Endo A N171A, that larger structures, such as a Man<sub>9</sub>GlcNAc oxazoline could be added to a pentapeptide (Scheme 2b). In the same report the N175A Endo M mutant was similarly used to attach complex biantennary glycans to the HIV C34 peptide.<sup>50</sup>

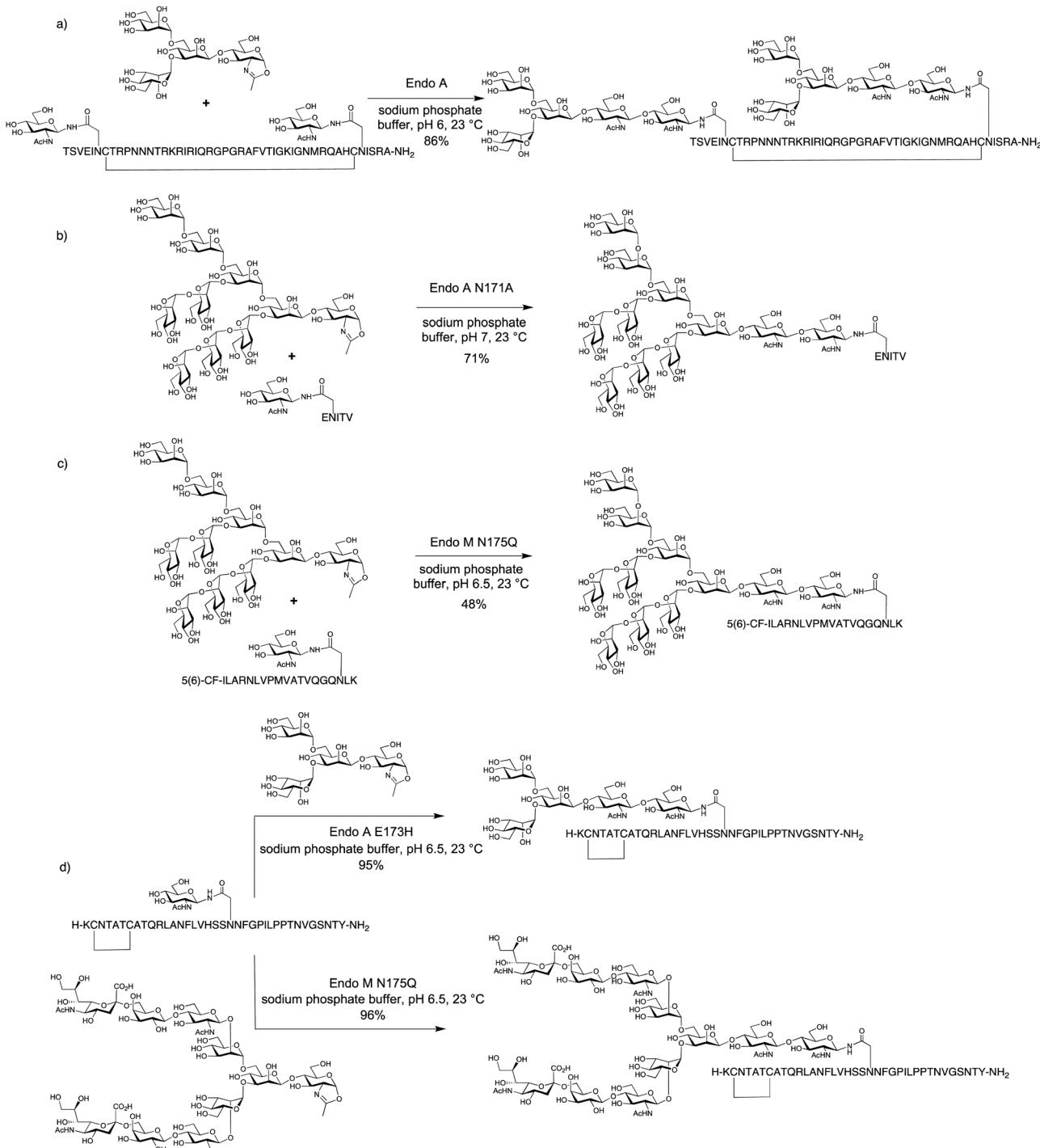
A combination of solid phase peptide synthesis (installing GlcNAc and GalNAc at the pre-determined *N*- and O-glycosylation sites), and the use of an ENGase (either WT Endo A, or the Endo M N175A or Y217F mutants) to attach *N*-glycans to GlcNAc, and a galactosyl transferase to attach galactose to GalNAc, was developed by Wang<sup>110</sup> to produce CD52 glycopeptides bearing both *N*- and O-glycan structures. Wang<sup>111</sup> has also used the ENGases (WT Endo A and N175A Endo M) in combination with an *N*-glycosyl transferase (NGT), which catalyses the addition of a glucose residue to asparagine residues in peptides, for the production of other glycosylated versions of the HIV-1 gp41 polypeptide C34. Similarly, Nakahara<sup>112</sup> and co-workers used the Endo M N175Q mutant to glycosylate Saposin C with a truncated complex biantennary glycan; the Saposin C 80-mer peptide was assembled by the combined use of solid phase peptide synthesis and native chemical ligation (NCL).

Neither Endo A nor Endo M (nor their mutants) is capable of catalysing the glycosylation of GlcNAc residues with a core fucose. Wang has therefore developed the use of Endo F<sub>2</sub> and Endo F<sub>3</sub> for the production of core-fucosylated glycopeptides. Their application, together with a complex bi-antennary *N*-glycan oxazoline donor, was exemplified by the synthesis of a complex core-fucosylated CD52 antigen (52% yield with Endo F<sub>3</sub>, 25% yield with Endo F<sub>2</sub>).<sup>87</sup>

Fairbanks has applied mutant ENGases for the synthesis of putative glycopeptide vaccine candidates (Scheme 2c).<sup>113</sup> In particular it was shown that the conjugation of high mannose glycans to a human cytomegalovirus (CMV) tegument 19-mer peptide increased its binding to antigen presenting cells (APCs), presumably *via* interaction with the mannose receptor (MR) which is widely expressed on APCs. Furthermore an embedded 9-mer peptide epitope was cross-presented by the APCs to a cytotoxic T-lymphocyte clone (CTL, CD8<sup>+</sup>) provided that the site of glycosylation was outside the 9-mer epitope.

ENGases have also been used to access glycosylated variants of therapeutic peptides. Fairbanks used both the commercially available N175Q Endo M mutant, and the E173H Endo A mutant to access a variety of glycosylated variants of the anti-diabetes drug pramlintide,<sup>114</sup> which is a non-fibril forming analogue of human amylin (Scheme 2d). The rationale for this study was an attempt to improve the pharmacokinetic properties of the current drug, which is itself of low solubility and has a low *in vivo* half-life. The glycopeptides produced displayed both *in vitro* and *in vivo* activity, the latter demonstrated by a blood glucose tolerance test in rats. Further *in vitro* studies also





**Scheme 2** Selected examples of the convergent synthesis of glycopeptides using ENGases: (a) di-glycosylated 47-mer V3 domain of HIV gp120 made with WT Endo A; (b) high mannose ( $\text{Man}_9\text{GlcNAc}_2$ ) glycosylated pentapeptide made using the Endo A N171A mutant; (c) high mannose ( $\text{Man}_9\text{GlcNAc}_2$ ) glycosylated CMV tegument protein made using the Endo M N175Q mutant; (d) glycosylated variants of the 37-mer anti-diabetes drug pramlintide made using the Endo A F173H and Endo M N175Q mutants

correlated the position of glycosylation with the ability of the glycopeptide to agonise amylin receptors.<sup>115</sup>

### 3.2 Production of remodelled glycoproteins

RNase B has been the glycoprotein of choice for remodelling using various ENGases and different oxazolines. In the first step

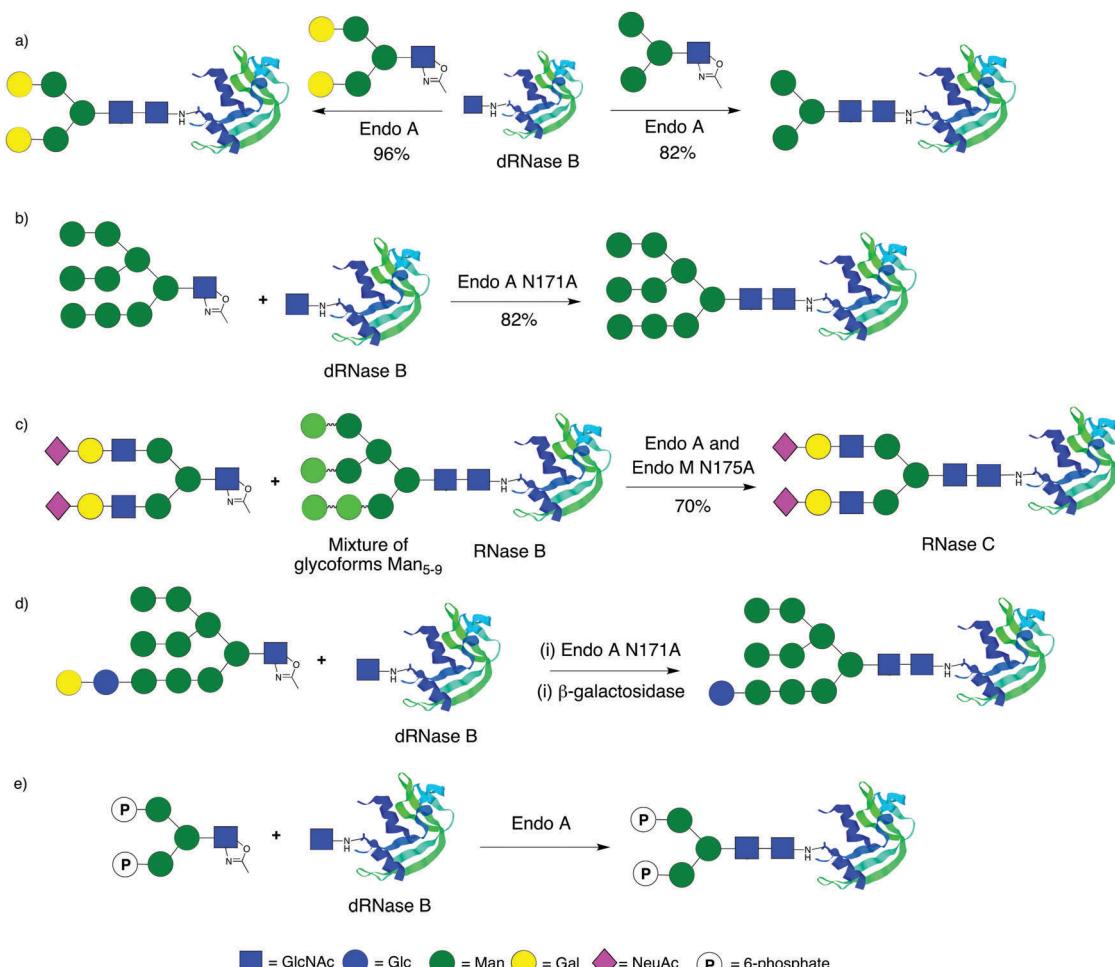
commercially available bovine RNase B (an inseparable mixture of high mannose glycoforms) is typically treated with Endo H, which cleaves the *N*-glycan at Asn-34 back to a single GlcNAc, giving de-glycosylated RNase B (dRNase B). This protein is then used as a substrate for the ENGase catalysed attachment of *N*-glycans. Transglycosylation using full length *N*-glycans is not

an effective method for glycoprotein production; even when using very large excesses of glycan yields are typically less than 5%.<sup>38,116</sup> Wang<sup>46</sup> was the first to demonstrate that the use of oxazolines was significantly more efficient than transglycosylation for the remodelling of RNase B (Scheme 3a). Using Endo A and a variety of different oxazolines he went on to produce a diverse array of RNase glycoforms,<sup>117</sup> including the attachment of azido-substituted glycans; these re-modelled glycoproteins could then be used as substrates for further modification by click chemistry. Building on the synthesis of glycopeptides, Wang showed that the glycosynthase mutants Endo A N171A and Endo M 175A were capable of producing re-modelled RNase B, bearing high mannose (Scheme 3b) or truncated complex biantennary glycans respectively, in excellent yield.<sup>50</sup> In an elegant example of 'one-pot remodelling' (Scheme 3c) Wang and Yamamoto showed that the combined use of WT Endo A (which cleaves high mannose glycans) and the N175A Endo M mutant (which is capable of catalysing the synthesis of complex biantennary glycans) together with a complex biantennary oxazoline derived from egg yolks allowed the direct conversion of RNase B (a mixture of high mannose

glycoforms) into RNase C (a single sialylated complex biantennary glycoform).<sup>118</sup>

Wang also produced<sup>119</sup> further high mannose glycoforms of RNase which additionally contained glucose as ligands for the lectin-like molecular chaperones calnexin and calreticulin using the Endo A N171A mutant (Scheme 3d). A notable point of this work was that galactose, which will not act as an acceptor for the ENGases, was used to 'protect' the glucose residue (which itself may be glycosylated by ENGases) at the non-reducing end of the oligosaccharide; following glycoprotein remodelling using the ENGase mutant, the galactose was removed with a  $\beta$ -galactosidase.

Protein transport to the lysosome is mediated by the mannose-6-phosphate receptors (M6PRs), which recognise high mannose glycans comprising mannose-6-phosphate (M6P) residues. The production of glycoproteins that possess optimal phosphorylation of mannose residues is therefore of particular significance for improving the efficacy of enzyme replacement therapies for the treatment of lysosomal storage disorders.<sup>7</sup> Fairbanks and co-workers recently reported<sup>120</sup> the first production of a phosphorylated glycoprotein using ENGase catalysis. In this work a



**Scheme 3** Selected examples of the synthesis of remodelled RNase B using ENGases: (a) truncated glycoforms produced using Endo A; (b) high mannose ( $\text{Man}_9\text{GlcNAc}_2$ ) glycoform produced using the Endo A N171A mutant; (c) one-pot remodelling of a mixture of RNase B glycoforms into RNase C using Endo A and the Endo M N175A mutant; (d) glucosylated high mannose ( $\text{GlcMan}_9\text{GlcNAc}_2$ ) glycoform produced using the Endo A N171A mutant; (e) mannose-6-phosphate containing glycoform produced using Endo A.



synthetic glycan bearing M6P residues was converted to an oxazoline and transferred to both a glycosyl amino acid and de-glycosylated RNase B (dRNase B) using WT Endo A (Scheme 3e). Shortly afterwards Wang and co-workers published<sup>121</sup> a similar approach, using both WT Endo A and the N171A Endo A mutant, and additionally reported surface plasmon resonance (SPR) studies of glycopeptide/protein binding to the immobilised cation independent mannose-6-phosphate receptor (CI-M6PR).

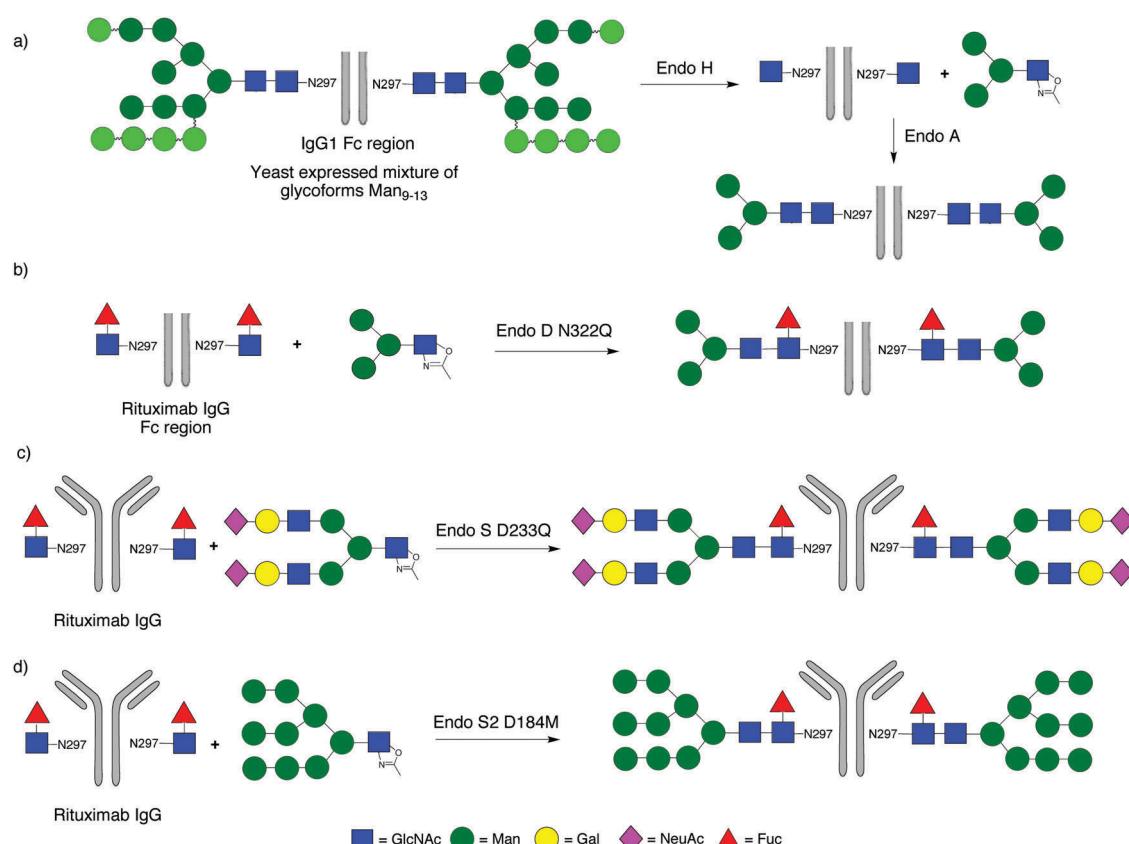
### 3.3 Production of remodelled mAbs

A significant focus of recent attention has been application of ENGase-mediated remodelling of the glycans attached to the Fc region of IgGs, with the objective of maximising the effector functions of clinically administered mAbs. One key aspect of this work is that the heterogeneous *N*-glycans attached to Asn-297 of the Fc region of IgGs expressed in mammalian cell lines are frequently fucosylated, precluding their remodelling by many of the ENGases.

In the first example (Scheme 4a) WT Endo A was used to remodel the Fc region of human IgG1, which was expressed in *Pichia pastoris*. The heterogeneous mixture of non-fucosylated yeast glycoforms was first trimmed back to a single GlcNAc at the Asn-297 residues of the homodimer by treatment with Endo H, and then extended with tetra- and hexasaccharide oxazolines.<sup>122</sup>

Alternatively high mannose glycoforms of IgG1 could be produced by expression in a mammalian cell line (CHO cells) in the presence of the glycosidase inhibitor kifunensine; again treatment with Endo H then trimmed back these glycans to a single GlcNAc at Asn-297. However, remodelling of the Fc region of IgG1 with the Endo M N175A and N175Q mutants was unsuccessful.<sup>123</sup> Similarly WT Endo A, which is high mannose specific, was also unable to transfer full-length complex biantennary glycans to the Fc region, though it could attach a variety of truncated structures, including those with a 'bisecting' GlcNAc.

In subsequent studies Wang turned attention to glycosylation of IgG fragments that were core fucosylated. Firstly the IgG Fc region of the therapeutic mAb rituximab (used for the treatment of leukaemia and lymphoma) was produced by papain digestion, and the heterogeneous *N*-glycans at Asn-297 of commercial rituximab, which are core-fucosylated biantennary complex type oligosaccharides, were trimmed back to Fuc $\alpha$ (1-6)GlcNAc using Endo S. Wang then demonstrated that the N322Q mutant of Endo D was capable of attaching the Man<sub>3</sub>GlcNAc tetrasaccharide oxazoline (Scheme 4b).<sup>71</sup> However the narrow substrate specificity of Endo D precluded its application to larger oligosaccharide structures. Given these synthetic limitations attention turned to the family GH18 ENGase Endo S; an ENGase which preferentially acts hydrolytically on complex type *N*-glycans attached to folded



**Scheme 4** Selected examples of the remodelling of IgGs using ENGases; (a) Endo A catalysed attachment of a truncated core glycan to the non-fucosylated Fc region of human IgG; (b) Endo D N322Q mutant catalysed attachment of a truncated core glycan to the fucosylated Fc region of the mAb rituximab; (c) Endo S D233Q mutant catalysed attachment of a complex biantennary glycan to fucosylated intact rituximab; (d) Endo S2 D184M mutant catalysed attachment of a high mannose glycan to fucosylated intact rituximab.



IgG's. The first report of the synthetic use of WT Endo S by Davis<sup>98</sup> demonstrated its ability to transfer a tetrasaccharide oxazoline to the fucosylated Fc region of human IgG. Shortly afterwards Wang<sup>99</sup> developed D233A and D233Q glycosynthase mutants of Endo S, and applied them for remodelling of rituximab with full-length complex biantennary structures (Scheme 4c). Firstly the heterogeneous *N*-glycans at Asn-297 were trimmed back to Fuc $\alpha$ (1-6)GlcNAc using WT Endo S. Then the D233A and D233Q Endo S mutants were shown to be capable of catalysing the attachment of a variety of glycans, including a full length complex biantennary glycan, as well as asialylated complex and core structures. In this study Wang also showed that the core fucose attached to the GlcNAc at Asn-297 of rituximab could be removed by a fucosidase after the hydrolytic action of Endo S, and that subsequently the D233A and D233Q Endo S mutants could also transfer glycans to the non-fucosylated GlcNAc. However, these Endo S mutants would not transfer high mannose structures.

The Wang Endo S D233Q mutant has subsequently been applied by others<sup>124</sup> for the production of a variety of defined versions of IgGs, including the therapeutic mAbs herceptin<sup>125</sup> and rituximab.<sup>126</sup> Wang has also demonstrated<sup>127</sup> the ability of the D126A mutant of Endo F<sub>3</sub> (numbering after signal peptide cleavage) for the glycosylation of the fucosylated Fc region of rituximab with an asialo triantennary complex type oxazoline to produce homogeneous intact antibody glycoforms; a transformation that mutants of Endo S were incapable of performing. Most recently Wang<sup>34</sup> has developed D184M and D184Q glycosynthase mutants of the new enzyme Endo S2, and shown that these have more relaxed substrate specificity than the corresponding Endo S mutants. For example they were capable of transferring all three major types of *N*-glycan (high mannose, hybrid and complex) to a range of IgGs, both with (Scheme 4d) and without core fucose in a highly efficient manner.

### 3.4 Scope and limitations

Although the use of ENGases is a powerful approach to access defined homogeneous glycoconjugates, there are currently some clear limitations. Firstly, and perhaps most obviously, their use only allows remodelling of *N*-glycan structures. As many mammalian glycoproteins, for example the therapeutic glycoprotein erythropoietin (EPO), also contain *O*-linked oligosaccharides, it is clear that ENGases cannot on their own be used to access versions comprising defined *O*-glycan structures. One possible solution is the use of glycosyl transferases to build more extended *O*-glycan structures from GalNAc-Ser/Thr, as has already been demonstrated by Wang for the production of glycopeptides bearing both defined *N*- and *O*-glycan structures.<sup>110</sup> Alternatively one may in the future envisage the further<sup>128</sup> synthetic development of *endo*- $\alpha$ -N-acetylgalactosaminidases, for example from family GH101, the use of which may perhaps allow the direct attachment of extended *O*-glycan structures to protein serine and threonine residues.

Another issue is that many glycoproteins possess several different *N*-glycans at different glycosylation sites. Possible strategies to overcome this hurdle are already evident. For example in the very first ENGase-mediated synthesis of a glycopeptide bearing different

*N*-glycans at two different sites Haneda and Toma used the differential hydrolytic and transglycosylation activities of Endo M and Endo A to transfer first a complex and then a high mannose glycan to different positions on the same peptide.<sup>129</sup> Although ENGases have not yet been applied for the fabrication of glycoproteins bearing multiple different *N*-glycan structures, similar strategies may be envisaged. A caveat is that most ENGases have only rather broad hydrolytic specificity and so differentiating between different *N*-linked glycosylation sites may not always be possible. This is particularly true from a synthetic perspective where one may expect limited control over the relative rates of ENGase-catalysed glycosylation of different GlcNAc residues at different *N*-linked glycosylation sites; in the Haneda/Toma case<sup>129</sup> it was perhaps fortuitous that one GlcNAc was considerably more reactive than the other, and so selectively was achievable.

Finally it is important to note that the use of ENGases together with oxazolines as donors is not without competing side reactions. For example, Davis<sup>125</sup> and co-workers recently reported that a non-ENGase catalysed non-specific glycation of the IgG herceptin occurred during attempted glycosylation when using a large excess of oxazoline. In this instance Davis and colleagues were able to find conditions which suppressed this competing glycation; namely the use of the Endo S D233Q mutant at higher enzyme loading (10% w/w) and by performing multiple additions of oxazoline (e.g. 7 additions of 15 equivalents at 15 min intervals) to reduce its peak concentration.<sup>125</sup> However, since analysis of the products of many ENGase-catalysed IgG and protein glycosylations has typically been performed solely on the basis of mass spectrometry, this type of glycation may have been previously overlooked (or at least underestimated), particularly in cases where large excesses of oxazoline have been used.

## 4. Conclusions

ENGase enzymes are potent biocatalysts that may be applied for the convergent synthesis of glycopeptides, and the re-modelling of the heterogeneous mixtures of glycoprotein glycoforms, including IgGs, that are invariably produced during expression. Two key developments have transformed the ENGases from being tools primarily for the Glycobiologist into highly effective and practical synthetic biocatalysts; namely the development of oxazolines as activated donors, and the production of glycosynthase ENGase mutants. An ever-increasing number of glycosynthase ENGase mutants is available. A summary of some of the most synthetically useful mutant enzymes produced so far is given in Table 2. At this point in time it is fair to say that there is no 'best enzyme for every job'; rather each mutant has a different substrate tolerance, both with respect to donor (*i.e.* oxazoline structure) and acceptor (*i.e.* functionalization of the GlcNAc, and what is it attached to). The situation is therefore rather complex, and the precise specificities of each enzyme need to be considered in detail before a catalyst is selected.

With respect to the cloning and expression of new enzymes, then considering the total number of ENGase genes already identified to date, it is clear that so far only the very surface has



been scratched. Given the broad diversity of organisms that produce ENGases, and the wide variety of *N*-linked glycoconjugates on which they act, it is certain that new enzymes with useful action will continue to be identified, expanding the synthetic toolkit that is available. In terms of future applications the use of ENGase remodelling of therapeutic mABs in order to optimise glycan structure and maximise effector functions and subsequent clinical investigations of these materials appears to be only a matter of time. Moreover other significant applications of ENGases, for example for the production of optimised glycoprotein therapeutics and the convergent synthesis of therapeutic glycopeptides with improved pharmacokinetic properties, all beckon.

## References

- 1 A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, *Essentials of Glycobiology*, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY, 2009.
- 2 A. Helenius and M. Aebi, *Science*, 2001, **291**, 2364–2369.
- 3 K. R. Rodgers and R. C. Chou, *Biotechnol. Adv.*, 2016, **34**, 1149–1158.
- 4 R. Jefferis, *Biotechnol. Prog.*, 2005, **21**, 11–16.
- 5 R. Jefferis, *Nat. Rev. Drug Discovery*, 2009, **8**, 226–234.
- 6 M. L. Reitman and S. Kornfeld, *J. Biol. Chem.*, 1981, **256**, 11977–11980.
- 7 Y. Zhu, X. Li, J. Kyazike, Q. Zhou, B. L. Thurberg, N. Raben, R. J. Mattaliano and S. H. Cheng, *J. Biol. Chem.*, 2004, **279**, 50336–50341.
- 8 R. Kornfield and S. Kornfield, *Annu. Rev. Biochem.*, 1985, **54**, 631–664.
- 9 N. Jenkins, R. B. Parekh and D. C. James, *Nat. Biotechnol.*, 1996, **14**, 975–981.
- 10 N. Sethuraman and T. A. Stadheim, *Curr. Opin. Biotechnol.*, 2006, **17**, 341–346.
- 11 (a) C. Unverzagt and Y. Kajihara, *Chem. Soc. Rev.*, 2013, **42**, 4408–4420; (b) D. P. Gamblin, E. M. Scanlan and B. G. Davis, *Chem. Rev.*, 2009, **109**, 131–163.
- 12 <http://www.cazy.org>. See: V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.*, 2013, **42**, D490–D495.
- 13 B. Henrissat and G. Davies, *Curr. Opin. Struct. Biol.*, 1997, **7**, 637–644.
- 14 J. Yin, L. Li, N. Shaw, Y. Li, J. K. Song, W. Zhang, C. Xia, R. Zhang, A. Joachimiak, H.-C. Zhang, L.-X. Wang, Z.-J. Liu and P. Wang, *PLoS One*, 2009, **4**, e4658.
- 15 (a) F. Maley, R. B. Trimble, A. L. Tarentino and T. H. Plummer, *Anal. Biochem.*, 1989, **180**, 195–204; (b) Y. Karamanos, S. Bourgerie, J. P. Barreaud and R. Julien, *Res. Microbiol.*, 1995, **146**, 437–443.
- 16 For some previous reviews see: (a) L.-X. Wang, *Carbohydr. Res.*, 2008, **343**, 1509–1522; (b) L.-X. Wang and W. Huang, *Curr. Opin. Chem. Biol.*, 2009, **13**, 592–600; (c) A. J. Fairbanks, *C. R. Chim.*, 2011, **14**, 44–58.
- 17 S. Kadokawa, K. Yamamoto, M. Fujisaki, K. Izumi, T. Tochikura and T. Yokoyama, *Agric. Biol. Chem.*, 1990, **54**, 97–106.
- 18 B. F. Schmidt, E. Ashizawa, A. S. Jarnagin, S. Lynn, G. Noto, L. Woodhouse, D. A. Estell and P. Lad, *Arch. Biochem. Biophys.*, 1994, **311**, 350–353.
- 19 D. W. Abbott, M. S. Macauley, D. J. Vocadlo and A. B. Boraston, *J. Biol. Chem.*, 2009, **284**, 11676–11689.
- 20 I. Tews, A. van Scheltinga, A. Perrakis, K. Wilson and B. Dijkstra, *J. Am. Chem. Soc.*, 1997, **119**, 7954–7959.
- 21 M. Fujita, S.-i. Shoda, K. Haneda, T. Inazu, K. Takegawa and K. Yamamoto, *Biochim. Biophys. Acta*, 2001, **1528**, 9–14.
- 22 For two personal perspectives on the parallel development of more extended oxazolines as donors for ENGase enzymes see: (a) L.-X. Wang, *Trends Glycosci. Glycotechnol.*, 2011, **23**, 33–52; (b) A. J. Fairbanks, *Pure Appl. Chem.*, 2013, **85**, 1847–1863.
- 23 M. Noguchi, T. Tanaka, H. Gyakushi, A. Kobayashi and S.-i. Shoda, *J. Org. Chem.*, 2009, **74**, 2210–2212.
- 24 M. Umekawa, T. Higashiyama, Y. Koga, T. Tanaka, M. Noguchi, A. Kobayashi, S.-i. Shoda, W. Huang, L.-X. Wang, H. Ashida and K. Yamamoto, *Biochim. Biophys. Acta*, 2010, **1800**, 1203–1209.
- 25 J. Q. Fan, K. Takegawa, S. Iwahara, A. Kondo, I. Kato, C. Abeygunawardana and Y. C. Lee, *J. Biol. Chem.*, 1995, **270**, 17723–17729.
- 26 E. Akaike, M. Tsutsumida, K. Osumi, M. Fujita, T. Yamanoi, K. Yamamoto and K. Fujita, *Carbohydr. Res.*, 2004, **339**, 719–722.
- 27 C. D. Heidecke, T. B. Parsons and A. J. Fairbanks, *Carbohydr. Res.*, 2009, **344**, 2433–2438.
- 28 For a recent review which also includes a discussion of ENGase glycosynthases see: P. M. Danby and S. G. Withers, *ACS Chem. Biol.*, 2016, **11**, 1784–1794.
- 29 L. F. Mackenzie, Q. R. Wang, R. A. J. Warren and S. G. Withers, *J. Am. Chem. Soc.*, 1998, **120**, 5583–5584.
- 30 C. Malet and A. Planas, *FEBS Lett.*, 1998, **440**, 208–212.
- 31 M. Umekawa, W. Huang, B. Li, K. Fujita, H. Ashida, L.-X. Wang and K. Yamamoto, *J. Biol. Chem.*, 2008, **283**, 4469–4479.
- 32 M. Umekawa, C. Li, T. Higashiyama, W. Huang, H. Ashida, K. Yamamoto and L.-X. Wang, *J. Biol. Chem.*, 2010, **285**, 511–521.
- 33 Y. Higuchi, Y. Eshima, Y. Huang, T. Kinoshita, W. Sumiyoshi, S.-i. Nakakita and K. Takegawa, *Biotechnol. Lett.*, 2017, **39**, 157–162.
- 34 T. Li, X. Tong, Q. Yang, J. P. Giddens and L.-X. Wang, *J. Biol. Chem.*, 2016, **291**, 16508–16518.
- 35 C. D. Heidecke, Z. Ling, N. C. Bruce, J. W. B. Moir, T. B. Parsons and A. J. Fairbanks, *ChemBioChem*, 2008, **9**, 2045–2051.
- 36 K. Takegawa, M. Nakoshi, S. Iwahara, K. Yamamoto and T. Tochikura, *Appl. Environ. Microbiol.*, 1989, **55**, 3107–3112.
- 37 J. Q. Fan, M. S. Quesenberry, K. Takegawa, S. Iwahara, A. Kondo, I. Kato and Y. C. Lee, *J. Biol. Chem.*, 1995, **270**, 17730–17735.
- 38 K. Takegawa, M. Tabuchi, S. Yamaguchi, A. Kondo, I. Kato and S. Iwahara, *J. Biol. Chem.*, 1995, **270**, 3094–3099.
- 39 K. Fujita and K. Takegawa, *Biochem. Biophys. Res. Commun.*, 2001, **282**, 678–682.



40 K. Takegawa, K. Yamabe, K. Fujita, M. Tabuchi, M. Mita, H. Izu, A. Watanabe, Y. Asada, M. Sano, A. Kondo, I. Kato and S. Iwahara, *Arch. Biochem. Biophys.*, 1997, **338**, 22–28.

41 K. Fujita and K. Takegawa, *Biochem. Biophys. Res. Commun.*, 2001, **283**, 680–686.

42 J. Q. Fan, L. H. Huynh, B. B. Reinhold, V. N. Reinhold, K. Takegawa, S. Iwahara, A. Kondo, I. Kato and Y. C. Lee, *Glycoconjugate J.*, 1996, **13**, 643–652.

43 Y. Tomabechi, M. A. Squire and A. J. Fairbanks, *Org. Biomol. Chem.*, 2014, **12**, 942–955.

44 B. Li, Y. Zeng, S. Hauser, H. Song and L.-X. Wang, *J. Am. Chem. Soc.*, 2005, **127**, 9692–9693.

45 H. Li, B. Li, H. Song, L. Breydo, I. V. Baskakov and L.-X. Wang, *J. Org. Chem.*, 2005, **70**, 9990–9996.

46 B. Li, H. Song, S. Hauser and L.-X. Wang, *Org. Lett.*, 2006, **8**, 3081–3084.

47 Y. Zeng, J. Wang, B. Li, S. Hauser, H. Li and L.-X. Wang, *Chem. – Eur. J.*, 2006, **12**, 3355–3364.

48 T. W. D. F. Rising, C. D. Heidecke, J. W. B. Moir, Z. Ling and A. J. Fairbanks, *Chem. – Eur. J.*, 2008, **14**, 6444–6464.

49 M. D. L. Suits, Z. Ling, R. J. Bingham, N. C. Bruce, G. J. Davies, A. J. Fairbanks, J. W. B. Moir and E. J. Taylor, *J. Mol. Biol.*, 2009, **389**, 1–9.

50 W. Huang, C. Li, B. Li, M. Umekawa, K. Yamamoto, X. Zhang and L.-X. Wang, *J. Am. Chem. Soc.*, 2009, **131**, 2214–2223.

51 S. Kadowaki, K. Yamamoto, M. Fujisaki, H. Kumagai and T. Tochikura, *Agric. Biol. Chem.*, 1988, **52**, 2387–2389.

52 K. Yamamoto, S. Kadowaki, M. Fujisaki, H. Kumagai and T. Tochikura, *Biosci., Biotechnol., Biochem.*, 1994, **58**, 72–77.

53 K. J. Yamamoto, S. Kadowaki, J. Watanabe and H. Kumagai, *Biochem. Biophys. Res. Commun.*, 1994, **203**, 244–252.

54 K. Haneda, T. Inazu, K. Yamamoto, H. Kumagai, Y. Nakahara and A. Kobata, *Carbohydr. Res.*, 1996, **292**, 61–70.

55 K. Yamamoto, K. Fujimori, K. Haneda, M. Mizuno, T. Inazu and H. Kumagai, *Carbohydr. Res.*, 1997, **305**, 415–422.

56 K. Haneda, T. Inazu, M. Mizuno, R. Iguchi, K. Yamamoto, H. Kumagai, S. Aimoto, H. Suzuki and T. Noda, *Bioorg. Med. Chem. Lett.*, 1998, **8**, 1303–1306.

57 M. Mizuno, K. Haneda, R. Iguchi, I. Muramoto, T. Kawakami, S. Aimoto, K. Yamamoto and T. Inazu, *J. Am. Chem. Soc.*, 1999, **121**, 284–290.

58 K. Fujita, K. Kobayashi, A. Iwamatsu, M. Takeuchi, H. Kumagai and K. Yamamoto, *Arch. Biochem. Biophys.*, 2004, **432**, 41–49.

59 K. Osumi, Y. Makino, E. Akaike, T. Yamanoi, M. Mizuno, M. Noguchi, T. Inazu, K. Yamamoto and K. Fujita, *Carbohydr. Res.*, 2004, **339**, 2633–2635.

60 T. W. D. F. Rising, T. D. W. Claridge, N. Davies, D. P. Gamblin, J. W. B. Moir and A. J. Fairbanks, *Carbohydr. Res.*, 2006, **341**, 1574–1596.

61 T. W. D. F. Rising, C. D. Heidecke, J. W. B. Moir, Z. Ling and A. J. Fairbanks, *Chem. – Eur. J.*, 2008, **14**, 6444–6464.

62 T. B. Parsons, J. W. B. Moir and A. J. Fairbanks, *Org. Biomol. Chem.*, 2009, **7**, 3128–3140.

63 T. W. D. F. Rising, T. D. W. Claridge, J. W. B. Moir and A. J. Fairbanks, *ChemBioChem*, 2006, **7**, 1177–1180.

64 Y. Tomabechi, Y. Odate, R. Izumi, K. Haneda and T. Inazu, *Carbohydr. Res.*, 2010, **345**, 2458–2463.

65 T. Katoh, T. Katayama, Y. Tomabechi, Y. Nishikawa, J. Kumada, Y. Matsuzaki and K. Yamamoto, *J. Biol. Chem.*, 2016, **291**, 23305–23317.

66 T. Muramatsu, *J. Biol. Chem.*, 1971, **246**, 5535–5537.

67 (a) N. Koide and T. Muramatsu, *J. Biol. Chem.*, 1974, **249**, 4897–4904; (b) T. Tai, K. Yamashita, M. Ogata-Arakawa, N. Koide, T. Muramatsu, S. Iwashita, Y. Inoue and A. Kobata, *J. Biol. Chem.*, 1975, **250**, 8569–8575; (c) N. Koide, M. Nose and T. Muramatsu, *Biochem. Biophys. Res. Commun.*, 1977, **75**, 838–844; (d) T. Muramatsu, N. Koide and K.-i. Maeyama, *J. Biochem.*, 1978, **83**, 363–370; (e) T. Muramatsu, *Methods Enzymol.*, 1978, **50**, 555–559.

68 H. Muramatsu, H. Tachikui, H. Ushida, X.-j. Song, Y. Qiu, S. Yamamoto and T. Muramatsu, *J. Biochem.*, 2001, **129**, 923–928.

69 S. Yamamoto, H. Muramatsu and T. Muramatsu, *Glycoconjugate J.*, 2005, **22**, 35–42.

70 T. B. Parsons, M. K. Patel, A. B. Boraston, D. J. Vocadlo and A. J. Fairbanks, *Org. Biomol. Chem.*, 2010, **8**, 1861–1869.

71 S.-Q. Fan, W. Huang and L.-X. Wang, *J. Biol. Chem.*, 2012, **287**, 11272–11281.

72 Y. Eshima, Y. Higuchi, T. Kinoshita, S. I. Nakakita and K. Takegawa, *PLoS One*, 2015, **10**, e0132859.

73 S. Murakami, Y. Takaoka, H. Ashida, K. Yamamoto, H. Narimatsu and Y. Chiba, *Glycobiology*, 2013, **23**, 736–744.

74 T. Kato, K. Fujita, M. Takeuchi, K. Kobayashi, S. Natsuka, K. Ikura, H. Kumagai and K. Yamamoto, *Glycobiology*, 2002, **12**, 581–587.

75 T. Suzuki, K. Yano, S. Sugimoto, K. Kitajima, W. Lennarz, S. Inoue, Y. Inoue and Y. Emori, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 9691–9696.

76 K. Fujita, H. Takami, K. Yamamoto and K. Takegawa, *Biosci., Biotechnol., Biochem.*, 2004, **68**, 1059–1066.

77 R. M. Fischl, J. Stadlmann, J. Grass, F. Altmann and R. Léonard, *Plant Mol. Biol.*, 2011, **77**, 275–284.

78 Y. Kimura, Y. Takeoka, M. Inoue, M. Maed and K. Fujiyama, *Biosci., Biotechnol., Biochem.*, 2014, **75**, 1019–1021.

79 C. A. Waddling, T. H. Plummer, A. L. Tarentino and P. Van Roey, *Biochemistry*, 2000, **39**, 7878–7885.

80 (a) A. L. Tarentino and F. Maley, *J. Biol. Chem.*, 1974, **249**, 811–817; (b) A. L. Tarentino, J. Thomas, H. Plummer and F. Maley, *J. Biol. Chem.*, 1974, **249**, 818–824.

81 R. B. Trimble and F. Maley, *Anal. Biochem.*, 1984, **141**, 515–522.

82 R. A. O'Neill, *J. Chromatogr. A*, 1996, **720**, 201–215.

83 R. J. Trumbly, P. W. Robbins, M. Belfort, F. D. Ziegler, F. Maley and R. B. Trimble, *J. Biol. Chem.*, 1985, **260**, 5683–5690.

84 V. Rao, C. Guan and P. V. Roey, *Structure*, 1995, **3**, 449–457.

85 V. Rao, T. Cui, C. Guan and P. Van Roey, *Protein Sci.*, 1999, **8**, 2338–2346.

86 R. B. Trimble, P. H. Atkinson, A. L. Tarentino, T. H. Plummer, F. Maley and K. B. Tomer, *J. Biol. Chem.*, 1986, **261**, 12000–12005.

87 W. Huang, J. Li and L.-X. Wang, *ChemBioChem*, 2011, **12**, 932–941.



88 J. H. Elder and S. Alexander, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, **79**, 4540–4544.

89 A. L. Tarentino, G. Quinones, W. P. Schrader, L. M. Changchien and J. T. H. Plummer, *J. Biol. Chem.*, 1992, **267**, 3868–3872.

90 A. L. Tarentino, G. Quinones, L. M. Changchien and T. H. Plummer, *J. Biol. Chem.*, 1993, **268**, 9702–9708.

91 P. Van Roey, V. Rao, T. H. Plummer and A. L. Tarentino, *Biochemistry*, 1994, **33**, 13989–13996.

92 R. B. Trimble and A. L. Tarentino, *J. Biol. Chem.*, 1991, **266**, 1646–1651.

93 B. Grimwood, T. Plummer and A. Tarentino, *Glycobiology*, 1998, **8**, 633–636.

94 J. P. Giddens, J. V. Lomino, M. N. Amin and L.-X. Wang, *J. Biol. Chem.*, 2016, **291**, 9356–9370.

95 M. Collin and A. Olsen, *EMBO J.*, 2001, **20**, 3046–3055.

96 M. Allhorn, A. Olsén and M. Collin, *BMC Microbiol.*, 2008, **8**, 3.

97 M. Collin and A. Olsen, *Infect. Immun.*, 2001, **69**, 7187–7189.

98 J. J. Goodfellow, K. Baruah, K. Yamamoto, C. Bonomelli, B. Krishna, D. J. Harvey, M. Crispin, C. N. Scanlan and B. G. Davis, *J. Am. Chem. Soc.*, 2012, **134**, 8030–8033.

99 W. Huang, J. Giddens, S.-Q. Fan, C. Toonstra and L.-X. Wang, *J. Am. Chem. Soc.*, 2012, **134**, 12308–12318.

100 B. Trastoy, J. V. Lomino, B. G. Pierce, L. G. Carter, S. Günther, J. P. Giddens, G. A. Snyder, T. M. Weiss, Z. Weng, L.-X. Wang and E. J. Sundberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 6714–6719.

101 J. Sjögren, W. B. Struwe, E. F. J. Cosgrave, P. M. Rudd, M. Stervander, M. Allhorn, A. Hollands, V. Nizet and M. Collin, *Biochem. J.*, 2013, **455**, 107–118.

102 J. Sjögren, E. F. J. Cosgrave, M. Allhorn, M. Nordgren, S. Björk, F. Olsson, S. Fredriksson and M. Collin, *Glycobiology*, 2015, **25**, 1053–1063.

103 I. Stals, S. Karkehabadi, S. Kim, M. Ward, A. Van Landschoot, B. Devreese and M. Sandgren, *PLoS One*, 2012, **7**, e40854.

104 I. Stals, B. Samyn, K. Sergeant, T. White, K. Hoorelbeke, A. Coorevits, B. Devreese, M. Claeysens and K. Piens, *FEMS Microbiol. Lett.*, 2010, **303**, 9–17.

105 T. Hamaguchi, T. Ito, Y. Inoue, T. Limpaseni, P. Pongsawasdi and K. Ito, *Glycobiology*, 2010, **20**, 420–432.

106 J. Garbe, J. Sjögren, E. F. J. Cosgrave, W. B. Struwe, M. Bober, A. I. Olin, P. M. Rudd and M. Collin, *PLoS One*, 2014, **9**, e91035.

107 M. Collin and V. A. Fischetti, *J. Biol. Chem.*, 2004, **279**, 22558–22570.

108 H. Li, S. Singh, Y. Zeng, H. Song and L. Wang, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 895–898.

109 L.-X. Wang, H. Song, S. Liu, H. Lu, S. Jiang, J. Ni and H. Li, *ChemBioChem*, 2005, **6**, 1068–1074.

110 W. Huang, X. Zhang, T. Ju, R. D. Cummings and L.-X. Wang, *Org. Biomol. Chem.*, 2010, **8**, 5224–5233.

111 J. V. Lomino, A. Naegeli, J. Orwenyo, M. N. Amin, M. Aeby and L.-X. Wang, *Bioorg. Med. Chem.*, 2013, **21**, 2262–2270.

112 H. Hojo, H. Tanaka, M. Hagiwara, Y. Asahina, A. Ueki, H. Katayama, Y. Nakahara, A. Yoneshige, J. Matsuda, Y. Ito and Y. Nakahara, *J. Org. Chem.*, 2012, **77**, 9437–9446.

113 J. D. McIntosh, M. A. Brimble, A. E. S. Brooks, P. R. Dunbar, R. Kowalczyk, Y. Tomabechi and A. J. Fairbanks, *Chem. Sci.*, 2015, **6**, 4636–4642.

114 Y. Tomabechi, G. Krippner, P. M. Rendle, M. A. Squire and A. J. Fairbanks, *Chem. – Eur. J.*, 2013, **19**, 15084–15088.

115 R. Kowalczyk, M. A. Brimble, Y. Tomabechi, A. J. Fairbanks, M. Fletcher and D. L. Hay, *Org. Biomol. Chem.*, 2014, **12**, 8142–8151.

116 K. Fujita and K. Yamamoto, *Biochim. Biophys. Acta*, 2006, **1760**, 1631–1635.

117 H. Ochiai, W. Huang and L.-X. Wang, *J. Am. Chem. Soc.*, 2008, **130**, 13790–13803.

118 W. Huang, Q. Yang, M. Umekawa, K. Yamamoto and L.-X. Wang, *ChemBioChem*, 2010, **11**, 1350–1355.

119 M. N. Amin, W. Huang, R. M. Mizanur and L.-X. Wang, *J. Am. Chem. Soc.*, 2011, **133**, 14404–14417.

120 P. Priyanka, T. B. Parsons, A. Miller, F. M. Platt and A. J. Fairbanks, *Angew. Chem., Int. Ed.*, 2016, **55**, 5058–5061.

121 T. Yamaguchi, M. N. Amin, C. Toonstra and L.-X. Wang, *J. Am. Chem. Soc.*, 2016, **138**, 12472–12485.

122 Y. Wei, C. Li, W. Huang, B. Li, S. Strome and L.-X. Wang, *Biochemistry*, 2008, **47**, 10294–10304.

123 G. Zou, H. Ochiai, W. Huang, Q. Yang, C. Li and L.-X. Wang, *J. Am. Chem. Soc.*, 2011, **133**, 18975–18991.

124 M. Kuroguchi, M. Mori, K. Osumi, M. Tojino, S.-I. Sugawara, S. Takashima, Y. Hirose, W. Tsukimura, M. Mizuno, J. Amano, A. Matsuda, M. Tomita, A. Takayanagi, S.-I. Shoda and T. Shirai, *PLoS One*, 2015, **10**, e0132848.

125 T. B. Parsons, W. B. Struwe, J. Gault, K. Yamamoto, T. A. Taylor, R. Raj, K. Wals, S. Mohammed, C. V. Robinson, J. L. P. Benesch and B. G. Davis, *Angew. Chem., Int. Ed.*, 2016, **55**, 2361–2367.

126 C.-W. Lin, M.-H. Tsai, S.-T. Li, T.-I. Tsai, K.-C. Chu, Y.-C. Liu, M.-Y. Lai, C.-Y. Wu, Y.-C. Tseng, S. S. Shivate, C.-H. Wang, P. Chao, S.-Y. Wang, H.-W. Shih, Y.-F. Zeng, T.-H. You, J.-Y. Liao, Y.-C. Tu, Y.-S. Lin, H.-Y. Chuang, C.-L. Chen, C.-S. Tsai, C.-C. Huang, N.-H. Lin, C. Ma, C.-Y. Wu and C.-H. Wong, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 10611–10616.

127 J. P. Giddens, J. V. Lomino, M. N. Amin and L.-X. Wang, *J. Biol. Chem.*, 2016, **291**, 9356–9370.

128 (a) H. Ashida, H. Ozawa, K. Fujita, S. Suzuki and K. Yamamoto, *Glycoconjuge J.*, 2010, **27**, 125–132; (b) H. Ashida, K. Yamamoto, T. Murata, T. Usui and H. Kumagai, *Arch. Biochem. Biophys.*, 2000, **373**, 394–400.

129 K. Haneda, M. Takeuchi, M. Tagashira, T. Inazu, K. Toma, Y. Isogai, M. Hori, K. Kobayashi, M. Takeuchi, K. Takegawa and K. Yamamoto, *Carbohydr. Res.*, 2006, **341**, 181–190.

