Review

Cell PRESS

Expanding the glycoengineering toolbox: the rise of bacterial *N*-linked protein glycosylation

Jenny L. Baker[1], Eda Çelik[1,2], and Matthew P. DeLisa[1]

[1] School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, USA
[2] Department of Chemical Engineering, Hacettepe University, Beytepe, TR-06800 Ankara, Turkey

Glycosylation is the most prevalent post-translational modification found on proteins, occurring in all domains of life. Ever since the discovery of asparagine-linked (*N*-linked) protein glycosylation pathways in bacteria, major efforts have been made to harness these systems for the creation of novel therapeutics, vaccines, and diagnostics. Recent advances such as the ability to produce designer glycans in bacteria, some containing unnatural sugars, and techniques for evolving glycosylation enzymes have spawned an entirely new discipline known as bacterial glycoengineering. In addition to their biotechnological and therapeutic potential, bacteria equipped with recombinant *N*-linked glycosylation pathways are improving our understanding of the *N*-glycosylation process. This review discusses the key role played by microorganisms in glycosciences, particularly in the context of *N*-linked glycosylation.

Protein glycosylation in bacteria

Glycosylation is one of the major ways in which proteins are post-translationally modified and involves linking of monosaccharides via glycosidic bonds to form a glycan that is covalently attached to a biomolecule, such as a protein or a lipid. Over 50% of eukaryotic proteins are predicted to be glycosylated [1], although yeasts have fewer total glycoproteins than multicellular eukaryotes [2]. The significance of glycosylation is further illuminated by the fact that approximately 70% of therapeutic proteins that are either approved by European and US regulatory agencies or are in clinical and preclinical development are glycoproteins [3]. Glycosylation can occur at several amino acid residues, most commonly through asparagine (*N*-linked) and serine or threonine (*O*-linked). Although there are currently no established rules for predicting the effect of glycosylation on protein folding or function, empirical evidence reveals numerous roles for this protein modification. For example, glycans can influence folding, stability, molecular interactions, and quality control [4–7]. Extracellular display of glycans plays a role in cell–cell recognition, adhesion, and host immune responses to pathogens [8]. Intentionally changing protein-associated carbohydrates can be used to tailor the pharmacokinetic properties of a protein, leading to drugs with enhanced *in vivo* activity, half-life, and resistance to proteolysis [9–13]. Glycosylation can also be used to target therapeutic proteins to specific cells or tissues [14] and to modulate their biological activities through interactions with specific receptors [15]. Conversely, the incorrect structure or position (on a protein backbone) of glycans can negatively affect pharmacokinetics, and in some cases trigger immunogenic responses [16]. For these reasons, expression systems used to produce therapeutic glycoproteins must be carefully chosen based on their ability to synthesize desired glycan structures. Although mammalian expression systems are currently the preferred host for producing therapeutic glycoproteins, glycoengineered bacteria are emerging as a viable alternative for producing glycoprotein therapeutics and vaccines [17–19] and are the focus of this review.

Although long established in eukaryotes, *N*-linked protein glycosylation in bacteria is a relatively recent discovery. The pathogenic ε-proteobacterium *Campylobacter jejuni* was the first bacterium found to have an *N*-linked glycosylation pathway [20], encoded by the protein glycosylation locus (*pgl*) (Figure 1A). A growing number of similar protein modification systems have been discovered in other ε- and δ-proteobacteria bacteria [21], but the *C. jejuni N*-glycosylation pathway remains the most extensively characterized [21,22]. It is further described in Box 1 and compared to the eukaryotic *N*-glycosylation process. To date, more than 60 periplasmic and membrane *N*-glycoproteins have been identified in *C. jejuni* [23,24], and it has been predicted that up to 150 proteins of various functions in this bacterium are *N*-glycosylated [24].

Shortly after its discovery, the *C. jejuni* glycosylation pathway was functionally transferred to *Escherichia coli*, bestowing on the latter organism the ability to produce *N*-linked glycoproteins (Figure 1B) [25–27]. With this system, proteins can be glycosylated at authentic sites or can be modified for glycosylation by the introduction of short glycosylation tags (GlycTags) containing a preferred *C. jejuni N*-glycosylation consensus sequence, D-Q-N-A-T [28]. When expressed in glycosylation-competent *E. coli*, recombinant proteins engineered with GlycTags are efficiently glycosylated [28,29]. This glycosylation is compatible with

Corresponding author: DeLisa, M.P. (md255@cornell.edu)
Keywords: bioorthogonal chemical reporters; biopharmaceuticals; therapeutics and vaccines; glycoconjugates; glycoproteins; post-translational modification.

0167-7799/$ – see front matter
 http://dx.doi.org/10.1016/j.tibtech.2013.03.003

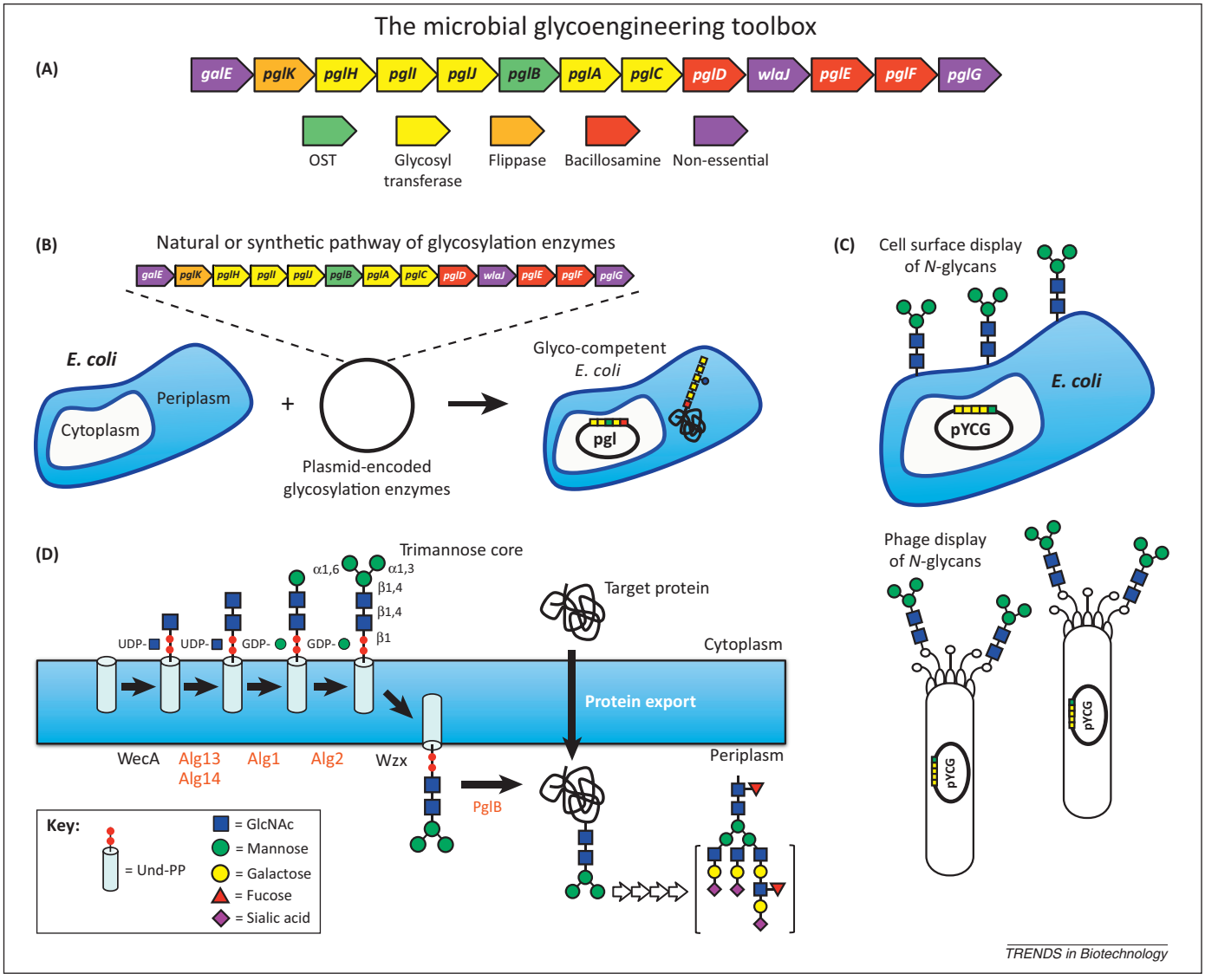

Figure 1. The expanding glycoengineering toolbox. **(A)** The 17-kb *pgl* locus of *Campylobacter jejuni* encodes the *N*-linked glycosylation machinery and has been fully reconstituted in *Escherichia coli*. The greater than expected number of bacterial glycosylation pathways and the diverse glycosyltransferases they encode are a rich source of unique enzymes for understanding and exploiting glycoconjugation reactions. **(B)** Functional transfer of the *C. jejuni* protein glycosylation pathway on plasmid pACYC-*pgl* (*pgl*) into *E. coli*. **(C)** Display of glycoproteins can occur on the surface of bacteria (top) or on filamentous phage particles (bottom). These technologies create a genotype–phenotype linkage that is needed for directed evolution of glycosylation enzymes, pathways, and host strains. Plasmid pYCG encodes the complete pathway for the biosynthesis of trimannosyl core glycans. **(D)** Schematic of the synthetic pathway for synthesis of a trimannosyl core glycan and its transfer to acceptor sites in target proteins. Enzyme names in black are native to *E. coli*; enzyme names in red are heterologous. The glycan in brackets to the right of open arrows depicts a terminally sialylated structure common in human glycoproteins. Although such a structure has not yet been produced in bacteria, it is a major target of glycoengineering efforts.

transport mechanisms that deliver the proteins to environments beyond the periplasm, such as the outer membrane, membrane vesicles, and the extracellular medium (Figure 1C) [28]. *N*-Linked glycoproteins have also been displayed on filamentous phage particles [29,30], opening up the route for glycophage display and its application to the engineering of glycophenotypes.

Although glycan transfer in bacterial *N*-linked glycosylation pathways typically occurs in the periplasm and involves a membrane-bound oligosaccharyltransferase (OST), *N*-glycan transfer to proteins can also proceed in the bacterial cytoplasm via soluble *N*-glycosyltransferases such as the HMW1C protein from *Actinobacillus pleuropneumoniae* [31]. Cytosolic HMW1C is able to transfer a glucose (Glc) or galactose (Gal) moiety directly to an asparagine within an N-X-S/T consensus sequence, whereas a separate glucosyltransferase is necessary to elaborate the *N*-linked glucose or galactose with up to six additional glucose units.

In addition to *N*-glycosylation, *O*-linked glycosylation pathways have been discovered in bacteria. For example, certain bacteria such as *Neisseria meningitidis* and *Neisseria gonorrhoeae* *O*-glycosylate their pilins in a process that is catalyzed by the OSTs PglL and PglO, respectively [32–35]. These pathways differ from most *O*-linked glycosylation pathways identified in eukaryotes in that the OST transfers the glycan *en bloc* from the Und-PP carrier onto target proteins in the periplasm [21]. In fact, bacterial *O*-glycosylation is mechanistically more similar to the bacterial *N*-linked process. Like its *N*-glycosylation counterpart, the *N. gonorrhoeae* *O*-linked glycosylation pathway is a general system capable of modifying a diverse group of membrane-associated proteins [35]. Although a more thorough discussion of *O*-glycosylation is outside the scope of