Editorial

Pharmaceutical BIOPROCESSING

Pharm. Bioprocess.
(2013) 1(3), 221–224

Expanding the bioprocessing toolbox of *Escherichia coli* through metabolic engineering and synthetic biology: an emerging glycosylation chassis

Benjamin Strutton, Jagroop Pandhal & Phillip Craig Wright*
ChELSI Institute, Department of Chemical & Biological Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK
*Author for correspondence:
E-mail: p.c.wright@sheffield.ac.uk

"The challenge to improve glycoprotein productivity in *Escherichia coli* and genetically engineer the cells to produce the correct human glycans is huge."

Keywords: biopharmaceuticals, *Escherichia coli,* glycosylation, metabolic engineering, MS, proteomics, recombinant proteins, synthetic biology

The production of recombinant proteins in biological host cell factories is a lucrative market worth over US$50 billion, with annual sales growing, on average, 20% per annum [1]. The choice of a host-cell production factory strongly depends on the type of product required. The most popular, and hence representative, species for bacterial, yeast and mammalian cell culture are *Escherichia coli*, *Saccharomyces cerevisiae* and Chinese hamster ovary (CHO) cells, respectively. In terms of recombinant DNA technology and scale-up for manufacture, these systems are the most understood.

Although relatively simple proteins, such as insulin, can be produced in all three of these cell systems, microbial cells are preferred as they are much less expensive to cultivate with much faster growth rates. However, once additional complexities are required, for example, intricate post-translational modifications of proteins, these simple cell systems do not have the correct modification machinery and therefore cannot be used. Or so many of the community still believe.

Over two-thirds of proteins in drug development are post-translationally modified by the addition of sugar groups [2]. The process, referred to as glycosylation, is crucial for the protein function and efficacy. CHO cells are the favored host as they produce a glycosylation pattern similar to humans, and therefore, the protein therapeutic product is less likely to cause an immunogenic response in patients [3]. This specific advantage has meant research into improving controllability, robustness and productivity in CHO cells has been intense. The majority of improvements in CHO cell protein yields have been at the process level, although, recently, the release of the full genetic code opens the door to more directed metabolic engineering to improve cellular characteristics [4]. One such feature is the heterogeneity in glycan patterns that occur in CHO-produced proteins, something that can be triggered by even a small change in bioreactor conditions [5].

However, we are in a synthetic age, where understanding the complexity of biological systems is being paralleled by attempts to synthetically build relatively simple cell chassis and adding desirable functional components: this is where the old laboratory favorite *E. coli* comes in. It is well accepted that these cells do not naturally perform N-glycosylation, the most common type of sugar-related post-translational modification in protein therapeutics. Yet, in 2002, the discovery and characterization of a bacterial N-glycosylation pathway in *Campylobacter jejuni*, culminated in its functional transfer into *E. coli* [6]. The module, a

10.4155/PBP.13.26 ISSN 2048-9145

12-gene pathway (pgl), includes enzymes involved in sugar biosynthesis and transfer. The *E. coli* cells were able to successfully add a sugar heptasaccharide to two glycosylation sites in AcrA, a target protein also taken from *C. jenuni*. Although yields of the glycoprotein were very poor and the glycan structure was bacterial, the product was homogenous. Perhaps this is unsurprising considering the host cell does not contain hundreds of glycosylation genes. Conceptually, the door had been opened to create a bacterial cell capable of human-type glycosylation and take the bioprocessing community by surprise.

The challenge to improve glycoprotein productivity in *E. coli* and genetically engineer the cells to produce the correct human glycans is huge. Although there are similarities between prokaryotic and eukaryotic N-glycosylation, there are distinct differences too. For example, prokaryotes have an extended N-glycosylation site sequon (by two amino acids) [7]. In addition, N-glycosylation in prokaryotes is truly post-translational; the protein folds and subsequently the sugars are added; however, in eukaryotes it is more co-translational. Bearing these differences in mind, progress has been rapid.

By removing the undesirable genes from the *C. jejuni* glycosylation module, researchers were able to demonstrate that WecA, a protein native to *E. coli* and involved in lipopolysaccharide synthesis, was able to attach the same first saccharide, GlcNAc, as present in human therapeutic proteins. After further addition of GalNAc glycans *in vivo*, the protein was purified and chemical synthesis employed to trim the sugar structure to the initial saccharide, whilst adding the eukaryotic core N-pentasaccharide, $Man_3GlcNAc_2$, *in vitro* [8]. Following on from this and the advancements in complex eukaryotic protein production in *E. coli* [9], some of the prokaryotic glycosyltransferases were replaced with the expression of those from *S. cerevisiae* [10] that were known to be expressible in *E. coli* [11–13]. The resulting glycosylation construct was able to sequentially build the initial five saccharides of the core human type glycan, $mannose_3$-*N*-$acetylglucosamine_2$, using both eukaryotic and prokaryotic glycosyltransferases, successfully transferring it to the consensus bacterial recognition site, via the *C. jejuni* oligosaccharyltransferase, PglB [14]. This achievement shows the potential in engineering the *C. jejuni* pgl construct to accommodate various other external genes from eukaryotic organisms such as yeast, whilst maintaining essential *C. jejuni* genes such as *PglB*, and *PglK* to achieve successful bacterial glycosylation but with a potentially eukaryotic glycan. The pgl construct found in *C. jejuni* is constantly being manipulated to suit our needs and produce the eukaryotic glycan in *E. coli*. Once successful, this glycosylation construct could be transferred into other expression systems and ultimately into the 'synthetic cell' if glycosylation is required in the final product.

/// Attempts to increase understanding and improve glycosylation efficiency are essential for the *Escherichia coli* glycosylation toolbox option to become an industrial reality. ///

The bacterial glycoengineering field is ever expanding; research is being conducted with various target proteins, for example, antibody fragments and endotoxins [15,16], in different *E. coli* strains, whilst using a variety of glycosylation modules [17]. Attempts to increase understanding and improve glycosylation efficiency are essential for the *E. coli* glycosylation toolbox option to become an industrial reality. Chen *et al.* varied the glycosylation consensus sequence and found that DQNAT was the optimal acceptor substrate for PglB [18]. In an attempt to reduce metabolic burden, codon optimizing the PglB was seen to increase glycosylation by approximately 100% [19]. At present, glycosylation efficiencies and overall glycoprotein yields are relatively poor, often in the region of 1–20% and just milligrams per liter, respectively [6,8,20]. An inverse metabolic engineering strategy, where extra chromosomal *E. coli* fragments were present in the cell, identified a selection of native proteins that when expressed could increase glycosylation efficiency [17]. In more of a forward metabolic engineering approach, enzymes and pathways leading to potential bottlenecks were highlighted using a proteomic discovery methodology that incorporated probabilistic modeling [20]. By driving flux through pathways to aid the metabolic process of producing precursors for the glycosylation machinery, the amount of glycosylated protein being produced increased by as much as 300% [20]. These changes to the host cell would ultimately have to be combined to quantify their cumulative impact. An important aspect to note here is that the methodology used to quantify glycoprotein production and calculate glycosylation efficiencies, needs to be more transparent to allow useful comparisons. Detailed culture conditions are crucial for this

evaluation and western blots are only semi-quantitative at best. Moreover, it is common that absolute protein or glycoprotein quantities are not calculated or revealed. Advances in MS are allowing for more accurate site assignments and absolute quantization [17,20] and these modern approaches should be used over semi-quantitative methods to enhance our understanding and provide robust comparisons.

What is clear is that the demand for protein therapeutics is rising. Concurrently, there are shifts in the outlook for the biopharmaceutical industry, for example, the growing popularity of biosimilars and small molecule drugs. At the same time, technology to gain deeper insight into the workings of host cells as well as tools in synthetic biology are developing rapidly. Ultimately, standing still in the ever-changing landscape of the biopharmaceutical industry could have ominous consequences, and expansion of the *E. coli* toolbox remains a financially attractive option.

Financial & competing interests disclosure

The authors acknowledge the UK Biotechnology and Biological Sciences Research Council for research funding under the BRIC program. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1 Matasci M, Hacker DL, Baldi L, Wurm FM. Recombinant therapeutic protein production in cultivated mammalian cells: current status and future prospects. *Drug Discov. Today* 5(2), e37–e42 (2008).

2 Sethuraman N, Stadheim TA. Challenges in therapeutic glycoprotein production. *Curr. Opin. Biotechnol.* 17(4), 341–346 (2006).

3 Comstock LE, Kasper DL. Bacterial glycans: key mediators of diverse host immune responses. *Cell* 126(5), 847–850 (2006).

4 Xu X, Nagarajan H, Lewis NE *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* 29(8), 735–741 (2011).

5 Hossler P, Khattak SF, Li ZJ. Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* 19(9), 936–949 (2009).

6 Wacker M, Linton D, Hitchen PG *et al.* N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* 298(5599), 1790–1793 (2002).

7 Kowarik M, Young NM, Numao S *et al.* Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* 25(9), 1957–1966 (2006).

8 Schwarz F, Huang W, Li C *et al.* A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nat. Chem. Biol.* 6(4), 264–266 (2010).

9 Skretas G, Carroll S, Defrees S, Schwartz M, Johnson K, Georgiou G. Expression of active human sialyltransferase ST6GalNAcI in *Escherichia coli*. *Microb. Cell Fact.* 8(1), 50 (2009).

10 Valderrama-Rincon JD, Fisher AC, Merritt JH *et al.* An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. *Nat. Chem. Biol.* 8(5), 434–436 (2012).

11 O'Reilly MK, Zhang G, Imperiali B. *In vitro* evidence for the dual function of Alg2 and Alg11: essential mannosyltransferases in N-linked glycoprotein biosynthesis. *Biochemistry* 45(31), 9593–9603 (2006).

12 Couto JR, Huffaker TC, Robbins PW. Cloning and expression in *Escherichia coli* of a yeast mannosyltransferase from the asparagine-linked glycosylation pathway. *J. Bio. Chem.* 259(1), 378–382 (1984).

13 Wang X, Weldeghiorghis T, Zhang G, Imperiali B, Prestegard JH. Solution structure of Alg13: the sugar donor subunit of a yeast N-acetylglucosamine transferase. *Structure* 16(6), 965–975 (2008).

14 Kowarik M, Young NM, Numao S *et al.* Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* 25(9), 1957–1966 (2006).

15 Ihssen J, Kowarik M, Dilettoso S, Tanner C, Wacker M, Thony-Meyer L. Production of glycoprotein vaccines in *Escherichia coli*. *Microb. Cell Fact.* 9, 61 (2010).

16 Schwarz F, Huang W, Li C *et al.* A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nat. Chem. Biol.* 6(4), 264–266 (2010).

17 Pandhal J, Woodruff L, Jaffe S *et al.* Inverse metabolic engineering to improve *Escherichia coli* as an N-glycosylation host. *Biotechnol. Bioeng.* doi:10.1002/bit.24920 (2013) (Epub ahead of print).

18 Chen MM, Glover KJ, Imperiali B. From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in *C. jejuni*. *Biochemistry (Mosc.)* 46(18), 5579–5585 (2007).

19 Pandhal J, Desai P, Walpole C, Doroudi L, Malyshev D, Wright PC. Systematic metabolic engineering for improvement of glycosylation efficiency in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 419(3), 472–476 (2012).

20 Pandhal J, Ow SY, Noirel J, Wright PC. Improving N-glycosylation efficiency in *Escherichia coli* using shotgun proteomics, metabolic network analysis, and selective reaction monitoring. *Biotechnol. Bioeng.* 108(4), 902–912 (2011).