6 Jaber, M. *et al*. (1996) Essential role of β-adrenergic receptor kinase-1 in cardiac development and function. *Proc. Natl. Acad. Sci. U. S. A.* 93, 12974–12979
7 Rockman, H.A. *et al*. (1998) Control of myocardial contractile function by the level of β-adrenergic receptor kinase-1 in gene-targeted mice. *J. Biol. Chem.* 273, 18180–18184
8 Ungerer, M. *et al*. (1993) Altered expression of β-adrenergic receptor kinase and β1-adrenergic receptors in the failing heart. *Circulation* 87, 454–463
9 Rockman, H.A. *et al*. (1998) Expression of a beta-adrenergic receptor kinase 1 inhibitor prevents the development of myocardial failure in gene-targeted mice. *Proc. Natl. Acad. Sci. U. S. A.* 95, 7000–7005
10 Harding, V.B. *et al*. (2001) Cardiac beta ARK1 inhibition prolongs survival and augments beta blocker therapy in a mouse model of severe heart failure. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5809–5814
11 Freeman, K. *et al*. (2001) Alterations in cardiac adrenergic signaling and calcium cycling differentially affect the progression of cardiomyopathy. *J. Clin. Invest.* 107, 967–974
12 Williams, M.L. and Koch, W.J. (2004) Viral-based myocardial gene therapy approaches to alter cardiac function. *Annu. Rev. Physiol.* 66, 49–75
13 Iaccarino, G. *et al*. (1998) Reciprocal *in vivo* regulation of myocardial G protein-coupled receptor kinase expression by β-adrenergic receptor stimulation and blockade. *Circulation* 98, 1783–1789

0167-7799/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tibtech.2004.08.004

Tailor-made glycoproteins

Linda C. Hsieh–Wilson

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

Posttranslational modifications are a fundamental mechanism for the regulation of cellular physiology and function. A recent paper by Zhang *et al.* provides a novel strategy for the generation of homogeneous glycoproteins. The ability to install covalent modifications site-specifically into proteins holds tremendous promise for deciphering the role of posttranslational modifications and has exciting implications for the development of protein therapeutics.

Given that the complexity of higher organisms is encoded in a remarkably small number of genes, attention has focused on other mechanisms to account for the diversity of biological systems. Protein posttranslational modifications (PTMs) represent a major source of variation – the covalent attachment of phosphate, lipid, and other chemical groups to amino acid side chains extends the capabilities of proteins and provides selective and temporal control over protein function. At a cellular level, PTMs have been shown to regulate vital processes such as transcription, cell-cycle progression, and programmed cell death.

Protein glycosylation, the most abundant and complex form of PTM, requires an estimated 1% of mammalian genes [1] and distinguishes eukaryotic cells from simpler prokaryotic systems. Recent studies are challenging early views of carbohydrates as non-specific, static structures and are revealing exciting physiological roles for glycosylation. For example, Wei *et al.* have shown that HIV evades the immune system by evolving a dynamically changing shield of carbohydrates [2]. Glycosaminoglycans, another class of carbohydrates, play crucial roles during development, with genetic disruption of glycosaminoglycan biosynthesis associated with severe growth defects [3]. The complex sulfation patterns present in glycosaminoglycans, which are tissue-specific, tightly regulated, and crucial for growth factor activation, suggest an intricate relationship between carbohydrate structure and biological function [3]. In the intracellular context, dynamic forms of glycosylation have been identified, such as the addition of β-*N*-acetylglucosamine to serine or threonine residues of proteins (*O*-GlcNAc glycosylation) [4]. Recent studies have linked *O*-GlcNAc glycosylation to the regulation of transcription factors such as cAMP-responsive element binding protein (CREB) [5] and the proteasome [6], raising the exciting possibility that this carbohydrate modification plays a fundamental role in cellular communication and protein function.

Despite many important contributions, the diverse roles of glycosylation and other covalent modifications are only beginning to be understood. Detailed studies of their biological effects have been hindered by the dynamic nature and complexity of PTMs *in vivo*. For instance, the presence of a given modification can depend on the cell type or status (e.g. stimulated vs non-stimulated, normal vs diseased). It can also be restricted to subcellular compartments or protein subpopulations and can occur in combination with other PTMs. Except for protein phosphorylation, in which an aspartic or glutamic acid can sometimes mimic the phosphorylated residue, no natural amino acid can serve as an effective surrogate for glycosylation and other PTMs. Moreover, efforts to purify modified proteins for biochemical studies generally require abundant sources and extensive purification. Even when available, the purified proteins are frequently heterogeneous, bearing incomplete or multiple modifications.

A recent paper by Schultz and colleagues provides a novel solution to these problems and should considerably

Corresponding author: Linda C. Hsieh–Wilson (lhw@caltech.edu).
Available online 28 August 2004

enhance our ability to analyze and manipulate glycoprotein structure and function [7]. Zhang *et al.* hijacked the translational machinery of *Escherichia coli* to selectively incorporate a glycosylated amino acid (serine modified by β-GlcNAc, Figure 1) into the protein myoglobin. The strategy produced a homogeneous, structurally defined glycoprotein with the carbohydrate installed at a pre-selected site.

Hijacking the translational machinery

The key to their approach was the evolution of an orthogonal synthetase-tRNA pair that genetically encodes a glycosylated amino acid in response to the amber codon TAG (Figure 1). Synthetases usually charge tRNA molecules with one of the twenty naturally occurring amino acids. To coax a synthetase into accepting an unnatural glycosylated amino acid, residues within the tyrosine-binding site of *Methanococcus jannaschii* tyrosyl tRNA synthetase (TyrRS) were randomized to create a large library of mutants. This library was then subjected to a positive selection to enrich TyrRS mutants competent to transfer the glycosylated amino acid to the *M. jannaschii* suppressor tRNA, followed by a negative selection to delete clones that transfer endogenous amino acids. After several rounds of selection, a TyrRS mutant was identified that selectively incorporated β-GlcNAc-serine into myoglobin. With the evolved TyrRS synthetase, milligram quantities of homogeneous glycoprotein could be generated by co-expression of the synthetase, suppressor tRNA, and TAG-mutated myoglobin genes in *E. coli* in medium containing the glycosylated amino acid. Notably, several laboratories have demonstrated the power of hijacking the translational machinery to provide insights into protein structure and function [8–12]. By extending the strategy now to glycoproteins, Zhang *et al.* illustrate the breadth of the approach and its potential to transform our understanding of PTMs.

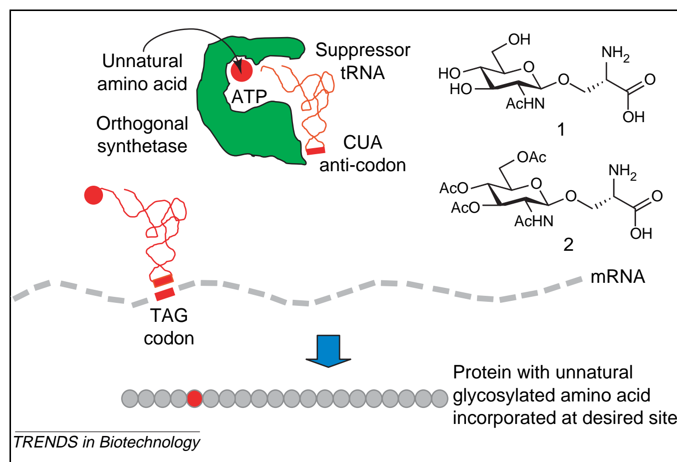

Figure 1. Strategy for the selective incorporation of glycosylated amino acids into proteins. An orthogonal tRNA synthetase capable of charging a suppressor tRNA with the glycosylated amino acid **1** was selected from a library of mutants. Milligram quantities of homogeneous glycoprotein were obtained upon co-expression of the synthetase, suppressor tRNA, and TAG-mutated myoglobin genes in *E.coli* in the presence of acetylated amino acid **2**. Acetylation enhanced transport of the glycosylated amino acid across cell membranes.

The promise of homogeneous, modified proteins

The approach developed by Schultz and co-workers has the advantages of high selectivity, efficiency, and yield. No other glycoforms or unmodified forms of myoglobin were generated. Because the overproduction of proteins in *E. coli* is well established, the approach can also be optimized to permit large-scale production of desired glycoproteins. In principle, the same synthetase could be used to incorporate a glycosylated amino acid into other proteins of interest or into multiple sites within the same protein. These glycosylated amino acids could then be further elaborated to more complex carbohydrate structures using glycosyltransferase enzymes, as illustrated by a recent study from the Schultz and Wong groups [13]. In the future, the selection strategy could be applied to evolve tRNA synthetases for other PTMs such as phosphorylation, acetylation, and methylation. Finally, extension of the approach to other codon strategies [11,12,14] could permit the introduction of multiple, different PTMs into the same protein.

The strategy reported by Zhang *et al.* represents one of several methods under active investigation for the preparation of homogeneously modified proteins. Peptides and small proteins are accessible via chemical synthesis, but issues of cost and homogeneity complicate application to larger proteins or commercial production. Chemical methods have also been developed for the introduction of PTM mimics via site-directed mutagenesis and chemoselective labeling reagents [15,16]. Semi-synthetic approaches, such as expressed protein ligation, have permitted modified synthetic peptides containing phosphorylated, glycosylated or lipidated amino acids to be coupled to recombinant proteins [17]. In a recent impressive achievement, Kochendoerfer *et al.* used native chemical ligation to construct the 166-amino acid erythropoietin protein with two monodisperse polymers attached at natural glycosylation sites [18]. The resulting protein displayed potent hematopoietic activity and an extended time action *in vivo*.

The ability to generate homogeneous proteins bearing specific modifications has tremendous potential for advancing our understanding of PTMs. The incorporation of covalent modifications at specific sites circumvents the challenges of heterogeneous biochemical preparations. The strategy should also permit studies of the interplay among different modifications, such as the influence of one modification on the location or kinetics of another. For instance, phosphorylation and *O*-GlcNAc glycosylation can occur at distinct sites within the protein, yet functionally oppose one another [4,5]. In such cases, it will be interesting to examine whether glycosylation antagonizes the kinetics of phosphorylation and vice versa. With the development of methods for incorporating multiple PTMs into proteins, the combinatorial effects of distinct modifications within a given protein can also be addressed [19]. Such methods would be useful, for example, to deconvolute the influence of histone acetylation, methylation, and phosphorylation on transcription. Finally, recent extension of the unnatural amino acid technology to yeast and mammalian cells [20–22] could permit the expression of modified, activated forms of

proteins or dominant-negative mutants to understand PTMs in a cellular context.

Methods for generating homogenously modified proteins also have implications for protein therapeutics. Nearly all therapeutic glycoproteins exist as heterogeneous mixtures of glycoforms, which can have profound effects on the stability, solubility, and pharmacological properties of the protein [23]. The ability to install carbohydrates at specific sites could transform the way in which therapeutic glycoproteins are discovered, developed, and manufactured. In this regard, parallels can be drawn to small molecule pharmaceuticals, in which the ability to evaluate single enantiomers established new standards for the industry. If similar stringencies are applied to protein therapeutics, the development of methods for the production of homogeneous proteins will become a major focus. Finally, the potential to fine-tune the properties of proteins (e.g. potency, stability, time action), much in the same way the medicinal chemist improves the performance of small molecule drugs, could lead to significant breakthroughs. By providing a new method for producing homogeneous glycoproteins, the work of Zhang *et al.* represents an important milestone along this path.

References

1 Lowe, J.B. and Marth, J.D. (2003) A genetic approach to mammalian glycan function. *Annu. Rev. Biochem.* 72, 643–691
2 Wei, X.P. *et al.* (2003) Antibody neutralization and escape by HIV-1. *Nature* 422, 307–312
3 Selleck, S.B. (2000) Proteoglycans and pattern formation - sugar biochemistry meets developmental genetics. *Trends Genet.* 16, 206–212
4 Slawson, C. and Hart, G.W. (2003) Dynamic interplay between O-GlcNAc and O-phosphate: the sweet side of protein regulation. *Curr. Opin. Struct. Biol.* 13, 631–636
5 Lamarre-Vincent, N. and Hsieh-Wilson, L.C. (2003) Dynamic glycosylation of the transcription factor CREB: A potential role in gene regulation. *J. Am. Chem. Soc.* 125, 6612–6613
6 Zhang, F.X. *et al.* (2003) O-GIcNAc modification is an endogenous inhibitor of the proteasome. *Cell* 115, 715–725
7 Zhang, Z.W. *et al.* (2004) A new strategy for the synthesis of glycoproteins. *Science* 303, 371–373
8 Beene, D.L. *et al.* (2003) Unnatural amino acid mutagenesis in mapping ion channel function. *Curr. Opin. Neurobiol.* 13, 264–270
9 Mendel, D. *et al.* (1995) Site-directed mutagenesis with an expanded genetic code. *Annu. Rev. Biophys. Biomol. Struct.* 24, 435–462
10 Link, A.J. *et al.* (2003) Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* 14, 603–609
11 Hohsaka, T. and Sisido, M. (2002) Incorporation of non-natural amino acids into proteins. *Curr. Opin. Chem. Biol.* 6, 809–815
12 Anderson, R.D. *et al.* (2002) Fluorescence resonance energy transfer between unnatural amino acids in a structurally modified dihydrofolate reductase. *J. Am. Chem. Soc.* 124, 9674–9675
13 Liu, H.T. *et al.* (2003) A method for the generation of glycoprotein mimetics. *J. Am. Chem. Soc.* 125, 1702–1703
14 Anderson, J.C. *et al.* (2004) An expanded genetic code with a functional quadruplet condon. *Proc. Natl. Acad. Sci. U. S. A.* 101, 7566–7571
15 Davis, B.G. (2003) Chemical modification of biocatalysts. *Curr. Opin. Biotechnol.* 14, 379–386
16 Hang, H.C. and Bertozzi, C.R. (2001) Chemoselective approaches to glycoprotein assembly. *Acc. Chem. Res.* 34, 727–736
17 Muir, T.W. (2003) Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* 72, 249–289
18 Kochendoerfer, G.G. *et al.* (2003) Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. *Science* 299, 884–887
19 Khidekel, N. and Hsieh-Wilson, L.C. (2004) A 'molecular switchboard' - covalent modifications to proteins and their impact on transcription. *Org. Biomol. Chem.* 2, 1–7
20 Chin, J.W. *et al.* (2003) An expanded eukaryotic genetic code. *Science* 301, 964–967
21 Monahan, S.L. *et al.* (2003) Site-specific incorporation of unnatural amino acids into receptors expressed in mammalian cells. *Chem. Biol.* 10, 573–580
22 Kohrer, C. *et al.* (2003) A possible approach to two different unnatural site-specific insertion of amino acids into proteins in mammalian cells via nonsense suppression. *Chem. Biol.* 10, 1095–1102
23 Haselbeck, A. (2003) Epoetins: differences and their relevance to immunogenicity. *Curr. Med. Res. Opin.* 19, 430–432

0167-7799/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tibtech.2004.08.009

An array of diverse microbial genomes

Melinda M. Pettigrew

Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT, 06520-8034, USA

Methods for comparing gene frequencies across large, epidemiologically defined bacterial collections are limited. A novel microarray technology has been developed called 'library on a slide'. In this technology, hundreds of entire microbial genomes are arrayed, rather than sequences of a single genome or sets of genes. These slides can then be probed for the presence of specific genes allowing researchers to draw inferences regarding important differences between related strains that differ in their pathogenic potential.

Within a given bacterial species, individual strains often exhibit incredible levels of diversity and can differ in genome make-up by up to 20% [1]. Genome sizes can vary by as much as 1 Mb among *E. coli* isolates, translating into a difference of 800 genes between the largest and smallest strains [2]. Zhang *et al.* [3] at the University of Michigan have designed a novel DNA microarray technology for exploring genetic diversity among hundreds of closely related bacterial

Corresponding author: Melinda M. Pettigrew (melinda.pettigrew@yale.edu).
Available online 28 August 2004