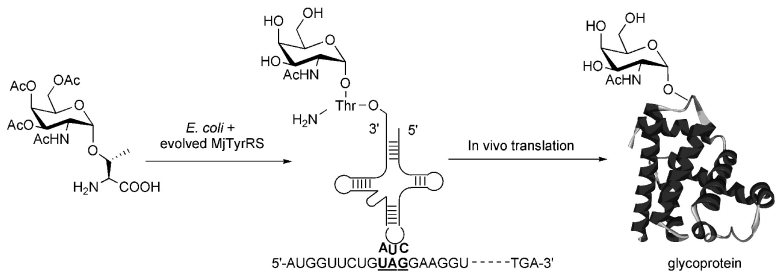
Communication

Site-Specific Incorporation of the Mucin-Type *N*-Acetylgalactosamine-□-*O*-threonine into Protein in *Escherichia coli*

Ran Xu, Sarah R. Hanson, Zhiwen Zhang, Yu-Ying Yang, Peter G. Schultz, and Chi-Huey Wong

J. Am. Chem. Soc., **2004**, 126 (48), 15654-15655• DOI: 10.1021/ja044711z • Publication Date (Web): 11 November 2004



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 3 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
COMMUNICATIONS

Published on Web 11/11/2004

Site-Specific Incorporation of the Mucin-Type *N*-Acetylgalactosamine-α-*O*-threonine into Protein in *Escherichia coli*

Ran Xu, Sarah R. Hanson, Zhiwen Zhang, Yu-Ying Yang, Peter G. Schultz,* and Chi-Huey Wong*

Department of Chemistry, and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 Torrey Pines Road, La Jolla, California 92037

Received September 1, 2004; E-mail: wong@scripps.edu

Glycosylation is one of the most prevalent posttranslational modifications of proteins and plays an important role in many biological processes.[1] Natural glycoproteins usually exist as heterogeneous glycoforms, and this complicates studies of their structure and function.[2] Consequently, considerable effort has focused on the methods to generate homogeneous glycoproteins including chemical and enzymatic synthesis, in vitro translation, and pathway engineering.[3] Recently, Zhang et al. demonstrated that *N*-acetylglucosamine (GlcNAc) *β*-linked to the hydroxyl group of serine can be site-specifically incorporated into proteins in *Escherichia coli* during protein translation in vivo.[4] This process required the evolution of an orthogonal *Methanococcus jannaschii* Tyr-tRNA synthetase (MjTyrRS) and Tyr-tRNA[10,11] pair that does not cross-react with endogenous *E. coli* tRNAs and aminoacyl-tRNA synthetases, but selectively inserts the desired glycosyl amino acid into proteins in response to the amber codon TAG.[4]

Herein, we report that *N*-acetylgalactosamine α-*O*-threonine (GalNAc-α-Thr, **1**) can also be genetically encoded in *E. coli* using an evolved MjTyrRS pair (Figure 1). The GalNAc saccharide α-linked to the hydroxyl group of threonine or serine represents the core unit, also known as the Tn antigen, in mucin-type glycoproteins. In eukaryotes, mucins comprise the most prevalent type of *O*-glycans, constituting a polydisperse group of glycoproteins and proteoglycans involved in inflammation and cellular recognition.[12] A wide variety of mucin-type core structures are generated from the GalNAc-α-serine/threonine motif by glycosylation at the C-3 and/or C-6 hydroxyl groups of GalNAc. It is essential for the study of mucin-type glycans to first establish this key protein−carbohydrate linkage so that it can then be further elaborated into a homogeneous glycoform of interest.

To enhance the cell membrane permeability of the unprotected glycosyl amino acid **1**, peracetylated GalNAc-α-Thr (**2**) was chemically synthesized using established methods.[13] Previous experiments with glycosyl amino acids have shown that the acetyl groups will be cleaved by nonspecific esterases once in the cytosol of cells.[4] A mutant MjTyrRS synthetase specific for **1** was evolved from two previously described MjTyrRS mutant libraries[4] using a positive selection based on suppression of: (1) the nonessential mutation Asp112TAG in chloramphenicol acetyl transferase gene, which confers antibiotic resistance, and (2) the nonessential mutations Met1TAG and Gln107TAG in T7 RNA polymerase gene, which drives the expression of a green fluorescence protein, GFPuv.[5,14] A negative selection based on suppression of amber codons at three permissive positions in the toxic barnase gene, Gln2, Asp44, and Gly65, was used to remove MjTyrRS mutants that accept endogenous natural amino acids.[11] After four rounds of positive selection and three rounds of negative selection, several MjTyrRS mutants were isolated that allowed cell growth at high concentrations of chloramphenicol when supplied with **2** (Supporting Information, Figure 1). The isolated MjTyrRS clones that accept

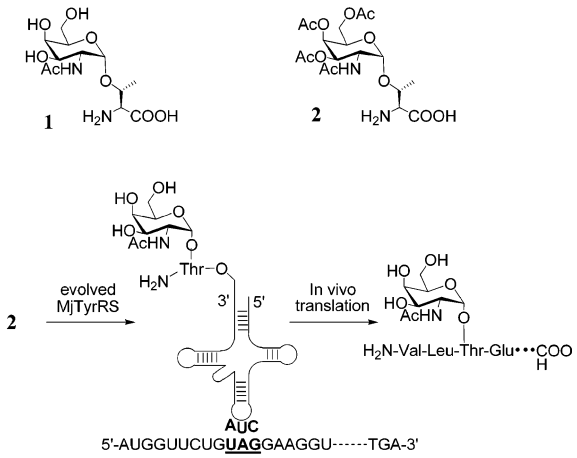


Figure 1. GalNAc-α-threonine (**1**) can be incorporated during biosynthesis into recombinant glycoproteins in *E. coli*. Synthetic glycosyl amino acid **1** is charged onto Tyr-tRNA$_{CUA}$ via evolved MjTyrRS (AH1 and C10F) and is site-specifically incorporated into glycoproteins by readthrough of nonsense amber codon.

GalNAc-α-Thr were sequenced and have the following mutations: AH1: Tyr32Phe and Ala67Pro; C8F: Tyr32Gln, Ala67Pro, Gly163Cys, and Ala167Val; C10F: Tyr32Ala, Ala67Ser, His70Pro, and Leu98Ile; D10B: Tyr32Leu, Ala67Thr, His70Lys, Val149Ile, Gln155Ser, Asp158Val, and Ala167Val. Notably, Tyr32, which establishes one of the two key hydrogen bonds to the native tyrosine substrate, is mutated in all of the GalNAc-α-Thr mutants.[15] The other major hydrogen bond donor, Asp158, is mutated in the GlcNAc-*β*-Ser synthetase, I-90.[4,15] Most of the other mutations in the glycosyl-accepting synthetases were found to occur outside the active-site, based on modeling. Current efforts toward obtaining crystals of glycosyl amino acid-bound mutant MjTyrRS should yield insights into how these mutations affect substrate binding and should allow the generation of additional focused mutant libraries.

To confirm that the phenotypes demonstrated by the above MjTyrRS mutants result from the incorporation of **2** in response to TAG, a myoglobin mutant was coexpressed with each of the evolved MjTyrRS clones AH1 and C10F and the mutant Tyr-tRNA$_{CUA}$ gene in the presence of **2**. The codon for the fourth amino acid in the myoglobin gene was mutated to the amber codon TAG, and a 6X His tag was added to the C-terminal of the gene to facilitate the purification of the full-length protein by Ni^{2+}-NTA affinity chromatography. In the presence of **2**, clones AH1 and C10F yielded 2 and 4 mg/L of purified protein, respectively, while the yield of wild-type myoglobin was approximately 5.5 mg/L under identical conditions. Full-length protein was not produced in the absence of either tRNA$_{CUA}$ or mutant synthetase; however, Coomassie-stained SDS-PAGE analysis revealed light myoglobin bands (roughly 10% of the positive sample) when grown in the absence of **2**, indicating some residual background suppression (Figure 2).

10.1021/ja044711z CCC: $27.50

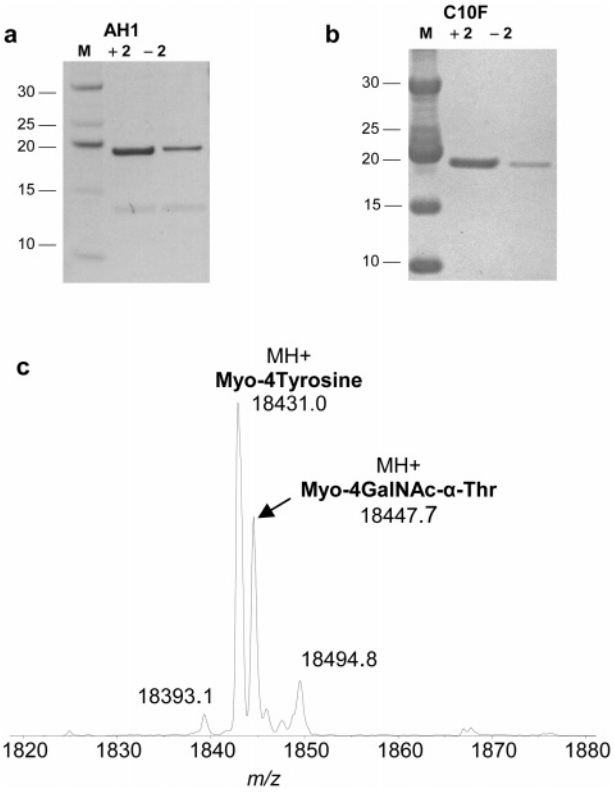


Figure 2. Protein levels for the expression of the myoglobin 4TAG mutant gene with the evolved AH1 (a) and C10F (b) MjTyrRS are higher in the presence (+) than in the absence (−) of **2**. M: protein markers (kDa). (c) High-resolution ESI-TOF of myoglobin expressed with C10F and glycosyl amino acid **2** shows peaks for GalNAc-α-Thr incorporation (expected average MH^+ 18448.8) and tyrosine incorporation (expected average MH^+ 18431.2). Sample mixtures were separated by reverse-phase HPLC first to remove mainly small molecules before mass analysis

The background activity found with clones C10F and AH1 is consistent with their high chloramphenicol-resistant phenotypes in the absence of amino acid, which at 70 μg/uL is greater than 2 times the background suppression of the previously identified GlcNAc-β-Ser clone (30 μg/uL).[4] Analysis by coupled reverse-phase liquid chromatography electrospray ionization time-of-flight (LC ESI-TOF) mass spectrometry revealed that myoglobin samples grown in the presence of **2** afforded a mixture of tyrosine and GalNAc-α-Thr-incorporated myoglobin (Figure 2c). Interestingly, tyrosine-incorporated myoglobin exists as the full-length protein with a mass value of 18431.0 (expected average MH^+ 18431.2), while glycomyoglobin was observed with the initial methionine cleaved (18447.7, expected average MH^+ 18448.8). A signal corresponding to tri-*O*-acetyl-GalNAc-modified myoglobin was not observed, confirming that the *O*-acetyl groups in **2** were hydrolyzed in vivo. Future determination of the structures of the mutant synthetases may allow new libraries to be designed that lead to clones with improved specificities and activities. Nonetheless, the current clones AH1 and C10F allow for the generation of sufficient quantities of glycoprotein to be purified in useful amounts.

Further confirmation of GalNAc-α-Thr incorporation was obtained using an enzyme-linked lectin-binding assay using lectins from *Vicia villosa* (VVL) and *Dolichos biflorus* (DBL). VVL is well-known for its specificity toward the Tn antigen, while DBL is frequently used to detect α-linked GalNAc residues.[16] Binding of biotinylated lectins to myoglobin and glycomyoglobin samples was measured by streptavidin−alkaline phosphatase activity on the colorimetric *p*-nitrophenyl phosphate substrate. When treated with VVL and DBL, 10- and 5-fold signal increases, respectively, were observed for glycomyoglobin versus equivalent concentrations of tyrosine-myoglobin (Supporting Information, Figure 2). These data support that incorporation of GalNAc-α-Thr into myoglobin has been accomplished during in vivo biosynthesis.

In summary, the glycosyl amino acid, GalNAc-α-Thr (**1**), has been site-specifically incorporated into a protein in *E. coli*. GalNAc-α-Thr forms the structural foundation of mucin-type *O*-glycosylation, the most common *O*-linked glycosylation in eukaryotic proteins. We have developed a general approach to co-translationally incorporate glycosyl amino acids into proteins in vivo which provides an expedient route to prepare homogeneous glycoproteins. The monosaccharide attached to protein in its natural configuration provides a starting point for further extension of the carbohydrate chain using purified or recombinant enzyme sources.[3a,4] On the basis of this work, we hope to develop a general biosynthetic approach for preparing glycoproteins with defined carbohydrate chains at positions of interest and to facilitate the structural and functional studies of glycoproteins.

Acknowledgment. We are grateful to Professor J. A. Loo at UCLA and Professors Yu-Ju Chen and Chung-Lin Liao at Academia Sinica, Taiwan, for high-resolution MS data. S.R.H. is grateful for pre-doctoral support from the ARCS foundation. We thank NIH for financial support of this research.

Supporting Information Available: Experimental details. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683−720. (b) Sears, P.; Wong, C.-H. *Cell Mol. Life Sci.* **1998**, *54*, 223−252.
(2) Rudd, P. M.; Dwek, R. A. *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32*, 1−100.
(3) (a) Sears, P.; Wong, C.-H. *Science* **2001**, *291*, 2344−2350. (b) Grogan, M. J.; Pratt, M. R.; Marcaurelle, L. A.; Bertozzi, C. R. *Annu. Rev. Biochem.* **2002**, *71*, 593−634. (c) Arslan, T.; Mamaev, S. V.; Mamaeva, N. V.; Hecht, S. M. *J. Am. Chem. Soc.* **1997**, *119*, 10877−10887. (d) Hamilton, S. R.; Bobrowicz, P.; Bobrowicz, B.; Davidson, R. C.; Li, H.; Mitchell, T.; Nett, J. H.; Rausch, S.; Stadheim, T. A.; Wischnewski, H.; Wildt, S.; Gerngross, T. U. *Science* **2003**, *301*, 1244−1246.
(4) Zhang, Z.; Gildersleeve, J.; Yang, Y.-Y.; Xu, R.; Loo, J. A.; Uryu, S.; Wong, C.-H.; Schultz, P. G. *Science* **2004**, *303*, 371−373.
(5) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498−500.
(6) Zhang, Z. W.; Wang, L.; Brock, A.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2002**, *41*, 2840−2842.
(7) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 9026−9027.
(8) Liu, H.; Wang, L.; Brock, A.; Wong, C.-H.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 1702−1703.
(9) Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 56−61.
(10) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *J. Am. Chem. Soc.* **2000**, *122*, 5010−5011.
(11) Wang, L.; Schultz, P. G. *Chem. Biol.* **2001**, *8*, 883−890.
(12) Hanisch, F. A. *Biol. Chem.* **2001**, *382*, 143−149.
(13) Grundler, G.; Schmidt, R. R. *Liebigs Ann. Chem.* **1984**, 1826−1847.
(14) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nat. Biotechnol.* **2002**, *20*, 1044−1048.
(15) Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S. *Nat. Struct. Biol.* **2003**, *10*, 425−432.
(16) Wu, A. M. *FEBS Lett.* **2004**, *562*, 51−58.

JA044711Z