

## Chemical Synthesis of the $\beta$ -Subunit of Human Luteinizing (hLH) and Chorionic Gonadotropin (hCG) Glycoprotein Hormones

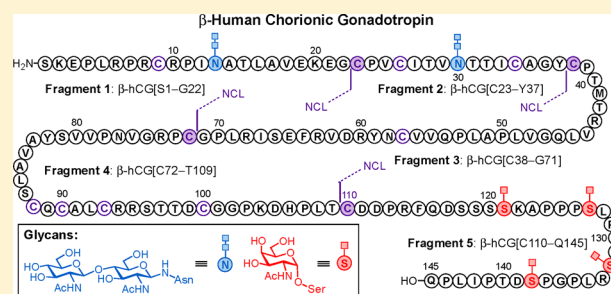
Alberto Fernández-Tejada,<sup>†,||</sup> Paul A. Vadola,<sup>†,||,§</sup> and Samuel J. Danishefsky<sup>\*,†,‡</sup>

<sup>†</sup>Laboratory for Bioorganic Chemistry, Molecular Pharmacology and Chemistry Program, Sloan Kettering Institute for Cancer Research, 1275 York Avenue, New York, New York 10065, United States

<sup>‡</sup>Department of Chemistry, Columbia University, 3000 Broadway, New York, New York 10027, United States

### S Supporting Information

**ABSTRACT:** Human luteinizing hormone (hLH) and human chorionic gonadotropin (hCG) are human glycoprotein hormones each consisting of two subunits, an identical  $\alpha$ -subunit and a unique  $\beta$ -subunit, that form noncovalent heterodimers. Structurally,  $\beta$ -hCG shares a high degree of sequence similarity with  $\beta$ -hLH, including a common N-glycosylation site at the N-terminus but differs mainly in the presence of an extended C-terminal portion incorporating four closely spaced O-linked glycans. These glycoproteins play important roles in reproduction and are used clinically in the treatment of infertility. In addition, the role of hCG as a tumor marker in a variety of cancers has also attracted significant interest for the development of cancer vaccines. In clinical applications, these hormones are administered as mixtures of glycoforms due to limitations of biological methods in producing homogeneous samples of these glycoproteins. Using the powerful tools of chemical synthesis, the work presented herein focuses on the highly convergent syntheses of homogeneous  $\beta$ -hLH and  $\beta$ -hCG bearing model glycans at all native glycosylation sites. Key steps in these syntheses include a successful double Lansbury glycosylation en route to the N-terminal fragment of  $\beta$ -hCG and the sequential installation of four O-linked glycosyl-amino acid cassettes into closely spaced O-glycosylation sites in a single, high-yielding solid-supported synthesis to access the C-terminal portion of the molecule. The final assembly of the individual glycopeptide fragments involved a stepwise native chemical ligation strategy to provide the longest and most complex human glycoprotein hormone ( $\beta$ -hCG) as well as its closely related homologue ( $\beta$ -hLH) as discrete glycoforms.



## INTRODUCTION

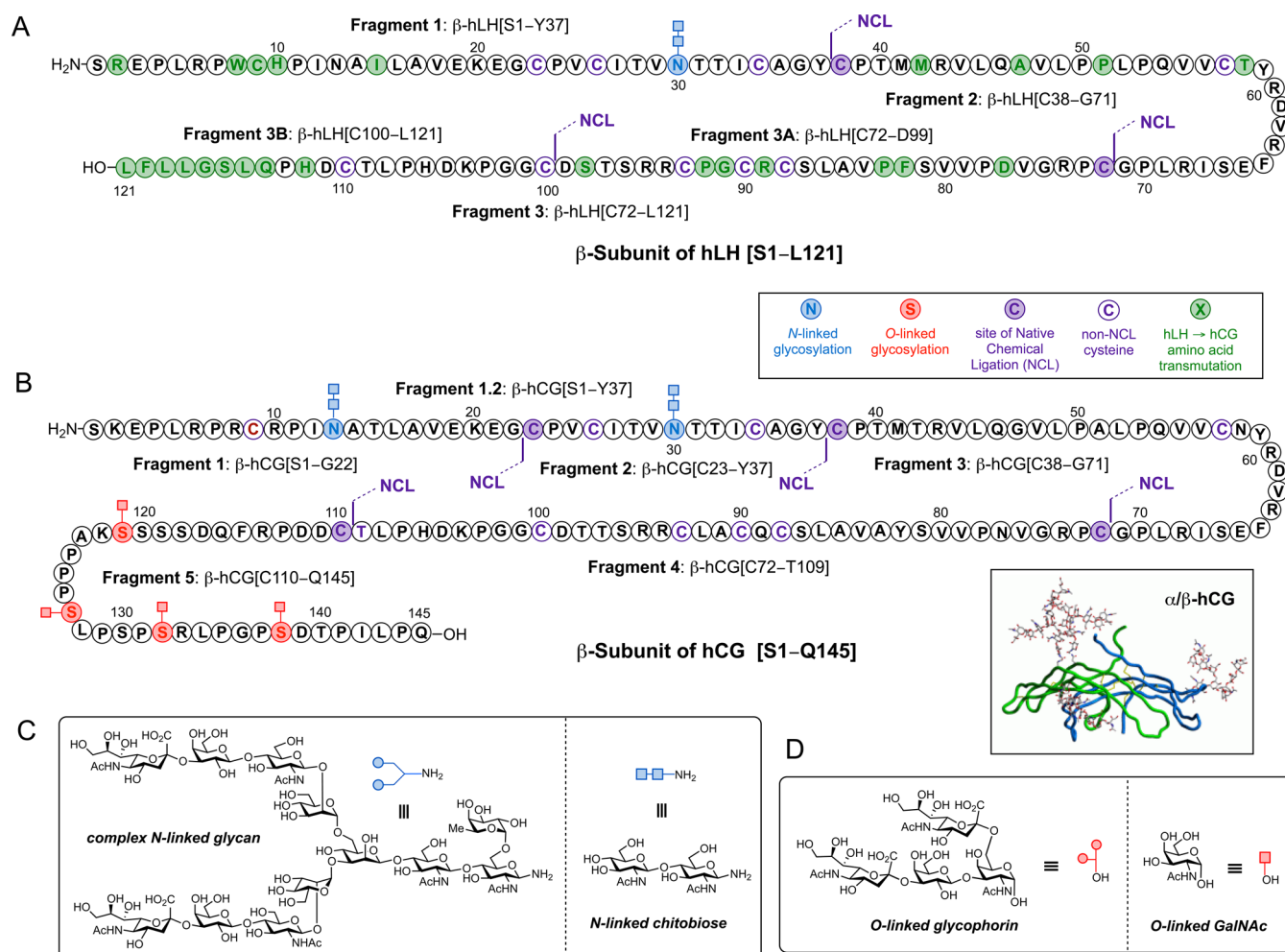
Protein glycosylation is one of the most frequent and relevant post-translational modifications.<sup>1</sup> It is present in more than half of all proteins in nature<sup>2</sup> and plays a key role in the modulation of protein properties and function.<sup>3</sup> One of the fastest growing areas in the pharmaceutical industry is the field of therapeutic glycoproteins.<sup>4</sup> They are currently being marketed as natural mixtures of glycoforms (i.e., conserved peptide backbone but highly variable as to the site and pattern of glycosylation). Our laboratory has a long-standing interest in the synthesis of homogeneous glycoproteins, as exhibited in our recently reported chemical synthesis of an erythropoietin glycoform with the wild-type polypeptide backbone and well-defined N- and O-glycosides.<sup>5</sup> We have also been interested in the human glycoprotein hormones (hGPH), which are composed of two noncovalently associated subunits, a common  $\alpha$ -subunit ( $\alpha$ -hGPH) and a hormone-specific  $\beta$ -subunit, each containing diverse oligosaccharides at defined glycosylation sites. Recently, we described the synthesis of the  $\alpha$ - and  $\beta$ -subunits of human follicle-stimulating hormone (hFSH or follitropin).<sup>6</sup> Herein, we have focused our synthetic efforts on the gonadotropins, human luteinizing hormone (hLH or lutropin) and human chorionic gonadotropin (hCG).

The gonadotropic hormones hLH and hCG are required for normal development and secretory activity of the gonads and stimulate the endocrine and gametogenic functions.<sup>7</sup> hLH is produced by gonadotroph cells in the pituitary gland and plays a key role in the reproductive process, including regulating the menstrual cycle, triggering ovulation and development of the corpus luteum, and stimulating the production of testosterone.<sup>8</sup> hCG is produced primarily in the human placenta and has similar physiological functions to hLH (i.e., upregulating progesterone and testosterone production and inducing ovulation and spermatogenesis).<sup>9</sup> It also plays a role in the protection of the fertilized embryo and development of the fetus during pregnancy.<sup>10</sup> In clinical settings, marketed forms of these hormones (often co-administered) are used in human reproductive medicine for the treatment of infertility and in *in vitro* fertilization.<sup>11</sup>

Interestingly, the  $\beta$ -subunit of hCG ( $\beta$ -hCG) is also overexpressed as particular glycoforms in certain types of tumors (epithelial, pancreatic, colorectal) and has been found to inhibit apoptosis in cancer tissues and promote metastasis.<sup>12</sup> Clinical

Received: April 9, 2014

Published: May 7, 2014



**Figure 1.** (A) Fully elaborated  $\beta$ -subunit of hLH with chitobiose at the N-glycosylation site (Asn30) and proposed disconnection positions. (B) Homogeneous full-length sequence of  $\beta$ -hCG bearing two chitobiose (Asn13 and Asn30) and four GalNAc moieties (Ser121, Ser127, Ser132, and Ser138) at the N- and O-glycosylation sites, respectively. Key ligation sites and envisioned peptide fragments for final assembly are also indicated. 3D structure of  $\alpha/\beta$ -hCG heterodimer (inset). (C) Structure of N-linked carbohydrates at defined glycosylation sites on hLH and hCG polypeptide backbone; elaborated dodecasaccharide (left) and chitobiose as model glycan (right). (D) Structure of representative O-linked glycans present in hCG peptide sequence: glycoporphin tetrasaccharide (left) and simpler GalNAc.

trials using anti- $\beta$ -hCG cancer vaccines have been carried out in patients with epithelial, pancreatic, and colorectal cancer.<sup>13</sup> Thus,  $\beta$ -hCG constitutes a promising target for cancer immunotherapy and antitumor vaccine development.<sup>14</sup> Furthermore, antibodies against  $\beta$ -hCG and one type of hCG glycoform (hyperglycosylated hCG) have been shown to inhibit tumor cell growth and metastasis.<sup>15</sup> There have also been studies, though conflicting, on the antiproliferative properties of hCG against the Kaposi's Sarcoma (KS) in HIV patients as well as its biological activity against HIV-1.<sup>16</sup> These conflicting studies may be due to the heterogeneity of different clinical-grade preparations of hCG, which contain mixtures of complex glycoforms.<sup>17</sup> Since certain glycoforms are more active than others, access to specific, discrete glycoforms by chemical synthesis can confer significant therapeutic advantages. Moreover, the synthesis of chemically pure glycoproteins bearing defined glycans can help determine the specific roles of glycosylation in biosynthesis and function.

Structurally, the  $\beta$ -hLH subunit consists of 121 residues with one N-linked glycan at Asn30 (Figure 1A).<sup>7</sup> The more complex  $\beta$ -subunit of hCG shares 85% sequence homology with  $\beta$ -hLH.<sup>7,18</sup> However, it has a unique serine-rich, 30 amino acid carboxy terminal extension with 4 additional sites of O-

glycosylation at these residues. Thus,  $\beta$ -hCG is made up of a 145-amino acid protein backbone bearing two N-linked glycans at Asn13 and Asn30 and four O-linked glycans at Ser121, Ser127, Ser132, and Ser138 (Figure 1B).<sup>19</sup> The carbohydrate content represents about one-third of the molecular weight of hCG, and glycosylation is of structural and functional importance, affecting both its half-life in circulation and signal transduction induced by this hormone.<sup>20</sup> Numerous investigations, mostly on hCG, suggest that the sugar chains are not required for hormone-receptor binding but play a critical role in receptor activation.<sup>21</sup> The major structural differences in the Asn-linked carbohydrates of hLH and hCG are in the terminal sugar moieties of the sialylated, biantennary, complex-type N-glycans (Figure 1C). The observed diversity of terminal glycosylation sequences suggests that the peripheral sugar residues are not essential for the biological action of these hormones.<sup>22</sup> This observation, together with the fact that structures lacking N-glycosylation exhibit almost complete loss of hormone function,<sup>23</sup> indicates that the underlying core structures may play a more relevant role in the biological activity of both hormones.<sup>24</sup> However, the entire  $\beta$ -hCG carboxy-terminus, which contains four O-glycosylation sites with mucin-type O-glycans (Figure 1D), has been found to

be a common epitope for hCG-based monoclonal antibodies, which can inhibit tumor cell growth and metastasis.<sup>25</sup> Detection of hCG isoforms by these antibodies has been used clinically to screen for Down Syndrome, to identify pregnancy disorders, and to monitor trophoblastic disease.<sup>26</sup>

In this work, we report the first total synthesis of the two longest  $\beta$ -subunits of the human glycoprotein hormones ( $\beta$ -hLH and  $\beta$ -hCG), which were accessed as pure glycoproteins bearing defined carbohydrate structures. The successful approach is exemplified herein with the use of the disaccharide chitobiose at the two N-glycosylation sites and the monosaccharide *N*-acetyl-galactosamine (GalNAc) at the four O-glycosylation sites as model glycans for our glycoprotein assembly. The system has been synthesized with protected cysteines anticipating folding and its association with the previously synthesized, common  $\alpha$ -subunit of the human glycoprotein hormones.<sup>6a</sup> Based on our extensive experience with similar glycoprotein targets, final cysteine deprotection and subsequent folding of the corresponding  $\beta$ -subunits are expected to be successfully accomplished in forthcoming studies to provide fully synthetic materials for future biological evaluation.

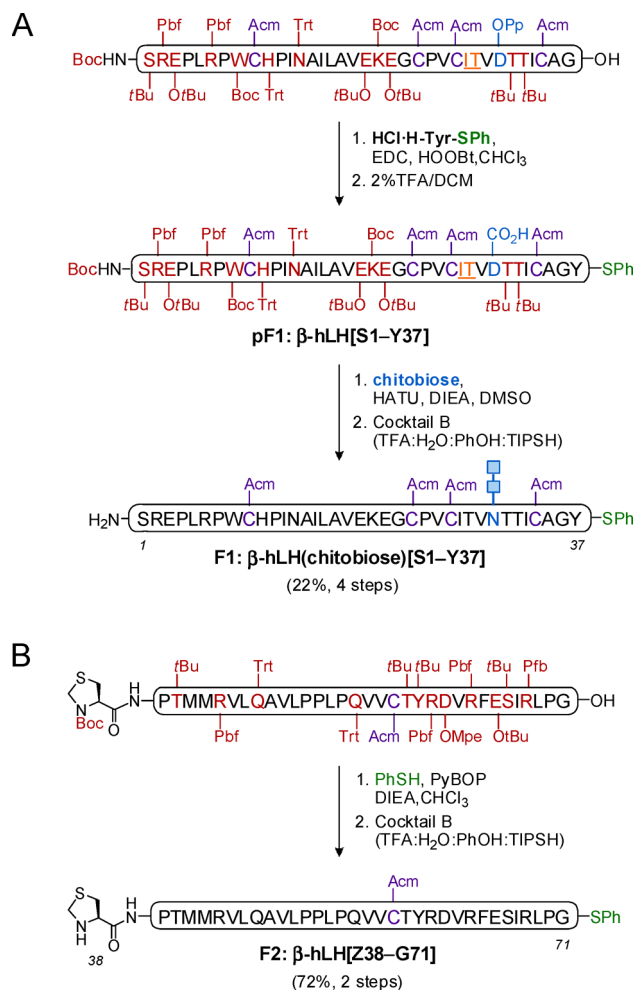
## RESULTS AND DISCUSSION

The complexity of the hCG  $\beta$ -subunit stems not only from its size (the longest of all human glycoprotein hormones) and the presence of the two N-glycosylation sites but also from the position of its unique, closely spaced four O-glycosylation sites at the carboxy-terminus of the molecule. Fortunately, the relatively high content of cysteine residues in the peptide backbone at fairly regular intervals enables the assembly of both full-length glycoproteins from appropriately selected peptide fragments using the native chemical ligation (NCL) reaction.<sup>27</sup> Our plan for the construction of  $\beta$ -hLH and  $\beta$ -hCG was based on maximum convergence, and thus, we identified the key ligation sites depicted in Figure 1.  $\beta$ -hLH was envisioned to arise from three fragments of between 30 and 50 amino acids long, whereas our synthetic strategy for hCG relied on the assembly of five peptide segments, each bearing the corresponding N-/O-linked glycans in a highly modular fashion.

Given the high sequence homology between both targets and the relatively simpler structure of  $\beta$ -hLH, in comparison with  $\beta$ -hCG, we chose to start with the synthesis of the former, en route to  $\beta$ -hCG. In addition to providing access to hLH as one of the important glycoproteins within the family, this synthetic sequence could also serve as a proof of concept for the feasibility of the approach in terms of the following strategic steps: glycosylation via Lansbury aspartylation,<sup>28</sup> global deprotection of the glycopeptide segments, and the final union of these fragments by NCL.

**Chemical Synthesis of the  $\beta$ -Subunit of hLH as a Validation of the Synthetic Strategy.** For the preparation of  $\beta$ -hLH, we initially envisioned a retrosynthetic disconnection of the molecule in three fragments:  $\beta$ -hLH(chitobiose)[S1–Y37],  $\beta$ -hLH[C38–G71], and  $\beta$ -hLH[C72–L121] (Figure 1A). The corresponding peptide fragments were obtained by Fmoc-based solid-phase peptide synthesis (SPPS) with acid-labile protecting groups on the amino acid side chains, with the exception of the non-NCL-participating cysteine residues, in which case the acetamidomethyl (Acm) group was chosen. For the glycopeptide segment ( $\beta$ -hLH(chitobiose)[S1–Y37]), a pseudoproline dipeptide (Ile27-Thr28) was introduced in the sequence to improve the synthetic efficiency of the subsequent glycosylation at Asn30 under Lansbury conditions.<sup>29</sup> In addition, the aspartic

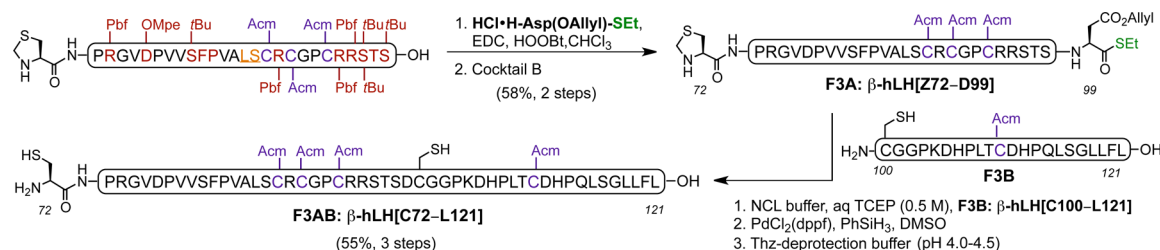
## Scheme 1. Synthesis of N-linked Glycopeptide Fragment 1: $\beta$ -hLH(chitobiose)[S1–Y37] (A) and Peptide Fragment 2: $\beta$ -hLH[Z38–G71] (B)<sup>a</sup>



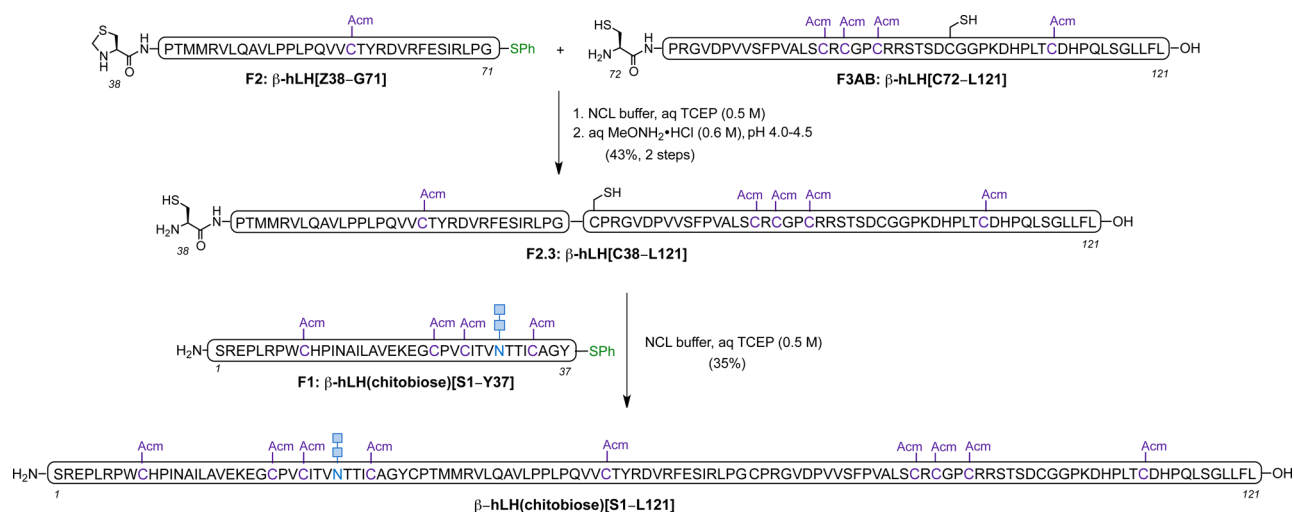
<sup>a</sup>Amino acid residues bearing acid-labile protecting groups shown in deep red, pseudoproline dipeptides shown in orange (underlined), and AcM-protected Cys residues shown in purple. Chitobiose glycan shown in blue, and thioester functionalities shown in green. Cocktail B: 88% TFA, 5% H<sub>2</sub>O, 5% phenol (PhOH), 2% triisopropylsilane (TIPSH).

acid that will bear the chitobiose moiety was protected with a quasi-orthogonal *O*-2-phenylisopropyl ester (*O*-2-Ph<sup>i</sup>Pr or OPp) group that can be removed selectively in the presence of *t*Bu-based protecting groups by treatment with 1–2% trifluoroacetic acid (TFA) in dichloromethane.<sup>30</sup> Thus, after cleavage from the resin, the C-terminal carboxylic acid of the protected peptide was first coupled to tyrosine phenylthioester (HCl·H-Tyr-SPh) under Sakakibara conditions,<sup>31</sup> which are known to be epimerization free, and subsequent selective deprotection of Asp30 provided the free carboxylic acid side chain (**pF1**: $\beta$ -hLH[S1–Y37]). Next, HATU-mediated Lansbury aspartylation with chitobiose glycosyl amine followed by treatment with Cocktail B (88% TFA, 5% H<sub>2</sub>O, 5% phenol, 2% TIPSH) for removal of all acid-labile protecting groups on the amino acid side chains afforded the corresponding glycopeptide fragment  $\beta$ -hLH(chitobiose)[S1–Y37] in 22% yield (over four steps) after only one HPLC purification (Scheme 1A). For the synthesis of the second fragment  $\beta$ -hLH[Z38–G71], the N-terminal cysteine was protected as a thiazolidine (Thz, Z), and the C-



Scheme 2. Synthesis of  $\beta$ -hLH[C72–L121] by NCL Between  $\beta$ -hLH[Z72–D99] and  $\beta$ -hLH[C100–L121] Fragments<sup>a</sup>

<sup>a</sup>Amino acid residues bearing acid-labile protecting groups shown in deep red, pseudoproline dipeptides shown in orange (underlined), Ac-protected Cys residues shown in purple, and thioester functionality shown in green. NCL buffer: guanidine·HCl (6 M), Na<sub>2</sub>HPO<sub>4</sub> (0.2 M), tris(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl) (0.02 M), 4-mercaptophenylacetic acid (MPAA) (0.2 M), aqueous TCEP solution (0.5 M), pH 7.2–7.4. Thz-deprotection buffer: MeONH<sub>2</sub>·HCl (0.3 M), guanidine·HCl (6 M), pH 4.0–4.5. Cocktail B: 88% TFA, 5% H<sub>2</sub>O, 5% phenol (PhOH), 2% triisopropylsilane (TIPSH).

Scheme 3. Final Assembly by NCL of Full-Length  $\beta$ -Subunit of hLH Bearing Chitobiose at the N-Glycosylation Site<sup>a</sup>

<sup>a</sup>NCL buffer: guanidine·HCl (6 M), Na<sub>2</sub>HPO<sub>4</sub> (0.2 M), tris(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl) (0.02 M), 4-mercaptophenylacetic acid (MPAA) (0.2 M), aqueous TCEP solution (0.5 M) pH 7.2–7.4.

terminal glycine residue was converted to glycine phenylthioester using PyBOP as a coupling agent. Finally, global deprotection under acidic conditions with Cocktail B provided  $\beta$ -hLH[Z38–G71] in 72% yield (over two steps) (Scheme 1B).

The last segment required for the assembly was  $\beta$ -hLH[C72–L121] (Scheme 2). This fragment is devoid of glycosylation sites, and our initial approach was to synthesize it entirely by SPPS. However, despite the incorporation of three pseudoproline dipeptides<sup>29</sup> and the use of the Hmb (2-hydroxy-4-methoxybenzyl) protecting group,<sup>32</sup> aspartimide formation was predominant, and we were unable to obtain useful yields of this peptide on solid support. We therefore approached this synthesis through NCL of two smaller peptide fragments of roughly similar size. Thus, protected peptides  $\beta$ -hLH[Z72–S98] and  $\beta$ -hLH[Z100–L121] were accessed using SPPS following cleavage from the resin. In the first case, the terminal serine residue (Ser98) was coupled to allyl-protected aspartic acid ethylthioester [HCl·H-Asp(OAllyl)-SEt]<sup>33</sup> under Sakakibara conditions, and subsequent treatment with Cocktail B gave  $\beta$ -hLH[Z72–D99] in 58% yield (over two steps). The fully deprotected peptide  $\beta$ -hLH[C100–L121] was obtained in two steps (64%), involving global acid deprotection (Cocktail B) and conversion of the N-terminal thiazolidine (Thz100) to cysteine using methoxyamine hydrochloride (MeONH<sub>2</sub>·HCl) at pH 4.0–4.5.<sup>34</sup> The two small fragments were then coupled under

NCL conditions (phosphate buffer solution containing guanidine·HCl and TCEP·HCl at pH 7.2–7.4) with 4-mercaptophenylacetic acid (MPAA) as an additive, and the allyl protecting group on the Asp99 side chain was subsequently removed [PdCl<sub>2</sub>(dppf), phenylsilane, DMSO]. Finally, the N-terminal Thz group was cleaved using a buffer solution of MeONH<sub>2</sub>·HCl and guanidine·HCl in water at pH 4.0–4.5 to provide  $\beta$ -hLH[C72–L121] in 55% yield (over three steps) (Scheme 2).

The final assembly of the individual peptide fragments commenced with the coupling of segment  $\beta$ -hLH[C72–L121] with  $\beta$ -hLH[Z38–G71] through NCL, followed by removal of the Thz group to free the N-terminal cysteine required for the final ligation, in a one-flask procedure. Purification by HPLC provided the desired peptide  $\beta$ -hLH[C38–L121] in 43% yield (two steps). Finally,  $\beta$ -hLH[C38–L121] was coupled to the chitobiose-containing glycopeptide  $\beta$ -hLH(chitobiose)[S1–Y37] under NCL conditions to afford the full-length  $\beta$ -subunit of hLH  $\beta$ -hLH(chitobiose)[S1–L121] in 35% yield following HPLC purification (Scheme 3).

**Chemical Synthesis of the  $\beta$ -Subunit of hCG Bearing Four O-Linked Glycans.** Having identified a successful route that featured ligation steps proceeding with complete conversion in reasonable reaction time frames, we set out to synthesize the

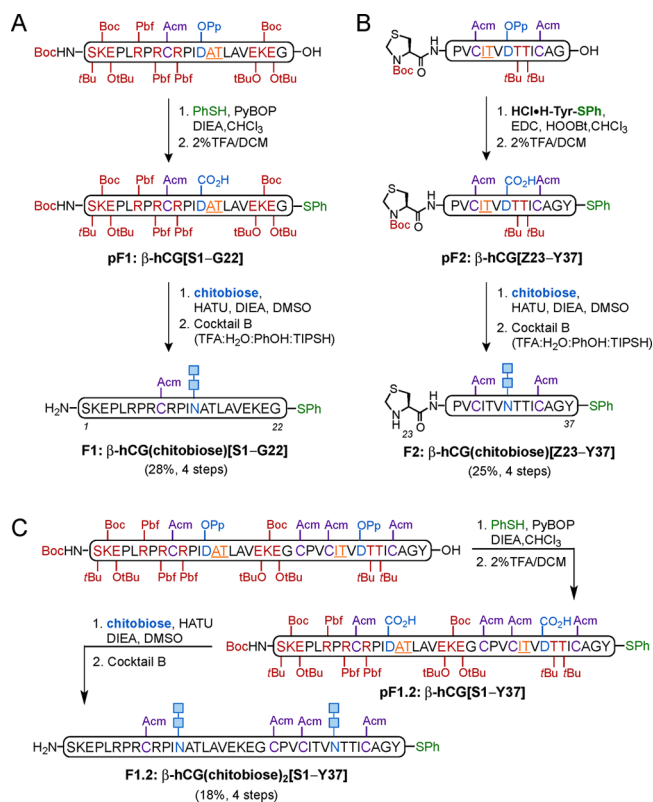
larger and more complex  $\beta$ -subunit of hCG incorporating two N-linked glycans and four closely spaced O-glycosylation sites.

Our initial approach to  $\beta$ -hCG relied on the key disconnections shown in Figure 1B, whereby five simpler fragments, each bearing the corresponding N-/O-glycans could be merged together via NCL. This highly modular strategy would allow us to investigate the effect of introducing various defined carbohydrates at different sites on the protein scaffold. As for the glycosylation points, each of the N-linked glycans was coupled to the complementary peptide fragment via Lansbury aspartylation, as shown earlier for  $\beta$ -hLH. Interestingly, a more challenging approach was also considered, consisting of the double incorporation, simultaneously, of the two N-linked carbohydrates on a longer peptide backbone comprising both N-glycosylation sites. With respect to the more complicated O-glycopeptide fragment, we envisioned the installation of all four O-linked glycans on solid support as glycosyl-serine “cassettes” during the synthesis of the carboxy-terminal segment.

Our initial target glycoform presents the readily available disaccharide chitobiose as a model N-linked glycan for the native complex biantennary dodecasaccharide (Figure 1B and C), similar to that previously described for hLH. We began our approach with the synthesis of protected **pF1: $\beta$ -hCG[S1–G22]** and **pF2: $\beta$ -hCG[Z23–Y37]** using standard Fmoc-based SPPS followed by cleavage from resin, C-terminus derivatization, and selective aspartic acid deprotection (Scheme 4A and B). In the event, the glycine residue (Gly22) in Fragment 1 was reacted with PhSH under PyBOP-mediated coupling conditions to give the corresponding C-terminal thioester, whereas Fragment 2 was subjected to single amino acid attachment with **HCl·H-Tyr-SPh** under epimerization-free Sakakibara conditions (EDC, HOObt). In both cases, selective removal of the O<sub>2</sub>P protecting group (at Asp 13 and Asp30, respectively) revealed the free aspartic acid to be coupled with chitobiose amine. Thus, **pF1: $\beta$ -hCG[S1–G22]** and **pF2: $\beta$ -hCG[Z23–Y37]** were then subjected to the two-step Lansbury aspartylation/Cocktail B deprotection protocol described earlier<sup>29a</sup> to afford glycopeptide fragments  **$\beta$ -hCG(chitobiose)[S1–G22]** and  **$\beta$ -hCG[Z23–Y37]** in 28% and 25% yield, respectively, (over four steps) after HPLC purification. Since the same N-glycan (chitobiose) was to be installed at both fragments, a double Lansbury aspartylation strategy was also explored for the synthesis of the entire bis-glycosylated segment  **$\beta$ -hCG(chitobiose)<sub>2</sub>[S1–Y37]** (Scheme 4C). First, SPPS of the protected fragment, followed by cleavage from the resin, C-terminal thioesterification (PhSH, PyBOP), and subsequent selective Asp(O<sub>2</sub>P) deprotection (2% TFA/DCM) provided peptide segment **pF1.2: $\beta$ -hCG[S1–Y37]** having the two aspartic acid (Asp13 and Asp30) side chains free for further coupling with two chitobiose units. In the event, 6 equiv of chitobiose amine, with respect to the peptide, and 6 equiv of HATU as the coupling agent were employed to give an encouraging 18% yield (over four steps) after global deprotection with Cocktail B. A 9% yield of monoglycosylated product was also isolated under these reaction conditions. Importantly, in the case of incorporating the same glycan unit, this approach represents a significant improvement over the synthesis of the two smaller individual fragments in terms of convergency and overall synthetic efficiency. Thus, this strategy was preferentially applied to access the corresponding  **$\beta$ -hCG(chitobiose)<sub>2</sub>[S1–Y37]** fragment (Scheme 4C).

Fragment 3 was prepared in a very straightforward manner, beginning with SPPS, followed by cleavage from the resin and thioesterification of the C-terminus with PhSH. After depro-

#### Scheme 4. Synthesis of N-Glycopeptide Fragments 1: $\beta$ -hCG(chitobiose)[S1–G22] (A) and 2: $\beta$ -hCG(chitobiose)[Z23–Y37] (B) and Convergent Approach to Bis-Glycosylated N-Glycopeptide Fragment 1.2: $\beta$ -hCG(chitobiose)<sub>2</sub>[S1–Y37] via a Double Lansbury Aspartylation (C)<sup>a</sup>

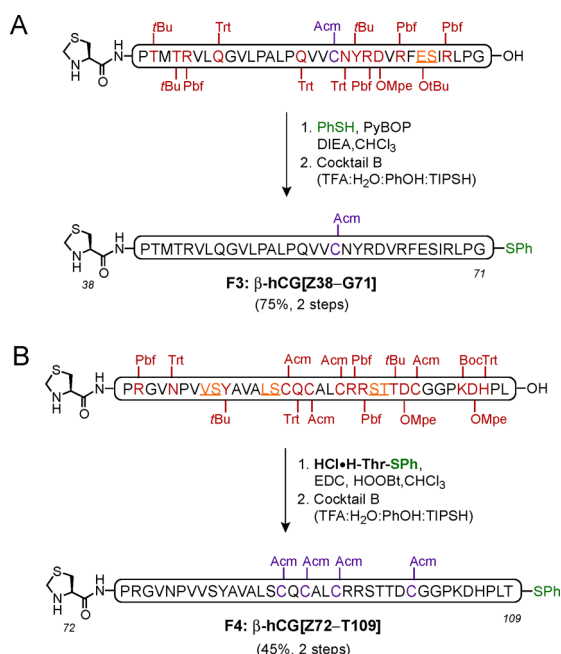


<sup>a</sup>Amino acid residues bearing acid-labile protecting groups shown in deep red, pseudoproline dipeptides shown in orange (underlined), and Ac<sub>2</sub>O-protected Cys residues shown in purple. Chitobiose glycan shown in blue and thioester functionalities shown in green. Cocktail B: 88% TFA, 5% H<sub>2</sub>O, 5% phenol (PhOH), 2% triisopropylsilane (TIPSH).

tection with Cocktail B,  **$\beta$ -hCG[Z38–G71]** was isolated in 75% yield over two steps (52% overall from resin) (Scheme 5A). The synthesis of Fragment 4,  **$\beta$ -hCG[Z72–T109]**, was initially hampered by significant aspartimide formation. Fortunately, the use of a piperidine/oxyma pure cocktail for the Fmoc-deprotection during the SPPS entirely suppressed this problem, and after single amino acid attachment (**HCl·H-Thr-SPh**)<sup>6a</sup> at the C-terminus, followed by subsequent treatment with Cocktail B, the desired peptide fragment  **$\beta$ -hCG[Z72–T109]** was obtained in 45% yield (over two steps) with no aspartimide formation (Scheme 5B).

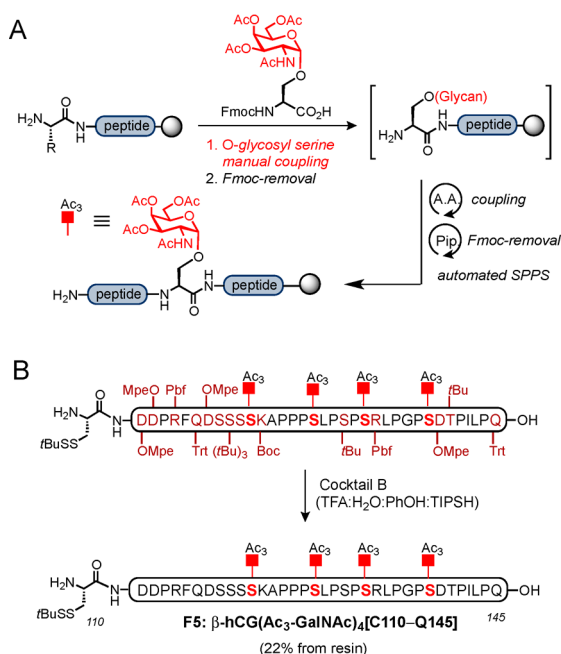
Finally, access to the last segment  **$\beta$ -hCG(Ac<sub>3</sub>-Gal-Nac)<sub>4</sub>[C110–Q145]** required a dramatically different approach than that described above. The presence of four O-linked glycans prohibits the late-stage attachment of these carbohydrates to the fully elaborated peptide. Alternatively, a more linear strategy, termed the “cassette-based” approach was followed,<sup>35</sup> wherein the O-glycan, linked to a properly protected serine residue, is directly introduced in the peptide sequence as a conveniently protected glycosyl amino acid building block (“cassette”), during solid-supported synthesis (Scheme 6A). This process not only allowed for assembly of this complex glycopeptide fragment but

**Scheme 5. Synthesis of Peptide Fragments 3:  $\beta$ -hCG[Z38–G71] (A) and 4:  $\beta$ -hCG[Z72–T109] (B)<sup>a</sup>**



<sup>a</sup>Amino acid residues bearing acid-labile protecting groups shown in deep red, pseudoproline dipeptides shown in orange (underlined), AcM-protected Cys residues shown in purple, and thioester functionalities shown in green. Cocktail B: 88% TFA, 5% H<sub>2</sub>O, 5% phenol (PhOH), 2% triisopropylsilane (TIPSH).

**Scheme 6. Cassette Approach Employed for Incorporation of the O-Linked Glycans on Solid Support (A) and Synthesis of Glycopeptide Fragment 5:  $\beta$ -hCG(Ac<sub>3</sub>-GalNAc)<sub>4</sub>[C110–Q145] Bearing N-acetylgalactosamine (GalNAc) at the Four O-Glycosylation Sites (B)**



also, in principle, enables the installation of different O-linked glycans by simply applying the desired cassette during peptide backbone elongation on SPPS.

We then targeted the synthesis of this fragment bearing the  $\alpha$ -N-acetylgalactosamine moiety (GalNAc) on all four serine residues, which corresponds to the “Tn antigen” structure (Scheme 6B). This system is the minimal common constituent of larger O-linked glycans and has been shown to be relatively abundant at Ser121, 127, 132, and 138 in wild-type  $\beta$ -hCG.<sup>24</sup> Following known procedures, we synthesized the conveniently protected, glycosylated Fmoc-Ser building block bearing O-acetyl groups in the GalNAc residue and having a free carboxylic acid on the serine.<sup>35a</sup> With this cassette in hand, we performed the synthesis of the required fragment via automated SPPS, with manual coupling of the corresponding glycosyl-serine residue (1.5 equiv) under PyAOP/HOAt coupling conditions. Gratifyingly, after treatment with Cocktail B, we were able to obtain the entire fragment  $\beta$ -hCG(Ac<sub>3</sub>-GalNAc)<sub>4</sub>[C110–Q145] with O-acetyl protected sugars in a single SPPS with great purity in 22% overall yield after HPLC purification (Scheme 6B). Importantly, this successful synthesis was amenable to scale-up, which afforded up to more than 50 mg of the most complex fragment of this synthetic version of  $\beta$ -hCG bearing four closely spaced Tn antigen structures.

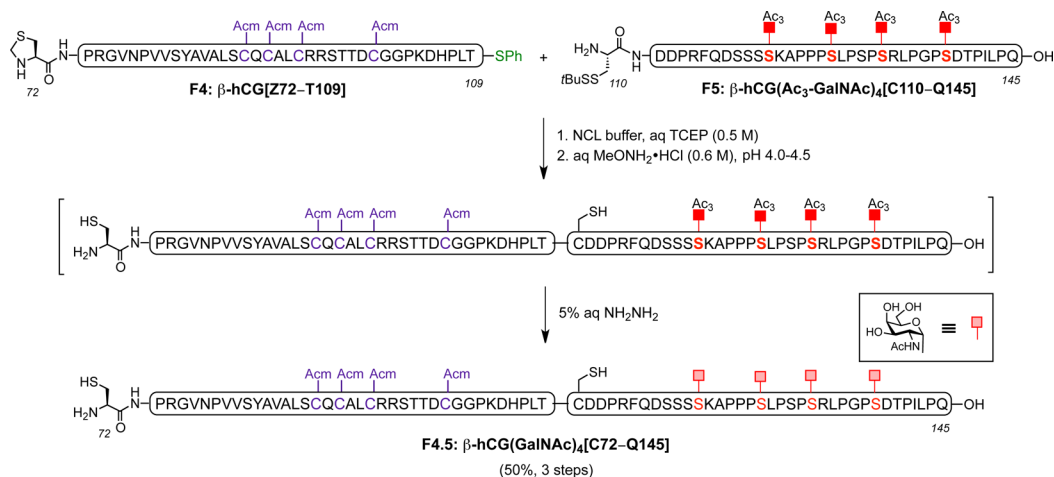
This strategy is likely to be applicable to the installation of alternative O-linked glycan cassettes at the four O-glycosylation sites for the assembly of different glycoforms. In this regard, a number of glycosyl amino acids incorporating more elaborated glycans, such as tumor-associated carbohydrate antigens (TF, STn, STF) or mucin-related core structures 1–4, have been synthesized through the cassette approach. Importantly, using these O-linked glycosylated building blocks, the cassette methodology has been successfully applied for the solution- and solid-phase synthesis of a series of mucin-derived O-glycopeptides bearing larger oligosaccharides.<sup>36</sup> These examples constitute an important precedent for the applicability of this strategy toward more complex glycosylated versions of  $\beta$ -hCG[C110–Q145]. Practical synthetic access to the carboxy-terminus fragment of  $\beta$ -hCG is an important step in itself as it has been found to be a common epitope for hCG-based monoclonal antibodies.<sup>25</sup> Screening of the specificity of different glycoforms and their role in binding to these antibodies or the development of new glycoforms could be used as a powerful tool in important clinical applications.

With all the prerequisite fragments in hand, assembly of the  $\beta$ -subunit of hCG relied on the coupling of the individual peptide segments by NCL, starting with fragments  $\beta$ -hCG[Z72–T109] and  $\beta$ -hCG(Ac<sub>3</sub>-GalNAc)<sub>4</sub>[C110–Q145]. Thus, these two fragments were joined together under standard NCL conditions, using MPAA as an additive. Upon completion of the reaction as monitored by ultraperformance liquid chromatography (UPLC), the terminal Thz (Z) protecting group was removed using MeONH<sub>2</sub>·HCl in a one-flask procedure. Subsequently, size-exclusion centrifugal filtration of the previous crude mixture followed by treatment with 5% aqueous hydrazine led to clean and complete removal of the acetate groups to provide  $\beta$ -hCG(GalNAc)<sub>4</sub>[C72–Q145] in 50% yield over three steps (Scheme 7).

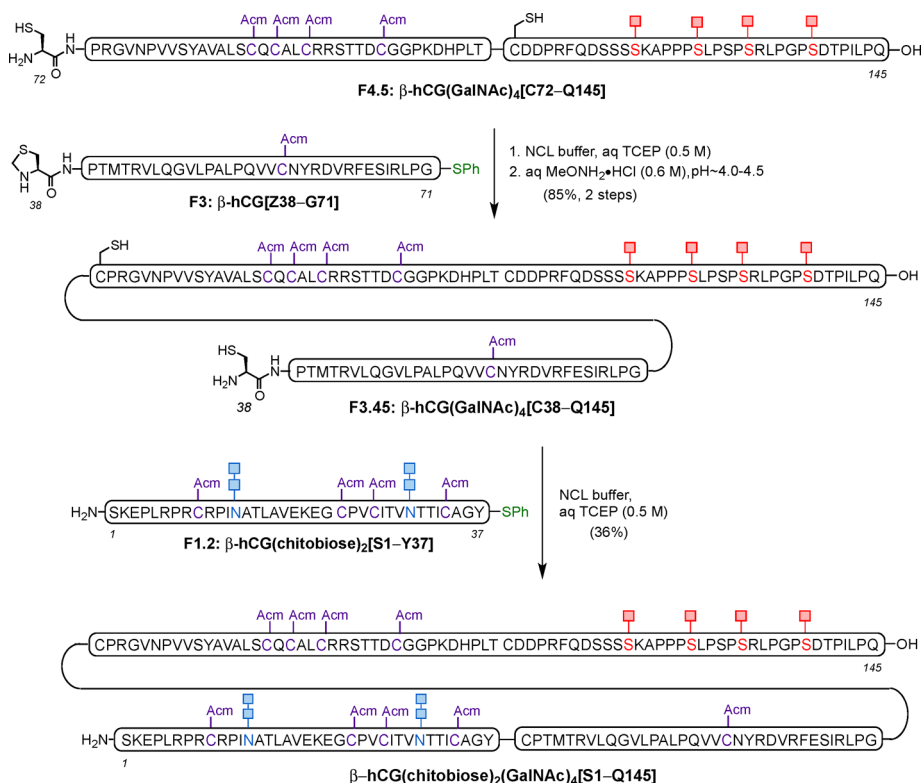
With the *in situ* ligation and Thz opening followed by acetate deprotection successfully performed in the synthesis of half of the full-length glycoprotein, we next used another NCL to bring  $\beta$ -hCG(GalNAc)<sub>4</sub>[C72–Q145] and  $\beta$ -hCG[Z38–G71] together. The coupling reaction was completed within 3 h with no starting material remaining, as assessed by LC-MS. Then, MeONH<sub>2</sub>·HCl was added, and the pH was adjusted to 4.0–4.5 to induce conversion of the N-terminal Thz to cysteine. Happily,



Scheme 7. Synthesis of Glycopeptide Fragment 4.5:  $\beta$ -hCG(GalNAc)<sub>4</sub>[C72–Q145] by NCL of  $\beta$ -hCG[Z72–T109] and  $\beta$ -hCG(Ac<sub>3</sub>-GalNAc)<sub>4</sub>[C110–Q145] Followed by Thz Opening and Subsequent De-O-Acetylation

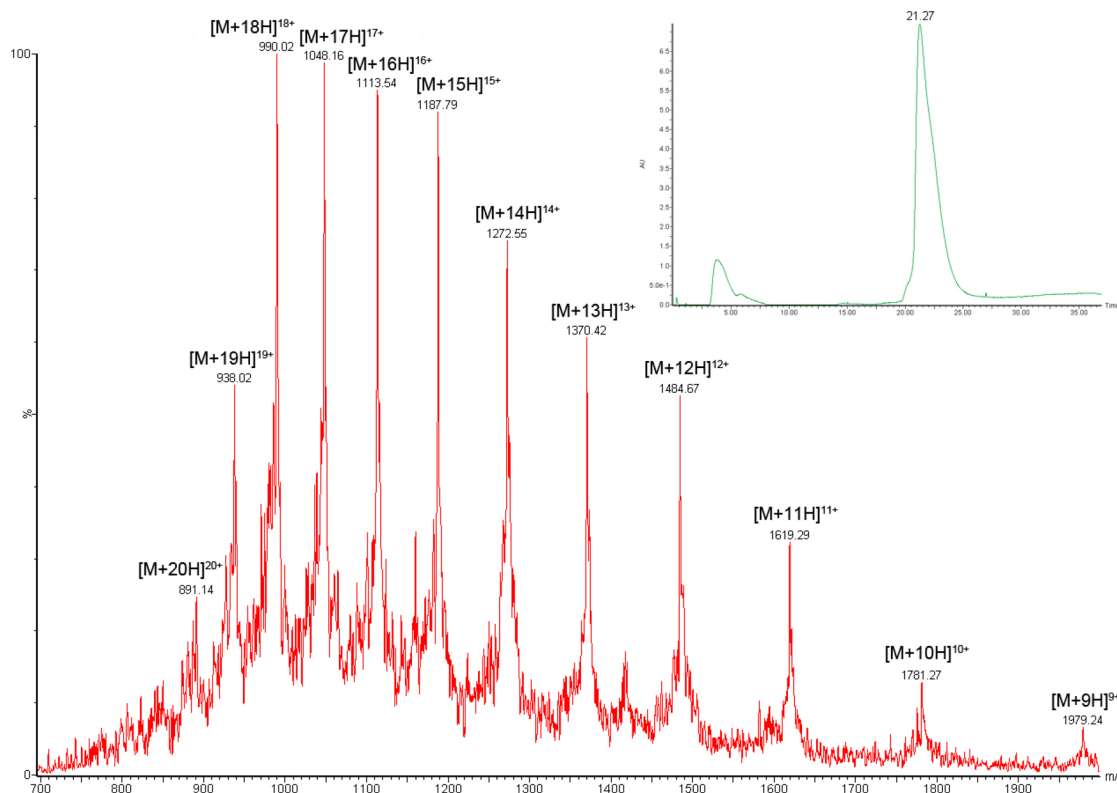


Scheme 8. Final Ligation of the Glycopeptide Fragments to Access Full-Length, Homogeneously Glycosylated  $\beta$ -Subunit of hCG



after size-exclusion centrifugal filtration, this procedure enabled good recovery of  $\beta$ -hCG(GalNAc)<sub>4</sub>[C38–Q145] for the final ligation reaction.<sup>37</sup> In the critical ligation event, O-glycosylated fragment  $\beta$ -hCG(GalNAc)<sub>4</sub>[C38–Q145] was combined with N-glycopeptide  $\beta$ -hCG(chitobiose)<sub>2</sub>[S1–Y37] under standard NCL conditions to give the target glycoprotein  $\beta$ -hCG(chitobiose)<sub>2</sub>(GalNAc)<sub>4</sub>[S1–Q145] after 4 h. Upon HPLC purification, the primary sequence of the 145-residue  $\beta$ -subunit of hCG containing homogeneous N- and O-linked glycans was obtained in a gratifying 33% yield (Scheme 8)<sup>38</sup> [see Figure 2 for mass spectrum and LC trace (UV) of  $\beta$ -hCG containing chitobiose at Asn13 and Asn30, and N-acetylgalactosamine at Ser121, Ser127, Ser132, and Ser138].

Thus, compound  $\beta$ -hCG(chitobiose)<sub>2</sub>(GalNAc)<sub>4</sub>[S1–Q145], bearing not only two N-linked carbohydrates but also, and even more challenging, four O-linked sugars, represents the largest human glycoprotein hormone to have been synthesized in homogeneous form using strictly chemical means. While the current work has involved chitobiose and GalNAc as representative N-/O-linked model glycans, this demonstration of feasibility through the modular glycoform assembly presented herein opens the door to the application of this technology to the synthesis of a library of homogeneous glycoproteins via chemical synthesis by installing alternative glycans on the corresponding peptide fragments



**Figure 2.** Mass spectrum and UV trace of  $\beta$ -hCG(chitobiose)<sub>2</sub>(GalNAc)<sub>4</sub>[S1–Q145] glycoprotein. Calcd for C<sub>759</sub>H<sub>1239</sub>N<sub>213</sub>O<sub>252</sub>S<sub>13</sub>, 17797.08 Da (average isotopes) [M + 9H]<sup>9+</sup> *m/z* 1978.45, found 1979.24; [M + 10H]<sup>10+</sup> *m/z* 1780.71, found 1781.27; [M + 11H]<sup>11+</sup> *m/z* 1618.92, found 1619.29; [M + 12H]<sup>12+</sup> *m/z* 1484.09, found 1484.67; [M + 13H]<sup>13+</sup> *m/z* 1370.01, found 1370.42; [M + 14H]<sup>14+</sup> *m/z* 1272.22, found 1272.55; [M + 15H]<sup>15+</sup> *m/z* 1187.47, found 1187.79; [M + 16H]<sup>16+</sup> *m/z* 1113.32, found 1113.54; [M + 17H]<sup>17+</sup> *m/z* 1047.89, found 1048.16; [M + 18H]<sup>18+</sup> *m/z* 989.73, found 990.02; [M + 19H]<sup>19+</sup> *m/z* 937.69, found 938.02; [M + 20H]<sup>20+</sup> *m/z* 890.85, found 891.14.

## CONCLUSION

In summary, we have synthesized the glycosylated primary sequence of two complex glycoprotein hormones ( $\beta$ -hLH and  $\beta$ -hCG) in homogeneous form using the current innovations of peptide and glycopeptide chemistry. The first approach described for the preparation of  $\beta$ -hLH served to validate the synthetic strategy en route to the more complex hCG  $\beta$ -subunit. The synthesis of the latter featured two challenging aspects that were successfully executed. First, a double Lansbury glycosylation was accomplished to provide the N-terminal fragment of the molecule in a highly modular fashion. Second, practical access to the carboxy-terminus was gained by sequential installation of four O-linked glycosyl-amino acid cassettes into closely spaced O-glycosylation sites in a single and high-yielding solid-supported synthesis. This O-linked glycopeptide fragment was then successfully advanced by further coupling, deprotection, and subsequent ligations with the remaining peptide segments to provide the full-length  $\beta$ -subunit of hCG. Interestingly, the highly modular assembly exemplified herein sets the stage for accessing more complex, chemically pure  $\beta$ -hCG glycoforms by synthesizing collections of each individual fragment containing a number of different, elaborated glycans and bringing them together following the successful ligation strategy outlined in this work. While the late-stage processes involving final AcM removal and oxidation/folding remain to be validated for  $\beta$ -hLH and  $\beta$ -hCG, building on our previous successes with erythropoietin and other glycoprotein targets, we have confidence in the ability to generate correctly folded  $\beta$ -subunits of these glycoprotein hormones. This important prospect should allow future

biological studies of these well-defined glycoforms to better understand the specific roles of certain glycans in the function and bioactivity of hCG in clinical settings. This knowledge may, in turn, lead to the development of more efficacious and improved therapeutics.

## ASSOCIATED CONTENT

### Supporting Information

A detailed description of general experimental procedures, including spectroscopic and analytical data for new compounds is provided. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

s-danishesky@ski.mskcc.org

### Present Address

<sup>§</sup>Department of Chemistry, DePaul University, 1110 West Belden Avenue, Chicago, IL 60614.

### Author Contributions

<sup>||</sup>These authors contributed equally to this work.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health (CA103823 to S.J.D.). A.F.-T. thanks the European Commission (Marie Curie International Outgoing Fellowship) for funding. Postdoctoral support is acknowledged by P.A.V. (NIH,



NRSA:1F32GM100567-01). We thank Dr. George Sukenick, Ms. Sylvi Rusli, and Ms. Hui Fang (Sloan Kettering Institute NMR core facility) for MS analysis. We also thank Dr. Lisa Ambrosini Vadola for her assistance with the preparation of this manuscript.

## REFERENCES

- (1) Walsh, G.; Jefferis, R. *Nat. Biotechnol.* **2006**, *24*, 1241–1252.
- (2) Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta* **1999**, *1473*, 4–8.
- (3) Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E. In *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2009.
- (4) Ghaderi, D.; Zhang, M.; Hurtazo-Ziola, N.; Varki, A. *Biotechnol. Genet. Eng. Rev.* **2012**, *28*, 147–176.
- (5) (a) Wang, P.; Dong, S.; Shieh, J. H.; Peguero, E.; Hendrickson, R.; Moore, M. A. S.; Danishefsky, S. J. *Science* **2013**, *342*, 1357–1360. (b) Wang, P.; Dong, S.; Brailsford, J. A.; Iyer, K.; Townsend, S. D.; Zhang, Q.; Hendrickson, R. C.; Shieh, J. H.; Moore, M. A. S.; Danishefsky, S. J. *Angew. Chem., Int. Ed.* **2012**, *51*, 11576–11584.
- (6) (a) Aussedat, B.; Fasching, B.; Johnston, E.; Sane, N.; Nagorny, P.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2012**, *134*, 3532–3541. (b) Nagorny, P.; Sane, N.; Fasching, B.; Aussedat, B.; Danishefsky, S. J. *Angew. Chem., Int. Ed.* **2012**, *51*, 975–979.
- (7) Pierce, J. G.; Parsons, T. F. *Ann. Rev. Biochem.* **1981**, *50*, 465–495.
- (8) (a) Sairam, M. R. In *Hormonal proteins and peptides*; Li, C. H., Ed., Academic Press: New York, 1983; Vol. 11, pp 1–79. (b) Ryan, R. J.; Keutmann, H. T.; Charlesworth, M. C.; McCormick, D. J.; Milius, R. P.; Calvo, F. O.; Vutyavanich, T. *Recent Prog. Horm. Res.* **1987**, *43*, 383–429.
- (9) Fiddes, J. C.; Talmage, K. *Recent Prog. Horm. Res.* **1984**, *40*, 43–74.
- (10) (a) Lei, Z. M.; Rao, C. V.; Kornyei, J.; Licht, P.; Hiatt, E. S. *Endocrinology* **1993**, *132*, 2262–2270. (b) Keutmann, H. T.; Charlesworth, M. C.; Mason, K. A.; Ostrea, T.; Johnson, L.; Ryan, R. J. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 2038–2042.
- (11) Stenman, U.-H.; Tiitinen, A.; Alftan, H.; Valmu, L. *Hum. Reprod. Update* **2006**, *12*, 769–784.
- (12) (a) Acevedo, H. F.; Tong, J. Y.; Hartsock, R. J. *Cancer* **1995**, *76*, 1467–1475. (b) Geissler, M.; Wands, G.; Gesien, A.; de la Monte, S.; Bellet, D.; Wands, J. R. *Lab. Invest.* **1997**, *76*, 859–871. (c) Butler, S. A.; Iles, R. K. In *Human Chorionic Gonadotropin*; Cole, L. A., Ed.; Elsevier: New York, 2010; Vol. 14; pp 149–167.
- (13) (a) Triozzi, P. L.; Martin, E. W.; Gouchnour, D.; Aldritch, W. *Ann. N.Y. Acad. Sci.* **1993**, *630*, 358–359. (b) Triozzi, P. L. *Int. J. Oncol.* **1994**, *5*, 1447–1453. (c) Moulton, H. M.; Yoshihara, P. H.; Mason, D. H.; Iversen, P. L.; Triozzi, P. L. *Clin. Cancer Res.* **2002**, *8*, 2044–2051. (d) Morse, M. A.; Chapman, R.; Powderly, J.; Blackwell, K.; Keler, T.; Green, J.; Riggs, R.; He, L. Z.; Ramakrishna, V.; Vitale, L.; Zhao, B.; Butler, S. A.; Hobeika, A.; Osada, T.; Davis, T.; Clay, T.; Lyerly, H. K. *Clin. Cancer Res.* **2011**, *17*, 4844–4853.
- (14) (a) Delves, P. J.; Iles, R. K.; Roitt, I. M.; Lund, T. *Mol. Cell. Endocrinol.* **2007**, *260–262*, 276–281. (b) Talwar, G. P.; Vyas, H. K.; Purswani, S.; Gupta, J. C. *J. Reprod. Immunol.* **2009**, *83*, 158–163. (c) Purswani, S.; Talwar, G. P. *Vaccine* **2011**, *29*, 2341–2348. (d) Butler, S. A.; Iles, R. K. *Human Chorionic Gonadotropin (hCG)*; Cole, L. A., Ed.; Elsevier: Burlington, MA, 2010; pp 153–172.
- (15) (a) Yu, N.; Xu, W.; Jiang, Z.; Cao, Q.; Chu, Y.; Xiong, S. *Immunol. Lett.* **2007**, *114*, 94–102. (b) Cole, L. A.; Dai, D.; Butler, S. A.; Leslie, K. K.; Kohorn, E. I. *Gynecol. Oncol.* **2006**, *102*, 145–150.
- (16) (a) Lunardi-Iskandar, Y.; Bryant, J. L.; Zeman, R. A.; Lam, V. H.; Samaniego, F.; Besnier, J. M.; Hermans, P.; Thierry, A. R.; Gill, P.; Gallo, R. C. *Nature* **1995**, *375*, 64–68. (b) Gill, P.; Lunardi-Iskandar, Y.; Louie, S.; Tulpule, A.; Zheng, T.; Espina, B. M.; Besnier, J. M.; Hermans, P.; Levine, A. M.; Bryant, J. L.; Gallo, R. C. *N. Engl. J. Med.* **1996**, *335*, 1261–1269. (c) Harris, P. J. *Lancet* **1995**, *346*, 118–119. (d) Lunardi-Iskandar, Y.; Bryant, J. L.; Blattner, W. A.; Hung, C. L.; Flamand, L.; Gill, P.; Hermans, P.; Birken, S.; Gallo, R. C. *Nat. Med.* **1998**, *4*, 428–434.
- (17) Cole, L. A. *Clin. Chim. Acta* **2012**, *413*, 48–65.
- (18) Lund, T.; Delves, P. J. *Rev. Reprod.* **1998**, *3*, 71–76.
- (19) Morgan, F. J.; Birken, S.; Canfield, R. E. *J. Biol. Chem.* **1975**, *250*, 5247–5258.
- (20) Lustbader, J. W.; Lobel, L.; Wu, H.; Elliott, M. M. *Recent Prog. Horm. Res.* **1998**, *53*, 395–424.
- (21) (a) Moyle, W. R.; Bahl, O. P.; März, L. *J. Biol. Chem.* **1975**, *250*, 9163–9169. (b) Kalyan, N. K.; Bahl, O. P. *J. Biol. Chem.* **1983**, *258*, 67–74.
- (22) Weisshaar, G.; Hiyama, J.; Renwick, A. G. C.; Nimtz, M. *Eur. J. Biochem.* **1991**, *195*, 257–268.
- (23) Sairam, M. R. *FASEB J.* **1989**, *3*, 1915–1926.
- (24) Valmu, L.; Alftan, H.; Hotakainen, K.; Birken, S.; Stenman, U. H. *Glycobiology* **2006**, *16*, 1207–1218.
- (25) (a) Birken, S.; Yershova, O.; Myers, R. V.; Bernard, M. P.; Moyle, W. *Mol. Cell. Endocrinol.* **2003**, *204*, 21–30. (b) Iversen, P. L.; Mourich, D. V.; Moulton, H. M. *Curr. Opin. Mol. Ther.* **2003**, *5*, 156–160.
- (26) (a) Cole, L. A.; Omrani, A.; Cermik, A.; Singh, R. O.; Mahoney, M. J. *Prenatal Diagn.* **1998**, *18*, 926–933. (b) Kovalevskaya, G.; Birken, S.; Kakuma, T.; Ozaki, N.; Sauer, M.; Lindheim, S.; Cohen, M.; Kelly, A.; Schlatterer, J.; O'Connor, J. F. *J. Endocrinol.* **2002**, *172*, 497–506. (c) Cole, L. A.; Butler, S. J. *Reprod. Med.* **2002**, *47*, 433–444.
- (27) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776–779.
- (28) Cohen-Anisfeld, S. T.; Lansbury, P. T., Jr. *J. Am. Chem. Soc.* **1993**, *115*, 10531–10537.
- (29) (a) Wang, P.; Aussedat, B.; Vohra, Y.; Danishefsky, S. J. *Angew. Chem., Int. Ed.* **2012**, *51*, 11571–11575. (b) Ullman, V.; Rädisch, M.; Boos, I.; Freund, J.; Pöhner, C.; Schwarzingler, S.; Unverzagt, C. *Angew. Chem., Int. Ed.* **2012**, *51*, 11566–11570.
- (30) Ocampo, S. M.; Albericio, F.; Fernández, I.; Vilaseca, M.; Eritja, R. *Org. Lett.* **2005**, *7*, 4349–4352.
- (31) Sakakibara, S. *Biopolymers* **1995**, *37*, 17–28.
- (32) Mergler, M.; Dick, F.; Sax, B.; Weiler, P.; Vorherr, T. *J. Pept. Sci.* **2003**, *9*, 36–46.
- (33) Aussedat, B.; Vohra, Y.; Park, P. K.; Fernández-Tejada, A.; Alam, S. M.; Dennison, S. M.; Jaeger, F. H.; Anasti, K.; Stewart, S.; Blinn, J. H.; Liao, H. X.; Sodroski, J. G.; Haynes, B. F.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2013**, *135*, 13113–13120.
- (34) Villain, M.; Vizzavona, J.; Rose, K. *Chem. Biol.* **2001**, *8*, 673–679.
- (35) (a) Kuduk, S. D.; Schwarz, J. B.; Chen, X.-T.; Glunz, P. W.; Sames, D.; Ragupathi, G.; Livingston, P. O.; Danishefsky, S. J. *J. Am. Chem. Soc.* **1998**, *120*, 12474–12485. (b) Herzner, H.; Reipen, T.; Schultz, M.; Kunz, H. *Chem. Rev.* **2000**, *100*, 4495–4537.
- (36) (a) Schwarz, J. B.; Kuduk, S. D.; Chen, X.-T.; Sames, D.; Glunz, P. W.; Danishefsky, S. J. *J. Am. Chem. Soc.* **1999**, *121*, 2662–2673. (b) Mathieux, N.; Paulsen, H.; Meldal, M.; Bock, K. *J. Chem. Soc., Perkin Trans. 1* **1997**, 2359–2368. (c) Gaidzik, N.; Westerlind, U.; Kunz, H. *Chem. Soc. Rev.* **2013**, *42*, 4431–4442.
- (37) Size-exclusion centrifugal filtration was used when possible, as it provides better recovery rates of the extremely valuable glycopeptides.
- (38) To avoid random disulfide bond formation within the fully deprotected glycoprotein sequence, the AcM protecting groups of the cysteines are left intact until the  $\beta$ -hCG subunit is ready to be folded and associated with the  $\alpha$ -hGPH subunit.