# Total Synthesis of the α-Subunit of Human Glycoprotein Hormones: Toward Fully Synthetic Homogeneous Human Follicle-Stimulating Hormone

**ARTICLE** *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · FEBRUARY 2012

Impact Factor: 12.11 · DOI: 10.1021/ja2111459 · Source: PubMed

Published in final edited form as:
*J Am Chem Soc*. 2012 February 22; 134(7): 3532–3541. doi:10.1021/ja2111459.

# Total Synthesis of the α-Subunit of the Human Glycoprotein Hormones (hGPH): Toward Fully Synthetic Homogeneous Human Follicle-Stimulating Hormone (hFSH)

**Baptiste Aussedat**[†], **Bernhard Fasching**[†], **Eric Johnston**[†], **Neeraj Sane**[†], **Pavel Nagorny**[†], and **Samuel J. Danishefsky**[†,‡,*]

[†]Laboratory for Bioorganic Chemistry, Sloan-Kettering Institute for Cancer Research, 1275 York Avenue, New York, NY 10065

[‡]Department of Chemistry, Columbia University, Havemeyer Hall, 3000 Broadway, New York, NY 10027

## Abstract

Described herein is the first total chemical synthesis of the unique α-subunit of the human glycoprotein hormone (α-hGPH). Unlike the biologically–derived glycoprotein hormones, which are isolated as highly complex mixtures of glycoforms, α-hGPH obtained by chemical synthesis contains discrete homogeneous glycoforms. Two such systems have been prepared. One contains the disaccharide chitobiose at the natural *N*-glycosylation sites. The other contains dodecamer oligosaccharides at these same sites. The dodecamer sugar is a consensus sequence incorporating the key features associated with human glycoproteins.

## Introduction

The glycosylation of proteins is one of the most common post-translational modifications.[1] Encountered in ca. 50% of proteins,[2] glycosylation is understood to play a pivotal role in numerous vital processes, including protein folding, cell-cell recognition, immune surveillance, hormone activities and inflammatory reactions.[3] As a consequence, the elucidation of the role of defined glycan structures on the conformation, stability, and functional activity of glycoproteins has drawn considerable attention.[4] However, such studies can be quite complicated. The installation of a glycan moiety on a protein is not "templated" *via* genetic information, and thus results in the cellular production of a mixture of glycoforms.[1] Even contemporary state-of-the-art analytical and chromatographic techniques cannot deliver useful quantities of discrete protein glycoforms for further study. This situation, in addition to the inherent challenge of the problem, has served to encourage the development of new methods for the synthesis of homogeneous glycoproteins.[5]

Our laboratory has been actively involved in this area of research for almost a decade.[6] In particular, we have focused our efforts on erythropoietin (EPO) and on a specific family of heterodimeric glycoproteins: the human glycoprotein hormones (hGPH). Among the members of this family are human thyroid-stimulating hormone (hTSH or thyrotropin) as well as the gonadotropins human luteinizing hormone (hLH or lutropin), human chorionic

[*]Corresponding Author: Samuel J. Danishefsky, 1275 York Avenue, New York, NY 10065, s-danishefsky@ski.mskcc.org, Fax: 212-772-8691.

ASSOCIATED CONTENT

General experimental procedures, including spectroscopic and analytical data for new compounds. This material is available free of charge via the Internet at http://pubs.acs.org."

gonadotropin (hCG) and human follicle-stimulating hormone (hFSH or follitropin). These four hormones are composed, in their respective active forms, of two non-covalently linked subunits – a common α-subunit (α-hGPH) and a hormone-specific β-subunit, each of which is characterized by defined glycosylation sites (Scheme 1).[7]

As our first target, we selected the human follicle-stimulating hormone (hFSH), which is involved in the regulation and maintenance of essential reproductive processes, such as gametogenesis, follicular development and ovulation. Used for over four decades for its therapeutic value in the treatment of anovulatory disorders as well as in assisted reproductive technologies, such as intrauterine insemination (IUI),[8] hFSH is today one of the most utilized hormones. However, serious side effects have been reported with its use[9] and its potential implication in the development and proliferation of certain cases of cancer is under debate.[10] These recently encountered complications might arise from a dose-to-dose variability in the relative quantities of each glycoform. Some glycoforms may be more active than others or, conceivably, trigger side effects in preference to the desired effect. In order to determine their specific roles and to overcome these problems, the chemical synthesis and subsequent testing of defined glycoforms seems to be the most straightforward approach.

A recent report disclosed the presence of the hFSH-receptor (hFSHR) on the endothelial surface of blood vessels at the periphery of a wide range of tumors.[11] This discovery renders hFSHR a very promising target for a new antitumor strategy. The ability to gain access to fully-synthetic hFSH with defined glycan structures would allow a better understanding of the impact of specific glycoforms on the binding of this hormone to its receptor.

We identified the achievement of the chemical total synthesis of homogeneous cysteine-rich α-hGPH as the first milestone in generating synthetic glycoprotein hormones (hGPH) in general, and hFSH in particular. The α-hGPH subunit consists of 92 residues with two *N*-linked glycans at Asn52 and Asn78 (Scheme 2).[12] The access to a homogeneous *N*-linked protein glycoform relies on the availability of useful quantities of the desired glycan with a defined structure. To date, most complex chemically-accessed glycoproteins presenting mature oligosaccharides at glycosylation sites have been obtained by semisynthesis from a glycan moiety isolated from egg-yolk.[13,14] Although these complex-type carbohydrates could be isolated on scales of hundreds of milligrams, they lack potentially important naturally occurring glycan modifications – such as α1,6 fucosylation – and display α2,6 sialylation, whereas α2,3 sialylation also appears ubiquitously on natural *N*-linked glycans.[15] Thus, we identified dodecasaccharide **1** (Scheme 2), which incorporates these two specific structural features, as a consensus high-mannose sialic acid-containing *N*-glycoside. Though the synthesis of **1** is highly convergent, in the end it does require a total of 71 steps.[16] Our goal was to introduce **1** in the form of *N*-linked glycosides at Asn52 and Asn78 en route to the first fully synthetic homogeneous complex glycoprotein.

Ideally, the homogeneous glycoprotein might be assembled by the direct coupling of **1** to the entire peptidic chain, giving the *N*-linked glycoprotein with maximal convergence. However, because aspartylation with **1** cannot be effectively conducted on a peptide as large as the entire α-subunit, we elected instead to disconnect the subunit retrosynthetically into smaller fragments that should be more suitable for the critical coupling reaction. Following proper attachment of the glycan to the appropriate peptide fragments, the resulting glycopeptides would be combined using the powerful strategy of native chemical ligation (NCL)[17] to furnish the fully elaborated subunit.

Two major strategies have been suggested to generate glycopeptides featuring an *N*-linked carbohydrate motif. The more linear approach is termed the “cassette-based” strategy,

wherein the glycan, linked to a properly-protected asparagine residue, is directly introduced in the peptidic sequence during solid-supported synthesis.[18] The second, more concise, approach involves installation of the glycan anomeric amine on an unprotected aspartic acid side chain using the Lansbury aspartylation reaction.[19] Due to its greater convergence and the high value of the dodecasaccharide (**1**), our program to synthesize homogeneous glycoproteins adopted the latter approach. *Herein we report the first total synthesis of two fully-synthetic, defined glycoforms of the α-subunit of human glycoprotein hormone (α-hGPH) bearing simple chitobiose units, as well as complex, sialylated and fucosylated biantennary N-linked dodecasaccharides.*

## Results and Discussion

The inherent complexity of the hGPH α-subunit stems from its size, the position of the natural glycosylation sites in the sequence, and the structure of the dodecasaccharide **1** to be installed at these positions (Scheme 2). In order to develop a synthetic program which is responsive to the complexities of this challenging macromolecule, we investigated several synthetic routes. As the synthesis of **1** remains a time-intensive undertaking, requiring 71 synthetic steps, we chose to work first with a far simpler system, chitobiose **4**, as a model glycan (Scheme 3). Thus, as a first step en route to **3**, we targeted the hGPH α-subunit with chitobiose at positions Asn52 and Asn78 (**5**). Indeed, it was hoped that the use of this easily accessible disaccharide would facilitate learning about the feasibility of each strategic step: glycosylation *via* Lansbury reaction,[19] the global deprotection of glycopeptidic segments, and the union of glycopeptidic fragments by native chemical ligation (NCL).

## A first-generation synthetic strategy for synthesizing hGPH α-subunit featuring chitobiose at the glycosylation sites

Our first vision for the assembly of **5** emphasized maximum convergency. In these terms, we identified the amide bond joining Thr58 and Cys59 as an ideal site for retrosynthetic disconnection of the targeted glycoprotein into two large glycopeptidic fragments: α-hGPH[1–58] and α-hGPH[59–92]. Each fragment would ideally be generated from the corresponding peptidic fragment, bearing a suitably exposed aspartic acid residue, and the glycan amine. In order to enable such a union, the remaining reactive residues required an orthogonal protection scheme. These protecting groups were to be maintained through Fmoc solid phase peptide synthesis as well as the acidic cocktail B (TFA/Water/TIPS/Phenol) treatment required to remove most of the side-chain protecting groups. Ultimately, we selected Allyl and Alloc protecting groups, anticipating that they could be readily removed in the presence of the glycan and thioester functionalities in a single palladium catalyst-mediated synthetic event. Although the optimization efforts are not described here, it was eventually found that the combination of $Pd(dppf)Cl_2$ and phenylsilane comprised a particularly effective means for this global deprotection.

Our journey addressed to the synthesis of α-hGPH began by targeting the sequence α-hGPH[59–92] bearing chitobiose (**7**). To this end, the model peptide **10** was synthesized on solid support,[20] with Alloc protection on Lys63, Lys75 and Lys91, allyl ester protection on Glu77, and with N-terminal Cys59 masked as a thiazolidine (Scheme 4). Unfortunately, upon exposure of peptide **10** and model glycan amine **4** to aspartylation conditions (HATU, DIEA, DMSO) only a trace amount of the expected glycopeptide **11** was generated. Instead, the major product resulted from the intramolecular attack of the adjacent His79 amidic nitrogen atom on the HATU-activated ω-carbonyl group of Asp78, leading to an unreactive aspartimide-containing peptide, **12**.

We therefore revised our approach to α-hGPH[59–92], envisioning a retrosynthetic disconnection at Ala81-Cys82, thereby suggesting a new set of building blocks: α-hGPH[59–81] and α-hGPH[82–92] (Scheme 5). Our first attempt to synthesize α-hGPH[59–80] on solid support led to an apparently very hydrophobic peptide, which rendered its derivatization in solution difficult. LC-MS analysis of the crude product showed that two side products caused by deletion of tripeptide Tyr65-Arg67 and pentapeptide Lys63-Arg67 co-eluted with the desired product, rendering HPLC purification extremely tedious.

Hoping to solve these issues, we evaluated the use of several amide surrogates including pseudoproline dipeptides,[21] isoacyl dipeptides[22] and Dmb dipeptides[23]. Incorporation of Fmoc-Val-Thr($\psi^{Me,Me}$Pro)-OH in position 69 improved the synthesis somewhat, and use of Fmoc-Gly-(Dmb)Gly-OH in position 73 improved the solubility of the protected peptide, making the derivatization more facile. However, both deletion products remained. Hence, non-HPLC purification methods were investigated. Following cleavage from the resin, the fully-protected peptide was purified by silica gel-flash chromatography. The protected peptide, thus apparently free of deletion products, was derivatized at its C-terminus by the introduction of an alanine-thiophenyl ester (**14**)[24] under Sakakibara's epimerization-free conditions (EDC, HOOBt)[25] and the acid-labile protecting groups were subsequently removed. This sequence unveiled the aspartic acid side-chain for coupling and removed all but the allyl and Alloc protecting groups, which are strategically included to allow condensation of **15** with glycan amine **4**. Thus, **15** and **4** were treated under the aforementioned Lansbury conditions. Afterward, addition of Pd(dppf)$Cl_2$ and $PhSiH_3$ to the reaction mixture resulted in removal of all Alloc and allyl ester functionalities. This convenient single-flask procedure provided glycopeptide **16** in 29% yield.

Glycopeptide **16** was then merged with peptide **17** using standard NCL conditions (6 M guanidine·HCl, 0.2 M $Na_2HPO_4$, 0.02 M TCEP·HCl, 0.02 M TCEP, Water, pH 7.2) with 4-mercaptophenylacetic acid (MPAA) as an additive (0.2 M).[26] After 3 h, the ligation was complete, the pH was adjusted and methoxyamine was added to induce conversion of the N-terminal thiazolidine to cysteine (0.3 M $MeONH_2$·HCl, pH 4, 4 h).[27] Happily, this sequence enabled access to **18** in 31% yield.

Encouraged by these results, we turned our attention to the synthesis of α-hGPH[1–58] (**6**). Due to the marginal performance of the large peptide **10** in the aspartylation step, we planned to synthesize **6** through smaller fragments, hoping to realize much better reactivity during the glycan/amine attachment. Given its length and given the fact that the cysteine residues are located relatively far from the site of glycosylation, synthesis of this fragment is a particular challenge. Fortunately, Tam and co-workers had reported a methionine-based ligation, wherein homocysteine (hCys) was employed in the NCL-type transthioesterification/S-N shift reaction and was subsequently methylated.[28] To the best of our knowledge, this technique had not yet been reported in the context of a glycoprotein synthesis. Furthermore, we envisioned utilization of a Kent-type kinetically-controlled ligation (KCL)[29] in order to produce α-hGPH[1–58] poised for further NCL with α-hGPH[59–92]. The methionine ligation was untested in the context of a KCL. Accordingly, it was of particular interest to target the Thr46-Met47 ligation site for the assembly of α-hGPH[1–58].

We first focused on the synthesis of building blocks **19** and **20.** Peptidic fragment α-hGPH[1–45] was synthesized on solid support, derivatized in solution under modified Sakakibara's conditions to incorporate threonine thiophenyl ester (**21**), then deprotected to afford, after HPLC purification, fragment **19** in 54% yield with excellent purity. We then prepared **22** (Scheme 7) incorporating Alloc-hCys(S*t*Bu)-OH (**23**) at the N-terminus and threonine thioethyl ester (**24**) at the C-terminus in order to enable subsequent NCL of α-

hGPH[1–58] with α-hGPH[59–92]. After SPPS, derivatization and deprotection in solution, **22** was available in reasonable purity, but its extremely poor solubility hampered HPLC purification efforts. Only very small quantities were isolated. Fortunately, the purity of the crude peptide could be improved by water precipitation. Following lyophilization, **22** was employed in the Lansbury reaction[19] with the anomeric chitobiose amine (**4**). Subsequent removal of the allyl-based protecting groups using $Pd(dppf)Cl_2$ afforded **20** after HPLC purification, in an overall yield of 35% from the resin. Notably, no cleavage of the homocysteine disulfide protecting group was observed.

With **19** and **20** in hand, we investigated their ligation. Under NCL conditions, the reaction proceeded in 5 h to afford the desired product α-hGPH[1–58] (**25**). In addition to **25**, we observed two side products, **26** and **27**, resulting from transthioesterification of MPAA and thiolactamization of the homocysteinyl thiol on the **25** thioethyl ester, respectively.[30] Surprisingly, given their low abundance, both **26** and **27** co-eluted with the desired product, making the HPLC saponification difficult. Accordingly, the crude product was treated with EtSH in order to restore the thioalkyl ester at the C-terminus. In the optimal case, the mixture was treated with EtSH in DMSO in the presence of DIEA, allowing the formation of **25** as the only observable product after 4 h.[31] Subsequent HPLC purification provided **25** in high purity in 36% yield over two steps.

Encouraged by this result, we turned our attention to the methylation of the homocysteinyl thiol to obtain the native methionine residue in position 47 (Scheme 8). In order to avoid potential overmethylation, as is sometimes observed with the use of *p*-nitrobenzene sulfonate alkylating agents, and potential thioester hydrolysis due to the high pH associated with the use of this reagent, we began our investigations by employing methyl iodide-based conditions based on the method of Ruvo et al, with slight modifications.[32] Thus, as anticipated by Ruvo, **25** was dissolved in $NaHCO_3$ buffer at pH 8.3 and treated with an excess of MeI (in DMF) and TCEP to maintain a reduced thiol. The Ruvo procedure involved quenching after 1 minute through addition of dithiotreitol (DTT) and adjustment of the pH to 2. In the case at hand, DTT was not seen to be a viable quenching agent, given the likelihood of transthioesterification of the C-terminal ethyl thioester with DTT. As an alternative, EtSH was employed to quench the reaction. Unfortunately, these slightly modified conditions afforded only traces of the desired product (**6**), some cyclized product (**27**), and a large amount of the intact starting material after the 1 min reaction. LC-MS analysis after 5 min showed trace **25** and **27** as well as overmethylated products (2 methyls). After 10 minutes, LC-MS analysis showed complete consumption of starting material, accompanied by an increased amount of overmethylated product. In order to avoid overmethylation as well as formation of cyclic product **27**, we returned to the shorter reaction time of 1 to 2 minutes, but increased the amount of methyl iodide employed ten-fold (to 1000 equiv), while increasing the pH to 8.6 to improve the nucleophilicity of the homocysteinyl thiol. Gratifyingly, these conditions resulted in almost complete conversion of **25** to **6** with minimal formation of cyclized and overmethylated byproducts. Quenching the reaction with a solution of EtSH in DMF and acidifying to pH 2 with a solution of aqueous TFA enabled, after immediate HPLC purification, isolation of the product in 51% yield. This approach allowed us to synthesize up to 7 mg of **6**.

To confirm the selectivity of the alkylation, we first treated the homocysteinyl-containing fragment (**25**) with radical initiator VA-044 and EtSH in the presence of TCEP.[33] After a few hours, **25** was converted to compound **28**, incorporating an Abu residue at position 47 without affecting either the glycan or the thioester moieties. In order to rule out the possibility that a random methylation had occurred on another nucleophilic function (the ε-amino group of a lysine residue, for example), we applied these conditions to **6**. Even after a prolonged reaction time (14 hours), the starting material remained intact, and no loss of

sulfur was observed. We concluded, from these two experiments, that the methyl group had been incorporated selectively in position 47.

With α-hGPH[59–92] and properly activated α-hGPH[1–58] fragments in hand, we turned our attention to an NCL reaction that would unite these fragments to form α-hGPH bearing *N*-linked chitobiose at the native glycosylation sites (Scheme 9). Happily, standard NCL conditions resulted in formation of the desired product. However, a non-negligible portion of α-hGPH[1–58] transesterified with MPAA remained intact after 18 h, which overlapped with the desired product in terms of retention time. Subsequent HPLC purification gave access to **29** in 20% yield. This result marks the first synthesis of the full sequence of α-hGPH bearing a single defined glycan. This glycoprotein with a model glycan thus serves as our first homogeneous α-hGPH analog.

Although we were pleased to have a validated potential approach to access α-hGPH with complex glycan (**3**), several drawbacks to this route motivated us to further optimize our synthetic plan while still using the chitobiose model system. We were particularly concerned by our inability to gain access to **22** in the high standard of purity required for the Lansbury aspartylation with the complex glycan.[34] Furthermore, the challenging reactivity of **6** and the prolonged reaction time of the NCL involving **6** and **18** were also problematic. The incomplete reaction observed during the NCL would result in significant loss of **6**, bearing dodecasaccharide – a serious drawback given the complexities of its synthesis. Additionally, the stability of glycopeptidic fragments bearing complex sialylated and fucosylated glycan **1** in a prolonged reaction time in NCL media was unknown. The approach to reach fragment α-hGPH[59–92] appeared satisfactory. Hence, we focused our attention on a more efficient synthetic route to assemble α-hGPH[1–58].

## Second generation synthetic strategy of α-hGPH featuring chitobiose at the asparagine glycosylation sites

Careful inspection of the sequence of α-hGPH[1–58] led to a modified retrosynthetic approach, presented in Scheme 10. We identified Gly30-Cys31 as an ideal site for NCL. In particular, the rate enhancement that would be afforded by employing a C-terminal glycine thioester was anticipated to substantially improve the material throughput of the synthesis. Furthermore, the C to N assembly of **18**, **31** and **32** would allow the presentation of Thr58 as a thiophenyl ester rather than a thioethyl ester. Here again, we envisioned a significant advantage from the increased reactivity of the aryl thioester. Thus, although this approach appears to be more linear, it seemed to offer significant advantages over our first route.

As with α-hGPH[59–80], synthesis of α-hGPH[31–57] on solid support turned out to be highly problematic. Different deletion products, which co-eluted with the product during HPLC purification, were detected. The synthesis was improved by using pseudoproline dipeptides, where applicable, by employing prolonged couplings, and by adding a capping step. The thioester was subsequently appended in solution providing **33** in 13% isolated yield (Scheme 11). Lansbury reaction[19] of chitobiose amine **4** with this peptide proceeded very efficiently, accompanied by only minimal amounts of aspartimide formation. The allyl protecting groups were cleaved in the same reaction vessel by employing $Pd(dppf)Cl_2$ and $PhSiH_3$ over 30 to 40 min, to yield **31** in 56% yield over two steps. This result was extremely encouraging for us as this aspartylation reaction had been expected to be the most challenging step of our new approach. Indeed only one example of a Lansbury reaction performed with a peptide of such length and chitobiose amine **4** had been reported.[35] We were also surprised by the similar reactivity of both peptides **22** and **33** towards chitobiose amine. Despite being twice as large (28 residues versus 12), **33** showed reactivity equal to, if not better than, **22**.

The last fragment (**32**) was isolated in 38% yield after solid-supported synthesis, conversion of the thiophenyl ester, and deprotection. With segments **18**, **31**, and **32** in hand, we began the final ligations. Fortunately, the ligation between **18** and **31** was achievable, thereby providing access to the desired product with complete consumption of both partners in 4 h. Subsequent *in situ* deprotection of Cys31 afforded, after purification, the bis-glycosylated segment, **34**, in 39% yield. Finally, attachment of the first thirty residues of α-hGPH to bis-glycosylated fragment **34** through NCL with **32** offered the target compound **30** in only 3 h with complete conversion. Upon purification, homogeneous chitobiose-containing α-hGPH was isolated in 31% yield.

This second-generation approach to α-hGPH bearing chitobiose did indeed serve to solve all of the problems that we encountered in the first approach. Thus, we proceeded with confidence to the synthesis of α-hGPH bearing the biantennary dodecasaccharide **1**. At this point, we had identified a route to α-hGPH that featured ligation steps that proceed with complete conversion in reasonable reaction timeframes. Thus, we were able to reliably prepare ten milligram quantities of **33** and **15** in high purity. As a result, we were well-poised to investigate the key fragment coupling with complex glycan **1**.

## Synthetic strategy for α-hGPH featuring dodecasaccharide at the glycosylation sites

At the outset of our work, the coupling of a sterically hindered glycan, such as **1**, with large peptides (in our case, consisting of 28 and 23 residues, respectively) was thought to be not just very challenging, but in fact impossible (Scheme 12). Indeed, due to the lower steric accessibility of the aspartic acid side chain in large peptides, the activated oxoester was believed to be too short-lived to react with a hindered glycan anomeric amine, which is not particularly nucleophilic. In fact, competitive intramolecular formation of aspartimide was expected to be dominant. Nevertheless, following the Kochetkov amination procedure,[36] glycan **2** (Scheme 2) was treated with saturated aqueous ammonium bicarbonate at 40 °C, and after repeated lyophilization, the dodecasaccharide bearing a β-anomeric amine was obtained in quantitative isolated yield. In order to conserve the valuable glycan **1**, the stoichiometry of the reaction was modified. Whereas in the Lansbury reaction with **4**, the glycan amine was used in excess (3:1, **4**:peptide), here the ratio was reversed (1:3, **1**:peptide).[37] Using these modified conditions, we investigated the attachment of **1** to **33**. The purity of the two reactants was critical to the success of this reaction. Thus, freshly prepared glycan amine and newly repurified peptide were combined for 1 h to produce the desired glycopeptide **35** in 42% yield. This result exceeded our expectations in terms of both the quantity and the quality of glycopeptide recovered. In order to avoid an unnecessary purification of this advanced product, we investigated the *in situ* deprotection of the lysine and glutamate side chains using $Pd(dppf)Cl_2$ and $PhSiH_3$. We were surprised to observe that, unlike the reaction involving chitobiose-appended peptide, the Alloc and allyl ester functions could not be removed in a tandem operation after attachment of **1**, as these conditions led to decomposition of the product. Fortunately, treatment of HPLC-purified **35** with $Pd(dppf)Cl_2$ and $PhSiH_3$ afforded **36** in 30% yield. Repetition of this sequence, using HPLC to isolate **35**, enabled, after size exclusion chromatography and HPLC, access to 6 mg of this glycopeptide **36** bearing dodecasaccharide and a C-terminal thiophenyl ester.[38]

Encouraged by these results, we began investigation of the glycosylation of **15** with dodecasaccharide **1**. Despite its shorter length, the yield for the union of **15** with **1** proceeded slightly less efficiently than the construct **33**, occuring in 26% isolated yield. This difference in yield was, in retrospect, not surprising given the difference in the steric bulk of the amino acid adjacent to the key aspartic acid residue. Indeed, the imidazole side chain of this residue might encourage the kinetic formation of the aspartimide. In practice, glycopeptide **37** was

obtained as a tetramethyl-guanidine adduct, presumably due to reaction of the histidine residue with HATU. As in the previous example, the Allyl/Alloc deprotection had to be carried out in a stepwise manner because the single reaction vessel protocol led to product decomposition. Size exclusion chromatography afforded **38** ready for NCL.

In the critical NCL event, **38** was combined with peptide **17** following the previously reported conditions (*vide supra*). The resultant clean formation of fragment **39** following deprotection of the thiazolidine (Scheme 13). Importantly, the tetramethylguanidinium adduct was cleaved from the His79 sidechain under the ligation conditions. Following desalting *via* gel filtration and HPLC purification, **39** was obtained in 21% yield over three steps. The stability of the dodecasaccharide over all of these steps was encouraging. With both **39** and **36** in hand, we began investigation of their merger. Under NCL conditions, we observed the formation of the key doubly glycosylated fragment **40**, albeit at a slower rate than was obsverved for the chitobiose analog **34** (8 h vs. 4 h). We attributed this slower reactivity to the proximity of the bulky glycan to the ligation site, which increased the effective steric hindrance of the thioester and attenuated the rate of the transesterification step. The cysteine residue was unveiled *in situ* upon addition of hydroxylamine and HCl. The reaction mixture was then desalted by gel filtration, purified by HPLC, and **40** was obtained in 33% yield. At this point we were ready to perform the ultimate ligation reaction. In the event, **40** was reacted for 3 h with freshly repurified **32**, leading to the formation of the primary sequence of the 92-residue human-glycoprotein hormone α-subunit featuring two defined, complex sialylated and fucosylated dodecasaccharide domains (**41**) in 28% yield.[39] This accomplishment marks the first synthesis of protected α-hGPH, as a single glycoform of the protein containing a realistic asparagine–linked, high mannose, fucose and sialic acid oligosaccharide.

## Conclusion

In summary, we have described the synthesis of two glycoforms of the α-subunit of the human-glycoprotein hormones in a homogeneous manner by chemical synthesis following two novel synthetic approaches. In the first approach, the key step is homocysteine-based NCL with subsequent methylation. Our work marks the first use of this strategy in glycoprotein synthesis. We have also shown that this concept could be efficiently extended to glycopeptide synthesis and employed in a kinetically-controlled ligation. Furthermore, using our recently described mild desulfurization method, we confirmed the selectivity of the alkylation in an original manner. Our second approach features the attachment of glycan **1** on the large thioester containing fragments **33** and **15**. The purity of both the glycan and the peptide were crucial to enabling Lansbury reaction on these large polypeptide fragments. Importantly, to the best of our knowledge, **36** represents the most complex *N*-linked glycopeptide formed by a direct glycosylation of a peptide to date. We are currently studying the technologically complex late stage issues in preparing homogeneous folded heterodimer[40] of hFSH in the laboratory by strictly chemical means.[41]

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

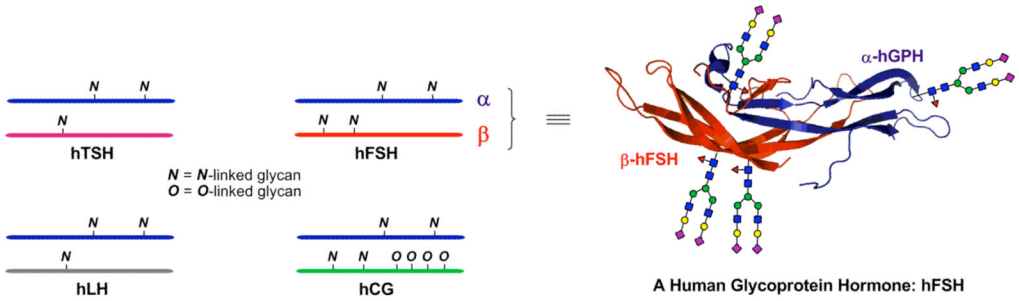
## Acknowledgments

The authors are grateful for support from the NIH (CA103823 to S.J.D.), and spectroscopic assistance from Dr. George Sukenick, Hui Fang, and Sylvi Rusli of SKI's NMR core facility. Special thanks to Dr. Jennifer Stockdill for helpful discussions and editorial consultation.

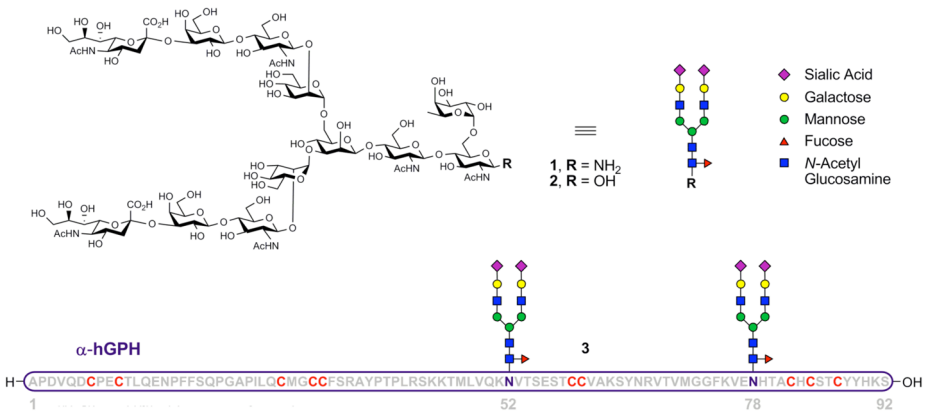
## References


1. Walsh G, Jefferis R. Nat Biotechnol. 2006; 24:1241–1252. [PubMed: 17033665]
2. Apweiler R, Hermjakob H, Sharon N. BBA-Gen Subjects. 1999; 1473:4–8.
3. Dwek RA. Chem Rev. 1996; 96:683–720. [PubMed: 11848770]
4. (a) Wormald MR, Petrescu AJ, Pao YL, Glithero A, Elliott T, Dwek RA. Chem Rev. 2002; 102:371–386. [PubMed: 11841247] (b) Solis D, Jiménez-Barbero J, Kaltner H, Romero A, Siebert HC, Von der Lieth CW, Gabius HJ. Cells Tissues Organs. 2001; 168:5–23. [PubMed: 11114583]
5. Gamblin DP, Scanlan EM, Davis BG. Chem Rev. 2008; 109:131–163. [PubMed: 19093879]
6. Kan C, Danishefsky SJ. Tetrahedron. 2009; 65:9047–9065. [PubMed: 20161377]
7. Pierce J, Parsons TF. Ann Rev Biochem. 1981; 50:465–495. [PubMed: 6267989]
8. (a) Peters H, Byskov AG, Himelstein-Braw R, Faber M. J Reprod Fertil. 1975; 45:559–566. [PubMed: 128630] (b) Chappel SC, Howles C. Hum Reprod. 1991; 6:1206–1212. [PubMed: 1752920] (c) Mannaerts B, Uilenbroek J, Schot P, De Leeuw R. Endocr Rev. 1983; 51:72–81.
9. Pang SC. Women's Health. 2005; 1:87–95.
10. Salehi F, Dunfield L, Phillips KP, Krewski D, Vanderhyden BC. J Toxicol Environ Health, Part B. 2008; 11:301–321.
11. Radu A, Pichon C, Camparo P, Antoine M, Allory Y, Couvelard A, Fromont Gl, Hai MTV, Ghinea N. N Engl J Med. 2010; 363:1621–1630. [PubMed: 20961245]
12. (a) Shome B, Parlow AF. J Clin Endocrinol Metab. 1974; 39:199–202. [PubMed: 4835135] (b) Rathnam P, Saxena BB. J Biol Chem. 1975; 250:6735–6746. [PubMed: 1158880]
13. Seko A, Koketsu M, Nishizono M, Enoki Y, Ibrahim HR, Juneja LR, Kim M, Yamamoto T. BBA-Gen Subjects. 1997; 1335:23–32.
14. (a) Yamamoto N, Ohmori Y, Sakakibara T, Sasaki K, Juneja LR, Kajihara Y. Angew Chem Int Ed. 2003; 42:2537–2540.(b) Yamamoto N, Tanabe Y, Okamoto R, Dawson PE, Kajihara Y. J Am Chem Soc. 2007; 130:501–510. [PubMed: 18085777] (c) Piontek C, Varón Silva D, Heinlein C, Pöhner C, Mezzato S, Ring P, Martin A, Schmid F, Unverzagt C. Angew Chem Int Ed. 2009; 48:1941–1945.(d) Hirano K, Macmillan D, Tezuka K, Tsuji T, Kajihara Y. Angew Chem Int Ed. 2009; 48:9557–9560.(e) Heinlein C, Varón Silva D, Tröster A, Schmidt J, Gross A, Unverzagt C. Angew Chem Int Ed. 2011; 50:6406–6410.
15. (a) Ma B, Simala-Grant JL, Taylor DE. Glycobiology. 2006; 16:158–184.(b) Varki, A.; Cummings, RD.; Esko, JD., et al. Essentials of Glycobiology. 2. Cold Spring Harbor Laboratory Press; Cold Spring Harbor (NY): 2009.
16. Nagorny P, Fasching B, Li X, Chen G, Aussedat B, Danishefsky S. J Am Chem Soc. 2009; 131:5792–5799. [PubMed: 19341309]
17. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH. Science. 1994; 266:776–779. [PubMed: 7973629]
18. Yamamoto N, Takayanagi A, Yoshino A, Sakakibara T, Kajihara Y. Chem Eur J. 2007; 13:613–625. [PubMed: 16977655]
19. (a) Cohen-Anisfeld ST, Lansbury PT Jr. J Am Chem Soc. 1993; 115:10531–10537.(b) Offer J, Quibell M, Johnson T. J Chem Soc, Perkin Trans 1. 1996:175–182.(c) Chen R, Tolbert TJ. J Am Chem Soc. 2010; 132:3211–3216. [PubMed: 20158247]
20. For synthetic considerations, we used peptide **10** as a model for peptide **9** in order to avoid the need for specific protection of the C- and N-terminus.
21. Guichou JF, Patiny L, Mutter M. Tetrahedron Lett. 2002; 43:4389–4390.
22. Sohma Y, Sasaki M, Hayashi Y, Kimura T, Kiso Y. Chem Commun. 2004:124–125.
23. Zahariev S, Guarnaccia C, Zanuttin F, Pintar A, Esposito G, Maravic G, Krust B, Hovanessian AG, Pongor S. J Peptide Sci. 2005; 11:17–28. [PubMed: 15635723]
24. Wieland T, Koppe H. Ann. 1954; 588:15–17.
25. (a) Kuroda H, Chen YN, Kimura T, Sakakibara S. Int J Pept Protein Res. 1992; 40:294–299. [PubMed: 1478787] (b) Sakakibara S. Biopolymers. 1995; 37:17–28. [PubMed: 7880964]
26. Johnson ECB, Kent SBH. J Am Chem Soc. 2006; 128:6640–6646. [PubMed: 16704265]

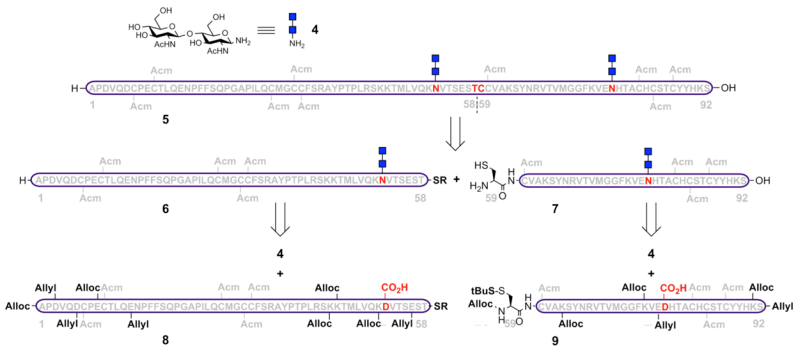
27. a) Villain M, Vizzanova J, Rose K. Chem & Biol. 2001; 123:526–533.(b) Bang D, Kent SBH. Angew Chem Int Ed. 2004; 43:2534–2538.
28. Tam JP, Yu Q. Biopolymers. 1998; 46:319–327. [PubMed: 9754028]
29. Bang D, Pentelute BL, Kent SBH. Angew Chem Int Ed. 2006; 45:3985–3988.
30. Beside the formation of side-products **26** and **27**, no cyclization or oligomerization of compound **20** was detected during this reaction.
31. We have observed that regenerating the thioester with EtSH instead of MESNa alleviates the formation of thiolactone, a potential issue during the methylation of homocysteine. C-terminal epimerization cannot be ruled out under these conditions despite the fact that only a single peak was observed by UPLC. However this result is inconsequential since this approach was not used to access the final target (**41**).
32. Saporito A, Marasco D, Chambery A, Botti P, Monti SM, Pedone C, Ruvo M. Biopolymers. 2006; 83:508–518. [PubMed: 16888773]
33. Wan Q, Danishefsky SJ. Angew Chem Int Ed. 2007; 46:9248–9252.
34. While challenging, the convergence of the synthesis allows us to prepare hundreds of milligrams of **1**.
35. Kan C, Trzupek JD, Wu B, Wan Q, Chen G, Tan Z, Yuan Y, Danishefsky SJ. J Am Chem Soc. 2009; 131:5438–5443. [PubMed: 19334679]
36. Likhosherstov LM, Novikova OS, Derevitskaja V, Kochetkov NK. Carbohydr Res. 1986; 146:C1–C5.
37. Additionally, 10 equivalents of HATU and 13 equivalents of DIEA, rather than 2.5, were added relative to the glycan.
38. Gel filtration was used when possible, as it provides better recovery rates of the extremly valuable glycopeptides.
39. In order to prevent random disulfide bond formation, the Acm protecting groups are left intact until the α-hGPH subunit is ready to be folded and associated with the hFSH β-subunit.
40. The synthesis of the required β-subunit of hFSH was recently disclosed: Nagorny P, Sane N, Fasching B, Aussedat B, Danishefsky S. Angew Chem Int Ed. 201110.1002/anie.201107482
41. It should be noted that in a slightly modified version of the route described above, a construct containing chitobiose at natural *N*-glycosylation site (Asn52 and 78) but with Acm groups on Cys7, 10, 28, 32, 60, 82, 84 and 87, the Acm deprotecion was smoothly accomplished, producing the free glycoprotein corresponding to deprotected **30**.



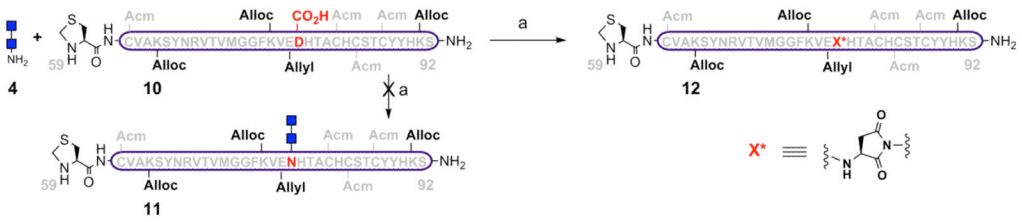

**Scheme 1.**
Human Glycoprotein Hormones (hGPH) family




**Scheme 2.**
Homogeneous α-hGPH bearing complex human dodecasaccharides (3).

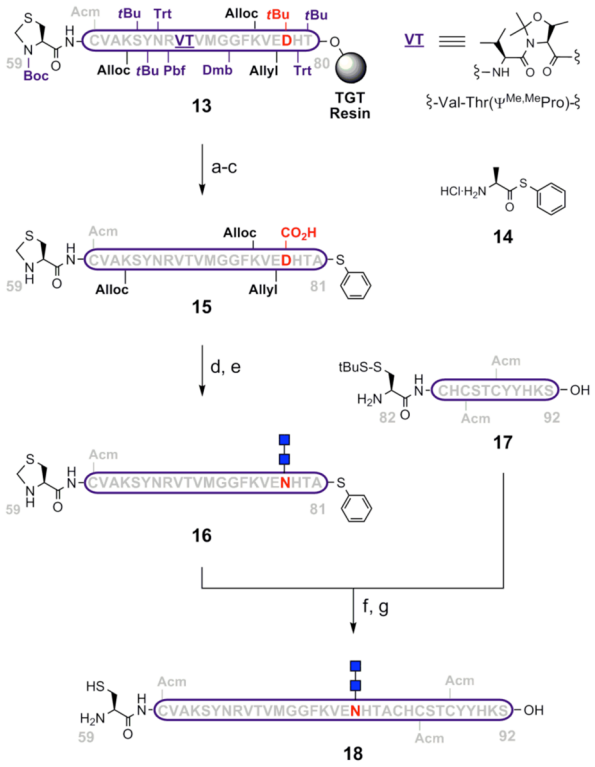


**Scheme 3.**
First retrosynthetic analysis of α-hGPH.




**Scheme 4. Aspartylation reaction**[a]

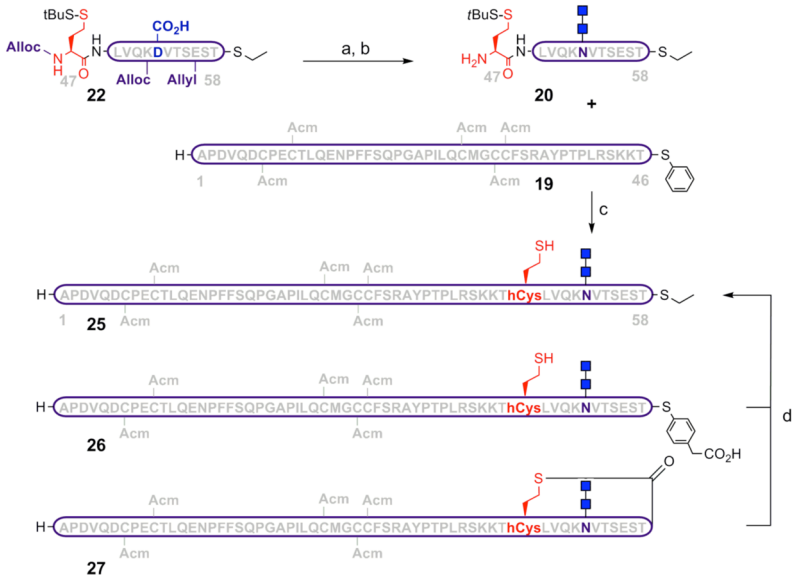
[a]Key: (a) Glycan amine **4**, HATU, DIEA, DMSO.




**Scheme 5. Synthesis of α-hGPH[59–92] with disaccharide**[a]

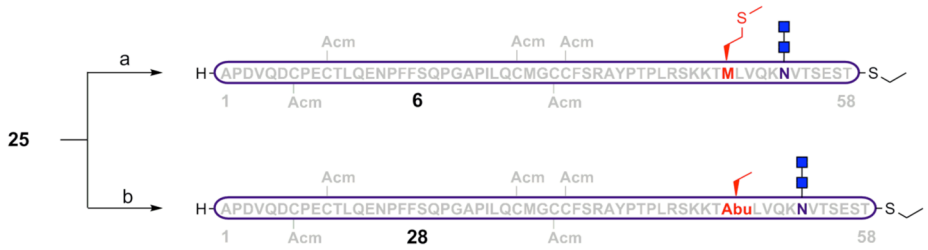
[a]Key: (a) 20% AcOH, 20% TFE, 60% $CH_2Cl_2$, 31%; (b) EDC, HOOBt, H-Ala-SPh·HCl (**14**), $CH_2Cl_2$/$CHCl_3$; (c) 88% TFA, 5% $H_2O$, 5% phenol, 2% *i*$Pr_3SiH$, 22% over two steps; (d) Glycan amine **4**, HATU, DIEA, DMSO; (e) Pd(dppf)$Cl_2$, $PhSiH_3$, DMSO, 29% over two steps; (f) Guanidine·HCl, $Na_2HPO_4$, TCEP·HCl, MPAA, $H_2O$; (g) $MeONH_2$·HCl, TCEP·HCl, $H_2O$, 31% over two steps.


**Scheme 6.**
Retrosynthetic analysis of α-hGPH[1–58].



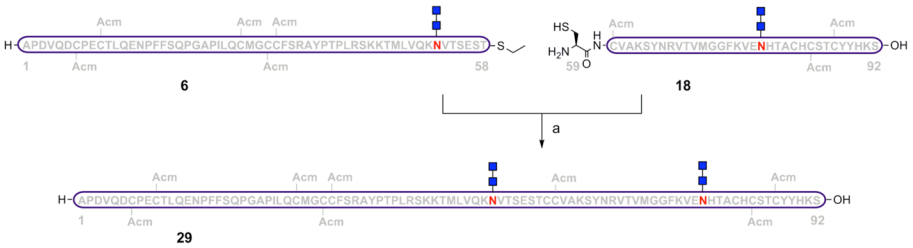

**Scheme 7. Synthesis of α-hGPH[1–58] with disaccharide**[a]

[a]Key: (a) Glycan amine **4**, HATU, DIEA, DMSO; (b) $Pd(dppf)Cl_2$, $PhSiH_3$, DMSO, 35% over two steps; (c) Guanidine·HCl, $Na_2HPO_4$, TCEP·HCl, MPAA, $H_2O$; (d) EtSH, DIEA, DMSO, 36% over two steps.



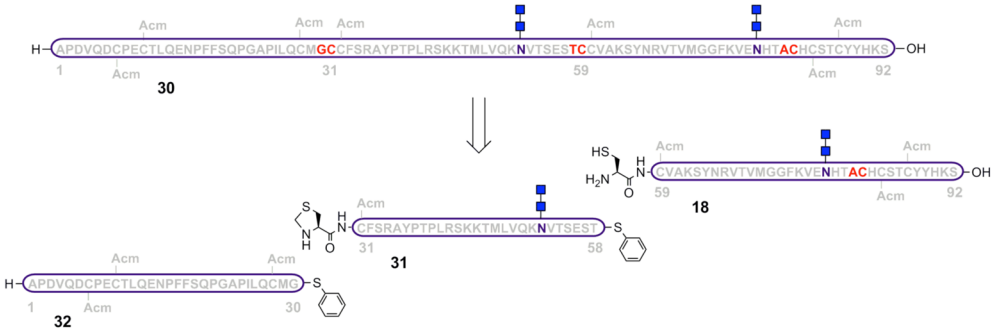

**Scheme 8. Synthesis of α-hGPH[1–58] with disaccharide**[a]

[a]Key: (a) MeI, $NaHCO_3$, TCEP·HCl, DMF/$H_2O$, 51%; (b) VA-044, TCEP, EtSH, $H_2O$/acetonitrile/DMF, 37 °C.

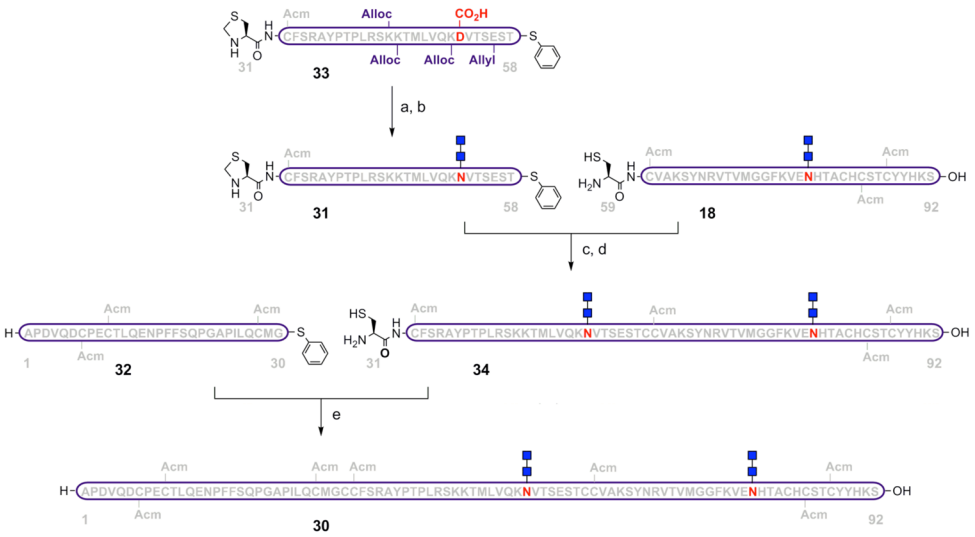



**Scheme 9. Access to *α*-hGPH with disaccharide[a]**

[a]Key: (a) Guanidine·HCl, $Na_2HPO_4$, TCEP·HCl, MPAA, $H_2O$, 20%.

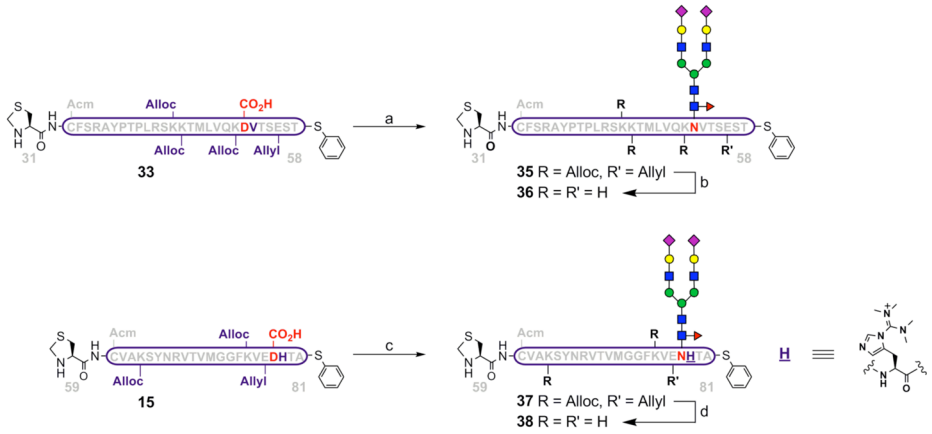



**Scheme 10.**
Second retrosynthetic analysis of α-hGPH.



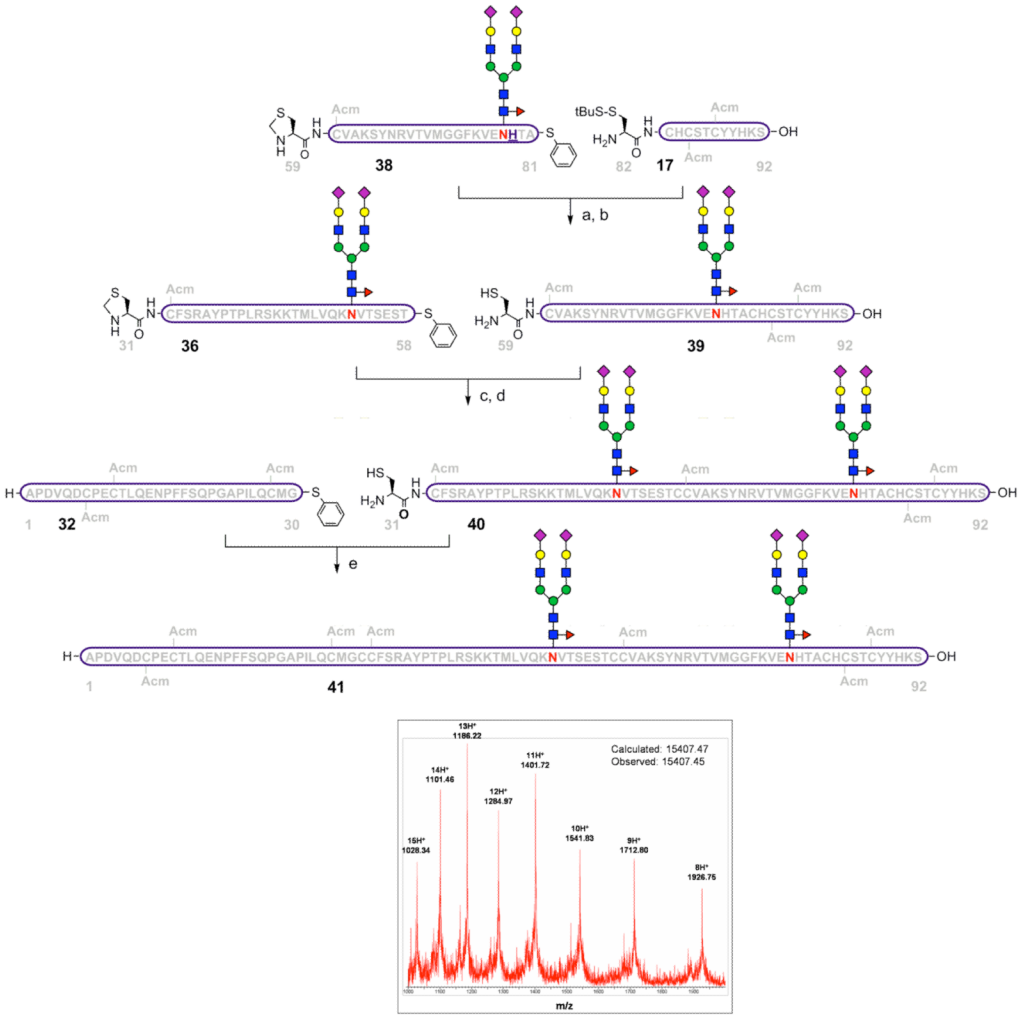

**Scheme 11. Synthesis of α-GPH[31–58] and merging of the fragments using NCL to access α-hGPH with disaccharide**

[a]Key: (a) Glycan amine **4**, HATU, DIEA, DMSO; (b) $Pd(dppf)Cl_2$, $PhSiH_3$, DMSO, 56% over two steps; (c) Guanidine·HCl, $Na_2HPO_4$, TCEP·HCl, MPAA, $H_2O$; (d) $MeONH_2$·HCl, TCEP·HCl, $H_2O$, 39% over two steps; (e) Guanidine·HCl, $Na_2HPO_4$, TCEP·HCl, MPAA, $H_2O$, 31%.




**Scheme 12. Lansbury aspartylation of α-hGPH[31–58] and α-hGPH[59–81] with complex glycan 1**

[a]Key: (a) Glycan amine **1**, HATU, DIEA, DMSO, 42%; (b) $Pd(dppf)Cl_2$, $PhSiH_3$, DMSO, 30%; (c) Glycan amine **1**, HATU, DIEA, DMSO, 26%; (d) $Pd(dppf)Cl_2$, $PhSiH_3$, DMSO, used as such in the next step.




**Scheme 13. Merging of dodecasaccharide containing fragments to access α-hGPH. ESI-MS spectrum of 41 (inset)**

[a]Key: (a) Guanidine·HCl, $Na_2HPO_4$, TCEP·HCl, MPAA, $H_2O$; (b) $MeONH_2$·HCl, TCEP·HCl, $H_2O$, 21% over three steps; (c) Guanidine·HCl, $Na_2HPO_4$, TCEP·HCl, MPAA, $H_2O$; (d) $MeONH_2$·HCl, TCEP·HCl, $H_2O$, 33% over two steps. (e) Guanidine·HCl, $Na_2HPO_4$, TCEP·HCl, MPAA, $H_2O$, 28%.