Published in final edited form as:
Angew Chem Int Ed Engl. 2012 January 23; 51(4): 975–979. doi:10.1002/anie.201107482.

Probing the Frontiers of Glycoprotein Synthesis: The Fully Elaborated β-subunit of the Human Follicle-Stimulating Hormone**

Dr. Pavel Nagorny[†], **Dr. Neeraj Sane**[†], **Dr. Bernhard Fasching**, **Dr. Baptiste Aussedat**, and **Prof. Dr Samuel J. Danishefsky**
Laboratory for Bioorganic Chemistry, Sloan-Kettering Institute for Cancer Research, 1275 York Avenue, New York, New York 10065, and Department of Chemistry, Columbia University, Havemeyer Hall, 3000 Broadway, New York, New York 10027(USA), Fax: (+1)212-772-8691

Samuel J. Danishefsky: s-danishefsky@ski.mskcc.org

The human follicle-stimulating hormone (hFSH) belongs to a family of hormones responsible for the maintenance of essential reproductive processes (gonadotropins).[1] FSH is produced in the anterior pituitary, and the binding of FSH to its receptor stimulates the maturation of follicles and the production of estrogen in females, and maintains spermatogenesis in males.[2] Consequently, FSH is clinically used in the treatment of anovulatory disorders associated with infertility.[3] Administration is usually in the form of subcutaneous injections, often once a day, over prolonged periods of time. Side effects of this treatment range from allergic reactions and nausea, to mood swings and fatigue.[4] Presently, FSH is mainly derived from recombinant technologies, specifically, from Chinese Hamster Ovary (CHO) cells.[5] The material so obtained is a complex mixture of hormone glycoforms, i.e. highly heterogeneous *vis-à-vis* the carbohydrates on the peptide backbone.[6] In normal adult humans, the FSH receptor (FSH-R) is expressed only on the ovarian granulosa cells of females and the testicular Sertoli cells of males. However, in a recent discovery it was found that the FSH-R is ubiquitously expressed on the endothelial cells of the peripheries of the tumors of the breast, prostrate, colon, pancreas, kidney, stomach, testis and ovary.[7] Earlier studies in mice have indicated that the effect of FSH on the growth of tumors is at least as potent as that of epidermal growth factor (EGF).[8] Although there was a concentration dependency of this effect, what might be an interesting study of the relative roles of the various glycoforms is presently stymied by the unavailability of homogeneous forms of FSH.

Structurally, FSH is a heterodimeric glycoprotein composed of two non-covalently associated subunits (α and β) (**1**, Figure 1).[9] Each of the subunits contains two *N*-linked oligosaccharides—the α-subunit at Asn^{52} and Asn^{78}, and the β-subunit at Asn^{7} and Asn^{24}, which are incorporated in the rough endoplasmic reticulum (RER) through co-translational modifications of the peptide backbone. The structures of the oligosaccharides play a crucial role in the proper folding, subunit assembly, secretion, and activation of the target receptor and, ultimately, the metabolic fate of the molecule.[10] Clearly, a method for gaining access to homogeneous glycoforms of FSH would be highly desirable for establishing a structure-

**This work was supported by NIH grant CA103823 (to SJD). We thank Ms. Rebecca Wilson for assistance with the preparation of the manuscript.

Correspondence to: Samuel J. Danishefsky, s-danishefsky@ski.mskcc.org.

†Authors have contributed equally to the work.

Supporting information for this article is available on the WWW under http://www.angewandte.org

activity relationship (SAR). Due to a lack of viable techniques for separating such complex mixtures of glycoforms,[11] chemical and chemoenzymatic methods[12] have emerged as a viable option for the preparation of homogeneous glycoproteins.

The seminal work by Kent and coworkers in developing the native chemical ligation (NCL) [13] reaction has led to the synthesis of very challenging protein targets such as a ubiquitin diastereomer,[14] RNase,[15] integral membrane kinase,[16] and tetraubiquitin protein,[17] among others, that are inaccessible by conventional peptide synthesis. The scope of the NCL reaction has been vastly expanded now, to enable ligations at Met, Ala, Phe, Ser, Val, Thr, Lys, Leu and Pro.[18] Yet, the ability to gain access to complex glycoproteins through *de novo* chemical synthesis remains a rather daunting task.[12] The difficulties can be attributed in part to the challenges of obtaining complex carbohydrates, particularly those containing sialic acid, fucose and high mannose residues, by chemical synthesis, in sufficient quantities. With respect to a possible synthesis of FSH, we had recently described an approach to the synthesis of the β-subunit using the disaccharide chitobiose as a model building block for our glycoprotein assembly.[19] Herein, we present the results of a rather ambitious undertaking, i.e. the synthesis of the β-subunit of FSH (**2**) containing a consensus sequence oligosaccharide at each of the two *N*-linkage sites, endowed with high mannose, fucose and sialic acid presentation, using natural motifs in all of the glycosidic linkages. The system has been synthesized with protected cysteines anticipating folding and association with the α-subunit.[20] The particular biantennary consensus dodecasaccharide chosen for this purpose was found to be abundant in batches of recombinant FSH displaying high bioactivity. The Manα**1**—**6**[Manα**1**—**3**]Manβ**1**—**4**GlcNAcβ**1**—**4**GlcNAc core of this oligosaccharide is also found to exist on other glycoprotein hormones such as, Chorionic gonadotropin (hCG), Luteinizing hormone (hLH), and Thyroid-stimulating hormone (hTSH) as well a s α-fetoprotein (associated with human hepatocellular carcinoma).[21] Consequently, the protocol presented here should also be extendable to the synthesis of these and other glycoproteins.

The β-subunit of FSH consists of 111 amino acids and the *N*-linked sugars are present at Asn^{7} and Asn^{24}.[9a] The abundance of cysteine residues in the peptide backbone, at fairly regular intervals, speaks to the possibility of assembling the glycoprotein using NCL. The key disconnections are depicted in Scheme 1. The two key ligation sites chosen were Phe^{19}-Cys^{20} and Trp^{27}-Cys^{28}. To enable the incorporation of the more precious glycopeptides in the final stages of assembly, the construction was performed from the *C*- to the *N*-terminus of the subunit. In order to engage in NCL, the *C*-terminus of each individual peptide fragment was functionalized as a thioester by single amino acid coupling. The peptide fragments were obtained by Fmoc-based solid phase peptide synthesis (SPPS) and the protecting groups on the amino acid side-chains during SPPS were chosen in a way that the aspartic acid residue that will bear the dodecasaccharide was protected orthogonally to those residues that were likely to interfere during the glycan attachment. Of the twelve cysteines on the protein, the nine that were not required for NCL were protected with acetamidomethyl (Acm) protecting groups. Additionally, these cysteine side chain protections prevent undesirable cross-linkages, *via* oxidation, during the course of the synthesis. Cleavage from the resin and selective deprotection of the aspartic acid side chain provided a free carboxylic acid, which was then coupled with the glycosylamine **7** by HATU-mediated Lansbury aspartylation.[22] The dodecasaccharide **7** was obtained by chemical synthesis, following the program adumbrated in Scheme 2.[19,23]

The strategy exploited the symmetry in the dodecasaccharide, in that a bis-glycosylation reaction on the hexasaccharide acceptor **13** with the trisaccharide donor **12** was executed in the late stages of the synthesis. The hexasaccharide core **13** was obtained through a highly convergent series of glycosylations using the known building blocks **8–11**. In each case, the

protecting groups were carefully selected to maximize the diastereoselectivity during the glycosylations and to minimize the number of deprotection steps that would need to be performed towards the end of the synthesis. Global deprotection of all the protecting groups present on the sugar moieties provided the dodecasaccharide **14**. The anomeric hydroxyl group was converted into a primary amine under Kochetkov amination conditions[19,24] by treatment with saturated ammonium bicarbonate solution. Excess ammonium bicarbonate was removed by repeated lyophilization of the material with minimal exposure to moisture.[25]

The protected [βFSH$^{1–18}$] **15** was obtained *via* Fmoc-based SPPS starting from the commerically available Fmoc-Arg(Pbf)-TGT resin, Pseudoproline dipeptides (denoted by ψ) were incorporated into the peptide synthesis sequence to improve the yield of this aggregation-prone fragment.[26] The *C*-terminal carboxylic acid of the peptide was coupled to phenylalanine phenylthioester **16** under Sakakibara conditions,[27] which are known to be epimerization-free. Cocktail B[28] treatment removed all the acid labile protecting groups on the amino acid side chains. This three step protocol afforded [βFSH$^{1–19}$] **17** in 21% yield after HPLC purification. The material so obtained was treated with the glycosyl amine **7** under HATU-mediated coupling conditions. The aspartylation proceeded in 73% conversion as indicated by LC-MS, along with a trace amount of undesired aspartimide formation. Treatment with $Pd(PPh_3)_4$ and phenylsilane removed all the allyl and alloc protecting groups on the amino acid side chains, in a one-flask procedure. The mixture was purified by HPLC to provide the [βFSH$^{1–19}$] glycopeptide **5** in 17% yield (averaged over 9 trials).

Similarly, [βFSH$^{20–26}$] **18** was obtained by Fmoc-based SPPS on Fmoc-Thr(*t*Bu)-TGT resin and cleavage from the resin. Treatment of peptide **18** with tryptophan phenylthioester **19**, and subsequent deprotection with Cocktail B, provided [βFSH$^{20–27}$] **6** in 66% yield (3 steps) (Scheme 4). Lansbury aspartylation of this peptide with the glycosyl amine **7** proceeded in 50% conversion with some aspartimide formation (ca. 8%). The [βFSH$^{20–27}$] glycopeptide **4** was isolated by HPLC purification in 27% yield (averaged over 5 trials).

The last segment required for the assembly was [βFSH$^{28–111}$] (**3**, Scheme 5). Fragment **3** is devoid of glycosylation sites and could have been obtained entirely by SPPS. However, we were unable to obtain useful yields of this peptide on solid support and we therefore opted to accomplish its synthesis *via* native chemical ligation of two smaller peptide fragments of roughly equal length. Gly^{65}-Cys^{66} was chosen as the ligation site, for two reasons. First, ligations at glycines are particularly effective, likely due to the lack of steric bulk at the α-position. Addionally, conversion of terminal glycine carboxylic acids to their thioesters is not subject to epimerization. Thus, two shorter peptide fragments were synthesized. [βFSH$^{28–65}$] was obtained by SPPS on Fmoc-Gly-TGT resin. The terminal glycine residue is converted to glycine phenylthioester and treated with Cocktail B to remove all the acid labile protecting groups, providing fragment **20** in 48% yield (over 3 steps). The fully deprotected peptide [βFSH$^{66–111}$] **21** was obtained on solid support using Fmoc-Glu(O*t*Bu)-TGT resin, cleavage and treatment with Cocktail B. The two fragments, **20** and **21**, were coupled under native chemical ligation conditions, using thiophenol as an additive. Upon completion of the reaction as monitored by LC-MS, the terminal Thz protecting group was removed using methoxylamine hydrochloride, in a one-flask procedure, to provide [βFSH$^{28–111}$] **3** in 22% yield after HPLC purification.

The final assembly of the individual peptide fragments commenced with the coupling of fragment [βFSH$^{20–27}$] **4** with the fragment [βFSH$^{28–111}$] **3** under NCL conditions in a pH 7.3 buffer, Scheme 5. Subsequently, the *N*-terminal Thz group was cleaved using methoxylamine hydrochloride at pH 4.8, to free the *N*-terminal cysteine required for the final ligation, in a one-flask procedure. Purification by HPLC provided the desired

glycopeptide [βFSH$^{20–111}$] **22** bearing the dodecasaccharide, in 26% yield (averaged over 3 trials).

Finally, the glycopeptide [βFSH$^{1–19}$] **5** was coupled to the glycopeptide [βFSH$^{20–111}$] **22** under NCL conditions. Gratifyingly, we obtained the full chain β-subunit of FSH (**2**) containing the two dodecasaccharides in 27% yield (averaged over 3 trials) after HPLC purification. Interestingly, the ligation was facile despite the fact that the two bulky dodecasaccharides are merely 16 residues apart during the ligation event. Also noteworthy, is the fact that the final compound **2** represents the largest realistically glycosylated glycoprotein with all natural-type linkages to have been synthesized in a homogeneous state, using strictly chemical methods.

Much encouraged by this demonstration of feasibility, we continue on the journey through which we hope to reach hFSH itself. There remains deprotection of the Acm protecting groups[29] on the cysteine side chains such as to allow for folding and subsequent association with the α-subunit. The total synthesis of the latter is already well advanced. The results of these ongoing studies will be disclosed in a timely way.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Fevold HL, Hisaw FL, Leonard SL. Am J Physiol. 1931; 97:291–301.
2. Smith PE. J Am Med Assoc. 1927; 88:158–161.
3. Howles CM. Hum Reprod Update. 1996; 2:172–191. [PubMed: 9079412]
4. Pang SC. Women's Health. 2005; 1:87–95.
5. Lobo, RA. Comprehensive Gynecology. 5. Katz, VL., et al., editors. MOSBY; Philadelphia: 2007. p. 1001-1037.
6. Gervais A, Hammel YA, Pelloux S, Lepage P, Baer G, Carte N, Sorokine O, Strub JM, Koerner R, Leize E, Van Dorsselaer A. Glycobiology. 2003; 13:179–189. [PubMed: 12626416]
7. Radu A, Pichon C, Camparo P, Antoine M, Allory Y, Couvelard A, Fromont G, Hai MTV, Ghinea N. N Engl J Med. 2010; 363:1621–1630. [PubMed: 20961245]
8. Li Y, Ganta S, Cheng C, Craig R, Ganta RR, Freeman LC. Mol Cell Endocrinol. 2007; 267:26–37. [PubMed: 17234334]
9. a) Pierce JG, Parsons TF. Annu Rev Biochem. 1981; 50:465–495. [PubMed: 6267989] b) Baenziger JU, Green ED. Biochim Biophys Acta. 1988; 947:287–306. [PubMed: 3130893] c) Fan QR, Hendrickson WA. Nature. 2005; 433:269–277. [PubMed: 15662415]
10. a) Hartree AS, Renwick AG. Biochem J. 1992; 287:665–679. [PubMed: 1445230] b) Ulloa-Aguirre A, Timossi C. Hum Reprod Update. 1998; 4:260–283. [PubMed: 9741710]
11. Rudd PM, Dwek RA. Crit Rev Biochem Mol Biol. 1997; 32:1–100. [PubMed: 9063619]
12. (a) Gamblin DP, Scanlan EM, Davis BG. Chem Rev. 2009; 109:131–163. [PubMed: 19093879] b) Kunz H, Unverzagt C. Angew Chem. 1988; 100:1763–1765.Angew Chem Int Ed Engl. 1988; 27:1697–1699.c) Offer J, Quibell M, Johnson T. J Chem Soc, Perkin Trans 1. 1996:175–182.d) Macmillan D, Bertozzi CR. Tetrahedron. 2000; 56:9515–9525.e) Yamamoto N, Ohmori Y, Sakakibara T, Sasaki K, Juneja LR, Kajihara Y. Angew Chem. 2003; 115:2641–2644.Angew Chem Int Ed. 2003; 42:2537–2540.f) Hirano K, Macmillan D, Tezuka K, Tsuji T, Kajihara Y. Angew Chem. 2009; 121:9721–9724.Angew Chem Int Ed. 2009; 48:9557–9560.g) Chen R, Tolbert TJ. J Am Chem Soc. 2010; 132:3211–3216. [PubMed: 20158247]
13. Dawson PE, Muir TW, Clark-Lewis I, Kent SB. Science. 1994; 266:776–779. [PubMed: 7973629]
14. Bang D, Makhatadze GI, Tereshko V, Kossiakoff AA, Kent SB. Angew Chem. 2005; 117:3920–3924.Angew Chem Int Ed. 2005; 44:3852–3856.

15. (a) Piontek C, Varón Silva D, Heinlein C, Pöhner C, Mezzato S, Ring P, Martin A, Schmid FX, Unverzagt C. Angew Chem. 2009; 121:1974–1978.Angew Chem Int Ed. 2009; 48:1941–1945.b) Piontek C, Ring P, Harjes O, Heinlein C, Mezzato S, Lombana N, Pöhner C, Püttner M, Varón Silva D, Martin A, Schmid FX, Unverzagt C. Angew Chem. 2009; 121:1968–1973.Angew Chem Int Ed. 2009; 48:1936–1940.
16. Lahiri S, Brehs M, Olschewski D, Becker CFW. Angew Chem. 2011; 123:4074–4078.Angew Chem Int Ed. 2011; 50:3988–3992.
17. Kumar KSA, Bavikar SN, Spasser L, Moyal T, Ohayon S, Brik A. Angew Chem. 2011; 123:6261–6265.Angew Chem Int Ed. 2011; 50:6137–6141.
18. (a) Tam JP, Yu Q. Biopolymers. 1998; 46:319–327. [PubMed: 9754028] b) Yan LZ, Dawson PE. J Am Chem Soc. 2001; 123:526–533. [PubMed: 11456564] c) Wan Q, Danishefsky SJ. Angew Chem. 2007; 119:9408–9412.Angew Chem Int Ed. 2007; 46:9248–9252.d) Crich D, Banerjee A. J Am Chem Soc. 2007; 129:10064–10065. [PubMed: 17658806] e) Okamoto R, Kajihara Y. Angew Chem. 2008; 120:5482–5486.Angew Chem Int Ed. 2008; 47:5402–5406.f) Haase C, Rohde H, Seitz O. Angew Chem. 2008; 120:6912–6915.Angew Chem Int Ed. 2008; 47:6807–6810.g) Chen J, Wan Q, Yuan Y, Zhu J, Danishefsky SJ. Angew Chem. 2008; 120:8649–8652.Angew Chem Int Ed. 2008; 47:8521–8524.h) Yang R, Pasunooti KK, Li F, Liu XW, Liu CF. J Am Chem Soc. 2009; 131:13592–13593. [PubMed: 19728708] i) Kumar KSA, Haj-Yahya M, Olschewski D, Lashuel HA, Brik A. Angew Chem. 2009; 121:8234–8238.Angew Chem Int Ed. 2009; 48:8090–8094.j) Chen J, Wang P, Zhu J, Wan Q, Danishefsky SJ. Tetrahedron. 2010; 66:2277–2283. [PubMed: 20798898] k) Harpaz S, Siman P, Kumar KS, Brik A. ChemBioChem. 2010; 11:1232–1235. [PubMed: 20437446] l) Tan Z, Shang S, Danishefsky SJ. Angew Chem. 2010; 122:9690–9693.Angew Chem Int Ed. 2010; 49:9500–9503.m) Shang S, Tan Z, Dong S, Danishefsky SJ. J Am Chem Soc. 2011; 133:10784–10786. [PubMed: 21696220]
19. Nagorny PN, Fasching B, Li X, Chen G, Aussedat B, Danishefsky SJ. J Am Chem Soc. 2009; 131:5792–5799. [PubMed: 19341309]
20. Samaddar M, Catterall JF, Dighe RR. Protein Expr Purif. 1997; 10:345–355. [PubMed: 9268682]
21. (a) Gharib SD, Wierman ME, Shupnik MA, Chin WW. Endocr Rev. 1990; 11:177–199. [PubMed: 2108012] b) Sairam MR. FASEB J. 1989; 3:1915–1926. [PubMed: 2542111] c) Mutsaers JHGM, Kamerling JP, Devos R, Guisez Y, Fiers W, Vliegenthart JFG. Eur J Biochem. 1986; 156:651–654. [PubMed: 3084257] d) Damm JBL, Kamerling JP, van Dedem GWK, Vliegenthart JFG. Glycoconjugates. 1987; 4:129–144.e) Hokke CH, Berwer AA, van Dedem GWK, van Ostrum JV, Kamerling JP, Vliegenthart JFG. FEBS Lett. 1990; 275:9–14. [PubMed: 2124546] f) Green ED, Baenziger JU. J Biol Chem. 1988; 263:25–35. [PubMed: 3121609] g) Yamashita K, Hitoi A, Tsuchida Y, Nishi S, Kobata A. Cancer Res. 1983; 43:4691–4695. [PubMed: 6192908]
22. (a) Cohen-Anisfeld ST, Lansbury PT Jr. J Am Chem Soc. 1993; 115:10531–10537.b) Miller JS, Dudkin VY, Lyon GJ, Muir TW, Danishefsky SJ. Angew Chem. 2003; 115:447–450.Angew Chem Int Ed. 2003; 42:431–434.
23. (a) Danishefsky SJ, Hu S, Cirillo PF, Eckhardt M, Seeberger PH. Chem-Eur J. 1997; 3:1617–1628.b) Dudkin VY, Miller JS, Danishefsky SJ. Tetrahedron Lett. 2003; 44:1791–1793.c) Wu B, Hua Z, Warren JD, Ranganathan K, Wan Q, Chen G, Tan Z, Chen J, Endo A, Danishefsky SJ. Tetrahedron Lett. 2006; 47:5577–5579.
24. Likhosherstov LM, Novikova OS, Derevitskaja V, Kochetkov NK. Carbohydr Res. 1986; 146:C1–C5.
25. Isbell HS, Frush HL. J Org Chem. 1958; 23:1309–1319.
26. Wöhr T, Wahl F, Nefzi A, Rohwedder B, Sato T, Sun X, Mutter M. J Am Chem Soc. 1996; 118:9218–9227.
27. Sakakibara S. Biopolymers. 1995; 37:17–28. [PubMed: 7880964]
28. Cocktail B (2% triisopropylsilane/5% phenol/5% H_2O/88% TFA).
29. On the chitobiose model system [Ref. 19], we have observed that the Acm groups on the nine protected cysteines can be removed using AgOAc, as analyzed by LC-MS.

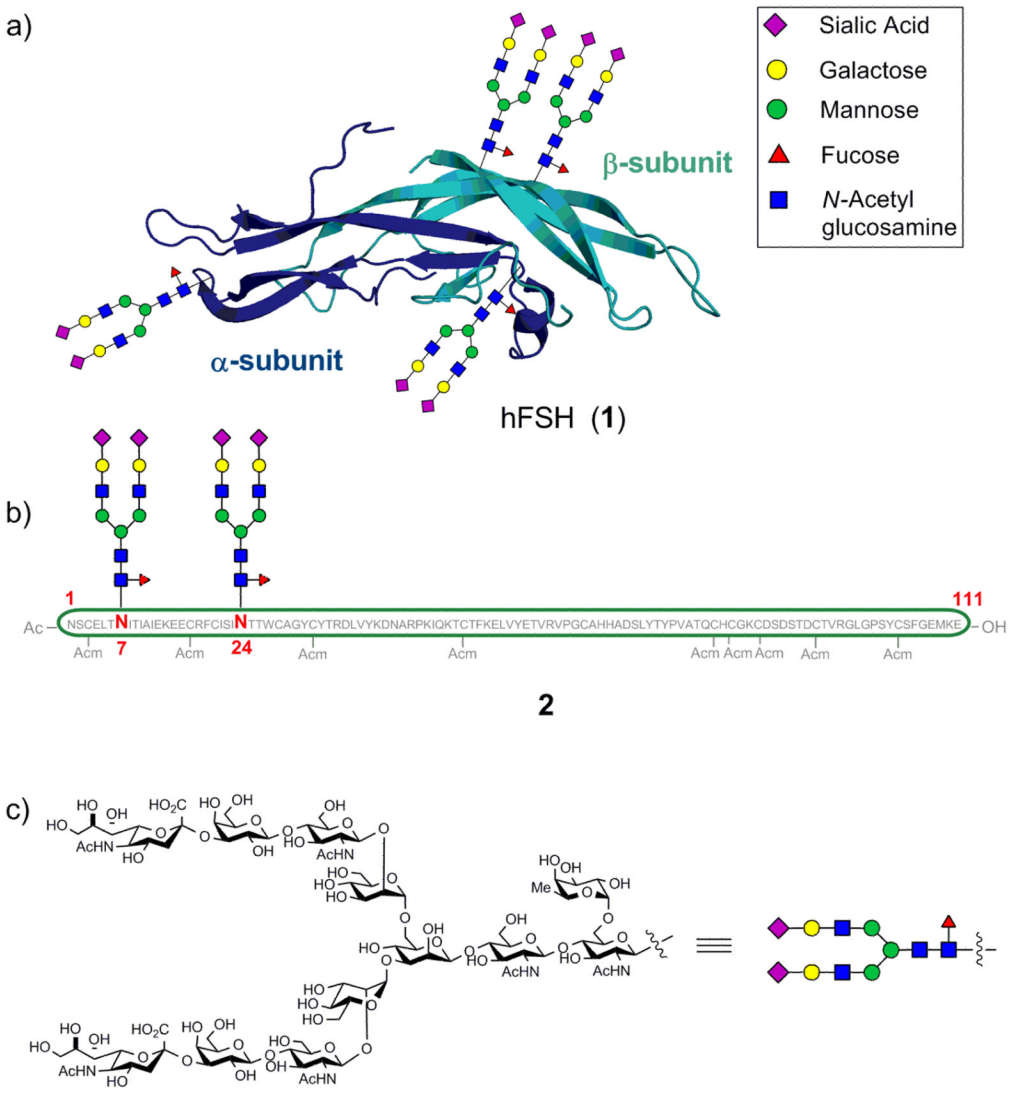
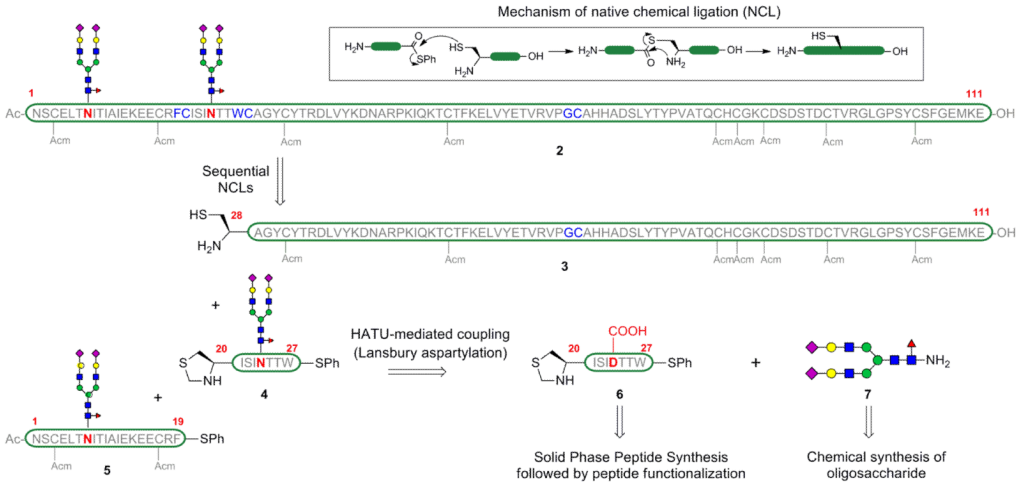
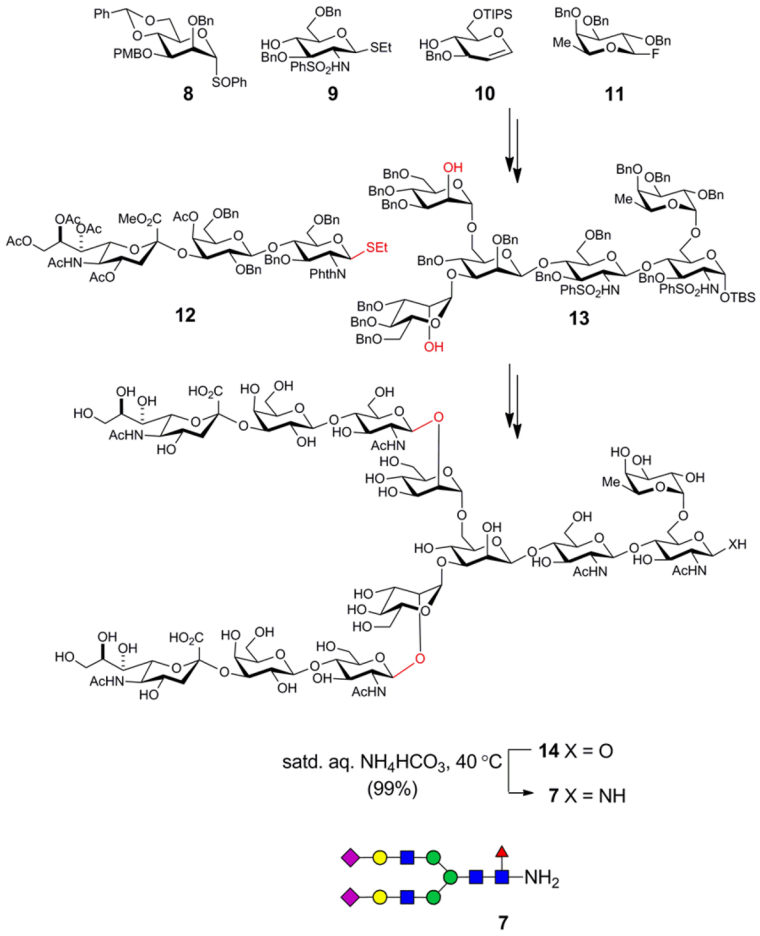


Figure 1.

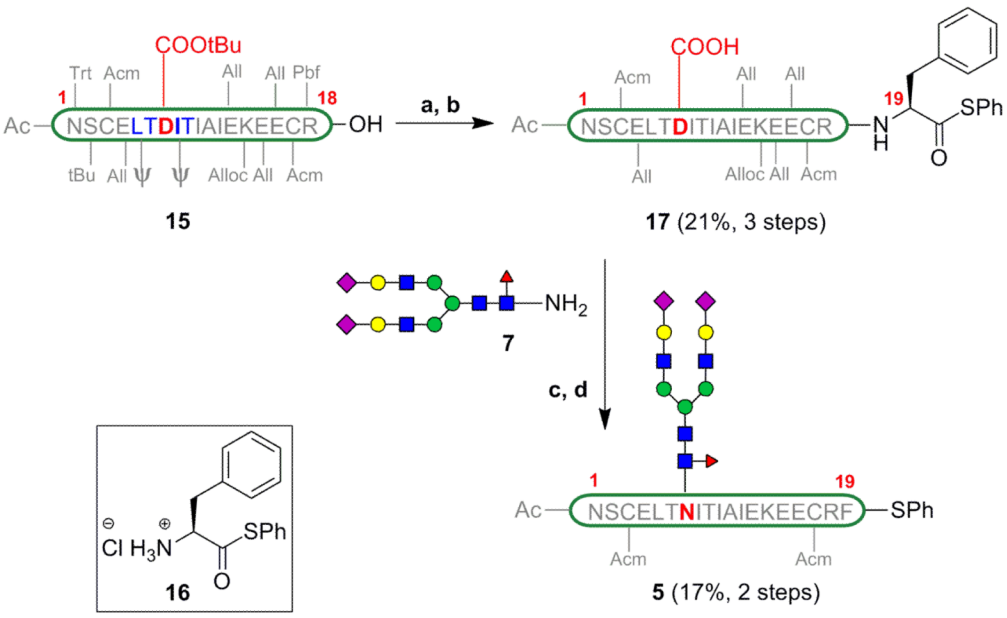
(a) Structure of hFSH heterodimer (**1**) with glycans (see inset legend). (b) β-subunit (**2**) displaying the *N*-linked consensus sequence oligosaccharide at the wild-type sites. (c) Structure of the consensus sequence oligosaccharide.



Scheme 1.
Retrosynthetic strategy for the construction of the fully elaborated β-subunit of hFSH (**2**).

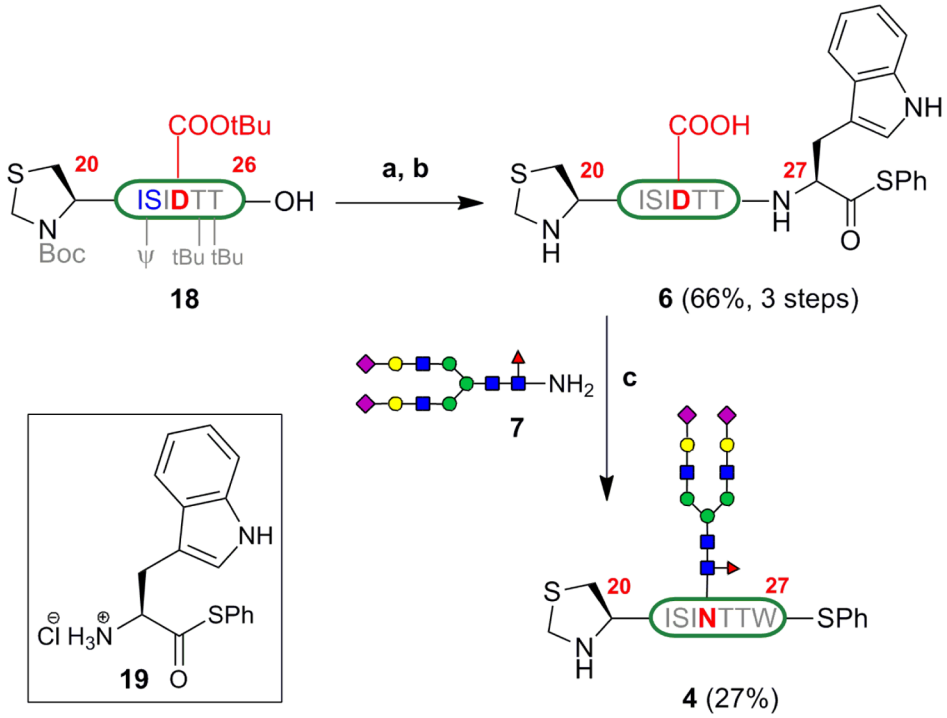


Scheme 2.
Schematic representation of the synthesis of the dodecasaccharide **7**.



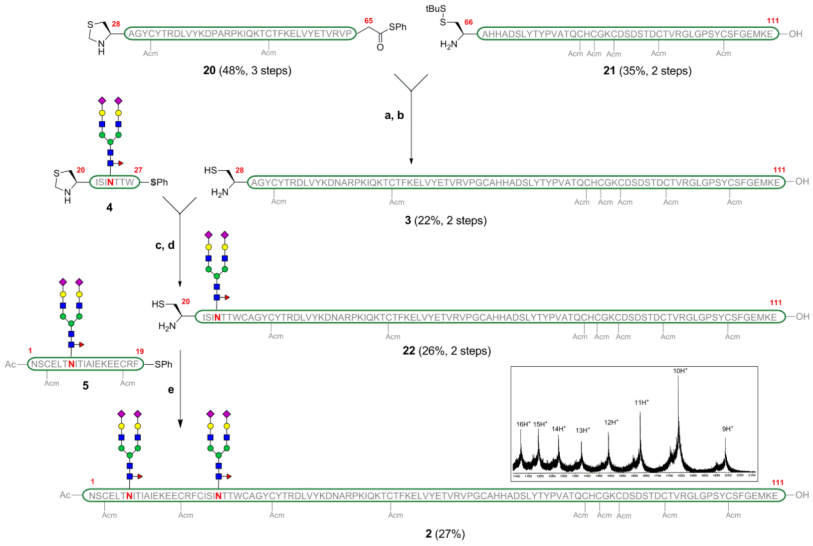
Scheme 3.

Synthesis of [βFSH$^{1–19}$] **5** (a) **16**, HOOBt, EDC, TFE/$CHCl_3$ 1:3; (b) Cocktail B; (c) **7**, HATU, DIEA, DMSO; (d) $Pd(PPh_3)_4$, $PhSiH_3$, NMP.



Scheme 4.

Synthesis of [βFSH$^{20–27}$] **4** (a) **19**, HOOBt, EDC, TFE/$CHCl_3$ 1:3 (b) Cocktail B (c) **7**, HATU, DIEA, DMSO.



Scheme 5.

Final ligation of the glycopeptide fragments. (a) **20**, **21**, PhSH, Gnd·HCl, Na_2HPO_4, TCEP·HCl, H_2O (pH = 7.4) (b) NH_2OMe·HCl, Gnd·HCl, TCEP·HCl (pH = 4.8) (c) **4**, PhSH, Gnd·HCl, Na_2HPO_4, TCEP·HCl, H_2O (pH = 7.4) (d) NH_2OMe·HCl, Gnd·HCl, TCEP·HCl (pH = 4.8). (e) **5**, PhSH, Gnd·HCl, Na_2HPO_4, TCEP·HCl, H_2O (pH = 7.3). Mass spectrum of compound **2** (inset).