Homogeneous Glycoprotein Synthesis Utilizing Bifunctional Glycosyl α-Amino Thioacid

by

Kota Nomura

Referees: Dr. Yasuhiro Kajihara,
Dr. Hironobu Hojo, Dr. Koichi Fukase and Dr. Yukishige Ito

Department of Chemistry
Graduate School of Science
Osaka University

2022

Table of contents

Homogeneous Glycoprotein Synthesis Applying Bifunctional Glycosyl α-Amino Thioacid

（二官能性糖鎖結合型αアミノチオアシッドを利用した糖タンパク質精密合成）

Abstract

Glycosylation is a major modification of secreted and cell surface proteins, however the resultant glycans show a considerable heterogeneity in their structures. To understand the biological functions arising from each glycoform, the preparation of homogeneous glycoproteins is essential for extensive biological experiments. To establish a more robust and rapid synthetic route of homogeneous glycoproteins, several key reactions based on α-amino thioacids have been studied. Extensive studies found that diacyl disulfide coupling (DDC) between a glycosyl asparagine thioacid **B** and a peptide thioacid **A** yielded a glycopeptide thioacid **C** (Figure 1).

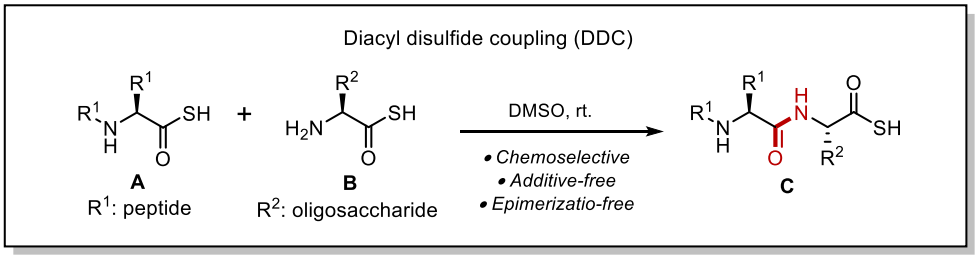


Figure 1. Reaction scheme of diacyl disulfide coupling (DDC).

This efficient coupling reaction enabled me to develop a new glycoprotein synthesis method, chemical glycan insertion strategy, which can insert glycosyl asparagine **E** between two unprotected peptides **D, F** (Figure 2). The first coupling is DDC between a peptide thioacid **D** and a glycosyl asparagine thioacid **E**. Because the resultant glycopeptide has a thioacid form at its C terminus, the thioacid capture ligation (TCL) can apply to the coupling of glycopeptide thioacid and another peptide **F** having *p*-nitropyridylsulfide (Npys) group at its N-terminus to afford the full-length glycoprotein **G**. Subsequent desulfurization of glycoproteins **G** can finally yields a natural form of glycoproteins **H**.

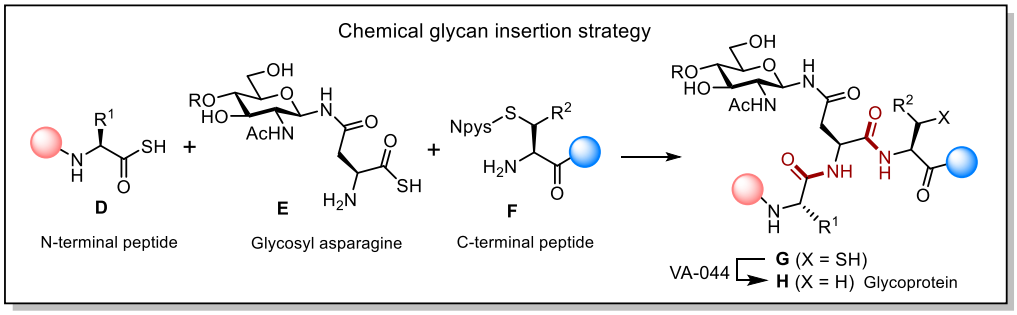


Figure 2. Chemical glycan insertion strategy for the synthesis of glycoproteins.

Applying this strategy, bioactive cytokines, interleukin 3 (IL3, **1**), CC chemokine ligand 1 (CCL1, **2**) and serine protease inhibitor kazal type 13 (SPINK13, **3**) having a biantennary sialyloligosaccharide respectively were synthesized within a few steps (Figure 3). Previous

glycoprotein synthesis methods required valuable glycosyl asparagine in the early stage and subsequent multiple glycoprotein synthesis routes. However, the developed new concept can generate glycoproteins within a few steps from peptides and glycosyl asparagine thioacids.

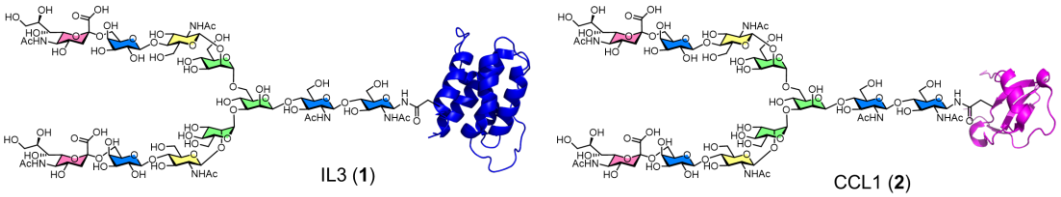


Figure 3. Synthetic examples utilizing chemical glycan insertion strategy.

In terms of the synthesis of SPINK13 **3**, DDC and TCL should be performed with the low reactive proline and bulky valine and the synthesis is therefore challenging. The first a glycosyl asparagine thioacid **4** was coupled to an N-terminal peptide prolyl thioacid **5** using DDC to afford the glycopeptide thioacid **6** in 28% yield. Then, the resultant glycopeptide thioacid **6** was coupled with C-terminal peptide having the Npys group at β-mercaptovaline **7** utilizing TCL to give a full-length glycoprotein **8** in 59% yield. After desulfurization and deprotection of Acm protecting groups convergently afforded the linear glycoprotein polypeptide **9** and stepwise dialysis folding conditions gave the desired glycoproteins folded SPINK13 **3** (Figure 4).

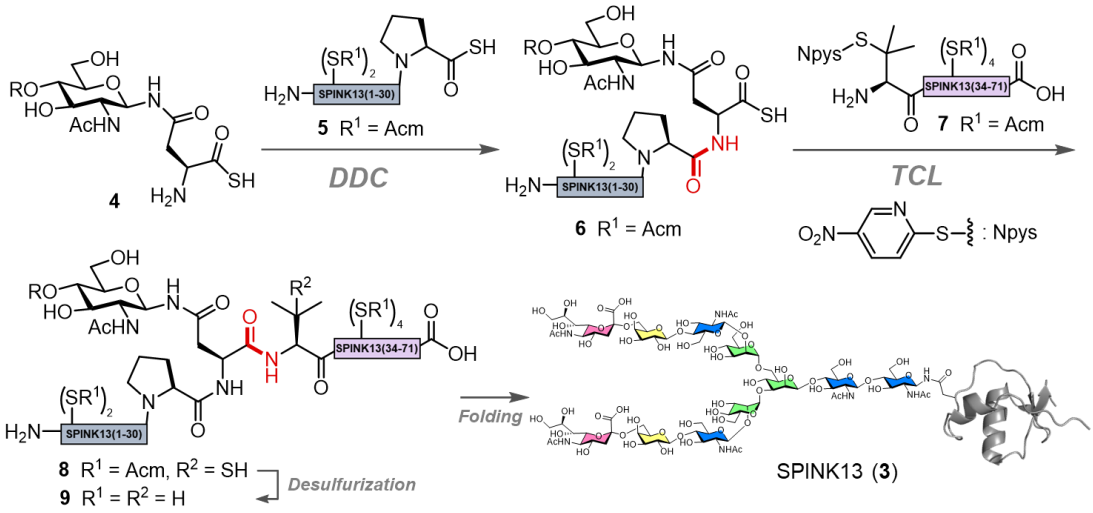


Figure 4. Synthetic scheme of serine protease inhibitor Kazal type 13 (SPINK13, **3**).

Furthermore, using homogeneous glycoproteins IL3 **1** and SPINK13 **3** synthesized by the developed strategy as chemical probes, several in vitro bioassays could be performed. These achievements indicate that our new synthetic strategy of homogeneous glycoproteins can be efficiently applied to not only the synthesis of chemical probes but also the various bioassays for the elucidation of the glycan structure-bioactivity relationship.

Abbreviation

ADC: antibody drug conjugate

AIBN: 2,2'-(diazene-1,2-diyl)bis(2-methylpropane nitrile)

ATC: amino thioacid coupling

Ac: acetyl

Acm: acetamidomethyl

Acr: acridine

Ala: alanine

Ar: aryl

Arg: arginine

Asn; asparagine

Asp: aspartic acid

Aux: auxiliary

BCP: B-cell progenitor

BME: 2-mercaptoethanol

BSA: bistrimethylsilylacetamide

Bn: benzyl

Boc-OSu:N-(tert-Butoxycarbonyloxy)succinimide

Bz: benzoyl

CAN: ceric ammonium nitrate

CCL1: CC chemokine ligand 1

CD: circular dichroism

CFCs: colony forming cells

CLP: common lymphoid progenitor,

CMP: common myeloid progenitor

CMP: cytidine monophosphate

CNX: calnexin

COSY: correlation spectroscopy

CPaseY: carboxypeptidase Y

CRT: calreticulin

CSFs: colony-stimulating factors

Cys: cysteine

DBU: 1,8-diazabicyclo[5.4.0]undec-7-ene

DCC: N,N'-Dicyclohexylcarbodiimide

DCM: dichloromethane

DDC: diacyl disulfide coupling

DIBAL: diisobutylaluminium hydride

DIC: N,N'-diisopropylcarbodiimide

DIPEA: N,N-diisopropylethylamine

DIPSI: decoupling in the presence of scalar interactions

DKP: diketopiperazine

DMAP: 4-dimethylaminopyridine

DMF: N,N-dimethylformamide

DMSO: dimethyl sulfoxide

DTT: DL-dithiothreitol

EDT: 1,2-ethanedithiol

EPO: erythropoietin (EPO),

ER: endoplasmic reticulum

ESI: electrospray ionization

Endo-A: endo-β-N-acetylglucosaminidases from *Arthrobacter protophormiae*

Endo-M: endo-β-N-acetylglucosaminidases from *Mucor hiemalis*

FCMA: formimidate carboxylate mixed anhydride

Fmoc:(9H-Fluoren-9-yl)methyl carbonochloridate

GM-CSF: granulocyte macrophage colony-stimul ating factors

GMP: granulocyte–macrophage progenitor

GPCR: G protein-coupled receptor

Gal: galactose

GalNAc: N-acetyl galactosamine

Gdn: guanidine

Glc: glucose

GlcNAc: N-acetyl glucosamine

Gln: glutamine

Glu: glutamic acid

Gly: glycine

HATU: O-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate

HBTU: 3-[Bis(dimethylamino)methyliumyl]-3H-benzotriazol-1-oxide hexafluorophosphate

HCTU: O-(6-Chlorobenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate

HMBC: heteronuclear multiple-bond correlation spectroscopy

HMPB: 4-(4-Hydroxymethyl-3-methoxyphenoxy) butyric acid

HOBt: 1-hydroxybenzotriazole

HPLC: high performance liquid chromatography

HRMS: high-resolution MS

HSC: hematopoietic stem cell

HSQC: heteronuclear single-quantum correlation spectroscopy

His: histidine

ICOS: inducible T cell costimulator

IL3: interleukin 3

INF: interferon

IgG: Immunoglobulin G

Ile: isoleucine

KHMDS: potassium hexamethyldisilazide

Leu: leucine

Lys: lysine

MBA: 4-mercaptobenzoic acid

MEP: megakaryocyte erythroid progenitor

MMPs: matrix metalloproteinases

MPA: 3-mercaptopropionic acid

MPAA: 4-mercaptophenylacetic acid

MS: mass spectrometry

MSNT: 1-(2-Mesitylenesulfonyl)-3-nitro-1H-1,2,4-triazole

Man: mannose

Mes: mesitylene

Mesna: sodium 2-mercaptoethanesulfonate

Met: methionine

Mob: 4-methoxybenzyl

Mtr: 4-methoxy-2,3,6-trimethylbenzenesulphonyl

NCL: native chemical ligation

NIS: N-Iodosuccinimide

NMR: nuclear magnetic resonance

Neu5Ac: N-acetylneuraminic acid

NOESY: nuclear overhauser effect spectroscopy

NTA: nitrosothioacid

Npys: nitropyridyl thiol

Ns: 2-Nitrobenzenesulfonyl

OST: oligosaccharyl transferase

PMB: p-methoxybenzyl

PTM: post-transrational modification

Ph: phenyl

Phe: phenylalanine

Pro: proline

PyBOP: (Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate

RNase: ribonuclease

RP: reverse phase

Rt: room temperature

SGP: sialylglycopeptide

SPINK13: serine protease inhibitor Kazal type 13

SPPS: solid phase peptide synthesis

STL: serine threonine ligation

Boc: tert-butoxycarbonyl

SUMO: small ubiquitin-related modifier

Sec: selenocysteine

Ser: serine

TBTU: 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethylaminium tetrafluoroborate

TCEP: tris(2-carboxyethyl)phosphine)

TCL: thioacid capture ligation

TFA: 2,2,2-trifluoroacetic acid

TfOH: trifluoromethanesulfonic Acid

TFE: 2,2,2-trifluoroethanol

THF: tetrahydrofuran

TIPS: triisopropylsilane

TOCSY: total correlation spectroscopy

TPO: thrombopoietin

Thr: threonine

Thz: thiazolidine

Tris: 2-Amino-2-hydroxymethyl-propane-1,3-diol

Trp: tryptophane

Trt: trityl

Ts: tosyl

Tyr: tyrosine

VA-044: 1,2-Bis(2-(4,5-dihydro-1H-imidazol-2-yl)propan-2-yl)diazene dihydrochloride

Val: valine

iPr: iso-propyl

tBu: tert-butyl

uPA: urokinase plasminogen activator

uPAR: urokinase plasminogen activator receptor

Chapter 1. General introduction: background and synthesis of glycoprotein

1-1. Posttranslational modification

In post-genomic era, post- and co-translational modifications (P/C-TMs) of proteins are known as more essential elements for protein function activation.[1] There are many types of P/C-TMs such as N-acetylation, methylation, phosphorylation, ubiquitination, hydroxylation, glycosylation, sulfation, kynurenine formation and N-terminal pyroglutamate formation (Figure 1-1). Recent research has revealed that these modifications regulate a wide range of biological processes, including several diseases such as cancer, Alzheimer's disease, and Parkinson's disease.[2-5] These modification reactions on proteins have been known for a long time, however in recent years, they have frequently focused on as not only essential basic biochemistry subjects but also as targets for drug discovery in the pharmaceutical industry. Since these P/C-TMs of proteins are extremely diverse in vivo, it is difficult to isolate homogeneous proteins modified from living organisms and analyze the function at the molecular level.

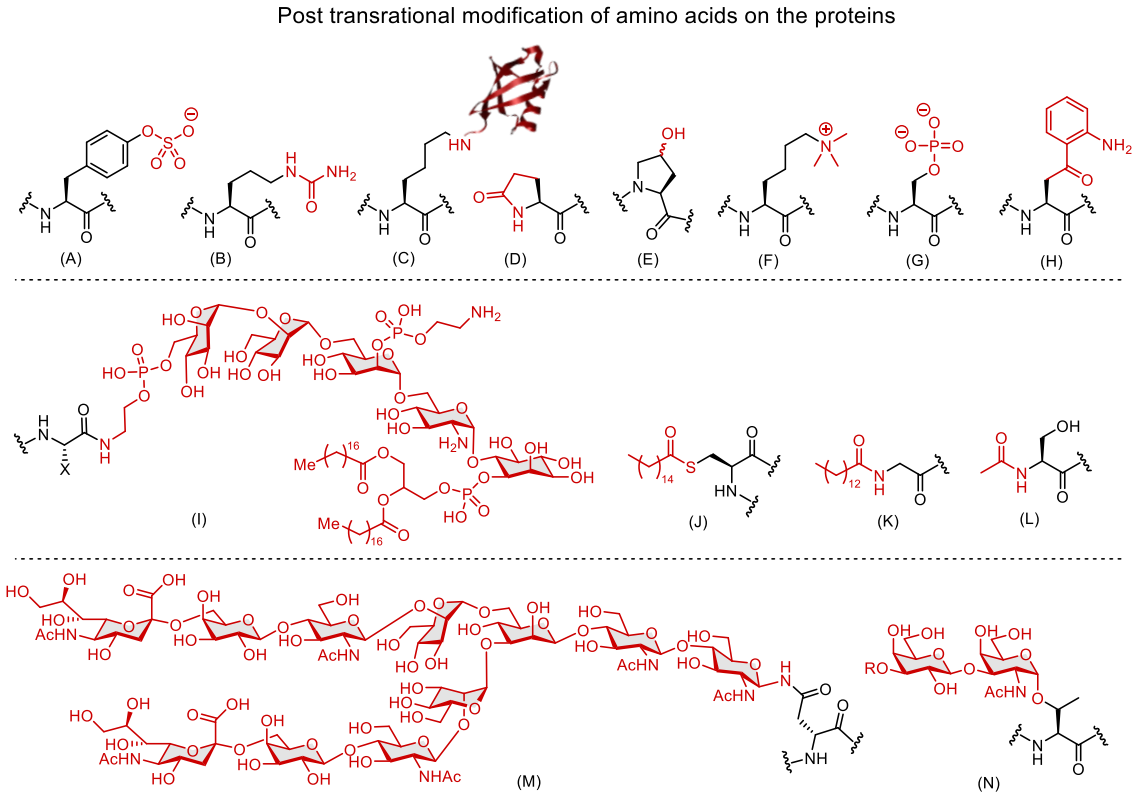


Figure 1-1. Examples of post- and co-transrational modifications (P/C-TMs) of proteins.

[(A) sulfation at Tyr. (B) citrullination at Arg. (C) ubiquitination at Lys. (D) pyroglutamate at N-terminal Gln. (E) hydroxylation at Pro. (F) methylation at Lys and Arg. (G) phosphorylation at Ser, Tyr, Thr. (H) kynurenine formation at Trp. (I) GPI anchor at C-terminus of peptides. (J) S-palmitoylation at Cys. (K) *N*-Myristylation at *N*-terminal Gly. (L) *N*-Acetylation at *N*-terminus of peptides. (M) *N*-Glycosylation at Asn. (N) *O*-Glycosylation at Thr and Ser.]

1-2. Glycosylation of protein

In protein modifications, glycosylation is a major modification, whereas the resultant glycans show considerable heterogeneity in their structure (Figure 1-2 A).[6] Glycoproteins are found on the cell surface and in body fluids, and these proteins are modified with serine/threonine-linked *O*-glycans or asparagine-linked *N*-glycans (Figure 1-2 B).[7] In case of *N*-glycans, the oligosaccharides are almost attached to the nitrogen of asparagine side chain which is present as a part of consensus sequences; Asn-Xaa-Ser/Thr (Xaa: any amino acid except for Pro). Structures of N-glycans are mainly classified with three types such as complex type, high-mannose type, and hybrid type saccharide. All of them have a common core structure [Man-α-1,3-(Man-α-1,6)-Man-β-1,4-GlcNAc-β-1,4-GlcNAc-β-Asn] and partially modified by fucosylation, phosphorylation and sulfation. The biosynthesis of glycans is regulated by the substrate specificity of glycosyltransferases and glycosidases in endoplasmic reticulum (ER) and Golgi apparatus, resulting in considerably diverse glycan structures (glycoforms). Under these circumstances, we cannot identify which glycans play an important role in sustaining specific biological events.[8-11]

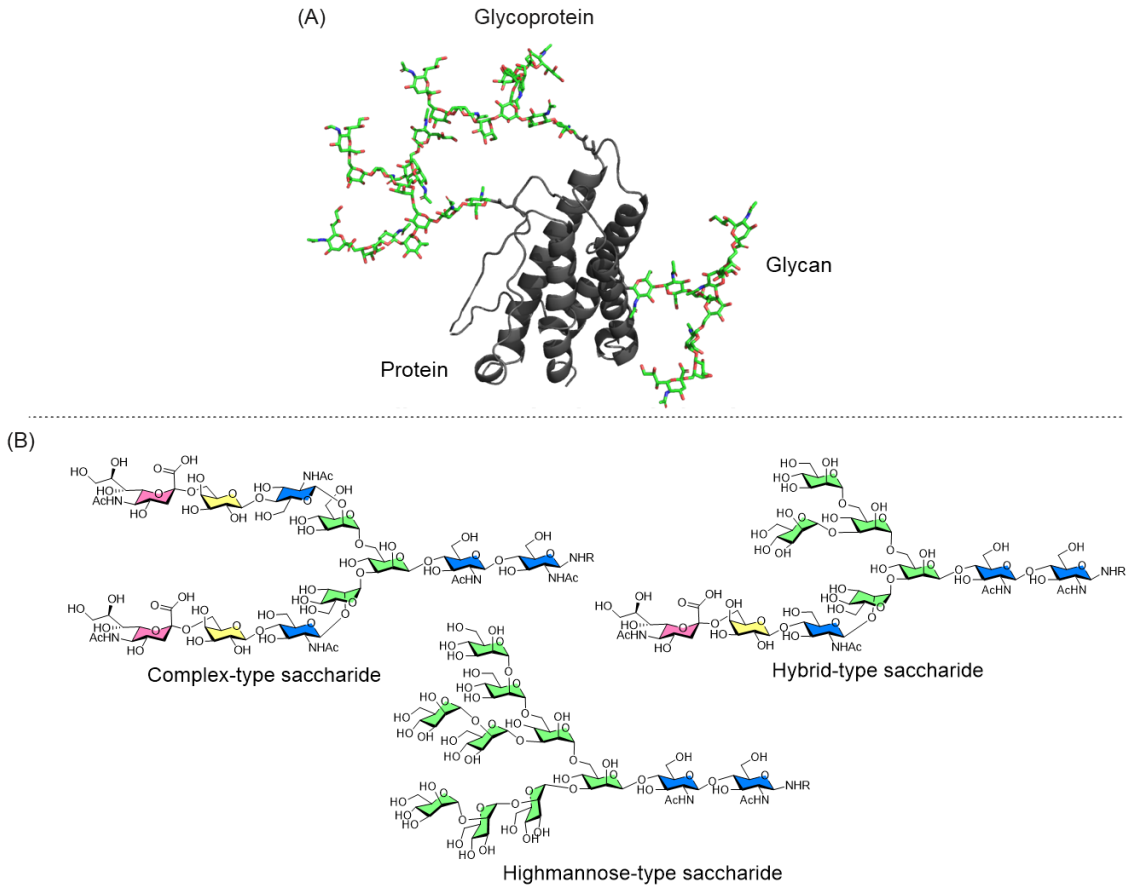


Figure 1-2. Glycoprotein and glycan presented on the cell surface. Diversity of glycan structures presented on the glycoproteins. Complex, Hybrid and High mannose type saccharides.

1-3. Biosynthesis of glycoprotein

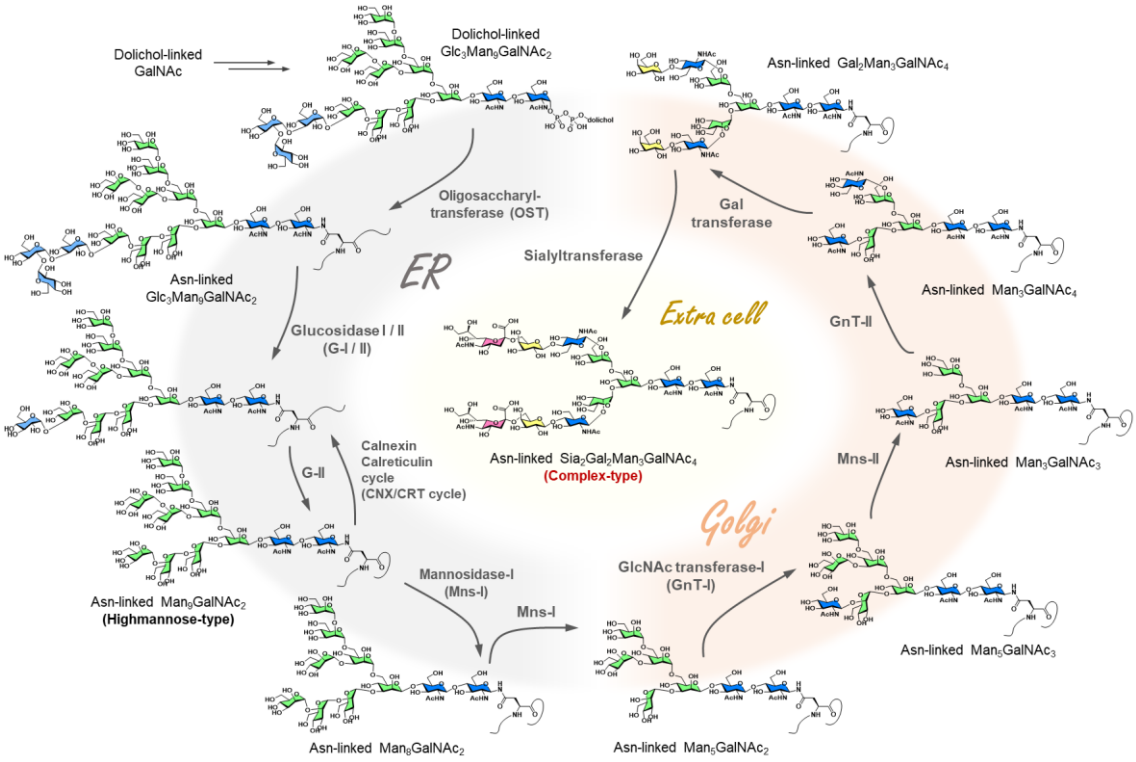


Figure 1-3. Biosynthesis of *N*-glycan of the glycoprotein.

Glycan structures on glycoproteins are constructed by the addition of monosaccharides to proteins in a cotranslational or posttranslational manner by various types of enzymes (glycosyltransferases, glycosidases and other carbohydrate modifying enzymes). In the case of O-linked glycans, N-acetyl-α-D-galactosamine (GalNAc) is attached to the hydroxy groups of Thr and Ser, and further glycosylation in the Golgi apparatus results in the formation of major eight types of core structures of O-linked glycans.[12] In contrast, N-linked glycans, oligosaccharyltransferase (OST) transfer a $Glc_3Man_9GlcNAc_2$ glycan unit to the nitrogen atom of the amide group of Asn in the protein consensus sequence Asn-Xaa-Ser/Thr in the endoplasmic reticulum (ER) from $Glc_3Man_9GlcNAc_2$ glycan-PP-dolichols, where PP is pyrophosphate (Figure 1-3).[13] The glycan precursors are trimmed by glucosidase I/II (G-I/II), converting to mono-glucosylated saccharide $Glc_1Man_9GlcNAc_2$ glycan, which is a key compound to the calnexin-calreticulin (CNX/CRT) cycle in the ER protein quality control system (folding system).[14] This folding process is monitored by several enzymes and chaperones to identify misfolded and correctly folded glycoproteins.[15] Correctly folded glycoproteins are transported to the Golgi apparatus, while misfolded glycoproteins are transferred to the cytoplasm for degradation. If the protein is properly folded, glucose by G-II is removed and converted to $Man_9GlcNAc_2$ glycan. After removal of glucose, mannose of oligosaccharide is trimmed by

mannosidase-I (Mns-I) and converted to $Man_8GlcNAc_2$ glycans which are transported to the Golgi apparatus where further glycosylation takes place by GlcNAc transferase, galactosyltransferase and sialyltransferase. Consequently, the diversity of glycan structures increases to high mannose type, complex type, and hybrid type glycans mainly (Figure 1-2 B).[16] In terms of complex-type structure, which is a multi-branched structure with a sialic acid-containing motif capping the end and a clearly modified core pentasaccharide (Figure 1-4). Finally, they are transported to the cell surface as membrane proteins and to the extra cell as secretory proteins. In this process, uncountable repertoire of oligosaccharides is produced. Therefore, it is difficult to elucidate to which glycans are involved in transport, secretion, and bioactivity.

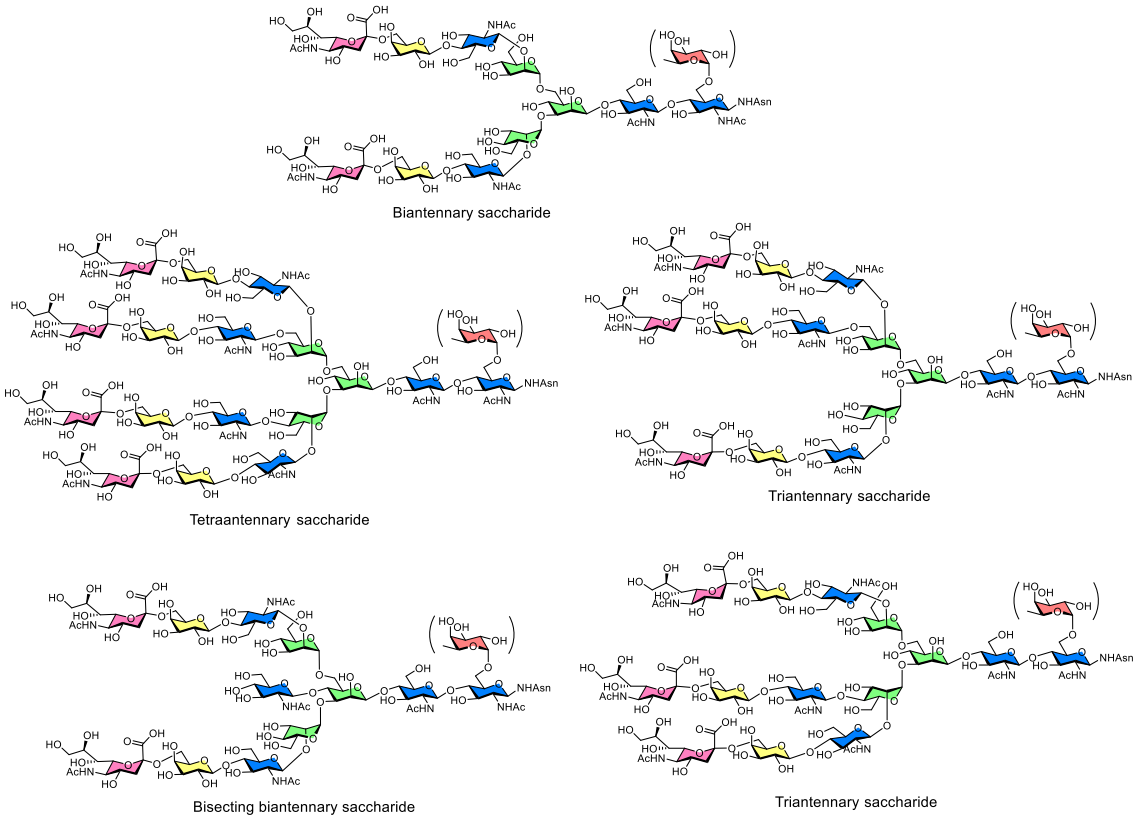


Figure 1-4. Divergency of complex saccharide. The saccharide shown here are all complex type saccharides. There are biantennary, tetraantennary, triantennary and bisecting types.

The diverse glycosylation processes increase the diversity of the resultant glycan structures of glycoproteins and make the properties of proteins more complex. However, the structural heterogeneity of glycosylation makes the functional analysis of glycoproteins very difficult. Unfortunately, homogeneous glycoprotein synthesis in protein expression methods is not practical, and current purification techniques is extremely difficult to isolate sufficient amounts of proteins depending on different glycan structures. The synthesis of glycoproteins with uniform glycan

structures is very important for functional analysis of glycoproteins and biochemical experiments, and methods to synthesize glycoproteins with chemically and biologically uniform structures are being actively studied. In this context, many interesting synthetic methods have been developed and summarized in many reviews.

1-4. Synthetic strategy of glycoprotein

To understand the biological processes arising from each glycoform, the preparation of homogeneous native glycoproteins is essential for extensive biological experiments. This section describes how glycoproteins having homogeneous glycan structures can be synthesized. Basically, to synthesize glycoproteins, the chemical synthesis method can be applied. The most common method for glycopeptide synthesis is solid phase peptide synthesis (SPPS) using resin. Although several biologically active glycopeptide chains have been synthesized by relatively convenient method based on SPPS, this method cannot be applied to glycoproteins directly. The first reason is that the yield of the condensation reaction of glycans and fully protected peptides is very low. The second reason is low solubility of the protected long-chain peptides in aqueous and organic solvents. The third reason is *O*-glycoside linkages of glycans are decomposed by strong acidic deprotection conditions for global deprotection of peptides such as TFA. To solve these problems in glycoprotein synthesis, we can synthesize glycoproteins by using the following synthetic strategies (Figure 1-5). The synthesis methods depend on how the glycan building blocks are attached to the protein. In this chapter, the synthetic strategies of glycoproteins are classified into three types according to the linkages to be cleaved for the retro synthetic analysis: direct enzymatic glycosylation (Figure 1-5; *Chapter 1-6*), chemoenzymatic remodeling of glycoproteins (Figure 1-5; *Chapter 1-7*), total chemical synthesis and semisynthesis (Figure 1-5; *Chapter 1-9*) following the synthesis of glycopeptides (*Chapter 1-8*). The syntheses of homogeneous glycan building blocks are also reviewed (*Chapter 1-5*).

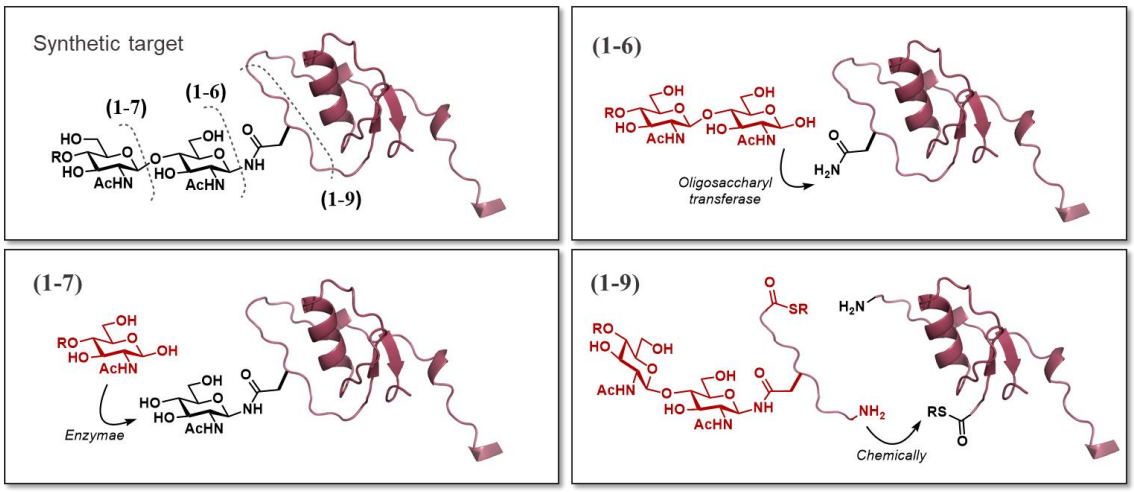


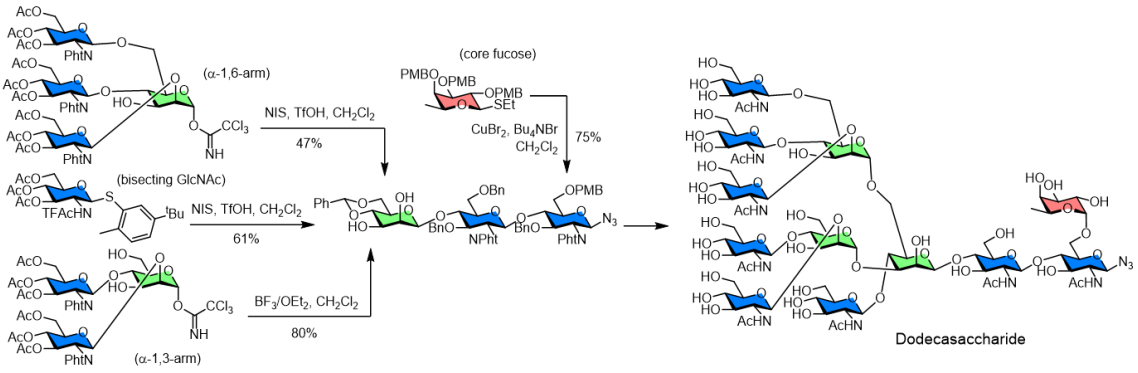
Figure 1-5. Synthetic strategy for the native glycoprotein having a homogeneous glycan.

In addition to the synthesis of native glycoprotein forms, glycoprotein analogs with non-natural linkages have also been developed by many groups. For example, Davis groups achieved the

convergent synthesis of glycoprotein analog by using a dehydroalanine (Dha).[17] Kajihara grouups applied a bromo acetamide functional group for the rapid synthesis of glycoprotein analog having unnatural linkage at N-glycosylation sites.[18-19] However, this chapter focuses on the synthesis of glycoproteins with natural-type binding linkage.

1-5. Synthesis of glycan building blocks

As mentioned in the previous section, N-linked glycans are structurally very complex and therefore, synthesis of diverse oligosaccharide forms needs to regulate difficult conversion steps such as α/β and regioselective glycosylation with high yield in order to use enough amount of glycans for the synthesis of glycoproteins.[20] However, in recent years, fortunately, many convergent chemical and enzymatic glycosylation protocols for the synthesis of complex saccharide have been developed. For example, highly branched N-glycans are synthesized by some groups. Eller and coworkers reported the first chemical synthesis of highly branched pentaantennary N-glycans (dodecasaccharide) and its derivatives with bisecting and core fucosyl modifications (Figure 1-6).[21] The synthetic method involved oligosaccharyl building block such as tetrasaccharide mannosyl donors (α-1,6-arm), trisaccharide mannosyl donors (α-1,3-arm), GlcNAc thioglycoside donors (bisecting parts) and core fucosyl thioglycoside and these building blocks were sequentially coupled to core trisaccharide acceptors for the synthesis of the target dodecasaccharide under the convergent manner. The glycosylation reactions toward the hindered acceptors were systematically optimized after the identification of key protecting groups which reduce the proximal and peripheral crowding. The use of a universally applicable modular set of building blocks became the base to facilitate the general chemical synthesis of branched N-glycans for their N-linked glycoprotein synthesis.

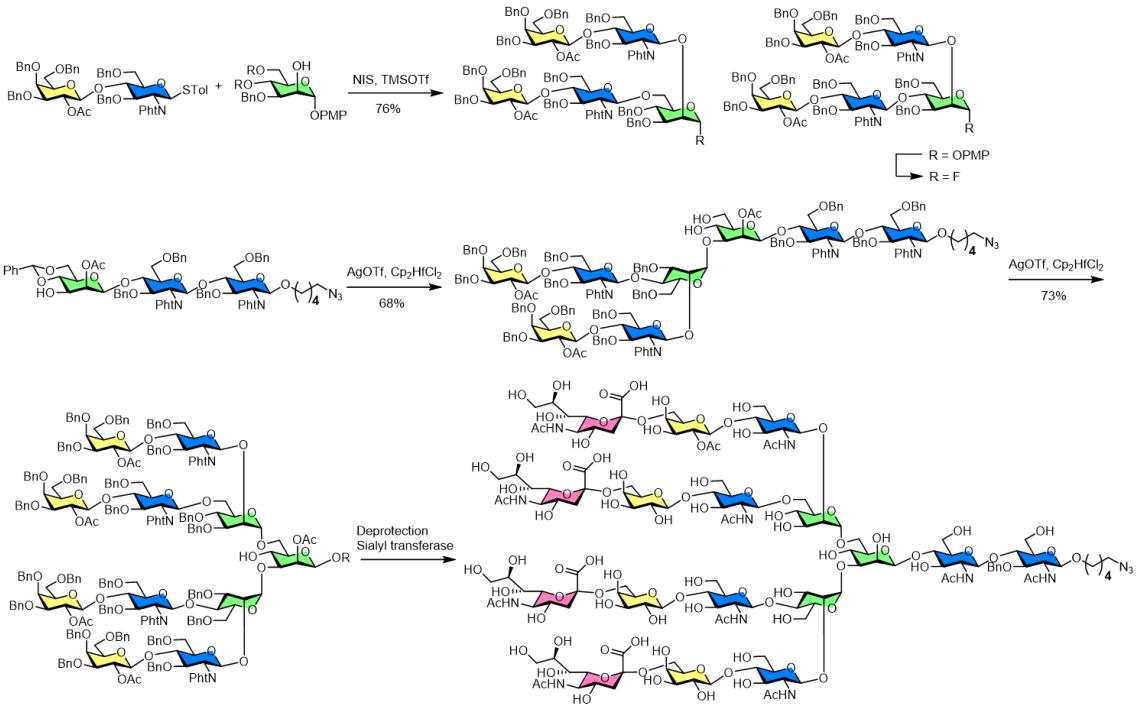


Eller, S.; Schuberth, R.; Gundel, G.; Seifert, J.; Unverzagt, U. *Angew. Chem. Int. Ed.* **2007**, *46*, 4173

Figure 1-6. The Unverzagt group: synthesis of dodecasaccharide.

Almost all of the complex-type glycans in living organisms are known to have biantennary, triantennary, and tetraantennary glycan structures, as described in the previous chapter. By constructing middle size glycans as building blocks convergently, Shivatare and coworkers achieved efficient chemical synthesis of tetraantennary, triantennary, and biantennary glycans (Figure 1-7).[22] They reported unified convergent strategy for the rapid production of multi-antennary complex type N-glycans with and without terminal N-acetylneuraminic acid residues connected via the α-2,6 or α-

2,3 linkages. They chemically synthesized the target asialo glycans by glycosylation of a pentasaccharide fluoroglycoside with a branched structure to a core trisaccharide acceptor utilizing Koenigs–Knorr glycosylation. Finally, α-2,6-sialylation was carried out using sialyltransferases, and a library of sialyl complex-type glycans was successfully synthesized. Using sialyltransferases to install sialic acids could minimize synthetic steps through the use of intermediates to simplify the complicated procedures associated with conventional sialic acid chemistry. Finally, their synthetic complex oligosaccharides were compiled to create a glycan array for the profiling of HIV-1 broadly neutralizing antibodies PG9 and PG16 that were isolated from HIV infected donors. From the study of antibody PG16, they identified potential natural and unnatural glycan ligands, which facilitated the design of carbohydrate-based immunogens and haptens for the HIV vaccine development.

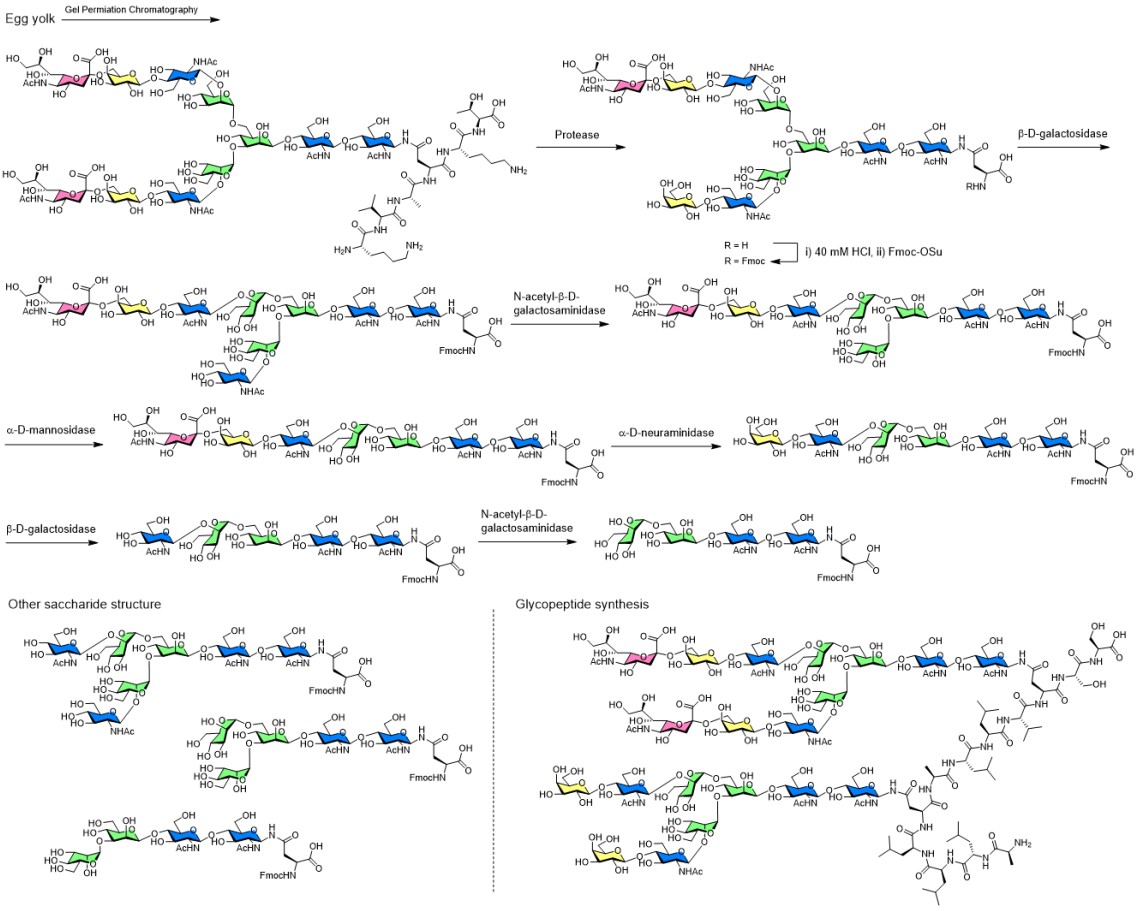


Shivatare, S. S.; Chang, S. H.; Tsai, T.; Ren, C. T.; Chuang, H. Y.; Hsu, L.; Lin, C. H.; Li, S. T.; Wu, C. Y.; Wong. C. H. *J. Am. Chem. Soc.* **2013**, *135*, 15382

Figure 1-7. The Wong group: synthesis of tetraantennary saccharide.

Furthermore, Seko and coworkers demonstrated that the biantennary complex type N-linked glycan can be readily isolated from chicken egg yolks in1997.[23] Applying this methodology, Kajihara and Fukae succeeded to synthesize glycan library using enzymes (Figure 1-8).[24-25] They reported the preparation of 24 species of homogeneous asparagine-linked oligosaccharides (Asn-oligosaccharides) from asparagine-linked biantennary complex-type sialylundecasaccharide obtained from egg yolk. Their synthetic strategy aimed at adapting branch specific exo-glycosidases digestion (α-D-neuraminidase, β-D-galactosidase, N-acetyl-β-D-glucosaminidase and α-D-mannosidase) of the

individual asialo-branch after preparation of monosialyloligosaccharides obtained from sialylundecasaccharid by acid hydrolysis of NeuAc. To perform branch specific exo-glycosidase digestion, isolation of pure monosialyloligosaccharides obtained was essential. They examined chemical protection with hydrophobic protecting (Fmoc and benzyl) groups to isolate two kinds of monosialyloligosaccharides. Finally, they obtained 24 Asn-linked oligosaccharides (100 mg scale) within a few weeks by branch specific exo-glycosidase digestions. In addition, they succeeded solid-phase synthesis of glycopeptide having Asn-linked sialyl-undeca- and asialo-nonasaccharides obtained by their enzymatic approaches.



Kajihara, Y.; Suzuki, S.; Yamamoto, N.; Sasaki, K.; Sakakibara, T.; Juneja, L. R. *Chem. Eur. J.* **2004**, *10*, 971

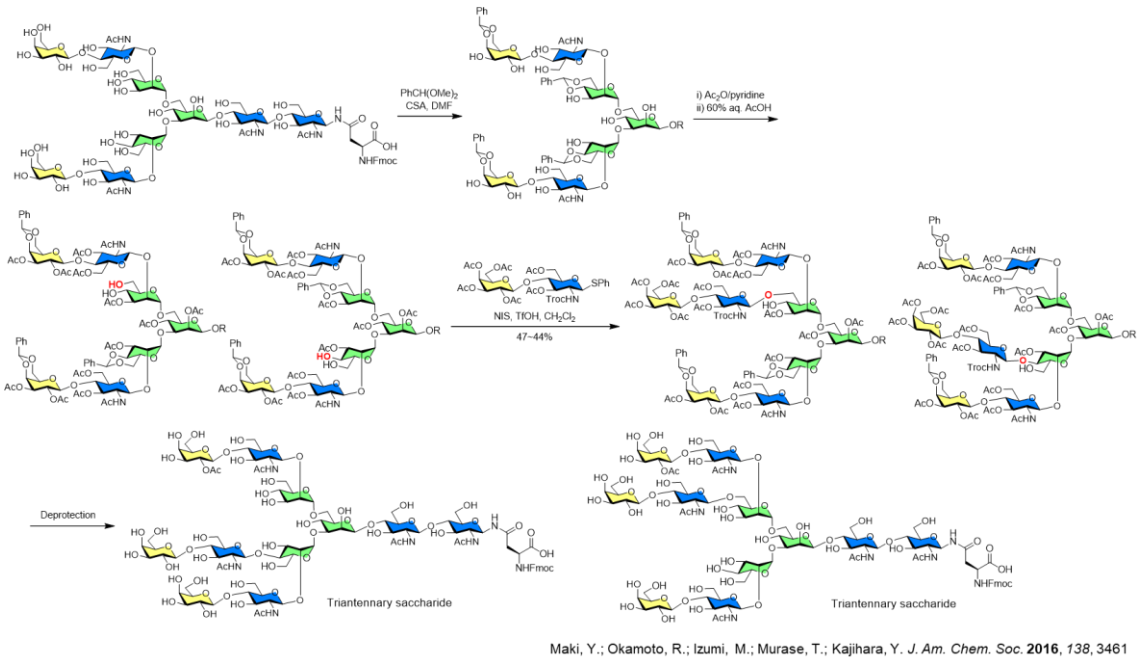
Fukae, K.; Yamamoto, N.; Hatakeyama, Y.; Kajihara, Y. *Glycoconj. J.* **2004**, *21*, 243

Seko, A.; Koketsu, M.; Nishizono, M.; Enoki, Y.; Ibrahim, H. R.; Juneja, L. R.; Kim, M.; Yamamoto, T. *Biochim. biophys. acta.* **1997**, *1335*, 23

Figure 1-8. Chemoenzymatic deglycosylation of saccharide and the synthesis of 24 species of homogeneous asparagine-linked oligosaccharides from sialylglycopeptide isolated from egg yolk.

Moreover, Maki and coworkers succeeded to synthesize triantennary complex type N-linked glycan convergently from asparagine-linked biantennary complex-type sialylundecasaccharide

obtained from egg yolk (Figure 1-9).[26] They reported a highly efficient strategy to access the representative two types of triantennary oligosaccharides through only 9- or 10-step chemical conversions from a biantennary oligosaccharide. Four benzylidene acetals were introduced to the terminal two galactosides and two core mannosides of the biantennary asialononasaccharide bearing 24 hydroxy groups, followed by protection of the remaining hydroxy groups with acetyl groups. Selective removal of one of the benzylidene acetals gave two types of suitably protected glycosyl acceptors.[27] Glycosylation toward the individual acceptors with protected Gal-β-1,4-GlcN thioglycoside and subsequent deprotection steps successfully yielded two types of complex-type triantennary oligosaccharides.



Maki, Y.; Okamoto, R.; Izumi, M.; Murase, T.; Kajihara, Y. *J. Am. Chem. Soc.* **2016**, *138*, 3461

Figure 1-9. Synthesis of triantennary complex type saccharide from asparagine-linked biantennary sialylundecasaccharide obtained from egg yolk by applying selective protection of benzylidene protecting groups.

As have introduced many synthetic approaches of *N*-glycan, introduction of these synthetic glycans into proteins using various techniques enabled us to synthesize glycoproteins with homogeneous glycan structures. In the following sections, this thesis will introduce five approaches of the synthesis of glycoproteins having homogeneous glycoform.

1-6. Direct enzymatic glycosylation

In the biosynthesis of glycoproteins, the key reaction is the enzymatic transfer of glycans from oligosaccharyl dolichol pyrophosphate to Asn located in the protein consensus sequence Asn-Xaa-Ser/Thr by oligosaccharyltransferase (OST), as described in the previous chapter. In other words, by manipulating OST *in vitro* and introducing glycans to the consensus sequence of prepared protein, glycoproteins can in principle be synthesized very easily. As for the mechanism of how OST transfer glycans to asparagine, Xu and Dempski found that glycans were transferred to the chemically synthesized benzoyl protected tripeptide Bz-[^{13}C,^{15}N]-Asn-Leu-Thr-NH_2 as a substrate of OST (Figure 1-10).[28] Their experiments demonstrated that glycopeptides could be synthesized by OST.[29]

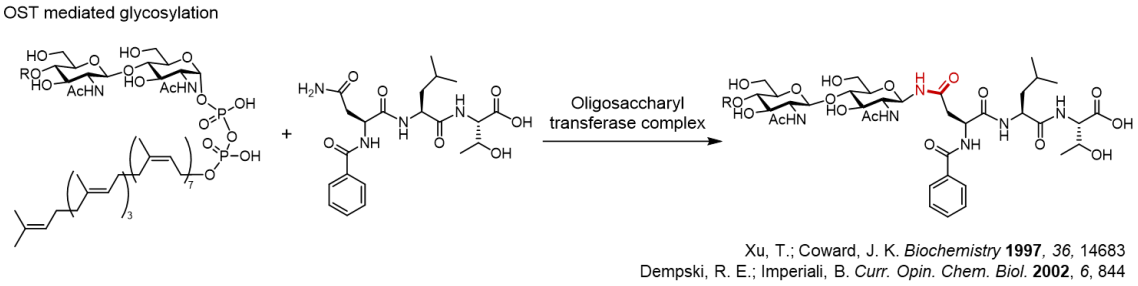


Figure 1-10. Oligosaccharyltransferase transfer N-linked glycans by dolichol-linked glycans and afford glycopeptides in one step.

However, the following points make it difficult to use OST for the *in vitro* glycosylation. The first is that the OST complex is comparably unstable in the laboratory. The second is that OST complex consists of a membrane-bound enzyme, which is highly complicated due to nine hydrophobic protein subunits are involved in the activity. The OST complex, which forms a supramolecular transrational machinery with ribosomes and the Sec61 complex. Especially in mammals, the formation of the OST complex is crucial for the oligosaccharide transferring activity. The third reason is that OST cannot transfer glycans to the folded proteins.[30]

In contrast, PglB, which is a single OST unit of gram-negative bacteria *Campylobacter jejuni*, can catalyze the N-glycan transfer reactions without forming a supramolecular complex. Catalyzing by PglB, the glycan is transferred the heptasaccharide; GalNAc-α-1,4-GalNAc-α-1,4-(Glc-β-1,3)-GalNAc-α-1,4-GalNAc-α-1,4-GalNAc-α-1,3-Bac, where Bac is abbreviation of bacillosamine (2,4-diacetamido-2,4,6-trideoxyglucose). Imperiali groups achieved the *in vitro* glycosylation using PglB and dolichol-linked disaccharides (Figure 1-11).[31-32] The preparation of additional undecaprenyl-linked glycan variants revealed the ability of PglB to transfer a wide variety of saccharides. They also

found the consensus sequence (D/E-X-N-X-S/T, X can be any amino acids except proline) for the glycosylation of PglB using a library of peptide substrates and a quantitative radioactivity-based *in vitro* assay (Figure 1-11).[32] In the case of transfer with a folded protein, the differences between variations of the glycosylation sequences were also found to be consistent with the trends observed from liner peptides. According to their attempts, they resulted that PglB glycosylated a folded protein where glycosylation sites are presented in flexible parts in the consensus sequence. However, they also found that PglB cannot be used for the glycosylation of chitobiose motifs, which is the mammalian substrate for the eukaryotic OT process. Thus, new glycosyltransferase, PglB, cannot efficiently transfer mammalian N-glycan from the corresponding glycolipid.

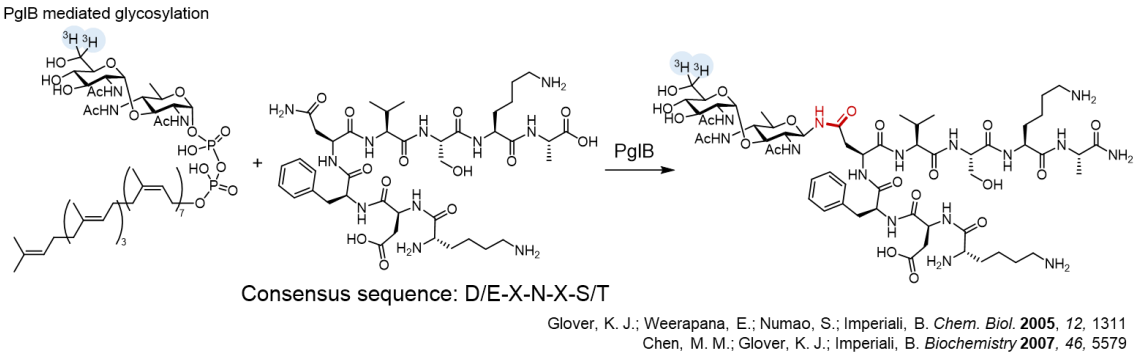


Figure 1-11. Transfer of glycans by using PglB and dolichol-linked bacterial glycan to asparagine residues of the peptides having the consensus sequence.

Further studies are needed to apply PglB-based glycosylation to mammalian-type glycoproteins. Rational protein engineering based on the crystal structure and directed evolution might accelerate the substrate specificity of PglB. Schwarz and coworkers characterized an alternative bacterial pathway wherein a cytoplasmic *N*-glycosyltransferase (NGT) uses nucleotide-activated monosaccharides (UDP-Glc) as donors to modify asparagine residues of peptides and proteins (Figure 1-12).[33] NGT was found that an inverting glycosyltransferase and recognizes the consensus sequence (N-X-S/T) and exhibited same acceptor specificity as eukaryotic OST in different catalytic mechanism.

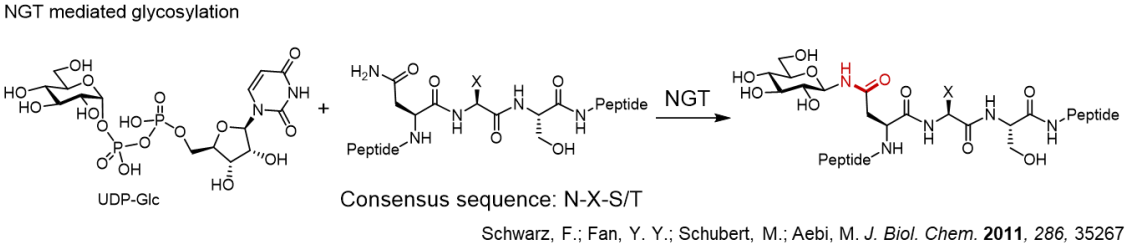
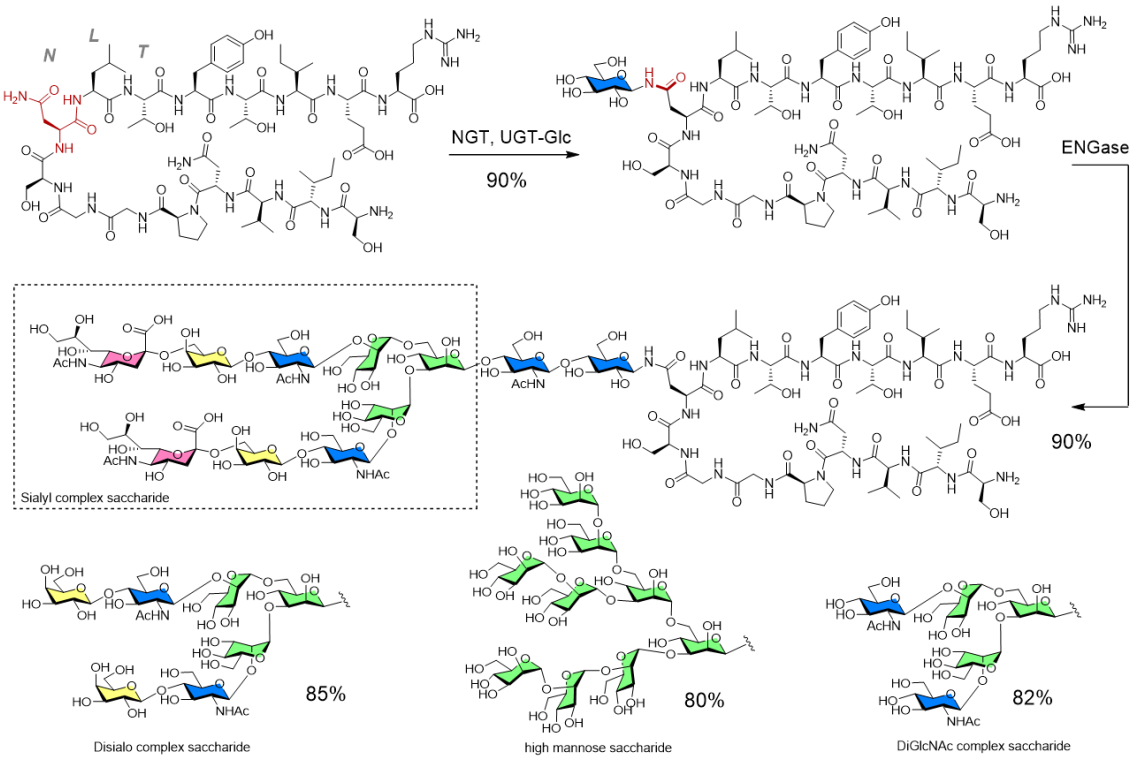


Figure 1-12. NGT mediated glycosylation with UDP-Glucose.

Applying NGT glycosylation, Lomino and coworkers synthesized glycopeptides having complex type sialyloligosaccharides combining with endoglycosidase (ENGase) catalyzed transglycosylation.[34] In detail their method consists of two enzymatic glycosylation steps (Figure 1-13). First step is introduction of a glucose to Asn at the consensus sequence using NGT with UDP-Glc. Second step is the attachment of a complex saccharide to the glucose by ENGase catalyzed transglycosylation, which is introduced in the next chapter. They demonstrated the enzymatic syntheses of full-sized glycopeptides having various glycan structures including a complex sialyloligosaccharide, complex asialooligosaccharide, a high mannose type oligosaccharide and a diGlcNAc saccharide in high yield respectively from mono glucosylated glycopeptide prepared by NGT and UDP-Glc shown in the following figure.



Lomino, J. V.; Naegeli, A.; Orwenyo, J.; Amin, M. N.; Aebi, M.; Wang, L. X. *Bioorg. Med. Chem.* **2013**, *21*, 2262

Figure 1-13. Two-step enzymatic glycosylation of polypeptides using NGT and ENGase.

As shown in the above examples, the complete glycosylation of human-type glycans such as complex type oligosaccharides have not been developed *in vitro* experimental conditions yet. However direct enzymatic glycosylation of polypeptide having a consensus sequence might be a very efficient method for the homogeneous glycoprotein synthesis. Because this is a highly green synthetic method that does not require the number of multiple synthetic processes like other chemical methods, and by encoding the consensus sequence in the artificial RNA sequence, the protein obtained by the

E. coli expression can be arbitrarily converted into a glycoprotein. In the next chapter, I will discuss different types of enzyme-based glycoprotein synthesis methods as well.

1-7. Chemoenzymatic remodeling of glycoproteins

As mentioned in the *chapter 1-5*, the glycan structures of glycoproteins that can be isolated from natural or recombinant source are known to be heterogeneous. Thus, if a heterogeneous glycoprotein is remodeled into homogeneous glycoproteins *in vitro*, a large amount of glycoproteins having homogeneous glycans are obtained by natural source. To solve these problems, the concept of "glycan remodeling" was proposed (Figure 1-14).[35] Remodeling approach consists of two key reactions. First reaction is enzymatic trimming to remove the heterogeneous *N*-glycans from glycoproteins, leaving the innermost GlcNAc residue at the glycosylation site. Second required reaction is sugar chain elongation by glycosyltransferases or endoglycosidases.

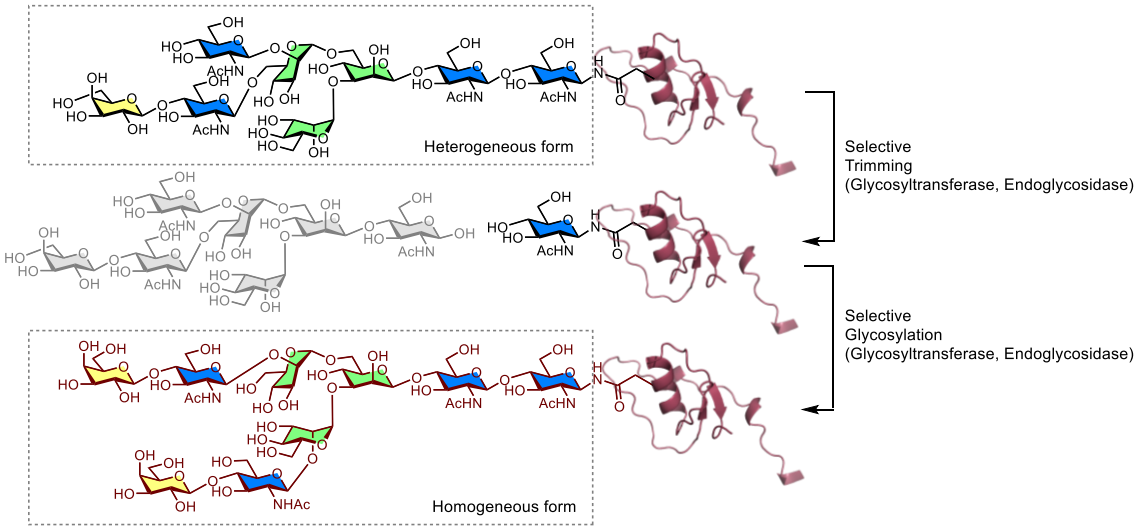
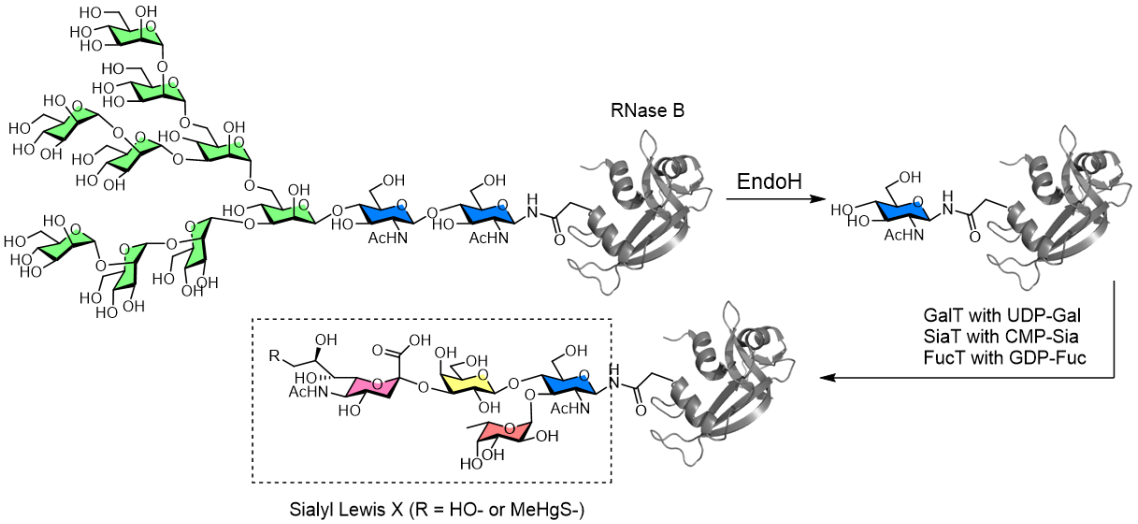


Figure 1-14. Synthetic scheme of glycan remodeling, which consists of two reactions, enzymatic trimming, and enzymatic glycosylation.

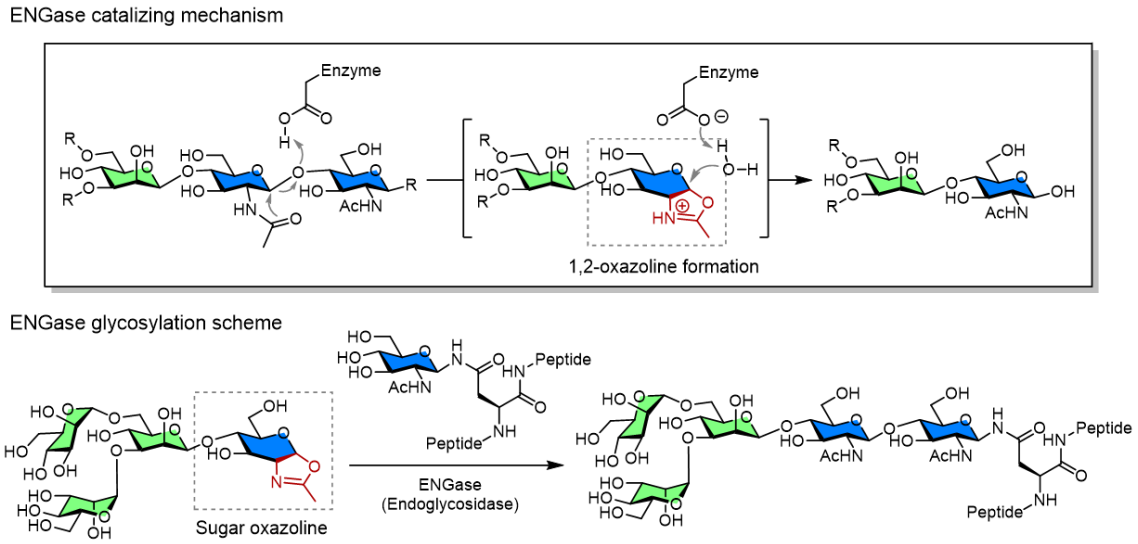
As an early example of glycan remodeling, Witte and coworkers transformed ribonuclease B (RNase B), which consists of 124 amino acids and has heterogeneous high mannose glycans at Asn34, to a single homogeneous glycoform having a sialyl Lewis X tetrasaccharide moiety (Figure 1-15).[36] High manse type oligosaccharides on RNase B were removed for leaving only the innermost GlcNAc to the site of glycosylation by endoglycosidase H (Endo H) to afford homogeneous form of GlcNAc-RNase. Then several glycosyltransferases including β-1,4-galactosyltransferase (GalT), α-1,3-fucosyltransferase V (FucT) and α-2,3-sialyltransferase (SialT) were used to build a sialyl Lewis X and mercury derivatives respectively. As the result they completed stepwise remodeling of RNase B enzymatically.



Witte, K.; Sears, P.; Martin, R.; Wong, C. H. *J. Am. Chem. Soc.* **1997**, *119*, 2114

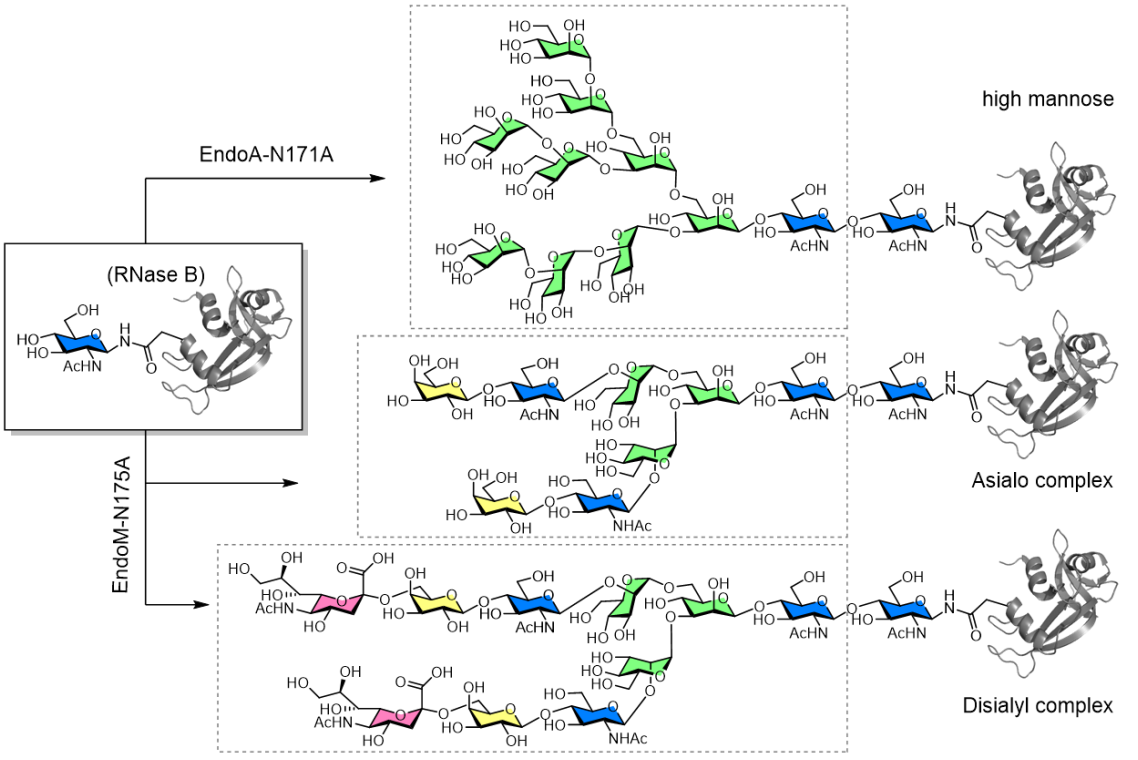
Figure 1-15. The chemoenzymatic synthesis of RNase B having Lewis X applying glycan remodeling.

In contrast to the stepwise synthesis of RNase B, the alternative remodeling approaches was subsequently developed. In the former procedure the mono glycan donors are elongated in a stepwise manner using specific transferases, however in the later approach the target oligosaccharide block is transferred in a single step in a convergent manner by using transglycosylation activity of the endo-β-N-acetylglucosaminidases (ENGase). The single step remodeling approach was enabled by two robust developments. One is the discovery of a glycan active donor, an 1,2-oxazoline motif of a target glycan can be transferred to a glycopeptide carrying mono GlcNAc by transglycosylation. This is devised by Fujita and coworkers.[37] This oxazoline structure is a mimic of the structure of the transition state form of sugar donors when endoglycosidase transfer glycans to acceptors (Figure 1-16).[38] Li and coworkers also studied the donor substrate scope.[39] Another development is glycosynthases based on endo-β-N-acetylglucosaminidases from *Arthrobacter protophormiae* (Endo-A) and *Mucor hiemalis* (Endo-M), which suppress side reactions such as hydrolysis of target glycans while taking the highly active sugar oxazolines as substrates for transglycosylation. Umekawa and coworkers also achieved the more robust mutated ENGases (EndoA-N171N, EndoM-N175A) are able to use for the extension of glycan structures such as sialyl glycan or disialyl glycan, allowing the synthesis of sialyl glycoproteins,[40] RNase B having high a mannose type oligosaccharide by EndoA-N171N, an asialo oligosaccharide and a sialyl saccharide by EndoM-N175A respectively (Figure 1-17).[41-42] In addition, they also succeeded in semisynthesis of complex type sialyloligosaccharyl oxazoline using 2-chloro-1,3-dimethylimidazolinium chloride (DMC)[43] developed by Noguchi and coworkers (Figure 1-18). [41, 44] Their first syntheses of glycoproteins accelerated the research related to glycan remodeling strategy. Many glycoproteins have been synthesized by ENGase derivatives including Endo-S/D. [35]



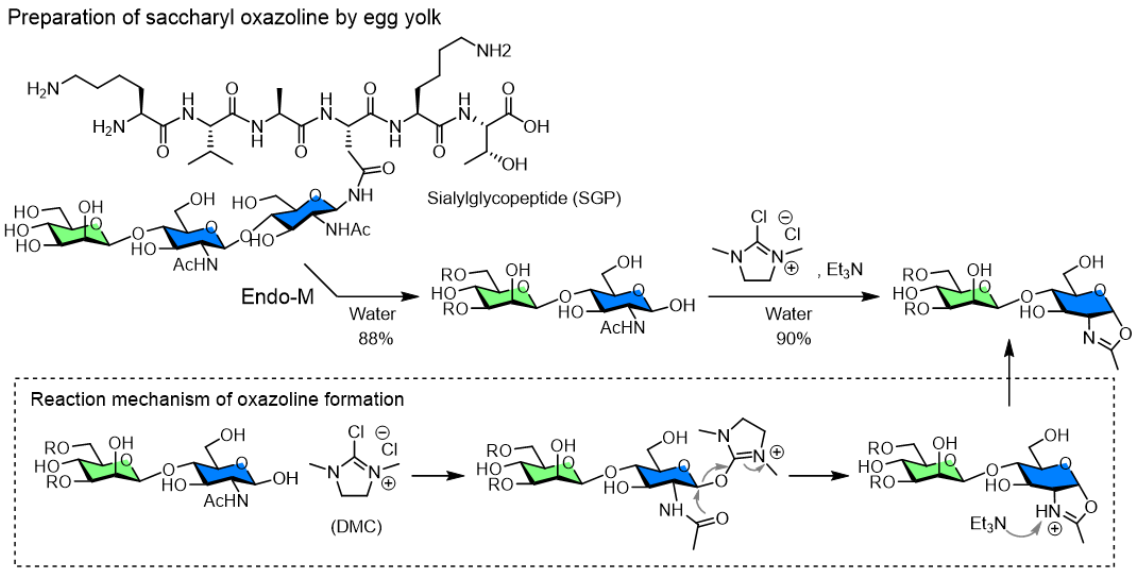
Fujita, M.; Shoda, S.; Haneda, K.; Inazu, T.; Takegawa, K.; Yamamoto, K. *Biochim. Biophys. Acta.* **2001**, *1528*, 9
Li, B.; Zeng, Y.; Hauser, S.; Song, H.; Wang, L. X. *J. Am. Chem. Soc.* **2005**, *127*, 9692
Wan, L. X. *Carbohydrate Research* **2008**, *343*, 1509

Figure 1-16. Glycosylation remodeling applying ENGase catalyzed glycosylation.



Huang, W.; Yang, Q.; Umekawa, M.; Yamamoto, K.; Wang, L. X. *ChemBioChem* **2010**, *11*, 1350
Huang, W.; Li, C.; Li, B.; Umekawa, M.; Yamamoto, K.; Zhang, X.; Wang, L. X. J. Am. Chem. Soc. **2008**, *131*, 2214

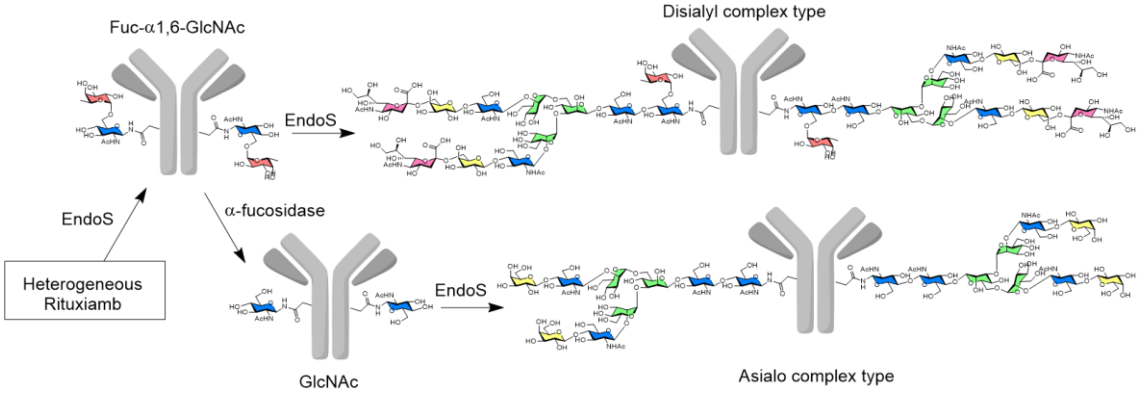
Figure 1-17. Chemoenzymatic synthesis of RNase B having various glycoform by ENGase family, high mannose, asialo and disialyl complex types.



Huang, W.; Yang, Q.; Umekawa, M.; Yamamoto, K.; Wang, L. X. *ChemBioChem* **2010**, *11*, 1350
Noguchi, M.; Tanaka, T.; Gyakushi, H.; Kobayashi, A.; Shoda, S. *J. Org. Chem.* **2001**, *74*, 2210

Figure 1-18. Preparation of oligosaccharyl oxazoline from egg yolk and oxazoline formation by DMC.

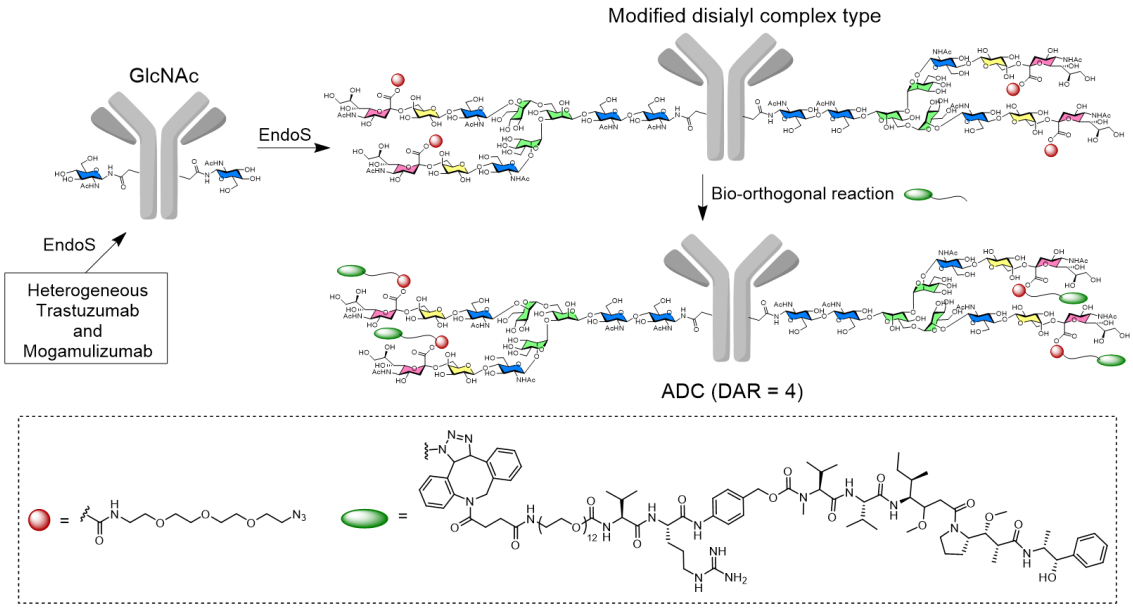
A series of glycoprotein remodeling using ENGases were utilized for glycoprotein synthesis and especially for the synthesis of IgG antibodies.[45-46] Furthermore, recently it was impressive that glycan remodeling was applied to antibody drug conjugates (ADCs).[47] Huang and coworkers reported an efficient chemoenzymatic method for site-selective glycoengineering of a Fc fragment of intact IgG antibody (Figure 1-19).[45] Due to their attempts, two types of glycosynthase mutants, EndoS-D233A and D233Q, were designed from Endo-S (an ENGase from *Streptococcus pyogenes*) and these mutants transferred glycans to the Fc deglycosylated IgG convergently. Consequently, rituximab (heterogeneous monoclonal antibody) was converted to homogeneous glycoforms, including a sialyl, a nonfucosylated complex type glycoform.



Huang, W.; Giddens, J.; Fan, S. Q.; Toonstra, C.; Wang, L. X. *J. Am. Chem. Soc.* **2012**, *134*, 12308

Figure 1-19. Glycan remodeling of Rituximab by Wang and coworkers.

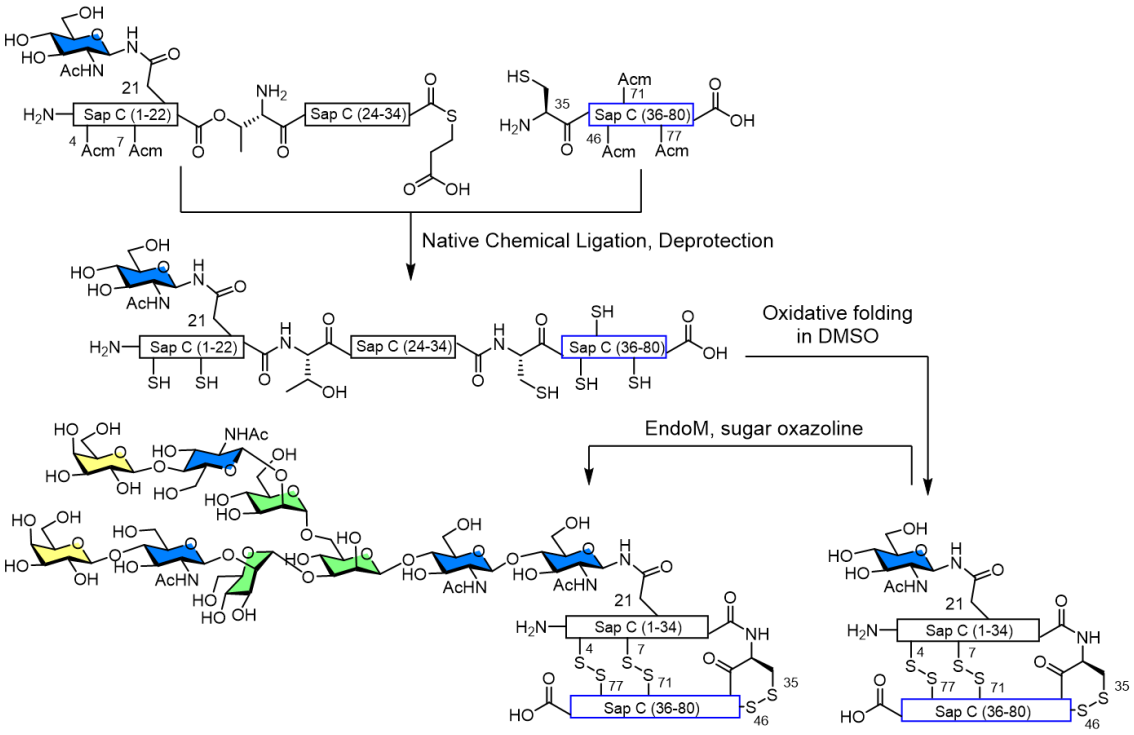
Recently Manabe and coworkers successfully applied Endo-S based glycan remodeling to antibody-drug conjugates (ADCs).[47] As a method, they transferred an oxazoline derivative having a complex type sialyl glycan consisting of a sialic acid and a linker with an azide group, to antibodies (at Asn297) for example trastuzumab (with fucose) and mogamulizumab (without fucose) utilizing Endo S (Figure 1-20). As a result, four azide groups were presented on the sugar chains of the modified monoclonal antibodies, and the desired drug molecules, a potent cytotoxic agent monomethyl auristatin E, were successfully conjugated to the antibodies by using azide click reaction site specifically. The prepared antibody-drug conjugate exhibited cytotoxicity against HER2-expressing cells. This study applies glycan remodeling to an innovative strategy using glycans as a new scaffold.



Manabe, S.; Yamaguchi, Y.; Matsumoto, K.; Fuchigami, H.; Kawase, T.; Hirose, K.; Mitani, A.; Sumiyoshi, W.; Kinoshita, T.; Abe, J.; Yasunaga, M.; Matsumura, Y.; Ito, Y. *Bioconjugate Chem.* **2019**, *30,* 1343

Figure 1-20. Application of glycan remodeling for the antibody-drug conjugate.

This chemoenzymatic glycan remodeling strategy can also be combined with NCL for the total chemical synthesis of glycoproteins. Hojo and coworkers succeeded in synthesizing the glycosyl saposin C, a hydrophobic lipid-binding protein (Figure 1-21).[48] Saposin C carrying mono GlcNAc was firstly synthesized by sequential NCL and the synthetic octasaccharide oxazoline was then transferred to this glycoprotein using the Endo-M N175Q, and the saposin C carrying the complex-type nonasaccharide was successfully obtained. Asahina also completed to the synthesis of an immunoglobulin domain of Tim-3 carrying a complex type *N*-Glycan.[49]



Hojo, H.;Tanaka, H.; Hagiwara, M.; Asahina, Y.; Ueki, A.; Katayama, H.; Nakahara, Y.; Yoneshige, A.; Matsuda, J.; Ito, Y.; Nakahara, Y. *J. Org. Chem.* **2012**, *77*, 9437

Figure 1-21. Chemoenzymatic synthesis of saposin C by utilizing ENGase glycan remodeling.

In addition to this synthesis of glycoproteins, there are various chemical methods to synthesize monosaccharide containing glycoproteins and then remodel the glycans. As shown in these examples, glycan remodeling has been used in a wide range of fields and has become a main strategy for glycoprotein synthesis along with chemical synthesis methods described in the next chapter.

1-8. Chemical synthesis of glycopeptides and fragment ligations

Other approaches include chemical and semisynthetic methods using native chemical ligation (NCL). Chemical methods employ both glycopeptides and peptides prepared by solid-phase peptide synthesis (SPPS) or sophisticated other protocols and those segments are sequentially coupled by using NCL (Figure 1-22).

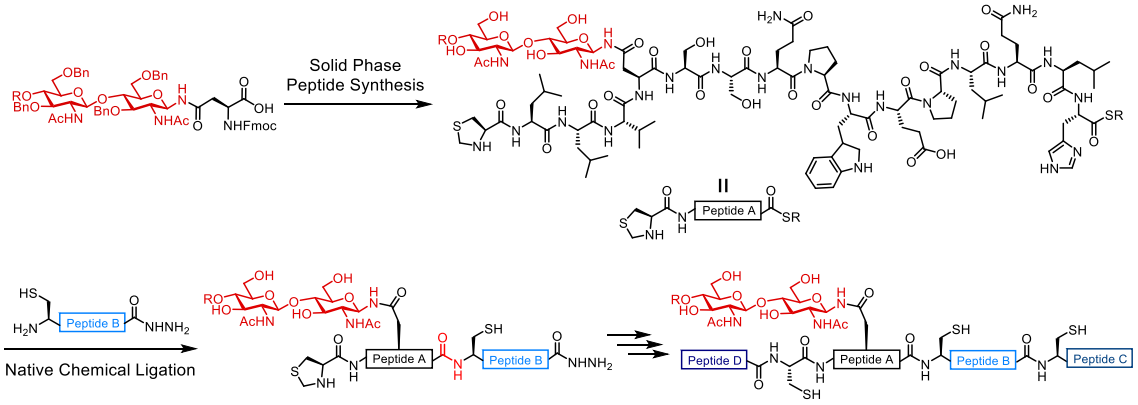
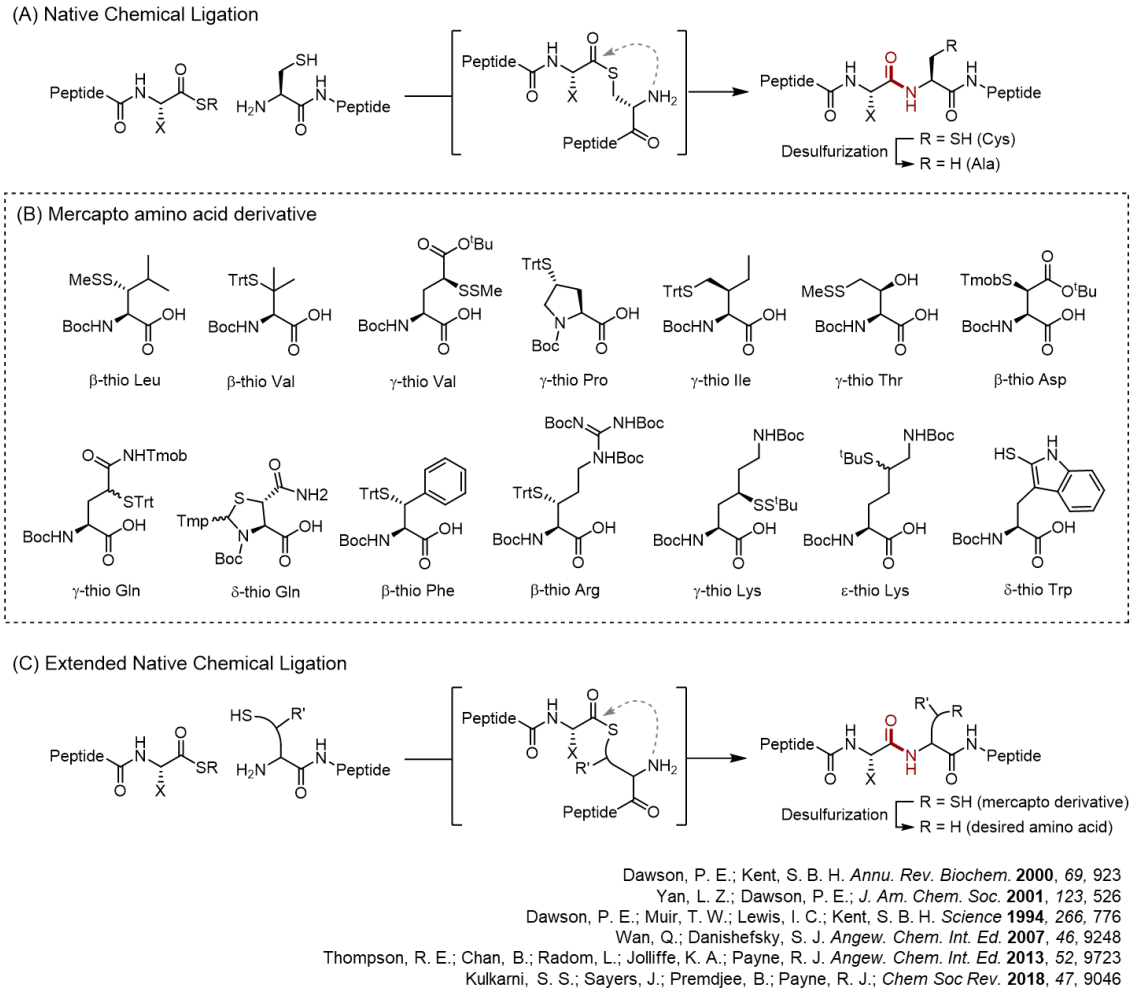


Figure 1-22. General scheme of the synthesis of glycoprotein.

Native Chemical Ligation (NCL) developed by Dawson and coworkers is indispensable for the chemical synthesis of glycoproteins (Figure 1-23 A).[50-52] NCL is a chemoselective condensation reaction between two partners, a peptide thioester and an N-terminal cysteinyl peptide fragment to afford a natural amide bond. In principle, NCL occurs through a two-step process including capture and intramolecular S-N acyl shift. The capture step relies on a unique chemoselective reaction, such as a thiol-thioester exchange, that brings two unprotected peptides together. This step induces a subsequent intramolecular S-N acyl rearrangement reaction, leading to the regioselective amide bond formation (Figure 1-23 A). The unique mechanism of NCL enables couplings between unprotected peptides that have potentially reactive functional groups such as hydroxyl, carboxyl and amino groups on their amino acid residues. However, using original NCL, the ligated peptides have Cys residue beside the ligated site. The Cys residue, which remains at the ligation site following the reaction, can be manipulated by hydrogenation or radical-based desulfurization to convert Cys residues to Ala, which were reported by Dawson and Danishefsky groups(Figure 1-23 C).[53-54] This method has been widely adopted, for example in the total chemical synthesis of several complex proteins and glycoproteins. Moreover, Payne groups achieved mutated Cys residue to other amino acids using β, γ, δ-mercapto amino acid derivatives (Figure 1-23 B).[55-56] After ligation, the thiol group is selectively removed by radical desulfurization to restore the respective native amino acid residues in the sequence (Figure 1-23 C). Applying synthetic β-mercapto amino acid derivatives to Cys residue of NCL, they demonstrated that desulfurization of this residue within ligation products was possible to convert to

any desired amino acids. Currently many groups developed the synthetic method of mercapto amino acid derivatives.[57] Almost all the mercapto amino acids were successfully synthesized in the sophisticated reactions (Figure 1-23 B). In this background theoretically any amino acid is usable for the ligation reactions. These NCL is a base of the synthesis of glycoproteins.



Dawson, P. E.; Kent, S. B. H. *Annu. Rev. Biochem.* **2000**, *69*, 923
Yan, L. Z.; Dawson, P. E.; *J. Am. Chem. Soc.* **2001**, *123*, 526
Dawson, P. E.; Muir, T. W.; Lewis, I. C.; Kent, S. B. H. *Science* **1994**, *266*, 776
Wan, Q.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2007**, *46*, 9248
Thompson, R. E.; Chan, B.; Radom, L.; Jolliffe, K. A.; Payne, R. J. *Angew. Chem. Int. Ed.* **2013**, *52*, 9723
Kulkarni, S. S.; Sayers, J.; Premdjee, B.; Payne, R. J.; *Chem Soc Rev.* **2018**, *47*, 9046

Figure 1-23. Native chemical ligation. (A) The mechanism of native chemical ligation. (B) Mercapto amino acid derivatives synthesized many groups. (C) Extended native chemical ligation with applying mercapto amino acid derivatives and the desulfurization protocol.

Glycopeptides can also be efficiently prepared by sophisticated convergent methods. The synthesis of glycopeptides has expanded with the establishment of SPPS methods (Figure 1-24).[58] In the SPPS, the C-terminus of an amino acid, whose N-terminus and side chains are protected, was firstly condensed on the surface of a resin. Next, only the N-terminal protecting group of the amino acids are deprotected. Then the resulting amino group is condensed with the C-terminus of the second amino acids, which have also protecting groups at N-terminal and sidechain, to afford protected

dipeptides. The same procedure can be repeated continuously to synthesize peptide chains with protected side chains. Finally, after deprotection of the side chains of peptides, the peptides can be cleaved from the resin to quickly obtain the desired peptide chain. There are mainly two types in SPPS protocol, Boc method and Fmoc method. Boc method uses a tert-butoxycarbonyl (Boc) group as the N-terminal protecting group, which is deprotected with an acid such as TFA or sulfuric acid. On the other hand, the Fmoc method uses 9-fluorenylmethoxycarbonyl (Fmoc) group as the N-terminal protecting group, and each deprotection is performed with piperidine. With these methods, polypeptide chains (~20aa) can be synthesized rapidly. Applying SPPS, many groups have synthesized bioactive glycopeptides, which is termed "cassette based" strategy.[20, 59] The details of such solid-phase peptide synthesis of glycopeptides will be described in detail in the next chapter about the synthesis of glycoproteins. In this chapter, the conventional and recent synthetic strategies of glycopeptides are focused.

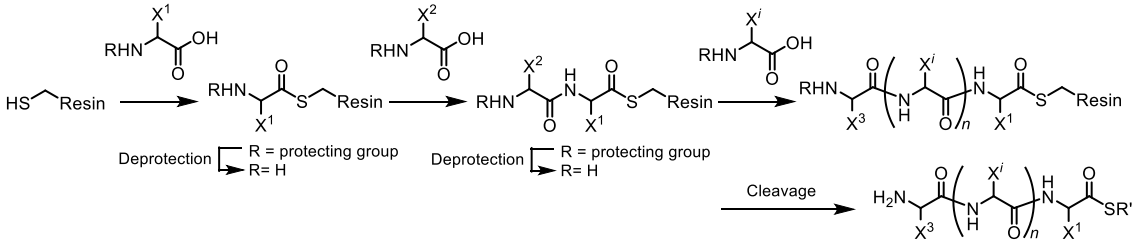
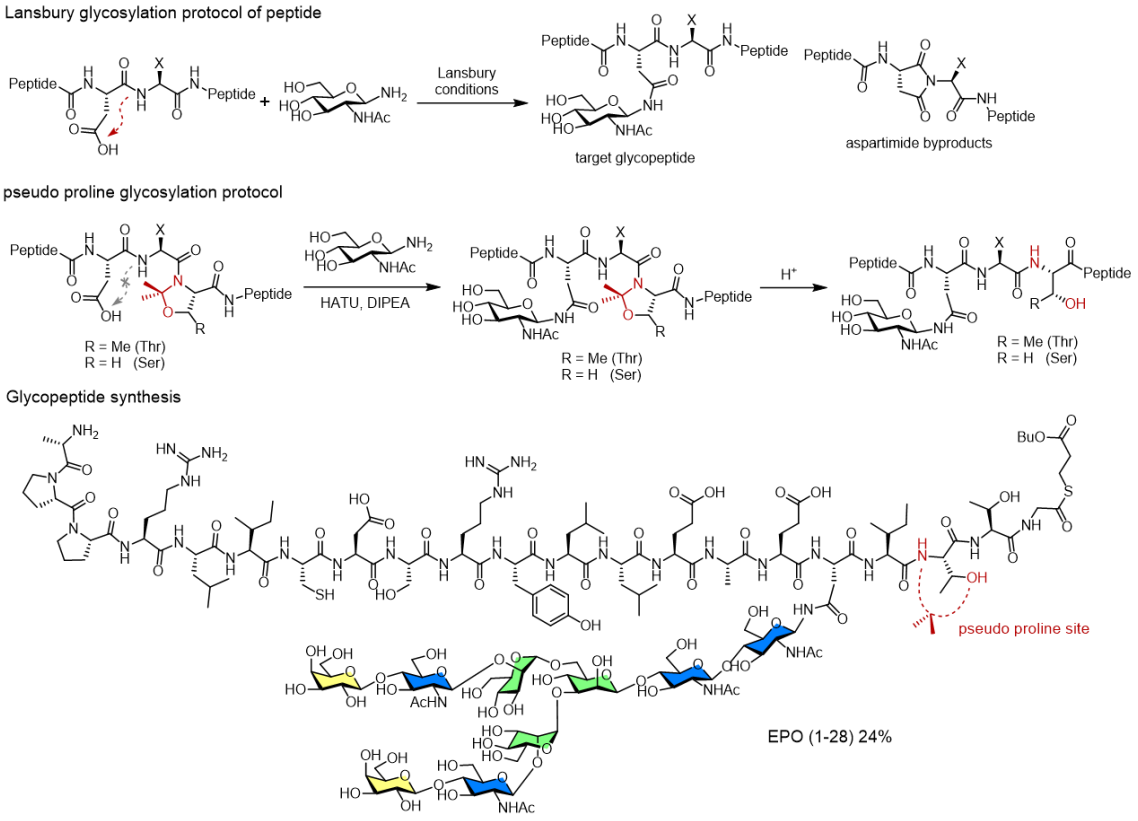


Figure 1-24. General scheme of solid phase peptide synthesis (SPPS). The peptide elongation was progressed by multistep deprotection and condensation on the resin.

However, when the carbohydrate component of glycopeptide is complex, the overall efficiency of this linear SPPS strategy is significantly compromised by low reaction yields obtained during the glycosyl amino acid coupling step and the subsequent elongation of the next amino acid. Moreover, any sialic acid motifs in the glycan must be protected during SPPS. Glycopeptides of up to 20 amino acids in length may be generated through this method.

As second generation of the synthesis of glycopeptide is a convergent strategy, and some groups developed unique synthetic methods. The Unverzagt, and the Danishefsky groups employed the convergent aspartylation approach pioneered by Lansbury and co-workers, which is termed "Lansbury aspartylation" strategy.[60-61] In this protocol, a protected peptide, bearing the free aspartyl residue, is merged with the glycosyl amine to generate a glycopeptide segment by using condensation reagents. However, this convergent method has a drawback that is the formation of aspartimide. Then they found that placement of a dimethyl group of pseudoproline motif, derived from Ser or Thr, at the (n+2) position dramatically suppress the undesired aspartimide formation at the n position (Figure 1-25).[62-63] As the result the pseudoproline-assisted Lansbury aspartylation was then applied to the

erythropoietin (EPO) 1-28 sequence by Ullmann and coworkers (Figure 1-25).[62] Using PyBOP glycosylamine was coupled to peptide having free aspartic group, then the glycopeptide was cleaved from resin, thioesterification and deprotection was carried out to afford EPO 1-28 in a yield of 37% with no detectable aspartimide.



Ullmann, V.; Radisch, M.; Boos, I.; Freund, J.; Pçhner, C.; Schwarzinger, S.; Unverzagt, C. *Angew. Chem. Int. Ed.* **2012**, *51*, 11566

Wang, P.; Aussedat, B.; Vohra, Y.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* ***2012***, *51*, 11571

Figure 1-25. Pseudoproline assisted Lansbury aspartylation and the application for the glycopeptide (EPO 1-28) synthesis.

Then, in addition to the direct amide bond formation reaction using the pseudoproline motif, more moderate N-glycosylation reactions using thioesters and thioacids are introduced here (Figure 1-26). At first Wang and coworkers developed a method for the joining of peptides to a complex type oligosaccharide *via* a *N*-linkage by using thioacid moiety (Figure 1-26 A).[64] The ω-aspartylation was conducted by coupling fully deprotected glycosylamine with a peptide including thioacid at the ω-aspartate carboxylic acid. In the presence of HOBT and mild oxidant, complex components are combined in encouraging yields to produce structurally and defined *N*-linked glycopeptides through an intermediate, active peptide ester having HOBt. They progressed to synthesize more than 10 glycopeptides and average yield was about 40%. However, since the key to this protocol is combining

sugar chains to peptides without the use of conventional condensation reagents, it was reported as a revolutionary method at the time.

In the same manner, Joseph and coworkers attempted the activation of thioacids by Cu(II) for the selective amide bond formation reaction and developed chemoselective Cu(II)-HOBt promoted chemical ligation of glycosylamines and peptide thioacids to give N-glycosylated peptides (Figure 1-26 B).[65] As will be discussed in the next chapter, thioacids can be activated by metal ions, especially silver and copper ions. They took advantage of thioacid Cu-mediated activation to selectively activate thioacids in the peptide side chain and convert them to HOBt-active esters by adding HOBt to attack the amino group. They progressed to synthesize more than 6 glycopeptides including dodecasaccharide and average yield was about 60%.

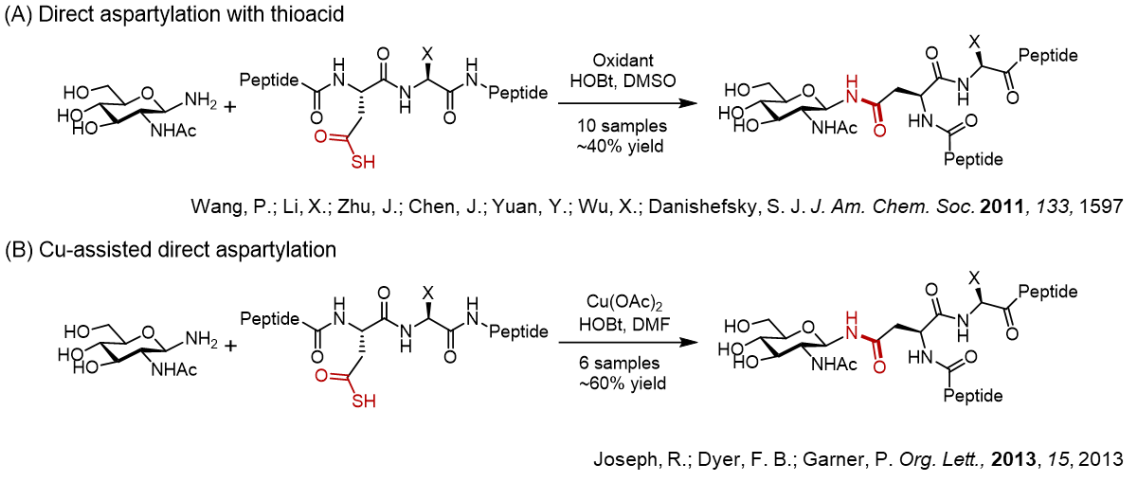


Figure 1-26. Direct aspartylation of glycosylamine with thioacid. (A) Direct aspartylation with thioacids using HOBt. (B) Direct Cu-assisted aspartylation with thioacids.

In contrast to thioacid, thioester is used for direct glycosylation of N-linked glycopeptide (Figure 1-27). Du and coworkers developed the efficient N-linked glycosylation reaction between glycosylamines and *p*-nitrophenyl thioester peptides (Figure 1-27 A).[66] The reaction conditions are mild and compatible with the C-terminal free carboxylic acid group and the unprotected N-linked sialyloligosaccharide. The reaction mechanism is a nucleophilic attack of the N-terminal amino group on the thioester of peptides under base conditions. By means of this convergent strategy, 12 species of N-glycopeptide fragments containing an N-terminal thiazolidine and a C-terminal thioester was synthesized, which is available for the synthesis of long glycopeptides and glycoproteins using the protocol of native chemical ligation. The yield of this direct amination is about 80%.

Recently Schowe and coworkers reported combined protocol. They described a general and robust method for the incorporation of aspartates with a thioacid side chain into peptides applying pseudoproline assisted N-glycosylation (Figure 1-27 B).[67] peptides having thioacid at Asp (n) and pseudoproline at Ser/Thr (n+2) were readily converted to complex N-glycopeptides by using a fast and chemoselective deprotection/ligation procedure. Their method allows for chemoselective aspartylation of glycosylamine with using thioacids, while at the same time inhibiting aspartimidation with the effect of pseudoproline motif. Using this combined method, they succeeded in synthesizing seven glycopeptides in average 50% yield, which is a clear improvement over the simple thioacid method without pseudoproline developed by Wang and coworkers.[63]

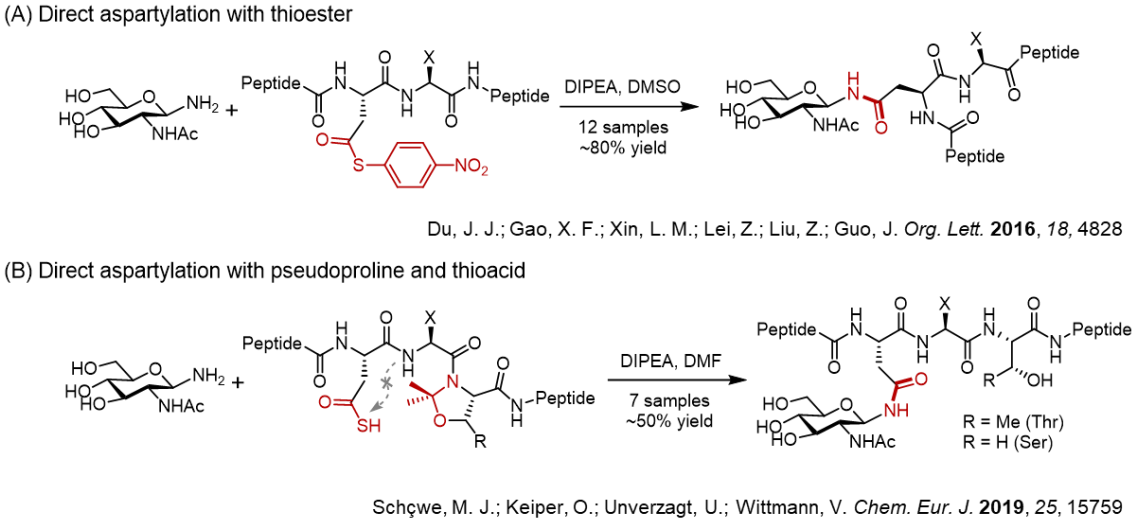
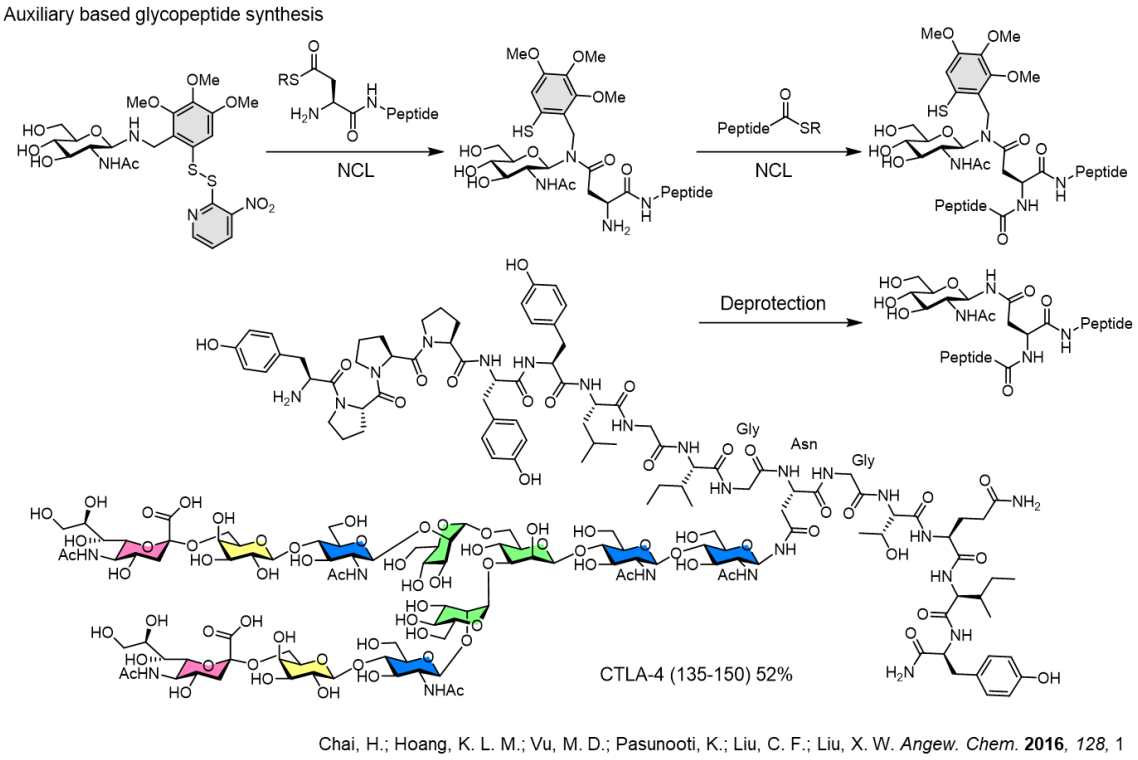


Figure 1-27. Direct aspartylation developed by using thioester and thioacid. (A) Direct aspartylation with thioesters. (B) Direct aspartylation with pseudoproline and thioacids.

In contrast to the second generation of glycopeptide synthesis based on direct amidation reactions, a completely novel approach has been developed by Chai and coworkers (Figure 1-28).[68] They studied a practical approach towards N-glycopeptide synthesis using an auxiliary-mediated dual native chemical ligation (NCL). The first NCL connects an N-linked glycosyl auxiliary to the thioester side chain of an N-terminal aspartate glycopeptide. This intermediate undergoes a second NCL with a C-terminal thioester glycopeptide. Mild cleavage provides the desired N-glycopeptide. They firstly studied the scope of NCL with modified glycosyl amine (Figure 1-28). Trimethoxybenzyl auxiliary was found to be the most proficient in the ligation and in its removal. The 3-nitro-2-pyridinesulfenyl group is well-known for rapid disulfide exchange under NCL condition. This scaffold is therefore capable of performing ligation to produce the glycan-linked N-terminal asparagine glycopeptide, followed by a second ligation with C-terminal thioester glycopeptide to provide the glycopeptide.

Subsequently, the auxiliary is readily cleaved following treatment with TFA and deacetylation to furnish glycopeptide. As the result they attempted to adopt auxiliary mediated method into the synthesis of a homogeneous glycopeptide, cytotoxic T-lymphocyte-associated antigen-4 (CTLA-4) fragment (135-150). Finally, they obtained glycopeptide CTLA-4 (135-150) segment in 52% overall conversion after purification.



Chai, H.; Hoang, K. L. M.; Vu, M. D.; Pasunooti, K.; Liu, C. F.; Liu, X. W. *Angew. Chem.* **2016**, *128*, 1

Figure 1-28. Auxiliary-mediated dual native chemical ligation (NCL) and the total chemical synthesis of CTLA-4 (135-150) segment.

Maki and coworkers also developed liquid phase glycopeptide synthesis.[69] In this method, partially protected polypeptide was cleaved, and the N-terminus of the protected peptides were condensed with α carboxy acid of the glycosyl asparagine with a triantennary glycan using a condensing reagent in the liquid phase (Figure 1-29). The N-terminal Fmoc group was then deprotected by piperidine and condensed with the protected peptide in the liquid phase using a condensing reagent again. The target glycopeptide was finally deprotected by TFA cocktail to afford the desired glycopeptide with a triantennary oligosaccharide (Figure 1-29). They further succeeded in synthesizing erythropoietin with a complex type triantennary sialyloligosaccharide. The sialic acid was introduced byα-2,6-sialyltransferase. Since this reaction is carried out in the liquid phase, the condensation rate of peptide and glycosyl asparagine was higher than that of SPPS. And since the condensation can be carried out at a low temperature, the rate of epimerization can be reduced.

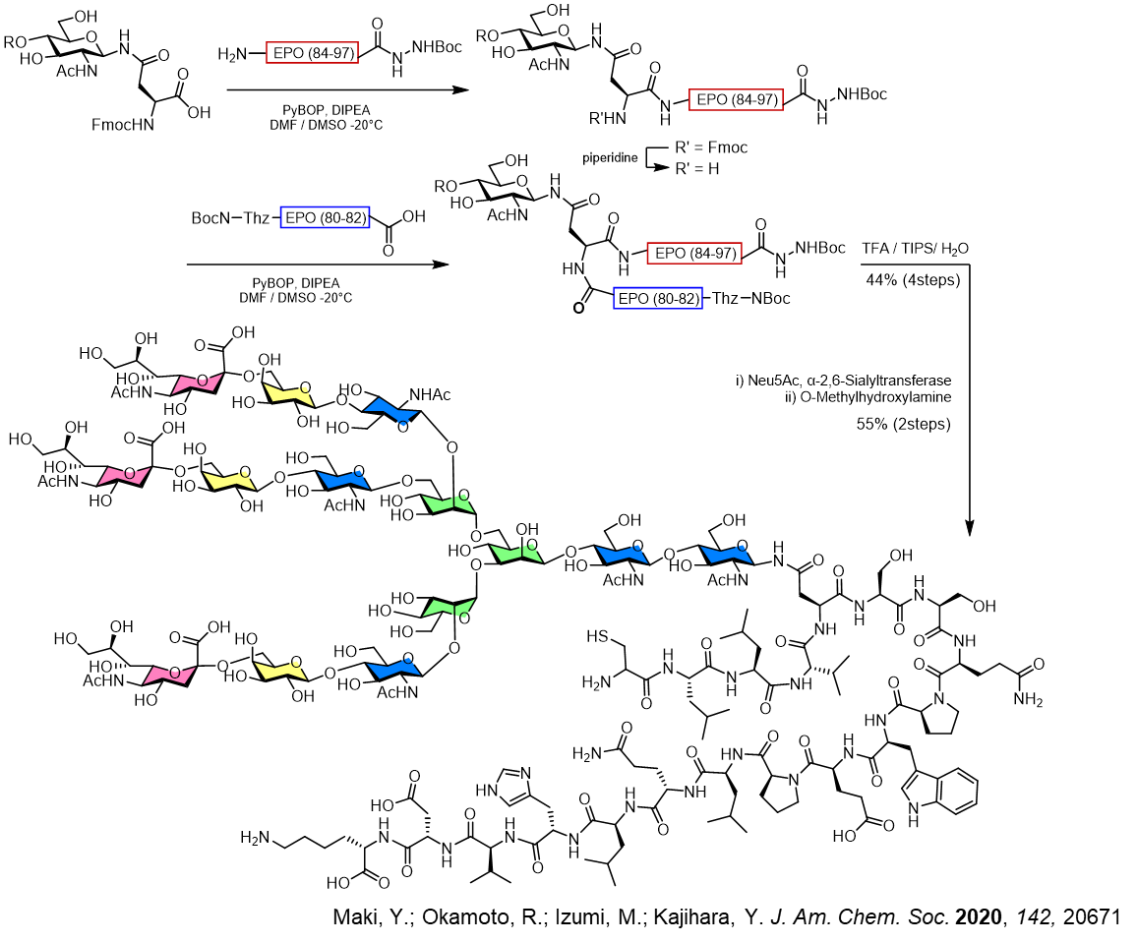
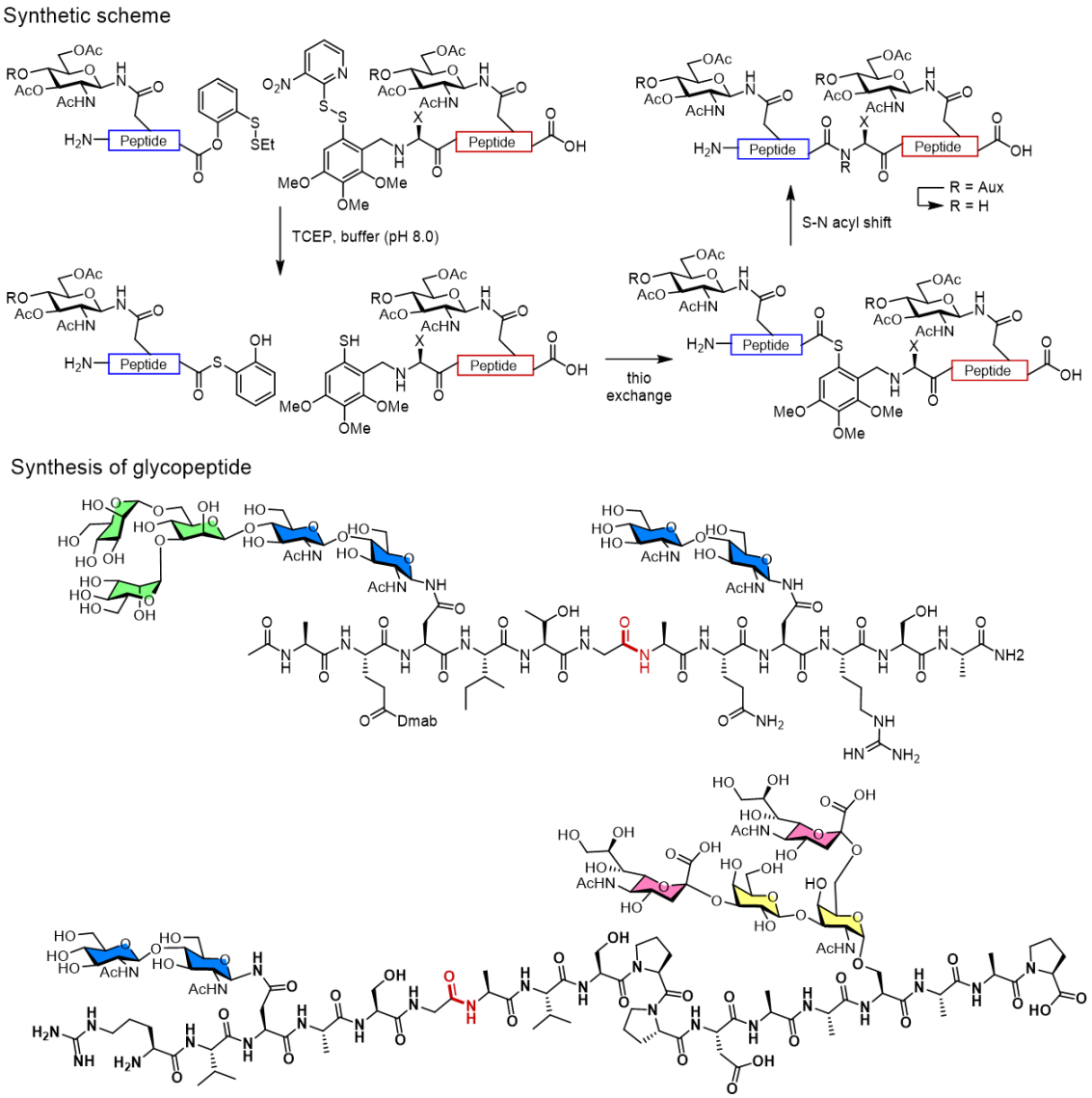


Figure 1-29. Glycopeptide synthesis having triantennary complex saccharide using a condensing reagent in the liquid phase.

Next, this section introduces how the resulting glycopeptide fragments can be combined with other peptides. The fundamental route for this approach is NCL.[52] As mentioned earlier in this chapter, NCL allows chemoselective coupling between peptide thioesters and cysteinyl peptides without any protecting group. By repeating this sequential NCL, in principle, any glycoprotein can be synthesized. However, the percentage of cysteine in various glycoproteins is extremely low (~1.7%).[50] Therefore, if the sequence of the desired protein does not contain cysteine, it is highly difficult to synthesize. The following three methods have been developed to solve this problem. The first is the synthesis using an auxiliary group, the second is a sugar-assisted ligation method, and the third is a ligation using β- or γ- mercapto amino acids. First, the synthesis method of glycopeptide using auxiliary groups is described.

Wu and coworkers achieved the synthesis of glycopeptide, which was a component of erythropoietin, using auxiliary groups (Figure 1-30).[70] They prepared an auxiliary bearing

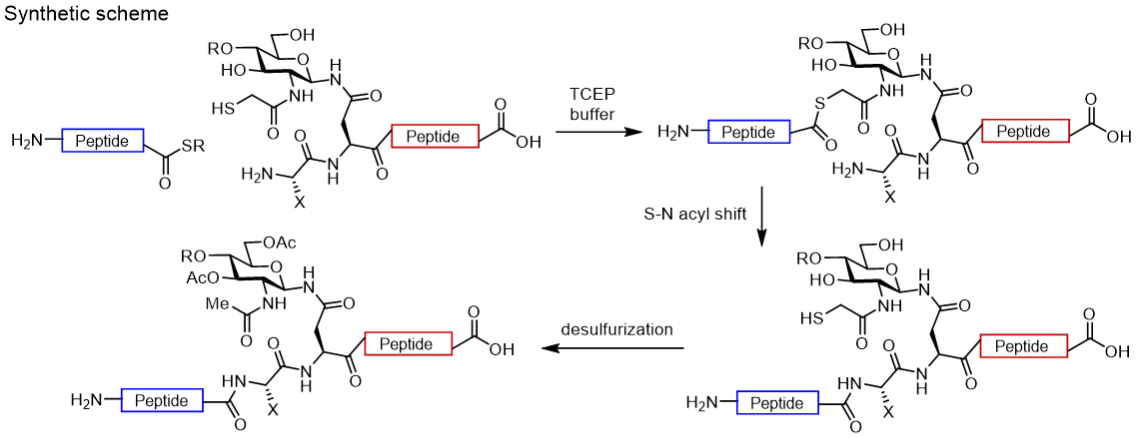
glycopeptides and C-terminal phenolic ester peptides, which can be converted to active thioester in the reducing conditions and examined a ligation. As expected, upon exposure to TCEP in sodium phosphate buffer solution (pH 8.0), glycopeptides underwent reductive disulfide cleavage. Presumably, the C-terminal glycopeptide then suffered an O-S acyl shift to generate the activated thioester. Both peptides were temporarily joined through a thioester exchange reaction. Following intramolecular acyl shift, the fully glycopeptides including O-glycan and N-glycan were isolated in higher yields. This auxiliary based method is a great achievement, showing that glycopeptides can be synthesized without any cysteine residues (Figure 1-30).



Wu, B.; Chen, J.; Warren, J. D.; Chen, G.; Hua, Z.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2006**, *45*, 4116

Figure 1-30. Auxiliary based glycopeptide synthesis. Using auxiliary based method glycopeptides having various saccharides were synthesized without any cysteine residues.

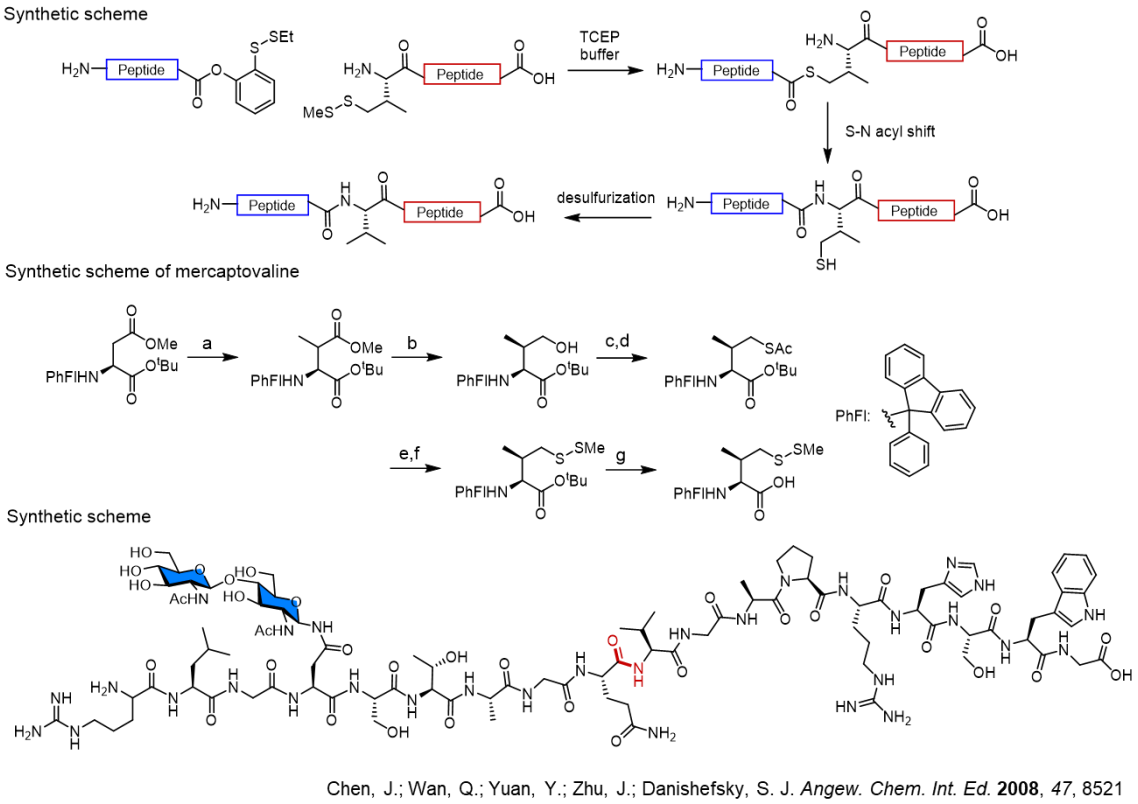
Second ligation method for the synthesis of glycopeptide without using cysteine residue is sugar assisted ligation developed by Brik and coworkers.[71] They reported that the sugar moiety of a glycopeptide could mimic the thiol capture function at the ligation site once it is modified with a thiol at the C-2 position. Moreover, the thiol groups can be reduced to the acetamide moiety, furnishing the unmodified glycopeptide (Figure 1-31). Based on NCL, the thioester exchange proceeds between the peptide thioesters and the peptides having sugar motif. The desired glycopeptide is obtained by the SN acyl shift occurred in the thioester intermediates. Since this glycopeptide has an extra thiol group in C-2 position, the desired C-2 N-acetyl glycopeptide can be obtained by desulfurization reaction. Although this strategy can be applied only to the synthesis of glycopeptides with GlcNAc or GalNAc sugar motif, it is an important reaction that allows coupling without using cysteine residues.



Brik, A.; Yang, Y. Y.; Ficht, S.; Wong, C. H. *J. Am. Chem. Soc.* **2006**, *128*, 5626

Figure 1-31. Sugar assisted ligation to the synthesis of glycopeptide based on NCL.

The third methodology for the synthesis of glycopeptide without cysteine residues is a ligation with mercapto amino acid derivatives. As mentioned in former chapter, introduction of β-, γ- or δ-mercapto amino acids to the N-terminus of a peptide allowed for a ligation with a peptide-thioester under the standard NCL conditions to afford amide bonds in any sites. Using this expansion method of NCL, Chen and coworkers achieved glycopeptide synthesis using γ-mercaptovaline (Figure 1-32).[72] In their synthesis, γ-hydroxy valine was obtained from *L*-aspartic acid. They modified the route thereby enabling the preparation of a γ-thiol valine derivative. Methylation and regioselective reduction furnished alcohols as a mixture of diastereomers, which were readily separated by chromatography. After mesylation followed by the treatment with the thioacetic acid in the presence of DBU, acetylated thiol was added. The latter process gave the target compound, in a straightforward manner (Figure 1-32). Synthesized γ-mercaptovaline was introduced into peptide by SPPS. They prepared C-terminal phenolic ester peptides as a counter peptide. NCL with mercaptovaline progressed within 3 hours and afforded glycopeptide in 89% yield (Figure 1-32).



Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2008**, *47*, 8521

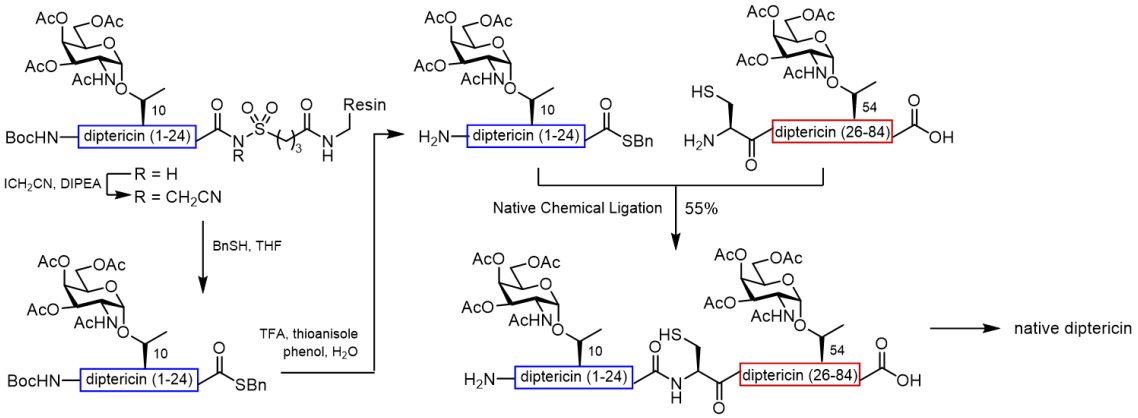
Figure 1-32. glycopeptide synthesis using γ-mercaptovaline

reaction conditions; a) KHMDS, MeI, THF, -78˚C, 3 h, 97%; b) DIBAL-H, THF, -35˚C, 1 h, 83%; c) MsCl, Et3N, CH_2Cl_2, 0˚C, 1 h; d) AcSH, DBU, DMF, RT, 16 h, 73% in 2 steps; e) 1M NaOH, MeOH, 0˚C, 10 min; f) MMTS, Et_3N, CH_2Cl_2, RT, 30 min, 86% in 2 steps; g) HCl in EtOAc, RT, 82%

In addition to these examples, NCL strategy has been further expanded by the application of the ligation-desulfurization strategy and applied for many syntheses of glycopeptides. Efficient syntheses are reviewed by both the Payne and the Kajihara groups.[6, 73]

1-9. Total chemical synthesis and semisynthesis

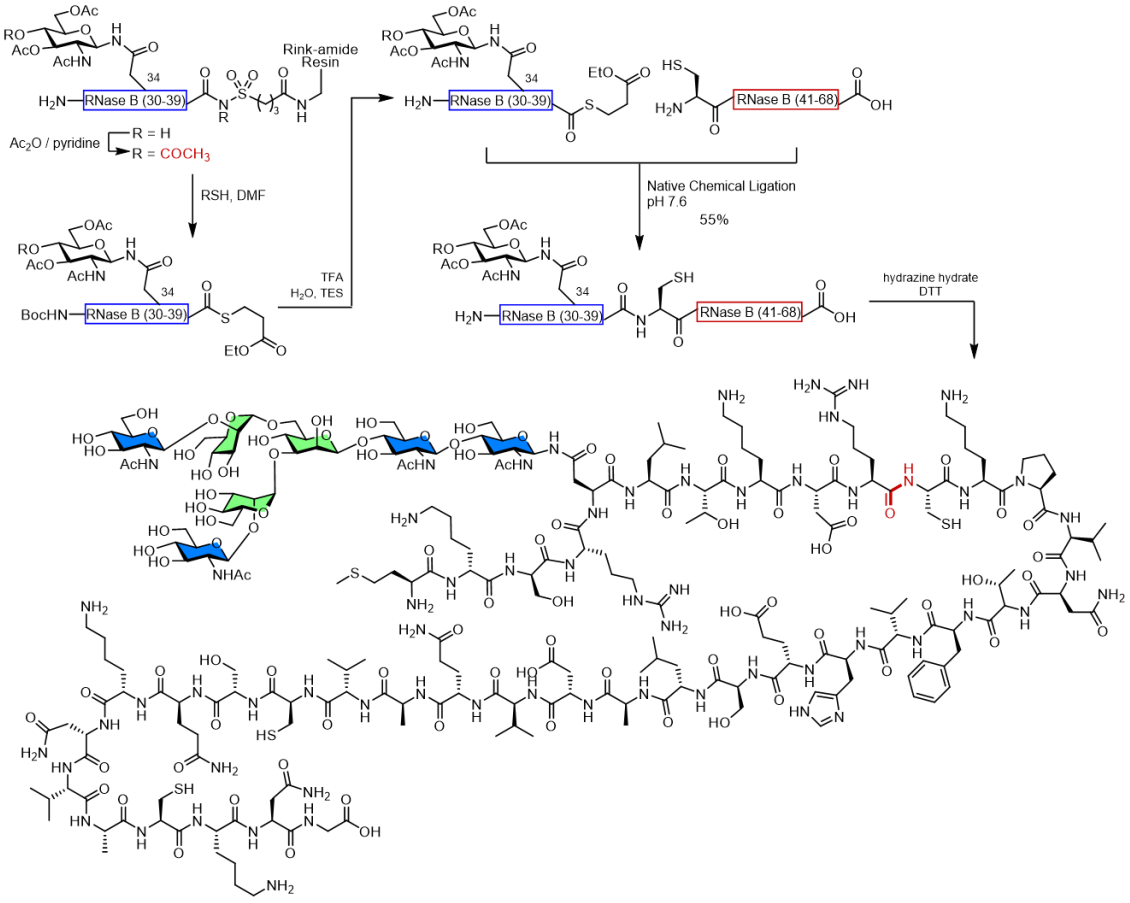
The earliest synthetic examples of chemical synthesis of glycoproteins were achieved by Shin and coworkers.[74] Applying NCL they succeeded total chemical synthesis of a glycoprotein diptericin (82 amino acids) having two triacetyl α-D-GalNAc residues (Figure 1-33). Because the original sequence of diptericin did not contain Cys residues, a G25C mutation was introduced to enable a retrosynthetic disconnection at the site. In addition to the synthesis of diptericin, they developed an alkane sulfonamide "safety-catch" linker that enabled the synthesis of glycopeptide thioester via the Fmoc-based SPPS protocol. In this "safety-catch" linker method, peptides are extended on a sulfonamide-linked resin, the sulfonamide group is alkylated to activate the C-terminus, and finally a thiol is added to obtain a peptide thioester (Figure 1-33). The conventional SPPS method of assembling peptides as thioesters is difficult to apply to the Fmoc SPPS using nucleophile such as piperidine. Therefore, the "safety catch" linker method, which can synthesize peptide thioesters by Fmoc SPPS, has become a frequently used method by many groups.



Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121*, 11685

Figure 1-33. Safety-catch linker method and the synthesis of diptericin glycoform.

Mezzato and coworkers achieved the first synthesis of a glycopeptide thioester having a complex type N-glycan applying modified "safety-catch" linker strategy (Figure 1-34).[75] They have established a general and efficient SPPS of complex glycopeptide thioesters. They also developed the synthetic method of glycopeptides in higher yields by a unique thioesterification method based on the "safety catch" linker method developed by Bertozzi and coworkers.[74] It was found that activation of glycosyl-asparagine in situ gives the highest coupling yields and that the OH groups can be acetylated without activating the safety-catch linker. Based on their preparation method of glycopeptide thioester, they succeeded to synthesize glycosylated RNase B (30-68) with heptasaccharides (Figure 1-34).

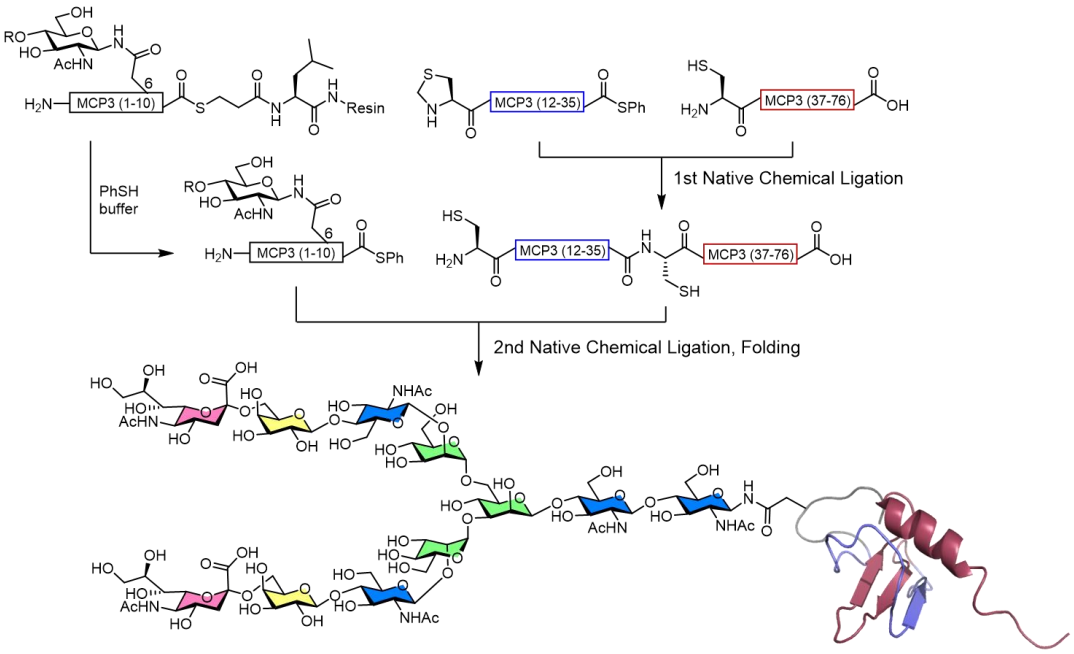


Mezzato, S.; Schaffrath, M.; Unverzagt, U. *Angew. Chem. Int. Ed.* **2005**, *44*, 1650

Figure 1-34. The first total chemical synthesis of glycopeptide, glycosylated RNase B (30-68) with heptasaccharides applying "safety-catch" linker.

Next Yamamoto and coworkers achieved the first total chemical synthesis of a glycoprotein. They achieved the synthesis of chemokine monocyte chemotactic protein-3 (MCP-3), a CC-chemokine that consists of 76-amino acid and a complex type N-glycan (Figure 1-35).[76] Using two sequential NCL of three unprotected peptides segments they synthesized the native glycoprotein MCP-3. Importantly, they applied "cassette based" approach and the synthesis required the development of methods for the generation of sialylglycopeptide thioesters. For the sialylglycopeptide thioester segment, they examined and successfully implemented approaches using Fmoc SPPS protocols and Boc SPPS protocols (Figure 1-35). In case of the Fmoc protocols, they applied condensation phenyl thiol to C-terminus of peptides directly. In case of the Boc approach utilized minimal side chain protection and direct thiolysis of the resin bound peptide. Using these strategies, they successfully ligated chemically synthesized peptide segments by NCL and examined oxidative folding conditions to the linear glycopeptides to afford a glycoprotein having a homogeneous complex-type sialyloligosaccharide and two disulfide bonds. Finally, they confirmed the positions of disulfide bonds

using enzymatic digestion. CD spectrum, MS spectrum and simple ELISA assay indicated that they completed to synthesize desired glycosylated MCP-3. Their attempt to synthesize homogeneous glycoproteins by chemical synthesis has been leading many groups to enter the race of glycoprotein synthesis still now.

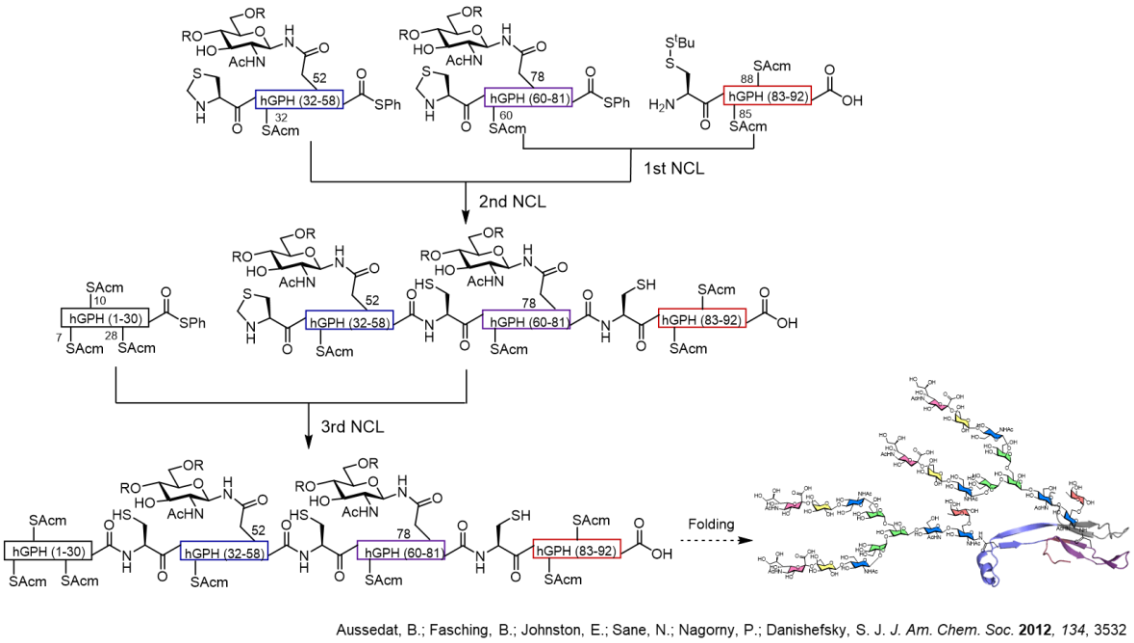


Yamamoto, N.; Tanabe, Y.; Okamoto, R.; Dawson, P. E.; Kajihara, Y. *J. Am. Chem. Soc.* **2008**, *130*, 501

Figure 1-35. Total synthesis of MCP-3 having biantennary sialyloligosaccharides.

Followed by the first synthesis of glycoproteins, MCP-3, many groups succeeded in the chemical synthesis of glycoprotein. For example, Aussedat and coworkers reported the chemical synthesis of protected follicle-stimulating hormone (hFSH), a member of the family of specific family of heterodimeric glycoproteins: the human glycoprotein hormones (hGPH) (Figure 1-36).[77] Actually, these hormone proteins consist of two non-covalently linked subunits a common α-subunit and a hormone-specific β-subunit. This time they only synthesized protected α-subunit of hFSH (92 amino acids), which is involved in the regulation and maintenance of essential reproductive processes, such as gametogenesis, follicular development, and ovulation. They prepared complex type dodecasaccharide amine having α1,6 fucosylation and α2,6 sialylation and applied "Lansbury aspartylation" strategy to synthesize glycopeptide segments. To synthesize glycopeptides, they first used HATU and DIPEA for disaccharide chitobiose glycosylamines and peptides having free carboxylic acid in the side chain. Instead, the major product resulted to an unreactive aspartimide peptide and the yield of followed kinetically controlled ligation decreased due to side reactions. To solve these problems, they revised synthetic route for several times and applied pseudoproline strategy to afford desired glycopeptide having chitobiose instead of dodecasaccharide. Using resulting strategy,

glycopeptide with dodecasaccharide was synthesized in 26% by Lansbury reaction. Consequently, followed by the following figure the protected α-hGPH having complex sialylated and fucosylated dodecasaccharide was completed to synthesize. However, they could not achieve folding of the linear α-hGPH and examine bioassay (Figure 1-36).

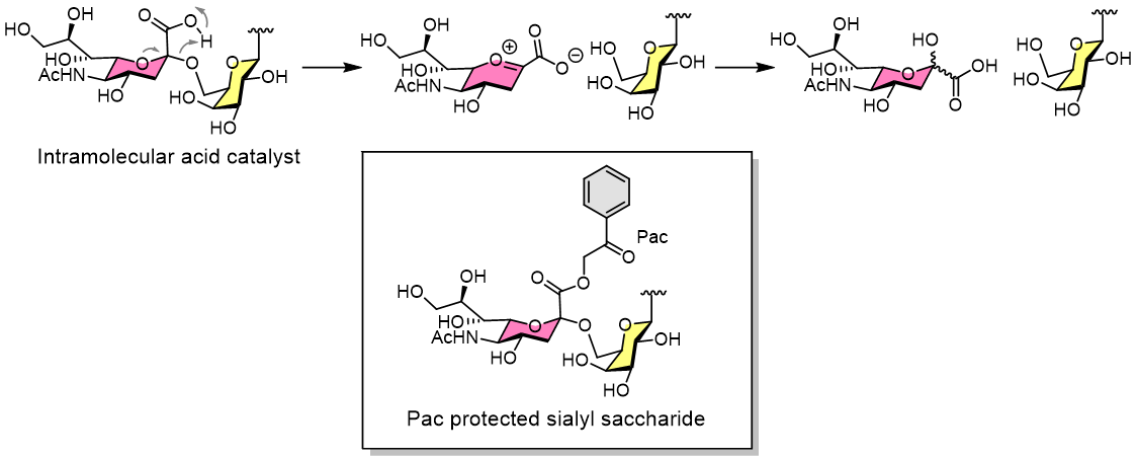


Aussedat, B.; Fasching, B.; Johnston, E.; Sane, N.; Nagorny, P.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2012**, *134*, 3532

Figure 1-36. Protected α-hGPH having complex type oligosaccharide synthesized by Danishefsky.

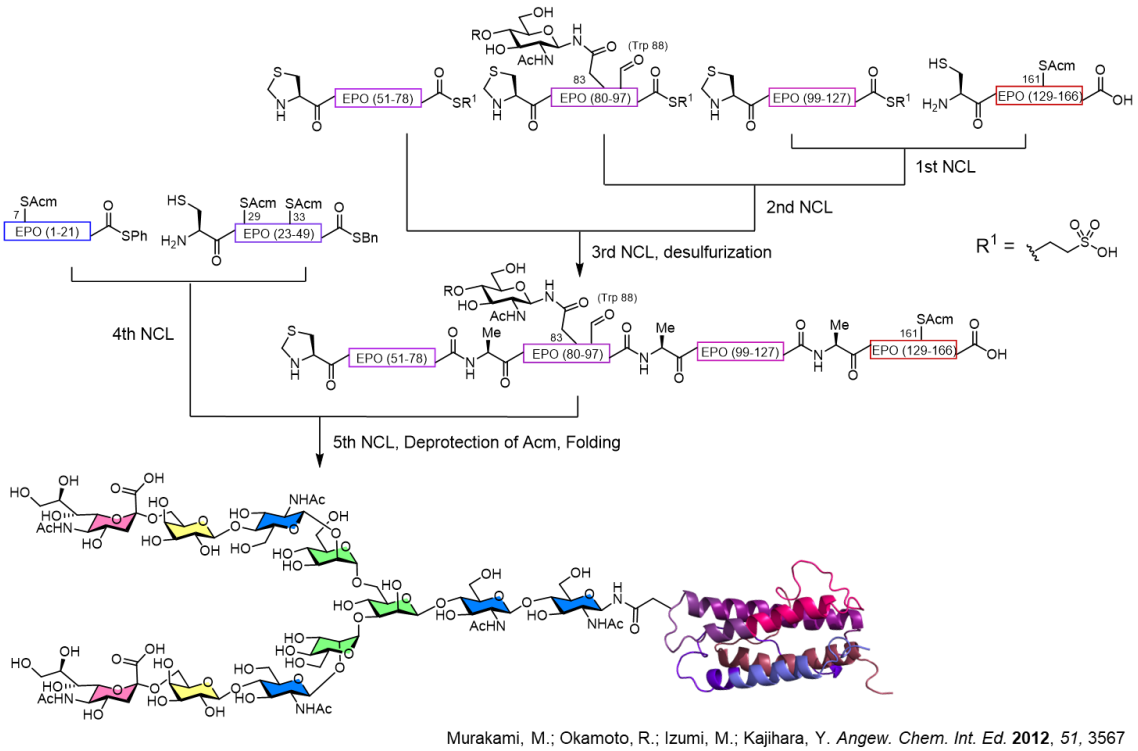
In the same year, Murakami and coworkers achieved the total synthesis of folded erythropoietin having a complex type sialyloligosaccharide using more efficient approach (Figure 1-38).[78] They adopted SPPS to assemble glycopeptides followed by "cassette based" strategy, however the problem was how to protect the carboxylic acid of the sialic acid. The sialic acid residue is known to be extremely labile under acidic conditions and easily hydrolyzed to generate asialo oligosaccharide. They hypothesized that the carboxylic acid group in sialic acid residue might function as an acid catalyst and facilitate the hydrolysis of the glycosidic bond and found sialyl saccharide with protected carboxylic acid by phenacyl ester (Pac) could be used for the Boc SPPS under acidic conditions (Figure 1-37).[78] This protection allowed the generation of the glycopeptide thioester using a thiol linker. Applying Pac protection in SPPS strategy, they completed the chemical synthesis of an erythropoietin (EPO) glycoform (166 amino acids) having one sialyloligosaccharide at Asn83 (Figure 1-38). The four native cysteines were protected by Acm protecting group and alternatively five alanine residues were selected as ligation sites for the six fragments. Glycopeptide thioester was synthesized by Pac based Boc SPPS. After global side chain deprotection, the Pac protected glycopeptide thioester was released from the resin by thiolysis. Four segments were ligated sequentially by NCL. The glycopeptide 50-

166 was subjected to afford desulfurized glycopeptide of the three free thiol groups. A final ligation and deprotection of Acm protecting group of four native cysteines gave the liner EPO. Following oxidative folding protocols afforded an EPO with one disialyloligosaccharide. CD spectra and MS spectra suggested correct folding.



Murakami, M.; Okamoto, R.; Izumi, M.; Kajihara, Y. *Angew. Chem. Int. Ed.* **2012**, *51*, 3567

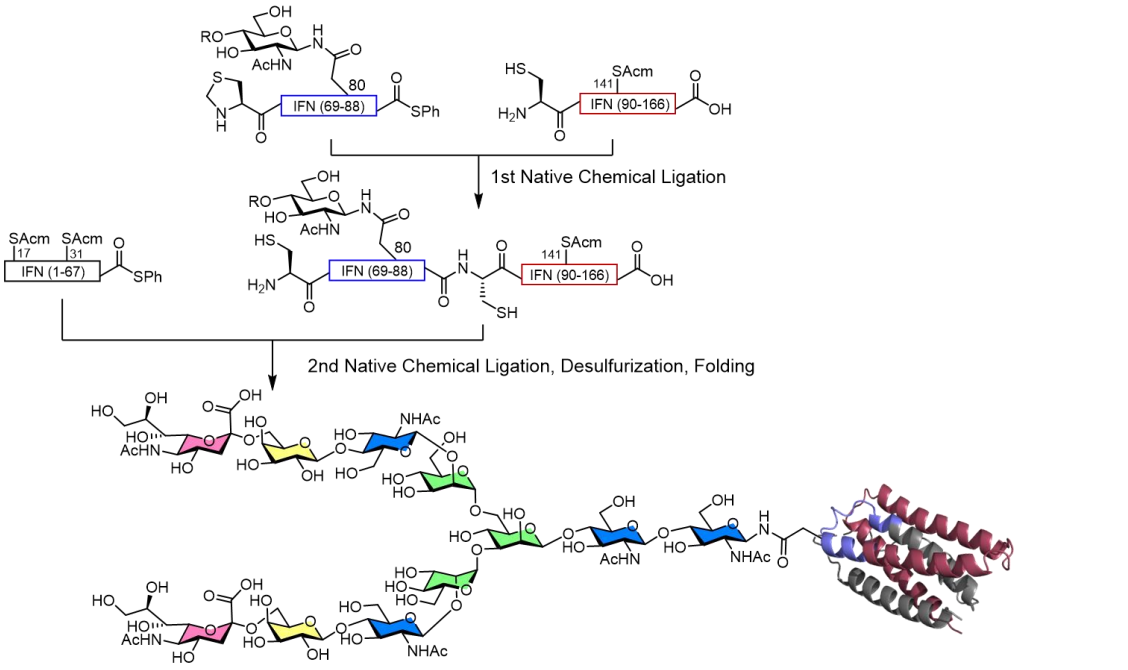
Figure 1-37. Hydrolyzation by intramolecular acid catalyst and the structure of phenacyl protection.



Murakami, M.; Okamoto, R.; Izumi, M.; Kajihara, Y. *Angew. Chem. Int. Ed.* **2012**, *51*, 3567

Figure 1-38. Synthetic strategy of the total synthesis of erythropoietin by efficient phenacyl protection.

Sakamoto and coworkers synthesized glycosylated human interferon-β (IFNβ), a cytokine which possesses immunomodulating, antiviral, and cytostatic activities (Figure 1-39).[79] IFN-β is classified into the subtypes IFNβ-1a and IFNβ-1b, which plays an important role in host cell defense. Especially IFNβ-1a is a glycoprotein, which exhibits significantly more potent immunomodulation, antiviral, and antitumor activities compared to IFNβ-1b. IFNβ-1a consists of 166 amino acids and has N-linked glycan at Asn80. In their synthetic strategy, glycopeptide was synthesized by standard SPPS protocol. Then three segments were sequentially coupled by NCL and followed by desulfurization to give full length glycosyl polypeptide. After deprotection of the Acm protecting groups, stepwise dialysis folding afforded the folded glycosyl IFNβ-1a. Finally, they confirmed the positions of disulfide bonds using enzymatic digestion and mass/mass analysis. CD spectra, MS spectrum and ELISA assay indicated that they completed to synthesize desired glycosylated IFNβ-1a. Their chemically synthesized sialyl IFNβ-1a also exhibited potent antitumor activity in vivo.

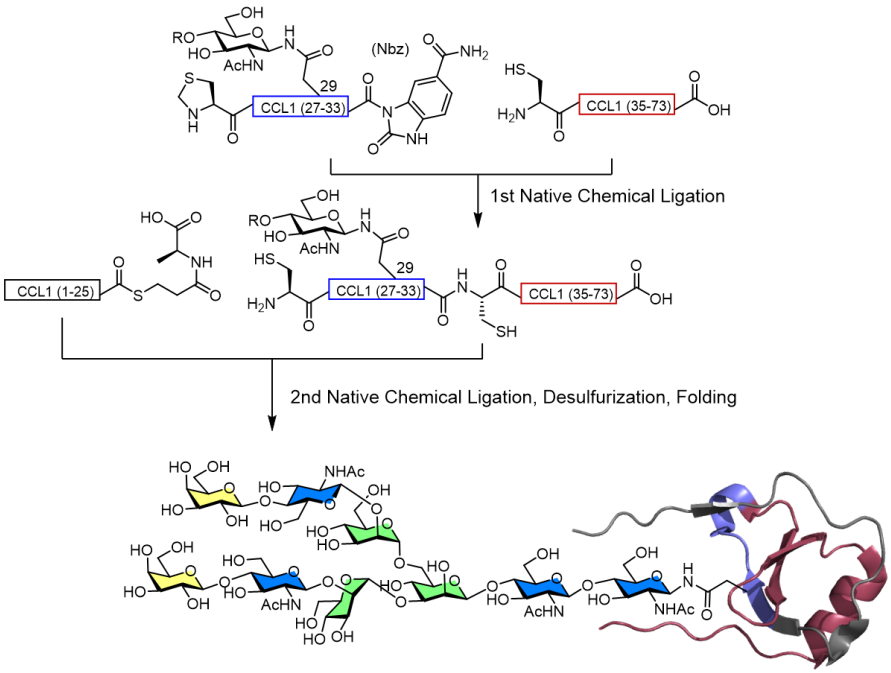


Sakamoto, I.; Tezuka, K.; Fukae, K.; Ishii, K.; Taduru, K.; Maeda, M.; Ouchi, M.; Yoshida, K.; Nambu, Y.; Igarashi, J.; Hayashi, N.; Tsuji, T.; Kajihara, Y. *J. Am. Chem. Soc.* **2012**, *134*, 5428

Figure 1-39. Synthetic strategy of the total synthesis of interferon-β by Sakamoto and coworkers.

Two years later, Okamoto and coworkers achieved the total synthesis of glycosyl CC chemokine ligand 1 (CCL1) and non-glycosyl CCL1 respectively (Figure 1-40).[80] CCL1 is one of a glycosylated cytokines secreted by activated T-cells and acts as a chemo attractant for monocytes. CCL1 consists of 73 amino acids and has N-linked glycan at Asn29. According to their synthetic scheme they synthesized glycopeptide having a complex type nonasaccharide by standard Fmoc SPPS

protocol with the Nbz linker for the synthesis of the glycopeptide thioester building block. Then three segments were sequential ligated under the NCL conditions to afford the linear full-length glycopeptide. Subsequent dialysis folding afforded the folded glycosyl CCL1. Surprisingly they confirmed the structure of glycosyl CCL1 by X-ray crystallographic analysis under the unique quasi racemic crystallization.[81] Finally, chemotaxis assays of these glycoproteins and the corresponding non-glycosylated proteins were carried out. The results were correlated with the chemical structures of the glycoprotein molecules. Consequently, their attempt became the first investigations of the effect of glycosylation on the chemotactic activity of the CCL1.

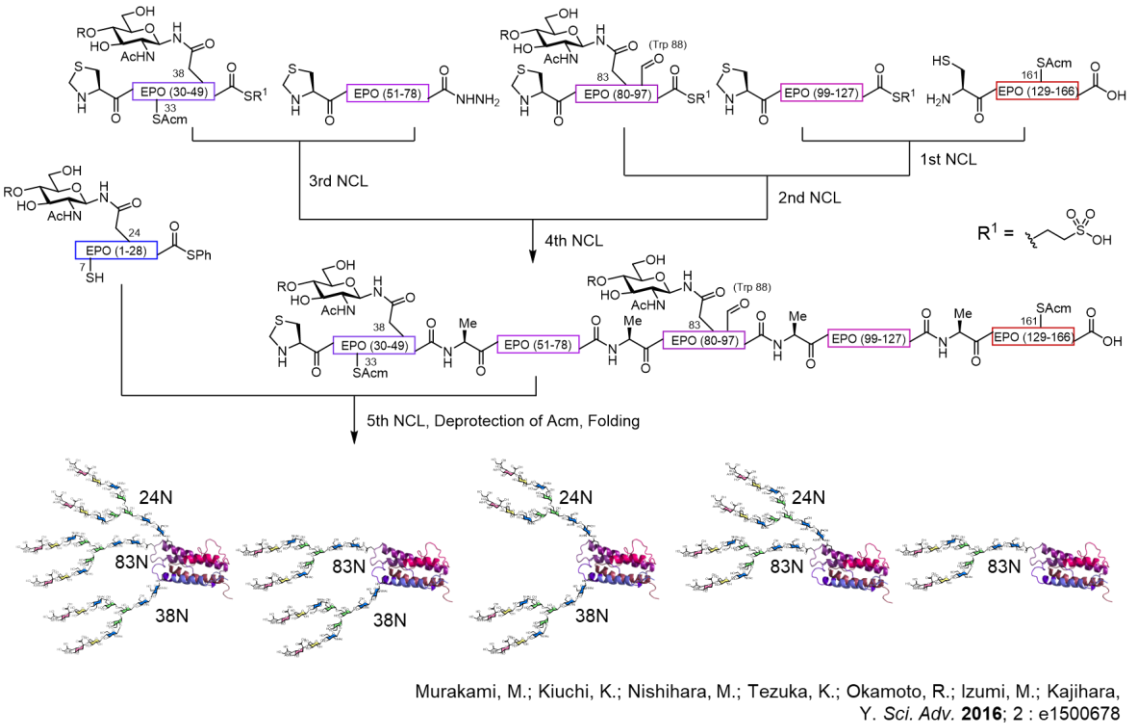


Okamoto, R.; Mandal, K.; Ling, M.; Luster, A. D.; Kajihara, Y.; Kent. S. B. H. *Angew. Chem. Int. Ed.* **2014**, *53*, 5188

Figure 1-40. Synthetic strategy of the total synthesis of CCL1 having a biantennary oligosaccharide.

In 2016, Murakami and coworkers succeeded in the synthesis of glycoproteins of which N-glycosylation numbers and positions were varied from one to three.[82] In this synthesis, they installed complex type biantennary sialylglycan at Asn24, Asn38, Asn83 and as a result they synthesized five EPO glycoforms varying in glycosylation positions and the numbers of human-type biantennary sialyloligosaccharides (Figure 1-41). They synthesized glycopeptide having complex type Pac protected sialylsaccharide by a standard Boc SPPS protocol for the synthesis of the glycopeptide thioester building block. Then six segments were sequentially coupled by NCL and followed by desulfurization. Finally, deprotection of Acm protecting groups afforded the linear full length glycosyl polypeptide. After folding under the stepwise dialysis conditions were performed to have five kinds

of the folded glycosyl EPO glycoforms. Based on an *in vivo* erythropoiesis assay in mice, all types of EPO displayed biological activity, especially EPO having three glycan chains exhibited the highest *in vivo* activity. Furthermore, they discussed structure and activity relationship. Consequently, this example became the first work to give an insight into the relationship between glycosylation and hematopoietic activity.

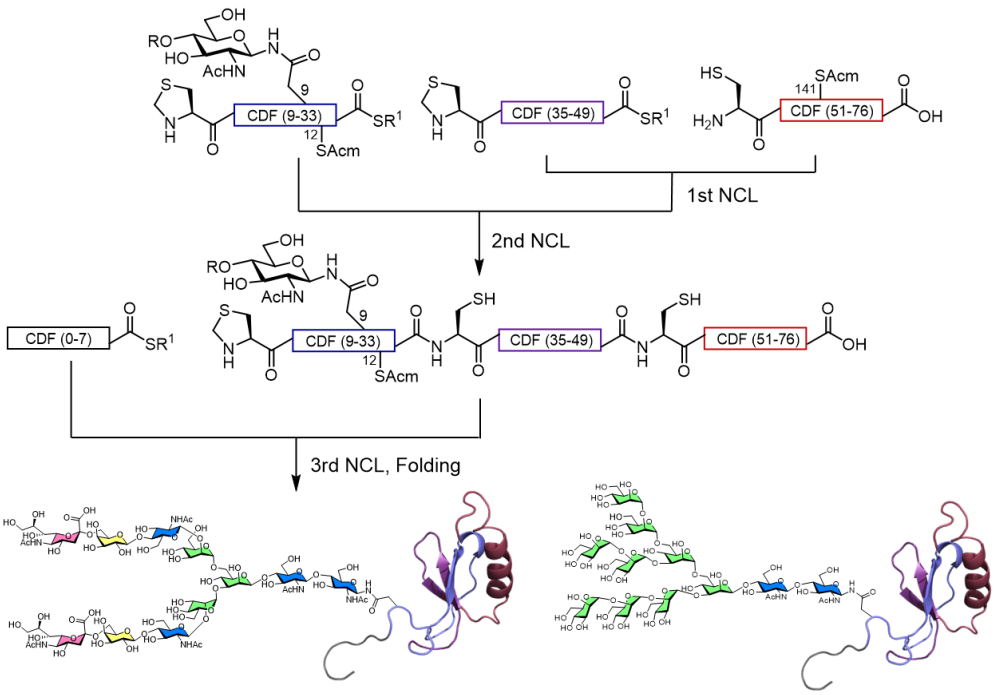


Murakami, M.; Kiuchi, K.; Nishihara, M.; Tezuka, K.; Okamoto, R.; Izumi, M.; Kajihara, Y. *Sci. Adv.* **2016**; 2 : e1500678

Figure 1-41. Total synthesis of five glycoforms of EPO varying in glycosylation positions and the numbers of human-type biantennary sialyloligosaccharides.

For elucidating the effect of oligosaccharide on glycoprotein, Hien and coworkers synthesized fractalkine (CDF) having complex type and high mannose type oligosaccharide (Figure 1-42).[83] Furthermore, they conducted structural analyses of the glycoproteins using stable isotope labeling with position specific and gradient of ^{15}N-isotope concentration (SGI). CDF consists of 76 amino acids and has N-linked glycan at Asn9. According to their synthetic scheme they synthesized glycopeptide thioesters having complex type nonasaccharide and high-mannose type saccharide by either standard Fmoc or Boc protocol respectively. Then four segments were sequential coupled by NCL and then deprotection were performed. Finally, deprotection of Acm protecting groups afforded the full-length glycopeptide. After oxidative folding process afforded the folded two types of glycosyl CDF, in which thirteen ^{15}N-labeled amino acids were inserted at specific positions. They also synthesized non-glycosyl CDF in the same manner. CD spectra and enzymatic disulfide bond mapping

confirmed that both glycosylated CDF and non-glycosylated CDF were correctly folded. The SGI labeling protocol enabled us to assign specific amino acids to the ^{1}H-^{15}N HSQC spectra and investigated the dynamics of local conformational properties of three synthetic CDFs. According to the HSQC combined with TOCSY and NOESY methods, glycosylation did not affect conformational changes of CDF at room temperature. However, T_1 values indicated that the glycan motif influenced dynamics at the local conformation. As the result, temperature varied circular dichroism CD spectra and T_1 values indicated that oligosaccharides might inhibit protein fluctuation and stabilize protein structure. Their work became a landmark for the application of the synthetic glycoproteins.



Hien, N. M.; Izumi, M.; Sato, H.; Okamoto, R.; Kajihara, Y. *Chem. Eur. J.* **2017**, *23*, 6579

Figure 1-42. Total chemical synthesis of CDF and structural analysis by using specific and gradient of ^{15}N-isotope concentration.

Recently Li and coworkers achieved the total chemical synthesis of Interleukin 17A (IL17A) glycoforms including complex type biantennary sialyloligosaccharide, mono GlcNAc and chitobiose disaccharide efficiently (Figure 1-43).[84] IL17A (132 amino acids) is a glycoprotein that possesses a cystine linked homodimeric structure and has N-linked glycan at Asn45. For the synthetic route, they divided the reduced form of IL17A into four segments that could be assembled using three times NCL. Followed by "cassette base" strategy, a glycopeptide hydrazide had a β-mercapto aspartic acid at N-terminus was synthesized by Fmoc SPPS. They also applied Endo-M-catalyzed transglycosylation to obtain a glycopeptide having complex type biantennary sialyloligosaccharide. Next, they explored the

ligations to assemble the IL17A. Then four segments sequential NCL and followed desulfurization were progressed convergently. Finally, deprotection of Acm protecting groups afforded the linear full-length IL17A segment. After cysteine/cystine with arginine folding conditions afforded the folded two types of glycosyl IL17A glycoforms. To confirm the disulfide linkages of the folded protein, trypsin digestion was performed. Moreover, CD spectrum and MS spectra indicated that synthesized IL17A correctly folded. In vitro biological activity of the IL-17A isoforms was evaluated by utilizing a normal human adult dermal fibroblasts (NHDF) assay, and glycoform with sialyl undecasaccharides displays much weaker stimulatory effect. Further surface plasmon resonance (SPR) and hydrogen/deuterium exchange (HDX) mass spectroscopic experiments suggested that the complex type glycan impeded the binding between IL17A and its receptor.

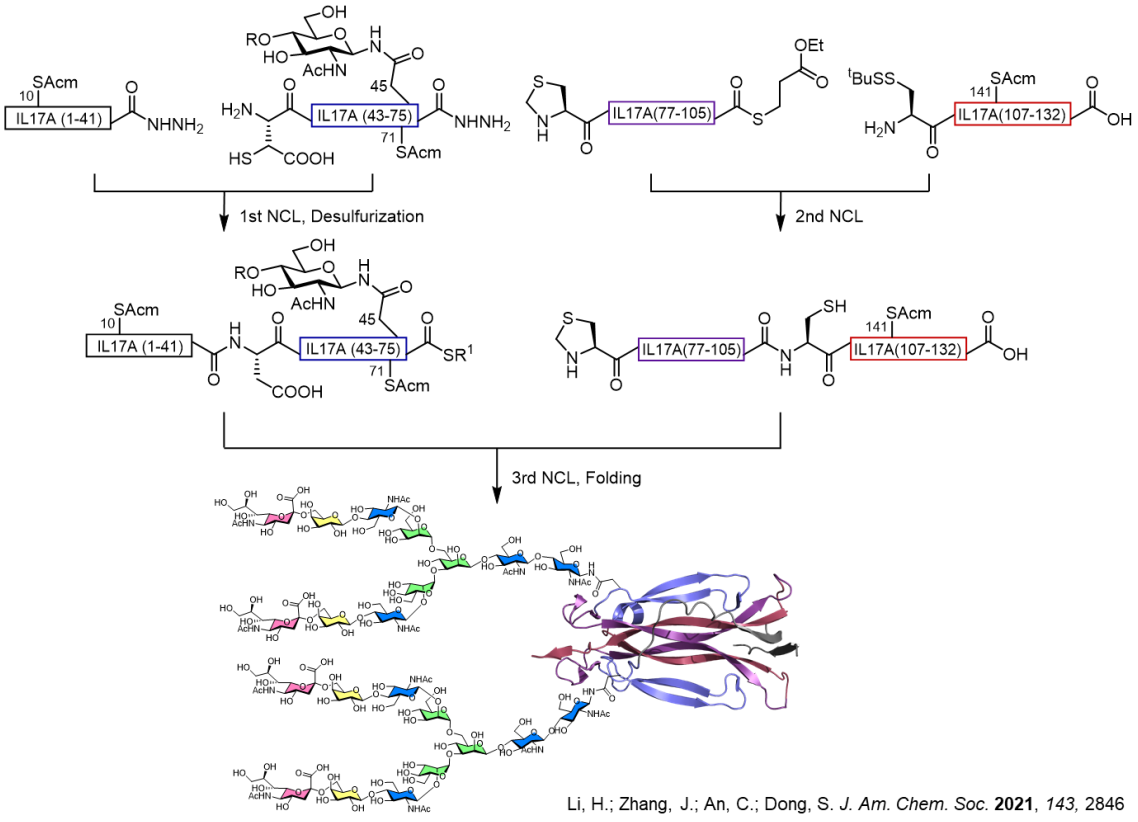
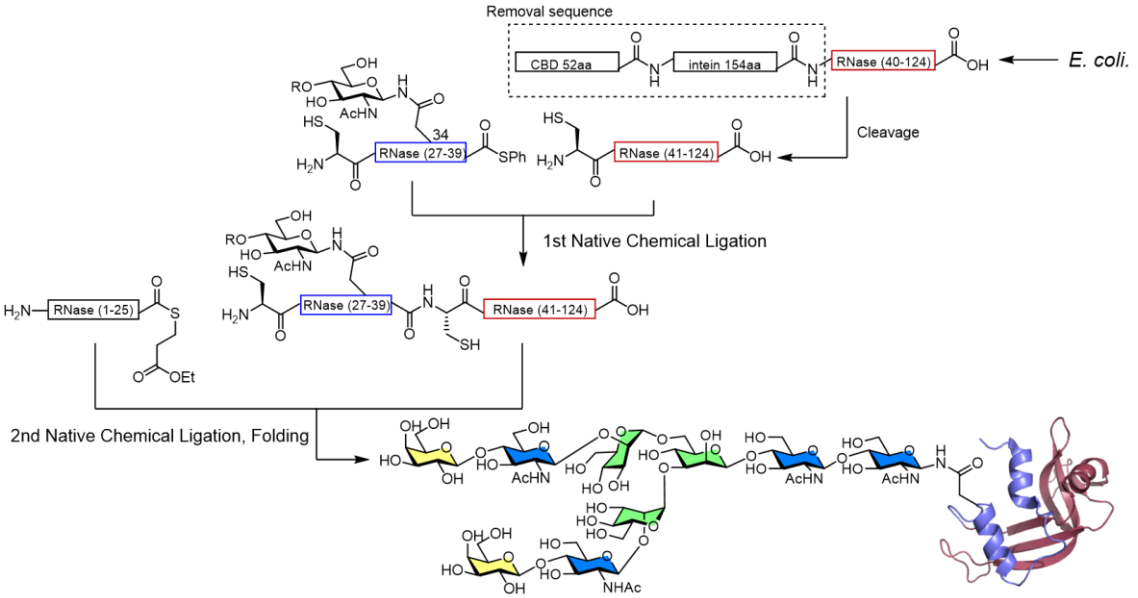


Figure 1-43. Total chemical synthesis of IL17A dimer having two biantennary sialyloligosaccharides.

Moreover, expressed protein ligation (EPL) developed by Muir has been successfully utilized for glycoprotein synthesis, which is termed semisynthesis.[85-86] The first semisynthesis of glycoprotein was Ribonuclease C (RNase C) synthesized by Piontek and coworkers (Figure 1-44).[87-88] RNase C (124 amino acids) is a glycoform of bovine RNase having complex type glycan at asn34. RNase C has eight cysteines engaged in four disulfide bridges. They designed and divided RNase C

into three segments that could be assembled using three NCL. Glycopeptide thioester having complex type asialooligosaccharide was synthesized by Fmoc SPPS procedure and “safety catch” linker strategy. Another key fragment was the peptide segments 40-124, which was obtained recombinantly from *E.coli.* using chitin binding purification and intein cleavage. Then sequential convergent NCL by three segments were progressed to afford the linear full-length glycopeptide. Folding was allowed to proceed for four days, and the protein was subsequently isolated by gel filtration in 71% yield. The folded synthetic RNase C was assayed for enzymatic activity. It showed a relative activity of 56% when compared with a sample of commercial RNase A. In addition to the activity, a CD spectrum indicated synthetic RNase C correctly folded.



Piontek, C.; Ring, P.; Harjes, O.; Heinlein, C.; Mezzato, S.; Lombana, N.; Puhner, C.; Puttner, M.; Silva, D. V.; Martin, A.; Schmid, F. X.; Unverzagt, C. *Angew. Chem. Int. Ed.* **2009**, *48*, 1936
Piontek, P.; Silva, D. V.; Heinlein, C.; Puhner, C.; Mezzato, S.; Ring, P.; Martin, A.; Schmid, F. X.; Unverzagt, U. *Angew. Chem. Int. Ed.* **2009**, *48*, 1941

Figure 1-44. The first semisynthesis of RNase C having a biantennary complex oligosaccharide by applying *E. coli* expression system.

More recently Okamoto and coworkers achieved semisynthesis of several glycoproteins.[89-90] They also developed thioesterification method to utilize the semisynthesis of homogeneous glycoproteins. Thioesterification at the C terminus of recombinant peptides is an essential process for the semisynthesis of proteins. The thioesterification methods using such as intein system[91] or specific protease including sortase[92] developed by Pentelute groups and CPaseY[93] developed by Otaka groups were utilized for the recombinant polypeptides, however the resultant peptide thioesters were obtained in a low yield due to the aggregation of hydrophobic areas of expressed peptides. To solve these problems, Okamoto and coworkers developed reactions for chemical thioesterification of recombinant

peptides by using the unique both of C-terminal and internal Cys modification strategies. In 2012 they achieved internal Cys modification, by S-thiocarbonylation (Figure 1-45 a).[94] They designed a new thioester synthesis consisting of three steps: the selective S-thiocarbonylation of a β-thiol group of cysteine residue to provide a phenyl xanthate group by using O-phenylchlorothionoformate. Treatment of the resultant peptide with N-acetylguanidine to give peptide-thioester. They also found that N-Boc protection of both amino groups of the Lys side chains, and the N terminus are necessary to avoid side reaction by the N-acetylguanidine treatment. In addition to this efficient reaction, peptidyl N-acetylguanidine is much less reactive for NCL than a corresponding thioester, they could be directly used for one-pot sequential NCL. Using their thioesterification approach, they succeeded to synthesize semisynthesis of interleukin 13, which is introduced in next paragraph (Figure 1-46).[89]

Two years later, Kajihara and coworkers developed another thioesterification methods for the semisynthesis of glycoproteins by using cyanylation of Cys residue.[95] Reaction mechanism is shown in the following figure (Figure 1-45 b). First, the thiol group of the cysteine residue is cyanylated by cyanylation reagents. This intermediate is comparably stable, and in the presence of nucleophiles such as hydrazine or alkyl thiols, the amide nitrogen of the cysteine attacks the *sp* carbon of cyanyl groups nucleophilically, subsequently forming a five-membered ring intermediate, which gave the corresponding peptide hydrazide or peptide thioester efficiently. In their reports, they demonstrated the preparation of polypeptide thioester consists of 94 amino acids by *E.coli* expression followed by cyanylation, hydrazinolysis, and thiolysis. Their attempts to synthesize expressed peptides thioester utilizing internal Cys residue succeeded to synthesize several homogeneous glycoproteins. However, these thioesterification protocols by Cys residue required multiple reaction steps and selective modification of the C-terminal Cys was also difficult in peptides including internal Cys residues at that time.

To solve these problems Okamoto and coworkers found two types of highly chemoselective peptides thioesterification reactions in 2022.[90] These new thioesterification reactions progressed without any protecting groups, applying an intramolecular N-S acyl shift reaction at a C-terminal Cys residue (Figure 1-45 c, d). They made a well improved thioesterification based of the Aimoto[96-97] and Macmillan[98] groups. The first method is chemical transformation of peptide-Cys using bis(2-sulfanylethyl)amine (SEA) efficiently yielded peptide-SEA, which can be utilized as a peptide thioester surrogate (Figure 45 c).[90] Reaction starts from conversion of peptide-Cys to peptide-S-Cys-thioester by an intramolecular N-S acyl shift at the C-terminal Cys and the trans-thioesterification of the peptide-S-Cys-thioester with an external thiol such as SEA to afford peptide-SEA (peptide thioester surrogate). Fortunately, the equilibrium state is biased toward peptide-SEA because peptide-SEA was oxidized in air to form C-terminal disulfide bonds, which cannot be converted to peptide-S-Cys-

thioester. Alternatively, they also found a peptide having cysteinyl glycyl cysteine (peptide-CGC) can be used for the synthesis of peptide thioesters from unprotected polypeptides (Figure 1-45 d).[90] Peptides-CGC could be transformed to peptide-thioesters through peptides with CG-thioester units at the C terminus *via* an N-S acyl shift at Cys. Then peptide-CG-thioester can be converted to peptide thioesters *via* diketopiperazine (DKP) formation and a subsequent N-S acyl shift at Cys. Both methods were successfully applied to recombinant polypeptides and utilized for NCL. In the next paragraph, the several examples of the semisynthesis applying Cys based thioesterification protocols developed by Okamoto and coworkers will be introduced.

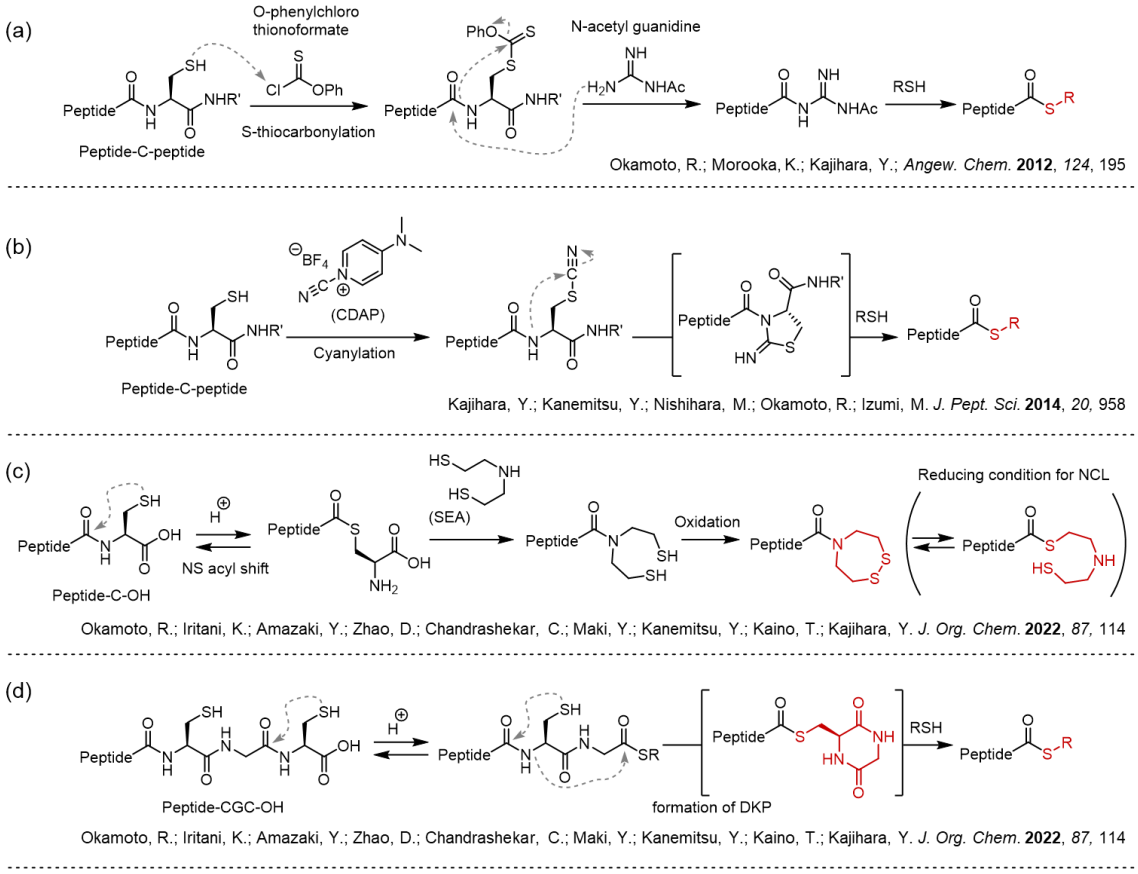


Figure 1-45. Cys based peptide α-thioesterification reactions developed by Kajihara and coworkers. (a) Peptide thioesterification using S-thiocarbonylation, "guanidine method" (b) Peptide thioesterification by cyanylation. (c) C-terminal Cys selective thioesterification applying peptide-C sequence. (d) C-terminal Cys selective thioesterification applying peptide-CGC sequence.

Applying their developed thioesterification approaches, Okamoto and coworkers synthesized interleukin 13 (IL13) in 2014 (Figure 1-46), inducible T cell costimulator (ICOS) in 2022 (Figure 1-47).[89-90] At the first they reported a new semisynthetic strategy of IL13 in which the middle

region is glycosylated.[89] IL13 is a glycoprotein that consists of 112 amino acids and has a complex-type oligosaccharide at Asn52. According to their synthetic strategy of IL-13, they employed four peptide segments and three sequential NCL to obtain a full-length glycoprotein. Then IL13 segment 1-27 and 55-112 were simultaneously prepared by *E.coli.* expression system and cleaved fusion proteins by CNBr, which can selectively cleave an amide bond between the Met and Cys residues; furthermore, the Met residue is converted into the homoserine lactone (Hsl) (Figure 1-46). They utilized cleavage reaction on the amide bond between artificially inserted Met and Cys with CNBr to afford segment 1-27 and 55-112 simultaneously in high yield. Then segment 1-27 was converted to peptide thioesers by using their developed "guanidine method" and obtained peptide α thioester. On the other hands, glycopeptide thioester was synthesized by Fmoc standard SPPS protocol followed by their reported manner. After preparation of all segments by semisynthetic approaches, sequential NCL was convergently progressed to afford the linear full-length IL13 segment. Folding process under stepwise dialysis with redox reagents yielded folded IL13. ELISA and CD spectrum indicated that semisynthesized IL13 correctly folded having α-helix rich conformation.

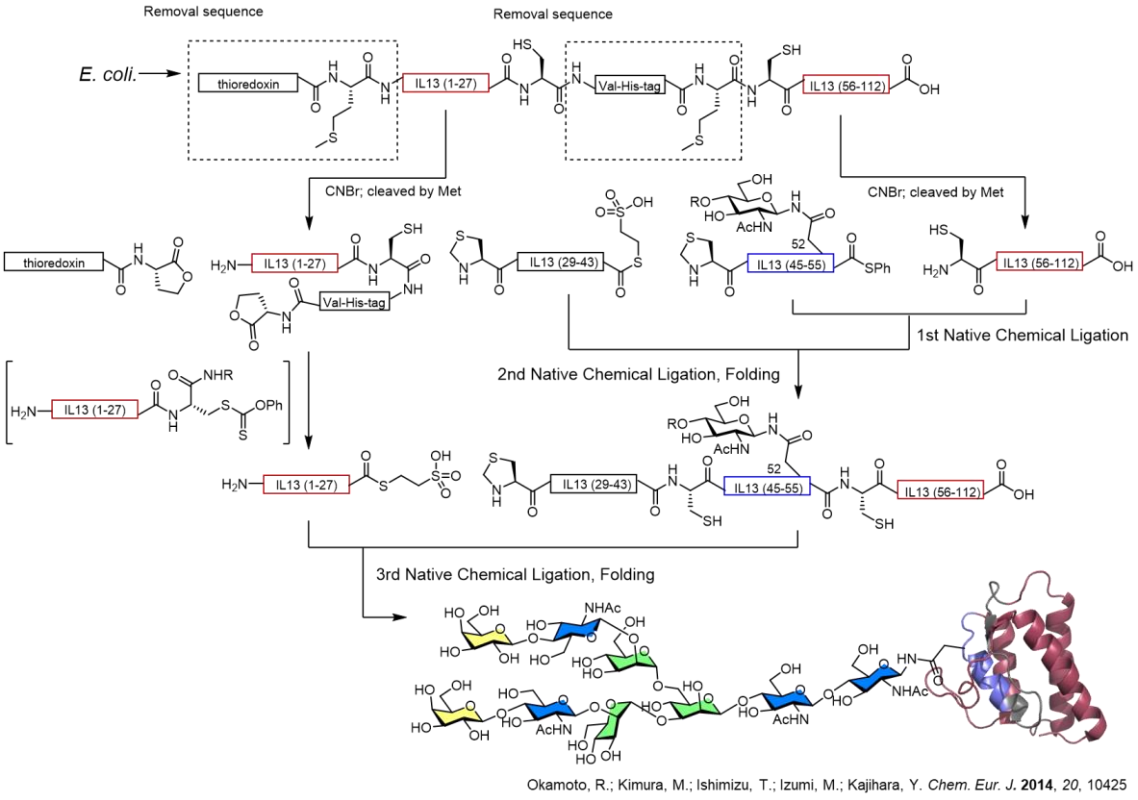


Figure 1-46. Semisynthesis of IL13 using chemical expressed thioesterification method. The key of the synthesis is semisynthesized simultaneously two peptide segments.

In the semisynthesis of ICOS, Okamoto and coworkers applied peptide-Cys thioesterification by using SEA (Figure 1-47).[90] ICOS is a costimulator for the activation of T cells and has N-glycosylation sites in the extracellular domain. Thus, their synthetic ICOS consists of 106 amino acids and has two biantennary sialyl oligosaccharide at Asn 89 and Asn110. A convergent synthetic strategy was designed for glycosyl ICOS using a recombinant polypeptide thioesterification as a key reaction. They employed three peptide segments and two sequential NCL to obtain a full-length glycoprotein. ICOS segment 1-62 was co-expressed with Small Ubiquitin-related Modifier (SUMO) and six His (His Tag) sequences by *E.coli.* expression system. Thioesterification applied SEA was progressed convergently after cleavage of SUMO sequences by SUMO protease to afford peptide-SEA. Obtained peptide-SEA was subjected to NCL with glycopeptide prepared from glycopeptides thioester 63-88 and cysteinyl peptide 89-106 synthesized by a standard Boc SPPS protocol. In terms of glycopeptide-thioester, they employed Fmoc-SPPS. The resultant all segments was sequentially coupled by NCL to obtain a full-length glycopeptide of ICOS. Using stepwise dialysis conditions, folding process with redox reagents yielded folded glycosyl ICOS (Figure 1-47). ELISA, CD and MS spectrum indicated that semisynthesized ICOS correctly folded structure having β-sheet rich conformation. By using both chemical synthesis and semisynthesis as appropriate, it might be possible to synthesize glycoproteins with larger molecular weights

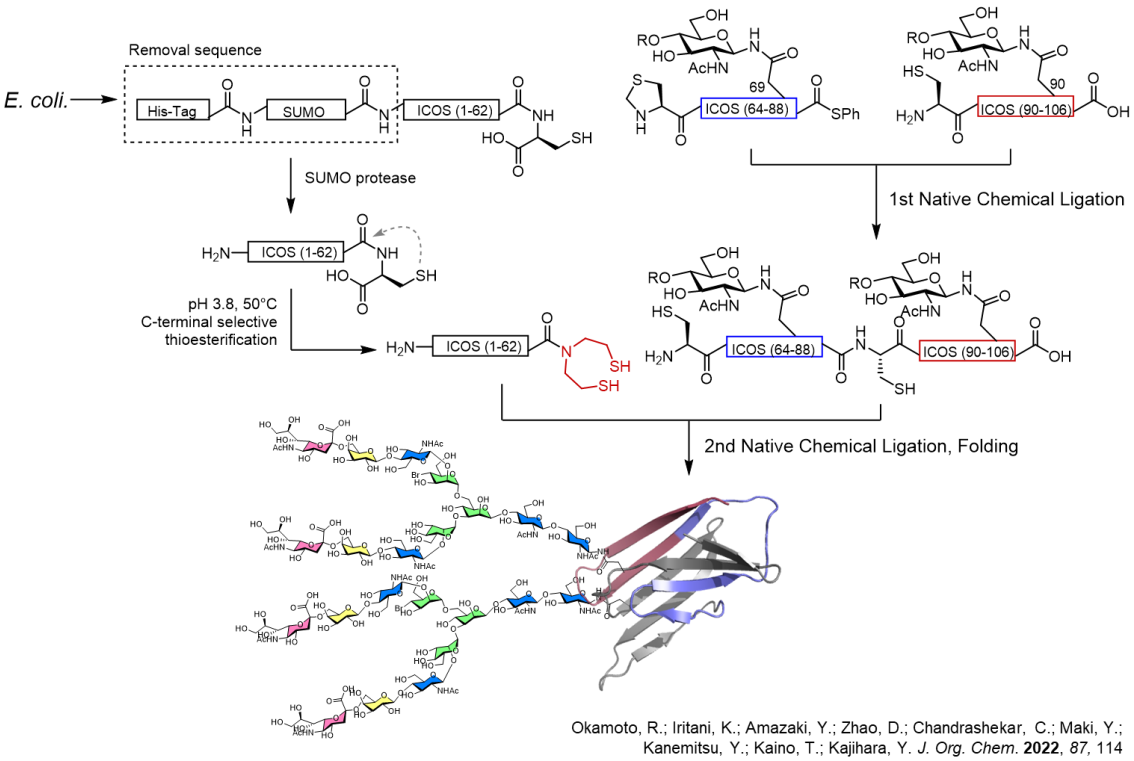


Figure 1-47. Semisynthesis of ICOS by utilizing novel thioesterification approach by peptide-SEA.

1-10. Essential problems and improvement for the novel glycoprotein synthesis

Superior robustness and efficiency need to be developed for the synthesis of homogeneous glycoproteins and their analogs. Chemical approaches are relatively time-consuming due to requiring multiple synthetic steps and need an appropriate amount of valuable N- or O-glycans because glycopeptide syntheses are set at the early stage of total synthesis. To accelerate the elucidation studies of glycan functions, we need to solve these synthesis drawbacks.

The author of this thesis has considered a new method to synthesize glycoproteins by direct insertion of glycosyl amino acids into the target peptide backbone of glycoproteins (Figure 1-48). In this chemical insertion strategy, many drawbacks are solved. The major outcome is that there is no need to synthesize glycopeptides by SPPS or Lansbury aspartylation (*chapter 1-9*), which decrease the reaction yield, in addition to this, there is no need to use enzymatic elongation of glycans. There has never been a strategy to introduce such a single residue amino acid into a peptide backbone of target glycoprotein. In this strategy, because glycosyl amino acid can be used at the late stage of synthesis, we can reduce the excessive consumption of glycans. However, conventional methodology using simple condensation reagents need peptide-side chain protection, resulting in multi-step total synthesis. Therefore, we needed to develop a new reaction for the insertion of glycosyl amino acids into the target peptide backbone of glycoproteins.

Therefore, the author of this thesis aspired to develop a more chemoselective amide bond formation reaction, which can be used toward non-protected peptides. After examining various functional groups to find selective reactions, thioacids was found to be the most promising functional group. Thioacids have been used for a variety of amide bond forming reactions. In *chapter 1-11*, I will describe the reactivity and selectivity of thioacids.

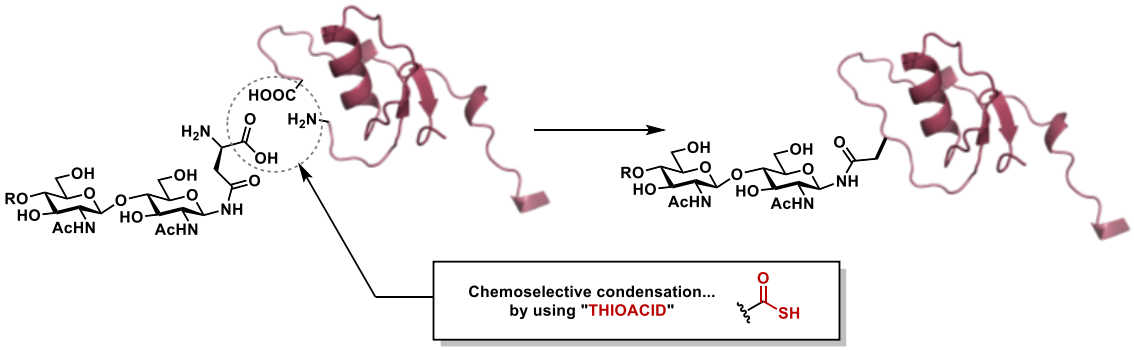


Figure 1-48. Novel concept of the synthetic strategy of homogeneous glycoprotein; Chemical insertion of glycosyl amino acid applying the chemoselective amide bond formation with thioacid functional groups.

1-11. Chemoselective coupling reactions with thioacid functional group

To set up alternative synthesis methods, The author of this thesis studied several key reactions based on amino thioacids. Thioacid (-COSH) was originally used as acylation functionalities for condensation and ligation reactions with azide, aziridine, dithiocarbamate, isocyanide, sulfonamide, nitropyridyldisulfide, S-nitroso intermediates, Sanger and Mukaiyama reagent, N,O-Bis(trimethylsilyl)acetamide, isonitrile. The author of this thesis has interest into the potent nucleophilicity and low pKa (~3) value of thioacids compared with carboxylic acids. These reactions can be divided into two main categories (Figure 1-49).[99] The first is a reaction in which the thioacid is activated by reagents and attacked by nucleophile such as amine chemoselectively. The second is a reaction in which the amine is activated as electrophile, which is selectively condensed with thioacid through intramolecular S-N acyl shift. In both reactions, the key is the differentiation of thioacid from other functional groups such as an amine or carboxylic acid groups. It is the most essential that both reactions proceed in a mild manner by a completely different reaction pathway than the amide bond formation reaction using conventional condensation reagents such as N,N'-Dicyclohexylcarbodiimide (DCC) or benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate (PyBOP). These chemoselective amide bond formation reactions applying thioacid are used for the total chemical synthesis of proteins and site-selective modification of proteins. Herein, the recent comprehensive review of thioacid-based chemoselective amide bond formation reactions and applications.

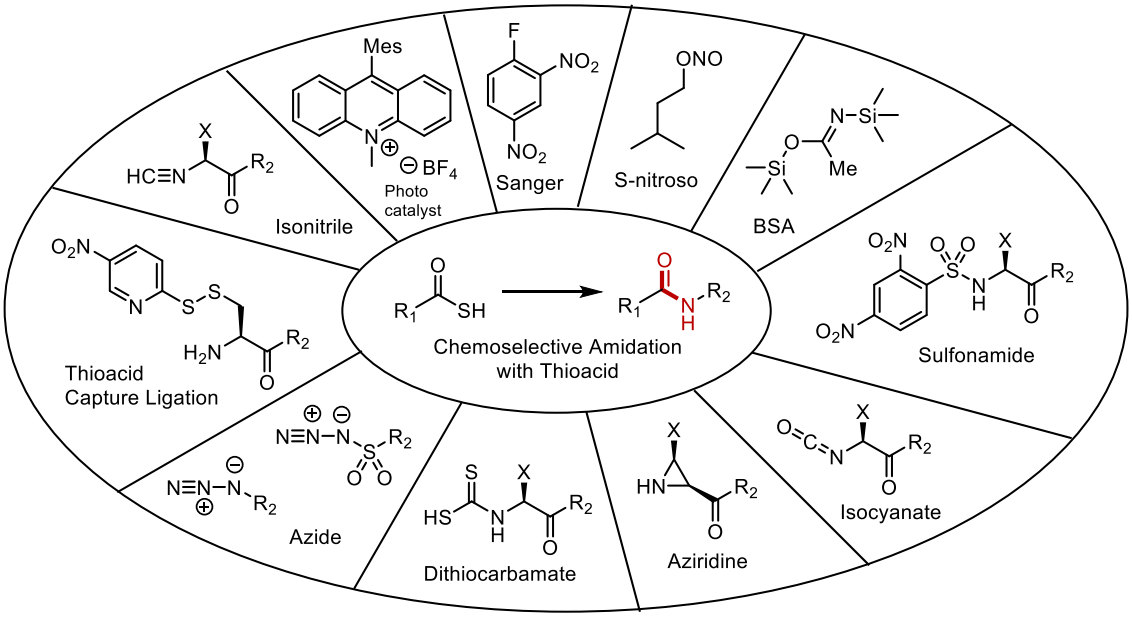
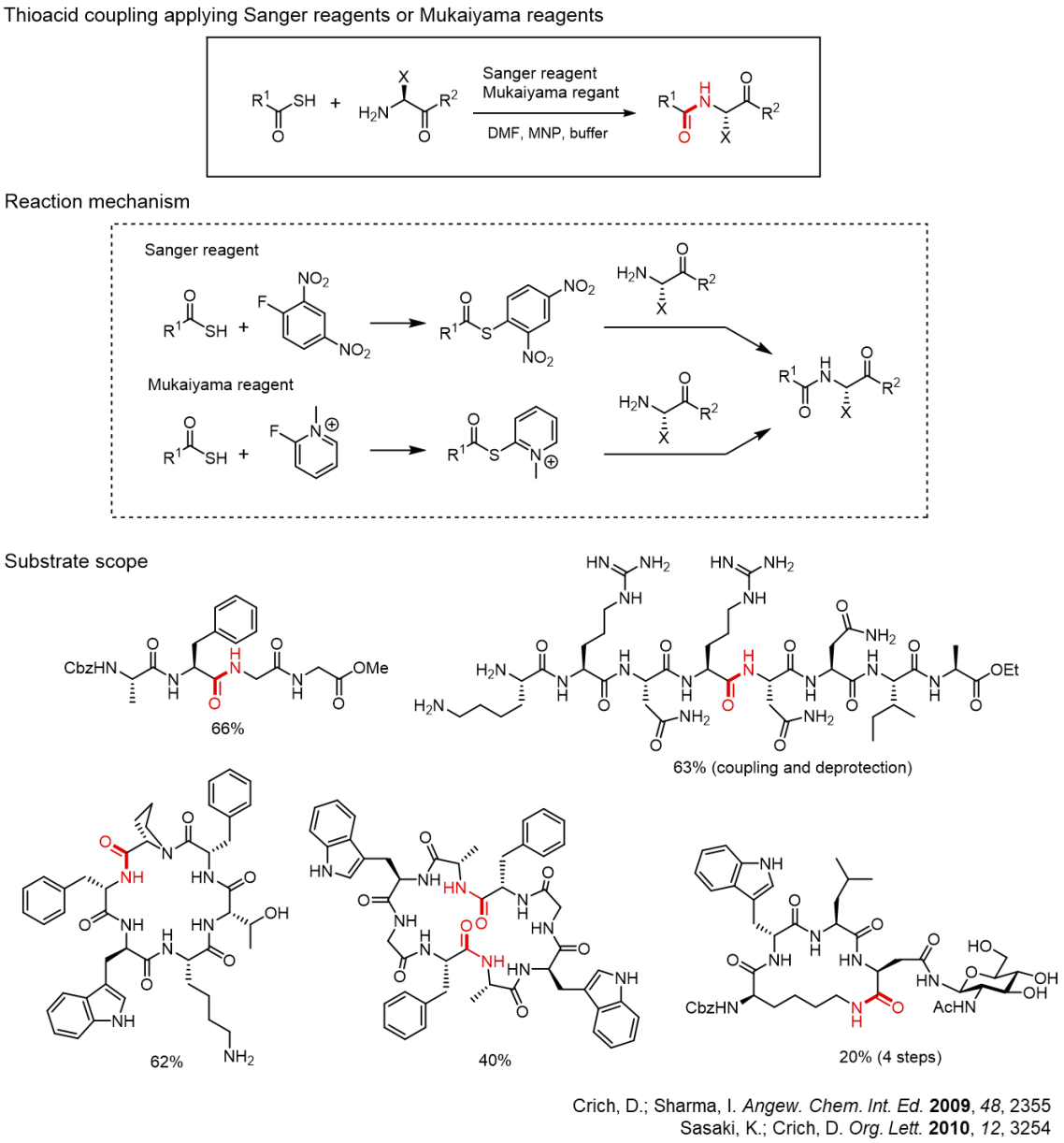


Figure 1-49. an overview of functional group selective amide bond formation reactions based on thioacids. Examples are reactions with dithiocarbamate, isocyanide, sulfonamide, nitropyridyldisulfide, aziridine, azide, S-nitroso intermediates, Sanger reagent, BSA, isonitrile.

1-11-1. Thioacid coupling applying Sanger reagent and Mukaiyama reagent

Thioacids are used in amide bond formation reactions. Crich and coworkers demonstrated that thioacid formed active thioesters with Sanger and Mukaiyama reagents (Figure 1-50).[100-101] Under weak base conditions, the thioacid becomes a thiocarboxylate anion, which causes ipso-substitution for the Sanger and Mukaiyama reagents, and the thioacid is readily converted to the active thioester. The active thioester undergoes nucleophilic substitution reaction with amine, resulting in an amide bond formation. Using this reaction, the synthesis of octapeptides and several cyclic peptides containing glycan were attempted at room temperature in DMF and buffer solutions. This reaction does not use conventional condensing reagents.

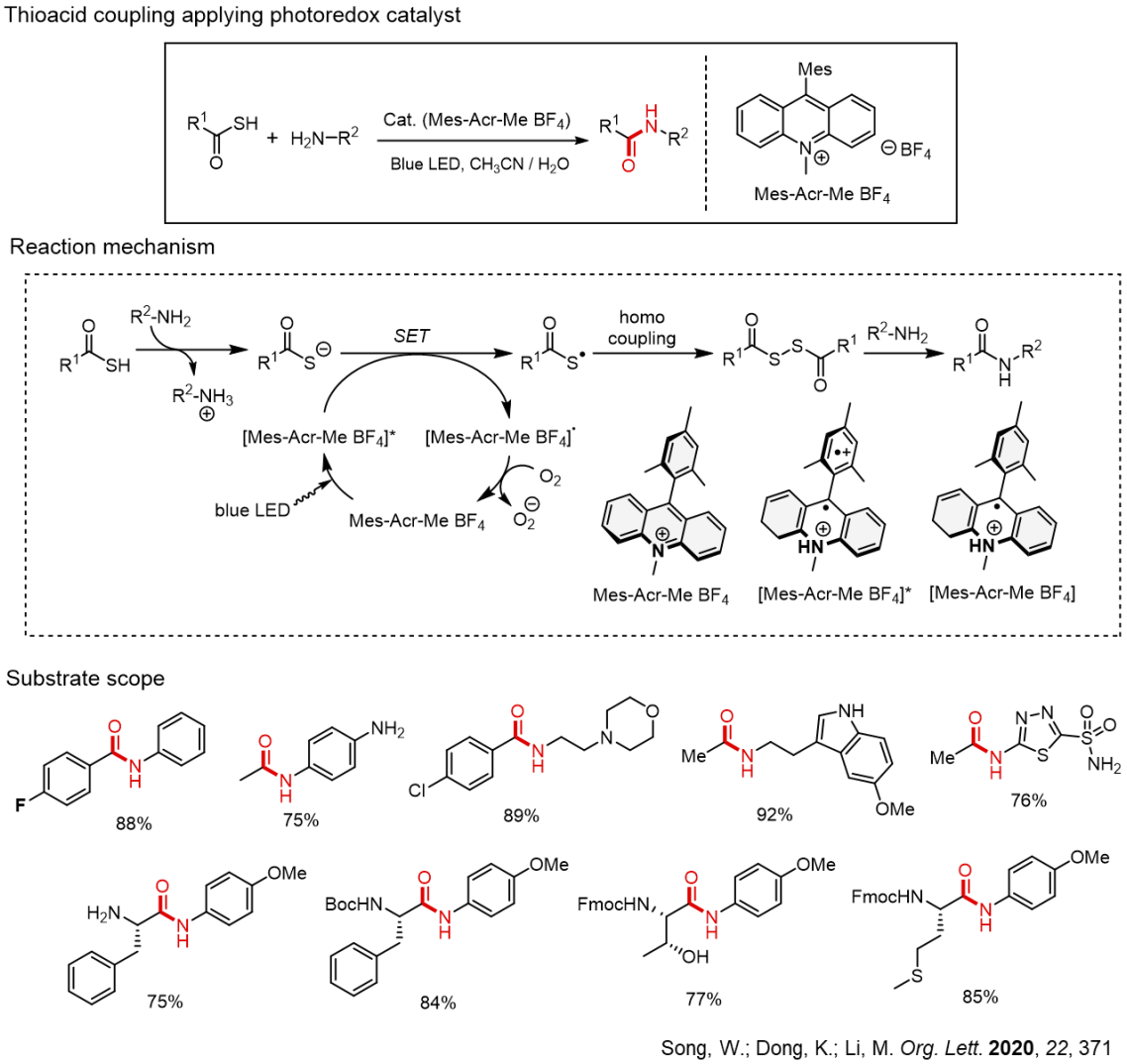


Crich, D.; Sharma, I. *Angew. Chem. Int. Ed.* **2009**, *48*, 2355
Sasaki, K.; Crich, D. *Org. Lett.* **2010**, *12*, 3254

Figure 1-50. Thioacid coupling applying Sanger reagent and Mukaiyama reagent.

1-11-2. Thioacid coupling applying organic photoredox catalyst

Recently a metal-, base-, and additive-free amide bond formation reaction were developed under an "organic photoredox catalyst" by Song and coworkers (Figure 1-51).[102] They achieved photoredox catalyst that assists amidation of thioacid without affecting other functional groups such as alcohols, phenols, ethers, esters, halogens, or heterocycles. Catalytic activation of thioacid has been established in reaction systems using cadmium sulfide and rubidium catalysts so far. Song's method using organic photoredox catalyst, Mes-Acr-Me BF_4, featured a broad substrate scope, good compatibility with water and air, and high yields. The potential utilities were demonstrated by the synthesis of important drug molecules such as paracetamol, melatonin, moclobemide, and acetazolamide. This is the first example of an organocatalyzed amide bond formation that has been shown to have high potential for application in protein synthesis. However, this report did not examine epimerization in amide bonds and its application to peptides. More information is expected in the future.

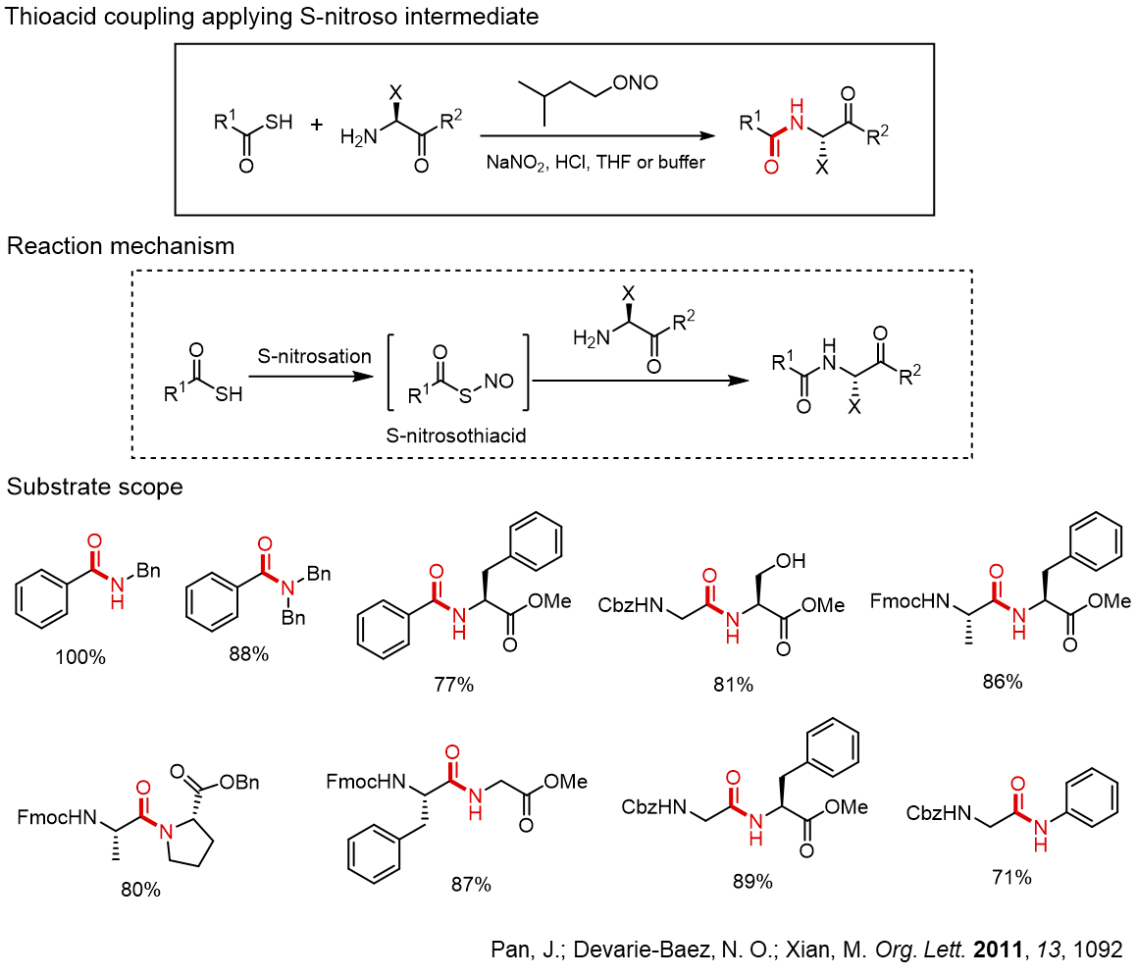


Song, W.; Dong, K.; Li, M. *Org. Lett.* **2020**, *22*, 371

Figure 1-51. Thioacid coupling applying organic photoredox catalyst.

1-11-3. Thioacid coupling applying S-nitrosothioacid (NTA)

The third reaction that activates thioacid is an amide bond formation strategy from simple thioacid and amine through S-nitrosothioacid (NTA) intermediates, which is generated with organonitrite (R-ONO) developed by Pan and coworkers (Figure 1-52).[103] Although it is known that S-nitrosothioacid are unstable moieties, their chemistry, especially synthetically useful reactions, had not been well studied so far. Optimizing conditions, they found nitrosation as a novel strategy for thioacid activation mediated by reactive NTA intermediates. Compared to other amide formation methods for example DDC or PyBOP, this reaction only utilized readily available organonitrite such as commercially available amyl nitrite or sodium nitrite as activators. It took place under very mild reaction conditions, and the reaction rate was extremely fast. It also showed excellent selectivity toward amines over hydroxyl groups and succeeded to synthesize dipeptides library under mild conditions and aqueous buffer solutions including THF, DMF and PBS were usable for solvents. Although this report did not examine epimerization and applications for the longer peptide chains, they thought that this S-nitrosation protocol should be promising for peptide coupling/ ligation and selective N-acylation.

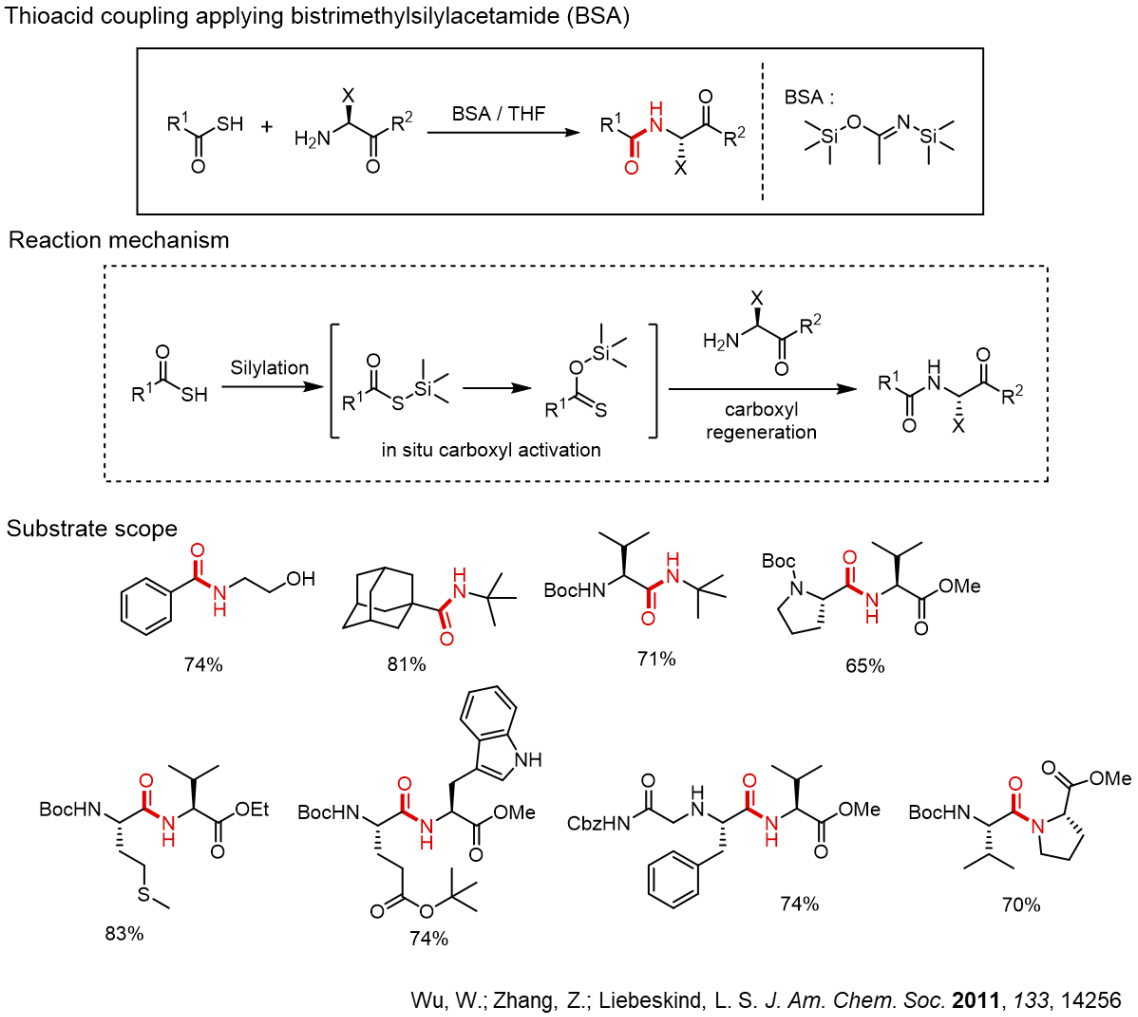


Pan, J.; Devarie-Baez, N. O.; Xian, M. *Org. Lett.* **2011**, *13*, 1092

Figure 1-52. Thioacid coupling applying S-nitrosothioacid (NTA).

1-11-4. Thioacid coupling applying bistrimethylsilylacetamide (BSA)

The activation of thioacids using silylation reagents was achieved by Wu and coworkers.[104] In their report, the novel reactivity of O-silylthionoesters with amine nucleophiles to generate oxoamides (rather than thioamides) was described (Figure 1-53). A robust first-generation of trimethylsilylation protocol using bistrimethylsilylacetamide (BSA) combined with the unique reactivity of the O-silylthionoesters toward primary and secondary amines to generate oxoamides provides the simplest means of activating a thioacid for an amide bond formation at neutral pH conditions. This unique carboxyl activation relied on the known spontaneous formation of O-silylthionoesters from their S-silylthiol ester isomers by a thermodynamically driven tautomerization of the triorganosilicon group from sulfur to oxygen. Since thionoesters are significantly more reactive toward nucleophiles than thiol esters, the S-to-O silatropy and the attendant formation of oxoamides serves as a novel and mild *in situ* activation of the carboxyl function for nucleophilic addition. Though they did not evaluate the epimerization, they succeeded in the synthesis of dipeptides library.



Wu, W.; Zhang, Z.; Liebeskind, L. S. *J. Am. Chem. Soc.* **2011**, *133*, 14256

Figure 1-53. Thioacid coupling applying bistrimethylsilylacetamide (BSA).

1-11-5. Thioacid coupling with thio-FCMA

Wu and coworkers developed a novel amide bond formation reaction using isonitrile (Figure 1-54).[105] Their group had previously developed a microwave assisted isonitrile-mediated amidation from carboxylic acid. They applied isonitrile-mediated amidation to thioacid and achieved total synthesis of cyclosporine A. Both type 1 and type 2 amidations (Figure 1-54) are utilized in their report, allowing access to epimeric cyclosporin A from single amino thioacids. In type 1, the amide bond is fashioned when an exogenous amine intercepts the presumed formimidate carboxylate mixed anhydride (FCMA) intermediate. In this format, an isonitrile serves to activate thioacid. Use of a sterically encumbered "throwaway" isonitrile can be quite helpful in favoring the required bimolecular acylation pathway over internal S-N rearrangement, which would lead to an amide bond. In type 2, thioacid and isonitrile are combined to provide an N-formyl amide. The analogous transformation of thioacids into N-thioformyl amides is even more facile, proceeding at ambient temperature. Under suitable reduction protocols afforded their N-methylated counterparts.

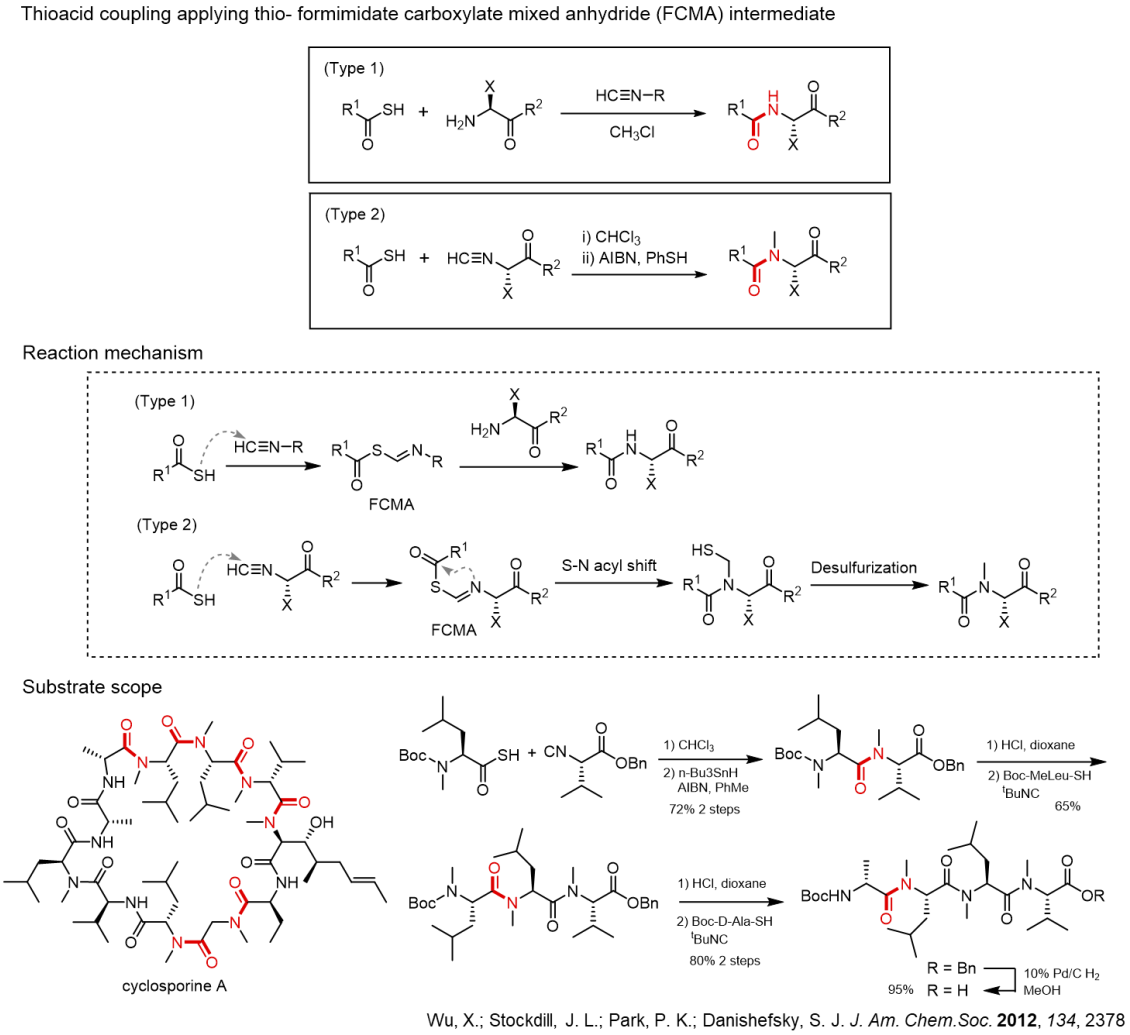
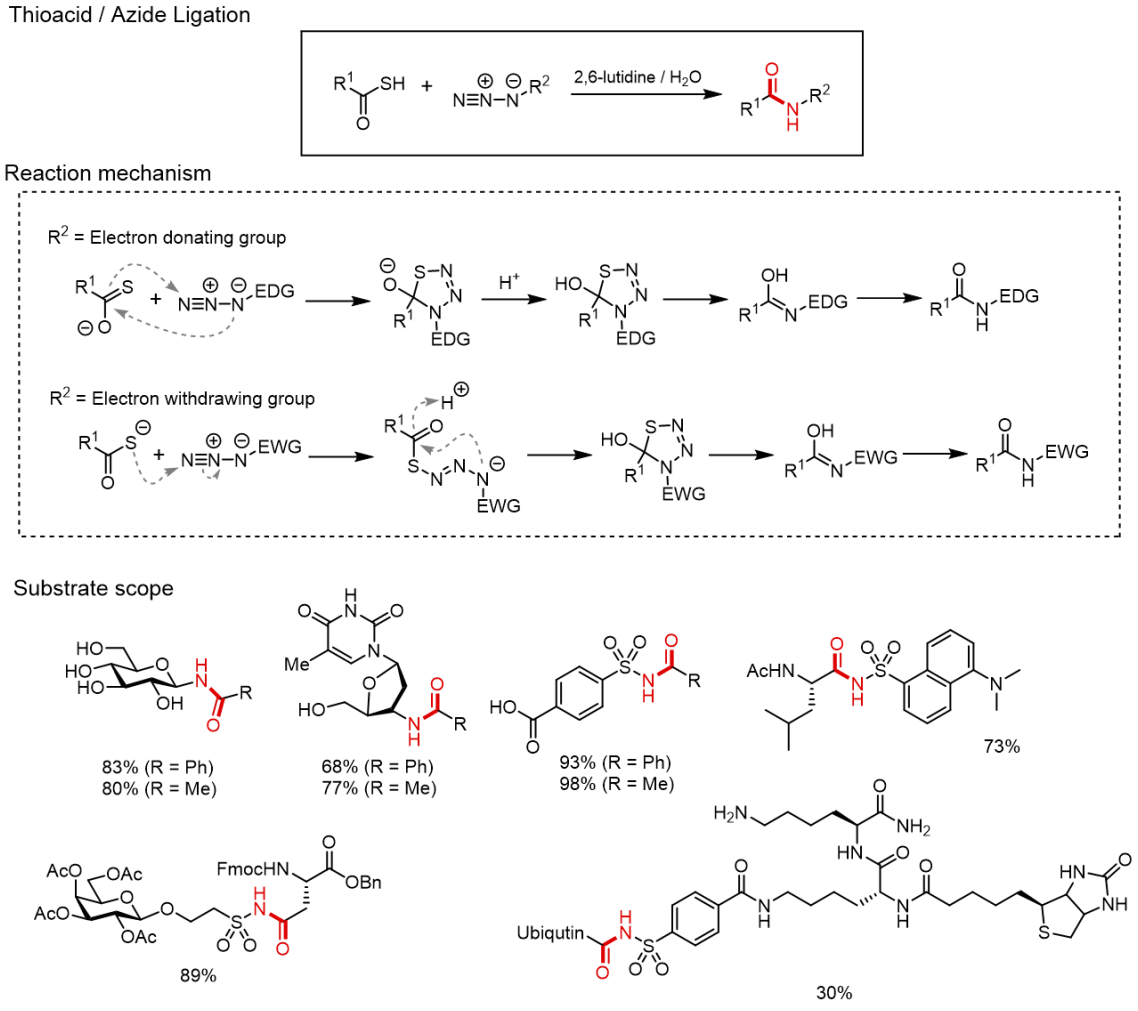


Figure 1-54. Thioacid coupling applying thio-FCMA.

1-11-6. Thioacid / Azide coupling

Williams groups reported an amide synthesis strategy based on a fundamental mechanistic revision of the reaction of thioacid and organic azide (Figure 1-55).[106] The reaction mechanism was later elucidated by their group.[107] For relatively electron-rich azides, the nitrogen-sulfur and nitrogen-carbon connectivity of the thiatriazoline intermediate is formed in a single step by anion-accelerated [3+2] cycloaddition. Subsequent protonation and loss of nitrous sulfide via retro-[3+2] cycloaddition gives the amide product. Highly electron-poor azides first form the nitrogen-sulfur bond to give a linear intermediate. In a separate step, formation of the nitrogen-carbon bond and protonation gives the thiatriazoline intermediate. Retro-[3+2] cycloaddition gives the amide product. Moreover, Rohmer and coworkers reported the application of this thioacid / azide coupling to not only syntheses of several classes of complex amides in nonpolar and water but also ubiquitin modification.[108]

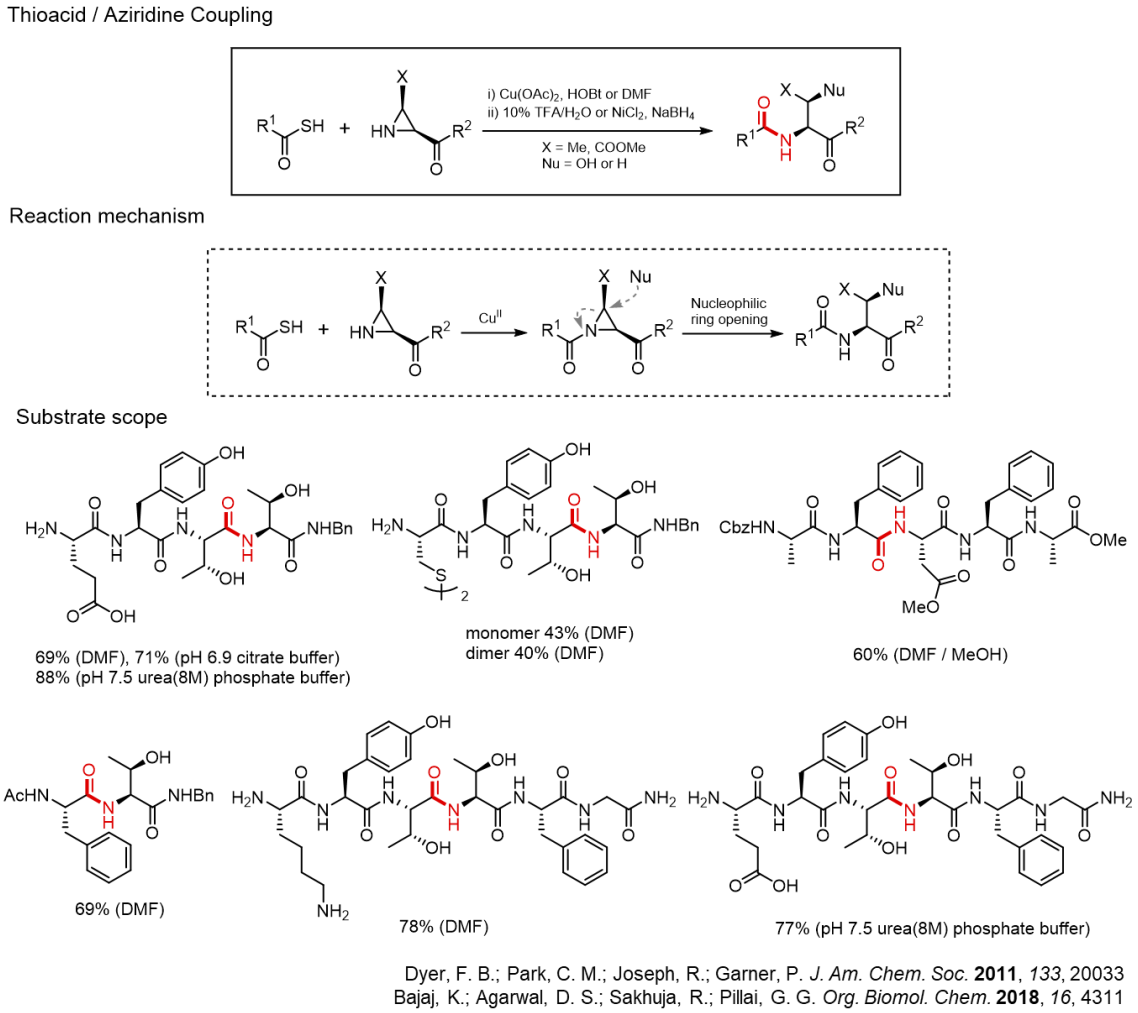


Shangguan, N.; Katukojvala, S.; Greenberg, R.; Williams. L. J. *J. Am. Chem. Soc.* **2003**, *125*, 7754
Kolakowski, R. V.; Shangguan, N.; Sauers, R. R.; Williams. L. J. *J. Am. Chem. Soc.* **2006**, *128*, 5695
Rohmer, K.; Mannuthodikayil, J.; Wittmann, V. *Isr. J. Chem.* **2015**, *55*, 437

Figure 1-55. Thioacid / Azide coupling.

1-11-7. Thioacid / Aziridine coupling

A synthesis of aziridine-containing peptides via the Cu(II)-promoted coupling of unprotected peptide thioacids and NH aziridine-2-carbonyl peptides was reported by Dyer and coworkers (Figure 1-56).[109] The unique reactivity of the resulting N-acylated aziridine-2-carbonyl peptides facilitates their subsequent regioselective and stereoselective nucleophilic ring-opening to give unprotected peptides that are specifically modified at the ligation site. The aziridine-mediated peptide ligation concept is exemplified using H_2O as the nucleophile, producing a Xaa-Thr linkage. The overall process is compatible with a variety of unprotected amino acid functionality, especially the N-terminal and Lys side chain amines. Later, this regioselective coupling applying active 3-member aziridine scaffold was extended from Thr to Glu by Bajaj and coworkers.[110] However, they examined coupling with only polar solvents except aqueous buffer. Although these reactions have not been applied to long chain peptides and examined side reactions such as epimerization, reactivity of thioacid and aziridine is highly selective for functional groups and might be applied to protein synthesis.

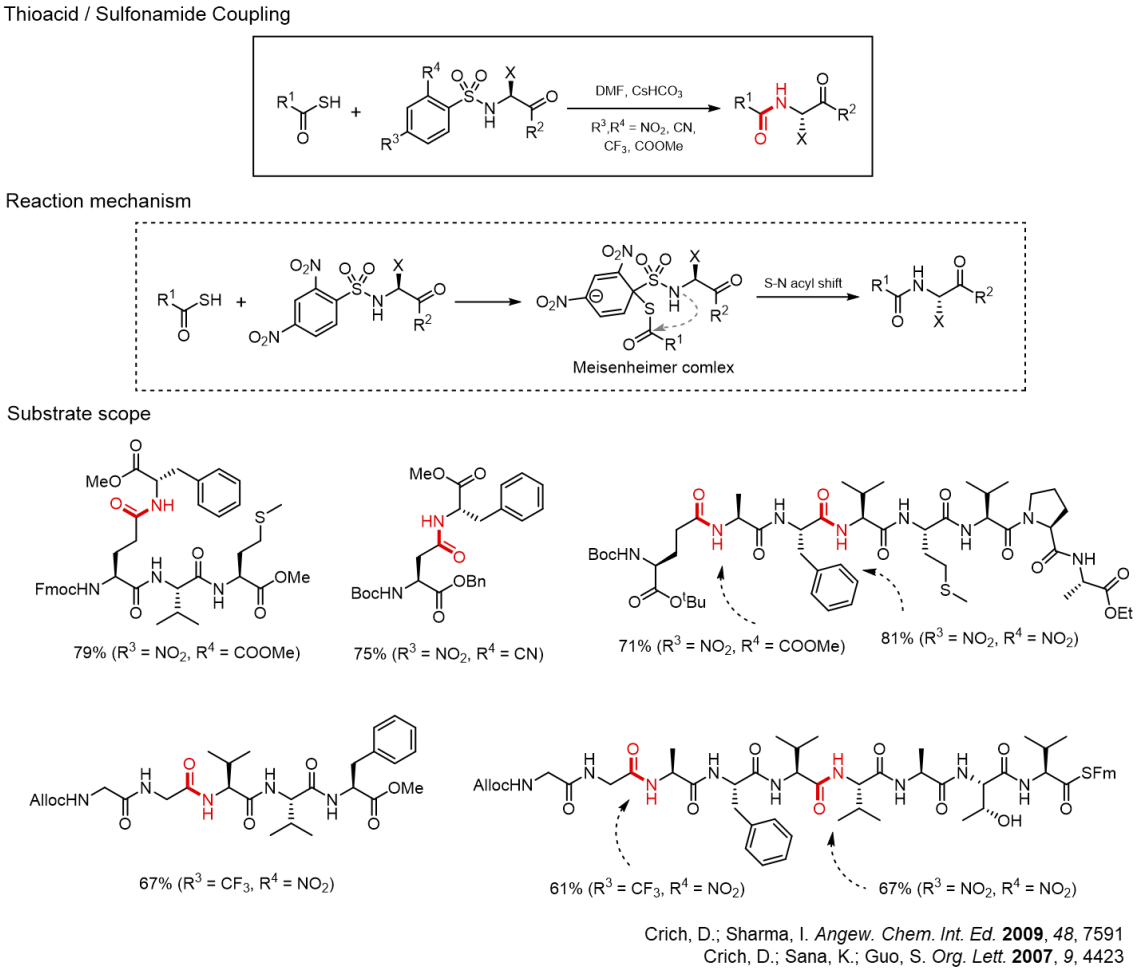


Dyer, F. B.; Park, C. M.; Joseph, R.; Garner, P. *J. Am. Chem. Soc.* **2011**, *133*, 20033
Bajaj, K.; Agarwal, D. S.; Sakhuja, R.; Pillai, G. G. *Org. Biomol. Chem.* **2018**, *16*, 4311

Figure 1-56. Thioacid / Aziridine coupling.

1-11-8. Thioacid / Sulfonamide coupling

Crich and coworkers reported thioacids reacted at room temperature in DMF in the presence of cesium carbonate with benzenesulfonamide with an electron-withdrawing substituent for example 2,4-dinitrobenzenesulfonamide (DNS) or 2-nitro-4-trifluorobenzenesulfonamide (FNS) to give amides (Figure 1-57).[111-112] The reaction mechanism is thought to be that the addition of the thioacid to the ipso-position of the benzenesulfonamide, resulting in an S-N acyl transfer. Such activation and elimination of sulfonamides by thiol is the similar reaction mechanism observed in 2-nitrobenzenesulfonamides (Ns-amides) reported by Fukuyama.[113] They reported the addition of cesium carbonate makes the reaction proceed more efficiently, and thioacid / sulfonamide coupling reaction was applied to the ligation of partially protected peptides. Nucleophilic activity of thioacids is different depending on the type of electron-withdrawing substituent of the benzenesulfonamide. For this reason, sulfonamide-based block synthesis of peptides was achieved using FNS and DNS-amide. By first reacting the highly electrophilic DNS peptide with a peptide thioacid in 67% yield, and finally reacting the less electrophilic FNS with a thioacid in 61% yield, the desired peptide was synthesized.

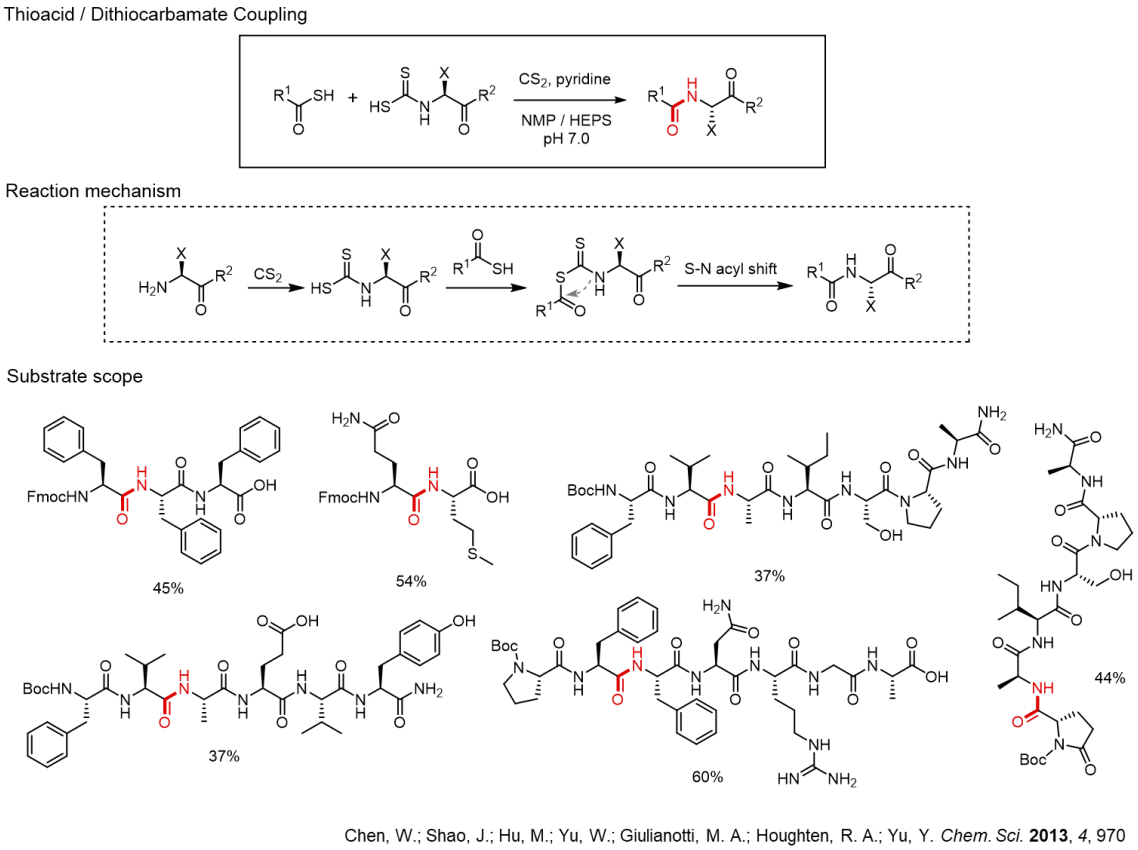


Crich, D.; Sharma, I. *Angew. Chem. Int. Ed.* **2009**, *48*, 7591
Crich, D.; Sana, K.; Guo, S. *Org. Lett.* **2007**, *9*, 4423

Figure 1-57. Thioacid / Sulfonamide coupling.

1-11-9. Thioacid / Dithiocarbamate coupling

A novel coupling of thioacids and dithiocarbamate of terminal amines was developed by Chen and coworkers (Figure 1-58).[114] This strategy is same manner of other coupling reactions with azide, aziridine and sulfonamide, which activates N-terminal amines before condensation. This approach enables the traceless removal of CS_2 to directly generate the desired amide bond and is compatible with a range of unprotected side chains of amino acid. Reaction starts with CS_2 and N-terminal amino group of the peptides in a buffer solution including pyridine to afford dithiocarbamate terminal peptides. Condensation of dithiocarbamates with thioacids forms highly active intermediates, and subsequent intramolecular S-N acyl shift occurs to form the desired amide bond. According to their substrate scope, even when functional groups such as guanidino group of Arg and amide group of Gln are present in peptides, reaction proceeds in a thioacid selectively. There is a possibility that the reaction can be applied to the chemical synthesis of long chain peptides by protecting the amino group although the yield is low (30~50%). Unfortunately, epimerization was not examined in detail. The ability to produce amide or peptides by a traceless removal of the auxiliary is a relatively significant virtue of the method rather than other thioacid coupling. Meanwhile, it was realized as applications of this new peptide-bond-forming reaction that the synthesis of novel endomorphin (EM) derivatives with various binding potencies.



Chen, W.; Shao, J.; Hu, M.; Yu, W.; Giulianotti, M. A.; Houghten, R. A.; Yu, Y. *Chem. Sci.* **2013**, *4*, 970

Figure 1-58. Thioacid / Dithiocarbamate coupling.

1-11-10. Thioacid / Isocyanate coupling

Crich and Sasaki reported that thioacids reacted at room temperature with isocyanates and isothiocyanates to give amide bonds in excellent yield (Figure 1-59).[115] The reaction begins with a nucleophilic attack of the thioacid on the *sp* carbon of the isocyanate, producing a thioisocyanate intermediate. This intermediate readily undergoes S-N acyl rearrangement with loss of COS or CS_2 to give the final product and subsequent removal of thiocarbonyl group under acidic conditions give the desired amide bond. They conveniently prepared thioacid by cleavage of 9-fluorenylmethyl with piperidine or trimethoxybenzyl thioesters with TFA / triethylsilane and examined coupling reactions with thioacids and isocyanate to afford neoglycopeptides. They reported that both aromatic and aliphatic isocyanates well accepted the nucleophilic attack of thiol (function correctly in this chemistry) and many exapmles showed modest yields. In addition to this amidation generates symmetrical urea as a major byproduct. This reaction has been tested only in organic solvents and was not applied to reactions in aqueous solutions. The rate of epimerization is also unknown, making it difficult to apply the reaction to the chemical synthesis of peptides and proteins. However, the reaction mechanism of the thioacid is unique. The key intermediate in this reaction has a similar structure to that of isonitrile-condensation reaction developed by Danishfsky and coworkers.[105]

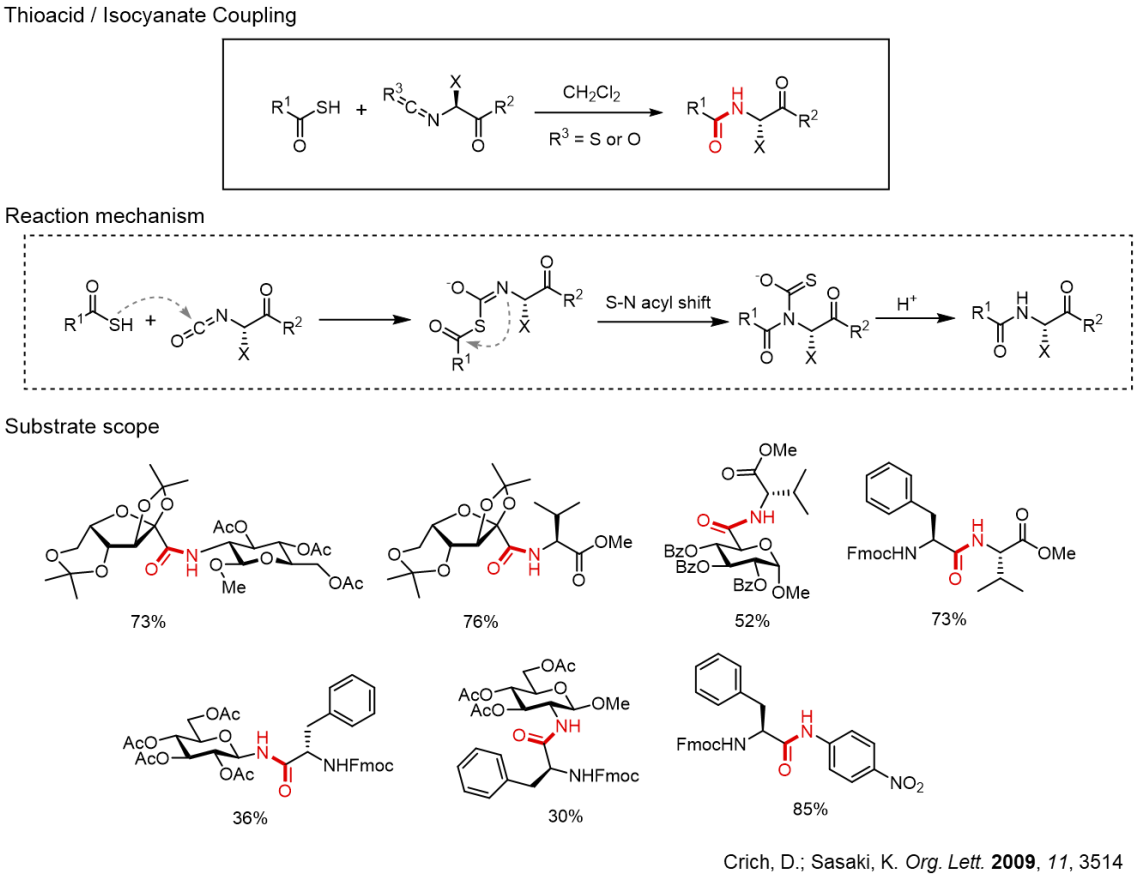
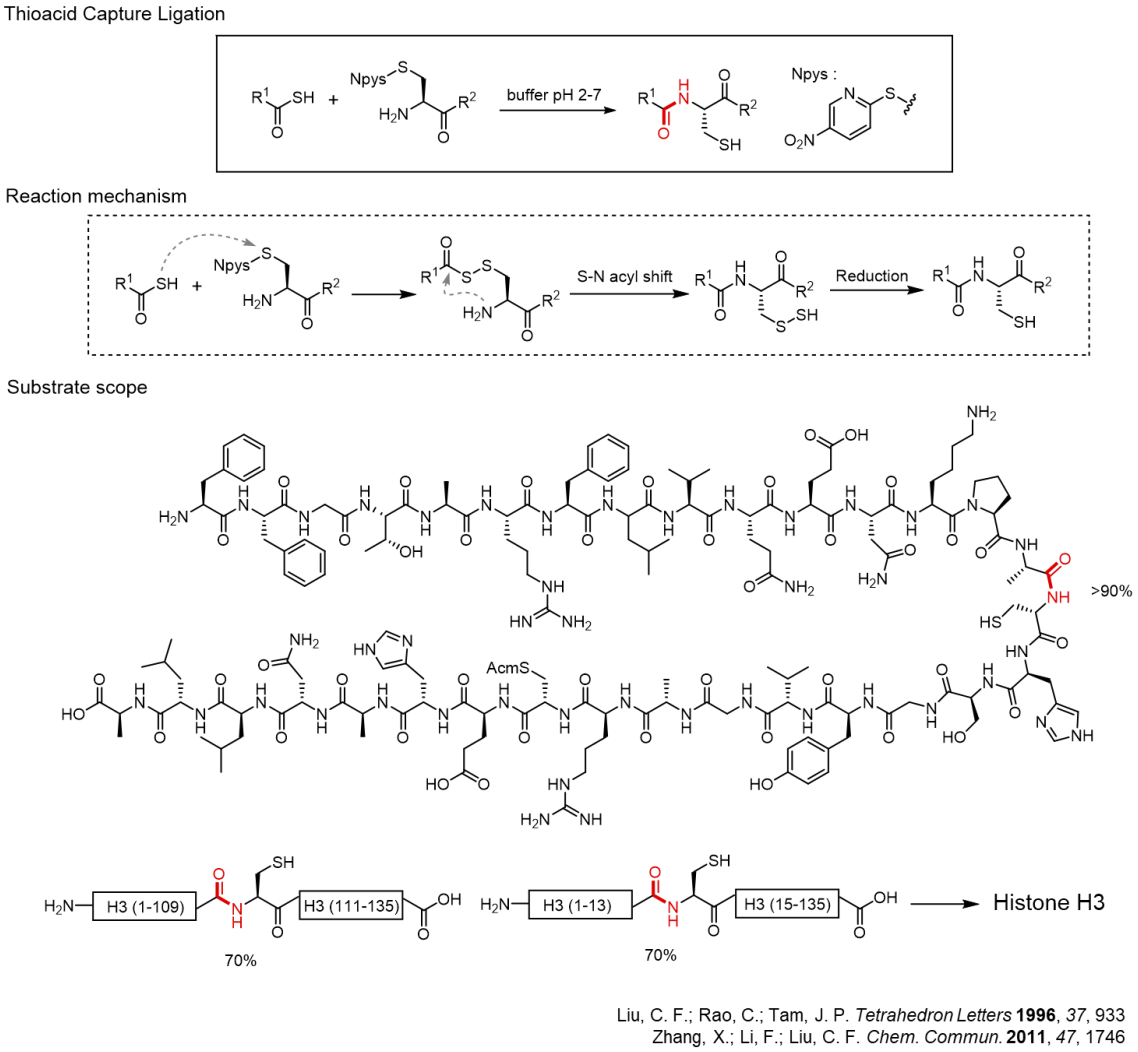


Figure 1-59. Thioacid / Isocyanate coupling.

1-11-11. Thioacid Capture Ligation (TCL)

Liu and coworkers have developed a novel thioacid capture ligation (TCL), which couples a cysteinyl peptide of which thiol was activated by nitropyridyldisulfide (Npys) group and an unprotected peptide thioacid (Figure 1-60).[116] Reaction starts from disulfide exchange reaction of Npys-thiol with thioacid in a buffer (pH 2-7). Key of this ligation is a highly efficient capture reaction between a C-terminal thioacid of peptide and an activated Npys-disulfide on the N-terminal Cys of counter peptide to form an acyl disulfide intermediate which then undergoes rapid intramolecular S-N acylation to generate an amide bond. The Cys residue at the ligation site can be converted into alanine by the reductive desulfurization, indicating this reaction can be performed at either cysteine or alanine sites. TCL is a highly chemoselective ligation that can be used to afford peptides without using a protecting group. They succeeded in synthesizing peptides (32aa). TCL can be also applied to expressed peptides by *E.coli* and Liu groups achieved the total synthesis of Histone3 in two steps by using TCL.[117]

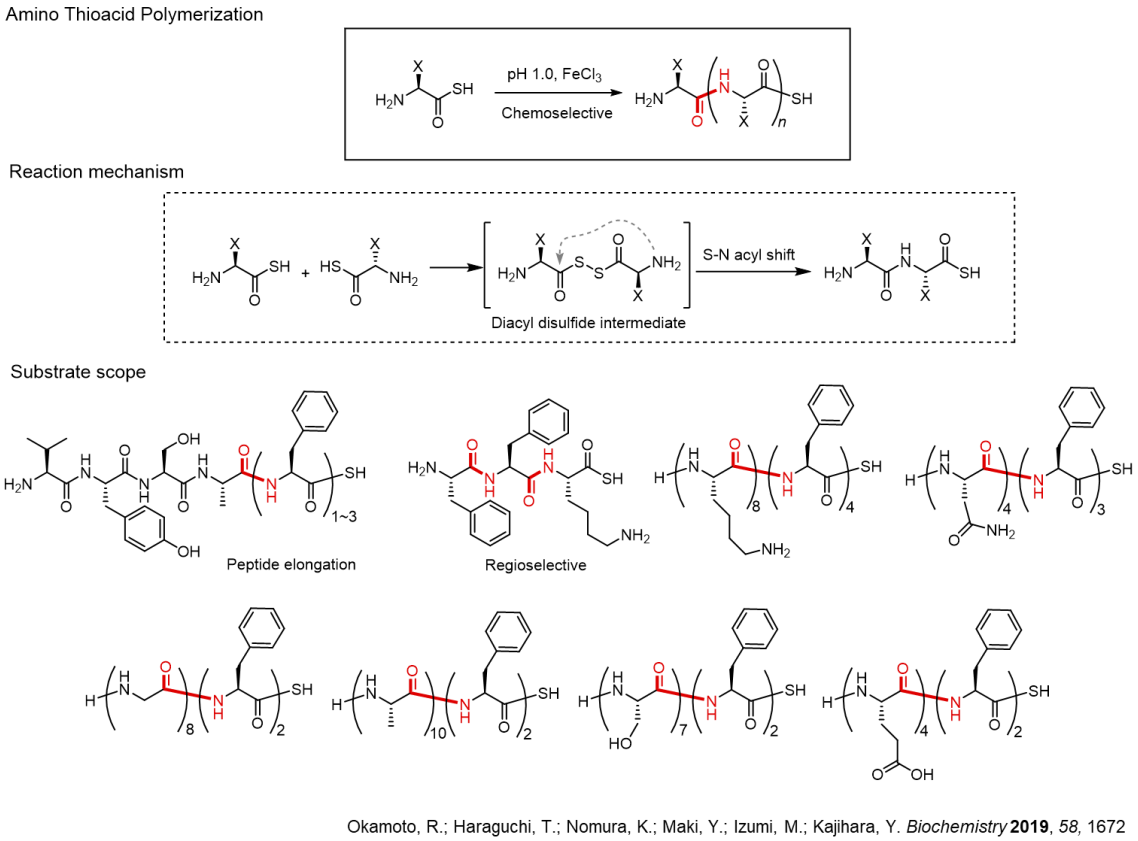


Liu, C. F.; Rao, C.; Tam, J. P. *Tetrahedron Letters* **1996**, *37*, 933
Zhang, X.; Li, F.; Liu, C. F. *Chem. Commun.* **2011**, *47*, 1746

Figure 1-60. Thioacid Capture Ligation (TCL).

1-12. Development of the new chemoselective amide formation reaction

In the previous chapter, it was shown that thioacids are applied in many chemoselective amide bond formations. Our group also studied chemoselective amide bond forming reactions based on thioacid.[118-119] First, we found that moderate oxidation of α-amino thioacids with iron (II) oxide induces selective α-amide bond formation and polymerization (Figure 1-61).[118] This polymerization reactions progress with nonprotected amino thioacids via oxidation processes in acidic solutions. The thioacid group is oxidized to form a diaminoacyl disulfide intermediates, which is a potent acyl donor. This diacyl disulfide bond then yields polypeptides via intramolecular S-N acyl transfer. The diacyl disulfide formation of thioacids has also been reported by several groups. This amidation through diaminoacyl disulfide intermediate is highly chemoselective and succeeded to synthesize polypeptides including twelve residues and even the thioacid form of lysine, which has a free ε-amino group, generated a regioselective α-peptide bond. Interestingly, this oligomerization reaction proceeded even in the presence of hematite (Fe_2O_3) a prebiotic ore, thus suggesting a plausible prebiotic peptide bond forming reaction. A rational α-peptide bond formation, without protecting groups, is difficult to achieve in chemical synthesis. In this background this polymerization can also become the key reaction for the alternative rapid peptide synthesis method.

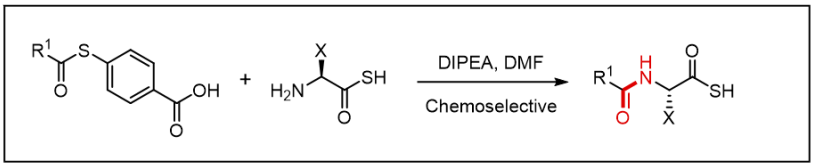


Okamoto, R.; Haraguchi, T.; Nomura, K.; Maki, Y.; Izumi, M.; Kajihara, Y. *Biochemistry* **2019**, *58,* 1672

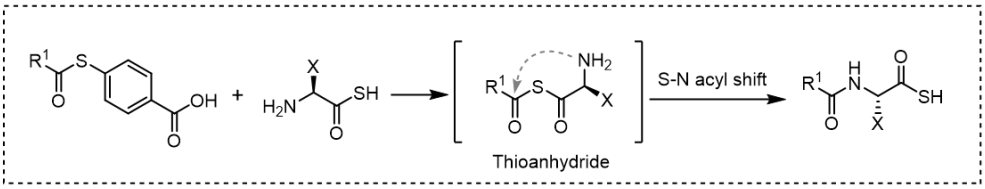
Figure 1-61. Amino thioacid polymerization.

In addition to polymerization, I successfully developed an alternative chemoselective amide bond formation utilizing amino thioacid coupling (ATC) [119], which couples an α-amino group of an amino thioacid to the C-terminus of an unprotected counter peptide aryl thioester (Figure 1-62). This reaction potentially proceeds through two reaction steps: a thioacid-thioester exchange and a subsequent intramolecular S-N acyl transfer reaction. The first step is analogous to thiol-thioester exchange reaction of NCL. In the case of ATC, the nucleophilic addition of the thioacid to thioester gives thioanhydride intermediate. These active intermediates induce subsequent intramolecular S-N acyl transfer reaction resulting in an α-peptide bond via a plausible 5-membered ring transition state. By using ATC, I successfully demonstrated the coupling of amino acids including nonproteinogenic amino acids without protecting groups as well as coupling reagents. After optimization of reaction conditions, The ATC reactions progress with peptide aryl thioester having 4-methylbenzenethiol (MBT) and 4-mercaptobenzoic acid (MBA) in the highest yield. They performed ATC using other amino thioacids including a nonproteinogenic amino acid, propargylglycine (PAG). These experiments revealed that all ATC reactions successfully gave the desired products in good yields.

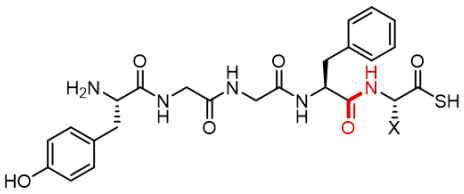
Amino Thioacid Coupling (ATC)



Reaction mechanism



Substrate scope



entry	R (Xaa)	product	Yield*2 (%)
1	H (Gly)	YGGF-G-SH	97
2	CH_3 (Ala)	YGGF-A-SH	92
3	$CH(CH_3)_2$ (Val)	YGGF-V-SH	91
4	CH_2OH (Ser)	YGGF-S-SH	91
5	$CH(CH_3)OH$ (Thr)	YGGF-T-SH	89
6	$(CH_2)_2SCH_3$ (Met)	YGGF-M-SH	88
7	CH_2Ph (Phe)	YGGF-F-SH	83
8	CH_3PhOH (Tyr)	YGGF-Y-SH	78
9	$CH_2CH(CH_3)_2$ (Leu)	YGGF-L-SH	73
10	$(CH_2)_2COOH$ (Glu)	YGGF-E-SH	68
11	CH_2CCH (PAG)	YGGF-Pag-SH	43

Okamoto, R.; Nomura, K.; Maki, Y.; Kajihara, Y. *Chem. Lett.* **2019**, *48*, 1391

Figure 1-62. Amino Thioacid Coupling (ATC)

The author of this thesis thought that amino thioacids are essential functional groups for the chemoselective formation of amide bonds. In particular, the intramolecular S-N acyl shift described above is a key reaction to realize the chemical insertion of glycans as described in the previous section. Besides promoting terminal amino group and carboxylic acid selective reactions, such intramolecular rearrangement using thioacids can also suppress epimerization by passing through a mild activation intermediate. In this background based on the amide bond formation reaction using amino thioacids, the author of this thesis has developed a new method for glycoprotein synthesis (Figure 1-63), which will be discussed in *chapter 1-13*.

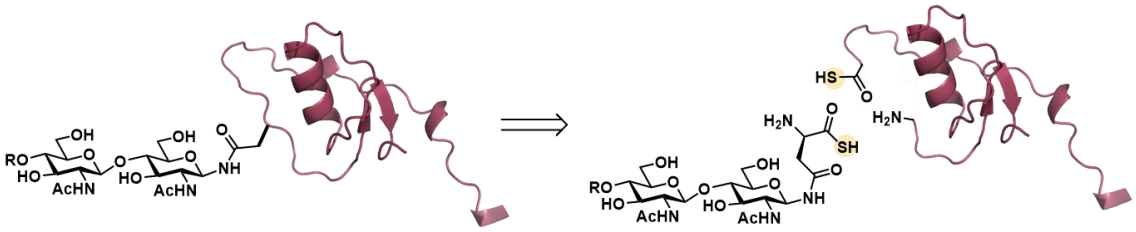


Figure 1-63. Chemical glycan insertion strategy for the synthesis of homogeneous glycoprotein.

1-13. Chemical glycan insertion for the synthesis of glycoprotein

By studying amino thioacid polymerization, I could control polymerization reactions and found an efficient amide formation reaction: diacyl disulfide coupling (DDC) (Figure 1-64 B). DDC could form an α-amide bond with a glycosyl asparagine thioacid and a peptide α-thioacid via a disulfide bond intermediate without any polymerization reactions. This DDC enabled us to set up a highly convergent glycoprotein synthesis strategy, chemical glycan insertion, that can couple two peptides with the N- and C-termini of glycosyl asparagine thioacid. I employed glycosyl asparagine thioacid as the junction point for the coupling of N- and C-terminal peptides. The first coupling is DDC between a peptide thioacid and glycosyl asparagine thioacid. Because the resultant glycopeptide has a thioacid form at its C-terminus, I could apply thioacid capture ligation (TCL) for the coupling of the resultant glycopeptide thioacid and another peptide having a disulfide functional group at its N-terminus to afford the full-length glycoprotein backbone (Figure 1-64 C). This route enables to assemble the entire glycoprotein backbone in two steps, chemical glycan insertion strategy (Figure 1-64 A).[120] In addition, valuable glycosyl amino thioacids can be used at the late stage of the synthesis of glycoproteins.

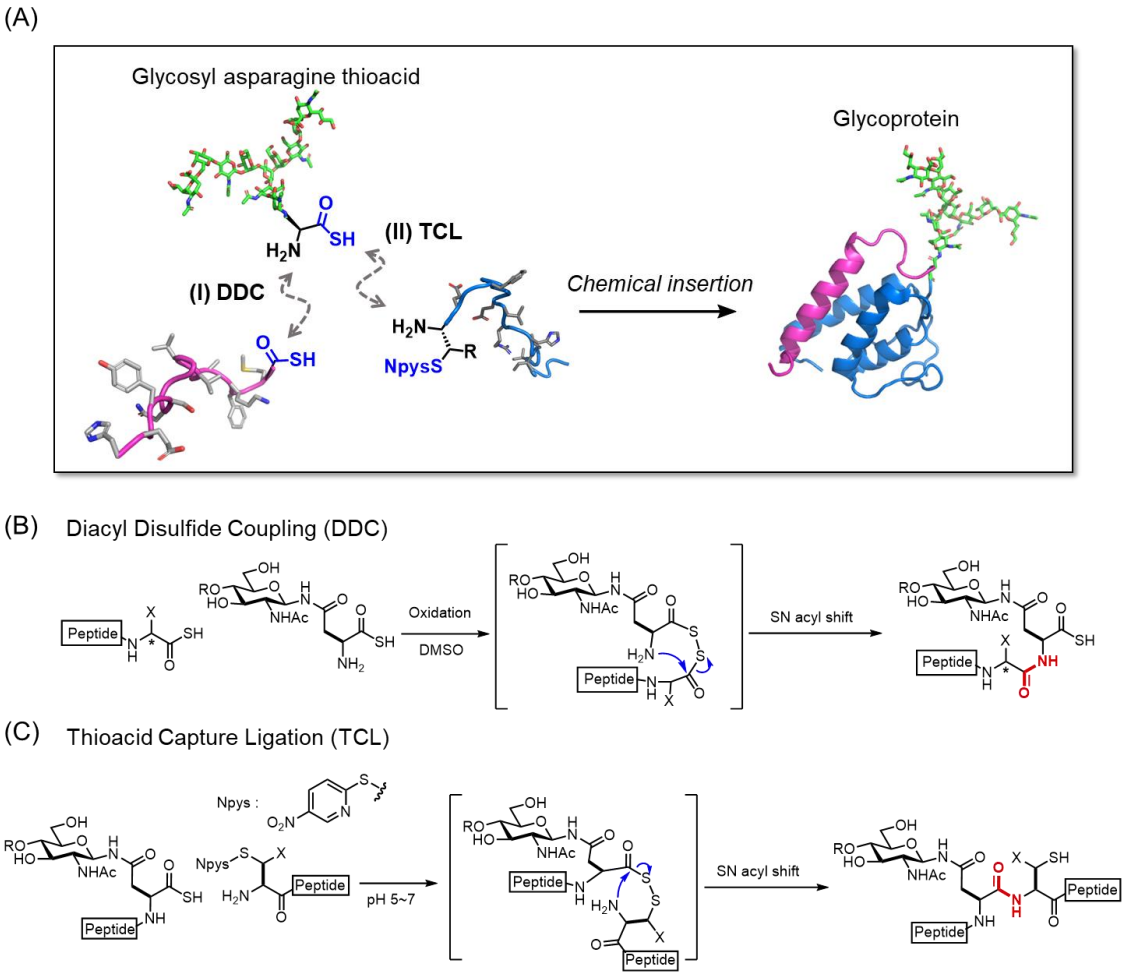


Figure 1-64. Chemical glycan insertion for the synthesis of glycoprotein. (A) General synthetic scheme (B) Diacyl disulfide coupling (DDC) and the mechanism. (C) Thioacid capture ligation (TCL) and the mechanism.

1-14. Contents of the thesis: the aim of this research

Herein, I would describe the characterization of diacyl disulfide activation of amino thioacids and the efficient ability of glycosyl asparagine thioacid to be used for robust glycoprotein semisynthesis. To demonstrate my strategy, I examined the total syntheses of two cytokine glycoproteins, namely, CC chemokine ligand 1 (CCL1) and interleukin 3 (IL3) and serine protease inhibitor Kazal type 13 (SPINK13) having homogeneous biantennary sialyloligosaccharides. In particular, the synthesis of IL3 could employ a long-chain polypeptide expressed in E. coli. This dissertation consists of five chapters described below.

Chapter 1.
Introduction of the synthesis of glycoproteins. Current method to synthesize glycoproteins having homogeneous glycan structures and problems were described in this chapter. The chemoselective amide bond formations applying the functional groups, thioacids, were also introduced for the development of novel coupling reaction to synthesize glycoproteins.

Chapter 2.
I describe the development of diacyl disulfide coupling (DDC), which form an α-amide bond with an α-amino thioacid and a peptide thioacid in an oxidative condition. In this chapter, the syntheses of glycopeptides have been also achieved by the chemical glycan insertion strategy applied DDC.

Chapter 3
This chapter describes that total synthesis of CCL1 carrying a sialyl biantennary complex-type oligosaccharide, which is a chemotactic cytokine expressed by cells that induce directional movement of leukocytes, by the developed synthetic strategy.

Chapter 4
This chapter describes that semisynthesis of IL3 carrying a sialyl biantennary complex-type oligosaccharide, which is a cytokine produced by T cells as a regulator of hematopoiesis. Furthermore ,bioassays were also performed using the obtained glycosyl IL3 and non-glycosyl IL3.

Chapter 5
This chapter describes that total synthesis of SPINK13 carrying a complex-type oligosaccharide, which is a secreted protein obtained from the hepatoblastoma cell line. DDC with prolyl thioacids and bioassays using synthesized glycosylated SPINK13 were also performed in this chapter.

Chapter 2. Novel synthetic strategy for the glycoprotein: Development of DDC

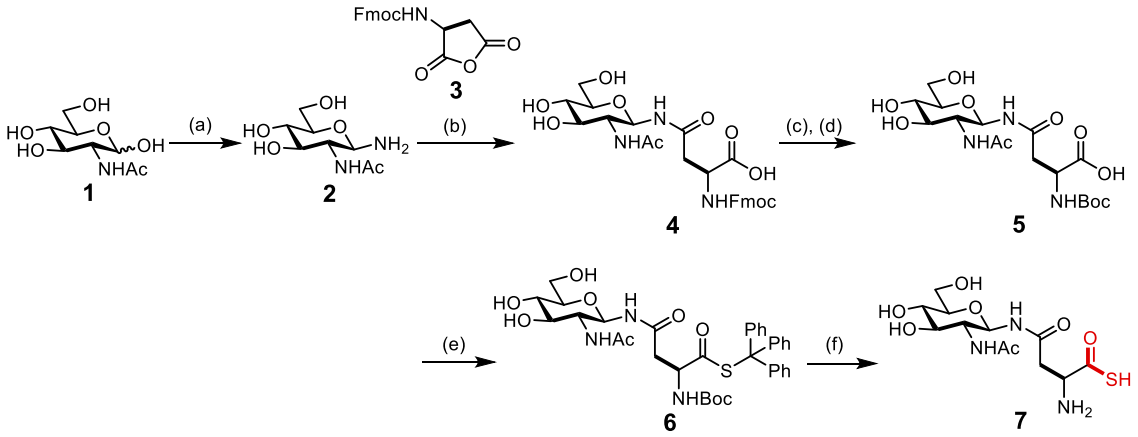
2-1. Introduction of chapter 3

In *chapter 2*, I describe the reaction development of DDC, which form an α-amide bond with an amino α-thioacid and a peptide thioacid in an oxidative condition. The thioacid group is oxidized to form a diacyl disulfide bond, which is a potent acyl donor. This diacyl disulfide bond then yields an amide bond *via* intramolecular S-N acyl transfer. DDC is a robust α-amide bond formation reaction in that it progresses highly chemoselectively without any condensation reagents and the products have no epimerization. This efficient coupling reaction enabled to develop a new glycoprotein synthesis method, chemical insertion strategy, which can insert glycosyl asparagine thioacids between two unprotected peptides. In this *chapter 2*, I will discuss the development of DDC and chemical insertion strategy by utilizing DDC toward the convergent synthesis of glycoproteins.

2-2. Synthesis of glycosyl asparagine thioacid carrying sialyl complex saccharide and GlcNAc

I commenced from the synthesis of glycosyl asparagine thioacids derivatives such as carrying mono-N-acetylglucosamine, GlcNAc **7** (Scheme 2-1) and sialyl complex type oligosaccharide **11** (Scheme 2-2) for the development of DDC. Glycosyl asparagine thioacids carrying mono-GlcNAc **7** was prepared by from GlcNAc **1**. The synthesis of a precursor β-glycosylamine **2** by reported procedure by Sakata and coworkers[121], using methanolic ammonia and 1.0 equivalent of ammonium bicarbonate and stepwise evaporation *in vacuo* (61% yield). For the synthesis of N-Fmoc-(β-GlcNAc)-L-asparagine **4** the procedure reported by Selivanov and coworkers[122] using a N-Fmoc-aspartic anhydride **3** and glycosyl amines **2** was applied to glycosyl amines without protecting groups. β-glycosyl amine **2** was converted to N-Fmoc-(β-GlcNAc)-L-asparagine **4** with N-Fmoc aspartic anhydride **3** in DMSO (51%) in β-selective manner regarding the position of aspartic anhydride. The β-regioselectivity of a nucleophilic attack on the aspartic anhydride was induced by polarity of solvents. After converting protecting groups from the Fmoc groups into the Boc groups, the condensation of triphenyl methane thiol to N-Boc-(β-GlcNAc)-L-asparagine **5** was performed with PyBOP and DIPEA at -15 °C to give compound **6** (22% yield). Then, deprotection of both the Boc group and triphenyl methane thiol groups were performed with trifluoroacetic acid (TFA) and triisopropylsilane (TIPS) yielded the desired glycosyl asparagine thioacid carrying mono-GlcNAc **7** (95% yield). The structure of compound **7** was confirmed by nuclear magnetic resonance (NMR) spectroscopy and high-resolution mass spectrometry (HRMS).

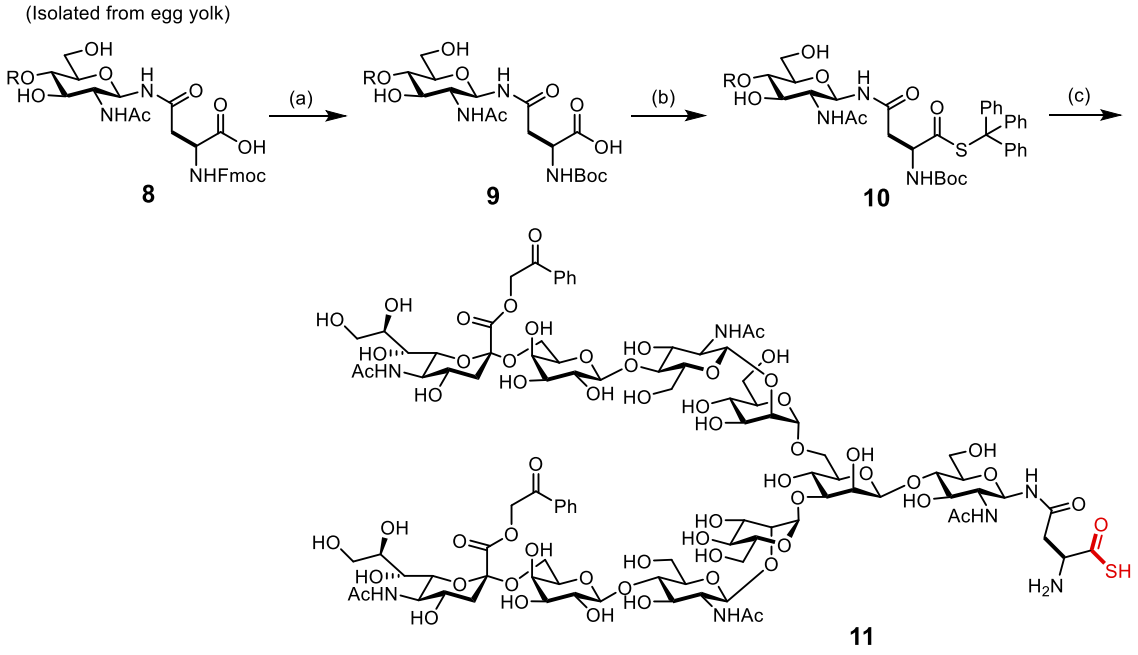
Scheme 2-1. The synthesis of glycosyl asparagine thioacid carrying mono GlcNAc



Reaction conditions; (a) NH_3-MeOH, $(NH_4)_2CO_3$, 3 d, 40°C, 61%. (b) **3**, DMSO, 3 h, rt. 51%. (c) piperidine, DMF, 30 min, rt. (d) Boc-OSu, DIPEA, DMF, 12 h, rt. 74% (2 steps). (e) PyBOP, DIPEA, triphenyl methane thiol, 1 h, -15 °C, 22%. (f) TFA, TIPS, 10 min, rt. 95%.

Glycosyl asparagine thioacids carrying sialyl biantennary complex type oligosaccharide **11** was prepared by from the Asn linked decasaccharide building block, Boc-Asn(sialylglycan)-OH **8** (Scheme 2-2). The substrate N-Fmoc-Asn-(sialylglycan)-OH **8** was prepared from the egg yolk followed with the reported method by Seko and coworkers[23]. After converting protecting groups from the Fmoc groups into the Boc groups and phenacyl protection of carboxylic groups of sialic acids by the procedure reported by Murakami and coworkers[78], the condensation of triphenyl methane thiol to N-Boc-Asn-(sialylglycan)-OH **9** was performed with PyBOP and DIPEA at -15 °C to afford compound **10** (57% yield). Then, deprotection of both the Boc group and triphenyl methane thiol groups were performed with TFA and TIPS yielded the desired glycosyl asparagine thioacid carrying sialyl biantennary complex type oligosaccharide **11** (92% yield). The structure of compound **11** was confirmed by nuclear magnetic resonance (NMR) spectroscopy and high-resolution mass spectrometry (HRMS).

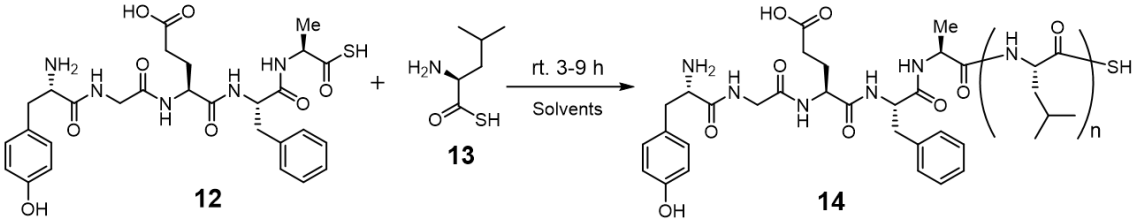
Scheme 2-2. The synthesis of glycosyl asparagine thioacid with complex-type sialyl oligosaccharide



Reaction conditions: (a) piperidine, DMF, 30 min, rt. (b) Boc-OSu, DIPEA, DMF, 12 h, rt. 40% (2 steps). (c) PyBOP, DIPEA, triphenyl methane thiol, 3 h, -15 °C, 57%. (d) TFA, TIPS, 10 min, rt. 92%.

2-3. Development of DDC and substrate scope

To realize glycopeptide synthesis by DDC, I commenced this study to understand the characteristic nature of thioacid polymerization under neutral conditions[118]. Our previous polymerization reactions with the oxidating conditions were progressed in acidic aqueous solutions (pH ~ 1.0) in the presence of Fe(III), and these reactions converted α-amino thioacids to polypeptides in moderate yield. Then, there were two problems in applying this reaction to the synthesis of glycopeptides. The first is that acidic aqueous conditions are not suitable for the coupling reaction of a glycosyl asparagine thioacid carrying sialyl complex type oligosaccharide because a sialyl linkage is labile in an acidic aqueous condition. The second is that the elongation of the polymerization reactions cannot be strictly controlled. Our focused reaction involves the polymerization of α-amino thioacids, not the condensation reaction of the desired amines with the polypeptides. Since our goal was to condense one glycosyl asparagine thioacid to a peptide thioacid, I needed to suppress the polymerization as much as possible. However, the lack of progress in the polymerization reaction simultaneously means that the yield of the condensation reaction is low. Therefore, I first performed optimization with several amino thioacids to suppress polymerization reactions under neutral conditions. Firstly, we optimized solvents and additives (Figure 2-2). To a solution of peptide thioacid (15.0 mM) in various solution was added leucine thioacid **13** (45.0 mM, 3.0 equiv.) at rt. and the resultant mixture was stirred.

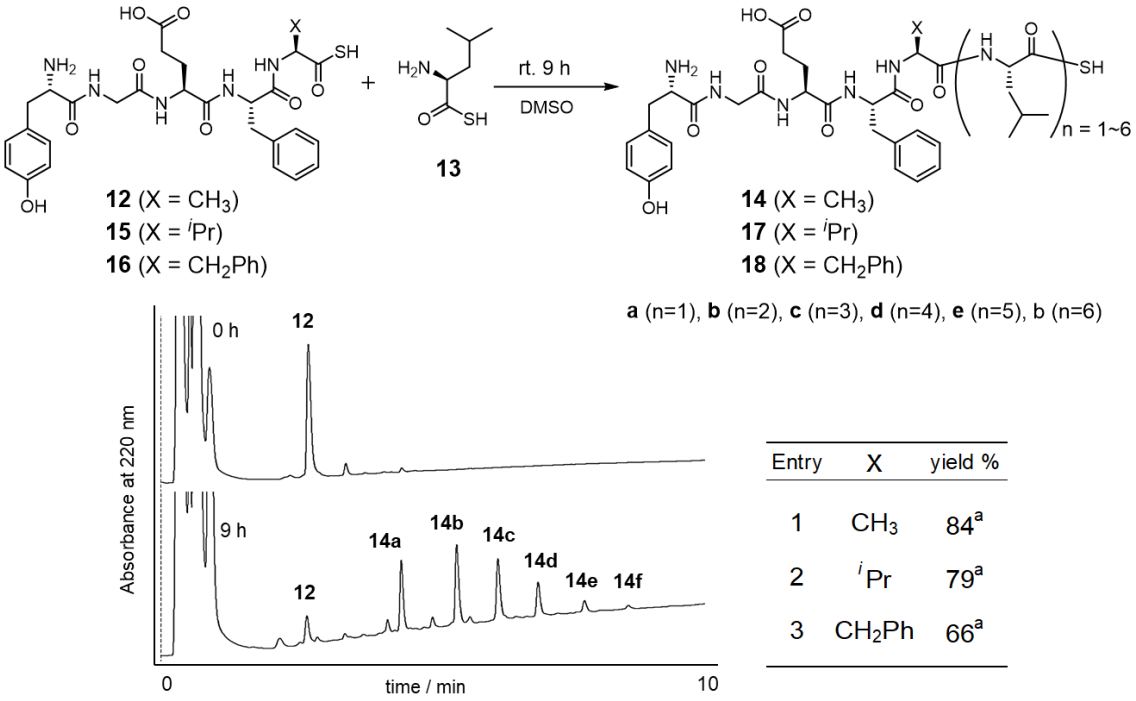


entry	solvent	additive	time / h	conversion yield (%) n = 1	2	3	4
1	6M Gdn-HCl, 0.2 M Sodiumphosphate	$FeCl_3$ aq (pH 1.5)	6	14	1	-	-
2	DMF	-	9	5	18	19	15
3	TFE	-	9	39	26	9	2
4	DMSO	-	9	9	21	26	21
7	DMSO	CAN	9	27	25	13	5
8	DMSO	DTT	3	-	-	-	-
9	DMSO	VA-044	9	26	29	17	8

Reaction conditions; Peptide thioacid **12** (15.0 mM), leucine thioacid **13** (45.0 mM, 3.0 equiv.) rt.

Figure 2-2. Diacyl disulfide coupling (DDC) in various organic solvents.

When we used aqueous conditions for the reactions, polymerization or decomposition were major results; there-fore, we used organic solvents for subsequent studies. With these optimizations, the reactions were found to be feasible in DMSO (Figure 2-2, entry 2-4). In addition, when DTT was added to DMSO, the reactions did not yield products (Figure 2-2, entry 8). These results indicated that disulfide bond formation seemed to be an essential process. Since the yield of the condensation reaction in DMSO without additives was the highest, then we examined the reaction conditions by varying amino acid sequences of peptide thioacids in DMSO (Figure 2-3). To a solution of peptide thioacid **12**, **15** and **16** which has methyl groups CH_3, *iso*-propyl groups ^{i}Pr and benzyl groups CH_2Ph respectively (15.0 mM) in DMSO was added leucine thioacid **13** (45.0 mM, 3.0 equiv.) at rt. and the resultant mixture was stirred for 9 hours.



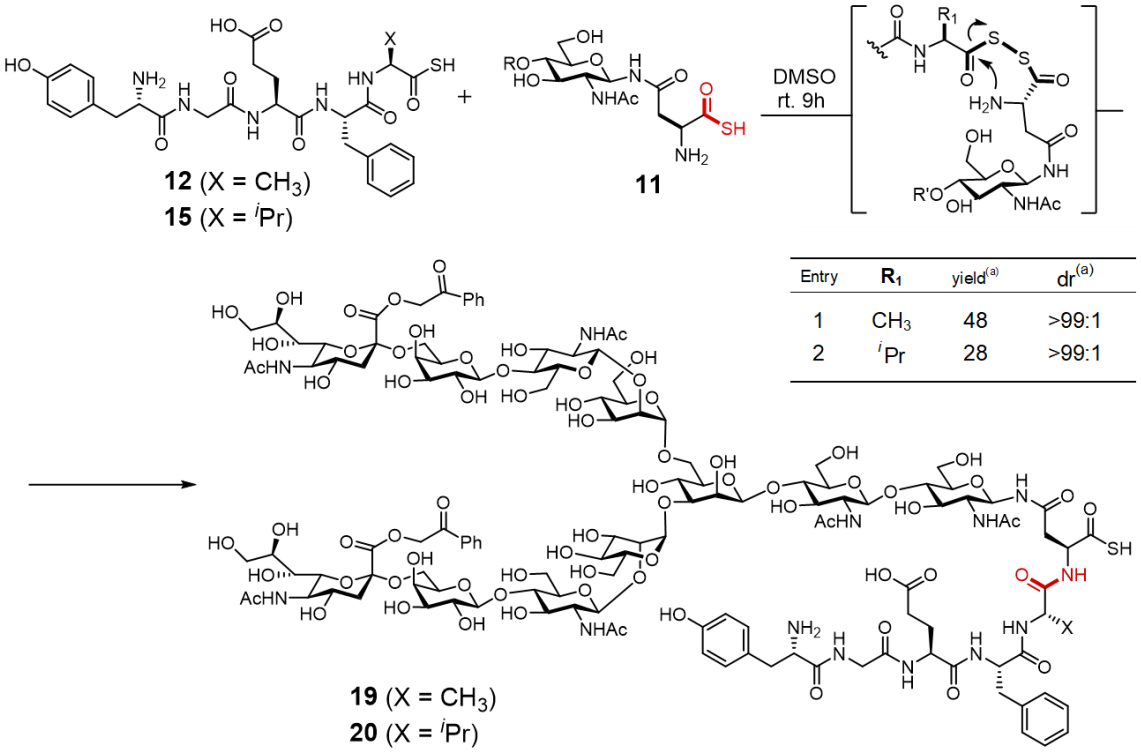
Entry	X	yield %
1	CH_3	84[a]
2	^{i}Pr	79[a]
3	CH_2Ph	66[a]

Reaction conditions; Peptide thioacid **12**, **15** and **16** (15.0 mM), leucine thioacid **13** (45.0 mM) rt.

Figure 2-3. Diacyl disulfide coupling (DDC) with having varied C-terminal amino acids.

Consequently, although we examined the reaction conditions by varying the amino acid sequences of peptide thioacids and solvents, polymerization was the major result, and we did not find the conditions, in which the extension of mono amino acids is a major product. We then examined this reaction for the glycosyl asparagine thioacids having biantennary sialyl complex type oligosaccharide **11**. We hoped that glycan motif might prevent amino thioacids from forming diacyl disulfide intermediate and suppress the polymerization of glycosyl asparagine thioacid having saccharide.

When we employed glycosyl asparagine thioacid **11** and peptide thioacid **12** for DDC, surprisingly, the desired glycopeptide thioacid formed chemoselectively without any polymerization. This reaction employed two equivalents of glycosyl asparagine thioacid **11** toward peptide thioacid **12** (one equivalent) with Ala at the C-terminus (Figure 2-4). The reaction velocity was slow, but the desired monoglycosylated peptide thioacid **19** formed as a major product after 9 h. A longer reaction time did not improve the yield, and the decomposition of glycosyl asparagine thioacid **11** was observed as a major result after 24 h. In terms of this unexpected result, we considered that the steric hindrance of glycans slowed the polymerization reaction. It is also interesting that DDC is thioacid selective and produces α-amide bonds even in the presence of free hydroxy groups of glycans, amine and carboxyl groups of peptides. The coupling reactions of glycosyl asparagine thioacid **11** were then examined with another peptide thioacid **15** having a bulky Val at the C-terminus. Although the yield was decreased, α-amide bond formation proceeded with glycosyl asparagine thioacid **11**, and the reaction did not yield undesired polymerization. Extensive optimizations indicated that 2.0 equiv of glycosyl asparagine thioacid **11** combined with peptide thioacid **15** showed moderate yields.



Entry	R_1	yield[a]	dr[a]
1	CH_3	48	>99:1
2	^{i}Pr	28	>99:1

Figure 2-4. Diacyl disulfide coupling (DDC) with glycosyl asparagine thioacid.

Thus, it has been shown that DDC proceeds well with glycosyl asparagine thioacids in the thioacid selectively even in the presence of other functional groups. Such functional group selectivity indicates that the reaction can be applied to longer polypeptide chains.

2-4. Proposed reaction mechanism

The proposed mechanism of DDC is shown in Figure 2-5. In DMSO both of thioacids including peptide α-thioacids (**A**) and glycosyl asparagine thioacids (**B**) are oxidized to form a diacyl disulfide intermediates (**C**), which is a potent acyl donor. The diacyl disulfide formation of thioacids has also been reported by several groups[123-124]. Through a spontaneous intramolecular rearrangement (1,5 S-N acyl transfer), a bifunctional α-amine of a glycosyl asparagine thioacid attack the carbonyl carbon and the intermediate (**C**) yields a glycosylated peptide α-dithioacid (**D**) including α-amide bonds. Finally, reduction of glycosylated peptide dithioacid (**D**) afforded the desired final glycosylated peptide thioacid (**E**) with loss of H_2S gas using DTT or TCEP.

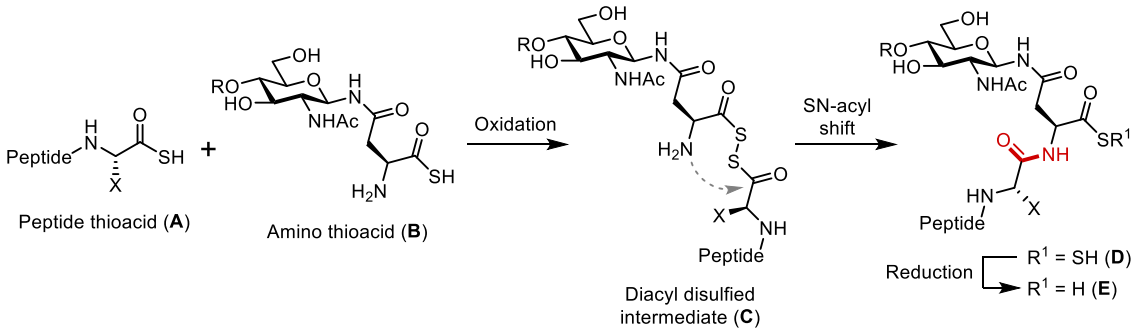


Figure 2-5. Reaction mechanism of DDC.

For a more detailed understanding of the reaction mechanism, the pathway of side reactions was considered. The main side reactions of DDC are the hydrolysis of the starting material (**A**, **B**) and the target glycopeptide thioacid (**E**). These hydrolyses are considered to be generated as shown in the following Figure 2-6. First of all, the desired reaction pathway is that a peptide thioacid (**A**) and a glycosyl asparagine thioacid (**B**) form a heterodimer (**C**), and the resulting diacyl disulfide intermediate (**C**) progresses an intramolecular arrangement to afford the desired α-amide bond (**D**). The compound (**D**) is a dithioacid with SH added to the target glycosylated peptide thioacid (**E**). This compound (**D**) is relatively unstable until it is reduced by a reducing reagent and can be easily hydrolyzed to give a carboxylic acid. This reaction pathway explains the formation of the hydrolyzed decomposition of the target glycosyl peptide thioacids (**E**). Next, the hydrolysis of the starting material (**A**, **B**) is also considered. Since both of starting materials have thioacids, peptide thioacids (**A**) and glycosyl asparagine thioacids (**B**) are able to form homodimers respectively. In a case of peptide homodimers, the resulting diacyl disulfide intermediates (**A**-**A**) are labile for the nucleophile, they are quickly hydrolyzed by a trace amount of water molecules. This intermediate (**A**-**A**) also undergo a disulfide exchange reaction with a glycosyl asparagine thioacid (**B**), in which case the desired heterodimer is formed and then the target molecule (**D**) is generated. On the other hands, in a case of

forming homodimers consisted of glycosyl asparagine thioacids (**B**), diacyl disulfide intermediates (**B-B**) are generated. Since the glycosyl asparagine thioacid has a bifunctional α-amine, an intramolecular rearrangement can undergo to afford a glycosyl dipeptide thioacid (**F**) as in the undesired pathway. However, no glycosyl dipeptide (**F**) has been observed in DDC. Therefore, it can be said that the considerable pathway forming glycosyl dipeptides does not exist. This phenomenon is probably due to the steric hindrance of the two oligosaccharyl structures in the transition state of a transfer reaction. Another byproduct of DDC is aspartimide (**G**) of a glycosyl asparagine thioacid (**B**). This is thought to be caused by the nitrogen of the amide bond of innermost GlcNAc, which attack the active carbonyl carbon in the intermediate (**B-B**) consisted of glycosyl asparagine thioacids as a secondary amine.

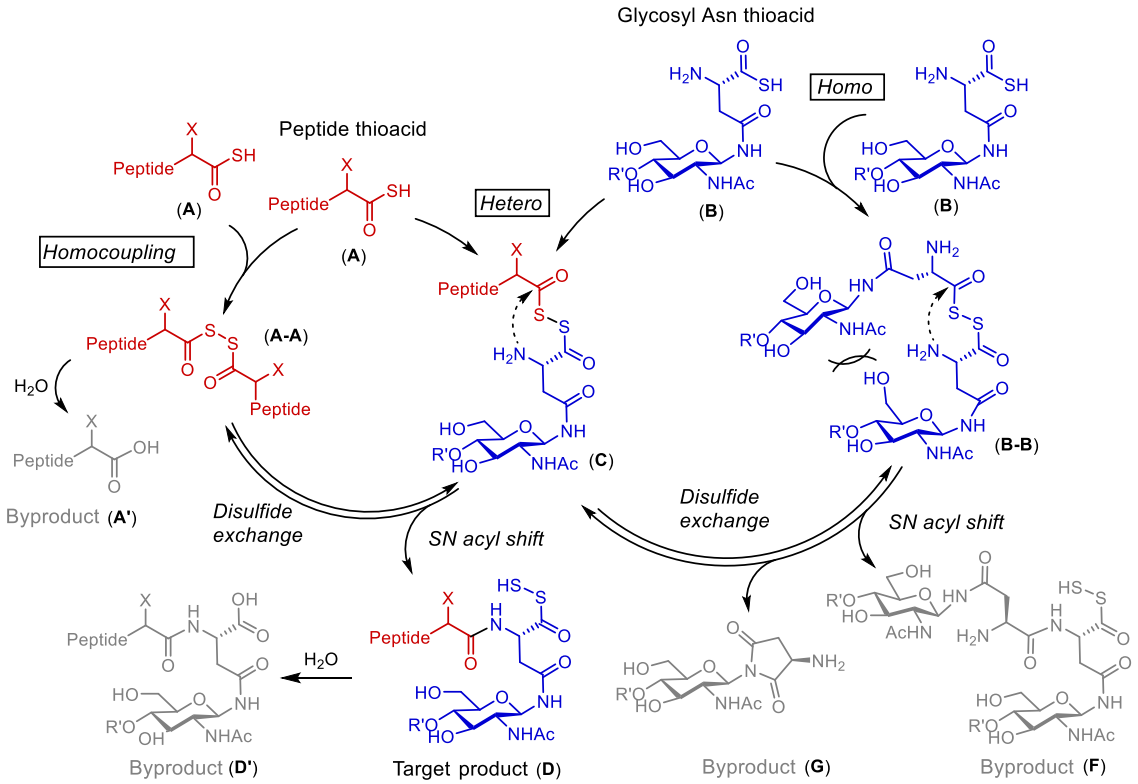


Figure 2-6. Considerable reaction pathway for DDC.

2-4. Application of DDC for the chemical insertion strategy

Because glycosyl asparagine thioacid **11** can be coupled with peptide thioacids by DDC, we then examined the second reaction according to the chemical insertion strategy. We applied thioacid capture ligation (TCL)[116-117] as the second reaction with the resulting glycosyl peptide thioacid (**A**) and Cys activated *p*-nitropyridyl disulfide (Npys) derivative (**B**) to afford glycopeptide (**C**) (Figure 2-7).

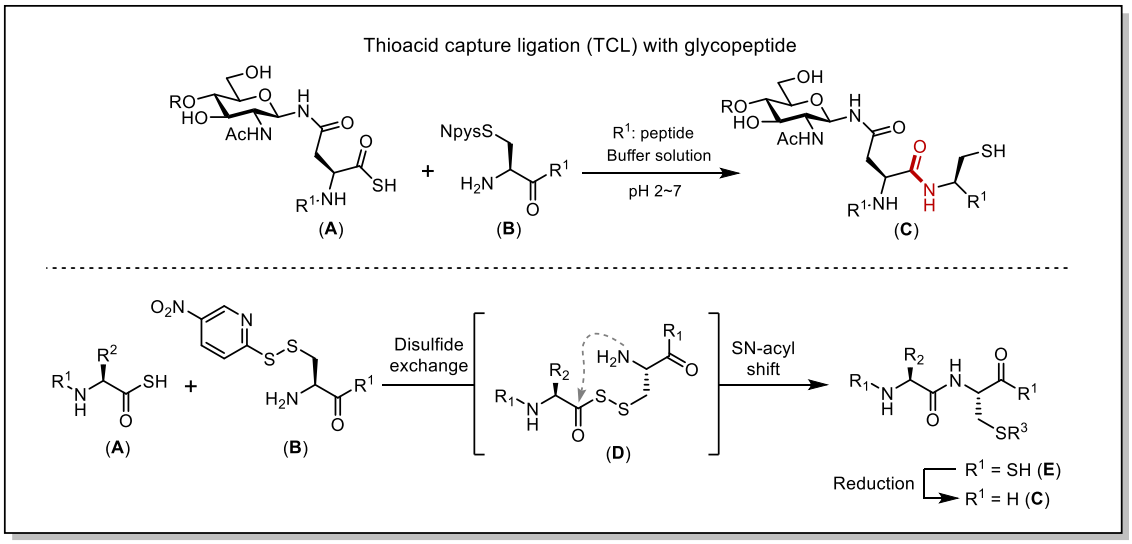


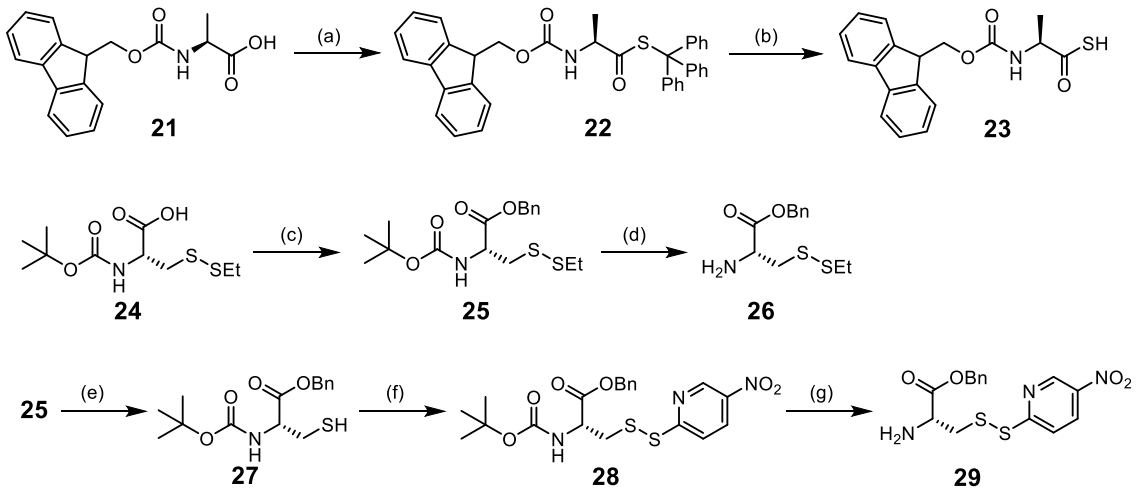
Figure 2-7. Thioacid capture ligation (TCL) for the synthesis of glycopeptide.

TCL is the reaction developed by Tam and coworkers, which couples unprotected peptide thioacids (**A**) and cysteinyl peptides with *p*-nitropyridyl disulfide (Npys) groups (**B**). Reaction starts from disulfide exchange with thioacid and Npys-thiol in a buffer. Key of this ligation method is a highly efficient capture reaction between a C-terminal thioacid and an activated Npys-disulfide on the N-terminal Cys to form an acyl disulfide intermediate (**D**) which then undergoes rapid intramolecular S-N acylation to generate an α-amide bond (**E**). The final product with a native Cys residue (**C**) at the ligation site is obtained after a simple thiolytic reduction reaction with reductants. Since there is no examples of the application of TCL for the synthesis of glycopeptides, especially peptide thioacids having a huge glycan motif at the peptide C-terminus, we first tested the effectiveness of this reaction utilizing the synthesis of a simple glycopeptide **31** consisted of three amino acids (Scheme 2-4).

In this case, we firstly used mono amino acid coupling at both the N- and C-terminus of glycosyl asparagine thioacid **11** to give a glycopeptide **31** for the feasible structural analysis comparing to the glycopeptide **31** synthesized by the conventional strategy. The coupling of glycosyl asparagine thioacid **11** was experimented with N-Fmoc-Ala-SH **23** and Cys derivatives having Npys group, H-Cys(Npys)-OBn **28**. N-Fmoc-Ala-SH **23** was prepared by a commercially available N-Fmoc-Ala-OH

21 (Scheme 2-3). The condensation of triphenyl methane thiol to N-Fmoc-Ala-OH **21** was performed with PyBOP and DIPEA at -20 °C to give compound **22** (73% yield). Then, deprotection of both the Boc group and triphenyl methane thiol groups were performed with TFA and TIPS yielded the desired N-Fmoc-Ala-SH **23** (87% yield). Next H-Cys(Npys)-OBn **28** was prepared by a commercially available N-Boc-Cys(SEt)-OH **24** (Scheme 2-3). The esterification of benzyl groups to N-Boc-Cys(SEt)-OH **24** was performed with benzyl bromide (BnBr) to give compound **25** (40% yield). Then reduction and disulfide formation of compound **25** were progressed by 2,2'-dithiobis(5-nitropyridine) and afforded N-Boc-Cys(Npys)-OBn **28** (75% yield). N-protected Cys derivative 28 was treated with TFA to yield the desired H-Cys(Npys)-OBn **28** (99% yield).

Scheme 2-3. The syntheses of amino acid derivatives.

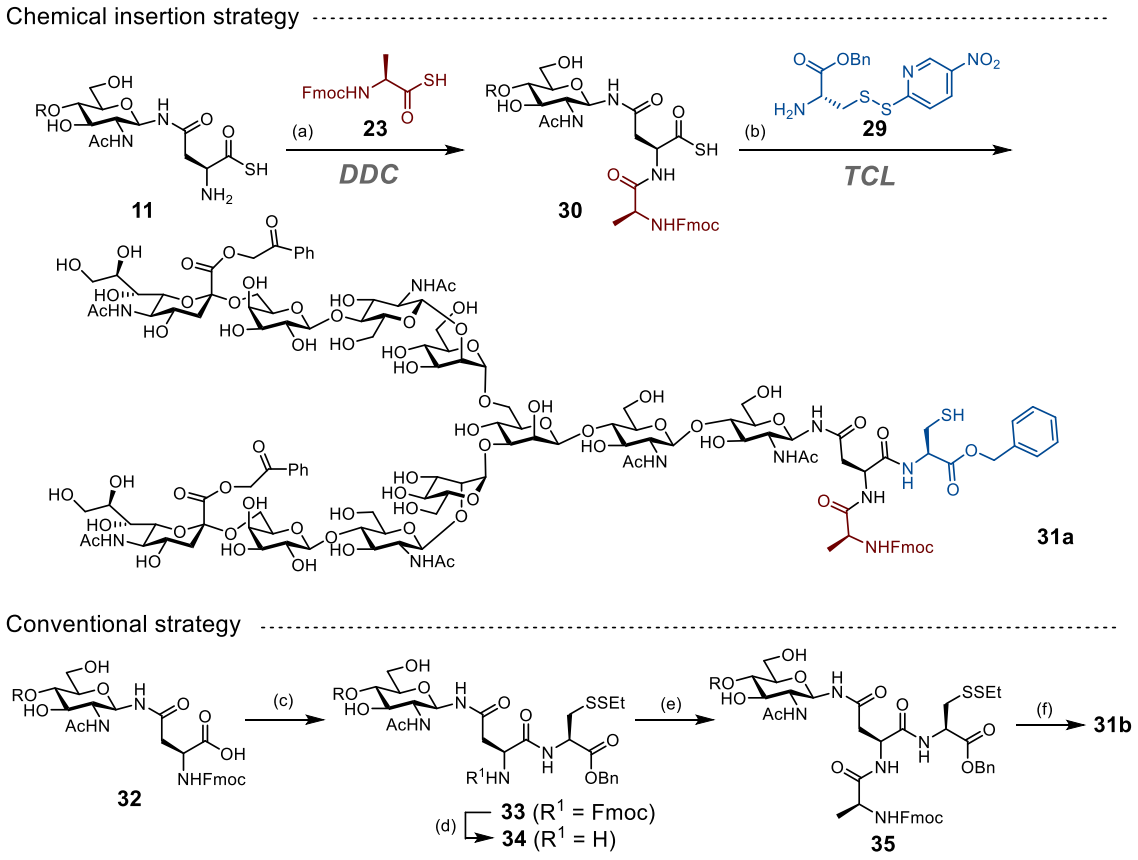


Reaction conditions; (a) PyBOP, DIPEA, triphenyl methane thiol, 3 h, -20 °C. 73%. (b) TFA, TIPS, 10 min, rt. 87%. (c) BnBr, DBU, DCM, 3h, rt. 40%. (d) TFA, rt. 99%. (e) DTT, DIPEA, MeOH, 1 h rt. (f) 2,2'-dithiobis(5-nitropyridine), MeOH, 12 h, rt. 75% (2 steps). (g) TFA, rt. 99%.

Having successfully synthesized the compounds that that consists of glycopeptides **31**, we synthesized the glycopeptides **31a** followed by chemical insertion strategy. Furthermore, in order to confirm the structure of glycopeptide **31a** (Fmoc-*L*-Ala-*L*-Asn(glycan)-*L*-Cys-OBn) synthesized by chemical insertion strategy (using glycosyl asparagine thioacid **11**), we compared it with the authentic sample **31b** prepared by a conventional strategy (Scheme 2-4). In the chemical insertion strategy, to a solution of glycosyl asparagine thioacid **11** in dry DMSO was added Fmoc-*L*-Ala-SH **23** (red colored) at room temperature using DDC. The resultant mixture was stirred for 6 hours to afford Fmoc-*L*-Ala-*L*-Asn(glycan)-SH **30**. After completion of DDC, to a resulting mixture of compound **30** was added a phosphate buffer solution (pH 5.7) containing the activated H-Cys(Npys)-OBn **29** (blue colored) and then the mixture was stirred for 0.5 hours to afford Fmoc-*L*-Ala-*L*-Asn(glycan)-*L*-Cys-OBn **31a** (24%

yield) with TCL. On the other hands, in the conventional strategy, glycopeptides carrying sialyl biantennary complex type oligosaccharide **31b** was prepared by from the Asn linked decasaccharide building block, Boc-Asn(sialylglycan)-OH **32**. The condensation of H-Cys(SEt)-OBn **26** to the substrate N-Fmoc-Asn-(sialylglycan)-OH **32** was performed with PyBOP and DIPEA at -15 °C to afford a Fmoc protecting glycosyl dipeptide **33**, then the Fmoc protecting groups were removed to afford compound **34**. Glycosyl dipeptide **34** was coupled with commercially available N-Fmoc-Ala-OH **21** with PyBOP and DIPEA at -15 °C to afford a glycosyl tripeptide **35** (47% yield). Then, thiolytic reduction removed ethane thiol from compound **35** yielded the desired glycosyl tripeptide carrying sialyl biantennary complex type oligosaccharide **31b** (91% yield).

Scheme 2-4. The synthesis of glycopeptide using chemical insertion and conventional strategy.



Reaction conditions; (a) DMSO, 6 h, rt. (b) DMSO, phosphate buffer (pH 5.7), CH_3CN, 12 h, rt. 24% (2 steps). (c) **26**, PyBOP, DIPEA, DMF, -20 ºC. (d) HOBt, hexamethyleneimine, 1-methylpyroridine, DMF, 1 h, rt. (e) **21**, PyBOP, DIPEA, DMF, -20 ºC. 47% (3 steps) (f) TCEP, TFA, CH_3CN, 1 h, 91%.

We next confirmed the structure of glycopeptide **31** synthesized with glycosyl asparagine thioacid **11** by extensive NMR analyses. We compared the structure of glycopeptide **31** with the authentic glycopeptide **31** synthesized by a conventional protocol with PyBOP. After isolation of both

products, ^{1}H NMR, HSQC, HMBC, NOESY, and DIPSI experiments were conducted. The basic assignments were performed by HMBC connectivity (Figure 2-8 A: blue arrows). The representative cross peak between benzyl methylene protons and the carbonyl carbon of cysteine was the starting point, and subsequent connectivity was successfully observed from the cysteines to Fmoc-protected alanine, as shown in Figure 2-8 (blue arrows). In addition to HMBC, assignments of NOESY (Figure 2-8 D: red arrows) and DIPSI (Figure 2-8 C: red arrows) signals supported the structure of both glycopeptides synthesized by the chemical insertion strategy and conventional coupling strategy. All DIPSI, NOESY, and ^{1}H NMR spectrum perfectly overlapped between glycopeptide **31** (red contour plot) and its authentic sample (black contour plot) (Figure 2-8 C–D, 2-9). As a result, the structures of both glycopeptides were completely identical in NMR analyses. When we analyzed the signal pattern, byproduct signals corresponding to the epimerized peptides were not observed (Figure 2-8, 2-9).

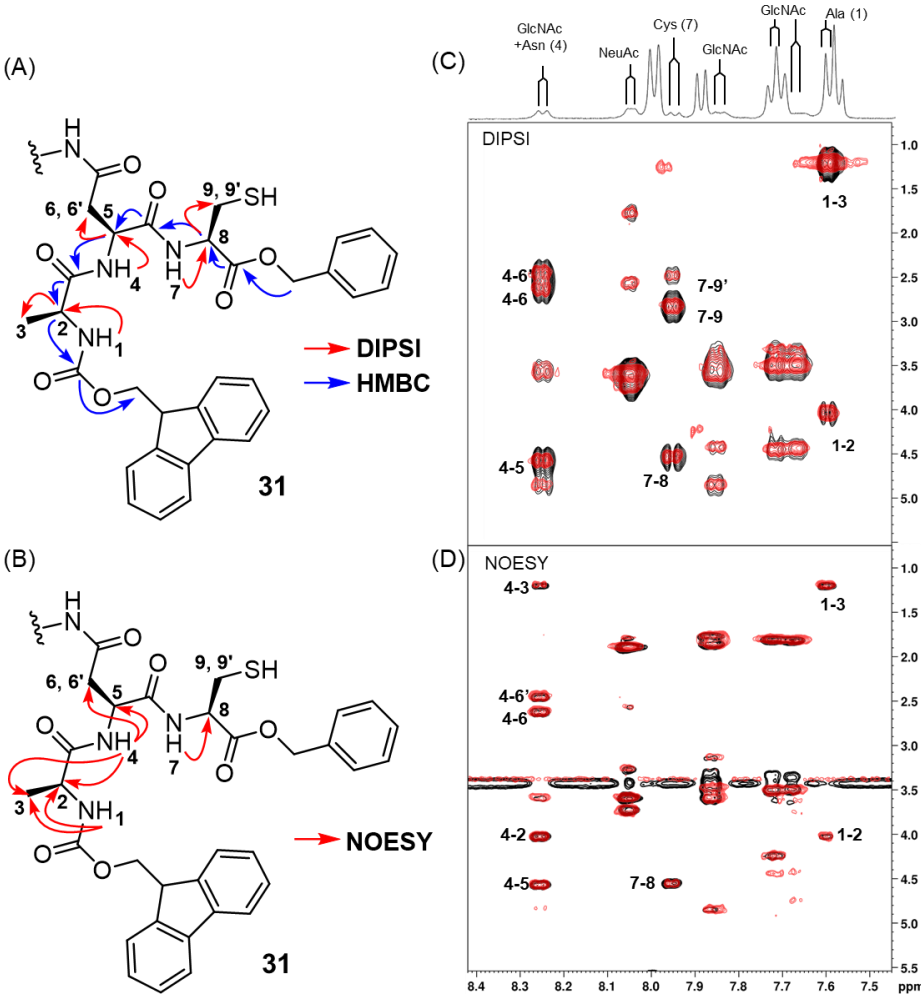


Figure 2-8. NMR analyses of Fmoc-Ala-Asn(glycan)-Cys-OBn **31**. (A) DIPSI (red arrows) and HMBC (blue arrows) connectivity. (B) The observed NOESY correlations are shown by red arrows. (C) DIPSI spectrum by 700 MHz NMR. Black color contour plot indicates authentic **31b** and red color contour plot indicates **31a** prepared by chemical insertion. (D) NOESY spectrum by 700 MHz NMR.

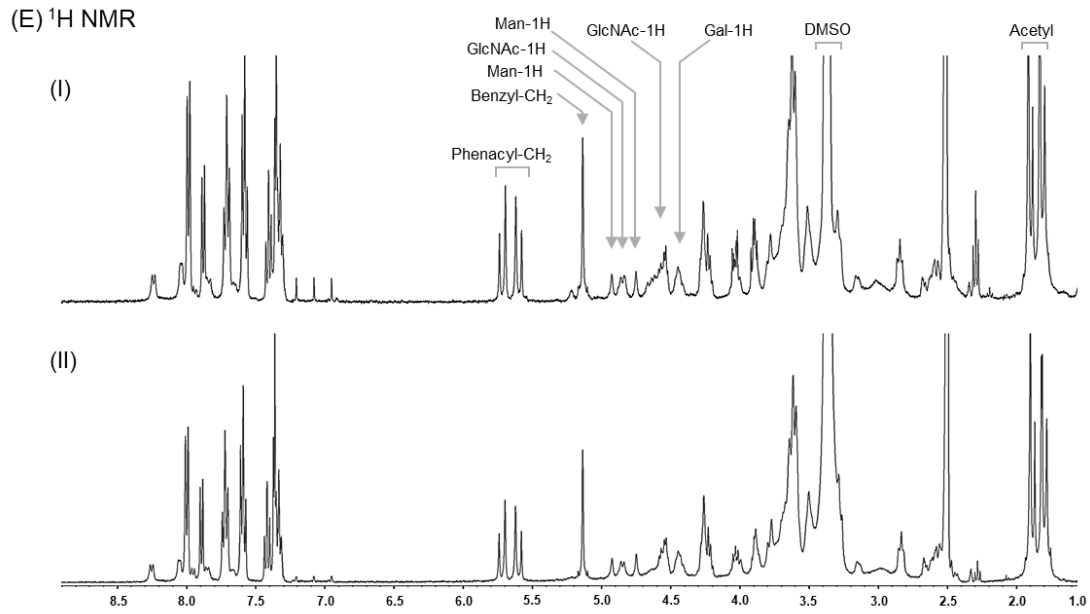


Figure 2-9. NMR analysis of glycopeptide. (E) 1H NMR spectra of Fmoc-Ala-Asn(glycan)-Cys-OBn 16 (I): synthesized by the thioacid-mediated strategy, (II): synthesized by PyBOP strategy.

From the results of NMR analysis, it was confirmed that the synthesized glycopeptides correctly formed α-amide bonds by using DDC and TCL. Thus, in principle, the glycan insertion strategy is capable of obtaining glycopeptides by two condensation reactions (Figure 2-10). Using this strategy, DDC with N-terminal peptide thioacids (**A**) and glycosyl asparagine thioacids (**B**) and following TCL with C-terminal cysteinyl peptides (**D**) afford full-length glycoprotein (**E**) efficiently.

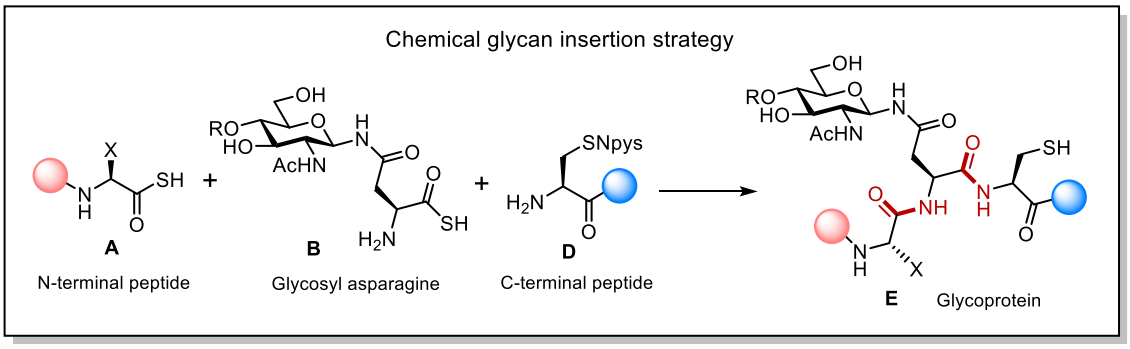
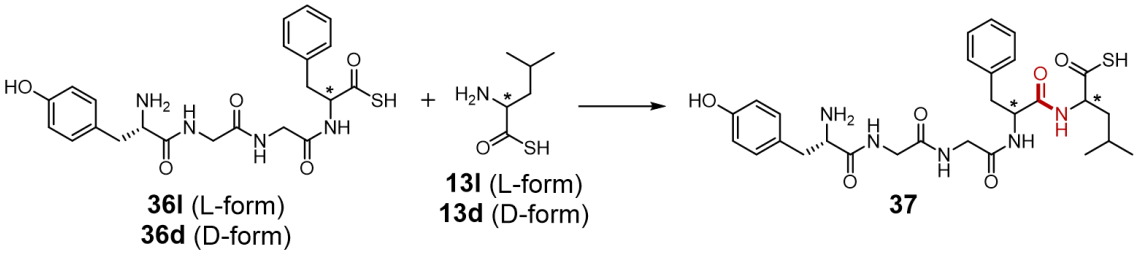


Figure 2-10. The chemical glycan insertion strategy.

2-5. Epimerization analysis

Furthermore, we investigated the epimerization amount of amino thioacids on α-amide formation in detail. We examined the condensation reactions of the *L* and *D* forms of leucine thioacid **13l** (Leu^{L}-SH) and **13d** (Leu^{D}-SH) with peptide thioacids **36l** having either the *L* form of phenylalanine (Tyr-Gly-Gly-Phe^{L}-SH) or **36d** having *D* form of phenylalanine at the C-terminus (Tyr-Gly-Gly-Phe^{D}-SH) . Consequently, we prepared all patterns of isomers (Tyr-Gly-Gly-Phe^{L}-Leu^{L}-SH, Tyr-Gly-Gly-Phe^{L}-Leu^{D}-SH, Tyr-Gly-Gly-Phe^{D}-Leu^{L}-SH, and Tyr-Gly-Gly-Phe^{D}-Leu^{D}-SH) by DDC and compared the retention time of the product with reversed-phase HPLC. All four peptide thioacids were well separated by HPLC analysis (Figure 2-11). Based on these results, we confirmed that the individual DDC reaction did not yield any of the isomers, which indicated DDC progressed with no epimerization.



36l (L-form)
36d (D-form)

13l (L-form)
13d (D-form)

37

Petide thioacid	Amino thioacid	LC chromatogram
L	*L*	I
L	*D*	II
D	*L*	III
D	*D*	IV

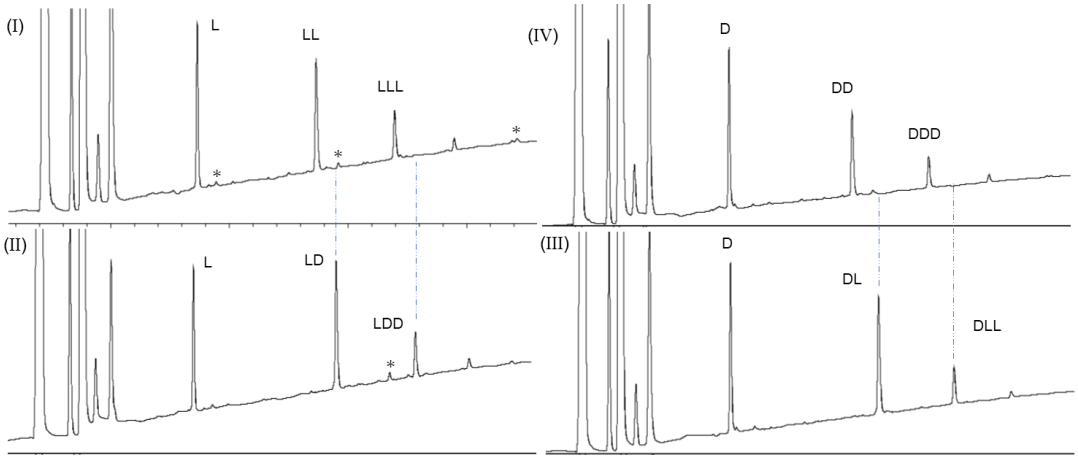
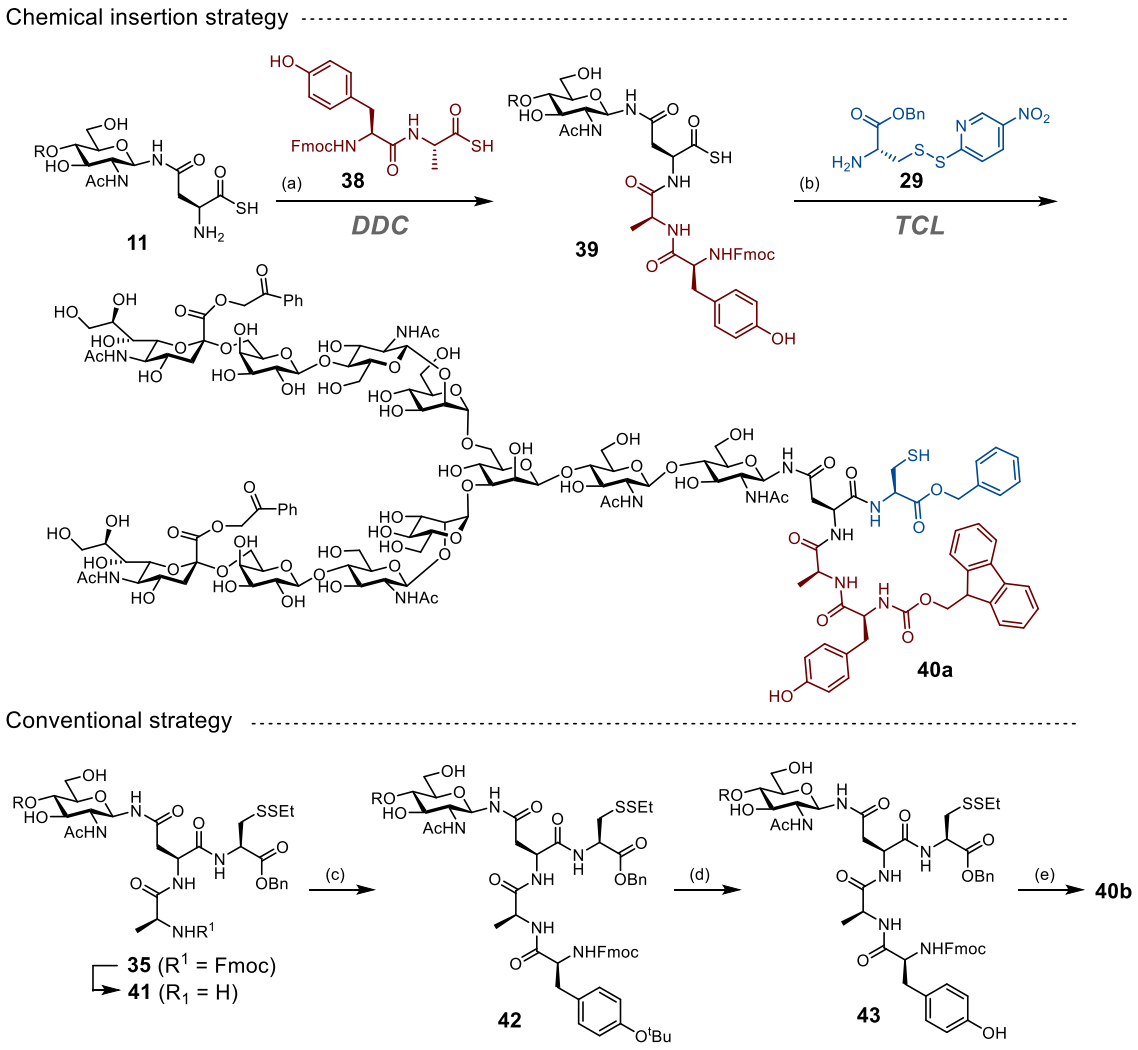


Figure 2-11. LC profiles for monitoring epimerization. Asterisk* was not peptides. L; H$_2$N-Tyr-Gly-Gly-Phe^{L}-SH, D; H$_2$N-Tyr-Gly-Gly-Phe^{D}-SH, LL; H$_2$N-Tyr-Gly-Gly-Phe^{L}-Leu^{L}-SH, LD; H$_2$N-Tyr-Gly-Gly-Phe^{L}-Leu^{D}-SH, DL; H$_2$N-Tyr-Gly-Gly-Phe^{D}-Leu^{L}-SH, DD; H$_2$N-Tyr-Gly-Gly-Phe^{D}-Leu^{D}-SH, LLL; H$_2$N-Tyr-Gly-Gly-Phe^{L}-Leu^{L}-Leu^{L}-SH, LDD; H$_2$N-Tyr-Gly-Gly-Phe^{L}-Leu^{D}-Leu^{D}-SH, DLL; H$_2$N-Tyr-Gly-Gly-Phe^{D}-Leu^{L}-Leu^{L}-SH, DDD; H$_2$N-Tyr-Gly-Gly-Phe^{D}-Leu^{D}-Leu^{D}-SH

Scheme 2-5. The synthesis of glycopeptide using chemical insertion and conventional strategy.



Reaction conditions; (a) DMSO, 9 h, rt. 47%. (b) Phosphate buffer (pH 5.7), CH_3CN, 3 h, rt. 43%. (c) Fmoc-*L*-Tyr-OH, PyBOP, DIPEA, DMF, -20 ºC. 33%. (d) TFA, TIPS, 30 min, rt. (e) TCEP, CH_3CN, H_2O, 1 h, rt. 80% (2 steps).

To confirm the epimerization amount under potential oxazolone formation, we had to examine DDC using peptide thioacids, which consist of at least dipeptide thioacids instead of mono amino thioacids. This is because dipeptide is the smallest component to afford oxazolone formation and induce the epimerization to peptides. In this case, we first tested the epimerization yield of DDC utilizing the synthesis of a simple glycopeptide **40a** (Fmoc-*L*-Tyr-*L*-Ala-*L*-Asn(glycan)-*L*-Cys-OBn) consisted of four amino acids by the chemical insertion strategy and examined the feasible structural analysis comparing to the glycopeptide **40b** synthesized by the conventional strategy (Scheme 2-5). In order to synthesize glycopeptide **40a**, to a solution of glycosyl asparagine thioacid **11** in dry DMSO was added Fmoc-*L*-Tyr-*L*-Ala-SH **38** (red colored) at room temperature using DDC. The resultant

mixture was stirred for 9 hours to afford Fmoc-*L*-Tyr-*L*-Ala- *L*-Asn(glycan)-SH **39** (47% yield). After completion of DDC, to a resulting mixture of compound **39** was added a phosphate buffer solution (pH 5.7) containing the activated H-Cys(Npys)-OBn **29** (blue colored) and then the mixture was stirred for 3 hours to afford the desired glycosyl tetrapeptide Fmoc-*L*-Tyr-*L*-Ala-*L*-Asn(glycan)-*L*-Cys-OBn **40a** with TCL (43% yield). On the other hands, in the conventional strategy, glycopeptides carrying sialyl biantennary complex type oligosaccharide **40b** was prepared from a glycosyl tripeptide building block, N-Fmoc-*L*-Ala-*L*-Asn(sialylglycan)-*L*-Cys(Set)-OH **35**. the Fmoc protecting group of compounds **35** was removed by hexamethyleneimine to afford compound **41**. The condensation of commercially available N-Fmoc-Tyr-OH to the substrate **41** was performed with PyBOP and DIPEA at -15 °C to afford a Fmoc protecting glycosyl tetrapeptide **42** (33% yield). The tert-butyl group of Tyr was deprotected by TFA acidic conditions to give glycosyl tetra peptide **43** and then, thiolytic reduction removed ethane thiol from compound **43** yielded the desired glycosyl tetrapeptide carrying sialyl biantennary complex type oligosaccharide **40b** (80% yield).

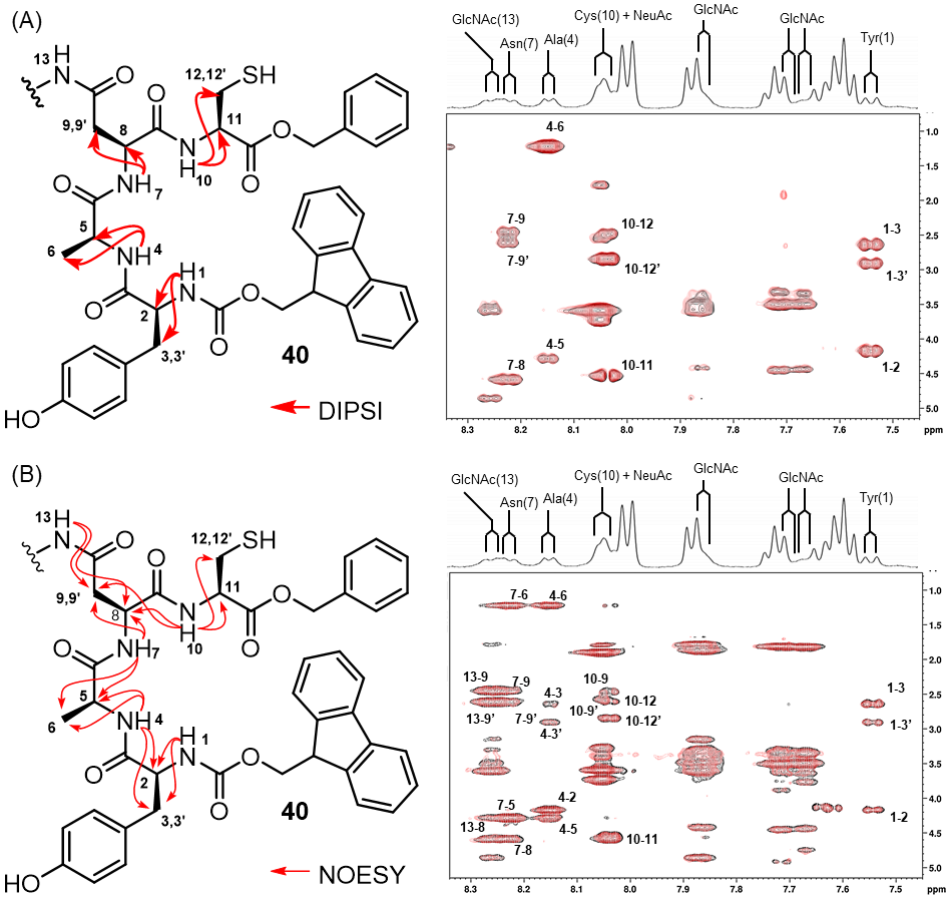


Figure 2-12. NOESY, DIPSI and HSQC of glycopeptide Fmoc-Tyr-Ala-Asn(glycan)-Cys-OBn **63**. (A) DIPSI connectivity is shown by red arrows. DIPSI spectrum by 700 MHz NMR. Black contour plot indicates authentic **40b** and red contour plot indicates **40a** prepared by thioacid-mediated strategy. (B) NOESY correlations are shown by red arrows. NOESY spectrum by 700 MHz NMR. Black contour plot indicates authentic **40b** and red contour plot indicates **40a** prepared by chemical insertion .

The structure of the resultant glycosyl tetrapeptide **40a** (Fmoc-Tyr-Ala-Asn(glycan)-Cys-OBn) was also confirmed by extensive NMR analyses (Figure 2-12). All signals of the DDC/TCL product were found to be identical to those of the authentic product **40b** synthesized by a conventional protocol with PyBOP. We observed a single HSQC cross peak between the α-proton and its carbon signal of all amino acids, but we could not exclude the possibility of the complete overlap of the signals caused by epimerization (Figure 2-12). Therefore, we further examined the monitoring of epimerization under DDC reaction conditions.

We synthesized Fmoc-*L*-Tyr-*L*-Ala-SH **38l** and Fmoc-*L*-Tyr-*D*-Ala-SH **38d** and left both substrates in DDC conditions. We found the best conditions for the separation of both HPLC profiles,

and these data clearly indicated no epimerization (Figure 2-13). Combined with NMR analyses supported that DDC was less prone to cause epimerization by oxazolone formation.

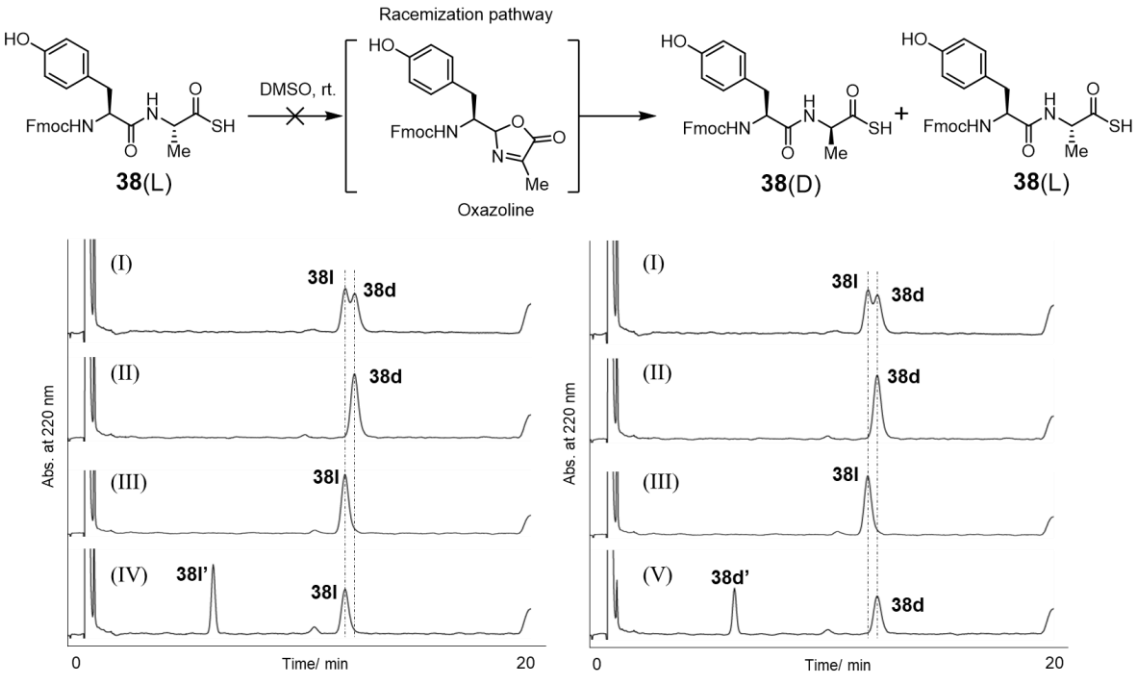


Figure 2-13. LC for the monitoring of epimerization of **38l** and **38d** in DMSO. The peaks corresponding with **38l'** (IV) and **38d'** (V) are generated by hydrolysis.

2-6. Extension of chemical glycan insertion strategy

In *chapter 2-4*, I succeeded to device the chemical glycan insertion strategy. However, in that synthetic strategy, the peptides only can have cysteine residues or alanine residues, which can be obtained by desulfurization of cysteines, beside glycosylated asparagine. Thus, I had to come up with an alternative method to extend the amino acids beside asparagine from cysteines and alanines. Then, I investigated two types of extended chemical glycan insertion strategy. First is the use of C-terminal peptides (**D**) having the β-mercapto amino acids with the Npys groups for TCL (Figure 2-14 A). Ligations using β-mercapto amino acid derivatives have been so far achieved by many groups. The β-mercapto amino acids can form α-amide bonds through thiol capture reactions and subsequent intramolecular SN acyl rearrangements (*chapter 1-8*).[57] The desulfurization of β-mercapto groups of full-length glycopeptides (**E**) after TCL will yield glycoproteins (**F**) having any amino acids beside glycosylated asparagine residues. The second strategy is that the use of C-terminal peptide (**I**) having a selenocysteine with the Npys group for TCL (Figure 2-14 B). Selenocysteine have been used for alternative native chemical ligations so far since the deselenization applying single electron transfer by TCEP after ligation converts a selenocysteine to a serine efficiently.[125-126] Utilizing the second deselenization strategy, we can obtain a glycoprotein (**K**) having serine beside glycosylated asparagine residue from full-length glycopeptide (**J**). The first strategy is described in *chapter 5-4*. Herein, I describe the second strategy.

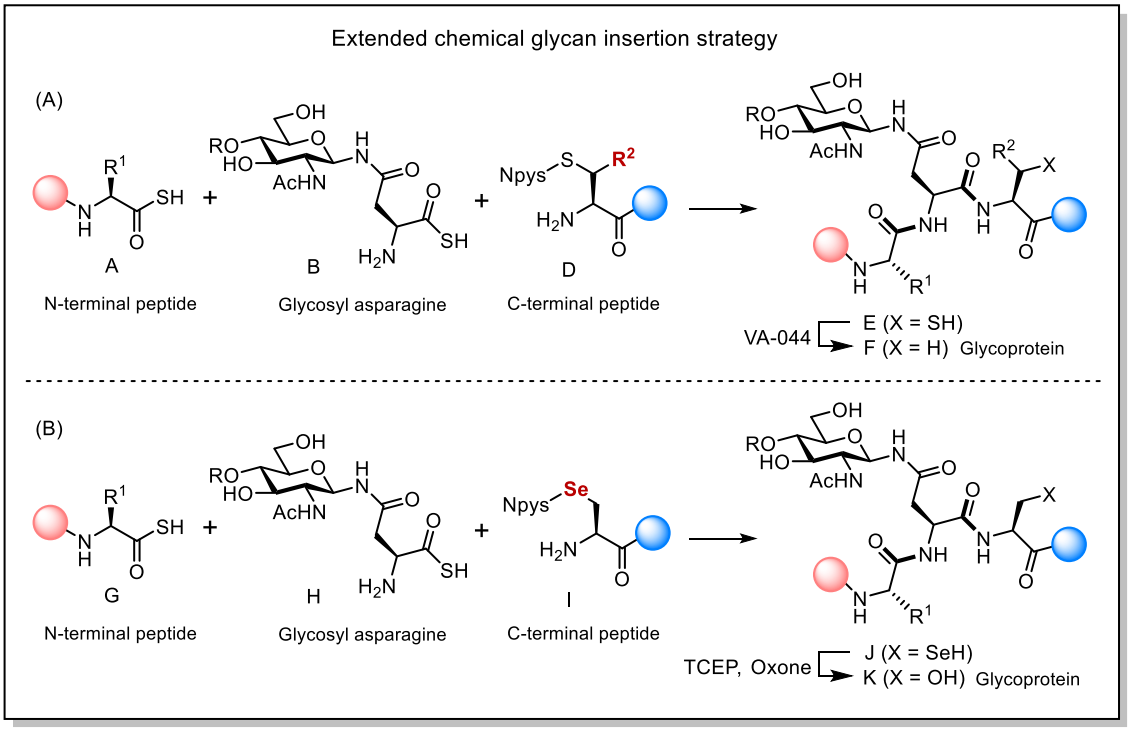
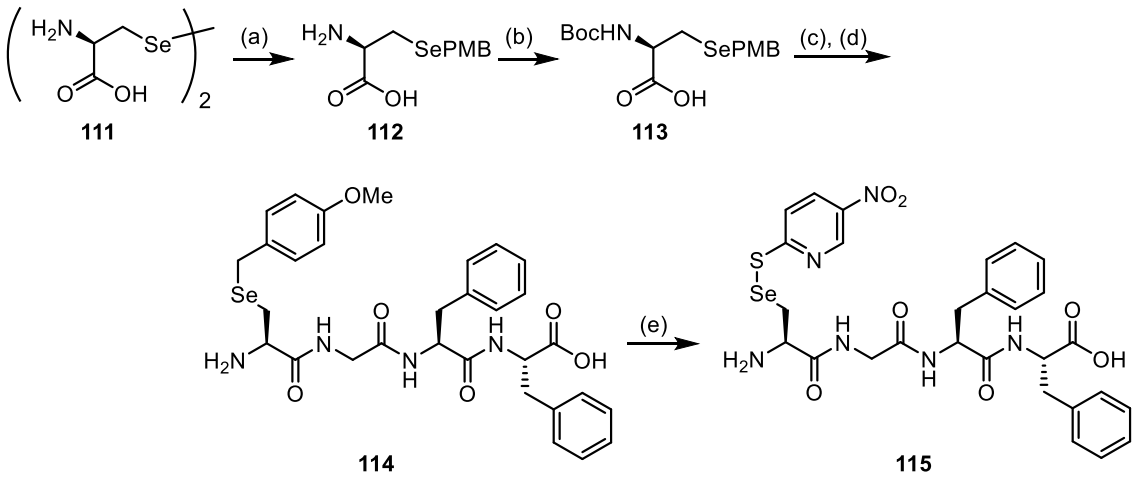


Figure 2-14. Extended chemical glycan insertion strategy. (A) Strategy with β-mercapto amino acid derivatives. (B) Strategy with selenocysteine derivatives.

To investigate TCL with peptides containing the Npys activated selenocysteine at N-terminus, the synthesis of a selenopeptide **115** was progressed (Scheme 2-6). Following the previous report[127], H-*L*-Sec(PMB)-OH **112** was synthesized from *L*-selenocystine **111**. L-selenocystine was reduced to selenocysteine with $NaBH_4$ in an aqueous condition and subsequently subjected to protection by *p*-methoxybenzyl chloride (PMB-Cl) to yield H-*L*-Sec(PMB)-OH **112**. Then Boc protection afforded N-Boc-*L*-Sec(PMB)-OH **113** with 83% isolated yield. Selenocysteine derivatives **112** were subjected to Fmoc-SPPS and cleaved by resin in an acidic condition to yield PMB-protected selenopeptide **114**. Then following the previous report,[128] deprotection of the PMB groups was performed with 2,2′-dithiobis(5-nitropyridine) and thioanisole to afford selenopeptide with the Npys group **115** with 54% isolated yield. The structure of compound **115** was confirmed by nuclear magnetic resonance (NMR) spectroscopy and high-resolution mass spectrometry (HRMS).

Scheme 2-6. Synthesis of peptide containing selenocysteine.



Reaction conditions; (a) PMB-Cl, $NaBH_4$, NaOH, EtOH, 4 h, 0˚C. (b) Boc_2O, $NaHCO_3$, CH_3CN, 12 h, rt. 83% (2 steps). (c) Fmoc-SPPS. (d) TFA, TIPS. (e) TFA, H2O, thioanisole (93:5:2), 2,2'-dithiobis(5-nitropyridine) (30 eq. 5mM), 1h, rt.

In order to perform TCL with glycosyl asparagine thioacids and selenopeptides, TCL was performed using model peptide **115** having the Npys group at N-terminal selenocysteine and glycosyl asparagine thioacid **11** having a complex type oligosaccharide. This reaction employed one equivalent of glycosyl asparagine thioacid **11** toward two equivalent of peptide **115** (Figure 2-15). The reaction was traced in 5 hours, however the desired glycopeptide **116** was not observed in pH 5.5 buffer. As the result the hydrolysis of glycosyl asparagine thioacid **11** and dimerization of selenopeptide **115** quantitatively proceeded as side reactions. Although, TCL was performed at various pH conditions. the target molecule **116** was not generated.

Then, we hypothesized that the reaction proceeded through a different pathway than expected (Figure 2-15). According to the reaction mechanism of TCL, thioacid **11** nucleophilically attacks a selenium atom of a selenopeptide **115**, and a peptide is captured by the thioacid (S-Se bond form), resulting in S-N acyl intramolecular rearrangement progress to afford an amide bond (Figure 2-15; red dashed arrows). However, in our experiments, thioacid **11** nucleophilically attacked a sulfide atom of *p*-nitropyridilsulfide of selenopeptide **115**, which became a strongly activated acyl disulfide intermediate **119** and was immediately hydrolyzed to afford compound **118** (Figure 2-15; blue dashed arrows). This side reactions resulted in a formation of diselenide byproducts **117** from the selenopeptide **115** and hydrolysis of the thioacid **11**.

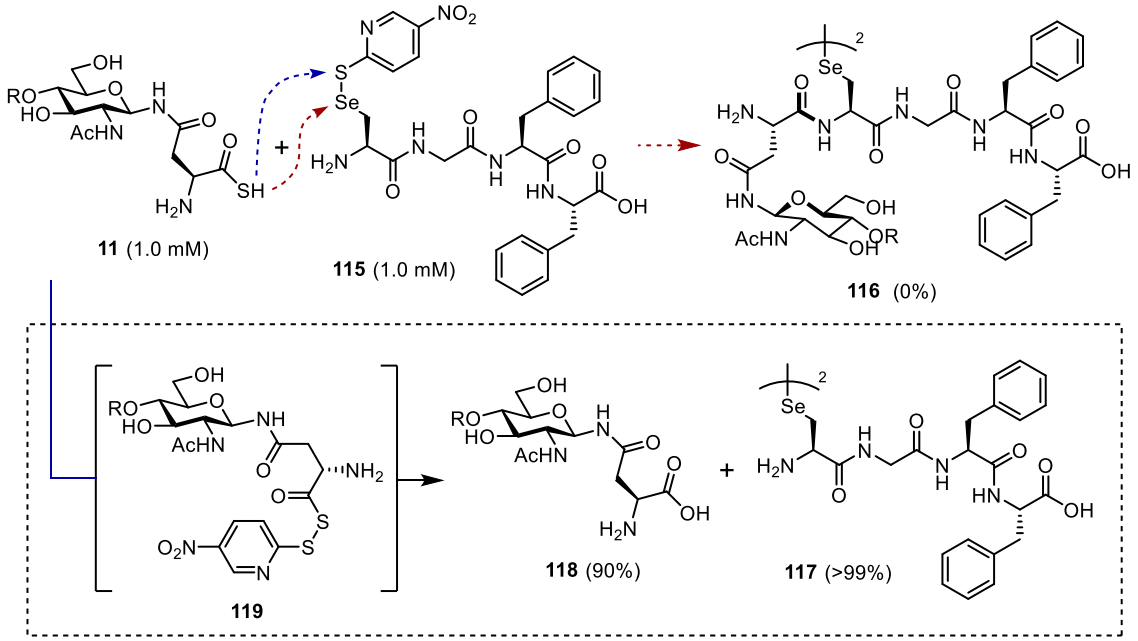


Figure 2-15. TCL with peptides having Npys activated selenocysteine **116**.

These results indicated the limitations of TCL. In other words, utilizing the chemical glycan insertion strategy, an amino acid beside the C-terminus of glycosylated asparagine cannot be adapted by serine and threonine at present. In the future, it will be necessary to use auxiliary groups to realize efficient TCL.

2-7. Discussion

We found that glycosyl asparagine thioacid **11** unexpectedly did not yield polymerization but afforded the desired glycopeptide thioacid. This phenomenon was also observed with glycosyl asparagine thioacid carrying mono GlcNAc shown in *Chapter 5*. Even the monosaccharide did not yield any polymerization of glycosyl asparagine thioacids. Then regarding the suppression of polymerization with glycosyl asparagine thioacid, we hypothesized that the inhibition of polymerization seemed to be due to the steric repulsion between monosaccharides or oligosaccharides in the S-N acyl rearrangement from the diacyl disulfide intermediate **46** or the formation of the diacyl disulfide intermediate **45** itself (Figure 2-16). In addition, we considered that fluctuation in the glycan contributed to the reduction in polymerization due to steric hindrance.

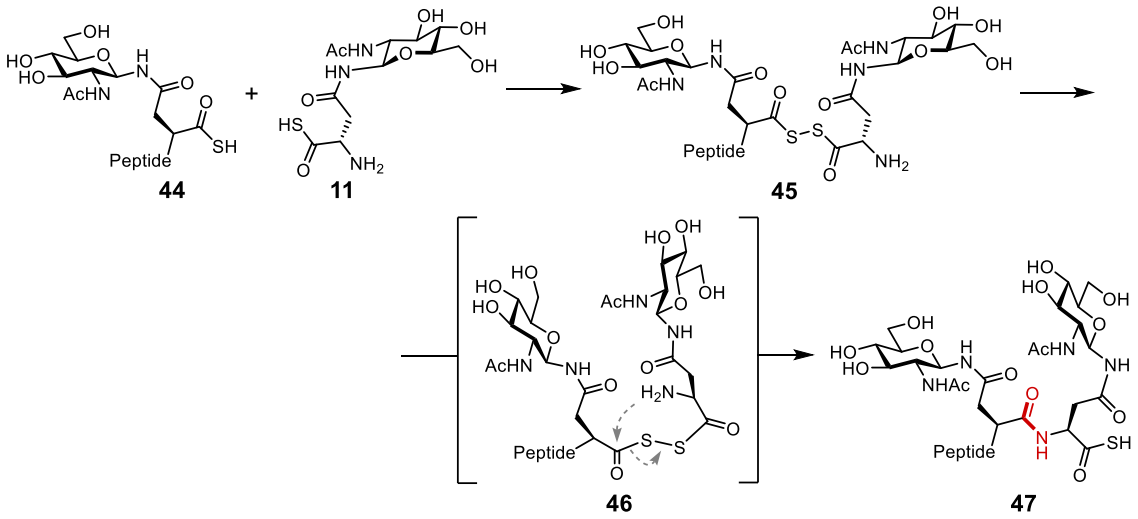


Figure 2-16. Mechanism of the inhibition of polymerization by glycans of glycosyl asparagine thioacid.

The diacyl disulfide intermediate consisting of peptide thioacid and glycosyl asparagine thioacid can form, and then a subsequent S-N acyl shift occurs to yield glycopeptide thioacid. As shown in Figure 2-3, the peptide thioacid with valine at the C-terminus suppressed the yield compared with those with alanine at the same C-terminus. These results supported our hypothesis that steric hindrance interferes with the formation of the diacyl disulfide intermediate and S-N acyl shift to yield glycopeptide thioacid.

On the other hands, the anhydride form of ATC between glycosyl asparagine thioacid **11** and peptide thioester **48** did not give the desired glycopeptide thioacid using longer peptides. [119] In this case, shown in the Figure 2-17, steric hindrance might interfere with S-to-N migration via a five-

membered ring intermediate **49**. The retardation of S-to-N migration by steric hindrance or sugar moiety fluctuation led to the resultant anhydride remaining in a solution or organic solvents, and it might have immediately decomposed. Therefore, the anhydride activation method might not give the desired glycopeptide thioacid. On the other hand, the six-membered ring **52** formed by a disulfide bond seemed to be suitable for S-to-N migration since the three covalent bonds C-S-S-C showed a specific angle and flexibility[129].

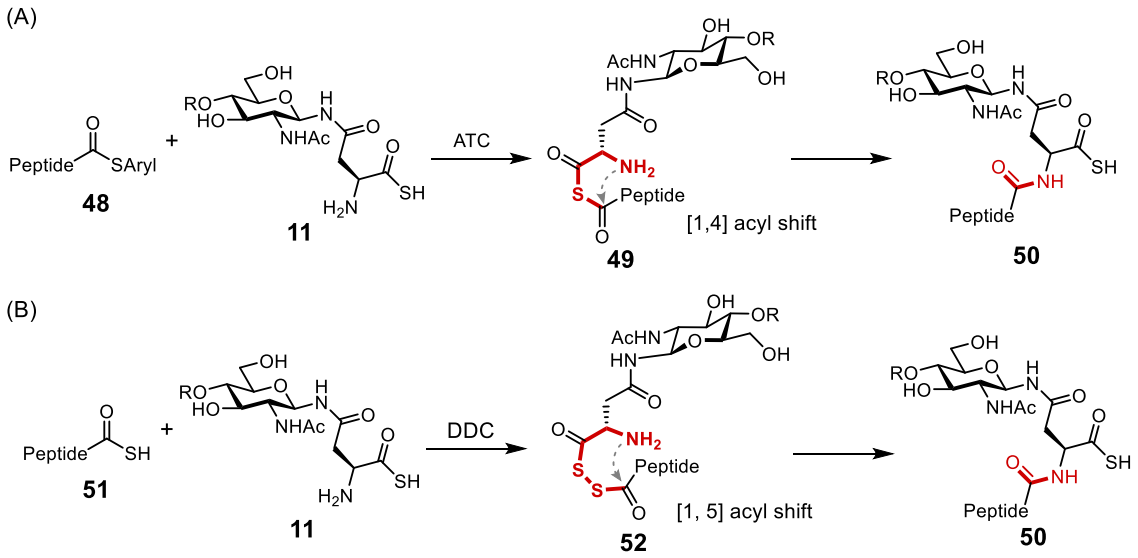


Figure 2-17. (A) Anhydride formation **49** of ATC with peptide-thioester **48** and glycosyl asparagine thioacid **11**. A subsequent [1,4]-acyl shift yields an amide bond. (B) Diacyl disulfide formation with peptide thioacid **51** and glycosyl asparagine thioacid **11** by DDC. Subsequent [1,5]-acyl shift yields an amide bond.

DDC can yield glycopeptides without epimerization at the N-terminal amino acid of peptide thioacid and glycosyl asparagine. It is known that peptide thioester generates epimerization under basic conditions during NCL or some other ligation protocols, whereas peptide thioacid has a different chemical characteristic nature. Because the diacyl disulfide bond does not show a potent electron withdrawing effect in DDC under neutral or acidic conditions, we considered that epimerization can be considerably reduced.

2-8. Summary

We established a method to convert glycosyl asparagine carrying monosaccharide and biantennary complex type oligosaccharides into thioacid form. We also have developed a robust α-amide bond formation reaction, DDC (Figure 2-18). Furthermore, by optimizing the conditions, we succeeded in DDC using glycosyl asparagine thioacids (**B**) to yield glycopeptide thioacids (**C**). HPLC and NMR analyses of the glycopeptides obtained by DDC confirmed that the reaction progressed without epimerization through the formation of oxazolone.

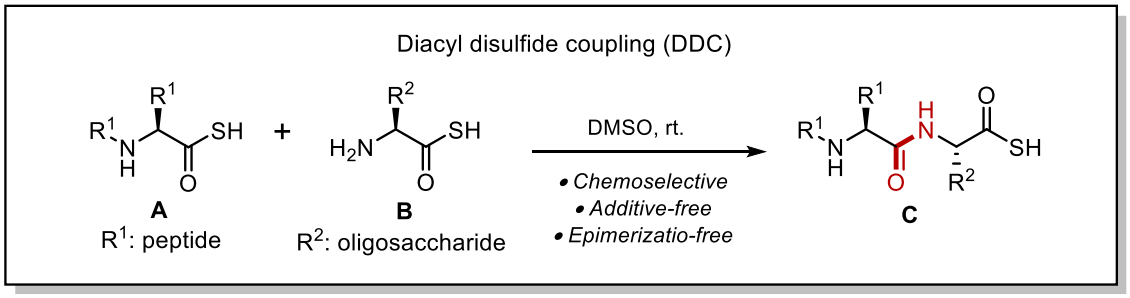


Figure 2-18. Diacyl disulfide coupling (DDC)

Since the glycopeptides synthesized by DDC have thioacids, we succeeded in further condensation of another peptides by TCL. This led to the discovery of the chemical glycan insertion strategy shown in the Figure 2-19, that can couple two peptides (**A**, **D**) with the N- and C-terminus of glycosyl asparagine thioacid (**B**). We employed glycosyl asparagine thioacid as the junction point for the coupling of N- and C-terminal peptides. The first coupling is DDC between a peptide thioacid (**A**) and glycosyl asparagine thioacid (**B**). Because the resultant glycopeptide has a thioacid form at its C-terminus, we could apply TCL for the coupling of the resultant glycopeptide thioacid and another peptide having a Npys group at its N-terminus (**D**) to afford the full-length glycoprotein backbone (**E**). This route enabled us to assemble the entire glycoprotein backbone in two steps. In addition, valuable glycosyl amino thioacids could be used at the late stage of the synthesis of glycoproteins.

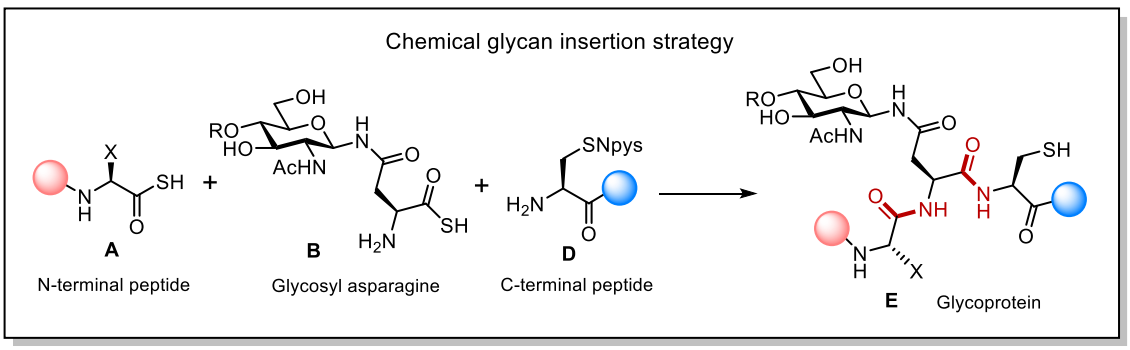


Figure 2-19. Chemical glycan insertion strategy for the synthesis of glycoproteins homogeneously.

Chapter 3. Synthesis of glycoprotein 1: CC chemokine ligand 1 (CCL1)

3-1. Introduction of chapter 3

After we studied the optimization of a new α-amide formation reaction employing a peptide thioacid and glycosyl asparagine thioacid, we examined glycoprotein syntheses by the new strategy shown in Figure 2-16. We achieved the syntheses of three bioactive glycoproteins including the CC chemokine ligand 1 (CCL1) (*Chapter 3*), interleukin 3 (IL3) (*Chapter 4*) and serine protease inhibitor Kazal type 13 (SPINK13) (*Chapter 5*) with a homogeneous biantennary sialyl oligosaccharides by the chemical insertion strategy. First, we synthesized CCL1 carrying a sialyl biantennary complex type oligosaccharide.

Chemokines are chemotactic cytokines expressed by cells that induce directional movement of leukocytes and exert the biological functions through the specific G protein-coupled receptor (GPCR) presented on the cell membrane called chemokine receptors[130]. Since the chemokines are known as pathogenesis of inflammatory diseases and viral infection, many biochemical studies targeting chemokines have been investigated to date[131]. In terms of the structure of chemokines, their molecular weight is approximately 8-12 kDa and consist of a N-terminal loop, three β-sheets and one C-terminal α-helix as common domains.[132] Especially, N-terminal loops is responsible for the binding to their receptors and activate them.[133] In detail, chemokines are classified into four main subfamilies such as CC, CXC, CX_3C and C motif chemokines depending on the formation of key two disulfide bonds (Figure 3-1).[132] The CC motif chemokines have two cysteines presented on the N-terminal side are contiguous. The CXC subfamily have two adjacent cysteines separated by one amino acid. The C motif chemokines have one N-terminal cysteine and one complementary cysteine in the downstream. The CX_3C motif chemokines such as fractalkine[134] have three amino acids between the first two cysteine residues.

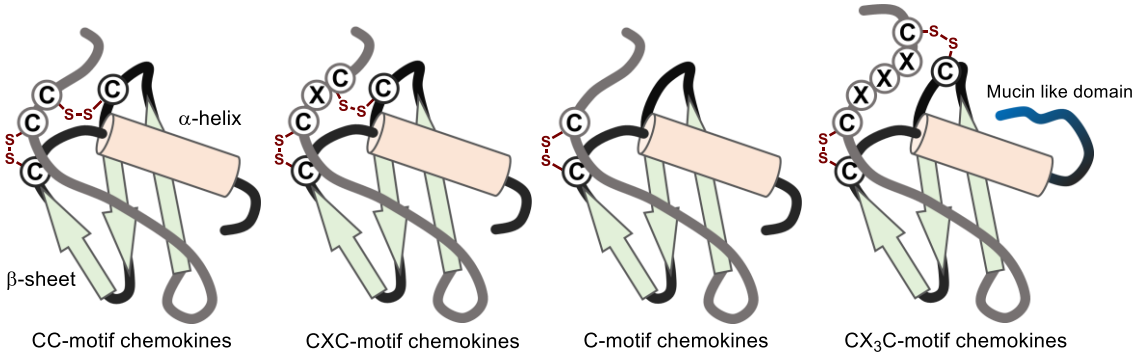


Figure 3-1. Classification of the chemokines depending on the position of the two disulfide bonds. Light green arrows indicate three anti parallel β-sheets. A light red stick indicates an α-helix. Chemokines classified as CX_3C motif also have mucin like domain at the C-terminus of chemokines.

Our synthetic target CCL1 belongs to a family of CC motif chemokines and has a complex-type, N-linked sialyloligosaccharide at the 29th position. Since the previous chemical synthesis of CCL1 with a homogeneous complex type of glycan employed SPPS of the glycopeptide segment[89], glycosylated CCL1 was a suitable target to demonstrate our approach without SPPS of glycopeptides.

3-2. Synthetic strategy of CCL1

In this experiment, we examined the synthesis of CCL1 **53** by the chemical insertion strategy. Based on glycan insertion strategy, we divided CCL1 to three components, glycosyl asparagine thioacids **11** and two peptide segments (Lys[1]-Ala[28]-COSH **54** and Cys[30](Npys)-Lys[73] **55** prepared by SPPS. We designed to synthetic scheme, which insert a glycosyl asparagine thioacid **11** between peptide **54** and peptide **55** using two chemoselective reactions, DDC and TCL (Figure 3-2).

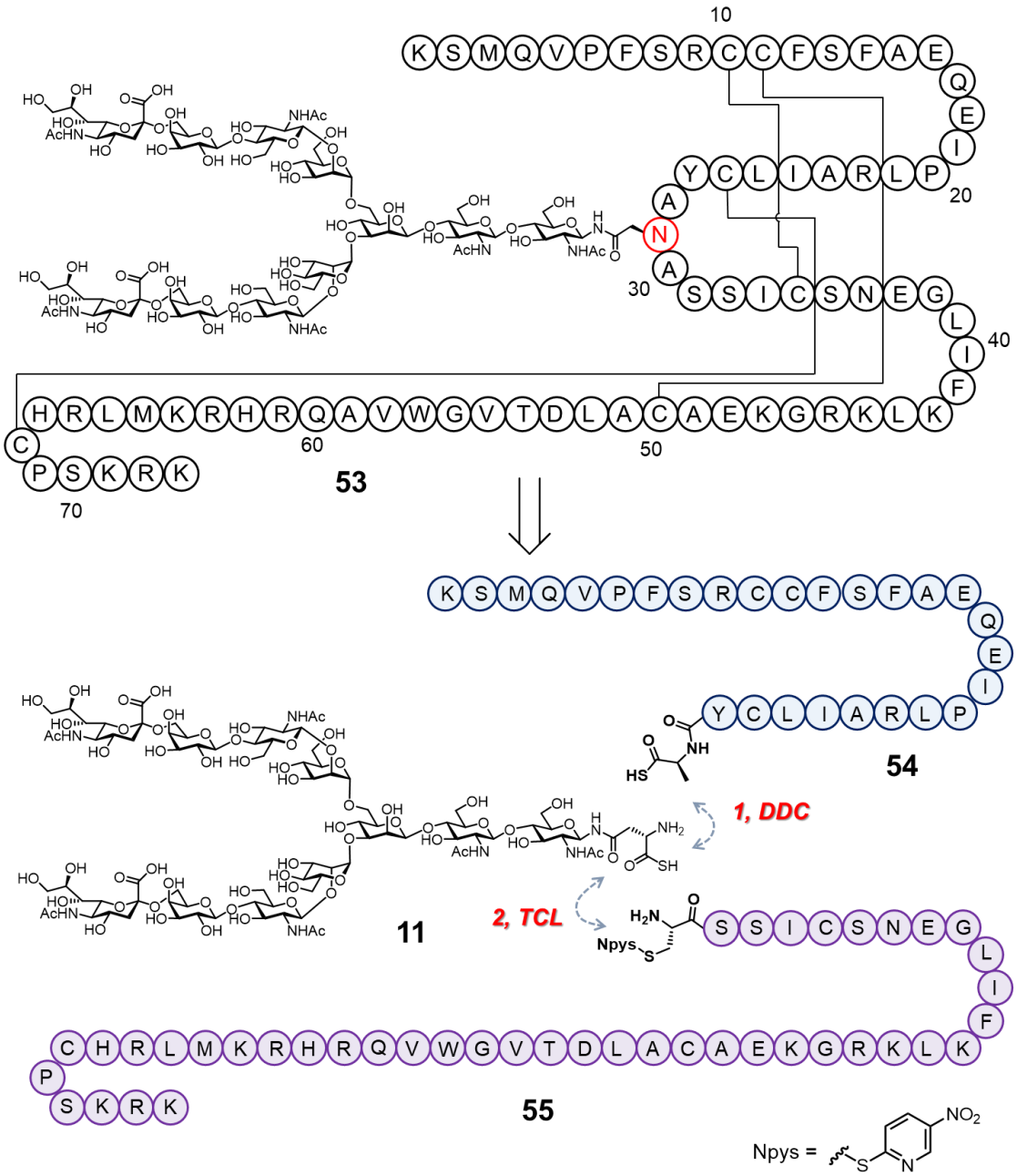


Figure 3-2. Retro synthetic analysis of folded glycosyl CCL1 **53** having biantennary sialyl saccharide.

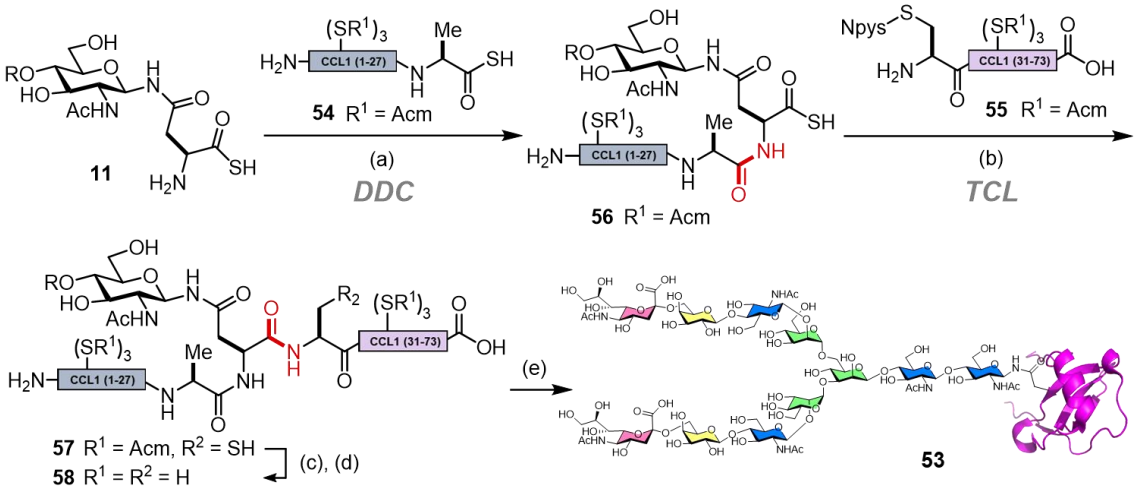
Although the native CCL1 sequence contains Thr at the 30th position, we used a Cys at the 30th position for TCL and a subsequent desulfurization protocol to convert Cys to Ala. We considered that mutation of Thr to the resultant Ala may not have a negative effect on folding processes because the N(glycan)-Thr-Ser sequence is located near the flexible loop position. This flexibility around the glycosylation site has been reported using NMR analysis[83].

According to the developed strategy, segments **54** and **55** were prepared by the Boc[135] and Fmoc[136] SPPS conditions (average yield = 5%), respectively. Then, all internal cysteines of peptides were protected by the acetamidomethyl (Acm) protecting groups because their thiol groups can prevent essential disulfide bond formation in DDC and TCL.

3-3. Synthesis of CCL1

After preparation of peptides, CCL1 **53** having a biantennary sialyl complex type oligosaccharide was synthesized (Scheme 3-1). First a glycosyl asparagine thioacid **11** was coupled to an N-terminal peptide thioacid **54** using DDC to afford the glycopeptide thioacid **56**. Then, the resultant glycopeptide thioacid was coupled with C-terminal peptide having the Npys group **55** utilizing TCL to give a full-length glycoprotein **57**. After desulfurization of Cys^{30} and deprotection of Acm protecting groups afforded the linear glycoprotein polypeptide **58** convergently and oxidative folding conditions gave the desired glycoproteins **53**.

Scheme 3-1. Synthesis of CCL1



Reaction conditions; (a) **54**, DMSO, 7 h, rt. 27%. (b) **55**, phosphate buffer (pH 5.7), 24 h, >90%. (c) VA-044, Mesna, TCEP, phosphate buffer (pH 7.0),14 h, >90%. (d) $PdCl_2$, then piperidine, BME, phosphate buffer (pH 7.0), rt. 71%. (e) Folding, 33%.

At the first, DDC was performed with two equivalents of peptide thioacid **54** and glycosyl asparagine thioacid **11** (15 mM) in 12 hours at room temperature and quenched by the reductant DTT to give a glycopeptide thioacid, Lys^{1}-Asn^{29}(glycan)-COSH **56** with a 27 % isolated yield (Figure 3-3). The reaction progressed through diacyl disulfide intermediates **59** and following SN acyl shift afforded an α-amide bond **60**, then dithioacid form was reduced to yield glycopeptide thioacids **56**. The reaction was monitored by LCMS, and the resultant product was confirmed by LCMS without any aggregation of peptide **54** or **56** in DMSO. The DMSO experiments did not need denaturation conditions for short peptide substrates. The main byproduct was a hydrolyzed form of a starting material peptide thioacid **54**. In order to prevent hydrolysis reactions of peptide thioacids **54** by water molecules, the reaction was progressed under dehydrated solvents using molecular sieves (4A), however the yield of the

reaction was not improved. Although, we also attempted to accelerate the reaction by exposing the reaction system to air, the yield did not improve as well.

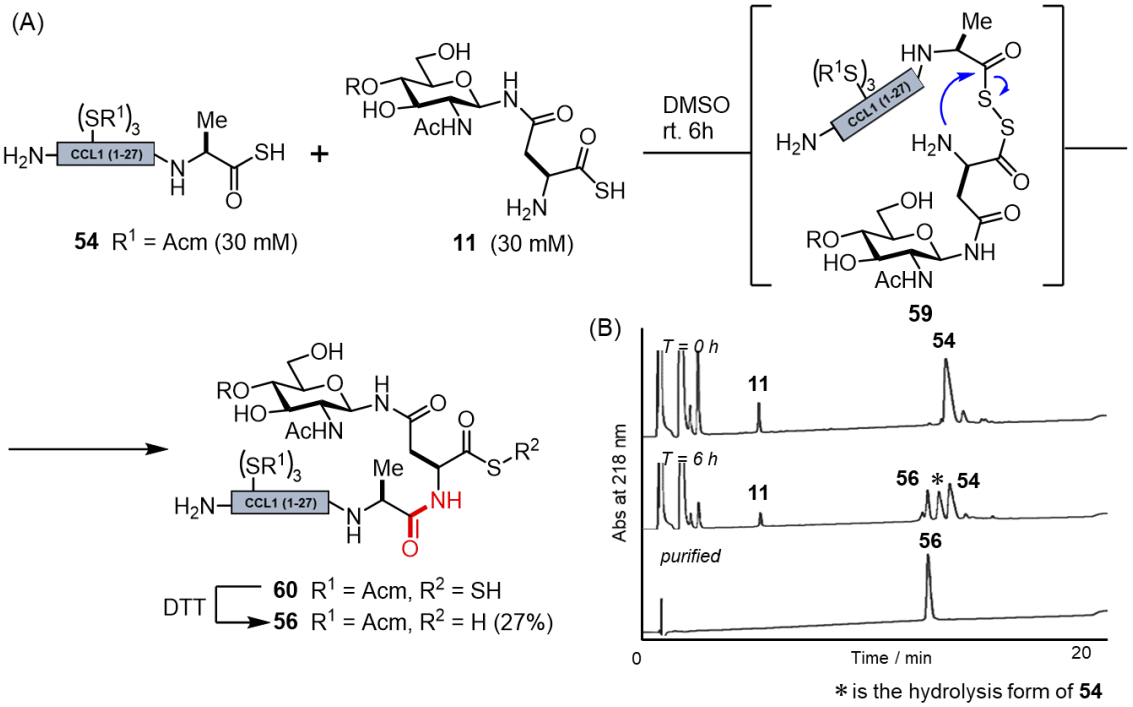


Figure 3-3. (A) DDC with peptide thioacids **54** and glycosyl asparagine thioacids **11**. (B) HPLC monitoring of DDC between glycosyl asparagine thioacid **11** and peptide thioacid **54**. The asterisk indicates peptide-COOH generated by hydrolysis of peptide-thioacid **54**.

After isolation of **56**, TCL was performed as shown in Figure 3-4. The C-terminal peptide Lys[1]-Asn[29](glycan)-COSH **56** and two equivalents of peptide **55** (1.0 mM) were coupled by TCL in a buffer solution (0.2 M sodium phosphate, pH 5.7) containing 6 M guanidine-HCl and subsequently reduced by DTT to yield the protected full-length CCL1 peptide Lys[1]-Asn[29](glycan)-Lys[73] **57** (>90% isolated yield) (Figure 3-4). The reaction progressed through acyl disulfide intermediates **61** and following SN acyl shift afforded an α-amide bond **62**, then disulfide bond was reduced to yield a full-length glycopeptide **57**. The reaction was monitored by LCMS, and the resultant product was confirmed by LCMS. Consequently, the TCL reaction proceeded in high yield even with peptide thioacids having bulky saccharides at C-terminus. In the case of NCL, a ligation with peptide thioester having complex type oligosaccharides at C-terminus was performed by Nishikawa and coworkers, however it was found that NCL did not proceed well due to the side reaction such as an aspartimidation[137]. Even in the case of serine threonine ligation (STL)[138], developed by Liu and coworkers, it has been confirmed that ligation using a peptide salicylaldehyde ester having a glycan near the N and C-terminus does not proceed well to afford ligated peptides. These two examples were attributed to the steric repulsion of a saccharide motif These indicate that TCL is more robust than

other peptide ligation reactions. Furthermore, TCL applying a pyridyl group instead of the Npys group afforded the glycopeptide **57** at less than 50% isolated yield. These facts suggest that the ligation progress quantitatively with glycopeptide is due to the extremely high reactivity of the thioacid groups and the high leaving ability of the Npys group.

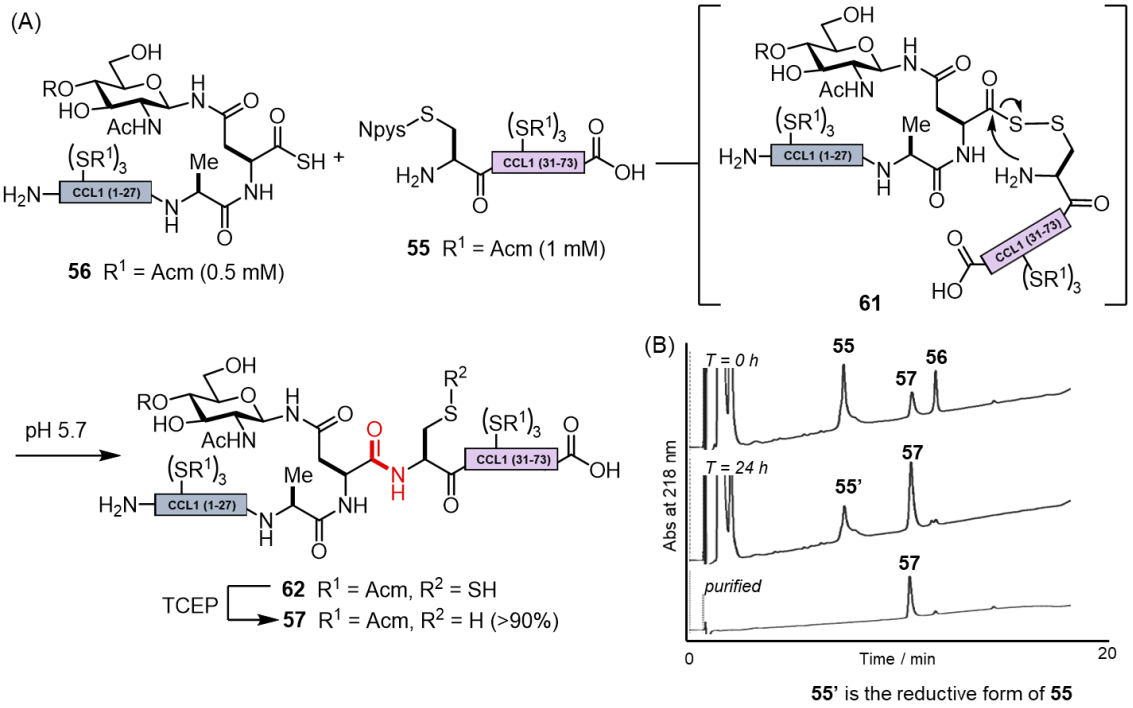


Figure 3-4. (A) TCL with glycopeptide thioacids **56** and peptide **55**. (B) HPLC monitoring of TCL between glycopeptide thioacids **56** and peptide **55**. The compound **55'** indicates a reductive form (having no disulfide at Cys^{30}) of peptide **55**.

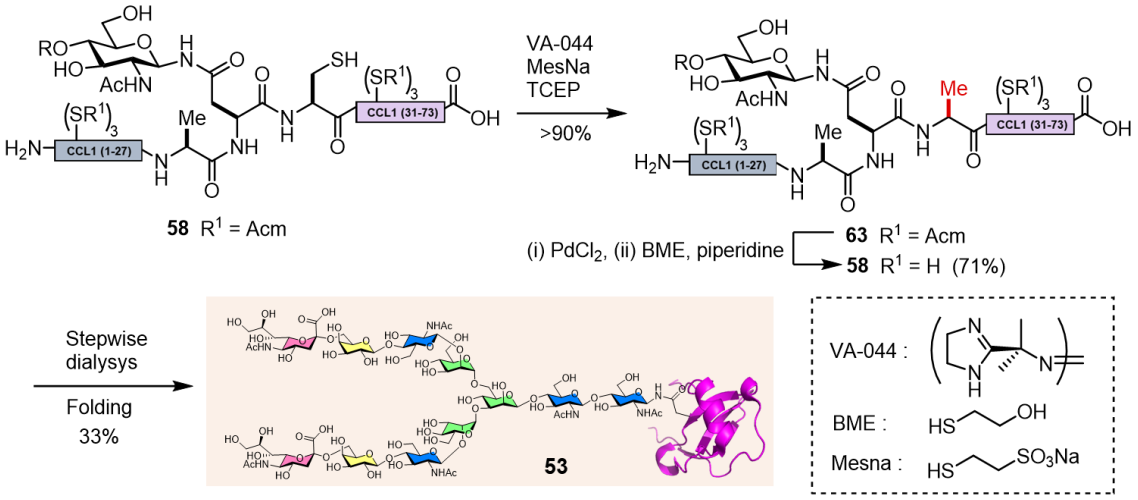


Figure 3-5. Desulfurization, deprotection and folding of full-length glycopeptide **58**.

Desulfurization of the 30th Cys of a full-length glycoprotein **58** with a radical initiator[54] was performed to yield a desulfurized glycopeptide **63** which was converted from Cys to Ala at 30th position in >90% isolated yield (Figure 3-5). Subsequently deprotection of the Acm protecting groups of Cys with $PdCl_2$ [139-140]and the phenacyl protecting group of sialyloligosaccharide with piperidine and 2-mercaptoethanol (BME) were performed to yield glycosyl CCL1 polypeptide **58** in 71% isolated yield (Figure 3-5). Finally, oxidative folding under redox conditions[80] at pH 8.0 yielded folded CCL1 **53** with a biantennary sialyl oligosaccharide at the 29th position in 33% isolated yield.

After isolation of the folded CCL1 **53**, enzyme (chymotrypsin) digestion was performed to analyze disulfide bond positions (Figure 3-7). The circular dichroism (CD) spectrum and the high-resolution mass(10575.9479, average isotopes) supported the correctly folded structure of glycosyl CCL1 **53** (Figure 3-6). These results indicated that the feasibility of synthetic processes dramatically improved compared with that of previous synthesis.

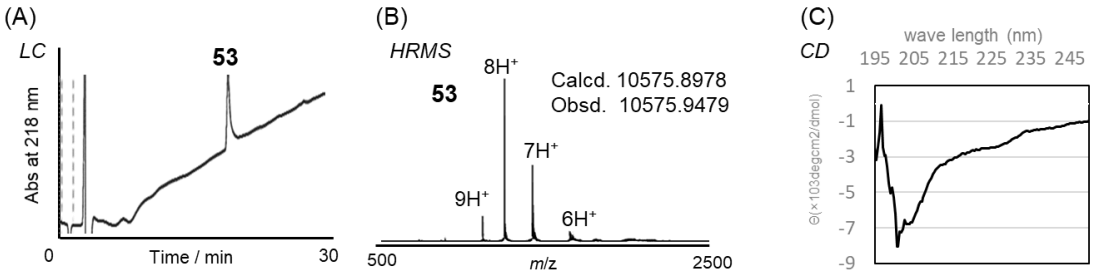


Figure 3-6. (A) HPLC spectrum, (B) HRMS spectrum, CD spectrum of purified folded CCL1 **53**.

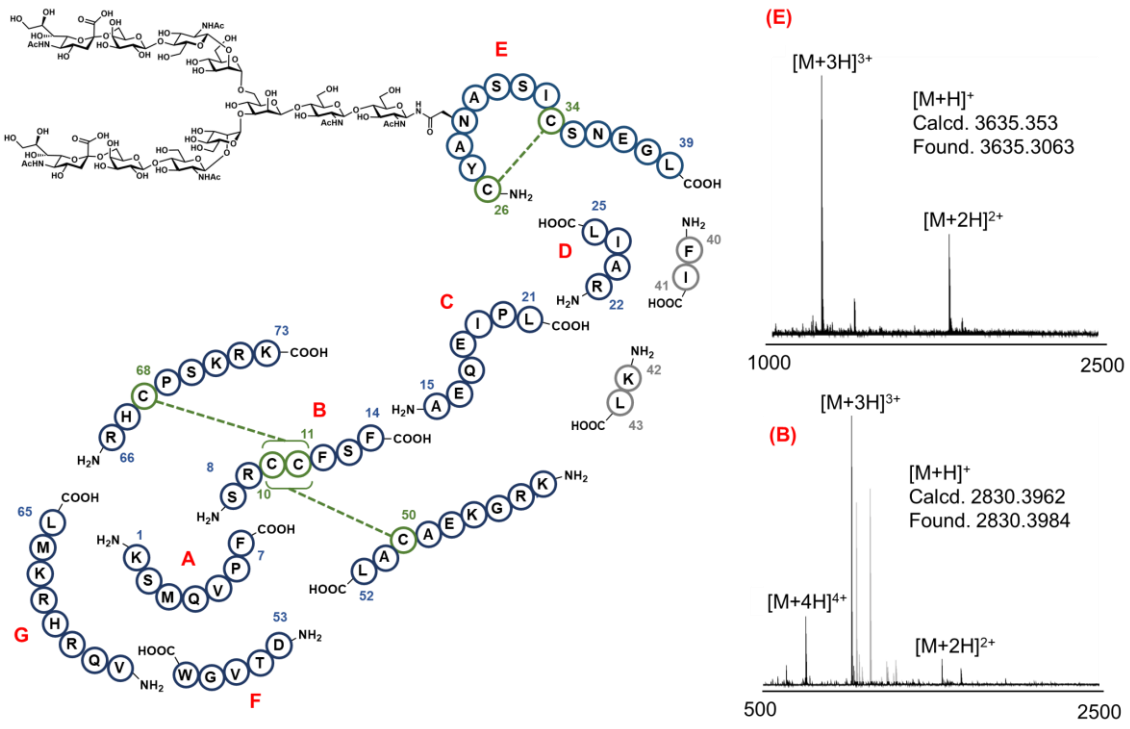


Figure 3-7. Disulfide mapping using the enzyme (chymotrypsin) digestion. MS spectra of the digested fragments of folded CCL1 **53**. Fragment A: Lys^{1}-Phe^{7}. Fragment B: Ser^{10}-Phe^{14} forming disulfide bonds with Lys^{44}-Leu^{52} and Arg^{66}-Lys^{73}. Fragment C: Ala^{15}-Leu^{21}. Fragment D: Arg^{22}-Leu^{25}. Fragment E: Cys^{26}-Leu^{39} forming an internal disulfide bond. Fragment F: Asp^{53}-Trp^{57}. Fragment G: Val^{58}-Leu^{65}.

3-4. Summary

We achieved a convergent synthesis of glycosyl CCL1 **53** having a biantennary sialyl oligosaccharide by using a chemical glycan insertion strategy applied two key reactions, DDC and TCL. This synthetic route did not require SPPS of glycopeptides and succeeded in synthesizing glycoproteins rapidly. Surprisingly, the time taken from the design of the synthetic route to folding experiments in the synthesis of CCL1 was only two weeks. The most time-consuming part of the synthesis is the synthesis of peptide thioacids and peptides having the Npys groups, which were successfully synthesized in a short time by applying the rapid peptide synthesis method developed by Pentelute and coworkers[136]. This synthesis of CCL1 was the first example indicating the effectiveness of the chemical glycan insertion strategy.

Chapter 4. Synthesis of glycoprotein 2: Interleukin 3

4-1. Introduction of chapter 4

After successful synthesis of glycosyl CCL1, we set out to synthesize interleukin 3 (IL3), a cytokine produced by T cells as a regulator of hematopoiesis (Figure 5-1). IL3 was a secreted cytokine first screened and identified as a novel hematopoietic growth factor activity produced by a *gibbon* T cell line using a mammalian cell expression cloning system by Clark and coworkers in 1986 [141]. In terms of the hematopoiesis and the role of cytokines *in vivo*, IL3 has multiple roles for the hematopoietic differentiation [141-143]. Hematopoiesis differentiation pathways[144-146] are shown in the Figure 5-1. IL-3 is especially known to be a very potent primer of basophils in that exposure to IL-3 is associated with large increases in antigen or IgE driven release of leukotrienes and IL4 from basophils.

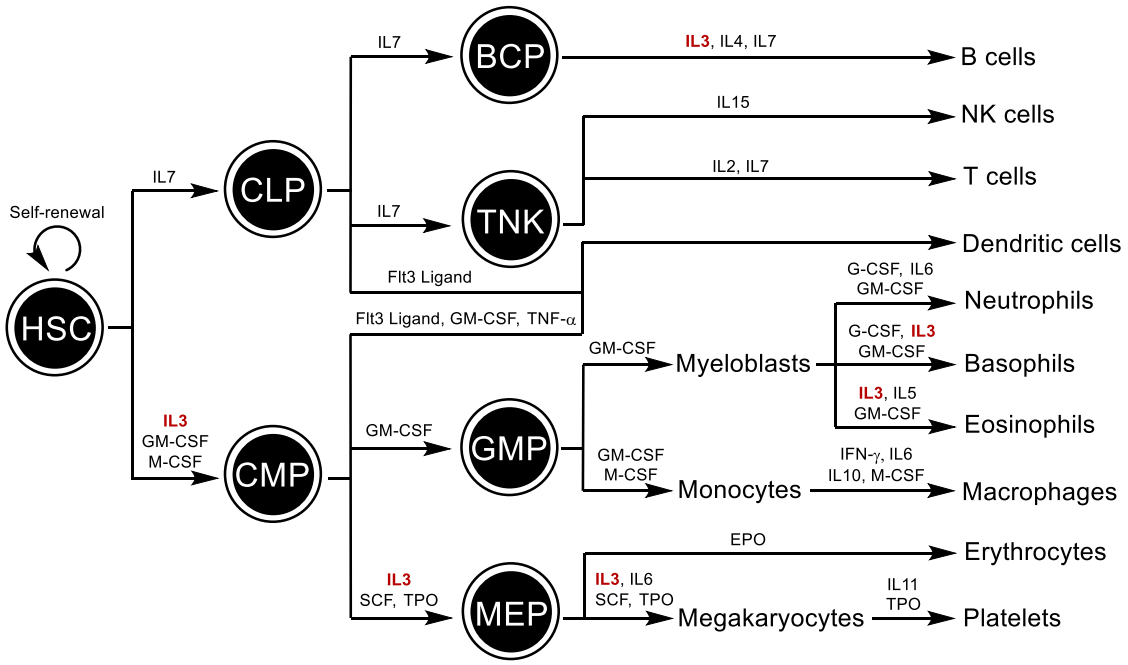


Figure 4-1. Hematopoiesis and the role of cytokines *in vivo*. B-cell progenitor (BCP), common lymphoid progenitor (CLP), common myeloid progenitor (CMP), hematopoietic stem cell (HSC), granulocyte–macrophage progenitor (GMP), megakaryocyte erythroid progenitor (MEP), colony-stimulating factors (CSFs), erythropoietin (EPO), thrombopoietin (TPO), colony forming cells (CFCs), granulocyte–macrophage colony-stimulating factors (GM-CSF). IL3 is indicated as red colored.

IL3 consists of five α-helices and one disulfide bond and molecular weight is 18 kDa (133 amino acids), which is classified as a middle-size cytokine. Normally, the IL3 protein is known to have a complex-type, N-linked sialyloligosaccharide at the 15th asparagine. Glycosylated IL3 is an ideal target to apply our synthetic strategy since there are no synthetic examples to date.

4-2. Synthetic strategy of IL3

In this experiment, we examined the synthesis of IL3 **64** having a biantennary sialyl oligosaccharide by the chemical insertion strategy. Based on glycan insertion strategy, we divided IL3 to three components, glycosyl asparagine thioacids **11** and two peptide segments (Ala^{1}-Ala^{15}-COSH **65** and Cys^{17}(Npys)-Phe^{133} **66**. We designed to synthetic scheme, which insert a glycosyl asparagine thioacid **11** between peptide **65** and peptide **66** using two chemoselective α-amide bond formation reactions, DDC and TCL (Figure 4-2).

Although the native IL3 sequence contains Val at the 14th position, we used alanine at the 14th position to improve DDC yield after observing a low yield. We considered that mutation of Thr to the resultant Ala may not have a negative effect on folding processes because the N(glycan)-Thr-Ser sequence is located near the flexible loop position. This flexibility around the glycosylation site has been reported using NMR analysis[83]. According to the developed strategy, peptide **65** was prepared by the Boc^{135} SPPS conditions (average yield = 5%). And peptide **66** was attempted to prepare from *E.coli.* peptide expression system.

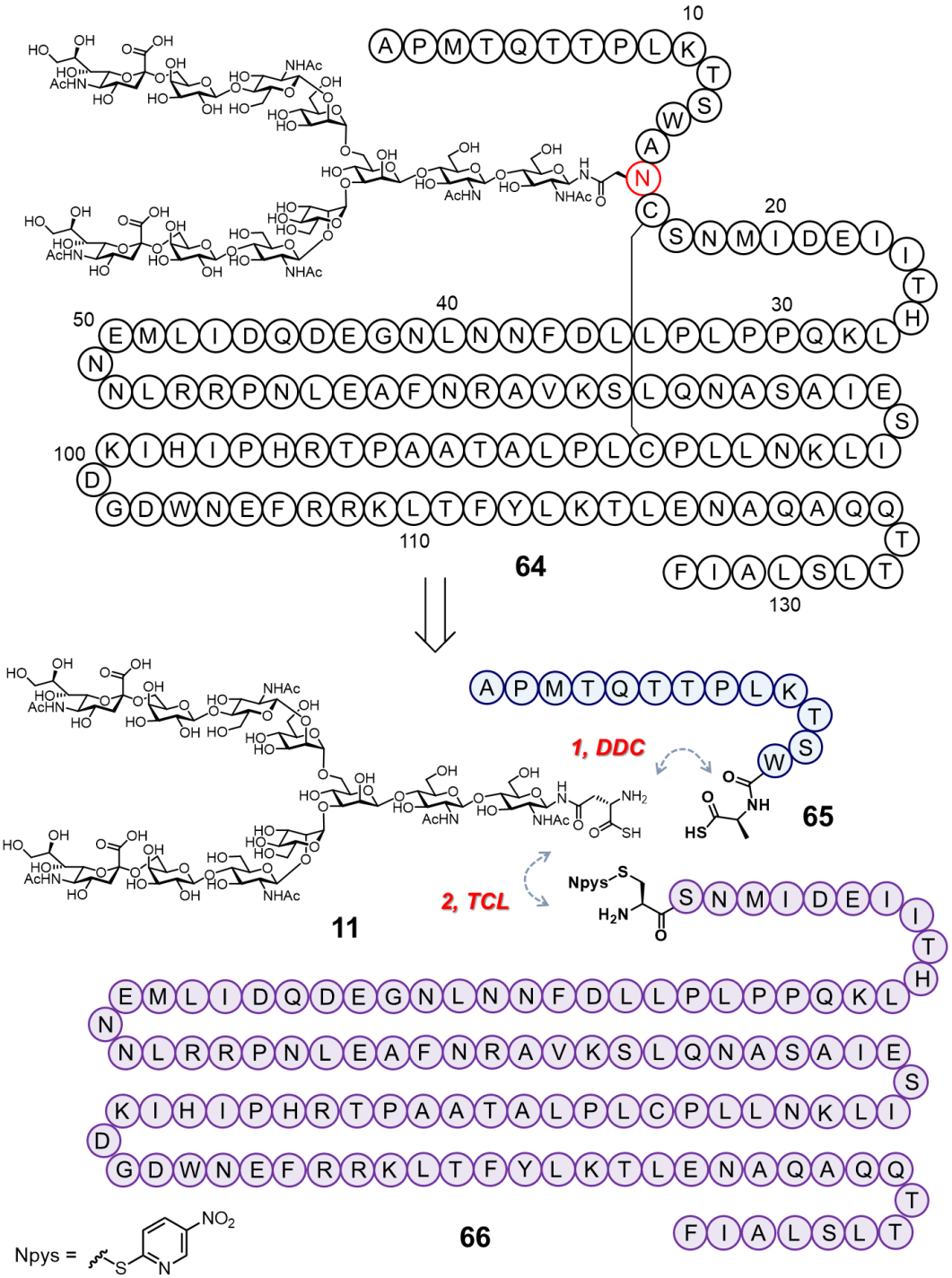
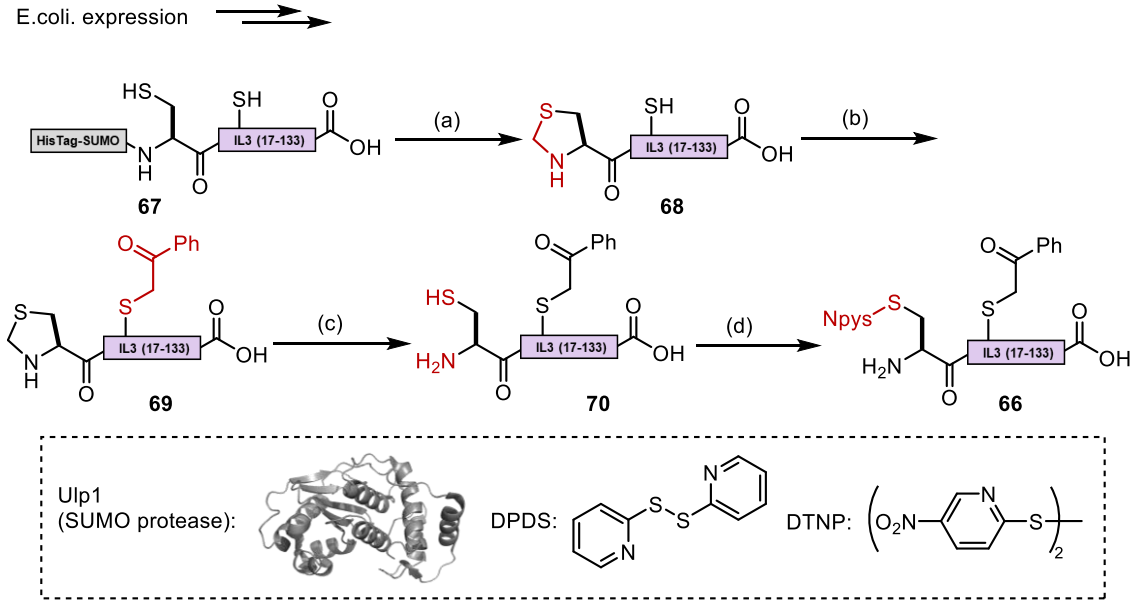


Figure 4-2. Retro synthetic analysis of folded glycosyl IL3 **64** having biantennary sialyl saccharide.

4-3. Synthesis of IL3

The long-chain polypeptide expression by *E. coli*. is an innovative development to prepare large quantities of peptides without using SPPS[86]. Therefore, we decided to apply this *E. coli.* expression system for the synthesis of IL3 **64**. However, in order to utilize the obtained C-terminal peptides in the chemical insertion strategy, it is necessary to protect the internal cysteines and to introduce the Npys leaving groups of the N-terminal cysteine. Then we synthesized C-terminal peptide **66** using the selective thiazolidine formation of the N-terminal Cys and phenacyl protections of the internal Cys (scheme 4-1).

Scheme 4-1. Semisynthesis of C-terminal peptide 66 having the Npys group.



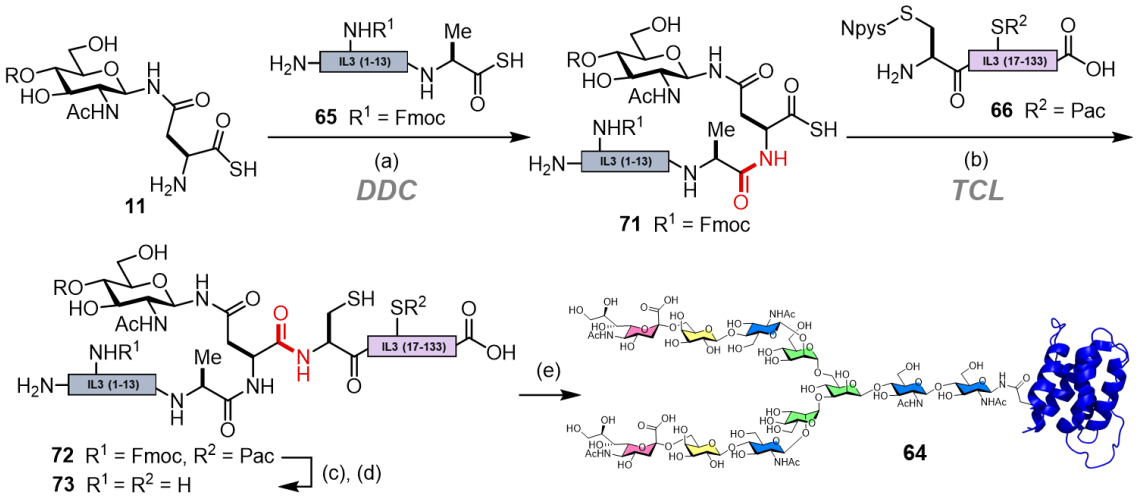
Reaction conditions; (a) Ulp1, formaldehyde, phosphate buffer (pH 8.0), 24 h, 30 ˚C. (b) PacBr, DMF, H_2O, 2.5 h, rt. (c) DPDS, then DTT, 2.5 h, 37˚C 47% (2 steps). (d) 2,2′-dithiobis(5-nitropyridine), H_2O, CH_3CN, 45 min, rt. >90%.

C-terminal peptides **66** was semisynthesized in >40% yield (total 4 steps) by an *E. coli.* expression system and several modification steps. Firstly, the SUMO and His-Tag sequences were cleaved from the isolated polypeptide **67** using SUMO protease (Ulp1) in a buffer solution. The His-Tag sequence was inserted in order to selectively recover only the target polypeptide **67** from the *E. coli*. solution using Ni column. By adding formaldehyde to the SUMO cleavage conditions, peptide **67** selectively formed a thiazolidine at the N-terminus of peptide and afforded N-terminal protected peptide **68**. The internal cysteine of the resulting peptide **68** was protected with the Pac group, and then thiazolidine was opened by DPDS to yield peptide **70** in 47% isolated yield. The thiazolidine ring

opening reaction was developed by Katayama and coworkers[147]. Finally, the target peptide **66** was obtained by forming an Npys disulfide against the thiol of the N-terminal cysteine in >90% isolated yield.

After preparation of peptides, IL3 **64** having a biantennary sialyl complex type oligosaccharide was synthesized (Scheme 4-2). First a glycosyl asparagine thioacid **11** was coupled to an N-terminal peptide thioacid **65** using DDC to afford the glycopeptide thioacid **71**. Then, the resultant glycopeptide thioacid was coupled with C-terminal peptide having the Npys group **66** utilizing TCL to give a full-length glycoprotein **72**. After deprotection of Fmoc protecting groups afforded the linear glycoprotein polypeptide **73** convergently and oxidative folding conditions gave the desired glycoproteins **64**.

Scheme 4-2. Synthesis of IL3



Reaction conditions; (a) **65**, DMSO, 6 h, rt. 34%. (b) **66**, phosphate buffer (pH 5.7), 24 h, >90%. (c) piperidine, CH_3CN, H_2O, 10 min, 28 ˚C. (d) Zinc, MPA, phosphate buffer (pH 5.7), 2 h, rt. 62%. (e) stepwise dialysis folding.

At the first, DDC was performed with two equivalents of peptide thioacid **65** and glycosyl asparagine thioacid **11** (15 mM) in 6 hours at room temperature and quenched by the reductant DTT to yield a glycopeptide thioacid, Ala[1]-Asn[15](glycan)-COSH **71** with a 34 % isolated yield (Figure 4-3). The reaction progressed through diacyl disulfide intermediates **74** and following SN acyl shift afforded glycopeptide having an α-amide bond **75**, then dithioacid form was reduced to yield glycopeptide thioacids **71**. The reaction was monitored by LCMS, and the resultant product was confirmed by LCMS. The main byproduct was a hydrolyzed form of a starting material peptide thioacid **65**.

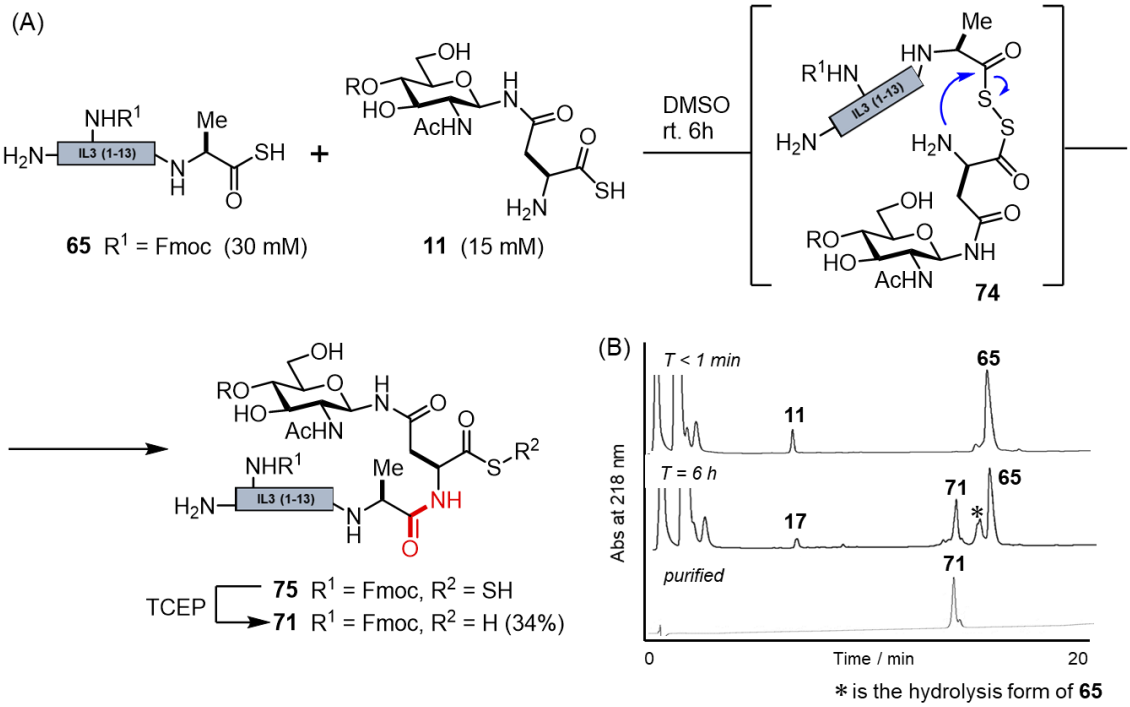


Figure 4-3. (A) DDC with peptide thioacids **65** and glycosyl asparagine thioacids **11**. (B) HPLC monitoring of DDC between glycosyl asparagine thioacid **11** and peptide thioacid **65**. The asterisk indicates peptide-COOH generated by hydrolysis of peptide-thioacid **65**.

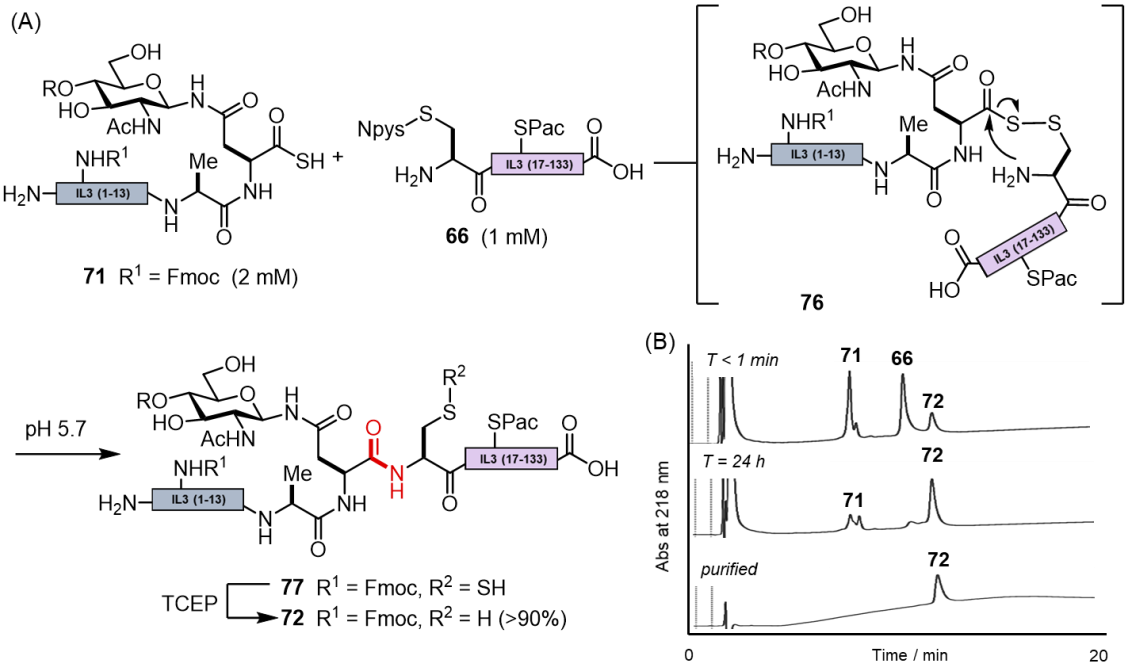


Figure 4-4. (A) TCL with glycopeptide thioacids **71** and peptide **66**. (B) HPLC monitoring of TCL between glycopeptide thioacids **71** and peptide **66**.

After isolation of **65**, TCL was performed as shown in Figure 4-4. Two equivalents of C-terminal peptide Ala[1]-Asn[15](glycan)-COSH **71** and semisynthesized peptide **66** (1.0 mM) were coupled by TCL in a buffer solution (0.2 M sodium phosphate, pH 5.7) containing 6 M guanidine-HCl and subsequently reduced by DTT to yield the protected full-length IL3 peptide Ala[1]-Asn[16](glycan)-Phe[133] **72** (>90% isolated yield) (Figure 4-4). The reaction progressed through acyl disulfide intermediates **76** and following SN acyl shift afforded an α-amide bond **77**, then disulfide bond was reduced to yield full-length glycopeptide **72**.The reaction was monitored by LCMS, and the resultant product was confirmed by LCMS. Consequently, the TCL reaction also proceeded in high yield even with peptide thioacids having bulky saccharides at C-terminus like the case of the synthesis of CCL1.

The full-length glycosyl IL3 Ala[1]-Asn[16](glycan)-Phe[133] **72** was subjected to deprotection of the formyl protecting group and Fmoc protecting group with piperidine, and then the removal of the internal phenacyl group with zinc reduction[148] afforded glycosyl IL3 polypeptide **73** in 62% isolated yield (Figure 4-5). The oxidative folding of a linear glycoprotein **73** employed stepwise dialysis conditions[78] under 6, 3, and 1 M guanidine-HCl solution and a final 10 mM Tris-HCl buffer solution, where a 3 M guanidine-HCl solution contained a mixture of 4.0 mM cysteine and 0.5 mM cystine for disulfide bond formation. The folding processes were monitored by LCMS, and this condition successfully afforded folded glycosyl IL3 **64** with a biantennary sialyl oligosaccharide at the 15th position in 90% isolated yield. After isolation of the folded IL3 **64**, CD spectrum and the high-resolution mass (17268.3508, average isotopes) supported the correctly folded structure of glycosyl IL3 **53** (Figure 4-6).

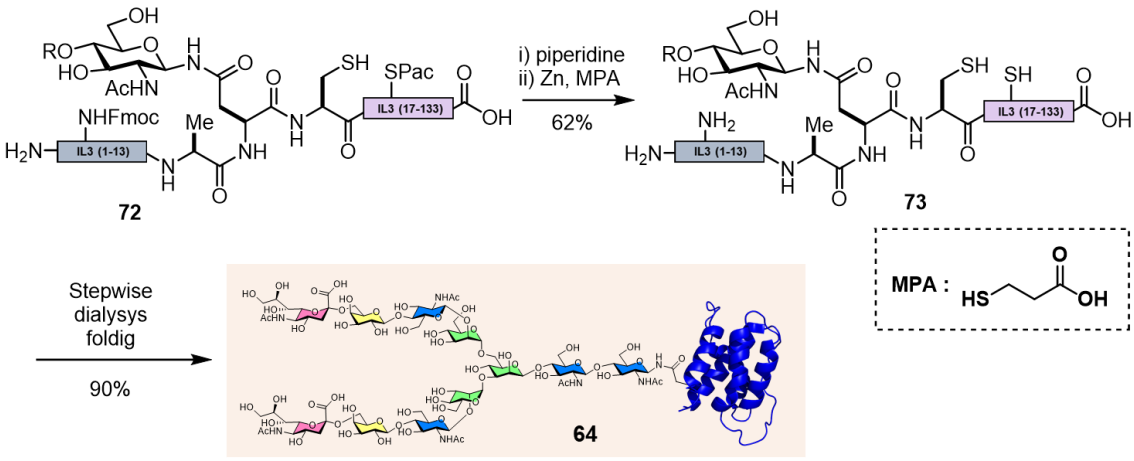


Figure 4-5. Deprotection and folding of full-length glycopeptide **72**.

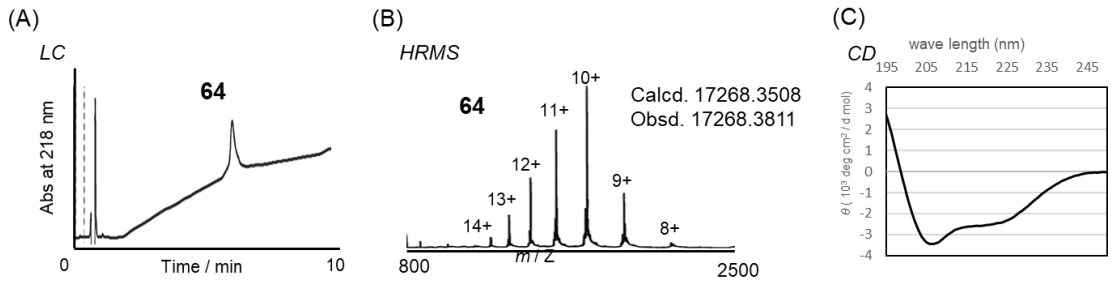


Figure 4-6. (A) HPLC spectrum, (B) HRMS spectrum, CD spectrum of purified folded IL3 **64**.

The *in vitro* bioassay with glycosyl IL3 **64** was performed based on the cell proliferation of TF-1 cell[142], and the activity of **64** was confirmed to be similar to that of commercially available non-glycosylated IL3 **78** expressed in *E. coli.* (Figure 4-7). The biological assay and analytical data, such as the CD and HRMS spectrum supported that glycosyl IL3 **64** employed the native folded structure.

4-4. Biological analysis and discussion

The *in vitro* bioassay with glycosyl IL3 **64** was performed based on the cell proliferation of TF-1 cell [142]. At the result, the activity of glycosyl IL3 **64** was confirmed to be similar to that of commercially available non-glycosylated IL3 **78** expressed in *E. coli.* (Figure 4-7).

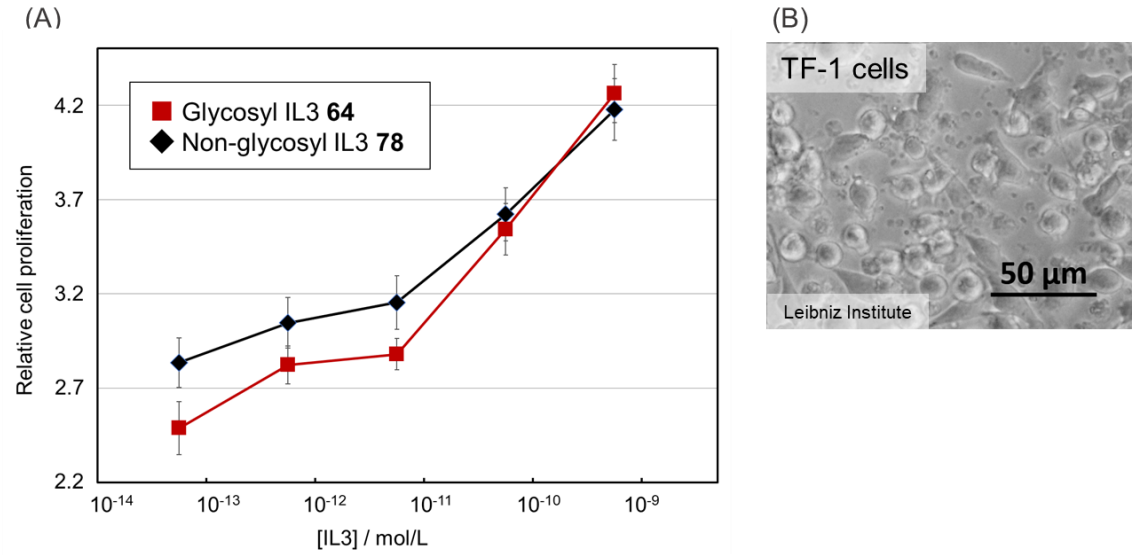


Figure 4-7. (A) TF-1 Cell proliferation assay. TF-1 cells were cultured for 72 hours. Red and black color shows the activity of synthetic glycosylated IL3 **64** and non-glycosyl IL3 **78** expressed in *E. coli.* respectively. Error bars are SD (n = 6). Relative proliferation was calculated by luminescence. (B) a photo of TF-1 cells by *Leibniz Institute*.

In terms of cell proliferation activity, we focused on glycosyl IL3 **64**. Glycosyl IL3 showed a steep slope at a high concentration compared with that of non-glycosyl IL3 **78**. This interesting profile was reproducible. The native N-glycosylation position we implemented was located at the interface between IL3 and its receptor (Figure 4-8). Therefore, the N-glycan appeared to change some chemical properties of IL3 rather than interfering with the formation of the IL3 / IL3 receptor complex (PDB:5UV8).

We hypothesized that the following theory. When IL3 bind to the IL3 receptors, the water molecules that exist between them need to be excluded from the protein binding interface (Figure 4-9; I→II). Hydration shells of N-glycans that can absorb water molecules such as sweet sugar syrup can exclude water molecules between the IL3 / IL3 receptors to accelerate the formation of the complex (Figure 4-9; II→III). In other words, we assumed that water molecules at the interface between the IL3 / IL3 receptor were allowed to be adsorbed by the hydration property of adjacent N-

glycans. This unique water flux may accelerate the formation of the IL3 / IL3 receptor complex. This hydration shell hypothesis is applicable to erythropoietin and its receptor (PDB:1EER), and Maki and coworkers have noted the function of N-glycan hydration shells[69].

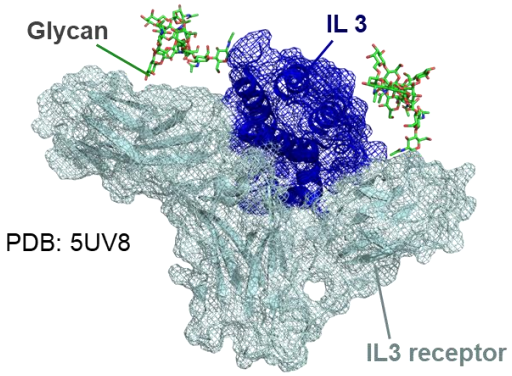


Figure 4-8. 3D structure of IL3 / IL3 receptor. Blue colored structure is the surface model of IL3, and light blue colored structure is the surface model of IL3 receptor. Green colored structure indicates N-linked glycan simulated based on PDB:5UV8.

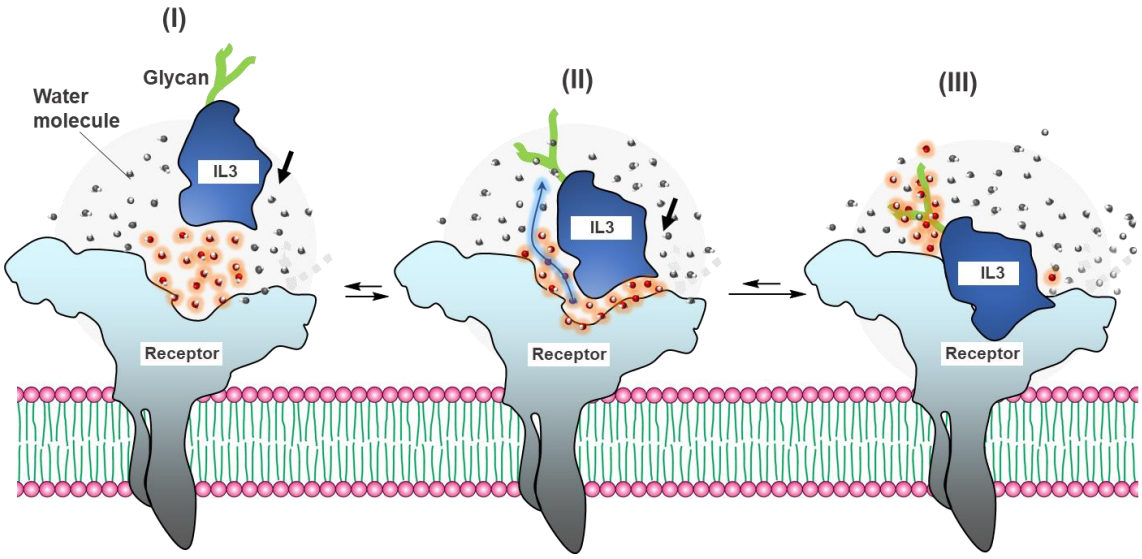


Figure 4-9. The mechanism of the binding of IL3 and IL3 receptor. When IL3 approaches to IL3 receptor (I→II), glycan moiety of IL3 accelerate the formation of the complex (II→III).

In our *in vitro* experiment, the presence or absence of glycans did not affect the activity of IL3. Regarding this, we had to consider IL3 receptor. IL3 receptor also has a N-glycans at the interface with IL3 (Figure 4-8), and these N-glycans may also be able to absorb water molecules and exclude water molecules from the interface (Figure 4-10; II→III). Therefore, we hypothesized that there is

only a slight difference in the biological activity of glycosylated and non-glycosylated IL3 due to the glycan of the IL3 receptors.

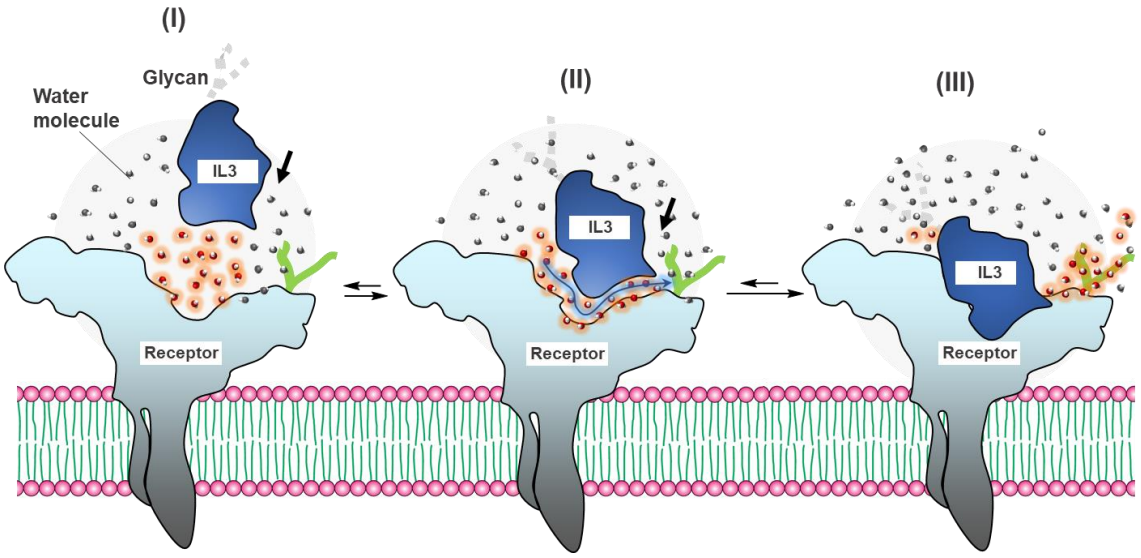


Figure 4-10. The mechanism of the binding of IL3 and IL3 receptor. When IL3 approaches to IL3 receptor (I→II), glycan moiety of IL3 receptor accelerate the formation of the complex (II→III).

In the case of the IL7 / IL7 receptor complex, two N-glycans of the receptor located at the interface with IL7 dramatically changed the binding affinity (PDB:3DI3, 3DI2)[149]. We have originally found many examples in which N-glycosylation determined by genetic sequence analysis based on the glycosylation consensus sequence is often located at the interface between the IL / IL receptor (PDB:1P9M, 6LFM, 1F45, 4GS7, 3TGX, etc.). However, the functions of these N-glycans are still unclear.

4-5. Summary

In *Chapter 4*, we have achieved the semisynthesis of IL3 **64** using the chemical glycan insertion strategy, which is the first synthesis of glycosyl IL3 having homogeneous glycoform. In the semisynthesis of IL3 **64**, we succeeded in efficiently using peptides isolated from *E. coli* for the synthesis of glycoproteins. Bioassays were performed using the obtained glycosyl IL3 **64** and non-glycosyl IL3 **78**. Consequently, new functions of the glycans were discussed. According to our hypothesis, it was suggested that glycan has a role as not only extending the lifespan of proteins but also absorbing water molecules and controlling the protein-protein interactions.

Chapter 5. Synthesis of glycoprotein 3: serine protease inhibitor kazal-type 13

5-1. Introduction of chapter 5

Serine protease inhibitor Kazal type 13 (SPINK13) was a secreted protein first screened and obtained from the hepatoblastoma cell line (HepG2) by Feng and coworkers in 2014 [150]. SPINK13 consists of two domains: a typical Kazal domain with 12-71 amino acids and a Kazal domain with 1-11 amino acid residues. It is known that the Kazal domain has an α-helix, a 3-stranded anti-parallel β-sheet and three disulfide bonds (Figure 5-1A). The Kazal inhibitor has six cysteine residues engaged in disulfide bonds arranged as shown in the following schematic representation (Figure 5-1B). As well as other serine protease inhibitor Kazal type classified as a member of MEROPS inhibitor family I1, clan IA, SPINK13 inhibited the hydrolysis activity of trypsin [150]. Recently this new secreted protein has been studied by some groups as a therapeutic reagent and biomarker for cancer cells due to its function as a tumor suppressor in ovarian[151], hepatocellular[152], and renal[153] cell cancers.

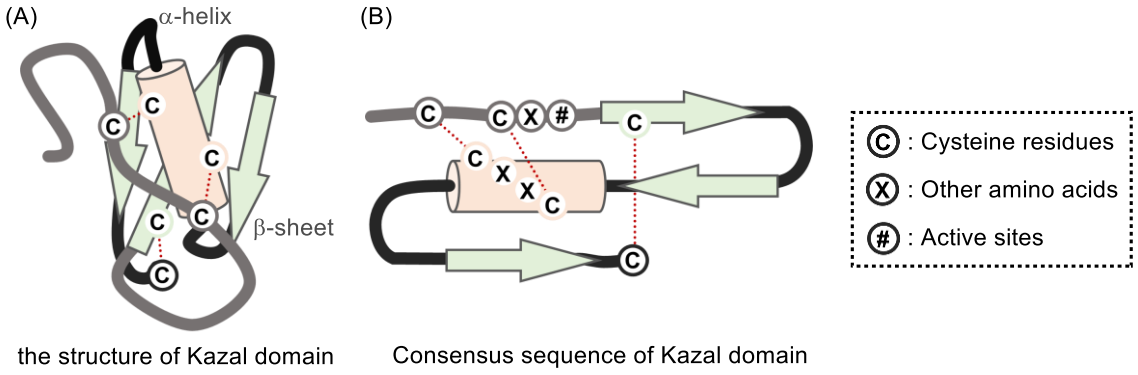


Figure 5-1. (A) Kazal domain, (B) The consensus sequence of Kazal type serine protease inhibitor.

The point that makes SPINK13 unique from other proteins of the SPINK family is that SPINK13 has a consensus sequence for glycosylation in the sequence. Many secreted proteins are usually condensed with N-liked glycans in the consensus sequence (Asn-Xaa-Ser/Thr) in the ER [154], and the Asn32 of SPINK13 seems to have a N-linked sialyl complex type oligosaccharide. In addition, computational predictions by NetNGlyc[155], a glycosylation simulation that has been extensively studied in recent years, show that the possibility of glycosylation at Asn32 is extremely high (>75%). However, there have been no studies discussing the relationships between SPINK13 and the actual protease inhibitory effects caused by glycans remain unclear. The chemical synthesis of SPINK13 glycosides with homogeneous glycan structures is essential to evaluate the biological events caused by glycans of SPINK13. Then, we attempted to synthesize a glycosylated SPINK13 using the chemical insertion strategy. SPINK13 consists of 71 amino acids and three disulfide bonds. The 32th asparagine is in the consensus sequence. Interestingly, the 32th Asn is in the active part that reacts with the catalytic triad (His57, Asp102, Ser195) of the serine proteases when SPINK13 works as an inhibitor.

5-2. Synthetic strategy of SPINK13

In this experiment, we examined the synthesis of SPINK13 **79** having a biantennary sialyl oligosaccharide by the chemical insertion strategy. Based on glycan insertion strategy, we divided SPINK13 to three components, glycosyl asparagine thioacids **11** and two peptide segments (Gly^{1}-Pro^{31}-COSH **80** and βVal^{33}(Npys)-Asp^{71} **81**. βVal means β-mercapto Val. We designed to synthetic scheme, which insert a glycosyl asparagine thioacid **11** between two peptide **80** and **81** (Figure 5-2).

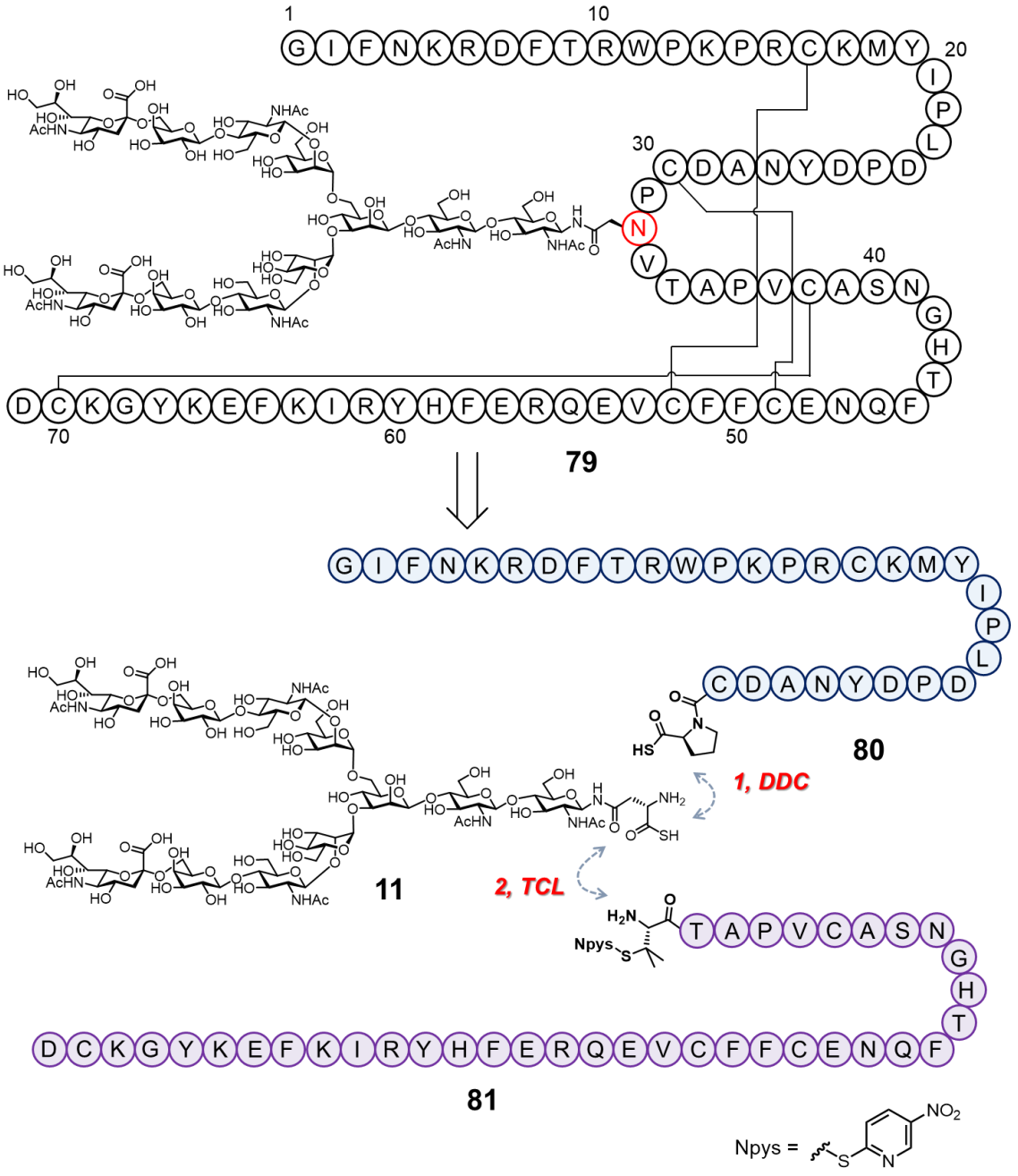


Figure 5-2. Retro synthetic analysis of glycosyl SPINK13 **79** having biantennary sialyl saccharide.

Although the strategy for the synthesis of SPINK13 is same as that for the synthesis of CCL1 and IL3 conceptually, however the amino acids besides the glycosyl Asn at 32th position are completely different. We classified the chemical insertion strategy into two types, type I (CCL1 and IL3) and type II (SPINK13) (Figure 5-2). In type I, a glycosyl asparagine thioacid (**B**) and a peptide thioacid (**A**) are first condensed by DDC, and then cysteinyl peptide (**D**) was coupled by TCL to afford full-length glycoprotein (**E**). Finally, the desired glycoprotein (**F**) can be obtained by desulfurization.

On the other hands, in type II, the first condensation reaction, DDC, undergoes with peptide prolyl thioacid (**G**) and a glycosyl asparagine thioacid (**H**) (Figure 5-2). As I will refer in the next section, it is often difficult to couple peptides in proline ligation sites due to the structural effects of prolines. Therefore, it was necessary to investigate DDC for prolyl thioacids (**H**). The second condensation reaction, TCL, undergoes with a glycopeptide thioacid (**H**) and β-mercapto valine (**I**), which has a mercapto group in the β-position of valine. Since TCL has so far been studied only on cysteine residues, it was necessary to investigate whether thioacids form disulfide bonds with mercapto groups at the β-position in bulky amino acids such as valine. In other words, the synthesis of SPINK13, which is a synthetic strategy (type II) of sandwiching glycosyl asparagine thioacids (**H**) between proline (**G**) and valine (**I**), is much more challenging than type I synthesis such as CCL1 and IL3 because of the difficulty of the two condensation reactions of DDC and TCL.

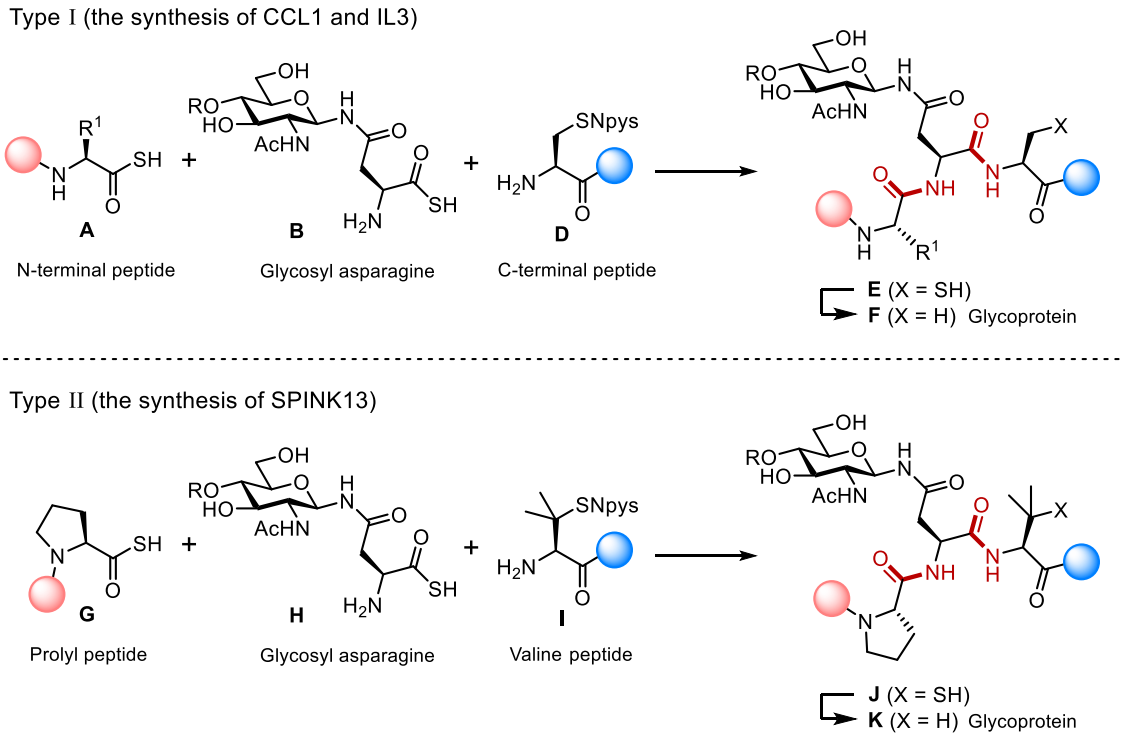


Figure 5-2. Two types of the chemical glycan insertion strategy.

5-3. DDC with proline thioacid

Peptides having proline residues at the C-terminus tends to be avoided for ligation reactions of the chemical synthesis of protein due to the low condensation yields. This is attributed to the unique structure of the proline scaffold. It is often observed the deletion of amino acids with a peptide segment having prolyl thioester in the NCL. Otaka and coworkers reported the deletion side reactions proceed through the formation of diketopiperazine at the peptide C-terminus followed by the elimination of two amino acid residues (Xaa-Pro) (Figure 5-3)[156]. Proposed reaction mechanisms are shown in the figure 5-3. Firstly, under NCL conditions, the peptide thioester (**A**) was converted to diketopiperazine-linked peptide (**B**) followed by the nucleophilic attack of a nitrogen atom of the amide bond at (n+2) position to the carbonyl carbon of a prolyl thioester via intramolecular cyclization. Then elimination reactions of thiol proceed to afford the intermediate (**B**). This diketopiperazine-linked intermediate (**B**) undergoes nucleophilic attack by the coexisting thiol, which leads to the elimination of the diketopiperazine (**D**) to give the deleted peptide thioester (**C**) with two residues removed. Such a side reaction was also found by several groups[157-158], and Hojo and coworkers established a peptide thioesterification method by the cysteinyl prolyl ester (CPE)[159], the reaction mechanism is similar to the side reaction of prolyl thioester. Moreover, peptide thioester (**A**) and two residues deleted peptide thioester (**C**) are difficult to separate by HPLC, and thus byproducts (**C**) become impurities in the target material.

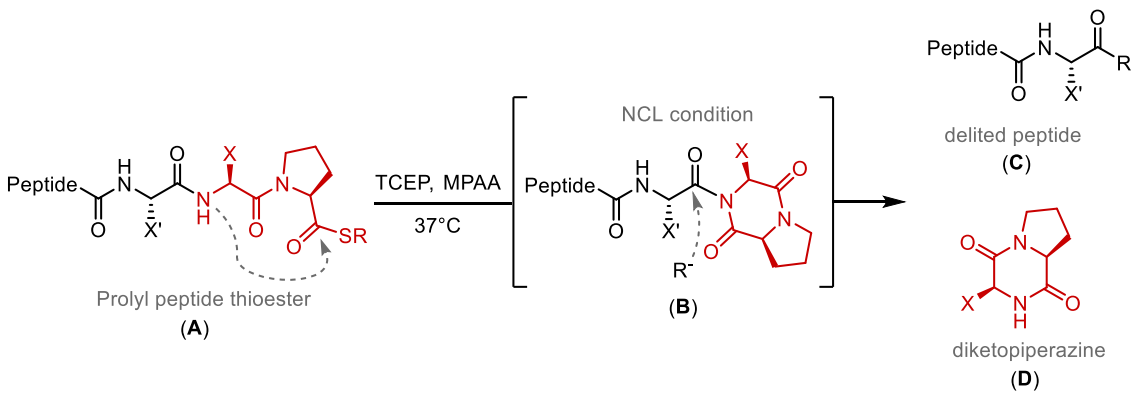


Figure 5-3. Elimination reaction by forming diketopiperazine (DKP) with prolyl peptide thioester.

In order to perform DDC with glycosyl asparagine thioacids and prolyl thioacids, we investigated DDC using model prolyl thioacid **82** and glycosyl asparagine thioacid **7** having mono GlcNAc. This reaction employed one equivalent of glycosyl asparagine thioacid **7** with GlcNAc toward prolyl thioacid **82** (one equivalent) (Figure 5-4). The reaction was proceeded through the formation of diacyl disulfide intermediates **83** and the desired monoglycosylated peptide thioacid **84** formed followed by intramolecular S to N acyl rearrangement as a major product after 9 h in 45%

convergent yield. The reaction progressed without generating other byproducts to yield an α-amide bond. Surprisingly, in contrast to the previous DDC, the hydrolysis of prolyl peptide thioacids **82** was observed in a little amount (~5%), which is the major byproducts of DDC in other cases. Furthermore, comparing to DDC with peptide having Ala or Val at C-terminus (Figure 2-4), the reaction yield of DDC using prolyl thioacids was moderate though the prolyl thioacid has a bulkier backbone especially in the diacyl disulfide intermediates **83**.

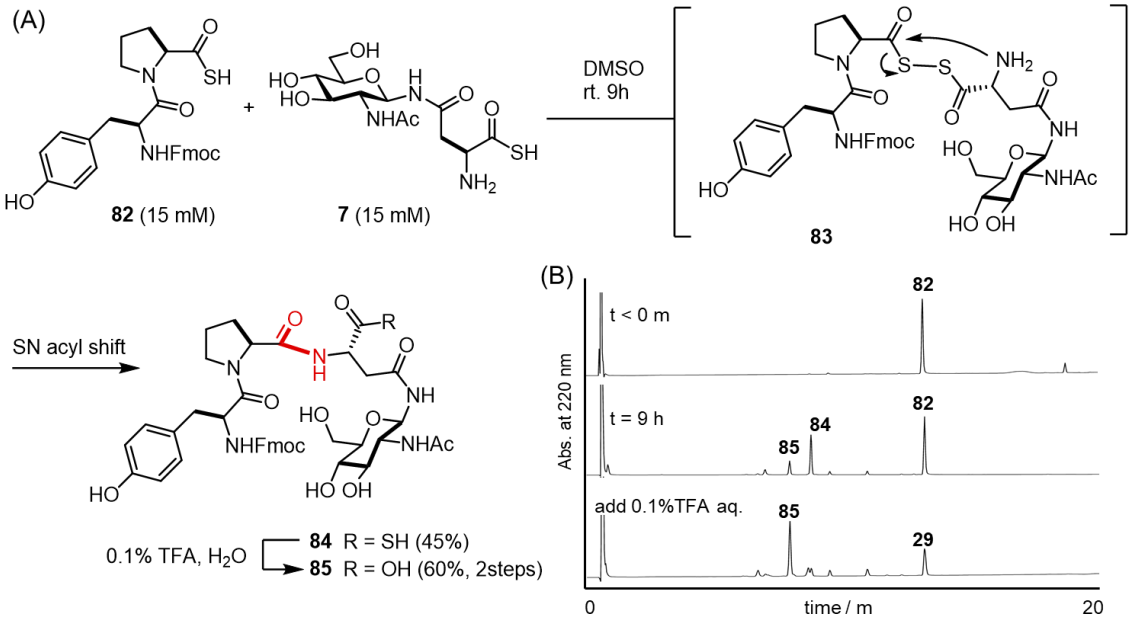


Figure 5-4. DDC for prolyl peptide thioacid. (A) DDC with prolyl peptide thioacids **82** and glycosyl asparagine thioacids **7** having mono GlcNAc. (B) HPLC monitoring of DDC for prolyl thioacids.

Next, we evaluated the epimerization of DDC with using prolyl thioacid **82**. In this case, glycopeptide thioacids **84** are unsuitable for analysis by LC because asparaginyl thioacids are unstable functional groups in an aqueous solution and are easily hydrolyzed or converted to aspartimides. Therefore, we hydrolyzed the glycopeptide thioacids **84** to obtain glycopeptide **85** by adding acidic water containing 0.1% TFA after the DDC reaction (Figure 5-4). The hydrolysis reaction proceeded rapidly and afforded glycopeptide **85** having mono GlcNAc in 60% yield (2 steps). Finally, the retention time of LC spectrum of glycopeptide **85** was compared with those of the authentic glycopeptide **(R)-85** and **(L)-85** synthesized by a conventional SPPS protocol (Figure 5-5). **(R)-85** is consisted of N-Fmoc-*L*-Tyr-*D*-Pro-*L*-Asn(GlcNAc)-OH (Figure 5-5 BI), **(S)-85** is consisted of N-Fmoc-*L*-Tyr-*L*-Pro-*L*-Asn(GlcNAc)-OH (Figure 5-5 BII). The LC analysis showed that the LC peaks of the glycopeptides **85** obtained by DDC had the same retention time as **(S)-85**, and the absence of other peaks at the retention time of **(R)-85** confirmed that epimerization did not proceed in DDC

with prolyl thioacid **82**. This result indicated that DDC with prolyl thioacid proceeded in moderate yield and epimerization does not proceed. Thus, the results of epimerization experiments using prolyl thioacids showed that prolyl thioacids can be used in DDC and our chemical glycan insertion strategy.

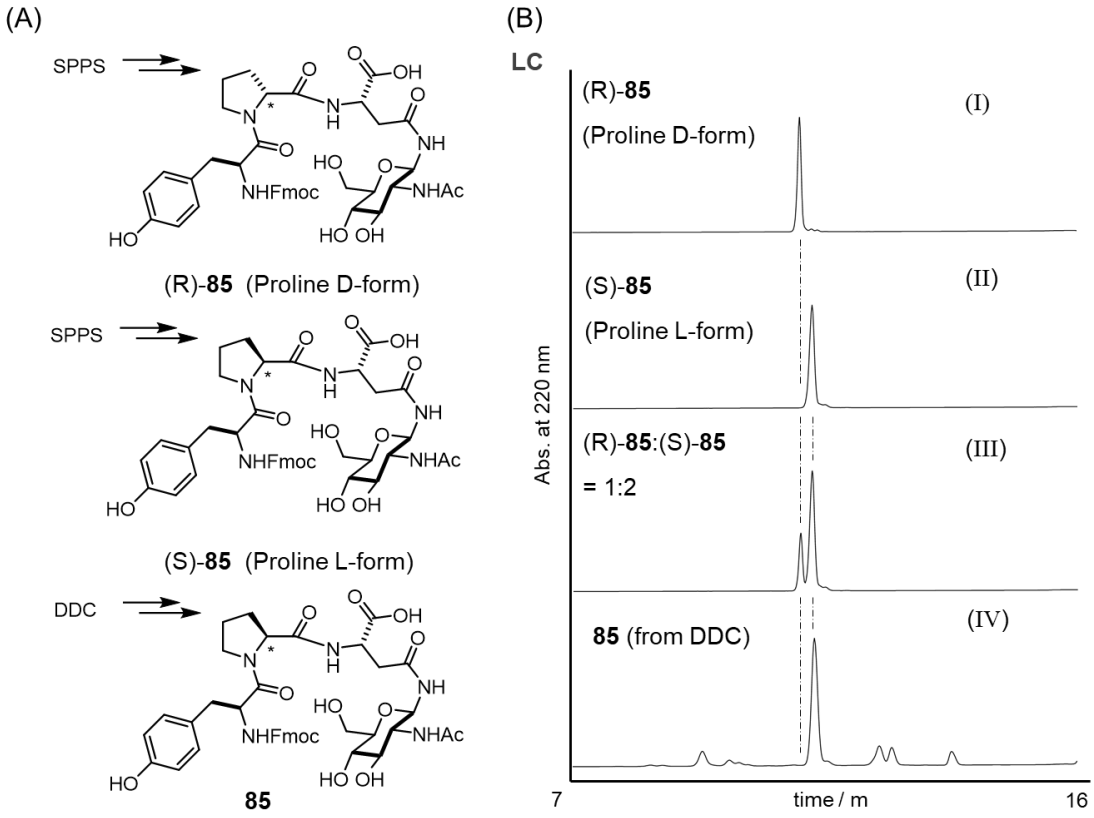
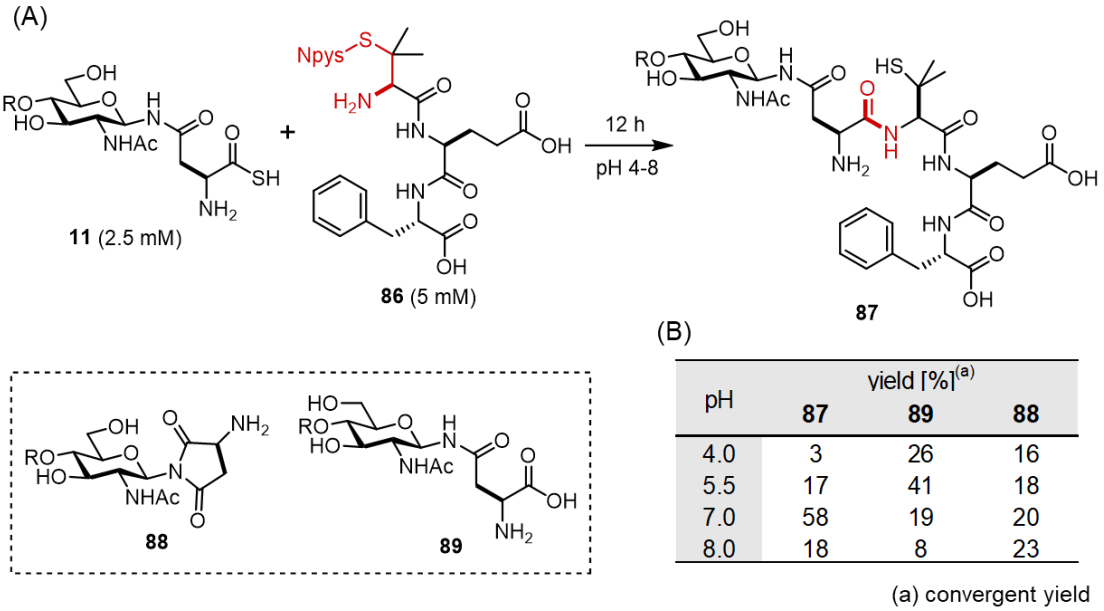


Figure 5-5. LC profiles for monitoring epimerization of DDC. (A) Structure of glycopeptide **(R)-85** synthesized by SPPS, **(S)-85** synthesized by SPPS, **(S)-85** synthesized by DDC. (B) LC profiles of **(R)-85** synthesized by SPPS, **(S)-85** synthesized by SPPS, **85** synthesized by DDC.

5-4. TCL with β-mercaptovaline

The amino acid with a mercapto group at the β-position and the peptide ligations using β-mercapto amino acid derivatives have been so far achieved by many groups. The β-mercapto amino acids can form α-amide bonds through thiol capture reactions and subsequent intramolecular SN acyl rearrangements. However, there are no examples which examined TCL using β-mercapto amino acids, which is a key reaction in our synthetic strategy. Therefore, we first tested TCL with a valine derivative having a mercapto group at the β-position (β-mercapto valine).



pH	yield [%][(a)] 87	89	88
4.0	3	26	16
5.5	17	41	18
7.0	58	19	20
8.0	18	8	23

(a) convergent yield

Figure 5-6. TCL with peptide **86** having β-mercapto valine at N-terminus. (A) TCL reaction with peptide **86** having β-mercapto valine at N-terminus and glycosyl asparagine thioacid **11**. (B) Convergent yield of TCL in varied pH conditions. Compound **88** is an aspartimide of glycosyl asparagine thioacid **11** and compound **89** is hydrolyzed form of glycosyl asparagine thioacid **11**.

In order to perform TCL with glycosyl asparagine thioacids and varinyl peptide, we investigated TCL using model peptide **86** having the Npys group at β-position of N-terminal valine and glycosyl asparagine thioacid **11** having a complex type oligosaccharide. This reaction employed one equivalent of glycosyl asparagine thioacid **11** toward two equivalent of peptide **82** (Figure 5-6). The reaction was proceeded in 12 hours and the desired glycopeptide **87** was formed followed by intramolecular S to N acyl rearrangement as a major product after in 17% convergent yield at pH 5.5 (Figure 5-6B). In case of the TCL for the synthesis of CCL1 (*Chapter 3*) and IL3 (*Chapter 4*) with cysteinyl peptides at pH 5~6 proceeded in high yield (>90%). However, in the case of β-mercapto

valine, the yield was very low in the same conditions. At the same time, the hydrolysis of glycosyl asparagine thioacid **11** proceeded as a side reaction. Then, TCL was performed at various pH conditions (Figure 5-6B). The results showed that the hydrolysis of thioacids **11** progressed more at lower pH conditions, and the formation of aspartimides of thioacids **11** progressed at higher pH conditions. Consequently, the reaction proceeded with the highest yield by aqueous buffer (pH 7.0) .

In terms of the reaction yield in various pH conditions, the nucleophilicity of a thioacid and the ability to form disulfide bonds should be considered (Figure 5-7). Since the pKa of thioacids is about 3, the reaction can be carried out at a lower pH with conventional cysteinyl peptide having the Npys groups. However, with a peptide **86** having dimethyl group at the β-position of N-terminal amino acids such as valine, the reaction did not proceed well in the low pH conditions (Figure 5-7 red colored). This might be because the disulfide exchange reaction between glycosyl asparagine thioacid **11** and peptide **86** did not proceed due to the steric repulsion of dimethyl groups of valines and the followed acyl disulfide intermediate (**A**) was not generated. On the other hand, in high pH conditions TCL also did not work well due to the generation of byproducts such as aspartimides **88** of glycosyl asparagine thioacid **11** (Figure 5-7 blue colored). This is because the nucleophilicity of the nitrogen atom of the innermost GlcNAc amide was increased by the high pH conditions, and acyl disulfide intermediates (**A**) cyclized intramolecularly to afford a glycosyl asparagine aspartimide **88**. According to the experimental results, it was found that the yield of TCL using amino acid derivatives with a external mercapto group at the β-position was lower than that of conventional TCL, and the aqueous condition (pH 7.0) gave the highest yield to afford an α-amide bond.

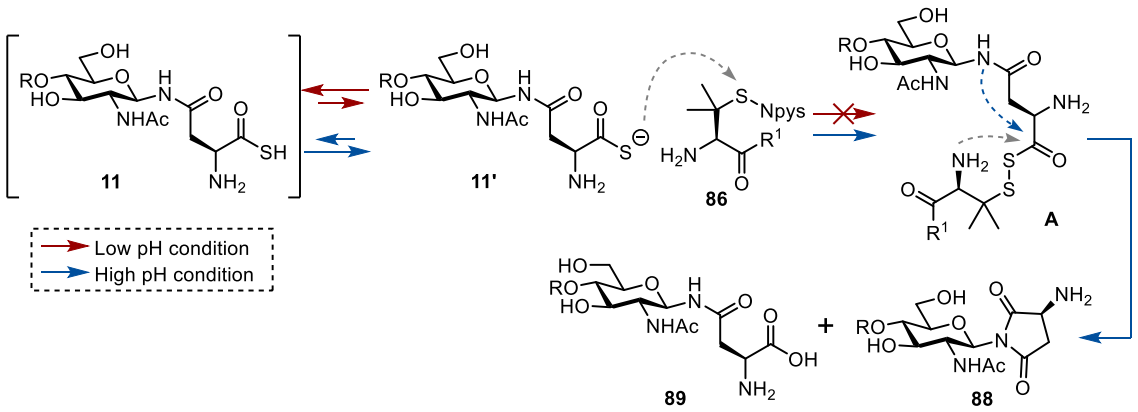
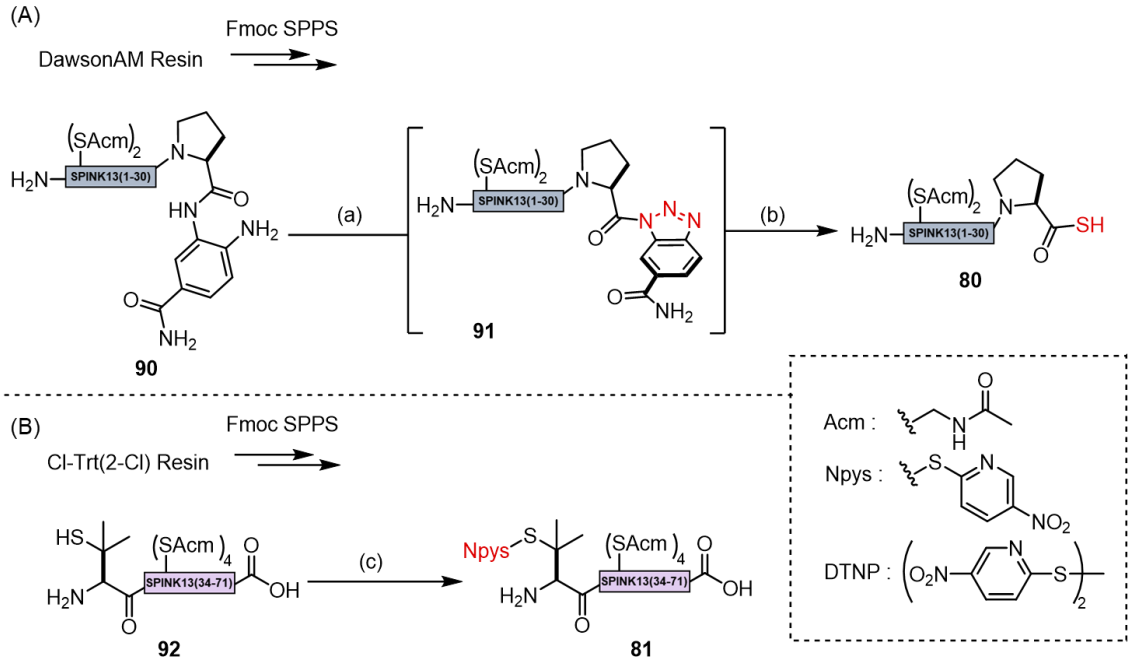


Figure 5-7. Proposed reaction mechanism of TCL in low pH conditions (red colored) and in high pH conditions (blue colored).

5-5. Synthesis of SPINK13

Experiments including DDC with prolyl peptide thioacids and TCL with peptide having β-mercapto valine at N-terminus enabled us to the synthesis of SPINK13 using chemical glycan insertion strategy. Therefore, we first attempted to synthesize two peptides **80** and **81**sandwiching a glycosyl asparagine thioacid **11** according to our synthetic strategy (Figure 5-2). Both of N and C terminal precursor peptides **90** and **92** were synthesized by the rapid Fmoc SPPS protocol[136] (Scheme 5-1).

Scheme 5-1. The synthesis of peptide thioacid 80 and peptide having β-mercapto Npys valine 81.



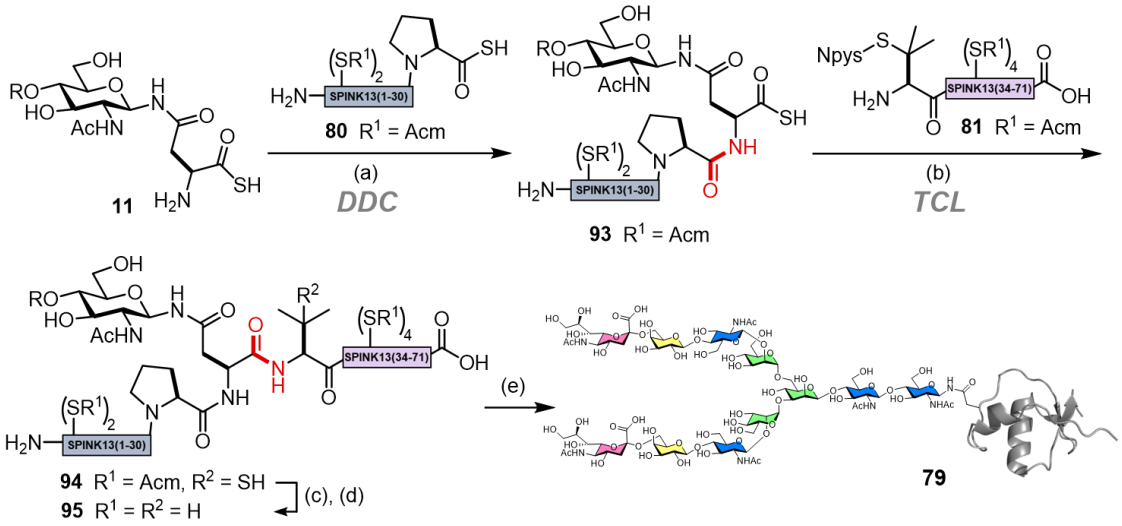
Reaction conditions; (a) $NaNO_2$, phosphate buffer (pH 5.0), 10 min, -10˚C. (b) Na_2S, phosphate buffer (pH 7.0), 10 min, rt. 85% (2 steps). (c) 2,2′-dithiobis(5-nitropyridine), H_2O, CH_3CN, 12 h, rt. 83%.

N-terminal peptide prolyl thioacid **80**, Gly^1-Pro^{31}-COSH, was synthesized from prolyl peptide *o*-aminoanilides **90** (Scheme 5-1). Peptide *o*-aminoanilides **90** was mildly activated by $NaNO_2$ at -10 ˚C in aqueous solution for 10 minutes to afford peptide benzotriazole intermediates **91**, and then subsequent addition of sodium sulfide furnished peptide thioacids **80** in 85% isolated yield (Scheme 5-1). The methods of the synthesis of peptide thioacids from peptide having carboxylic acids[160-161], thioesters[162], hydrazides[163], benzotriazoles[164-165] and intein[166] have achieved by many groups. However, we believe that our protocol is more robust because the desired peptide thioacids was able to be obtained from stable peptide *o*-aminoanilides on demand in about 20 minutes. In addition, in our synthetic method the formation of diketopiperazines does not occur. Combined with Fmoc rapid SPPS

protocol, we can obtain the desired peptide thioacids in a few hours. C-terminal peptide **81**, βVal[33](Npys)-Asp[71], was synthesized from peptide **92**. the target C-terminal peptide **81** was obtained by forming an Npys disulfide against the thiol of the N-terminal β-mercapto valine in 83% isolated yield. Compared to the formation of the Npys disulfide bond with cysteinyl peptides in the synthesis of IL3 and CCL1, the peptides with β-mercapto valine took longer to form disulfide bonds. This is attributed to the dimethyl group at the β-position.

After preparation of peptides, SPINK13 **79** having a biantennary sialyl complex type oligosaccharide was synthesized (Scheme 5-2). First a glycosyl asparagine thioacid **11** was coupled to an N-terminal peptide prolyl thioacid **80** using DDC to afford the glycopeptide thioacid **93**. Then, the resultant glycopeptide thioacid **93** was coupled with C-terminal peptide having the Npys group at β-mercapto valine **81** utilizing TCL to give a full-length glycoprotein **94**. After desulfurization and deprotection of Acm protecting groups afforded the linear glycoprotein polypeptide **95** convergently and stepwise dialysis folding conditions gave the desired glycoproteins **79**.

Scheme 5-2. Synthesis of SPINK13



Reaction conditions; (a) **80**, DMSO, 12 h, rt. 28%. (b) **81**, phosphate buffer (pH 7.0), 8 h, 59%. (c) VA-044, Mesna, TCEP, phosphate buffer (pH 7.0),14 h, 80%. (d) $PdCl_2$, then piperidine, BME, phosphate buffer (pH 7.0), rt. 78%. (e) Stepwise dialysis folding, 65%.

At the first, DDC was performed with two equivalents of peptide prolyl thioacid **80** and glycosyl asparagine thioacid **11** (15 mM) in 12 hours at room temperature and quenched by the reductant TCEP to yield a glycopeptide thioacid, Gly[1]-Asn[32](glycan)-COSH **93** with a 28 % isolated yield (Figure 5-8). The reaction progressed through diacyl disulfide intermediates **96** and following S-N acyl shift afforded glycopeptide having an α-amide bond **97**, then dithioacid form was reduced to

yield glycopeptide thioacids **93**. The reaction was monitored by LCMS, and the resultant product was confirmed by LCMS. The main byproduct was a hydrolyzed form of a starting material peptide thioacid **80**. Surprisingly, in contrast to the previous DDC, the hydrolysis of prolyl peptide thioacids **80** was not observed, which is the major byproducts of DDC in other cases such as the synthesis of CCL1 and IL3. This might be attributed to the unique character of prolyl thioacids.

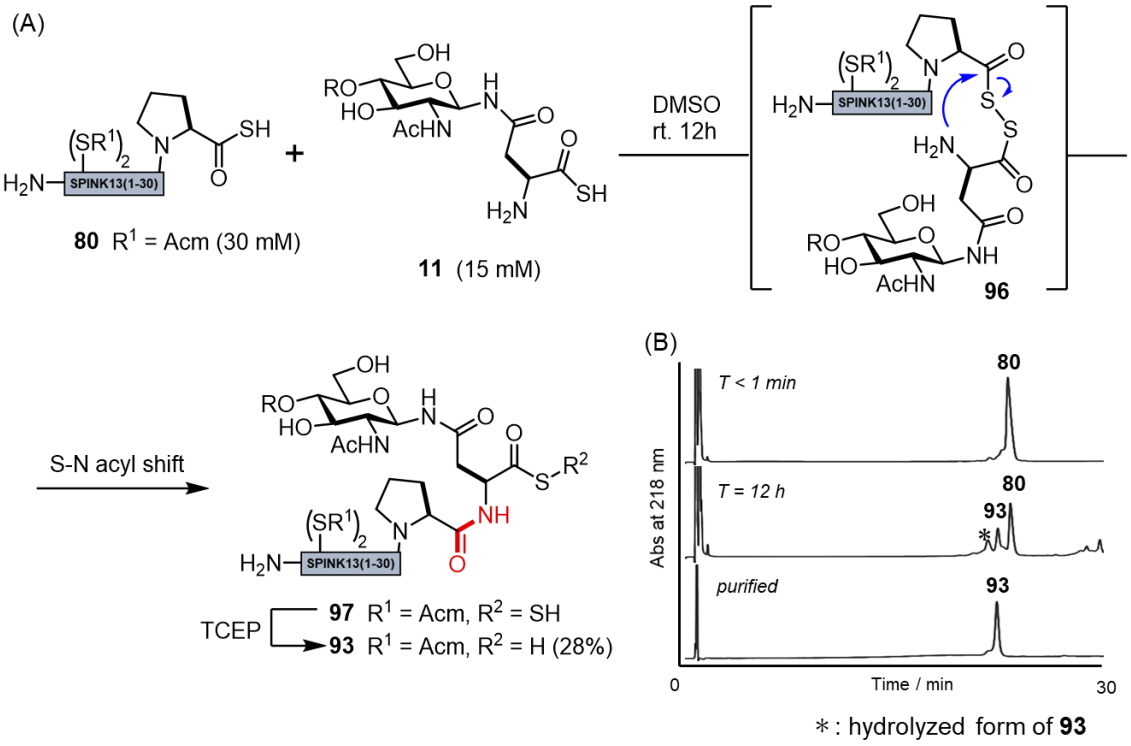


Figure 5-8. (A) DDC with peptide prolyl thioacids **80** and glycosyl asparagine thioacids **11**. (B) HPLC monitoring of DDC between glycosyl asparagine thioacid **11** and peptide prolyl thioacid **80**. The asterisk indicates glycopeptide-COOH generated by hydrolysis of target **93**.

After isolation of **93**, TCL was performed as shown in Figure 5-9. C-terminal peptide Gly[1]-Asn[32](glycan)-COSH **93** and two equivalents of peptide having the Npys group at β-mercapto valine **81** (5.0 mM) were coupled by TCL in a buffer solution (0.2 M sodium phosphate, pH 7.0) containing 6 M guanidine-HCl and subsequently reduced by TCEP to yield the protected full-length SPINK13 peptide Gly[1]-Asn[32](glycan)-Asp[71] **94** (59% isolated yield) (Figure 5-9). The reaction progressed through acyl disulfide intermediates **98** and following S-N acyl shift afforded an α-amide bond **99**, then disulfide bond was reduced to yield full-length glycopeptide **94**.The reaction was monitored by LCMS, and the resultant product was confirmed by LCMS. Consequently, the TCL reaction also proceeded in moderate yield at pH 7.0. Although TCL was performed in other pH conditions, the yield of full-length glycoprotein **94** was not improved.

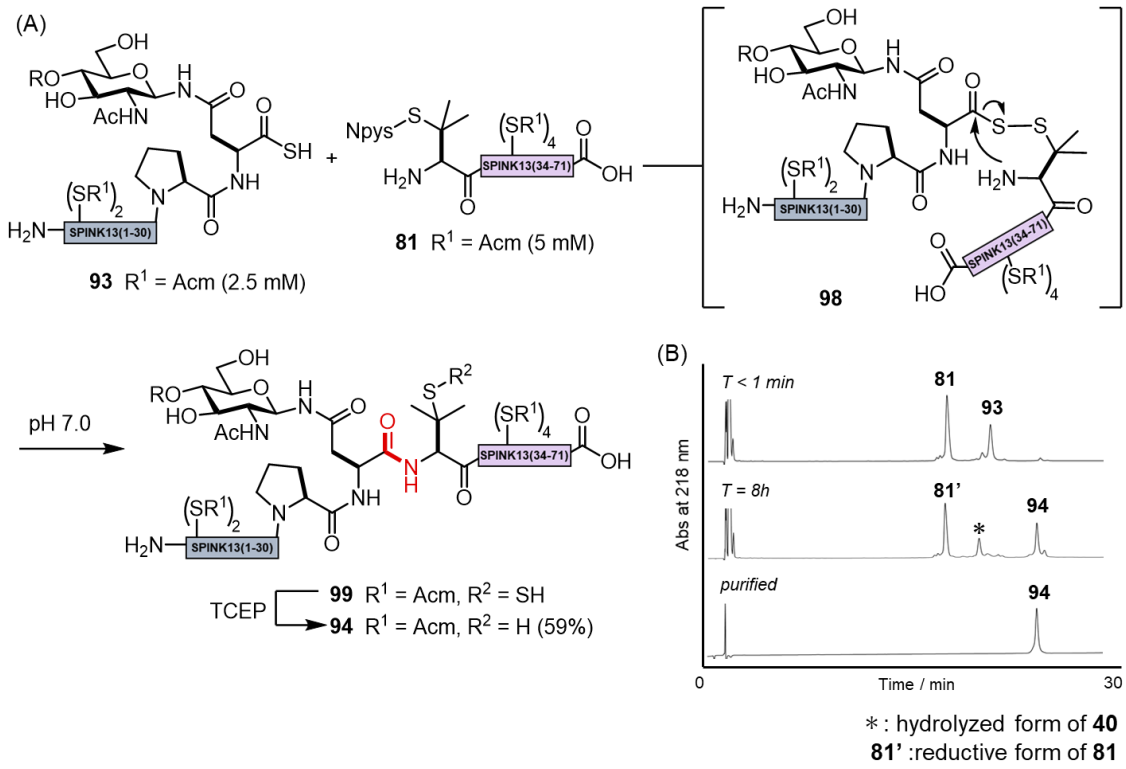


Figure 5-9. (A) TCL with glycopeptide thioacids **93** and peptide **81** having the Npys group at β-mercapto valine. (B) HPLC monitoring of TCL between glycopeptide thioacids **93** and peptide **81**. The compound **81'** indicates a reductive form (having no disulfide at Val[33]) of peptide **81**. The asterisk indicates an aspartimide and a hydrolyzed form of glycopeptide thioacid **93**.

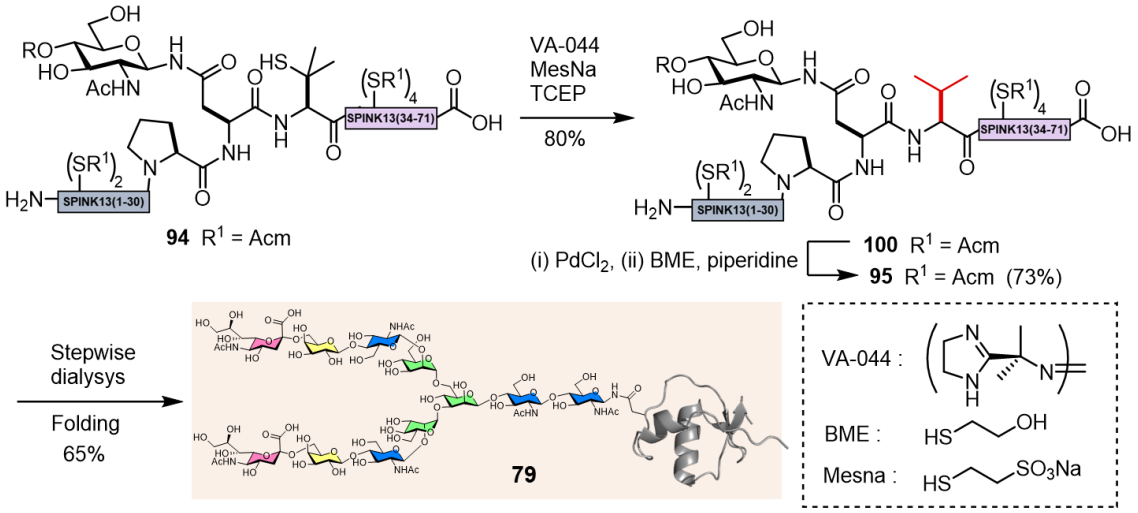


Figure 5-10. Desulfurization, deprotection and folding of full-length glycopeptide **94**.

Desulfurization of the 33th β-mercapto valine of a full-length glycoprotein **94** with a radical initiator[54] was performed to yield a desulfurized glycopeptide **100** which was converted from βVal to Val at 33th position in 80% isolated yield (Figure 5-10). Subsequently deprotection of the Acm protecting groups of Cys with $PdCl_2$[139-140] and the phenacyl protecting group of sialyloligosaccharide with piperidine and 2-mercaptoethanol (BME) were performed to yield the linear glycosyl SPINK13 polypeptide **95** in 73% isolated yield (Figure 5-10). Finally, stepwise dialysis folding under redox conditions in the same manner for the synthesis of IL3 at pH 8.0 yielded folded SPINK13 **79** with a biantennary sialyl oligosaccharide at the 32th position in 65% isolated yield. After isolation of the folded SPINK13 **79**, CD spectrum and the high-resolution mass (10680.6854, average isotopes) supported the correctly folded structure of glycosyl SPINK13 **79** (Figure 5-11).

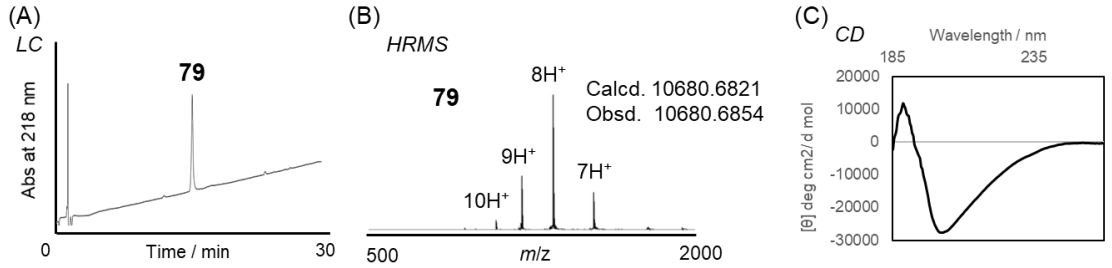
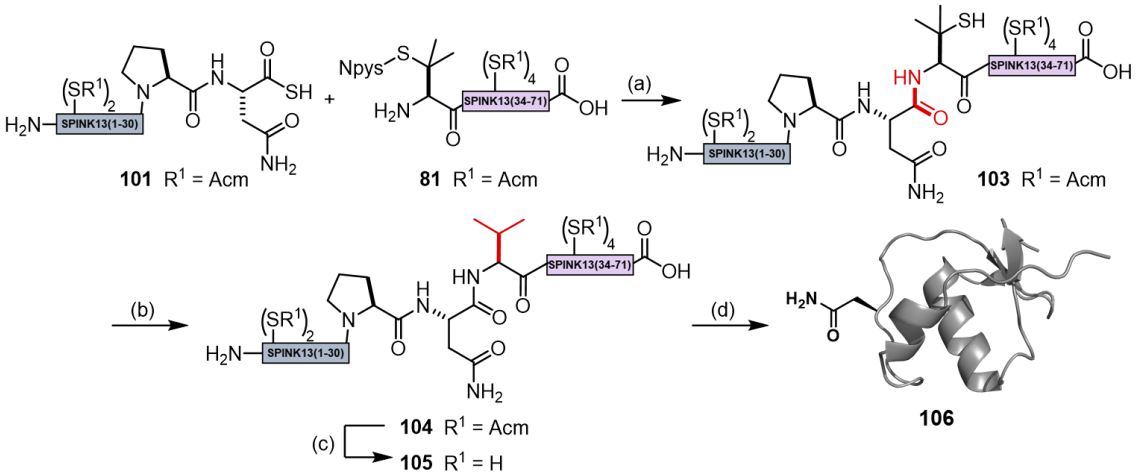


Figure 5-11. (A) HPLC spectrum, (B) HRMS spectrum, CD spectrum of purified folded SPINK13 **79**.

Scheme 5-3. Synthesis of non-glycosyl SPINK13 **106**.



Reaction conditions; (a) Phosphate buffer (pH 7.0), 8 h, 32%. (b) VA-044, Mesna, TCEP, phosphate buffer (pH 7.0),14 h, 43%. (c) $PdCl_2$, then piperidine, BME, phosphate buffer (pH 7.0), rt. 80%. (d) Stepwise dialysis folding, 66%.

After the synthesis of glycosyl SPINK13 **79**, we also synthesized non-glycosyl SPINK13 **106** without a biantennary sialyl oligosaccharide for bioassay using SPINK13 (Scheme 5-3). Peptide

thioacid **101**, which was synthesized by Fmoc SPPS in the same manner of peptide prolyl thioacid **80**, was coupled with C-terminal peptide having the Npys group at β-mercapto valine **81** utilizing TCL to give a full-length protein **103** in 32% isolated yield. Desulfurization of the 33th β-mercapto valine of a full-length protein **103** with a radical initiator[54] was performed to yield a desulfurized peptide **104** in 43% isolated yield (Scheme 5-3). Subsequently deprotection of the Acm protecting groups of Cys with $PdCl_2$[139-140] were performed to yield the linear SPINK13 polypeptide **105** in 80% isolated yield. Finally, stepwise dialysis folding under redox conditions yielded folded non-glycosyl SPINK13 **106** in 66% isolated yield.

5-6. Biological analysis and discussion

Wang and coworkers found the *in vivo* antitumor activity of SPINK13 by using transwell assay and the nude mice tumor model[152]. This inhibitor activity for the cancer cells is due to that SPINK13 inhibits the metastasis and invasion of tumor cells. Tumor metastasis is a multistep process to invade the extracellular matrix and migrate to other organs through the hematogenous and lymphatic routes[167]. In the metastasis of cancer cells, the key step is the proteolytic degradation of the extracellular matrix to promote tumor cell invasion and migration[168]. According to hitherto research, urokinase type plasminogen activator (uPA), a member of serine protease, and urokinase type plasminogen activator receptor (uPAR) system regulate the cascade of the tumor metastasis mainly, mediating proteolysis during cancer invasion and metastasis[169]. In cancer cells, the expression of pro-uPA is much higher than other cells, and uPAR transform a pro-uPA to an active uPA by forming the complex (Figure 5-12 I). This uPA / uPAR complex cleave and activate plasminogen to afford plasmin, which also belongs to the family of serine proteases (Figure 5-12 II). The resultant plasmin activates the precursor of matrix metalloproteinases (pro-MMPs) and generate matrix metalloproteinases (MMPs), zinc containing endopeptidases, by the proteolysis[170] (Figure 5-12 III). MMPs leads to the breakage of the extracellular matrix and then tumor cells can migrate to the blood vessel (Figure 5-12 IV).

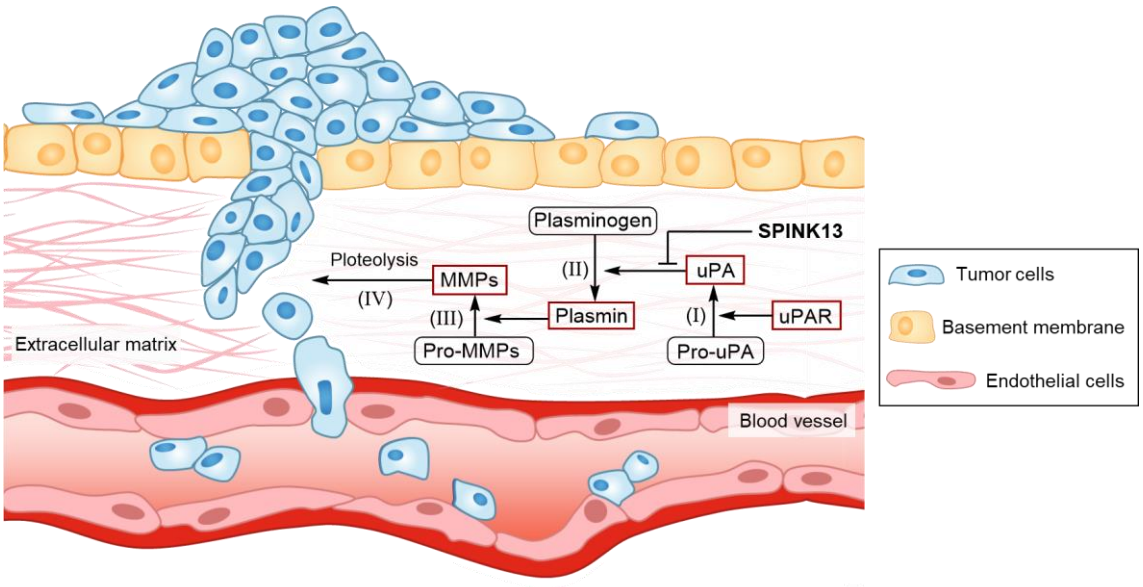


Figure 5-14. The mechanism of tumor cancer cells invasion and metastasis. Urokinase plasminogen activator (uPA), Urokinase plasminogen activator receptor (uPAR), Matrix metalloproteinases (MMPs). Tumor cells invasion progress through the multistep including the activation of uPA by uPAR (I) , the activation of plasmin by plasminogen (II), the activation of MMPs by pro-MMPs (III) and the proteolysis of the extracellular matrix (IV). SPINK13 can be therapeutically used as inhibitor of uPA.

In terms of the substrate, uPA consists of three domains: a growth factor domain, a Kringle domain (KD), and a serine protease domain, which is a substrate of SPINK13 (Figure 5-12). Although there are no examples of inhibition experiments using uPA and SPINK13 *in vitro*, a protease inhibition experiment using trypsin with a similar structure to uPA was performed and showed a slight inhibitory activity[150], suggesting that inhibition of uPA is responsible for the antitumor activity of SPINK13.

In this background, we performed an *in vitro* bioassay using glycosyl SPINK13 **79** having biantennary sialyl complex type oligosaccharides and chemically synthesized non-glycosyl SPINK13 **106**. Since SPINK13 inhibits growth and metastasis of hepatocellular carcinoma *in vivo*[152], we administered SPINK13 **79** and **106** to Huh7 liver tumor cell lines (hepatocellular carcinoma) and HEK293 cell lines (Human Embryonic Kidney cells 293), respectively, and incubated for 24 hours. At the result, the activity of glycosyl SPINK13 **79** was confirmed to be similar to that of non-glycosyl SPINK13 **106** Although both of SPINK13 derivatives did not show the antitumor activity in number using Huh7 cell lines (Figure 5-13 A), in contrast, the normal HEK293 cells decreased in a little amount (Figure 5-13 B).

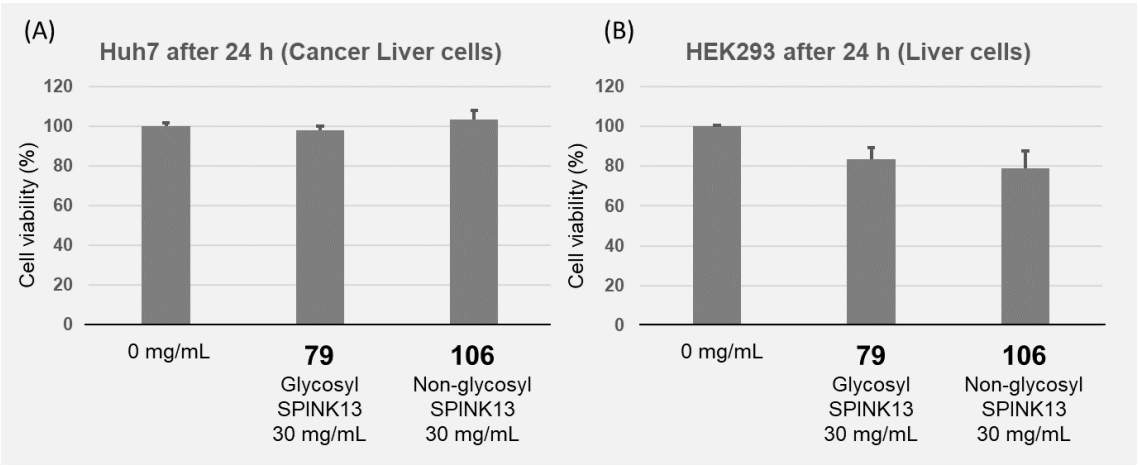


Figure 5-13. *In vitro* biological activity of SPINK13. The experiments were performed with triplicate samples. (A) Cell viability of Huh7. (B) Cell viability of HEK293.

This result indicates that SPINK13 does not have a growth inhibitory effect on cancer cells. In other words, the inhibition of tumor cells by SPINK13 is due to the inhibition of cancer cell invasion into other tissues. This is similar to the result of Wang and coworkers that the tumor reduction is mainly due to the inhibition of uPA by SPINK13.[152] Interestingly, in the system using HEK293 cells, the cells decreased in a little amount. This low cytotoxicity of SPINK13 will be a great advantage when SPINK13 is used as a uPA inhibitor of anticancer drugs for cancers that secrete MMPs well such as sarcomas (HT1080)[171] in the future.

The difference in antitumor activity with / without glycans was not shown to be significant in this experiment. However, since glycosylated Asn32 of SPINK13 presents close to the catalytic sites (catalytic triad, His57, Asp102, Ser195) of uPA, the glycan motif clearly interferes with the function of SPINK13 as an inhibitor. Furthermore, our structural analysis using AlphaFold[172] revealed that Asn32 is incorporated into the protease (trypsin), as shown in the figure 5-14. Therefore, it is possible that unknown glycan cleavage system might work as switching of the inhibitory activity of SPINK13 when it acts as a uPA inhibitor *in vivo*.

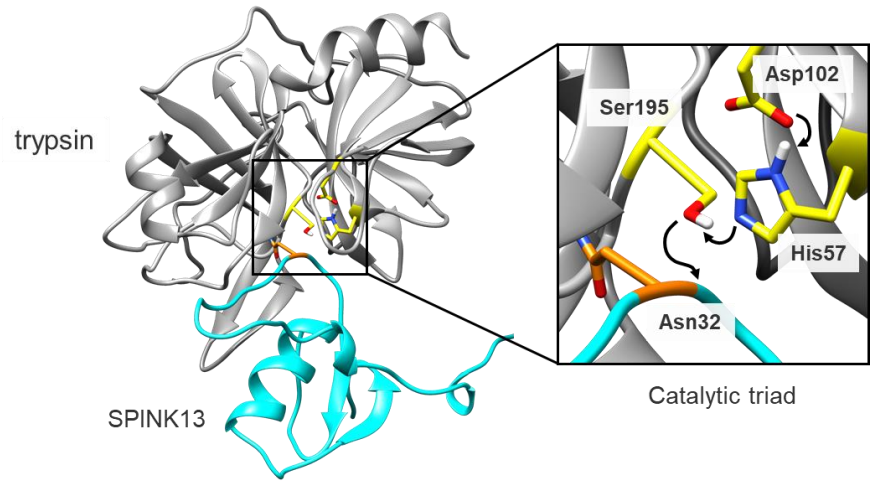


Figure 5-14. The results of computational simulation of the complex structure of SPINK13 and serine protease (trypsin) using AlphaFold. Catalytic sites (catalytic triad, His57, Asp102, Ser195).

5-7. Summary

In *Chapter 5*, we first investigated DDC using prolyl thioacids (**A**) and glycosyl asparagine thioacids (**B**) having mono GlcNAc (Figure 5-14 I). As a result, we confirmed that DDC with prolyl thioacids (**A**) proceeds in moderate yield to afford condensed glycopeptide (**C**) consisting of an α-amide bond without epimerization (Figure 5-14 I). For the synthesis of prolyl thioacids, we have developed a new synthetic method of peptide thioacids using peptide *o*-aminoanilides. Subsequently, we demonstrated for the first time that TCL proceeds with glycopeptide thioacid (**D**) and peptide (**E**) having β-mercapto valine at N-terminus, which has a mercapto group at the β-position and yielded full-length glycopeptide (**F**) (Figure 5-14 II). We also found that the yield of TCL using amino acids with β-mercapto amino acids decreased depending on pH compared to conventional TCL.

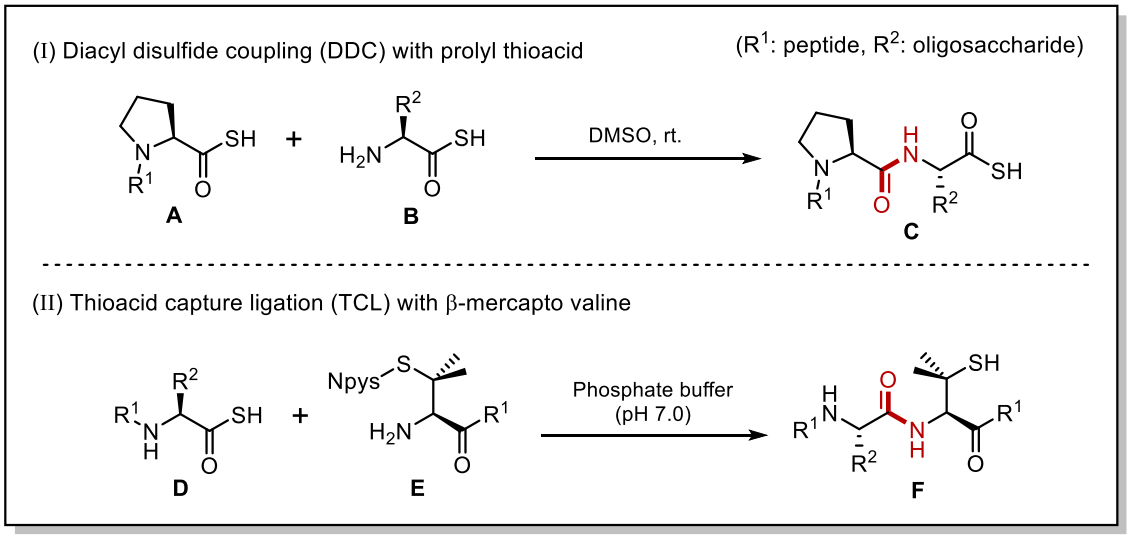


Figure 5-14. (I) DDC with prolyl thioacid. (II) TCL with β-mercapto valine.

Finally, we extended the chemical glycan insertion strategy by applying DDC with prolyl thioacids and TCL with β-mercapto valine having a mercapto group at the β-position (Figure 5-14). This led to the discovery of the applied chemical glycan insertion strategy shown in the Figure 5-15, that can couple two peptides (**A**, **D**) with the N- and C-terminus of glycosyl asparagine thioacid (**B**). The first coupling is DDC between a peptide thioacid (**A**) and glycosyl asparagine thioacid (**B**) and TCL for the coupling of the resultant glycopeptide thioacid and another peptide (**D**) having a Npys group at β-mercapto amino acids to afford the full-length glycoprotein backbone (**E**). Desulfurization of a full-length glycoprotein (**E**) afforded a glycoprotein which has a natural linkage (**F**). Furthermore, we achieved the synthesis of glycosyl SPINK13 **79** using the applied chemical glycan insertion strategy. Bioassays were also performed using glycosyl and non-glycosyl SPINK13.

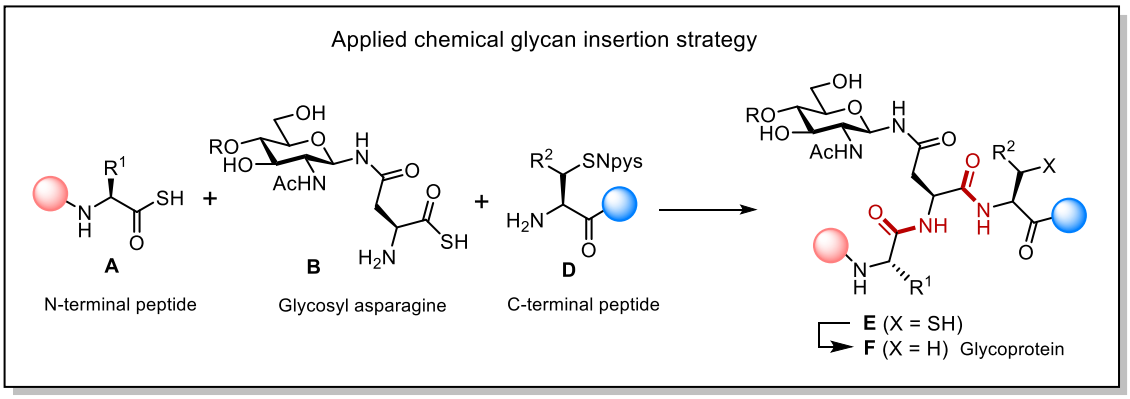


Figure 5-15. Applied chemical glycan insertion strategy for the synthesis of glycoproteins.

Chapter 6. Conclusion

We established diacyl disulfide coupling (DDC) as a robust α-amide bond formation reaction. Furthermore, we succeeded in DDC using glycosyl asparagine thioacids to yield glycopeptide thioacids. Since the glycopeptides synthesized by DDC have thioacids, we succeeded in further condensation of another peptides by thioacid capture ligation (TCL). This led us to the discovery of the chemical glycan insertion strategy.

We developed a highly convergent glycoprotein synthesis strategy, chemical glycan insertion strategy, that entails the coupling of two peptides with the N- and C-terminus of glycosyl asparagine thioacid chemoselectively (Figure 6). In the chemical glycan insertion strategy, the first coupling is DDC between a peptide thioacid (**C**) and glycosyl asparagine thioacid (**B**). Because the resultant glycopeptide has a thioacid form at its C terminus, we could apply TCL for the coupling of the resultant glycopeptide thioacid and another peptide (**D**) having the Npys group at its β-position of N-terminus to afford the full-length glycoprotein backbone (**A**) in 2 steps. Applying this strategy bioactive cytokines, CC-chemokine ligand 1 (CCL1), interleukin 3 (IL3) and serine protease inhibitor Kazal type 13 (SPINK13) were totally synthesized within a few steps. Our new synthetic strategy for the synthesis of homogeneous glycoproteins can be applied for the synthesis of not only N- and O-glycoproteins but also other posttranslational modifications (PTMs).

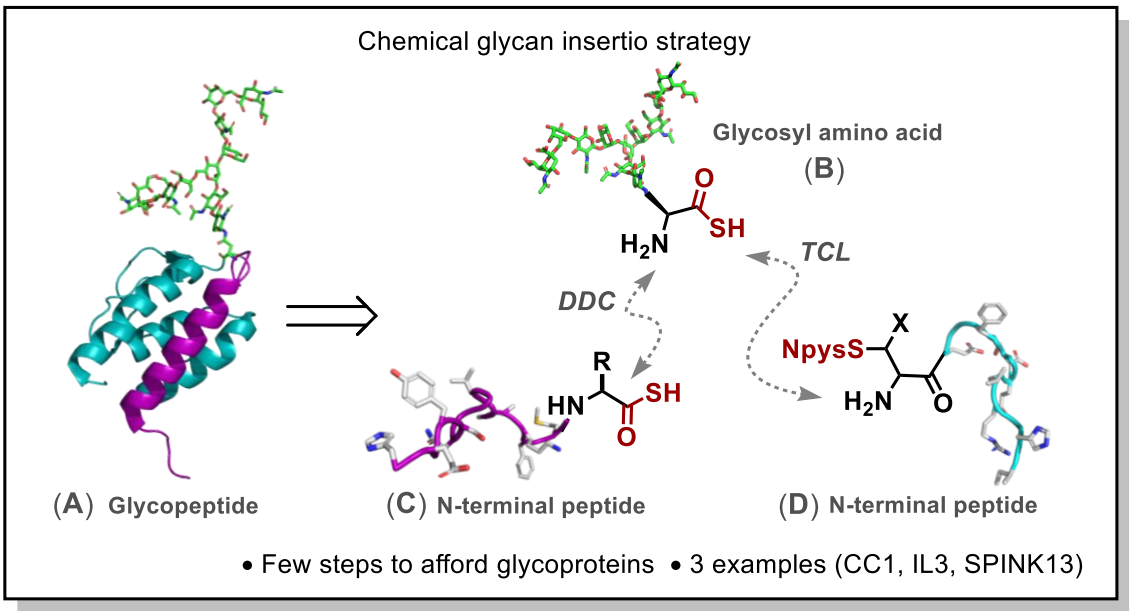


Figure 6. Chemical insertion strategy for the synthesis of glycoproteins.

Reference

1. Walsh, G.; Jefferis, R., Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol* **2006,** *24* (10), 1241-52.

2. Okayama, A.; Miyagi, Y.; Oshita, F.; Ito, H.; Nakayama, H.; Nishi, M.; Kurata, Y.; Kimura, Y.; Ryo, A.; Hirano, H., Identification of Tyrosine-Phosphorylated Proteins Upregulated during Epithelial–Mesenchymal Transition Induced with TGF-β. *Journal of Proteome Research* **2015,** *14* (10), 4127-4136.

3. Nagata, K.; Kawakami, T.; Kurata, Y.; Kimura, Y.; Suzuki, Y.; Nagata, T.; Sakuma, Y.; Miyagi, Y.; Hirano, H., Augmentation of multiple protein kinase activities associated with secondary imatinib resistance in gastrointestinal stromal tumors as revealed by quantitative phosphoproteome analysis. *J Proteomics* **2015,** *115*, 132-42.

4. Ishigami, A.; Masutomi, H.; Handa, S.; Nakamura, M.; Nakaya, S.; Uchida, Y.; Saito, Y.; Murayama, S.; Jang, B.; Jeon, Y. C.; Choi, E. K.; Kim, Y. S.; Kasahara, Y.; Maruyama, N.; Toda, T., Mass spectrometric identification of citrullination sites and immunohistochemical detection of citrullinated glial fibrillary acidic protein in Alzheimer's disease brains. *J Neurosci Res* **2015,** *93* (11), 1664-74.

5. Ishigami, A.; Ohsawa, T.; Hiratsuka, M.; Taguchi, H.; Kobayashi, S.; Saito, Y.; Murayama, S.; Asaga, H.; Toda, T.; Kimura, N.; Maruyama, N., Abnormal accumulation of citrullinated proteins catalyzed by peptidylarginine deiminase in hippocampal extracts from patients with Alzheimer's disease. *J Neurosci Res* **2005,** *80* (1), 120-8.

6. Unverzagt, C.; Kajihara, Y., Recent advances in the chemical synthesis of N-linked glycoproteins. *Curr Opin Chem Biol* **2018,** *46*, 130-137.

7. Wong, C.-H., Protein Glycosylation: New Challenges and Opportunities. *The Journal of Organic Chemistry* **2005,** *70* (11), 4219-4225.

8. Varki, A., Biological roles of glycans. *Glycobiology* **2017,** *27* (1), 3-49.

9. Wolfert, M. A.; Boons, G. J., Adaptive immune activation: glycosylation does matter. *Nat Chem Biol* **2013,** *9* (12), 776-84.

10. Sinclair, A. M.; Elliott, S., Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. *J Pharm Sci* **2005,** *94* (8), 1626-35.

11. Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A., Glycosylation and the immune system. *Science* **2001,** *291* (5512), 2370-6.

12. Van den Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G., Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol* **1998,** *33* (3), 151-208.

13. Lizak, C.; Gerber, S.; Numao, S.; Aebi, M.; Locher, K. P., X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **2011,** *474* (7351), 350-355.

14. Kiuchi, T.; Izumi, M.; Mukogawa, Y.; Shimada, A.; Okamoto, R.; Seko, A.; Sakono, M.; Takeda, Y.; Ito, Y.; Kajihara, Y., Monitoring of Glycoprotein Quality Control System with a Series of Chemically

Synthesized Homogeneous Native and Misfolded Glycoproteins. *Journal of the American Chemical Society* **2018,** *140* (50), 17499-17507.

15. Helenius, A.; Aebi, M., Roles of N-Linked Glycans in the Endoplasmic Reticulum. *Annual Review of Biochemistry* **2004,** *73* (1), 1019-1049.

16. Shirakawa, A.; Manabe, Y.; Fukase, K., Recent Advances in the Chemical Biology of N-Glycans. *Molecules* **2021,** *26* (4).

17. Dadová, J.; Galan, S. R.; Davis, B. G., Synthesis of modified proteins via functionalization of dehydroalanine. *Curr Opin Chem Biol* **2018,** *46*, 71-81.

18. Hirano, K.; Macmillan, D.; Tezuka, K.; Tsuji, T.; Kajihara, Y., Design and synthesis of a homogeneous erythropoietin analogue with two human complex-type sialyloligosaccharides: combined use of chemical and bacterial protein expression methods. *Angew Chem Int Ed Engl* **2009,** *48* (50), 9557-60.

19. Yamamoto, N.; Sakakibara, T.; Kajihara, Y., Convenient synthesis of a glycopeptide analogue having a complex type disialyl-undecasaccharide. *Tetrahedron Letters* **2004,** *45* (16), 3287-3290.

20. Unverzagt, C.; Kajihara, Y., Chemical assembly of N-glycoproteins: a refined toolbox to address a ubiquitous posttranslational modification. *Chem Soc Rev* **2013,** *42* (10), 4408-20.

21. Eller, S.; Schuberth, R.; Gundel, G.; Seifert, J.; Unverzagt, C., Synthesis of pentaantennary N-glycans with bisecting GlcNAc and core fucose. *Angew Chem Int Ed Engl* **2007,** *46* (22), 4173-5.

22. Shivatare, S. S.; Chang, S. H.; Tsai, T. I.; Ren, C. T.; Chuang, H. Y.; Hsu, L.; Lin, C. W.; Li, S. T.; Wu, C. Y.; Wong, C. H., Efficient convergent synthesis of bi-, tri-, and tetra-antennary complex type N-glycans and their HIV-1 antigenicity. *J Am Chem Soc* **2013,** *135* (41), 15382-91.

23. Seko, A.; Koketsu, M.; Nishizono, M.; Enoki, Y.; Ibrahim, H. R.; Juneja, L. R.; Kim, M.; Yamamoto, T., Occurrence of a sialylglycopeptide and free sialylglycans in hen's egg yolk. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1997,** *1335* (1-2), 23-32.

24. Kajihara, Y.; Suzuki, Y.; Yamamoto, N.; Sasaki, K.; Sakakibara, T.; Juneja, L. R., Prompt chemoenzymatic synthesis of diverse complex-type oligosaccharides and its application to the solid-phase synthesis of a glycopeptide with Asn-linked sialyl-undeca- and asialo-nonasaccharides. *Chemistry* **2004,** *10* (4), 971-85.

25. Fukae, K.; Yamamoto, N.; Hatakeyama, Y.; Kajihara, Y., Chemoenzymatic synthesis of diverse asparagine-linked alpha-(2,3)-sialyloligosaccharides. *Glycoconj J* **2004,** *21* (5), 243-50.

26. Maki, Y.; Okamoto, R.; Izumi, M.; Murase, T.; Kajihara, Y., Semisynthesis of Intact Complex-Type Triantennary Oligosaccharides from a Biantennary Oligosaccharide Isolated from a Natural Source by Selective Chemical and Enzymatic Glycosylation. *J Am Chem Soc* **2016,** *138* (10), 3461-8.

27. Maki, Y.; Nomura, K.; Okamoto, R.; Izumi, M.; Mizutani, Y.; Kajihara, Y., Acceleration and Deceleration Factors on the Hydrolysis Reaction of 4,6-O-Benzylidene Acetal Group. *The Journal of Organic Chemistry* **2020,** *85* (24), 15849-15856.

28. Xu, T.; Coward, J. K., 13C- and 15N-labeled peptide substrates as mechanistic probes of

oligosaccharyltransferase. *Biochemistry* **1997,** *36* (48), 14683-9.

29. Dempski, R. E.; Imperiali, B., Oligosaccharyl transferase: gatekeeper to the secretory pathway. *Current Opinion in Chemical Biology* **2002,** *6* (6), 844-850.

30. Wang, L. X.; Amin, M. N., Chemical and chemoenzymatic synthesis of glycoproteins for deciphering functions. *Chem Biol* **2014,** *21* (1), 51-66.

31. Chen, M. M.; Glover, K. J.; Imperiali, B., From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in C. jejuni. *Biochemistry* **2007,** *46* (18), 5579-85.

32. Glover, K. J.; Weerapana, E.; Numao, S.; Imperiali, B., Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni. *Chem Biol* **2005,** *12* (12), 1311-5.

33. Schwarz, F.; Fan, Y. Y.; Schubert, M.; Aebi, M., Cytoplasmic N-glycosyltransferase of Actinobacillus pleuropneumoniae is an inverting enzyme and recognizes the NX(S/T) consensus sequence. *J Biol Chem* **2011,** *286* (40), 35267-74.

34. Lomino, J. V.; Naegeli, A.; Orwenyo, J.; Amin, M. N.; Aebi, M.; Wang, L. X., A two-step enzymatic glycosylation of polypeptides with complex N-glycans. *Bioorg Med Chem* **2013,** *21* (8), 2262-2270.

35. Wang, L. X.; Lomino, J. V., Emerging technologies for making glycan-defined glycoproteins. *ACS Chem Biol* **2012,** *7* (1), 110-22.

36. Witte, K.; Sears, P.; Martin, R.; Wong, C.-H., Enzymatic Glycoprotein Synthesis: Preparation of Ribonuclease Glycoforms via Enzymatic Glycopeptide Condensation and Glycosylation. *Journal of the American Chemical Society* **1997,** *119* (9), 2114-2118.

37. Fujita, M.; Shoda, S.-i.; Haneda, K.; Inazu, T.; Takegawa, K.; Yamamoto, K., A novel disaccharide substrate having 1,2-oxazoline moiety for detection of transglycosylating activity of endoglycosidases. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2001,** *1528* (1), 9-14.

38. Wang, L. X., Chemoenzymatic synthesis of glycopeptides and glycoproteins through endoglycosidase-catalyzed transglycosylation. *Carbohydr Res* **2008,** *343* (10-11), 1509-22.

39. Li, B.; Zeng, Y.; Hauser, S.; Song, H.; Wang, L. X., Highly efficient endoglycosidase-catalyzed synthesis of glycopeptides using oligosaccharide oxazolines as donor substrates. *J Am Chem Soc* **2005,** *127* (27), 9692-3.

40. Umekawa, M.; Huang, W.; Li, B.; Fujita, K.; Ashida, H.; Wang, L. X.; Yamamoto, K., Mutants of Mucor hiemalis endo-beta-N-acetylglucosaminidase show enhanced transglycosylation and glycosynthase-like activities. *J Biol Chem* **2008,** *283* (8), 4469-79.

41. Huang, W.; Yang, Q.; Umekawa, M.; Yamamoto, K.; Wang, L. X., Arthrobacter endo-beta-N-acetylglucosaminidase shows transglycosylation activity on complex-type N-glycan oxazolines: one-pot conversion of ribonuclease B to sialylated ribonuclease C. *Chembiochem* **2010,** *11* (10), 1350-5.

42. Ochiai, H.; Huang, W.; Wang, L. X., Expeditious chemoenzymatic synthesis of homogeneous N-

glycoproteins carrying defined oligosaccharide ligands. *J Am Chem Soc* **2008,** *130* (41), 13790-803.

43. Noguchi, M.; Tanaka, T.; Gyakushi, H.; Kobayashi, A.; Shoda, S., Efficient synthesis of sugar oxazolines from unprotected N-acetyl-2-amino sugars by using chloroformamidinium reagent in water. *J Org Chem* **2009,** *74* (5), 2210-2.

44. Huang, W.; Groothuys, S.; Heredia, A.; Kuijpers, B. H.; Rutjes, F. P.; van Delft, F. L.; Wang, L. X., Enzymatic glycosylation of triazole-linked GlcNAc/Glc-peptides: synthesis, stability and anti-HIV activity of triazole-linked HIV-1 gp41 glycopeptide C34 analogues. *Chembiochem* **2009,** *10* (7), 1234-42.

45. Huang, W.; Giddens, J.; Fan, S. Q.; Toonstra, C.; Wang, L. X., Chemoenzymatic glycoengineering of intact IgG antibodies for gain of functions. *J Am Chem Soc* **2012,** *134* (29), 12308-18.

46. Fan, S. Q.; Huang, W.; Wang, L. X., Remarkable transglycosylation activity of glycosynthase mutants of endo-D, an endo-beta-N-acetylglucosaminidase from Streptococcus pneumoniae. *J Biol Chem* **2012,** *287* (14), 11272-81.

47. Manabe, S.; Yamaguchi, Y.; Matsumoto, K.; Fuchigami, H.; Kawase, T.; Hirose, K.; Mitani, A.; Sumiyoshi, W.; Kinoshita, T.; Abe, J.; Yasunaga, M.; Matsumura, Y.; Ito, Y., Characterization of Antibody Products Obtained through Enzymatic and Nonenzymatic Glycosylation Reactions with a Glycan Oxazoline and Preparation of a Homogeneous Antibody-Drug Conjugate via Fc N-Glycan. *Bioconjug Chem* **2019,** *30* (5), 1343-1355.

48. Hojo, H.; Tanaka, H.; Hagiwara, M.; Asahina, Y.; Ueki, A.; Katayama, H.; Nakahara, Y.; Yoneshige, A.; Matsuda, J.; Ito, Y.; Nakahara, Y., Chemoenzymatic synthesis of hydrophobic glycoprotein: synthesis of saposin C carrying complex-type carbohydrate. *J Org Chem* **2012,** *77* (21), 9437-46.

49. Asahina, Y.; Kamitori, S.; Takao, T.; Nishi, N.; Hojo, H., Chemoenzymatic synthesis of the immunoglobulin domain of Tim-3 carrying a complex-type N-glycan by using a one-pot ligation. *Angew Chem Int Ed Engl* **2013,** *52* (37), 9733-7.

50. Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoet, M.; Monbaliu, J. M.; Melnyk, O., Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem Rev* **2019,** *119* (12), 7328-7443.

51. Dawson, P. E.; Kent, S. B., Synthesis of native proteins by chemical ligation. *Annu Rev Biochem* **2000,** *69*, 923-60.

52. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266* (5186), 776-9.

53. Yan, L. Z.; Dawson, P. E., Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J Am Chem Soc* **2001,** *123* (4), 526-33.

54. Wan, Q.; Danishefsky, S. J., Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew Chem Int Ed Engl* **2007,** *46* (48), 9248-52.

55. Thompson, R. E.; Chan, B.; Radom, L.; Jolliffe, K. A.; Payne, R. J., Chemoselective Peptide

Ligation–Desulfurization at Aspartate. *Angewandte Chemie International Edition* **2013,** *52* (37), 9723-9727.

56. Kulkarni, S. S.; Sayers, J.; Premdjee, B.; Payne, R. J., Rapid and efficient protein synthesis through expansion of the native chemical ligation concept. *Nature Reviews Chemistry* **2018,** *2* (4), 0122.

57. Wong, C. T. T.; Tung, C. L.; Li, X., Synthetic cysteine surrogates used in native chemical ligation. *Molecular BioSystems* **2013,** *9* (5), 826-833.

58. Merrifield, R. B., Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **1963,** *85* (14), 2149-2154.

59. Asahina, Y.; Komiya, S.; Ohagi, A.; Fujimoto, R.; Tamagaki, H.; Nakagawa, K.; Sato, T.; Akira, S.; Takao, T.; Ishii, A.; Nakahara, Y.; Hojo, H., Chemical Synthesis of O-Glycosylated Human Interleukin-2 by the Reverse Polarity Protection Strategy. *Angew Chem Int Ed Engl* **2015,** *54* (28), 8226-30.

60. Anisfeld, S. T.; Lansbury Jr, P. T., A convergent approach to the chemical synthesis of asparagine-linked glycopeptides. *The Journal of Organic Chemistry* **1990,** *55* (21), 5560-5562.

61. Cohen-Anisfeld, S. T.; Lansbury, P. T., A practical, convergent method for glycopeptide synthesis. *Journal of the American Chemical Society* **1993,** *115* (23), 10531-10537.

62. Ullmann, V.; Radisch, M.; Boos, I.; Freund, J.; Pohner, C.; Schwarzinger, S.; Unverzagt, C., Convergent solid-phase synthesis of N-glycopeptides facilitated by pseudoprolines at consensus-sequence Ser/Thr residues. *Angew Chem Int Ed Engl* **2012,** *51* (46), 11566-70.

63. Wang, P.; Aussedat, B.; Vohra, Y.; Danishefsky, S. J., An advance in the chemical synthesis of homogeneous N-linked glycopolypeptides by convergent aspartylation. *Angew Chem Int Ed Engl* **2012,** *51* (46), 11571-5.

64. Wang, P.; Li, X.; Zhu, J.; Chen, J.; Yuan, Y.; Wu, X.; Danishefsky, S. J., Encouraging progress in the omega-aspartylation of complex oligosaccharides as a general route to beta-N-linked glycopolypeptides. *J Am Chem Soc* **2011,** *133* (5), 1597-602.

65. Joseph, R.; Dyer, F. B.; Garner, P., Rapid formation of N-Glycopeptides via Cu(II)-promoted glycosylative ligation. *Org Lett* **2013,** *15* (4), 732-5.

66. Du, J. J.; Gao, X. F.; Xin, L. M.; Lei, Z.; Liu, Z.; Guo, J., Convergent Synthesis of N-Linked Glycopeptides via Aminolysis of omega-Asp p-Nitrophenyl Thioesters in Solution. *Org Lett* **2016,** *18* (19), 4828-4831.

67. Schowe, M. J.; Keiper, O.; Unverzagt, C.; Wittmann, V., A Tripeptide Approach to the Solid-Phase Synthesis of Peptide Thioacids and N-Glycopeptides. *Chemistry* **2019,** *25* (69), 15759-15764.

68. Chai, H.; Le Mai Hoang, K.; Vu, M. D.; Pasunooti, K.; Liu, C. F.; Liu, X. W., N-Linked Glycosyl Auxiliary-Mediated Native Chemical Ligation on Aspartic Acid: Application towards N-Glycopeptide Synthesis. *Angew Chem Int Ed Engl* **2016,** *55* (35), 10363-7.

69. Maki, Y.; Okamoto, R.; Izumi, M.; Kajihara, Y., Chemical Synthesis of an Erythropoietin Glycoform Having a Triantennary N-Glycan: Significant Change of Biological Activity of Glycoprotein by Addition of a Small Molecular Weight Trisaccharide. *J Am Chem Soc* **2020,** *142* (49), 20671-20679.

70. Wu, B.; Chen, J.; Warren, J. D.; Chen, G.; Hua, Z.; Danishefsky, S. J., Building complex glycopeptides: Development of a cysteine-free native chemical ligation protocol. *Angew Chem Int Ed Engl* **2006,** *45* (25), 4116-25.

71. Brik, A.; Yang, Y. Y.; Ficht, S.; Wong, C. H., Sugar-assisted glycopeptide ligation. *J Am Chem Soc* **2006,** *128* (17), 5626-7.

72. Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J., Native chemical ligation at valine: a contribution to peptide and glycopeptide synthesis. *Angew Chem Int Ed Engl* **2008,** *47* (44), 8521-4.

73. Payne, R. J.; Wong, C. H., Advances in chemical ligation strategies for the synthesis of glycopeptides and glycoproteins. *Chem Commun (Camb)* **2010,** *46* (1), 21-43.

74. Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R., Fmoc-Based Synthesis of Peptide-αThioesters: Application to the Total Chemical Synthesis of a Glycoprotein by Native Chemical Ligation. *Journal of the American Chemical Society* **1999,** *121* (50), 11684-11689.

75. Mezzato, S.; Schaffrath, M.; Unverzagt, C., An orthogonal double-linker resin facilitates the efficient solid-phase synthesis of complex-type N-glycopeptide thioesters suitable for native chemical ligation. *Angew Chem Int Ed Engl* **2005,** *44* (11), 1650-4.

76. Yamamoto, N.; Tanabe, Y.; Okamoto, R.; Dawson, P. E.; Kajihara, Y., Chemical synthesis of a glycoprotein having an intact human complex-type sialyloligosaccharide under the Boc and Fmoc synthetic strategies. *J Am Chem Soc* **2008,** *130* (2), 501-10.

77. Aussedat, B.; Fasching, B.; Johnston, E.; Sane, N.; Nagorny, P.; Danishefsky, S. J., Total synthesis of the alpha-subunit of human glycoprotein hormones: toward fully synthetic homogeneous human follicle-stimulating hormone. *J Am Chem Soc* **2012,** *134* (7), 3532-41.

78. Murakami, M.; Okamoto, R.; Izumi, M.; Kajihara, Y., Chemical synthesis of an erythropoietin glycoform containing a complex-type disialyloligosaccharide. *Angew Chem Int Ed Engl* **2012,** *51* (15), 3567-72.

79. Sakamoto, I.; Tezuka, K.; Fukae, K.; Ishii, K.; Taduru, K.; Maeda, M.; Ouchi, M.; Yoshida, K.; Nambu, Y.; Igarashi, J.; Hayashi, N.; Tsuji, T.; Kajihara, Y., Chemical synthesis of homogeneous human glycosyl-interferon-beta that exhibits potent antitumor activity in vivo. *J Am Chem Soc* **2012,** *134* (12), 5428-31.

80. Okamoto, R.; Mandal, K.; Ling, M.; Luster, A. D.; Kajihara, Y.; Kent, S. B., Total chemical synthesis and biological activities of glycosylated and non-glycosylated forms of the chemokines CCL1 and Ser-CCL1. *Angew Chem Int Ed Engl* **2014,** *53* (20), 5188-93.

81. Okamoto, R.; Mandal, K.; Sawaya, M. R.; Kajihara, Y.; Yeates, T. O.; Kent, S. B., (Quasi-)racemic X-ray structures of glycosylated and non-glycosylated forms of the chemokine Ser-CCL1 prepared by total chemical synthesis. *Angew Chem Int Ed Engl* **2014,** *53* (20), 5194-8.

82. Murakami, M.; Kiuchi, T.; Nishihara, M.; Tezuka, K.; Okamoto, R.; Izumi, M.; Kajihara, Y., Chemical synthesis of erythropoietin glycoforms for insights into the relationship between glycosylation

pattern and bioactivity. *Sci Adv* **2016,** *2* (1), e1500678.

83. Minh Hien, N.; Izumi, M.; Sato, H.; Okamoto, R.; Kajihara, Y., Chemical Synthesis of Glycoproteins with the Specific Installation of Gradient-Enriched (15) N-Labeled Amino Acids for Getting Insights into Glycoprotein Behavior. *Chemistry* **2017,** *23* (27), 6579-6585.

84. Li, H.; Zhang, J.; An, C.; Dong, S., Probing N-Glycan Functions in Human Interleukin-17A Based on Chemically Synthesized Homogeneous Glycoforms. *J Am Chem Soc* **2021,** *143* (7), 2846-2856.

85. Muir, T. W.; Sondhi, D.; Cole, P. A., Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A* **1998,** *95* (12), 6705-10.

86. Muir, T. W., Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem* **2003,** *72*, 249-89.

87. Piontek, C.; Ring, P.; Harjes, O.; Heinlein, C.; Mezzato, S.; Lombana, N.; Pohner, C.; Puttner, M.; Varon Silva, D.; Martin, A.; Schmid, F. X.; Unverzagt, C., Semisynthesis of a homogeneous glycoprotein enzyme: ribonuclease C: part 1. *Angew Chem Int Ed Engl* **2009,** *48* (11), 1936-40.

88. Piontek, C.; Varon Silva, D.; Heinlein, C.; Pohner, C.; Mezzato, S.; Ring, P.; Martin, A.; Schmid, F. X.; Unverzagt, C., Semisynthesis of a homogeneous glycoprotein enzyme: ribonuclease C: part 2. *Angew Chem Int Ed Engl* **2009,** *48* (11), 1941-5.

89. Okamoto, R.; Kimura, M.; Ishimizu, T.; Izumi, M.; Kajihara, Y., Semisynthesis of a post-translationally modified protein by using chemical cleavage and activation of an expressed fusion polypeptide. *Chemistry* **2014,** *20* (33), 10425-30.

90. Okamoto, R.; Iritani, K.; Amazaki, Y.; Zhao, D.; Chandrashekar, C.; Maki, Y.; Kanemitsu, Y.; Kaino, T.; Kajihara, Y., Semisynthesis of a Homogeneous Glycoprotein Using Chemical Transformation of Peptides to Thioester Surrogates. *J Org Chem* **2022,** *87* (1), 114-124.

91. Ludwig, C.; Schwarzer, D.; Zettler, J.; Garbe, D.; Janning, P.; Czeslik, C.; Mootz, H. D., Semisynthesis of proteins using split inteins. *Methods Enzymol* **2009,** *462*, 77-96.

92. Ling, J. J.; Policarpo, R. L.; Rabideau, A. E.; Liao, X.; Pentelute, B. L., Protein thioester synthesis enabled by sortase. *J Am Chem Soc* **2012,** *134* (26), 10749-52.

93. Komiya, C.; Shigenaga, A.; Tsukimoto, J.; Ueda, M.; Morisaki, T.; Inokuma, T.; Itoh, K.; Otaka, A., Traceless synthesis of protein thioesters using enzyme-mediated hydrazinolysis and subsequent self-editing of the cysteinyl prolyl sequence. *Chem Commun (Camb)* **2019,** *55* (49), 7029-7032.

94. Okamoto, R.; Morooka, K.; Kajihara, Y., A synthetic approach to a peptide alpha-thioester from an unprotected peptide through cleavage and activation of a specific peptide bond by N-acetylguanidine. *Angew Chem Int Ed Engl* **2012,** *51* (1), 191-6.

95. Kajihara, Y.; Kanemitsu, Y.; Nishihara, M.; Okamoto, R.; Izumi, M., Efficient synthesis of polypeptide-alpha-thioester by the method combining polypeptide expression and chemical activation for the semi-synthesis of interferon-gamma having oligosaccharides. *J Pept Sci* **2014,** *20* (12), 958-63.

96. Kawakami, T.; Sumida, M.; Nakamura, K. i.; Vorherr, T.; Aimoto, S., Peptide thioester

preparation based on an N-S acyl shift reaction mediated by a thiol ligation auxiliary. *Tetrahedron Letters* **2005,** *46* (50), 8805-8807.

97. Nakamura, K. i.; Sumida, M.; Kawakami, T.; Vorherr, T.; Aimoto, S., Generation of an S-Peptide via an N–S Acyl Shift Reaction in a TFA Solution. *Bulletin of the Chemical Society of Japan* **2006,** *79* (11), 1773-1780.

98. Kang, J.; Macmillan, D., Peptide and protein thioester synthesis via N→S acyl transfer. *Organic & Biomolecular Chemistry* **2010,** *8* (9), 1993-2002.

99. N, N.; Thimmalapura, V. M.; Hosamani, B.; Prabhu, G.; Kumar, L. R.; Sureshbabu, V. V., Thioacids – synthons for amide bond formation and ligation reactions: assembly of peptides and peptidomimetics. *Organic & Biomolecular Chemistry* **2018,** *16* (19), 3524-3552.

100. Crich, D.; Sharma, I., Epimerization-free block synthesis of peptides from thioacids and amines with the Sanger and Mukaiyama reagents. *Angew Chem Int Ed Engl* **2009,** *48* (13), 2355-8.

101. Sasaki, K.; Crich, D., Cyclic peptide synthesis with thioacids. *Org Lett* **2010,** *12* (14), 3254-7.

102. Song, W.; Dong, K.; Li, M., Visible Light-Induced Amide Bond Formation. *Org Lett* **2020,** *22* (2), 371-375.

103. Pan, J.; Devarie-Baez, N. O.; Xian, M., Facile amide formation via S-nitrosothioacids. *Org Lett* **2011,** *13* (5), 1092-4.

104. Wu, W.; Zhang, Z.; Liebeskind, L. S., In situ carboxyl activation using a silatropic switch: a new approach to amide and peptide constructions. *J Am Chem Soc* **2011,** *133* (36), 14256-9.

105. Wu, X.; Stockdill, J. L.; Park, P. K.; Danishefsky, S. J., Expanding the limits of isonitrile-mediated amidations: on the remarkable stereosubtleties of macrolactam formation from synthetic seco-cyclosporins. *J Am Chem Soc* **2012,** *134* (4), 2378-84.

106. Shangguan, N.; Katukojvala, S.; Greenberg, R.; Williams, L. J., The reaction of thio acids with azides: a new mechanism and new synthetic applications. *J Am Chem Soc* **2003,** *125* (26), 7754-5.

107. Kolakowski, R. V.; Shangguan, N.; Sauers, R. R.; Williams, L. J., Mechanism of thio acid/azide amidation. *J Am Chem Soc* **2006,** *128* (17), 5695-702.

108. Rohmer, K.; Mannuthodikayil, J.; Wittmann, V., Application of the Thioacid-Azide Ligation (TAL) for the Preparation of Glycosylated and Fluorescently Labeled Amino Acids. *Israel Journal of Chemistry* **2015,** *55* (3-4), 437-446.

109. Dyer, F. B.; Park, C. M.; Joseph, R.; Garner, P., Aziridine-mediated ligation and site-specific modification of unprotected peptides. *J Am Chem Soc* **2011,** *133* (50), 20033-5.

110. Bajaj, K.; Agarwal, D. S.; Sakhuja, R.; Pillai, G. G., Aziridine based electrophilic handle for aspartic acid ligation. *Org Biomol Chem* **2018,** *16* (23), 4311-4319.

111. Crich, D.; Sana, K.; Guo, S., Amino acid and peptide synthesis and functionalization by the reaction of thioacids with 2,4-dinitrobenzenesulfonamides. *Org Lett* **2007,** *9* (22), 4423-6.

112. Crich, D.; Sharma, I., Triblock peptide and peptide thioester synthesis with reactivity-

differentiated sulfonamides and peptidyl thioacids. *Angew Chem Int Ed Engl* **2009,** *48* (41), 7591-4.

113. Fukuyama, T.; Jow, C.-K.; Cheung, M., 2- and 4-Nitrobenzenesulfonamides: Exceptionally versatile means for preparation of secondary amines and protection of amines. *Tetrahedron Letters* **1995,** *36* (36), 6373-6374.

114. Chen, W.; Shao, J.; Hu, M.; Yu, W.; Giulianotti, M. A.; Houghten, R. A.; Yu, Y., A traceless approach to amide and peptide construction from thioacids and dithiocarbamate-terminal amines. *Chem. Sci.* **2013,** *4* (3), 970-976.

115. Crich, D.; Sasaki, K., Reaction of thioacids with isocyanates and isothiocyanates: a convenient amide ligation process. *Org Lett* **2009,** *11* (15), 3514-7.

116. Tam, J. P.; Lu, Y. A.; Liu, C. F.; Shao, J., Peptide synthesis using unprotected peptides through orthogonal coupling methods. *Proc Natl Acad Sci U S A* **1995,** *92* (26), 12485-9.

117. Zhang, X.; Li, F.; Liu, C. F., Synthesis of histone H3 proteins by a thioacid capture ligation strategy. *Chem Commun (Camb)* **2011,** *47* (6), 1746-8.

118. Okamoto, R.; Haraguchi, T.; Nomura, K.; Maki, Y.; Izumi, M.; Kajihara, Y., Regioselective alpha-Peptide Bond Formation Through the Oxidation of Amino Thioacids. *Biochemistry* **2019,** *58* (12), 1672-1678.

119. Okamoto, R.; Nomura, K.; Maki, Y.; Kajihara, Y., A Chemoselective Peptide Bond Formation by Amino Thioacid Coupling. *Chemistry Letters* **2019,** *48* (11), 1391-1393.

120. Nomura, K.; Maki, Y.; Okamoto, R.; Satoh, A.; Kajihara, Y., Glycoprotein Semisynthesis by Chemical Insertion of Glycosyl Asparagine Using a Bifunctional Thioacid-Mediated Strategy. *J Am Chem Soc* **2021,** *143* (27), 10157-10167.

121. Kato, M.; Uno, T.; Hiratake, J.; Sakata, K., alpha-Glucopyranoimidazolines as intermediate analogue inhibitors of family 20 beta-N-acetylglucosaminidases. *Bioorg Med Chem* **2005,** *13* (5), 1563-71.

122. Ibatullin, F. M.; Selivanov, S. I., Reaction of N-Fmoc aspartic anhydride with glycosylamines: a simple entry to N-glycosyl asparagines. *Tetrahedron Letters* **2009,** *50* (46), 6351-6354.

123. Liu, R.; Orgel, L. E., Oxidative acylation using thioacids. *Nature* **1997,** *389* (6646), 52-4.

124. Wang, P.; Danishefsky, S. J., Promising general solution to the problem of ligating peptides and glycopeptides. *J Am Chem Soc* **2010,** *132* (47), 17045-51.

125. Dery, S.; Reddy, P. S.; Dery, L.; Mousa, R.; Dardashti, R. N.; Metanis, N., Insights into the deselenization of selenocysteine into alanine and serine. *Chem Sci* **2015,** *6* (11), 6207-6212.

126. Premdjee, B.; Andersen, A. S.; Larance, M.; Conde-Frieboes, K. W.; Payne, R. J., Chemical Synthesis of Phosphorylated Insulin-like Growth Factor Binding Protein 2. *J Am Chem Soc* **2021,** *143* (14), 5336-5342.

127. Koide, T.; Itoh, H.; Otaka, A.; Yasui, H.; Kuroda, M.; Esaki, N.; Soda, K.; Fujii, N., Synthetic study on selenocystine-containing peptides. *Chem Pharm Bull (Tokyo)* **1993,** *41* (3), 502-6.

128. Harris, K. M.; Flemer Jr, S.; Hondal, R. J., Studies on deprotection of cysteine and selenocysteine

side-chain protecting groups. *Journal of Peptide Science* **2007,** *13* (2), 81-93.

129. Mali, S. M.; Gopi, H. N., Thioacetic Acid/NaSH-Mediated Synthesis of N-Protected Amino Thioacids and Their Utility in Peptide Synthesis. *The Journal of Organic Chemistry* **2014,** *79* (6), 2377-2383.

130. Luster, A. D., Chemokines--chemotactic cytokines that mediate inflammation. *N Engl J Med* **1998,** *338* (7), 436-45.

131. Koenen, R. R.; Weber, C., Therapeutic targeting of chemokine interactions in atherosclerosis. *Nat Rev Drug Discov* **2010,** *9* (2), 141-53.

132. Miller, M. C.; Mayo, K. H., Chemokines from a Structural Perspective. *Int J Mol Sci* **2017,** *18* (10), 2088.

133. Baggiolini, M.; Dewald, B.; Moser, B., Human chemokines: an update. *Annu Rev Immunol* **1997,** *15*, 675-705.

134. Haskell, C. A.; Cleary, M. D.; Charo, I. F., Unique role of the chemokine domain of fractalkine in cell capture. Kinetics of receptor dissociation correlate with cell adhesion. *J Biol Chem* **2000,** *275* (44), 34183-9.

135. Izumi, M.; Murakami, M.; Okamoto, R.; Kajihara, Y., Safe and efficient Boc-SPPS for the synthesis of glycopeptide-α-thioesters. *Journal of Peptide Science* **2014,** *20* (2), 98-101.

136. Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. A.; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A. M.; Jensen, K. F.; Pentelute, B. L., Rapid Flow-Based Peptide Synthesis. *ChemBioChem* **2014,** *15* (5), 713-720.

137. Kajihara, Y.; Nishikawa, R.; Maki, Y.; Okamoto, R., Studies in glycopeptide synthesis. *ARKIVOC* **2020,** *2021* (4), 230-240.

138. Liu, H.; Li, X., Serine/Threonine Ligation: Origin, Mechanistic Aspects, and Applications. *Accounts of Chemical Research* **2018,** *51* (7), 1643-1655.

139. Jbara, M.; Laps, S.; Morgan, M.; Kamnesky, G.; Mann, G.; Wolberger, C.; Brik, A., Palladium prompted on-demand cysteine chemistry for the synthesis of challenging and uniquely modified proteins. *Nature Communications* **2018,** *9* (1), 3154.

140. Maity, S. K.; Jbara, M.; Laps, S.; Brik, A., Efficient Palladium-Assisted One-Pot Deprotection of (Acetamidomethyl)Cysteine Following Native Chemical Ligation and/or Desulfurization To Expedite Chemical Protein Synthesis. *Angewandte Chemie International Edition* **2016,** *55* (28), 8108-8112.

141. Yang, Y.-C.; Ciarletta, A. B.; Temple, P. A.; Chung, M. P.; Kovacic, S.; Witek-Giannotti, J. S.; Leary, A. C.; Kriz, R.; Donahue, R. E.; Wong, G. G.; Clark, S. C., Human IL-3 (multi-CSF): Identification by expression cloning of a novel hematopoietic growth factor related to murine IL-3. *Cell* **1986,** *47* (1), 3-10.

142. Schweiger, A.; Stern, D.; Lohman, I. C.; Baldini, M.; Martinez, F. D.; Halonen, M., Differences in proliferation of the hematopoietic cell line TF-1 and cytokine production by peripheral blood leukocytes

induced by 2 naturally occurring forms of human IL-3. *J Allergy Clin Immunol* **2001,** *107* (3), 505-10.

143. Urdal, D. L.; Price, V.; Sassenfeld, H. M.; Cosman, D.; Gillis, S.; Park, L. S., Molecular characterization of colony-stimulating factors and their receptors: human interleukin-3. *Ann N Y Acad Sci* **1989,** *554*, 167-76.

144. Cheng, H.; Zheng, Z.; Cheng, T., New paradigms on hematopoietic stem cell differentiation. *Protein Cell* **2020,** *11* (1), 34-44.

145. Robb, L., Cytokine receptors and hematopoietic differentiation. *Oncogene* **2007,** *26* (47), 6715-23.

146. Socolovsky, M.; Lodish, H. F.; Daley, G. Q., Control of hematopoietic differentiation: lack of specificity in signaling by cytokine receptors. *Proc Natl Acad Sci U S A* **1998,** *95* (12), 6573-5.

147. Katayama, H.; Morisue, S., A novel ring opening reaction of peptide N-terminal thiazolidine with 2,2′-dipyridyl disulfide (DPDS) efficient for protein chemical synthesis. *Tetrahedron* **2017,** *73* (25), 3541-3547.

148. Matveenko, M.; Hackl, S.; Becker, C. F. W., Utility of the Phenacyl Protecting Group in Traceless Protein Semisynthesis through Ligation-Desulfurization Chemistry. *ChemistryOpen* **2018,** *7* (1), 106-110.

149. McElroy, C. A.; Dohm, J. A.; Walsh, S. T. R., Structural and biophysical studies of the human IL-7/IL-7Ralpha complex. *Structure* **2009,** *17* (1), 54-65.

150. Lun, Y. Z.; Wang, X. L.; Feng, J., Purification and identification of the Kazal domain of a novel serine protease inhibitor, Hespintor, through a bacterial (Escherichia coli) expression system. *Int J Mol Med* **2014,** *34* (1), 321-6.

151. Cai, S.; Zhang, P.; Dong, S.; Li, L.; Cai, J.; Xu, M., Downregulation of SPINK13 Promotes Metastasis by Regulating uPA in Ovarian Cancer Cells. *Cell Physiol Biochem* **2018,** *45* (3), 1061-1071.

152. Wei, L.; Lun, Y.; Zhou, X.; He, S.; Gao, L.; Liu, Y.; He, Z.; Li, B.; Wang, C., Novel urokinase-plasminogen activator inhibitor SPINK13 inhibits growth and metastasis of hepatocellular carcinoma in vivo. *Pharmacol Res* **2019,** *143*, 73-85.

153. Xu, W. H.; Shi, S. N.; Wang, J.; Xu, Y.; Tian, X.; Wan, F. N.; Cao, D. L.; Qu, Y. Y.; Zhang, H. L.; Ye, D. W., The Role of Serine Peptidase Inhibitor Kazal Type 13 (SPINK13) as a Clinicopathological and Prognostic Biomarker in Patients with Clear Cell Renal Cell Carcinoma. *Med Sci Monit* **2019,** *25*, 9458-9470.

154. Bause, E.; Hettkamp, H., Primary structural requirements for N-glycosylation of peptides in rat liver. *FEBS Lett* **1979,** *108* (2), 341-4.

155. Pitti, T.; Chen, C. T.; Lin, H. N.; Choong, W. K.; Hsu, W. L.; Sung, T. Y., N-GlyDE: a two-stage N-linked glycosylation site prediction incorporating gapped dipeptides and pattern-based encoding. *Sci Rep* **2019,** *9* (1), 15975.

156. Nakamura, T.; Shigenaga, A.; Sato, K.; Tsuda, Y.; Sakamoto, K.; Otaka, A., Examination of native chemical ligation using peptidyl prolyl thioesters. *Chem Commun (Camb)* **2014,** *50* (1), 58-60.

157. Durek, T.; Alewood, P. F., Preformed selenoesters enable rapid native chemical ligation at intractable sites. *Angew Chem Int Ed Engl* **2011,** *50* (50), 12042-5.

158. Raibaut, L.; Seeberger, P.; Melnyk, O., Bis(2-sulfanylethyl)amido peptides enable native chemical ligation at proline and minimize deletion side-product formation. *Org Lett* **2013,** *15* (21), 5516-9.

159. Kawakami, T.; Aimoto, S., The use of a cysteinyl prolyl ester (CPE) autoactivating unit in peptide ligation reactions. *Tetrahedron* **2009,** *65* (19), 3871-3877.

160. Matsumoto, T.; Sasamoto, K.; Hirano, R.; Oisaki, K.; Kanai, M., A catalytic one-step synthesis of peptide thioacids: the synthesis of leuprorelin via iterative peptide-fragment coupling reactions. *Chem Commun (Camb)* **2018,** *54* (86), 12222-12225.

161. Sureshbabu, V.; Vishwanatha, T.; Samarasimhareddy, M., Facile N-Urethane-Protected α-Amino/Peptide Thioacid Preparation Using EDC and Na2S. *Synlett* **2011,** *2012* (01), 89-92.

162. Sureshbabu, V.; Madhu, C.; Vishwanatha, T., An Efficient Synthesis of Nα-Protected Amino and Peptide Acid Aryl Amides via Iodine-Mediated Oxidative Acylation of Nα-Protected Amino and Peptide Thioacids. *Synthesis* **2013,** *45* (19), 2727-2736.

163. Chen, C.; Huang, Y.; Xu, L.; Zheng, Y.; Xu, H.; Guo, Q.; Tian, C.; Li, Y.; Shi, J., Thiol-assisted one-pot synthesis of peptide/protein C-terminal thioacids from peptide/protein hydrazides at neutral conditions. *Org Biomol Chem* **2014,** *12* (46), 9413-8.

164. Elagawany, M.; Hegazy, L.; Elgendy, B., Catalyst- and organic solvent-free synthesis of thioacids in water. *Tetrahedron Letters* **2019,** *60* (30), 2018-2021.

165. Katritzky, A.; Khaybullin, R.; Panda, S.; Al-Youbi, A., A Facile Synthesis of Thioacids from N-Acylbenzotriazoles. *Synlett* **2013,** *25* (02), 247-250.

166. Kinsland, C.; Taylor, S. V.; Kelleher, N. L.; McLafferty, F. W.; Begley, T. P., Overexpression of recombinant proteins with a C-terminal thiocarboxylate: implications for protein semisynthesis and thiamin biosynthesis. *Protein Sci* **1998,** *7* (8), 1839-42.

167. Barkan, D.; Green, J. E.; Chambers, A. F., Extracellular matrix: a gatekeeper in the transition from dormancy to metastatic growth. *Eur J Cancer* **2010,** *46* (7), 1181-8.

168. Mahmood, N.; Mihalcioiu, C.; Rabbani, S. A., Multifaceted Role of the Urokinase-Type Plasminogen Activator (uPA) and Its Receptor (uPAR): Diagnostic, Prognostic, and Therapeutic Applications. *Front Oncol* **2018,** *8*, 24.

169. Pillay, V.; Dass, C. R.; Choong, P. F. M., The urokinase plasminogen activator receptor as a gene therapy target for cancer. *Trends in Biotechnology* **2007,** *25* (1), 33-39.

170. Jabłońska-Trypuć, A.; Matejczyk, M.; Rosochacki, S., Matrix metalloproteinases (MMPs), the main extracellular matrix (ECM) enzymes in collagen degradation, as a target for anticancer drugs. *J Enzyme Inhib Med Chem* **2016,** *31* (sup1), 177-183.

171. Stanton, H.; Gavrilovic, J.; Atkinson, S. J.; d'Ortho, M. P.; Yamada, K. M.; Zardi, L.; Murphy, G., The activation of ProMMP-2 (gelatinase A) by HT1080 fibrosarcoma cells is promoted by culture on a

fibronectin substrate and is concomitant with an increase in processing of MT1-MMP (MMP-14) to a 45 kDa form. *J Cell Sci* **1998,** *111 (Pt 18)*, 2789-98.

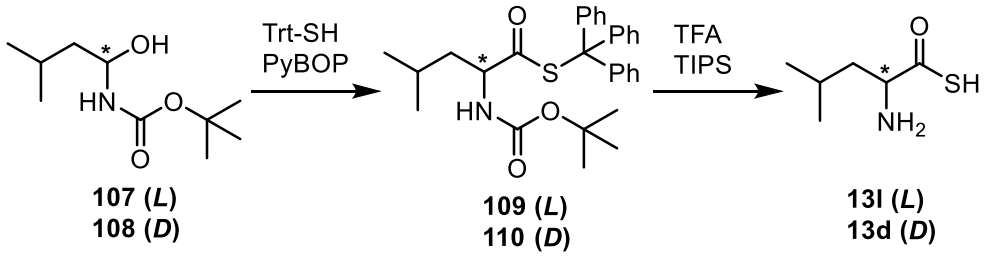
172. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021,** *596* (7873), 583-589.

173. Izumi, M.; Murakami, M.; Okamoto, R.; Kajihara, Y., Safe and efficient Boc-SPPS for the synthesis of glycopeptide-alpha-thioesters. *J Pept Sci* **2014,** *20* (2), 98-101.

Experimental section

1. Synthesis of amino acid derivatives

Scheme E1. Synthesis of leucine thioacids **13l** and **13d**.



Boc-*L*-Leu-STrt (**109**)

To a solution of Boc-*L*-Leu-OH **107** (633 mg, 3.35 mmol) in dry DCM (33.5 mL) was added triphenylmethanethiol (1.85 g, 6.70 mmol) and PyBOP (3.50 g, 6.70 mmol) under argon atmosphere and cooled down at -20 ºC for 10 min. To the solution was added diisopropylethylamine (DIPEA, 1.17 mL, 6.7 mmol) and the resulting mixture was stirred for 1.5 hours at -20 ºC. The mixture was quenched with saturated aqueous ammonium chloride solution and extracted with DCM. The combined organic layers were dried over $MgSO_4$, filtered and concentrated *in vacuo*. Purification of the resultant residue on silica gel flash column chromatography (ethyl acetate /hexane = 1:9) afforded Boc-*L*-Leu-STrt **109** (1.24 g, 83 %) as a white amorphous. $[\alpha]_D^{23}$ = -21.0 (c = 0.50 in dichloromethane) ^{1}H NMR (400 MHz, $CDCl_3$) δ: 7.24 (m), 4.75 (d, 1H), 4.30 (dd, 1H), 1.55 (m, 1H), 1.43 (s, 9H) 1.29 (m, 1H), 0.87 (d, 6H) ppm; ^{13}C NMR (100 MHz, $CDCl_3$) δ: 198.7, 155.2, 143.6, 129.9, 127.9, 127.7, 127.1, 80.1, 70.2, 59.1, 41.5, 28.3, 24.6, 22.9, 21.8 ppm; HRMS (ESI): calcd for $C_{30}H_{35}NO_3SNa$ $[M + Na]^+$ 512.2236 found for 512.2228.

H-*L*-Leu-SH (**13l**)

To the solution of Boc-*L*-Leu-STrt **109** (1.24g, 2.78 mmol) in TFA (12.0 mL) was added triisopropylsilane (TIPS, 0.6 mL) and stirred for 10 min at room temperature. The resulting mixture was evaporated *in vacuo* and added chilled diethyl ether (Et_2O, 40 mL) to afford white precipitate. The suspension was centrifuged, and Et_2O was removed by decantation. The isolated precipitate was then washed with chilled Et_2O twice as the same manner. Lyophilization of the precipitate afforded H-*L*-Leu-SH **13l** (278 mg, 75 %) as a white powder. $[\alpha]_D^{23}$ = -16.8 (c = 0.50 in MeOH) ^{1}H NMR (400 MHz, D_2O) δ: 3.84 (dd, 1H), 1.76 (m, 1H), 1.54 (m, 2H), 0.84 (d, 6H) ppm; ^{13}C NMR (100 MHz, D_2O) δ: 215.8, 61.6, 41.6, 24.2, 22.1, 20.8 ppm; HRMS (ESI): m/z calcd for $C_6H_{13}NOSNa$ $[M + Na]^+$ 170.0616, found for 170.0612.

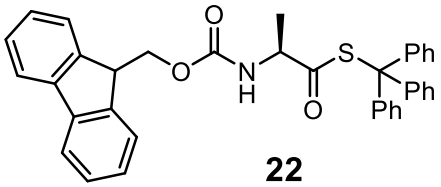
Boc-*D*-Leu-STrt (110)

Synthesis of Boc-*D*-Leu-STrt **110** was performed by the same manner with the synthesis of **109**, namely Boc-*D*-Leu-OH **108** (249 mg, 1.08 mmol) afforded Boc-*D*-Leu-STrt **110** (301 mg, 51%) as a white powder. $[\alpha]_D^{23}$ = +29.5 (c = 0.82 in dichloromethane), ^{1}H NMR (400 MHz, $CDCl_3$) δ: 7.25 (m), 4.75 (d, 1H), 4.31 (dd, 1H), 1.56 (s, 2H), 1.46 (s, 9H), 1.30 (m, 1H), 0.88 (d, 6H) ppm; ^{13}C NMR (100 MHz, $CDCl_3$) δ: 189.7, 155.3, 143.6, 129.9, 127.8, 127.7, 127.1, 80.1, 70.2, 59.1, 41.5, 28.3, 24.6, 22.9, 21.8 ppm; HRMS (ESI): calcd for $C_{30}H_{35}NO_3SNa$ $[M + Na]^+$ 512.2236 found for 512.2228.

H-*D*-Leu-SH (41)

Synthesis of H-*D*-Leu-SH **13d** was performed in the same manner to the synthesis of **13l**, consequently Boc-*D*-Leu-STrt **110** (114 mg, 0.284 mmol) afforded H-*D*-Leu-SH **13d** (26.2 mg, 76%) as a white powder. $[\alpha]_D^{23}$ = +16.2 (c = 0.50 in MeOH) ^{1}H NMR (400 MHz, D_2O) δ: 3.86 (dd, 1H), 1.79 (m, 1H), 1.55 (m, 2H), 0.86 (d, 6H); ^{13}C NMR (100 MHz, D_2O) δ: 215.8, 61.6, 41.6, 24.2, 22.1, 20.1 ppm; HRMS(ESI): m/z calcd for $C_6H_{13}NOSNa$ $[M + Na]^+$ 170.0616, found for 170.0612

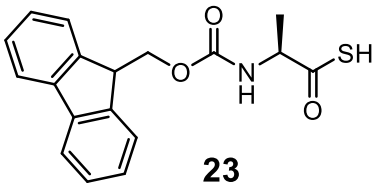
Fmoc-*L*-Ala-STrt (22)



22

To a solution of Fmoc-*L*-Ala-OH **21** (501 mg, 1.61 mmol), Triphenylmethanethiol (1.31 g, 4.73 mmol), PyBOP (2.50 g, 4.81 mmol) in dichloromethane (DCM, 16.0 mL) was added diisopropylethylamine (DIPEA, 0.84 mL, 4.8 mmol) at -20 ºC under argon atmosphere. The resulting mixture was stirred for 1.5 hours at -20 ºC. The mixture was quenched with saturated aqueous ammonium chloride solution and extracted with DCM. The combined organic layers were dried over $MgSO_4$, filtered and concentrated *in vacuo*. Purification of the resultant residue by flash column chromatography on silica gel (ethyl acetate/ hexane = 1:9) afforded Fmoc-*L*-Ala-STrt **22** (668 mg, 73%) as a white amorphous. $[\alpha]_D^{23}$ = -40.8 (c = 0.70 in dichloromethane) ^{1}H NMR (400 MHz, $CDCl_3$) δ: 7.78 (d, 2H), 7.61 (d, 2H), 7.41 (t, 3H), 7.26 (m, 17H), 5.25 (d, 1H), 4.47 (m, 3H), 4.36 (m, 1H), 4.23 (t, 1H), 1.32 (d, 3H) ppm. ^{13}C NMR (100 MHz, $CDCl_3$) δ: 198.0, 143.5, 141.3, 129.8, 127.8, 127.2, 125.1, 120.0, 70.7, 56.7, 47.2, 19.2 ppm. HRMS (ESI): calcd for $C_{37}H_{31}NO_3Sna$ $[M+Na]^+$ 592.1923, found for 592.1928

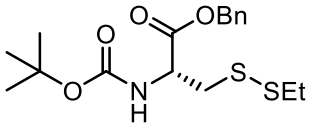
Fmoc-*L*-Ala-SH (**23**)



23

To a solution of Fmoc-*L*-Ala-STrt **22** (101 mg, 0.178 mmol) in TFA (5.0 mL) was added triisopropylsilane (TIPS, 0.25 mL, 1.2 mmol, 6.7 equiv.) at room temperature. The resulting mixture was stirred for 10 min and evaporated *in vacuo*. To the residue, CH_3CN and water was added for lyophilization. The residue was purified by preparative HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O: 0.1 % TFA in CH_3CN = 50 : 50 to 00 : 100 over 60 min at the flow rate of 2.5 mL/min) to afforded Fmoc-*L*-Ala-SH **23** (50.2 mg, 87%) as a white powder. $[\alpha]_D^{23}$ = -4.4 (c = 0.52 in MeOH), ^{1}H NMR (400 MHz, DMSO-D6) δ: 7.95 (d, 2H), 7.90 (d, 2H), 7.73 (t, 2H), 7.42 (t, 2H), 7.34 (ddd, 2H), 4.36 (m, 1H), 4.24 (m, 3H), 1.30 (d, 3H) ppm; ^{13}C NMR (100 MHz, DMSO-D6) δ: 197.4, 156.3, 144.1, 141.2, 128.1, 127.5, 125.7, 120.6, 57.2, 49.7, 47.1, 17.5 ppm; HRMS (ESI): calcd for $C_{18}H_{17}NO_3SNa$ $[M+Na]^+$ 350.0827, found for 350.0806

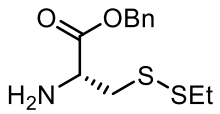
Boc-*L*-Cys(SEt)-OBn (**25**)



25

To a solution of Boc-*L*-Cys(SEt)-OH **24** (498 mg, 1.34 mmol) in dry DCM (10.0 mL) was added benzyl bromide (0.40 mL, 3.4 mmol) at room temperature under argon atmosphere. The mixture was cooled down to 0 ºC and stirred for 10 min. To the reaction mixture was added DBU (0.31 mL, 2.1 mmol) and stirred at 0 ºC. The resulting mixture was warmed up to room temperature. The mixture was quenched with saturated aqueous ammonium chloride and extracted with DCM. The combined organic layers were washed with saturated brine, dried over with $MgSO_4$, filtered and concentrated *in vacuo*. The residue was purified by flash column chromatography on silica gel (ethyl acetate/ hexane = 1: 7) to afford Boc-*L*-Cys(SEt)-OBn **25** (259 mg, 40%) as a colorless oil. $[\alpha]_D^{23}$ = +19.4 (c = 0.61 in dichloromethane), ^{1}H NMR (400 MHz, $CDCl_3$) δ: 7.35 (s, 5H), 5.39 (d, 1H), 5.18 (s, 2H), 4.64 (m, 1H), 3.15 (dd, 2H), 2.67 (q, 2H), 1.44 (s, 9H), 1.27 (t, 3H) ppm; ^{13}C NMR (100 MHz, $CDCl_3$) δ: 170.7, 155.1, 135.2, 128.4, 80.2, 67.5, 53.2, 41.2, 32.7, 28.3, 14.3 ppm; HRMS (ESI): calcd for $C_{17}H_{25}NO_4S_2Na$ $[M+Na]^+$ 394.1123, found for 394.1144

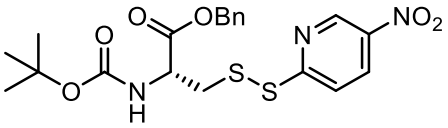
H-*L*-Cys-OBn (26)



26

Compound Boc-*L*-Cys(SEt)-OBn **25** (23.1 mg, 0.0623 mmol) was dissolved in TFA (5.0 mL) at room temperature. The resulting mixture was stirred for 10 min and evaporated *in vacuo*. To this suspension was added CH_3CN and H_2O (1.0 mL and 1.0 mL respectively) for lyophilization afforded H-*L*-Cys(SEt)-OBn **26** (16.7 mg, quantitative) as a reddish syrup. $[\alpha]_D^{23}$ = -23.5 (c = 0.59 in MeOH), ^{1}H NMR (400 MHz, DMSO-D6) δ: 7.38 (m, 5H), 5.25 (m, 2H), 4.42 (dd, 1H), 3.22 (dd, 2H), 2.51 (q, 2H), 1.23 (t, 3H) ppm; ^{13}C NMR (100 MHz, DMSO-D6) δ: 168.6, 135.4, 128.8, 67.9, 51.8 38.3, 31.8, 14.6 ppm; HRMS (ESI): calcd for $C_{12}H_{18}NO_2S_2$ $[M+H]^+$ 272.0779, found for 272.0771.

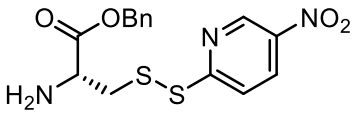
Boc-*L*-Cys(Npys)-OBn (28)



28

To a solution of Boc-*L*-Cys(SEt)-OBn **25** (124 mg, 0.333 mmol) in MeOH (5.0 mL) was added 1,4-dithiothreitol (513 mg, 3.33 mmol) and DIPEA (0.030 mL, 0.17 mmol) at room temperature. The resulting mixture was stirred for 1 h and afforded Boc-*L*-Cys-OBn **27** under argon atmosphere. To the solution of crude **27** in MeOH (5.0 mL) was added 2,2'-dithiobis(5-nitropyridine) (2.0 g, 6.5 mmol) dissolved in MeOH (5.0 mL) and stirred under argon atmosphere for 12 h at room temperature. The filtrate was collected by filtration and evaporated *in vacuo*. The resulting residue was purified by flash column chromatography on silica gel (ethyl acetate/ hexane = 1: 5) to afford Boc-*L*-Cys(NPys)-OBn **28** (117 mg, 75%) as a yellow oil. $[\alpha]_D^{23}$ = +7.6 (c = 0.67 in dichloromethane), ^{1}H NMR (400 MHz, $CDCl_3$) δ: 9.31 (dd, 2H), 8.41 (m, 2H), 7.70 (m, 2H), 7.33 (m, 5H), 6.19 (d, 1H), 5.17 (dd, 2H), 4.64 (m, 1H), 3.47 (dd, 1H), 3.34 (dd, 1H), 1.48 (s, 9H) ppm; ^{13}C NMR (100 MHz, $CDCl_3$) δ: 165.0, 145.2, 142.3, 134.9, 132.1, 131.6, 128.6, 119.8, 67.7, 53.1, 41.8, 28.3 ppm; HRMS (ESI): calcd for $C_{20}H_{23}N_3O_6S_2Na$ $[M+Na]^+$ 488.0926, found for 488.0909.

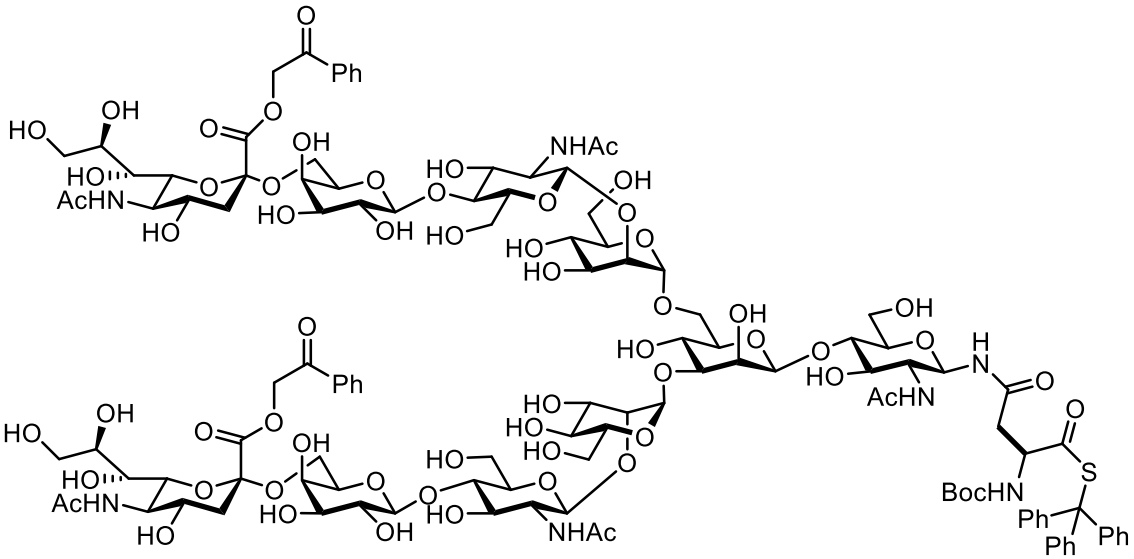
H-*L*-Cys(Npys)-OBn (**29**)



29

To the solution of Boc-*L*-Cys(NPys)-OBn **28** (18.1 mg, 38.7 μmol) in TFA (10.0 mL) at room temperature. The resulting mixture was stirred for 5 min and evaporated *in vacuo*. To this suspension was added CH_3CN-H_2O (1.0 mL-1.0 mL) and then lyophilized to afford H-*L*-Cys(NPys)-OBn **29** (14.0 mg, quantitative) as a yellow powder. $[\alpha]_D^{23}$ = +36.8 (c = 0.51), ^{1}H NMR (400 MHz, DMSO-D6) δ: 9.35 (dd, 2H), 8.41 (m, 2H), 7.74 (m, 2H), 7.51 (d, 1H), 7.34 (m, 5H), 5.24 (s, 1H), 4.26 (s, 2H), 3.59 (s, 2H) ppm; ^{13}C NMR (100 MHz, DMSO-D6) δ: 165.0, 145.3, 142.8, 133.9, 131.9, 128.8, 119.8, 68.9, 51.8, 38.9 ppm; HRMS (ESI): calcd for $C_{15}H_{16}N_3O_4S_2$ $[M+H]^+$ 366.0582, found for 366.0588.

Boc-Asn(sialyloligosaccharide phenacyl ester)-STrt (**10**)

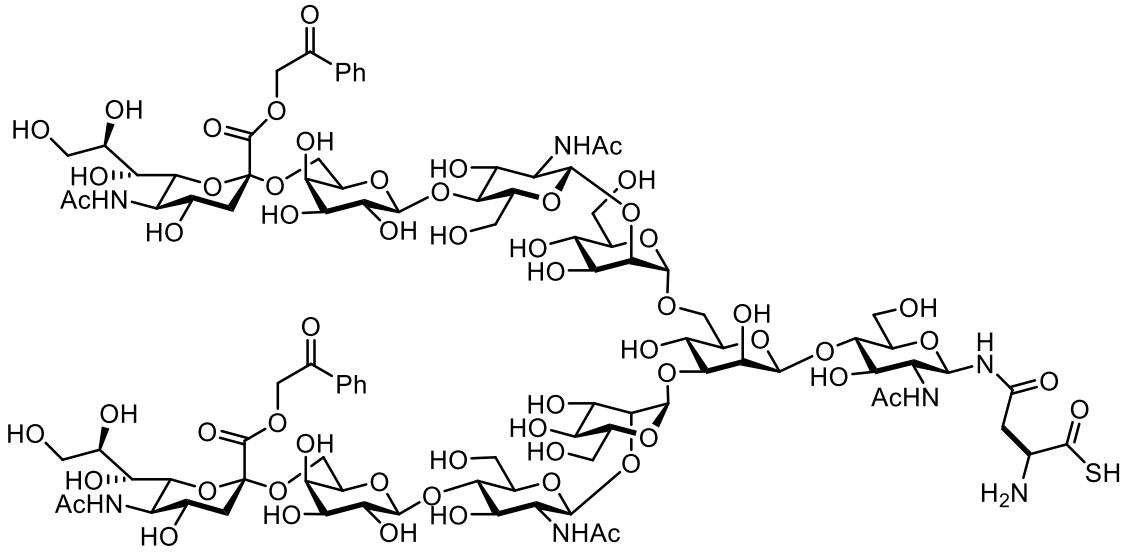


10

Boc-Asn(sialyloligosaccharide phenacyl ester)-OH **9** was prepared by our previous method.[82] To a solution of compound **9** (18.0 mg, 6.7 μmol) in DMF (486 μL) was added triphenylmethanethiol (47.2 mg, 0.17 mmol), PyBOP (29.5 mg, 56.6 μmol) and DIPEA (10.0 μL , 57.3 μmol) and this mixture was stirred at -15 ℃ under argon atmosphere. After 1.5 h, chilled Et_2O (20 mL) was added to this mixture and then precipitate was collected by centrifugation. The product in the precipitate was purified by preparative HPLC (Capcell Pak C18 Φ10 × 250 mm, 50 mM ammonium acetate aqueous solution: CH_3CN = 65: 35 to 20: 80 over 80 min at the flow rate of 3 mL/min). The fractions containing the product were pooled and then passed through a short cation exchange resin-column (Dowex-50Wx8(H^+), Φ0.5 cm x 5 cm) and the eluent was lyophilized to obtain Boc-Asn(sialyloligosaccharide

phenacyl ester)-STrt **10** (11.0 mg, 57%). ^{1}H NMR (400 MHz, DMSO-D6, D_2O) δ: 7.88 (d, 4H, Phenyl), 7.65 (t, 2H, Phenyl), 7.51 (t, 4H, Phenyl), 7.22 (t, 6H, trityl), 7.10 (d, 3H, trityl), 5.55 (s, 4H, PhenacylCH_2), 4.94 (s, 1H, Man-H-1), 4.84 (m, 1H, J = 13.0 Hz, AsnGlcNAc-H-1), 4.78 (m, 1H, Man-H-1), 4.54 (s, 1H, Man-H-1), 4.30 (m, 2H, Gal-H-1), 4.20 (m, 1H, Asn-αH), 2.55 (d, 2H, NeuAc-H-3_{eq}), 1.88-1.84 (each s, Acetyl), 1.74 (dd, 2H, NeuAc-H-3_{ax}), 1.29 (m, 9H, Boc-*t*butyl) ppm; HRMS (ESI): calcd for $C_{128}H_{179}N_8O_{67}S$ $[M+H]^+$ 2932.0566, found for 2932.0443.

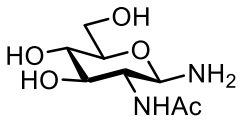
H-Asn(sialyloligosaccharide phenacyl ester)-SH (**11**)



11

To the solution of Boc-Asn(sialyloligosaccharide phenacyl ester)-STrt **10** (8.5 mg, 2.9 μmol) in TFA (750 μL) was added TIPS (38 μL) at room temperature. After 30 min, chilled diethyl ether (Et_2O, 10 mL) was added to this mixture and then precipitate was collected by centrifugation and dried in *vacuo* with vacuum pump (0.1 Pa) for 1h to obtain H-Asn(sialyloligosaccharide phenacyl ester)-SH **11** (6.9 mg, 92%). ^{1}H NMR (400 MHz, D_2O) δ: 7.90 (d, 4H, Phenyl), 7.68 (t, 2H, Phenyl), 7.52 (t, 4H, Phenyl), 5.64 (s, 4H, PhenacylCH_2), 5.06 (s, 1H, Man-H-1), 4.97 (d, 1H, J = 9.5 Hz, AsnGlcNAc-H-1), 4.87 (s, 1H, Man-H-1), 4.53 (m, 3H, GlcNAc-H-1), 4.52 (d, 2H, J = 7.60 Hz, Gal-H-1), 4.17 (s, 1H, Man-H-2), 4.11 (m, 1H, Man-H-2), 4.10 (m, 1H, Asn-αH), 4.03 (m, 1H, Man-H-2), 3.86 (m, 2H, NeuAc-H-4), 3.77 (m, 1H, AsnGlcNAc-H-2), 3.68 (m, 3H, GlcNAc-H-2), 3.61 (m, 2H, NeuAc-H-5), 3.47 (m, 2H, Gal-H-2), 2.99 (m, 2H, Asn-βH), 2.68 (dd, 2H, NeuAc-H-3_{eq}), 1.99-1.94 (each s, 6H, Acetyl), 1.87 (m, 2H, NeuAc-H-3_{ax}) ppm; HRMS (ESI): calcd for $C_{104}H_{157}N_8O_{65}S$ $[M+H]^+$ 2589.8946, found for 2589.8790.

2-Acetamido-2-deoxy-β-D-glucopyranosylamine (**2**)

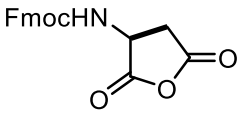


2

To N-Acetyl-D-glucosamine **1** (5.00 g, 22.6 mmol) and $(NH_4)CO_3$ (1.8 g, 22.8 mmol) were added to 2M ammonia methanol solution (30 mL) and the reaction mixture was stirred under Ar at 40 ˚C for 2 days. The white suspension became a transparent solution in 2 days. Et_2O (100 mL) was added to the reaction mixture at 4 ˚C, and the precipitate was collected by filtration and washed with EtOH and Et_2O. The crude product was dissolved in MeOH (100 mL) and was evaporated to decompose the carbamate. Evaporation was repeated for five times to give pure glycosylamine **2** whitish solid (3.57 g, 71%).

^{1}H NMR (400 MHz, D_2O) δ 4.15 (d, 1H, J = 9.1Hz), 3.88 (dd, 1H), 3.70 (dd, 1H), 3.61 (dd, 1H), 3.49 (m, 1H), 3.4 (m, 2H), 2.04 (s, 3H) ppm; ^{13}C NMR (100 MHz, D_2O) δ: 174.7, 84.2, 76.8, 74.5, 70.1, 60.9, 56.4, 22.3 ppm. MS (ESI): Calcd for $C_8H_{16}N_2NaO_5$ $[M+Na]^+$ 243.10, found for 243.10

N-Fmoc-L-aspartic anhydride (**3**)

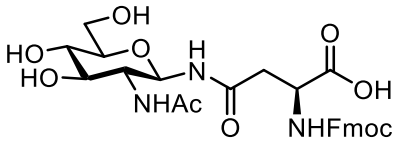


3

Compound N-Fmoc-*L*-Asp(O^{t}Bu)-OH (5.00 g, 12.2 mmol) was dissolved in TFA (5.0 mL) at room temperature. The resulting mixture was stirred for 1.5 h and evaporated *in vacuo*. The crude product was dissolved in MeOH (100 mL) and was evaporated three times to afford a white solid. To the resulting product was dissolved in acetic anhydride (15 mL) and the reaction mixture was stirred at 110 ˚C for 5 min. Then the reaction mixture quickly cooled to room temperature to afford a whitish crystalline solid. After filtration and washing with Et_2O and dried *in vacuo* to give pure N-Fmoc-L-aspartic anhydride 3 white crystalline solid (2.50 g, 61%).

^{1}H NMR (400 MHz, DMSO-D6) δ 8.19 (d, 1H, J = 8.0 Hz), 7.90 (d, 2H, J = 7.5 Hz), 7.68 (d, 2H, J = 7.5 Hz), 7.43 (t, 2H, J = 7.3, Hz), 7.34 (t, 2H, J = 7.3 Hz), 4.67 (dt, 1H, J1 = 9.5 Hz, J2 = 6.7 Hz), 4.43 (dd, 1H, J = 10.6 Hz), 4.40 (dd, 1H, J = 10.6 Hz), 4.26 (dd, 1H, J = 6.3 Hz), 3.25 (dd, 1H, J = 18.4 Hz), 2.89 (dd, 1H, J = 17.8 Hz) ppm; ^{13}C NMR (100 MHz, DMSO-D6) δ: 172.6, 170.3, 156.3, 144.1, 144.0, 141.3, 128.1, 127.5, 127.4, 125.5, 125.4, 120.6, 120.5, 66.5, 50.8, 47.0. 35.2 ppm. MS (ESI): calcd for $C_{19}H_{15}NO_5$ $[M+H]^+$ 338.10 found for 338.10

N-[N-Fmoc-L-aspart-4-oyl]-β-D-glucopyranosylamine (4)

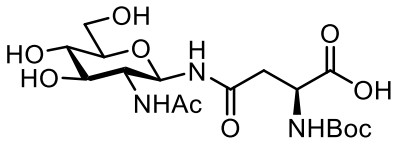


4

To a stirred solution of compound **2** (400 mg, 1.82 mmol) in DMSO (3 mL), N-L-Fmoc-aspartic anhydride **3** (300 mg, 0.890 mmol) was added. The reaction mixture was gently stirred at room temperature for 3.5 h. After stirring, the reaction mixture was congealed by liquid nitrogen and lyophilized to give brown solid. The crude product was purified by flash column chromatography (3:5, MeOH/AcOEt) to afford the compound **4** (244 mg, 51%) as a white solid.

^{1}H NMR (400 MHz, DMSO-D6 + 5%D_2O) δ 7.93 (d, 1H, J = 8.7 Hz, CONH), 7.90 (d, 2H, J = 7.4 Hz, Ar-H Fmoc), 7.70 (d, 2H, J = 7.2 Hz, Ar-H Fmoc), 7.42 (dd, 2H, J = 7.4, Hz, Ar-H Fmoc), 7.33 (dd, 2H, J = 7.4 Hz, Ar-H Fmoc), 4.82 (d, 1H, J = 9.7 Hz, H-1), 4.26 (dd, 2H, CH_2-Fmoc), 4.21 (dd, 1H, CH-Fmoc), 4.19 (dt, 1H, J = 7.4 Hz, α-CH), 3.66 (dd, 1H, J = 11.0 Hz, H-6a), 3.50 (dd, 1H, H-2), 3.44 (dd, 1H, J1 = 11.5 Hz, J2 = 4.3 Hz, H-6b), 3.34 (dd, 1H, J = 7.4 Hz, H-3), 3.11 (dd, 1H, H-5), 3.09 (dd, 1H, J = 10.1 Hz, H-4), 2.57 (dd, 1H, βa-CH_2), 2.52 (dd, 1H, βb-CH_2), 1.80 (s, 3H) ppm; ^{13}C NMR (100 MHz, DMSO-D6 + 5%D_2O) δ: 173.9, 170.6, 156.1, 144.3, 144.2, 141.1, 128.1 (Ar-Fmoc), 127.6 (Ar-Fmoc), 125.7 (Ar-Fmoc), 120.5 (Ar-Fmoc), 79.0 (C-1), 74.6 (C-3), 70.7 (C-4), 66.1 (Fmoc-CH_2), 61.2 (C-6), 54.9 (C-2), 51.5 (α-CH), 47.1 (Fmoc-CH), 23.2 (acetyl) ppm. MS (ESI): calcd for $C_{27}H_{31}N_3NaO_{10}$ $[M+Na]^+$ 580.19, found for 580.19.

N-[N-Boc-L-aspart-4-oyl]-β-D-glucopyranosylamine (5)

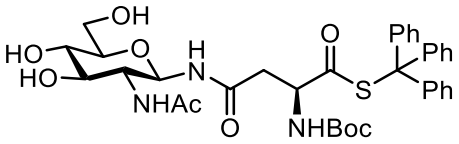


5

Compound **4** (142 mg, 0.25 mmol) was dissolved in 20% piperidine / DMF (8.0 mL) at room temperature. After 30 min Et_2O (10 mL) was added to this mixture and then precipitate was collected by centrifugation. The precipitate was then dissolved in DMF (4.0 mL) and BocOSu (880 mg, 4.04 mmol) and DIPEA (220 μL, 1.26 mmol) was added to this mixture. After 12 h, Et_2O (10 mL) was added to this mixture and then precipitate was collected by centrifugation. The crude product was dissolved in toluene and was coevaporated three times to remove DMF. The crude product was purified by flash column chromatography (1:1, MeOH/AcOEt) to afford the compound **5** (83 mg, 74%) as a white solid.

^{1}H NMR (400 MHz, DMSO-D6 + 5%D_2O) δ 4.80 (d, 1H, J = 9.8 Hz), 3.94 (dt, 1H, J = 7.1 Hz), 3.66 (dd, 1H, J = 11.0 Hz), 3.50 (dd, 1H), 3.44 (dd, 1H, J1 = 11.5 Hz, J2 = 4.3 Hz), 3.34 (dd, 1H, J = 7.4 Hz), 3.11 (m, 2H), 2.57 (dd, 1H), 2.52 (dd, 1H), 1.80 (s, 3H), 1.37 (s, 9H) ppm; ^{13}C NMR (100 MHz, DMSO-D6 + 5%D_2O) δ: 173.5, 171.2, 155.4, 78.9, 78.8, 78.3, 74.5, 70.7, 61.1, 54.8, 52.2, 28.6, 25.6, 23.2 ppm. MS (ESI): calcd for $C_{17}H_{29}N_3NaO_{10}$ $[M+Na]^+$ 458.18, found for 458.18

N-Boc-L-Asn(β-D-glucopyranosylamine)-STrt (**6**)

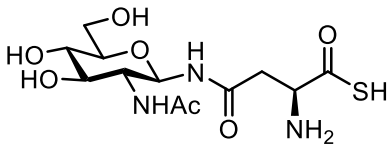


6

To a solution of compound **5** (45 mg, 0.10 mmol) in DMF (2.8 mL) was added triphenylmethanethiol (300 mg, 1.09 mmol), PyBOP (300 mg, 0.58 mmol) and DIPEA (112 μL , 6.42 mmol) and this mixture was stirred at -15 ℃ under Ar atmosphere. After 6 h, chilled Et_2O (20 mL) was added to this mixture and then precipitate was collected by centrifugation. The product in the precipitate was purified by preparative HPLC (Capcell Pak C18 Φ10 × 250 mm, 0.1% TFA aqueous solution: 0.1% TFA CH_3CN = 65: 35 to 30: 70 over 80 min at the flow rate of 3 mL/min) to give compound **6** (16 mg, 22%) as a white powder.

^{1}H NMR (400 MHz, DMSO-D6 + 5%D_2O) δ 7.29 (m, 11H), 7.13 (m, 7H), 4.80 (d, 1H, J = 9.8 Hz), 4.26 (dt, 1H, J = 6.2 Hz), 3.64 (dd, 1H, J = 11.0 Hz), 3.49 (dd, 1H, J = 9.5 Hz), 3.42 (dd, 1H, J1 = 11.5 Hz, J2 = 4.3 Hz), 3.33 (dd, 1H, J = 7.4 Hz), 3.11 (dd, 1H, J = 8.5 Hz), 3.08 (dd, 1H, J = 9.5 Hz), 2.39 (dd, 2H, J = 6.5 Hz), 1.79 (s, 3H), 1.40 (s, 9H) ppm; ^{13}C NMR (100 MHz, DMSO-D6 + 5%D_2O) δ: 198.6, 170.7, 169.4, 155.5, 143.8, 129.8, 128.2, 79.5, 79.1, 79.0, 74.7, 70.7, 69.1, 61.1, 57.8, 54.9, 36.8, 28.6, 23.2 ppm. MS (ESI): calcd for $C_{36}H_{44}N_3O_9S$ $[M+H]^+$ 694.28, found for 694.28

H-L-Asn(β-D-glucopyranosylamine)-SH (7)



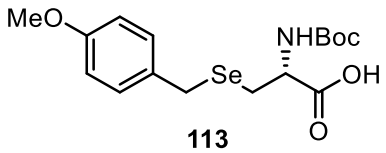
7

To the solution of N-Boc-L-Asn(β-D-glucopyranosylamine)-STrt **6** (9.2 mg, 13 μmol) in TFA (3.0 mL) was added TIPS (150 μL) at room temperature. After 30 min, chilled diethyl ether (Et_2O, 10 mL) was added to this mixture and then precipitate was collected by centrifugation and dried in *vacuo* with vacuum pump (0.1 Pa) for 1h to obtain H-L-Asn(β-D-glucopyranosylamine)-SH **7** (4.3 mg, 92%).

$[\alpha]_D^{23}$ = +12.0 (c = 0.50)

^{1}H NMR (400 MHz, D_2O) δ 5.07 (d, 1H, J = 9.8 Hz, H-1), 4.21 (dt, 1H, J = 4.8 Hz, α-CH), 3.89 (dd, 1H, J1 = 1.6 Hz, J2 = 12.0 Hz, H-6a), 3.83 (dd, 1H, J = 10.0 Hz, H-2), 3.75 (dd, 1H, J1 = 4.5, J2 = 12.2 Hz, H-6b), 3.61 (dd, 1H, J = 8.5 Hz, H-3), 3.53 (dd, 1H, H-5), 3.48 (dd, 1H, J = 9.7 Hz, H-4), 3.16 (dd, 1H, J = 6.2 Hz, βa-CH_2), 3.05 (dd, 1H, J = 6.2 Hz, βa-CH_2), 2.04 (s, 3H, Acetyl) ppm; ^{13}C NMR (100 MHz, D_2O) δ: 211.8 (COSH), 175.0 (CONH), 172.6 (CONH), 78.1 (C-1), 77.6 (C-5), 74.3 (C-3), 69.6 (C-4), 60.6 (C-6), 59.0 (C-α), 54.1 (C-2), 36.7 (C-β), 22.4 (acetyl) ppm. MS (ESI): calcd for $C_{12}H_{22}N_3O_7S$ $[M+H]^+$ 352.12, found for 352.12.

N-Boc-L-Sec(PMB)-OH (**113**)



113

To L-selenocystine **1** (500 mg, 1.5 mmol) in 0.5 M NaOH aq. (8.0 mL) and EtOH (2.0 mL) was added $NaBH_4$ (400 mg, 15.0 mmol). The reaction mixture was gently stirred at 0 ˚C for 10 min then the yellow color disappeared. To the reaction mixture was added additional 2 M NaOH aq. (4.0 mL) and p-methoxybenzyl chloride (0.82 mL, 6.0 mmol). The reaction was stirred for 4 h at room temperature and, then the reaction was quenched by adjusting pH 1.0 by HCl followed by evaporation with toluene to yield white solid. The resulting precipitate was dissolved in CH_3CN 20 mL and saturated $NaHCO_3$ solution 20 mL. The reaction was stirred for 15 min at room temperature, followed by the addition of Boc_2O (753 mg, 0.80 mL, 3.5 mmol). After stirring for 12 h, the reaction was quenched with 10% citric solution (60 mL). To the resulting mixture was added EtOAc (50 mL). The organic layer was separated, and the aqueous layer was extracted with EtOAc (2 × 50 mL). The combined organic layers were dried over Na_2SO_4, filtered, and concentrated *in vacuo*. The crude product was purified by flash column chromatography (1:2, EtOAc/hexane, 2% AcOH) to afford the compound **113** (967 g, 83%) as a colorless oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.20 (d, 2H, J = 8.6 Hz, ArH-methoxy), 6.82 (d, 2H, J = 8.6 Hz, ArH-methoxy), 5.28 (s, 1H, NH), 4.61 (dt, 1H, α-H), 3.78 (dd, 2H, methoxy-CH_2), 3.78 (s, 3H, methoxy-CH_3), 2.92 (d, 2H, J = 4.8 Hz, β-H), 1.45 (s, 9H, CH_3) ppm; ^{13}C NMR (100 MHz, $CDCl_3$) δ: 158.6 (CONH), 130.5 (COOH), 130.0 (Ar), 114.1 (Ar), 55.3 (methoxy-CH_3), 53.2 (α-C), 28.3 (Boc-C), 27.6 (methoxy-CH_2), 25.2 (Boc-CH_3) ppm. MS (ESI): calcd for $C_{16}H_{22}NNaO_5Se$ $[M+Na]^+$ 411.06, found for 411.06

2. Synthesis of model peptide and model glycopeptide

Model peptide thioacid (12), (15) and (16)

The model peptide **12**, **15** and **16** segments were synthesized on Aminomethyl ChemMatrix® resin (50 μmol scale) by our improved safe Boc solid-phase peptide synthesis (SPPS).[2] The peptide chain was elongated by standard Boc-based protocol.[173] Firstly to a resin (50.0 μmol, 1.0 equiv) was added anhydrous dimethyl formamide (DMF, 1.0 mL) containing HBTU (72.0 mg, 0.190 mmol, 3.8 equiv.), S-trityl-3-mercaptopropionic acid (70.0 mg, 0.200 mmol, 4.0 equiv.) and diisopropylethylamine (DIPEA, 70.0 μL, 0.400 mmol, 8.0 equiv.) and stirred for 30 min. All reagents were filtered and the resultant resin was washed with DCM and DMF. Trityl group was removed by treatment of the resin with a mixture containing TFA (4.5 mL) and TIPS (250 μL) for 5 min at room temperature. After preparation of 3-mercaptopropionic acid-linker on a resin, Boc amino acids were elongated. The first amino acid (0.200 mmol, 4.0 equiv.) was pre-activated with HBTU (0.190 mmol, 3.8 equiv.) and DIPEA, (0.400 mmol, 8.0 equiv.) in anhydrous DMF (1.0 mL) for 1 min and then the activated amino acid was added to the suspension of the resin. The suspension was stirred for 15 min at room temperature. After the coupling of a first amino acid, the resin was washed with DCM and DMF and then Boc group was deprotected with TFA (5.0 mL) for 5 min twice. The following couplings of other Boc amino acids were performed by the same manner with the first amino acid coupling. After all coupling steps were complete, the fully protected peptide was deprotected with a cocktail containing TFA (5.0 mL), thioanisole (0.50 mL), ethanedithiol (0.25 mL) and trifluromethanesulfonic acid (0.5 mL). Then the mixture was allowed to 0°C and stirred for 1 hour. During this time, the solution color slowly changed from red to bright yellow. After deprotection of side chain protecting groups, an acid cocktail solution was removed by the filtration and the resultant resin was washed with DCM and DMF. Peptide thioacid was released from the resin by a thiolysis condition using a buffer solution (5.0 mL, pH 8.0) containing 6.0 M Gdn-HCl, 0.2 M sodium phosphate and 0.2 M Na_2S for 3 hours at room temperature. Purification of the resultant peptide thioacid was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 90 : 10 to 50 : 50 over 40 min at the flow rate of 2.5 mL/min) afforded model peptide thioacids **12, 15** and **16**.

H_2N-YGEFA-COSH (12)

HRMS(ESI): *m*/z calcd. For $C_{28}H_{36}N_5O_8S$: $[M+H]^+$ 602.2206, found for 602.2272. (Figure E1a)

H_2N-YGEFV-COSH (15)

HRMS(ESI): *m*/z calcd. For $C_{30}H_{40}N_5O_8S$: $[M+H]^+$ 630.2519, found for 630.2585. (Figure E1b)

H_2N-YGEFF-COSH (16)

HRMS(ESI): *m*/z calcd. For $C_{34}H_{40}N_5O_8S$: $[M+H]^+$ 678.2519, found for 678.2577. (Figure E1c)

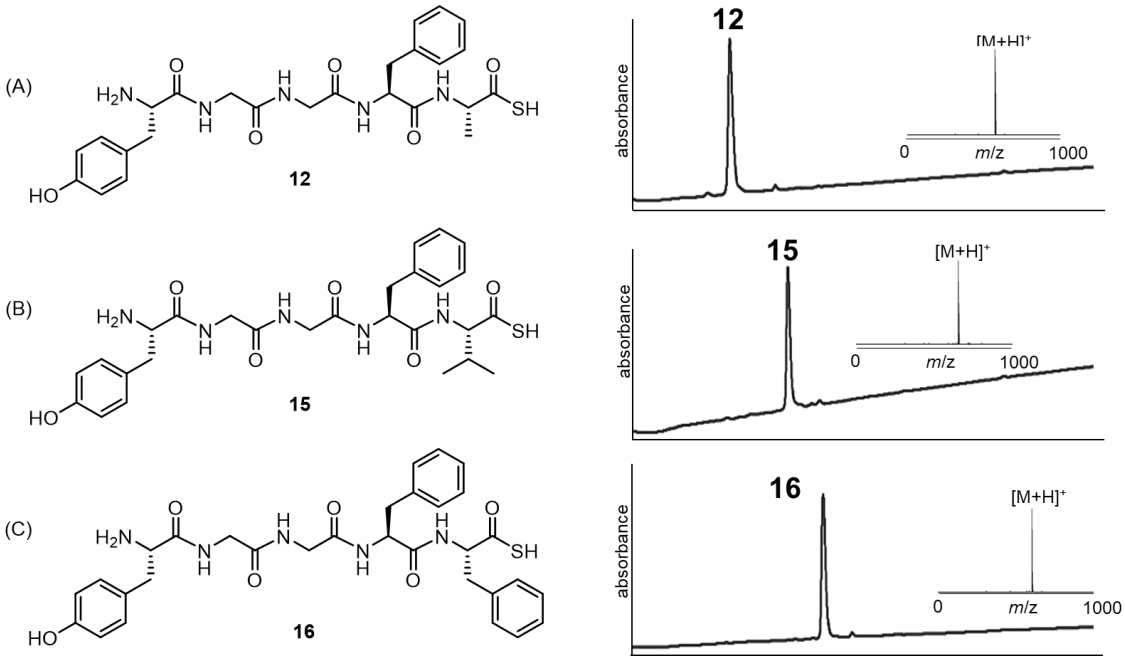


Figure E1. Synthesis of peptide thioacids **12**, **15** and **16**. HPLC profiles of the purified peptide thioacids and mass data.

Model peptide thioacid (38l, 38d)

For the monitoring of epimerization under the potential oxazolone formation, preparation of Fmoc-Tyr-Ala thioacids **38l** and **38d** were performed in the same manner to the preparation of model peptide thioacids **12**. Purification of the resultant product was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 70 : 30 to 30 : 70 over 40 min at the flow rate of 2.5 mL/min) afforded Fmoc-Tyr-*(D,L)*-Ala-COSH **38l** and **38d**.

Fmoc-Tyr-*L*-Ala 38l

HRMS(ESI): *m*/z calcd. For $C_{27}H_{27}N_2O_5S$: $[M+H]^+$ 491.1641, found for 491.1642. (Figure E15 A)

Fmoc-Tyr-*D*-Ala 38d

HRMS(ESI): *m*/z calcd. For $C_{27}H_{27}N_2O_5S$: $[M+H]^+$ 491.1641, found for 491.1628. (Figure E15 A)

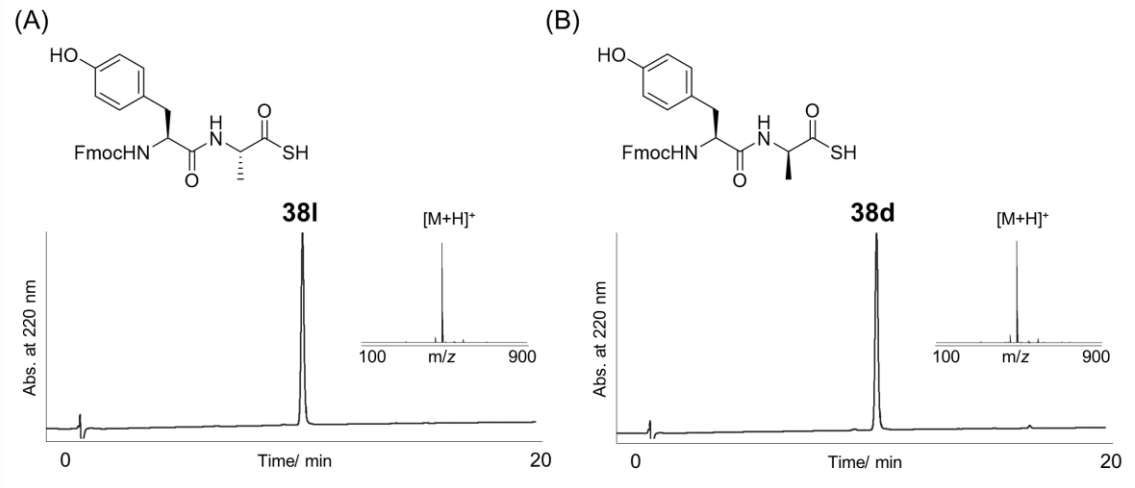


Figure E15. LC and MS of **38l** and **38d**

Model peptide thioacid (**82**)

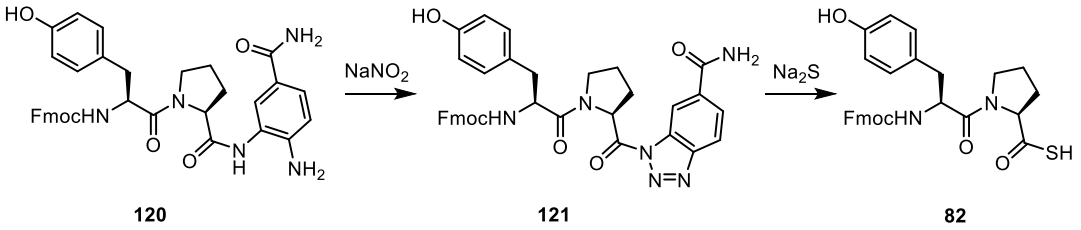
The model peptide **82** was synthesized on Dawson AM resin (150 μmol scale) by Fmoc SPPS (Scheme E2). After swelling a resin by DCM and DMF for 1 h, to a resin (150.0 μmol, 1.0 equiv) was added 20% piperidine / DMF (5.0 mL) for the removal of the Fmoc groups and stirred for 7 min two times. All reagents were filtered, and the resultant resin was washed with DCM and DMF. Then Fmoc-L-Pro-OH (303 mg, 0.90 mmol) was pre-activated with HATU (342 mg, 0.90 mmol) and DIPEA, (235 μL, 0.15 mmol) in anhydrous DMF (3.0 mL) for 1 min and then the activated amino acid was added to a resin. The suspension was gently agitated for 1 h at room temperature. After the coupling of a first amino acid, the resin was washed with DCM and DMF and then Fmoc group was deprotected with 20% piperidine/DMF for 7 min twice. Then Fmoc-L-Tyr(O^{t}Bu)-OH (413 mg, 0.90 mmol) was pre-activated with HATU (342 mg, 0.90 mmol) and DIPEA, (235 μL, 0.15 mmol) in anhydrous DMF (3.0 mL) for 1 min and then the activated amino acid was added to a resin. The suspension was gently agitated for 30 min at room temperature followed by washing with DCM. After all coupling steps were complete, the fully protected dipeptide was cleaved with a cocktail containing TFA/H2O/TIPS (95:2.5:2.5, 5mL), then the reaction mixture was allowed at room temperature and kept for 1 hour. After cleavage, the resultant solution was filtered and evaporated *in vacuo* to give a colorless oil. To this suspension was added CH_3CN-H_2O (1.0 mL-1.0 mL) and then lyophilized to afford dipeptide *o*-aminoanilide **120** (40 mg, 42%) as a white powder.

MS (ESI): calcd for $C_{36}H_{36}N_5O_6$ $[M+H]^+$ 634.27, found for 634.27.

To peptide *o*-aminoanilides **120** (40 mg, 42%) was added 0.2 M sodium phosphate buffer (pH 3.0) containing 6 M Gdn-HCl (6.0 mL) and CH_3CN (1.0 mL) and the reaction mixture was cooled at -10 ˚C. To the resultant suspension was added 1.2 M $NaNO_2$ aqueous solution (243 μL, 3.0 eq) and

subsequently the reaction was stirred for 5 min at -10 °C to afford peptide benzotriazole intermediates **121**, followed by the addition of 0.2 M sodium phosphate buffer (pH 7.8) containing 6 M Gdn-HCl and 0.2 M Na_2S (10.0 mL, 20.0 eq). Then a white suspension changed to a yellow transparent solution. After stirring for 30 min at room temperature, the product was purified by preparative HPLC (Capcell Pak C18 Φ10 × 250 mm, 0.1% TFA aqueous solution: 0.1% TFA CH_3CN = 70: 30 to 30: 70 over 90 min at the flow rate of 3 mL/min) to give compound **82** (12 mg, 15%) as a white powder.
MS (ESI): calcd for $C_{29}H_{29}N_2O_5S$ $[M+H]^+$ 517.18, found for 517.18. (Figure E2)

Scheme E2. The synthesis of dipeptide thioacid 82.



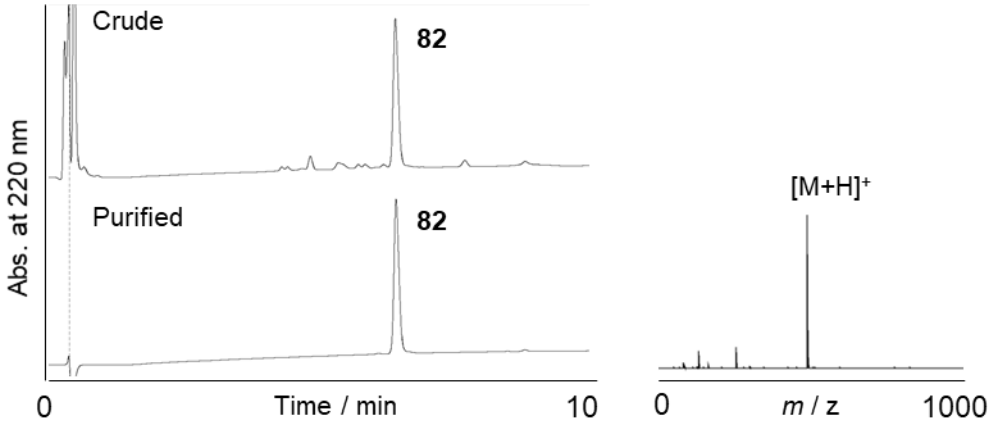
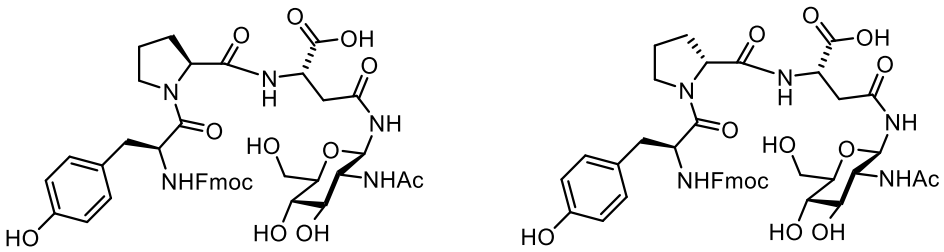


Figure E2. LC and MS spectrum of dipeptide thioacid **82**.

Scheme E3. Synthesis of model glycotripeptide (S)-85 and (R)-85



(S)-85 (*L-form*) **(R)-85** (*D form*)

Model glycopeptide ((S)-85, (R)-85))

The model glycopeptides **(S)-85** and **(R)-85** (Scheme E3) were synthesized on trityl chloride resin (60 μmol scale) by Fmoc solid-phase peptide synthesis (SPPS) respectively. To a resin (60.0 μmol) was added N-[N-Fmoc-L-aspart-4-oyl]-β-D-glucopyranosylamine **4** (45 mg, 80 μmol) and DIPEA (48 μL, 280 μmol) in DMF (1.0 mL) and stirred for 3 h. All reagents were filtered, and the resultant resin was washed with DCM and DMF, then the Fmoc group was deprotected with 20% piperidine/DMF (3 mL) for 7 min twice. Then Fmoc-*L*-Pro-OH (101 mg, 300 μmol) was pre-activated with DIC (47 μL, 300 μmol) and HOBt (40 mg, 300 μmol) in anhydrous DMF (3.0 mL) for 1 min and then the activated amino acid was added to a resin. The suspension was gently agitated for 1 h at room temperature. After the coupling of an amino acid, the resin was washed with DCM and DMF and then Fmoc group was deprotected with 20% piperidine/DMF for 7 min twice. Then Fmoc-*L*-Tyr(O^{t}Bu)-OH (138 mg, 300 μmol) was pre-activated with DIC (47 μL, 300 μmol) and HOBt (40 mg, 300 μmol) in anhydrous DMF (3.0 mL) for 1 min and then the activated amino acid was added to a resin. The suspension was gently agitated for 1 h at room temperature followed by washing with DCM. After all coupling steps were complete, a fully protected peptide was cleaved with a cocktail containing TFA/H2O/TIPS (95:2.5:2.5, 5 mL), then the reaction mixture was allowed at room temperature and kept for 1 hour. After cleavage, the resultant solution was filtered and evaporated *in vacuo* to give a colorless oil. The reaction mixture was purified by preparative HPLC (Capcell Pak C18 Φ10 × 250 mm, 0.1% TFA aqueous solution: 0.1% TFA CH_3CN = 75: 25 to 50: 50 over 60 min at the flow rate of 3 mL/min) to give compound **(S)-85** (6.7 mg, 14%) as a white powder. Synthesis of compound **(R)-85** was performed by the same manner with the synthesis of **(S)-85**, using Fmoc-*D*-Pro-OH instead of Fmoc-*L*-Pro-OH and afforded compound **(R)-85** (12.1 mg, 25%) as a white powder.

(S)-85

^{1}H NMR (400 MHz, DMSO-D6 + 5%D_2O) δ 8.28 (d, 1H, J = 8.7 Hz, GlcNAc-NH-1), 7.98 (d, 1H, J = 8.2 Hz, Asn-α-amide), 7.88 (d, 2H, J = 7.4 Hz, Ar-H Fmoc), 7.85 (d, 1H, GlcNAc-NH-2), 7.69 (d, 1H, Tyr-α-amide), 7.64 (d, 2H, J = 7.2 Hz, Ar-H Fmoc), 7.42 (dd, 2H, J = 7.4, Hz, Ar-H Fmoc), 7.32

(dd, 2H, J = 7.4 Hz, Ar-H Fmoc), 7.14 (d, 2H, J = 8.5 Hz, Ar-H Tyr), 6.66 (d, 2H, J = 8.4 Hz, Ar-H Tyr), 4.83 (d, 1H, J = 9.4 Hz, H-1), 4.53 (dt, 1H, J = 5.6 Hz, αH-Asn), 4.41 (dt, 1H, J = 3.4 Hz, αH-Pro), 4.32 (dt, 1H, αH-Tyr), 4.16 (dd, 1H, CH-Fmoc), 4.15 (d, 2H, CH_2-Fmoc), 3.65 (dd, 1H, J = 11.7 Hz, H-6a), 3.54 (m, 2H, βH-Pro), 3.54 (dd, 1H, H-2), 3.45 (dd, 1H, H-6b), 3.34 (dd, 1H, J = 7.5 Hz, H-3), 3.11 (dd, 1H, H-5), 3.10 (dd, 1H, J = 10.1 Hz, H-4), 2.88 (dd, 1H, βaH-Tyr), 2.69 (dd, 1H, βbH-Tyr), 2.58 (dd, 1H, βaH-Asn), 2.51 (dd, 1H, βbH-Asn), 2.01 (m, 1H, δaH-Pro), 1.90 (m, 2H, γH-Pro), 1.89 (m, 1H, δbH-Pro), 1.83 (s, 3H, acetyl) ppm; ^{13}C NMR (100 MHz, DMSO-D6 + 5%D_2O) δ: 172.8 (Asn-αCO), 171.7 (Tyr-αCO), 171.0 (Pro-αCO, GlcNAc-2-NHCO), 170.1 (Asn-βCO), 156.3 (Fmoc-OCO), 156.1 (Tyr-Ar), 144.2 (Fmoc-Ar), 141.1 (Fmoc-Ar), 130.7 (Tyr-Ar), 128.7 (Tyr-Ar), 128.1 (Fmoc-Ar), 127.6 (Fmoc-Ar), 125.8 (Fmoc-Ar), 120.5 (Fmoc-Ar), 115.3 (Tyr-Ar), 79.3 (C-1), 79.2 (C-5), 74.8 (C-3), 70.6 (C-4), 66.2 (Fmoc-CH_2), 61.2 (C-6), 59.7 (αC-Pro), 55.0 (αC-Tyr, C-2), 48.6 (αC-Asn), 47.0 (Fmoc-CH, βC-Pro), 37.4 (βC-Asn), 36.0 (βC-Tyr), 29.4 (δC-Pro), 24.7 (γC-Pro), 23.3 (acetyl) ppm. MS (ESI): calcd for $C_{41}H_{48}N_5O_{13}$ $[M+H]^+$ 818.32, found for 818.32. (Figure E3a)

(R)-85

^{1}H NMR (400 MHz, DMSO-D6 + 5%D_2O) δ 7.87 (d, 2H, J = 7.4 Hz, Ar-H Fmoc), 7.70 (d, 2H, J = 7.2 Hz, Ar-H Fmoc), 7.58 (dd, 2H, J = 7.4, Hz, Ar-H Fmoc), 7.36 (m, 2H, Ar-H Fmoc), 7.09 (d, 2H, J = 8.5 Hz, Ar-H Tyr), 6.63 (d, 2H, J = 8.4 Hz, Ar-H Tyr), 4.80 (d, 1H, J = 9.4 Hz, H-1), 4.68 (dd, 1H, J = 5.4 Hz), 4.50 (dd, 1H, J = 5.4 Hz), 4.34 (dt, 1H, J = 3.4 Hz), 4.24 (m, 3H), 4.16 (dd, 1H, CH-Fmoc), 4.15 (d, 2H, CH_2-Fmoc), 3.65 (dd, 1H, J = 11.7 Hz, H-6a), 3.54 (m, 2H, βH-Pro), 3.54 (dd, 1H, H-2), 3.45 (dd, 1H, H-6b), 3.34 (dd, 1H, J = 7.5 Hz, H-3), 3.11 (dd, 1H, H-5), 3.10 (dd, 1H, J = 10.1 Hz, H-4), 2.88-2.31 (m, βH-Asn, Pro, Tyr), 2.01 (m, 1H, δaH-Pro), 1.90 (m, 2H, γH-Pro), 1.89 (m, 1H, δbH-Pro), 178(s, 3H, acetyl) ppm; ^{13}C NMR (100 MHz, DMSO-D6 + 5%D_2O) δ: 172.8 (Asn-αCO), 171.7 (Tyr-αCO), 171.0 (Pro-αCO, GlcNAc-2-NHCO), 170.1 (Asn-βCO), 156.3 (Fmoc-OCO), 156.1 (Tyr-Ar), 144.2 (Fmoc-Ar), 141.1 (Fmoc-Ar), 130.7 (Tyr-Ar), 128.7 (Tyr-Ar), 128.1 (Fmoc-Ar), 127.6 (Fmoc-Ar), 125.8 (Fmoc-Ar), 120.5 (Fmoc-Ar), 115.3 (Tyr-Ar), 79.3 (C-1), 79.2 (C-5), 74.8 (C-3), 70.6 (C-4), 66.2 (Fmoc-CH_2), 61.2 (C-6), 59.7 (αC-Pro), 55.0 (αC-Tyr, C-2), 48.6 (αC-Asn), 47.0 (Fmoc-CH, βC-Pro), 37.4 (βC-Asn), 36.0 (βC-Tyr), 29.4 (δC-Pro), 24.7 (γC-Pro), 23.3 (acetyl) ppm. MS (ESI): calcd for $C_{41}H_{48}N_5O_{13}$ $[M+H]^+$ 818.32, found for 818.32. (Figure E3b)

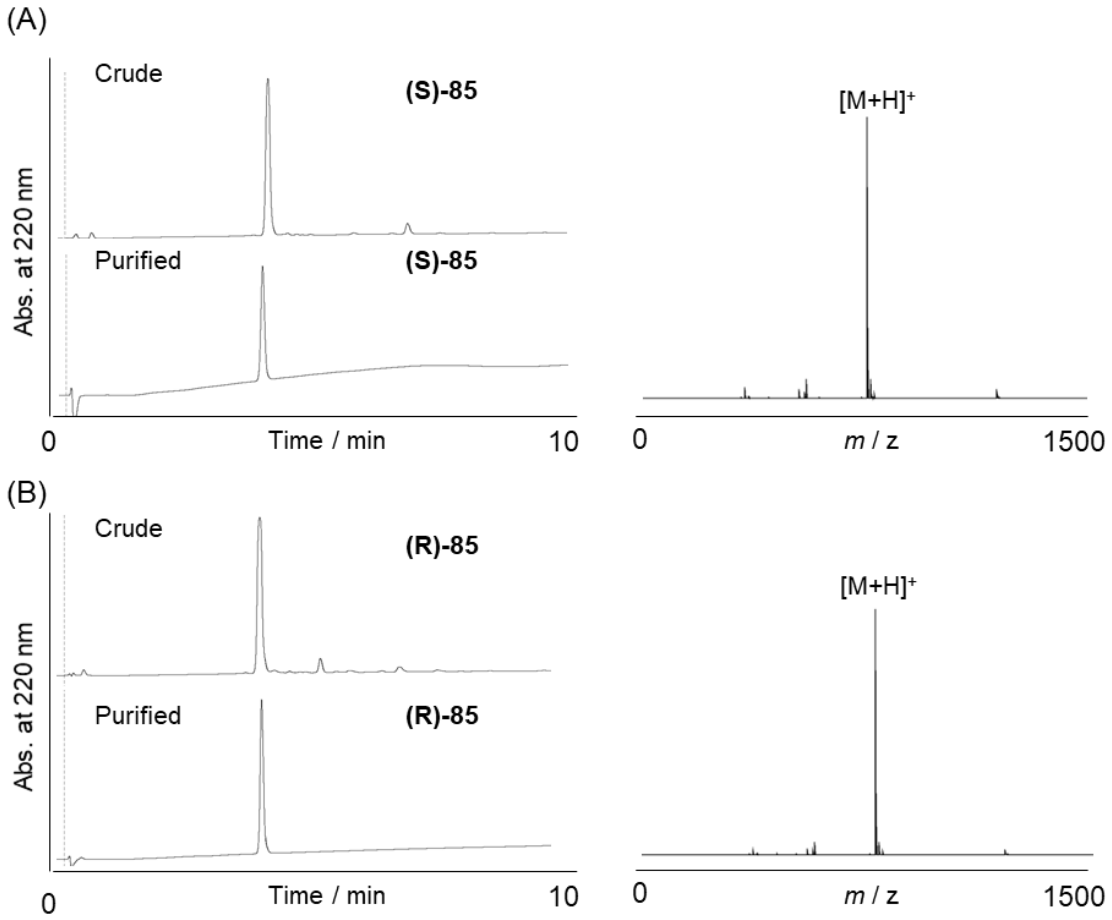


Figure E3. (A) LC and MS spectrum of model glycopeptide **(S)-85**, (B) LC and MS spectrum of model glycopeptide **(R)-85**

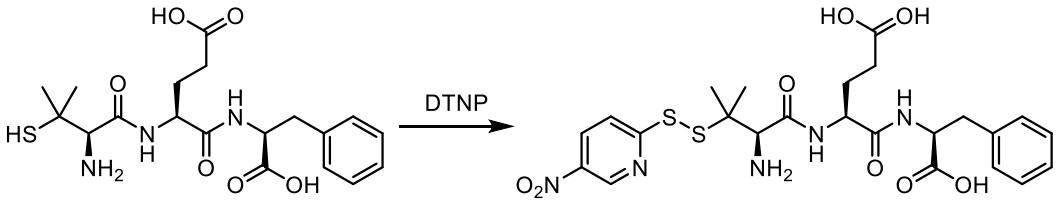
Model peptide having β-mercapto valine (86)

The model glycopeptides **86** were synthesized on trityl chloride resin (100 μmol scale) by our improved Fmoc solid-phase peptide synthesis (SPPS). To a resin (100 μmol) was added Fmoc-L-Phe-OH (78 mg, 200 μmol) and DIPEA (80 μL, 500 μmol) in DCM (1.0 mL) and stirred for 2 h. All reagents were filtered, and the resultant resin was washed with DCM and DMF, then the Fmoc group was deprotected with 20% piperidine/DMF (5 mL) for 7 min twice. Then Fmoc-*L*-Glu(O^{t}Bu)-OH (170 mg, 400 μmol) was pre-activated with anhydrous DMF solution (2.0 mL) containing 0.2M HBTU and 0.2 M HOBt and DIPEA (110 μL) for 1 min and then the activated amino acid was added to a resin. The suspension was gently agitated for 45 min at room temperature. After the coupling of an amino acid, the resin was washed with DCM and DMF and then the Fmoc group was deprotected with 20% piperidine/DMF for 7 min twice. Then Fmoc-*L*-Pen(Trt)-OH (250 mg, 400 μmol) was pre-activated with anhydrous DMF solution (2.0 mL) containing 0.2M HBTU and 0.2 M HOBt and DIPEA (110 μL) for 1 min and then the activated amino acid was added to a resin. The suspension was gently

agitated for 45 min at room temperature. After the coupling of an amino acid, the resin was washed with DCM and DMF and then the Fmoc group was deprotected with 20% piperidine/DMF for 7 min twice followed by washing with DCM. After all coupling steps were complete, a fully protected tripeptide was cleaved with a cocktail containing TFA/H2O/TIPS (95:2.5:2.5, 10 mL), then the reaction mixture was allowed at room temperature and kept for 1 hour. After cleavage, the resultant solution was filtered and evaporated *in vacuo* to give a colorless oil. To the reaction mixture was added chilled Et_2O (10 mL) and yield tripeptide **122** as a white precipitate was washed by Et_2O twice and dried *in vacuo*. The precipitate was dissolved in CH_3CN-H_2O (1.0 mL-4.0 mL) containing 0.1% TFA and to the solution of crude **122** was added 2,2'-dithiobis(5-nitropyridine) (620 mg) and stirred under argon atmosphere for 12 h at room temperature. During the reaction a colorless solution changed to a yellow. The reaction mixture was purified by preparative HPLC (Proteo C4 Φ10 × 250 mm, 0.1% TFA aqueous solution: 0.1% TFA CH_3CN = 80: 20 to 50: 50 over 60 min at the flow rate of 2.5 mL/min) to give compound **86** (39 mg, 67%) as a white powder.

MS (ESI): calcd for $C_{24}H_{30}N_5O_8S_2$ $[M+H]^+$ 580.15, found for 580.15. (Figure E4)

Scheme E4. Synthesis of model peptide having β-mercapto valine 86



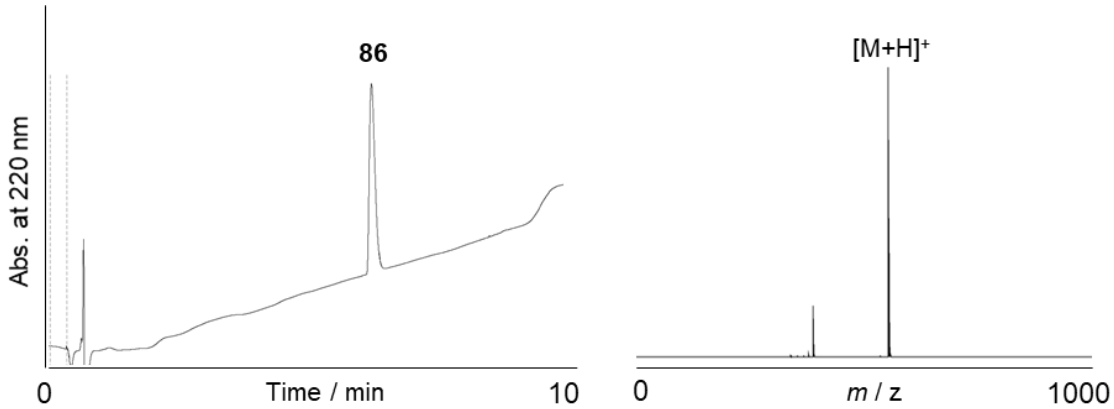


Figure E4. LC and MS spectrum of model peptide **86**

3. Procedures for DDC and TCL with model peptide (Chapter 2 and 5)

Reaction conditions of DDC with leucine thioacid 13l (*Chapter 2*)

To a solution of peptide thioacid **12**, **15** and **16** (15.0 mM) in solution was added leucine thioacid **13l** (45.0 mM, 3.0 equiv.) at 23 ºC and the resultant mixture was stirred for 9 hours. The reaction progress was monitored by RP-HPLC. All reactions generally proceeded to the completion within this time range. After the reaction, D,L-dithiothreitol (DTT, 50.0 equiv.) was added to the reaction to quench the reaction. The yellow color solution became colorless and then the amount of products was estimated from the area of each UV-peak at 220 nm in the LC chromatograms (Figure E5-E7).

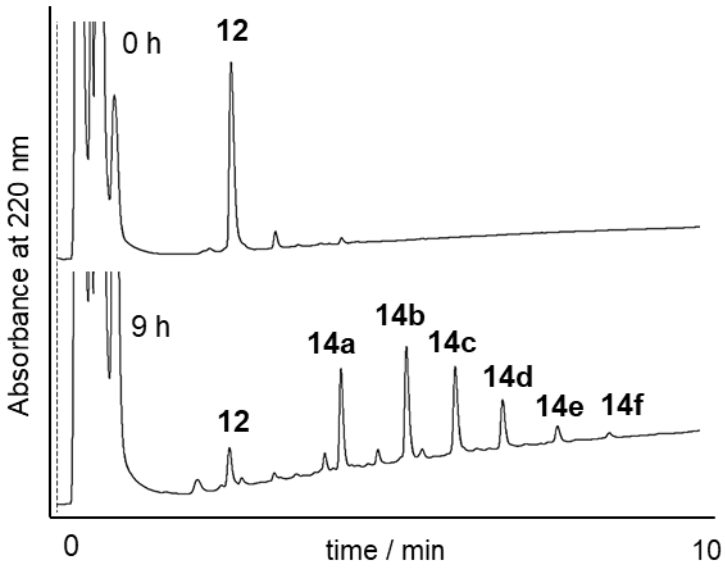


Figure E5. LC chromatogram for diacyl disulfide coupling between Tyr-Gly- Glu-Phe-*L*-Ala-COSH **12** and H_2N-*L*-Leu-COSH **13l**.

HRMS(ESI): calcd. For $C_{28}H_{36}N_5O_8S$ (Tyr-Gly-Glu-Phe-*L*-Ala-COSH) $[M+H]^+$ 602.2206, found, 602.2264, calcd. For $C_{34}H_{47}N_6O_9S$ (Tyr-Gly-Glu-Phe-*L*-Ala-*L*-Leu-COSH) $[M+H]^+$ 715.3047, found, 715.3089, calcd for $C_{40}H_{58}N_7O_{10}S$ (Tyr-Gly-Glu-Phe-*L*-Ala-(*L*-Leu)$_2$-COSH) $[M+H]^+$ 828.3888, found, 828.3932, calcd. For $C_{46}H_{69}N_8O_{11}S$ (Tyr-Gly-Glu-Phe-*L*-Ala-(*L*-Leu)$_3$-COSH) $[M+H]^+$ 941.4728, found, 941.4767, calcd. For $C_{52}H_{80}N_9O_{12}S$ (Tyr-Gly-Glu-Phe-*L*-Ala-(*L*-Leu)$_4$-COSH) $[M+H]^+$ 1054.5569, found, 1054.5608. calcd. For $C_{58}H_{91}N_{10}O_{13}S$ (Tyr-Gly-Glu-Phe-*L*-Ala-*L*-Leu-(*L*-Leu)$_5$-COSH) $[M+H]^+$ 1167.6410, found, 1167.6465, calcd. For $C_{64}H_{102}N_{11}O_{14}S$ (Tyr-Gly-Glu-Phe-*L*-Ala-*L*-Leu-(*L*-Leu)$_6$-COSH) $[M+H]^+$ 1280.7250, found, 1280.7294, calcd. For $C_{70}H_{113}N_{12}O_{15}S$ (Tyr-Gly-Glu-Phe-*L*-Ala-*L*-Leu-(*L*-Leu)$_7$-COSH) $[M+H]^+$ 1393.8091, found, 1393.8135. calcd. For $C_{28}H_{36}N_5O_9$ (Tyr-Gly-Glu-Phe-*L*-Ala-COOH) $[M+H]^+$ 586.2435, found, 586.2519, calcd. For

$C_{34}H_{47}N_6O_{10}$ (Tyr-Gly-Glu-Phe-*L*-Ala-*L*-Leu-COOH) $[M+H]^+$ 699.3275, found 699.3348, calcd. For $C_{40}H_{58}N_7O_{11}$ (Tyr-Gly-Glu-Phe-*L*-Ala-(*L*-Leu)$_2$-COOH) $[M+H]^+$ 812.4116, found, 812.4188, calcd. For $C_{46}H_{69}N_8O_{12}$ (Tyr-Gly-Glu-Phe-*L*-Ala-(*L*-Leu)$_3$-COOH) $[M+H]^+$ 925.4957, found 925.5017.

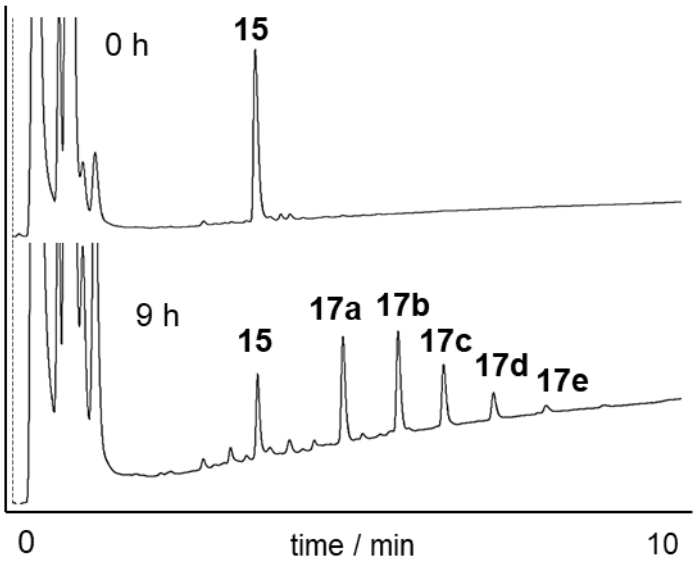


Figure E6. LC chromatogram for diacyl disulfide coupling between Tyr-Gly- Glu-Phe-*L*-Val-COSH **15** and H_2N-*L*-Leu-COSH **13l**.

HRMS(ESI): calcd. For $C_{30}H_{40}N_5O_8S$ (Tyr-Gly-Glu-Phe-*L*-Val-COSH) $[M+H]^+$ 630.2519, found, 630.2576, calcd. For $C_{36}H_{51}N_6O_9S$ (Tyr-Gly-Glu-Phe-*L*-Val-*L*-Leu-COSH) $[M+H]^+$ 743.3360, found, 743.3399, calcd for $C_{42}H_{62}N_7O_{10}S$ (Tyr-Gly-Glu-Phe-*L*-Val-(*L*-Leu)$_2$-COSH) $[M+H]^+$ 856.4201, found, 856.4239, calcd. For $C_{48}H_{73}N_8O_{11}S$ (Tyr-Gly-Glu-Phe-*L*-Val-(*L*-Leu)$_3$-COSH) $[M+H]^+$ 969.5041, found, 969.5080, calcd. For $C_{54}H_{84}N_9O_{12}S$ (Tyr-Gly-Glu-Phe-*L*-Val-(*L*-Leu)$_4$-COSH) $[M+H]^+$ 1082.5882, found, 1082.5925. calcd, for $C_{60}H_{95}N_{10}O_{13}S$ (Tyr-Gly-Glu-Phe-*L*-Val-(*L*-Leu)$_5$-COSH) $[M+H]^+$ 1195.6723, found, 1195.6765. calcd. For $C_{30}H_{40}N_5O_9$ (Tyr-Gly-Glu-Phe-*L*-Val-COOH) $[M+H]^+$ 614.2748, found, 614.2804, calcd. For $C_{36}H_{51}N_6O_{10}$ (Tyr-Gly-Glu-Phe-*L*-Val-*L*-Leu-COOH) $[M+H]^+$ 727.3588, found 727.3633.

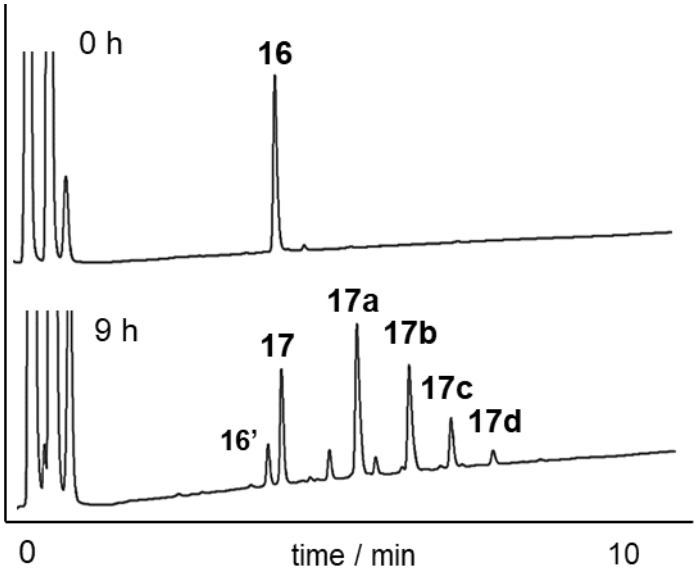


Figure E7. LC chromatogram for diacyl disulfide coupling between Tyr-Gly- Glu-Phe-*L*-Phe-COSH 1**6** and H_2N-*L*-Leu-COSH **13l**.

HRMS(ESI): calcd. For $C_{34}H_{40}N_5O_8S$ (Tyr-Gly-Glu-Phe-*L*-Phe-COSH) $[M+H]^+$ 678.2519, found, 678.2573, calcd. For $C_{40}H_{51}N_6O_9S$ (Tyr-Gly-Glu-Phe-*L*-Phe-*L*-Leu-COSH) $[M+H]^+$ 791.3360, found, 791.3415, calcd for $C_{46}H_{62}N_7O_{10}S$ (Tyr-Gly-Glu-Phe-*L*-Phe-(*L*-Leu)$_2$-COSH) $[M+H]^+$ 904.4201, found, 904.4263, calcd. For $C_{52}H_{73}N_8O_{11}S$ (Tyr-Gly-Glu-Phe-*L*-Phe-(*L*-Leu)$_3$-COSH) $[M+H]^+$ 1017.5041, found, 1017.5093, calcd. For $C_{58}H_{84}N_9O_{12}S$ (Tyr-Gly-Glu-Phe-*L*-Phe-(*L*-Leu)$_4$-COSH) $[M+H]^+$ 1130.5882, found, 1130.5950. calcd. For $C_{34}H_{40}N_5O_9$ (Tyr-Gly-Glu-Phe-*L*-Phe-COOH) $[M+H]^+$ 662.2748, found, 662.2814, calcd. For $C_{40}H_{51}N_6O_{10}$ (Tyr-Gly-Glu-Phe-*L*-Phe-*L*-Leu-COOH) $[M+H]^+$ 775.3588, found 775.3637, calcd. For $C_{46}H_{62}N_7O_{11}$ (Tyr-Gly-Glu-Phe-*L*-Phe-(*L*-Leu)$_2$-COOH) $[M+H]^+$ 888.4429, found, 888.4484.

Reaction conditions of DDC with glycosylated asparagine thioacid 11 (*Chapter 2*)

To a solution of peptide thioacid **4** and **5** (15.0 mM, 1.0 equiv.) in anhydrous DMSO was added glycosyl asparagine thioacid **3** (30.0 mM) at 23 ºC and the resultant mixture was stirred for 9 hours. The reaction progress was monitored by RP-HPLC. All reactions generally proceeded to the completion within this time range. After completion of the reaction, DTT (50.0 equiv.) was added to the mixture for quenching the reaction. The yellow color of the solution became colorless, and the mixture afforded the target glycopeptides. All data obtained are shown in Figure 2 of main text (Figure 2-4). Glycopeptide **19** was confirmed by HRMS (ESI): calcd for $C_{132}H_{190}N_{13}O_{73}S$ $[M+H]^+$ 3157.1276, found for 3157.1210. Glycopeptide **20** was confirmed by HRMS (ESI): calcd for $C_{134}H_{194}N_{13}O_{73}S$ $[M+H]^+$ 3185.1589, found for 3185.1621.

Reaction conditions of TCL with peptide 86 (*Chapter 5*)

To a solution of glycosyl asparagine thioacid **11** (1.5 mM) in sodium phosphate buffer (pH 4.0, 5.5, 7.0) was added peptide having β-mercapto valine at N-terminus **86** (3.0 mM, 2.0 equiv.) at 23 ºC and the resultant mixture was stirred for 24 hours. The reaction progress was monitored by RP-HPLC. All reactions generally proceeded to the completion within this time range. After the reaction (50.0 equiv.) was added to the reaction to quench the reaction. The amount of products was estimated from the area of each UV-peak at 220 nm in the LC chromatograms (Figure 5-6). Glycopeptide **87** was confirmed by HRMS (ESI): calcd for $C_{123}H_{182}N_{11}O_{71}S$ $[M+H]^+$ 2981.0690, found for 2981.0688

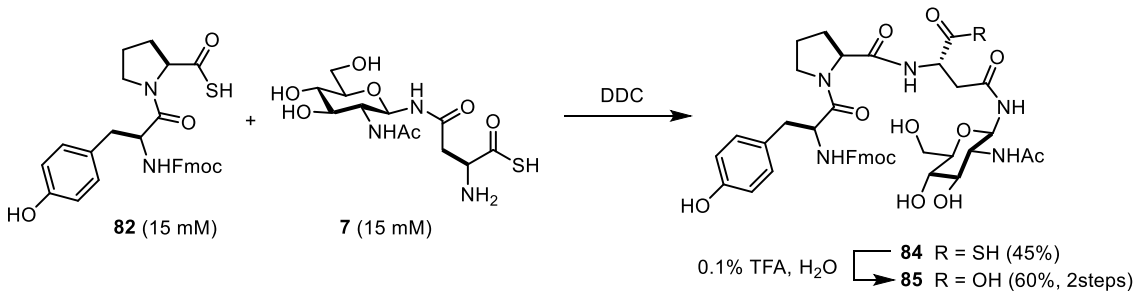
4. Monitoring the epimerization in DDC *(Chapter 2 and 5)*

Evaluation of the epimerization of DDC with leucine thioacids

To a solution of peptide thioacids **36l** having either the *L* form of phenylalanine (Tyr-Gly-Gly-Phe^{L}-SH) or **36d** having *D* form of phenylalanine at the C-terminus (Tyr-Gly-Gly-Phe^{D}-SH) with either *D*-Phe or *L*-Phe at the C-terminus (15.0 mM) in anhydrous DMSO was added either *L*-Leu thioacid **13l** or *D*-Leu thioacid **13d** (45.0 mM). The percentage of each product was compared on chromatography. The items **I**, **II**, **III**, and **IV** shown in the table indicate the combinations of peptide thioacids and leucine thioacids for DDC and the items (I)-(IV) show the HPLC profile of the reactions, respectively. Asterisks * in the HPLC profiles were not peptide derivatives. As shown in HPLC profiles (**I-IV**), no epimerization was observed for both the C-termini of peptide thioacid and leucine thioacid (Figure 2-11).

Evaluation of the epimerization of DDC with prolyl thioacids

Scheme E5. Synthesis of glycopeptide 85 by using DDC



To a solution of peptide thioacid **82** (30.0 mM) in DMSO was added glycosyl asparagine thioacid **7** having mono GlcNAc **7** (30.0 mM, 1.0 equiv.) at 23 ºC and the resultant mixture was stirred for 9 hours to give monoglycosylated peptide thioacid **84** with 45% convergent yield (Figure 5-4). Glycopeptide **84** was confirmed by HRMS (ESI): $C_{41}H_{48}N_5O_{12}S$ $[M+H]^+$ 834.3020, found for 834.3580. The reaction progress was monitored by RP-HPLC after quenching by excess of TCEP. The amount of products was estimated from the area of each UV-peak at 220 nm in the LC chromatograms. To the reaction mixture was added 0.1% TFA H_2O (excess amount) without reductants to afford glycopeptide **85** (Scheme E5) in 60% isolated yield (2 steps) (Figure E8 A). Glycopeptide **85** was confirmed by MS (ESI): calcd for $C_{41}H_{48}N_5O^{13}$ $[M+H]^+$ 818.3249, found for 818.3263.

Finally, the retention time of LC spectrum of glycopeptide **85** was compared with those of the authentic glycopeptide **(R)-85** and **(L)-85** synthesized by a SPPS protocol. **(R)-85** is consisted of N-Fmoc-*L*-Tyr-*D*-Pro-*L*-Asn(GlcNAc)-OH, **(S)-85** is consisted of N-Fmoc-*L*-Tyr-*L*-Pro-*L*-Asn(GlcNAc)-OH .

The LC analysis showed that the LC peaks of the glycopeptides **85** obtained by DDC had the same retention time as **(S)-85**, and the absence of other peaks at the retention time of **(R)-85** confirmed that epimerization did not proceed in DDC with prolyl thioacid **82** (Figure E8 B).

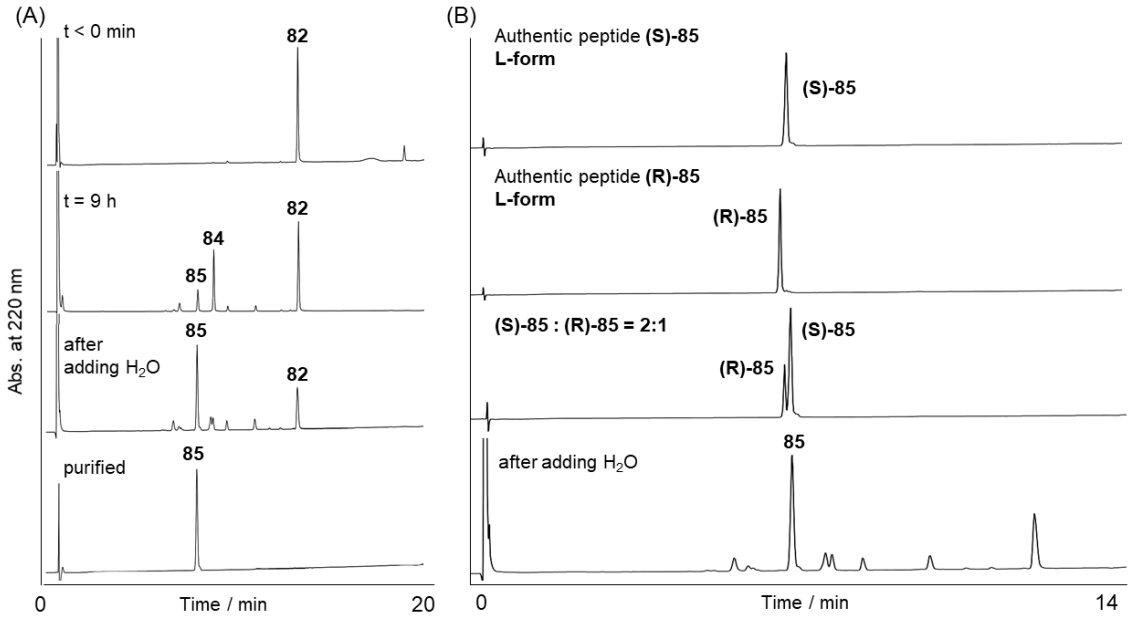


Figure E8. DDC for prolyl peptide thioacid. (A) HPLC monitoring of DDC with prolyl peptide thioacids **82** and glycosyl asparagine thioacids **7** having mono GlcNAc. (B) LC profiles of **(R)-85** synthesized by SPPS, **(S)-85** synthesized by SPPS, **85** synthesized by DDC.

5. Synthesis of glycotripeptide 31 and NMR analyses (Chapter 2)

A conventional strategy

To a solution of Fmoc-*L*-Asn(glycan)-OH **32** (5.3 mg, 1.9 μmol)[1] in dry DMF (400 μL) was added cysteine derivative **26** (2.1 mg, 7.7 μmol), PyBOP (3.6 mg, 6.9 μmol), and DIPEA (1.1 μL, 6.3 μmol) at -20 ºC under argon atmosphere. The resulting mixture was stirred and warmed up to 0 ºC. The reaction was monitored by RP-HPLC. After the completion of the reaction, the mixture was added chilled Et_2O (1.0 mL) to afford white precipitation. The suspension was centrifuged, and Et_2O was removed by decantation to afford white precipitate of Fmoc-*L*-Asn(glycan)-*L*-Cys(SEt)-OBn **33**. The separated precipitate was then washed with chilled Et_2O twice as the same manner. HRMS(ESI): *m*/z calcd. for $C_{131}H_{182}N_9O_{69}S_2$ $[M+H]^+$ 3049.0372, found for 3049.0297.

To the crude **33** was added DMF cocktail (430 μL) containing 1-methylpyroridine (15.0 μL), hexamethyleneimine (0.4 μL) and HOBt (0.4 mg) and stirred for 1 hour. The resulting mixture was added over 1.0 mL of chilled diethyl ether (Et_2O) to afford white precipitate. The suspension was centrifuged, and Et_2O was removed by decantation to afford H-*L*-Asn(glycan)-*L*-Cys(SEt)-OBn **34** as a white powder. This mixture was directly used for the next reaction without further purification. HRMS(ESI): *m*/z calculated for $C_{116}H_{172}N_9O_{69}S_2$ $[M+H]^+$ 2826.9692, found for 2826.9772 .

To a solution of the resultant **34** in dry DMF (400 μL) Fmoc-*L*-Ala-OH **21** (2.5 mg, 8.0 μmol) and PyBOP (3.7 mg, 7.1 μmol) were added diisopropylethylamine (DIPEA, 1.1 μL, 6.3 μmol) at -20 ºC under argon atmosphere. The resulting mixture was stirred and warmed up to 0 ºC. The reaction was monitored by RP-HPLC. The resulting mixture was purified by preparative HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 70 : 30 to 45 : 55 over 50 min at the flow rate of 1 mL/min, RT = 35 min) to afford Fmoc-*L*-Ala-*L*-Asn(glycan)-*L*-Cys(SEt)-OBn **35** (2.8 mg, 47%). HRMS (ESI): *m*/z calculated for $C_{134}H_{187}N_{10}O_{70}S_2$ $[M+H]^+$ 3120.0744, found for 3120.0630.

To a solution of the glycosyl tripeptide **35** (2.8 mg, 0.90 μmol) in CH_3CN/H_2O (0.5 mL and 3.5 mL respectively) containing 0.1% of trifluoroacetic acid was added TCEP (100 mg, 0.349 mmol) at room temperature. The resulting mixture was stirred for 1 hour. The reaction was monitored by RP-HPLC. The mixture was purified by preparative HPLC (Proteonavi Φ2.0 × 250 mm, 0.1 % TFA aq : 0.1 % TFA in CH_3CN = 70 : 30 to 45 : 55 over 50 min at the flow rate of 1 mL/min, RT = 27 min) to afford Fmoc-*L*-Ala-*L*-Asn(glycan)-*L*-Cys-OBn **31b** (2.5 mg, 91%). HRMS (ESI): *m*/z calculated for $C_{132}H_{183}N_{10}O_{70}S$ $[M+H]^+$ 3060.0788, found for 3060.0841.

Chemical glycan insertion strategy

To a solution of H-*L*-Asn(glycan)-SH **11** (1.8 mg, 0.70 μmol) in dry DMSO (42.0 μL) was added Fmoc-*L*-Ala-SH **23** (1.0 mg, 2.9 μmol) at room temperature. The resultant mixture was stirred for 6 hours to afford Fmoc-*L*-Ala- *L*-Asn(glycan)-SH **30**. DDC was monitored by RP-HPLC. HRMS(ESI): *m*/z calculated for $C_{122}H_{172}N_9O_{68}S$ $[M+H]^+$ 2882.9920, found for 2882.9833.

After completion of the DDC, to a resulting mixture of compound **30** was added a buffer solution (6.0 M Gdn-HCl in 0.2 M sodium phosphate, pH 5.7) containing the activated H-*L*-Cys(Npys)-OBn **29** (2.2 mg, 6.0 μmol) and then the mixture was stirred for 0.5 hours. To the mixture was added CH_3CN (150 μL) and stirred for additional 12 hours at room temperature. The reaction was monitored by RP-HPLC. The mixture was purified by preparative HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA aq : 0.1 % TFA in CH_3CN = 70 : 30 to 45 : 55 over 50 min at the flow rate of 1 mL/min, RT = 27 min) to afford Fmoc-*L*-Ala-*L*-Asn(glycan)-*L*-Cys-OBn **31a** (0.5 mg, 24% :DDC/TCL 2 steps yield). ^{1}H NMR (700 MHz, DMSO-D6) δ: 8.25 (d, Asn-NH), 8.25 (d, GlcNAc-NH-Asn), 8.05 (d, NeuAc-NH), 8.00 (d, 4H, Phenacyl), 7.95 (d, 1H, Cys-NH), 7.90 (d, Fmoc), 7.85 (d, 1H, GlcNAc-NH), 7.71 (m, Phenacyl, Fmoc), 7.67 (d, 2H, GlcNAc-NH),7.60 (t, Phenacyl, Fmoc), 7.59 (m, 1H, Ala-NH), 7.42-7.34 (each t, 5H, Benzyl), 5.72, 5.61 (each d, each 2H, PhenacylCH_2), 5.14 (s, 2H, BenzylCH_2), 4.92 (s, 1H, Man-H-1), 4.85 (d, 1H, *J* = 9.9 Hz, AsnGlcNAc-H-1), 4.75 (s, 1H, Man-H-1), 4.58 (m, 1H, Asn-αH), 4.56 (m, 1H, Cys-αH), 4.54 (b, 1H, Man-H-1), 4.45 (bd, 3H, GlcNAc-H-1), 4.25 (bd, 2H, Gal-H-1), 4.03 (m, 1H, Ala-αH), 3.73 (m, 2H, NeuAc-H-4), 2.84 (dd, Cys-βH), 2.54 (dd, Asn-βH), 2.58 (dd, 2H, NeuAc-H-3_{eq}) , 2.50 (dd, Cys-βH'), 1.90 (s, 6H, Acetyl), 1.87, 1.82, 1.81, 1.79 (each s, each 3H, Acetyl), 1.80 (m, 2H, NeuAc-H-3_{ax}), 1.22 (bd, 3H, Ala-βH). HRMS(ESI): *m*/z calc for $C_{132}H_{183}N_{10}O_{70}S$ $[M+H]^+$ 3060.0788 found for 3060.0841.

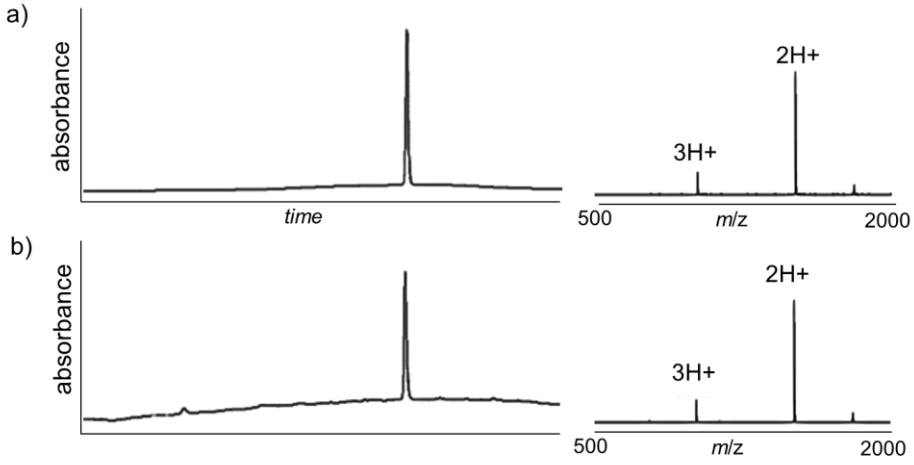
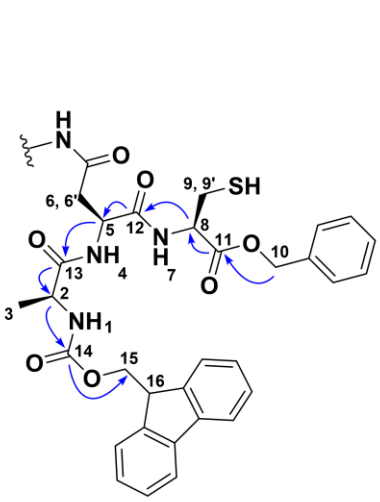
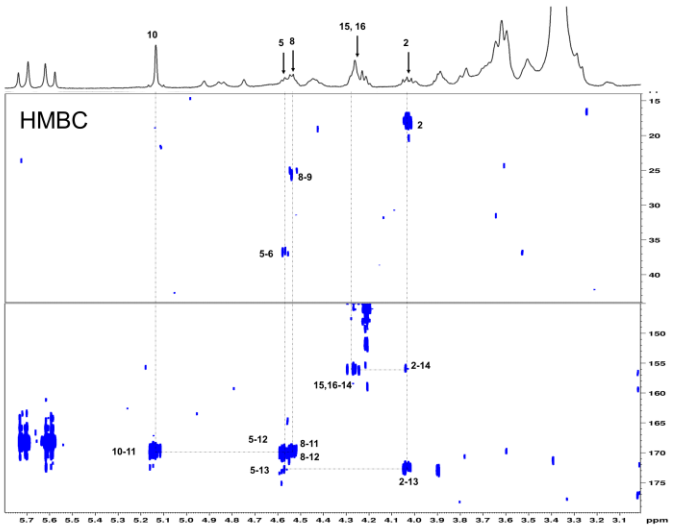
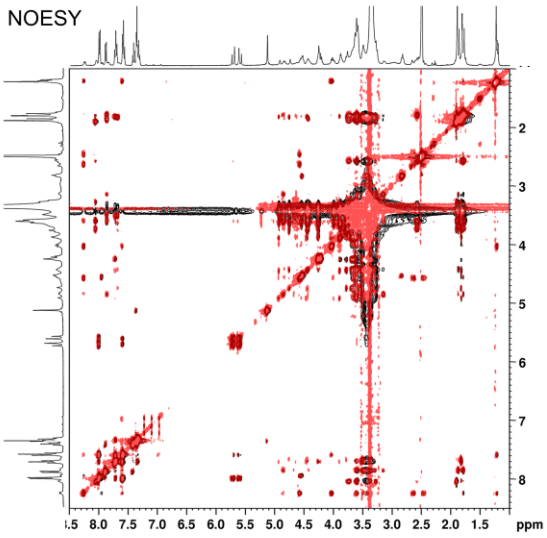
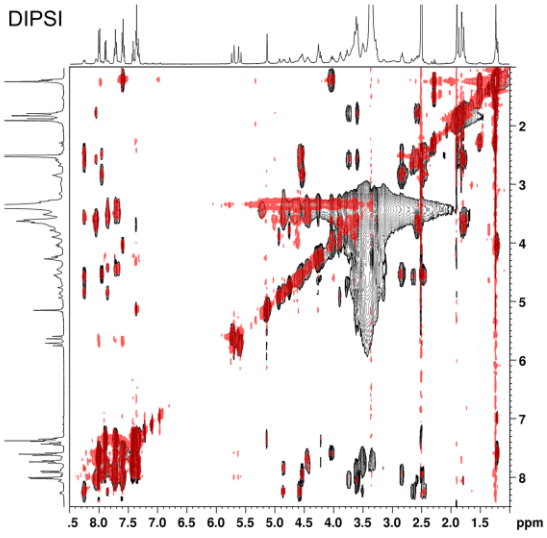


Figure E9. LC and mass spectra of glycopeptide prepared by two strategies. The profile (a) shows the LC of glycopeptide **31b** synthesized by a conventional condensation strategy. The profile (b) shows the LC of glycopeptide **31a** synthesized with DDC (Chemical insertion strategy).




HMBC
10
5
8
15, 16
2
8-9
5-6
2-14
15,16-14'
5-12
8-11
8-12
10-11
5-13
2-13
ppm


NOESY
ppm


DIPSI
ppm

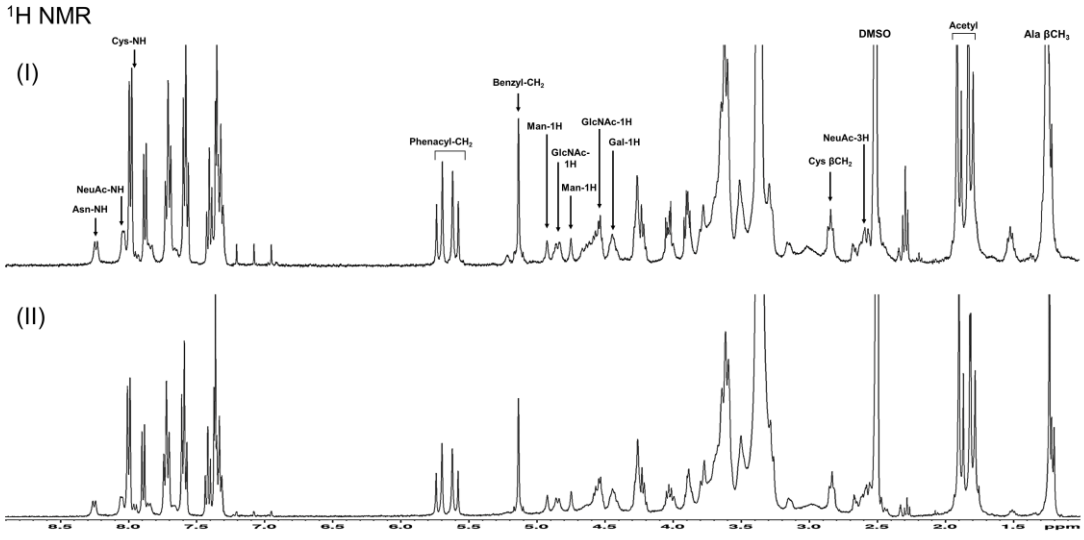

^{1}H NMR
(I)
Cys-NH
NeuAc-NH
Asn-NH
Phenacyl-CH$_2$
Benzyl-CH$_2$
Man-1H
GlcNAc-1H
GlcNAc-1H
Man-1H
Gal-1H
NeuAc-3H
Cys βCH$_2$
DMSO
Acetyl
Ala βCH$_3$
(II)
ppm

Figure E10. HMBC, NOESY, DIPSI and ^{1}H-NMR spectra of the glycopeptide **31** prepared by two strategies. HMBC spectrum shows the connectivity of long range ^{1}H-^{13}C coupling patterns. NOESY and DISPSI are superimposed spectra with glycopeptides synthesized by the two strategies (black contour plot: **31b** conventional strategy; red contour plot: **31a** chemical insertion strategy) The spectrum (I) shows ^{1}H-NMR spectrum of glycopeptide **31b** synthesized by a conventional condensation strategy and the spectrum (II) shows ^{1}H-NMR spectrum of glycopeptide **31a** synthesized with DDC (chemical insertion strategy). Since both NMR spectra are identical, both glycopeptides were confirmed to be identical.

6. Synthesis of glycosylated tetrapeptide 40 and NMR analyses (Chapter 2)

A conventional strategy

To compound Fmoc-*L*-Ala-*L*-Asn(glycan)-*L*-Cys(SEt)-OBn **35** (8.5 mg, 2.7 μmol) was added DMF cocktail (430 μL) containing 1-methylpyroridine (15.0 μL), hexamethyleneimine (0.40 μL) and HOBt (0.4 mg) and the mixture was stirred for 2.5 hour. To the resulting mixture was added the chilled diethyl ether (Et_2O, ca 10 mL) to afford white precipitate. The suspension was centrifuged, and Et_2O was removed by decantation to afford H-*L*-Ala-*L*-Asn(glycan)-*L*-Cys(SEt)-OBn **41** as a white powder. This mixture was directly used for the next reaction without further purification. HRMS(ESI): *m*/z calculated for $C_{119}H_{177}N_{10}O_{68}S_2$ $[M+H]^+$ 2898.0141, found for 2898.0114.

To the resultant precipitate **41** was added dry DMF (200 μL) containing Fmoc-*L*-Tyr(tBu)-OH (7.3 mg, 16 μmol), and then PyBOP (3.7 mg, 7.2 μmol.) was added. To this solution was then added diisopropylethylamine (DIPEA, 1.1 μL, 6.3 μmol) at -20 ºC under argon atmosphere. The reaction was monitored by RP-HPLC. The resulting Fmoc-*L*-Tyr(tBu)-*L*-Ala-*L*-Asn(glycan)-*L*-Cys(SEt)-OBn **42** was purified by preparative HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 60 : 40 to 35 : 65 over 60 min at the flow rate of 2.5 mL/min, RT = 31 min) to afford compound **42** (3.0 mg, 33%). HRMS (ESI): *m*/z calc for $C_{147}H_{204}N_{11}O_{72}S_2$ $[M+H]^+$ 3339.2081, found for 3339.2234.

For the deprotection of t-butyl group, to the solution of compound **42** (3.0 mg, 0.91 μmol) in TFA (1.0 mL) was added TIPS (50 μL) at room temperature and the resultant mixture was left for 30 min. After the completion of reaction, the mixture was concentrated in *vacuo* to obtain white powder of Fmoc-*L*-Tyr-*L*-Ala-*L*-Asn(glycan)-*L*-Cys(SEt)-OBn **43**. This mixture was directly used for the next reaction without further purification. HRMS (ESI): *m*/z calc for $C_{143}H_{196}N_{11}O_{72}S_2$ $[M+H]^+$ 3283.1455, found for 3283.1443.

To a solution of the resultant **43** in CH_3CN/H_2O (1.0 mL and 1.0 mL respectively) was added TCEP (0.12 g, 0.42 mmol) at room temperature. The resulting mixture was stirred for 1 hour. The reaction was monitored by RP-HPLC. The mixture was purified by preparative HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA aq : 0.1 % TFA in CH_3CN = 70 : 30 to 35 : 65 over 60 min at the flow rate of 2.5 mL/min, RT = 28 min) to afford Fmoc-*L*-Tyr-*L*-Ala-*L*-Asn(glycan)-*L*-Cys-OBn **40b** (2.9 mg, 80%). HRMS (ESI): *m*/z calc for $C_{141}H_{192}N_{11}O_{72}S$ $[M+H]^+$ 3223.1421, found for 3223.1295.

Chemical glycan insertion strategy

To a solution of **11** (2.0 mg, 0.77 μmol) in dry DMSO (51.0 μL) was added Fmoc-*L*-Tyr-*L*-Ala-SH **38** (1.1 mg, 2.2 μmol) at room temperature. The resultant mixture was stirred for 9 hours. The reaction was monitored by RP-HPLC. After completion of the reaction, the mixture was quenched with a buffer solution (0.2 M sodium phosphate, pH 7.0, 0.75 mL) containing 6.0 M Gdn-HCl and 0.1 M 1,4-dithiothreitol. The mixture was purified by preparative HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA aq : 0.1 % TFA in CH_3CN = 80 : 20 to 45 : 55 over 60 min at the flow rate of 2.5 mL/min, RT = 33 min) to afford Fmoc-*L*-Tyr-*L*-Ala-*L*-Asn(glycan)-SH **39** (1.1 mg, 47%). HRMS (ESI): *m*/z calc for $C_{131}H_{181}N_{10}O_{70}S$ $[M+H]^+$ 3046.0632, found for 3046.0464.

To a solution of the resultant **39** (1.1 mg, 0.36 μmol) in a buffer solution (0.2 M sodium phosphate, pH 5.7, 400 μL) containing 6.0 M Gdn-HCl was added CH_3CN (60.0 μL) containing activated cysteine derivative **29** (1.6 mg, 4.4 μmol) and the mixture was stirred at 23 ºC for 3 hours. The reaction progress was periodically monitored by RP-HPLC. After completion of the reaction, the resulting mixture was quenched with a buffer solution (0.2 M sodium phosphate, pH 7.0, 0.75 mL) containing 6.0 M Gdn-HCl and 0.2 M 1,4-dithiothreitol. Purification of the resultant buffer was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O: 0.1 % TFA in CH_3CN = 70 : 30 to 35 : 65 over 60 min at the flow rate of 2.5 mL/min, RT = 28 min) afforded Fmoc-*L*-Tyr-*L*-Ala-*L*-Asn(glycan)-*L*-Cys-OBn **40a** (0.5 mg, 43%).

^1H NMR (700 MHz, DMSO-D6) δ: 8.26(d, GlcNAc-NH-Asn), 8.22(d, Asn-NH), 8.15(d, Ala-NH), 8.05 (d, NeuAc-NH), 8.04(d, Cys-NH), 8.00 (d, 4H, Phenacyl), 7.88 (d, Fmoc), 7.87(d, GlcNAc-NH), 7.72 (m, Phenacyl, Fmoc), 7.70(d, GlcNAc-NH), 7.67(d, GlcNAc-NH), 7.60 (m, Phenacyl, Fmoc), 7.54(d, Tyr-NH), 7.43-7.28 (m, Benzyl), 7.09(d, Tyr-aromatic), 6.63(d, Tyr-aromatic), 5.72, 5.61 (each d, each 2H, PhenacylCH_2), 5.14 (s, 2H, BenzylCH_2), 4.93 (s, 1H, Man-H-1), 4.85 (d, 1H, AsnGlcNAc-H-1), 4.75 (s, 1H, Man-H-1), 4.60(m, Asn-αH), 4.54(m, Cys-αH), 4.54 (m, Man-H-1), 4.45 (m, 3H, GlcNAc-H-1), 4.28(m, Ala-αH), 4.27 (m, 2H, Gal-H-1), 4.18(m, Tyr-αH), 2.92(dd, Tyr-βH'), 2.85(dd, Cys-βH'), 2.65(dd, Tyr-βH), 2.62(dd, Asn-βH'), 2.51(m, Cys-βH), 2.46(dd, Asn-βH), 1.90 (s, 6H, Acetyl), 1.87, 1.82, 1.81, 1.79 (each s, each 3H, Acetyl), 1.22(bd, 3H, Ala-βH). HRMS (ESI): *m*/z calc for $C_{141}H_{192}N_{11}O_{72}S$ $[M+H]^+$ 3223.1421, found for 3223.1215.

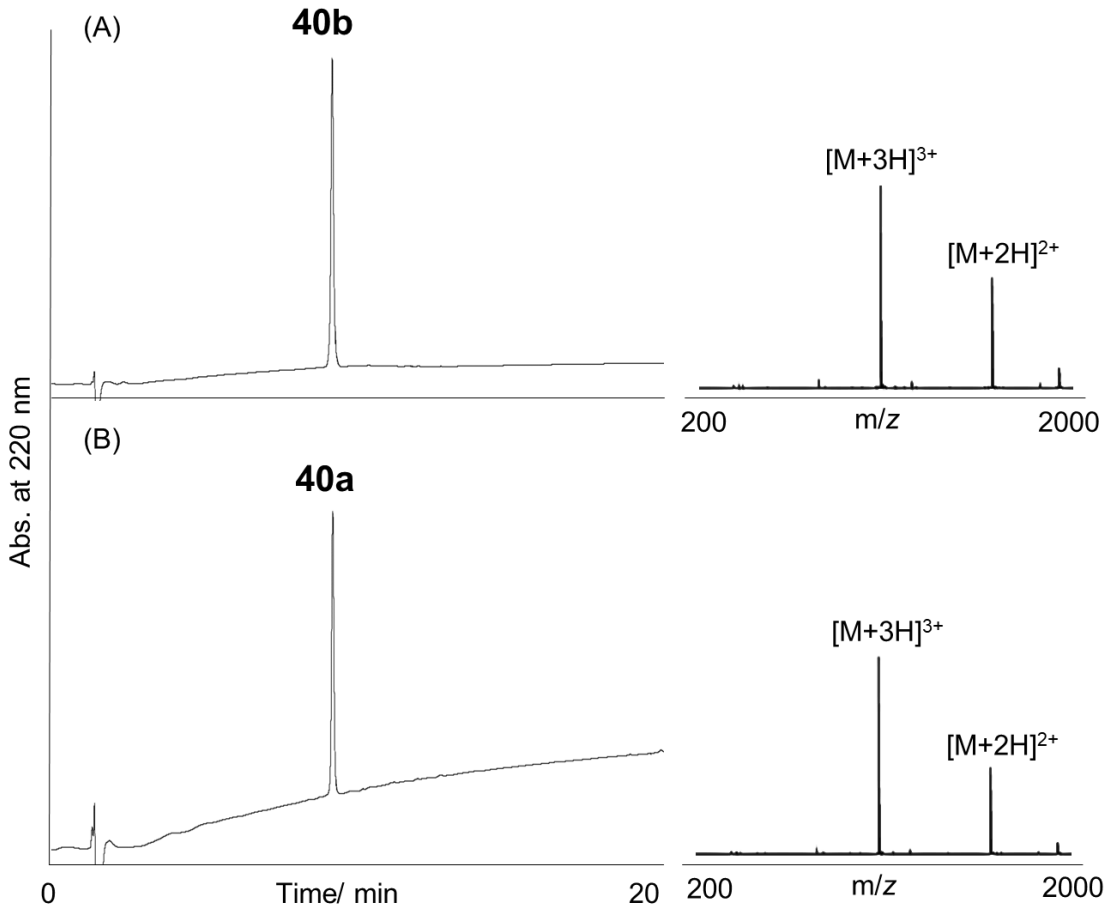


Figure E11. LC and mass spectra of glycopeptide prepared by two strategies. (A) LC of glycopeptide **40b** synthesized by a conventional condensation strategy. (B) LC of glycopeptide **40a** synthesized with DDC (Chemical glycan insertion strategy).

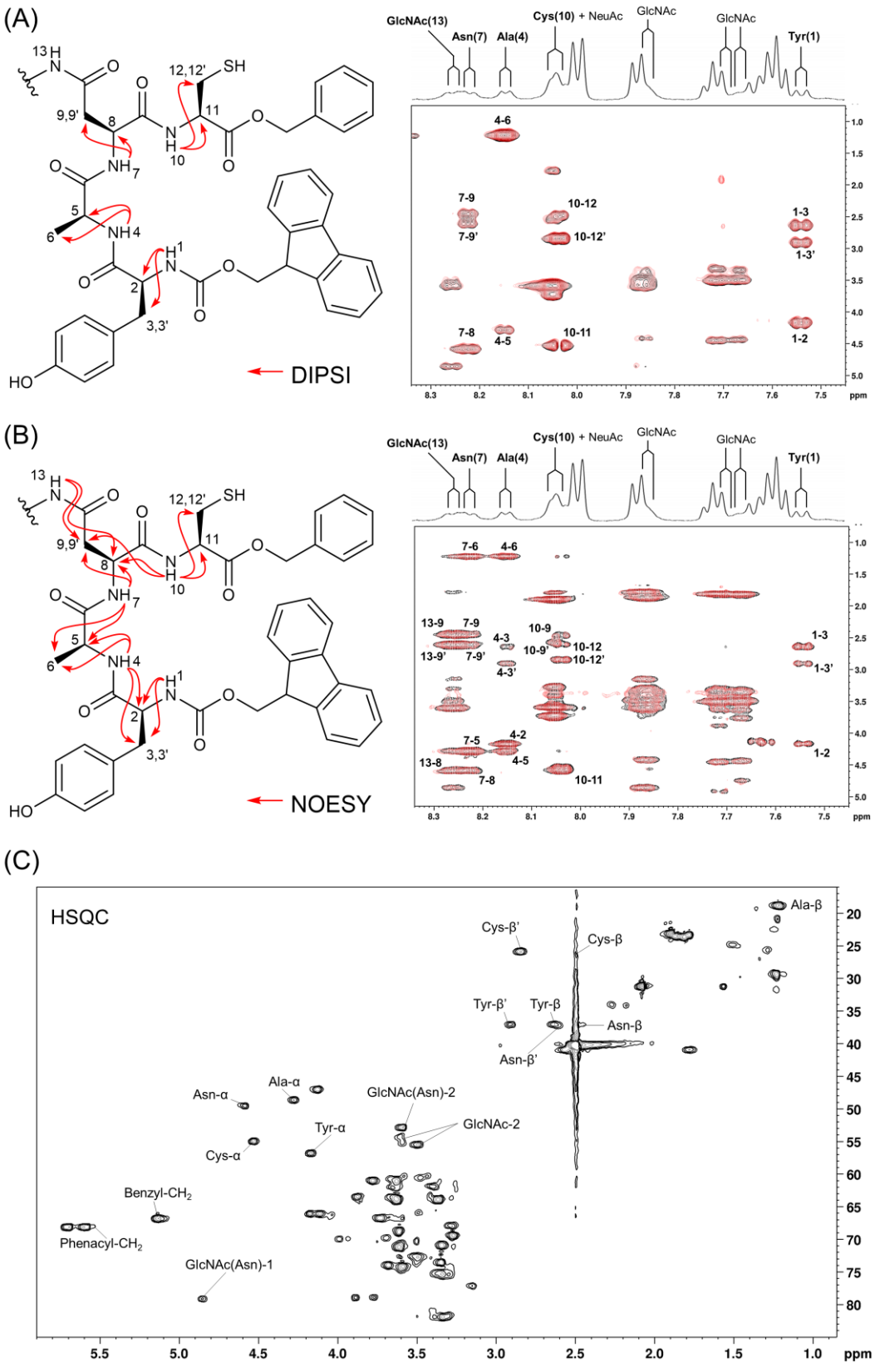


Figure E12. NOESY, DIPSI and HSQC of glycopeptide Fmoc-Tyr-Ala-Asn(glycan)-Cys-OBn **40**. (A)

DIPSI connectivity is shown by red arrows. DIPSI spectrum by 700 MHz NMR. Black contour plot indicates authentic **40b** and red contour plot indicates **40a** prepared by thioacid-mediated strategy. (B) NOESY correlations are shown by red arrows. NOESY spectrum by 700 MHz NMR. Black contour plot indicates authentic **40b** and red contour plot indicates **40a** prepared by thioacid-mediated strategy. (C) HSQC spectrum of **40a** by 700 MHz NMR. The cross peaks of α, β protons and carbons were assigned by HSQC and DIPSI spectra.

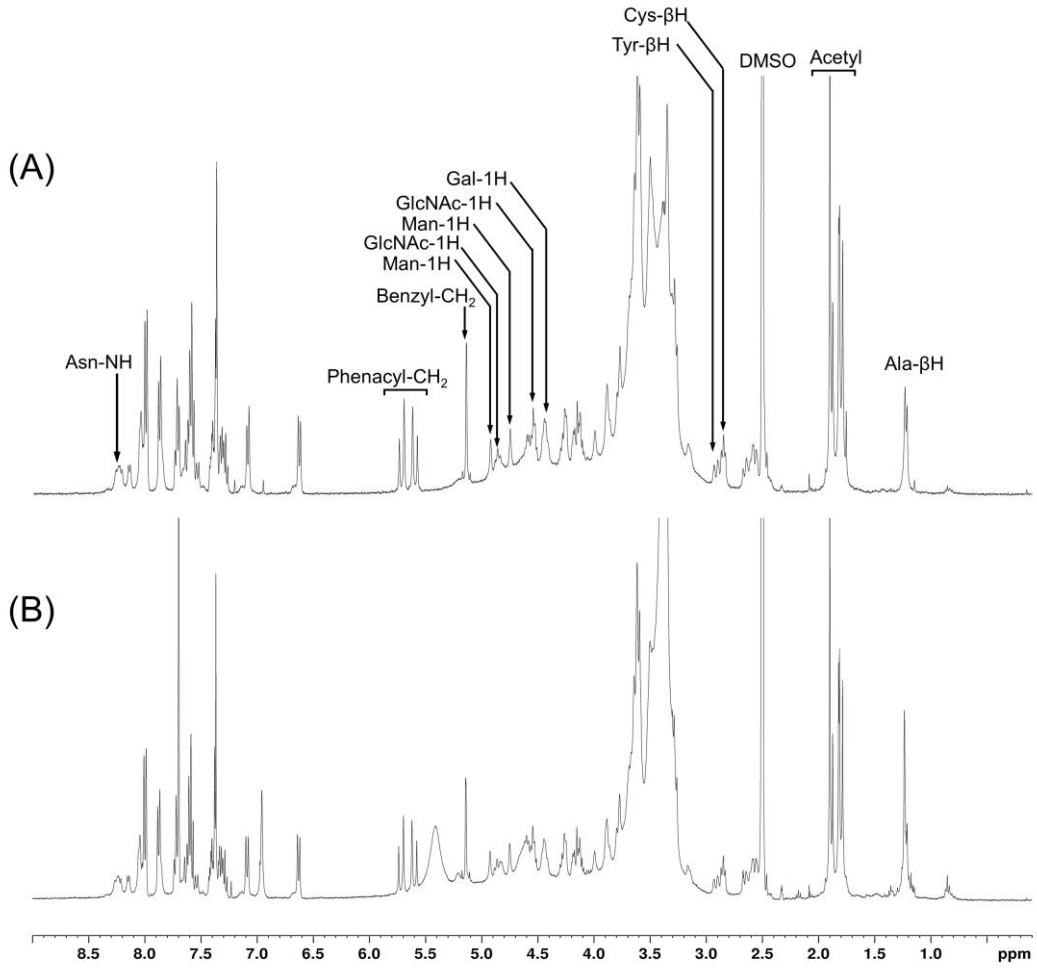


Figure E13. ^{1}H-NMR spectra of the glycopeptide **40** prepared by two strategies. (A) ^{1}H-NMR spectrum of glycopeptide **40b** synthesized by a conventional condensation strategy. (B) ^{1}H-NMR spectrum of glycopeptide **40a** synthesized by DDC (chemical glycan insertion strategy). Since both NMR spectra are identical, both glycopeptides were confirmed to be identical.

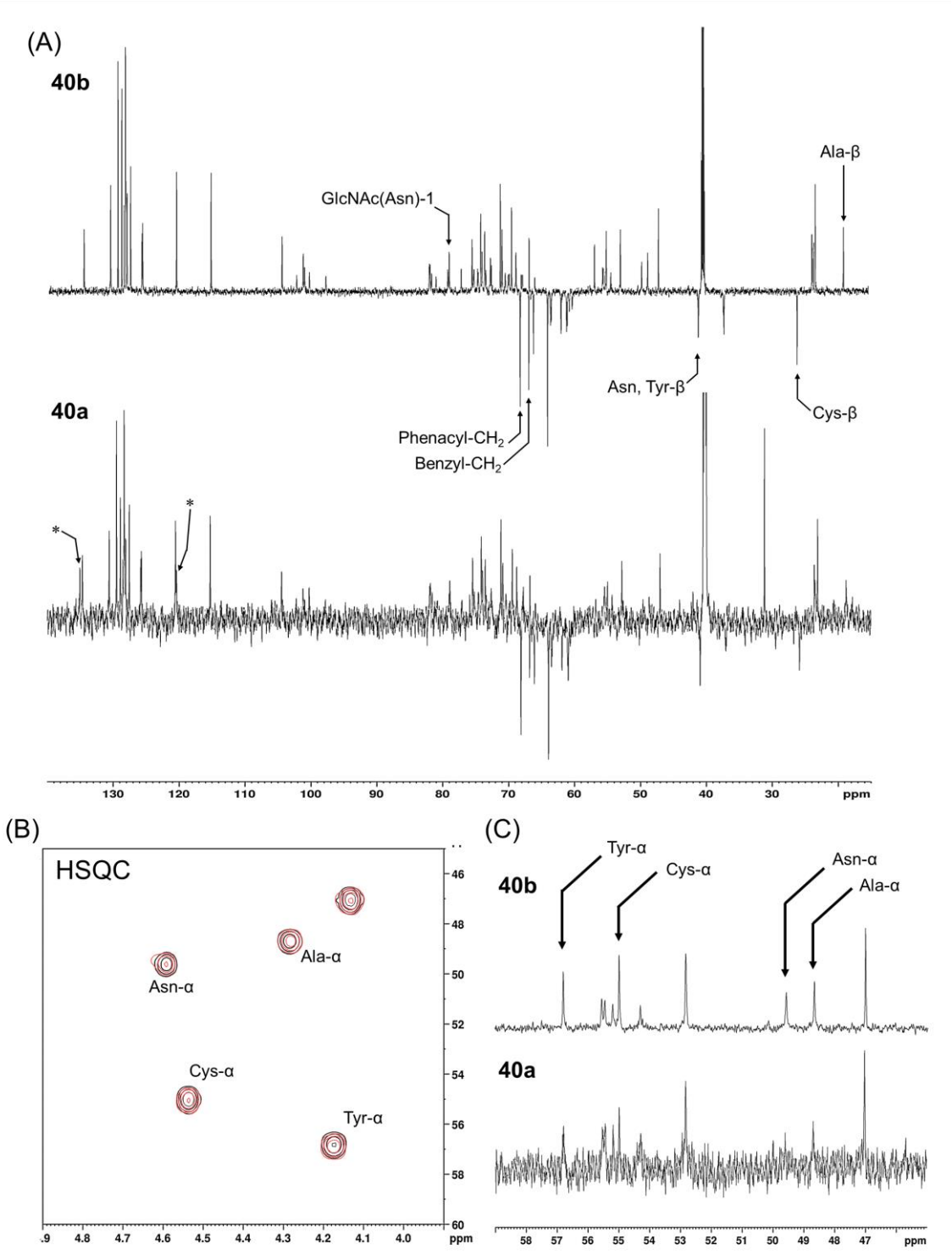


Figure E14. (A) DEPT ^{13}C-NMR spectrum (175 MHz NMR) of glycopeptide **40b** and **40a** synthesized with diacyl disulfide coupling (thioacid-mediated strategy). Asterisk* indicates impurities.

(B) HSQC spectrum of **40b** and **40a** focused on α-carbons. Black contour plot indicates authentic **40b** and red contour plot indicates **40a**. (C) DEPT ^{13}C-NMR spectrum of α-carbon of all amino acids. All ^{13}C NMR data indicate that both synthetic methodologies yielded an analytically single isomer.

7. Synthesis of CC Chemokine Ligand 1 (Chapter 3)

H_2N-Lys[1]-Ala[28]-COSH 54

The CCL1 segment A: H_2N-KSMQVPFSRC(Acm) C(Acm)FSFAEQEIP LRAILC(Acm)YA-COSH was synthesized on Aminomethyl ChemMatrix® resin (50 μmol scale) by Boc solid-phase peptide synthesis (SPPS). The protocol was performed with the same manner in the preparation of model peptide thioacid **12**. The Arg at the 28 position was replaced with Ala for improving the efficacy of diacyl disulfide coupling. Purification of the resultant solution including H_2N-Lys[1]-Ala[28]-COSH **54** was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O: 0.1 % TFA in CH_3CN = 70 : 30 to 55 : 45 over 90 min at the flow rate of 2.5 mL/min) afforded H_2N-Lys[1]-Ala[28]-COSH **54** (9.5 mg, 5.3 %). HRMS(ESI): m/z calc for $C_{154}H_{241}N_{40}O_{41}S_5$ $[M+H]^+$ 3466.6606, found for 3466.6761.

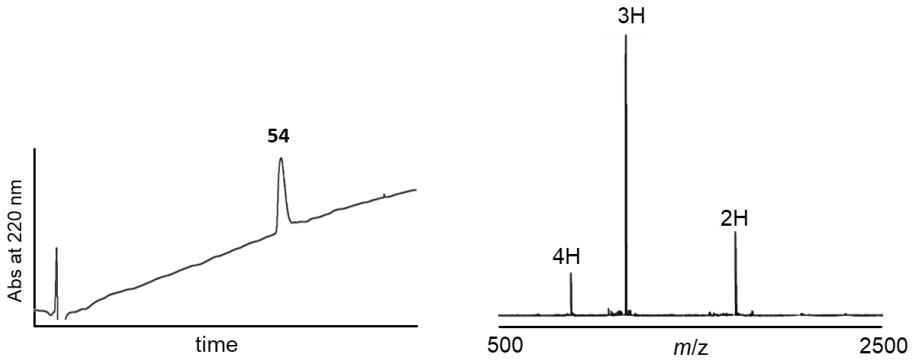


Figure E16. LC profile and MS spectrum of the purified H_2N-Lys[1]-Ala[28]-COSH **54**.

H_2N-Cys[30](Npys)-Lys[73]-COOH 55

The CCL1 segment B: H_2N-C(Npys)SSIC(Acm)SNEGL IFKLKRGKEA C(Acm)ALDTVGWVQ RHRKMLRHC(Acm)P SKRK-COOH was synthesized on Aminomethyl ChemMatrix® resin (50 μmol scale) by Rapid flow-based peptide synthesis developed by the Pentelute group.[3] The coupling of Aminomethyl ChemMatrix® resin (50 μmol scale) and 4-(4-Hydroxymethyl-3-methoxyphenoxy)butyric acid (HMPB, 30.0 mg, 0.125 mmol, 2.5 equiv.) was performed with 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethylaminium tetrafluoroborate (TBTU, 42.6 mg, 0.133 mmol, 2.7 equiv.) and N-ethylmorpholine (17.0 μL, 0.138 mmol, 2.8 equiv.) in dimethyl formamide (DMF, 1.0 mL) for 3 hours at room temperature. After washing of the resin with DMF, amino acid coupling reactions were performed. Fmoc-Lys(Trt)-OH (0.117 g, 0.25 mmol, 5.0 equiv.) was dissolved in DCM (1.0 mL) containing 1-(2-Mesitylenesulfonyl)-3-nitro-1H-1,2,4-triazole (MSNT, 76.0 mg 0.25 mmol, 5.0 equiv.) and N-methylimidazole (15.0 μL, 0.185 mmol, 3.7 equiv.) and the mixture was left for 1 min and then the reaction mixture was added to the resin. The suspension was stirred for 15 min at room temperature. Other amino acid was coupled with flow system.[3] After elongation of peptide, tert-butyl thiol group of the N-terminal cysteine was reduced trough DL-dithiothreitol (DTT, 5.0 equiv.).

To the resulting mixture was added 2,2′-Dithiobis(5-nitropyridine) (10.0 equiv.) and stirred for 1 hour. The peptide on the resin was released with trifluoracetic acid (5.0 mL) and concentrated in vacuo. Purification of the residue was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 80 : 20 to 50 : 50 over 40 min at the flow rate of 2.5 mL/min) afforded H_2N-Cys^{30}(Npys)-Lys^{73}-COOH **55** (9.0 mg, 3.3 %). HRMS(ESI): m/z calc for $C_{231}H_{383}N_{77}O_{63}S_6$ $[M+H]^+$ 5436.7535, found for 5436.7509.

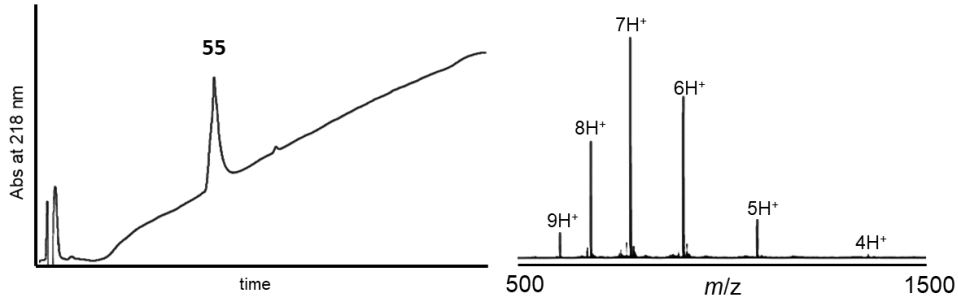


Figure E17. LC profile and MS spectrum of **55.**

H_2N-Lys^1-Asn(glycan)^{29}-COSH 56

(DDC)

To a solution of H_2N-Lys^1-Ala^{28}-COSH **54** (2.8 mg, 0.97 μmol, 2.0 equiv.) in anhydrous DMSO (27.0 μL) was added H_2N-Asn(glycan)-SH **11** (1.1 mg, 0.42 μmol, 1.0 equiv) at 23 ºC and stirred for 6 hours. The reaction progress was periodically monitored by RP-HPLC. After completion of the reaction, the mixture was quenched with a buffer solution (0.2 M sodium phosphate, pH 7.0, 0.50 mL) containing 6.0 M Gdn-HCl and 0.1 M 1,4-dithiothreitol. Purification of the resultant buffer solution was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O: 0.1 % TFA in CH_3CN = 70 : 30 to 40 : 60 over 40 min at the flow rate of 1.0 mL/min, RT = 15 min) afforded H_2N-Lys^1-Asn(glycan)^{29}-COOH **56** (0.7 mg, 27%). HRMS(ESI): m/z calc for $C_{258}H_{395}N_{48}O_{106}S_5$ $[M+H]^+$ 6021.5597, found for 6021.5709.

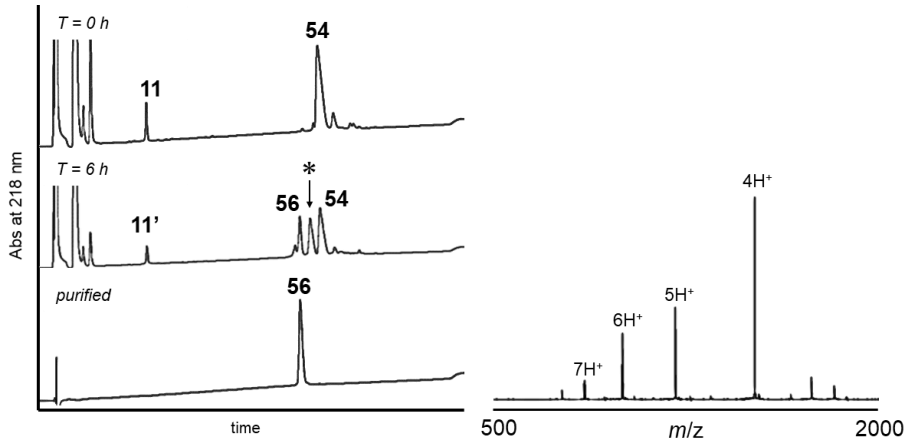


Figure E18. LC profile and MS spectrum of the purified **56**. Compound **11'** was found to be an aspartimide form of glycosyl asparagine thioacid **11**: HRMS(ESI): m/z calc for $C_{104}H_{155}N_8O_{65}$ $[M+H]^+$ 2555.9069, found for 2555.9294. Compound ***** asterisk indicates hydrolysis form of peptide thioacid **54**: HRMS(ESI): m/z calc for $C_{154}H_{241}N_{40}O_{42}S_4$ $[M+H]^+$ 3450.6835, found for 3450.6972.

H_2N-Lys^1-Cys^{10,11,26}(SAcm)-Asn(glycan)^{29}-Cys^{30}-Cys^{34,50,68}(SAcm)-Lys^{73}-COOH 57 (TCL)

To a solution of H_2N-Cys^{30}(Npys)-Lys^{73}-COOH **55** (0.60 mg, 0.11 μmol, 2.0 equiv.) in a buffer solution (0.2 M sodium phosphate, pH 5.7, 110.0 μL) containing 6.0 M Gdn-HCl was added H_2N-Lys^1-Asn(glycan)^{29}-COSH **56** (0.33 mg, 0.055 μmol, 1.0 equiv.) and the mixture was stirred at 23 ºC for 24 hours. The reaction progress was periodically monitored by RP-HPLC. After completion of the reaction, the resulting mixture was quenched with a buffer solution (0.2 M sodium phosphate, pH 7.0, 0.75 mL) containing 6.0 M Gdn-HCl and 0.1 M 1,4-dithiothreitol. Purification of the resultant buffer was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O: 0.1 % TFA in CH_3CN = 80 : 20 to 50 : 50 over 40 min at the flow rate of 1.0 mL/min, RT = 26 min) afforded H_2N-Lys^1-Cys^{10,11,26}(SAcm)-Asn(glycan)^{29}-Cys^{30}-Cys^{34,50,68}(SAcm)-Lys^{73}-COOH **57** (0.6 mg, >90%). HRMS(ESI): m/z calc for $C_{484}H_{774}N_{123}O_{167}S_9$ $[M+H]^+$ 11276.7498, found for 11276.7045 (average isotopes).

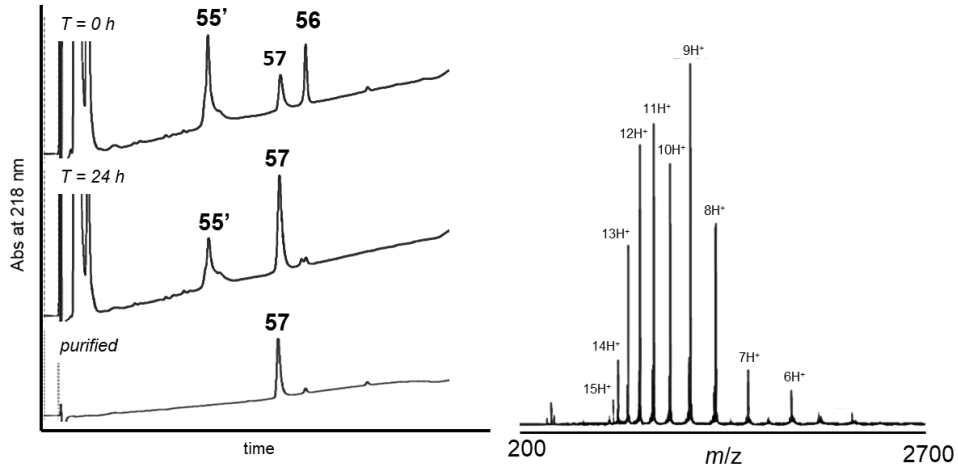


Figure E19. LC profile and MS spectrum of purified **57**. **55'** indicates reductive form without a disulfide bond at Cys^{30} of **55.**

H_2N-Lys^{1}-$Cys^{10,11,26}$(SAcm)-Asn(glycan)^{29}-Ala^{30}-$Cys^{34,50,68}$(SAcm)-Lys^{73}-COOH 63 (Desulfurization of Cys^{30})

To a powder of H_2N-Lys^{1}-$Cys^{10,11,26}$(SAcm)-Asn(glycan)^{29}-Cys^{30}-$Cys^{34,50,68}$(SAcm)-Lys^{73}-COOH **57** (0.3 mg, 0.03 μmol, 1.0 equiv.) was dissolved in a buffer solution (0.2 M sodium phosphate, pH 7.0, 1.0 mL) containing TCEP (130 mg, 0.45 mmol), 2,2'-Azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044, 1.6 mg, 4.9 μmol), sodium 2-sulfanylethanesulfonate (MESNA, 50.0 mg, 0.31 mmol) and 6.0 M Gdn-HCl. The resulting mixture was stirred for 14 hours at room temperature. Purification of the product was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 85 : 15 to 50 : 50 over 40 min at the flow rate of 1.0 mL/min, RT = 25 min) afforded H_2N-Lys^{1}-$Cys^{10,11,26}$(SAcm)-Asn(glycan)^{29}-Ala^{30}-$Cys^{34,50,68}$(SAcm)-Lys^{73}-COOH **63** (0.3 mg, >90%). HRMS(ESI): m/z calc for $C_{484}H_{774}N_{123}O_{167}S_8$ $[M+H]^+$ 11244.6898, found for 11244.6726 (average isotopes).

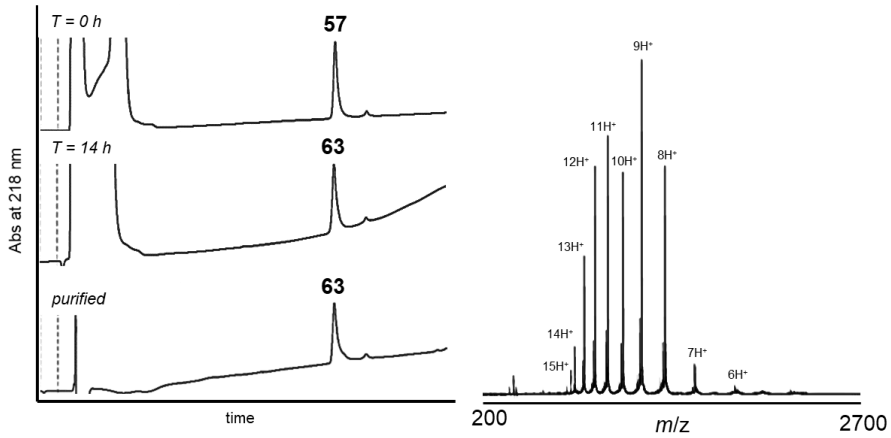


Figure E20 LC profile and MS spectrum of purified **63**.

H_2N-Lys[1]-Asn(glycan)[29]-Lys[73]-COOH 58

(Deprotection of Acm and Pac protecting group of glycan)

Palladium (II) chloride ($PdCl_2$, 0.6 mg, 3 μmol) was dissolved in a buffer solution (0.2 M sodium phosphate, pH 7.0, 170 μL) containing 6.0 M Gdn-HCl. To a solution of H_2N-Lys[1]-Cys[10,11,26](SAcm)-Asn(glycan)[29]-Ala[30]-Cys[34,50,68](SAcm)-Lys[73]-COOH **63** (0.3 mg, 0.03 μmol) in a buffer solution (0.2 M sodium phosphate, pH 7.0, 110 μL) containing 6.0 M Gdn-HCl was added above $PdCl_2$ solution (60.0 μL) at 30 ºC. The resultant mixture was stirred at 30 ºC for 1.5 hours. To the reaction mixture was added DTT (4.0 mg, 0.026 mmol) and the resultant mixture immediately afforded red precipitate. To the mixture was added 2-mercaptoethanol (BME, 3.0 μL, 0.043 mmol), piperidine (5.0 μL, 0.051 mmol) and a buffer solution (0.2 M sodium phosphate, pH 7.0, 52.0 μL) containing 6.0 M Gdn-HCl at 30 ºC. The reaction solution changed to transparent and was further stirred for 1.5 hours. The reaction was quenched by adjusting pH 7.0 with 5.0 M hydrochloride solution. Purification of the product was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 85 : 15 to 50 : 50 over 40 min at the flow rate of 1.0 mL/min, RT = 28 min) afforded H_2N-Lys[1]-Asn(glycan)[29]-Lys[73]-COOH **58** (0.2 mg, 71%) HRMS(ESI): m/z calc for $C_{450}H_{732}N_{117}O_{159}S_8$ $[M+H]^+$ 10581.9458, found for 10581.9196 (average isotopes).

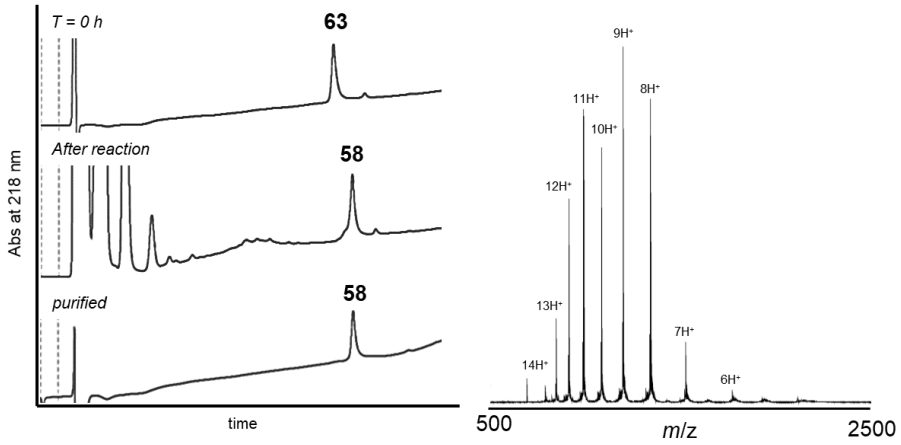


Figure E21 LC profile and MS spectrum of purified **58**.

Glycosyl CCL1 53

(Folding of H_2N-Lys[1]-Asn(glycan)[29]-Lys[73]-COOH 58)

The Linear Lys[1]-Asn(glycan)[29]-Lys[73] **58** (0.3 mg, 0.03 μmol) was dissolved in a buffer solution (0.2 M sodium phosphate, pH 8.0, 90.0 μL) containing 6.0 M Gdn-HCl and followed by addition of H_2O (90.0 μL). To the solution was added a buffer solution (150 mM Tris-HCl, pH 8.0, 400.0 μL) containing, 12.0 mM cysteine and 1.5 mM cystine. The resultant mixture was stirred for 2 hours at room temperature. To the solution was added H_2O (400.0 μL) and further stirred for 36 hours at room temperature. Folding processes were periodically monitored by LCMS. Purification of the product was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 85 : 15 to 50 : 50 over 40 min at the flow rate of 1.0 mL/min) afforded glycosyl CCL1 **53** (0.1 mg, 33%). HRMS(ESI): m/z calc for $C_{450}H_{726}N_{117}O_{159}S_8$ $[M+H]^+$ 10575.8978, found for 10575.9479 (average isotopes).

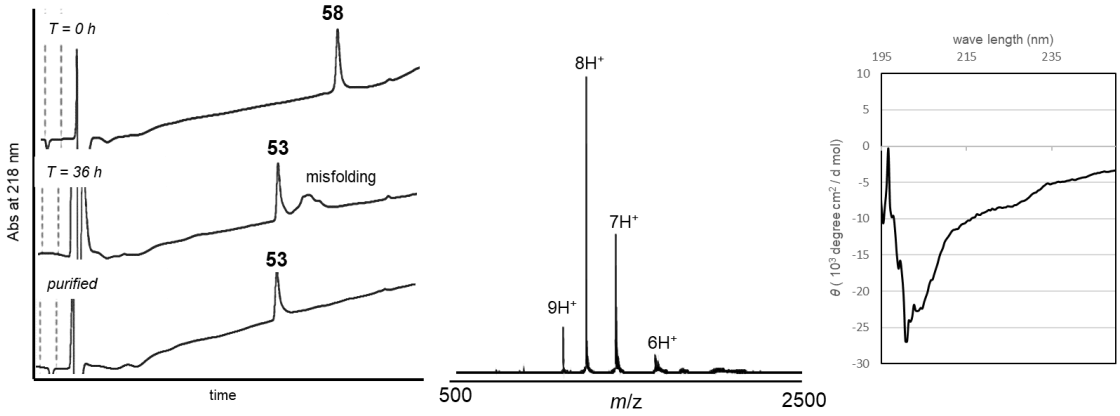


Figure E22. LC profile, MS spectrum and CD spectrum of purified **53**.

Characterization of folded CCL1 53

To confirm the disulfide bond positions, we employed a standard strategy using chymotrypsin digestion and subsequent mass analysis. The folded CCL1 **53** was dissolved in a buffer solution (200 mM tris-HCl, pH 8.0, 100 μL) and then an appropriate amount of α-chymotrypsin (ca. 7 μg) was added. After 15 h, the resultant fragments were analyzed by LC/MS.

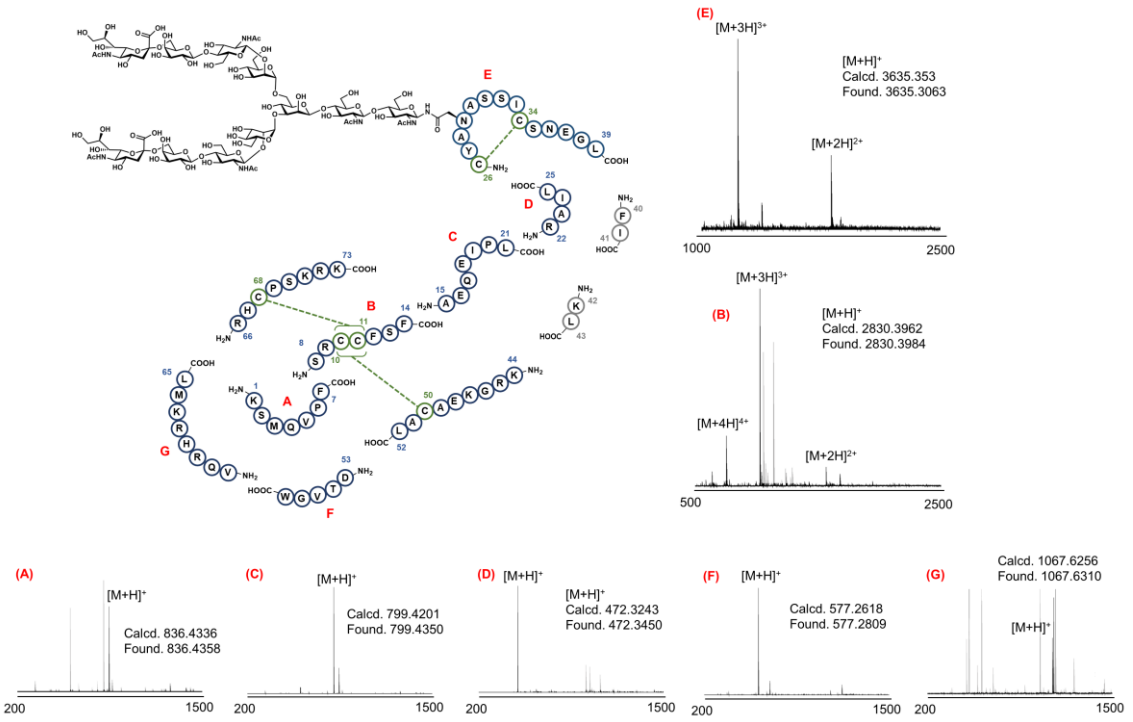


Figure E23. MS spectra of the digested fragments of folded CCL1 **22**. Fragment A: Lys[1]-Phe[7]. Fragment B: Ser[10]-Phe[14] forming disulfide bonds with Lys[44]-Leu[52] and Arg[66]-Lys[73]. Fragment C: Ala[15]-Leu[21]. Fragment D: Arg[22]-Leu[25]. Fragment E: Cys[26]-Leu[39] forming an internal disulfide bond. Fragment F: Asp[53]-Trp[57]. Fragment G: Val[58]-Leu[65].

8, Synthesis of interleukin 3 ***(Chapter 4)***

Synthesis of Ala^{1}-Ala^{14}-COSH 65

The IL3 segment A: H_2N- APMTQTTPLK(Fmoc)TSW(CHO)A-COSH **65** was synthesized on Aminomethyl ChemMatrix® resin (50 μmol scale) by Boc solid-phase peptide synthesis (SPPS). The protocol was performed with the same manner in the preparation of model peptide thioacid **12**. The Val at the 14 position was replaced with Ala for improving the efficacy of diacyl disulfide coupling. Purification of the resultant product was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 95 : 5 to 55 : 45 over 60 min at the flow rate of 3.0 mL/min) afforded H_2N-Ala^{1}-Ala^{14}-COSH **65** (4.4 mg, 4.9 %). HRMS(ESI): m/z calc for $C_{84}H_{120}N_{17}O_{23}S_2$ $[M+H]^+$ 1798.8184, found for 1798.8287.

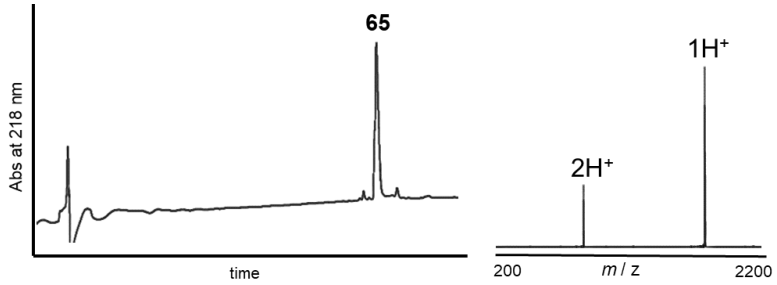


Figure E24. LC profile and MS spectrum of the purified **65**.

H_2N-HisTag-SUMO-Cys^{16}-Cys^{84}-Phe^{133}-COOH 67

Segment B was prepared using expression in *E. coli*. Plasmid DNA was transfected to BL21. The expressed amino acid sequence of **67** is MHHHHHHVNS LEMSDSEVNQ EAKPEVKPEV KPETHINLKV SDGSSEIFFK IKKTTPLRRL MEAFAKRQGK EMDSLRFLYD GIRIQADQTP EDLDMEDNDI IEAHREQIGG CSNMIDEIIT HLKQPPLPLL DFNNLNGEDQ DILMENNLRR PNLEAFNRAV KSLQNASAIE SILKNLLPCL PLATAAPTRH PIHIKDGDWN EFRRKLTFYL KTLENAQAQQ TTLSLAIF. The sequence of HisTag-SUMO was coded in Met^{1}-Gly^{110}. Between Met-HisTag and SUMO sequence were spacer peptide, VNSLE. All codons were optimized for *E. coli.* expression system. The cells were cultured in LB medium solution (10.0 mL) containing ampicillin (0.1 mg / mL) at 37 ºC with shaker 150 rpm overnight. To a LB ampicillin medium solution (1.0 L) was added the culture and the mixture was incubated at 37 ºC with 150 rpm. When the optical density (OD) at 600 nm reached 0.915, production of the recombinant protein was initiated by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration as 1.0 mM. After 24 hours cultivation, the cells were harvested by centrifugation at 4 ºC with 8000 rpm for 10 min 3 times. The precipitated cells were suspended with a lysis buffer solution (50.0 mM Tris-HCl, pH 8.0, 24.0 mL) containing, 0.2 M NaCl, 0.1% Triton-X and 5.0 mM manganese (II) chloride and then was treated by

microwave trituration. The suspension was centrifuged at 4 ℃ with 8000 rpm for 10 min and then the precipitate was suspended in a buffer solution (50.0 mM sodium phosphate 24.0 mL, pH8.0) containing 6.0 M Gdn-HCl, and 0.3 M NaCl. After microwave trituration again, the solution was centrifuged at 4 ºC with 8000 rpm for 10 min 3 times. Desired peptide **67** in a buffer solution was isolated by Ni-column purification and eluted with a buffer solution (50.0 mM sodium phosphate, pH8.0, 7.0 mL) containing 0.3 M NaCl, 6.0 M Gdn-HCl and 0.5 M imidazole. Desired product **67** in fractions was collected and then was dialyzed using microtube with Spectra/Por®7 Dialysis Membrane Pre-treated RC Tubing (MWCO: 8 kD) with H_2O (2.0 L) overnight. The resultant solution was then lyophilized to afford H_2N-HisTag-SUMO-Cys^{16}-Cys^{84}-Phe^{133}-COOH **67**. HRMS(ESI): *m*/z calc for $C_{1150}H_{1834}N_{329}O_{351}S_9$ [M+H]+ 26173.8138, found for 26173.7112 (average isotopes).

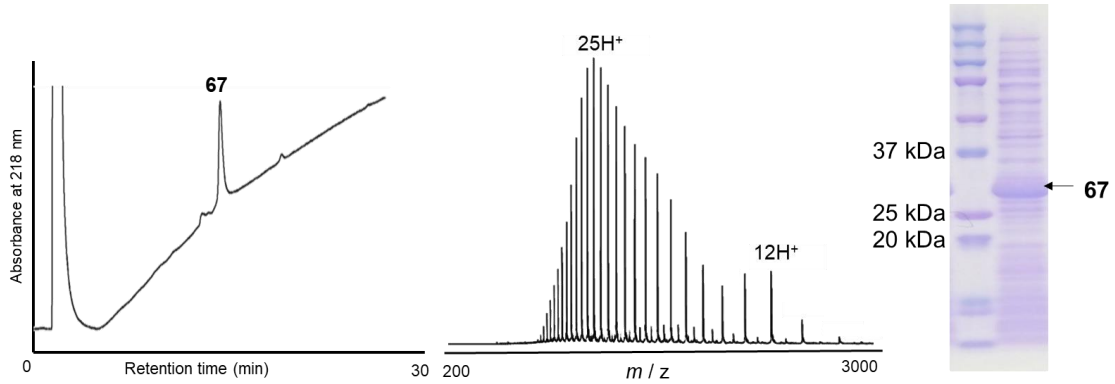


Figure E25. LC profile, MS spectrum of purified **54** and SDS-PAGE of crude **54**.

Thz^{16}-Cys^{84}-Phe^{133}-COOH 68

To a buffer solution (1.0 mL) containing SUMO protease (100 unit) and Tris-HCl (20.0 mM), DTT, (2.0 mM), formaldehyde (0.2 mM) was added H_2N-HisTag-SUMO-Cys^{16}-Cys^{84}-Phe^{133}-COOH **67** and pH of a solution was adjusted to 7.9. The resulting mixture was left for 24 hours at 30 ºC. The proteolysis reaction was periodically monitored by RP-HPLC. During SUMO protease reaction, the resultant peptide was protected with formaldehyde to afford thiazolidine form at the N terminus. After completion of the reaction, product was purified by RP HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA aq. : 0.1 % TFA in CH_3CN = 75 : 25 to 10 : 90 over 60 min at the flow rate of 3.0 mL/min) to afford Thz^{16}-Cys^{84}-Phe^{133}-COOH **68** (4.0 mg). HRMS(ESI): *m*/z calc for $C_{599}H_{962}N_{167}O_{376}S_4$ $[M+H]^+$ 13447.5178, found for 13447.5166 (average isotopes).

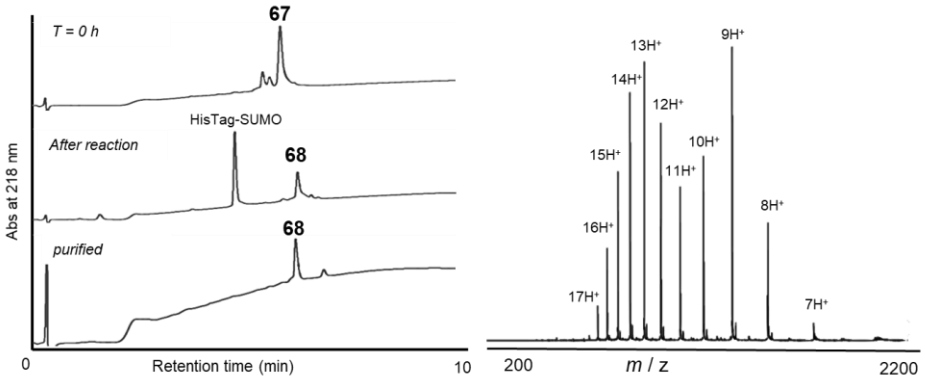


Figure E26. LC profile and MS spectrum of purified **68**.

H_2N-Cys[16](Npys)-Cys[84](Pac)-Phe[133]-COOH 66

To a solution of Thz[16]-Cys[84]-Phe[133]-COOH **55** (1.9 mg, 0.14 μmol, 1.0 equiv.) in H_2O (486 μL) and CH_3CN (20.0 μL) was added dimethylformamide (DMF, 1.1 μL) containing 0.2 M Phenacyl bromide (0.22 μmol, 1.6 equiv.) and then the mixture was stirred for 2.5 hours at room temperature to afford Thz[16]-Cys[84](SPac)-Phe[133]-COOH **69** (HRMS(ESI): m/z calc for $C_{607}H_{968}N_{167}O_{177}S_4$ $[M+H]^+$ 13565.6528, found for 13565.7079 (average isotopes). The reaction was monitored by RP-HPLC. This mixture was directly used for the next reaction without further purification.

To the reaction mixture was added a solution (H_2O-CH_3CN = 3:2, 0.50 mL) containing 0.1% trifluoroacetic acid (0.1 % v/v) and 2,2′-dipyridyl disulfide (DPDS, 17.0 mg, 77.3 μmol).[4] The resultant solution was stirred at 37 ºC for 2.5 hours. Then, to a reaction mixture was added DTT (120.0 mg, 0.778 mmol) and left for 2 hours at 23 ºC. The reaction was monitored by RP-HPLC. After completion of the reaction, the reaction mixture was purified by RP HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA aq. : 0.1 % TFA in CH_3CN = 80 : 20 to 20 : 80 over 45 min at the flow rate of 1.0 mL/min) to afford H_2N-Cys[16]-Cys[84](SPac)-Phe[133]-COOH **70** (0.9 mg, 47%). HRMS(ESI): *m*/z calc for $C_{606}H_{968}N_{167}O_{177}S_4$ $[M+H]^+$ 13553.6418, found for 13553.6066 (average isotopes).

To a solution of H_2N-Cys[16]-Cys[84](SPac)-Phe[133]-COOH **70** (0.7 mg, 0.05 μmol) in a solution (H_2O-CH_3CN = 4:1, 1.4 mL) containing trifluoroacetic acid (0.1 %, v/v) was added 2,2′-dithiobis(5-nitropyridine) (18.0 mg, 58.0 μmol). The reaction mixture was left for 5.5 hours at room temperature. The reaction was monitored by RP-HPLC. After completion of the reaction, the product was purified by RP HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA aq. : 0.1 % TFA in CH_3CN = 80 : 20 to 20 : 80 over 45 min at the flow rate of 1.0 mL/min) to afford H_2N-Cys[16](Npys)-Cys[84](SPac)-Phe[133]-COOH **66** (0.7 mg, >90%). HRMS(ESI): *m*/z calc for $C_{611}H_{970}N_{169}O_{179}S_5$ $[M+H]^+$ 13707.7848, found for 13707.8214 (average isotopes).

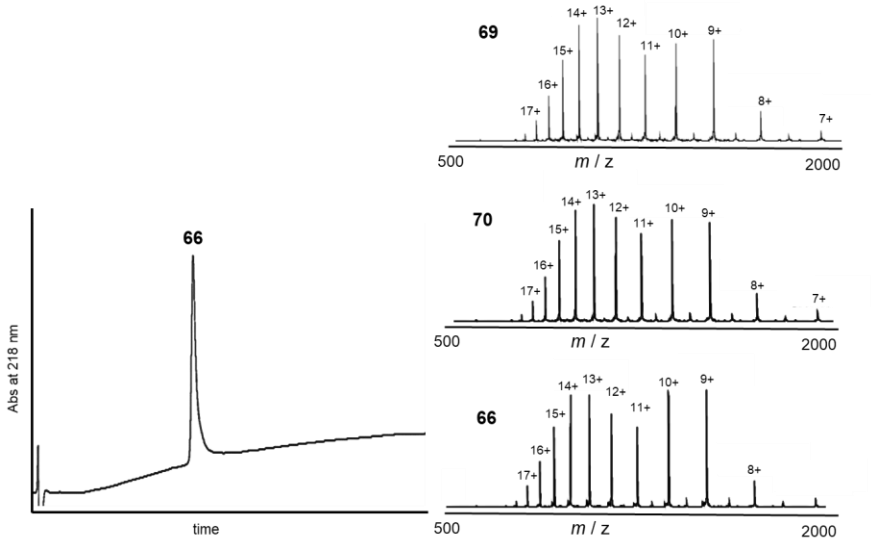


Figure E27. MS spectra of **69, 70** and **66** and LC profile of purified **66**.

H_2N-Ala^1-Asn(glycan)^{15}-COSH 71

(DDC)

To a solution of H_2N-Ala^1-Ala^{14}-COSH **65** (2.0 mg, 1.1 μmol, 2.0 equiv.) in anhydrous DMSO (37.0 μL) was added H_2N-Asn(glycan)-SH **11** (1.4 mg, 0.54 μmol, 1.0 equiv.) at 23 ºC and the mixture was stirred for 6 hours. The reaction progress was periodically monitored by RP-HPLC. After completion of the reaction, the resulting mixture was quenched with a buffer solution (0.2 M sodium phosphate, pH 7.0, 0.80 mL) containing 6.0 M Gdn-HCl and 0.1 M 1,4-dithiothreitol. Purification of product was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 70 : 30 to 40 : 60 over 30 min at the flow rate of 1.0 mL/min, RT = 14 min) to afford H_2N-Ala^1-Asn(glycan)^{15}-COOH **71** (0.8 mg, 34%). HRMS(ESI): m/z calc for $C_{188}H_{273}N_{25}O_{88}S_2$ $[M+H]^+$ 4353.7175, found for 4353.6961.

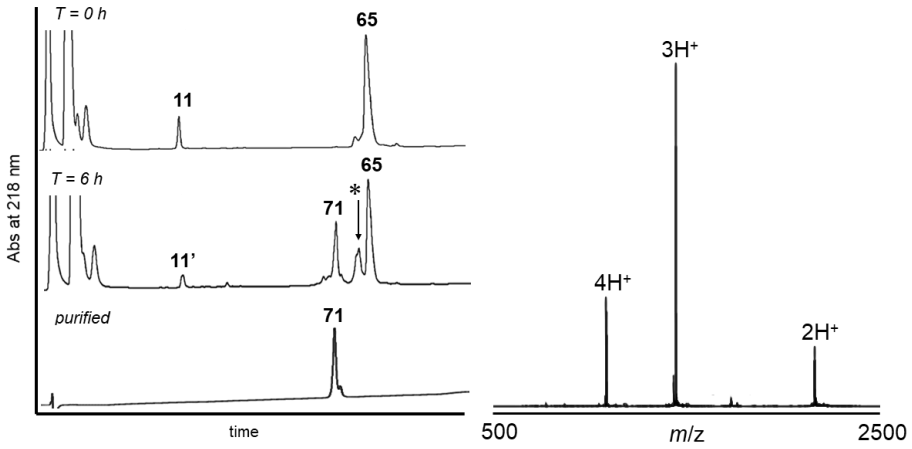


Figure E28 LC and MS of purified **71**. Compound **11'** was found to be an aspartimide form of glycosyl asparagine thioacid **11**; HRMS(ESI): m/z calc for $C_{104}H_{155}N_8O_{65}$ $[M+H]^+$ 2555.9069, found for 2555.9294. Compound * asterisk was a hydrolyzed form of peptide thioacid **65**; HRMS(ESI): m/z calc for $C_{84}H_{120}N_{17}O_{24}S$ $[M+H]^+$ 1782.8413, found for 1782.8589.

H_2N-Ala^1-Lys^{10}(Fmoc)-Trp^{13}(CHO)-Asn(glycan)^{15}-Cys^{84}(SPac)-Phe^{133}-COOH 72 (TCL)

To a solution of H_2N-Cys^{16}(Npys)-Cys^{84}(SPac)-Phe^{133}-COOH **66** (0.75 mg, 0.055 μmol, 1.0 equiv.) in a buffer solution (0.2 M sodium phosphate, pH 5.7, 55 μL) containing 6.0 M Gdn-HCl was added H_2N-Ala^1-Asn(glycan)^{15}-COOH **71** (0.45 mg, 0.10 μmol 2.0 equiv.) and stirred at 23 ºC for 24 hours. The reaction progress was monitored by RP-HPLC. After completion of the reaction, the resulting mixture was quenched with a buffer solution (0.2 M sodium phosphate, pH7.0, 0.75 mL) containing 6.0 M Gdn-HCl and 0.1 M 1,4-dithiothreitol. Purification of product was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 70 : 30 to 30 : 70 over 50 min at the flow rate of 1.0 mL/min, RT = 35 min) to afford H_2N-Ala^1-Lys^{10}(Fmoc)-Trp^{13}(CHO)-Asn(glycan)^{15}-Cys^{84}(SPac)-Phe^{133}-COOH **72** (0.9 mg, >90%). HRMS(ESI): m/z calc for $C_{794}H_{1239}N_{192}O_{265}S_5$ $[M+H]^+$ 17875.0248 found for 17874.9859 (average isotopes).

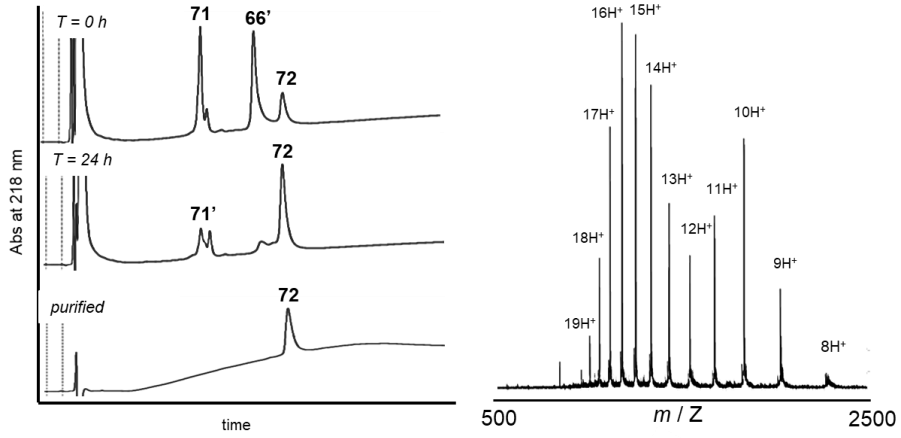


Figure E29 LC profile and MS spectrum of purified **72**. Compound **71'** was hydrolysis and aspartimide form of glycopeptide thioacid **71**; HRMS(ESI): m/z calc for $C_{188}H_{274}N_{25}O_{89}S$ $[M+H]^+$ 4337.7404, found for 4337.7416 and HRMS(ESI): m/z calc for $C_{188}H_{272}N_{25}O_{88}S$ $[M+H]^+$ 4319.7298, found for 4319.7242. **66'** indicates a reductive form without a disulfide bond at Cys^{16} of **66**.

H_2N-Ala^1-Asn(glycan)^{15}-Phe^{133}-COOH 73

(Deprotection of Fmoc, formyl and Pac group of peptide and saccharide)

To a solution of H_2N-Ala^1-Lys^{10}(Fmoc)-Trp^{13}(CHO)-Asn(glycan)^{15}-Cys^{84}(SPac)-Phe^{133}-COOH **72** (0.5 mg, 0.03 μmol) in H_2O- CH_3CN (45.0 μL-22.5 μL) was added piperidine (22.5 μL, 0.23 mmol). The reaction mixture was left for 10 min at 28 ºC. To the mixture was added H_2O (0.6 mL) containing 0.1% trifluoroacetic acid and then concentrated in vacuo to afford H_2N-Ala^1-Asn(glycan)^{15}-Cys^{84}(SPac)-Phe^{133}-COOH **123** as a white powder. HRMS(ESI): m/z calc for $C_{762}H_{1217}N_{192}O_{260}S_5$ $[M+H]^+$ 17388.5018, found for 17388.5541 (average isotopes).

To a solution of **123** in a buffer solution (0.2 M sodium phosphate, pH 5.7, 48.5 μL) containing 6.0 M Gdn-HCl was added 3-mercaptopropionic acid (8.5 μL) and activated zinc (6.8 mg). Then the reaction mixture was left for 2 hours at room temperature. The reaction progress was periodically monitored by RP-HPLC. After completion of the reaction, the reaction was quenched with a buffer solution (0.2 M sodium phosphate, pH 7.0, 115.0 μL) containing 0.2 M 1,4-dithiothreitol and 6.0 M Gdn-HCl. Zinc powder was filtered. Purification of product in the filtrate was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 80 : 20 to 30 : 70 over 50 min at the flow rate of 1.0 mL/min, RT = 32 min) to afford H_2N-Ala^1-Asn(glycan)^{15}-Phe^{133}-COOH **73** (0.3 mg, 62%). HRMS(ESI): m/z calc for $C_{754}H_{1211}N_{192}O_{259}S_5$ $[M+H]^+$ 17270.3668 found for 17270.3034 (average isotopes).

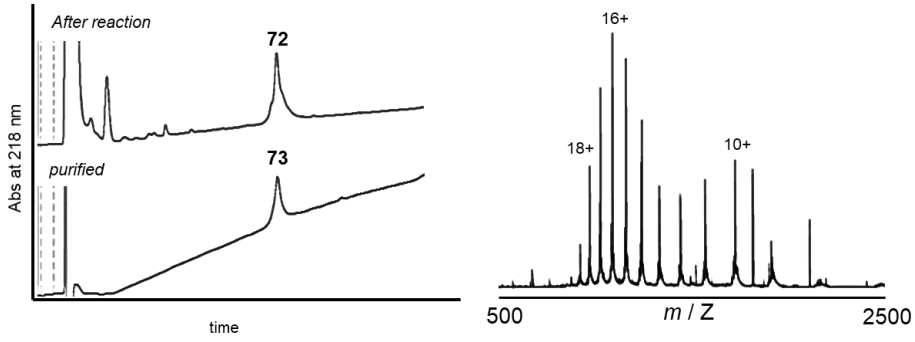


Figure E30. LC and MS of purified **27**.

Glycosyl IL3 64

(Folding of H_2N-Ala[1]-Asn(glycan)[15]-Phe[133]-COOH 73)

H_2N-Ala[1]-Asn(diPac)[15]-Phe[133]-COOH **73** (0.3 mg, 0.02 μmol) was dissolved in a buffer solution (100 mM Tris-HCl, pH 7.5, 0.50 mL) containing 6.0 M Gdn-HCl. This solution was poured into the dialysis tubing, Spectra/Por®7 Dialysis Membrane Pre-treated RC Tubing (MWCO: 8 kD) and then dialyzed against the first folding buffer (100mM Tris-HCl, pH 8.5, 1.0 L) containing 3.0 M Gdn-HCl, 4.0 mM cysteine hydrochloride and 0.5 mM cystine and the solution was left for 10 hours at 4 ºC. Then the external buffer solution was replaced with the second folding buffer (100 mM Tris-HCl, pH 8.0, 1.0 L) containing 1.0 M Gdn-HCl and the solution was left for 15 hours at 4 ºC. Then the external buffer solution was discarded and changed to the third folding buffer (10 mM Tris-HCl, pH 8.0, 1.0 L) and the solution was left for 24 hours at 4 ºC. The reaction progress was periodically monitored by RP-HPLC. Purification of the resultant residue was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 80 : 20 to 30 : 70 over 50 min at the flow rate of 1.0 mL/min) to afford glycosyl IL3 **64** (0.3 mg, analytically quantitative). HRMS(ESI): m/z calc for $C_{754}H_{1209}N_{192}O_{259}S_5$ $[M+H]^+$ 17268.3508, found for 17268.3811 (average isotopes).

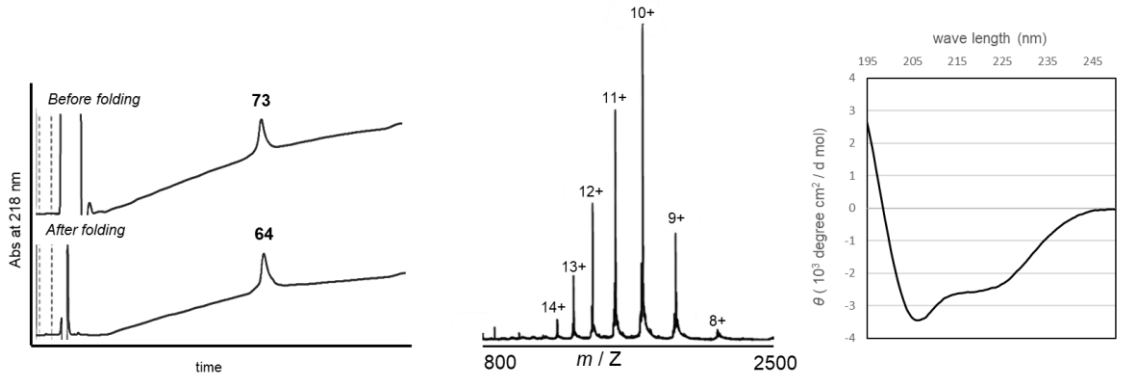


Figure E31. LC profile, MS spectrum and CD spectra of purified **64**.

Measurement of IL3 activity

The activity of IL3 was determined by cell proliferation of the human erythroblast cell line, TF-1 (ATCC, #CRL-2003)[5]. TF-1 cells were maintained in Roswell Park Memorial Institute (RPMI)-1640 medium supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich Japan, Tokyo, Japan) and 2 ng/mL recombinant granulocyte-macrophage colony-stimulating factor (GM-CSF, Wako-Fujifilm, Tokyo, Japan). For the proliferation assay, cells were plated at 1,000/well, and cultured without GM-CSF. Immediately after the plating, the commercially obtained recombinant human IL3 (5.6×10^{-10}, 5.6×10^{-11}, 5.6×10^{-12}, 5.6×10^{-13}, 5.6×10^{-14} mol/L) or equivalent amounts of the synthesized glycosyl human IL3 **64** was added to the culture medium and the cells were incubated for 72h. Cell proliferation was then measured using CellTiter-Glo 2.0 (Promega Japan, Tokyo, Japan) as a luminescent substrate by a GloMax microplate luminometer (Promega) as the manufacturer's instruction.

9. Synthesis of serine protease inhibitor 13 ***(Chapter 5)***

Synthesis of glycosylated SPINK13

H-Gly^{1}-Pro^{31}-SH 80

The SPINK13 segment : H_2N-GIFNKRDFTRWPKPRC(Acm)KMYIPLDPDYNADC(Acm)P-COSH **80** was synthesized on Dawson AM resin (100 μmol scale) by Fmoc SPPS. After swelling a resin by DCM and DMF for 1 h, to a resin (100.0 μmol, 1.0 equiv) was added 20% piperidine / DMF (5.0 mL) for the removal of the Fmoc groups and stirred for 7 min two times. All reagents were filtered, and the resultant resin was washed with DCM and DMF. Then Fmoc-L-Pro-OH (202 mg, 0.60 mmol) was pre-activated with HATU (228 mg, 0.60 mmol) and DIPEA, (157 μL, 0.90 mmol) in anhydrous DMF (2.0 mL) for 1 min and then the activated amino acid was added to a resin. The suspension was gently agitated for 1 h at room temperature. After the coupling of a first amino acid, the resin was washed with DCM and DMF. From next amino acid residues, the rapid flow-based peptide protocols[3] was performed. After all coupling steps were complete, the fully protected dipeptide was cleaved with a cocktail containing TFA/H2O/TIPS (95:2.5:2.5, 7.0 mL), then the reaction mixture was allowed at room temperature and kept for 1 hour. After cleavage, the resultant solution was filtered and evaporated *in vacuo* to give a colorless oil. Purification of the resultant solution including H-Gly^{1}-Pro^{31}-Dbz **90** was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O: 0.1 % TFA in CH_3CN = 80 : 20 to 65 : 35 over 90 min at the flow rate of 2.5 mL/min) afforded H-Gly^{1}-Pro^{31}-Dbz **90** (58.6 mg, 15 %). HRMS(ESI): m/z calc for $C_{183}H_{272}N_{51}O_{47}S_3$ $[M+H]^+$ 4031.9624, found for 4031.9653.

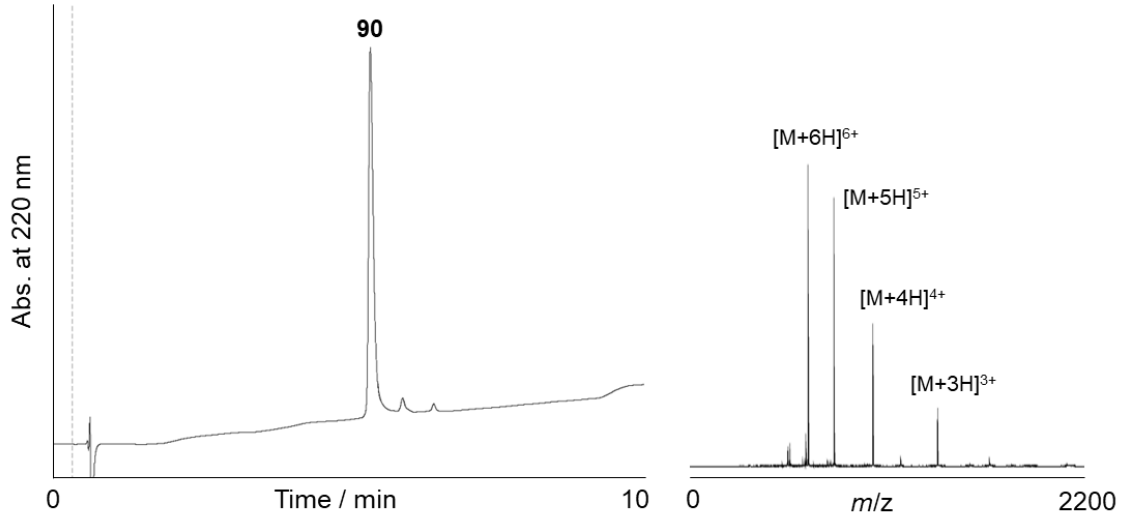


Figure E32. LC and MS spectrum of compound **90**

To peptide H-Gly^{1}-Pro^{31}-Dbz **90** (21.2 mg, 5.26 μmol) was added 0.2 M sodium phosphate buffer (pH

3.0) containing 6 M Gdn-HCl (1.5 mL) and the reaction mixture was cooled at -10 ˚C. To the resultant solution was added 0.4 M $NaNO_2$ aqueous solution (77 μL, 6.0 eq) and subsequently the reaction was stirred for 5 min at -10 °C to afford peptide benzotriazole intermediates **91**, followed by the addition of 0.2 M sodium phosphate buffer (pH 7.5) containing 6 M Gdn-HCl and 0.2 M Na_2S (1.5 mL, 60.0 eq). After stirring for 30 min at room temperature. Purification of the resultant solution was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O: 0.1 % TFA in CH_3CN = 80 : 20 to 65 : 35 over 90 min at the flow rate of 2.5 mL/min) afforded H-Gly[1]-Pro[31]-SH **80** (14.1 mg, 68 %). HRMS(ESI): m/z calc for $C_{176}H_{265}N_{48}O_{46}S_4$ $[M+H]^+$ 3914.8755, found for 3914.8775.

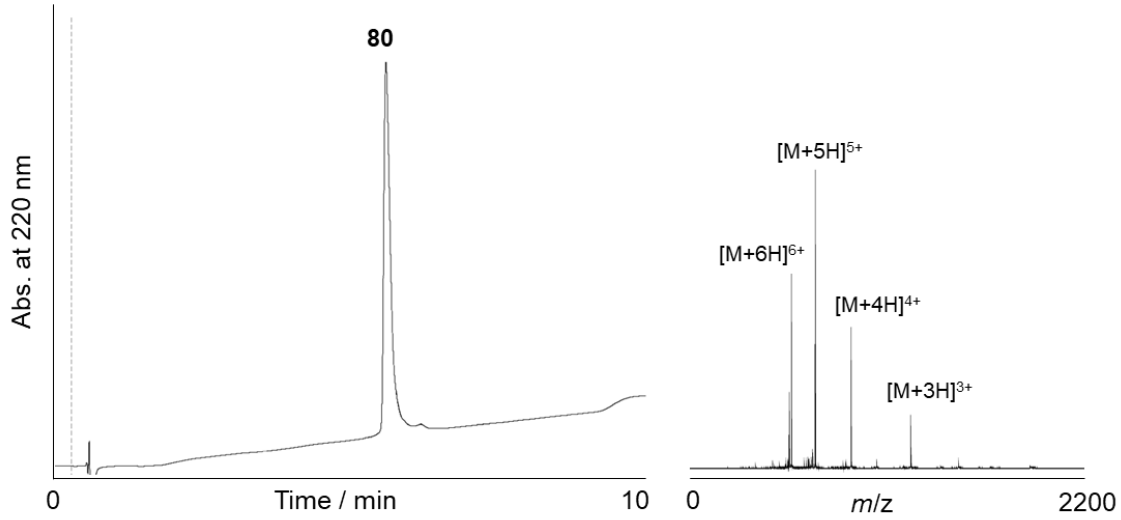


Figure E33. LC and MS spectrum of compound **80**

H-βVal[33](Npys)-Asp[71]-OH 81

The SPINK13 segment : H_2N-βV(Npys)TAPVC(Acm)ASNGHTFQNEC(Acm)FFC(Acm)VEQREF HYRIKFEKYGKC(Acm)D-COOH **81** was synthesized on Aminomethyl ChemMatrix® resin (50 μmol scale) by rapid flow-based peptide synthesis developed by the Pentelute group.[3] The coupling of Aminomethyl ChemMatrix® resin (50 μmol scale) and HMPB (30.0 mg, 0.125 mmol, 2.5 equiv.) was performed with TBTU (42.6 mg, 0.133 mmol, 2.7 equiv.) and N-ethylmorpholine (17.0 μL, 0.138 mmol, 2.8 equiv.) in dimethyl formamide (DMF, 1.0 mL) for 3 hours at room temperature. After washing of the resin with DMF and DCM, amino acid coupling reactions were performed. Fmoc-Asp(O^{t}Bu)-OH (0.103 g, 0.25 mmol, 5.0 equiv.) was dissolved in DCM (1.0 mL) containing MSNT (76.0 mg 0.25 mmol, 5.0 equiv.) and N-methylimidazole (15.0 μL, 0.185 mmol, 3.7 equiv.) and the mixture was left for 1 min and then the reaction mixture was added to the resin. The suspension was stirred for 15 min at room temperature. Other amino acid was coupled with flow system.[3] After all coupling steps were complete, the fully protected dipeptide was cleaved with a cocktail containing TFA/H2O/TIPS (95:2.5:2.5, 7.0 mL), then the reaction mixture was allowed at room temperature and

kept for 1 hour. After cleavage, the resultant solution was filtered and evaporated *in vacuo* to give a colorless oil. To the reaction mixture was added chilled Et_2O and a white precipitate was obtained. The precipitate and was dissolved in CH_3CN-H_2O (2.5 mL-2.5 mL) and to the solution was added 2,2'-dithiobis(5-nitropyridine) (323 mg) and stirred under argon atmosphere for 12 h at room temperature. During the reaction a colorless solution changed to a yellow. The reaction mixture was purified by preparative HPLC (Proteonavi C4 Φ10 × 250 mm, 0.1% TFA aqueous solution: 0.1% TFA CH_3CN = 80: 20 to 60: 40 over 70 min at the flow rate of 2.5 mL/min to give compound **81** (18.5 mg, 7%) as a white powder.

HRMS(ESI): m/z calc for $C_{223}H_{323}N_{62}O_{65}S_6$ $[M+H]^+$ 5101.2199, found for 5101.2203.

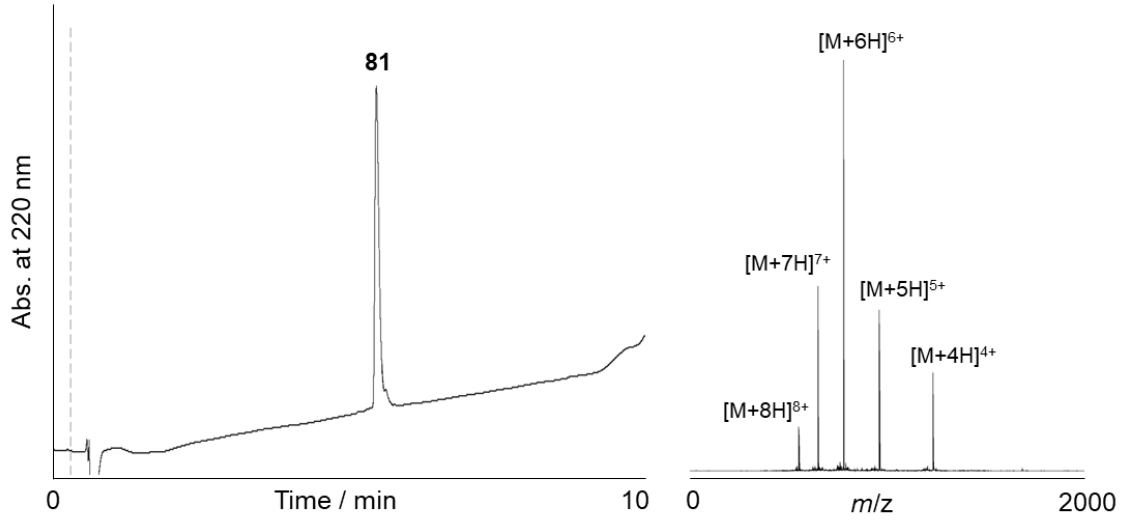


Figure E34. LC and MS spectrum of compound **81**

H-Gly^{1}-Asn(glycan)^{32}-SH 93

(DDC)

To a solution of H-Gly^{1}-Pro^{31}-SH **80** (8.2 mg, 2.2 μmol, 2.0 equiv.) in anhydrous DMSO (60.0 μL) was added H_2N-Asn(glycan)-SH **11** (2.9 mg, 1.1 μmol, 1.0 equiv.) at 23 ºC and stirred for 12 hours. The reaction progress was periodically monitored by RP-HPLC. After completion of the reaction, the mixture was quenched with a buffer solution (0.2 M sodium phosphate, pH 5.5, 1,0 mL) containing 6.0 M Gdn-HCl and 0.2 M DTT. Purification of the resultant buffer solution was performed by RP-HPLC (Proteonavi C4 Φ10 × 250 mm, 0.1% TFA aqueous solution: 0.1% TFA CH_3CN = 80: 20 to 60: 40 over 70 min at the flow rate of 2.5 mL/min to give H-Gly^{1}-Asn(glycan)^{32}-SH **93** (2.0 mg, 28%) as a white powder.

HRMS(ESI): m/z calc for $C_{280}H_{419}N_{56}O_{111}S_4$ $[M+H]^+$ 6469.7746, found for 6469.7755.

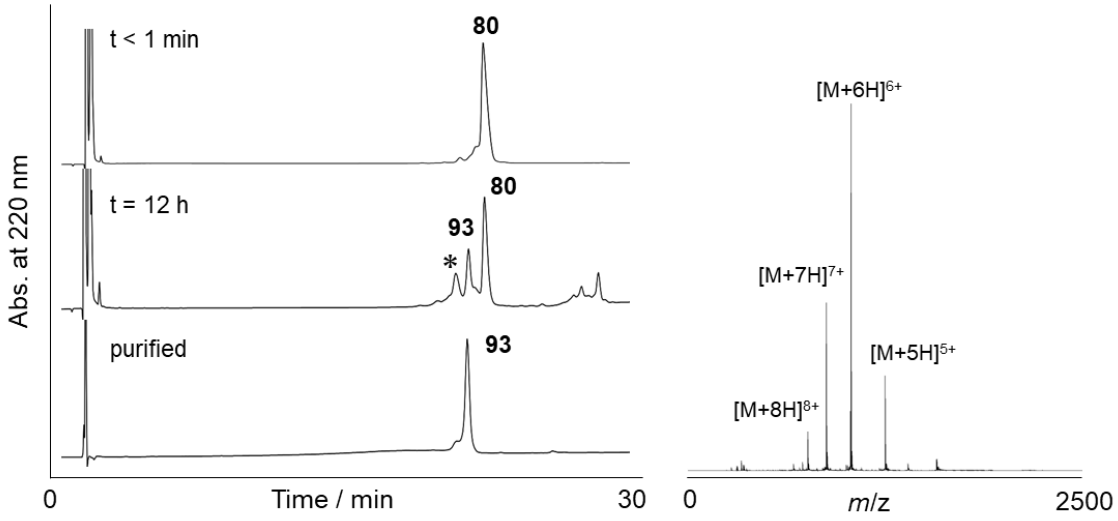


Figure E35. HPLC and MS monitoring of DDC between glycosyl asparagine thioacid **11** and peptide prolyl thioacid **80**. The asterisk indicates glycopeptide-COOH generated by hydrolysis of target **93**.

H-Gly^{1}-Cys^{16,30}(SAcm)-Asn(glycan)^{32}-βVal^{33}-Cys^{38,49,52,70}(SAcm)-Asp^{71}-OH 94

(TCL)

To a solution of H-βVal^{33}(Npys)-Asp^{71}-OH **81** (3.0 mg, 0.59 μmol, 2.0 equiv.) in a buffer solution (0.2 M sodium phosphate, pH 7.0, 111.0 μL) containing 6.0 M Gdn-HCl was added H-Gly^{1}-Asn(glycan)^{32}-SH **93** (1.8 mg, 0.28 μmol, 1.0 equiv.) and the mixture was stirred at 23 ºC for 8 hours. The reaction progress was periodically monitored by RP-HPLC. After completion of the reaction, the mixture was quenched with a buffer solution (0.2 M sodium phosphate, pH 5.5, 1,0 mL) containing 6.0 M Gdn-HCl and 0.2 M DTT. Purification of the resultant buffer solution was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1% TFA aqueous solution: 0.1% TFA CH_3CN = 80 : 20 to 65 : 35 over 60 min at the flow rate of 1.0 mL/min to give H-Gly^{1}-Cys^{16,30}(SAcm)-Asn(glycan)^{32}-βVal^{33}-Cys^{38,49,52,}

^{70}(SAcm)-Asp^{71}-OH **94** (1.9 mg, 59%) as a white powder.

HRMS(ESI): m/z calc for $C_{498}H_{737}N_{116}O_{174}S_8$ $[M+H]^+$ 11389.4920, found for 11389.4967 (average isotopes).

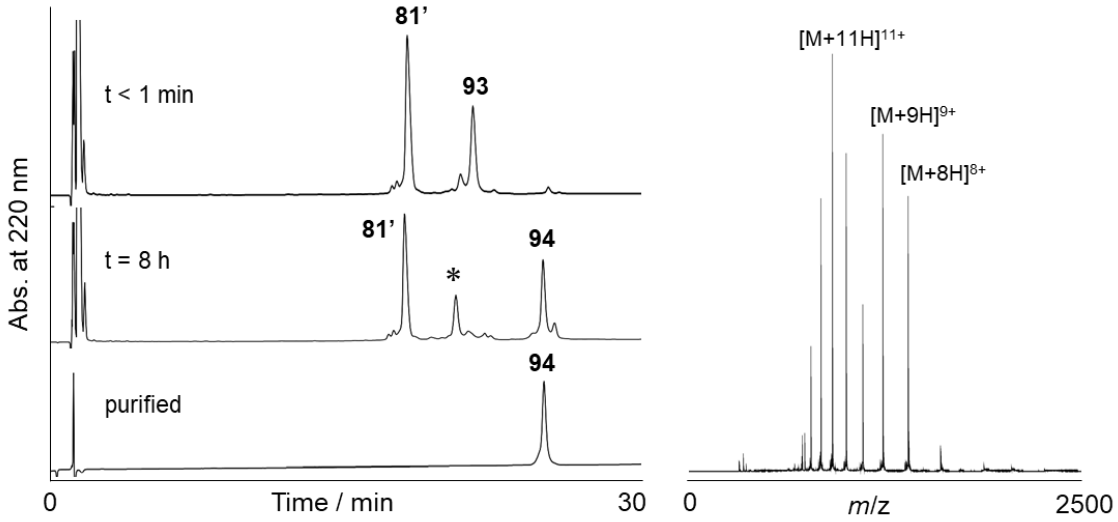


Figure E36. HPLC and MS monitoring of TCL between glycopeptide thioacids **93** and peptide **81**. The compound **81'** indicates a reductive form (having no disulfide at Val^{33}) of peptide **81**. The asterisk indicates an aspartimide and a hydrolyzed form of glycopeptide thioacid **93**.

H-Gly^1-Cys^{16,30}(SAcm)-Asn(glycan)^{32}-Val^{33}-Cys^{38,49,52,70}(SAcm)-Asp^{71}-OH 100 (Desulfurization of βVal^{33})

To a powder of H-Gly^1-Cys^{16,30}(SAcm)-Asn(glycan)^{32}-βVal^{33}-Cys^{38,49,52,70}(SAcm)-Asp^{71}-OH **94** (1.5 mg, 0.13 μmol) was dissolved in a buffer solution (0.2 M sodium phosphate, pH 7.0, 1.0 mL) containing TCEP (130 mg, 0.45 mmol), VA-044 (1.6 mg, 4.9 μmol), Mesna (50.0 mg, 0.31 mmol) and 6.0 M Gdn-HCl. The resulting mixture was stirred for 11 hours at room temperature. Purification of the product was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 80 : 20 to 65 : 35 over 60 min at the flow rate of 1.0 mL/min) afforded H-Gly^1-Cys^{16,30}(SAcm)-Asn(glycan)^{32}-Val^{33}-Cys^{38,49,52,70}(SAcm)-Asp^{71}-OH **100** (1.2 mg, 78%).

HRMS(ESI): m/z calc for $C_{498}H_{737}N_{116}O_{174}S_7$ $[M+H]^+$ 11357.4320, found for 11357.4333 (average isotopes).

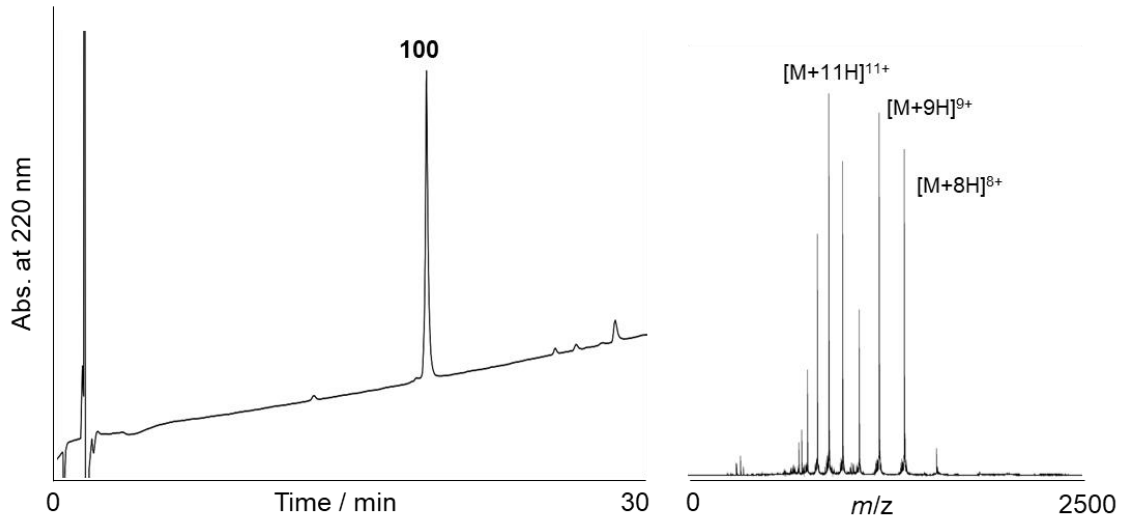


Figure E37. HPLC and MS spectrum of full-length glycopeptide **100**

H-Gly[1]- Asn(glycan)[32]- Asp[71]-OH 95

(Deprotection of Acm and Pac protecting group of glycan)

Palladium (II) chloride ($PdCl_2$, 3.1 mg) was dissolved in a buffer solution (0.2 M sodium phosphate, pH 7.0, 550 μL) containing 6.0 M Gdn-HCl. To a solution of H-Gly[1]-Cys[16,30](SAcm)-Asn(glycan)[32]-Val[33]-Cys[38,49,52,70](SAcm)-Asp[71]-OH **100** (1.2 mg, 0.11 μmol) in a buffer solution (0.2 M sodium phosphate, pH 7.0, 327 μL) containing 6.0 M Gdn-HCl was added above $PdCl_2$ solution (178.0 μL) at 30 ºC. The resultant mixture was stirred at 30 ºC for 1.5 hours. To the reaction mixture was added DTT (11.9 mg) and the resultant mixture immediately afforded red precipitate. To the mixture was added 2-mercaptoethanol (BME, 9.0 μL, 0.043 mmol), piperidine (15.0 μL, 0.051 mmol) and a buffer solution (0.2 M sodium phosphate, pH 7.0, 154.0 μL) containing 6.0 M Gdn-HCl at 30 ºC. The reaction solution changed to transparent and was further stirred for 1.5 hours. The reaction was quenched by adjusting pH 7.0 with 5.0 M hydrochloride solution. Purification of the product was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 80 : 20 to 65 : 35 over 60 min at the flow rate of 1.0 mL/min) afforded H-Gly[1]- Asn(glycan)[32]- Asp[71]-OH **95** (0.9 mg, 81%).
HRMS(ESI): m/z calc for $C_{464}H_{695}N_{110}O_{166}S_7$ $[M+H]^+$ 10694.6880, found for 10694.6891 (average isotopes).

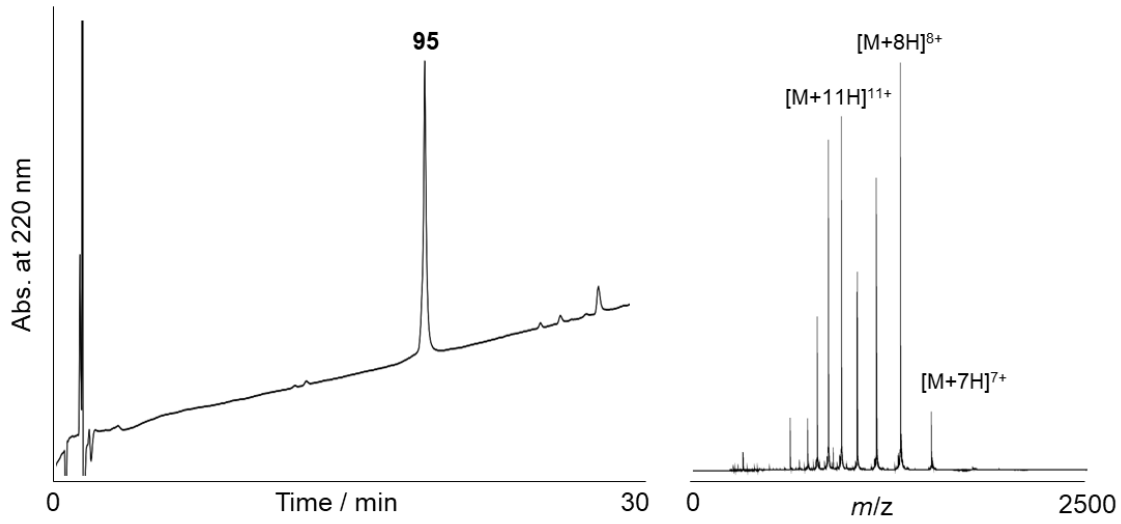


Figure E38. HPLC and MS spectrum of full-length glycopeptide **95.**

Glycosyl SPINK13 79

(Folding of H-Gly^{1}- Asn(glycan)^{32}- Asp^{71}-OH 95)

H-Gly^{1}- Asn(glycan)^{32}- Asp^{71}-OH **95** (0.9 mg, 0.08 μmol) was dissolved in a buffer solution (100 mM Tris-HCl, pH 7.5, 1.4 mL) containing 6.0 M Gdn-HCl. This solution was poured into the dialysis tubing, Spectra/Por®7 Dialysis Membrane Pre-treated RC Tubing (MWCO: 4.5 kD) and then dialyzed against the first folding buffer (100mM Tris-HCl, pH 8.5, 1.0 L) containing 3.0 M Gdn-HCl, 4.0 mM cysteine hydrochloride and 0.5 mM cystine and the solution was left for 16 hours at 4 ºC. Then the external buffer solution was replaced with the second folding buffer (100 mM Tris-HCl, pH 8.0, 1.0 L) containing 1.0 M Gdn-HCl and the solution was left for 11 hours at 4 ºC. Then the external buffer solution was discarded and changed to the third folding buffer (10 mM Tris-HCl, pH 8.0, 1.0 L) and the solution was left for 24 hours at 4 ºC. The reaction progress was periodically monitored by RP-HPLC. Purification of the product was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 80 : 20 to 65 : 35 over 60 min at the flow rate of 1.0 mL/min) afforded folded glycosylated SPINK13 **95** (0.5 mg, 59%).

HRMS(ESI): m/z calc for $C_{464}H_{689}N_{110}O_{166}S_7$ $[M+H]^+$ 10688.6400, found for 10688.6459 (average isotopes).

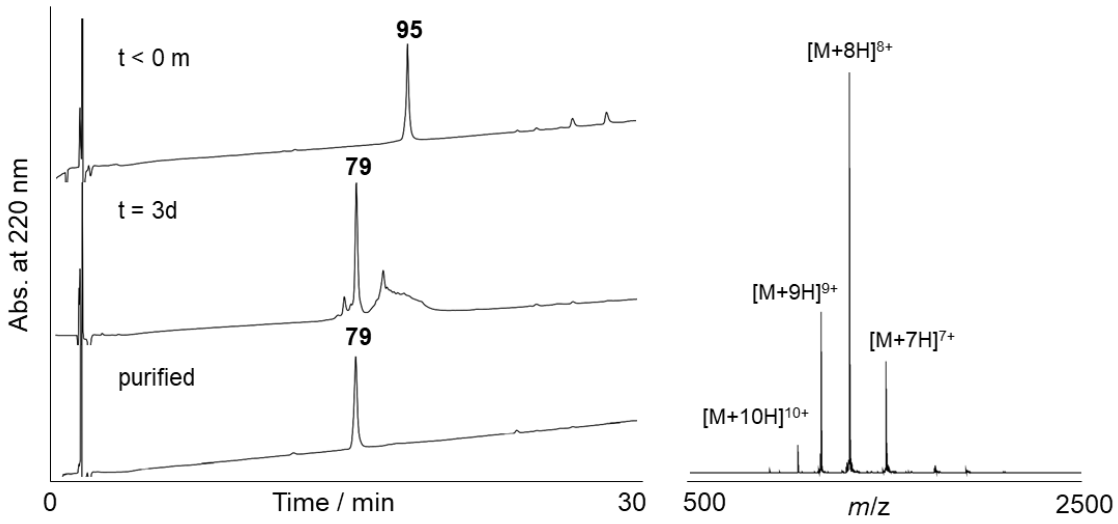


Figure E39. HPLC monitoring of folding of glycopeptide **95** and folded SPINK13 **79**. MS spectrum of folded SPINK13 **79.**

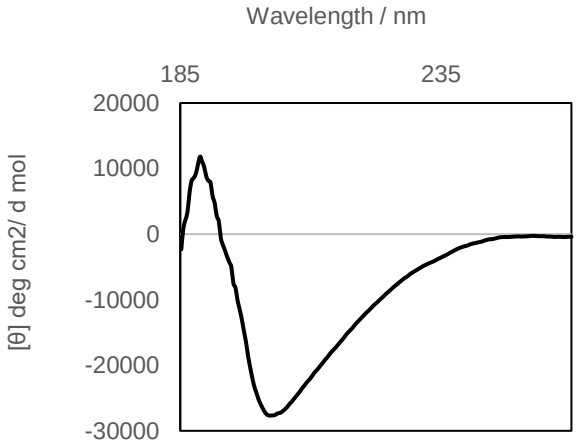


Figure E40. CD spectrum of folded SPINK13 **79.**

Synthesis of non-glycosylated SPINK13

H-Gly^{1}-Asn^{32}-SH 101

The non-glycosylated SPINK13 segment : H_2N-GIFNKRDFTRWPKPRC(Acm)KMYIPLDPDYNA DC(Acm)PN-COSH **101** was synthesized on Dawson AM resin (100 μmol scale) by Fmoc SPPS in the same manner of the synthesis of H-Gly^{1}-Pro^{31}-SH **80**. After swelling a resin by DCM and DMF for 1 h, to a resin (100.0 μmol, 1.0 equiv) was added 20% piperidine / DMF (5.0 mL) for the removal of the Fmoc groups and stirred for 7 min two times. All reagents were filtered, and the resultant resin was washed with DCM and DMF. Then Fmoc-L-Asp(O^{t}Bu)-OH (358 mg, 0.60 mmol) was pre-activated with HCTU (248 mg, 0.60 mmol) and DIPEA, (157 μL, 0.90 mmol) in anhydrous DMF (2.0 mL) for 1 min and then the activated amino acid was added to a resin. The suspension was gently agitated for 1 h at room temperature. After the coupling of a first amino acid, the resin was washed with DCM and DMF. From next amino acid residues, the rapid flow-based peptide protocols[3] was performed. After all coupling steps were complete, the fully protected dipeptide was cleaved with a cocktail containing TFA/H2O/TIPS (95:2.5:2.5, 7.0 mL), then the reaction mixture was allowed at room temperature and kept for 1 hour. After cleavage, the resultant solution was filtered and evaporated *in vacuo* to give a colorless oil. Purification of the resultant solution including H-Gly^{1}-Asn^{32}-Dbz **102** was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O: 0.1 % TFA in CH_3CN = 80 : 20 to 65 : 35 over 90 min at the flow rate of 2.5 mL/min) afforded H-Gly^{1}-Asn^{32}-Dbz **102** (61.6 mg, 15 %). HRMS(ESI): m/z calc for $C_{187}H_{278}N_{53}O_{49}S_3$ $[M+H]^+$ 4146.0053, found for 4146.0050.

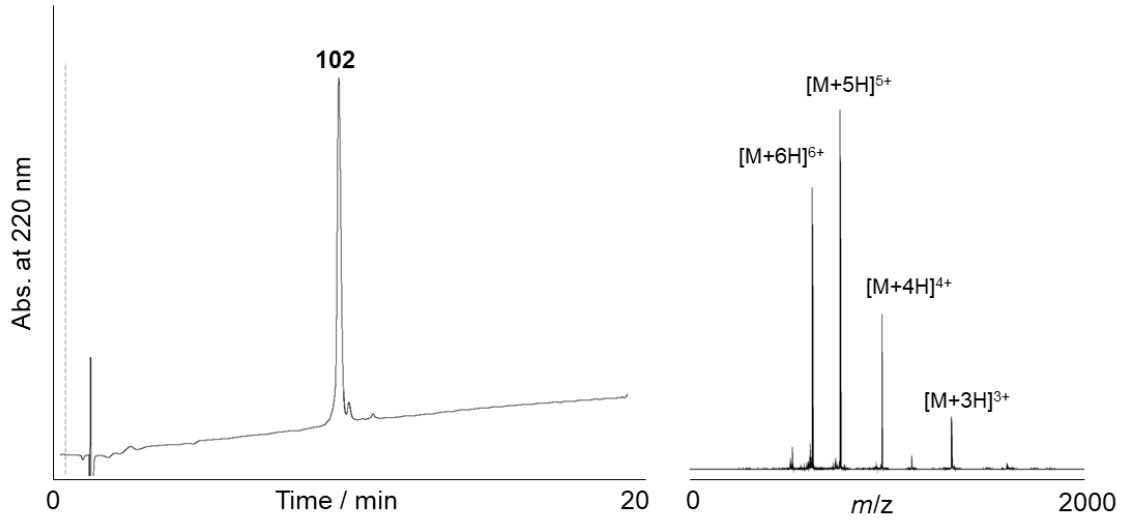


Figure E41. LC and MS spectrum of compound **102.**

To peptide H-Gly^{1}-Asn^{32}-Dbz **102** (22.5 mg, 5.43 μmol) was added 0.2 M sodium phosphate buffer (pH 3.0) containing 6 M Gdn-HCl (1.5 mL) and the reaction mixture was cooled at -10 ˚C. To the

resultant solution was added 0.4 M $NaNO_2$ aqueous solution (77 μL, 6.0 eq) and subsequently the reaction was stirred for 5 min at -10 °C to afford peptide benzotriazole intermediates, followed by the addition of 0.2 M sodium phosphate buffer (pH 7.5) containing 6 M Gdn-HCl and 0.2 M Na_2S (1.5 mL, 60.0 eq). After stirring for 30 min at room temperature. Purification of the resultant solution was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O: 0.1 % TFA in CH_3CN = 80 : 20 to 65 : 35 over 90 min at the flow rate of 2.5 mL/min) afforded H-Gly[1]-Asn[32]-SH **101** (4.4 mg, 20 %). HRMS(ESI): m/z calc for $C_{180}H_{270}N_{50}O_{48}S_4$ $[M+H]^+$ 4027.9106, found for 4027.9312.

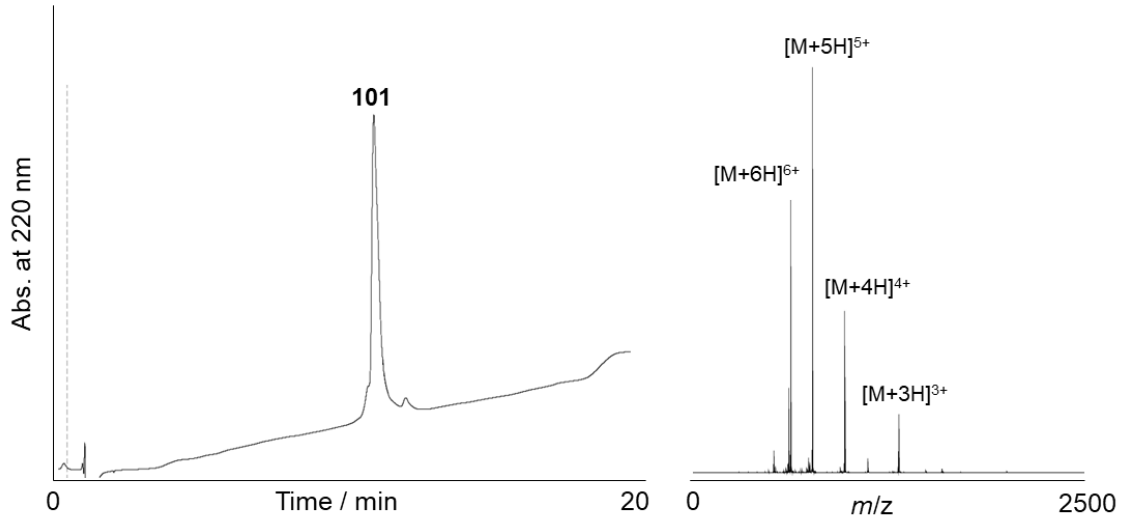


Figure E42. LC and MS spectrum of compound **101.**

H-Gly[1]-Cys[16,30](SAcm)-Asn[32]-βVal[33]-Cys[38,49,52, 70](SAcm)-Asp[71]-OH 103

To a solution of H-βVal[33](Npys)-Asp[71]-OH **81** (6.8 mg, 1.3 μmol, 2.0 equiv.) in a buffer solution (0.2 M sodium phosphate, pH 7.0, 111.0 μL) containing 6.0 M Gdn-HCl was added H-Gly[1]-Asn[32]-SH **101** (2.1 mg, 0.51 μmol, 1.0 equiv.) and the mixture was stirred at 23 ºC for 12 hours. The reaction progress was periodically monitored by RP-HPLC. After completion of the reaction, the mixture was quenched with a buffer solution (0.2 M sodium phosphate, pH 7.0, 1,0 mL) containing 6.0 M Gdn-HCl and 0.2 M TCEP. Purification of the resultant buffer solution was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1% TFA aqueous solution: 0.1% TFA CH_3CN = 80 : 20 to 65 : 35 over 90 min at the flow rate of 2.5 mL/min to give H-Gly[1]-Cys[16,30](SAcm)-Asn[32]-βVal[33]-Cys[38,49,52, 70](SAcm)-Asp[71]-OH **103** (1.5 mg, 32%) as a white powder.

HRMS(ESI): m/z calc for $C_{398}H_{588}N_{110}O_{111}S_8$ $[M+H]^+$ 8946.2210, found for 8946.2248 (average isotopes).

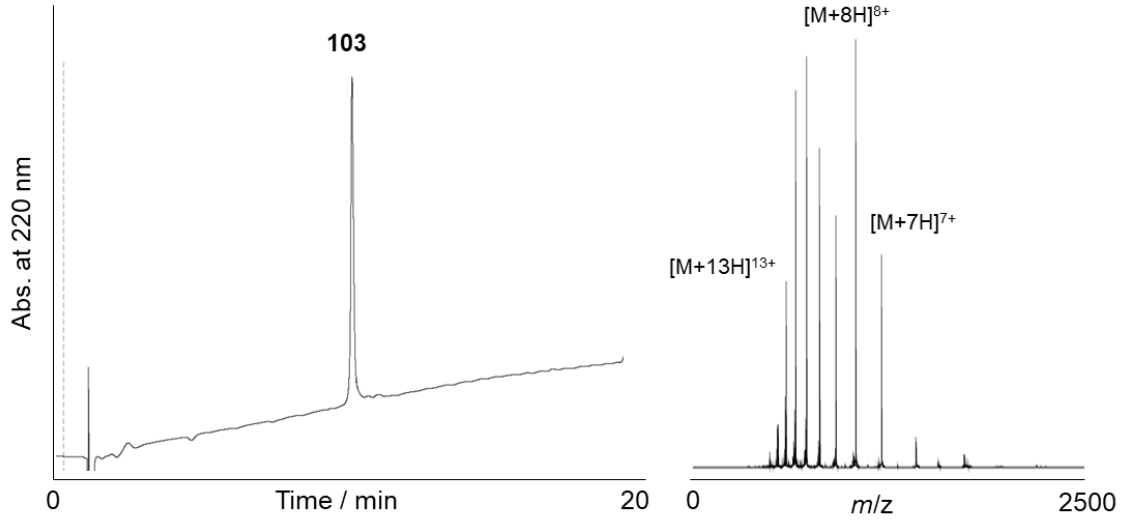


Figure E43. LC and MS spectrum of compound **103**

H-Gly[1]-Cys[16,30](SAcm)-Asn[32]-Val[33]-Cys[38,49,52,70](SAcm)-Asp[71]-OH 104
(Desulfurization of βVal[33])

To a powder of H-Gly[1]-Cys[16,30](SAcm)-Asn[32]-βVal[33]-Cys[38,49,52,70](SAcm)-Asp[71]-OH **103** (3.0 mg, 0.34 μmol) was dissolved in a buffer solution (0.2 M sodium phosphate, pH 7.0, 2.0 mL) containing TCEP (130 mg, 0.45 mmol), VA-044 (1.6 mg, 4.9 μmol), Mesna (50.0 mg, 0.31 mmol) and 6.0 M Gdn-HCl. The resulting mixture was stirred for 24 hours at room temperature. Purification of the product was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 80 : 20 to 65 : 35 over 90 min at the flow rate of 2.5 mL/min) afforded H-Gly[1]-Cys[16,30](SAcm)-Asn[32]-Val[33]-Cys[38,49,52,70](SAcm)-Asp[71]-OH **104** (1.3 mg, 43%).

HRMS(ESI): m/z calc for $C_{398}H_{588}N_{110}O_{111}S_7$ $[M+H]^+$ 8914.1610, found for 8914.1622 (average isotopes).

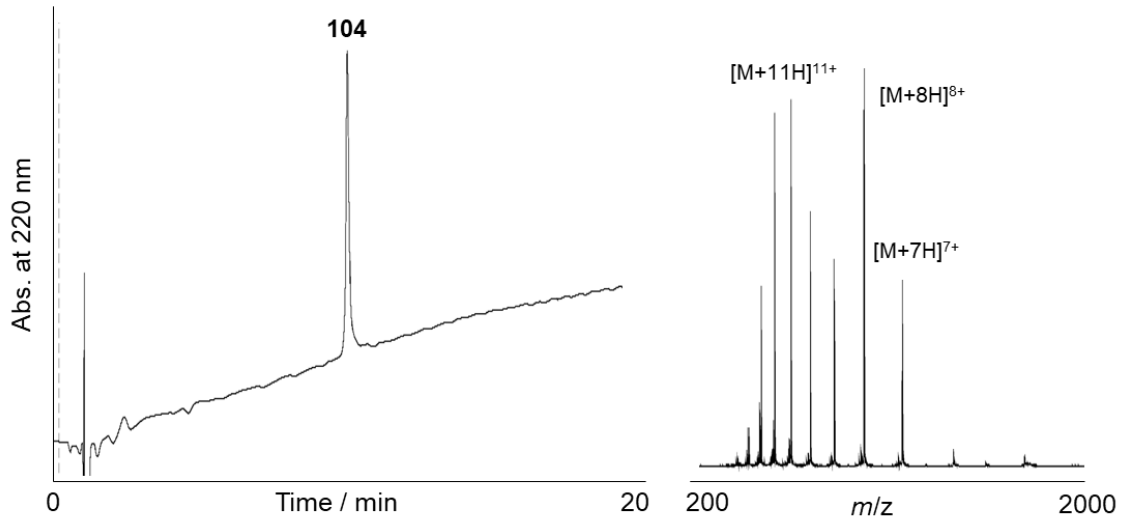


Figure E44. LC and MS spectrum of compound **104**

H-Gly[1]- Asn[32]- Asp[71]-OH 105

(Deprotection of Acm protecting group)

Palladium (II) chloride ($PdCl_2$, 4.0 mg) was dissolved in a buffer solution (0.2 M sodium phosphate, pH 7.0, 650 μL) containing 6.0 M Gdn-HCl. To a solution of H-Gly[1]-Cys[16,30](SAcm)-Asn[32]-Val[33]-Cys[38,49,52,70](SAcm)-Asp[71]-OH **104** (1.3 mg, 0.15 μmol) in a buffer solution (0.2 M sodium phosphate, pH 7.0, 464 μL) containing 6.0 M Gdn-HCl was added above $PdCl_2$ solution (252.0 μL) at 30 ºC. The resultant mixture was stirred at 30 ºC for 1.5 hours. To the reaction mixture was added DTT (17.0 mg) and the resultant mixture immediately afforded red precipitate. To the mixture was added piperidine (15.0 μL, 0.051 mmol) to dissolve the precipitate. The reaction solution subsequently changed to transparent. The reaction was quenched by adjusting pH 7.0 with 5 M HCl. Purification of the product was performed by RP-HPLC (Proteonavi Φ10 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 80 : 20 to 65 : 35 over 90 min at the flow rate of 2.5 mL/min) afforded H-Gly[1]- Asn[32]- Asp[71]-OH **105** (1.0 mg, 80%).

HRMS(ESI): m/z calc for $C_{380}H_{558}N_{104}O_{105}S_7$ $[M+H]^+$ 8487.6870, found for 8487.6851 (average isotopes).

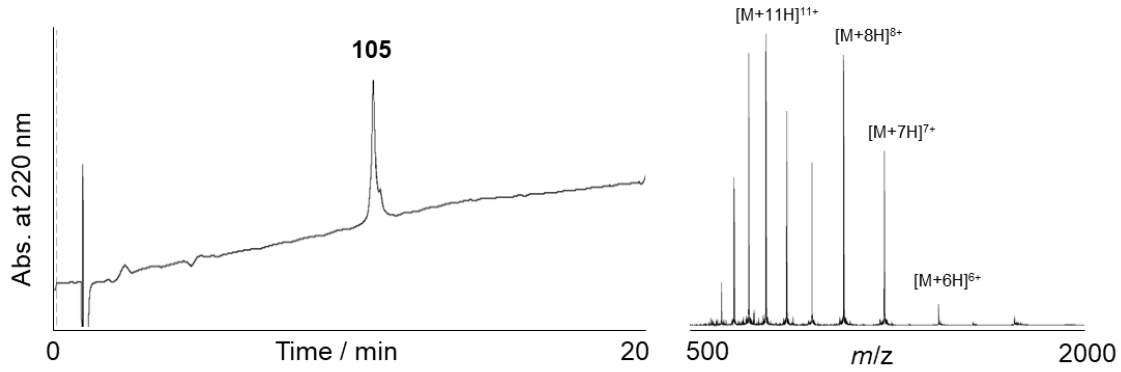


Figure E45. LC and MS spectrum of compound **105.**

Non-glycosyl SPINK13 106

(Folding of H-Gly^{1}- Asn^{32}- Asp^{71}-OH 105)

H-Gly^{1}- Asn^{32}- Asp^{71}-OH **105** (1.0 mg, 0.12 μmol) was dissolved in a buffer solution (100 mM Tris-HCl, pH 7.5, 2.0 mL) containing 6.0 M Gdn-HCl. This solution was poured into the dialysis tubing, Spectra/Por®7 Dialysis Membrane Pre-treated RC Tubing (MWCO: 4.5 kD) and then dialyzed against the first folding buffer (100mM Tris-HCl, pH 8.5, 1.0 L) containing 3.0 M Gdn-HCl, 4.0 mM cysteine hydrochloride and 0.5 mM cystine and the solution was left for 12 hours at 4 ºC. Then the external buffer solution was replaced with the second folding buffer (100 mM Tris-HCl, pH 8.0, 1.0 L) containing 1.0 M Gdn-HCl and the solution was left for 12 hours at 4 ºC. Then the external buffer solution was discarded and changed to the third folding buffer (10 mM Tris-HCl, pH 8.0, 1.0 L) and the solution was left for 24 hours at 4 ºC. The reaction progress was periodically monitored by RP-HPLC. Purification of the product was performed by RP-HPLC (Proteonavi Φ4.6 × 250 mm, 0.1 % TFA in H_2O : 0.1 % TFA in CH_3CN = 80 : 20 to 65 : 35 over 60 min at the flow rate of 1.0 mL/min) afforded folded non-glycosylated SPINK13 **106** (0.3 mg, 60%).

HRMS(ESI): m/z calc for $C_{380}H_{552}N_{104}O_{105}S_7$ $[M+H]^+$ 8481.6390, found for 8481.6393 (average isotopes).

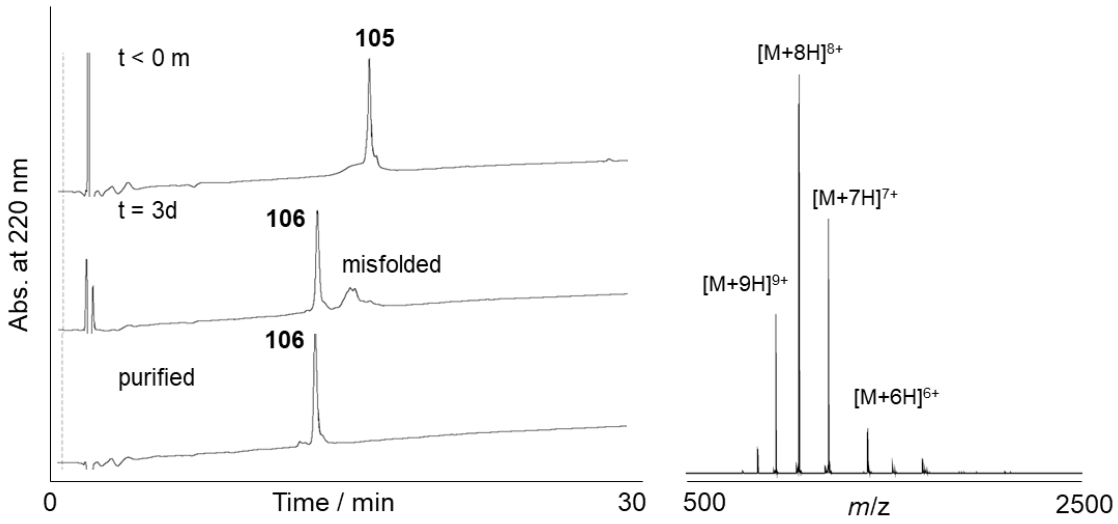


Figure E46. HPLC monitoring of folding of glycopeptide **105** and folded non-glycosylated SPINK13 **106**. MS spectrum of folded SPINK13 **106.**

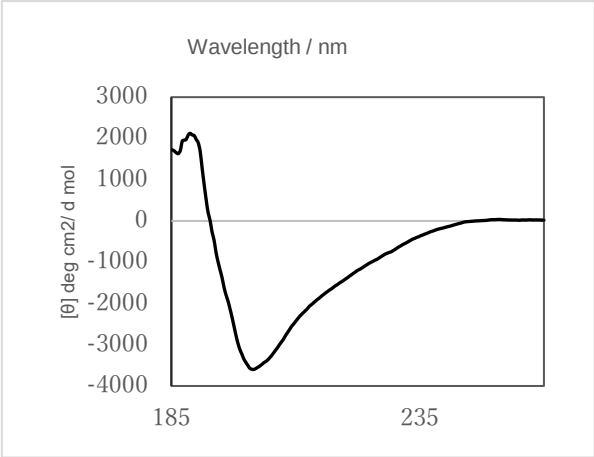


Figure E40. CD spectrum of folded SPINK13 **106.**

References of SI

(1) Murakami, M.; Kiuchi, T.; Nishihara, M.; Tezuka, K.; Okamoto, R.; Izumi, M.; Kajihara, Y. *Sci Adv.* **2016**, *2*, e1500678.

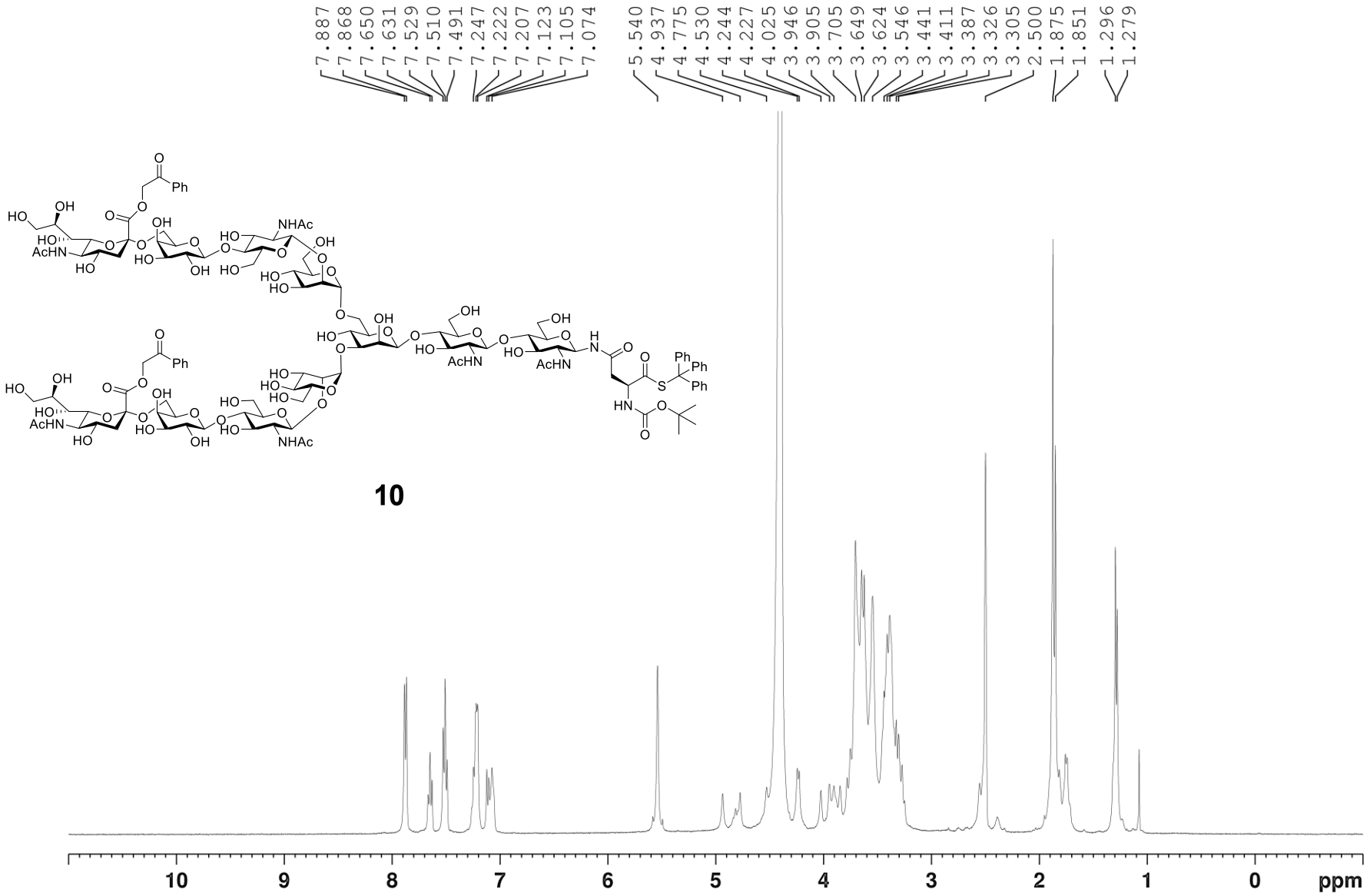
(2) a) Izumi, M.; Murakami, M.; Okamoto, R.; Kajihara, Y. *J Pept Sci.* **2014**, *20*, 98-101. b) Gates, Z. P.; Dhayalan, B.; Kent, S. B. *Chem Commun* **2016,** *52*, 13979-13982.

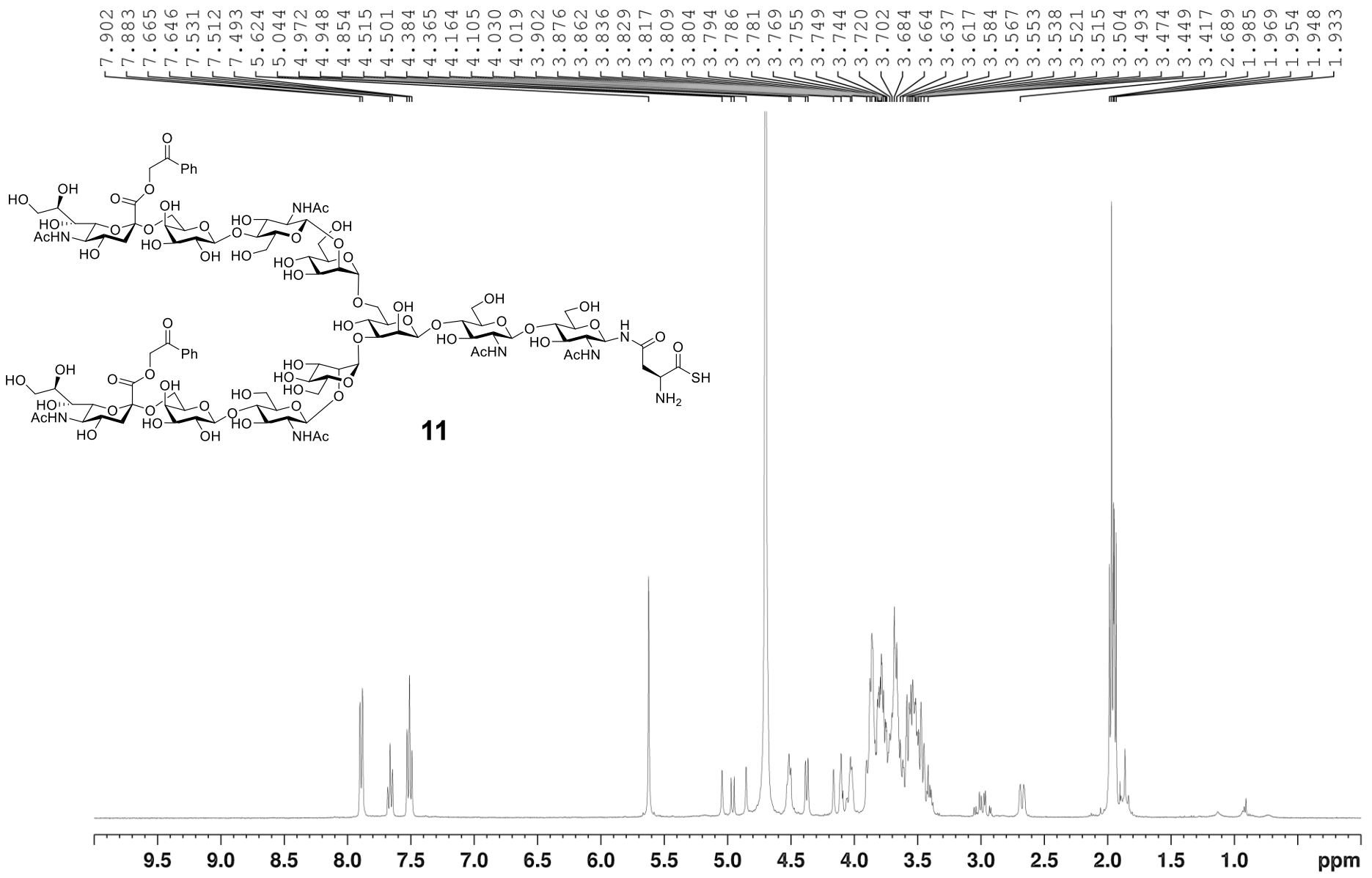
(3) Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. A.; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A. M.; Jensen, K. F.; Pentelute, B. L. *Chembiochem*. **2014**, *15*, 713-720.

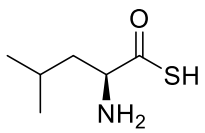
(4) Katayama, H.; Morisue, S. *Tetrahedron* **2017**, *73*, 3541-3547.

(5) Schweiger, A.; Stern, D.; Lohman, I. C.; Baldini, M.; Martinez, F. D.; Halonen, M. *J. Allergy Clin. Immunol.* **2001**, *107*, 505-510.

(6) Kajihara, Y.; Yoshihara, I.; Hirano, K.; Yamamoto, N. *Carbohydr. Res.* **2006**, *341*, 1333-40.

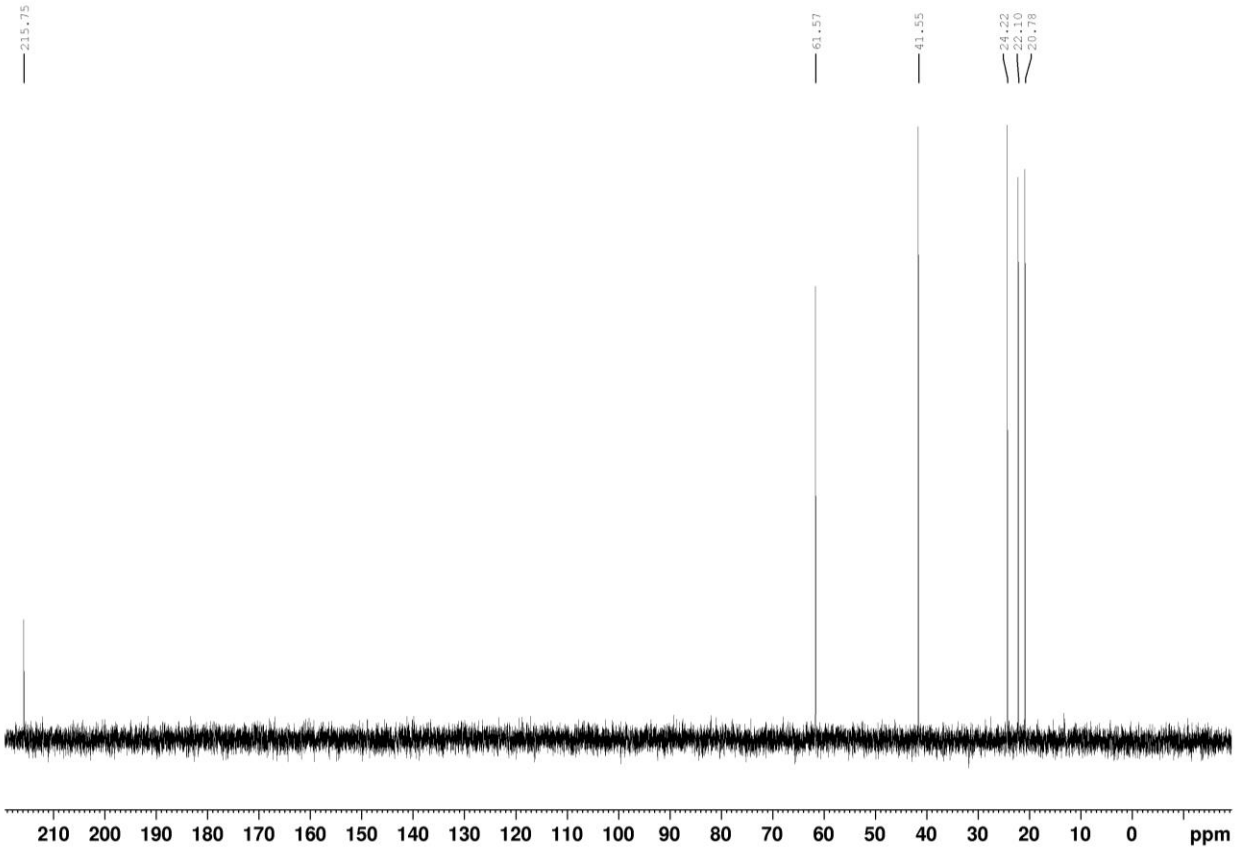

7.887
7.868
7.650
7.631
7.529
7.510
7.491
7.247
7.222
7.207
7.123
7.105
7.074
5.540
4.937
4.775
4.530
4.244
4.227
4.025
3.946
3.905
3.705
3.649
3.624
3.546
3.441
3.411
3.387
3.326
3.305
2.500
1.875
1.851
1.296
1.279
10
10
9
8
7
6
5
4
3
2
1
0
ppm

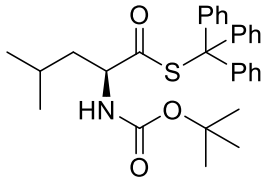

7.902
7.883
7.665
7.646
7.531
7.512
7.493
5.624
5.044
4.972
4.948
4.854
4.515
4.501
4.384
4.365
4.164
4.105
4.030
4.019
3.902
3.876
3.862
3.836
3.829
3.817
3.809
3.804
3.794
3.786
3.781
3.769
3.755
3.749
3.744
3.720
3.702
3.684
3.664
3.637
3.617
3.584
3.567
3.553
3.538
3.521
3.515
3.504
3.493
3.474
3.449
3.417
2.689
1.985
1.969
1.954
1.948
1.933
11
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
ppm


O
SH
NH2

13l

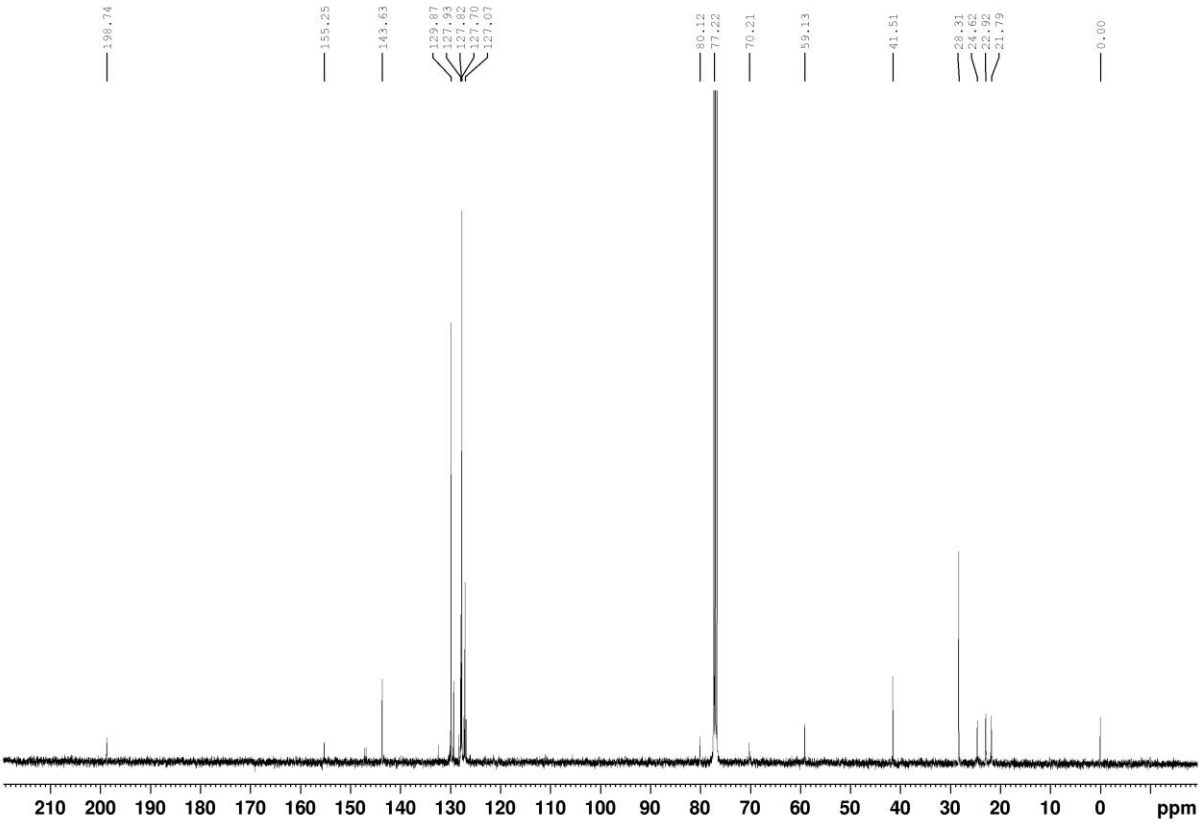
4.666
3.871
3.858
3.849
3.837
1.808
1.796
1.788
1.775
1.762
1.754
1.740
1.653
1.636
1.620
1.605
1.586
1.570
1.557
1.548
1.536
1.523
1.514
1.500
0.859
0.845
8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm
1.00
1.03
2.13
5.99

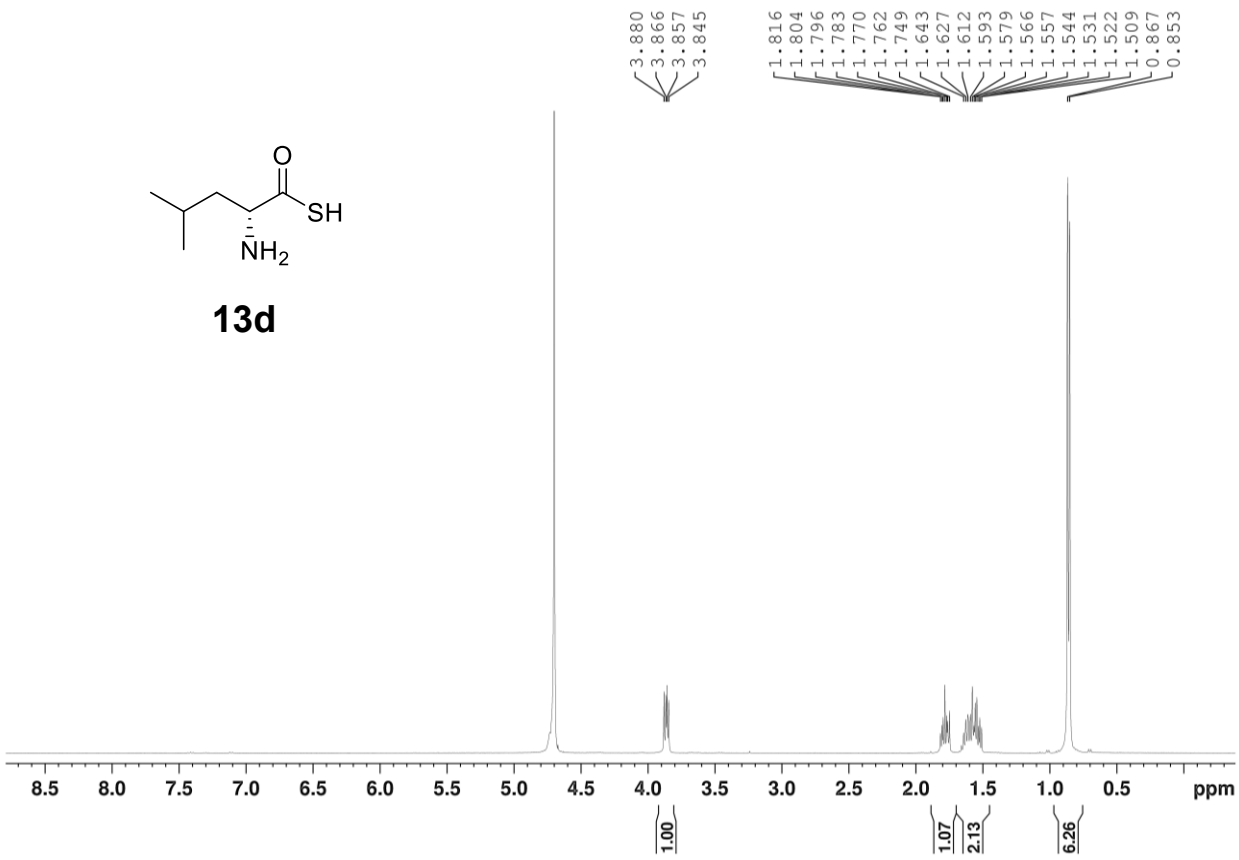

215.75
61.57
41.55
24.22
22.10
20.78
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

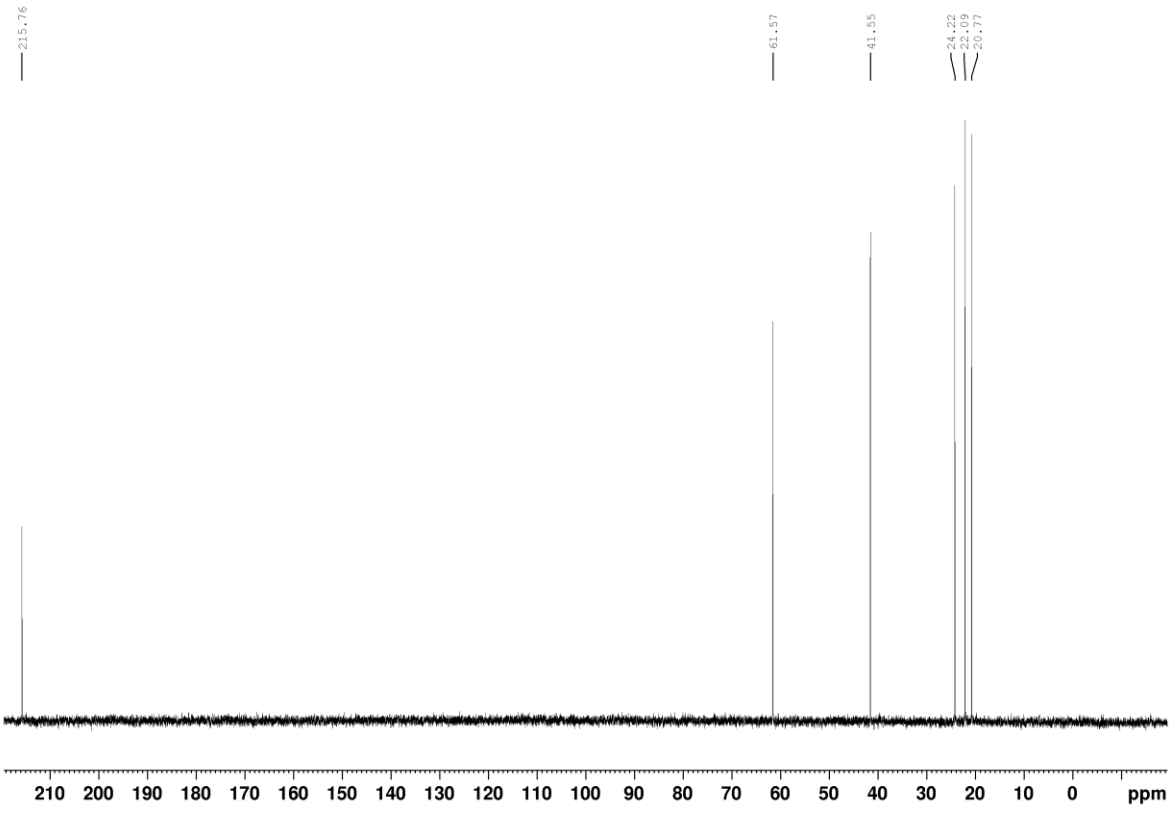


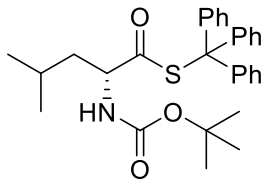
109

7.242
4.758
4.737
4.332
4.309
4.297
4.275
1.554
1.448
1.317
1.304
1.294
1.283
1.260
0.867
13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 ppm
1.00
1.00
1.60
12.24
1.61
7.40


198.74
155.25
143.63
129.87
127.93
127.82
127.70
127.07
80.12
77.22
70.21
59.13
41.51
28.31
24.62
22.92
21.79
0.00
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

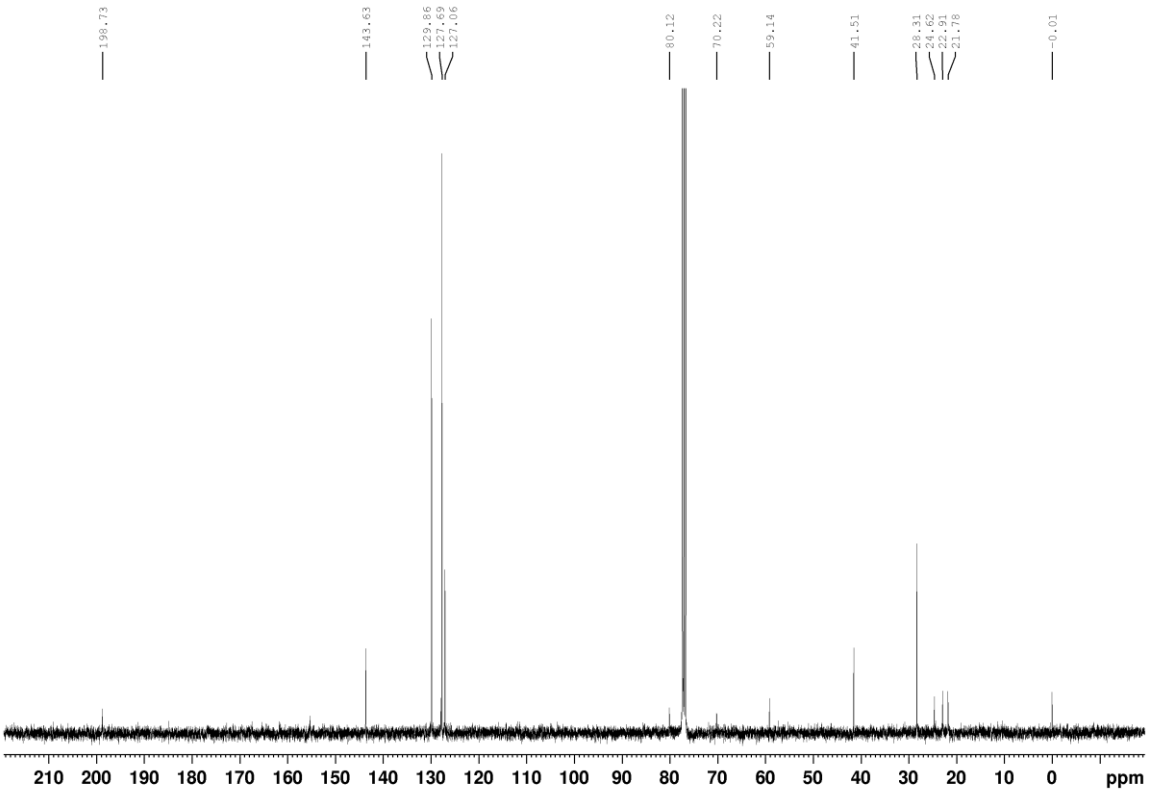

3.880
3.866
3.857
3.845
1.816
1.804
1.796
1.783
1.770
1.762
1.749
1.643
1.627
1.612
1.593
1.579
1.566
1.557
1.544
1.531
1.522
1.509
0.867
0.853
O
SH
NH2
13d
8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm
1.00
1.07
2.13
6.26

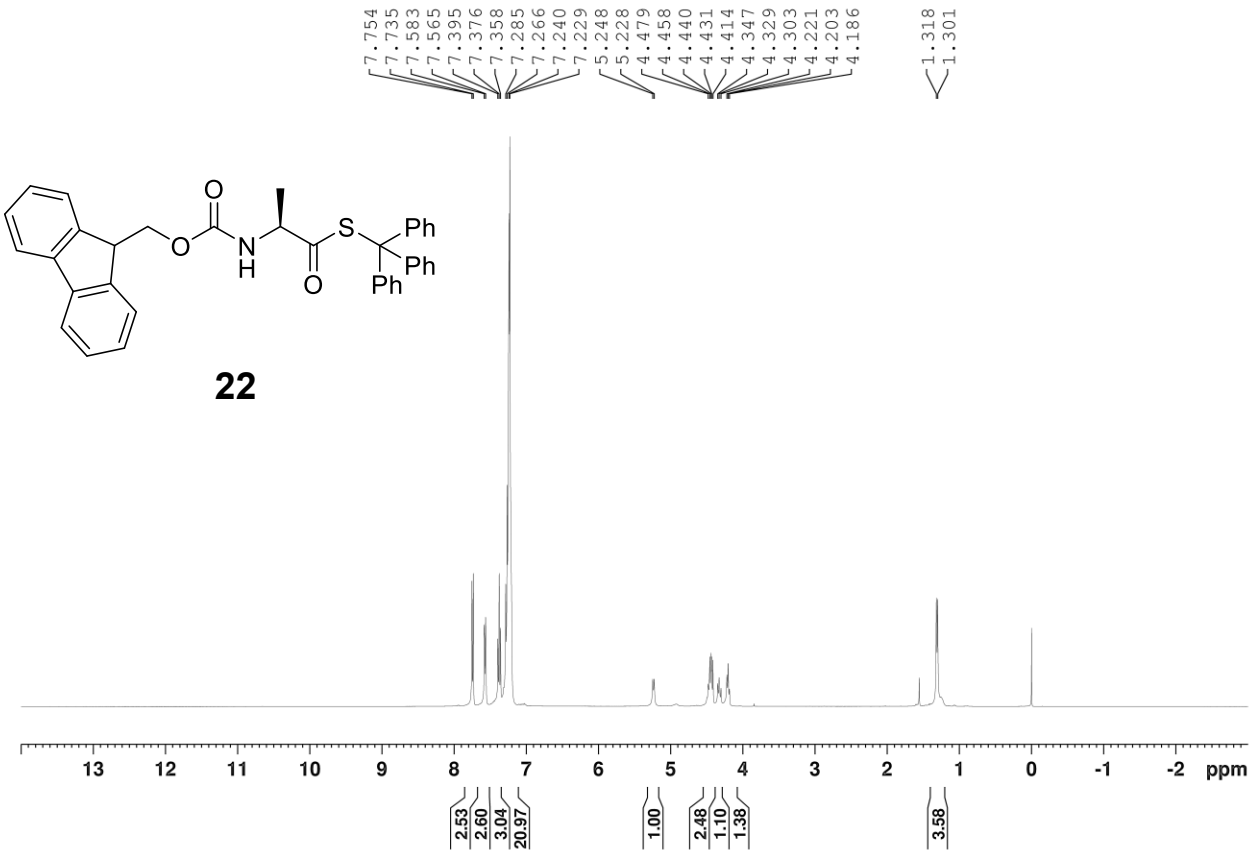

215.76
61.57
41.55
24.22
22.09
20.77
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm



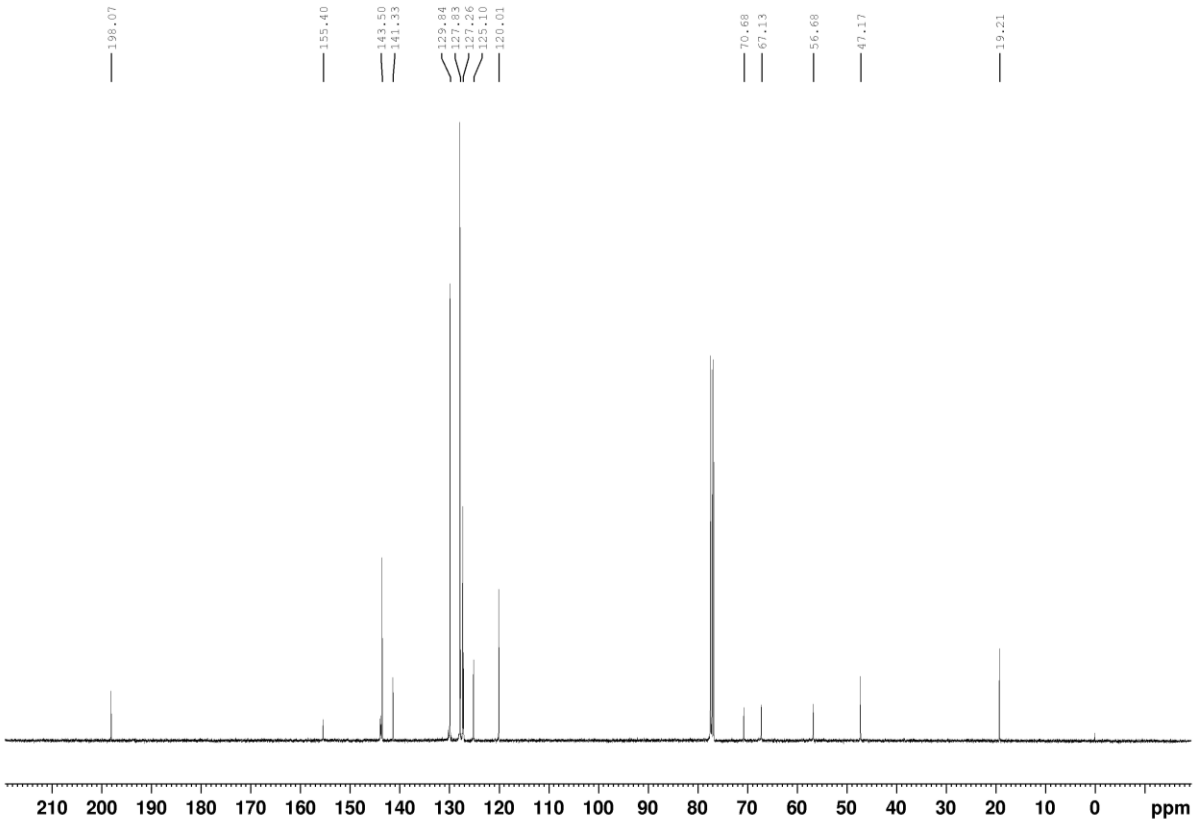
110

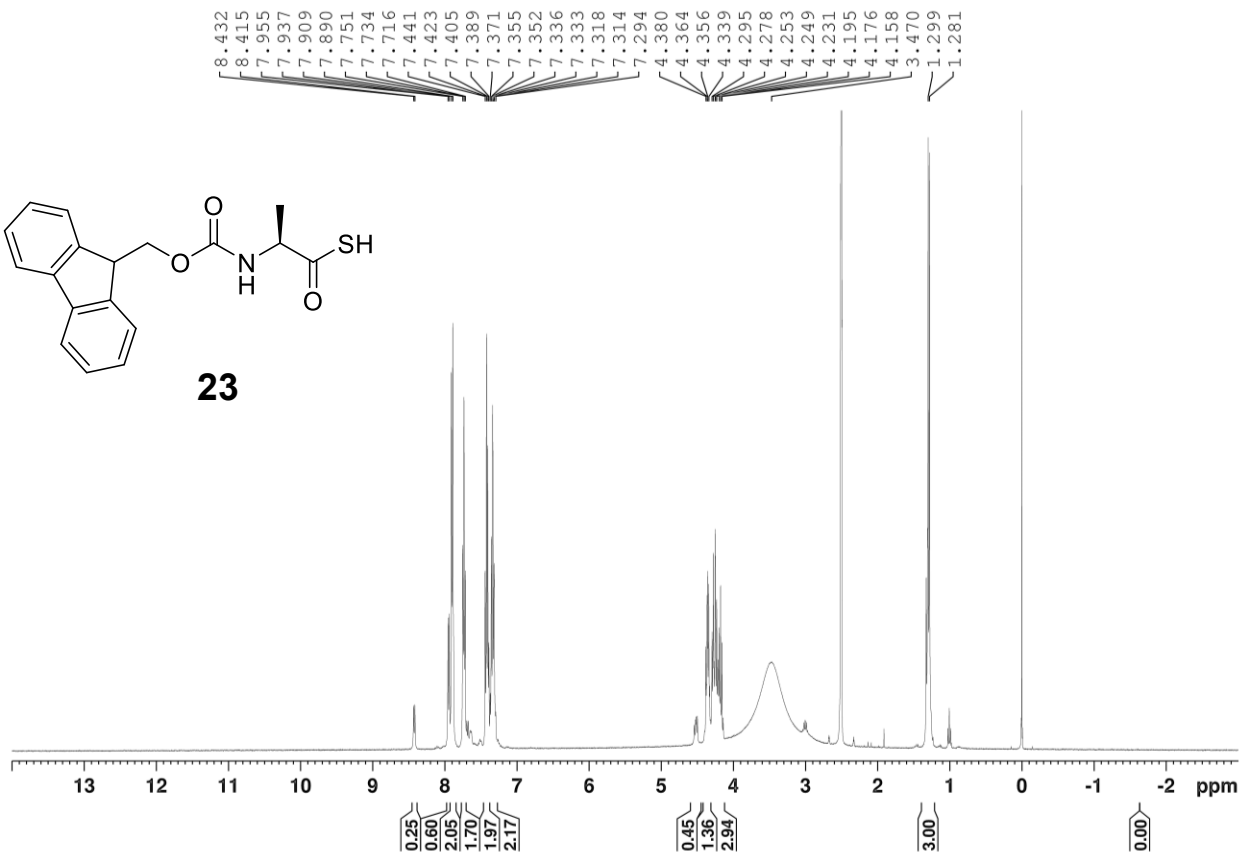
4.782
4.761
4.355
4.332
4.320
4.297
1.581
1.471
1.342
1.328
1.318
1.307
1.297
1.285
0.906
0.890
0.021
13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 ppm
1.00
1.04
2.58
10.98
1.48
7.53

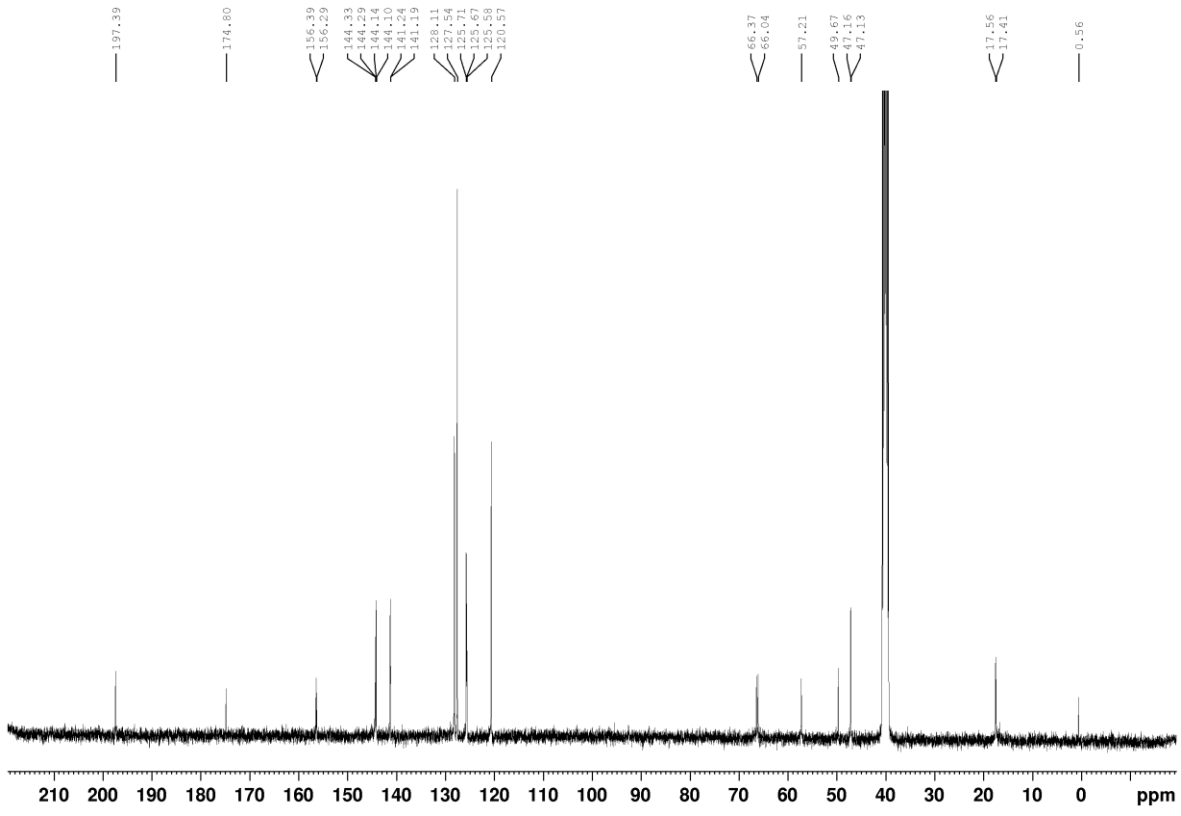

198.73
143.63
129.86
127.69
127.06
80.12
70.22
59.14
41.51
28.31
24.62
22.91
21.78
-0.01
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

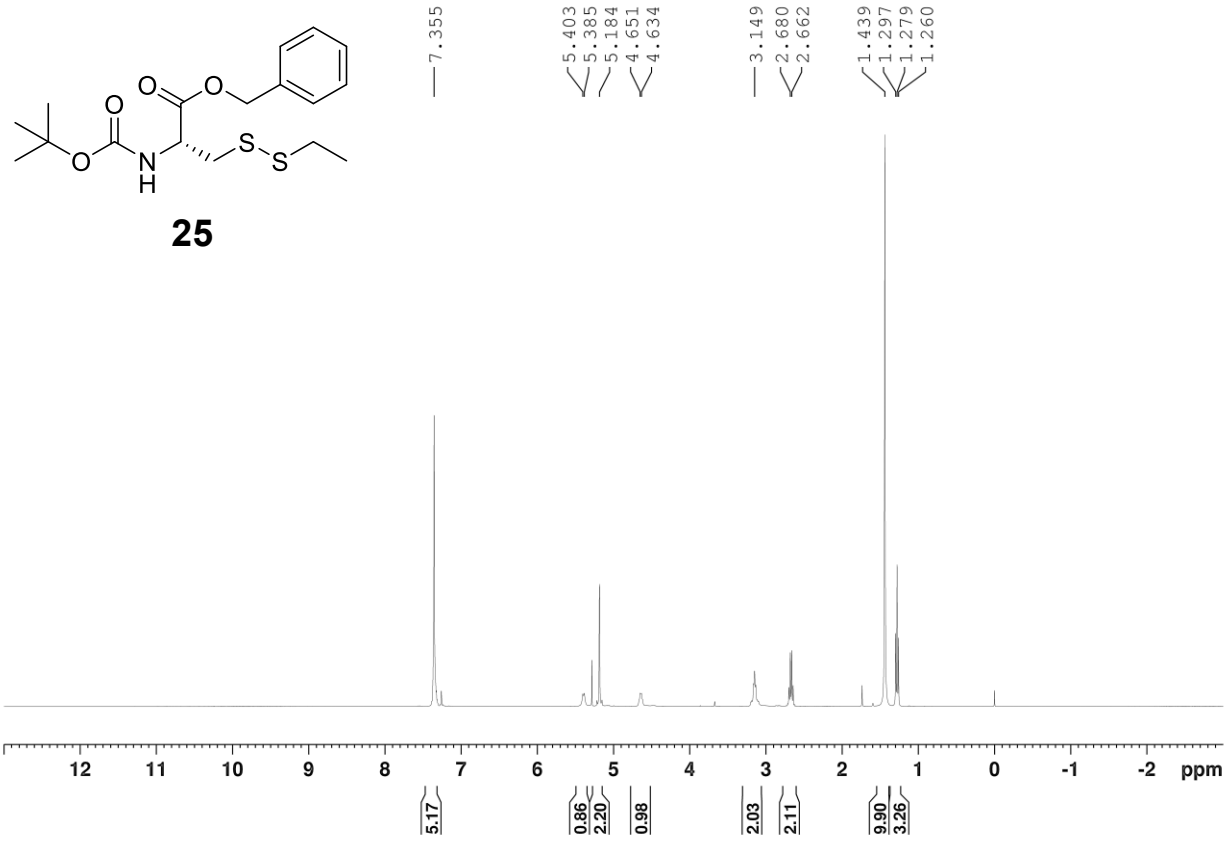

7.754
7.735
7.583
7.565
7.395
7.376
7.358
7.285
7.266
7.240
7.229
5.248
5.228
4.479
4.458
4.440
4.431
4.414
4.347
4.329
4.303
4.221
4.203
4.186
1.318
1.301
O
O
N
H
S
Ph
Ph
Ph
O
13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 ppm
2.53
2.60
3.04
20.97
1.00
2.48
1.10
1.38
3.58

22

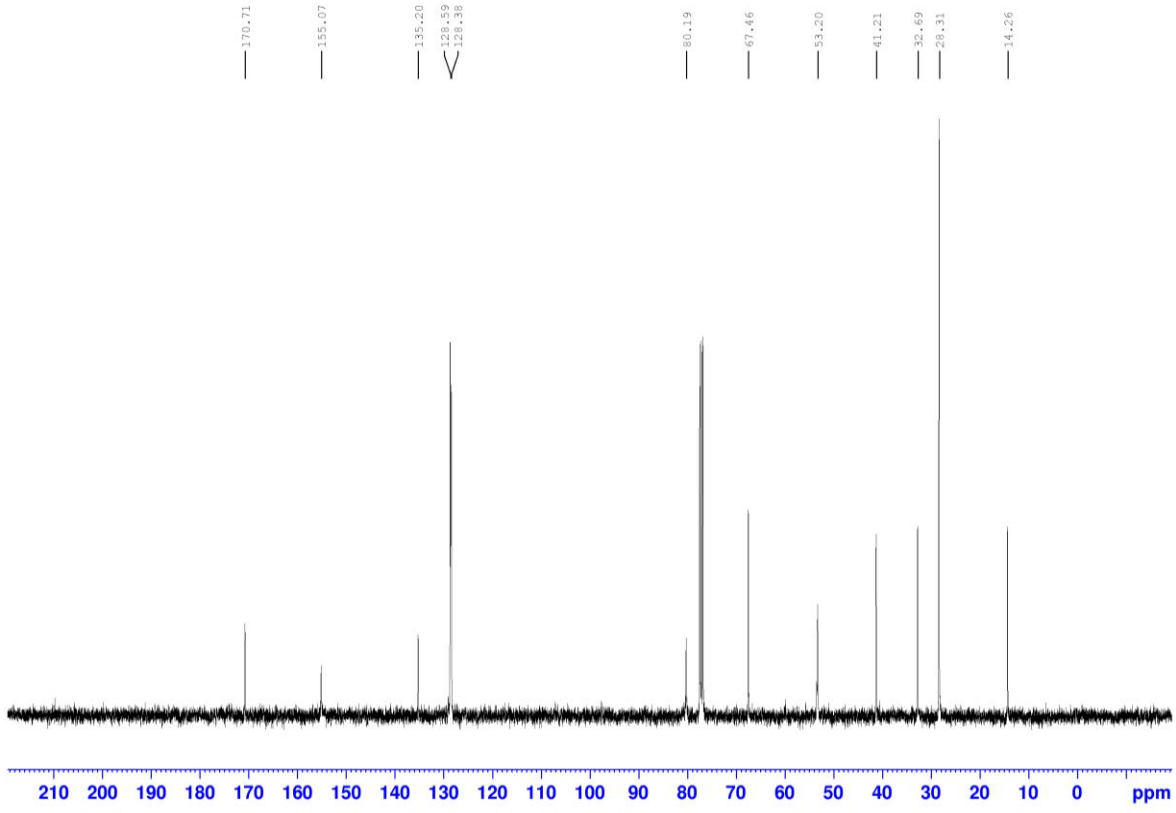

198.07
155.40
143.50
141.33
129.84
127.83
127.26
125.10
120.01
70.68
67.13
56.68
47.17
19.21
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

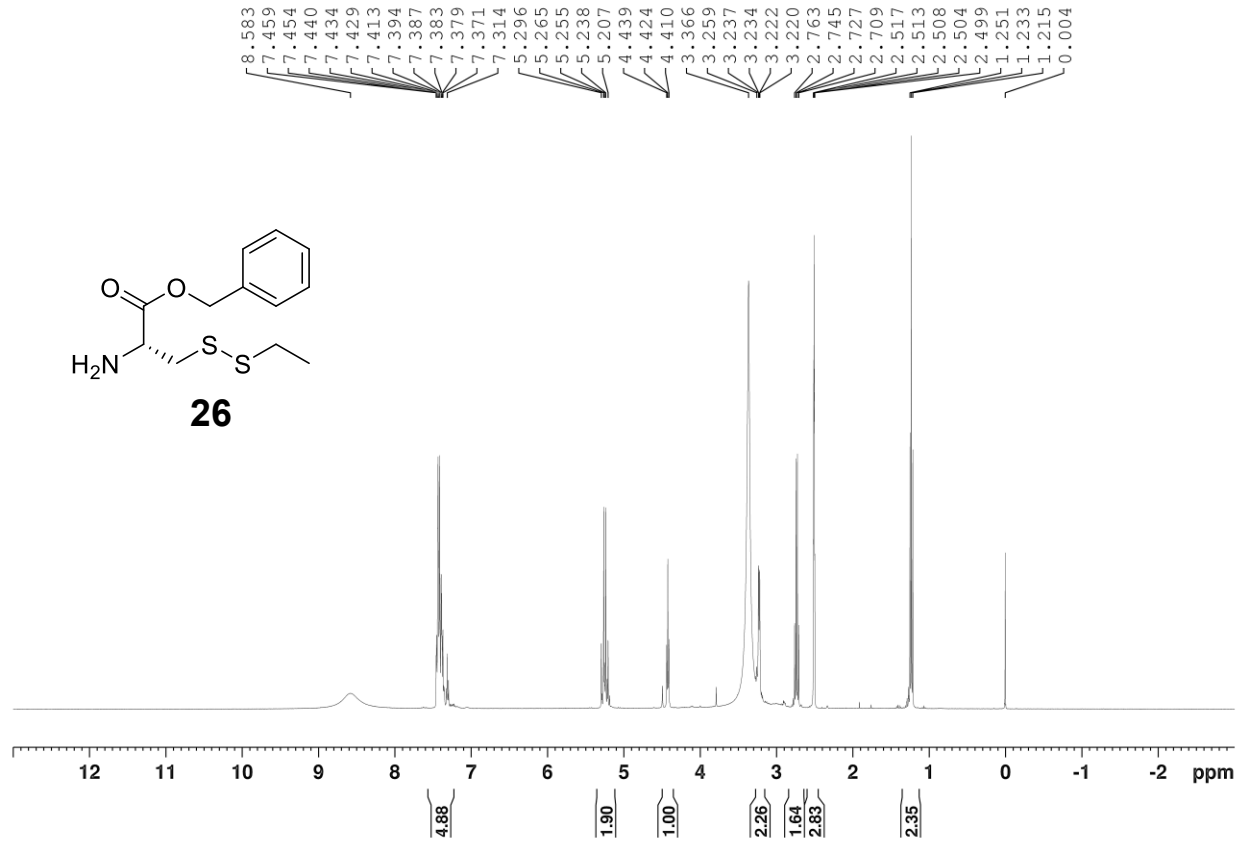

8.432
8.415
7.955
7.937
7.909
7.890
7.751
7.734
7.716
7.441
7.423
7.405
7.389
7.371
7.355
7.352
7.336
7.333
7.318
7.314
7.294
4.380
4.364
4.356
4.339
4.295
4.278
4.253
4.249
4.231
4.195
4.176
4.158
3.470
1.299
1.281
O
O
N
H
SH
O
23
13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 ppm
0.25
0.60
2.05
1.70
1.97
2.17
0.45
1.36
2.94
3.00
0.00

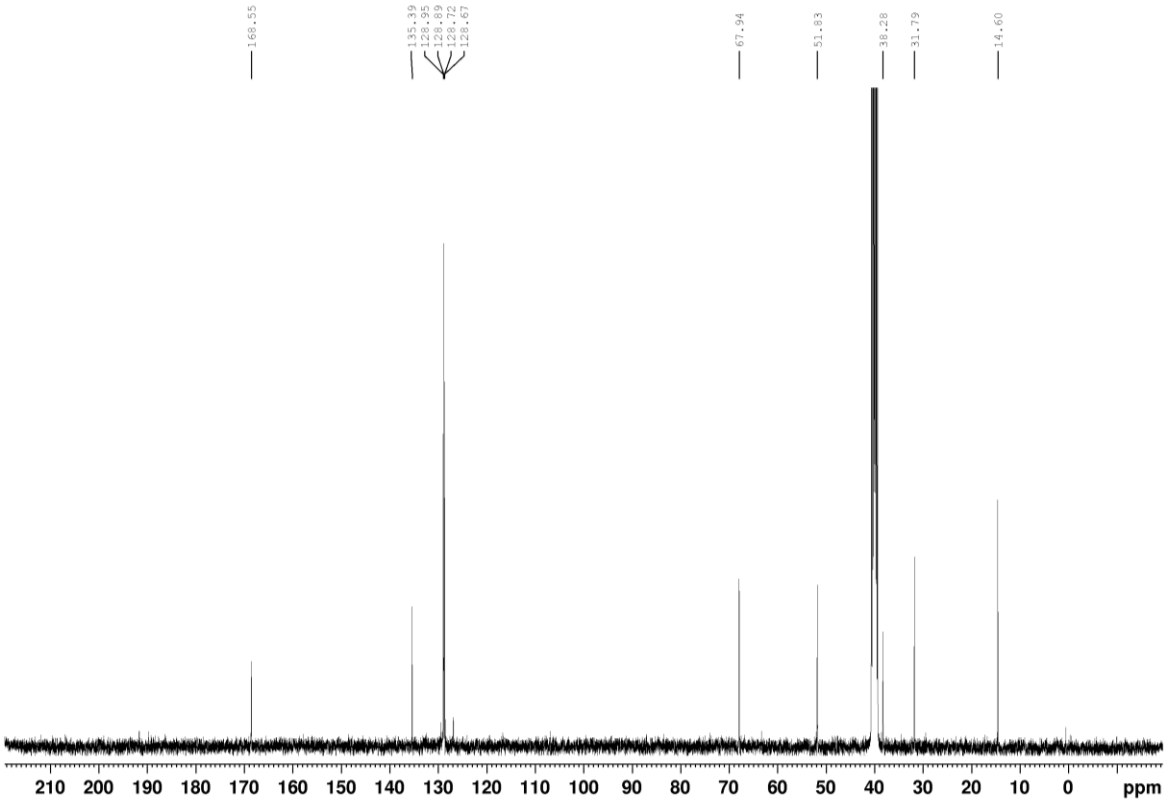

197.39
174.80
156.39
156.29
144.33
144.29
144.14
144.10
141.24
141.19
128.11
127.54
125.71
125.67
120.58
120.57
66.37
66.04
57.21
49.67
47.16
47.13
17.56
17.41
0.56
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

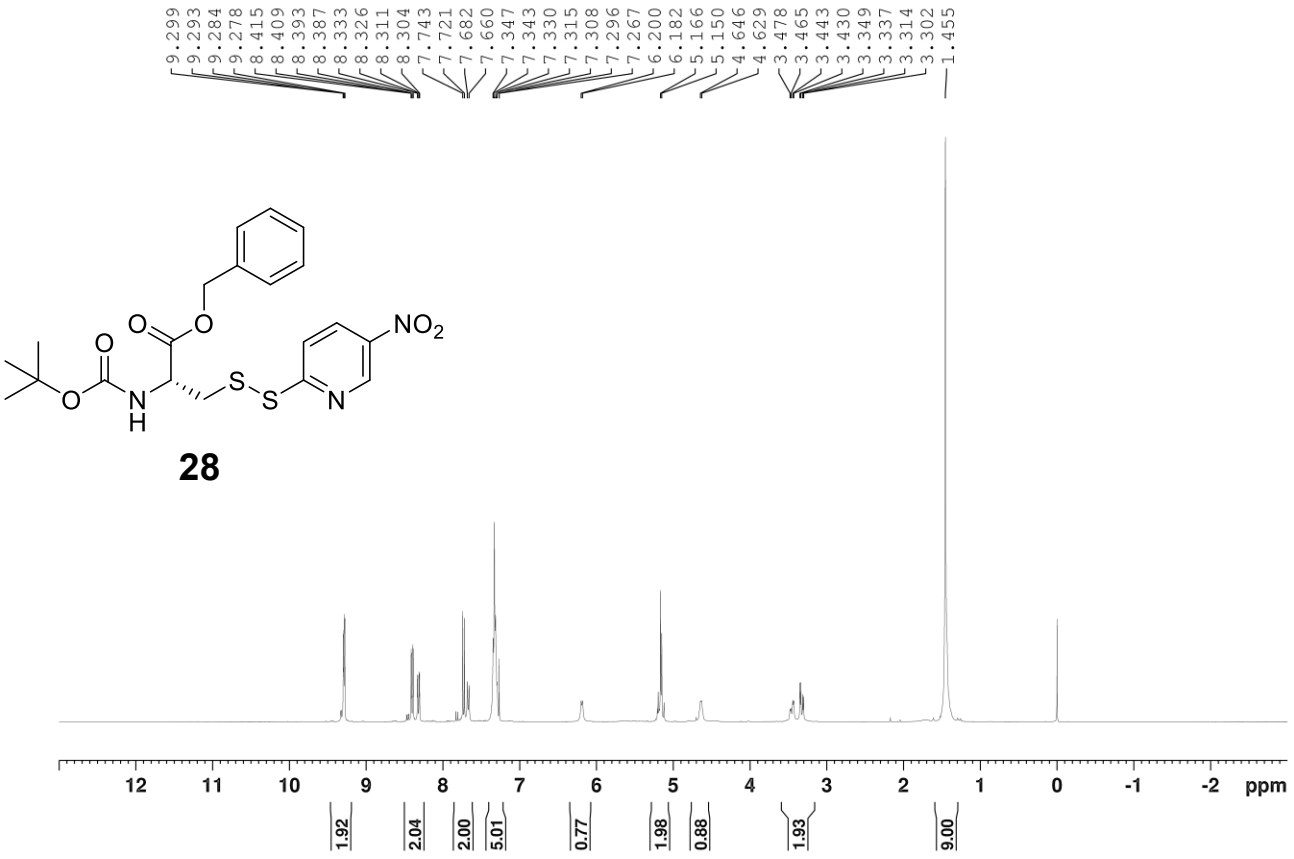

7.355
5.403
5.385
5.184
4.651
4.634
3.149
2.680
2.662
1.439
1.297
1.279
1.260
12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 ppm
5.17
0.86
2.20
0.98
2.03
2.11
9.90
3.26

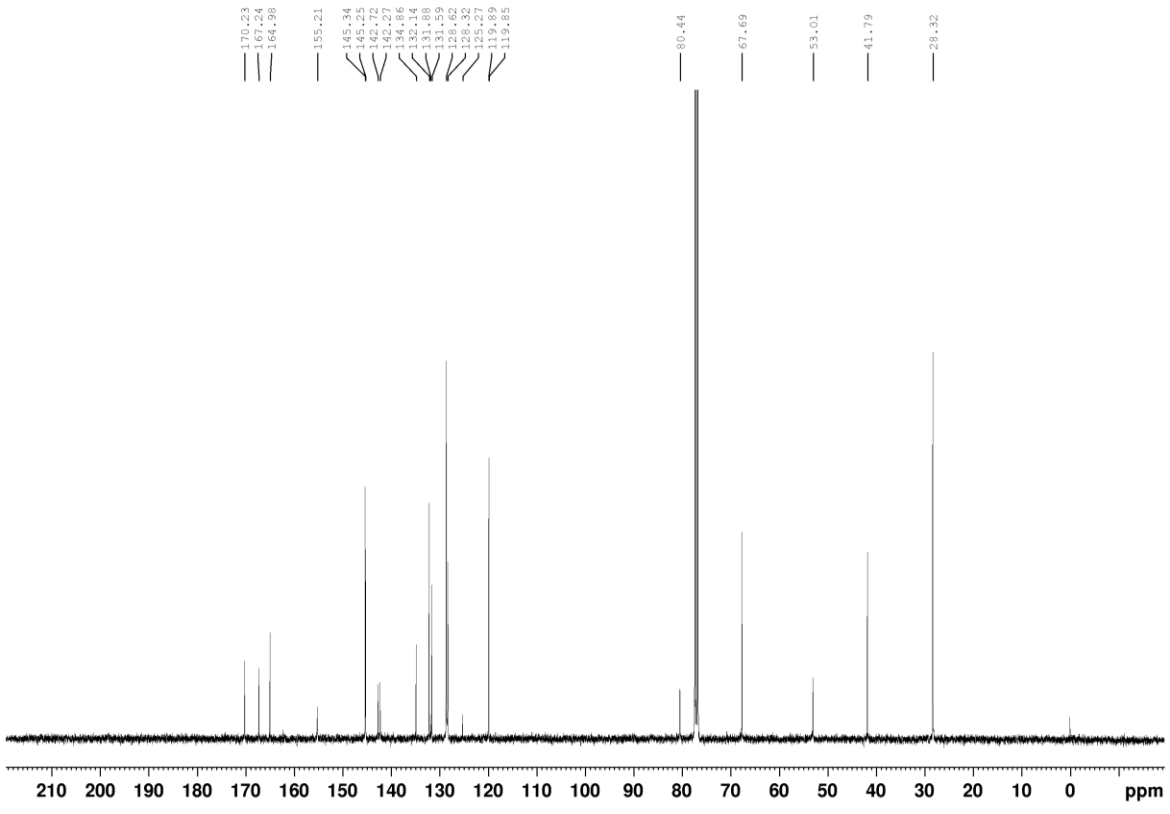
25

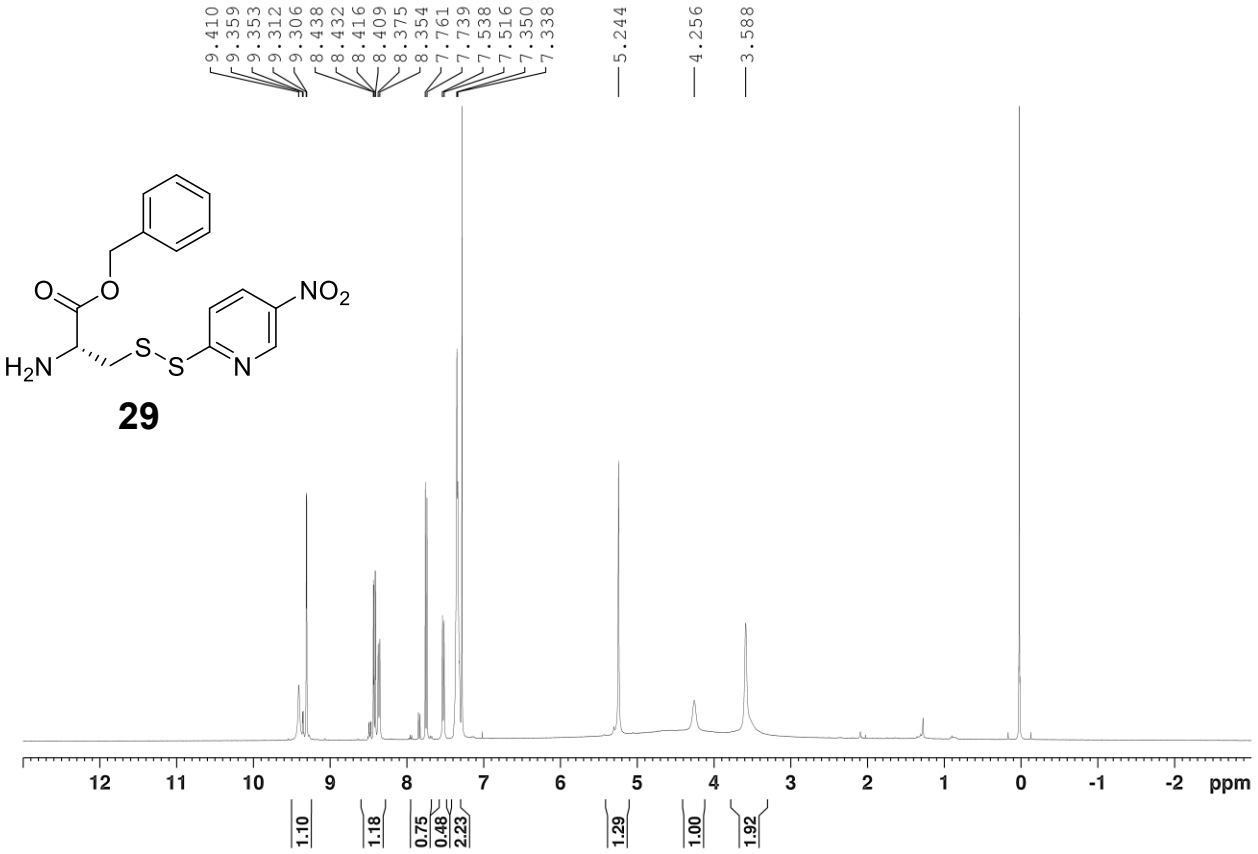

170.71
155.07
135.20
128.59
128.38
80.19
67.46
53.20
41.21
32.69
28.31
14.26
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

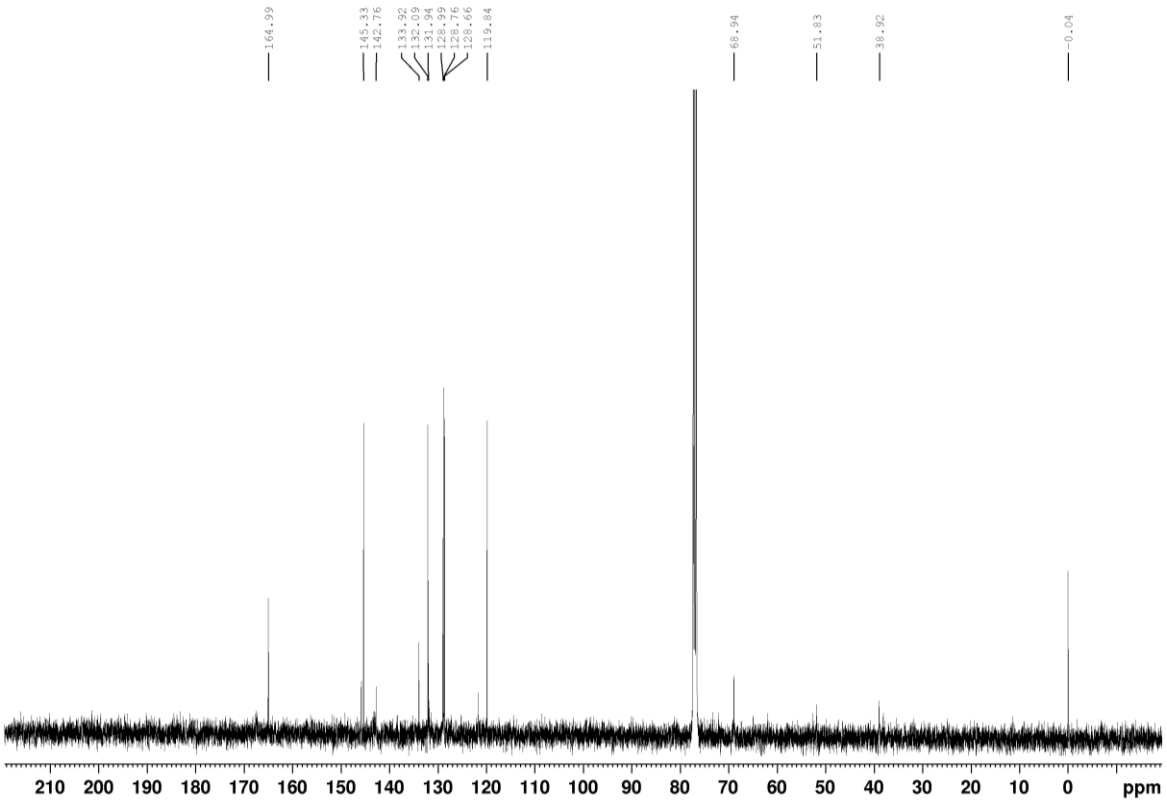

8.583
7.459
7.454
7.440
7.434
7.429
7.413
7.394
7.387
7.383
7.379
7.371
7.314
5.296
5.265
5.255
5.238
5.207
4.439
4.424
4.410
3.366
3.259
3.237
3.234
3.222
3.220
2.763
2.745
2.727
2.709
2.517
2.513
2.508
2.504
2.499
1.251
1.233
1.215
0.004
O
O
S
S
H2N
26
12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 ppm
4.88
1.90
1.00
2.26
1.64
2.83
2.35

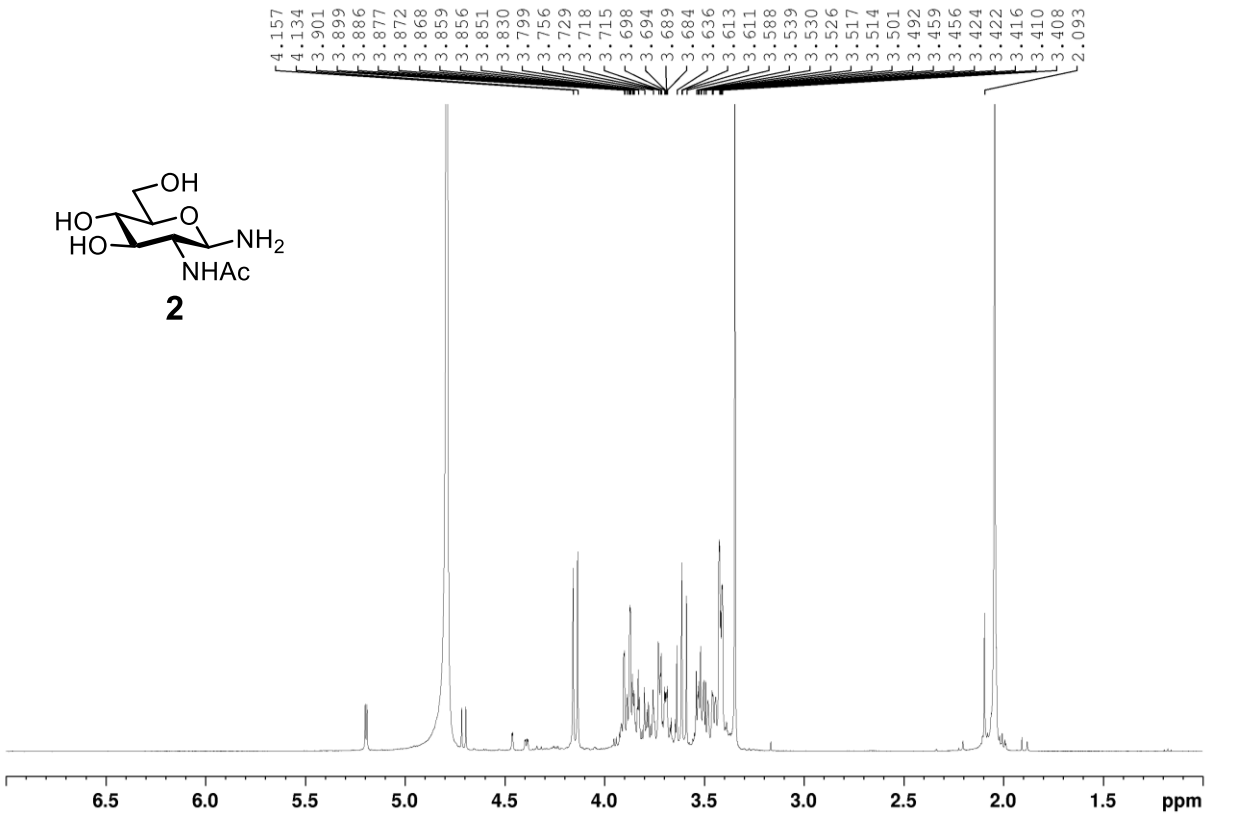

168.55
135.39
128.95
128.89
128.72
128.67
67.94
51.83
38.28
31.79
14.60
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

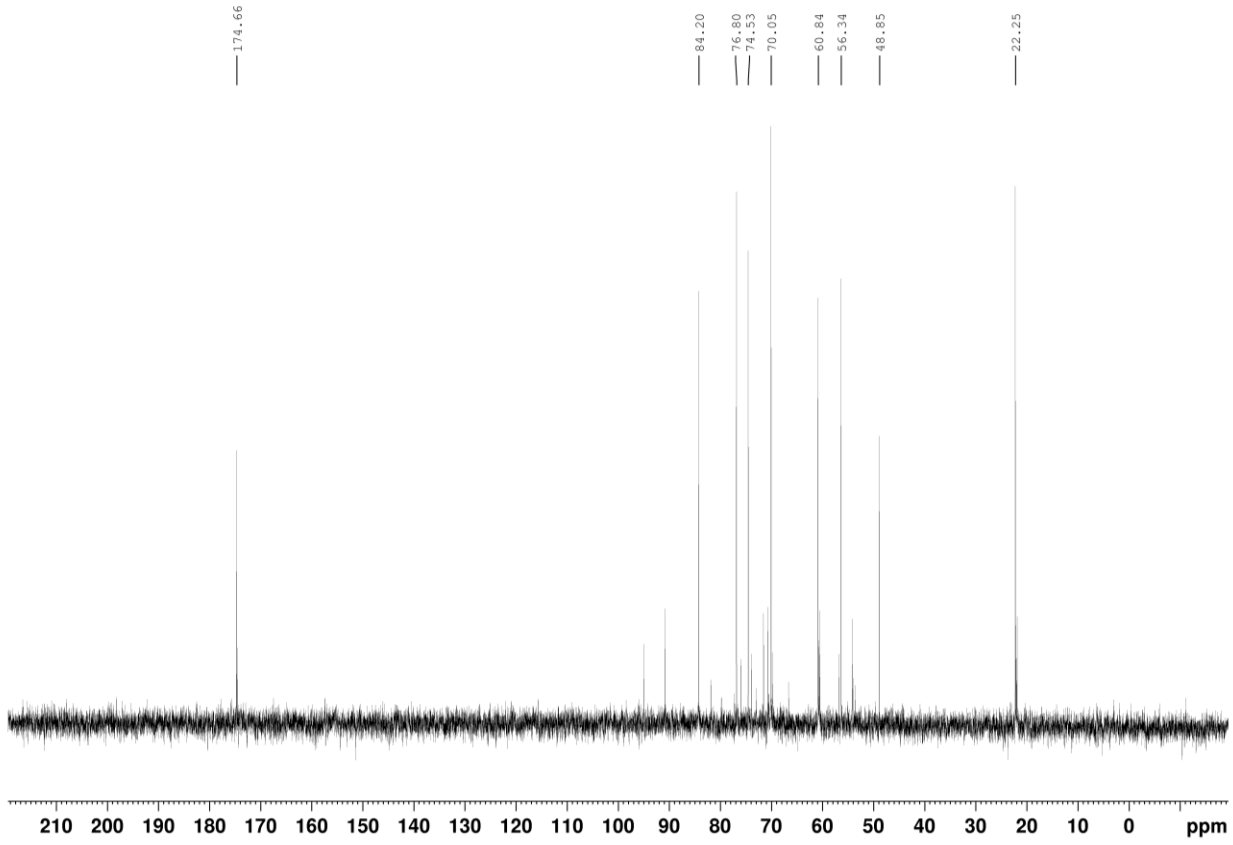

9.299
9.293
9.284
9.278
8.415
8.409
8.393
8.387
8.333
8.326
8.311
8.304
7.743
7.721
7.682
7.660
7.347
7.343
7.330
7.315
7.308
7.296
7.267
6.200
6.182
5.166
5.150
4.646
4.629
3.478
3.465
3.443
3.430
3.349
3.337
3.314
3.302
1.455
O
O
O
O
N
H
S
S
N
NO2
28
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
-2
ppm
1.92
2.04
2.00
5.01
0.77
1.98
0.88
1.93
9.00

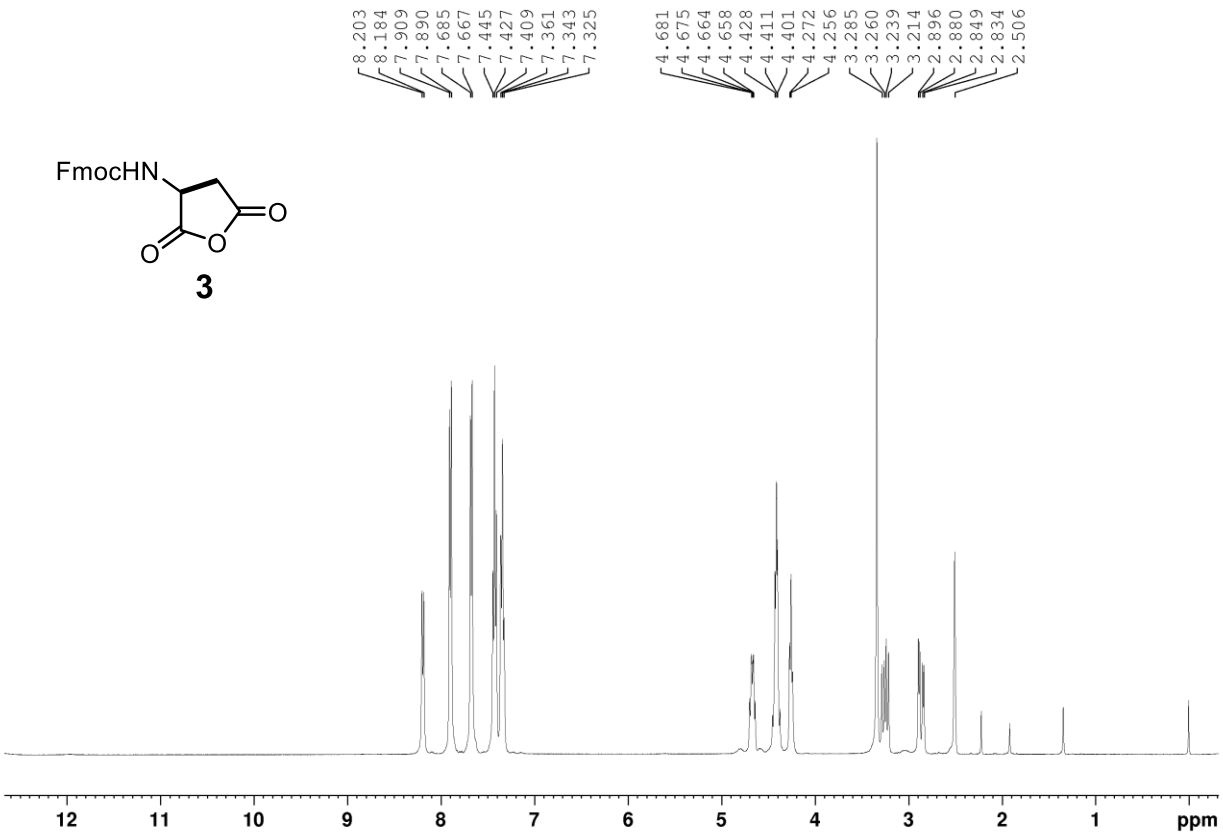

170.23
167.24
164.98
155.21
145.34
145.25
142.72
142.27
134.86
132.14
131.88
131.59
128.62
128.32
125.27
119.89
119.85
80.44
67.69
53.01
41.79
28.32
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
ppm

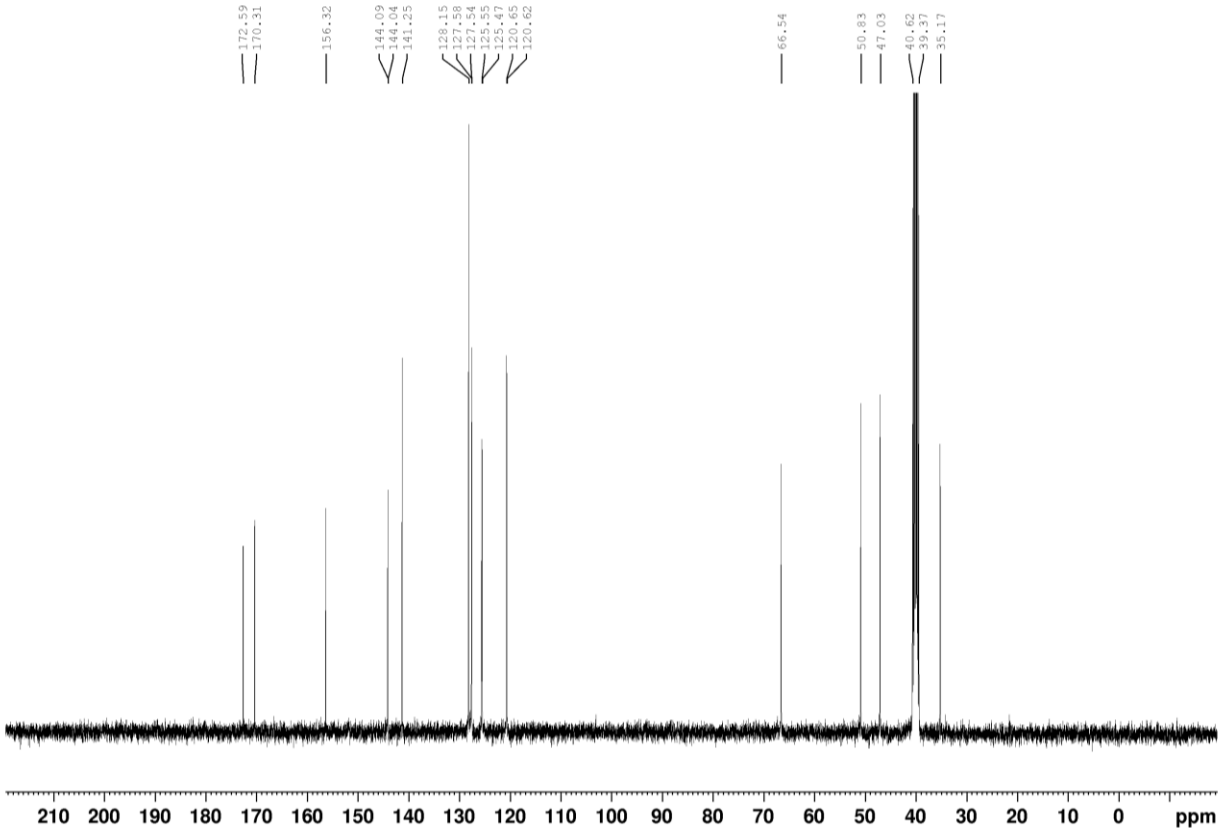

9.410
9.359
9.353
9.312
9.306
8.438
8.432
8.416
8.409
8.375
8.354
7.761
7.739
7.538
7.516
7.350
7.338
5.244
4.256
3.588
O
O
NO2
H2N
S
S
N
29
12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 ppm
1.10
1.18
0.75
0.48
2.23
1.29
1.00
1.92

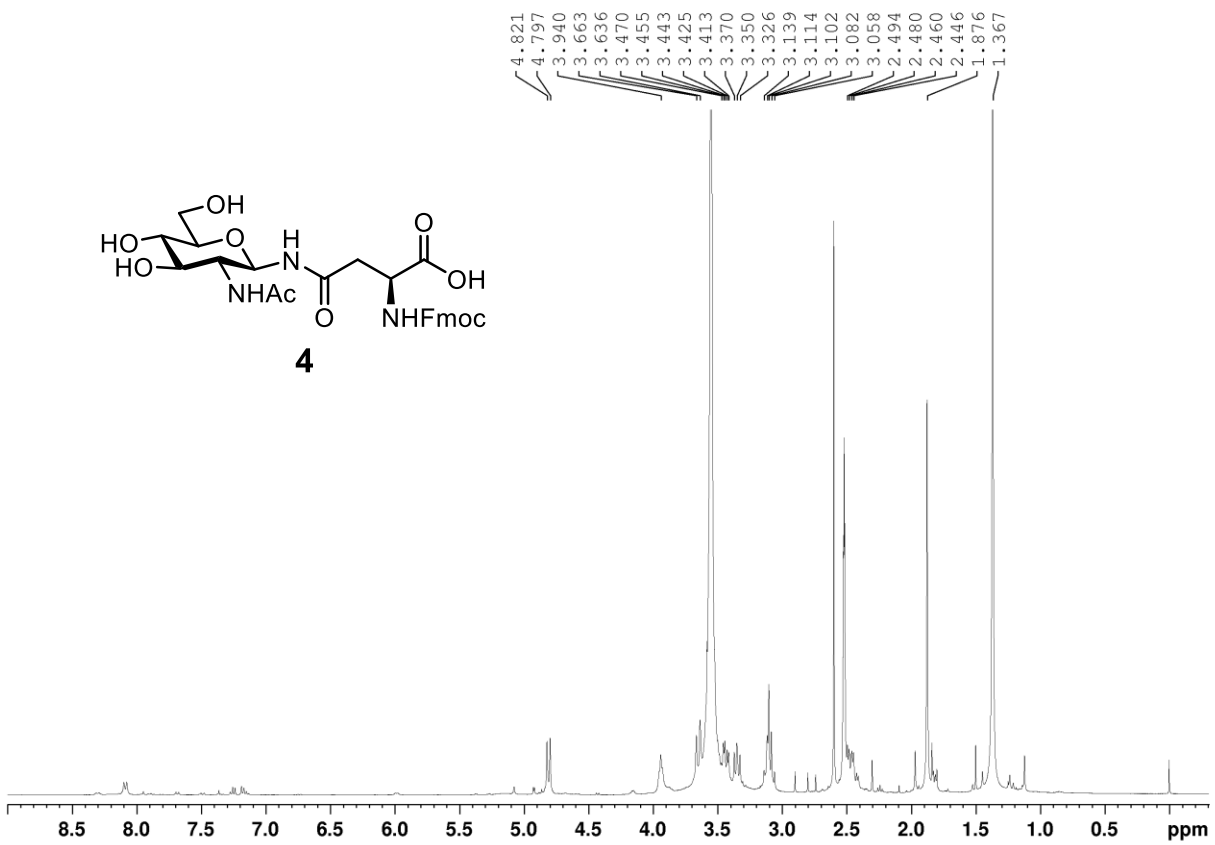

164.99
145.33
142.76
133.92
132.09
131.94
128.99
128.76
128.66
119.84
68.94
51.83
38.92
-0.04
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm


4.157
4.134
3.901
3.899
3.886
3.877
3.872
3.868
3.859
3.856
3.851
3.830
3.799
3.756
3.729
3.718
3.715
3.698
3.694
3.689
3.684
3.636
3.613
3.611
3.588
3.539
3.530
3.526
3.517
3.514
3.501
3.492
3.459
3.456
3.424
3.422
3.416
3.410
3.408
2.093
OH
HO
O
HO
NH2
NHAc
2
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
ppm

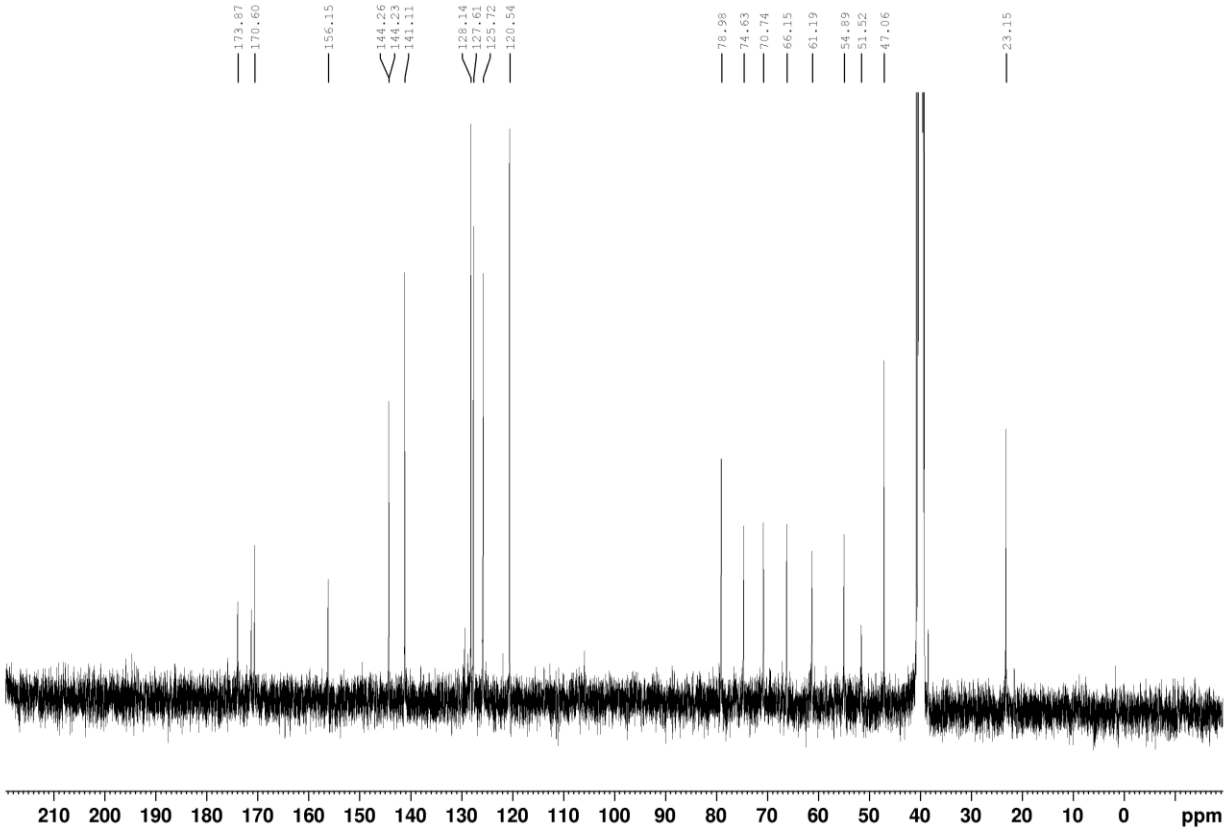

174.66
84.20
76.80
74.53
70.05
60.84
56.34
48.85
22.25
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
ppm

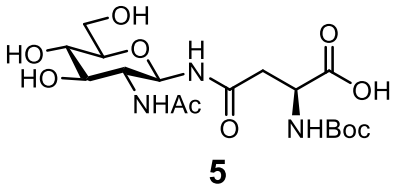

8.203
8.184
7.909
7.890
7.685
7.667
7.445
7.427
7.409
7.361
7.343
7.325
4.681
4.675
4.664
4.658
4.428
4.411
4.401
4.272
4.256
3.285
3.260
3.239
3.214
2.896
2.880
2.849
2.834
2.506
FmocHN
O
O
O
3
12
11
10
9
8
7
6
5
4
3
2
1
ppm


172.59
170.31
156.32
144.09
144.04
141.25
128.15
127.58
127.54
125.55
125.47
120.65
120.62
66.54
50.83
47.03
40.62
39.37
35.17
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
ppm


4.821
4.797
3.940
3.663
3.636
3.470
3.455
3.443
3.425
3.413
3.370
3.350
3.326
3.139
3.114
3.102
3.082
3.058
2.494
2.480
2.460
2.446
1.876
1.367
OH
HO
HO
O
H
N
NHAc
O
O
OH
NHFmoc
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
ppm

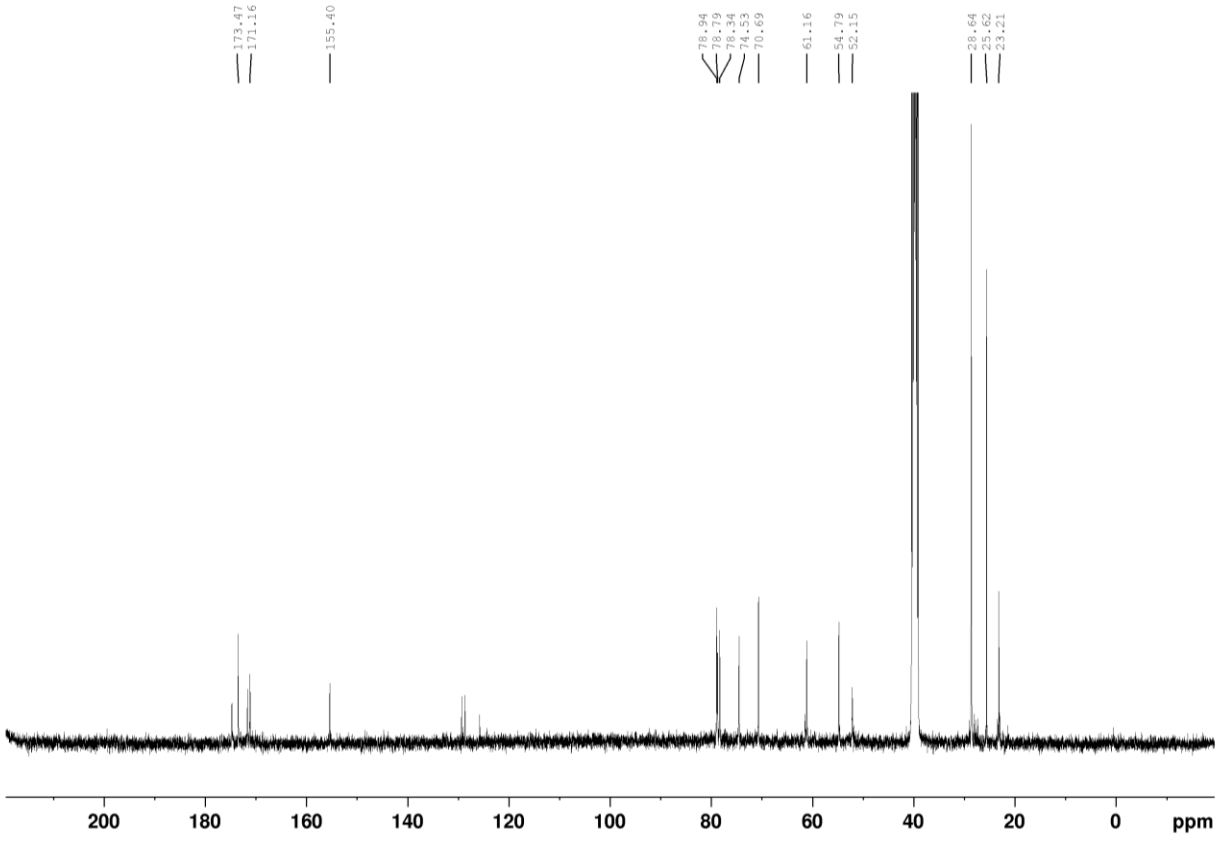
4

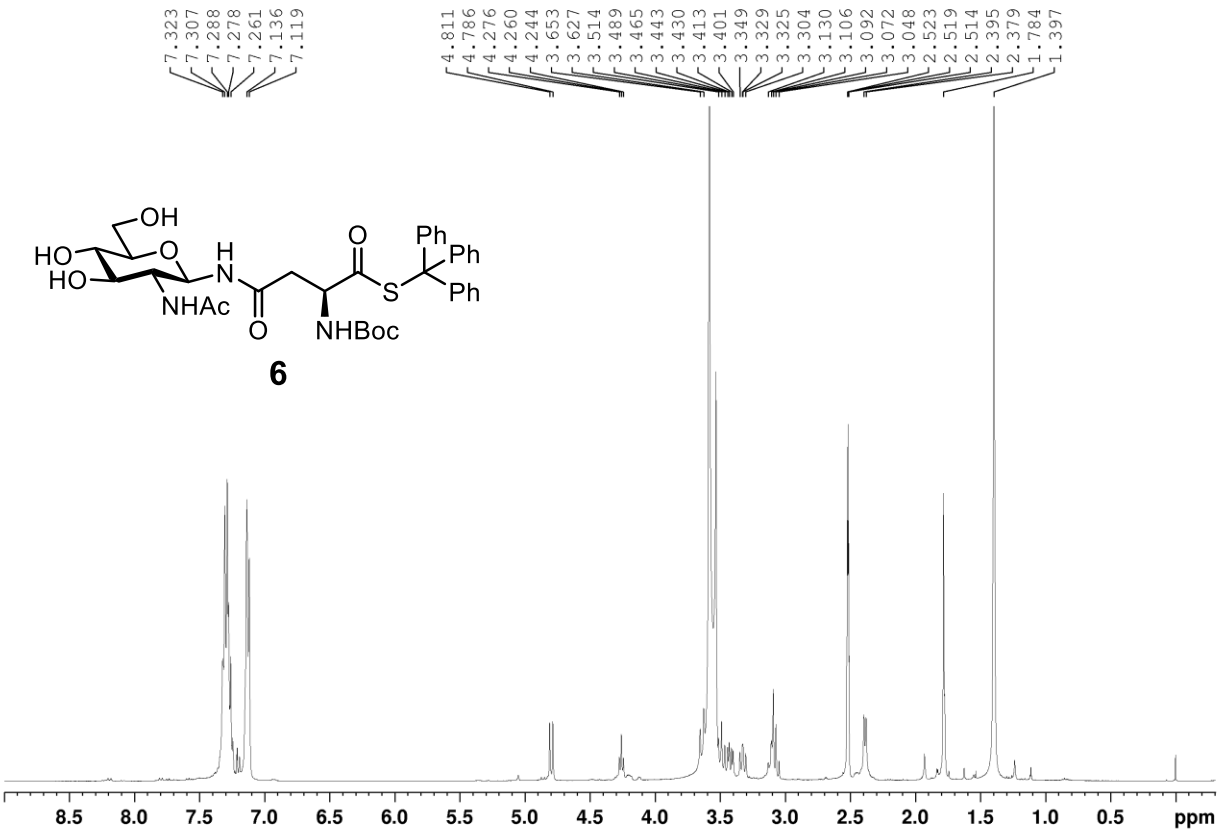

173.87
170.60
156.15
144.26
144.23
141.11
128.14
127.61
125.72
120.54
78.98
74.63
70.74
66.15
61.19
54.89
51.52
47.06
23.15
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
ppm


OH
HO
HO
O
H
N
NHAc
O
O
OH
NHBoc

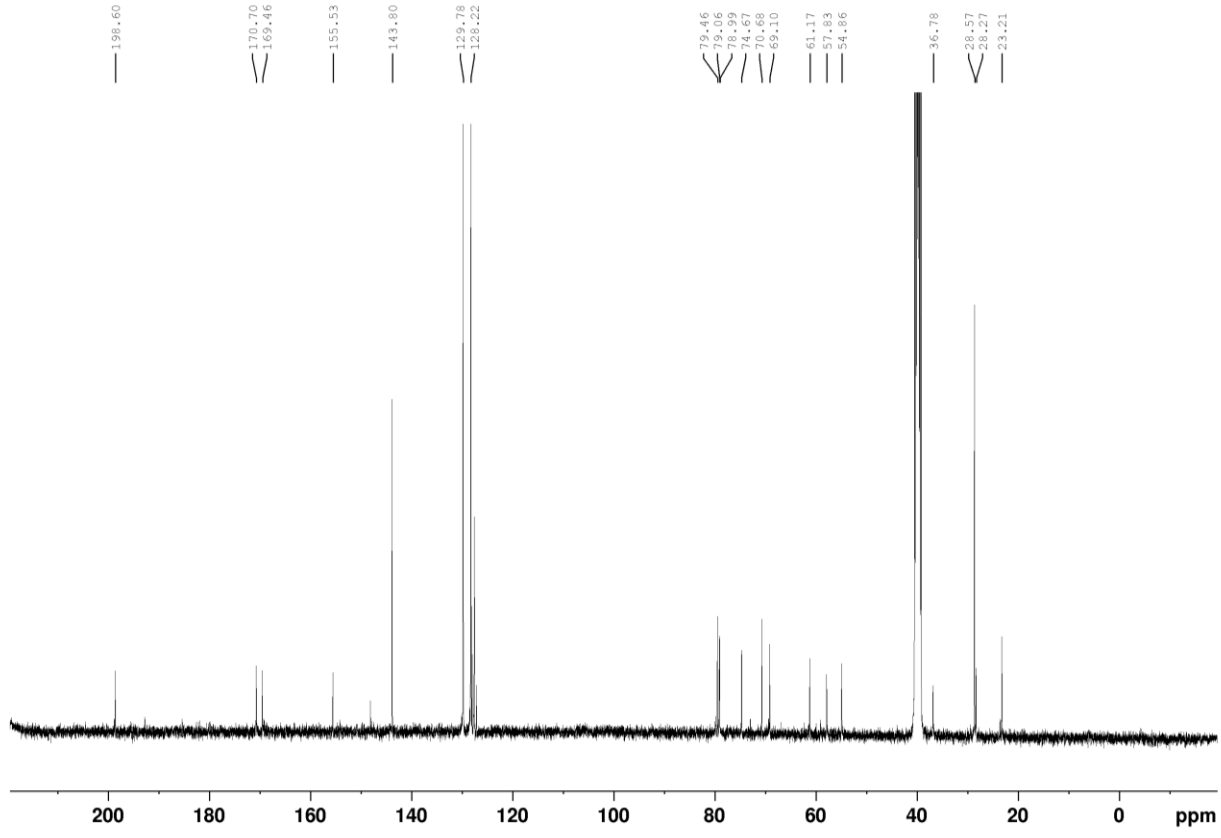
5

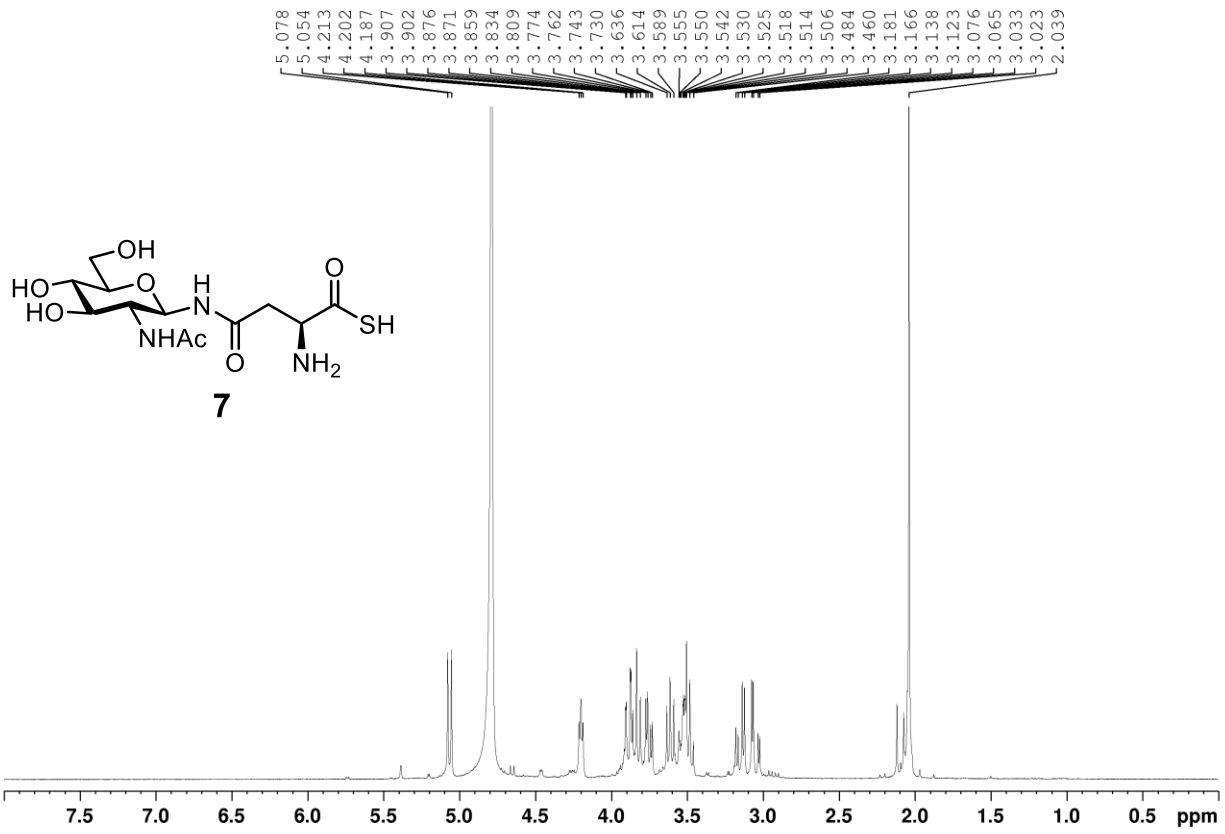
4.821
4.797
3.940
3.663
3.636
3.470
3.455
3.443
3.425
3.413
3.370
3.350
3.326
3.139
3.114
3.102
3.082
3.058
2.494
2.480
2.460
2.446
1.876
1.367
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
ppm

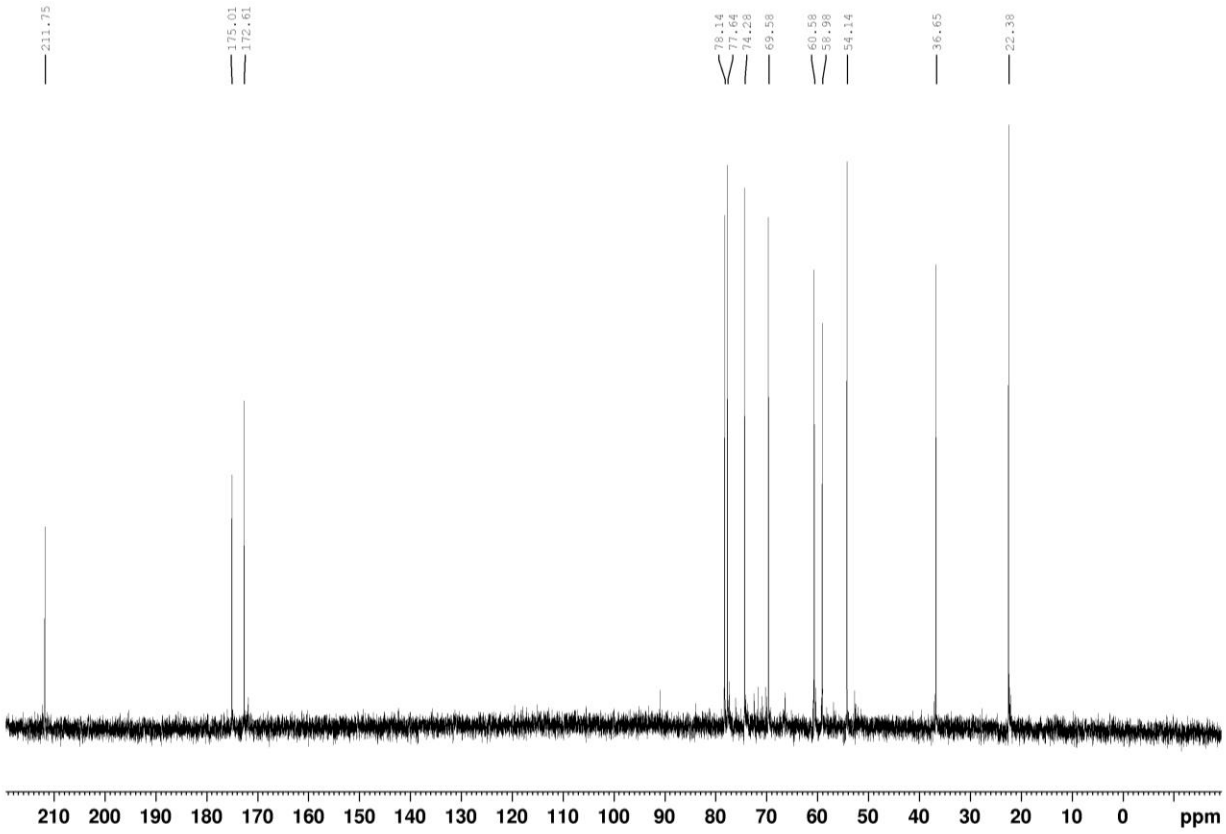

173.47
171.16
155.40
78.94
78.79
78.34
74.53
70.69
61.16
54.79
52.15
28.64
25.62
23.21
200
180
160
140
120
100
80
60
40
20
0
ppm

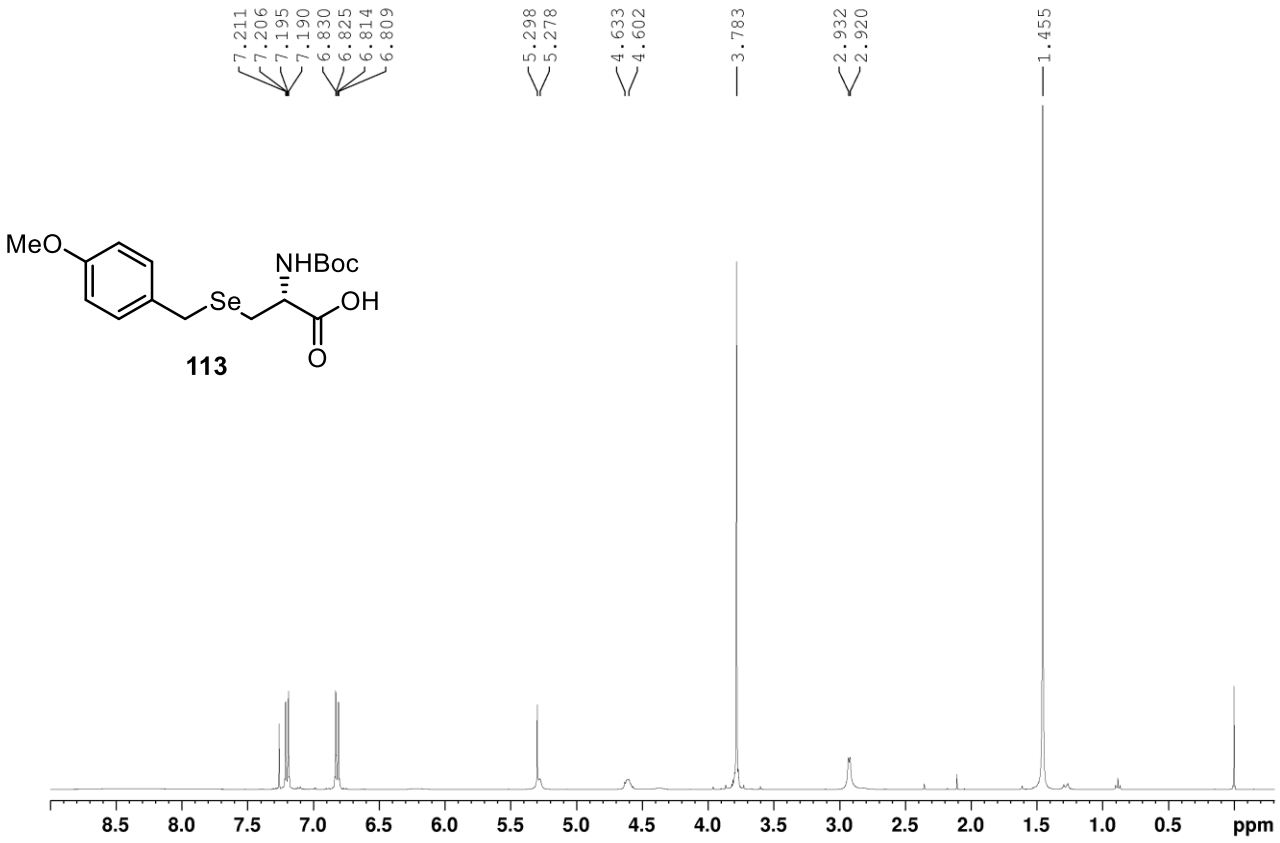

7.323
7.307
7.288
7.278
7.261
7.136
7.119
4.811
4.786
4.276
4.260
4.244
3.653
3.627
3.514
3.489
3.465
3.443
3.430
3.413
3.401
3.349
3.329
3.325
3.304
3.130
3.106
3.092
3.072
3.048
2.523
2.519
2.514
2.395
2.379
1.784
1.397
OH
HO
HO
O
H
N
NHAc
O
O
S
Ph
Ph
Ph
NHBoc
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
ppm

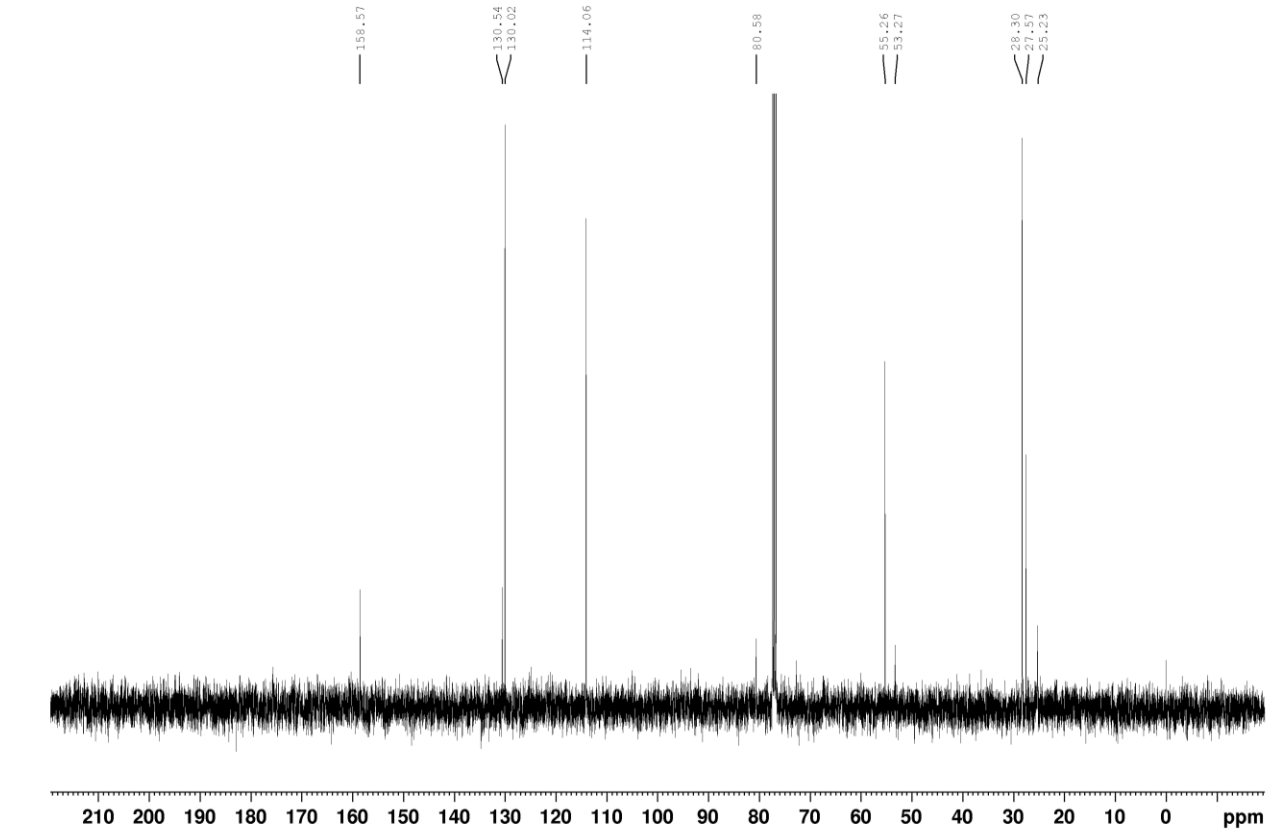
6

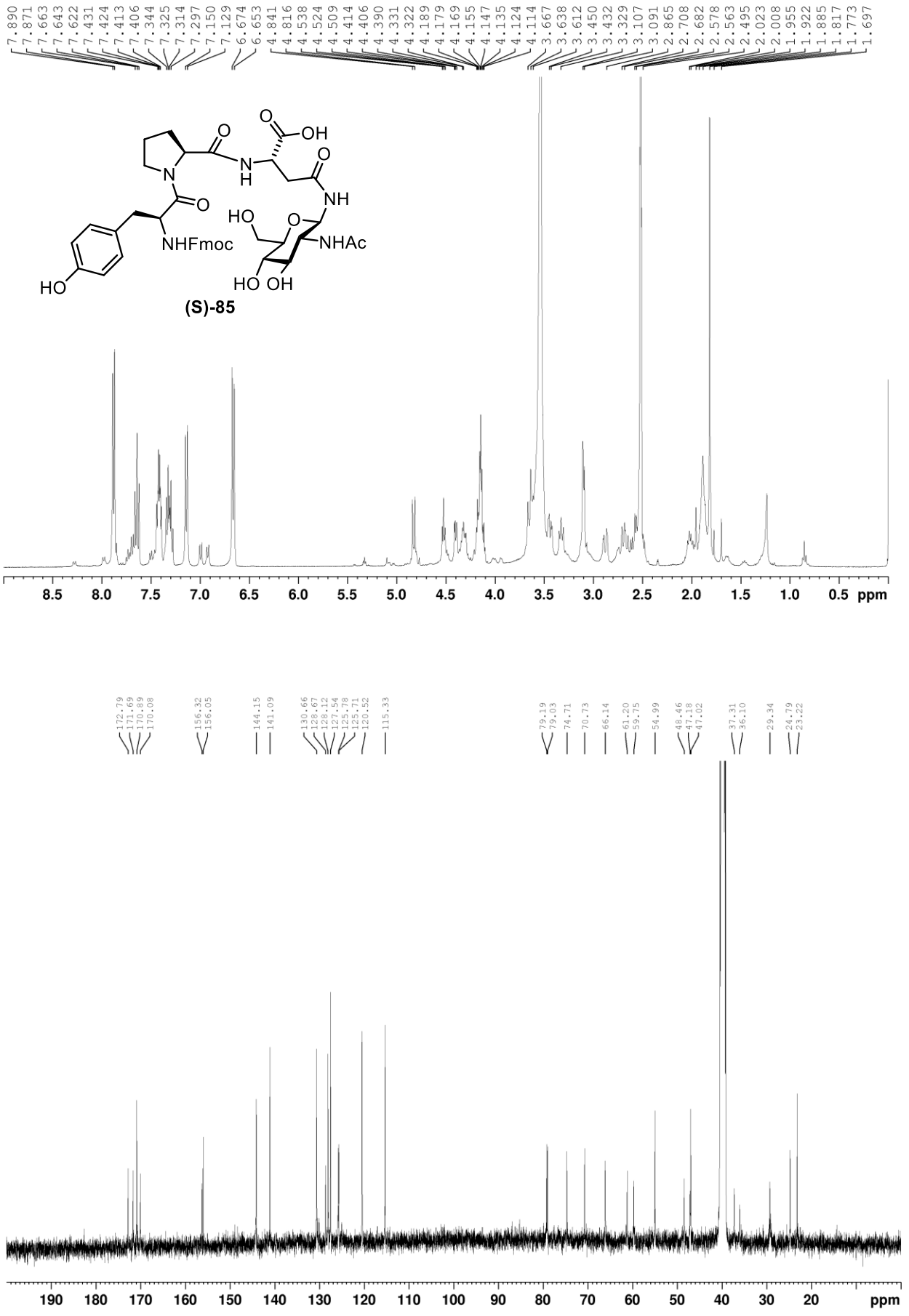

198.60
170.70
169.46
155.53
143.80
129.78
128.22
79.46
79.06
78.99
74.67
70.68
69.10
61.17
57.83
54.86
36.78
28.57
28.27
23.21
200
180
160
140
120
100
80
60
40
20
0
ppm

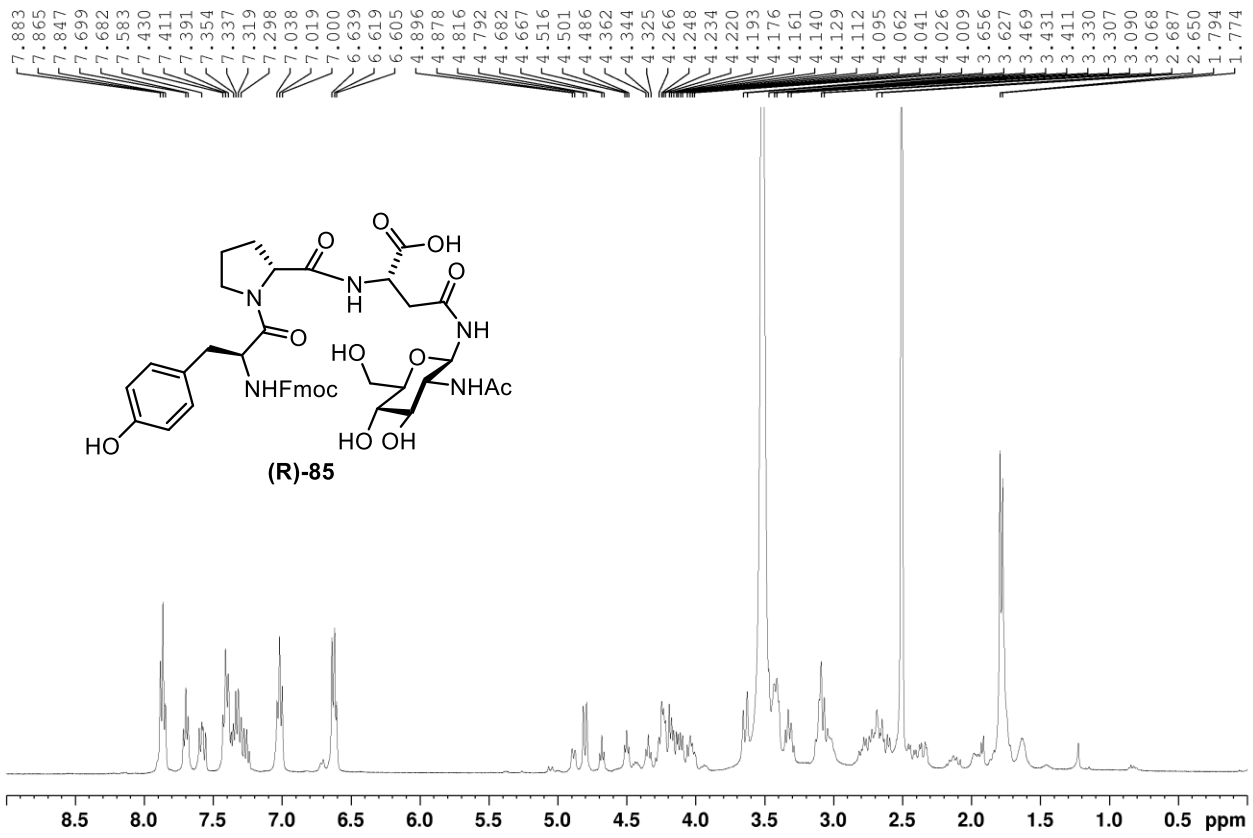

5.078
5.054
4.213
4.202
4.187
3.907
3.902
3.876
3.871
3.859
3.834
3.809
3.774
3.762
3.743
3.730
3.636
3.614
3.589
3.555
3.550
3.542
3.530
3.525
3.518
3.514
3.506
3.484
3.460
3.181
3.166
3.138
3.123
3.076
3.065
3.033
3.023
2.039
OH
HO
HO
O
H
N
NHAc
O
O
SH
NH2
7
7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm


211.75
175.01
172.61
78.14
77.64
74.28
69.58
60.58
58.98
54.14
36.65
22.38
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

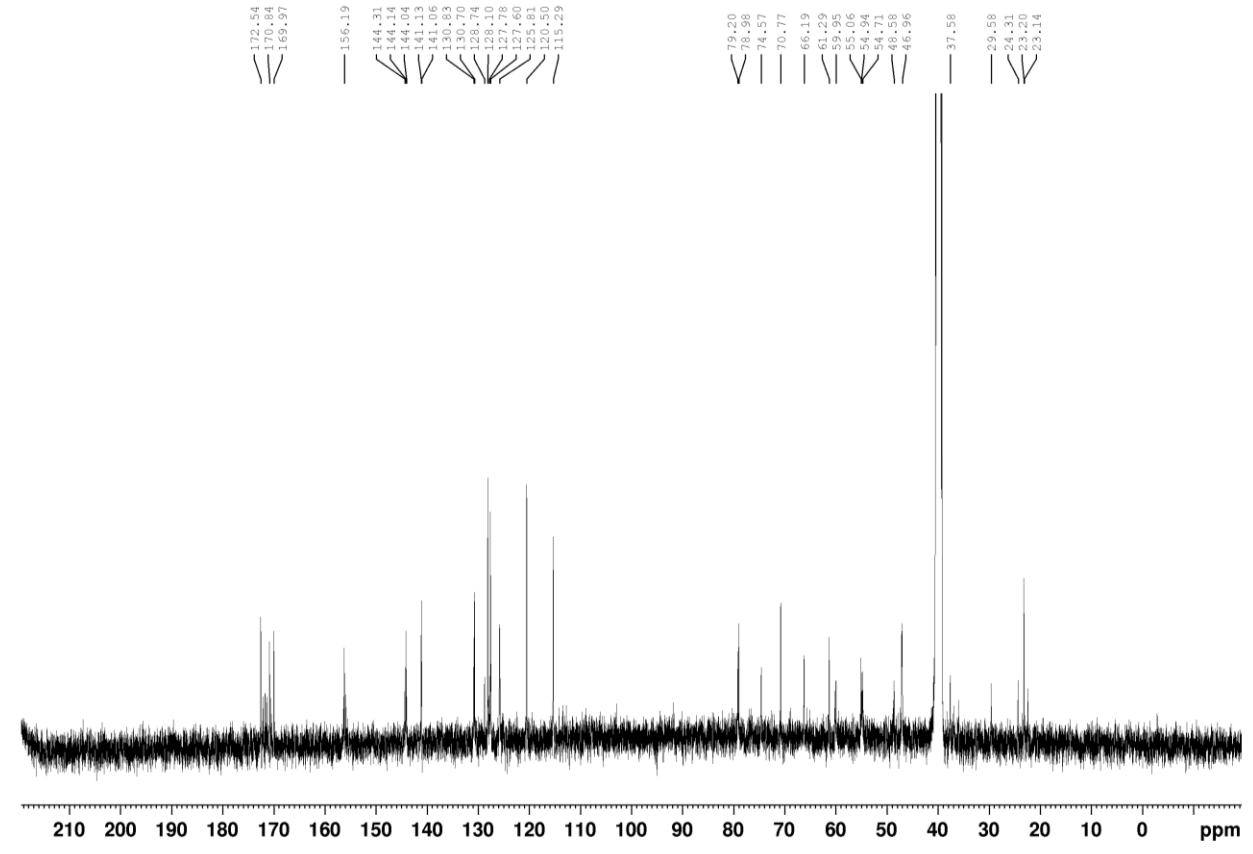

7.211
7.206
7.195
7.190
6.830
6.825
6.814
6.809
5.298
5.278
4.633
4.602
3.783
2.932
2.920
1.455
MeO
NHBoc
Se
OH
O
113
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
ppm


158.57
130.54
130.02
114.06
80.58
55.26
53.27
28.30
27.57
25.23
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
ppm


7.890
7.871
7.663
7.643
7.622
7.431
7.424
7.413
7.406
7.344
7.325
7.314
7.297
7.150
7.129
6.674
6.653
4.841
4.816
4.538
4.524
4.509
4.414
4.406
4.390
4.331
4.322
4.189
4.179
4.169
4.155
4.147
4.135
4.124
4.114
3.667
3.638
3.612
3.450
3.432
3.329
3.107
3.091
2.865
2.708
2.682
2.578
2.563
2.495
2.023
2.008
1.955
1.922
1.885
1.817
1.773
1.697
O
OH
O
N
H
N
O
NH
HO
O
NHFmoc
NHAc
HO
HO OH
(S)-85
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
ppm
172.79
171.69
170.89
170.08
156.32
156.05
144.15
141.09
130.66
128.67
128.12
127.54
125.78
125.71
120.52
115.33
79.19
79.03
74.71
70.73
66.14
61.20
59.75
54.99
48.46
47.18
47.02
37.31
36.10
29.34
24.79
23.22
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
ppm


7.883
7.865
7.847
7.699
7.682
7.583
7.430
7.411
7.391
7.354
7.337
7.319
7.298
7.038
7.019
7.000
6.639
6.619
6.605
4.896
4.878
4.816
4.792
4.682
4.667
4.516
4.501
4.486
4.362
4.344
4.325
4.266
4.248
4.234
4.220
4.193
4.176
4.161
4.140
4.129
4.112
4.095
4.062
4.041
4.026
4.009
3.656
3.627
3.469
3.431
3.411
3.330
3.307
3.090
3.068
2.687
2.650
1.794
1.774
O
OH
O
N
H
N
O
O
NH
HO
O
NHFmoc
NHAc
HO
HO OH
HO
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
ppm

(R)-85


172.54
170.84
169.97
156.19
144.31
144.14
144.04
141.13
141.06
130.83
130.70
128.74
128.10
127.78
127.60
125.81
120.50
115.29
79.20
78.98
74.57
70.77
66.19
61.29
59.95
55.06
54.94
54.71
48.58
46.96
37.58
29.58
24.31
23.20
23.14
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
ppm

List of publications

Publications related to this dissertation.

1, Nomura, K.; Maki, Y.; Okamoto, R.; Satoh, A.; Kajihara, Y. Glycoprotein Semisynthesis by Chemical Insertion of Glycosyl Asparagine Using a Bifunctional Thioacid-Mediated Strategy. *J. Am. Chem. Soc.* **2021**, *143*, 10157.

2, Okamoto, R.; Nomura, K.; Maki, Y.; Kajihara, Y. A Chemoselective Peptide Bond Formation by Amino Thioacid Coupling. *Chem. Lett.* **2019**, *48*, 1391.

Other publications

1, Maki, Y.; Nomura, K.; Okamoto, R.; Izumi, M.; Mizutani, Y.; Kajihara, Y. Acceleration and Deceleration Factors on the Hydrolysis Reaction of 4,6-O-Benzylidene Acetal Group. *J. Org. Chem.* **2020**, *85*, 15849.

2, Okamoto, R.; Haraguchi, T.; Nomura, K.; Maki, Y.; Izumi, M.; Kajihara, Y. Regioselective α-Peptide Bond Formation Through the Oxidation of Amino Thioacids. *Biochemistry* **2019**, *58*, 1672.

3, Nomura, K. Chemical Insertion of Glycan Extends the Scope of Homogeneous Glycoprotein Synthesis. *Glycoforum*. **2022**, *25*, A2.

Acknowledgement

I would like to appreciate all who supported me. Without them, this dissertation could not have been finished. I especially express my gratitude to my three supervisors, Prof. Yasuhiro Kajihara, Prof. Ryo Okamoto and Prof. Yuta Maki with whose passion I could finish these projects. Their deepest knowledge and extensive research experiences in both organic chemistry and biochemistry inspired me with many interesting ideas. I would also like to thank Prof. Ayano Satoh (Okayama Univ. Graduate School of Interdisciplinary Science and Engineering in Health Systems, *Chapter 4*, bioassay), Prof. Eiji Miyoshi and Dr. Tomoya Fukuoka (Osaka Univ. Graduate School of Medicine, *chapter 5*, bioassay) who willingly cooperated with my research. I am also grateful to JSPS (Japan Society for the Promotion of Science) (DC1 grant number: 20J20649) for their financial supports.

I owe a greatest deal to Dr. Hiroyuki Shibata, Dr. Yanbo Liu, Mr. Tatsuya Moriguchi, Mr. Youdong Yang, Dr. Junpei Abe, Dr. Donglin Zhao, Dr. Shun Ito and all other members of our laboratory and *GlyTech, Inc.* (2016 ~). At last, I am greatly indebted to my parents (1995 ~).

Kota Nomura, 2021, Toyonaka, Japan