Chemical modification of recombinant peptides with hydrophilic protecting group for expanding semisynthesis of glycoproteins

糖タンパク質の半合成に向けた親水性保護基を利用する発現ペプチドの化学修飾

By
Chaitra Chandrashekar

Department of Chemistry
Graduate School of Science
Osaka University

Osaka, Japan
2017

Table of Contents

Table of contents

Abstract

Glycosylation of proteins is the major post translational modifications (PTMs) that occurs in eukaryotes. An efficient way to study their structural and functional properties is by synthesis of homogeneous glycoproteins by total chemical synthesis or by semisynthesis. Total chemical synthesis allows desired modifications of proteins at the atomic-level, while it has several drawbacks such as size limitation of polypeptide, low yield, expensive starting material and is time consuming. Semisynthesis involves combination of recombinant expression and chemical synthesis that reduces the number of chemical conversion steps and increases yield in less time.

In order to perform chemical synthesis, Native Chemical Ligation (NCL) is a critical method to couple glycopeptide and peptide fragments together. NCL requires cysteine at the suitable position in a peptide backbone, but some glycoprotein targets don't meet this requirement which limits the use of NCL reaction. Recently, this limitation is overcome by using selenocysteine (Sec) or unnatural mercapto amino acids instead of Cys to perform NCL followed by subsequent deselenization/desulfurization.

However, these modifications cannot be applied to the expressed protein from *E.coli,* because specific modification at the *N*-terminal is difficult. Chemical modification of recombinant peptide can be performed only after its partial protection. The *tert*-butyloxycarbonyl (Boc) group is commonly used for partial protection which makes protected peptide hydrophobic. This causes solubility problems that influence the reactivity in many reactions that are to be performed in aqueous buffer conditions and also hinders efficient purification by reverse phase HPLC.

Therefore, I envisaged to establish a general semisynthetic strategy for the synthesis of glycoproteins by chemically modifying *N*-terminal of partially protected recombinant peptide using Sec to perform NCL. The major solubility problem of partially protected peptides using Boc group, is solved by replacing it with a hydrophilic carbohydrate-derived protecting group. The glucose- and galactose-derived protecting groups were previously used for α-amino group protection in chemical peptide synthesis. I employed carbohydrate-derived protecting groups to efficiently protect lysine side chains of recombinant peptide. Therefore, glucose-, galactose- and lactose- derived protecting groups were activated with imidazole and protection was optimized.

For establishing the strategy, a small model peptide **10** was obtained by *E.coli* expression as a fusion peptide **9** to SUMO fusion tag, which was cleaved by specific SUMO protease. The resultant desired short peptide **10** was partially protected with glucose- and lactose-derived protecting group. The increase in hydrophilicity of partially protected peptides was proved by comparing the retention time (t_R) of peptides in reverse phase HPLC using routine acidic solvent system and neutral solvent system. Later, glucose-protected peptide **12** was used to perform an efficient enzymatic cleavage under aqueous condition using TEV protease

to obtain partially protected peptide with free *N*-terminal **15**. This peptide could then be specifically modified at the *N*-terminal with Sec to obtain **19**. Using Sec is advantageous because Sec reacts more faster in NCL than Cys and moreover -SeH (selenol) can be selectively removed in the presence of –SH (thiol) groups. In the final step all the protecting groups were removed to give unprotected *N*-terminal selenocysteine modified recombinant peptide **20** which can be readily used for NCL.

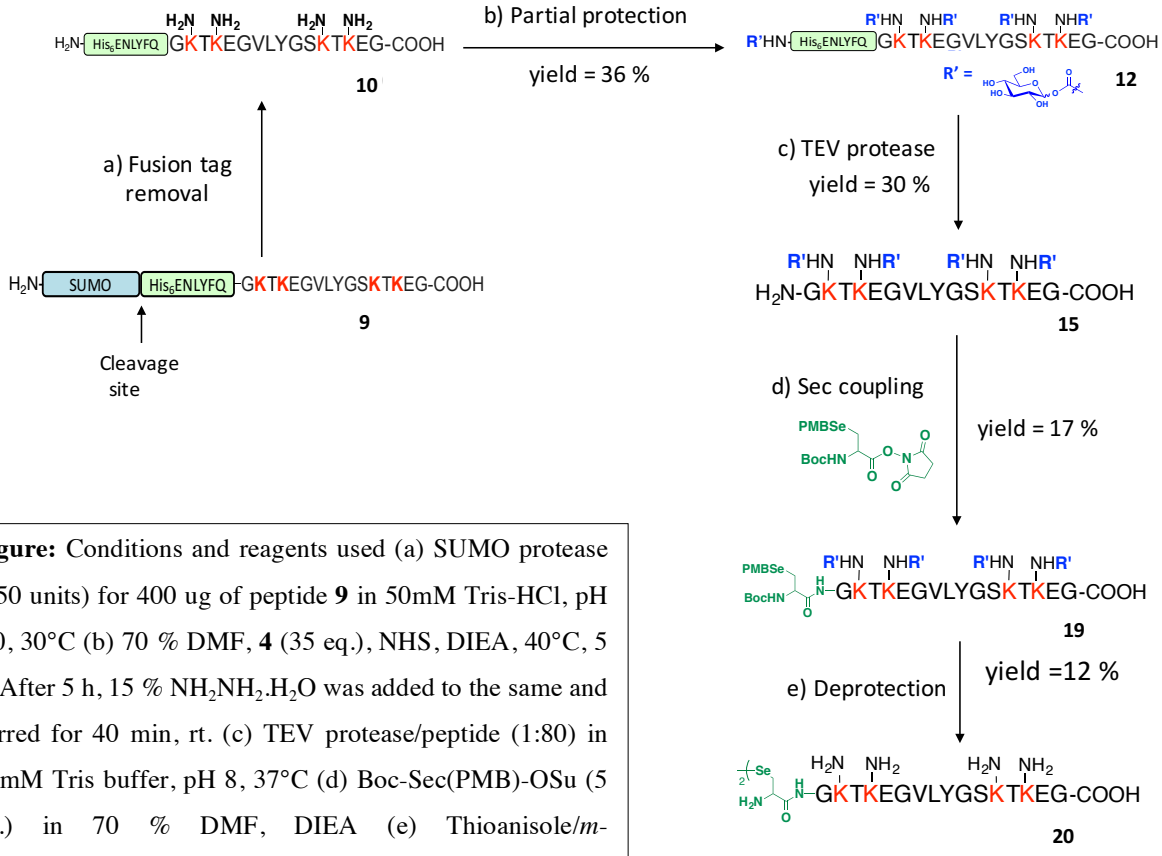


Figure: Conditions and reagents used (a) SUMO protease (150 units) for 400 ug of peptide **9** in 50mM Tris-HCl, pH 8.0, 30°C (b) 70 % DMF, **4** (35 eq.), NHS, DIEA, 40°C, 5 h. After 5 h, 15 % $NH_2NH_2.H_2O$ was added to the same and stirred for 40 min, rt. (c) TEV protease/peptide (1:80) in 25mM Tris buffer, pH 8, 37°C (d) Boc-Sec(PMB)-OSu (5 eq.) in 70 % DMF, DIEA (e) Thioanisole/*m*-cresol/TFA/TfOH (2:1:20:2), 0°C, 2 h.

Thus, an efficient semisynthetic strategy was developed for the specific modification of N-terminal of recombinant peptide. By replacing the hydrophobic Boc-protecting group with a hydrophilic carbohydrate-derived urethane protecting group, I solved the critical solubility problem associated with current method of partial protection of recombinant peptide. This improvement allows access to various modifications of recombinant short peptides that was previously difficult. I could successfully install Sec which is an efficient functional group for NCL, the most critical peptide-peptide ligation strategy. This will advance the field of semisynthesis of various post-translationally modified proteins such as glycoproteins, phosphorylated proteins, and other desired modifications for structural and functional studies.

Abbreviation

AaT: aminoacyltransferase

Ac: acetyl

Acm: acetamidomethyl

AGalOC: tetra-*O*-acetyl-D-galactopyranosyloxycarbonyl

AGlOC: tetra-*O*-acetyl-D-glucopyranosyloxycarbonyl

Ala: alanine

ALaOC: octa-*O*- acetyl-D-lactopyranosyloxycarbonyl

Asn: asparagine

Arg: arginine

Asp: aspartic acid

BHK: baby hamster kidney

CHO: Chinese hamster ovary

COPII: coat protein II

Boc: *tert*-butyloxycarbonyl

Bn: benzyl

Bz: benzoyl

CNX: calnexin

CRT: calreticulin

Cys: cysteine

DCM: dichloromethane

DIC: *N*, *N'*-diisopropylcarbodiimide

DIEA: *N*, *N'*-diisopropylethylamine

DMAP: 4-dimethylaminopyridine

DMF: *N*, *N'*-dimethylformamide

DMS: dimethyl sufide

DMSO: dimethyl sulfoxide

DNP: 2, 4-dinitrophenyl

Dol: dolicol

DTT: D, L -dithiothreitol

EDT: 1, 2-ethanedithiol

EPL: Expressed protein ligation

Abbreviation

EPO: erythropoietin

ER: endoplasmic reticulum

ESI: electrospray ionization

Fmoc: 9-fluorenylmethyloxycarbonyl

Gal: galactose

GalNAc: *N*-acetylgalactosamine

Glc: glucose

GlcNAc: *N*-acetyl-D-glucosamine

Gln: glutamine

Glu: glutamic acid

Gly: glycine

Gn.HCl: guanidine hydrochloride

GT: glycosyl transferase

HBTU: *O*-(benzotriazol-1-yl)-*N*, *N*, *N*', *N*'-tetramethyluronium hexafluorophosphate

Hcs: homocysteine

HCTU: *O*-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate

HEK: Human Embryonic Kidney

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

His: histidine

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

HOBt: 1-hydroxy-benzotriazole

HPLC: high performance liquid chromatography

HRMS: high resolution mass spectroscopy

LC: liquid chromatography

Ile: isoleucine

Leu: leucine

Lys: lysine

NCL: native chemical ligation

Man: mannose

MESNa: sodium 2-mercaptoethanesulfonate

Met: methionine

MPA: 3-mercaptopropionic acid

MPAA: 4-mercaptophenylacetic acid

MS: mass spectrometry

Abbreviation

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

NGNA: *N*-glycolylneuraminic acid

NMR: nuclear magnetic resonance

OST: oligosaccharyl transferase

Pbf: 2, 2, 4, 6, 7-pentamethyldihydrobezofuran-5-sulfonyl

PDB: protein data bank

Phe: phenylalanine

PMB: *para*-methoxy benzyl

Pro: proline

PTM: posttranslational modification

PyBOP: 1*H*-benzotriazol-1-yloxy-tri(pyrrolidino)phosphonium hexafluorophosphate

Sec: selenocysteine

Ser: serine

SPPS: solid phase peptide synthesis

TEV: tobacco etch virus

TFA: trifluoroacetic acid

TfOH: trifluoromethanesulfonic acid

THF: tetrahydrofuran

Thr: threonine

TIPS: Triisopropylsilane

tRNA: transfer ribonucleic acid

Trt: trityl

Trp: tryptophan

Tyr: tyrosine

UDP: uridine diphosphate

UV: ultraviolet

VA-044: 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride

Val: valine

Xyl: xylose

Chapter 1

General Introduction

1. Introduction of glycoproteins

Proteins are one of the major biological macromolecules that are essential for life as they play a critical role in cell structure and function. Almost all proteins undergo co- and post-translational modifications and is generally divided into two categories: a) covalent modification of amino acid side chains and b) cleavage of protein backbone at a specific peptide bond.[1] Most common covalent modifications are glycosylation, phosphorylation, acylation, methylation, ubiquitination, hydroxylation etc., (Figure 1-1) each having their own characteristic functions.[1]

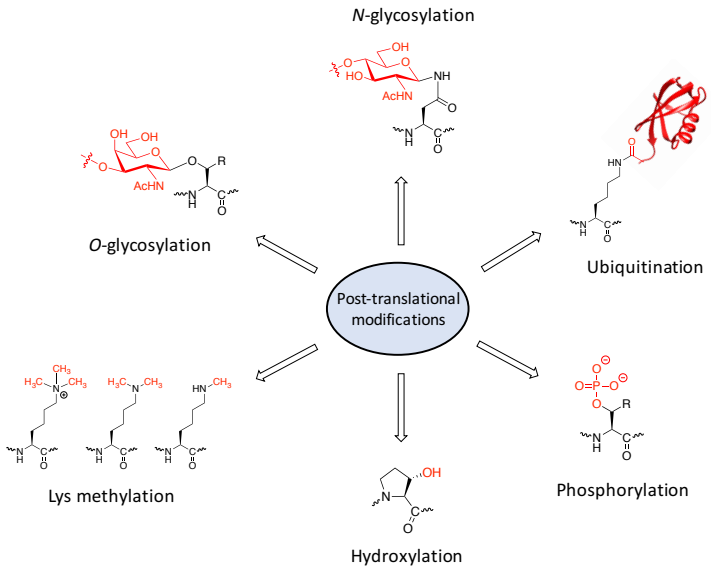


Figure 1-1: Common covalent post-translational modifications that regulate the protein level and function in the cells. (PDB no. of Ubiquitin 1UBQ)

1-1. Glycoproteins

Glycosylation is the major post translational modifications (PTMs) that refers to modification of protein backbone with oligosaccharides to form glycoproteins. About 50% of human proteins are glycosylated.[2] Oligosaccharides on proteins are mainly of two types: a) *O*-linked glycosylation which

refers to modification of hydroxy group on serine or threonine and b) *N*-linked glycosylation which refers to modification of asparagine side-chain of consensus sequence Asn-X-Thr/Ser (X is any amino acid, except for proline). *O*-Linked glycosylation in eukaryotes are generally short and less complex than *N*-linked glycosylation. There are eight different types of *O*-linked glycosylation that includes core 1 to core 8 (Figure 1-2)[3] all having GalNAc-α-1-O-Ser/Thr linkage in common.

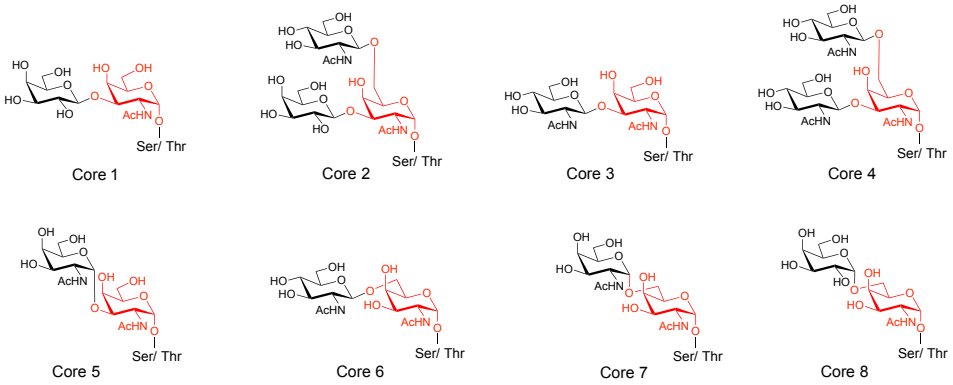


Figure 1-2: Core 1-8 *O*-glycans all having common GalNAc core which is highlighted in red.

N-Linked glycosylation being a major and more complex protein modification have a common pentasaccharide core $Man_3GlcNAc_2$ and are divided into three main types namely high mannose-type, hybrid-type and complex-type (Figure 1-3).[3]

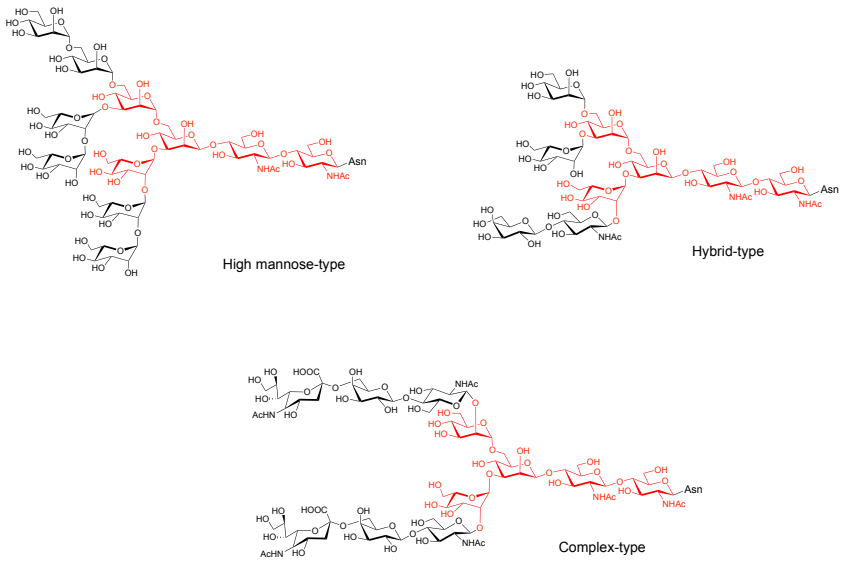


Figure 1-3: The three common types of *N*-glycans with $Man_3GlcNAc_2$ common core highlighted in red that are linked to Asn of consensus sequence Asn-X-Thr/Ser.

Chapter 1

1-2. Biosynthesis of glycoproteins

Biosynthesis of secretory proteins and membrane-bound proteins begins in the cytoplasm and the nascent peptide chain simultaneously enters the lumen of endoplasmic reticulum. Glycosylation occurs in the lumen of endoplasmic reticulum (ER) and also in Golgi apparatus. Almost all of membrane-bound proteins and secretory proteins have one or more carbohydrate chains.

1-2-1. Assembly of glycan and its transfer to nascent polypeptide

N-Linked glycosylation in ER occurs in two steps 1) assembly of a glycan molecule and 2) transfer of the glycan to a nascent polypeptide. In step 1, biosynthesis of a 14-sugar *N*-glycan precursor core takes place which is assembled on ER-membrane anchored lipid called dolichol phosphate (Dol-P). This step is initiated towards the cytoplasm of ER membrane and sequential assembly of heptasaccharide ($Man_5GlcNAc_2$) on Dol-P takes place that are catalyzed by multiple glycosyltransferases. It is then flipped towards the ER lumen and the remainder of the glycan are sequentially added to generate $Glc_3Man_9GlcNAc_2$-P-P-Dol tetradecaoligosaccharide (Figure 1-4).[4] In the second step, the membrane-bound oligosaccharyl transferase (OST) complex which is in proximity to the translocon pore secretory 61 (SEC61) catalyzes the transfer of 14-sugar *N*-glycan from Dol-P to the Asn of NXS/T of nascent polypeptide as soon as it emerges into ER lumen through SEC61 (Figure 1-4).[4]

1-2-2. *N*-Glycans in protein folding and quality control

Protein folding in ER begins co-translationally with the assistance of molecular chaperones and continues till it acquires a native structure, which are then trafficked to Golgi for further processing or trafficked to their destination. As the nascent polypeptide chain enters the lumen through SEC61 translocon complex it gets associated with the molecular chaperones that assist translocation and promote folding and also with those chaperones that target it for degradation.[4] The chaperones ensure that protein remains in the ER until its maturation is completed. As the transfer of *N*-glycan occurs, α-glucosidase I (GS-I) immediately acts on it and removes the terminal glucose that prevents rebinding of the processed *N*-glycan by OST and promotes binding to malectin, which can recognise misfolded proteins.[4] After the quality control check of malectin, a multi-subunit α-glucosidase II (GS-II) removes another glucose generating $Glc_1Man_9GlcNAc_2$

(Figure 1-4). This makes the nascent polypeptide gain affinity towards CNX/CRT-ERp57 chaperone system. Calnexin (CNX) is a membrane bound protein and Calreticulin (CRT), its soluble homolog consists a sugar-binding globular domain and a proline rich P-domain that weakly associates with ERp57, a protein disulphide isomerase. This chaperone system protects nascent chains from aggregation and facilitates in achieving a native conformation and correct disulphide pairings.[4]

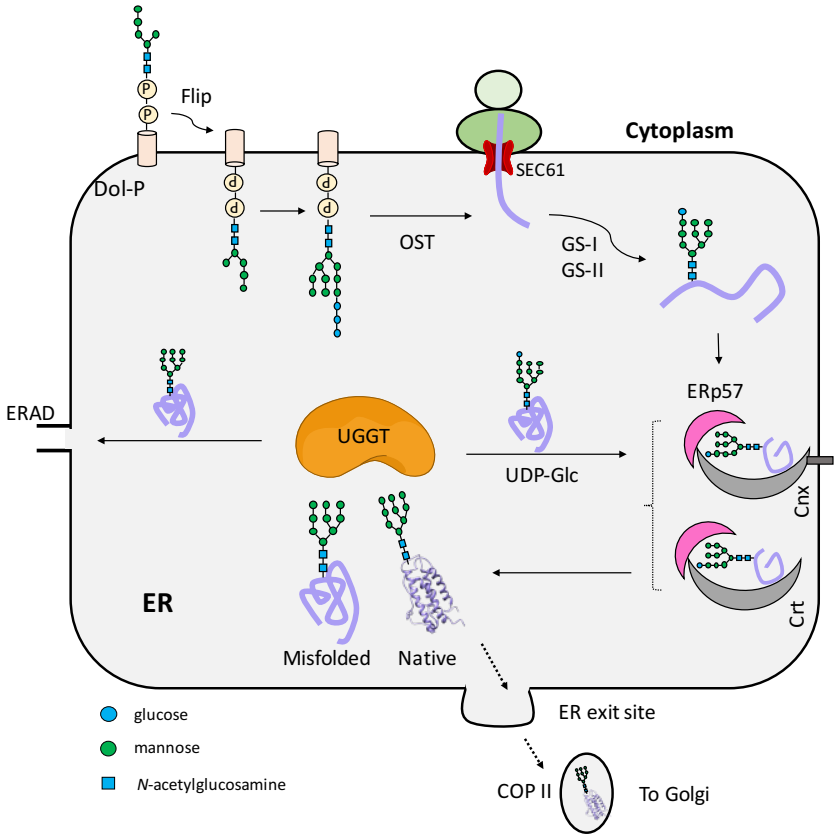


Figure 1-4: Biosynthetic pathway of N-glycosylation in ER involving transfer of N-glycan to nascent polypeptide, its processing and its association with chaperones that assist in folding.

A second deglucosylation by GS-II prevents rebinding to CNX/CRT-ERp57 complex and allows binding to UDP-Glc:glycoprotein glucosyltransferase 1 (UGGT1) which ascertains whether the protein has achieved its native conformation. Completely folded proteins are allowed to move from ER to Golgi for further processing if necessary. In contrast, non-native folds recognised by UGGT1 through solvent-exposed hydrophobic residues are reglucosylated by UDP-Glc, which again enters CNX/CRT-ERp57 cycle (Figure 1-4).[4] BiP is another chaperone that is abundantly present in ER which is peptide- dependent ATPase that

binds to proteins that lack oligosaccharides and assist them to attain native conformation.[4] If a protein fails to attain native conformation after repeated interactions with CNX/CRT, BiP or other chaperones of ER, it is targeted for degradation by ER-associated degradation (ERAD) pathway.

The transportation of proteins from ER occurs at ER exit sites by forming small membrane clusters contiguous with ER membrane and are coated with COPII coat. Incompletely folded cargo proteins and chaperones are generally excluded from the exit sites (Figure 1-4).[4]

O-Linked glycosylation does not have a consensus sequence and some Ser/Thr residues are specifically modified with GalNAc residues in the Golgi. Depending on the organism, unique *O*-glycosylations with mannose, glucose, xylose, fucose or GlcNAc occurs in the ER. The most abundant type of *O*-glycans are mucins initiated by GalNAc-Ser/Thr and the glycosaminoglycans initiated by Xyl-Ser. These can be further extended in the Golgi if the protein is trafficked there.[4]

1-2-3. Processing of glycoprotein in Golgi

Processing of glycoproteins in the Golgi takes place by glycosyl transferases that are distributed throughout the Golgi yielding hybrid-, complex- and high mannose-type *N*-glycans (Figure 1-3) and/or linear or branched *O*-GalNAc glycans (Figure 1-2). Glycoproteins with *N*-glycans that arrive in the *cis*-Golgi are of high mannose type and usually contain eight or nine Man residues (Figure 1-4). There are more than 250 glycosyltransferases (GTs) in the Golgi which catalyze the transfer of one sugar to another sugar on a glycan acceptor.[5] Since each GT transfers sugar to a specific acceptor generated by preceding GTs and in a particular manner, they must be appropriately distributed among the *cis*-, medial-, *trans*-Golgi or the *trans*-Golgi network (TGN) (Figure 1-5). α-mannosidases in the *cis*-Golgi remove Man residues to generate $Man_5GlcNAc_2Asn$ which then becomes the substrate for GT GlcNAcT-I present in medial-Golgi. GlcNAcT-I transfer GlcNAc to $Man_5GlcNAc_2Asn$ and initiates synthesis of hybrid and complex *N*-glycans. In case of hybrid *N*-glycans, the arm that receives GlcNAc is extended by Gal, sialic acid and/or other sugars to form hybrid *N*-glycans. In case of complex *N*-glycans, three Man residues of the core remain and is modified with two GalNAcs which may be further branched by addition of various sugars. Glycans that emerge from TGN act as tags that are essential for the proteins to be trafficked to their respective destinations.[5]

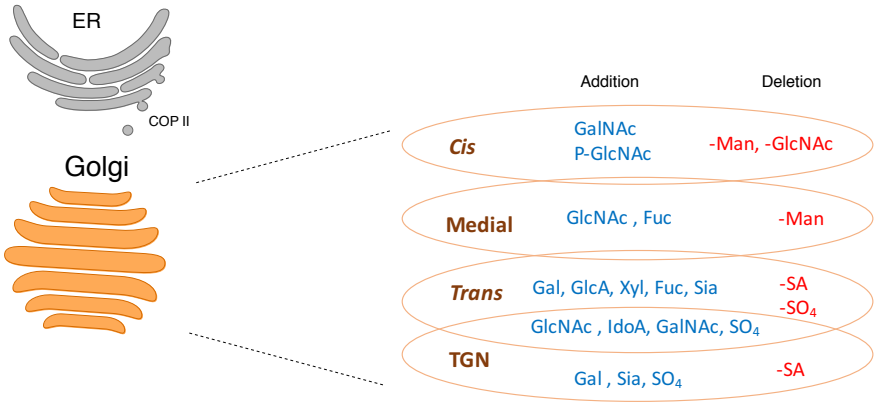


Figure 1-5: Processing of *N*-glycans in Golgi apparatus showing addition and deletion of various sugars in the various compartments of Golgi.

Modifications of GalNAc of Ser/Thr in Golgi occurs by addition of Gal, GlcNAc, sialic acid and fucose to form linear or branched *O*-GalNAc glycans.[5] Currently, much work is devoted in determining the consensus sequence for *O*-GalNAc addition.

1-3. Importance of glycoproteins

The study of structure and function of wide range of glycoproteins encoded by human genome answers many problems in biology and medicine, but is limited due to the micro-heterogeneity of glycoproteins. Oligosaccharides on glycoproteins have a potential role on protein stability, folding, trafficking and life time, immunogenicity, cell growth, cell-cell adhesion, as well as its primary functional activity.[6] The studies involving the effect of oligosaccharides on proteins is limited by the fact that glycosylation is not mediated by any template and thus characterized by its heterogeneity. There exists a lot of variations among the oligosaccharide units that are attached to the same polypeptide and are called glycoforms. Thus, the isolation of glycoproteins from natural sources is often tedious and impractical.

Therefore, protein chemists are working towards obtaining homogeneous natural glycoproteins, and not only that, they are also trying to create new proteins either by *de novo* design[6] or by altering the natural proteins. These can be possible only if one is able to access and manipulate proteins.

1-4. Obtaining glycoproteins by recombinant expression

Glycoproteins that are of therapeutical value like antibodies, blood factors, interferons etc., are often prepared by recombinant expression using cells with glycosylation capabilities, as it is cost-effective to get large quantities of glycoproteins. When genes encoding human proteins are recombinantly expressed in a host organism which does not naturally have this gene or gene fragment, the gene is called heterologous gene. the host organism/cell used for recombinant expression is called heterologous expression system. The biological activity of many therapeutic glycoproteins depends on their glycosylation status in terms of immunogenicity, product equivalence etc. However, glycoproteins obtained by recombinant expression are heterogeneous in their oligosaccharide structure, which has been a hindrance to identify the critical role of oligosaccharide for biological activity. In order to improve these drawbacks, several cell expression systems have been studied to prepare homogeneous glycoproteins. In eukaryotic expression systems, the initial steps of *N*-glycan synthesis that occurs in endoplasmic reticulum is conserved, while *N*-glycan processing and *O*-glycan biosynthesis in the Golgi apparatus differs.[6] This gives rise to a range of non-human type glycosylation patterns and currently, various eukaryotic systems are continuously engineered to get human type glycosylation.

Some examples of cell expression system are mentioned as follows.

1) The most common mammalian (non-human) cell lines used include Chinese hamster ovary cells (CHO), baby hamster kidney (BHK21) cells. However, these cell lines produce PTMs that are not expressed in humans and become immunogenic, e.g. galactose-α-1,3-galactose and *N*-glycolylneuraminic acid (NGNA) (Figure 1-6).[7] Recent advances in technology have allowed for increased productivity with human cell lines e.g. Human Embryonic Kidney 293 cells (HEK 293), that can produce proteins most similar to those synthesized naturally in humans.

2) Insect cells infected with baculovirus are also most efficient systems to produce glycoproteins, mainly for development of vaccines. The heterologous genes are well expressed and folded and the PTMs are often identical to that which occurs in mammalian cells. The major insect processed proteins have either high mannose or paucimannose residues that do not elongate further (Fi gure1-7).[8] Another possibility is that they produce a core α-1,3-fucosylated *N*-glycan and this is known to be allergenic in humans. However, there have been developments to engineer insect cells to produce proteins with human type glycosylation.

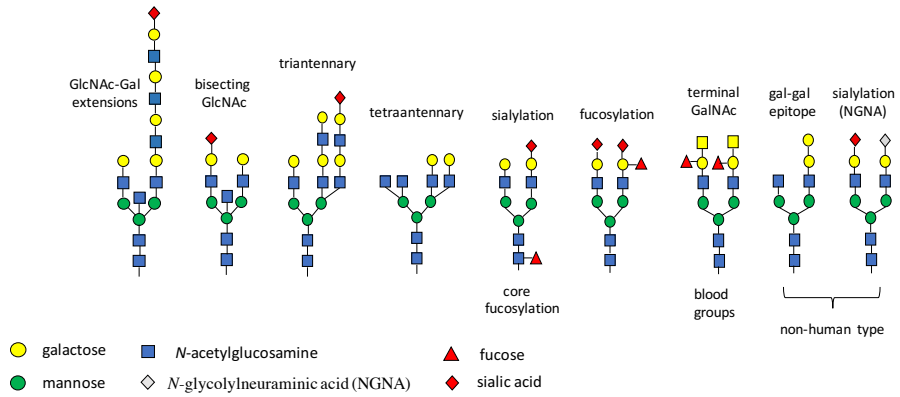


Figure 1-6: Diverse range of *N*-glycan modifications observed in mammalian cell lines including non-human type.

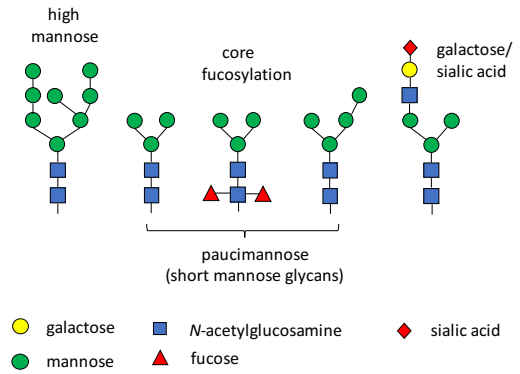


Figure 1-7: Major glycosylation pattern observed in insect cell lines are high mannose or paucimannose structures.

3) Yeast cells such as *Saccharomyces cerevisiae, Pichia pastoris* and *Kluyveromyces lactis* are preferred as they are relatively easy to grow in fermenters and thus, it is easy to scale-up and obtain glycoproteins in high yields. Their biosynthetic pathway is similar to higher eukaryotes, however, *N*-glycans significantly differ from those of mammalian cells and humans as they produce hypermannosylated structures (Figure 1-8).[9] Advances are being made to eliminate mannosylation and humanization of yeast glycosylation.

4) Plants are emerging as a powerful expression system for biopharmaceuticals due to their fast, high yield and low-cost production of pharmaceutical proteins. They show similar or even higher biological activity than protein homologs expressed in cultured mammalian cells. However, processing of *N*-glycan in Golgi gives rise to complex *N*-glycans with β-1,2-linked xylose, core α-1,3-linked fucose

which are immunogenic to human beings and research is going on to optimize glycan composition for low immunogenicity (Figure 1-8).[10]

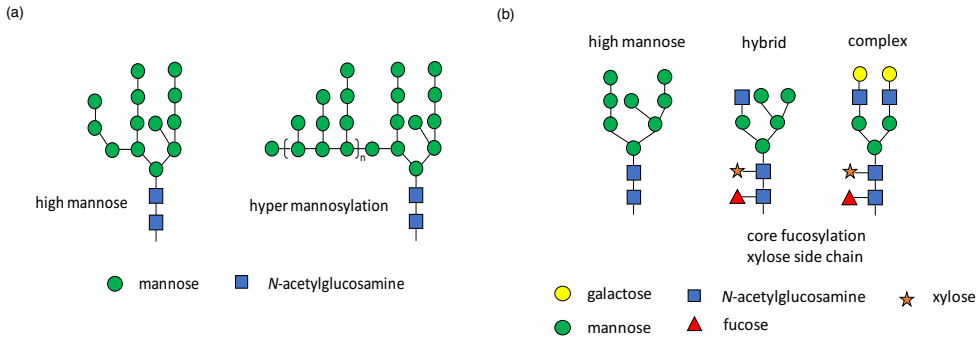


Figure 1-8: (a) *N*-Glycans obtained from yeast cell lines. (b) *N*-Glycans obtained from plants.

1-5. Preparation of homogeneous glycoprotein

Homogeneous glycoprotein can be obtained either by total chemical synthesis or semisynthesis, that are the most promising and efficient strategies. These strategies allow access to protein molecules having unnatural modifications, which are impossible to be obtained by standard ribosomal synthesis and also natural modifications. Total chemical synthesis as the name suggests, is the use of chemical methods to assemble proteins, while semisynthesis is the combination of chemical synthesis and recombinant expression and are explained in detail as follows.

1-6. Total Chemical Synthesis

Chemical synthesis of proteins allows modification of proteins at the atomic-level without limiting the number and kinds of modification which mainly involves solid phase synthesis of peptide fragments that are ligated together to get full-length protein (Figure 1-9).[11] Earlier days of peptide synthesis was performed by conventional solution based synthesis that gave highly pure peptides. But it was highly labor-intensive and required extensive knowledge and was limited. The introduction of solid phase peptide synthesis (SPPS) by Merrifield in 1963 greatly revolutionized the field of chemical protein synthesis.[11] SPPS involves stepwise addition of amino acids from *C*-terminal to *N*-terminal direction, where the first amino acid to be synthesized has its amino group protected and is connected *via* its carboxyl group to an insoluble solid support called resin. A molecule called linker connects the resin with the peptide. The amino protecting group of the first amino acid is removed after coupling to the linker and the second amino acid coupled. The peptide chain is

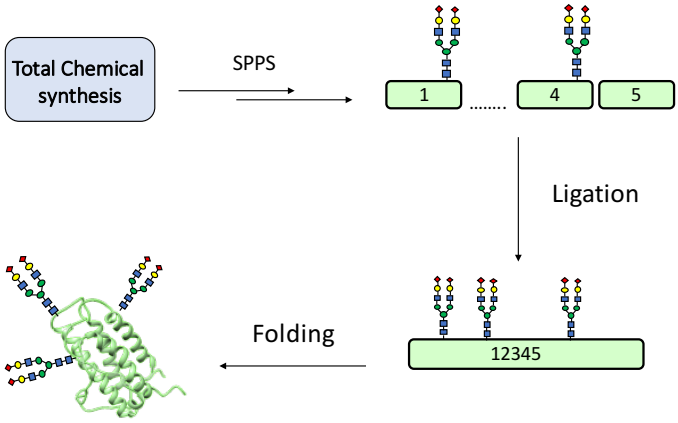


Figure 1-9: Total chemical synthesis of glycoproteins involves synthesizing the glycoprotein as fragments with desired modifications and ligating the fragments and folding to get the native form.

elongated by repetitive deprotection of amino group and coupling of amino acids with protected amino group and activated carboxyl group. After the assembly of peptide, all the protecting groups are removed in most cases and the covalent bond between peptide and linker is cleaved as shown in (Figure 1-10). Through SPPS,

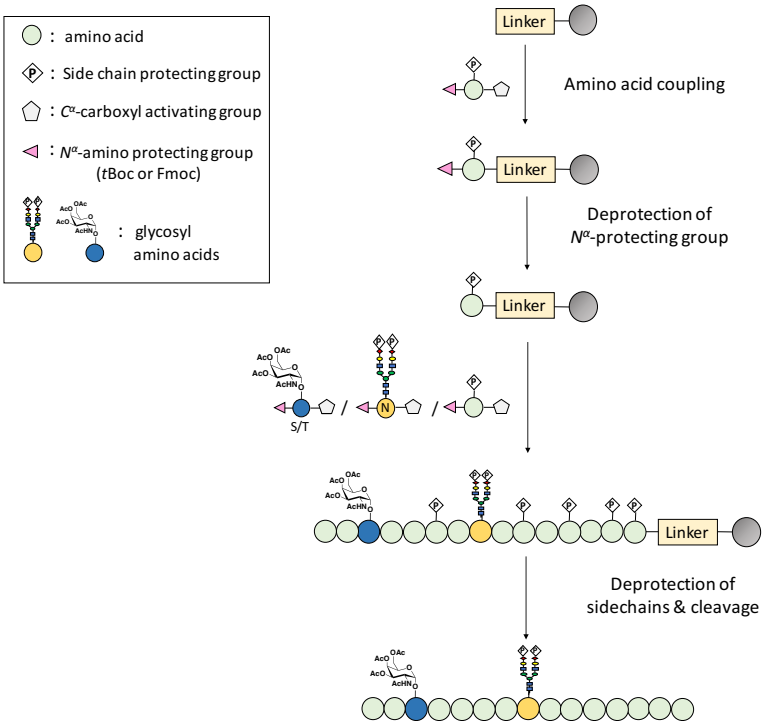


Figure 1-10: A general schematic representation of SPPS of (glyco)peptide.

the coupling reactions are forced to complete by using excess of soluble reagents that allows easy separation of the reagents and by-products during coupling by filtration. The dissolved peptide is separated from resin by filtration.

The most common strategy is to use N^{α}-protecting group *tert*-butoxycarbonyl (Boc) group that is acid sensitive in combination with benzyl-type semipermanent side-chain protection or a base-sensitive 9-fluorenylmethyloxycarbonyl (Fmoc) group in combination with *tert*-butyl type side-chain protection.[11] SPPS also enables linear chemical synthesis of glycopeptides using Asn linked *N*-glycan or Ser/Thr linked *O*-glycan.

1-6-1. Chemoenzymatic synthesis of oligosaccharides

Glycans isolated from naturally occurring sources show micro-heterogeneity and therefore chemical and chemoenzymatic synthesis of oligosaccharides is employed to obtain homogeneous form. Synthetic technologies for the chemical synthesis of large oligosaccharides has greatly advanced but, it is time-consuming and low yielding due to repetitive protection/deprotection of hydroxy groups and stepwise formation and purification of stereoselective glycosidic linkages. Therefore, semisynthetic method can afford large quantities of pure oligosaccharides. The isolation of Asn-linked biantennary complex-type sialylundecasaccharide and Asn-linked high mannose-type oligosaccharide in high yields from egg yolk is well established by the Kajihara group (Figure 1-11).[12] Recently, an efficient semisynthetic strategy to obtain complex-type triantennary oligosaccharides from biantennary oligosaccharide in just ten steps was reported by Maki from Kajihara group.[12] The Asn linked *N*-glycans can then be used for glycopeptide synthesis to get homogenous glycoproteins (Figure 1-10).

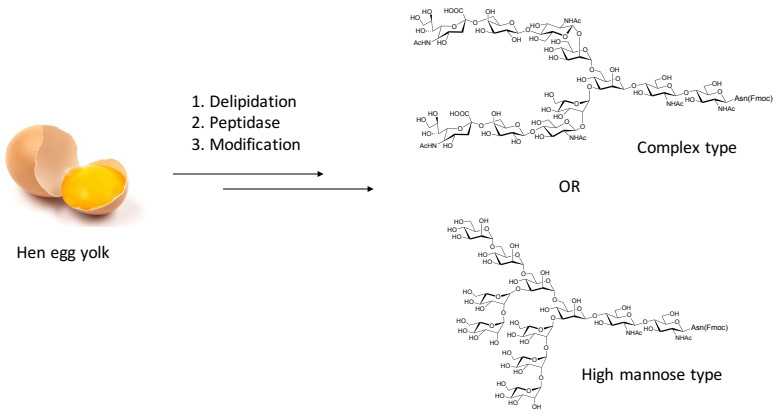


Figure 1-11: Isolation of complex type and high-mannose type *N*-glycans from hen egg yolk.

O-Glycans on the other hand can be introduced into peptide by various strategies. The most common approach is to use a suitably protected *O*-glycosyl amino acid building block in SPPS (Figure 1-10). If complex *O*-linked glycosyl amino acids are to be obtained, then the desired GalNAc-α-*O*-Ser/Thr is installed prior to elaboration of additional sugars from the core. Generally, Fmoc-based chemistry is preferred over Boc-based chemistry due to the acid sensitive glycosidic bonds. An alternative approach is to use GTs to transfer individual monosaccharides to preformed glycopeptides containing simple *O*-linked glycans (Figure 1-12).[13] Another attractive approach towards complex *O*-glycans, is to attach preformed oligosaccharides to simple glycopeptides by chemoselective ligation. This approach offers advantages similar to enzymatic approach, in addition, much broader range of substrates can be used as coupling partners (Figure 1-12).[13]

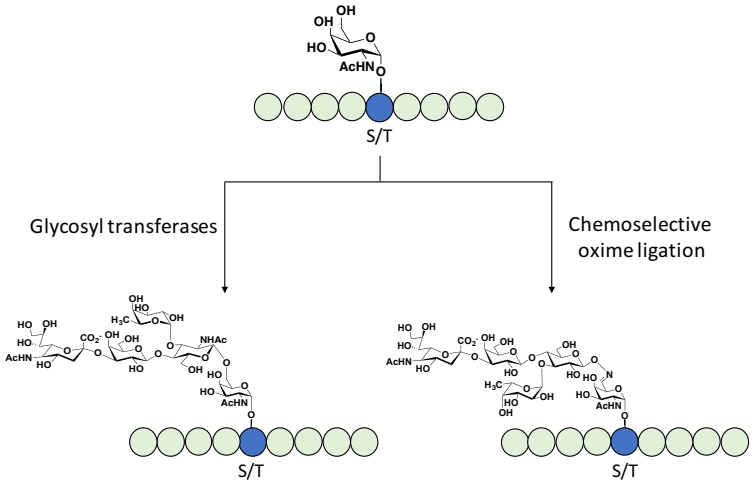


Figure 1-12: Elaboration of *O*-glycans on preformed simple *O*-glycopeptide by use of GTs or by chemoselective ligation, for example, ligation through oxime formation.

1-7. Native Chemical Ligation (NCL)

Native chemical ligation is the most efficient strategy to ligate unprotected peptide fragments which involves chemoselective reaction between *C*-terminal thioester peptide and *N*-terminal Cys peptide to get full-length proteins. In 1953, Wieland et al., discovered the synthesis of dipeptide in aqueous buffer which proceeded through a thioester intermediate.[14] This reaction was developed later in 1994, when the Kent group introduced a method to ligate large peptide fragments called native chemical ligation.[15] The reaction begins with a reversible transthioesterification between an α-thioester peptide and *N*-terminal Cys peptide at neutral pH followed by an irreversible and spontaneous intramolecular *S*→*N* acyl shift to generate a more

stable native amide bond at the ligation junction (Figure 1-13). The presence of internal Cys within peptide do not affect the reaction as transthioesterification is a reversible process, and only *N*-terminal Cys that has an α-amino group reacts irreversibly to form peptide bond. The rate of reaction depends on the amino acid at the *C*-terminal thioester (Gly reacts faster than bulky amino acids) and also on the nature of the thioester (aryl thioesters was found to react faster than alkyl thioesters).[16]

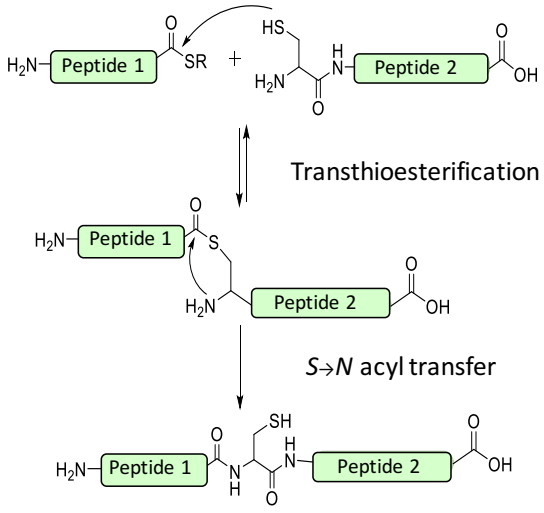


Figure 1-13: NCL mechanism involving a reversible transthioesterification followed by *S*→*N* acyl shift.

1-7-1. Limitations of NCL

NCL being the most promising and widely used methodology for the ligation of peptide fragments is limited by its requirement of Cys at the ligation site. The frequency of occurrence of Cys when primary structures of 1021 unrelated proteins were compared was found to be only 1.7 %,[17] thus limiting the application of NCL. Following the logic of NCL, various advancements have been made to perform ligation by using other amino acids as Cys surrogates or by using thiol auxiliaries.[11]

A strategy that involves Met instead of Cys at the ligation site was reported by Tam and Yu where homocysteine (Hcs) is used which is converted to Met after ligation and is called homocysteine/methionine ligation.[18] Transthioesterification with Hcs takes place followed by 1,5 *S*→*N* acyl transfer forming a six-member ring transition state unlike a five-member ring in conventional NCL. Then homocysteine is

converted to Met by methylation (Figure 1-14). This study by Tam revealed that regardless of ring size, acyl rearrangement takes place spontaneously due to the proximity-driven effective concentration of α-amino group of the amino acid. This laid the foundation for employing β-and γ- mercapto-amino acids as Cys surrogates.

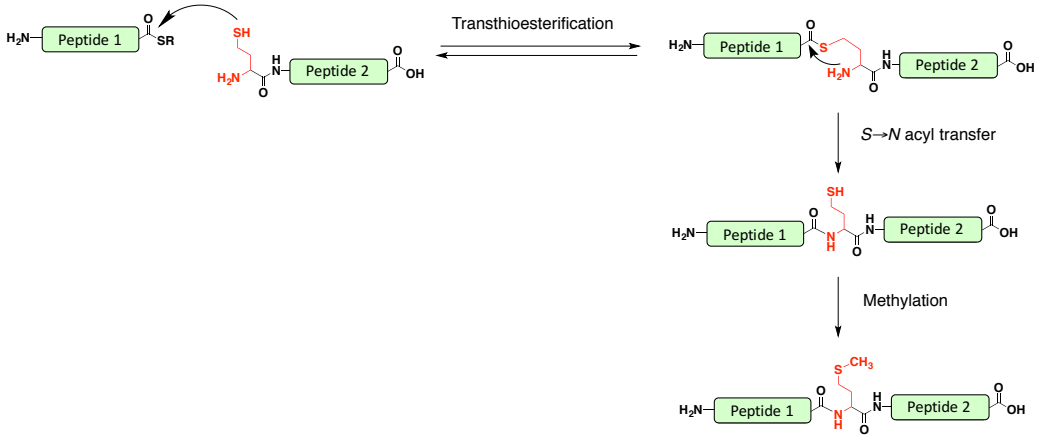


Figure 1-14: Methionine ligation mechanism involving use of homocysteine which is converted to Met after ligation.

1-8. Native chemical ligation-desulfurization

The introduction of desulfurization to convert Cys to Ala greatly expanded chemical protein synthesis by the use of various unnatural mercapto amino acids to perform NCL. A turning point to extend NCL to other amino acids at ligation site was achieved when Dawson and Yan introduced catalytic desulfurization to convert Cys to Ala.[19] This allowed access to more synthetic targets as Ala is more abundant in protein sequences i.e 7.8 %.[17] They used Raney Ni or Pd/Al_2O_3 under a H_2 atmosphere for desulfurization of Cys at ligation site.[19] However, this approach was limited due to the lack of selectivity resulting in hydrogenation of Trp and demethythiolation of Met residues and also low product recovery due to adsorption of peptides on metal surfaces. Later, Wan and Danishefsky reported an alternative to the metal-based desulfurization based on a free radical approach which involved TCEP and VA-044, a free radical initiator under aqueous conditions.[20] This led to the rapid development of ligations at various other amino acids over the years.

Several research groups have worked towards synthesis of various unnatural β-mercapto amino acids to mediate NCL-like ligation and unnatural γ-mercapto amino acids to mediate homocysteine-like ligation.[21] Till date β-mercaptophenylalanine (i),[22] β- and γ-mercaptovaline (ii),[23] γ- and δ-mercaptolysine (iii),[24] β-mercaptoleucine (iv),[25] γ-mercaptoproline (v),[26] β-mercaptothreonine (vi),[27] β-mercaptoarginine (vii),[28] γ-mercaptoglutamine (viii),[29] β-mercaptoaspartic acid (ix),[30] 2-mercaptotryptophan (x)[31] have been successfully reported to be used as Cys surrogates at the ligation site (Figure 1-15).

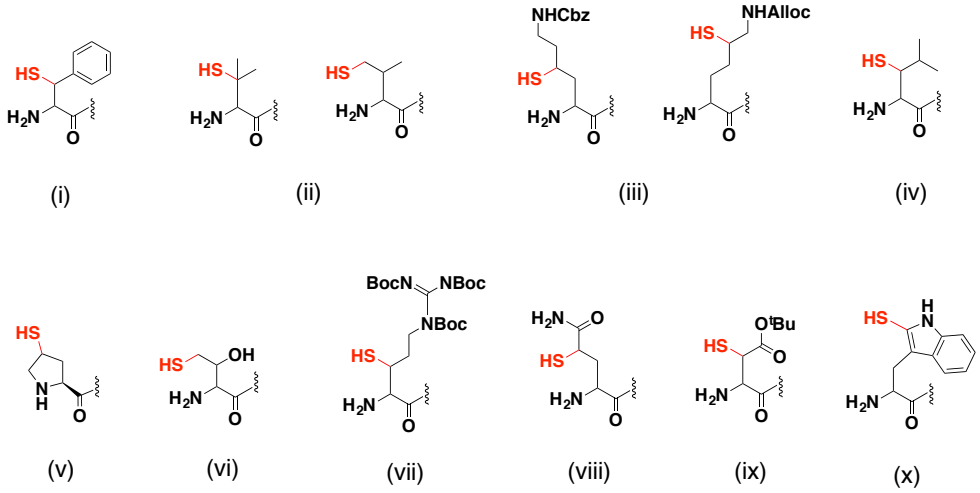


Figure 1-15: unnatural β-, γ- and δ- mercapto amino acids that act as Cys surrogates in NCL.

But still there exists a drawback, that is selective-desulfurization of thiol derived amino acids is not possible in the presence of Cys residues, if there is any. This issue was addressed by Pentelute and Kent by using Acm-protecting group on Cys, which could be removed after desulfurization thus allowing selectivity (Figure 1-16).[32] A potential approach to achieve selectivity in the presence of unprotected thiols is by using selenocysteine (Sec) for peptide ligation.

1-9. Native Chemical Ligation at Selenocysteine (Sec)

Another potential surrogate for Cys is Sec which allows a much faster chemoselective NCL and selective deselenization in the presence of free Cys due to its unique chemical properties. Selenocysteine (Sec or U) is the 21st natural amino acid that is present in a number of naturally occurring enzymes among

archaea, eubacteria and eukarya.[33] During biosynthesis of selenoproteins, Sec is inserted cotranslationally by its own $tRNA^{Sec}$ that recognizes the opal stop codon UGA assisted by selenocysteine insertion sequence (SECIS). The mechanism differs among eukaryotes and prokaryotes and therefore, Sec cannot be incorporated site-specifically by using heterologous bacterial expression with ease.[34] However, incorporation of Sec through chemical synthesis proves to be more efficient.

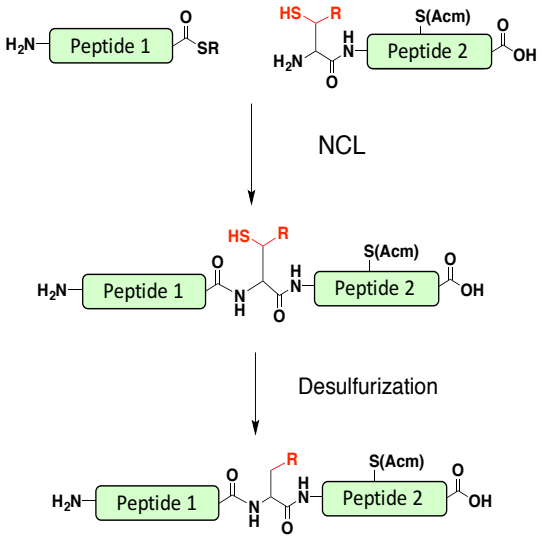


Figure 1-16: Mercapto amino acids in NCL followed by free-radical desulfurization using TCEP and VA-044 in the presence of Acm protected Cys.

1-9-1. Properties of Selenocysteine

The physicochemical properties of Sec are similar to Cys, however, with key differences in pKa and nucleophilicity that effects reactivity. The pKa of Cys is 8.25 and that of Sec is approximately 5.24-5.63 indicating that Sec is more acidic and exists primarily as a selenolate at physiological pH.[35] Selenolates have also proven to be more nucleophilic and a superior leaving group than their corresponding thiolates.[35] At lower pH (5.0-6.0), NCL can proceed chemoselectively at Sec and is 10^3 times faster than that with Cys.[35] Therefore, Sec has been employed in the synthesis of various peptide fragments and full-length proteins. Selective deselenization of Sec to Ala in the presence of Cys was demonstrated by Dawson and coworkers[36] by treating with excess of TCEP and DTT at ambient temperature which is similar to free-radical desulfurization (Figure 1-17).

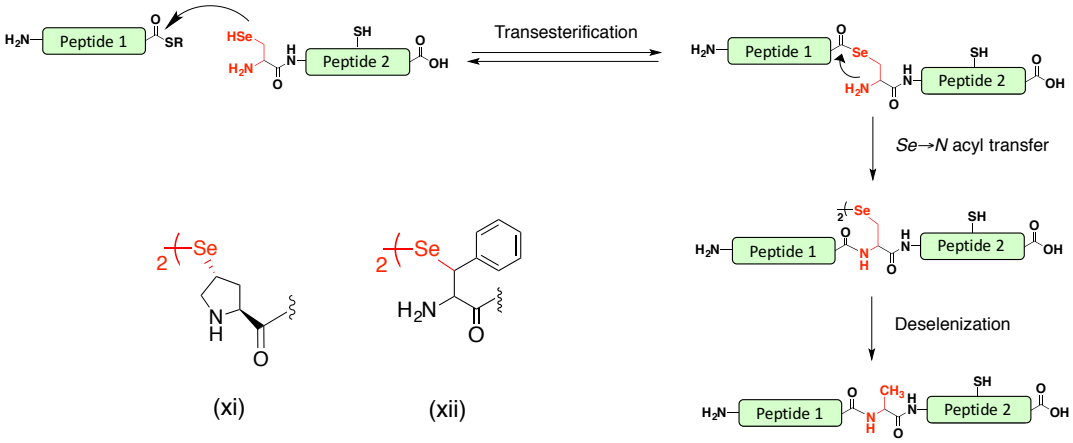


Figure 1-17: NCL with Sec followed by selective deselenization using TCEP and DTT in the presence of Cys to get Ala; other selenated amino acids γ-selenoproline (xi) and β-selenophenylalanine (xii)

The advances in NCL using Sec has further led to its expansion by exploring selenated amino acids γ-selenoproline (xi)[37] and β-selenophenylalanine (xii)[38] (Figure 1-17). Recently Payne reported oxidative deselenization of Sec to obtain Ser in a single step within unprotected peptides using oxone.[39]

1-10. Synthesis of peptide α-thioesters

Peptide α-thioesters are the key intermediates for the synthesis of proteins, which are readily obtained by Boc-SPPS and less readily by Fmoc-SPPS. Conventional protocol to obtain thioesters is by Boc-SPPS on a thioester functionalized resin. However, the harsh cleavage conditions using HF for final side-chain deprotection in Boc-SPPS is problematic for acid-sensitive glyco-, phospho-peptides etc. In those cases, a modified *in situ neutralization* Boc-SPPS[40] protocol can be employed where TfOH is used for final side-chain deprotection instead of HF and acid-labile glycopeptide α-thioesters can be obtained (Figure 1-18).

The use of Fmoc-SPPS for peptide α-thioester is limited due to the instability of thioester linkage under basic conditions used during coupling. The first attempt was to perform thioesterification of carboxylic group in solution phase of partially protected peptide obtained by Fmoc SPPS.[41] This suffered from the drawback of solubility and epimerization at C-terminal. Later, milder deblocking reagents for N^{α}-Fmoc

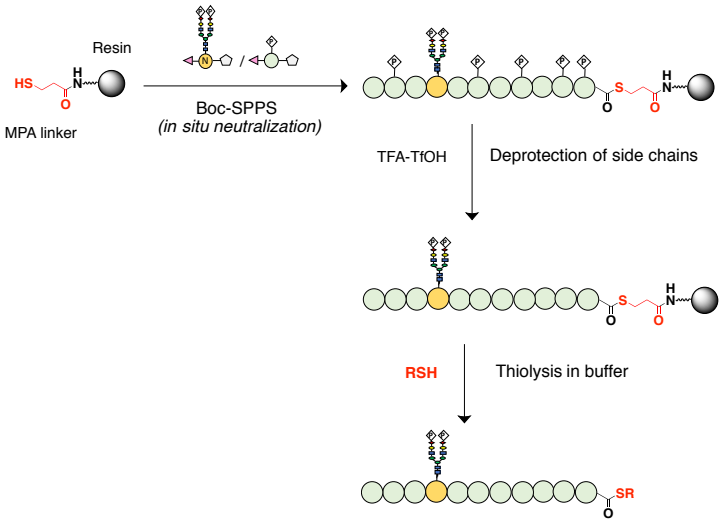


Figure 1-18: Schematic representation of glycopeptide α-thioester synthesis by Boc-SPPS using 3-mercaptopropionic acid (MPA) linker.

deprotection was employed and it suffered from lower yields.[42] Other strategies to obtain α-thioesters includes anchoring of peptide backbone via backbone amide linker (BAL) followed by C-terminal functionalization,[43] thiolysis of peptides bound to sulfonamide safety-catch linkers (Figure 1-19).[44] Recently, strategies to synthesize peptide thioester surrogates by Fmoc-SPPS are being developed that can be directly used for native chemical ligation.[45]

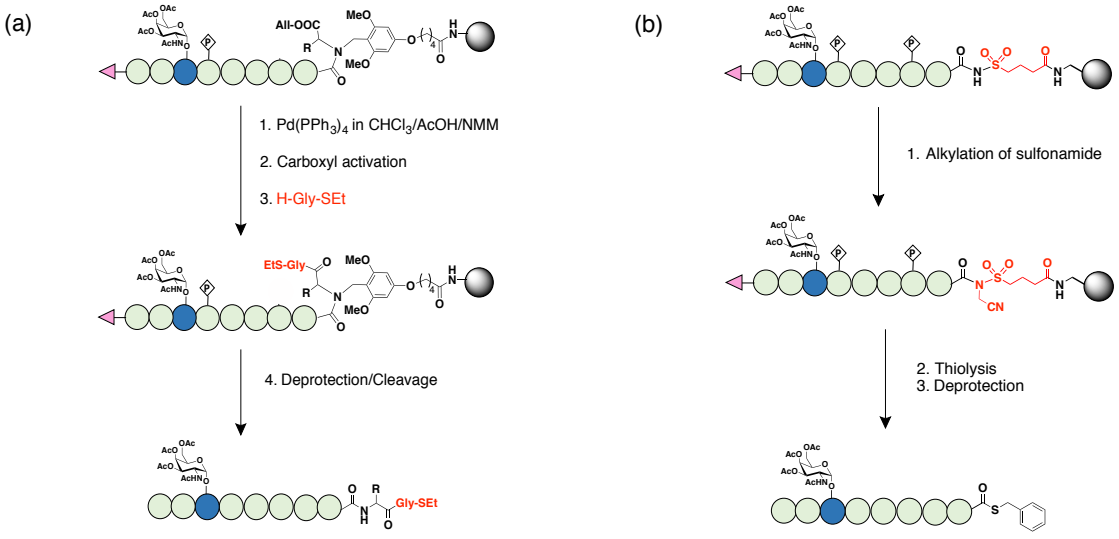


Figure 1-19: Schematic representation of (a) BAL strategy and (b) sulfonamide safety catch linker strategy for synthesis of glycopeptide α-thioester by Fmoc-SPPS.

1-11. Synthesis of N-terminal cysteine containing peptides

N-Cys peptides can be obtained easily by Boc- or Fmoc-based SPPS. It can also be synthesized as an unreactive thiazolidine for use in NCL, which on treatment with methoxyamine at pH 5.0[46] gives free *N*-terminal Cys (Figure 1-20).

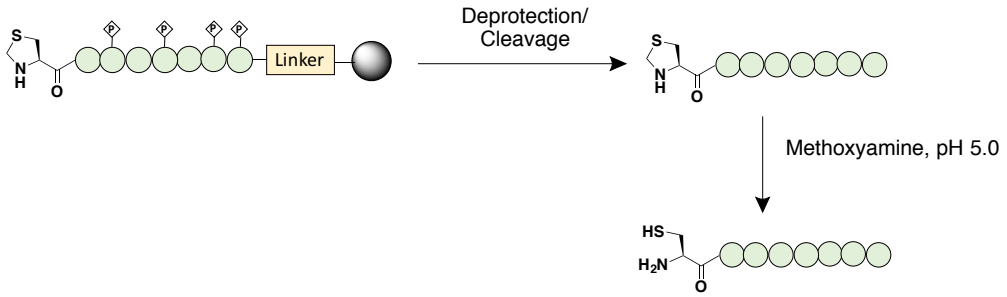


Figure 1-20: General schematic representation for obtaining *N*-Cys peptide by Boc- and Fmoc-SPPS.

1-12. Semisynthesis

Semisynthesis refers to a synthetic strategy where at least one of the building blocks is obtained by recombinant expression while other fragments are obtained by chemical synthesis, that reduces the number of chemical conversion steps and increased peptide yield achieved in a short time. Total chemical synthesis of larger proteins is cumbersome due to the restriction in size of peptide obtained by SPPS and also the difficulties associated with multiple NCL ligation steps, which results in low yield. These can be overcome by employing semisynthesis due is its ease in obtaining longer proteins in larger quantities and at a low cost. *E.coli* is frequently used as host since it is easier to handle, its inexpensive, well known genetics and availability of large number of compatible molecular tools which allows us to engineer the organism in a desired way.[47] Currently, the semisynthetic strategy for obtaining glycoproteins include a two fragment or three fragment approach based on the position of the amino acid to be chemically modified.[47] Two fragment approach is employed when the desired position of modification is within 50-60 amino acids of the N-terminus or the C-terminus which can be synthesized by SPPS (Figure 1-21).

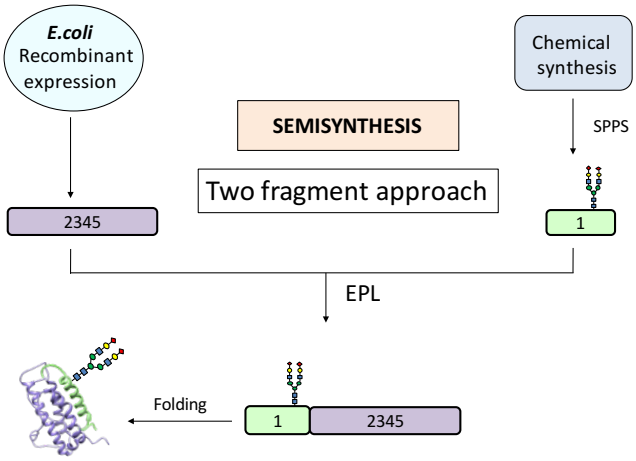


Figure 1-21: Representation of two fragment approach which is employed when either N- or C-terminus has to be modified.

Three fragment approach is employed when the desired position of modification is greater than 60 amino acids from either terminus (Figure 1-22). In this case, central peptide corresponding to the region of interest is obtained by chemical synthesis while the flanking segments are obtained by recombinant expression. The fragments are chemoselectively ligated together and is termed expressed protein ligation (EPL).

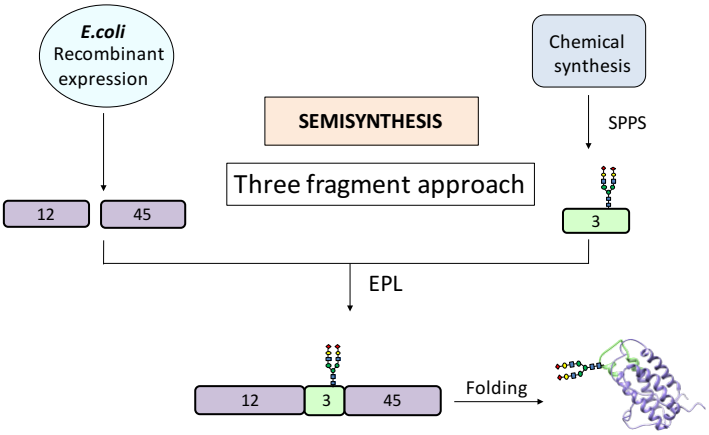


Figure 1-22: Representation of three fragment approach which is employed when middle region of target peptide has to be modified.

1-13. Expressed protein ligation

Expressed protein ligation (EPL) is a powerful methodology in protein semisynthesis that employs the principle of NCL.[47] In EPL, a synthetic peptide thioester can be ligated with a recombinant protein bearing *N*-terminal Cys residue allowing chemical modification at the *N*-terminus of target proteins or vice versa.

1-14. Recombinant polypeptide α-thioesters

Recombinant polypeptide α-thioesters are generally obtained by using engineered inteins and recently, strategies other than intein system are developed due to the difficulties in their handling. Inteins are self-processing domains that participate in naturally occurring post-translational process called protein-splicing.[47] In this process, a precursor protein undergoes a series of intramolecular rearrangements and internal reactions leading to precise removal of an internal segment called intein and ligating the flanking portions called exteins (Figure 1-23). Inteins are characterized by several conserved sequence motifs, while exteins have no sequence requirements.

Although the biological role of protein splicing is still not clear, the process has been extensively exploited. The inteins are mutated in a way that only first step of protein splicing occurs, for example Asn to Ala mutation at the C-terminal of intein (Figure 1-23). These mutated inteins when expressed with the desired proteins as *N*-terminal fusions can undergo intermolecular transthioesterification by external thiols giving rise to peptide α-thioesters.[47] The most common commercially available expression vectors with chitin binding domain (CBD) affinity purification tags are Sce VMA1 (*Saccharomyces cerevisiae* vacuolar ATPase) intein and Mxe GyrA (*Mycobacterium xenopi* DNA gyrase A) intein. Mxe GyrA intein can be efficiently refolded from chemical denaturants and can be reconstituted from bacterial inclusion bodies when compared to Sce VMA1.[47] The choice of external thiols to induce thiolysis depends on whether the peptide α-thioester is to be isolated prior to ligation. Ethanethiol produces ethyl α-thioester derivative which is quite stable allowing it to be purified and stored until use.[47] However, the unreactivity of simple alkyl thioesters during expressed protein ligation can be overcome by using thiol cofactors such as thiophenol or 2-mercaptoethanesulfonic acid (MESNA) that generate more reactive α-thioesters. Thiophenol and MESNA can be used in the initial protein-intein cleavage step if the peptide α-thioester need not be isolated.[47]

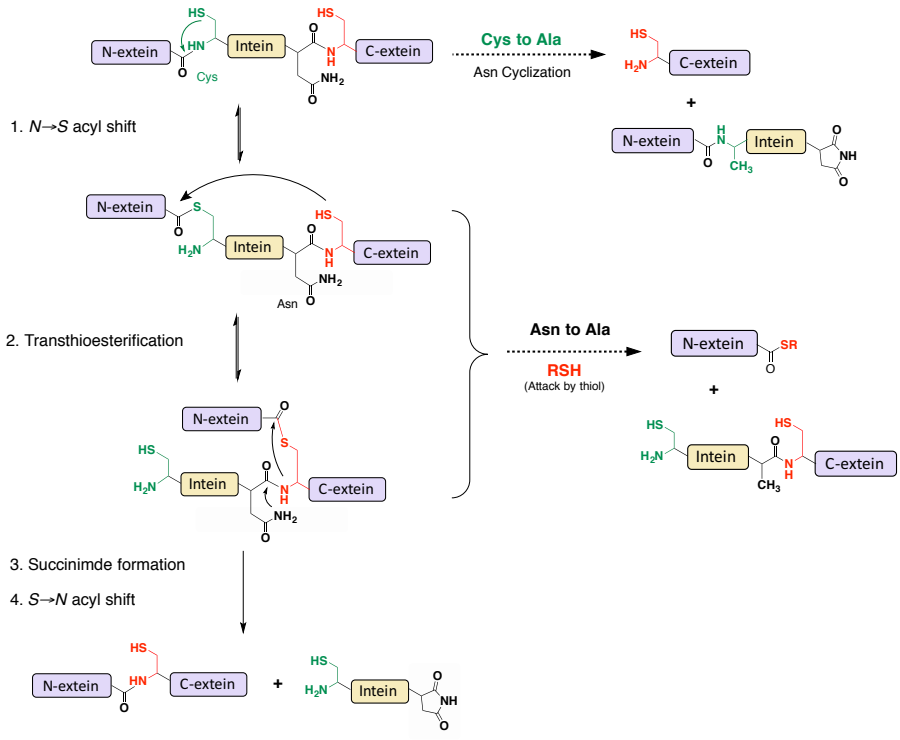


Figure 1-23: Mechanism of protein splicing. Mutation of Asn to Ala can generate C-terminal thioester peptide by using external thiols, while mutation of Cys to Ala generate N-terminal Cys peptide.

Recently, a chemical strategy has been explored by Okamoto from Kajihara group which can site specifically cleave and activate recombinant peptide to provide *N*-terminal peptide α-thioester.[48] In their new semisynthetic strategy towards Interleukin 13, they employed a three-fragment approach where a single fusion polypeptide provided both *N*-terminal α-thioester and *C*-terminal cysteine peptide and the central region bearing *N*-glycan was chemically synthesized (Figure 1-24). The method employed CNBr treatment of fusion polypeptide which selectively cleaves at the Met residue to obtain Cys peptide. The *N*-terminal fragment was modified by guanidine method to obtain α-thioester (Figure 1-24).

An alternative to recombinant peptide α-thioester was developed by the Lei Liu group by generating recombinant peptide α-hydrazides as ligation intermediates *via* genetic incorporation of an α-hydroxy acid.[49] They employed mutant pyrrolysyl-tRNA synthetase (PylRS), APCK-RS from *M.barkeri* to generate recombinant protein backbone oxoesters. The mutant PylRS can efficiently encode α-hydroxy acid into protein backbone by incorporating α-hydroxy acid analogues of pyrrolysine (Figure 1-25). Recombinant

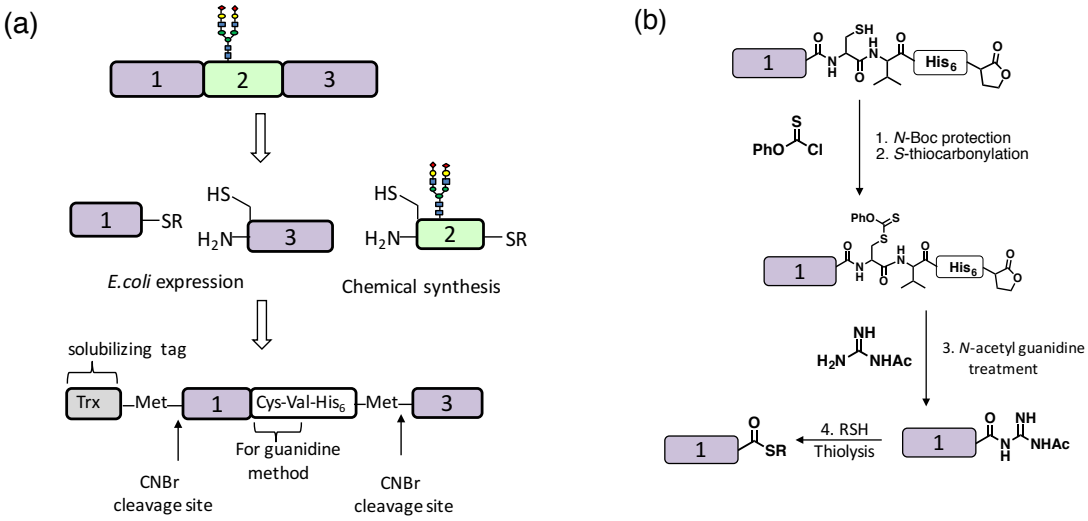


Figure 1-24: (a) retrosynthetic strategy for semisynthesis of IL-13. (b) Scheme of guanidine method to obtain thioester.

protein oxoesters obtained were converted to their corresponding α-hydrazides by treating with 4% aqueous hydrazine, which could be used for hydrazide-based protein ligation.[50]

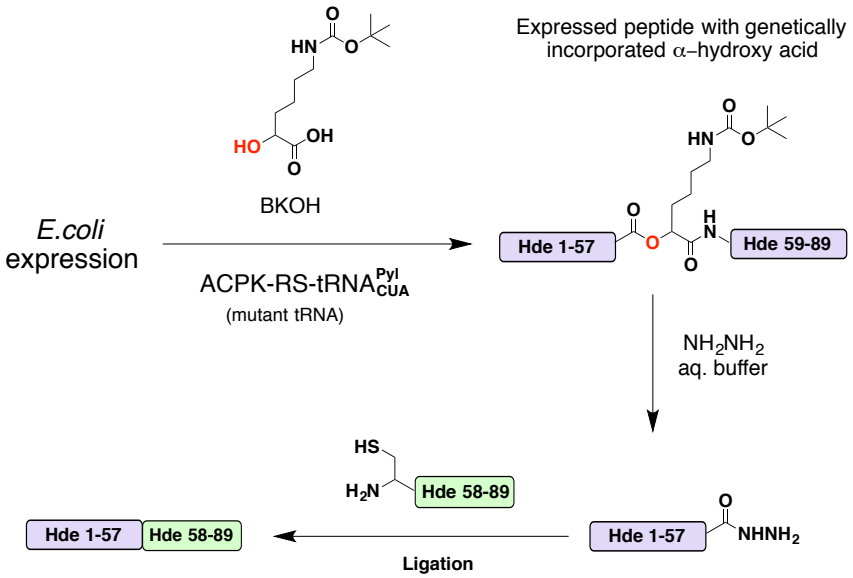


Figure 1-25: Scheme for semisynthesis of Hde by incorporating α-hydroxy acid using mutant pyrrolysyl-tRNA synthetase to obtain hydrazide peptide. BKOH is an analogue of pyrrolysine which gave good results and incorporation was efficient at 58th amino acid position for Hde.

1-15. Recombinant *N*-terminal Cys peptide

Recombinant *N*-terminal Cys peptide can be readily obtained by one of the following methods that includes enzymatic cleavage of a fusion protein, chemical cleavage using CNBr and mutated inteins (Figure 1-26).[47] The simplest form is to introduce Cys next to the initiating Met in the protein sequence which after translation is recognized by endogenous methionyl aminopeptidases (MetAPs)[51] that remove Met to give N-terminal Cys peptide *in vivo*. However, this process is inefficient resulting in lower yields.

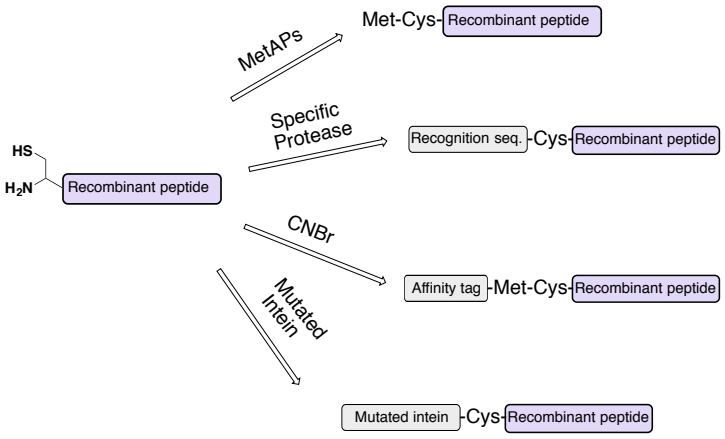


Figure 1-26: General approaches to obtain recombinant N-terminal Cys peptide.

More commonly practiced method is to perform *in vitro* cleavage using exogenous proteases. This includes enzymes like factor Xa,[52] tobacco etch virus (TEV) protease,[53] thrombin,[54] enterokinase each having their own specificities (table 1-1). TEV protease is found to be the most efficient enzyme as it has a highly specific recognition site and is active over a wide range of pH and temperature.

Protease	Recognition sequence
Factor Xa	IEGR’
TEV	ENLYFQ’
Thrombin	LVPR’GS
Enterokinase	DDDDK’

Table 1-1: Exogenous proteases and their recognition sequence. Proteases perform cleavage towards the C-terminal of recognition site, ’ represents cleavage site.

Chemical cleavage is performed by using CNBr which cleaves the C-terminus of Met residues.[55] The reaction begins with the nucleophilic attack by sulfur of Met towards electrophilic carbon of CNBr. This is followed by formation of a rigid five-membered ring which destabilizes the system leading to hydrolysis (Figure 1-27). Presence of Cys in peptide do not affect the reaction since thiol group of Cys behaves differently. (CNBr)

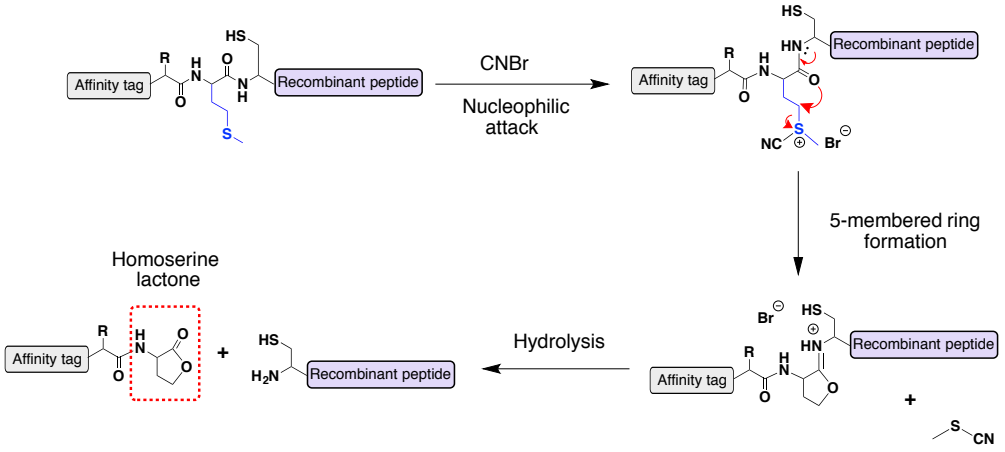


Figure 1-27: Mechanism of CNBr cleavage. Affinity tag helps in purification of the recombinant peptide.

Protein splicing can also be exploited to prepare recombinant *N*-terminal Cys proteins (Figure 1-23).[56] The inteins are mutated such that cleavage at the *C*-terminal splice junction is controlled in a pH- and temperature-dependent fashion. The main drawback in using mutated inteins for this purpose is spontaneous cleavage resulting in premature loss of intein.

1-16. Limitations and advances in semisynthesis

Recombinant expression of small peptides is limited due to their instability in bacterial host which is overcome by fusing an additional peptide called fusion tag. The foreign peptides are usually recognized as abnormal peptides and are consequently degraded.[57] This issue is overcome by expressing the desired small peptide fused to a host protein, i.e. as a hybrid polypeptide (Figure 1-28).[58]

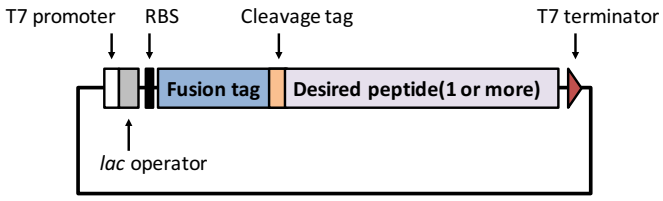


Figure 1-28: Representation of a fusion peptide for expression of small peptides in *E.coli*. Coding sequence for a desired peptide can be single or multiple copies can be linked tandemly.

Desired peptide is released from fusion tag by chemical or enzymatic cleavage. Fusion tags are known to improve yield, prevent proteolysis and increase solubility *in vivo* and also act as affinity tags to aid purification.[59] The most commonly used solubility tags are mentioned in table 1-2.[59]

Fusion Tag	Cleavage site	Size (kDa)
Maltose binding protein	ENLYFQ'GXX	40
Glutathione-S-transferase	ENLYFQ'GXX	26
Thioredoxin	ENLYFQ'GXX	12
NUS A	ENLYFQ'GXX	55
Ubiquitin	ENLYFQ'GXX	8
SUMO	GG'XXX	11.5

Table 1-2: Commonly used fusion tags having different properties, removed by proteases that cleave towards the C-terminal of recognition site, ' represents cleavage site. ENLYFQ is the recognition site of TEV protease, while SUMO fusion tag is cleaved by SUMO protease that recognize the tertiary structure of SUMO protein unlike other proteases.

Semisynthesis of human Ghrelin was demonstrated by Makino et al., using the fusion protein and modifying the recombinant peptide to get a free N-terminus while keeping the Lys side chains protected with Boc which was then condensed to chemically synthesized fragment (Figure 1-29).[60] They used β-galactosidase as fusion tag that forms inclusion bodies and two distinct protease cleavage sites. At first, fusion tag was removed by using OmpT protease then modified with Boc and followed by second protease cleavage using Kex2 (Figure 1-29).

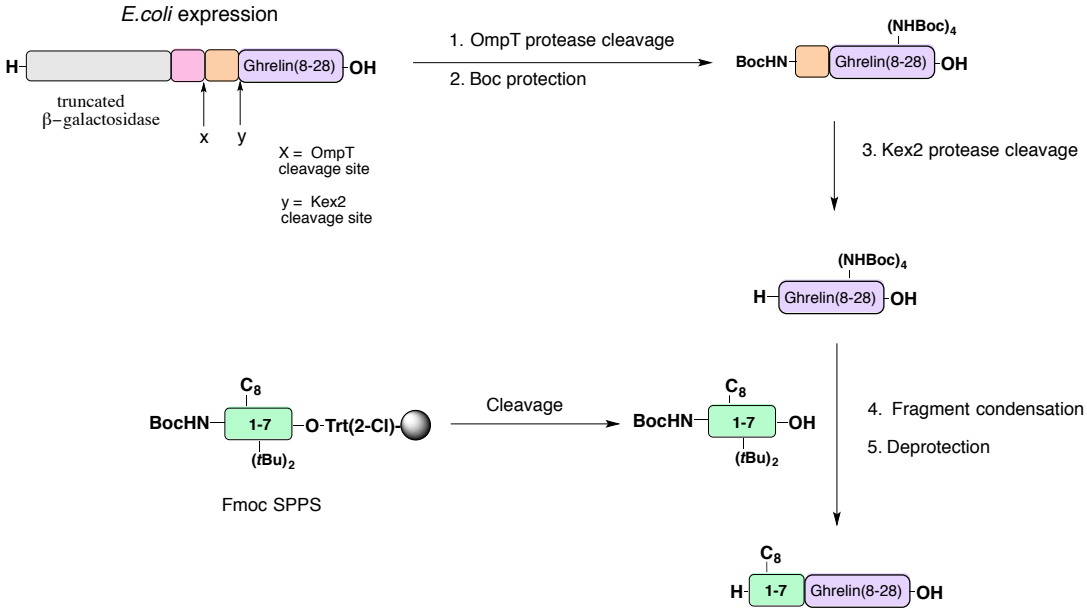


Figure 1-29: Semisynthetic strategy for human Ghrelin using fusion protein and chemically synthesized N-terminal peptide.

When expressing smaller peptides less than 50 amino acids as a fusion peptide, the desired peptide constitutes only a small part of the entire fusion peptide compared with the size of the fusion tag. In these cases, though recombinant expression gives higher yields, however, after the removal of fusion tag the desired peptide is obtained in considerably low amounts. To overcome this, the plasmid can be constructed such that the coding sequence of the desired gene is amplified by tandemly linking multiple copies of the desired gene that can enhance the stability of the protein in *E.coli* and at the same time increase the quantity of the desired peptide (Figure 1-28).[58]

The other major limitation is the requirement of Cys which is a less abundant amino acid.[17] Most proteins do not have Cys at convenient locations or lack Cys entirely which hinders the use of a recombinantly expressed peptide without mutation. Several groups are working towards the limitation of Cys at the ligation site. Aimoto's group have established a strategy to use expressed peptide as a building block for semisynthesis of a phosphorylated p21Max protein via thioester ligation method, which does not require Cys at the ligation site.[61] The method involves transamination of *N*-terminal amino acid to an α-oxoacyl group, followed by protection of Lys side chains with *t*-butyloxycarbonyl (Boc) and finally, removal

of α-oxoacyl group using diamine. While *O*-phosphoserine containing peptide thioester was synthesized via Boc-based SPPS. The two fragments are then condensed in the presence of silver ions and an activating reagent (Figure 1-30).

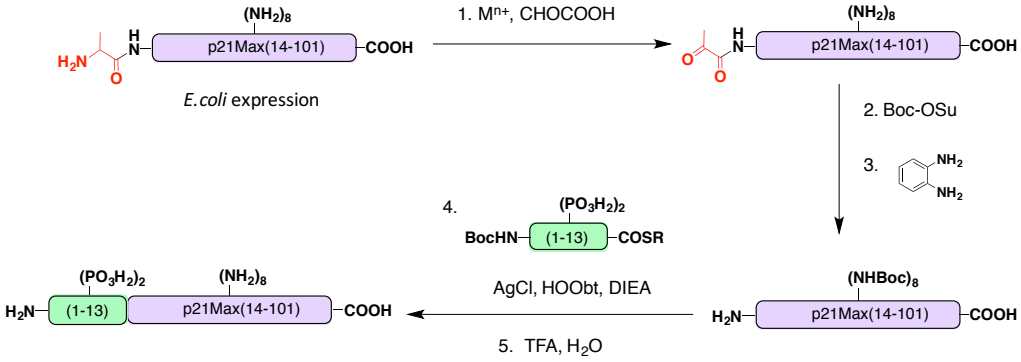


Figure 1-30: Semisynthetic strategy for synthesis of p21Max protein which involves selective excision of N-terminal amino acid by transamination and ligation to synthetic peptide by thioester method.

Recently, another strategy for *N*-terminal modification with Homocysteine (Hcs) towards *E.coli* expressed peptide was reported by the Petersson group who employed mutant aminoacyltransferase (AaT) that could transfer disulfide protected Hcs to the *N*-terminus of a protein under mild conditions.[62] The mutant AaT could recognize *N*-terminal Lys or Arg of the recombinant peptide and transfer Hcs *in vitro* which was followed by reduction of Hcs to get free thiol. This was used for NCL and later methylation of thiol was performed to get Met residue (Figure 1-31).

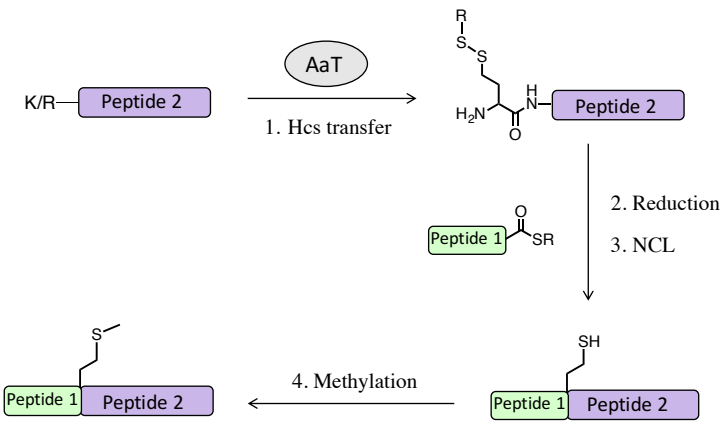


Figure 1-31: Semisynthesis of synuclein by functionalizing expressed protein with Hcs using AaT *in vitro* followed by reduction and NCL and methylation.

1-17. Chemical modification of a recombinant peptide

In order to expand semisynthesis by overcoming the limitation of Cys at ligation site, modification of recombinant peptide must be performed as shown in the above examples. To chemically modify a recombinant peptide, the most commonly employed strategy is to first protect the Lys side-chain, otherwise obtaining a homogeneous modification is impossible in the presence of other reactive functional groups (Figure 1-32).[60,61] In the process, the *N*-terminal amino group also gets protected. The widely employed protecting group for this purpose is *tert*-butyloxycarbonyl (Boc) which being a hydrophobic moiety, will make the peptide hydrophobic. This is a disadvantage as it makes the protected peptide insoluble under aqueous conditions and purification under routine HPLC conditions difficult and also reactions can be carried out only in suitable organic solvents. The extent of hydrophobicity that it imparts to a peptide depends on the number of Lys, the length of the peptide and also on the nature of the peptide.

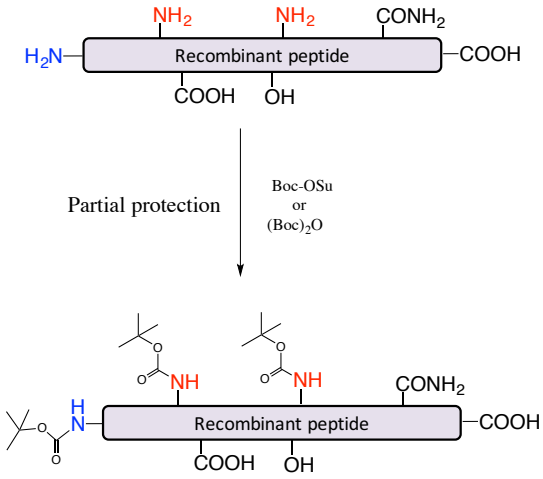


Figure 1-32: Partial protection of recombinant peptide with Boc-protecting group.

An attractive alternative for Boc and Fmoc hydrophobic protecting groups was developed by Waldmann et al. which are tetra-*O*-acetyl-D-glucopyranosyloxycarbonyl (AGlOC) and tetra-*O*-acetyl-β-D-galactopyranosyloxycarbonyl (AGalOC) protecting groups that are enzyme-labile offering high selectivity and milder conditions (pH 6-8, aqueous solvents and ambient temperatures) (Figure 1-33).[63] These urethane-derived protecting groups were employed for chemical synthesis of peptides from *N*-terminal to *C*-terminal direction unlike the conventional chemical peptide synthesis. At first, protected amino acid was synthesized through a one-pot procedure followed by catalytic hydrogenation and coupling of the second amino acid.

Removal of the AGlOC or AGalOC was performed as a two-step process without isolating the intermediates that involved removal of acetyl groups of sugar followed by removal of sugar. They could successfully demonstrate synthesis of tetrapeptide by their strategy.

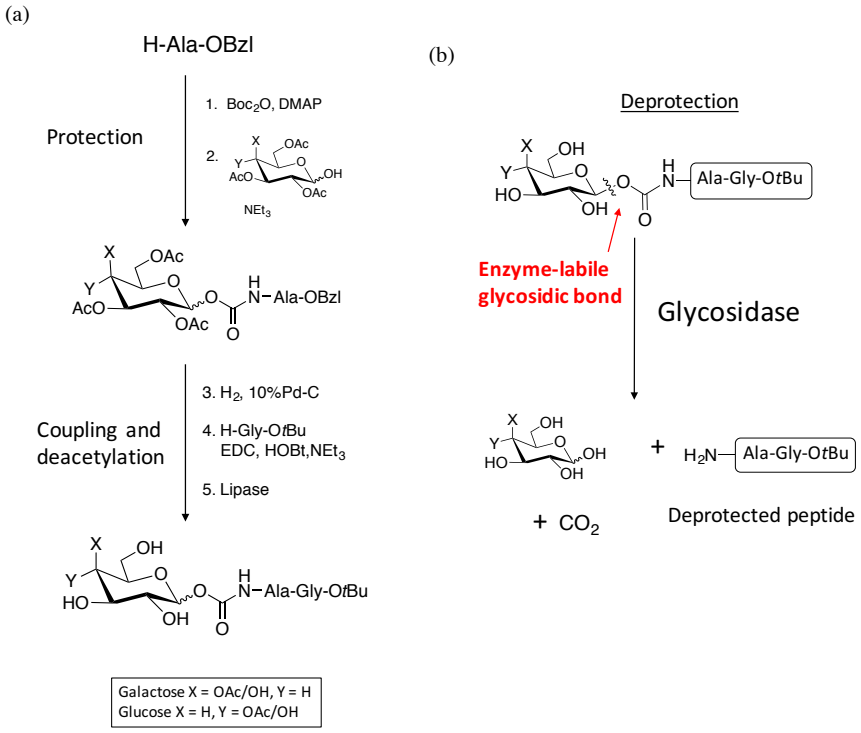


Figure 1-33: (a) use of carbohydrate protecting group for N^{α}-amino protecting group (b) cleavage mechanism by glycosidases.

Chapter 1

1-18. Problems associated with homogeneous glycopeptide synthesis.

Total chemical synthesis of homogeneous glycoproteins has several drawbacks such as size limitation of polypeptide, low crude purity of glycopeptide, low yield due to repetitive ligation steps, expensive starting material and time consuming. SPPS is limited to approximately 50 amino acid residue peptides. Moreover, not all peptides sequences can be easily synthesized by SPPS. Several of these problems can be solved by employing semisynthesis, through which mass production of peptides can be performed in a relatively short time with cheap starting materials and high purity. But it is still limited with the number and the sites of modifications that can be accessed, due to the low natural abundance of Cys and lack of methods to incorporate unnatural amino acids during *E.coli* expression, for example, incorporation of Sec or β-mercapto amino acids. Few examples that involve chemical modification of recombinant peptide at its *N*-terminal are already mentioned, since they use hydrophobic protecting group, in order to get full length semisynthetic protein, condensation reaction is performed with the chemically synthesized fragment in organic solvent like DMF in presence of activating reagents which might cause epimerization of *C*-terminal amino acid. Therefore, a strategy that overcomes the hydrophobicity problem of protected recombinant peptides and allows the most efficient ligation strategy NCL to be performed should expand the field of semisynthesis. Though recently reported *N*-terminal modification of recombinant peptide with Homocysteine (Hcs)[62] can overcome these problems, at present, not all unnatural amino acids can be transferred by using mutant tRNAs and moreover transfer is possible only with specific *N*-terminal amino acids in the recombinant peptide.

1-19. Objective of this doctoral thesis

My doctoral research is mainly focused on establishing a general semisynthetic strategy for the synthesis of glycoproteins with multiple modifications by addressing the issues of

1. Hydrophobicity associated with partially protected peptides.
2. Introducing unnatural amino acid/Sec at the *N*-terminus of recombinant peptide favoring NCL.
3. Low yield due to conventional chemical glycoprotein synthesis.

Thus, overcoming the drawbacks of total chemical synthesis and semisynthesis at the same time, by considering the advantages of both. The details regarding the novelty and difficulty are mentioned as follows.

An approach for semisynthesis of glycoproteins is shown in Figure 1-34 using erythropoietin (EPO) as the model. Erythropoietin(EPO) is a signaling glycoprotein that controls the fundamental process of erythropoiesis, regulating the production and maintenance of red blood cells. EPO is of 166 amino acids having four tightly conserved glycosylation sites, with three *N*-glycans at Asn-24, Asn-38, Asn-83 and *O*-glycan at Ser-126 and its total chemical synthesis is already achieved. [64]

For introducing oligosaccharides in the polypeptide backbone, the protein has to be divided into appropriate short peptide segments each ranging around 20-40 amino acids, as *E.coli* lacks the ability to insert oligosaccharide. These short peptides can be easily obtained by SPPS but when these peptides are to be obtained in high yield and at low cost, *E.coli* expression using a fusion peptide with tandem repeats is a best choice. The peptides prepared as fusion peptides can be converted to monomeric peptide units for example, by CNBr cleavage method.

For introducing the glycans, the peptide fragments obtained by *E.coli* expression should be coupled with the oligosaccharides in solution phase unlike glycopeptides obtained by SPPS. While *N*-glycan and *O*-glycan can be readily obtained by isolation from hen egg yolk and chemical/chemoenzymatic synthesis respectively as already mentioned, there are several factors that must be optimized in coupling aminoacyl-oligosaccharide to a recombinant peptide.

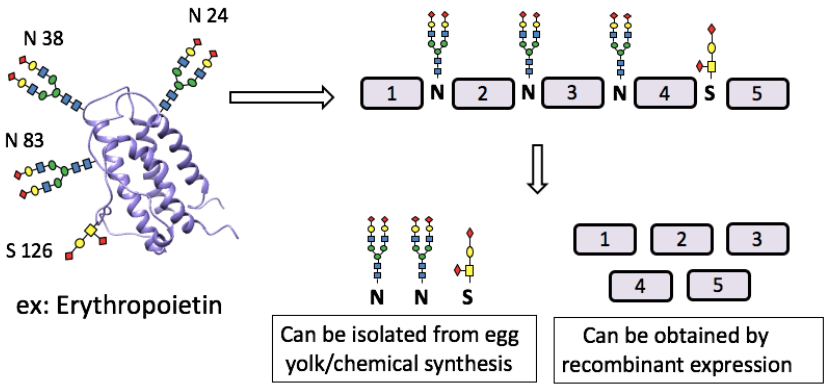


Figure 1-34: Semisynthetic approach using EPO as model

In a recombinant peptide, all the functional groups are free and trying to couple oligosaccharyl-amino acid only at the *N*-terminal of the peptide fragment is impossible without side-reactions in the presence of reactive functional groups. A way to overcome this is partial protection of peptide as mentioned previously and liberate free *N*-terminal amine.

In order to perform ligation efficiently, the best methodology is NCL over fragment condensation of peptides. For NCL, the peptide fragments must have a Cys at the *N*-terminus and when the target peptide lacks Cys at convenient positions it must have selenocysteine (Sec) or β-mercapto amino acid analogues. In addition to this modification, the other peptide fragment must be a thioester or hydrazide at the *C*-terminal after cleavage from fusion protein. Recently, the Kajihara group have reported a strategy towards activation of the *C*-terminal useful for this strategy. Therefore, having established a strategy to obtain recombinant *C*-terminal thioester peptide by our group, I further aimed to expand semisynthesis in a new direction by establishing a general strategy that involves easy modification of *N*-terminal of recombinant peptide with any desired amino acid that can be used for NCL.

In order to establish a new strategy, I chose a model peptide expressed in *E.coli* as a fusion peptide using SUMO tag and a hydrophilic protection group for peptide side chain. In general, fusion tag is removed by a specific protease. I planned to replace hydrophobic Boc protecting group with a carbohydrate based protecting group which renders hydrophilicity due to the hydroxy groups, thus making it easier to handle the peptide under aqueous buffer rather than organic solvents.

To get a free *N*-terminal amino group, I planned to examine TEV protease cleavage towards partially protected peptide and this reaction is favored by using hydrophilic protecting group, which allows the protected peptide to be soluble under aqueous conditions after which *N*-terminal can be modified in a desired way under milder reaction conditions.

I intended to modify *N*-terminal with Sec in my strategy, due to the fact that selenol can be readily removed in the presence of Cys thiol group and other advantages Sec possess over Cys, while this modification can be further extended in a desired way as previously mentioned. Moreover, coupling of a single amino acid, followed by NCL is more efficient than fragment condensation.

After successfully modifying a partially protected recombinant peptide with Sec, all the Lys side chain protecting groups were removed and thus obtained peptide can be used for the well-established NCL reaction.

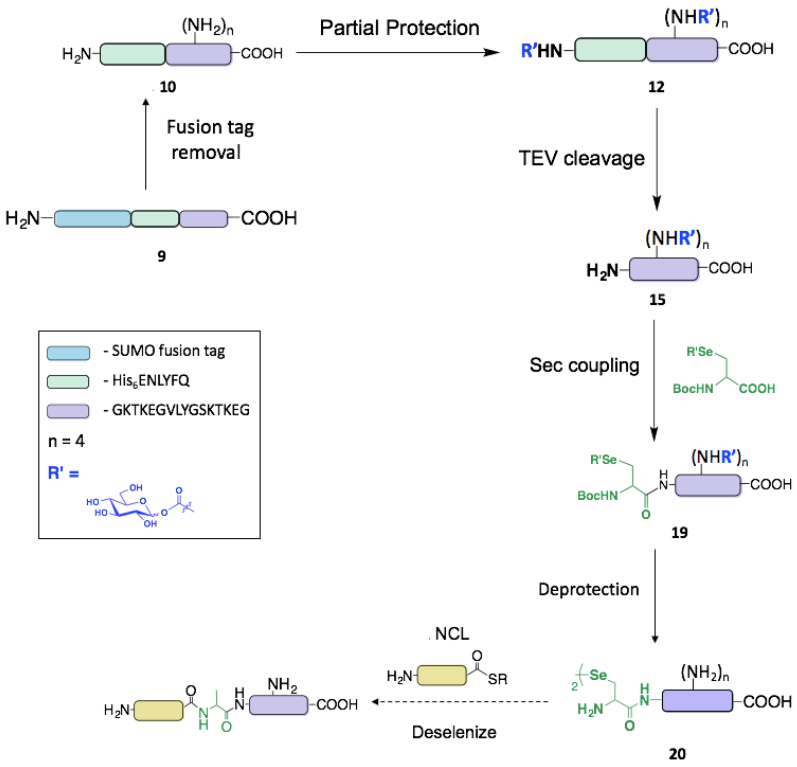


Figure 1-34: Schematic representation of my strategy

1-20. Contents of the thesis

As discussed, a strategy to obtain glycoprotein with multiple modifications by semisynthesis is highly advantageous in terms of time, yield and most importantly, cost of production. This needs exploration of more general ways to use recombinant peptide with any desired modifications. In *Chapter 2,* I will discuss the details about optimizing the conditions for employing hydrophilic protecting group for partial protection of peptide. This should enable me to perform enzymatic cleavage under aqueous conditions. I will discuss the details about protease cleavage and modification of *N*-terminal amino group with Selenocysteine, which thus should enable us to employ the well-established NCL.

References

1. (a) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. *Angew. Chem. Int. Ed.* **2005**, *44*, 7342. (b) Knorre, D.G.; Kudryashov, N.V.; Godovikova, T.S.; *Acta Naturae* **2009**, *3*, 29.
2. Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683.
3. (a) Hounsell, E. F.; Davies, M. J.; Renouf, D. V.; *Glycoconjugate J.* **1996**, *13*, 19. (b) Mariño, K.; Bones, J.; Kattla, J. J.; Rudd, P.M. *Nat. Chem. Biol.* **2010**, *6*, 713.
4. (a) Walsh, G.; Jefferis, R. *Nat. Biotechnol.* **2006**, *24*, 1241. (b) Ferris, S. P.; Kodali, V. K.; Kaufman, R. J.; *Dis. Model. Mech.* **2014**, *7*, 331. (c) Xu, C.; Ng, D. T. W.; *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 742. (d) Olivari, S.; Molinari, M.; *FEBS Lett.* **2007**, *581*, 3658. (e) Ellgaard, L.; Helenius, A. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 181.
5. (a) Stanley, P. *Cold Spring Harb Perspect Biol.* **2011**, *3*, 1. (b) Brockhausen, I.; Schachter, H.; Stanley, P.; *Essentials of Glycobiology* (ed. Varki A, et al.), **2009**, 115.
6. (a) Kobata, A.; *Eur. J. Biochem.* **1992,** *209*, 483. (b) Huang, P.-S.; Boyken, S. E.; Baker, D.; *Nature.* **2016***, 537*, 320
7. (a) Aricescu, A. R.; Owens, R. J.; *Curr. Opin. Struct. Biol.* **2013**, *23*, 345. (b) Dumont, J.; Euwart, D.; Mei, B.; Estes, S.; Kshirsagar, R.; *Crit. Rev. Biotechnol.* **2016**, *36*, 1110.
8. Shi, X.; Jarvis, D. J.; *Curr. Drug. Targets.* **2007**, *8*, 1116.
9. Hamilton, S. R.; Gerngross, T. U.; *Curr. Opin. Biol.* **2007**, *18*, 387.
10. Gomord, V. et al. *Plant Biotechnol. J.* **2010**, *8*, 564.
11. (a) Kent, S. B. H.; *Chem. Soc. Rev.*, **2009**, *38*, 338. (b) Merrifield, R. B.; *J. Am. Chem. Soc.* **1963**, *85*, 2149. (c) Hemantha, H. P. et al. *Tetrahedron.* **2012**, *68*, 9491. (d) Nilsson, B. L.; Soellner, M. B.; Raines, R. T.; *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 91. (e) Unverzagt, C.; Kajihara, Y.; *Chem. Soc. Rev.,* **2013**, *42*, 4408.
12. (a) Fukae, K.; Yamamoto, N.; Hatakeyama, Y.; Kajihara, Y.; *Glycoconjugate. J.* **2004**, *21*, 243. (b) Kajihara, Y.; Suzuki, Y.; Yamamoto, N.; Sasaki, K.; Sakakibara, T.; Juneja, L. R.; *Chem. Eur. J.* **2004**, *10*, 971. (c) Maki, Y.; Okamoto, R.; Izumi, M.; Murase, T.; Kajihara, Y.; *J. Am. Chem. Soc.* **2016**, *138*, 3461.
13. (a) Marcaurelle, L. A.; Bertozzi, C. R.; *Glycobiology* **2002**, *12,* 69.
14. Wieland, T.; Bokelmann, E.; Bauer, L.; Lang, H. U.; Lau, H.; *Liebigs. Ann. Chem.* **1953***, 583*, 129.
15. (a) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H.; *Science.* **1994**, *266*, 776. (b) Tam, J. P.; Lu, Y. A.; Liu, C. F.; Shao, J.; *Proc. Natl. Acad. Sci. USA.* **1995**, *92*, 12485.
16. Hackeng, T. M.; Griffin, J. H.; Dawson, P. E.; *Proc. Natl. Acad. Sci. USA.* **1999**, *96*,10068.

References

17. McCaldon, P.; Argos, P.; *Proteins*. **1988**, *4*, 99. [PubMed: 3227018]
18. Tam, J. P.; Yu, Q.; *Biopolymers*. **1998**, *46*, 319.
19. Yan, L. Z.; Dawson, P. E.; *J. Am. Chem. Soc*. **2001**, *123*, 526.
20. Wan, Q.; Danishefsky, S. J.; *Angew. Chem. Int. Ed*. **2007**, *46,* 9248.
21. Wong, C. T. T.; Tung, C. L.; Li, X.; *Mol. BioSyst*. **2013**, *9*, 826.
22. Crich, D.; Banerjee, A.; *J. Am. Chem. Soc*. **2007**, *129*, 10064.
23. (a) Haase, C.; Rohde, H.; Seitz, O.; *Angew. Chem. Int. Ed*. **2008**, *47*, 6807. (b) Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J.; *Angew. Chem. Int. Ed*. **2008**, *47*, 8521.
24. (a) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X.-W.; Liu, C.-F.; *J. Am. Chem. Soc*. **2009**, *131*, 13592. (b) Kumar, K. S. A.; Haj-Yahya, M.; Olschewski, D.; Lashuel, H. A.; Brik, A.; *Angew. Chem. Int. Ed*. **2009**, *48*, 8090.
25. Tan, Z.; Shang, S.; Danishefsky, S. J.; *Angew. Chem. Int. Ed*. **2010**, *49*, 9500.
26. Shang, S.; Tan, Z.; Dong, S.; Danishefsky, S. J.; *J. Am. Chem. Soc*. **2011**, *133*, 10784.
27. Chen, J.; Wang, P.; Zhu, J.; Wan, Q.; Danishefsky, S. J.; *Tetrahedron*. **2010**, *66*, 2277.
28. Malins, L. R.; Cergol, K. M.; Payne, R. J.; *ChemBioChem*. **2013**, *14*, 559.
29. Siman, P.; Karthikeyan, S. V.; Brik, A.; *Org. Lett*. **2012**, *14*, 1520.
30. Thompson, R. E.; Chan, B.; Radom, L.; Jolliffe, K. A.; Payne, R. J.; *Angew. Chem. Int. Ed*. **2013**, *14*, 559.
31. Malins, L. R.; Cergol, K. M.; Payne, R. J.; *Chem. Sci*. **2014**, *5*, 260.
32. (a) Pentelute, B. L.; Kent, S. B. H.; *Org. Lett*. **2007**, *9*, 687. (b) Veber, D.F.; Milkowski, J. D.; Varga, S.L.; Denkewalter, R. G.; Hirschmann, R.; *J. Am. Chem. Soc*. **1972**, *26*, 5456.
33. Stadtman, T. C.; *Annu. Rev. Biochem*. **1996**, *65*, 83.
34. (a) Low, S. C.; Berry, M. J.; *Trends. Biochem. Sci*. **1996**, *21*, 203. (b) Arnér, E. S. J.; Sarioglu, H.; Lottspeich, F.; Holmgren, A.; Böck, A.; *J. Mol. Biol*. **1999**, *292*, 1003.
35. (a) Muttenthaler, M.; Alewood, P. F.; *J. Pept. Sci*. **2008**, *14*, 1223. (b) Huber, R.; Criddle, R. S.; *Arch. Biochem. Biophys*. **1967**, *122*, 164. (c) Hondal, R. J.; Nilsson, B. L.; Raines, R. T.; *J. Am. Chem. Soc*. **2001**, *123*, 5140. (d) Malins, L. R.; Mitchell, N. J.; Payne, R. J.; *J. Pept. Sci*. **2014**, *20*, 64.
36. Metanis, N.; Keinan, E.; Dawson, P. E.; *Angew. Chem. Int. Ed*. **2010**, *49*, 7049.
37. Townsend, S. D.; Tan, Z.; Dong, S.; Shang, S.; Brailsford, J. A.; Danishefsky, S. J.; *J. Am. Chem. Soc*. **2012**, *134*, 3912.
38. Malins, L. R.; Payne, R. J.; *Org. Lett*. **2012**, *14*, 3142.
39. Malins, L. R.; Mitchell, N. J.; McGowan, S.; Payne, R. J.; *Angew. Chem. Int. Ed*. **2015**, *54*, 12716.

References

40. Izumi, M.; Murakami, M.; Okamoto, R.; Kajihara, Y.; *J. Pept. Sci.* **2014**, *20*, 98.
41. Futaki, S.; Sogawa, K.; Maruyama, J.; Asahara, T.; Niwa, M.; *Tetrahedron Lett.* **1997**, *38*, 6237.
42. Li, X. Q.; Kawakmi, T.; Aimoto, S.; *Tetrahedron Lett.* **1998**, *39*, 8669.
43. (a) Alsina, J.; Yokum, T. S.; Albericio, F.; Barany, G.; *J. Org. Chem.* **1999**, *64*, 8671. (b) Jensen, K. J.; Alsina, J.; Songster, M. F.; Vágner, J.; Albericio, F.; Barany, G.; *J. Am. Chem. Soc.* **1998**, *120*, 5441. (c) Alsina, J.; Albericio, F.; *Biopolymers*. **2003**, *71*, 454. (d) Tulla-Puche, J.; Getun, I. V.; Woodward, C.; Barany, G.; *Biochemistry*. **2004,** *43*, 1591. (e) Brask, J.; Albericio, F.; Jensen, K. J.; *Org. Lett.* **2003,** *5*, 2951.
44. (a) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R.; *J. Am. Chem. Soc.* **1999,** *121*, 11684. (b) Kenner, G. W.; McDermott, J. R.; Sheppard, R. C.; *Chem. Comm.* **1971**, 636. (c) Ingenito, R.; Bianchi, E.; Fattori, D.; Pessi, A.; *J. Am. Chem. Soc.* **1999,** *121*, 11369. (c) Backes, B. J.; Virgilio, A. V.; Ellman, J. A.; *J. Am. Chem. Soc.* **1996**, *118*, 3055. (d) Backes, B. J.; Ellman, J. A.; *J. Org. Chem.* **1999**, *64*, 2322.
45. Terrier, V. P.; Adihou, H.; Arnould, M.; Delmas, A. F.; Aucagne, V.; *Chem. Sci.* **2016**, *7*, 339.
46. Bang, D.; Kent, S. B. H.; *Angew. Chem. Int. Ed.* **2004**, *43*, 2534.
47. (a) Muir, T. W.; *Annu. Rev. Biochem.* **2003**, *72*, 249. (b) Berrade, L.; Camarero, J. A.; *Cell. Mol. Life Sci.* **2009**, *66*, 3909. (c) Aranko, A. S.; Wlodawer, A.; Iwaï, H.; *Protein Eng. Des. Sel.* **2014**, *27*, 263. (d) Shah, N. H.; Muir, T. W.; *Chem. Sci.* **2014**, *5*, 446 (e) Xu, M.-Q.; Evans, Jr. T. C.; *Methods*. **2001**, *24*, 257. (f) Wood, D. W.; Wu, W.; Belfort, G.; Derbyshire, V.; Belfort, M.; *Nature Biotechnol.* **1999**, *17*, 889. (g) Mootz, H. D.; *ChemBioChem*. **2009**, *10*, 2579. (h) Tolbert, T. J.; Wong, C.-H.;
48. Okamoto, R.; Kimura, M.; Ishimizu, T.; Izumi, M.; Kajihara, Y.; *Chem. Eur. J.* **2014**, *20*, 10425.
49. Li, Y.-M.; Yang, M.-Y.; Huang, Y.-C.; Li, Y.-T.; Chen, P. R.; Liu, L.; *ACS Chem. Biol.* **2012**, *7*, 1015.
50. Fang, M.-G.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L.; *Angew. Chem. Int. Ed.* **2011**, *50*, 7645.
51. Hirel, P. H.; Schmitter, M. J.; Dessen, P.; Fayat, G.; Blanquet, S.; *Proc. Natl. Acad. Sci. USA.* **1989**, *86*, 8247.
52. Quinlan, R. A.; Moir, R.D.; Stewart, M.; *J. Cell Sci.* **1989**, *93*, 71.
53. Tolbert, T. J.; Wong, C.-H.; *Angew. Chem. Int. Ed.* **2002**, *41*, 2171.
54. Liu, D.; Xu, R.; Dutta, K.; Cowburn, D.; *FEBS Lett.* **2008**, *582*, 1163.
55. Okamoto, R.; Kajihara, Y.; *Angew. Chem. Int. Ed.* **2008**, *47*, 5402.
56. (a) Evans, T. C.; Benner, J.; Xu, M.-Q.; *J. Biol. Chem.* **1999**, *274*, 3923. (b) Southworth, M. W.; Amaya, K.; Evans, T. C.; Xu, M.-Q.; Perler, F. B.; *Biotechniques*. **1999**, *27*, 110. (c) Mathys, S.;

References

Evans, T. C.; Chute, I. C.; Wu, H.; Chong, S.; Benner, J.; Liu, X. Q.; Xu, M. Q.; *Gene*. **1999**, *231*, 1.

57. (a)Goldschmidt, R.; *Nature*. **1970**, 228, 1151. (b) Lin, S.; Zabin, I.; *J. Biol. Chem*. **1972**, *247*, 2205.
58. Shen, S.-H.; *Proc. Natl. Acad. Sci. USA*. **1984**, *81*, 4627.
59. (a) Marblestone, J. G.; Edavettal, S. C.; Lim, Y.; Lim, P.; Zuo, X.; Butt, T. R.; *Protein Sci*. **2006**, *15*, 182. (b) Sørensen, H. P.; Mortensen, K. K.; *J. Biotechnol*. **2005**, *115*, 113. (c) Makrides, S. C.; *Microbiol. Rev*. **1996**, *60*, 512.
60. Makino, T.; Matsumoto, M.; Suzuki, Y.; Kitajima, Y.; Yamamoto, K.; Kuramoto, M.; Minamitake, Y.; Kangawa, K.; Yabuta, M.; *Biopolymers*. **2005**, *79*, 238.
61. Kawakami, T.; Hasegawa, K.; Teruya, K.; Akaji, K.; Horiuchi, M.; Inagaki, F.; Kurihara, Y.; Uesugi, S.; Aimoto, S.; *J. Pept. Sci*. **2001**, *7*, 474.
62. Tanaka, T.; Wagner, A. M.; Warner, J. B.; Wang, Y. J.; Petersson, E. J.; *Angew. Chem. Int. Ed*. **2013**, *52*, 6210.
63. (a) Kappes, T.; Waldmann, H.; *Carbohydr. Res*. **1998**, *305*, 341. (b) Gum, A. G.; Kappes, T.; Waldmann, H.; *Chem. Eur. J*. **2000**, *6*, 3714.
64. (a) Wang, P.; Dong, S.; Shieh, J.-H.; Peguero, E.; Hendrickson, R.; Moore, M. A.; Danishefsky, S. J.; *Science*. **2013**, *342*, 1357. (b) Murakami, M.; Kiuchi, T.; Nishihara, M.; Tezuka, K.; Okamoto, R.; Izumi, M.; Kajihara, Y.; *Sci. Adv*. **2016**, *2*, 1.

Chapter 2

2. Introduction

The specific modification of *N*-terminal α-amino group of peptide expressed in *E.coli* is still difficult and there is a need to develop methods for efficient modification towards peptide-peptide coupling for NCL. *N*-Terminal α-amino group has a lower pKa ~8 when compared to the higher pKa ~10 of side chain ε-amino group due to the electron withdrawing effect of geminal -CO-NH- group.[1] Selective modification of *N*-terminal α-amino group in the presence of Lys side chain ε-amino group must be carefully controlled by pH. Though selective modification of *N*-terminus with aldehydes is currently possible, they are known for protein labelling and cannot be further used for NCL.[1] In order to chemically modify a recombinant peptide and ligate with a chemically synthesized peptide, it is essential to protect side chains of Lys, since selectivity towards *N*-terminus of recombinant peptide cannot be achieved.[2] In this chapter, synthesis of hydrophilic carbohydrate-derived protecting group and optimization of protection and deprotection conditions towards Lys side chains are explained followed by enzymatic cleavage and *N*-terminal modification with Sec.

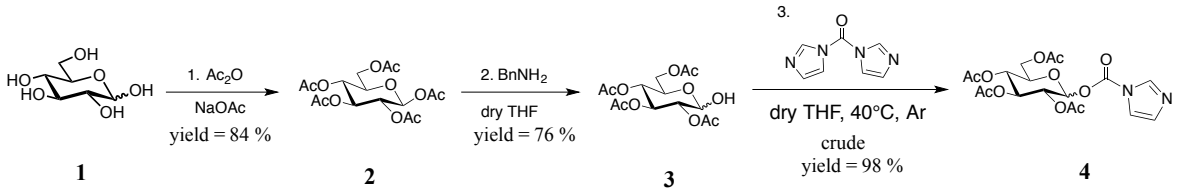
2-1. Carbohydrate-derived protecting group and its synthesis

Since the introduction of benzyloxycarbonyl group (Cbz) by Bergmann and Zervas in 1932,[3] an arsenal of carbamate amino protecting groups has been developed among which Boc group is commonly employed for partial protection of Lys side chains. Carbamate or Urethane type of protection is the most widely preferred type of protection for amino group as it is easy to protect, without causing additional byproducts and can be selectively removed.[3] As previously mentioned, Boc group being a hydrophobic protecting group enhances the hydrophobicity depending on the number of Boc groups and the length of the peptide. Hydrophobicity leads to solubility problems that greatly hinders various reactions in aqueous solvents and efficient purification by reverse phase HPLC. Thus, replacing hydrophobic Boc group with a hydrophilic protecting group becomes advantageous to deal with protected peptides.

A carbohydrate-derived carbamate protecting group is, therefore, an ideal hydrophilic protecting group as described in Chapter 1. The introduction of a carbohydrate-derived carbamate protecting group to peptide side chains need to be optimized. Efficient introduction of this protecting group can be performed

by activation with an imidazole. The synthesis of carbohydrate-derived carbamates is reported in the study of their behavior towards glycoside synthesis and a similar strategy[4] for the synthesis of AGlOC-imidazole **4** was employed as shown below.

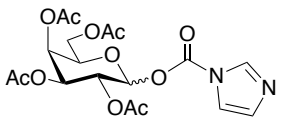
In the first step, all the hydroxyl groups of D-(+)-glucose **1** were protected with the acetyl groups following the established protocol that selectively gave β-D-glucose pentaacetate **2,** which was recrystallized using ethanol and used for the next step. In the second step, only the anomeric hydroxy group was



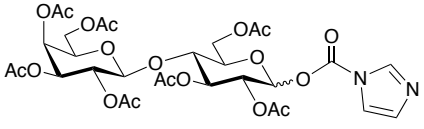
Scheme 2-1: Synthesis of AGlOC-imidazole protecting group **4**.[4,5]

selectively removed using benzyl amine that gave α:β anomeric mixture **3**,[4] which was purified and then used for coupling with carbonyldiimidazole to afford the protecting group **4** as a mixture of α:β anomers (6:4). The crude form was directly used for protection as the compound was not stable for silica-gel column purification.

The same protocol was extended to synthesize two new carbohydrate-derived imidazoles AGalOC-imidazole **5** and ALaOC-imidazole **6**. Unexpectedly, **5** was relatively unstable than **4** and **6**. Therefore, I selected glucose derivative **4** for establishing my semisynthetic strategy, since this sort of protection towards Lys side chains is not known so far.



5



6

2-2. Synthesis of model peptide for optimizing protection towards Lys

So far, the carbohydrate-derived carbamate protecting groups were not used for protection of Lys side chains and it was necessary to optimize conditions. Therefore, a model peptide **7** having two Lys was synthesized by Fmoc-based SPPS (Figure 2-1). For the synthesis, amino PEGA resin which is known for its robustness and high permeability in the swollen state was used. The linker 4-(hydroxymethyl)phenoxyacetic acid (HMPA), which is a standard TFA-labile linker for the preparation of peptide acids by Fmoc-SPPS was coupled to it. Coupling of first amino acid (Fmoc-Glu(O^{t}Bu)-OH) was performed using 2,4,6-mesitylene-sulfonyl-3-nitro- 1,2,4-triazolide (MSNT),[6] which is an efficient reagent for esterification of Fmoc-AA-OH onto unreactive linkers or enantiomerization prone amino acids. Deprotection of the Fmoc-amino group was performed using 20% piperidine in DMF. This was followed by coupling of other amino acids using 0.45 M HCTU/HOBt, 0.9 M DIEA in DMF and deprotection with 20% piperidine in DMF. After coupling of all amino acids, removal of side-chain protecting groups and cleavage from resin was performed in a single step by treating the resin with TFA/H_2O/TIPS (95:2.5:2.5). The peptide was precipitated in ice cold diethyl ether which was purified by reverse phase HPLC to afford desired model peptide **7** and characterized by ESI-MS as shown in Figure 2-1.

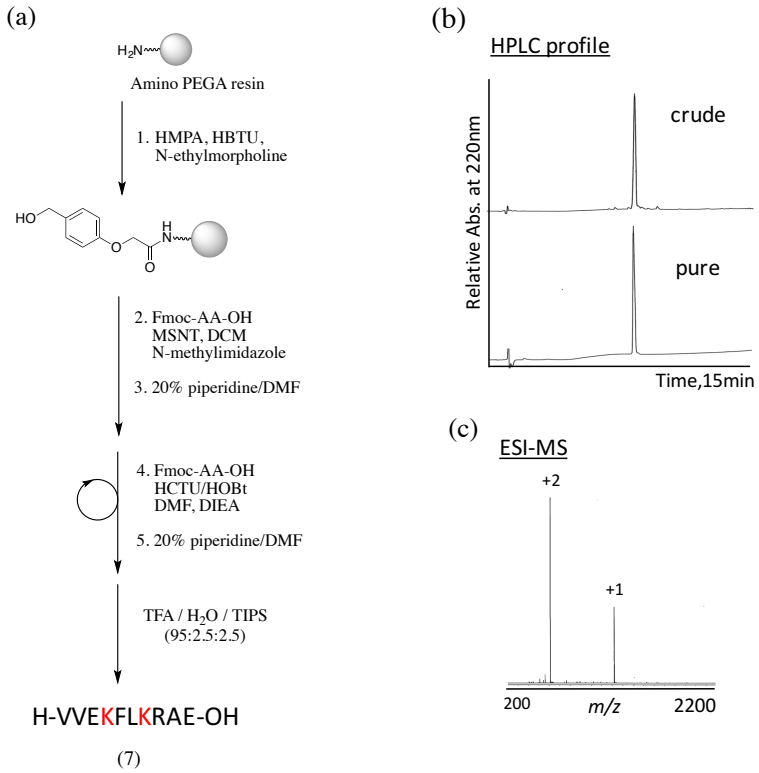
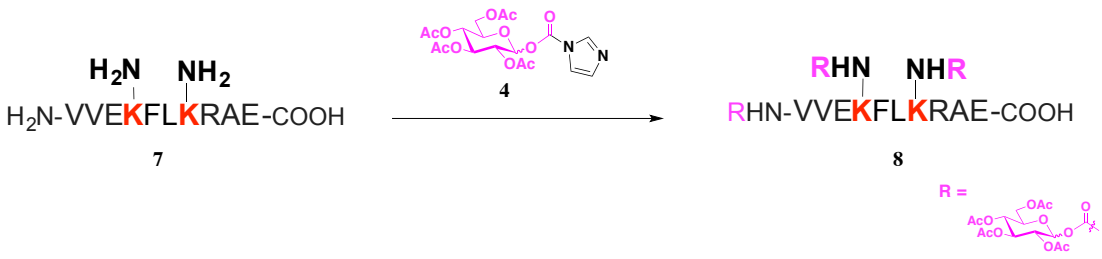


Figure 2-1: (a) Synthetic scheme for model peptide **7** by Fmoc-SPPS; (b) HPLC profile of crude and pure **7**; (c) Characterisation of **7** by ESI-MS $[M+H]^+$ calculated: 1218.72, observed: 1218.70.

2-3. Optimization of reaction conditions for protection of peptide **7**

At first, protection of peptide **7** side chain was performed using AGlOC-imidazole **4**, using the conventional condition used for Boc protection of peptide (entry 1). However, the protection was incomplete and the desired product could not be isolated. Therefore, on considering the reactivity of carbonyl group of amides and esters towards nucleophilic attack, commonly used ester activating reagents in amino acid coupling, i.e N-hydroxysuccinimide (NHS)[7] and hydroxybenzotriazole (HOBt)[8] were examined. These additives improved the reaction drastically (entry 2 & 3). In order to examine the reaction dependency on solvent, reaction was performed in DMSO and AcCN/H_2O solvent and AcCN/H_2O gave excellent results. Other conditions that were varied are mentioned in table 2-1. HOBt could not be used as an additive in AcCN/H_2O because of its insolubility. The reactions were monitored by reversed phase HPLC and the desired product **8** was purified by reversed phase HPLC and characterized by ESI-MS. The desired product **8** was further confirmed by HRMS by using Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometry that provides ultra-high resolution and highly accurate mass of molecules. The individual isolated yields of all reaction conditions are shown in table 2-1 and the reaction profile of the optimized condition (entry 8) is shown in Figure 2-2.

(a)



(b)

Entry	AGlOC-imi *	Solvent	Additive#	DIEA*	Temp.	t (h)	Isolated yield
1	20 eq	DMF	-	5 eq	30°C	6	incomplete protection
2	20 eq	DMF	NHS (2.5 eq)	5 eq	30°C	3	42 %
3	20 eq	DMF	HOBt (2.5 eq)	5 eq	30°C	3	45 %
4	20 eq	DMSO	NHS (2.5 eq)	5 eq	30°C	3	30 %
5	20 eq	DMSO	HOBt (2.5 eq)	5 eq	30°C	3	40 %
6	20 eq	AcCN:H_2O (2:1)	NHS (2.5 eq)	-	30°C	3	53 %
7	20 eq	AcCN:H_2O (2:1)	NHS (2.5 eq)	5 eq	30°C	3	57 %
8	**20 eq**	**AcCN:H_2O (2:1)**	**NHS (2.5 eq)**	**10 eq**	**30°C**	3	**75 %**
9	20 eq	AcCN:H_2O (2:1)	NHS (1.5 eq)	5 eq	30°C	3	57 %
10	20 eq	AcCN:H_2O (2:1)	NHS (3.5 eq)	5 eq	30°C	3	61 %
11	20 eq	AcCN:H_2O (2:1)	NHS (2.5 eq)	5 eq	40°C	3	68 %
12	60 eq	AcCN:H_2O (2:1)	NHS (2.5 eq)	5 eq	30°C	2	53 %
13	10 eq	AcCN:H_2O (2:1)	NHS (2.5 eq)	5 eq	30°C	2	39 %

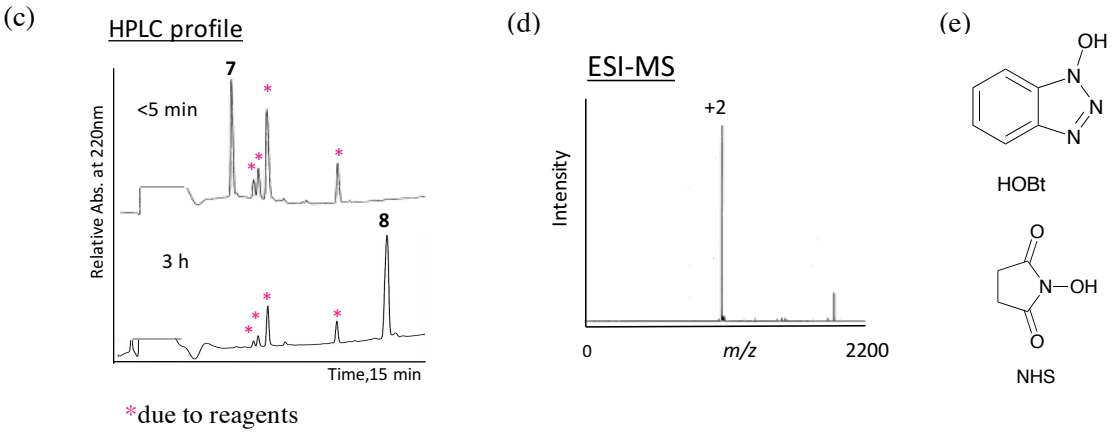
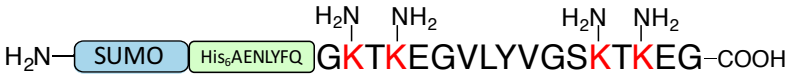


Figure 2-2: (a) Scheme for partial protection of **7** with AGlOC protecting group **4.** (b) Table 2-1: representing various optimization conditions employed for partial protection of **7**, * equivalents of **4** and DIEA were taken with respect to peptide, # equivalents of additive were taken with respect to **4.** (c) HPLC profile of reaction monitoring of optimized condition. (d) ESI-MS of desired product **8** $[M+H]^+$ calculated 2340.9751, observed: 2340.9752. **(e)** Structure of activating reagents.

2-4. Extending protection to peptide obtained by recombinant expression

Having optimized the condition for protection towards a smaller model peptide **7**, I wanted to examine protection towards a longer peptide obtained by *E.coli* expression and optimize further reaction conditions for establishing the strategy. For this purpose, another model peptide **10** having four Lys and a protease cleavage site was obtained by recombinant expression in *E.coli* as a fusion peptide **9** fused to SUMO protein tag (Figure 2-3). SUMO (Small Ubiquitin-related MOdifier) protein is a covalent modification of proteins found in eukaryotes with function similar to ubiquitin. SUMO fusion system facilitates efficient expression of recombinant proteins in *E.coli*.[9] SUMO fusion tag can be easily removed by SUMO protease which recognizes the tertiary structure of the SUMO protein and thus offers the distinct advantage over other fusion systems, which need an engineered TEV protease recognition site between fusion tag and desired peptide.[10] Moreover, since my strategy is based on enzymatic cleavage of partially protected peptide, for which I prefer to use TEV protease recognition site, the removal of fusion tag should be performed by a protease other than TEV protease.



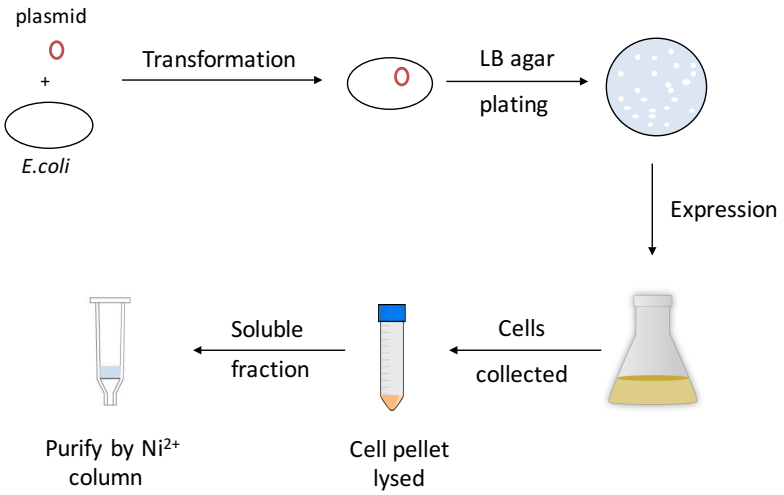
9

Figure 2-3: Representation of fusion peptide **9** with SUMO fusion tag, a His-tag and TEV protease cleavage site (ENLYFQ) followed by a model sequence.

2-4-1. Recombinant expression of fusion peptide **9**[11]

A pET vector encoding fusion peptide **9** was inserted into BL21(DE3) *E.coli* strain by heat-shock treatment and the process is called transformation. The plasmid contains the DNA sequence coding the desired peptide and ampicillin resistance gene that allows easy isolation of transformed cells by growing it on LB agar containing ampicillin. These cells were then used for obtaining starter culture for over-expression of fusion peptide **9** using IPTG as inducer. Fusion peptide **9** was isolated from the cells in the native folded form and is called soluble fraction (Figure 2-4). The purification of recombinant protein is usually facilitated by engineering the desired protein with a poly-His tag which has the ability to bind to immobilized metal ion matrix. The imidazole side chains of His-tag can form reversible coordinate bonds to divalent metal ions and Ni^{2+} shows highest affinity and selectivity. Thus, immobilized Ni^{2+} matrix was used for affinity purification of fusion peptide **9** in the native form. The peptide was dialyzed against water and further purified by reverse phase HPLC. The fusion peptide **9** was characterized by ESI-MS as shown in Figure 2-4.

(a)



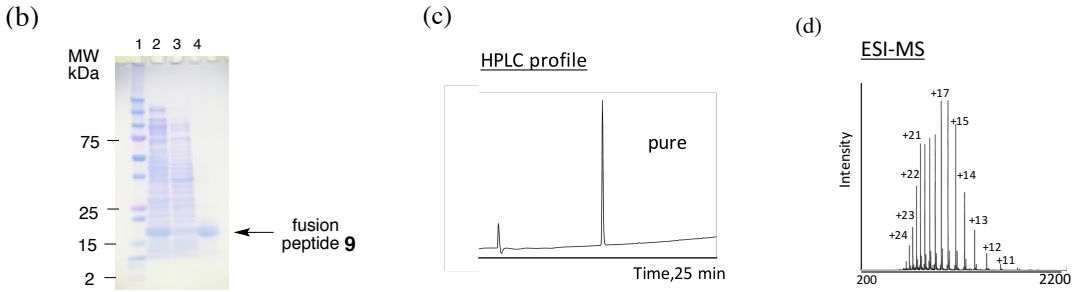


Figure 2-4: (a) Schematic representation of *E.coli* expression. (b) SDS PAGE lane 1: marker, lane 2: soluble fraction, lane 3: insoluble fraction, lane 4: after Ni^{2+}affinity purification. (c) HPLC profile of purified **9**. (d) ESI-MS characterization of **9** $[M+H]^+$ calculated 15766.70, observed: $[M+H]^+$ 15769.91

2-4-2. SUMO protease cleavage

Fusion peptide **9** was treated with SUMO protease, which is a highly active cysteinyl protease with high specificity. The assay was performed at the optimal temperature 30°C and pH 8.0 in 50 mM Tris-HCl buffer. The desired model peptide **10** having four Lys and a TEV protease cleavage site was obtained. The reaction was monitored by reverse phase HPLC and model peptide **10** was characterized by ESI-MS (Figure 2-5).

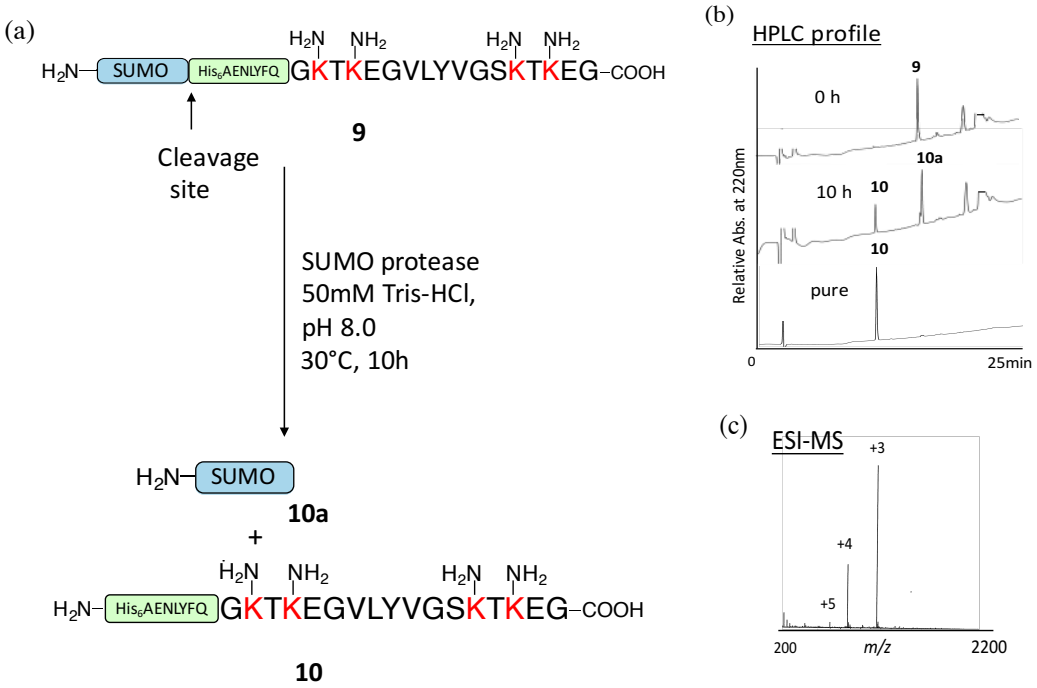


Figure 2-5: (a) Schematic representation of SUMO protease cleavage. (b) Reaction monitoring by HPLC. **(c)** ESI-MS characterization of **10** $[M+H]^+$ calculated 3469.73, observed: $[M+H]^+$ 3470.77.

2-5. Chemical synthesis of model peptide **10** by Fmoc-SPPS

The purpose of using fusion peptide **9** to obtain peptide **10** was to show that a peptide as small as 30 amino acids can be obtained by recombinant expression. Because model peptide **10** was not expressed as a tandem repeat the quantity of peptide **10** was low due to the relatively huge size of SUMO protein. In order to optimize the protection reaction and other reactions of the strategy, appropriate amount of the same model peptide **10** was necessary and hence it was also synthesized by Fmoc-based SPPS simultaneously. The product **10** was purified by reverse phase HPLC and characterized by ESI-MS (Figure 2-6).

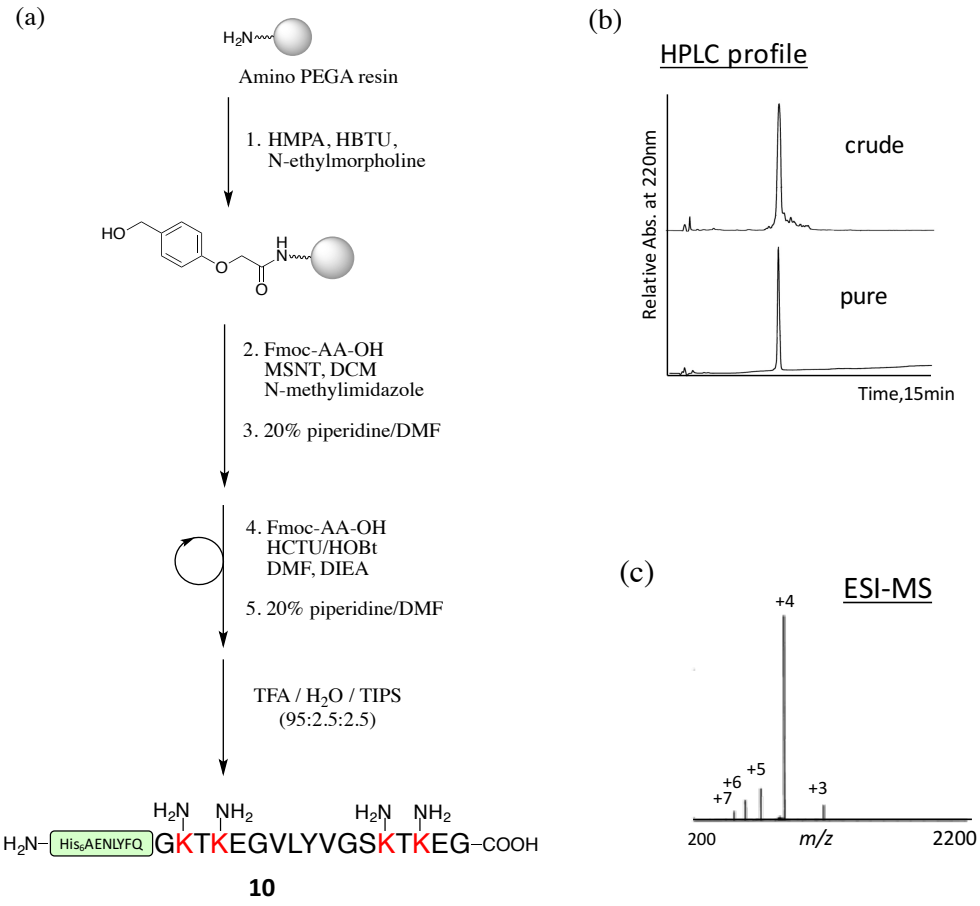


Figure 2-6: (a) Synthetic scheme for model peptide **10** by Fmoc-SPPS. (b) HPLC profile of crude and purified **10**. (c) characterisation by ESI-MS of **10** $[M+H]^+$ calculated 3469.73, observed: $[M+H]^+$ 3470.77.

2-6. Partial protection of Model peptide **10**

Partial protection of model peptide **10** with AGlOC-imidazole protecting group **4** was performed in AcCN/H_2O (3:2) at 40°C for 5 h using NHS as activating reagent and DIEA as base. The reaction was monitored by reverse phase HPLC and a by-product with additional protecting group was observed which might be due to reaction of His side chain, however, this was found to be unstable during the course of reaction. The protection reaction for **10** was not efficient in DMF and DMSO as the peptide was not completely soluble. The desired peptide **11** with all Lys and *N*-terminal protection thus obtained was purified by reverse phase HPLC and was characterized by ESI-MS. The product peptide **11** was confirmed by measuring HRMS.

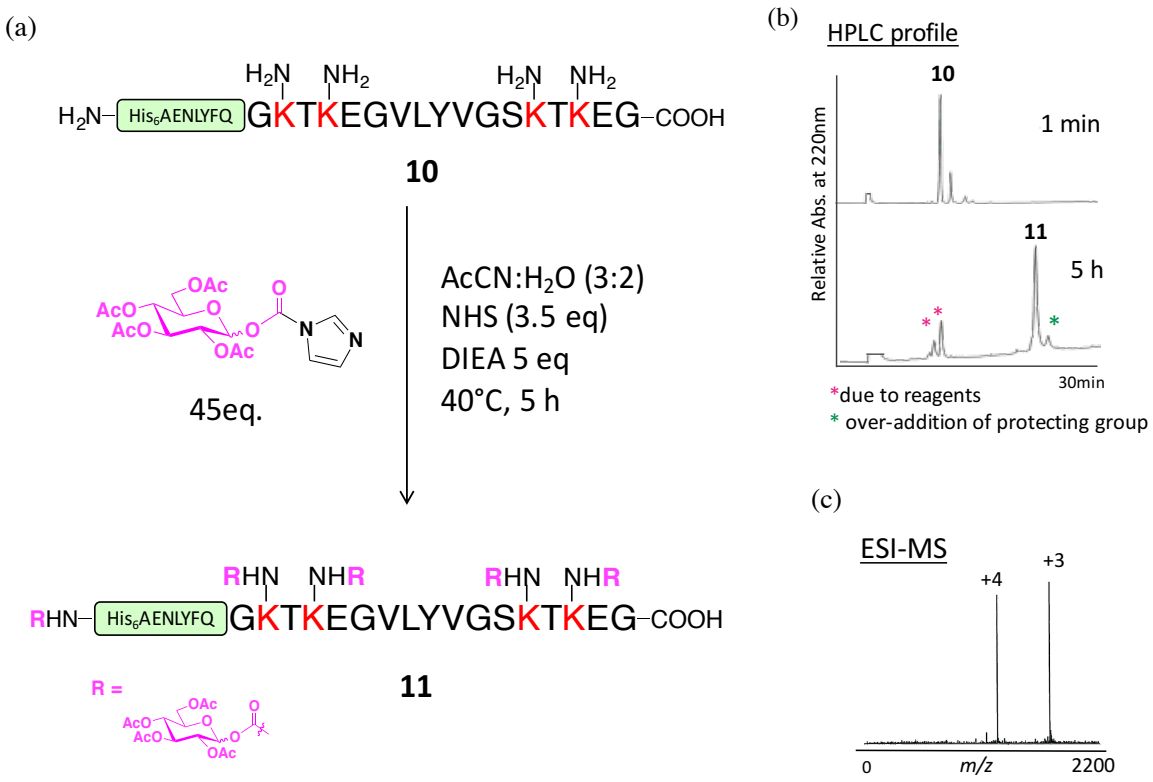


Figure 2-7: (a) Scheme for partial protection of **10** with AGlOC protecting group. (b) HPLC profile of reaction monitoring of optimized condition; isolated yield= 23.2%. (d) ESI-MS of desired product **11** $[M+H]^+$ calculated 5341.32, observed: $[M+H]^+$ 5341.31.

2-7. Deacetylation of the protecting groups

The AGlOC protected peptide **11,** had the glucose hydroxy groups protected with acetyl, but in order for **11** to show hydrophilic character and be able to dissolve into aqueous solutions, the acetyl groups must be selectively removed to liberate free hydroxyl groups. The hydrolysis of acetyl ester linkage to hydroxyl groups on sugars can be efficiently performed under basic conditions. Hydrazinolysis of acetyl esters was performed using 10% $NH_2NH_2.H_2O$ in DMF for 5 min at ambient temperature and the deacetylation was complete to give peptide **12.**[12] The product was precipitated in ice cold diethyl ether and then washed again with diethyl ether. Later, peptide **12** was purified by reversed phase HPLC and characterized by ESI-MS (Figure 2-8).

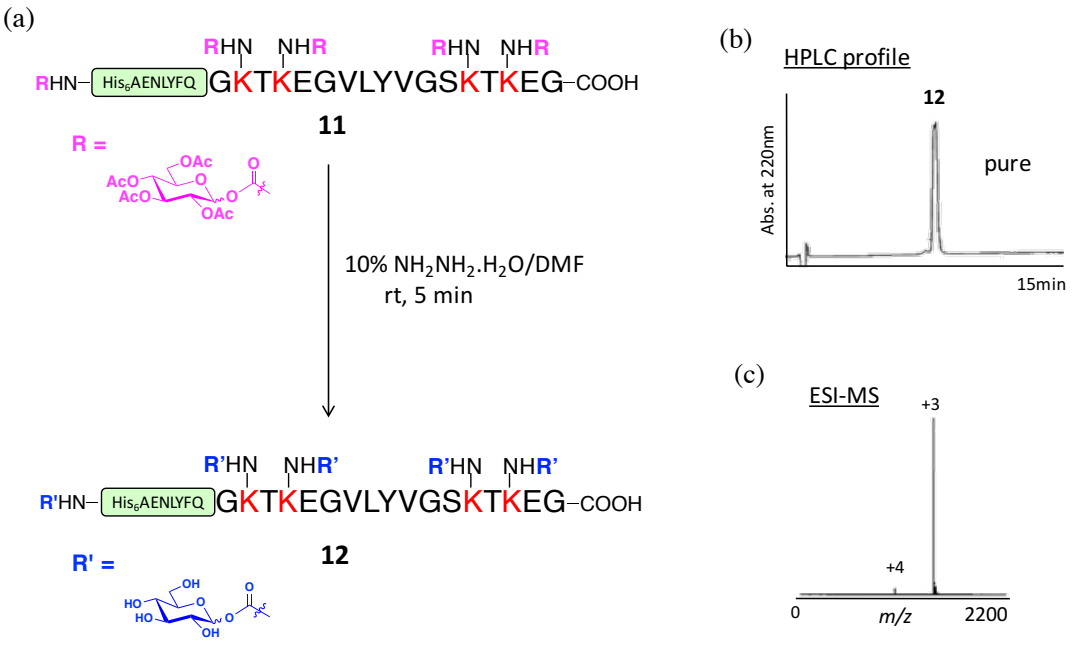


Figure 2-8: (a) Scheme for deacetylation of AGlOC protecting group in **11**. (b) HPLC profile of pure **12** isolated yield = 52.5 %. (d) ESI-MS of desired product **12** $[M+H]^+$ calculated: 4500.59, observed: $[M+H]^+$ 4500.40

2-8. One-pot protection and deacetylation of protecting group

Partial protection of peptide with AGlOC-imidazole **4** is more efficient if protection and deacetylation of the protecting group can be performed in one-pot, so that the intermediate hydrophobic protected peptide **11** with acetyl groups on glucose need not be dealt with. Therefore, to examine the efficiency of protection

and deacetylation as a one-pot reaction, conditions were optimized using model peptide **10**. At first, protection was performed in AcCN/H_2O (2:1) for 5 h at 40°C and reaction mixture was allowed to cool down to room temperature and 10% $NH_2NH_2.H_2O$ was added to the reaction mixture. The reaction mixture was stirred for 1 h at room temperature, but deacetylation was not complete. Then, protection was performed in 70 % DMF in H_2O for 5 h at 40°C and after completion of the protection, deacetylation was examined using 10%, 15% and 20% $NH_2NH_2.H_2O$ at room temperature and found that reactions with 15 % and 20 % $NH_2NH_2.H_2O$ gave good results. The reaction was monitored by reverse phase HPLC. The product **12** was purified by reverse phase HPLC and was characterized by ESI-MS. The data of the optimized condition is shown in Figure 2-9. The isolated yield of peptide **12** using one-pot protocol increased the isolated yield to 36% from 13% over two separate isolation steps. These results showed that one-pot protocol was useful and efficient.

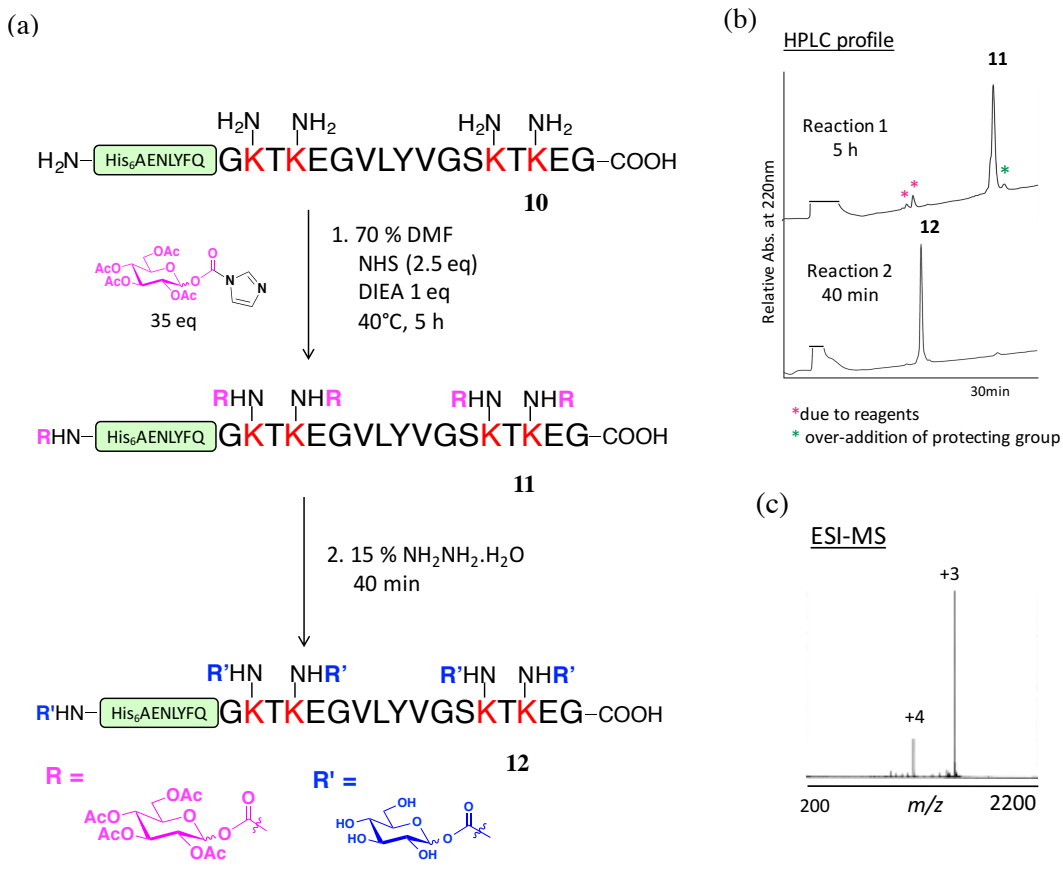


Figure 2-9: (a) Schematic representation of one-pot protection and deacetylation of AGlOC (b) HPLC profile of reaction monitoring; isolated yield = 36 %. (c) Characterization of **12** by ESI-MS $[M+H]^+$ calculated: 4500.94, observed: 4501.22.

Because of successful establishment of a one-pot protocol for protection of peptide **10** with **4**, the same strategy was used to protect **10** with lactose-derived protecting group **6** (Figure 2-10). This protocol also gave the desired peptide **14** *via* intermediate **13** in good yield.

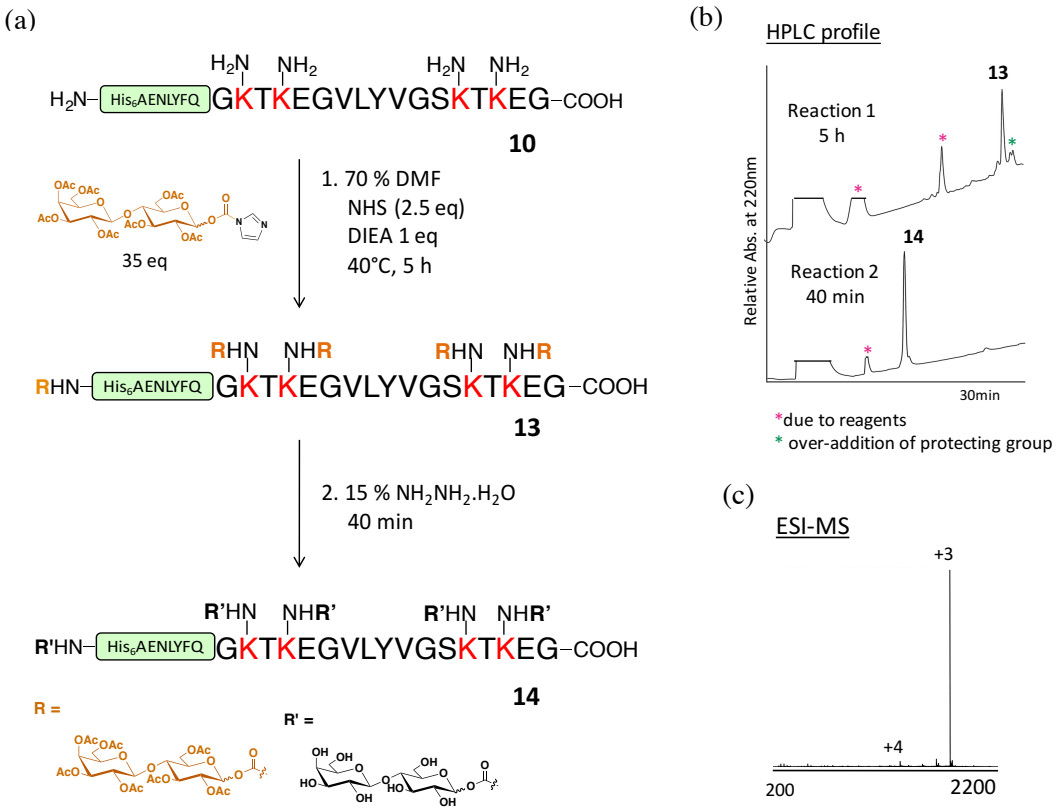


Figure 2-10: (a) Schematic representation of one-pot protection and deacetylation of AGlOC (b) HPLC profile of reaction monitoring isolated yield = 17 %. (c) Characterization of **14** by ESI-MS $[M+3H]^{3+}$ calculated: 1771.43, observed: 1771.49.

2-9. Comparison of hydrophilicity between unprotected peptide and protected peptides

In order to prove that the carbohydrate-derived protecting group increased the hydrophilicity of the protected peptide, it was essential to compare the hydrophilicity between unprotected peptide **10** and protected peptides **12** and **14**. The simple and most efficient way is to compare the retention time of the peptides in reverse phase HPLC. Retention time (t_R) is the amount of time a peptide/compound takes to pass through the column from the point of injection. It depends on the interaction of peptides with the hydrophobic stationary phase implying that higher the hydrophobicity of the peptide stronger the binding

affinity to column which increases the retention time of the peptide or the time spent in the column. Usually used solvent system for peptides/proteins in reverse phase HPLC is 0.1% TFA in H_2O and 90% AcCN/0.1% TFA in H_2O. When the retention time of unprotected and protected peptides were compared in the acidic buffering system, unprotected peptide **10** and protected peptide **12,** elutes around the same time, while **14** elutes earlier (Figure 2-11a). The similar retention time of unprotected peptide **10** and protected peptide **12** may be attributed to the protonation of free amines in **10**. The comparison of retention time was also performed in neutral solvent system 50mM NH_4OAc buffer, pH 7.0 and AcCN where unprotected peptide **10** showed higher retention time than **12** and **14** (Figure 2-11b). These data proved that hydroxyl groups of sugar increased the hydrophilicity of protected peptides and protection of peptide **10** with lactoside showed the most hydrophilic nature.

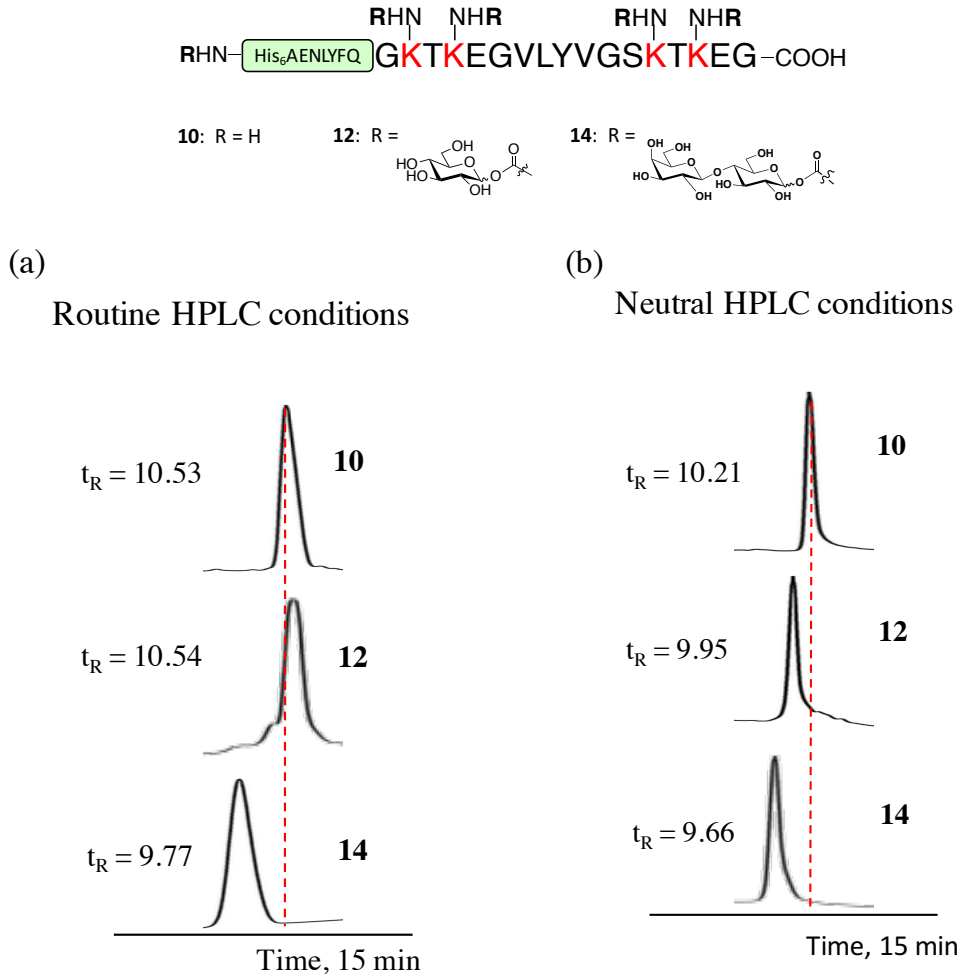


Figure 2-11: (a) Comparison of t_R using acidic solvent system. Solvent A: 0.1% TFA in H_2O, Solvent B: 90% AcCN/0.1% TFA in H_2O (80:20 to 60:40 over 15 min). (b) Comparison of t_R using neutral solvent system. Solvent A: 50mM NH_4OAc buffer, pH 7.0, Solvent B: AcCN (90:10 to 40:60 over 15 min).

2-10. Removal of protecting group

After successfully optimizing conditions for protection and comparing the hydrophilic nature, the next step was to optimize conditions for removal of the protecting group. Generally, carbamate protecting groups have more than one bond whose fission can result in deprotection of amine. Fission **A** is the most probable pathway, while fission **B** and **C** are less likely due to the low reactivity of urethane carbonyl group to nucleophiles (Figure 2-11).[3] Fission **A** can take place by different mechanisms based on the **R** group. If **R** group can form a stable carbocation or carbanion after cleavage, it generates a carbamic acid that spontaneously decomposes to the free amine liberating CO_2 (ex. Boc). This is usually achieved by catalytic hydrogenation or by acidolysis. If **R** group has an acidic methylene group β to the oxycarbonyl unit, then it undergoes β-elimination to generate free amine (ex. Fmoc), it is favored by base. Since, carbohydrates can form cation that can be stabilized, acidolysis can remove the carbohydrate-derived protecting groups.

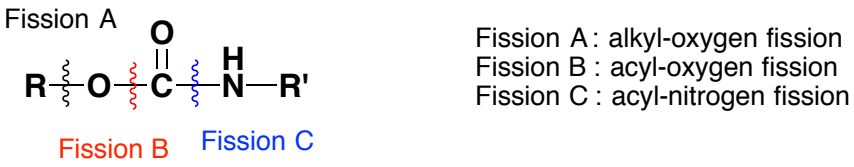


Figure 2-11: Representation of different cleavage positions in carbamate-type protecting group.

For optimizing conditions for removal of AGlOC group, glucosidases can also be employed as explained in Chapter 1 apart from acidolysis. At first, α- and β-glucosidases (1:1 mixture) were used as it proceeds under the mildest conditions, but unfortunately the reaction was slow and incomplete (table 2-2). Moreover, the enzymes were needed in the pure form free from contaminating peptidases and also needs an additional purification step to isolate the product peptide. Therefore, various acidic conditions were investigated (table 2-2) and stronger acidic condition in the presence of TfOH in TFA (entry 7) was found to give efficient results. The reaction was monitored and purified by reverse phase HPLC and desired product **12** was characterized by ESI-MS and the data for the optimized condition is shown in Figure 2-12.

Entry	Conditions	Temp.	Time	Results
1	α- and β-glucosidases*	37°C	48 h	Incomplete
2	20% TFA	50°C	20 h	Incomplete
3	TFA/TIPS/H_2O (95:2.5:2.5)	rt	2 h	Incomplete
4	TFA/TIPS/ H_2O/EDT (92.5:2.5:2.5:2.5)	rt	2 h	Incomplete
5	TFA/TIPS/H_2O/EDT (37:1:40:1)	rt	2 h	Incomplete
6	TFA/Thioanisole/H_2O (15:4:1)	rt	2 h	Incomplete
7	**DMS/*m*-cresol/TFA/TfOH (3.5:1:5:0.5)**	**0°C**	**0.5 h**	**Complete**

Table 2-2: Conditions for removal of protecting group, * 1:1 mixture of α- and β-glucosidases.

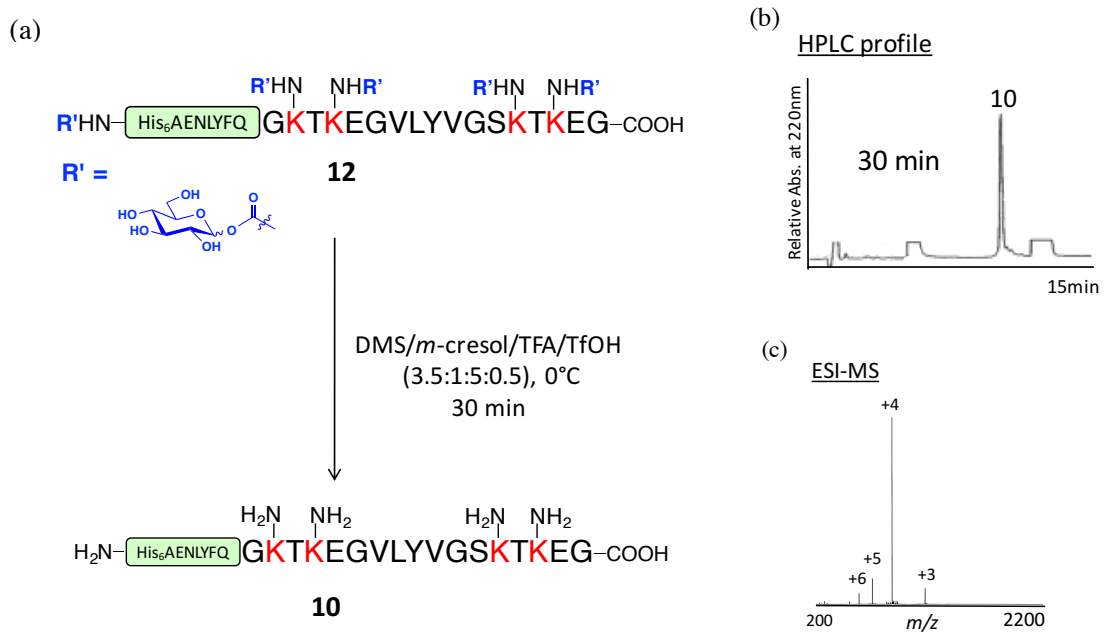


Figure 2-12: (a) Schematic representation of removal of protecting groups. (b) HPLC profile after 30 min, isolated yield = 47 %. (c) Characterization of **10** by ESI-MS $[M+H]^+$ calculated: 3469.84, observed: 3470.80.

2-11. TEV Protease cleavage

Protease cleavage of a protected peptide being already reported,[13] I intended to perform TEV protease cleavage towards the protected peptide, as it is highly specific, unlike chemical cleavage methods or other

proteases. Its specificity is signified by its six-amino acid recognition sequence ENLYFQ which is highly unlikely to be present in the desired peptide sequence. TEV protease cleaves towards the C-terminal of the recognition site under milder conditions of pH and temperature in aqueous buffer. The protease reaction was performed towards **12** to give a partially protected peptide with free *N*-terminus **15**. The reaction was monitored by reverse phase HPLC and characterized by ESI-MS and the data is shown for the optimized condition. The protecting groups in the carbamate-form on the peptide showed instability towards ESI-MS conditions and the intensity of these peaks varied with the ESI-MS conditions, while HPLC profile showed only a single peak. The product **15** was further confirmed by performing HRMS.

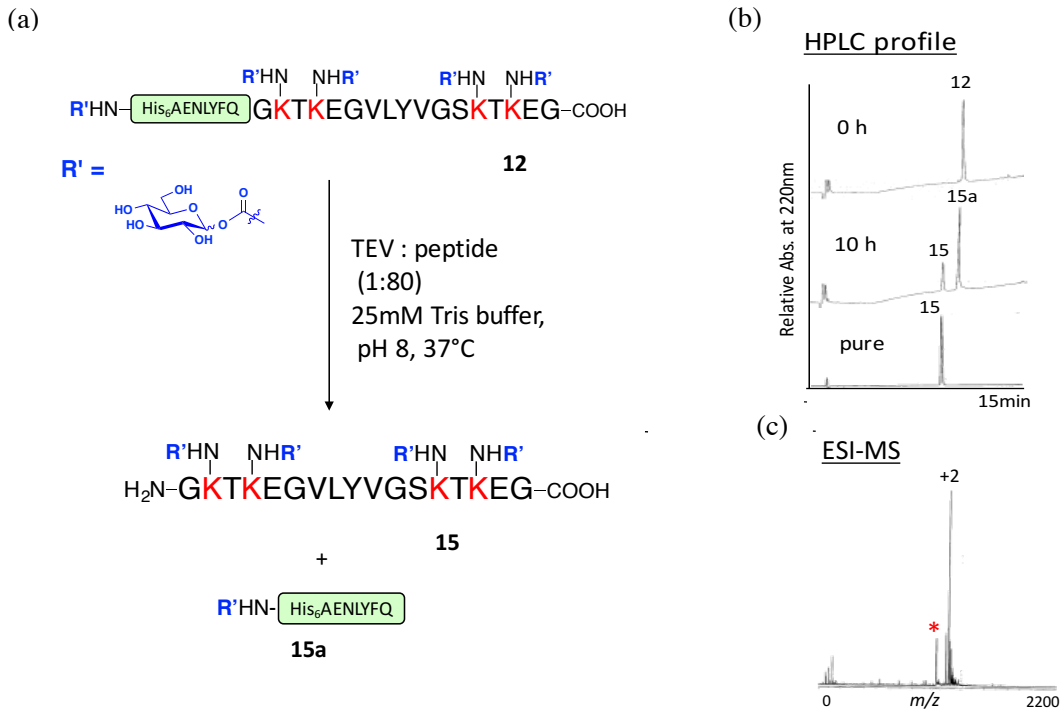


Figure 2-13: (a) Schematic representation of TEV protease cleavage (b) HPLC profile of reaction monitoring; isolate yield = 30 %. (c) ESI-MS of **15**, * peak arising due to instability of protecting group to ESI-MS that corresponds to -1 protecting group, $[M+H]^+$ calculated: 2605.64, observed: 2604.95

2-12. Modification of N-terminus with Selenocysteine.

Towards the free *N*-terminal of **15**, desired modification such as coupling of any mercapto amino acids, coupling of Sec, short peptide segment or oligosaccharyl amino acid can be performed. In my strategy, I wanted to modify **15** with Sec, as Sec has several advantages over mercapto amino acids or Cys which are

previously mentioned as a part of introduction. Boc-Sec(PMB)-OH **17** was synthesized by the protocol reported by Payne et al.,[14] starting from Boc-Ser-OH **16** (Figure 2-14).

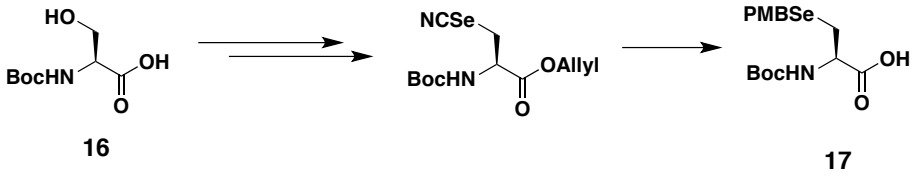


Figure 2-14: Synthesis of Boc-Sec(PMB)-OH from Boc-Ser-OH.

After synthesis of Boc-Sec(PMB)-OH **17,** I examined direct coupling of Boc-Sec(PMB)-OH **17** towards protected peptide. The commonly used activating reagents for *in situ* activation of Boc-Sec(PMB)-OH **17** to obtain active intermediates were used in DMF as solvent (table 2-3).[15] However, the coupling was very slow and low yielding giving rise to byproducts. Since, an activated intermediate was not observed in most cases, I decided to synthesize succinimide derivative Boc-Sec(PMB)-OSu **18** which was stable enough to be isolated.[15] Boc-Sec(PMB)-OSu **18** was synthesized before use by employing reagents DIC and NHS and isolated by simple work-up with water (Figure 2-16). The crude form of **18** was directly used for coupling with peptide **15** which successfully gave the desired product **19** having selenocysteine at the *N*-terminal in good yield (Figure 2-16). The reaction was monitored by reverse phase HPLC and characterized by ESI-MS. The protecting groups in the carbamate-form on the peptide showed instability towards ESI-MS conditions. As shown in figure 2-16b, HPLC profile showed only a single peak. The product **19** was further confirmed by performing HRMS.

Reagents	t (h)	% conversion (HPLC)
Boc-Sec(PMB)-OH, HBTU, HOBt, DIEA	20 h	< 1%
Boc-Sec(PMB)-OH, DIC, HOBt, DIEA	20 h	< 1%
Boc-Sec(PMB)-OH, DIC, NHS, DIEA	22 h	14 %
Boc-Sec(PMB)-OH, PyBOP, DIEA	17 h	50 %
Boc-Sec(PMB)-OH, HCTU, NHS, DIEA	17 h	10 %

Table 2-3: Table representing various *in situ* activation conditions[15] for Boc-Sec(PMB)-OH **17** and reagents used for direct coupling of **17** to **15**. Yields were estimated by HPLC.

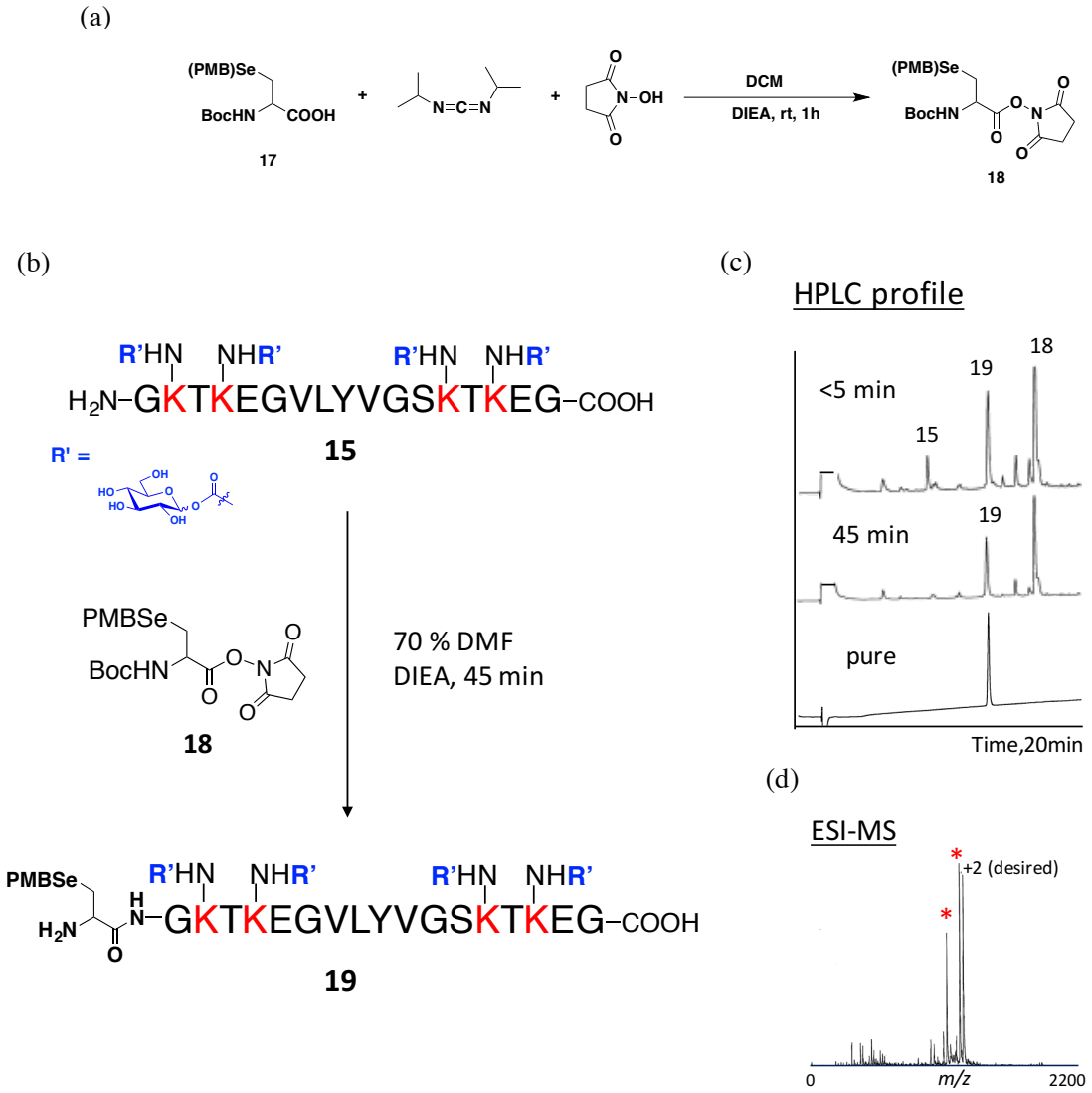


Figure 2-15: (a) Scheme for synthesis of **18.** (b) Schematic representation of Boc-Sec(PMB)-OSu coupling. (c) HPLC profile of reaction monitoring; isolated yield = 17 %. (d) ESI-MS of **15**, * peaks arising due to instability of protecting group to ESI-MS that corresponds to -1 protecting group and -44 Da, $[M+H]^+$ calculated: 2975.95, observed: 2974.99

2-13. Deprotection of peptide **19**

In order to use peptide **19** for NCL, all protecting groups must be removed. The Boc protecting group can be removed under acidic conditions using TFA, while removal of PMB group can be performed either using TfOH/TFA[16] or by oxidative deprotection using Iodine or DMSO in TFA.[17] Oxidative deprotection needs optimization for each peptide and moreover, it is problematic in the presence of oxidation

sensitive reagents. Since, removal of carbohydrate-derived protecting group also needs TfOH/TFA for efficient deprotection, I planned to remove all the protecting groups in a single step using TfOH/TFA in the presence of scavengers (figure 2-17).[16] All protecting groups were removed in 2 h at 0°C, and the unprotected peptide **20** was isolated by precipitating in ice cold ether. The peptide **20** was purified by reverse phase HPLC and characterized by ESI-MS as shown below. The unprotected peptide **20** can then be readily used to perform NCL reaction which is well-established.

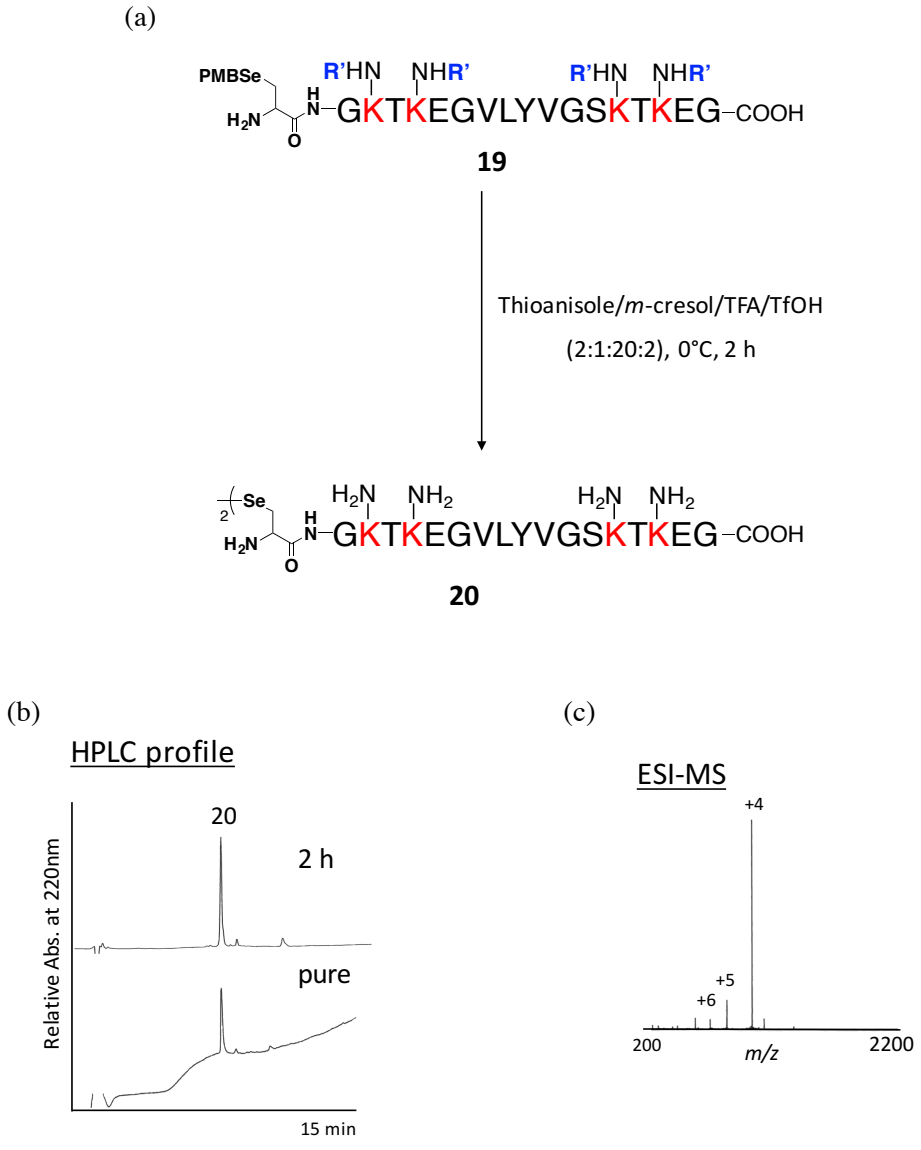


Figure 2-16: (a) Schematic representation of removal of all protecting groups of **19.** (b) HPLC profile of **20**; isolated yield = 12 %. (c) ESI-MS of **20** $[M+H]^+$ calculated: 3861.84, observed: 3861.90

.

References

1. (a) Chen, D.; Disotuar, M. M.; Xiong, X.; Wang, Y.; Chou, D. H-C.; *Chem. Sci*. **2017**, *8*, 2717. *Chem. Rev*. **1996**, *96*, 683. (b) Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B.; *Angew. Chem. Int. Ed*. **2006**, *45*, 5307. (c) Rutjes, F. P. J. T.; *Nat. Chem. Bio*. **2015**, *11*, 306.
2. (a) Makino, T.; Matsumoto, M.; Suzuki, Y.; Kitajima, Y.; Yamamoto, K.; Kuramoto, M.; Minamitake, Y.; Kangawa, K.; Yabuta, M.; *Biopolymers*. **2005**, *79*, 238. (b)Kawakami, T.; Hasegawa, K.; Teruya, K.; Akaji, K.; Horiuchi, M.; Inagaki, F.; Kurihara, Y.; Uesugi, S.; Aimoto, S.; *J. Pept. Sci*. **2001**, *7*, 474.
3. (a) Bergmann, M.; Zervas, L.; *Berichte der Deutschen Chemischen Gesellschaft* **1932**, *65*, 1192. (b) Sureshbabu, V. V.; Narendra, N.; *Amino Acids, Peptides and Proteins in Organic Chemistry*, **2011**, *4*, 1. (c) Isidro-Llobet, A.; Álvarez, M.; Albericio, F; *Chem. Rev.,* **2009**, *109*, 2455.
4. Ichikawa, Y.; Sim, M. M.; Wong, C.-H.; *J. Org. Chem*. **1992**, *57*, 2943.
5. Knoben, H.-P.; Schlüter, U.; Redlich, H.; *Carbohydr. Res*. **2004**, *339*, 2821.
6. (a) Blankmeyer-Menge, B.; Nimtz, M.; Frank, R.; *Tetrahedron Lett*. **1990**, *31*, 1701. (b) Harth-Fritschy, E.; Cantacuzène, D.; *J. Peptide res*. **1997**, *50*, 415.
7. Anderson, G. W.; Zimmerman, J. E.; Callahan, F.; *J. Am. Chem. Soc*. **1964**, *86*, 1839.
8. König, W.; Geiger, R.; *Chem. Ber*. 1970, **103**, 788.
9. Malakhov, M. P.; Mattern, M. R.; Malakhova, O. A.; Drinker, M.; Weeks, S. D.; Butt, T. R.; *J. Struct. Funct. Genomics* **2004,** *5,* 75.
10. Marblestone, J. G.; Edavettal, S. C.; Lim, Y.; Lim, P.; Zuo, X.; Butt, T. R.; *Protein Sci*. **2006**, *15*, 182.
11. Nallamsetty, S.; Waugh, D. S.; *Nat. Protoc*. **2007**, *2*, 383.
12. Bardají, E.; Torres, J. L.; Clapés, P.; Albericio, F.; Barany, G.; Rodríguez, R. E.; Sacristán, M. P.; Valencia, G.; *J. Chem. Soc., Perkin Trans*. **1991**, *1*, 1755.
13. Makino, T.; Matsumoto, M.; Suzuki, Y.; Kitajima, Y.; Yamamoto, K.; Kuramoto, M.; Minamitake, Y.; Kangawa, K.; Yabuta, M.; *Biopolymers*. **2005**, *79*, 238.
14. Malins, L. R.; Mitchell, N. J.; McGowan, S.; Payne, R. J.; *Angew. Chem. Int. Ed*. **2015**, *54*, 1271.
15. Montalbetti, C. A. G. N.; Falque, V.; *Tetrahedron*. **2005**, *61*, 10827.
16. Koide, T.; Itoh, H.; Otaka, A.; Yasui, H.; Kuroda, M.; Esaki, N.; Soda, K.; Fujii, N.; *Chem. Pharm. Bull*. **1993**, *41*, 502.
17. Koide, T.; Itoh, H.; Otaka, A.; Furuya, M.; Kitajima, Y.; Fujii, N.; *Chem. Pharm. Bull*. **1993**, *41*, 1596.

Conclusion

A strategy for semisynthesis of glycoproteins with multiple modifications was established which involves the following steps

i. recombinant expression of desired peptide as fusion peptide.
ii. partial protection of Lys side chains of recombinant peptide with hydrophilic protecting group
iii. liberating free *N*-terminal amine of protected peptide by TEV protease
iv. coupling of suitably protected selenocysteine to the *N*-terminal amine
v. removal of all protecting groups by TfOH in TFA to obtain unprotected peptide with selenocysteine at the *N*-terminal position.

Thus, the above mentioned chemical modifications enabled me to efficiently modify the *N*-terminus of recombinant peptide that can be readily used for NCL.

By using hydrophilic carbohydrate-derived protecting group, I could solve the critical problem of solubility associated with current method of partial protection of recombinant peptides. The current method employs hydrophobic Boc group that makes protected peptide hydrophobic. The hydrophobic property is a big hindrance as it becomes difficult to handle peptide for chemical reactions or purification due to insolubility. The replacement of hydrophobic protecting group with a hydrophilic carbohydrate-derived urethane protecting group not only solves the solubility problem but also allows access to various modifications of recombinant short peptides that was previously impossible.

The other development in this strategy towards semisynthesis is selective chemical modification of *N*-terminal of recombinant peptide that is favorable to perform NCL, which is the most efficient peptide ligation strategy.

The new strategy established here enables the preparation of short or long recombinant peptides having selenocysteine at the *N*-terminal. This will advance the field of semisynthesis of various post-translationally modified proteins such as glycoproteins, phosphorylated proteins, and other desired modifications for structural and functional studies.

Experimental

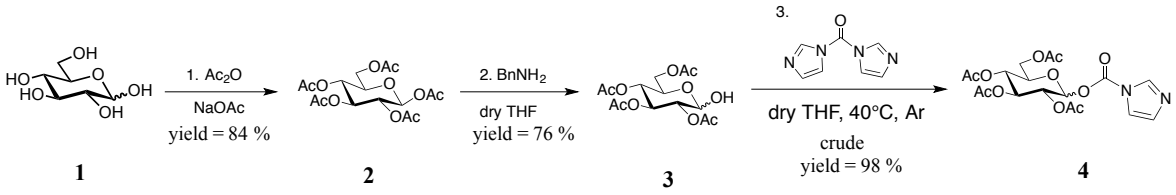
1. General methods and abbreviation

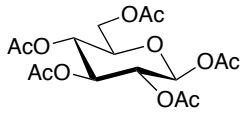
2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), Hydroxybenzotriazole (HOBt), 2-(6-Chloro-1-H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU), N-Hydroxysuccinimide (NHS) that were used was purchased from PEPTIDE INSTITUTE. INC. 1-(Mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole (MSNT), Amino PEGA resin, Fmoc-Gly-OH, Fmoc-Ala-OH, Fmoc-Val-OH, Fmoc-Leu-OH, Fmoc-Tyr(OtBu)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Lys(Boc)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Phe-OH, Fmoc-His(Trt)-OH, Boc-Ser-OH were purchased from Novabiochem. *N*-methylimidazole, Triisopropylsilane (TIPS), *N, N*-Diisopropylethylamine (DIPEA/DIEA), Dimethylsulfide (DMS), Trifluoromethanesulfonic acid (TfOH), *N, N'*-carbonyldidimidazole, D-(+)-Lactosemonohydrate were purchased from TOKYO CHEMICAL INDUSTRY CO., LTD. Piperidine, Thioanisole, 3-Mercaptopropionic Acid (MPA), Tetrahydrofuran (THF) were purchased from nacalai tesque. Sodium acetate, acetic anhydride, benzylamine, hydrazine monohydrate, *N, N'*-diisopropylcarbodiimide (DIC), Triethylamine, Sodium tetrahydroborate, Methanesulfonyl chloride, p-methoxybenzylchloride, Ammonium chloride, Methanol, Trifluoroacetic acid (TFA), Dimethyl sulfoxide (DMSO) were purchased from Wako Chemical Industries, Ltd. *N*-(3-Dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC), Potassium selenocyanate, D-(+)-Glucose, D-(+)-Galactose, α,β-Glucosidases, cOmplete His-tag purification resin were purchased from SIGMA-ALDRICH. Acetonitrile, *N, N'*-Dimethylformamide (DMF), Dichloromethane (DCM) were purchased from KANTO CHEMICAL CO., INC. The plasmid containing desired genes in pET vector for *E.coli* expression with *lac* operon, T7 promoter, ampicillin resistance, was purchased from INVITROGEN. *E.coli* BL21(DE3), SUMO protease were also purchased from INVITROGEN. TurboTEV protease was purchased from ACCELAGEN.

RP-HPLC analyses were performed using 0.1% aq. TFA and 90% aq. CH_3CN containing 0.1% TFA, and RP-UHPLC analyses were performed using 0.1% aq. HCOOH and 90% aq. CH_3CN containing 0.09% HCOOH, using CAPCELL PAK C18, Cadenza CD-C18 (Imtakt) or Proteonavi (Shiseido) for analytical or/and semipreparative HPLC. LC/MS and ESI-MS spectra were recorded on a Bruker Daltonics amaZon-mass spectrometer system and Bruker Daltonics Esquire 3000 mass spectrometer, respectively.

2. Experimental procedure and characterization

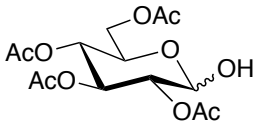
1. Synthesis of AGlOC-imidazole **4** [1,2]





2

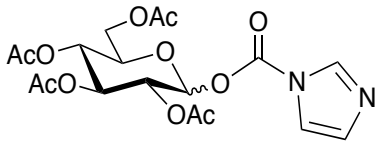
[(2R,3R,4S,5R,6S)-3,4,5,6-tetraacetyloxyoxan-2-yl]methyl acetate (2): A suspension of sodium acetate (4.6 g, 55.5 mmol) in acetic anhydride (100 mL, 1 mol) was taken in a 300 mL flask and was refluxed at 170°C for 10min with the condenser connected to a guard tube filled with $CaCl_2$, to which D-(+)-glucose monohydrate (10 g, 55.5 mmol) was added in parts due to effervescence. The reaction mixture gradually became clear by the end of addition of glucose, which was refluxed for 1 h and it turned pale brown. Reaction mixture was cooled down to room temperature and poured onto ice in a beaker which was stirred overnight. The precipitate obtained was filtered using CH_2Cl_2 and recrystallized from EtOH and petroleum ether to afford **2** as a white crystal (18.2 g, 46.6 mmol, 84 %).



3

(2*R*,3*R*,4*S*,5*R*)-2-(acetoxymethyl)-6-hydroxytetrahydro-2*H*-pyran-3,4,5-triyl triacetate 3: To a solution of **2** (10 mg, 25.6 mmol) in dry THF (20 mL), benzyl amine (4.9 mL, 45.7 mmol) was added to it and stirred at ambient temperature for 13 h. Reaction mixture was diluted with CH_2Cl_2 and then washed over 500 mM ice-cold aqueous HCl, followed by washing twice with saturated $NaHCO_3$. Finally, CH_2Cl_2 layer was washed over saturated NaCl solution and CH_2Cl_2 was concentrated to get a yellow semisolid which was

purified by flash column chromatography (EtOAc/Hexane = 4:6 to 9:1) to afford **3** as a pale yellow solid (6.7 g, 19.2 mmol, 76 %).

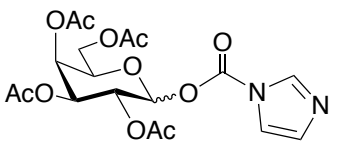


4

(3*R*,4*S*,5*R*,6*R*)-2-((1*H*-imidazole-1-carbonyl)oxy)-6-(acetoxymethyl)tetrahydro-2*H*-pyran-3,4,5-triyl triacetate 4: In a round bottom flask was taken **3** (0.20 g, 0.57 mmol) vacuum dried for 5min to which 1,1'-carbonyldiimidazole (0.33 g, 2.0 mmol) was added and saturated with argon. The reactants were dissolved in dry THF (10 mL) and stirred at 40°C for 3.5 h under Ar. Reaction mixture was then diluted with CH_2Cl_2 and workup was done using saturated NH_4Cl solution. The organic layer was dried over Na_2SO_4 and concentrated *in vacuo* carefully to afford **4** as a white solid (0.25 g, 0.56 mmol, 98 %, α:β = 4:5). The crude product was directly used for protection reactions due to its instability to flash column chromatography purifications. The product **4** stored at -30°C under Ar and was found to be stable upto 10 days. The formation of product **4** was confirmed by performing ^{1}H and identifying the characteristic NMR peaks for imidazole ring and H_α and H_β protons. ESI-MS: *m/z* calculated for $C_{18}H_{22}N_2O_{11}$: $[M+Na]^+$ 465.11, observed: $[M+Na]^+$ 465.03.

^{1}H (400 MHz, CD_3Cl, TMS) δ 8.22 (br s, 0.6 H, H_α(Imid)), 8.14 (br s, 0.4 H, H_β(Imid)), 7.48 (br s, 0.6 H, H_α(Imid)), 7.41 (br s, 0.4 H, H_β(Imid)), 7.16 (br s, 0.6 H, H_α(Imid)), 7.09 (br s, 0.4 H, H_β(Imid)), 6.50 (d, 0.6 H, 1H_α *J* = 3.39 Hz,), 5.83 (d, 0.4 H, 1H_β *J* = 7.71 Hz), 5.53-5.48 (dd, 0.7 H, *J* = 9.91 Hz), 5.37-5.27 (ddd, 0.9 H, *J* = 9.54, 9.35, 2.99 Hz), 5.24-5.17 (m, 1.7 H), 4.35-4.29 (m, 1.1 H), 4.20-4.11 (m, 2 H), 3.97-3.95 (ddd, 0.4 H) 2.10 (s, 3 H), 2.08 (s, 3 H), 2.07 (s, 3 H), 2.05 (s, 3 H).

2. Synthesis of AGalOC-imidazole **5.** [1,2]

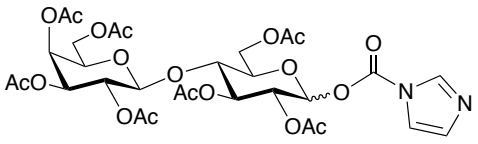


5

(3*R*,4*S*,5*S*,6*R*)-2-((1*H*-imidazole-1-carbonyl)oxy)-6-(acetoxymethyl)tetrahydro-2*H*-pyran-3,4,5-triyl triacetate (5): Synthesis of **5** was performed following the same protocol as given above for **4**, however isolation of 5 in the final reaction step was different. It was found that 5 was not stable when work-up was performed with saturated NH_4Cl solution, however, concentration of reaction mixture and fast filtration of reaction mixture over silica-gel afforded **5** as a white solid (20 % yield). The formation of product **5** was confirmed by performing ^1H and identifying the characteristic NMR peaks for imidazole ring and H_α and H_β protons. However, **5** was very reactive and found to decompose faster and thus, was not used as a protecting group. ESI-MS: *m/z* calculated for $C_{18}H_{22}N_2O_{11}$: $[M+Na]^+$ 465.11, observed: $[M+Na]^+$ 464.97.

^{1}H (400 MHz, CD_3Cl, TMS) δ 8.18 (br s, 0.6 H, H_α(Imid)), 8.15 (br s, 0.4 H, H_β (Imid)), 7.45 (m, 0.6 H, H_α(Imid)), 7.43 (m, 0.4 H, H_β(Imid)), 7.15 (br s, 0.6 H, H_α(Imid)), 7.09 (br s, 0.4 H, H_β(Imid)), 6.55 (d, 0.6 H, 1H_α, *J* = 3.53 Hz,), 5.77 (d, 0.4 H, 1H_β, *J* = 8.25 Hz), 5.58-5.57 (dd, 0.6 H, *J* = 2.94 Hz,), 5.50-5.46 (m, 0.9 H), 5.45-5.44 (m, 0.6 H), 5.17-5.14 (dd, *J* = 7.21, 3.27 Hz, 0.4 H) 4.40-4.37 (m, 0.6 H), 4.17-4.12 (m, 2 H), 2.20 (s, 3 H), 2.18 (s, 3 H), 2.06 (s, 3 H), 2.04 (s, 3 H), 2.02 (s, 3 H)

3. Synthesis of ALaOC-imidazole **6.** [1,2]



6

(2*S*,3*R*,4*S*,5*S*,6*R*)-2-(((2*R*,3*R*,4*S*,5*R*)-6-((1*H*-imidazole-1-carbonyl)oxy)-4,5-diacetoxy-2-(acetoxymethyl)tetrahydro-2*H*-pyran-3-yl)oxy)-6-(acetoxymethyl)tetrahydro-2*H*-pyran-3,4,5-triyl triacetate (6): Synthesis of 6 was performed following exactly the same protocol as given above for 4, except for reaction step 1 which was refluxed for 2 h. The product 6 stored at -30°C under Ar and was found to be stable up to 10 days. The characteristic NMR peaks in 1H NMR (400MHz, CDCl3) due to H_α 6.40 ppm and H_β 5.79 ppm were observed. ESI-MS: m/z calculated for $C_{30}H_{38}N_2O_{19}$: [M+K]+ 769.17, observed: [M+K]+ 769.67.

^{1}H(400 MHz, CD_3Cl, TMS) δ 8.22 (br s, 0.6 H, H_α(Imid)), 8.12 (br s, 0.4 H, H_β(Imid)), 7.49 (m, 10.6 H, H_α(Imid)), 7.40 (m, 0.4 H, H_β(Imid)), 7.16 (br s, 0.6 H, H_α(Imid)), 7.09 (br s, 0.4 H, H_β(Imid)), 6.43 (br s,

0.6 H, 1H$_\alpha$), 5.79 (d, 0.4 H, H$_\beta$, *J* = 9.1 Hz), 5.50-5.46 (dd, *J* = 10.23, 0.6 H), 5.36-5.30 (dd, 1.2 H), 5.20-5.18 (dd, *J* = 9.66 Hz, 0.4 H), 5.15-5.11 (dd, *J* = 9.66 Hz, 1.4 H), 4.98-4.95 (dd, *J* = 10.23 Hz, 0.9 H), 4.52-4.48 (m, 1.7 H), 4.45-4.43 (dd, *J* = 5.8 Hz, 1 H), 4.17-4.03 (m, 3.4 H), 3.92-3.87 (ddd, *J* = 8.75 Hz, 1.9 H), 2.16 (s, 3 H), 2.13 (s, 3 H), 2.08 (s, 3 H), 2.06 (s, 3 H), 2.05 (s, 3 H), 2.03 (s, 3 H), 1.97 (s, 3 H).

4. General procedure for Fmoc solid phase peptide synthesis

Fmoc-amino acids Gly, Ala, Val, Leu, Ser(tBu), Thr(tBu), Asp(OtBu), Glu(OtBu), Asn(Trt), Gln(Trt), Lys(Boc), Arg(Pbf), Phe, Tyr(OtBu), His(Trt) were used. Amino-PEGA resin was used to obtain peptide-carboxylic acids. At first, the resin was swollen for minimum of 2 h in DMF, followed by coupling of linker. For 50 μmol resin, coupling of HMPA linker (22.7 mg, 0.125 mmol) was performed twice using HBTU (47.4 mg, 0.125 mmol) and *N*-ethyl morpholine (15.8 μL, 0.125 mmol) in DMF (1.0 mL), activated for 1 min at ambient temperature and added to resin and gently stirred at ambient temperature for 1 h each. Coupling of first Fmoc-amino acid (Fmoc-AA-OH, 0.25 mmol) was performed in the presence of MSNT (74.1 mg, 0.25 mmol) and *N*-methylimidazole (15.3 μL, 0.19 mmol) in DCM (1.0 mL) which were activated for 1 min at ambient temperature and added to resin and gently stirred at ambient temperature for 1 h twice. The removal of Fmoc-group was performed using 1mL of 20% piperidine in DMF for 15min ambient temperature. Coupling from second amino acid onwards was performed using Fmoc-AA-OH (0.2 mmol), 0.45 M HCTU/HOBt cocktail in DMF (0.37 mL, 0.166 mmol each), 0.9 M DIPEA in DMF (0.37 mL, 0.33 mmol) and DMF (0.26 mL) for 30 min at ambient temperature. Removal of Fmoc group was performed using 20% piperidine in DMF for 15 min at ambient temperature. After the assembly of all amino acids in the similar manner, the resin was washed thoroughly with DCM and vacuum dried which was treated with 95% TFA:2.5% TIPS:2.5% H_2O (10 mL) cocktail at ambient temperature for less than 1 h or up to 3 h depending on the side-chain protecting groups. With acidic cocktail treatment deprotection of amino acid side chains and cleavage from resin takes place simultaneously. The solution was removed from the resin by filtration and concentrated *in vacuo*. The concentrated peptide mixture was added to ice-cold Et_2O in order to precipitate the peptide. The precipitate was washed two times with Et_2O and then purified by RP-HPLC.

5. Synthesis of Model peptide 7

Synthesis of model peptide 7 was performed by Fmoc-SPPS using amino PEGA resin in 50 μmol scale following the protocol described in general procedure for Fmoc-SPPS. Deprotection and cleavage was performed for 2.5 h at ambient temperature followed by precipitation in ice-cold Et_2O. The precipitate was purified by preparative HPLC using CAPCELL PAK C18 (Φ 10 x 250 mm, 0.1% TFA: 0.1% TFA in 90% AcCN = 85:15 to 55:45 over 60min with the flow rate 2.5 mL/min) and freeze-dried to obtain 7 as a white powder. ESI-MS: *m/z* calculated for $C_{56}H_{95}N_{15}O_{15}$: $[M+H]^+$ 1218.72, observed: 1218.70.

6. Optimization of reaction conditions for protection of peptide.

For all the reactions, model peptide **7** (1.5 mg, 1.2 umol) was taken, equivalents of AGlOC-imidazole **4** and DIEA were taken with respect to peptide, while NHS or HOBt was taken with respect to **4**. To a reaction mixture of AGlOC-imidazole and NHS dissolved in 0.5 mL of solvent, DIEA was added and stirred at ambient temperature for 2-3minutes. Later **7** dissolved in 0.25 mL of solvent was added to the reaction mixture and stirred at 30°C/40°C. Reaction was quenched using 1% of TFA or 1% MeOH in AcCN and reaction progress was monitored by reverse phase HPLC using Cadenza CD-C18 Φ4.6 × 10 mm, 0.1% TFA: 0.1% TFA in 90% AcCN (80:20 to 20:80 over 15min with a flowrate of 1 mL/min). After the reaction, the product **8** was purified by reverse phase HPLC using Proteonavi (Φ 10 x 250 mm) **HRMS(ESI):** *m/z* calculated for $C_{101}H_{149}N_{15}O_{48}$: $[M+H]^+$ 2340.9751, observed: 2340.9752.

7. Transformation of *E.coli* and Recombinant expression of fusion peptide **9**[3]

Sequence of fusion peptide **9**:

MHHHHHHVMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL
MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGHHHHHH
AENLYFQGKTKEGVLYVGSKTKEG

The plasmid containing desired genes in pET vector was transformed into *E.coli* BL21(DE3) strain by heat-shock treatment. 0.5 uL of plasmid (0.5 ug/uL) was added to 10 ul of *E.coli* cells and treated at 42°C for 45 sec and immediately transferred to ice bath and left for 5 min. Competent cells were obtained by plating the

transformed cells on LB agar plate containing ampicillin at 37°C overnight. One of the colony form the plate is used to make glycerol stock that can be used for preparation of starter culture for protein expression. Starter culture is obtained by inoculating 5 mL of LB broth containing ampicillin at 37°C overnight. This is then used to inoculate 1 L of LB media containing ampicillin in the final concentration of 100 mg/mL and incubated at 37°C until the optical density (OD) at 600 nm reached 0.75 OD. The culture media was then cooled to ambient temperature and induced with isopropylthiogalactoside (IPTG) at the final concentration of 100 mg/mL was added and incubated at 25°C for 24 h. *E.coli* cells were collected by centrifugation at 8000 rpm for 10 min at 4°C. The precipitated cells were gently dissolved in appropriate amount of lysis buffer on ice bath, with the composition 50 mM sodium phosphate, pH 7.8, 400 mM NaCl, 100 mM KCl, 10 % glycerol, 0.1 % triton X-100, 10 mM imidazole. The composition of lysis buffer was taken from the manual provided by Invitrogen. The cells were lysed by ultrasonication for 1 min with pauses on ice bath 4 times. This was subjected to centrifugation at 10000 rpm for 10 min at 4°C. Fusion peptide 9 being a soluble peptide remains in the buffer precipitating the cell debris which was discarded. Purification of Fusion peptide 9 was performed using cOmplete His-tag resin under native conditions using varying compositions of Buffer A (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl) and Buffer B (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 250 mM imidazole). The eluents containing **9** was confirmed by reverse phase-UHPLC and were subjected to dialysis against water for 15-20 h at 4°C and then freeze-dried to obtain **9** as a white powder (25 mg). This could be directly used for protease cleavage without additional purification step, but in my case I purified by using reverse phase HPLC using Proteonavi (Φ 10 x 250 mm, 0.1% TFA: 0.1% TFA in 90% MeCN = 70:30 to 30:70 for 60 min with 2.5 mL/min flow rate). ESI-MS: *m/z* calculated for $C_{689}H_{1075}N_{205}O_{211}S_5$: $[M+17H]^{17+}$ 928.22; $[M+18H]^{18+}$ 877.65, observed: $[M+17H]^{17+}$ 928.57; $[M+18H]^{18+}$ 877.03.

8. SUMO protease cleavage of fusion peptide **9**

Buffers provided by the SUMO protease kit from Thermo Fisher company were used for cleavage following their manual. Reaction mixture contained 200 uL of 10X SUMO protease buffer + salt (500 mM Tris-HCl, pH 8.0, 2% Igepal (NP-40), 1.5 M NaCl, 10 mM DTT), 400 ug of **9** in 1.65 mL of milliQ water to which 150 ul of SUMO protease (150 units) were added and incubated at 30°C for 10 h. Reaction was quenched using 6M Gn.HCl buffer and reaction progress was monitored by reverse phase HPLC using Proteonavi (Φ 4.6 x 250 mm, 0.1% TFA: 0.1% TFA in 90% AcCN = 85:15 to 15:85 for 20 min with 1 mL/min flow rate) and purified by reverse phase HPLC Proteonavi (Φ 4.6 x 250 mm, 0.1% TFA: 0.1% TFA in 90% MeCN = 85:15 to 15:85 for 20 min with 1 mL/min flow rate) to afford peptide **10** (**<** 50 ug). ESI-MS: *m/z* calculated for $C_{155}H_{230}N_{48}O_{44}$: $[M+H]^+$ 3469.73, observed: $[M+H]^+$ 3470.77.

9. Chemical synthesis of **10** by Fmoc SPPS

Synthesis of model peptide **10** was performed by Fmoc-SPPS using amino PEGA resin in 50 μmol scale, following the protocol described in general procedure for Fmoc-SPPS. Deprotection and cleavage was performed for 2.5 h at ambient temperature followed by precipitation in ice-cold Et_2O. The precipitate was purified by preparative CAPCELL PAK C18 (Φ 10 x 250 mm, 0.1% TFA: 0.1% TFA in 90% AcCN = 85:15 to 45:55 over 60min with the flow rate 2.5 mL/min) and freeze-dried to obtain **10** as a white powder. ESI-MS: *m/z* calculated for $C_{155}H_{230}N_{48}O_{44}$: $[M+H]^+$ 3469.73, observed: $[M+H]^+$ 3470.77.

10. Partial protection of **10**

To a reaction mixture of AGlOC-imidazole (114.8 mg, 0.26 mmol) and NHS (104.6 mg, 0.91 mmol) dissolved in 4 mL of AcCN, DIEA (5 ul, 28.7 umol) was added and stirred at ambient temperature for 2-3minutes. To this **10** (20 mg, 5.76 umol) dissolved in 2.6 mL of H_2O was added and stirred at 40°C for 5 h. Reaction was quenched using 1% of TFA and reaction progress was monitored by reverse phase HPLC using Proteonavi (Φ 4.6 x 250 mm, 0.1% TFA: 0.1% TFA in 90% MeCN = 75:25 to 15:85 for 30 min with 1 mL/min flow rate). After the reaction, the product **11** was purified by reverse phase HPLC using Proteonavi (Φ 10 mm × 250 mm, 0.1% TFA: 90% AcCN containing 0.1% TFA = 55:45 to 35:65 over 60 min, with the flow rate 2.5 mL/min) which afforded **11** as a white powder (7.1 mg, 1.4 umol, 23.2 %). **HRMS(ESI):** *m/z* calculated for $C_{230}H_{320}N_{48}O_{99}$: $[M+H]^+$ 5339.1499, observed: $[M+H]^+$ 5339.1554.

11. Deacetylation of AGlOC-protected peptide 11[4]

7.1 mg of **11** was treated with 2 mL of 10 % $NH_2NH_2.H_2O$ in DMF for 5 min and later precipitated and washed with Et_2O. Though the peptide doesn't dissolve completely but the reaction proceeds smoothly to give **12**, which was purified by reverse phase HPLC using CAPCELL PAK C18 (Φ 10 x 250 mm, 0.1% TFA: 0.1% TFA in 90% AcCN = 100:0 for 7 min 80:20 to 35:65 over 90min with the flow rate 2.5 mL/min) which afforded **12** as a white powder (3.1 mg, 0.68 umol, 52.5 %). **HRMS(ESI):** *m/z* calculated for $C_{190}H_{280}N_{48}O_{79}$: $[M+4H]^{4+}$ 1125.9858, observed: $[M+4H]^{4+}$ 1125.9922.

12. <u>One-pot protection and deacetylation of peptide **10** using **4**</u>

To a reaction mixture of AGlOC-imidazole (6.7 mg, 15.1 umol) and NHS (5.2 mg, 45.3 umol) dissolved in 210 uL of DMF, DIEA (0.5 ul, 2.87 umol) was added and stirred at ambient temperature for 2-3 min. To this **10** (1.5 mg, 0.43 umol) dissolved in 90 uL of H_2O was added and stirred at 40°C for 5 h. Reaction progress was monitored by reverse phase HPLC using Proteonavi (Φ 4.6 mm × 250 mm, 0.1% TFA: 90% AcCN containing 0.1% TFA = 90:10 to 10:90 over 30 min, with the flow rate 1 mL/min). After 5 h, 15 % $NH_2NH_2.H_2O$ with respect to total volume of reaction mixture was added and stirred at ambient temperature for 40 min and later, the reaction mixture was diluted and the product **12** was purified by reverse phase HPLC using CAPCELL PAK C18 (Φ 10 x 250 mm, 0.1% TFA: 0.1% TFA in 90% AcCN = 100:0 for 7 min 80:20 to 35:65 over 90min with the flow rate 2.5 mL/min) which afforded **12** as a white powder (0.7 mg, 0.15 umol, 36 %). **ESI-MS:** *m/z* calculated for $C_{190}H_{280}N_{48}O_{79}$: $[M+H]^+$ 4500.94, observed: 4501.22.

13. <u>One-pot protection and deacetylation of peptide **10** using **6**</u>

To a reaction mixture of ALaOC-imidazole (10.9 mg, 15.1 umol) and NHS (5.2 mg, 45.3 umol) dissolved in 210 uL of DMF, DIEA (0.5 ul, 2.87 umol) was added and stirred at ambient temperature for 2-3 min. To this **10** (1.5 mg, 0.43 umol) dissolved in 90 uL of H_2O was added and stirred at 40°C for 5 h. Reaction progress was monitored by reverse phase HPLC using Proteonavi (Φ 4.6 mm × 250 mm, 0.1% TFA: 90% AcCN containing 0.1% TFA = 90:10 to 10:90 over 30 min, with the flow rate 1 mL/min). After 5 h, 15 % $NH_2NH_2.H_2O$ with respect to total volume of reaction mixture was added and stirred at ambient temperature for 40 min and later, the reaction mixture was diluted and the product **14** was purified by reverse phase HPLC using CAPCELL PAK C18 (Φ 10 x 250 mm, 0.1% TFA: 0.1% TFA in 90% AcCN = 100:0 for 7 min 80:20 to 35:65 over 90min with the flow rate 2.5 mL/min) which afforded **14** as a white powder (0.4 mg, 0.075 umol, 17 %). **ESI-MS:** *m/z* calculated for $C_{220}H_{330}N_{48}O_{104}$: $[M+3H]^{3+}$ 1771.43, observed: 1771.49.

14. <u>Removal of protecting groups.</u>

i. For optimizing removal of protecting groups of **12** by glucosidases, varying amounts of **12** in 100 mM sodium phosphate buffer, pH 6.73, α-glucosidase, β-glucosidase, protease inhibitor cocktail

were incubated at 37°C from 12 h to 48 h. Reaction was quenched with 6N Gn.HCl buffer and monitored by reverse phase UHPLC using Cadenza CD-C18 (Φ 2 × 100 mm, 0.1% HCOOH : 0.1% HCOOH in 90% AcCN = 90:10 to 30:70 for 10 mins by 0.2 mL/min) however, an efficient reaction condition was not found.

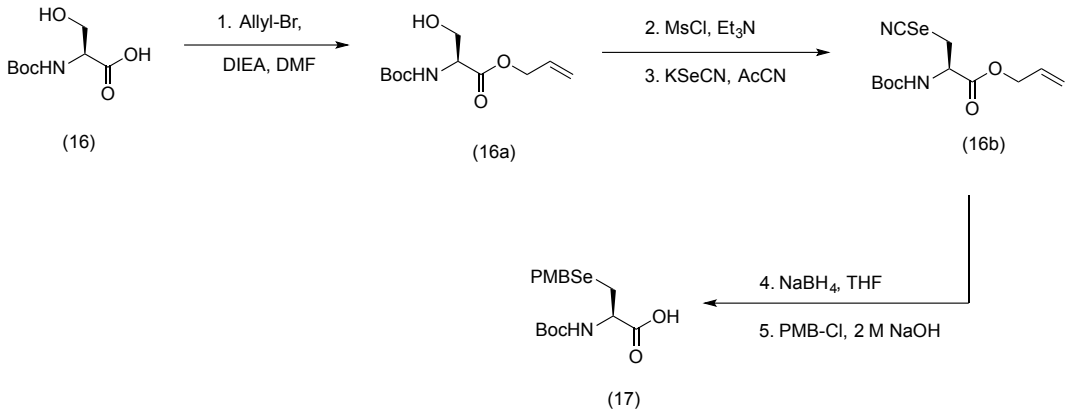
ii. For optimizing removal of protecting groups of **12** under acidic conditions, 200 ug of **12** was treated with various acidic cocktails, at the respective temperature and time as mentioned in table 2-2, the reactions were quenched by precipitating in ice-cold Et_2O. Under the optimized condition, though the reaction mixture remains turbid, reaction proceeds to completion. The product **10** was purified by reverse phase HPLC using CAPCELL PAK C18 (Φ 10 x 250 mm, 0.1% TFA: 0.1% TFA in 90% AcCN = 85:15 to 45:55 over 60min with the flow rate 2.5 mL/min) which afforded **10** as a white powder. **HRMS(ESI)**: *m/z* calculated for $C_{155}H_{230}N_{48}O_{44}$: $[M+H]^+$ 3468.7308, observed: $[M+H]^+$ 3468.7308.

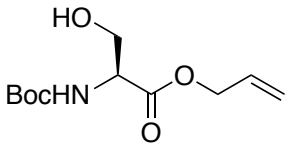
15. TEV protease cleavage of **12**

To the partially protected peptide **12** (5 mg, 0.44 umol), dissolved in 1 mL of 25 mM Tris-HCl buffer at pH 8.0, 500 units of TurboTEV protease (protease/peptide = 1:100 w/w) was added and incubated at 37°C for 10 h. Reaction was quenched with 6N Gn.HCl buffer and monitored by reverse phase HPLC using Cadenza CD-C18 (Φ 4.6 × 10 mm, 0.1% TFA: 0.1% TFA in 90% MeCN = 90:10 to 50: 50 over 15 min with a flow rate of 1ml/min) and purified by using Proteonavi (Φ 10 mm × 250 mm, 0.1% TFA: 90% AcCN containing 0.1% TFA = 90:10 to 50:50 over 60 min, with the flow rate 2.5 mL/min) using which afforded **13** as a white powder (0.86 mg, 0.33 umol, 30 %). **HRMS(ESI):** *m/z* calculated for $C_{106}H_{173}N_{21}O_{54}$: $[M+2H]^{2+}$ 1301.07911, observed: $[M+2H]^{2+}$ 1301.07673.

16. Modification of N-terminus with Selenocysteine

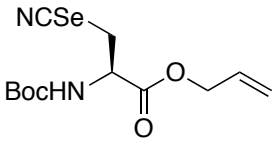
- Synthesis of selenocysteine[5]





16a

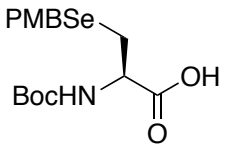
(*S*)-allyl 2-((*tert*-butoxycarbonyl)amino)-3-hydroxypropanoate (16a): Boc-Ser-OH (0.75 g, 3.65 mmol) was dissolved in DMF (21 mL) and cooled to 0°C. To this was added DIPEA (1.27 mL, 9.84 mmol) and Allyl-Br (0.63 mL, 5.2mmol). The reaction was warmed to ambient temperature and stirred for 8 h and then diluted with EtOAc which was washed with H_2O and saturated NaCl. The organic layer was dried over Na_2SO_4 and concentrated *in vacuo*. The pale yellow (0.82 g, 3.34 mmol, 92 %) obtained was directly used for next step without purification.



16b

(*R*)-allyl 2-((*tert*-butoxycarbonyl)amino)-3-selenocyanatopropanoate (16b): 16a (0.82 g, 3.34 mmol) was dissolved in CH_2Cl_2 (15 mL), cooled to 0°C and Et_3N (0.7 mL, 6.8 mmol) was added dropwise followed by MsCl (0.31 mL, 2.7 mmol). The reaction was stirred at 0°C for 1h, then diluted with CH_2Cl_2

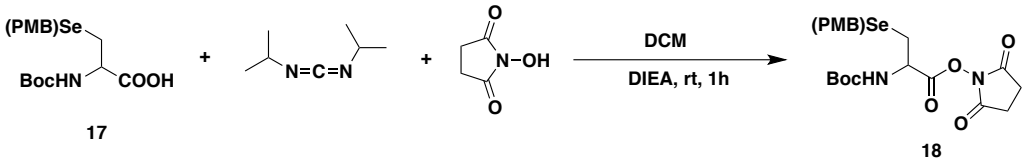
and washed with saturated NH_4Cl, followed by H_2O. The organic layer was dried over Na_2SO_4 and concentrated in vacuo to afford crude mesylate as a pale-yellow oil (1 g). Crude mesylate (1 g) was saturated with Ar and dissolved in CH_3CN (10 mL) and KSeCN (0.84 g, 5.8 mmol) was added in a single portion and stirred at 60°C for 48 h under Ar. The reaction mixture was diluted with CH_2Cl_2 which was washed with H_2O. The organic layer was dried over Na_2SO_4 and concentrated *in vacuo*. The crude residue was purified by flash column chromatography using silica gel (1:9 to 1:3 EtOAc/hexane) to afford **16b** as a yellow oil (0.34 g, 1 mmol). The NMR spectral properties were found to be in agreement with literature data.[5]



17

(*R*)-2-((*tert*-butoxycarbonyl)amino)-3-((4-methoxybenzyl)selanyl)propionic acid (17): Selenocyanate **16b** (0.34 g, 1 mmol) was dissolved in degassed dry THF (6 mL) at 0°C and $NaBH_4$ (79 mg, 2.1 mmol) was added followed by slow addition of degassed MeOH (1 mL) which leads to effervescence. After 15 min, PMB-Cl (0.55 mL, 4.16 mmol) was added and once the starting material is consumed which is confirmed by TLC, degassed 2 M NaOH (2.6 mL) was added and stirred at 0°C for 2.5 h. The pH of the reaction mixture is adjusted to 2.0 using 5 N HCl and extracted with EtOAc. The organic layer was dried over Na_2SO_4 and concentrated *in vacuo* to yield a clear oil which was purified via flash column chromatography using silica gel (2:8 to 8:2 EtOAc/hexane) to afford a pale yellow solid (0.12 g, 0.29 mmol, 29 %). The NMR spectral properties were found to be in agreement with literature data.[5]

- Synthesis of Boc-Sec(PMB)-OSu



To a mixture of Boc-Sec(PMB)-OH (5 mg, 12.8 umol), NHS (2.9 mg, 25.6 umol) in 100 uL of DCM, was added DIEA (1.1 ul, 6.4 umol) followed by DIC (3 ul, 19.2 umol) and stirred at ambient temperature for 1 h. Reaction was monitored by TLC. After 1 h, reaction was diluted with DCM and work-up was done with H_2O. The DCM organic layer was evaporated *in vacuo* to afford **18**. The crude **18** was directly used for coupling reaction. NMR analysis was not performed as it was a crude mixture. **ESI-MS:** *m/z* calculated for $C_{20}H_{26}N_2O_7Se$ $[M+Na]^+$ 509.07, observed : 509.31

- Coupling of Boc-Sec(PMB)-OSu to **15**

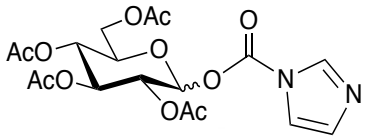
To a reaction mixture of Boc-Sec(PMB)-OSu (3.4 mg, 6.9 umol) and **15** (3.6 mg, 1.4 umol) in 250 ul of 70 % DMF in H_2O, DIEA (5 ul, 32 umol) was added and stirred at ambient temperature for 45 min. Reaction progress was monitored by reverse phase UHPLC using Cadenza CD-C18 (Φ4.6 × 10 mm, 0.1% HCOOH: 0.1% HCOOH in 90% MeCN = 90:30 to 20:80 for 20 mins by 0.2 mL/min). After the reaction, the product **11** was purified by reverse phase HPLC using Proteonavi (Φ 10 mm × 250 mm, 0.1% TFA: 90% AcCN containing 0.1% TFA = 85:15 to 35:65 over 40 min, with the flow rate 2.5 mL/min) which afforded **19** as a white powder (0.7 mg, 0.23 umol, 17 %). **HRMS(ESI):** *m/z* calculated for $C_{122}H_{194}N_{22}O_{58}Se$: $[M+H]^+$ 2973.2174, observed: $[M+H]^+$ 2973.2205.

17. Removal of protecting groups from **19.**[6]

Peptide **19** (0.7 mg, 0.23 umol) was treated with 100 ul of high acidic cocktail thioanisole/*m*-cresol/TFA/TfOH (2:1:20:2) at 0°C for 2 h in order for complete removal of all protecting groups including PMB. The peptide was precipitated using ice cold diethyl ether and washed several times with diethyl ether and purified by reverse phase HPLC using Proteonavi (Φ 4.6 mm × 250 mm, 0.1% TFA: 90% AcCN containing 0.1% TFA = 90:10 to 30:70 over 30 min, with the flow rate 1 mL/min) to afford **20** (0.12 mg, 12 %). **ESI-MS:** *m/z* calculated for $C_{162}H_{274}N_{44}O_{54}Se_2$: $[M+H]^+$ 3861.84, observed: $[M+H]^+$ 3861.90.

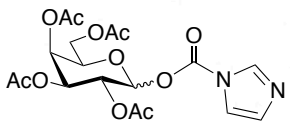
18. NMR Data

^{1}H NMR of **4**



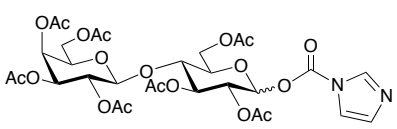
4

^{1}H NMR of **5**



5

^{1}H NMR of **6**


OAc
OAc
AcO
O
OAc
O
AcO
OAc
OAc
O
O
N
N

6

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

References

1. Ichikawa, Y.; Sim, M. M.; Wong, C.-H.; *J. Org. Chem.* **1992**, *57*, 2943.
2. Knoben, H.-P.; Schlüter, U.; Redlich, H.; *Carbohydr. Res.* **2004**, *339*, 2821.
3. Nallamsetty, S.; Waugh, D. S.; *Nat. Protoc.* **2007**, *2*, 383.
4. Bardají, E.; Torres, J. L.; Clapés, P.; Albericio, F.; Barany, G.; Rodríguez, R. E.; Sacristán, M. P.; Valencia, G.; *J. Chem. Soc., Perkin Trans.* **1991**, *1*, 1755.
5. Malins, L. R.; Mitchell, N. J.; McGowan, S.; Payne, R. J.; *Angew. Chem. Int. Ed.* **2015**, *54*, 1271.
6. Koide, T.; Itoh, H.; Otaka, A.; Yasui, H.; Kuroda, M.; Esaki, N.; Soda, K.; Fujii, N.; *Chem. Pharm. Bull.* **1993**, *41*, 502.

Acknowledgements

It's my great pleasure to acknowledge the support and encouragement that I received in accomplishing my PhD thesis. I am very thankful to God, for, because of Him, I'm able to be here today.

I would like to express my sincere gratitude towards Prof. Yasuhiro Kajihara for accepting me as his student and guiding me in learning science and for his constant support from day one.

I would like to thank Dr. Masayuki Izumi for his suggestions and Dr. Ryo Okamoto for always being approachable and discussions with him made me learn a lot about the subject and Dr. Yuta Maki for his support.

I would like to thank and appreciate the support of my committee members Prof. Koichi Fukase and Prof. Toshifumi Takao for their helpful suggestions and advices on my PhD work.

I thank all my lab members who worked with me in these three years for their kind support rendered whenever I was in need.

I thank JICA for providing me this wonderful opportunity and scholarship, and supporting my life in Japan.

I thank Nguyen Minh Hien, Fatma Yesil, Jyotheeswara Reddy and his family and all my Indian friends in Japan for all the get-togethers and being my family here.

Finally, I would like to express my deepest acknowledgement and gratitude towards the most important people of my life without whom I'm not what I am today. I thank my parents and my brother for their love, constant support and for being my backbone in all walks of my life. I thank Arjun Menon and his family for their love, encouragement, and for teaching me a lot about life and in helping me handle difficult situations without which surviving in Japan would have been an impossible task. I thank Namratha Nagesh and Harshitha S.C for being with me always and building up my confidence.