Study on Conformational Characteristics of Human NOTCH1 EGF Domains Induced by Dynamic *O*-glycosylation

(動的な糖鎖修飾が誘起するヒト NOTCH1 レセプター

EGF 様ドメインの立体構造特性に関する研究)

December 2020

Doctoral Thesis

Yasuhiro Yokoi

Laboratory of Advanced Chemical Biology

Graduate School of Life Science, Hokkaido University

Table of Contents

Abbreviations

ACN	acetonitrile
Ac	acetyl
Bu	butyl
COSY	correlation spectoroscopy
DCM	dichloromethane
DHB	dihydroxbenzoic acid
DMF	*N*,*N*-dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
DIC	*N*,*N*′-Diisopropylcarbodiimide
EDT	1,2-ethanedithiol
ESI	electro spray ionization
Et	ethyl
Fmoc	9-fluorenylmethyloxycarbonyl
Fuc	fucose
Gal	galactose
GalNAc	*N*-acetyl galactosamine
Glc	glucose
GlcNAc	*N*-acetyl glucosamine
GnF	GlcNAcβ1→3Fucα1→
HBTU	2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate

HOBt	1-hydroxybenzotriazole
HPLC	high performance liquid chromatography
HSQC	hetero-nuclear multiple quantum coherence
MALDI	matrix assisted laser desorption ionization
MS	mass
Me	methyl
Neu5Ac	*N*-acetyl neuraminic acid
NMR	nuclear magnetic resonance
NOE	nuclear overhauser effect
NOESY	NOE correlated spectroscopy
Sia	sialic acid
SPPS	solid phase peptide synthesis
TFA	trifluoroacetic acid
TIS	triisopropylsilane
TOCSY	totally correlated spectroscopy
TOF	time of flight
XXG	Xylα1→3Xylα1→3Glcβ1→
Xyl	xylose
m/z	mass/charge ratio

Chapter 1

General Introduction

1-1. Protein Glycosylation

Post-translational modification is a chemical modification of protein, which occur DNA-independent manner. Protein glycosylation is known as one of the most important post-translational modifications because over 50% of proteins are found to be glycosylated.[1] These glycosylations are performed by specific glycosyltransferase and sugar nucleotides as glycosyl donor in ER, Golgi apparatus or cytoplasm. These DNA-independent carbohydrate modifications produce protein structural diversity in terms of heterogeneous glycosylation sites as well as glycan structure, which generates protein functional diversity (Figure 1-1).[2]

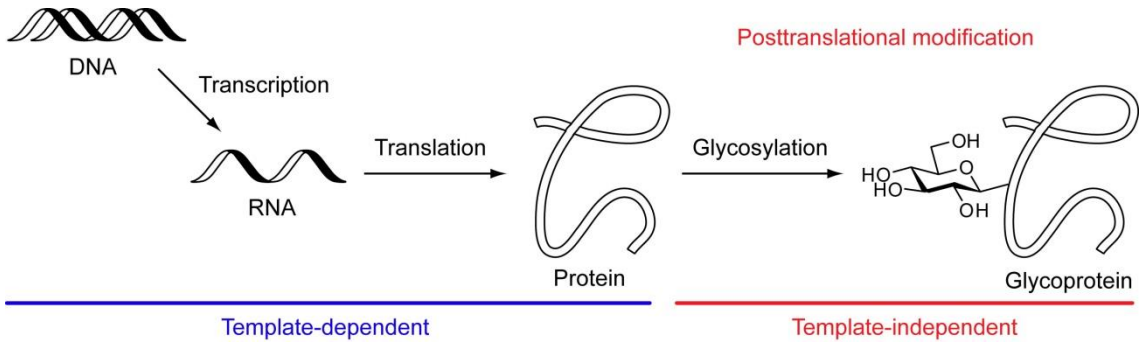


Figure 1-1. Protein expression and glycosylation

There are two major types of glycan structure, *N*-glycan and *O*-glycan. *N*-glycans are covalently attached to nitrogen atom on asparagines residue to form GlcNAcβ1→Asn, where Asn locates in Asn-X-Thr consensus sequence (X: any amino acid residue other than Pro).[3] *N*-glycans have pentasaccharide core structure, and it is majorly classified into complex, hybrid and high mannose type according to their branching and terminal structure. On the other hand, *O*-glycan is linked to Ser or Thr residue via oxygen atom, and various sugar residues are found to be attached. GalNAcα1→Thr/Ser modification is

one of the most abundant *O*-glycosylations, which is mainly found in mucin protein.[4] In addition, GlcNAcβ1→, Fucβ1→, Manα1→, Glcβ1→ and Xylβ1→ modifications at Thr/Ser have been found in several proteins (Figure 1-2).[5-8] These *O*-glycans are further elongated by specific sugar residues, producing heterogeneous glycan structure. For example, GalNAcα1→Thr/Ser can be further glycosylated to form Galβ1→3GalNAcα1→, which is called as core1 *O*-glycan.[9] In this context, *O*-glycans would have more complex structure than *N*-glycan in terms of structural diversity.

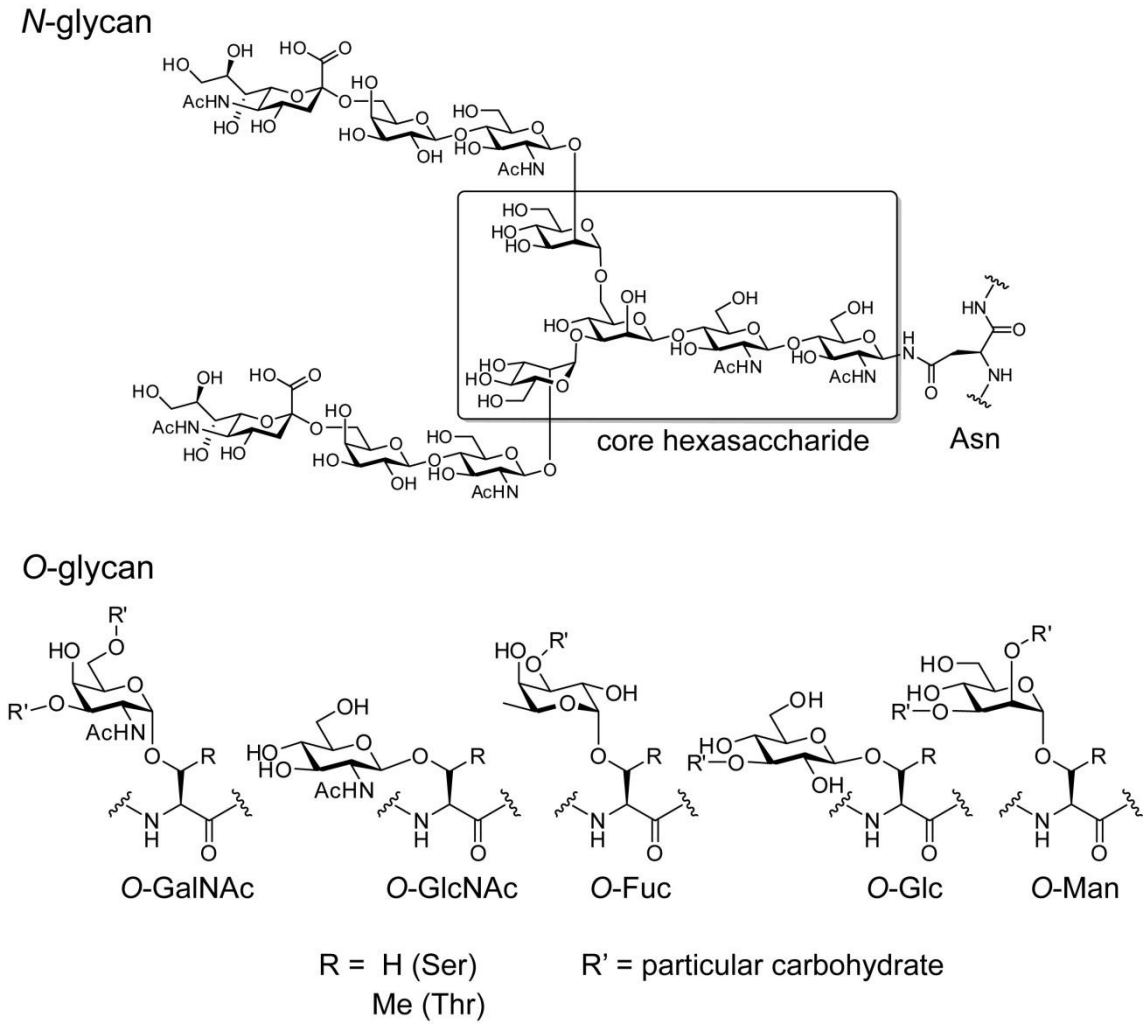


Figure 1-2. Chemical structure of typical *N*- and *O*-glycans

These *N*- and *O*-glycans relate to various biological phenomena involving cell-cell/protein-protein interaction, infection, immunity, quality control of protein expression, enhancement of protein stability, conformational alteration of protein backbone, and so on.[10] Aberrant glycosylation on protein as well as aberrant expression of glycoproteins relate to several diseases, such as cancer, leukemia, autoimmune disease etc.[11-13] Especially, hyper-branched *N*-glycans and truncated *O*-glycans have been found to be expressed in several cancer cells. These tumor-associated carbohydrate antigens have been applied to bio-marker of diagnosis and therapeutic target molecule. Therefore, glycoproteins have great potential to therapeutic applications to several diseases.[14]

The protein biological functions can be modulated by glycan structure, but "how glycosylation affect protein functions?" Investigation of molecular mechanism of glycosylation effects on protein functions could be important for application of glycopeptide towards diagnosis or therapy for cancer as well as several diseases.

Anti-freeze glycoprotein (AFGP) acquires anti-freeze property after modification of disaccharide, Galβ1→3GalNAcα1→, at threonine residue on repeating triplet (Thr-Ala-Ala). In prior research, chemical synthesis and NMR conformational analysis of AFGP demonstrated that synthetic *O*-glycosylated AFGP adopts polyproline type II helix structure in which disaccharide moieties are on the same side of the molecule.[15,16] This glycosylation-induced amphiphilic structure is presumed to be crucial for the AFGP specific activity (Figure 1-3A). In addition, NMR study of MUC1 glycopeptide revealed that glycan structure affected not only antibody recognition but also peptide backbone conformation.[17] Pro-Asp-Thr-Arg-Pro-Ala-Pro region of MUC1 tandem repeat carrying

core 1 glycan (Galβ1→3GalNAcα1→) at Thr residue has found to be minimal epitope against anti-KL6 monoclonal antibody which is applied for diagnosis of interstitial pneumonia, lung adenocarcinoma, breast cancer, colorectal adenocarcinoma, and hepatocellular carcinoma etc. Chemical synthesis and NMR conformational analysis of MUC1 glycopeptide derivatives revealed that disialylated core 2 [Neu5Acα2→3Galβ1→3(Neu5Acα2→3Galβ1→4GlcNAcβ1→6)GalNAcα1→] give conformational flexibility between two conformers of γ-turn and β-turn on PDTR region while naked and core 2-modified PDTR regions adopt γ-turn and β-turn, respectively. Thus, it is suggested that anti-KL6 mAb recognize carbohydrate structure and conformational alteration of peptide backbone induced by *O*-glycosylation at PDTR region (Figure 1-3B).

These results clearly indicate that regulation of protein function as well as antigenicity is demonstrated by glycosylation-induced conformational alteration. Therefore, investigation of glycosylation impacts on peptide backbone would be important for understanding of molecular mechanism of glycobiology. Furthermore, investigation of conformational motif/module which represents specific glycan-induced 3D-structure can become good probe to develop therapeutic as well as diagnostic methods.

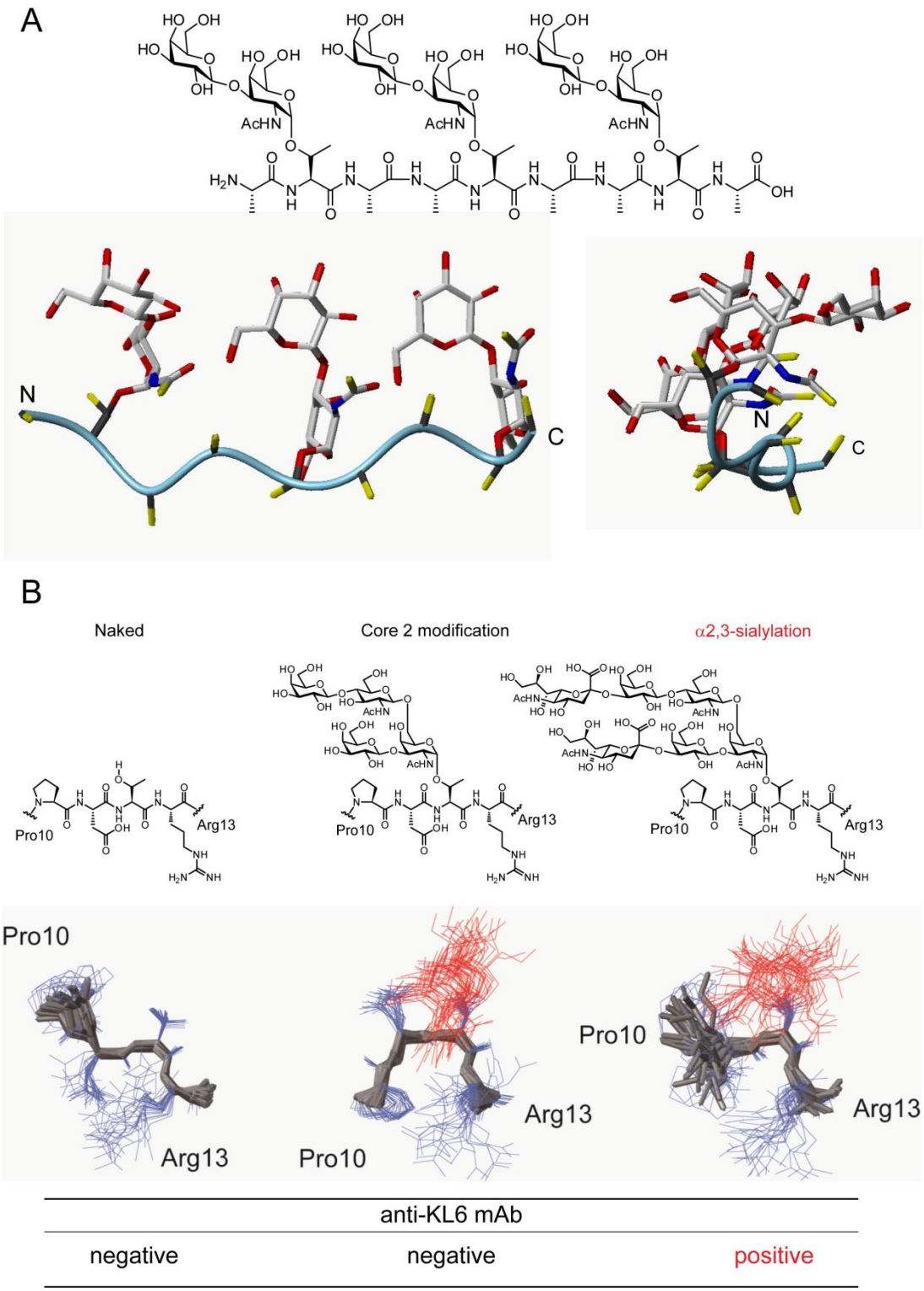


Figure 1-3. Glycan-induced conformational alteration. (A) synthetic AFGP adopts poly proline type-II helix. (B) PDTR sequence of MUC1 showed γ- and β-turn during glycosylation.

1-2. Organic Chemistry for Glycobiology

Organic chemistry of glycopeptide has been developed for powerful tool to understand the molecular mechanism of *O*-glycosylation impacts on conformational alteration as well as functional alteration because homogeneous glycopeptide is required to investigate functions of individual *O*-glycans.[18-22] Unlike the protein expressed with DNA-dependent manner, the difficulties in the production of structurally-well-defined glycopeptide often limit the proceeding of glycobiology. Since protein glycosylation is DNA-independent modification, biological approach will give heterogeneous glycopeptides in terms of glycan structure as well as glycosylation site.

To overcome this barrier, chemical approach to efficiently afford structurally-defined glycopeptide should be necessary.[15-17,23,24] In addition, this approach can modify some specific functional group on glycopeptide such as ketone-modification required for microarray immobilization which facilitate functional study of glycopeptides. In present thesis, we demonstrated organic chemical approach of glycobiology focusing on (1) establishment of efficient synthetic methodology to obtain structurally-well-defined glycopeptides and (2) NMR-based conformational analysis of synthetic glycopeptides to find conformational motif/modules induced by glycosylation on peptide backbone. These strategies will provide important relationship between glycan-induced conformational alteration and biological functions of glycopeptides, which can be applicable of glycopeptide to therapeutic target of glycosylation-associated disease.

In this thesis, unique *O*-glycans such as *O*-GlcNAcylation, *O*-fucosylation and *O*-glucosylation in human NOTCH1 receptor extracellular EGF-like domain are focused as

target molecule to investigate conformational and functional impacts.

1-3. Notch Signaling Pathway

The Notch signaling pathway is evolutionally conserved signal transduction system, which regulates cell-fate-determination processes in all metazoan species.[25] In adult mammals, Notch signaling directs neural and hematopoietic stem cell maintenance and differentiation, and development and regulation of many immune systems.[26-28] Notch pathway dysfunction is associated with several diseases, particularly cancer biology.[29,30] Notch signaling pathway is normally regulated by the receptor-ligand interactions, which is mediated by five ligands [Jagged1, Jagged2, Delta-like ligand 1 (DLL-1), DLL-3, and DLL-4] and the four Notch receptors (NOTCH1~4). The human NOTCH1 receptor has 36 tandemly connected EGF-like repeats in their extracellular domain (ECD), in which domains 10~12 are crucial for ligand binding. Importantly, the Notch signaling is triggered by ligand-receptor binding which induces proteolysis by ADAM10/17 metalloproteases and β-secretase to cleave stoichiometrically Notch intercellular domain (ICD). Cleaved Notch ICD moves into nucleus and regulates expression of Notch target gene.[31,32]

Many of EGF-like domains in the Notch extracellular domain can be modified by one or more of three-types of *O*-glycosylation, all of which have been shown to affect Notch function.[33-35] Six cysteine residues of the EGF-like domains are supposed to form three disulfide bonds (C1-C3, C2-C4, and C5-C6), namely the canonical EGF fold, and modified dynamically by *O*-glucosylation,[36] *O*-fucosylation,[37] and *O*-*N*-acetylglucosaminylation (*O*-GlcNAcylation)[38] in the consensus sequence between two cysteine residues, C1-X-S-X-(P/A)-C2,[39] C2-X-X-X-X-(T/S)-C3,[40] and C5-X-X-G-X-(T/S)-G-X-X-C6,[41] respectively. Accumulated results have provided evidence that

glycosylation of Notch ECDs in cultured cells can be initiated by catalytic reactions of glycosyltransferases with ECDs only after forming the canonical EGF folds.[42,43] In mammalian Notch receptors, α-*O*-fucose, β-*O*-glucose, and β-*O*-GlcNAc attached to Thr/Ser residues can be further extended with specific sugars, forming unusual glycoforms such as Neu5Acα2→3/6Galβ1→4GlcNAcβ1→3Fucα1→,[44] Xylα1→3Xylα1→3Glcβ1→,[45] and Neu5Acα2→3/6Galβ1→4GlcNAcβ1→,[46] respectively. In addition, recent studies communicated occurrence of other unique sugar modification profiles in EGF10 and EGF11 domains.[47,48]

Our interest has been focused on the significance of specific *O*-glycosylation states in the three-dimensional (3D) structure of the region including a central domain 12, because Fringe-mediated GlcNAc modification of the fucose residue at Thr466 may be a crucial process for the regulation of the selectivity and affinity with Notch ligands.[49] Chemical synthesis allowed for the first time structural characterization of variously *O*-glycosylated domain 12 modules, in which solution structure of the correctly folded human NOTCH1 EGF-like domain 12 is perfectly identical with that of mouse Notch1 EGF-like domain 12.[24] NMR studies uncovered roles of *O*-glycosylation and Ca^{2+} ion in the conformational stabilization of this domain. Notably, the results demonstrated that both GlcNAcβ1→3Fucα1→ at Thr466 and Xylα1→3Xylα1→3Glcβ1→ at Ser458 contribute to the stabilization of the β-hairpin involved in the antiparallel β-sheet structure of this pivotal domain through the interaction of sugars with neighboring amino acid residues, namely the intramolecular "sugar-bridge".[50] It was also indicated that *O*-glycosylation states of the linear precursors of domain 12 affect strongly glutathione-mediated *in vitro* oxidative folding process and efficiency. However, our understanding on the significance of a

variety of *O*-glycosylation patterns found in other important EGF-like domains in human NOTCH1 receptor, particularly the molecular basis as to how site-specific *O*-glycosylation affect the structural behavior of the domains involved in the essential region required for the ligand-binding remain to be elusive.

The present study reports significance of Ca^{2+} coordination and *O*-glycosylation states in the 3D structures of human NOTCH1 EGF-like domains 11 and 10 achieved by chemical synthesis, subsequent *in vitro* glutathione-mediated folding reaction, and comprehensive NMR experiments, because these two adjacent domains have also been considered to participate in the Notch receptor-ligand interactions by collaborating with domain 12. We conceived that specific *O*-glycosylation states on of EGF11 and EGF10 may affect molecular mechanism in receptor-ligand interactions by providing a conformational impact in each domain and/or the region between the domains. Indeed, recent structural studies have implied that EGF10 as well as EGF11 is in close to some Notch ligands, suggesting the presence of a possible contact site between EGF10 and ligands.[51-53] Novel synthetic EGF-like modules enable to decipher significance of the site-specific *O*-glycosylation in the structure of the Ca^{2+}-binding EGF11 and non-Ca^{2+}-binding EGF10 domains.

In chapter 2, to confirm efficacy of the *in vitro* oxidative folding and conformational stabilization of the functional EGF 11 and 10 domain depends on the *O*-glycosylation status, we synthesized model glycopeptides chemically and *in vitro* folding reactions from their unstructured status were assessed by monitoring HPLC profiles of the products. 40-mer linear human Notch1 EGF11 and 10 glycopeptide analogues were synthesized

rapidly and efficiently by microwave-assisted Fmoc solid-phase peptide synthesis protocol. Synthetic EGF modules having designated *O*-glycosylation states demonstrate that effect of *O*-glycosylation on the *in vitro* folding differs entirely between EGF10, EGF11, and EGF12 domains.

In chapter 3, to investigate the structural characteristics of the folded EGF11 and EGF10 domains, NMR structural analysis of each purified module was performed. Comprehensive NMR studies of synthetic EGF modules demonstrated that all peptide backbone adopted to anti-parallel β-sheet structure in Notch ligand binding region. In addition, NMR studies shed light on the new type "sugar bridges" crosslinking Thr-*O*-GlcNAc in the consensus sequence C5-X-X-G-X-(T/S)-G-X-X-C6 and an amino acid in the hinge region between the domains. Furthermore, *O*-glucose trisaccharide in the consensus sequence C1-X-S-X-(P/A)-C2 interacted with peptide chain and stabilize anti-parallel β-sheet in ligand binding region through carbohydrate-peptide interactions. These results suggested that various glycosylation in the EGF domain affects the molecular mechanism of receptor-ligand interaction by causing structural alterations in the domain itself or in the region between the two domains.

1-4. Reference

(1) Varki, a. Biological Roles of Oligosaccharides: All of the Theories Are Correct. *Glycobiology* **1993**, *3*, 97–130.

(2) Lis, H.; Sharon, N. Protein Glycosylation. Structural and Functional Aspects. *Eur. J. Biochem.* **1993**, *218*, 1–27.

(3) Kornfeld, R.; Kornfeld, S. Assembly of Asparagine-Linked Oligosaccharides. *Annu. Rev. Biochem.* **1985**, *54*, 631–664.

(4) Ten Hagen, K. G.; Fritz, T. T. A.; Tabak, L. A; Hagen, K. Ten. All in the Family: The UDP-GalNAc: Polypeptide *N*-Acetylgalactosaminyltransferases. *Glycobiology* **2003**, *13*, 1–16.

(5) Taylor, M. E.; Drickamer, K. Introduction to Glycobiology, Oxford University Press, **2011**.

(6) Endo, T. *O*-Mannosyl Glycans in Mammals. *Biochim. Biophys. Acta* **1999**, *1473*, 237–246.

(7) Harris, R. J.; Spellman, M. W. *O*-Linked Fucose and Other Post-Translational Modifications Unique to EGF Modules. *Glycobiology* **1993**, *3*, 219–224.

(8) Wells, L.; Hart, G. W. O-GlcNAc Turns Twenty: Functional Implications for Post-Translational Modification of Nuclear and Cytosolic Proteins with a Sugar. *FEBS Lett.* **2003**, *546*, 154–158.

(9) Tarp, M. A.; Clausen, H. Mucin-Type *O*-Glycosylation and Its Potential Use in Drug and Vaccine Development. *Biochim. Biophys. Acta - Gen. Subj.* **2008**, *1780*, 546–563.

(10) Varki, a. Biological Roles of Oligosaccharides: All of the Theories Are Correct. *Glycobiology* **1993**, *3*, 97–130.

(11) Brockhausen, I. Pathways of *O*-Glycan Biosynthesis in Cancer Cells. *Biochim. Biophys. Acta - Gen. Subj.* **1999**, *1473*, 67–95.

(12) Kannagi, R.; Izawa, M.; Koike, T.; Miyazaki, K.; Kimura, N. Carbohydrate-Mediated Cell Adhesion in Cancer Metastasis and Angiogenesis. *Cancer Sci.* **2004**, *95*, 377–384.

(13) Hollingsworth, M. A.; Swanson, B. J. Mucins in Cancer: Protection and Control of the Cell Surface. *Nat. Rev. Cancer* **2004**, *4*, 45–60.

(14) Yarema, K. J.; Bertozzi, C. R. Chemical Approaches to Glycobiology and Emerging Carbohydrate-Based Therapeutic Agents. *Curr. Opin. Chem. Biol.* **1998**, *2*, 49–61.

(15) Tachibana, Y.; Fletcher, G. L.; Fujitani, N.; Tsuda, S.; Monde, K.; Nishimura, S.-I. Antifreeze Glycoproteins: Elucidation of the Structural Motifs That Are Essential for Antifreeze Activity. *Angew. Chemie Int. Ed.* **2004**, *43*, 856–862.

(16) Izumi, R.; Matsushita, T.; Fujitani, N.; Naruchi, K.; Shimizu, H.; Tsuda, S.; Hinou, H.; Nishimura, S.-I. Microwave-Assisted Solid-Phase Synthesis of Antifreeze Glycopeptides. *Chem. - A Eur. J.* **2013**, *19*, 3913–3920.

(17) Matsushita, T.; Ohyabu, N.; Fujitani, N.; Naruchi, K.; Shimizu, H.; Hinou, H.; Nishimura, S.-I. Site-Specific Conformational Alteration Induced by Sialylation of MUC1 Tandem Repeating Glycopeptides at an Epitope Region for the Anti-KL-6 Monoclonal Antibody. *Biochemistry* **2013**, *52*, 402–414.

(18) Ohyabu, N.; Hinou, H.; Matsushita, T.; Izumi, R.; Shimizu, H.; Kawamoto, K.; Numata, Y.; Togame, H.; Takemoto, H.; Kondo, H.; Nishimura, S.-I. An Essential Epitope of Anti-MUC1 Monoclonal Antibody KL-6 Revealed by Focused Glycopeptide Library. *J. Am. Chem. Soc.* **2009**, *131*, 17102–17109.

(19) Buskas, T.; Thompson, P.; Boons, G.-J. Immunotherapy for Cancer: Synthetic Carbohydrate-Based Vaccines. *Chem. Commun. (Camb).* **2009**, *36*, 5335–5349.

(20) Vergati, M.; Intrivici, C.; Huen, N.-Y.; Schlom, J.; Tsang, K. Y. Strategies for Cancer Vaccine Development. *J. Biomed. Biotechnol.* **2010**, *2010*, 596432.

(21) Matsushita, T.; Takada, W.; Igarashi, K.; Naruchi, K.; Miyoshi, R.; Garcia-Martin, F.; Amano, M.; Hinou, H.; Nishimura, S.-I. A Straightforward Protocol for the Preparation of High Performance Microarray Displaying Synthetic MUC1 Glycopeptides. *Biochim. Biophys. Acta - Gen. Subj.* **2014**, *1840*, 1105–1116.

(22) Martínez-Sáez, N.; Castro-López, J.; Valero-González, J.; Madariaga, D.; Compañón, I.; Somovilla, V. J.; Salvadó, M.; Asensio, J. L.; Jiménez-Barbero, J.; Avenoza, A.; Busto, J. H.; Bernardes, G. L.; Peregrina, J. M.m Hurtado-Guerrero, R.; Corzana, F. Deciphering the Non-Equivalence of Serine and Threonine *O* -Glycosylation Points: Implications for Molecular Recognition of

the Tn Antigen by an Anti-MUC1 Antibody. *Angew. Chemie Int. Ed.* **2015**, *54*, 9830–9834.

(23) Hashimoto, R.; Fujitani, N.; Takegawa, Y.; Kurogochi, M.; Matsushita, T.; Naruchi, K.; Ohyabu, N.; Hinou, H.; Gao, X. D.; Manri, N.; Satake, H.; Kaneko, A.; Sakamoto, T.; Nishimura, S.-I.. An Efficient Approach for the Characterization of Mucin-Type Glycopeptides: The Effect of *O*-Glycosylation on the Conformation of Synthetic Mucin Peptides. *Chem. - A Eur. J.* **2011**, *17*, 2393–2404.

(24) Hiruma-Shimizu, K.; Hosoguchi, K.; Liu, Y.; Fujitani, N.; Ohta, T.; Hinou, H.; Matsushita, T.; Shimizu, H.; Feizi, T.; Nishimura, S.-I. Chemical Synthesis, Folding, and Structural Insights into *O*-Fucosylated Epidermal Growth Factor-like Repeat 12 of Mouse Notch-1 Receptor. *J. Am. Chem. Soc.* **2010**, *132*, 14857–14865.

(25) Bray, S. J. Notch signalling in context. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 722-735.

(26) Milner, L. A.; Bigas, A. Notch as a mediator of cell fate determination in hematopoiesis: Evidence and speculation. *Blood* **1999**, *93*, 2341-2448.

(27) Nyfeler, Y.; Kirch, R. D.; Mantei, N.; Leone, D. P.; Radtke, F.; Suter, U.; Taylor, V. Jagged1 signals in the postnatal subventricular zone are required for neural stem cell self-renewal. *EMBO J.* **2005**, *24*, 3504-3515.

(28) Radtke, F.; MacDonald, H. R.; Tacchini-Cottier, F. Regulation of innate and adaptive immunity by Notch. *Nat. Rev. Immunol.* **2013**, *13*, 427-437.

(29) Penton, A. L.; Leonard, L. D.; Spinner, N. B. Notch signaling in human development and disease. *Semin. Cell Dev. Biol.* **2012**, *23*, 450-457.

(30) Aster, J. C.; Pear, W. S.; Blacklow, S. C. The Varied Roles of Notch in Cancer. *Annu. Rev. Pathol.* **2017**, *12*, 245-275.

(31) Kopan, R.; Ilagan, M. X. G. The Canonical Notch Signaling Pathway: Unfolding the Activation Mechanism. *Cell* **2009**, *137*, 216-233.

(32) Andersson, E. R.; Sandberg, R.; Lendahl, U. Notch Signaling: Simplicity in Design, Versatility in Function. *Development* **2011**, *138*, 3593-3612.

(33) Haines, N.; Irvine, K. D. Glycosylation Regulates Notch Signaling. *Nat. Rev.*

Mol. Cell Biol. **2003**, *4*, 786-797.

(34) Jafar-Nejad, H.; Leonardi, J.; Fernandez-Valdivia, R. Role of glycans and glycosyltransferases in the regulation of Notch signaling. *Glycobiology* **2010**, *20*, 931-949.

(35) Takeuchi, H.; Haltiwanger, R. S. Significance of glycosylation in Notch signaling. *Biochem. Biophys. Res. Commun.* **2014**, *453*, 235-242.

(36) Acar, M.; Jafar-Nejad, H.; Takeuchi, H.; Rajan, A.; Ibrani, D.; Rana, N. A.; Pan, H.; Haltiwanger, R. S.; Bellen, H. J. Rumi is a CAP10 domain glycosyltransferase that modifies Notch and is required for Notch signaling. *Cell* **2008**, *132*, 247-258.

(37) Okajima, T.; Irvine, K. D. Regulation of Notch signaling by *O*-linked fucose. *Cell* **2002**, *111*, 893-904.

(38) Matsuura, A.; Ito, M.; Sakaidani, Y.; Kondo, T.; Murakami, K.; Furukawa, K.; Nadano, D.; Matsuda, T.; Okajima, T. *O*-Linked *N*-Acetylglucosamine Is Present on the Extracellular Domain of Notch Receptors. *J. Biol. Chem.* **2008**, *283*, 35486-35495.

(39) Rana, N. A.; Nita-Lazar, A.; Takeuchi, H.; Kakuda, S.; Luther, K. B.; Haltiwanger, R. S. *O*-Glucose Trisaccharide Is Present at High but Variable Stoichiometry at Multiple Sites on Mouse Notch1. *J. Biol. Chem.* **2011**, *286*, 31623-31637

(40) Shao, L.; Moloney, D. J.; Haltiwanger, R. Fringe Modifies *O*-Fucose on Mouse Notch1 at Epidermal Growth Factor-like Repeats within the Ligand-Binding Site and the Abruptex Region. *J. Biol. Chem.* **2003**, *278*, 7775-7782.

(41) Alfaro, J. F.; Gong, C.; Monroe, M. E.; Aldrich, J. T.; Clauss, T. R. W.; Purvine, S. O.; Wang, Z.; Camp, D. G.; Shabanowitz, J.; Stanley, P.; Hart, G. W.; Hunt, D. F.; Yang, F.; Smith, R. D. Tandem Mass Spectrometry Identifies Many Mouse Brain *O*-GlcNAcylated Proteins Including EGF Domain-Specific *O*-GlcNAc Transferase Targets. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 7280-7285.

(42) Li, Z.; Han, H.; Pak, J. E.; Satkunarajah, M.; Zhou, D; Rini, J. M. Recognition of EGF-like domains by the Notch-modifying *O*-

fucosyltransferase POFUT1. *Nat. Chem. Biol*. **2017**, *13*, 757-763.

(43) Takeuchi, H.; Schneider, M.; Williamson, D. B.; Ito, A.; Takeuchi, M.; Handford, P. A.; Haltiwanger, R. S. Two novel protein *O*-glucosyltransferases that modify sites distinct from POGLUT1 and affect Notch trafficking and signaling. *Proc. Natl. Acad. Sci. U S A.* **2018**, *115*, E8395-E8402.

(44) Moloney, D. J.; Shair, L. H.; Lu, F. M.; Xia, J.; Locke, R.; Matta, K. L.; Haltiwanger, R. S. Mammalian Notch1 is modified with two unusual forms of *O*-linked glycosylation found on epidermal growth factor-like modules. *J Biol Chem.* **2000**, *275*, 9604-9611.

(45) Whitworth, G. E.; Zandberg, W. F.; Clark, T.; Vocadlo, D. J. Mammalian Notch is modified by D-Xyl-α1-3-D-Xyl-α1-3-D-Glc-β1-*O*-Ser: Implementation of a method to study *O*-glucosylation. *Glycobiology* **2010**, *20*, 287–299.

(46) Ogawa, M.; Okajima, T. Structure and function of extracellular *O*-GlcNAc. *Curr. Opin. Struct. Biol.* **2019**, *56*, 72-77.

(47) Ogawa, M.; Senoo, Y.; Ikeda, K.; Takeuchi, H.; Okajima, T. Structural Divergence in *O*-GlcNAc Glycans Displayed on Epidermal Growth Factor-like Repeats of Mammalian Notch1. *Molecules* **2018**, *23*, 1745.

(48) Urata, Y.; Saiki, W.; Tsukamoto, Y.; Sago, H.; Hibi, H.; Okajima, T.; Takeuchi, H. Xylosyl Extension of O-Glucose Glycans on the Extracellular Domain of NOTCH1 and NOTCH2 Regulates Notch Cell Surface Trafficking. *Cells* **2020**, *9*, 1220.

(49) Taylor, P.; Takeuchi, H.; Sheppard, D.; Chillakuri, C.; Lea, S. M.; Haltiwanger, R. S.; Handford, P. A. Fringe-Mediated Extension of *O*-Linked Fucose in the Ligand-Binding Region of Notch1 Increases Binding to Mammalian Notch Ligands. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 7290–7295.

(50) Hayakawa, S.; Yokoi, Y.; Hinou, H.; Nishimura, S.-I. Chemical Synthesis Demonstrates That Dynamic *O*-Glycosylation Regulates the Folding and Functional Conformation of a Pivotal EGF12 Domain of the Human NOTCH1 Receptor. *Biochemistry* **2017**, *56*, 4379-4391.

(51) Luca, V. C.; Jude, K. M.; Pierce, N. W.; Nachury, M. V.; Fischer, S.; Garcia,

K. C. Structural basis for Notch1 engagement of Delta-like 4. *Science* **2015**, *347*, 847-853.

(52) Luca, V. C.; Kim, B. C.; Ge, C.; Kakuda, S.; Wu, D.; Roein-Peikar, M.; Haltiwanger, R. S.; Zhu, C.; Ha, T.; Garcia, K. C. Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity. *Science* **2017**, *355*, 1320-1324.

(53) Weisshuhn, P. C.; Sheppard, D.; Taylor, P.; Whiteman, P.; Lea, S. M.; Handford, P. A.; Redfield, C. Non-Linear and Flexible Regions of the Human Notch1 Extracellular Domain Revealed by High-Resolution Structural Studies. *Structure* **2016**, *24*, 555-566.

Chapter 2

Chemical Synthesis and In Vitro Oxidative Folding of Human NOTCH1 EGF11 and EGF10 Modules

2-1. Introduction

EGF12 and EGF11 are known to be key domains of human NOTCH1 because these domains have the minimal binding units required for physical interaction with major ligands.[1,2] Especially, on EGF12, which is modified with *O*-fucose glycan at Thr466 and *O*-glucose glycan at Ser458, several studies have demonstrated that *O*-fucose glycan regulates Notch-ligand interactions. Notably, Fringe *N*-acetylglucosaminyltransferase-mediated GlcNAc modification of the fucose residue at Thr466 strongly influence binding specificity and affinity of the EGF12 domain with ligands.[3-5] Although *O*-glucosylation-initiated modification at Ser458 does not directly participate in the Notch-ligand binding,[6,7] protein *O*-glucosyltransferase (POGLUT1),[8] which is responsible for the introduction of glucose to the Ser458 residue, may affect the processing of the human NOTCH1 receptor and regulate signaling by ICDs.[8,9] Overexpression of xylosyltransferases (GXYL1, GXYL2, and XXYLT1), which are responsible for the addition of xylose residues to the proximal *O*-glucose residue, results in negative regulation of Notch signaling in *Drosophila*.[10] These observations prompted us to assess hypothesis that dynamic *O*-glycosylation states affect significantly the structural stabilization of the central EGF12 domain, including on its folding process.

We have reported the chemical synthesis, *in vitro* folding, and NMR studies on the human NOTCH1 EGF12 modules having various *O*-glycoforms, such as Glcβ1→ or Xylα1→3Xylα1→3Glcβ1→ at Ser458 and Fucα1→ or GlcNAcβ1→3Fucα1→ at Thr466 residues.[11,12] The results demonstrated that several *O*-glycosylation states created by the synthetic EGF12 modules affect strongly the *in vitro* folding process and efficiency. For example, *O*-glucosylation at Ser458 facilitates proper folding of the EGF12 domain,

while non-glycosylated linear EGF12 peptide affords large amounts of misfolded products during *in vitro* oxidative folding. Moreover, *O*-fucosylation at Thr466 prior to *O*-glucosylation at Ser458 disturbs folding of EGF12 independent of Ca^{2+} ion, whereas modification of this Fucα1→ moiety with GlcNAc dramatically enhances folding efficiency. In addition, we also elicited that extension of the Glcβ1→ moiety with xyloses is a negative-regulation mechanism in the folding of EGF12 when synthesis of a trisaccharide (Xylα1 → 3Xylα1 → 3Glcβ1 →) at Ser458 dominates over the posttranslational modification at Thr466. We have demonstrated that efficacy of the canonical EGF folding and stabilization of the functional EGF 12 domain depends strongly on the *O*-glycosylation status both at Thr466 and Ser458 residues.[12] However, our understanding on the significance of a variety of *O*-glycosylation patterns found in other important EGF-like repeats in human NOTCH1 receptor, particularly the molecular basis as to how dynamic *O*-glycosylation affect the folding and conformation of each EGF domain remain unclear.

The present study reports significance of Ca^{2+} coordination and *O*-glycosylation in the 3D structures of human NOTCH1 EGF- like domains 11 and 10 achieved by *in vitro* GSH-mediated folding, because these two domains also participate in the Notch receptor-ligand interactions by collaborating with domain 12.[6,7] Indeed, recent structural studies have implied that EGF10 as well as EGF11 is in close to the ligands, especially DLL-4 proximity, suggesting the presence of a possible contact site between EGF10 and the ligands.[6,7,13] In addition, Our attention is further directed to molecular mechanism of the disulfide exchange reactions in the matured NOTCH1 EGF-like domains, since GSH is the most abundant antioxidant molecule having the redox potential inside cells that

present in millimolar concentration and kept reduced by NADPH and glutathione reductase.[14,15] Novel synthetic EGF modules allow for deciphering significance of the site-specific *O*-glycosylation in the folding of the Ca^{2+}-binding EGF11 and non-Ca^{2+}-binding EGF10 domains and underlying molecular mechanisms that may determine the integrity of the cell surface NOTCH1 receptor.

2-2. Result and Discussion

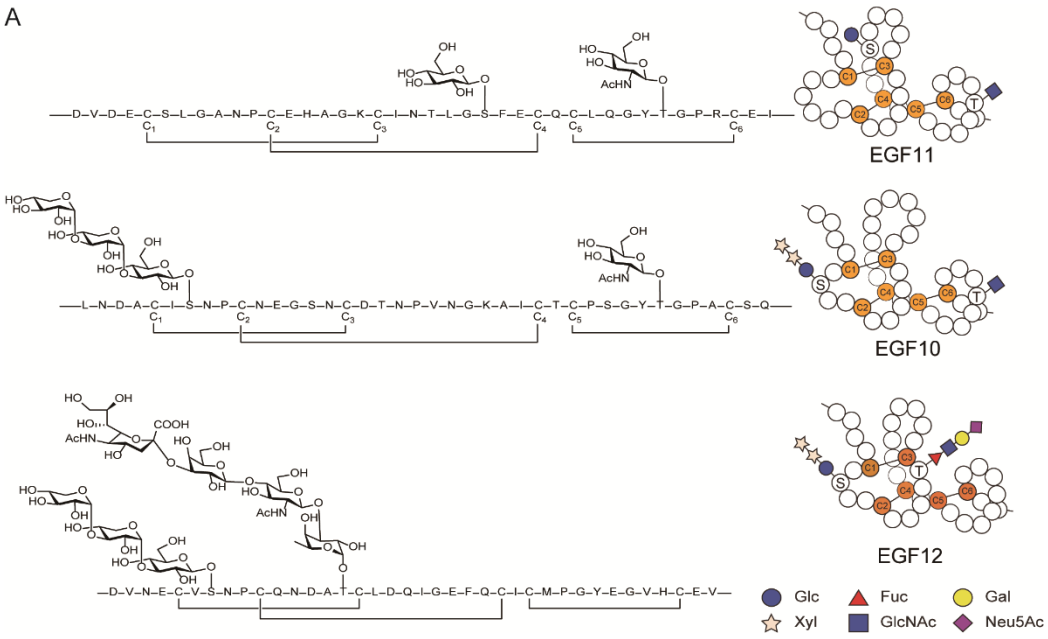
2-2-1. Chemical synthesis of human NOTCH1 EGF11 and EGF10 modules having designated O-glycosylation states.

Figure 2-1 illustrates correctly folded structures of human NOTCH1 extracellular domains 12, 11, and 10 having unique *O*-glycans and their distributions in the characteristic motifs between the two cysteine residues. EGF-like domain 11 has an *O*-GlcNAcylation site at Thr445 within the C5-X-X-G-X-(T/S)-G-X-X-C6,[16,17] whereas *O*-glucosylation site at Ser435 residue is involved in the novel consensus sequence, C3-X-N-T-X-G-S-F-X-C4.[17] On the other hand, *O*-glycans of domain 10, both *O*-GlcNAcylation at Thr405 and *O*-glucosylation at Ser378, occur within C5-X-X-G-X-(T/S)-G-X-X-C6 and C1-X-S-X-(P/A)-C2, two general consensus sequences reported previously.[16,18] Differences as well as similarities in the glycosylation sites and glycoforms of these consecutive EGF-like domains 10~12 greatly motivated us to assess effect of many glycosylation states on the structural characteristics of each domain. It is interesting to note that both domains 11 and 10 bear *O*-GlcNAcylation within C5-X-X-G-X-(T/S)-G-X-X-C6 motif while *O*-glucosylation occurs in entirely different motifs C3-X-N-T-X-G-S-F-X-C4 (domain 11) and C1-X-S-X-(P/A)-C2 (domain 10), respectively. These site-specific *O*-glucosylation of the domain 11 and domain 10 would lead to the distinctive effect on their structural behavior. Notably, it was demonstrated that Glc residue at Ser458 within C1-X-S-X-(P/A)-C2 motif of EGF12 domain is further extended with two xyloses to afford unique glycoform Xylα1→3Xylα1→3Glcβ1→. It seems likely that *O*-glucosylation at Ser378 within C1-X-S-X-(P/A)-C2 motif of domain 10 can also be modified with xyloses while modification of Glc at Ser435 residue involved in the novel consensus sequence of domain 11, C3-X-N-T-X-G-S-F-X-C4, has not been

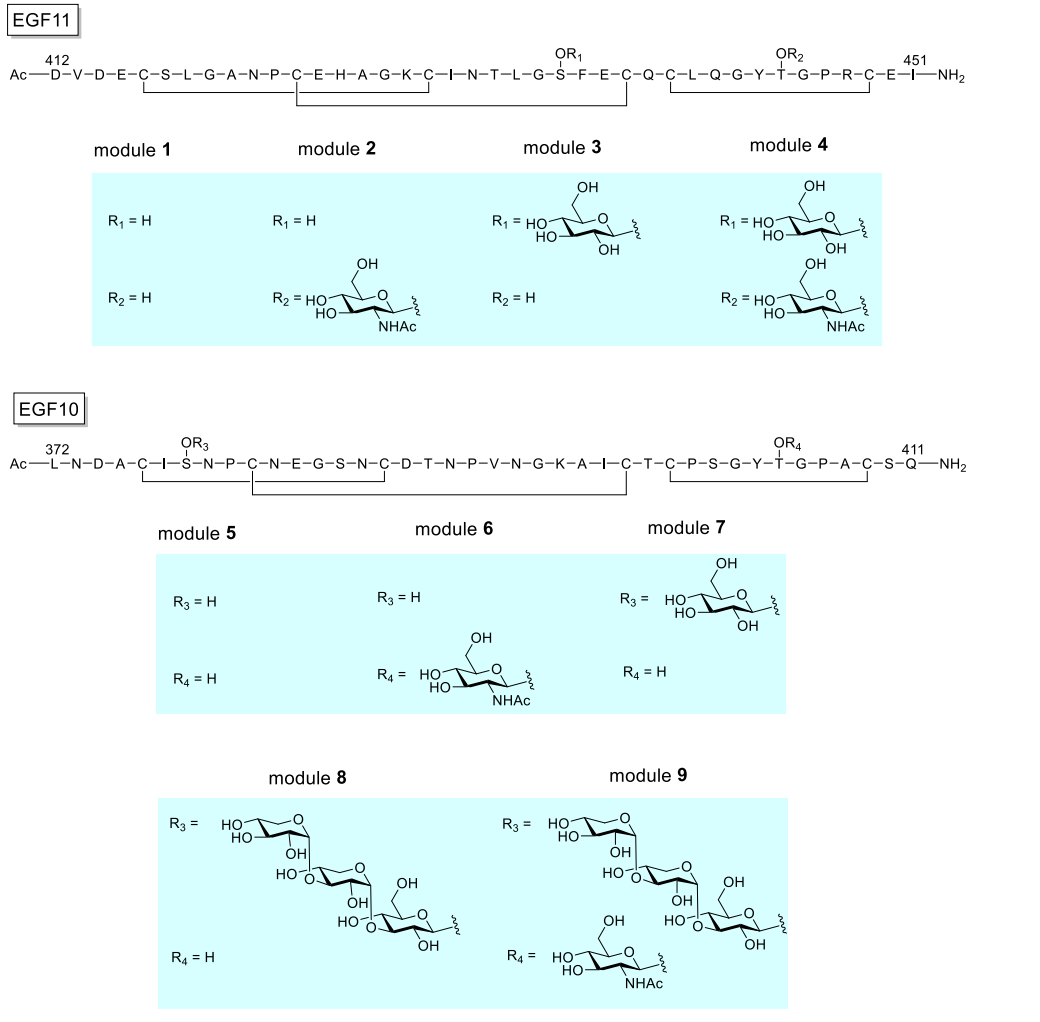
reported.

To investigate effect of variable *O*-glycosylation states at the distinctive sites on the *in vitro* folding process and conformation of the domains 11 and 10, we designed and synthesized domain 11 modules **1**~**4** and domain 10 modules **5**~**9** (Figure 2-1B) having possible *O*-glycosylation states generable in the endoplasmic reticulum (ER), in which the glycan biosynthesis may be conducted by the key enzymes such as POGLUT1~3,[16,17] EGF domain specific *O*-GlcNAc transferase (EOGT),[19,20] and xylosyltransferases[10] in the presence of sugar nucleotides. Figure 2-1C represents a standardized protocol for chemical synthesis of the linear peptide/glycopeptide precursors based on the microwave-assisted solid phase glycopeptide synthesis[11,12,21,22] by using key building blocks **10**~**12** having three distinct glycoforms found in domains 11 and 10. With this strategy, we synthesized efficiently all linear precursors of domains 11 and 10 modules **1**~**9** composed of 40 amino acid residues. The unstructured linear precursors having six thiol groups were subjected to *in vitro* glutathione-mediated oxidative folding reaction for the formation of intramolecular three disulfide bonds.[11,12,21]

A


EGF11
EGF10
EGF12
Glc
Fuc
Gal
Xyl
GlcNAc
Neu5Ac

B


EGF11
EGF10
module 1
module 2
module 3
module 4
module 5
module 6
module 7
module 8
module 9

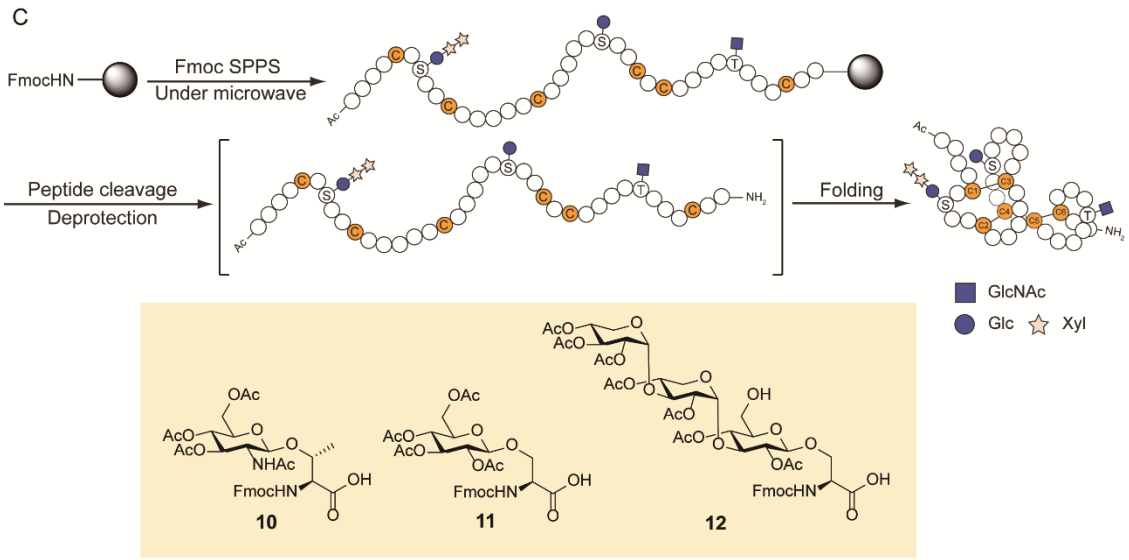


Figure 2-1. Synthetic human NOTCH1 EGF11 and EGF10 modules. (A) Chemical structures of EGF11, EGF10, and EGF12 domains having unique three-types of *O*-glycans and their distributions in four characteristic consensus motifs between the two cysteine residues. (B) Chemical structures of EGF modules **1**~**9** synthesized in this study. (C) A general synthetic strategy of EGF11 and EGF10 modules by using key building blocks **10**~**12**. This scheme illustrates a simplified synthetic procedure of a simplified EGF model to represent three distinctive glycosylation sites found in EGF11 and EGF10 domains. EGF11 and EGF10 domains share the *O*-GlcNAcylation sites within C5-X-X-G-X-(T/S)-G-X-X-C6, while *O*-Glucosylation occurs at two different sites within C3-X-N-T-X-G-S-F-X-C4 and C1-X-S-X-(P/A)-C2, respectively.

2-2-2. Folding of EGF11 domain proceeds in totally different manner from EGF10 and EGF12 domains.

To decipher how *O*-glycosylation states contribute to the *in vitro* folding process of each EGF domain, we first examined *in vitro* folding experiments of non-glycosylated linear EGF11 and EGF10 precursors according to the general conditions established previously.[21] It should also be noted that both temperature and Ca^{2+} ion influence significantly the folding efficiency of EGF12 domain. Figure 2-2A shows HPLC profiles of non-glycosylated linear EGF11 and EGF10 precursors (0.1 mg/mL) after oxidative folding reactions conducted for 24 h at 25 ºC or 4 ºC in the presence or absence of Ca^{2+} in 50 mM Tris-HCl buffer (pH 8.0) containing 1.0 mM reduced-type glutathione and 0.2 mM oxidized-type glutathione.

Surprisingly, *in vitro* oxidative folding process of non-glycosylated linear EGF11 precursor monitored by HPLC appeared to proceed smoothly and gave eventually only single major peak, even though the reactions were conducted under four different conditions, whereas the reactions of non-glycosylated linear EGF10 precursor yielded several distinctive peaks indicating a complicated product mixture in all experiments (Figure 2-2A). The results clearly show that molecular mechanism in the folding of non-glycosylated linear EGF11 precursor is completely different from those of non-glycosylated linear EGF10 precursor and non-glycosylated linear EGF12 precursor.

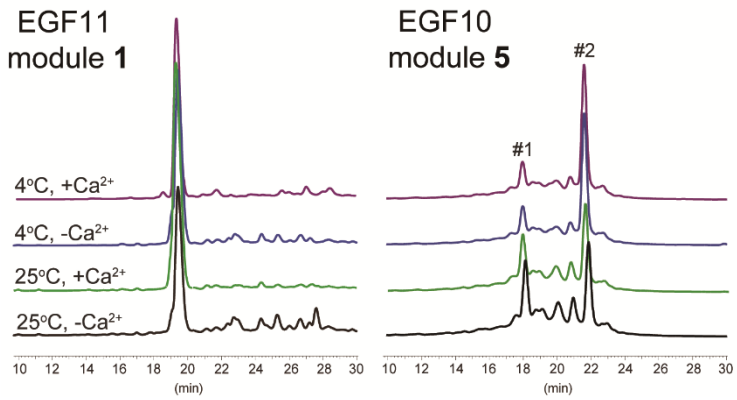
To determine the positions of disulfide bonds formed in the folding reaction, we isolated the major product generated by large-scale folding reaction of non-glycosylated linear EGF11 unstructured peptide at 25 ºC in the presence of Ca^{2+}. The mass value of the

isolated major product was found to decrease by 6 Da when compared with the starting non-glycosylated linear EGF11 precursor (Figure 2-2B), suggesting that the result is consistent with the predicted molecular mass of the properly folded EGF11 module **1** having three disulfide bonds. Furthermore, NOESY spectra of the isolated major product provided evidence of the clear correlation between the β-protons of Cys416-Cys429 (C1-C3), Cys423-Cys428 (C2-C4), and Cys430-Cys449 (C5–C6) as shown in Figure 2-2D. These results demonstrate that the folding reaction of non-glycosylated linear EGF11 precursor affords the properly folded EGF11 module **1** in high yields. Remarkably, it was also revealed that the folding efficacy of EGF11 module **1** is enhanced significantly at 4ºC (64.5%→83.2%) and in the presence of calcium ion (64.5%→92.1%) as shown in Figure 2-2F.

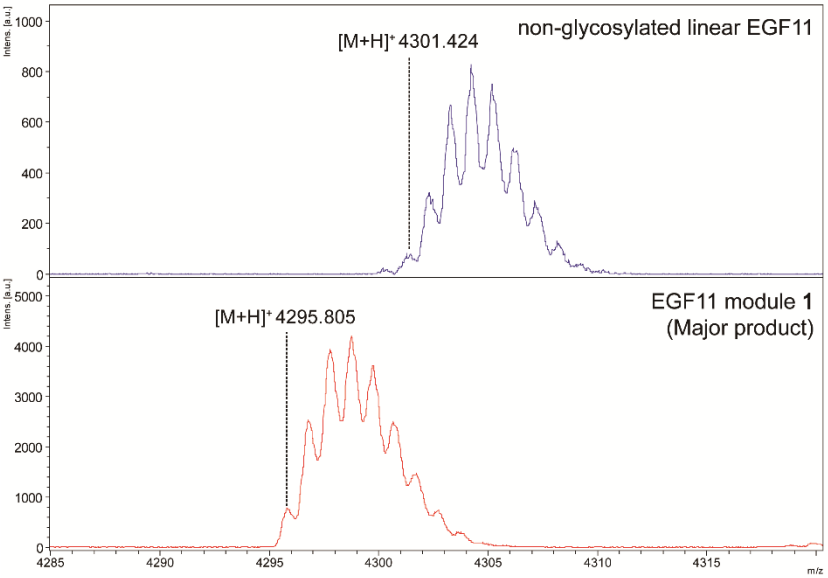
On the contrary, folding of non-glycosylated linear EGF10 precursor led to highly complicated mixture containing two major product peaks in HPLC eluted at 18 and 22 min, respectively (Figure 2-2A). The products isolated both from fractions #1 and #2 gave the expected mass value corresponding to the theoretical mass of properly folded EGF10 module **5** (Figure 2-2C), indicating the formation of two distinctively "folded" isomers with different disulfide bonding patterns. NOESY spectra uncovered that fraction #2 is correctly folded EGF10 module **5** having three disulfide bond pairs Cys376-Cys387 (C1-C3), Cys381-Cys398 (C2-C4), and Cys400-Cys409 (C5–C6) as shown in Figure 2-2E, while we could not isolate enough amount of the isomer at fraction #1 enabling NMR measurement. Although the folding conducted at 4 ºC enhances the yield of the properly folded EGF10 module **5** (39.7%→63.2%), Ca^{2+} ion does not influence efficacy in the

folding (39.7%→40.4%) as shown in Figure 2-2F. These results are consistent with the finding that EGF10 is not Ca^{2+}-binding domain.[23]

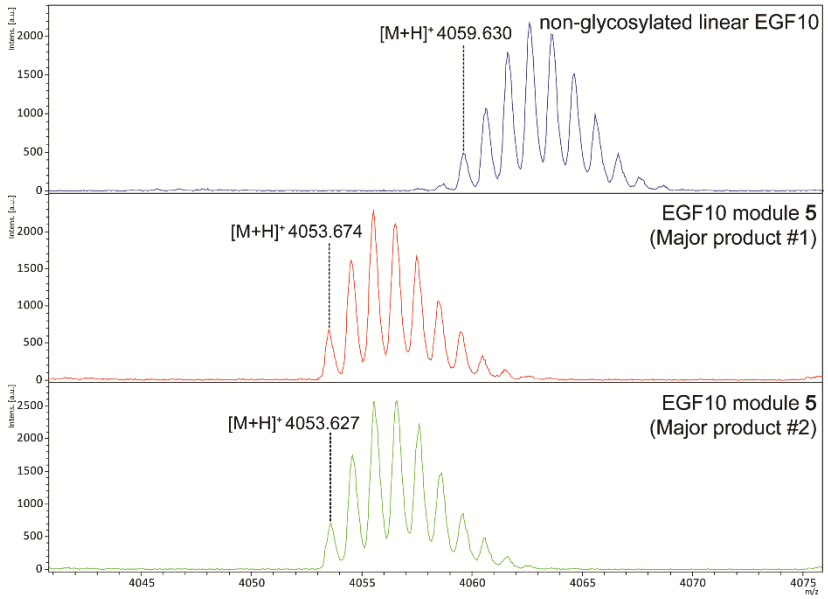
A

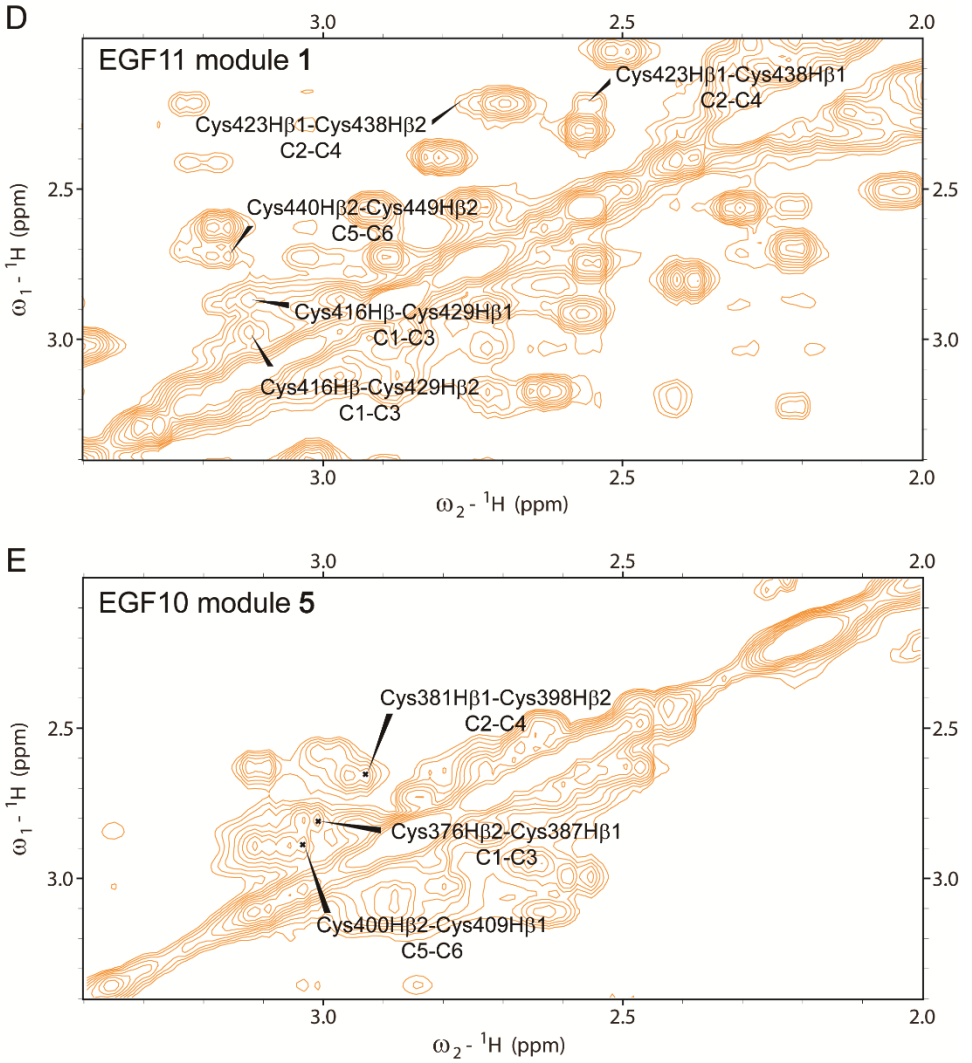

EGF11
module 1
EGF10
module 5
#1
#2
4°C, +Ca2+
4°C, -Ca2+
25°C, +Ca2+
25°C, -Ca2+
10 12 14 16 18 20 22 24 26 28 30
(min)

B


Intens. [a.u.]
[M+H]+ 4301.424
non-glycosylated linear EGF11
[M+H]+ 4295.805
EGF11 module 1
(Major product)
4285 4290 4295 4300 4305 4310 4315
m/z

C


Intens. [a.u.]
[M+H]+ 4059.630
non-glycosylated linear EGF10
[M+H]+ 4053.674
EGF10 module 5
(Major product #1)
[M+H]+ 4053.627
EGF10 module 5
(Major product #2)
4045 4050 4055 4060 4065 4070 4075
m/z


D
EGF11 module 1
Cys423Hβ1-Cys438Hβ1
C2-C4
Cys423Hβ1-Cys438Hβ2
C2-C4
Cys440Hβ2-Cys449Hβ2
C5-C6
Cys416Hβ-Cys429Hβ1
C1-C3
Cys416Hβ-Cys429Hβ2
C1-C3
ω1 - 1H (ppm)
ω2 - 1H (ppm)
E
EGF10 module 5
Cys381Hβ1-Cys398Hβ2
C2-C4
Cys376Hβ2-Cys387Hβ1
C1-C3
Cys400Hβ2-Cys409Hβ1
C5-C6
ω1 - 1H (ppm)
ω2 - 1H (ppm)

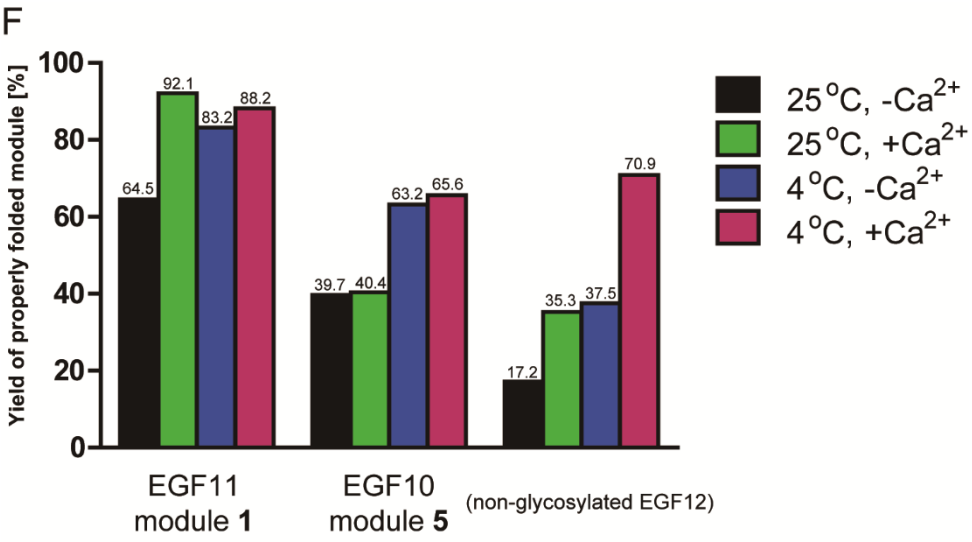

F
Yield of properly folded module [%]
64.5
92.1
83.2
88.2
39.7
40.4
63.2
65.6
17.2
35.3
37.5
70.9
25 °C, -Ca2+
25 °C, +Ca2+
4 °C, -Ca2+
4 °C, +Ca2+
EGF11
module 1
EGF10
module 5
(non-glycosylated EGF12)

Figure 2-2. Glutathione-mediated oxidative folding reactions of non-glycosylated linear EGF11 and EGF10 peptides. (A) HPLC profiles of the mixtures after folding reactions. Reactions were conducted under the following conditions for 24 h: 5 mM $CaCl_2$ at 4ºC (magenta), at 4ºC in the absence of Ca^{2+} (blue), 5 mM $CaCl_2$ at 25ºC (green), and at 25ºC in the absence of Ca^{2+} (black). (B and C) MALDI-TOFMS spectra of the major products isolated after the folding reactions and the starting EGF11 and EGF10 peptides, respectively. (D and E) NOESY spectra of properly folded non-glycosylated EGF11 module **1** and EGF10 module **5** (fraction #2). (F) The efficacy of the properly folded EGF modules **1** and **5** under the abovementioned four different conditions. The yield (%) of the correctly folded modules were calculated from the peak area corresponding to the product peak in the HPLC profiles. To compare differences in the folding mechanisms between three central EGF-like domains, the peak area of the correctly folded product in HPLC profiles of non-glycosylated linear EGF12 peptide[31] were calculated (see also Supplementary Information, Figure S2-3).

2-2-3. Effect of O-glycosylation on the folding of NOTCH1 EGF11 and EGF10 repeats.

Our interest was next focused on the effect of posttranslational glycosylation on the *in vitro* folding reaction of EGF11 and EGF10 domains. We have uncovered that the folding of synthetic linear EGF12 precursors is greatly influenced by the glycosylation states at Thr466 and Ser458 residues. Efficacy of the accurate folding of the EGF12 domain depends significantly on the sugar extension of Fucα1→ moiety at Thr466 residue by Fringe with GlcNAc residue. Inversely, modification at Ser458 residue with a trisaccharide, Xylα1→3Xylα1→3Glcβ1→, functions as a negative modulator during the folding reaction. It is also important to note that "sugar-bridge", the intramolecular interaction between carbohydrate moiety and the peptide region stabilizes β-hairpin structure within the antiparallel β-sheet conformation of the crucial EGF12 domain.[12] We hypothesized that formation of the three typical disulfide cross-linkages, C1-C3, C2-C4, and C5-C6, found commonly in EGF/EGF-like proteins might be affected significantly by *O*-glycosylation states of the synthetic EGF11 and EGF10 modules bearing designated glycosylation patterns.

To test this hypothesis, *in vitro* folding reactions of unstructured linear precursors of EGF11 modules **2**~**4** and EGF10 modules **6**~**9** were elucidated by monitoring HPLC profiles of the products generated under the conditions employed for the linear non-glycosylated precursors of EGF modules **1** and **5**. Contrary to our expectations, the HPLC profiles showed that folding of all glycosylated precursors of EGF11 modules **2**~**4** proceed smoothly and afford properly folded products in a similar manner to the results observed in the folding of non-glycosylated EGF11 module **1** (Figure 2-3A and Figure

S2-1). The results clearly demonstrate that folding of EGF11 domain is not affected by *O*-GlcNAcylation at Thr445 and *O*-glucosylation at Ser435 residues. Moreover, chelation of Ca^{2+} ion uniformly enhances folding efficiency of this domain independent of the glycosylation states as indicated in EGF11 modules **2** (67.6→92.5%), **3** (63.8→93.5%), and **4** (67.4→93.2%) when compared with the folding in the absence of Ca^{2+} as summarized in Figures 2-3B and 2-3C.

Our results provided evidence that molecular mechanisms in the folding of glycosylated linear EGF10 precursors **6**~**9** are totally different from those of EGF11 modules **2**~**4**. As shown in Figures 2-3A~2-3C and Figure S2-2, it was evident that *in vitro* glutathione-mediated oxidation of non-Ca^{2+}-binding EGF10 precursors **6**~**9** yields large amount of misfolded products (56.9~80.4%) both at 25ºC and 4ºC independent of Ca^{2+}ion. More importantly, these observations demonstrate that the modification at Ser378 with a trisaccharide (Xylα1→3Xylα1→3Glcβ1→) reduces equally the folding efficiency when compared with that of the non-glycosylated EGF10 repeat **5** as shown in the folding of modules **8** (40.4→20.2%) and **9** (40.4→21.7%) while other glycosylation states do not disturb the folding of this domain, **6** (40.4→38.8%) and **7** (40.4→43.1%). Considering that extension of the Glcβ1→ moiety with xyloses at Ser458 also impedes folding of EGF12 domain (Figure 2-3C and Figure S2-3),[31] it is concluded that posttranslational modification with this trisaccharide moiety at serine residue involved in the consensus sequence, C1-X-S-X-(P/A)-C2, is a general negative-regulation mechanism in the folding of Notch receptor EGF-like domains.

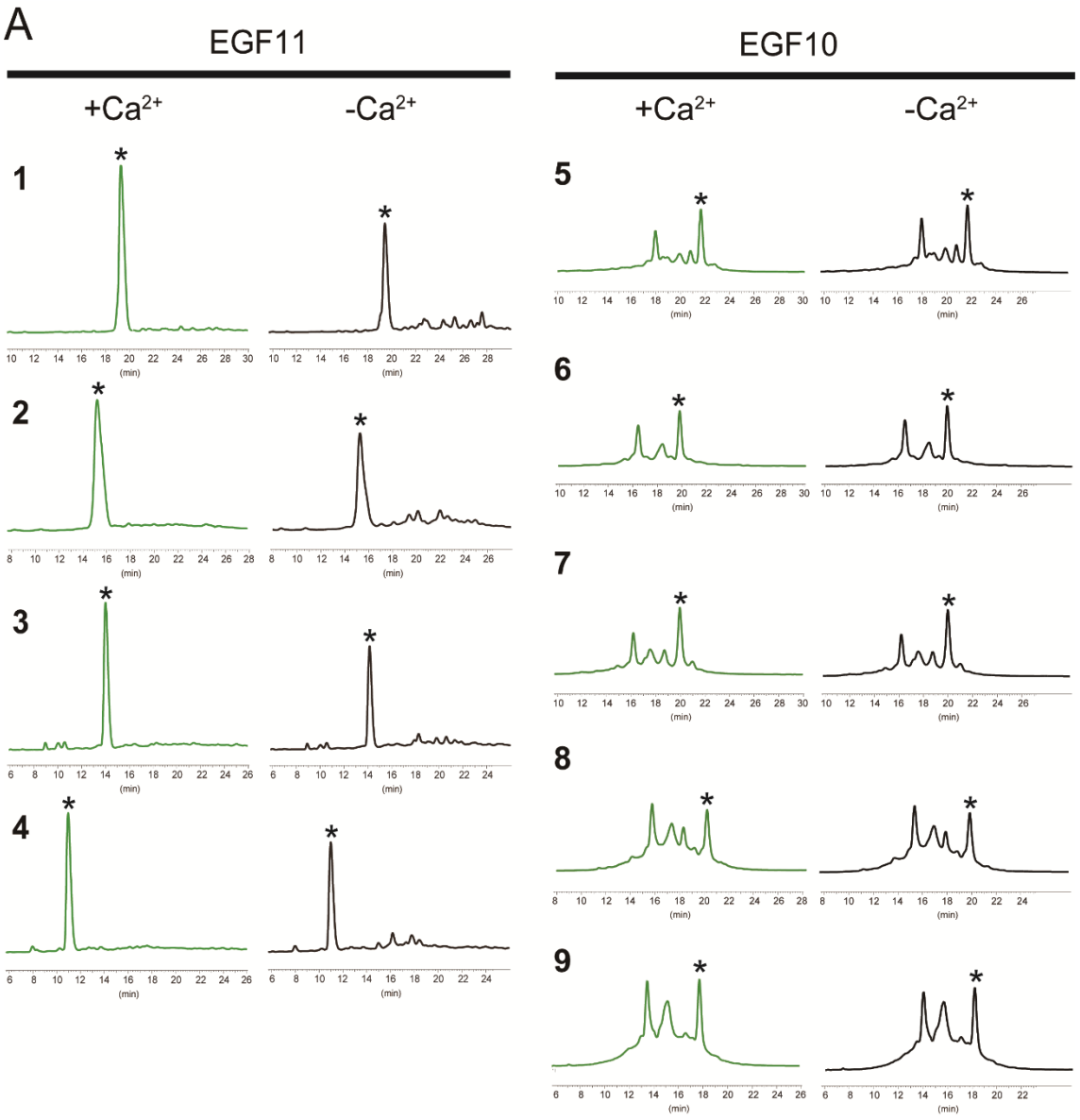
To decipher structural and functional roles of the individual EGF-like repeats in human

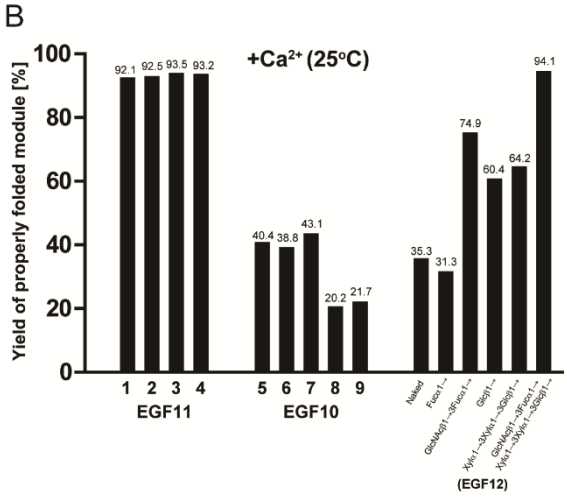
NOTCH1 receptor, it seemed likely that the optimal conditions giving the highest yield of the canonical EGF folds for individual domains would allow for the large-scale synthesis of the artificial EGF-like repeats harboring designated glycosylation states. As expected, the folding conducted at 25°C in the presence of Ca^{2+} facilitated highly efficient synthesis of EGF11 modules **1** (4.3 mg, 8.9%), **2** (4.0 mg, 7.9%), **3** (4.1 mg, 8.2%), and **4** (4.5 mg, 8.6%), respectively (Figure 2-3A, Figures S2-1, and S2-4). Here, the yields (%) mean overall yield of the canonical EGF folds in the whole synthetic process starting from the solid-phase synthesis of precursors to the isolation of the correctly folded products. On the other hand, the folding of EGF10 modules was found to proceed more efficiently at 4 °C than that conducted at 25 °C independent of Ca^{2+} ion (Figure 2-3A, Figures S2-2 and S2-4). Indeed, folding of the precursors of EGF10 modules at 4 °C allowed for the synthesis of the correctly folded modules **5** (2.5 mg, 5.5%), **6** (2.7 mg, 5.6%), **7** (2.6 mg, 5.5%), **8** (2.1 mg, 4.2%), and **9** (1.8 mg, 3.4%), respectively. Accordingly, NOE correlations between the Hβ atoms of the two cysteine residues of all these isolated products demonstrated correctly folded structures of newly synthesized EGF modules **1**~**9**, in which all NOE signals indicating properly formed disulfide bond pairs were detected in a range of 2.0~3.4 ppm (Figures S2-12 and S2-13).

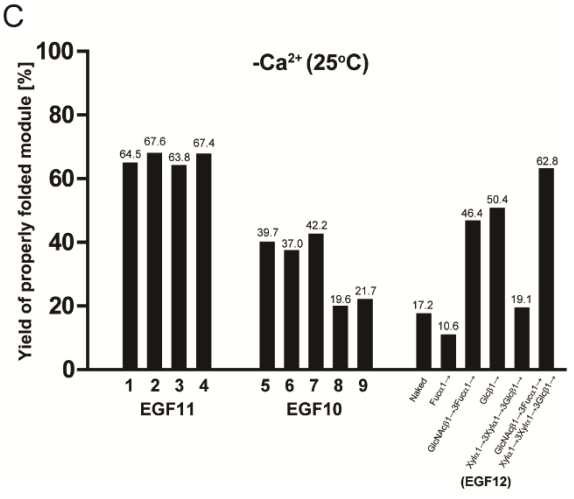
Our attention was directed to the presence of ubiquitous disulfide reduction mechanism for maintaining proteins in the reduced state, namely thioredoxin (TRX) and GSH that can be reduced by electrons from NADPH via thioredoxin reductase and glutathione reductase, respectively.[14] Remarkably, GSH is the most abundant antioxidant molecule showing the redox potential inside cells that present in millimolar concentration in the cytosol.[15] We hypothesized that NOTCH1 ECDs have redox active disulfide bonds that

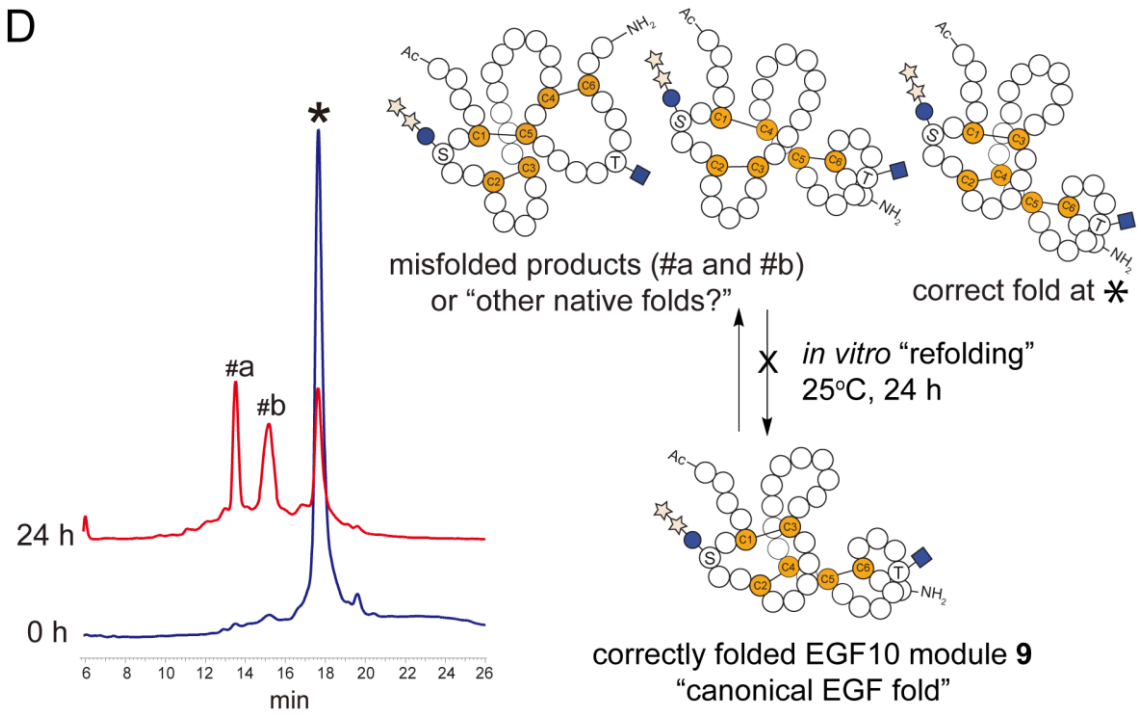
induce disulfide exchange reactions leading to the domain-specific equilibria between canonical EGF-like folds and misfolded/unfolded structures under a physiological reductant condition.

To test this hypothesis, we analyzed *in vitro* refolding processes of the purified canonical EGF folds of domains 10 in a general intracellular redox environment (millimolar concentration of GSH-mediated folding). Surprisingly, the results provided evidence that the correctly folded pure module **9** of the domain 10 having two different *O*-glycans (GlcNAcβ1 → at Thr405 and Xylα1 → 3Xylα1 → 3Glcβ1 → at Ser378) initiates unfolding and disulfide-shuffling reactions, and results in the mixture containing 56% of two major misfolded structures (peak #a = 28% and peak #b = 28%), and correctly folded module 9 (26%) when exposed to the general condition employed for *in vitro* GSH-mediated oxidative folding experiments (Figures 2-3D and 2-3E). Obviously, the HPLC profile was mostly identical with that of the mixture observed after folding from the linear precursor 9 of domain 10 under the same condition (Figure 2-3A), demonstrating that the folding of non-Ca-binding domain 10 is regulated negatively by the site-specific glycosylation at Ser378 residue with a trisaccharide moiety (Xylα1→3Xylα1→3Glcβ1 →) and leads eventually to the mixture of thermodynamically stable misfolded isomers with the same molecular mass. These results can propose presence of the equilibrium between canonical EGF-like folds and misfolded unusual structures in NOTCH1 extracellular 36 tandem repeating domains, in which Ca^{2+} ion and *O*-glycosylation affect strongly functional 3D structures in a domain-specific manner.


A
EGF11
EGF10
+Ca2+
-Ca2+
1
2
3
4
5
6
7
8
9
(min)


B
+Ca2+ (25°C)
Yield of properly folded module [%]
92.1 92.5 93.5 93.2
40.4 38.8 43.1 20.2 21.7
35.3 31.3 74.9 60.4 64.2 94.1
1 2 3 4
EGF11
5 6 7 8 9
EGF10
Naked
Fuc1→
GlcNAcβ1→3Fuc1→
Glc1→
Xyl1→3Xyl1→3Glcβ1→
GlcNAcβ1→3Fuc1→
Xyl1→3Xyl1→3Glcβ1→
(EGF12)


C
-Ca2+ (25°C)
Yield of properly folded module [%]
64.5 67.6 63.8 67.4
39.7 37.0 42.2 19.6 21.7
17.2 10.6 46.4 50.4 19.1 62.8
1 2 3 4
EGF11
5 6 7 8 9
EGF10
Naked
Fuc1→
GlcNAcβ1→3Fuc1→
Glc1→
Xyl1→3Xyl1→3Glcβ1→
GlcNAcβ1→3Fuc1→
Xyl1→3Xyl1→3Glcβ1→
(EGF12)


D
*
#a
#b
24 h
0 h
6 8 10 12 14 16 18 20 22 24 26
min
misfolded products (#a and #b)
or "other native folds?"
correct fold at *
in vitro "refolding"
25°C, 24 h
correctly folded EGF10 module 9
"canonical EGF fold"

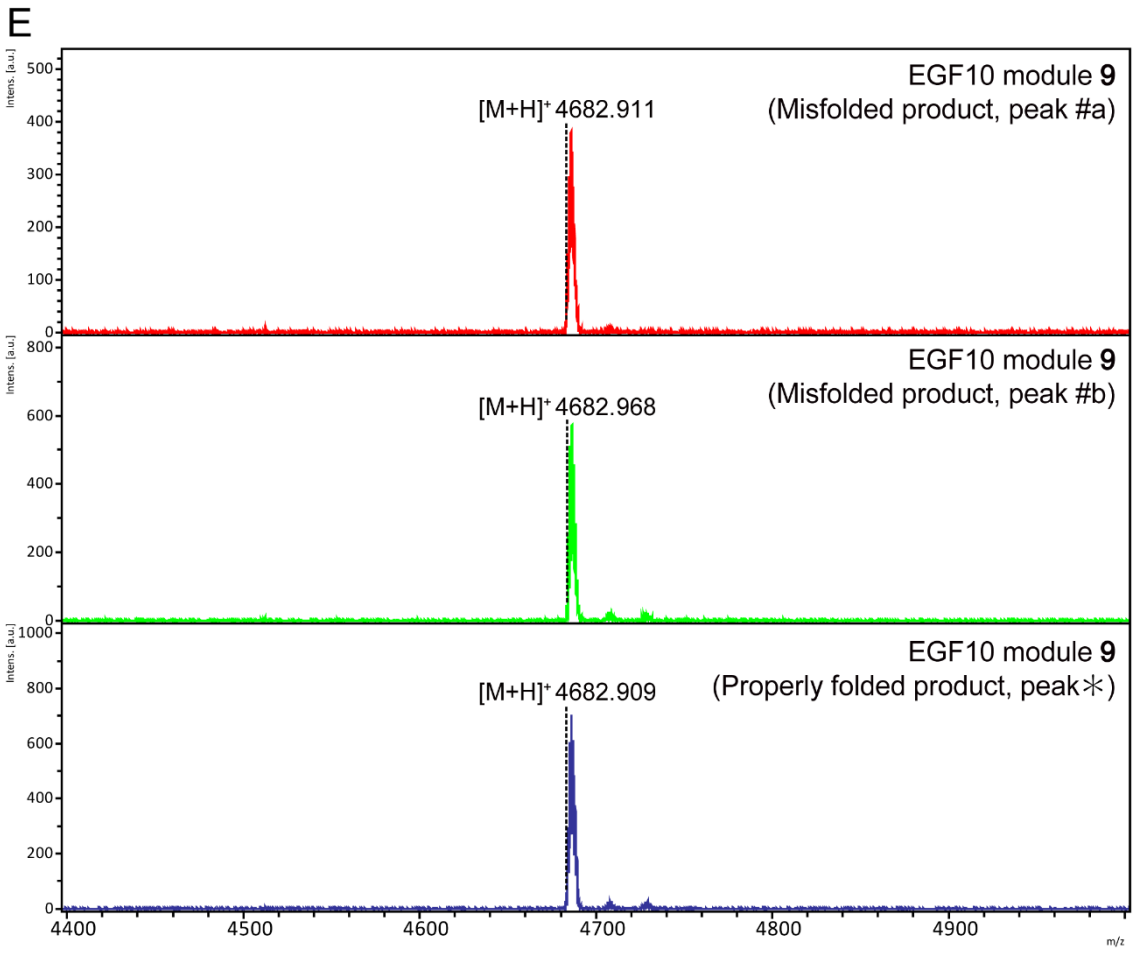

E
Intens. [a.u.]
EGF10 module 9
(Misfolded product, peak #a)
[M+H]+ 4682.911
EGF10 module 9
(Misfolded product, peak #b)
[M+H]+ 4682.968
EGF10 module 9
(Properly folded product, peak*)
[M+H]+ 4682.909
4400 4500 4600 4700 4800 4900
m/z

Figure 2-3. Glutathione-mediated oxidative folding reactions of linear EGF11 and EGF10 precursors **1**~**9**. (A) HPLC profiles of the folded products. Reactions were conducted for 24 h in 5 mM $CaCl_2$ at 25ºC or in the absence of Ca^{2+} under the general conditions described in the experimental section in Suppolementary Information. The peaks indicated by asterisk represent the fractions corresponding to the correctly folded products. (B and C) The efficiency of the properly folded EGF modules **1**~**9** and those of EGF12 modules was calculated from peak area of HPLC profiles reported in the preceding paper.[31] The yield (%) of correctly folded modules were calculated from the peak area corresponding to the product peak in the HPLC profiles. (D) HPLC profile of refolding products from the correctly folded EGF10 module **9**. The isolated pure EGF10 module **9** was re-subjected to dissolve into the general condition for *in vitro* oxidative glutathione-mediated folding reaction. Two misfolded EGF10 structures shown at peaks #a and #b were predicted according to the results identified in cases of the folding of synthetic EGF12 modules.[31] (E) MALDI-TOFMS of three major product peaks at #a, #b, and an asterisk in the HPLC profile in the refolding experiment from the correctly folded pure EGF10 module **9** shown in Figure 2-3D.

2-3. Conclusion

To investigate effect of variable *O*-glycosylation states at the distinctive sites on the *in vitro* folding process and conformation of the domains 11 and 10, we designed and synthesized domain 11 modules **1**~**4** and domain 10 modules **5**~**9** (Figure 2-1B) having possible *O*-glycosylation states generable in the endoplasmic reticulum (ER), in which the glycan biosynthesis may be conducted by the key enzymes such as POGLUT1~3,[16,17] EGF domain specific *O*-GlcNAc transferase (EOGT),[19,20] and xylosyltransferases[10] in the presence of sugar nucleotides. Figure 2-1C represents a standardized protocol for chemical synthesis of the linear peptide/glycopeptide precursors based on the microwave-assisted solid phase glycopeptide synthesis[11,12,21,22] by using key building blocks **10**~**12** having three distinct glycoforms found in domains 11 and 10. With this strategy, we synthesized efficiently all linear precursors of domains 11 and 10 modules **1**~**9** composed of 40 amino acid residues.

We succeeded in the construction of all modules **1**~**9** as canonical EGF folds while *in vitro* oxidative folding profiles and yields differ entirely between domain 11 and 10 (Figure 2-3A~C, Figures S2-1~S2-2). Surprisingly, *in vitro* folding of all precursors **1**~**4** of domain 11 proceeded efficiently and gave eventually only single major peak independent of *O*-glycosylation states, whereas the yields of these canonical EGF folds were enhanced dramatically in the presence of Ca^{2+} ion. In contrast, the folding of linear precursors **5**~**9** for domain 10 gave a complicated mixture in all experiments independent of Ca^{2+} ion. It was evident that *in vitro* glutathione-mediated oxidation of linear precursors **5**~**9** yields largely misfolded products (56.9~80.4%). These results demonstrate that the modification at Ser378 with a trisaccharide (Xylα1→3Xylα1→3Glcβ1→) reduces

equally the folding efficiency when compared with that of the non-glycosylated precursor **5** (40.4%) as shown in the folding of modules **8** (20.2%) and **9** (21.7%) while other glycosylation states do not affect the folding as shown in **6** (38.8%) and **7** (43.1%). Considering that extension of the Glcβ1→ moiety with xyloses at Ser458 also impedes folding of domain 12 (Figure 2-3B~C and Figure S2-3), it seems likely that modification of this trisaccharide moiety at serine residue involved in C1-X-S-X-(P/A)-C2 may have specific effect on the 3D structure of NOTCH ECDs.

More importantly, synthetic EGF10 module **9** provided evidence that the correctly folded pure EGF10 domain having two different *O*-glycans (GlcNAcβ1→ at Thr405 and Xylα1→3Xylα1→3Glcβ1→ at Ser378) initiates unfolding/refolding through disulfide-shuffling reactions and results in the product mixture containing 56% of two major misfolded structures (misfold at peak #a = 28%, misfold at peak #b = 28%, and correctly folded module **9** = 26%) when exposed to the general condition for *in vitro* glutathione-mediated oxidative folding experiments (Figures 2-3D and 2-3E). Obviously, the HPLC profile was mostly identical with that of the mixture after folding from the linear EGF10 precursor **9** (22%) under the same condition (Figure 2-3A), demonstrating that the folding of non-Ca^{2+}-binding EGF10 domain is regulated negatively by the site-specific glycosylation at Ser378 residue with a trisaccharide moiety (Xylα1→3Xylα1→3Glcβ1→) and leads eventually to the mixture of thermodynamically stable folded isomers.

It has become apparent that the disulfide bonds of mature proteins distributing in the human cell surface and extracellular space can be reduced in vivo to modify their functions. For example, membrane associated TRX secreted from $CD4^+$ human T cells

digests specifically one of the three disulfide bonds of CD4, leading to the enhanced affinity with gp120 of HIV through the conformational alteration.[24,25] It was also demonstrated that human β-defensin 1, an EGF-like antimicrobial peptide expressed by epithelial and hematopoietic cells, shows its antimicrobial activity in vivo only after cleavage of intramolecular disulfide linkages by membrane bound or secreted TRX.[26,27] TRX/GSH-mediated reduction of disulfide bonds may be a common mechanism to modulate matured protein functions by tuning the folding equilibria of active and inactive 3D structures. It seems likely that equilibria of folding states in ECDs may modulate cell surface avidity of "active Notch receptor" having an ability to interact with ligands. Given that X-ray crystal structures of recombinant Notch ECDs containing glycosylated domains 10-12 in complexed with ligands are identified as the canonical EGF folds,[6,7] it is thought that misfolded domain may impede ligand binding of Notch ECDs.

Our results provide a novel principle that GSH-mediated *in vitro* folding/refolding of NOTCH1 EGF-like repeats occurs in the domain-specific manner, in which an equilibrium of the 3D structures formed in each domain might depend on their thermodynamic stabilities during disulfide shuffling reactions under a reductant condition. It is important to note that misfolded structures generated through site-specific *O*-glycosylation states found in the domains 12 and 10 can be incorporated into the NOTCH1 ECD expressed on the cell surface. Given that 6 Ca^{2+}-binding EGF-like domains, notably domains 9, 16, 20, 21, 27, and 31, are also modified concurrently by *O*-Glc- and *O*-Fuc-initiated glycans at the same consensus sequences with those of domain 12, it is possible to predict that folding of these domains occurs in a mechanism similar to that of domain 12,[23] leading efficiently to the canonical EGF folds when the two

potential glycosylation sites are fully glycosylated. By contrast, it seems likely that folding of the non-Ca^{2+}-binding domains having a trisaccharide Xylα1→3Xylα1→3Glcβ1→ moiety in the consensus sequence C1-X-S-X-(P/A)-C2 such as domains 4, 10, 28, and 33[23] may give rise to misfolded structures in addition to the properly folded EGF-like structures as observed in the modules **8** and **9** of domain 10. Furthermore, the sequence alignment analysis by using CLUSTAL W uncovers that the Ca^{2+}-binding domains 8, 23, and 27 exhibit distinctively higher sequence similarity with domain 11 than other 32 NOTCH1 ECDs, suggesting that these domains also form dominantly the canonical EGF folds without influence of glycosylation. It is interesting to hypothesize that regions containing these domains may also have unidentified important roles in determining Notch signalling pathway because the domain 11 positioned in the central region of domains 10-13 contributes strongly to the integrity of a pivotal binding interface with some Notch ligands.[7, 28] However, we need to provide additional evidence that supports this principle by determining the structural and biochemical properties of other key EGF-like domains having characteristic *O*-glycosylation states as well as domains 10, 11, and 12.

2-4. Experimental Section

General methods and materials

All commercially available solvents and reagents were used without further purification. Rink Amide-ChemMatrix® resin was purchased from Biotage Japan Ltd. (Tokyo, Japan). N^{α}-Fmoc-amino acid derivatives except for glycosylated amino acids were purchased from Merck Millipore (Darmstadt, Germany). Fmoc-Thr(Ac_3GlcNAcβ1→)-OH was purchased from Medicinal Chemistry Pharmaceuticals, Co., Ltd. (Sapporo, Japan). Fmoc-Ser(Ac_4Glcβ1→)-OH and Fmoc-Ser[Ac_3Xylα(1→3)Ac_2Xylα(1→3)Ac_3Glcβ1→]-OH were synthesized according to the method reported previously.[11,22] 1-[Bis(dimethylamino)methyliumyl]-1*H*-benzotriazole-3-oxide hexafluorophosphate (HBTU), 1-hydroxybenzotriazole monohydrate (HOBt), *N*,*N*-diisopropylethylamine (DIEA), *N*,*N*-dimethylformamide (DMF) and 2,2,2-trifluoroacetic acid (TFA) were purchased from Watanabe Chemical Industries, Ltd. (Hiroshima, Japan). All mixing operation in peptide synthesis was performed by a vortex mixer. All solid-phase reactions for (glyco)peptides synthesis were performed manually in a polypropylene tube equipped with a filter (LibraTube®, Hipep Laboratories) under microwave irradiation. The microwave was irradiated during coupling reactions and Fmoc removal using temperature control at 50°C. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) were recorded with a Bruker UltraFlex III mass spectrometer in reflector positive or linear positive mode using matrix as 2,5-dihydroxybenzoic acid (DHB). Typically, the samples were dissolved in 1μL of 50% (*v*/*v*) aqueous acetonitrile and mixed with the same volume of 10 mg/mL DHB in 50% (*v*/*v*) aqueous acetonitrile containing 0.1% TFA. All ^{1}H- and ^{13}C-NMR spectra for identification of synthetic peptides were collected with 600 MHz Bruker AVANCE

(Bruker Biospin Co., Germany). Two-dimensional homonuclear double-quantum-filtered scalar-correlated spectroscopy (DQF-COSY), TOCSY with MLEV-17 sequence, and nuclear Overhauser enhancement spentroscopy (NOESY) spectra were recorded in the indirect dimension using State-TPPI phases cycling. Reverse-phase HPLC (RP-HPLC) separation and analysis were performed on a Hitachi system equipped with L-6250 intelligent pump and L-7400 UV detector using a reverse-phase C18 column Inertsil ODS-3 250×20 mmI.D. (GL Sciences Inc., Tokyo, Japan) or on a Hitachi system equipped with L-7100 pump and L-7405 UV detector using reverse-phase C18 column Inertsil ODS-3 250×4.6 mmI.D. (GL Sciences Inc., Tokyo, Japan).

Solid-phase synthesis of human NOTCH1 EGF11 and EGF10 modules

Novel EGF11 modules **1**~**4** and EGF10 modules **5**~**9** were synthesized as linear (glyco)polypeptides using sugar amino acid derivatives **10**~**12** according to the general procedure for the synthesis of EGF12 modules reported previously.[11,12,21] Rink Amide-ChemMatrix resin (0.45 mmol/g, 11.3 μmol) was swollen with dichloromethane (DCM) at ambient temperature for 1 h. The Fmoc group was removed with 20% piperidine in DMF (0.5 mL) for 3 min under microwave irradiation. *N*α-Fmoc-amino acid (4.0 equiv.) was coupled with 1-[bis(dimethylamino)methyliumyl]-1*H*-benzotriazole-3-oxide hexafluorophosphate (HBTU) (4.0 equiv.), 1-hydroxybenzotriazole monohydrate (HOBt) (4.0 equiv.), and *N*,*N*-diisopropylethylamine (DIEA) (6.0 equiv.) in DMF (113 μL) for 10 min under microwave irradiation. In the case of glycoamino acids, Fmoc-Thr(Ac_3GlcNAcβ1→)-OH (**10**) at Thr445 or Thr405 (1.2 equiv.), Fmoc-Ser(Ac_4Glcα1→)-OH (**11**) at Ser435 or Ser378 (1.2 equiv.), and Fmoc-Ser[Ac_3Xylα(1→3)Ac_2Xylα(1→3)Ac_3Glcβ1→]-OH (**12**) at Ser378 were coupled with HBTU (1.2 equiv.), HOBt (1.2

equiv.), and DIEA (3.0 equiv.) in DMF (34 μL) for 10 min under microwave irradiation. At the end of solid-phase synthesis, an *N*-terminal amino group of peptidyl-resin was capped with acetic anhydride with DIEA in DMF for 1 min at ambient temperature. For cleavage of (glyco)peptides from the resin and removal of acid-labile protective groups, a cocktail of TFA/ethanedithiol/H_2O/triisopropylsilane (94:2.5:2.5:1, *v*/*v*/*v*/*v*, 1 mL) was subjected to treatment with resin at ambient temperature for 2 h. The resin was filtered off and the peptide was precipitated using mixture solution of *tert*-butylmethyl ether (15 mL) and hexane (15 mL) in ice bath without removal of TFA solution. After centrifugation (3000 rpm at 4 °C for 15 min), the supernatant was carefully removed by decantation and the process was repeated three times. The precipitate was dissolved in 50% aqueous acetonitrile (2.5 mL) and then lyophilized. To deprotect the *O*-acetyl group of the sugar moiety, the dried precipitate was dissolved in methanol containing 6.5 mM of dithiothreitol (DTT) and the solution was adjusted and kept to pH12.5 with a 1 N NaOH aqueous solution. The reaction mixture was stirred for 1 h. In the case of glycopeptides having trisaccharide moiety Xylα1→3Xylα1→3Glcβ1→ (**8** and **9**), reaction mixture was stirred for 24 h in the same fashion. After being stirred at ambient temperature, the solution was neutralized with 30% acetic acid in methanol and the solvent was removed by rotary evaporator. Finally, the residual material was subjected to the purification using RP-HPLC to remove excess of DTT.

Disulfide bond formation of EGF11 and EGF10 modules under an optimized condition to obtain the canonical EGF folds

Linear EGF11 and EGF10 precursors **1**~**9** were dissolved in redox buffer (0.1 mg/mL) containing 50 mM Tris-HCl (pH 8.0), 1 mM reduced glutathione and 0.2 mM oxidized

glutathione. In the case of EGF11 precursors **1**~**4**, reaction was performed in the presence of 5 mM $CaCl_2$. The reaction mixtures were routinely stirred at 4°C for 24 h. Reaction was monitored by analytical RP-HPLC and the reaction mixture was finally acidified by addition of TFA (1% volume of reaction mixture), then lyophilized and subjected to RP-HPLC for purification.

Purification of synthetic EGF11 and EGF10 modules

Synthetic EGF11 modules **1**~**4** and EGF10 modules **5**~**9** were purified and isolated by using a Hitachi system equipped with L-6250 intelligent pump and L-7400 UV detector using a reverse-phase C18 column Inertsil ODS-3 250×20 mmI.D. Conditions for the purification by RP-HPLC of all EGF modules are as follows: ambient temperature and flow rate of 5 mL/min, UV detection at 220 nm, eluent A: 0.1% TFA in water, eluent B: 0.1% TFA in acetonitrile, and the linear gradient from 10 to 40% B over 50 min for all analogues. Analytical runs were performed on a Hitachi system equipped with L-7100 pump and L-7405 UV detector using reverse-phase C18 column Inertsil ODS-3 250×4.6 mmI.D. Conditions for the analysis by RP-HPLC of all EGF modules are as follows: ambient temperature and flow rate of 1 mL/min, UV detection at 220 nm, eluent A: 0.1% TFA in water, eluent B: 0.1% TFA in acetonitrile, and the linear gradient from 20 to 30% B over 30 min for EGF11 modules **1**~**4**, and the linear gradient from 18 to 28% B over 30 min for EGF10 modules **5**~**9**. Purification by RP-HPLC gave module **1** in 8.9% (4.3 mg, 1.0 μmol, t_R = 18.60 min), module **2** in 7.9% (4.0 mg, 0.9 μmol, t_R = 15.47 min), module **3** in 8.2% (4.1 mg, 0.9 μmol, t_R = 14.07 min), module **4** in 8.6% (4.5 mg, 1.0 μmol, t_R = 11.61 min), module **5** in 5.5% (2.5 mg, 0.6 μmol, t_R = 22.35 min), module **6** in 5.6% (2.7 mg, 0.6 μmol, t_R = 20.27 min), module **7** in 5.5% (2.6 mg, 0.6 μmol, t_R =

20.53 min), module **8** in 4.2% (2.1 mg, 0.5 μmol, t_R = 20.58 min), and module **9** in 3.4% (1.8 mg, 0.4 μmol, t_R = 18.54 min), respectively. MALDI-TOFMS for module **1**: $C_{177}H_{271}N_{51}O_{62}S_6$ $[M+H]^+$ calcd (*m*/*z*) 4295.8023, found (*m*/*z*) 4296.705; **2**: $C_{185}H_{284}N_{52}O_{67}S_6$ $[M+H]^+$ calcd (*m*/*z*) 4498.8817, found (*m*/*z*) 4499.701; **3**: $C_{183}H_{281}N_{51}O_{67}S_6$ $[M+H]^+$ calcd (*m*/*z*) 4457.8551, found (*m*/*z*) 4457.915; **4**: $C_{191}H_{294}N_{52}O_{72}S_6$ $[M+H]^+$ calcd (*m*/*z*) 4660.9345, found (*m*/*z*) 4660.916; **5**: $C_{161}H_{249}N_{49}O_{62}S_6$ $[M+H]^+$ calcd (*m*/*z*) 4053.6240, found (*m*/*z*) 4053.627; **6**: $C_{169}H_{262}N_{50}O_{67}S_6$ $[M+H]^+$ calcd (*m*/*z*) 4256.7034, found (*m*/*z*) 4256.822; **7**: $C_{167}H_{259}N_{49}O_{67}S_6$ $[M+H]^+$ calcd (*m*/*z*) 4215.6768, found (*m*/*z*) 4215.807; **8**: $C_{177}H_{275}N_{49}O_{75}S_6$ $[M+H]^+$ calcd (*m*/*z*) 4479.7614, found (*m*/*z*) 4479.861; **9**: $C_{185}H_{288}N_{50}O_{80}S_6$ $[M+H]^+$ calcd (*m*/*z*) 4682.8407, found (*m*/*z*) 4682.941. HPLC profiles and MALDI-TOFMS spectra of modules **1**~**9** were shown in Figures S2-6 and S2-7.

Refolding from the canonically folded EGF10 module under a general condition for in vitro redox folding system

Glycosylated EGF module **9** was dissolved in the redox buffer (0.1 mg/mL) containing 50 mM Tris-HCl (pH 8.0), 1 mM reduced glutathione, and 0.2 mM oxidized glutathione. The reaction mixture was kept for 24 h at 25°C, added with 1% TFA (v/v), and lyophilized. The products dissolved in the buffered solution was subjected to the analysis by RP-HPLC.

2-5. Reference

(1) Rebay, I.; Fleming, R. J.; Fehon, R. G.; Cherbas, L.; Cherbas, P.; Artavanis-Tsakonas, S. Specific EGF Repeats of Notch Mediate Interactions with Delta and Serrate: Implications for Notch as a Multifunctional Receptor. *Cell* **1991**, *67*, 687–699.

(2) Xu, A.; Lei, L.; Irvine, K. D. Regions of Drosophila Notch That Contribute to Ligand Binding and the Modulatory Influence of Fringe. *J. Biol. Chem.* **2005**, 280, 30158–30165.

(3) Brückner, K.; Perez, L.; Clausen, H.; Cohen, S. Glycosyltransferase Activity of Fringe Modulates Notch-Delta Interactions. *Nature* **2000**, *406*, 411–415.

(4) Moloney, D. J.; Panin, V. M.; Johnston, S. H.; Chen, J.; Shao, L.; Wilson, R.; Wang, Y.; Stanley, P.; Irvine, K. D.; Haltiwanger, R. S.; Vogt, T. F. Fringe Is a Glycosyltransferase That Modifies Notch. *Nature* **2000**, *406*, 369–375.

(5) Taylor, P.; Takeuchi, H.; Sheppard, D.; Chillakuri, C.; Lea, S. M.; Haltiwanger, R. S.; Handford, P. A. Fringe-Mediated Extension of *O*-Linked Fucose in the Ligand-Binding Region of Notch1 Increases Binding to Mammalian Notch Ligands. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 7290–7295.

(6) Luca, V. C.; Jude, K. M.; Pierce, N. W.; Nachury, M. V.; Fischer, S.; Garcia, K. C. Structural basis for Notch1 engagement of Delta-like 4. *Science* **2015**, *347*, 847-853.

(7) Luca, V. C.; Kim, B. C.; Ge, C.; Kakuda, S.; Wu, D.; Roein-Peikar, M.; Haltiwanger, R. S.; Zhu, C.; Ha, T.; Garcia, K. C. Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity. *Science* **2017**, *355*, 1320-1324.

(8) Fernandez-Valdivia, R.; Takeuchi, H.; Samarghandi, A.; Lopez, M.; Leonardi, J.; Haltiwanger, R. S.; Jafar-Nejad, H. Regulation of Mammalian Notch Signaling and Embryonic Development by the Protein *O*-Glucosyltransferase Rumi. *Development* **2011**, *138*, 1925–1934.

(9) Acar, M.; Jafar-Nejad, H.; Takeuchi, H.; Rajan, A.; Ibrani, D.; Rana, N. A.; Pan, H.; Haltiwanger, R. S.; Bellen, H. J. Rumi is a CAP10 domain glycosyltransferase that modifies Notch and is required for Notch signaling.

Cell **2008**, *132*, 247-258.

(10) Lee, T. V.; Sethi, M. K.; Leonardi, J.; Rana, N. a.; Buettner, F. F. R.; Haltiwanger, R. S.; Bakker, H.; Jafar-Nejad, H. Negative Regulation of Notch Signaling by Xylose. *PLoS Genet.* **2013**, *9*, e1003547.

(11) Hayakawa, S.; Koide, R.; Hinou, H.; Nishimura, S.-I. Synthetic Human NOTCH1 EGF Modules Unraveled Molecular Mechanisms for the Structural and Functional Roles of Calcium Ions and *O*-Glycans in the Ligand-Binding Region. *Biochemistry* **2016**, *55*, 776-787.

(12) Hayakawa, S.; Yokoi, Y.; Hinou, H.; Nishimura, S.-I. Chemical Synthesis Demonstrates That Dynamic *O*-Glycosylation Regulates the Folding and Functional Conformation of a Pivotal EGF12 Domain of the Human NOTCH1 Receptor. *Biochemistry* **2017**, *56*, 4379-4391.

(13) Fernandez-Valdivia, R.; Takeuchi, H.; Samarghandi, A.; Lopez, M.; Leonardi, J.; Haltiwanger, R. S.; Jafar-Nejad, H. Regulation of Mammalian Notch Signaling and Embryonic Development by the Protein *O*-Glucosyltransferase Rumi. *Development* **2011**, *138*, 1925–1934.

(14) Rubartelli, A.; Bajetto, A.; Allavena, G.; Wollman, E.; Sitia, R. Secretion of thioredoxin by normal and neoplastic cells through a leaderless secretory pathway. *J. Biol. Chem.* **1992**, *267*, 24161-24164.

(15) Arnér, E. S.; Holmgren, A. Physiological functions of thioredoxin and thioredoxin reductase. *Eur. J. Biochem.* **2000**, *267*, 6102-6109.

(16) Alfaro, J. F.; Gong, C.; Monroe, M. E.; Aldrich, J. T.; Clauss, T. R. W.; Purvine, S. O.; Wang, Z.; Camp, D. G.; Shabanowitz, J.; Stanley, P.; Hart, G. W.; Hunt, D. F.; Yang, F.; Smith, R. D. Tandem Mass Spectrometry Identifies Many Mouse Brain *O*-GlcNAcylated Proteins Including EGF Domain-Specific *O*-GlcNAc Transferase Targets. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 7280-7285.

(17) Takeuchi, H.; Schneider, M.; Williamson, D. B.; Ito, A.; Takeuchi, M.; Handford, P. A.; Haltiwanger, R. S. Two novel protein *O*-glucosyltransferases that modify sites distinct from POGLUT1 and affect Notch trafficking and signaling. *Proc. Natl. Acad. Sci. U S A.* **2018**, *115*, E8395-E8402.

(18) Rana, N. A.; Nita-Lazar, A.; Takeuchi, H.; Kakuda, S.; Luther, K. B.; Haltiwanger, R. S. *O*-Glucose Trisaccharide Is Present at High but Variable Stoichiometry at Multiple Sites on Mouse Notch1. *J. Biol. Chem.* **2011**, *286*, 31623-31637.

(19) Matsuura, A.; Ito, M.; Sakaidani, Y.; Kondo, T.; Murakami, K.; Furukawa, K.; Nadano, D.; Matsuda, T.; Okajima, T. *O*-Linked *N*-Acetylglucosamine Is Present on the Extracellular Domain of Notch Receptors. *J. Biol. Chem.* **2008**, *283*, 35486-35495.

(20) Ogawa, M.; Okajima, T. Structure and function of extracellular *O*-GlcNAc. *Curr. Opin. Struct. Biol.* **2019**, *56*, 72-77.

(21) Hiruma-Shimizu, K.; Hosoguchi, K.; Liu, Y.; Fujitani, N.; Ohta, T.; Hinou, H.; Matsushita, T.; Shimizu, H.; Feizi, T.; Nishimura, S. -I. Chemical Synthesis, Folding, and Structural Insights into O-Fucosylated Epidermal Growth Factor-like Repeat 12 of Mouse Notch-1 Receptor. *J. Am. Chem. Soc*. **2010** *132*, 14857-14865.

(22) Matsushita, T.; Hinou, H.; Kurogochi, M.; Shimizu, H.; Nishimura, S. I. Rapid Microwave-Assisted Solid-Phase Glycopeptide Synthesis. *Org. Lett.* **2005**, *7*, 877-880.

(23) Takeuchi, H.; Haltiwanger, R. S. Significance of glycosylation in Notch signaling. *Biochem. Biophys. Res. Commun.* **2014**, *453*, 235-242.

(24) Matthias, L. J.; Yam, P. T.; Jiang, X. M.; Vandegraaff, N.; Li, P.; Poumbourios, P.; Donoghue, N.; Hogg, P. J. Disulfide exchange in domain 2 of CD4 is required for entry of HIV-1. *Nat. Immunol.* **2002**, *3*, 727-732.

(25) Azimi, I.; Matthias, L. J.; Center, R. J.; Wong, J. W. H.; Hogg, P. J. Disulfide bond that constrains the HIV-1 gp120 V3 domain is cleaved by thioredoxin. *J. Biol. Chem.* **2010**, *285*, 40072-40080.

(26) Schroeder, B. O.; Wu, Z.; Nuding, S.; Groscurth, S.; Marcinowski, M.; Beisner, J.; Buchner, J.; Schaller, M.; Stange, E. F.; Wehkamp, J. Reduction of disulphide bonds unmasks potent antimicrobial activity of human β-defensin 1. *Nature* **2011**, *469*, 419-423.

(27) Jaeger, S. U.; Schroeder, B. O.; Meyer-Hoffert, U.; Courth, L.; Fehr, S. N.;

Gersemann, M.; Stange, E. F.; Wehkamp, J. Cell-mediated reduction of human β-defensin 1: a major role for mucosal thioredoxin. *Mucosal Immunol.* **2013**, *6*, 1179-1190.

(28) Weisshuhn, P. C.; Sheppard, D.; Taylor, P.; Whiteman, P.; Lea, S. M.; Handford, P. A.; Redfield, C. Non-Linear and Flexible Regions of the Human Notch1 Extracellular Domain Revealed by High-Resolution Structural Studies. *Structure* **2016**, *24*, 555-566.

2-6. Supporting Information

Figure S2-1. HPLC profiles of the folding of EGF11 modules. The mixtures after glutathione-mediated oxidative folding reactions of naked (non-glycosylated) linear EGF11 were subjected to HPLC analysis. Reactions were conducted under the following conditions for 24 h: 5 mM $CaCl_2$ at 4 ºC (magenta), 4 ºC in the absence of calcium ion (blue), 5 mM $CaCl_2$ at room temperature (green), and room temperature in the absence of calcium ion (black). The peak denoted with asterisk show the fractions containing properly folded EGF module.

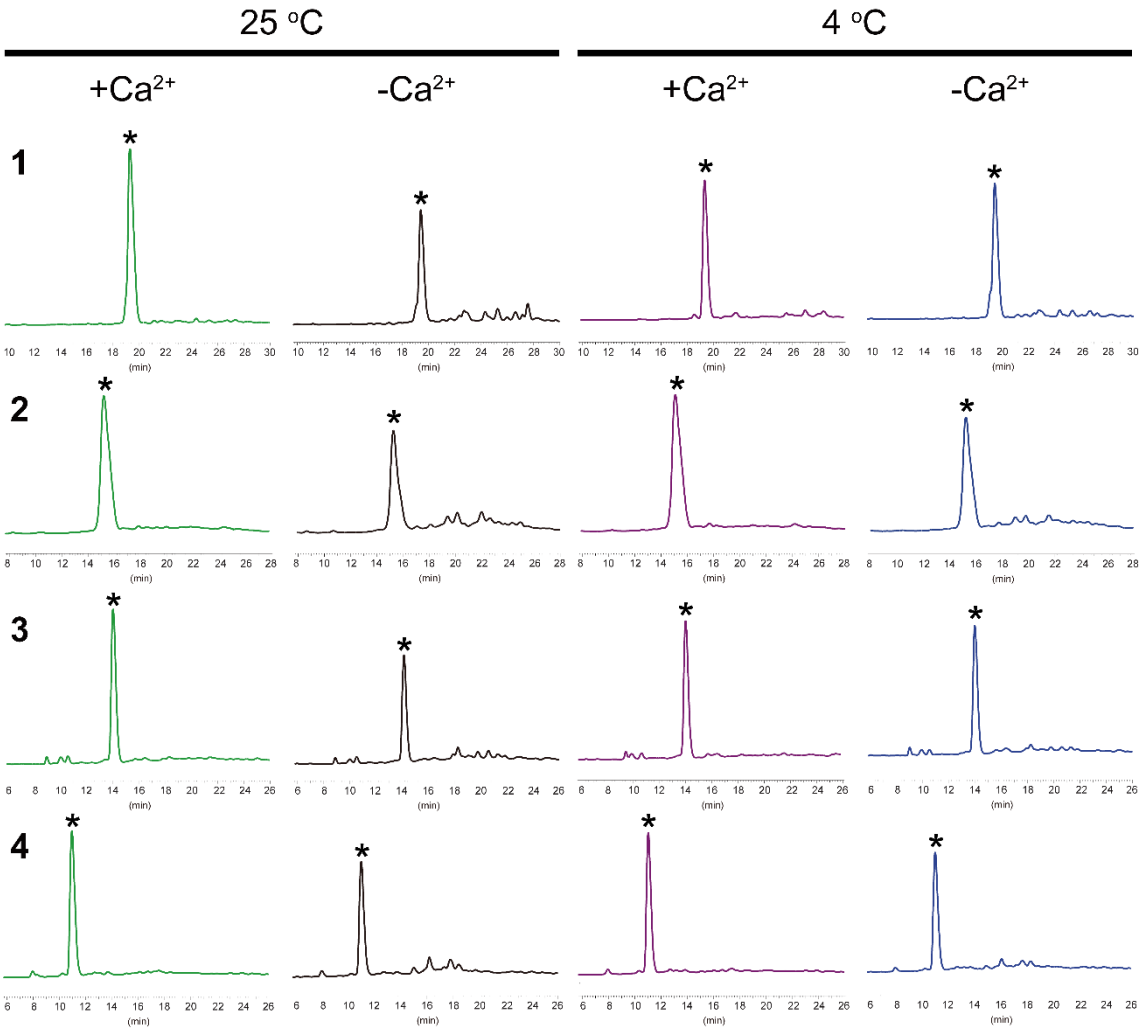


Figure S2-2. HPLC profiles of the folding of EGF10 modules. The mixtures after glutathione-mediated oxidative folding reactions of naked (non-glycosylated) linear EGF10 were subjected to HPLC analysis. Reactions were conducted under the following conditions for 24 h: 5 mM $CaCl_2$ at 4 ºC (magenta), 4 ºC in the absence of calcium ion (blue), 5 mM $CaCl_2$ at room temperature (green), and room temperature in the absence of calcium ion (black). The peak denoted with asterisk show the fractions containing properly folded EGF module.

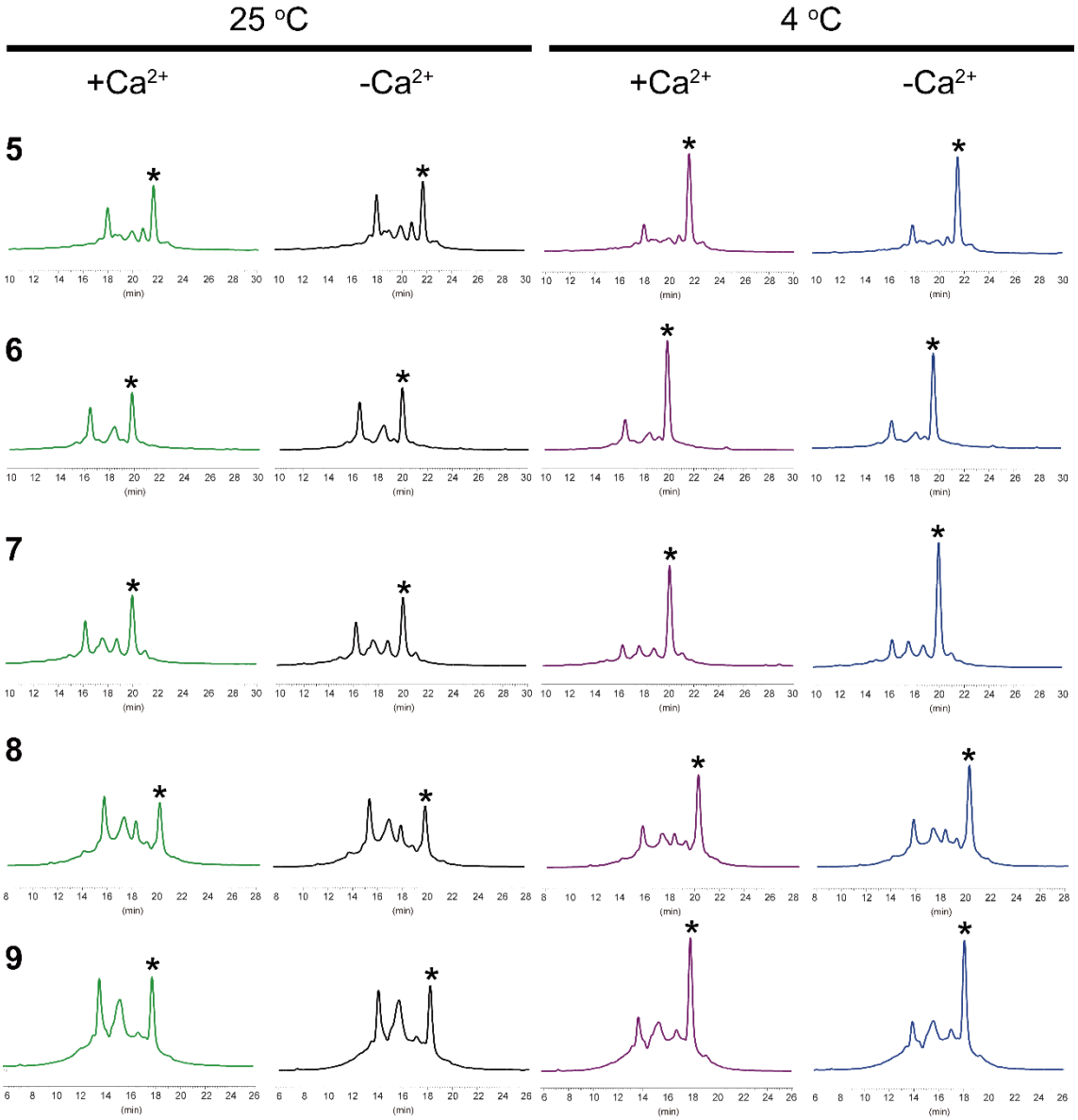


Figure S2-3. HPLC profiles of the folding of EGF12 modules. The mixtures after glutathione-mediated oxidative folding reactions of naked (non-glycosylated) linear EGF12 were subjected to HPLC analysis. Reactions were conducted under the following conditions for 24 h: 5 mM $CaCl_2$ at 4 ºC (magenta), 4 ºC in the absence of calcium ion (blue), 5 mM $CaCl_2$ at room temperature (green), and room temperature in the absence of calcium ion (black). The peak denoted with asterisk show the fractions containing properly folded EGF module. These results were cited from our previous paper; Hayakawa, S. *et al.*, *Biochemistry* **2017**, *56*, 4379-4391.

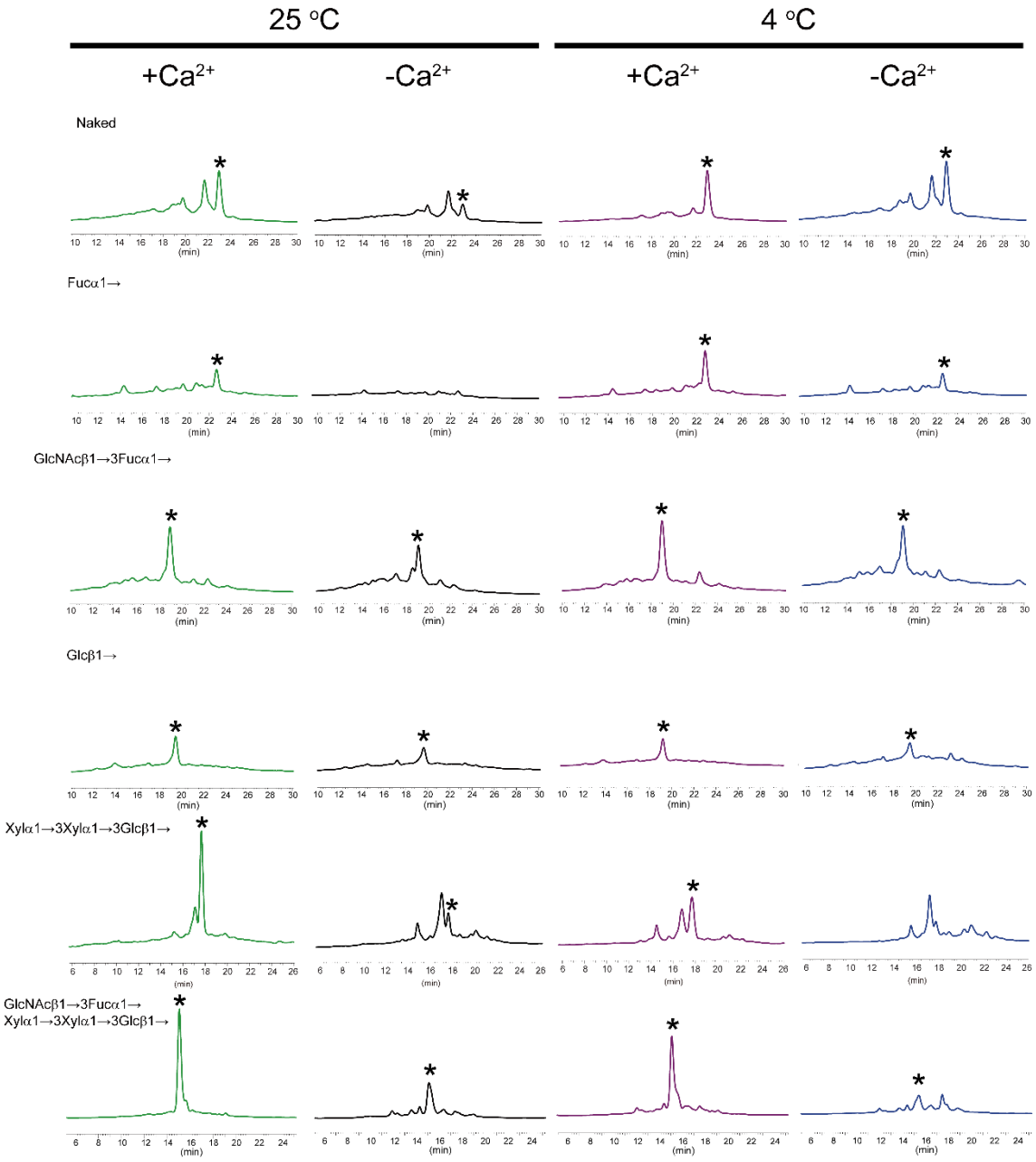


Figure S2-4. Canonical EGF folds in EGF10, 11, and 12. The ratio of properly folded product peak calculated by peak area of HPLC profiles. The results of EGF12 modules were cited and used after the arrangement from our previous paper; Hayakawa, S. *et al.*, *Biochemistry* **2017**, *56*, 4379-4391.

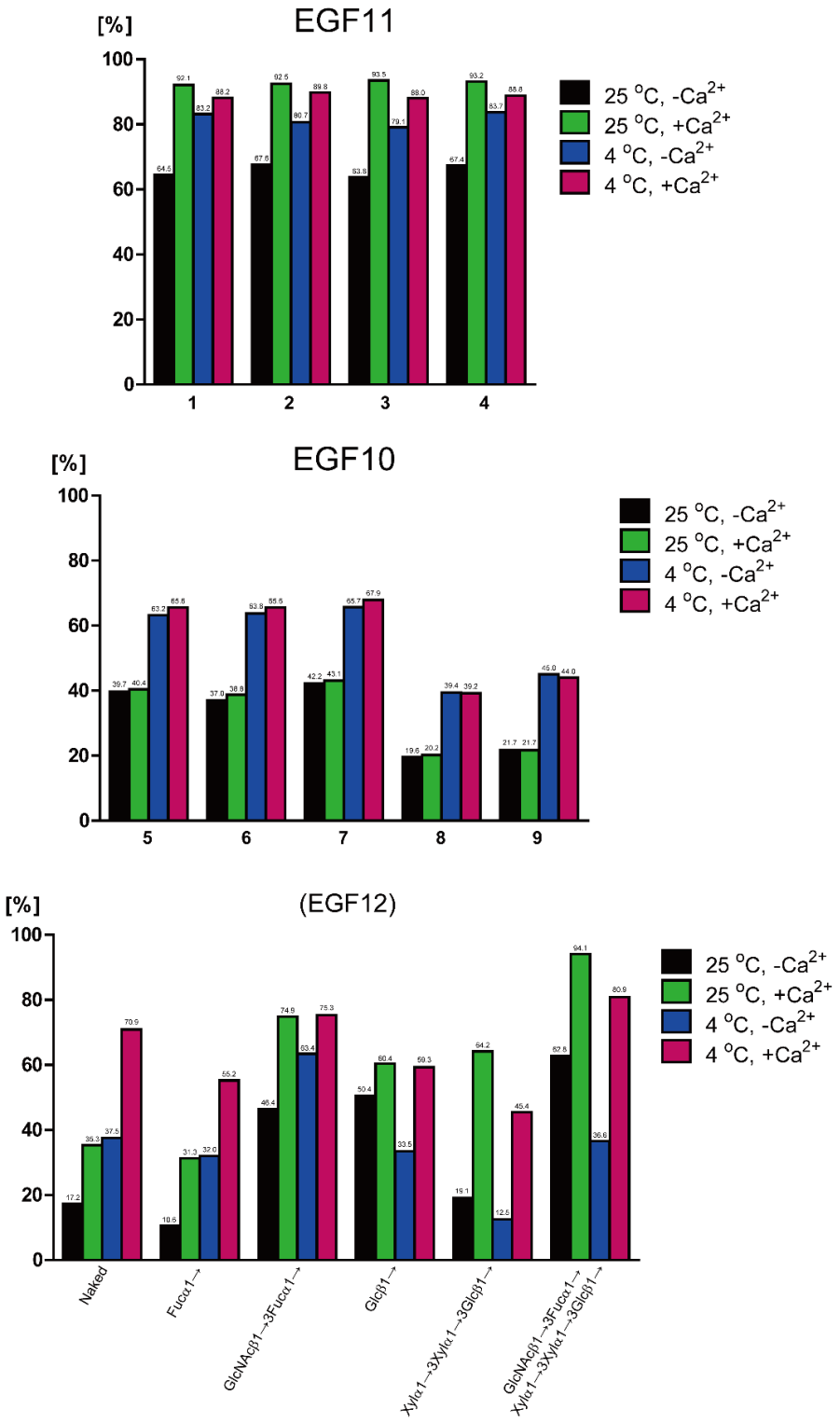
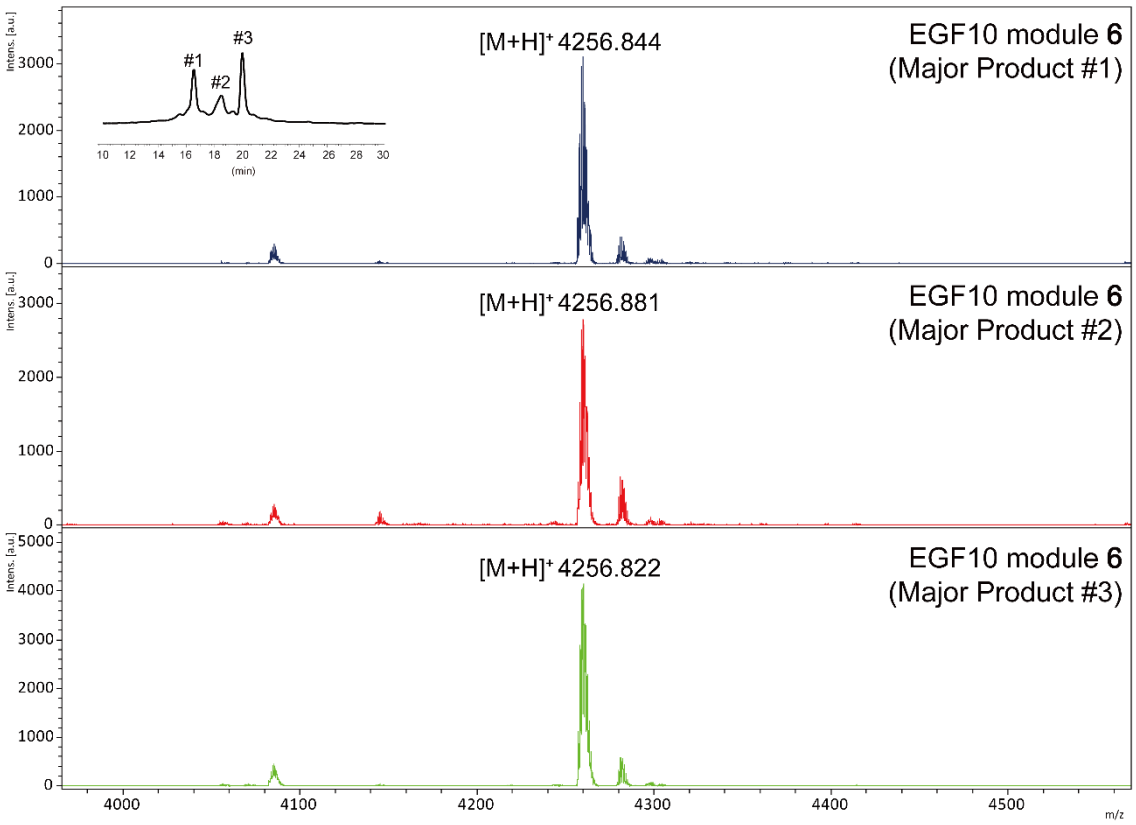
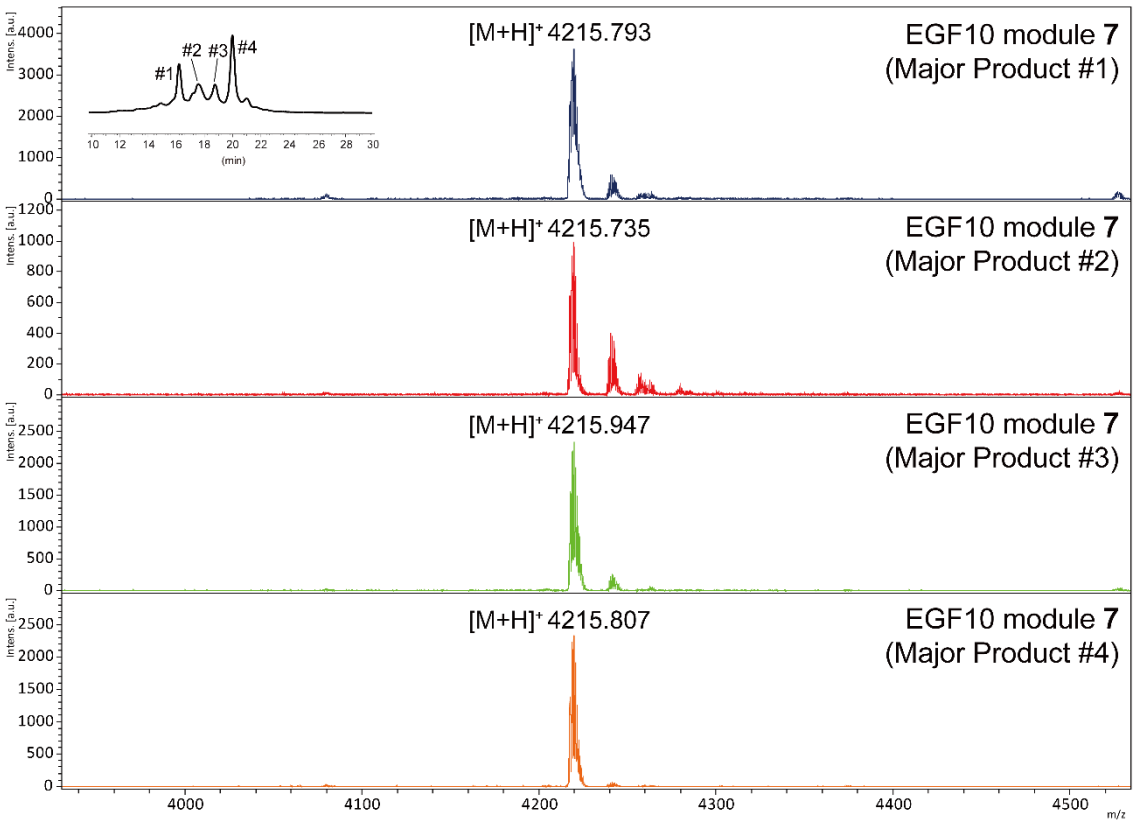


Figure S2-5. Noncanonical EGF folds in EGF10. MALDI-TOFMS spectra of the major products of EGF10 module **6**-**9**.

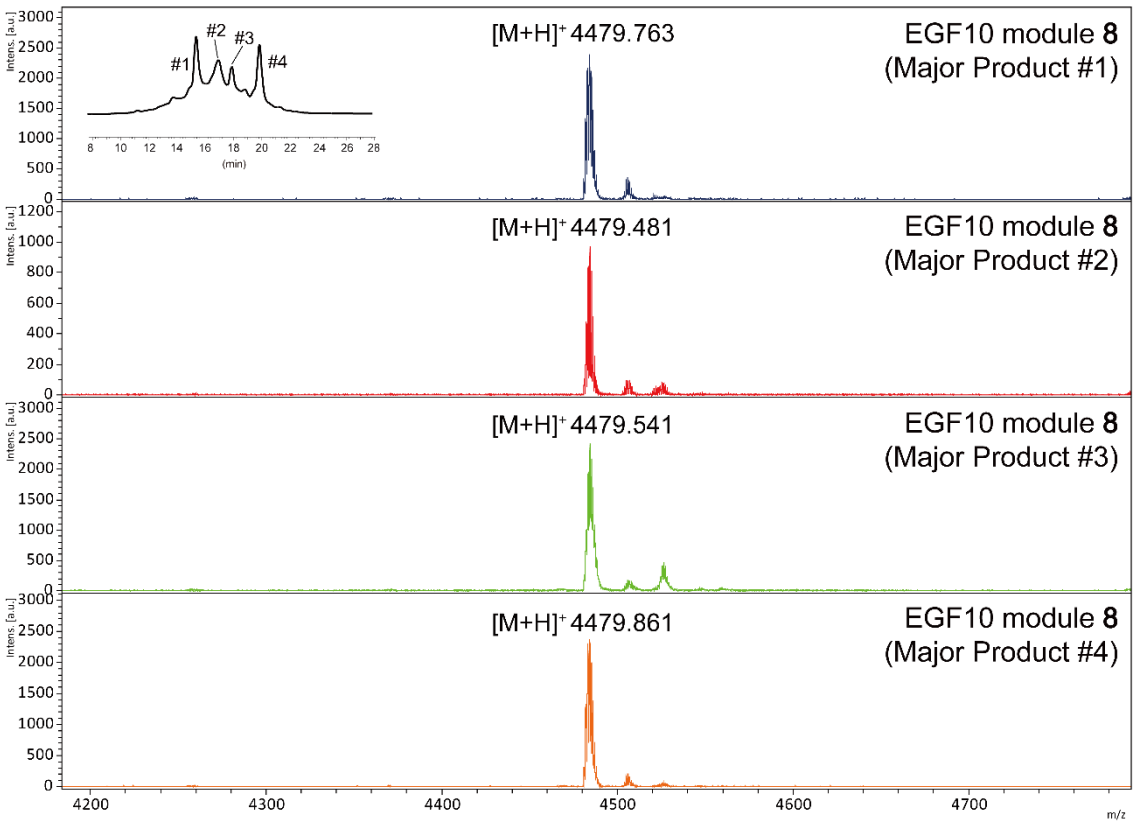
Module **6**.



Module **7**.


[M+H]+ 4215.793
EGF10 module 7
(Major Product #1)
[M+H]+ 4215.735
EGF10 module 7
(Major Product #2)
[M+H]+ 4215.947
EGF10 module 7
(Major Product #3)
[M+H]+ 4215.807
EGF10 module 7
(Major Product #4)
#1
#2
#3
#4
(min)
intens. [a.u.]
m/z

Module **8**.


#1
#2
#3
#4
(min)
[M+H]+ 4479.763
EGF10 module 8
(Major Product #1)
[M+H]+ 4479.481
EGF10 module 8
(Major Product #2)
[M+H]+ 4479.541
EGF10 module 8
(Major Product #3)
[M+H]+ 4479.861
EGF10 module 8
(Major Product #4)
Intens. [a.u.]
m/z

Module **9**.

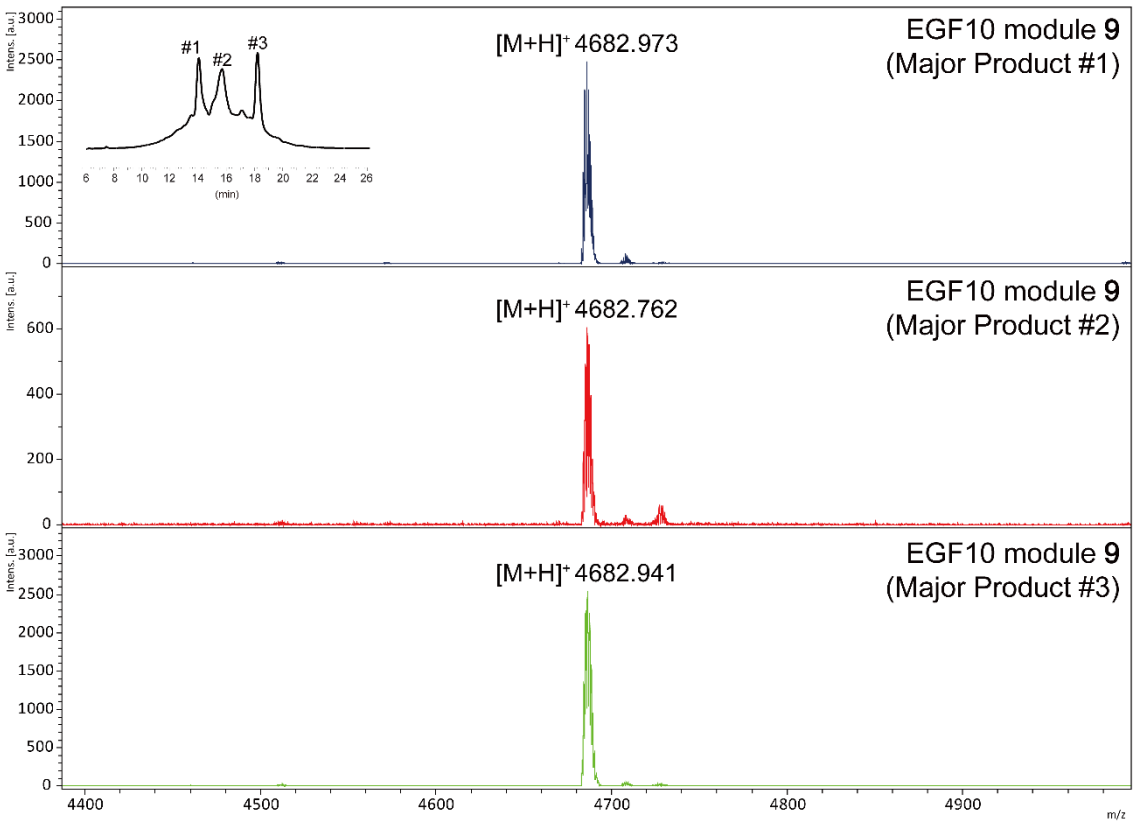
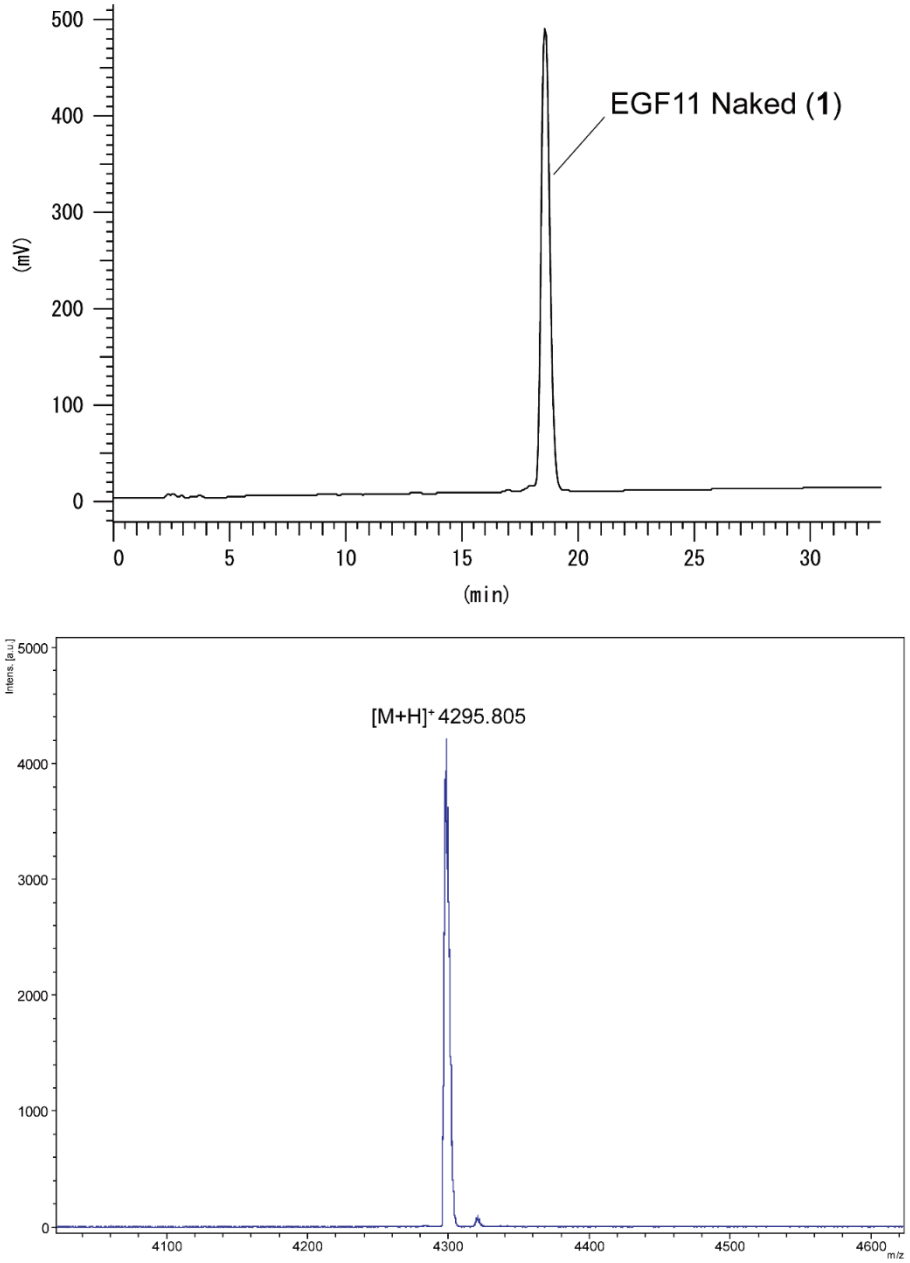
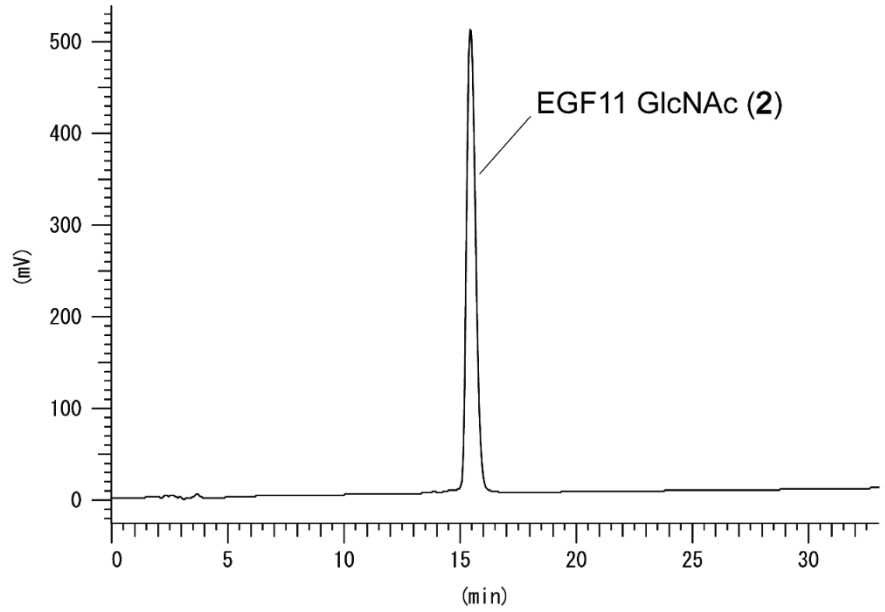

EGF10 module 9
(Major Product #1)
[M+H]+ 4682.973
#1
#2
#3
(min)
EGF10 module 9
(Major Product #2)
[M+H]+ 4682.762
EGF10 module 9
(Major Product #3)
[M+H]+ 4682.941
Intens. [a.u.]
m/z

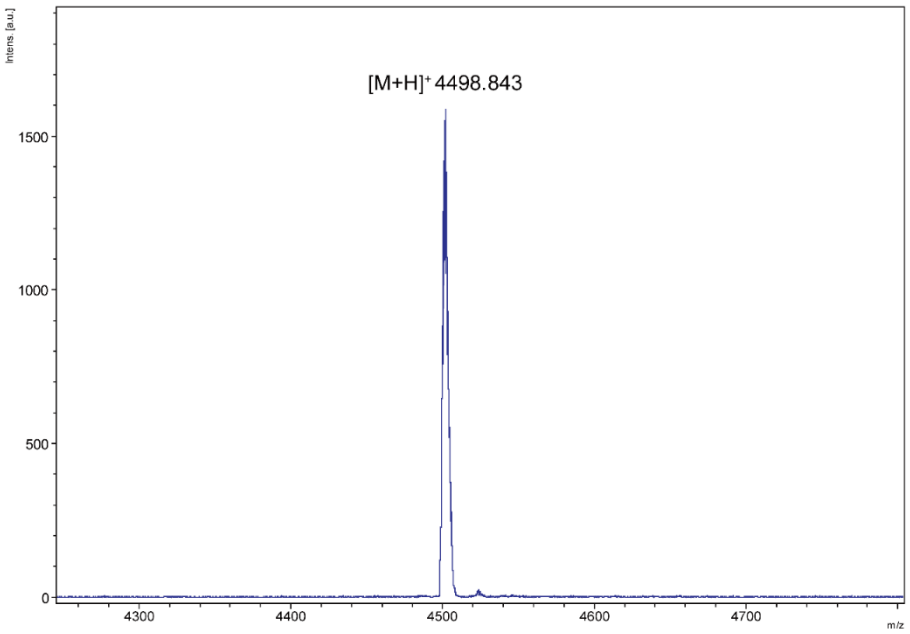
Figure S2-6. Characterization of isolated EGF11 modules. Analytical RP-HPLC (upper), MALDI-TOF MS spectra (bottom) of purified NOTCH1 EGF11 module **1-4**. Column: Inertsil ODS-3 4.5 μm (250×4.6 mmI.D.), Eluent A: water containing 0.1% TFA, Eluent B: acetonitrile containing 0.1% TFA. Gradient condition; eluent (A/B = 80/20) was employed, then the ratio of eluent B was increased lineally from 20% to 30% over 30 min with a flow rate of 1.0 mL/min.

Module **1**.

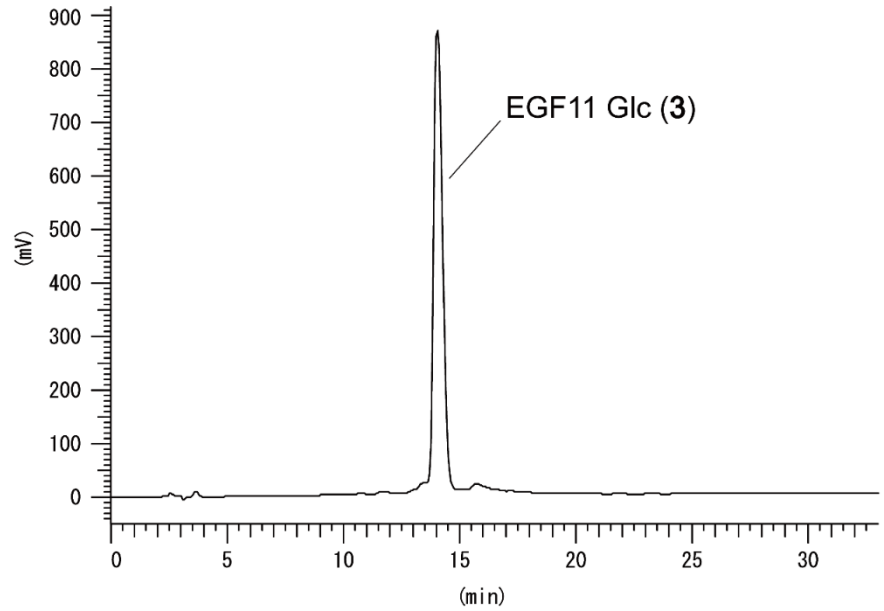


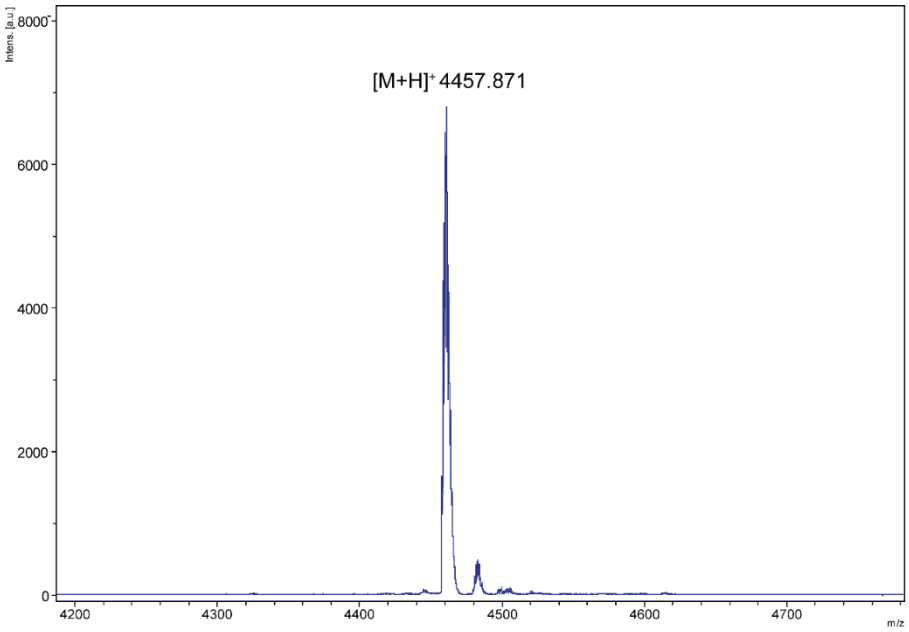
Module **2**.



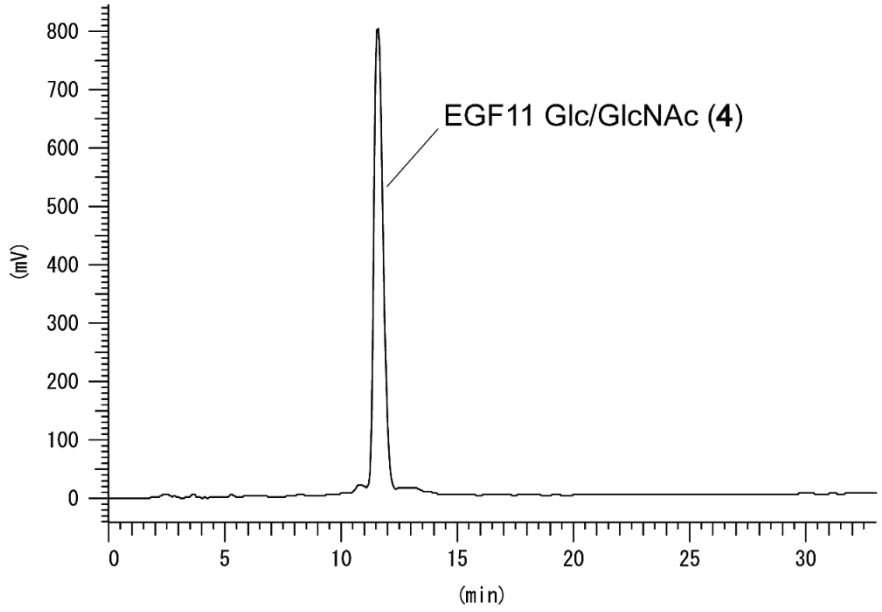


Module **3**.





Module **4**.



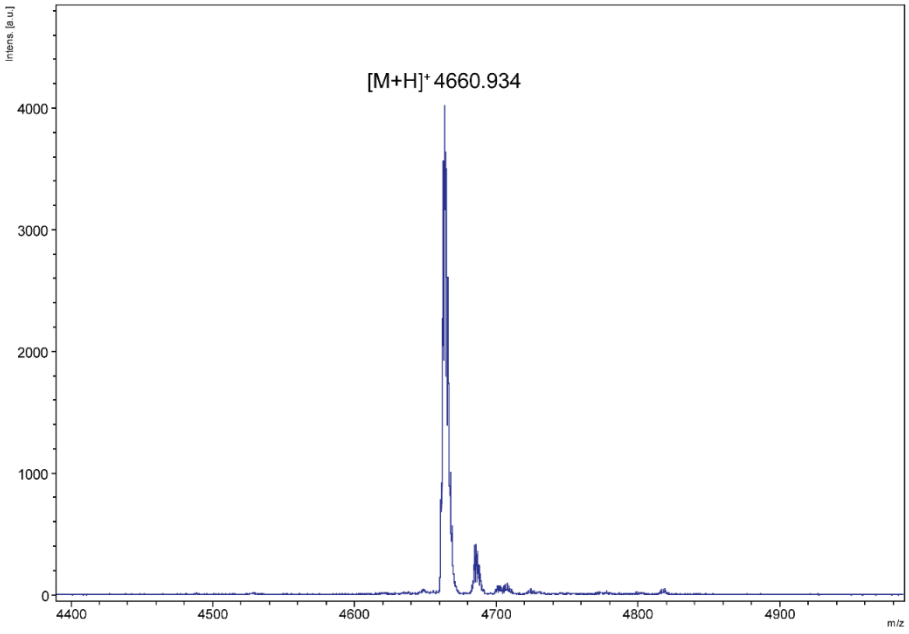
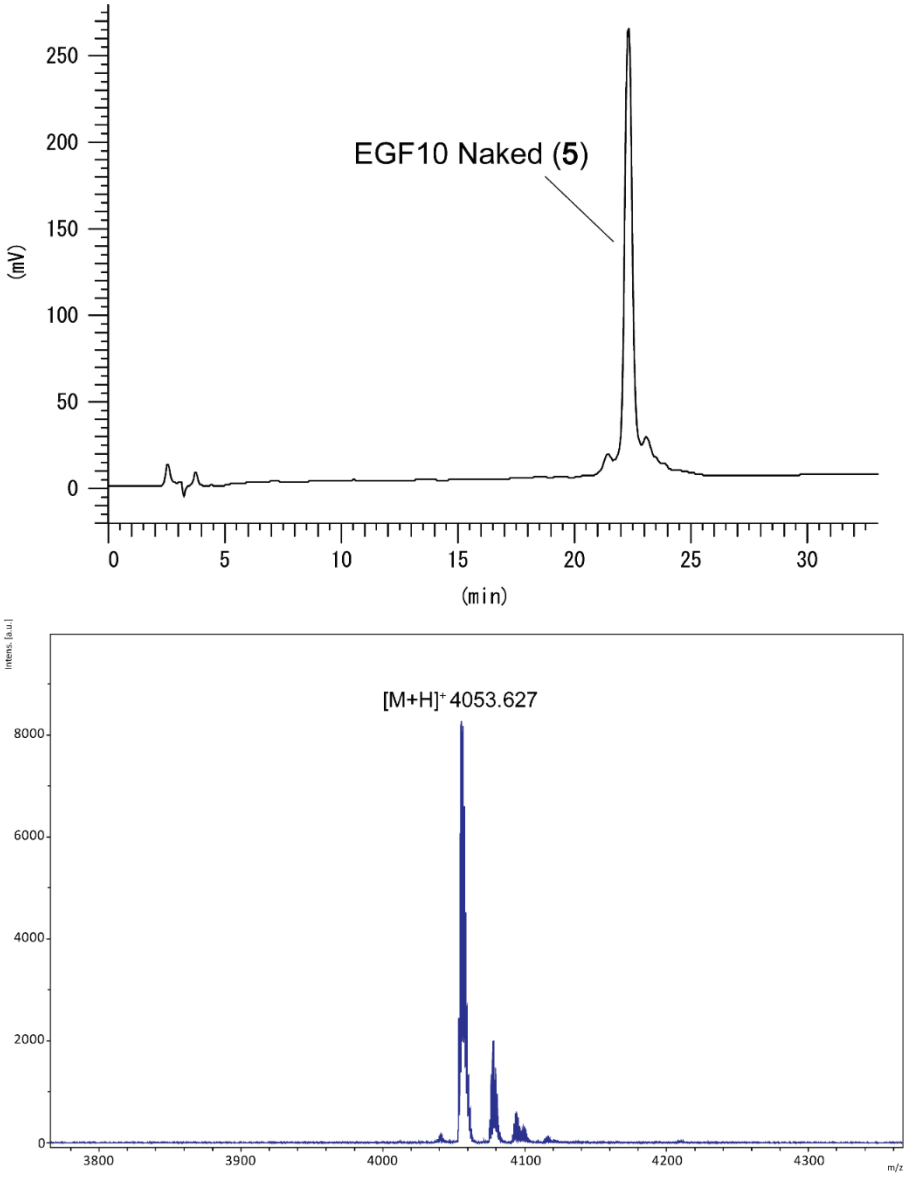
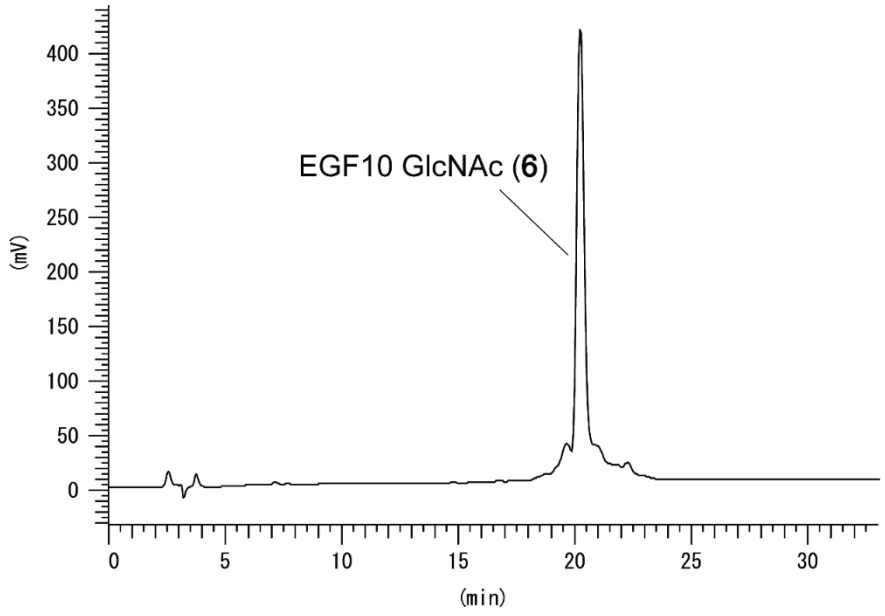


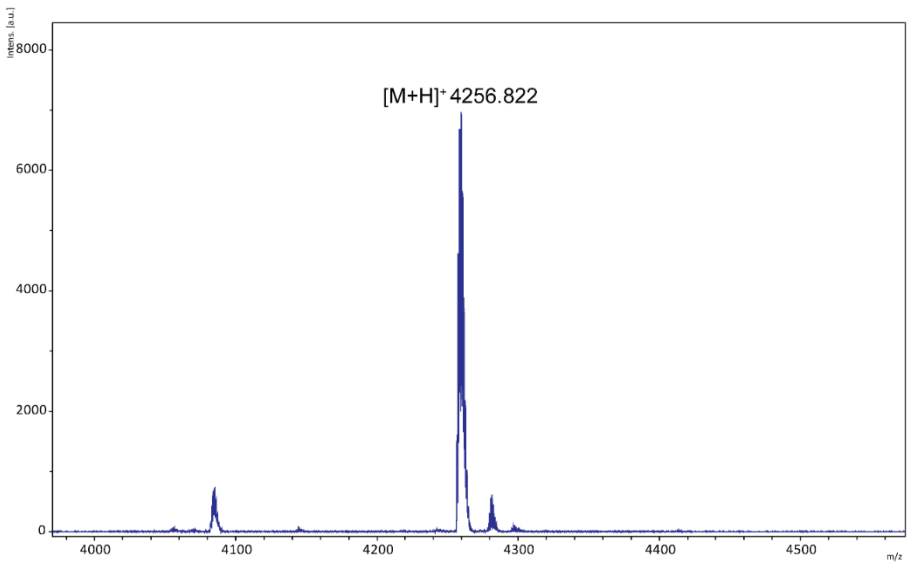
Figure S2-7. Characterization of isolated EGF10 modules. Analytical RP-HPLC (upper) and MALDI-TOF MS spectra (bottom) of purified NOTCH1 EGF10 module **5-9**. Column: Inertsil ODS-3 4.5 μm (250×4.6 mmI.D.), Eluent A: water containing 0.1% TFA, Eluent B: acetonitrile containing 0.1% TFA. Gradient condition; eluent (A/B = 82/18) was employed, then the ratio of eluent B was increased lineally from 18% to 28% over 30 min with a flow rate of 1.0 mL/min.

Module **5**.

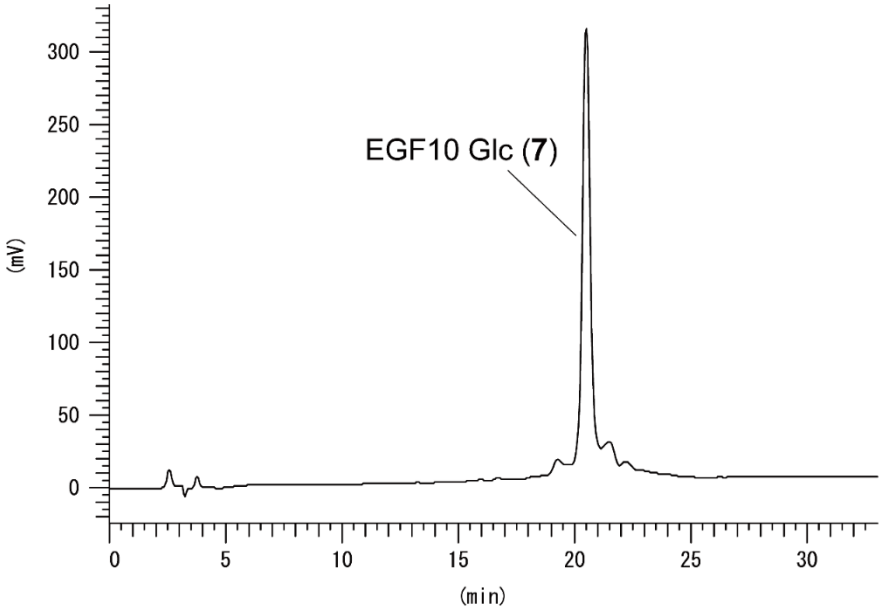


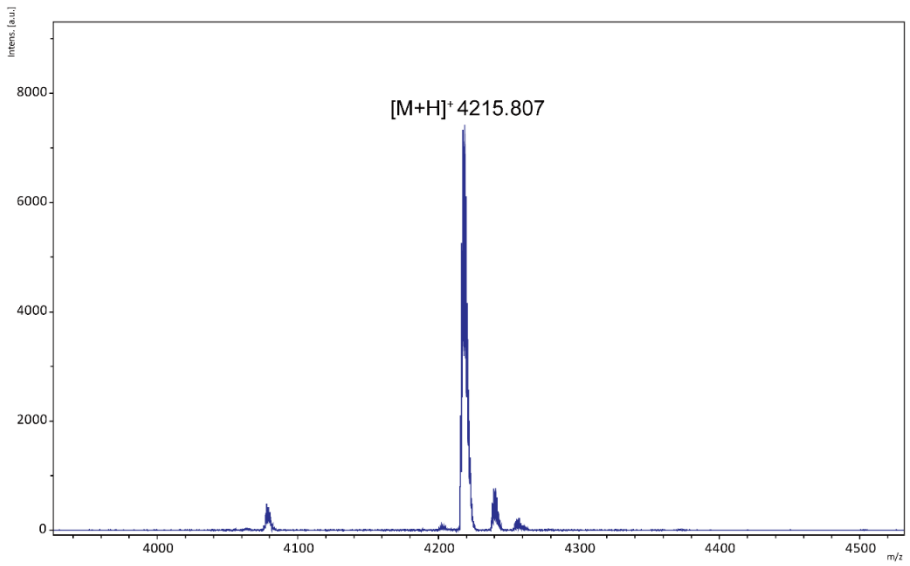
Module **6**.



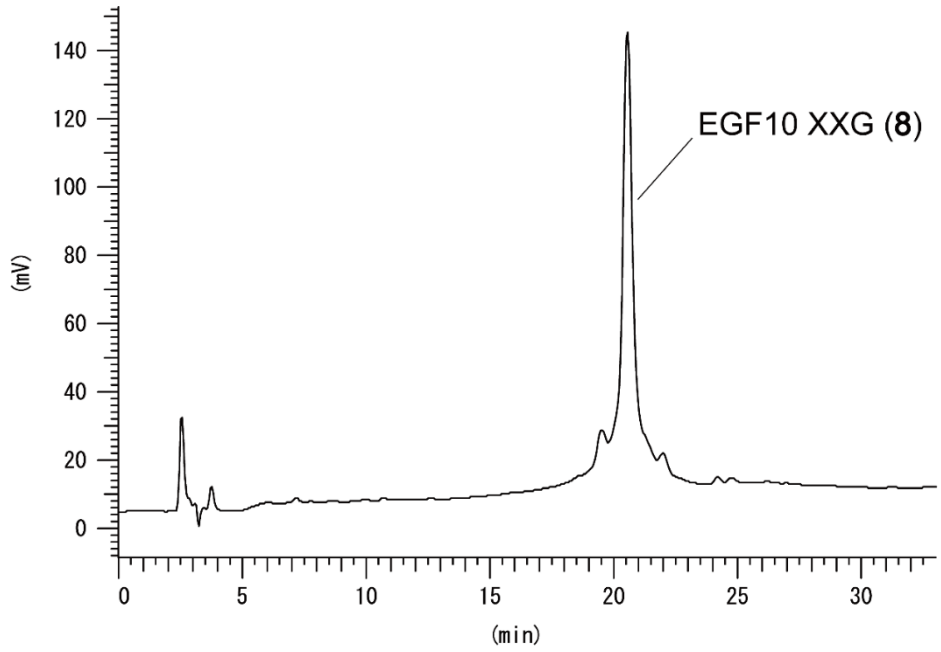


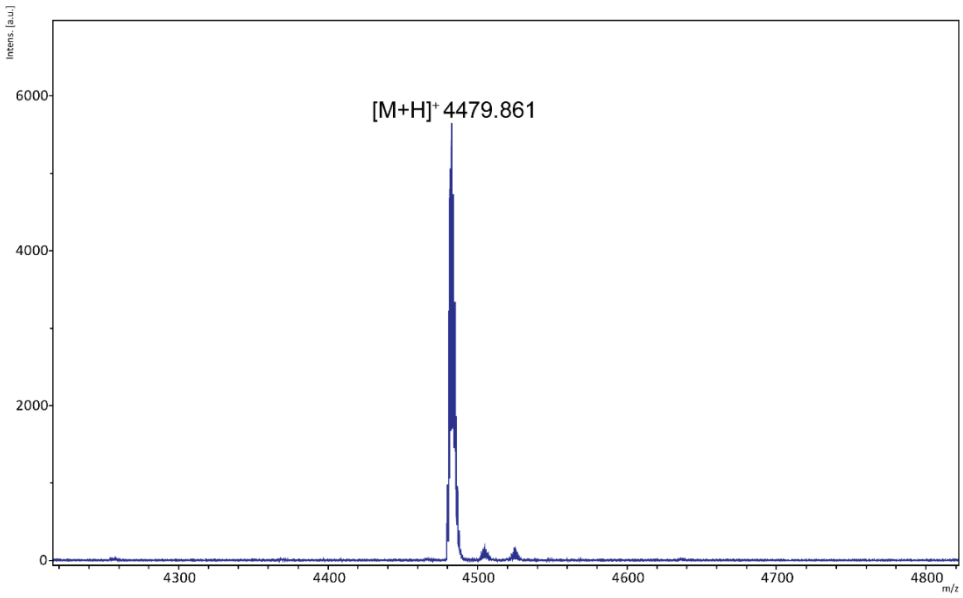
Module **7**.


EGF10 Glc (7)
(mV)
300
250
200
150
100
50
0
0
5
10
15
20
25
30
(min)

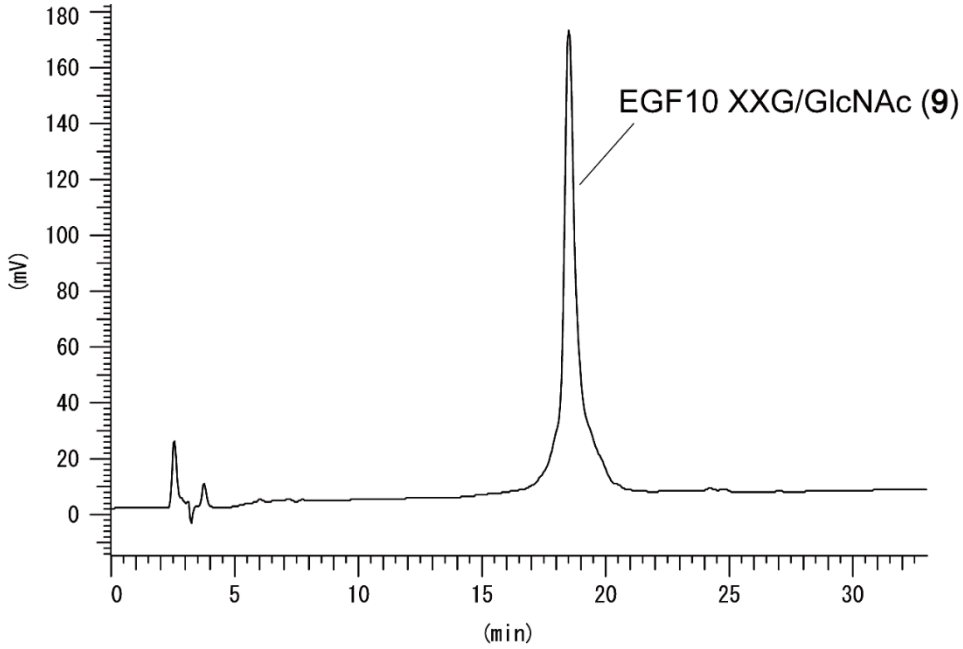

Intens. [a.u.]
[M+H]+ 4215.807
8000
6000
4000
2000
0
4000
4100
4200
4300
4400
4500
m/z

Module **8**.


EGF10 XXG (8)
(mV)
(min)
140
120
100
80
60
40
20
0
0
5
10
15
20
25
30


Intens. [a.u.]
[M+H]+ 4479.861
6000
4000
2000
0
4300
4400
4500
4600
4700
4800
m/z

Module **9**.



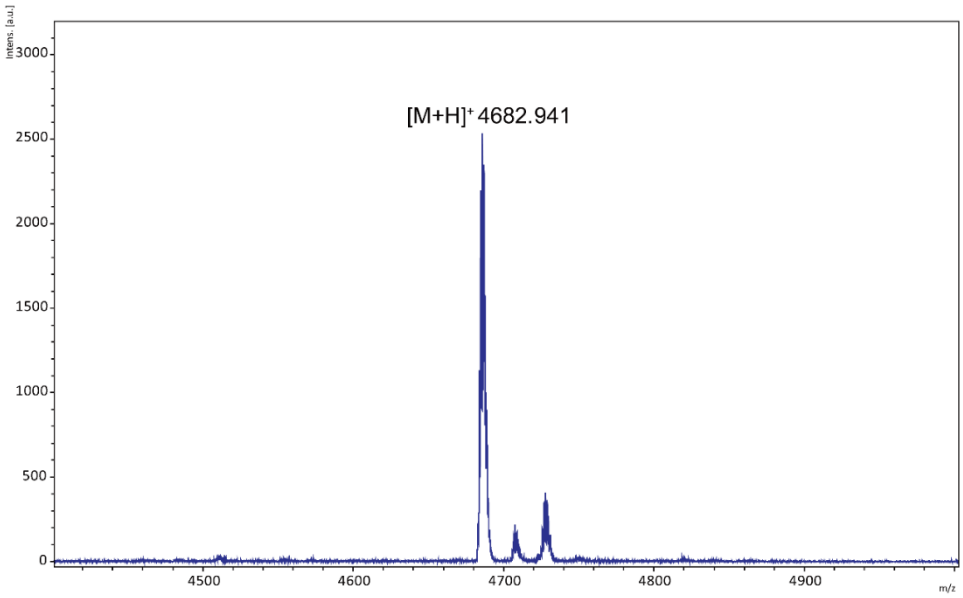


Figure S2-8. Monitoring the folding from linear EGF11 peptides. Analytical RP-HPLC (left) and MALDI-TOF MS spectra (right) of NOTCH1 EGF11 modules **1-4** isolated from the fractions represented by the asterisk which are shown in Figure 2-3, comparing with linear peptides. MALDI-TOF MS spectra show the decrease in the mass during oxidative folding reaction is six, indicating that three pairs of disulfide bonds are formed by this reaction. Column: Inertsil ODS-3 4.5 μm (250×4.6 mmI.D.), Eluent A: water containing 0.1% TFA, Eluent B: acetonitrile containing 0.1% TFA. Gradient condition; eluent (A/B = 80/20) was employed, then the ratio of eluent B was increased lineally from 20% to 40% over 30 min with a flow rate of 1.0 mL/min.

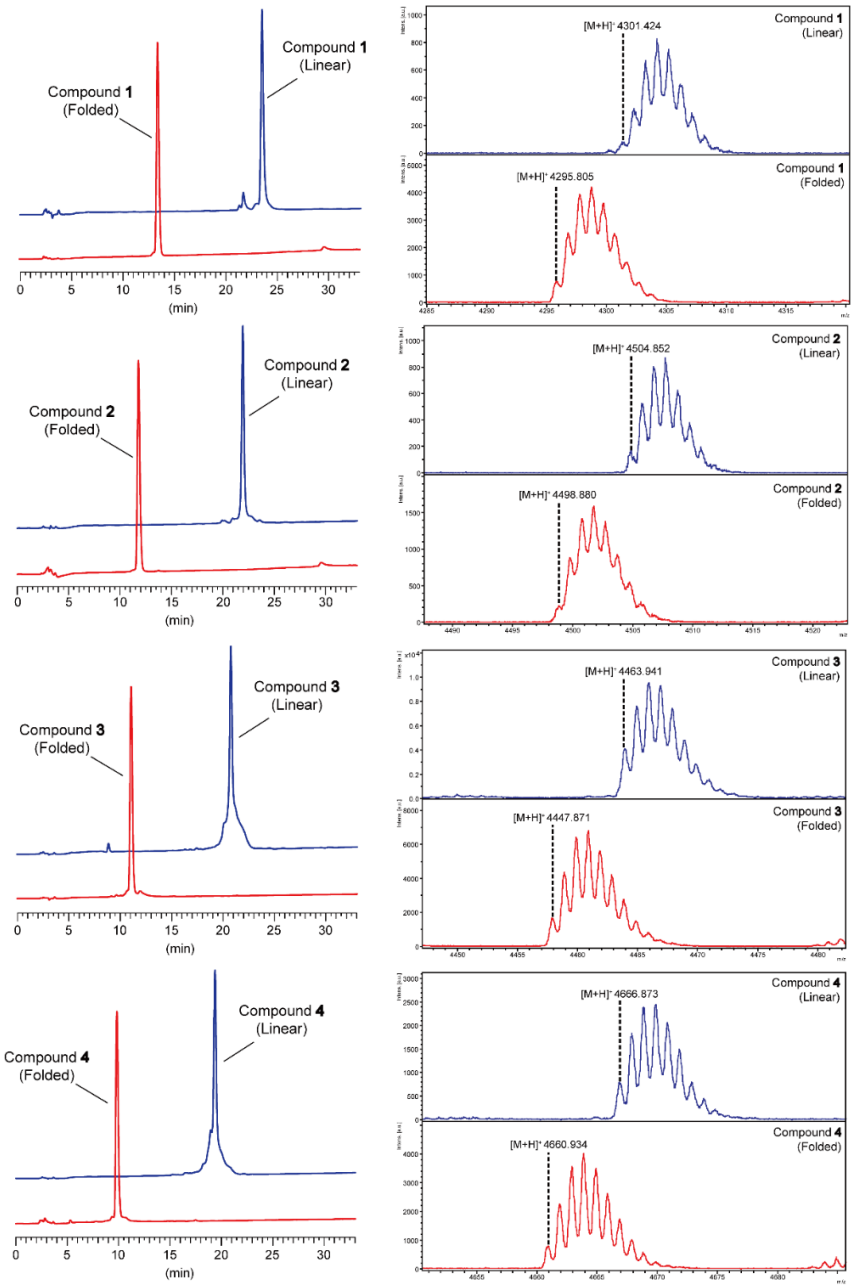


Figure S2-9. Monitoring the folding from linear EGF10 peptides. Analytical RP-HPLC (left) and MALDI-TOF MS spectra (right) of NOTCH1 EGF10 modules **5-9** isolated from the fractions represented by the asterisk which are shown in Figure 2-3, comparing with linear peptides. MALDI-TOF MS spectra show the decrease in the mass during oxidative folding reaction is six, indicating that three pairs of disulfide bonds are formed by this reaction. Column: Inertsil ODS-3 4.5 μm (250×4.6 mmI.D.), Eluent A: water containing 0.1% TFA, Eluent B: acetonitrile containing 0.1% TFA. Gradient condition; eluent (A/B = 85/15) was employed, then the ratio of eluent B was increased lineally from 15% to 35% over 30 min with a flow rate of 1.0 mL/min.

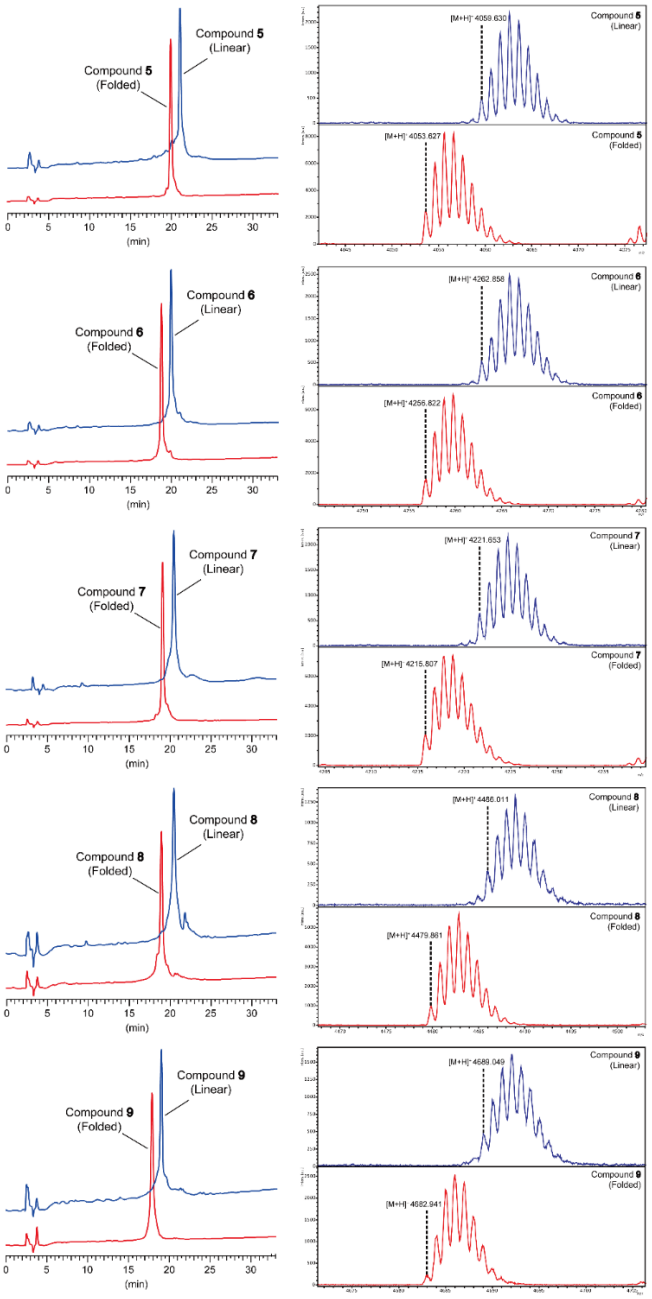
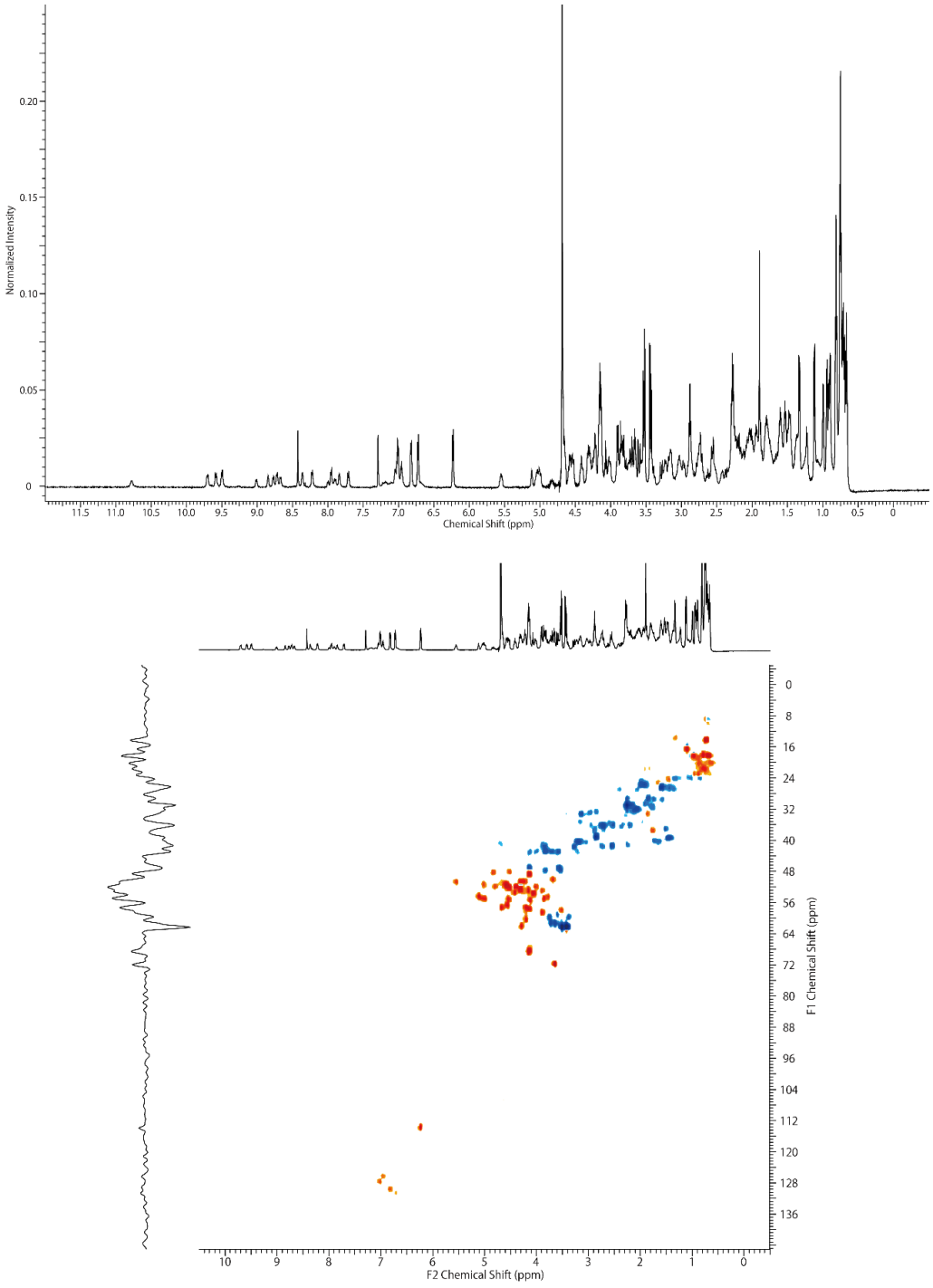
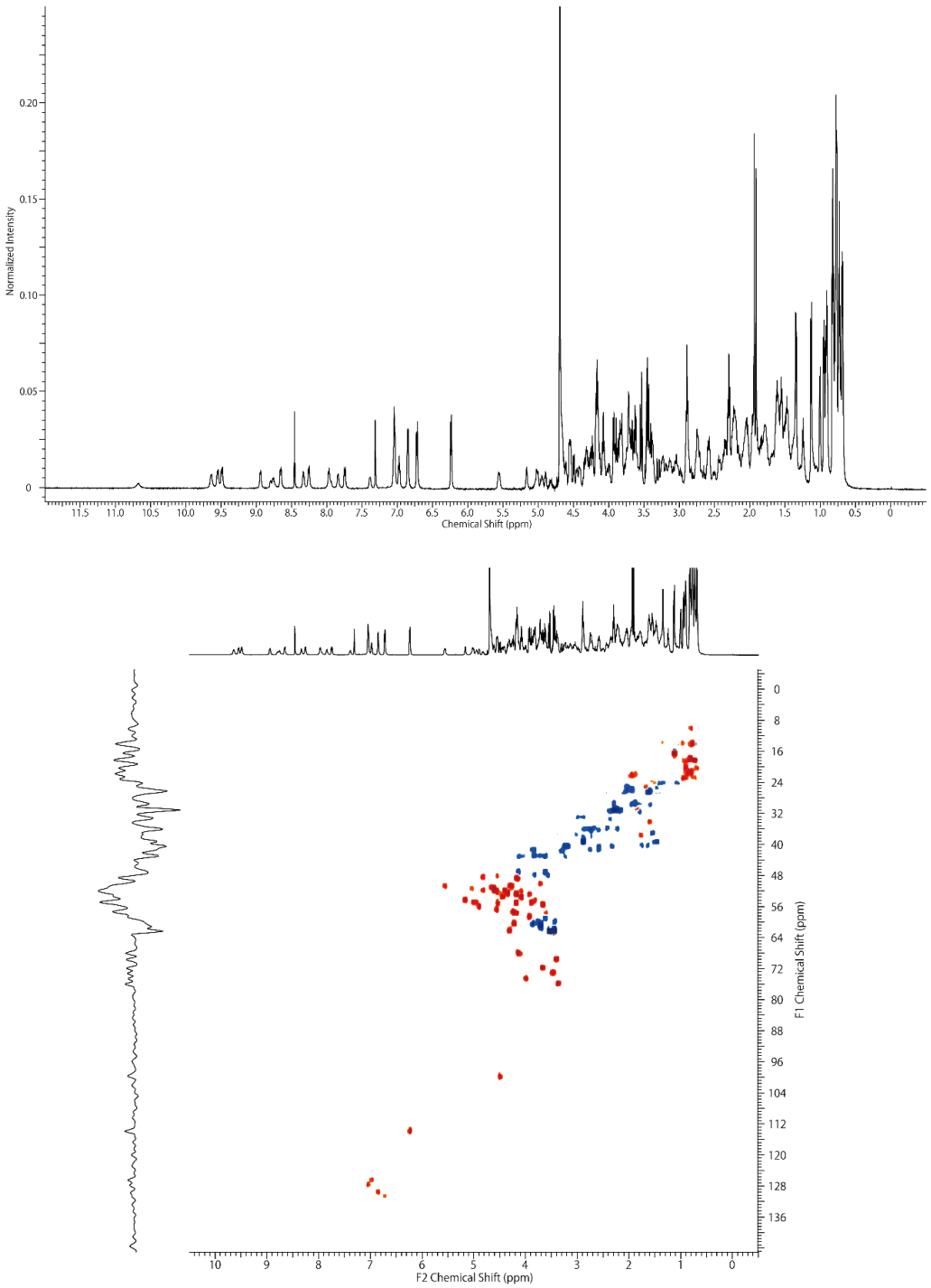


Figure S2-10. ^{1}H-NMR (upper) and HSQC (bottom) of EGF11 module **1-4** (D_2O, 600 MHz, at 298 K).

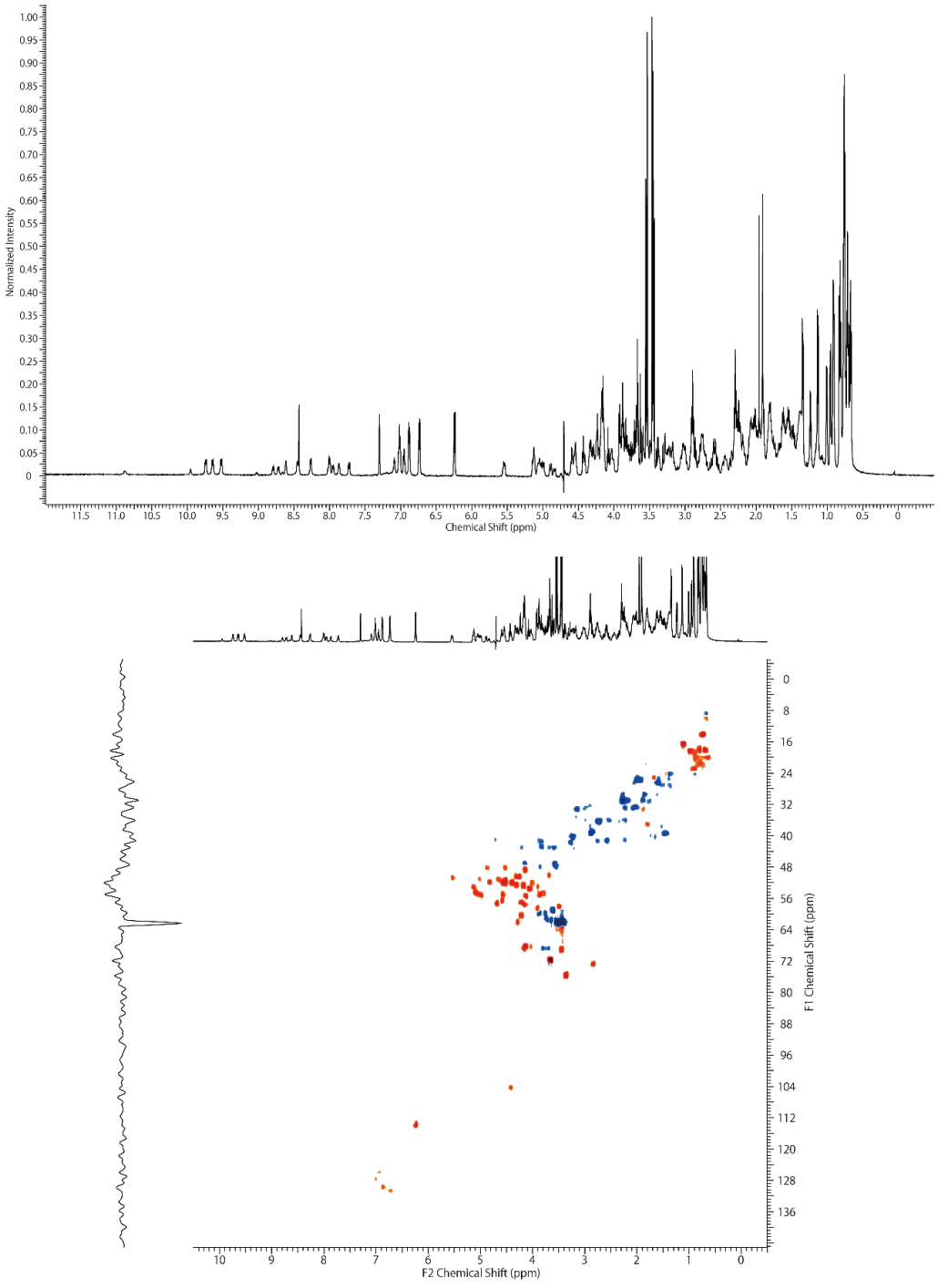
Module **1**.



Module **2**.


Normalized Intensity
0.20
0.15
0.10
0.05
0
11.5 11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0
Chemical Shift (ppm)
0 8 16 24 32 40 48 56 64 72 80 88 96 104 112 120 128 136
F1 Chemical Shift (ppm)
10 9 8 7 6 5 4 3 2 1 0
F2 Chemical Shift (ppm)

Module **3**.


Normalized Intensity
Chemical Shift (ppm)
F1 Chemical Shift (ppm)
F2 Chemical Shift (ppm)

Module **4**.

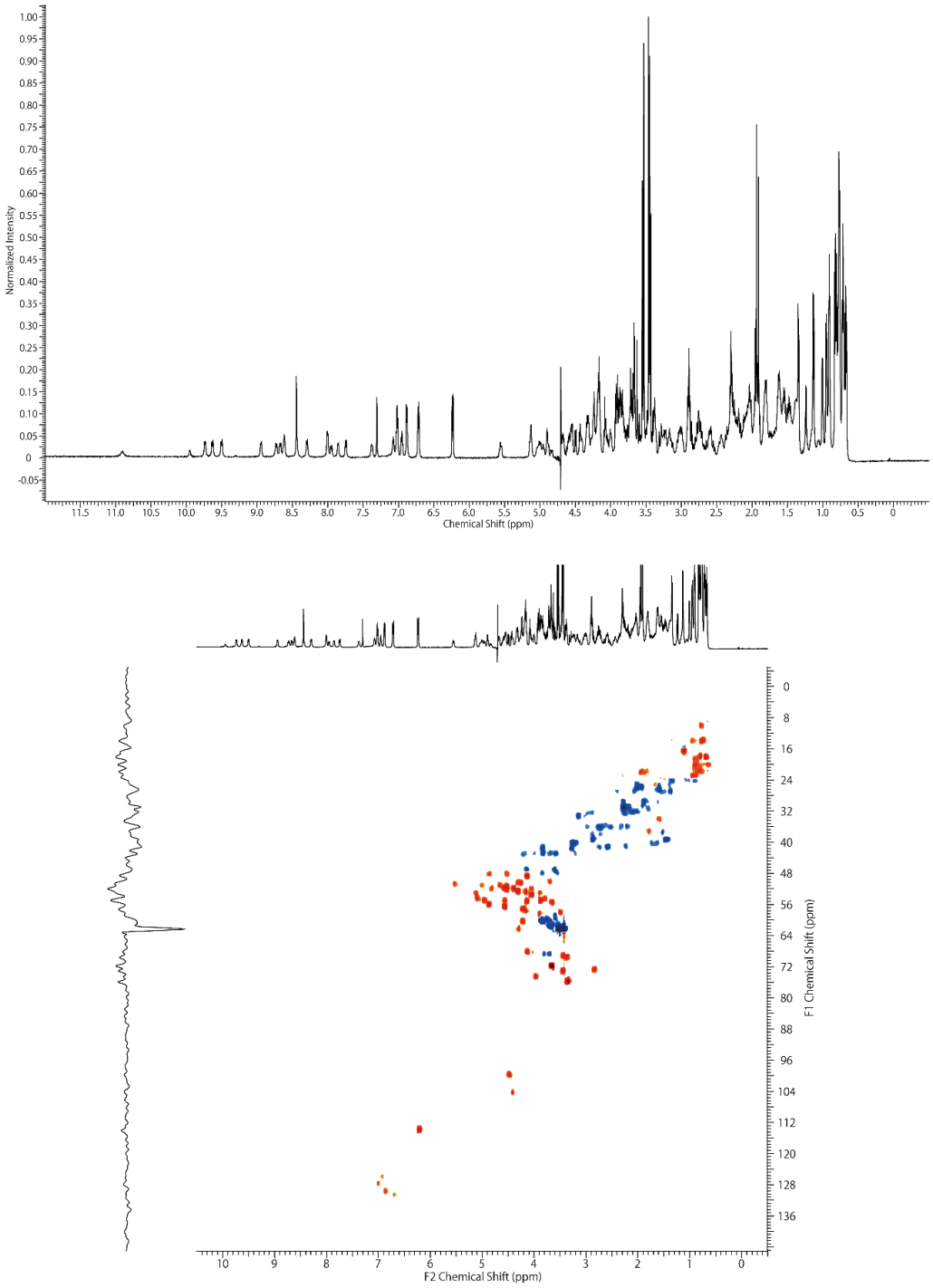
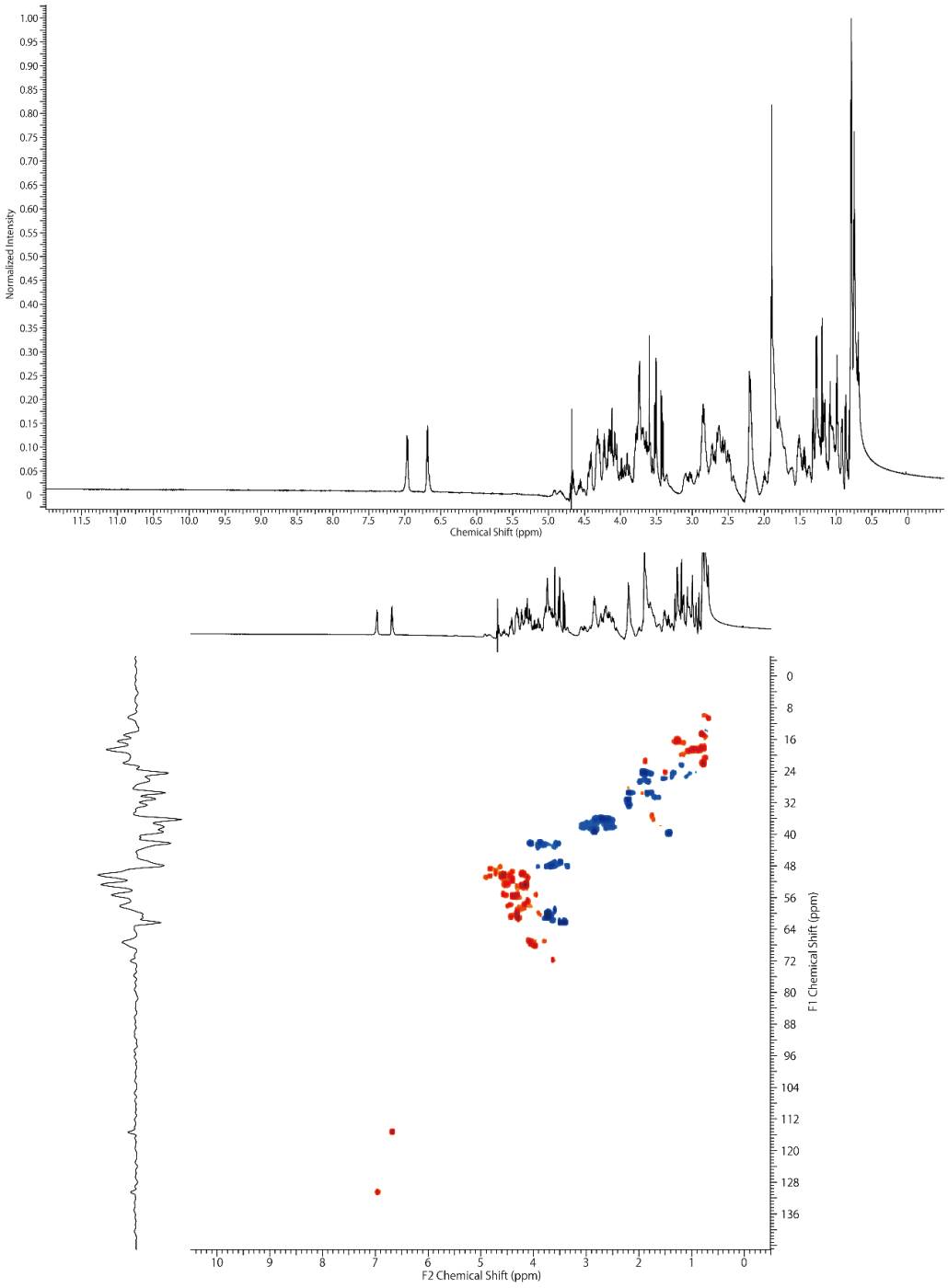
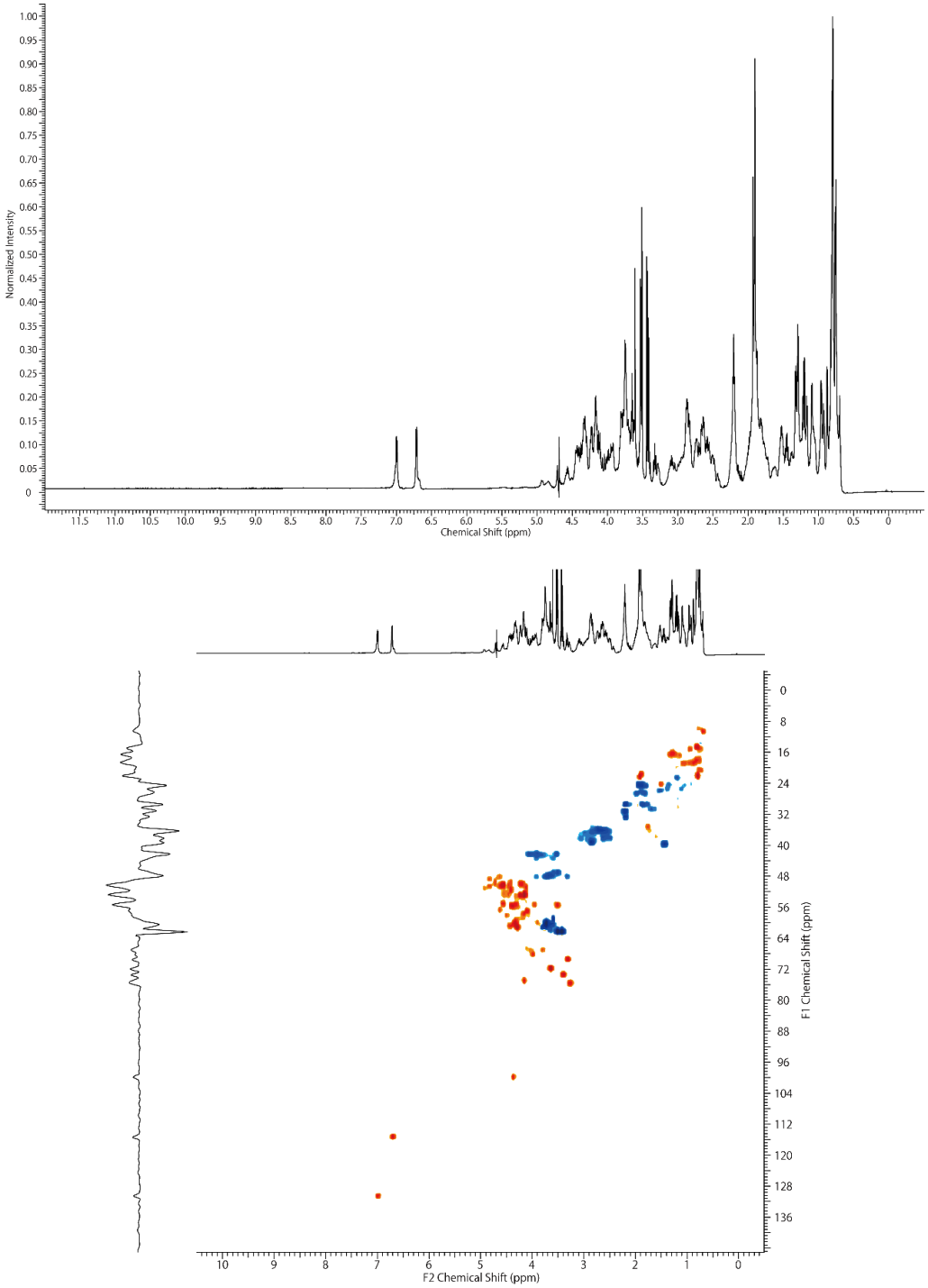

Normalized Intensity
Chemical Shift (ppm)
F2 Chemical Shift (ppm)
F1 Chemical Shift (ppm)

Figure S2-11. ^{1}H-NMR (upper) and HSQC (bottom) of EGF10 module **5-9** (D_2O, 600 MHz, at 298 K).

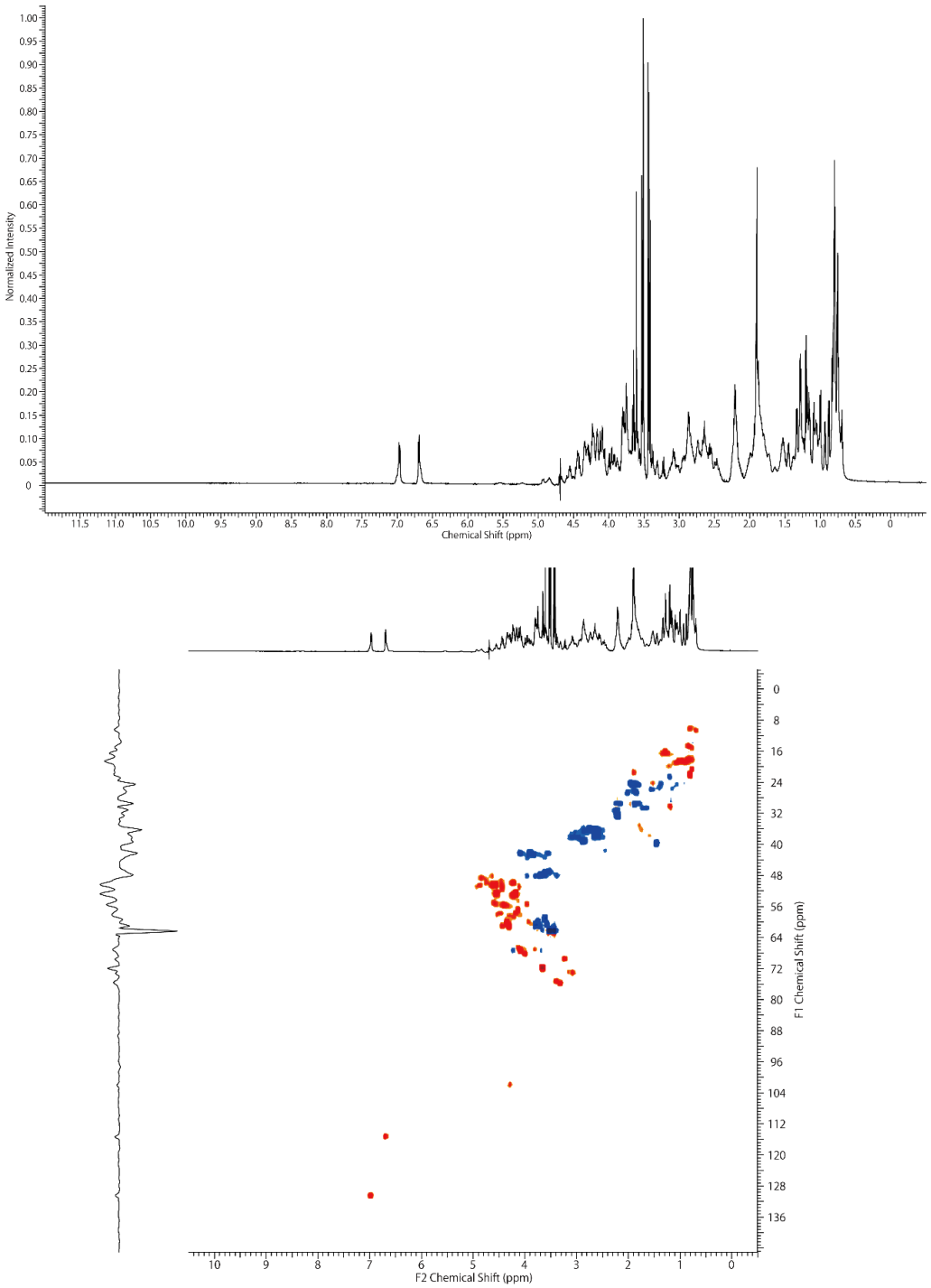
Module **5**.



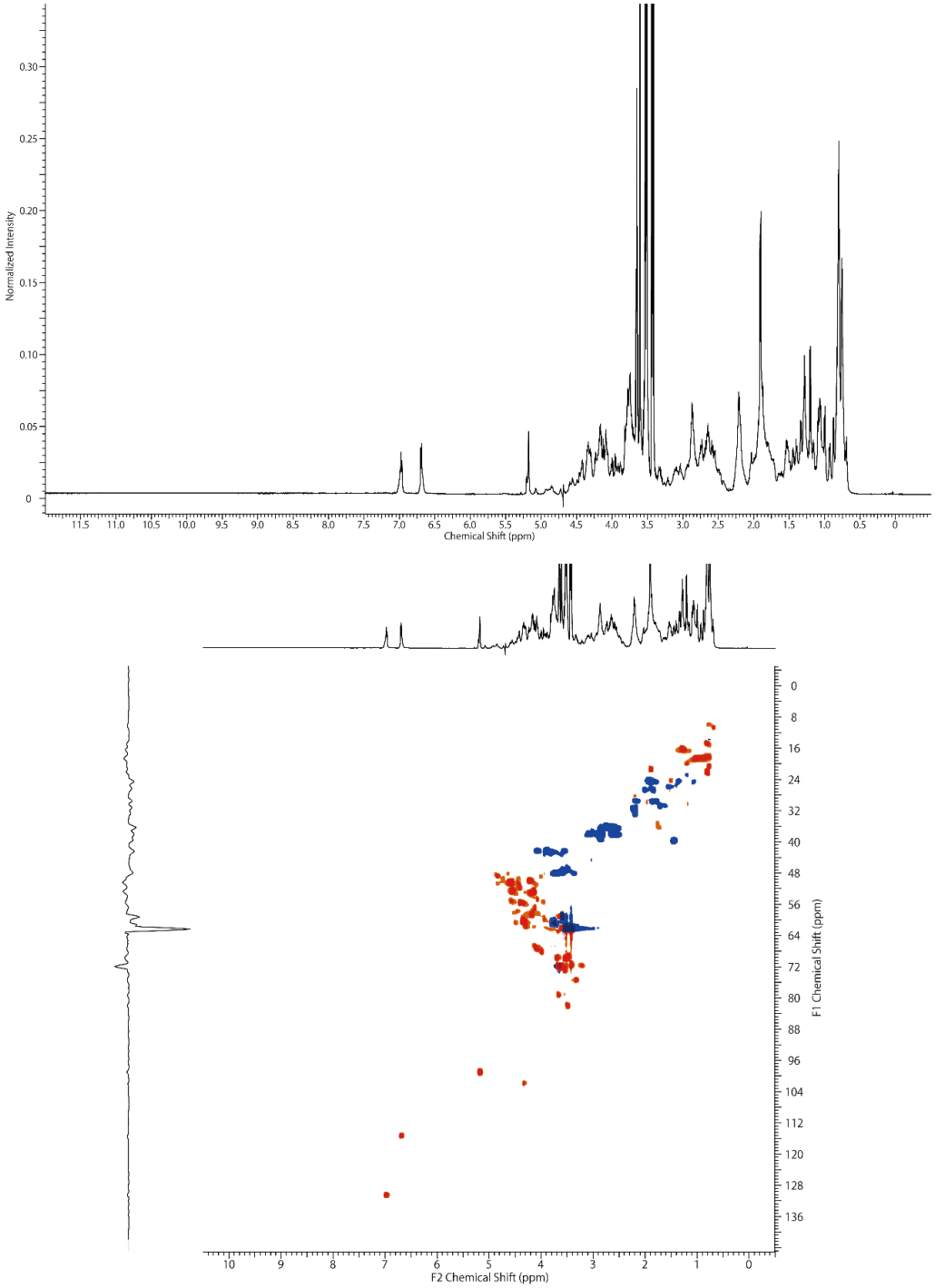
Module **6**.


Normalized Intensity
Chemical Shift (ppm)
F1 Chemical Shift (ppm)
F2 Chemical Shift (ppm)

Module **7**.


Normalized Intensity
Chemical Shift (ppm)
F1 Chemical Shift (ppm)
F2 Chemical Shift (ppm)

Module **8**.


Normalized Intensity
Chemical Shift (ppm)
F1 Chemical Shift (ppm)
F2 Chemical Shift (ppm)

Module **9**.

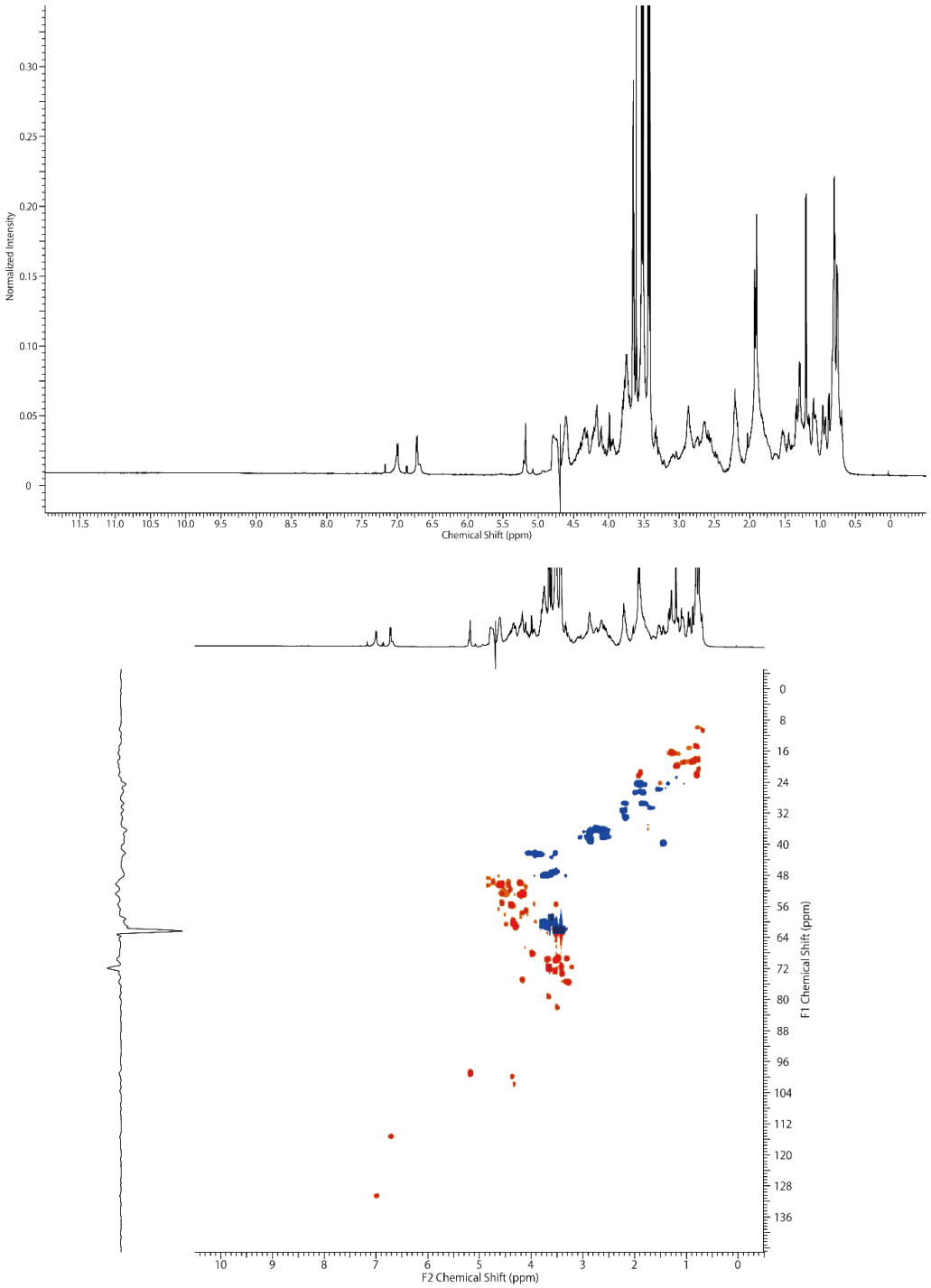
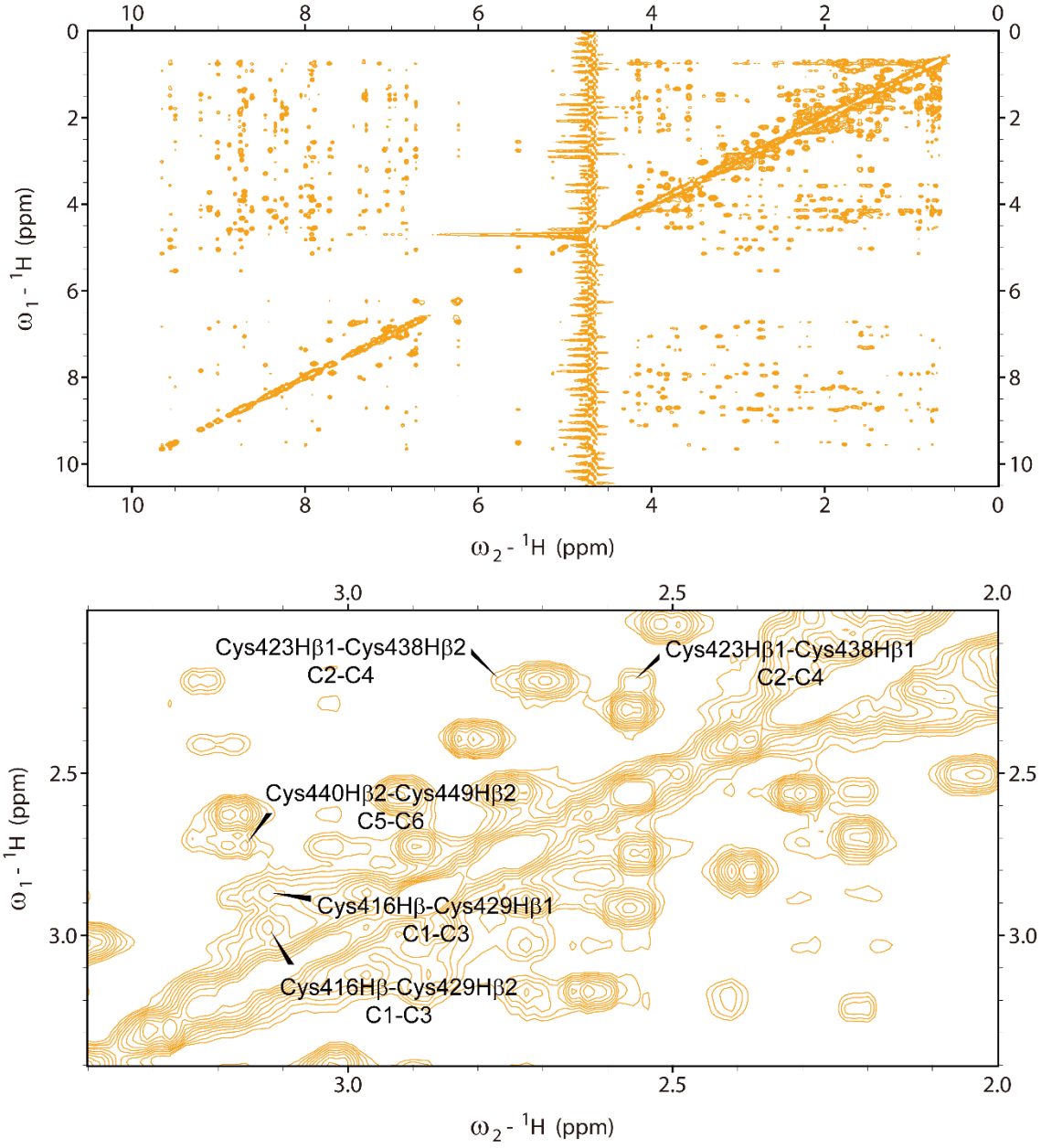
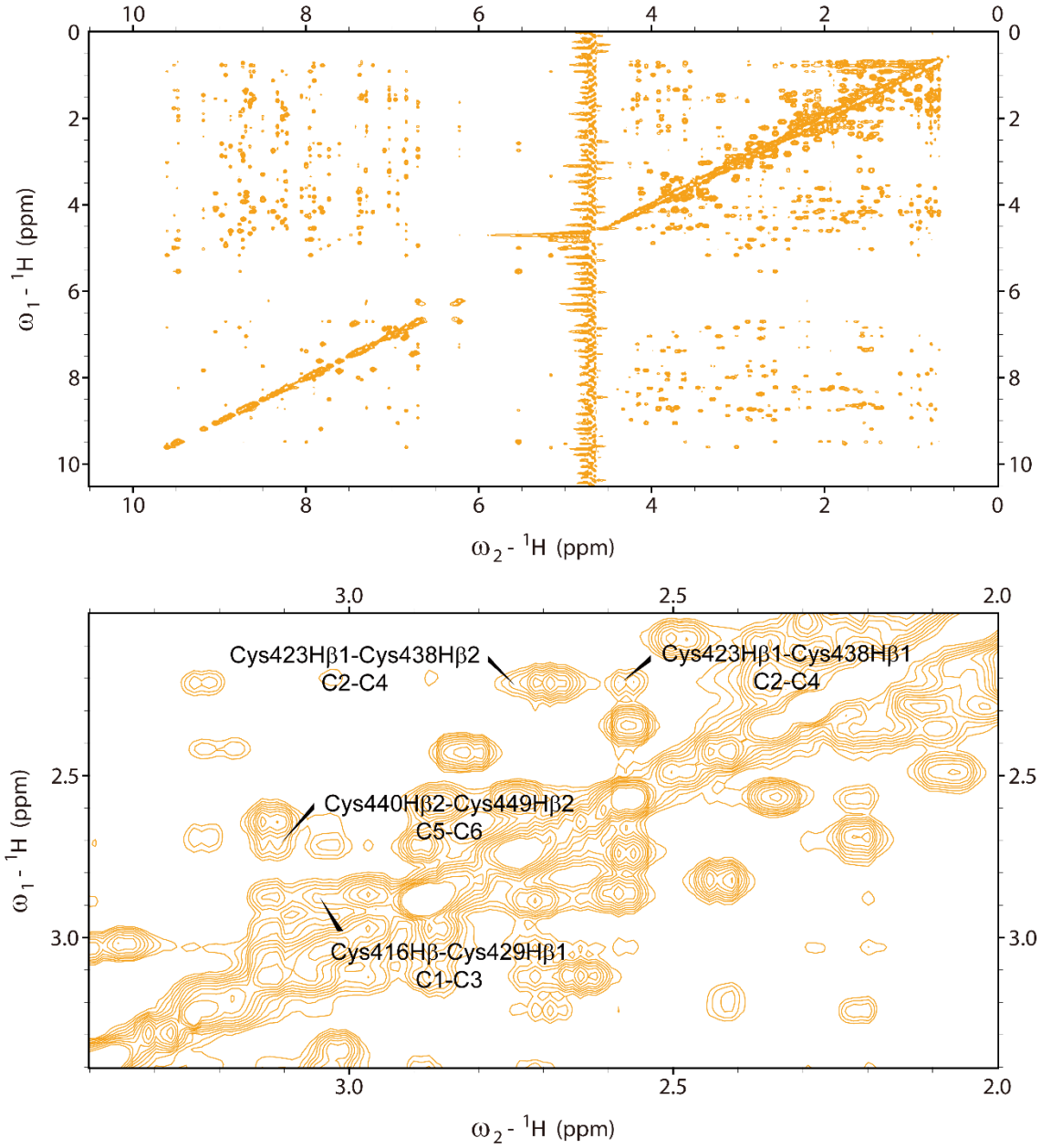

Normalized Intensity
Chemical Shift (ppm)
F2 Chemical Shift (ppm)
F1 Chemical Shift (ppm)

Figure S2-12. NOESY spectra of NOTCH1 EGF11 modules **1-4**. All compounds were isolated from the fractions represented by asterisk (upper) and enlarged NOEs from 2.0-3.4 ppm showing cross-linked cysteine residues between two β-protons (bottom).

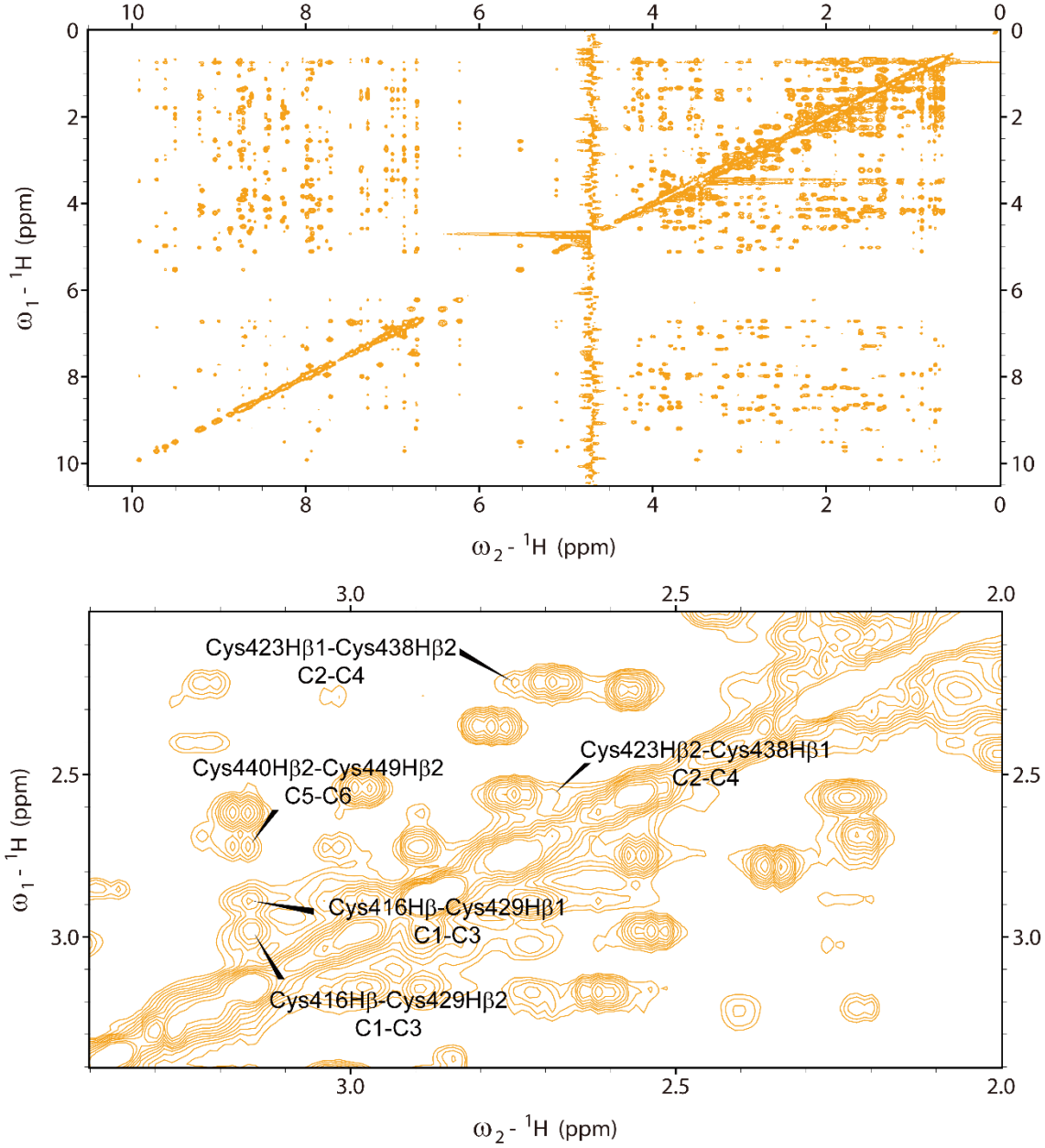
Module **1**.



Module **2**.


ω1 - 1H (ppm)
ω2 - 1H (ppm)
Cys423Hβ1-Cys438Hβ2
C2-C4
Cys423Hβ1-Cys438Hβ1
C2-C4
Cys440Hβ2-Cys449Hβ2
C5-C6
Cys416Hβ-Cys429Hβ1
C1-C3

Module **3**.


ω1 - 1H (ppm)
ω2 - 1H (ppm)
Cys423Hβ1-Cys438Hβ2
C2-C4
Cys423Hβ2-Cys438Hβ1
C2-C4
Cys440Hβ2-Cys449Hβ2
C5-C6
Cys416Hβ-Cys429Hβ1
C1-C3
Cys416Hβ-Cys429Hβ2
C1-C3

Module **4**.

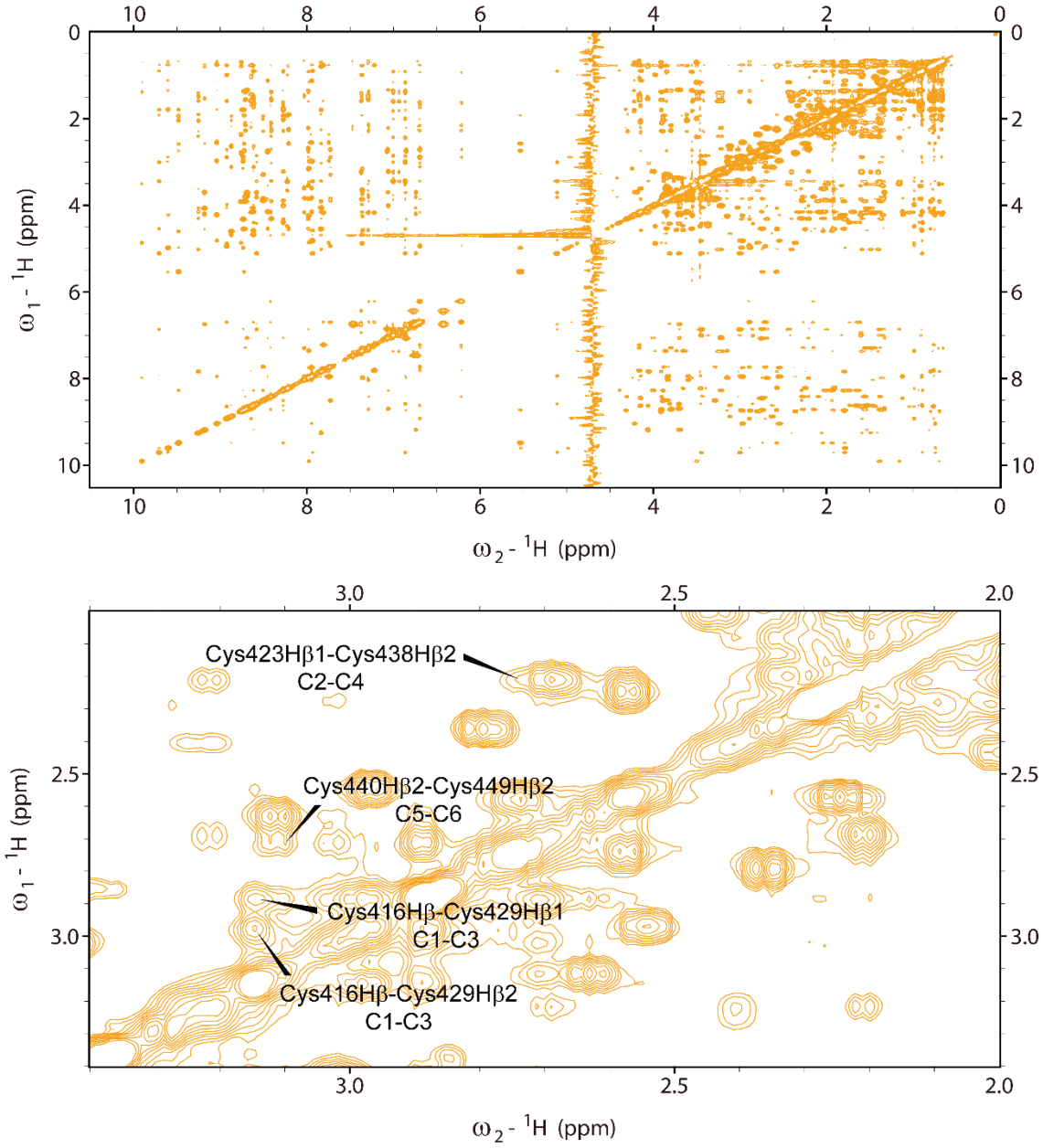
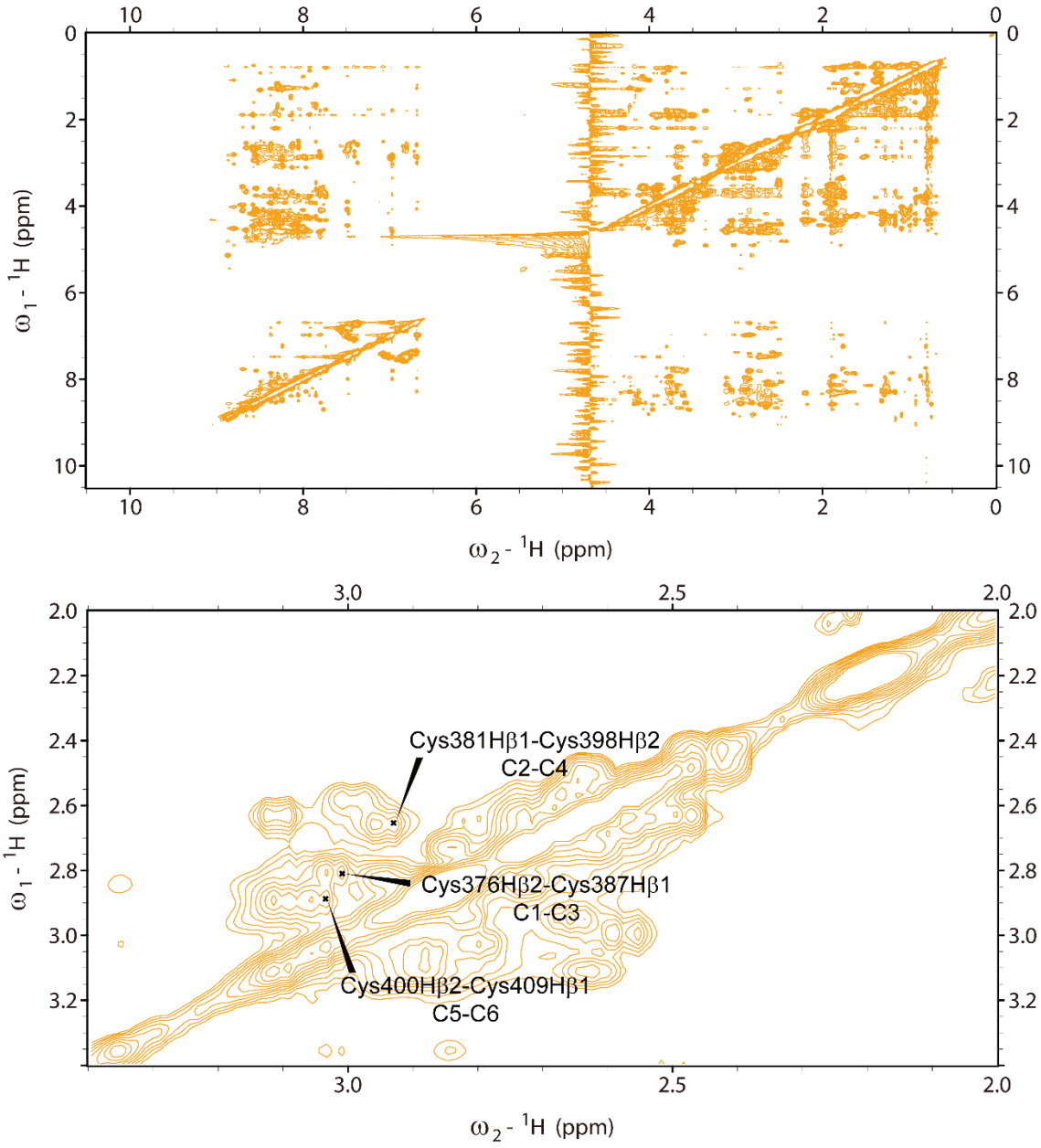
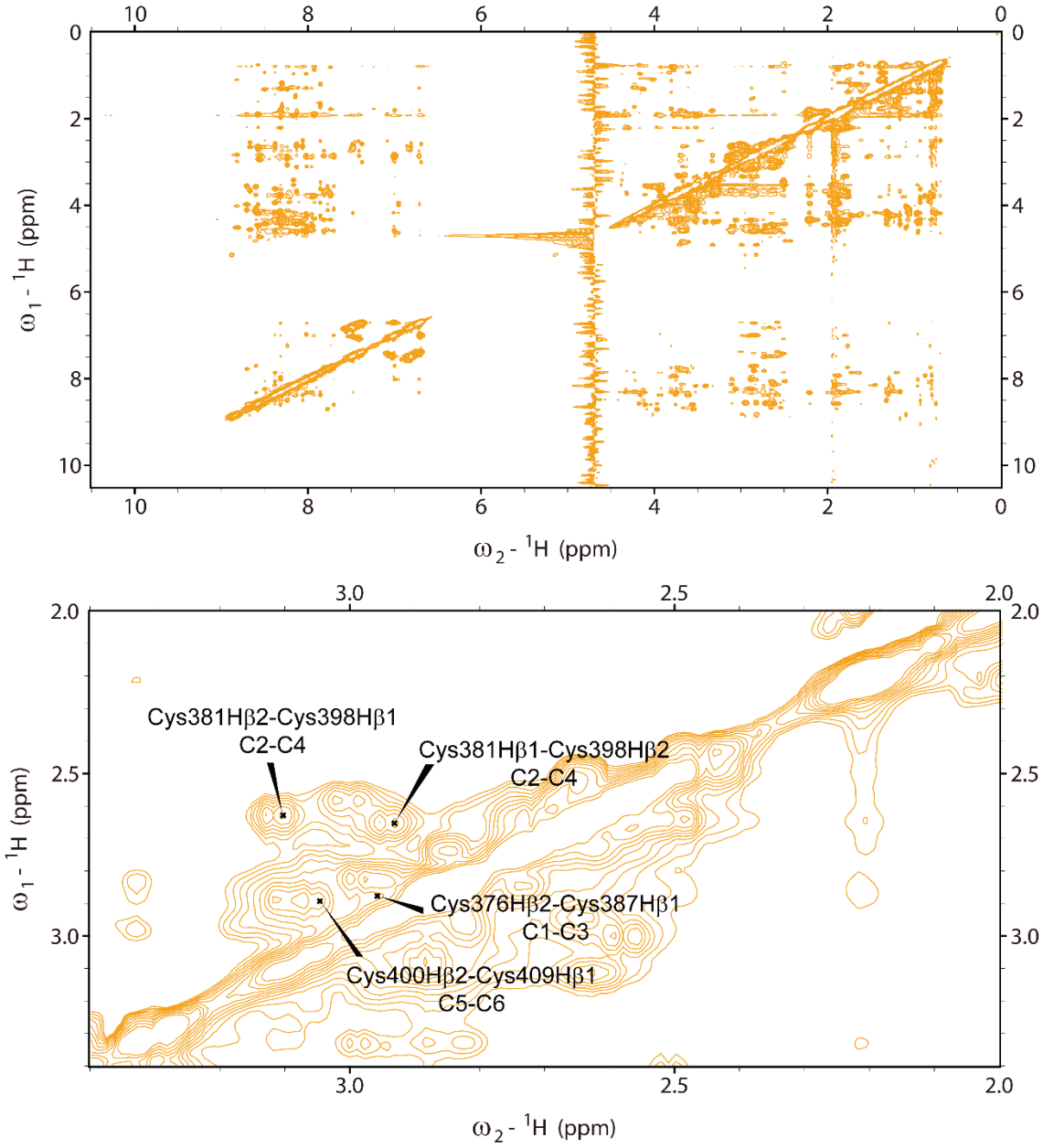

ω1 - 1H (ppm)
ω2 - 1H (ppm)
Cys423Hβ1-Cys438Hβ2
C2-C4
Cys440Hβ2-Cys449Hβ2
C5-C6
Cys416Hβ-Cys429Hβ1
C1-C3
Cys416Hβ-Cys429Hβ2
C1-C3

Figure S2-13. NOESY spectra of NOTCH1 EGF10 modules **5-9**. All compounds were isolated from the fractions represented by asterisk (upper) and enlarged NOEs from 2.0-3.4 ppm showing cross-linked cysteine residues between two β-protons (bottom).

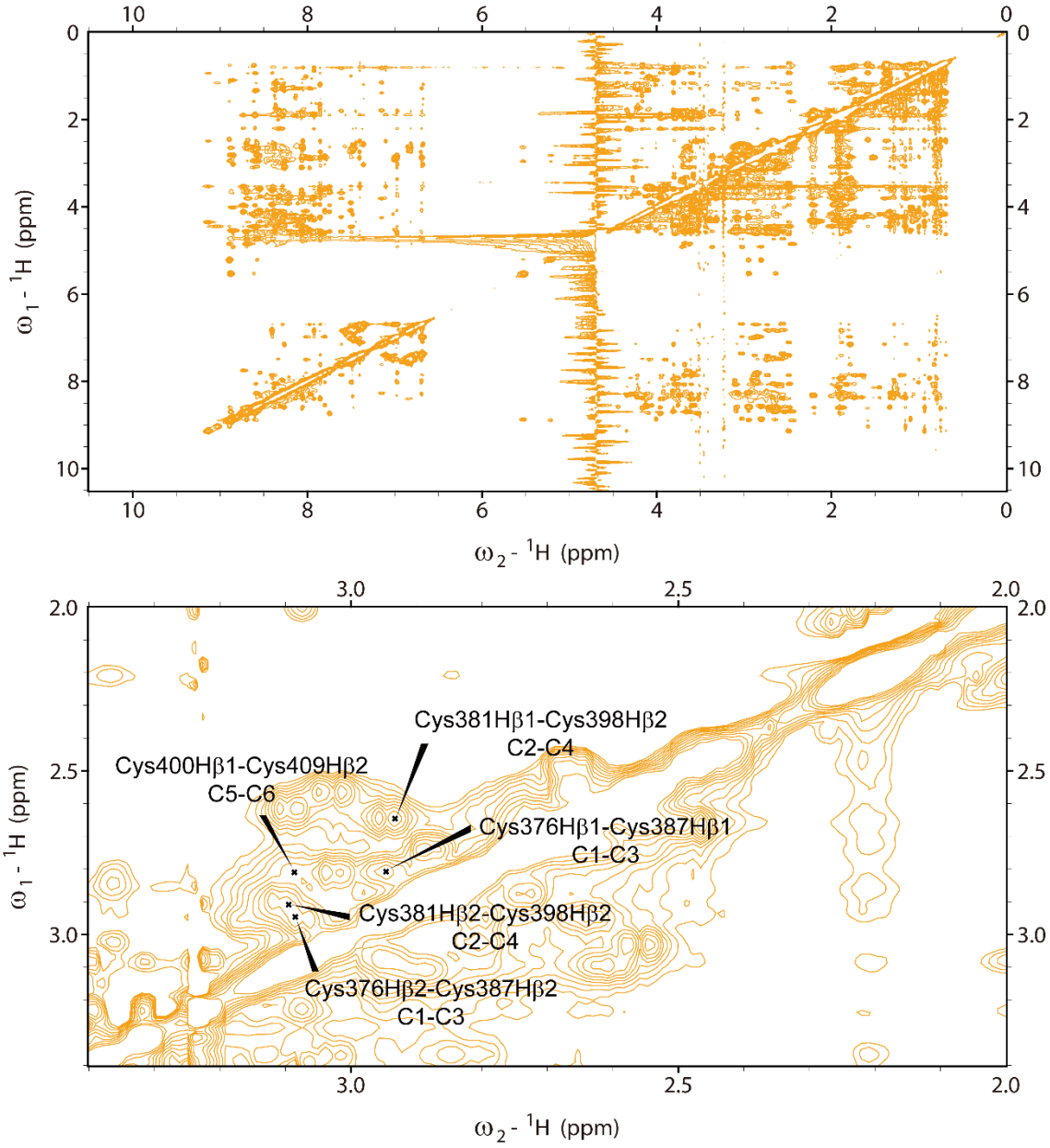
Module **5**.



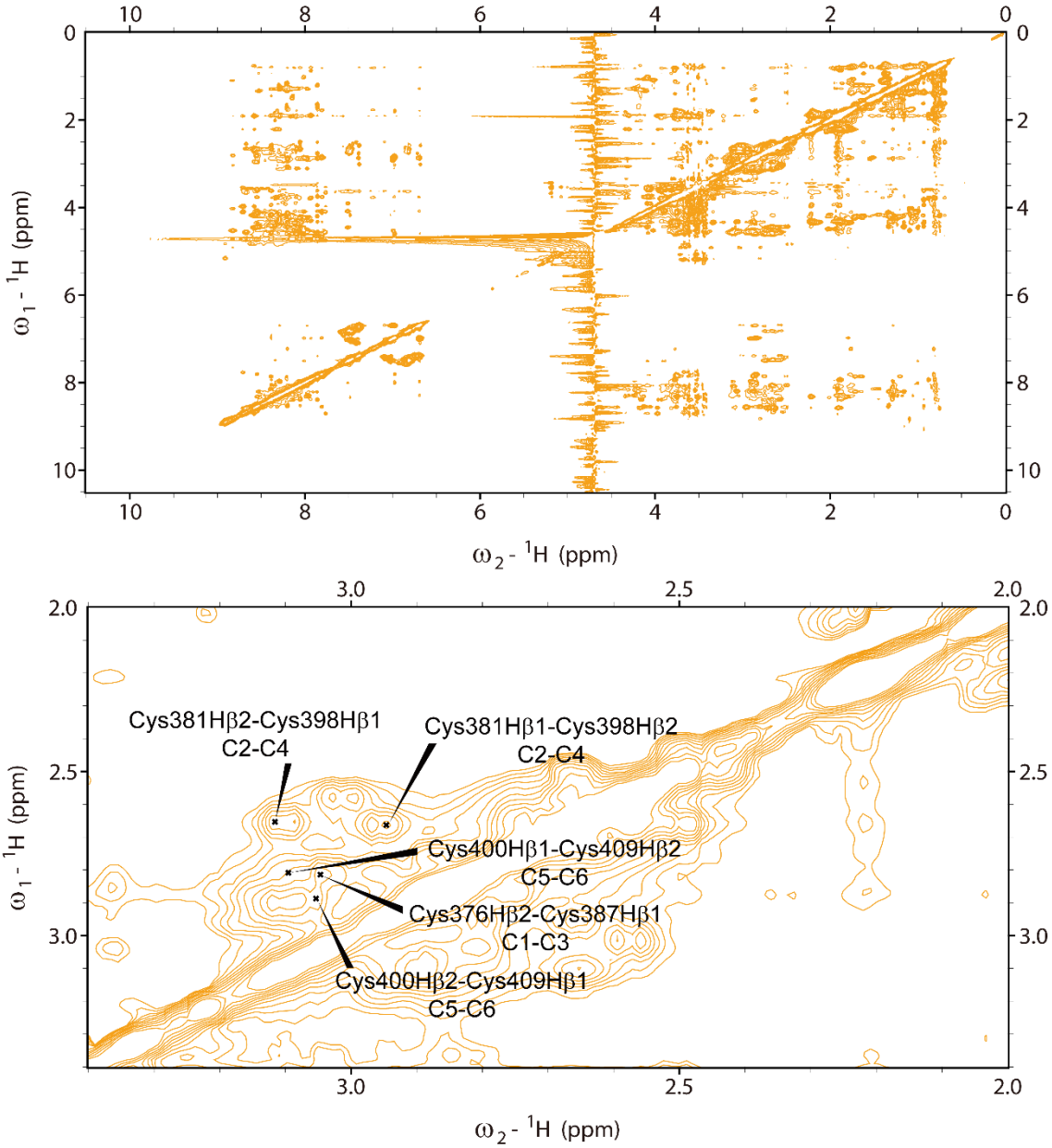
Module **6**.


ω1 - 1H (ppm)
ω2 - 1H (ppm)
Cys381Hβ2-Cys398Hβ1
C2-C4
Cys381Hβ1-Cys398Hβ2
C2-C4
Cys376Hβ2-Cys387Hβ1
C1-C3
Cys400Hβ2-Cys409Hβ1
C5-C6

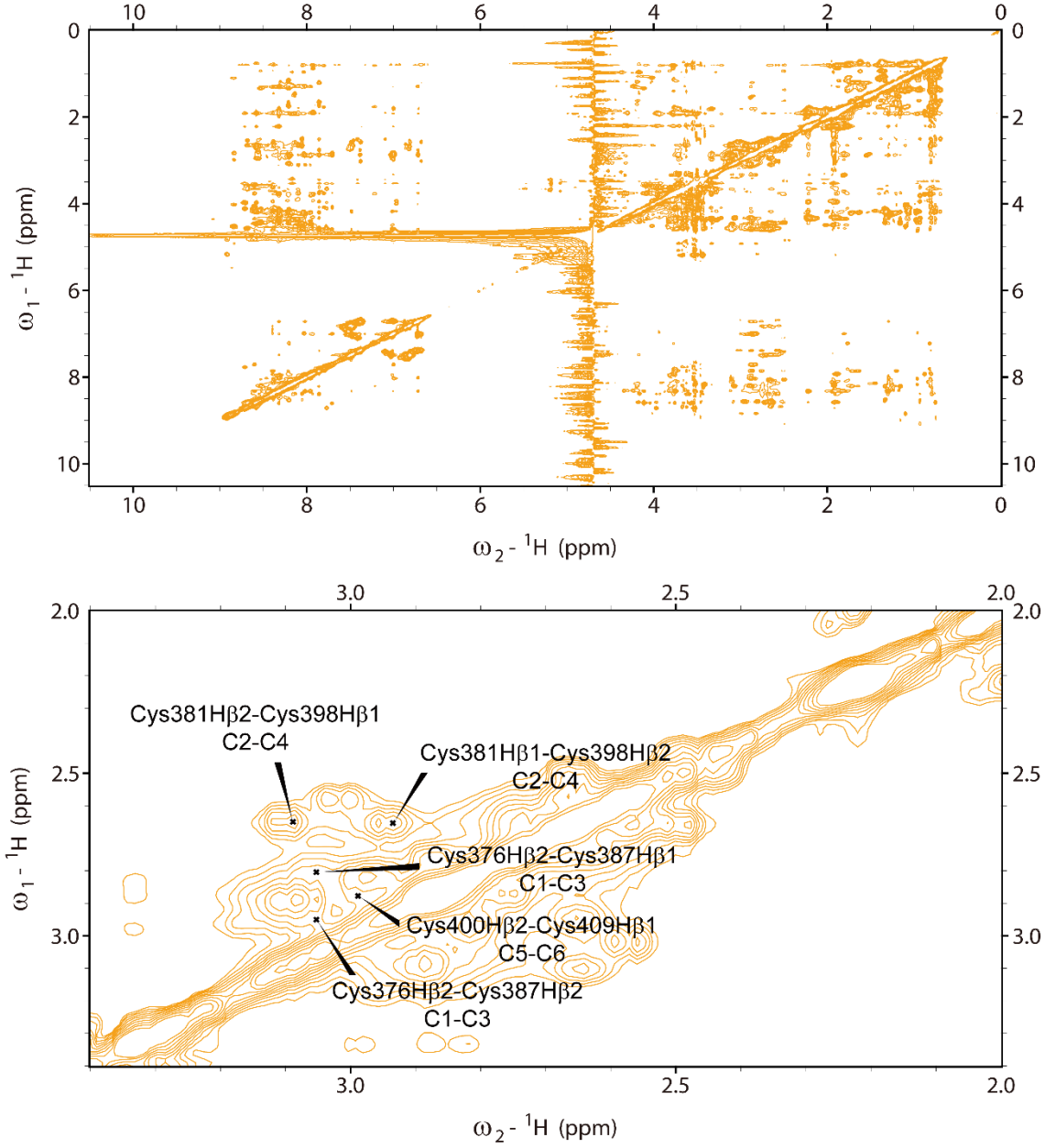
Module 7.


ω1 - 1H (ppm)
ω2 - 1H (ppm)
Cys381Hβ1-Cys398Hβ2
C2-C4
Cys400Hβ1-Cys409Hβ2
C5-C6
Cys376Hβ1-Cys387Hβ1
C1-C3
Cys381Hβ2-Cys398Hβ2
C2-C4
Cys376Hβ2-Cys387Hβ2
C1-C3

Module **8**.


ω1 - 1H (ppm)
ω2 - 1H (ppm)
Cys381Hβ2-Cys398Hβ1
C2-C4
Cys381Hβ1-Cys398Hβ2
C2-C4
Cys400Hβ1-Cys409Hβ2
C5-C6
Cys376Hβ2-Cys387Hβ1
C1-C3
Cys400Hβ2-Cys409Hβ1
C5-C6

Module **9**.


ω1 - 1H (ppm)
ω2 - 1H (ppm)
Cys381Hβ2-Cys398Hβ1
C2-C4
Cys381Hβ1-Cys398Hβ2
C2-C4
Cys376Hβ2-Cys387Hβ1
C1-C3
Cys400Hβ2-Cys409Hβ1
C5-C6
Cys376Hβ2-Cys387Hβ2
C1-C3

Chapter 3
Conformational Significance of O-glycosylations in NOTCH1 EGF Domains

3-1. Introduction

Next, our interest turned to the significance of dynamic *O*-glycosylation states in the three-dimensional (3D) structure of EGF-like domains 11 and 10. Chemical synthesis allowed for the structural characterization of variously *O*-glycosylated domain 12 modules, in which NMR structure of the correctly folded human NOTCH1 EGF-like domain 12 is perfectly identical with that of mouse Notch1 EGF-like domain 12.[1] NMR studies of the correctly folded EGF12 modules allowed us to understand that posttranslational *O*-glycosylation also contributes to the stabilization of the β-hairpin involved in the functional antiparallel β-sheet structure of this pivotal domain through the interaction of sugars with neighboring amino acid residues, namely the intramolecular "sugar-bridge" structure.[2] Three-dimensional (3D) structure of the correctly folded human NOTCH1 EGF12 domain is perfectly identical with that of mouse Notch1 EGF12 domain. We have demonstrated that conformational stabilization of the functional EGF 12 domain depends strongly on the *O*-glycosylation status both at Thr466 and Ser458 residues. However, our understanding on the significance of a variety of *O*-glycosylation patterns found in other important EGF-like repeats in human NOTCH1 receptor, particularly the molecular basis as to how dynamic *O*-glycosylation affect the folding and conformation of each EGF domain remain unclear. In the present study, our interest is focused on the significance of dynamic *O*-glycosylation in the folding and 3D structures of human NOTCH1 EGF11 and EGF10 domains. Novel synthetic EGF modules allow for deciphering significance of the site-specific *O*-glycosylation in the conformational alteration of the Ca^{2+}-binding EGF11 and non-Ca^{2+}-binding EGF10 domains and underlying molecular mechanisms that may determine the integrity of the cell surface NOTCH1 receptor.

3-2. Result and Discussion

3-2-1. Comprehensive NMR studies uncover domain-specific molecular mechanisms in the construction of functional conformations of glycosylated EGF-like repeats.

To gain the structural insights into the folded domains 11 and 10, we performed NMR measurements of the purified modules **1**~**9**. Two-dimensional NMR spectra such as TOCSY, NOESY, DQFCOSY, and ^{13}C-edited HSQC spectra were recorded in 90% H_2O/10% D_2O or 99.9% D_2O solution (Figures S2-10~2-13, S3-1, and S3-2). In 2D NMR, the sequential connectivity of the peptide backbone between Hα (i) and HN (i + 1) was detected in all amino acid residues except for proline in NOESY spectra, and all side chain protons as well as sugar protons were assigned using TOCSY and NOESY connection. ^{1}H-NMR and ^{13}C-NMR chemical shifts of the synthetic EGF11 and EGF10 modules **1**~**9** were summarized in Tables S3-1~S3-4.

As shown in the TOCSY spectra of domain 11 modules **1**~**3** (Figure S3-1) and **4** (Figure 3-1A), Ca^{2+} ion strongly influences the chemical shifts of Hα and HN protons within the β-sheet structure. Obviously, cross peak between Hα and HN protons of several amino acid residues could not be observed when NMR experiment was conducted in the absence of Ca^{2+}, indicating that Ca^{2+} ion coordination stabilizes the total structure of domain 11 including antiparallel β-sheet region. Conversely, in case of non-Ca^{2+}-binding domain 10 modules **5**~**9**, TOCSY spectra of module **5**~**8** (Figure S3-1) and **9** (Figure 3-1B) measured in the absence and presence of Ca^{2+} ion were completely identical, demonstrating that Ca^{2+} does not affect 3-1D structures of domain 10.

NOESY spectra of synthetic EGF11 modules in the presence of Ca^{2+} revealed that sugar

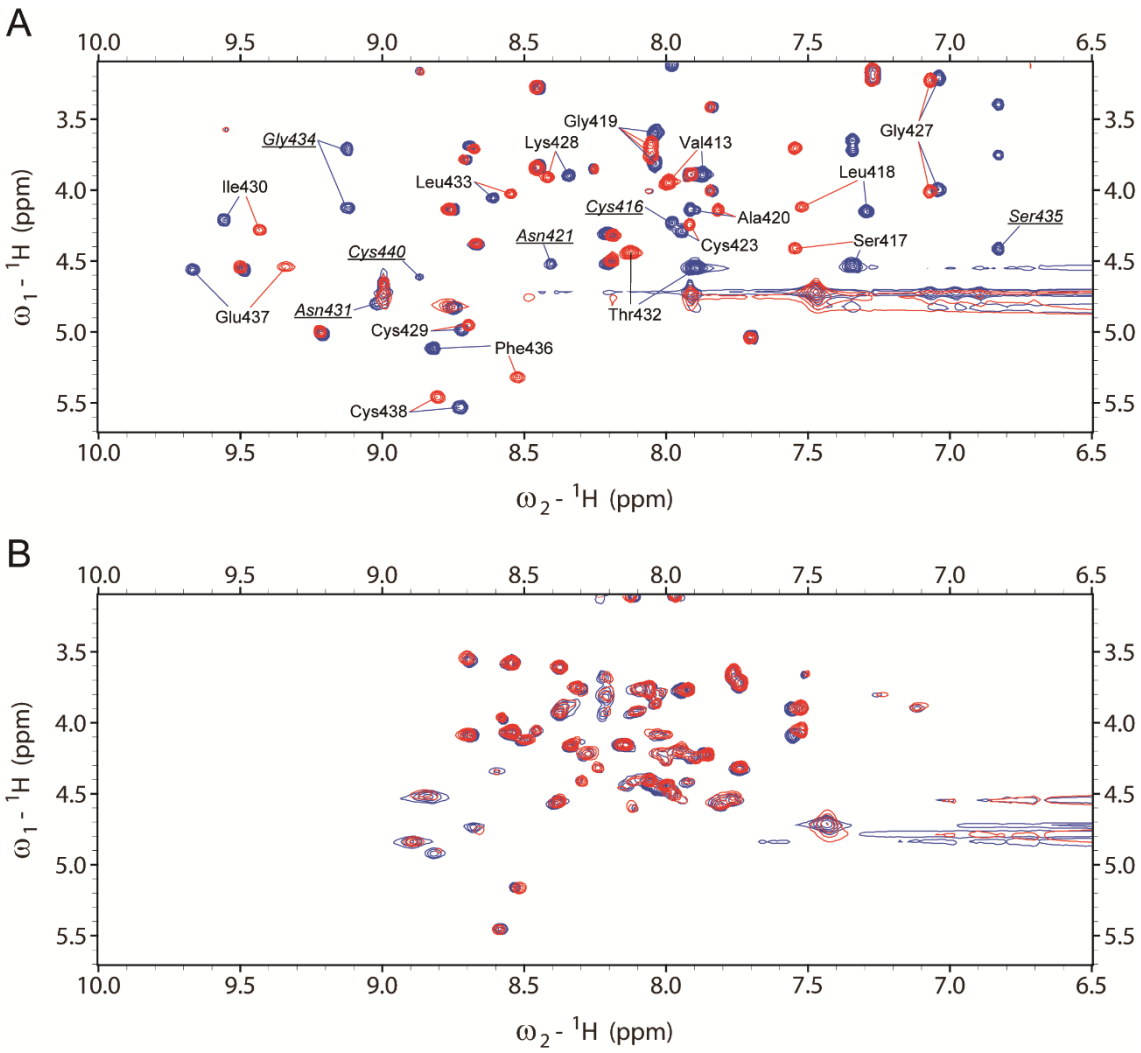
residues of modules **2** and **3** have characteristic NOE interactions with the peptide moiety (Figure S2-12). Notably, *O*-GlcNAc residue has NOE interactions with proximal Thr445 and with Ile451 located at *C*-terminus of EGF11 domain, namely hinge region between EGF11 and EGF12. It seemed likely that the intramolecular sugar-bridge is constructed by the hydrophobic interaction of Ile451 with *N*-acetyl methyl group of the *O*-GlcNAc residue. In contrast, *O*-glucose residue did not show any NOE connection with the peptide region while NOE interactions with the proximal Ser435 residue were observed. Furthermore, NOESY spectra of module **4** having two *O*-glycans, Glcβ1→ at Ser435 and GlcNAcβ1→ Thr445, exhibit quite similar NOE contacts to those observed in the modules **2** and **3**, concurrently.

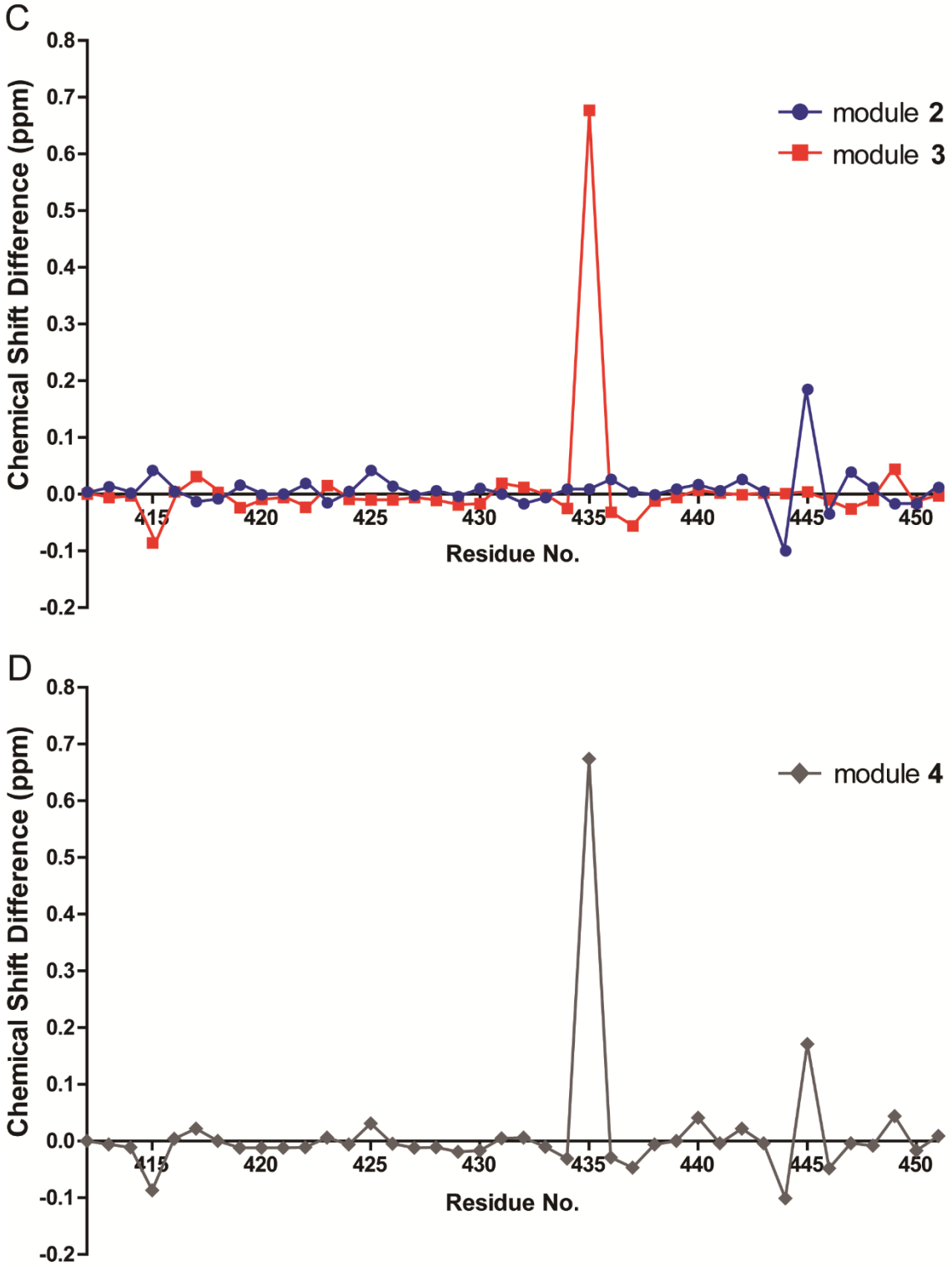
NOESY experiments of non-Ca^{2+}-binding EGF10 modules **6**~**9** also demonstrated the occurrence of characteristic NOEs between sugar residues and the peptide regions (Figure S2-13). As anticipated, EGF10 module **6** showed NOEs indicating the sugar-bridge from *O*-GlcNAc attached at Thr405 to Gln411 residue locating at *C*-terminus of this domain. Given that EGF11 modules **2** and **4** harboring GlcNAcβ1→ moiety at Thr445 form the sugar-bridge with Ile451 in a similar manner, we speculated that sugar-bridge formed by GlcNAcβ1→ moiety at Thr/Ser residue in the consensus sequence C5-X-X-G-X-(T/S)-G-X-X-C6 and an amino acid residue between the two consecutive EGF repeats can be a general structure that contributes to the stabilization of the connecting hinge region. On the other hand, NOE interactions in the *O*-glucose residue of module **7** were found only with the proximal Ser378 residue. However, it was demonstrated that module **8** exhibits specific NOE connections between the anomeric proton of the non-reducing distal xylose residue of the trisaccharide Xylα1→3Xylα1→3Glcβ1→ moiety at Ser378 residue

within the consensus sequence C1-X-S-X-(P/A)-C2 and the α-proton of Lys395 in the β-sheet region. It is important to note that the module **9** having two *O*-glycans, Xylα1→ 3Xylα1 → 3Glcβ1 → and GlcNAcβ1 →, also exhibits abovementioned key NOE interactions indicating the presence of two intramolecular "sugar-bridges" found separately in modules **6** and **8**.

To assess the impact of the individual *O*-glycosylation on the peptide backbone, the chemical shifts of α-protons (Hα) of EGF11 modules **2**~**4** and EGF10 modules **6**~**9** were compared with those of non-glycosylated EGF11 module **1** and EGF10 module **5**, respectively. In the EGF11 domain, it was indicated that *O*-glucosylation at Ser435 within the novel consensus sequence C3-X-N-T-X-G-S-F-X-C4[3] induces distinctly lower-field chemical shift (~0.7 ppm) of the α-proton of the Ser435 residue while that of the Thr445 residue caused by *O*-GlcNAcylation is only 0.2 ppm (Figures 3-1C and 3-1D). In contrast, changes in the chemical shift of the α-protons by *O*-glucosylation at Ser378 in C1-X-S-X-(P/A)-C2 as well as *O*-GlcNAcylation at Thr405 of the EGF10 domain were found to be approximately 0.2 ppm (Figures 3-1E and 3-1F), suggesting that *O*-glucosylation at Ser435 residue of EGF11 domain may have a distinctive role in the interaction of NOTCH1 receptor with ligands. Indeed, it seems likely that the glucose moiety and the proximal peptide region in EGF11 repeat are involved in the interface with DLL4 and Jag1.[4,5] It was also revealed that chemical shifts of the β-protons of Ile451 exhibit upper-field shift over 0.2 ppm in EGF11 modules **2** and **4** (Table S3-1). This result is consistent with NOESY of EGF11 modules **2** and **4** showing the correlation between GlcNAc attached to Thr445 and Ile451 (Figure S2-12), notably the sugar-bridge in the hinge region between EGF11 and EGF12. Interestingly, EGF10 modules **8** and **9** having xyloses

induced significantly ~0.2 ppm of the lower-field shift in the Hα proton of the Ala396 residue (Figures 3-1E and 3-1F), whereas the NOE experiment revealed the connection in the sugar-bridge between the nonreducing xylose and Lys395 involved in the antiparallel β-sheet (Figure S2-13).


A
B
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
3.5
4.0
4.5
5.0
5.5
ω1 - 1H (ppm)
ω2 - 1H (ppm)
Gly434
Ile430
Leu433
Lys428
Gly419
Val413
Gly427
Leu418
Cys416
Ala420
Ser435
Asn421
Cys423
Ser417
Cys440
Glu437
Asn431
Cys429
Thr432
Phe436
Cys438


C
Chemical Shift Difference (ppm)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
415
420
425
430
435
440
445
450
Residue No.
module 2
module 3
D
Chemical Shift Difference (ppm)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
415
420
425
430
435
440
445
450
Residue No.
module 4

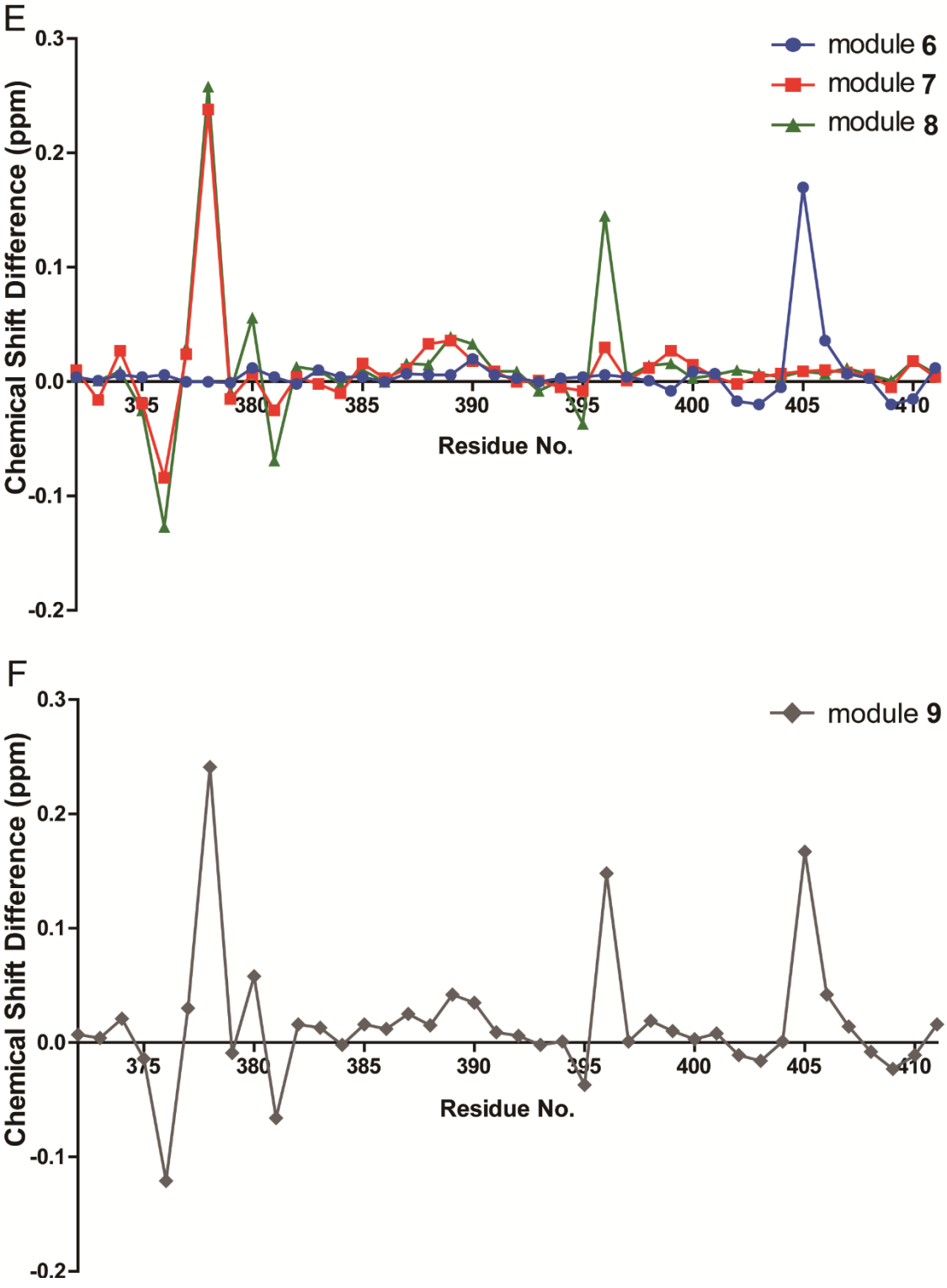

Chemical Shift Difference (ppm)
0.3
0.2
0.1
0.0
-0.1
-0.2
375
380
385
390
395
400
405
410
Residue No.
module 6
module 7
module 8
Chemical Shift Difference (ppm)
0.3
0.2
0.1
0.0
-0.1
-0.2
375
380
385
390
395
400
405
410
Residue No.
module 9

Figure 3-1. NMR studies of synthetic EGF modules elicit significance of site-specific *O*-glycosylation in the stabilization of functional conformation of EGF11 and EGF10 repeats. TOCSY spectra of folded non-glycosylated EGF11 (**1**) (A), and EGF10 (**5**) (B) in the absence (red) and presence (blue) of Ca^{2+}(see also TOCSY spectra of glycosylated EGF modules **2**~**4** and **6**~**9** shown in Figure S3-1 and Figure S3-2). (C) and (D) represent the differences in the chemical shifts of Hα protons between EGF11 modules **2**~**4** and non-glycosylated EGF11 (**1**). (E) and (F) show those between EGF10 modules **6**~**9** and non-glycosylated EGF10 (**5**).

3-2-2. 3D NMR structures highlight significance of the site-specific O-glycosylation in the correctly folded EGF11 and EGF10 domains.

The 3D structures of correctly folded non-glycosylated and fully glycosylated EGF11 modules **1** and **4**, and those of EGF10 modules **5** and **9** were generated by calculations using distance and angle restrictions estimated from two-dimensional NMR experiments (Tables 3-1 and 3-2). 3D NMR structures showed that all synthetic modules can adopt antiparallel β-sheet structures in the central area of domains 11 and 10 (Figure 3-2A). Judging from the well-converged structures of the non-glycosylated module **1** of domain 10 with fully glycosylated module **4** (Figure 3-2B), *O*-glycosylation appears to show only limited effect on the antiparallel β-sheet structures in the Ca^{2+}-binding domain 11. However, the difference between non-glycosylated module **5** of domain 10 and fully glycosylated module **9** was distinct in the antiparallel β-sheet structures involving β-hairpin of this domain, suggesting that the glycosylation influences significantly the conformation of non- Ca^{2+}-binding domain 10 (Figure 3-2B). The overlaying structure of correctly folded module **4** of domain 11 or module **9** of domain 10 elicited by solution NMR and the corresponding domains in the X-ray crystalline structure of recombinant human NOTCH1 EGF8~12 repeats[5] were shown to converge well while regions of the *N*-terminus of domain 11 and C-terminus of domain 10 do not appear to be converged between NMR and X-ray structures (Figure 3-2C).

As shown in Figure 3-2D, the 3D NMR structures clearly indicated importance of abovementioned key NOEs between *O*-glycans and peptide moieties observed in fully glycosylated module **4** and module **9**, respectively. We demonstrated the presence of the specific interaction between *O*-GlcNAc attached to Thr445 and Ile451 residing in the

hinge region between domains 11 and 12. Atomic distance of *N*-acetyl methyl carbon of the GlcNAc with Cβ of the Ile451 in the mean structure of fully glycosylated module **4** of domain 11 was found to be 3.8 Å, indicating that the *O*-GlcNAc moiety interacts hydrophobically with Ile451. This result suggests that *O*-GlcNAcylation may contribute to the regulation of flexibility/rigidity of this connecting region through interaction with Ile451. Similarly, GlcNAcylation at Thr405 of domain 10 within the same consensus sequence C5-X-X-G-X-(T/S)-G-X-X-C6 as GlcNAc at Thr445 facilitated the hydrophobic interaction with Gln411 residue in the hinge region between domains 10 and 11. Atomic distance of *N*-acetyl methyl carbon of the GlcNAc with Cβ of Gln411 was found to be 3.9 Å, demonstrating occurrence of the new type "sugar bridge" crosslinking Thr-*O*-GlcNAc in the consensus sequence C5-X-X-G-X-(T/S)-G-X-X-C6 and an amino acid in the hinge region between these domains, 445Thr-*O*-GlcNAc---Gln411 in domain 11 and 405Thr-*O*-GlcNAc---Ile451 in domain 10, respectively.

Intriguingly, no specific NOE correlation between the glucose attached to Ser435 and any amino acid residue is observed in module **4** of domain 11, suggesting that the *O*-glucosylation at Ser435 within the new consensus sequence C3-X-N-T-X-G-S-F-X-C4[3] may have distinct functional roles from that at C1-X-S-X-(P/A)-C2. As we have shown the importance of *O*-glucosylation and subsequent xylose extension at Ser458 in the domain 12,[2] the Xylα1→3Xylα1→3Glcβ1→ moiety at Ser378 within C1-X-S-X-(P/A)-C2 of the module **9** of domain 10 has quite similar characteristics to the trisaccharide *O*-glycan attached to Ser458 residue of the domain 12. Atomic distance of the non-reducing xylose residue of this trisaccharide moiety at Ser378 with Hα proton of the Lys395 residue in the mean structure of fully glycosylated module **9** was found to be ~4.0 Å,

indicating that the terminal xylose can interact hydrophobically with Lys395 (Figure 3-2D). It was likely that the *O*-glucosylation and xylose extension at Ser378 residue cover the hydrophobic pocket formed by Ala375, Pro460, and Ala474 residues (Figure 3-2E). This interaction may enhance solubility of the antiparallel β-sheet region by shielding this hydrophobic pocket from solvent exposure, as indicated by the difference in the chemical shift of Hα proton at Ala396 residue between non-glycosylated module **5** and fully glycosylated module **9** of domain 10 (Table S3-3). Altogether, the sugar bridge 378Ser-*O*-Glc-Xyl-Xyl---Lys395 contributes to the conformational stabilization of non-Ca^{2+}-binding domain 10, though this trisaccharide acts negatively in the *in vitro* folding process of this domain.

Table 3-1. Statistics of the NMR structures of synthetic NOTCH1 EGF modules.

	EGF11 Non-glycosylated (**1**)	EGF11 Glc GlcNAc (**4**)	EGF10 Non-glycosylated (**5**)	EGF10 XXG GlcNAc (**9**)
Average potential energies (kcal mol^{-1})[a]				
E_{total}	57.34 ± 1.47	63.52 ± 1.16	19.85 ± 0.94	35.76 ± 1.15
E_{bonds}	3.09 ± 0.18	3.07 ± 0.21	0.98 ± 0.09	1.42 ± 0.16
E_{angle}	14.91 ± 0.50	17.35 ± 0.93	8.83 ± 0.40	13.90 ± 0.66
E_{impr}[b,c]	1.65 ± 0.26	2.77 ± 0.39	0.80 ± 0.11	2.53 ± 0.30
E_{VDW}[b,c]	11.84 ± 0.80	12.68 ± 0.81	4.53 ± 0.46	8.10 ± 0.68
E_{NOE}[b,c]	25.46 ± 1.49	24.87 ± 2.06	4.60 ± 0.48	5.87 ± 0.57
E_{cdih}[b,c]	0.39 ± 0.20	0.55 ± 0.31	0.12 ± 0.04	0.24 ± 0.06
Deviations from idealized geometry				
Bond lengths (Å)	0.0023 ± 0.0001	0.0022 ± 0.0001	0.0014 ± 0.0001	0.0015 ± 0.0001
Bond angles (°)	0.31 ± 0.01	0.45 ± 0.01	0.25 ± 0.01	0.48 ± 0.01
Impropers (°)	0.19 ± 0.01	0.26 ± 0.01	0.14 ± 0.01	0.28 ± 0.02
Average pairwise r.m.s. deviation (Å)				
Cys416-Asn431, Ser435-Gln442 (EGF11) / Cys376-Thr389, Lys395-Ser402 (EGF10)				
Backbone atoms	0.99 ± 0.30	0.94 ± 0.25	0.96 ± 0.22	0.80 ± 0.21
Heavy atoms	1.63 ± 0.29	1.64 ± 0.29	1.50 ± 0.25	1.37 ± 0.26

[a] All energies and root mean square values were calculated by using programs CNS1.1 and MOLMOL, respectively.

[b] E_{impr}, E_{VDW}, E_{NOE} and E_{cdih} are the energy of improper torsion angles, the van der Waals repulsion energy, the square-well NOE potential energy, and the dihedral potential energy, respectively.

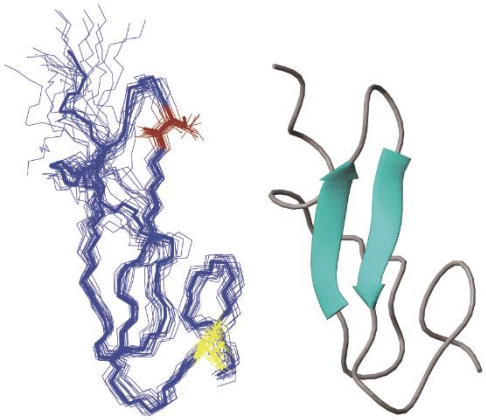
[c] The force constants for the calculations of E_{impr}, E_{VDW}, E_{NOE} and E_{cdih} were 4.0 kcal $mol^{-1}Å^{-4}$, 50 kcal $mol^{-1}Å^{-1}$ and 200 kcal $mol^{-1}rad^{-2}$, respectively.

Table 3-2. NMR restraints for structural calculation of synthetic EGF analogues.

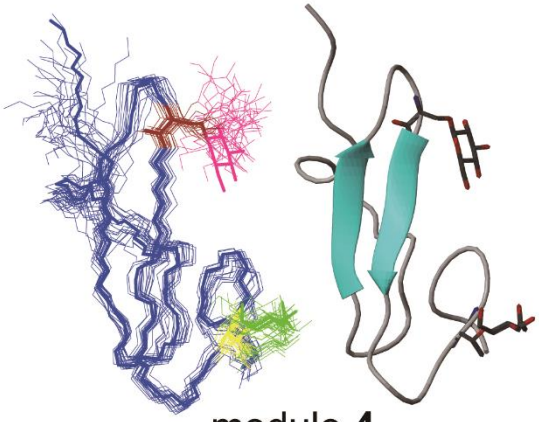
summary of restraints	EGF11 Non-glycosylated (**1**)	EGF11 Glc GlcNAc (**4**)	EGF10 Non-glycosylated (**5**)	EGF10 XXG GlcNAc (**9**)
distance restraints				
total	416	439	390	443
intra-residue				
peptide	263	259	262	260
glycan	0	16	0	37
sequential (|i-j|=1)				
peptide	82	82	74	73
glycan	0	0	0	2
medium-range (2<|i-j|<4)				
peptide	26	18	18	20
glycan	0	0	0	0
longer-range (|i-j|>5)				
peptide	45	52	36	38
glycan	0	0	0	0
peptide to glycan				
within the same glycosylated residue	0	10	0	9
glycans on other peptide residues	0	2	0	4
dihedral restraints				
total	29	43	22	48
peptide backbone	29	29	22	22
glycans	0	14	0	26

A

EGF11

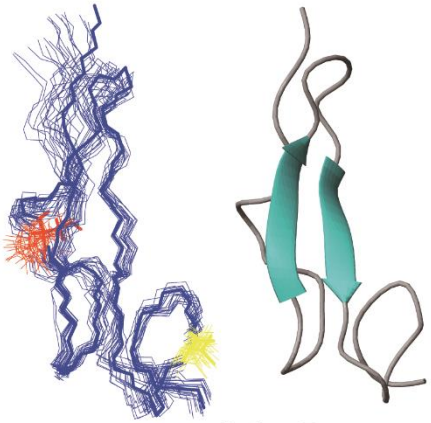


module **1**

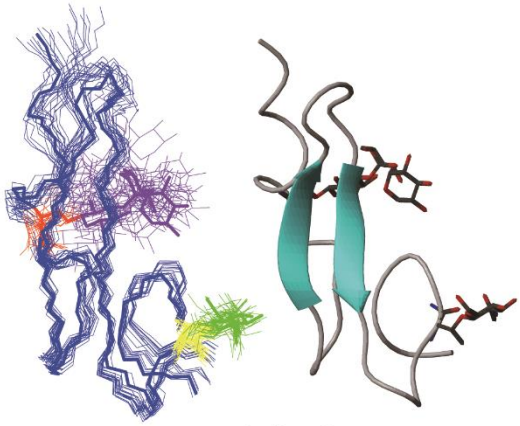


module **4**

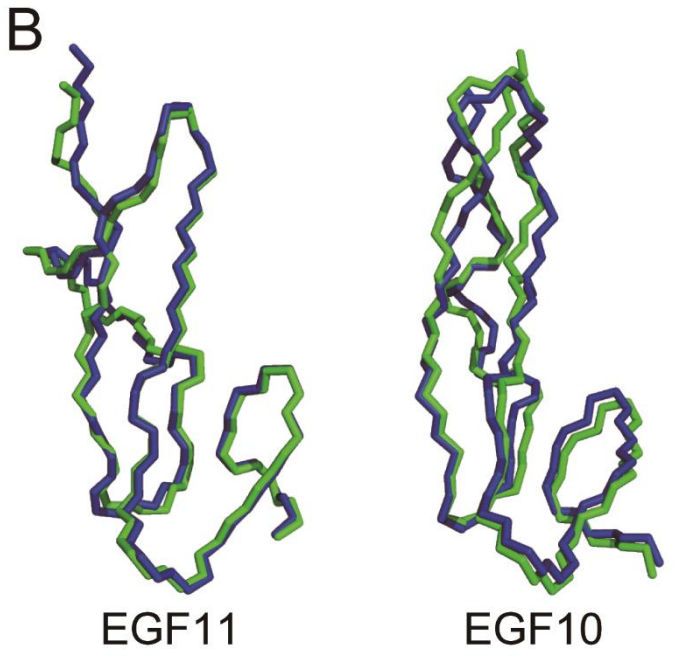
EGF10

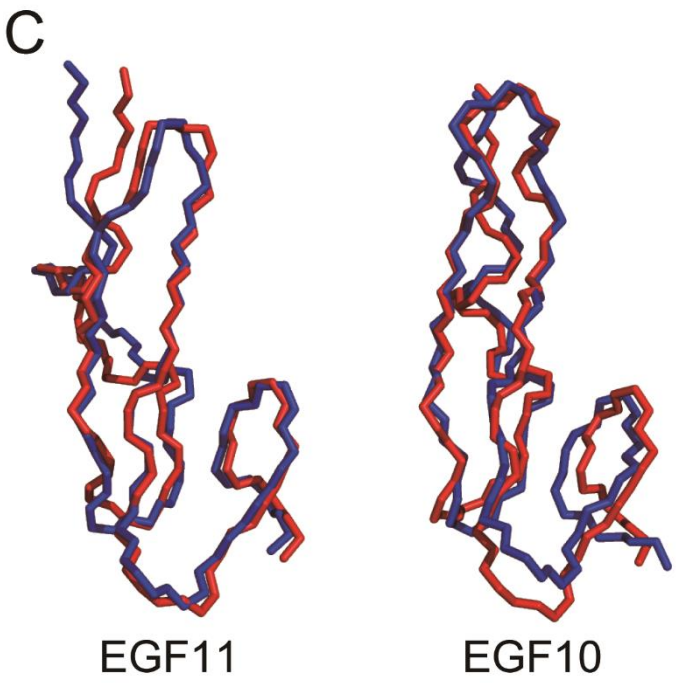


module **5**



module **9**


B
EGF11
EGF10


C
EGF11
EGF10

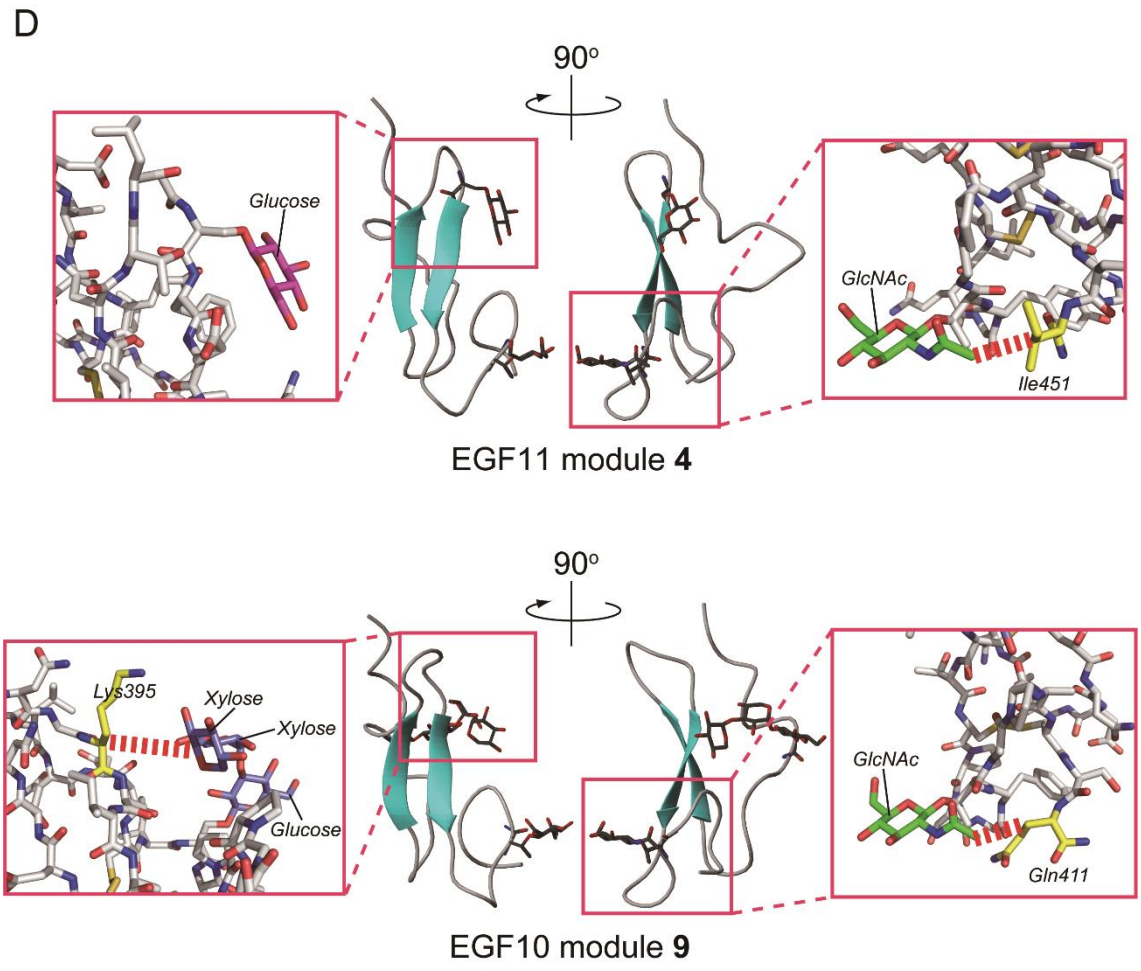

D
90°
Glucose
GlcNAc
Ile451
EGF11 module 4
90°
Lys395
Xylose
Xylose
Glucose
GlcNAc
Gln411
EGF10 module 9
E
90°
Ala375
Ala396
Pro380
Xyl-Xyl-Glc
GlcNAc
EGF10 module 5
EGF10 module 9

Figure 3-2. NMR structures of human NOTCH1 EGF11 and EGF10 domains. (A) Schematic representation of the most energy-minimized structure. The backbone structures of the 20 lowest-energy structures are shown at the left side. These structures are superimposed on the region of Cys416-Asn431 and Ser435-Gln442 (EGF11) / Cys376-Thr389 and Lys395-Ser402 (EGF10), and mean structures are represented by a bold line. The root-mean-square deviations of backbone atoms are 0.99 Å for non-glycosylated EGF11 module **1**, 0.94 Å for module **4**, 0.96 Å for module **5**, and 0.80 Å for module **9**. The structures are oriented with the *N*-terminus at the top. The peptide backbones are colored blue, and Ser435 and Thr445 in EGF11 (*O*-glycosylation sites) are colored brown and yellow, and Ser378 and Thr405 in EGF10 (*O*-glycosylation sites) are colored orange and yellow, respectively. Glucose residue at Ser435 and GlcNAc residues at Thr445 and Thr405 are colored pink and green, respectively. The Xylα1→3Xylα1→3Glcβ1→ moiety are colored purple. The ribbon structures of most energy-favored structures are shown at the right side. The □-sheet structures are predicted by arrows, and the heavy atoms of carbohydrate moieties and the underlying Ser/Thr residues are shown as sticks. Oxygen, nitrogen, and carbon atoms are colored red, blue, and black, respectively. (B) Comparison of NMR structures of human NOTCH1 non-glycosylated (green) and glycosylated (blue) EGF11 modules (left) and EGF10 modules (right). (C) The superimposed structure resolved by the present NMR study and X-ray crystallography of the human NOTCH1 EGF11 domain (left) and EGF10 domain (right). The blue stick is for the backbone of human EGF resolved in this study, and the cyan stick is for the EGF10 or EGF11 domain of human NOTCH1 EGF8~12 repeats generated by X-ray crystallography (PDB entry 5UK5).[5] (D) NMR structures focusing on the interactions between sugar moieties and amino acid residues, namely “sugar bridge”. (E) NMR structures focusing on the shielding effect of the glycosylation at Ser378 on the hydrophobic pocket. Backbone structures of the 20 lowest-energy structures of modules **5** and **9** are colored blue, Xylα1→3Xylα1→3Glcβ1→ moiety purple, and Ala375, Pro380, and Ala396 orange. The mean structure focused amino acid residues (Ala375, Pro380, and Ala396) and glycans are denoted by bold lines.

3-3. Conclusion

Comprehensive NMR studies of variously *O*-glycosylated models for NOTCH1 EGF-like domains 10~12 uncovered effects of site-specific *O*-glycosylation on the structural behavior of the individual domains, particularly, in the conformational stabilization by "sugar bridges". Figure 3-3 summarizes structural significance of the site-specific *O*-glycosylation and Ca^{2+} in these three crucial domains 10~12 of the NOTCH1 receptor. Novel synthetic EGF modules **1**~**9** accelerated NMR studies for deciphering importance of various *O*-glycosylation states in the structures of the Ca^{2+}-binding EGF11 and non-Ca^{2+}-binding EGF10 domains, respectively. Our results demonstrated for the first time the presence of the specific interaction of *O*-GlcNAc residue attached to Thr445 with Ile451 residing in the hinge region between EGF11 and EGF12 domains. Furthermore, GlcNAcylation at Thr405 of EGF10 domain within the same motif C5-X-X-G-X-(T/S)-G-X-X-C6 (GlcNAc residue attached to Thr405) also facilitated the interaction with Gln411 residue within the hinge region between EGF10 and EGF11 domains. These results clearly indicate that *O*-GlcNAcylation-mediated sugar bridge as well as Ca^{2+} coordination contributes to the regulation of flexibility and rigidity of the connecting regions of NOTCH1 extracellular domains. As expected, the sugar bridge 378Ser-O-Glc-Xyl-Xyl---Lys395 affects the conformational stabilization of non-Ca^{2+}-binding domain 10, whereas no NOE correlation between the glucose attached to Ser435 and any amino acid residue is observed in Ca^{2+}-binding domain 11. The result suggests that the *O*-glucosylation at Ser435 within C3-X-N-T-X-G-S-F-X-C4 may have unidentified roles in Notch receptor function, while mutation of Ser435 to Ala affects NOTCH1 function.[6]

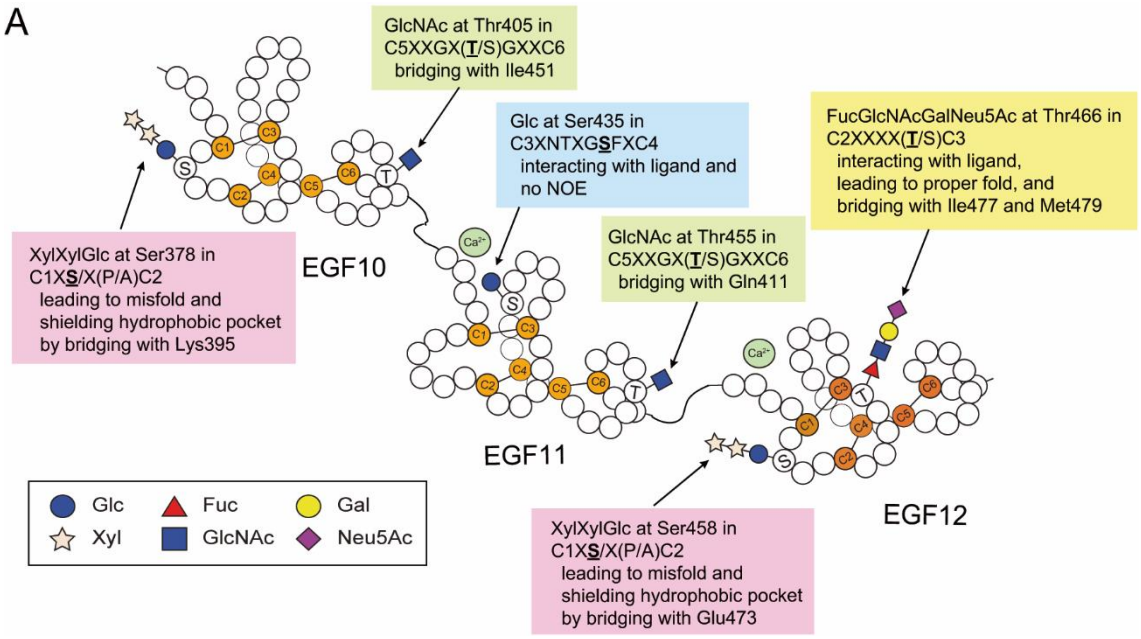


Figure 3-3. Significance of site-specific *O*-glycosylation in NOTCH1 extracellular EGF-like domains. Schematic representation for the functional roles of the site-specific *O*-glycosylation in the human NOTCH1 EGF-like repeats 10~12.

3-4. Experimental Section

General methods and materials

All ^{1}H- and ^{13}C-NMR spectra for identification of synthetic peptides were collected with 600 MHz Bruker AVANCE (Bruker Biospin Co., Germany). Two-dimensional homonuclear double-quantum-filtered scalar-correlated spectroscopy (DQF-COSY), TOCSY with MLEV-17 sequence, and nuclear Overhauser enhancement spentroscopy (NOESY) spectra were recorded in the indirect dimension using State-TPPI phases cycling.

NMR studies

NMR experiments of synthetic EGF modules **1**~**9** were performed on a Bruker Avance 600 spectrometer at 600.03 MHz for proton frequency equipped with a cryoprobe. EGF modules **1**~**9** were dissolved in 99.9% D_2O (300 μL) or in a mixture solution of 90% H_2O and 10% D_2O (300 μL) at the concentration of 2.0 mM, respectively. In the case of EGF11 modules **1**~**4**, each solution contained 25 mM $CaCl_2$. These solutions were adjusted at pH 5.0~5.3 by 0.1 N NaOHaq. These peptide solutions were packed in *Shigemi* thin-walled micro NMR tubes in NMR spectroscopy. All spectra were measured at 298 K. Data acquisition was performed with the Bruker TopSpin 2.1 software package, Two-dimensional[7] homonuclear DQF-COSY,[8,9] TOCSY[10] with MLEV-17 sequence[11] and NOESY spectra[12] were recorded in indirect dimension using States-TPPI phase cycling. Two-dimensional heteronuclear ^{13}C-edited HSQC and HSQC-TOCSY measurements were recorded with echo-antiecho mode. TOCSY experiments were applied for a spin-locking time of 60 msec, and NOESY experiments were carried out with mixing time of 100, 200 and 400 msec. The suspension of water signal was performed by presaturation

during 2 sec relaxation delay and by a 3-9-19 WATERGATE pulse sequence with a field gradient. TOCSY and NOESY spectra were acquired with 2048 by 512 frequency data point and were zero-filled to yield 2048 by 2048 data matrices. DQF-COSY with 16384 by 512 frequency data points was also recorded and zero-filled to yield 16384 by 16384 matrix to measure the coupling constant. The sweep widths of 7183.91 Hz were applied. Time domain data in both dimensions were multiplied by a sine bell window function with a 90^{o} phase shift prior to Fourier transformation. All NMR data were analyzed using a Sparky program[13] and were shown in the Supporting Information (Figures S2-10~S2-13, Figures S3-1~S3-2 and Tables S3-1~S3-4).

Structural calculation

Three-dimensional structures of synthetic EGF11 modules **1** and **4**, and EGF10 modules **5** and **9** were calculated using the CNS 1.1 program[14] with standard protocols for distance geometry-simulated annealing and refinement. Distance restraints for calculations were estimated from the cross-peak intensities in NOESY spectra with a mixing time of 400 msec. The estimated restraints were classified into four categories: strong (1.6~2.6 Å), medium (1.6~3.5 Å), weak (1.6~5.0 Å) and very weak (1.6~6.0 Å). In the first stage of structure determination, the structures of glycopeptides were calculated using only interproton distance information. After the validation of fulfilling distance restraints for the obtained structures, the restraints of the dihedral angle ϕ were adopted for further structural calculation. When the coupling constant $^{3}J_{HN\alpha}$ was more than 8.0 Hz, the dihedral angle ϕ was constricted to -120 ± 30°. The conformation of sugar ring was fixed to the chair conformation. All analyses of rmsd values and the solution structures of glycopeptides were performed with PROCHECK[15] and MOLMOL[16] program. The NMR

structures of all synthetic EGF modules were analyzed by MOLMOL and PyMOL program.

3-5. Reference

(1) Hiruma-Shimizu, K.; Hosoguchi, K.; Liu, Y.; Fujitani, N.; Ohta, T.; Hinou, H.; Matsushita, T.; Shimizu, H.; Feizi, T.; Nishimura, S. -I. Chemical Synthesis, Folding, and Structural Insights into O-Fucosylated Epidermal Growth Factor-like Repeat 12 of Mouse Notch-1 Receptor. *J. Am. Chem. Soc*. **2010** *132*, 14857-14865.

(2) Hayakawa, S.; Yokoi, Y.; Hinou, H.; Nishimura, S.-I. Chemical Synthesis Demonstrates That Dynamic *O*-Glycosylation Regulates the Folding and Functional Conformation of a Pivotal EGF12 Domain of the Human NOTCH1 Receptor. *Biochemistry* **2017**, *56*, 4379-4391.

(3) Takeuchi, H.; Schneider, M.; Williamson, D. B.; Ito, A.; Takeuchi, M.; Handford, P. A.; Haltiwanger, R. S. Two novel protein *O*-glucosyltransferases that modify sites distinct from POGLUT1 and affect Notch trafficking and signaling. *Proc. Natl. Acad. Sci. U S A.* **2018**, *115*, E8395-E8402.

(4) Luca, V. C.; Jude, K. M.; Pierce, N. W.; Nachury, M. V.; Fischer, S.; Garcia, K. C. Structural basis for Notch1 engagement of Delta-like 4. *Science* **2015**, *347*, 847-853.

(5) Luca, V. C.; Kim, B. C.; Ge, C.; Kakuda, S.; Wu, D.; Roein-Peikar, M.; Haltiwanger, R. S.; Zhu, C.; Ha, T.; Garcia, K. C. Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity. *Science* **2017**, *355*, 1320-1324.

(6) Takeuchi, H.; Schneider, M.; Williamson, D. B.; Ito, A.; Takeuchi, M.; Handford, P. A.; Haltiwanger, R. S. Two novel protein *O*-glucosyltransferases that modify sites distinct from POGLUT1 and affect Notch trafficking and signaling. *Proc. Natl. Acad. Sci. U S A.* **2018**, *115*, E8395-E8402.

(7) States, D. J.; Haberkorn, R. A.; Ruben, D. J. A Two-Dimensional Nuclear Overhauser Experiment with Pure Absorption Phase in Four Quadrants. *J. Magn. Reson.* **1982**, *48* (2), 286-292.

(8) Piantini, U.; Sorensen, O. W.; Ernst, R. R. Multiple Quantum Filters for Elucidating NMR Coupling Networks. *J. Am. Chem. Soc.* **1982**, *104* (24), 6800-6801.

(9) Rance, M.; Sørensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. Improved Spectral Resolution in Cosy ^{1}H NMR Spectra of Proteins via Double Quantum Filtering. *Biochem. Biophys. Res. Commun.* **1983**, *117* (2), 479-485.

(10) Braunschweiler, L.; Ernst, R. R. Coherence Transfer by Isotropic Mixing: Application to Proton Correlation Spectroscopy. *Journal of Magnetic Resonance (1969)*, 1983, *53*, 521-528.

(11) Bax, A.; Davis, D. G. MLEV-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. *Journal of Magnetic Resonance (1969)*, 1985, *65*, 355-360.

(12) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. Investigation of Exchange Processes by Two-Dimensional NMR Spectroscopy. *J. Chem. Phys.* **1979**, *71*, 4546-4553.

(13) Lee, W.; Westler, W. M.; Bahrami, A.; Eghbalnia, H. R.; Markley, J. I. PINE-SPARKY: Graphical Interface for Evaluating Automated Probabilistic Peak Assignments in Protein NMR Spectroscopy. *Bioinformatics* **2009**, *25* (16), 2085-2087.

(14) Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr. D. Biol. Crystallogr.* **1998**, *54* (Pt 5), 905-921.

(15) Laskowski, R. A.; Rullmannn, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR. *J. Biomol. NMR* **1996**, *8* (4), 477–486.

(16) Koradi, R.; Billeter, M.; Wüthrich, K. MOLMOL: A Program for Display and Analysis of Macromolecular Structures. *J. Mol. Graph.* **1996**, *14* (1), 51–55.

3-6. Supporting Information

Figure S3-1. TOCSY spectra of folded EGF11 module **2-4** in the absence (red) and presence (blue) of calcium ions (25 mM $CaCl_2$). Amino acid residues with underscores denote that cross peak between Hα and HN protons cannot be observed when NMR experiment was conducted in the absence of calcium ion.

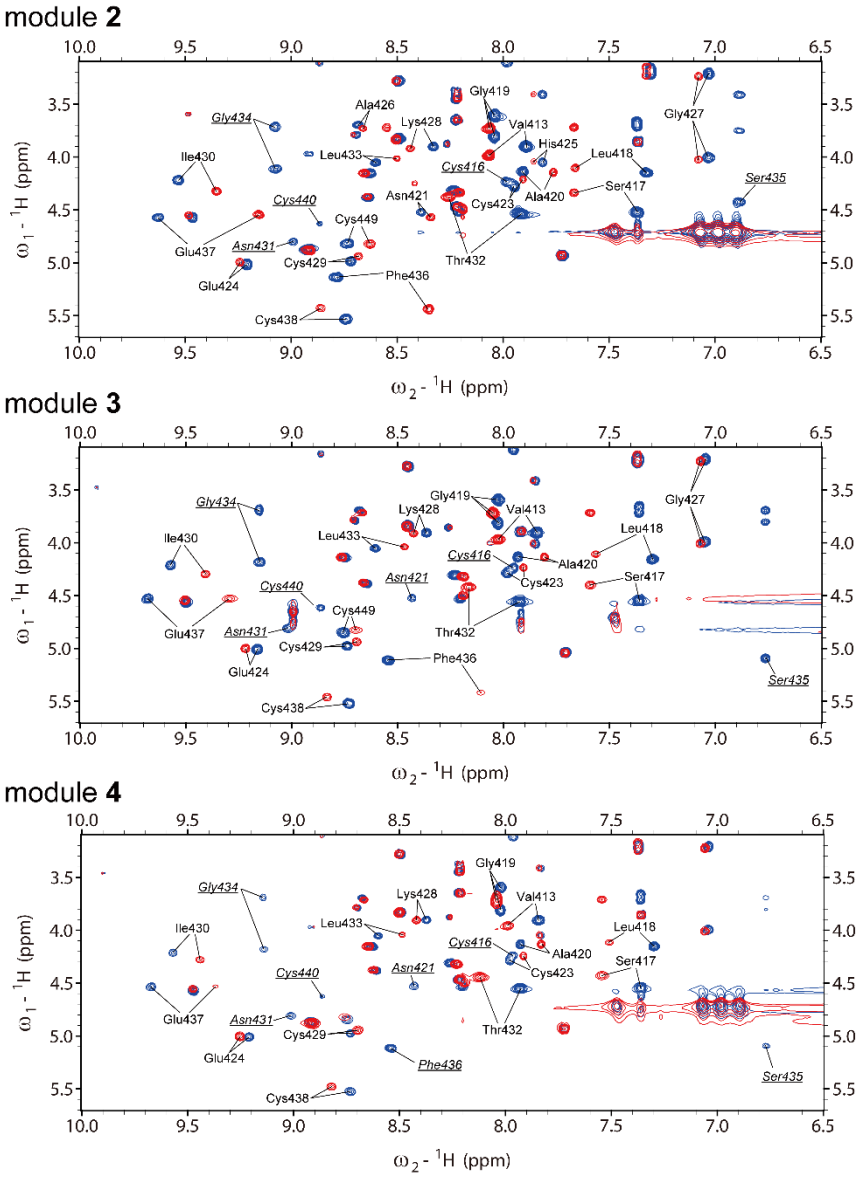


Figure S3-2. TOCSY spectra of folded EGF10 module **6-9** in the absence (red) and presence (blue) of calcium ions (25 mM $CaCl_2$). Amino acid residues with underscores denote that cross peak between Hα and HN protons cannot be observed when NMR experiment was conducted in the absence of calcium ion.

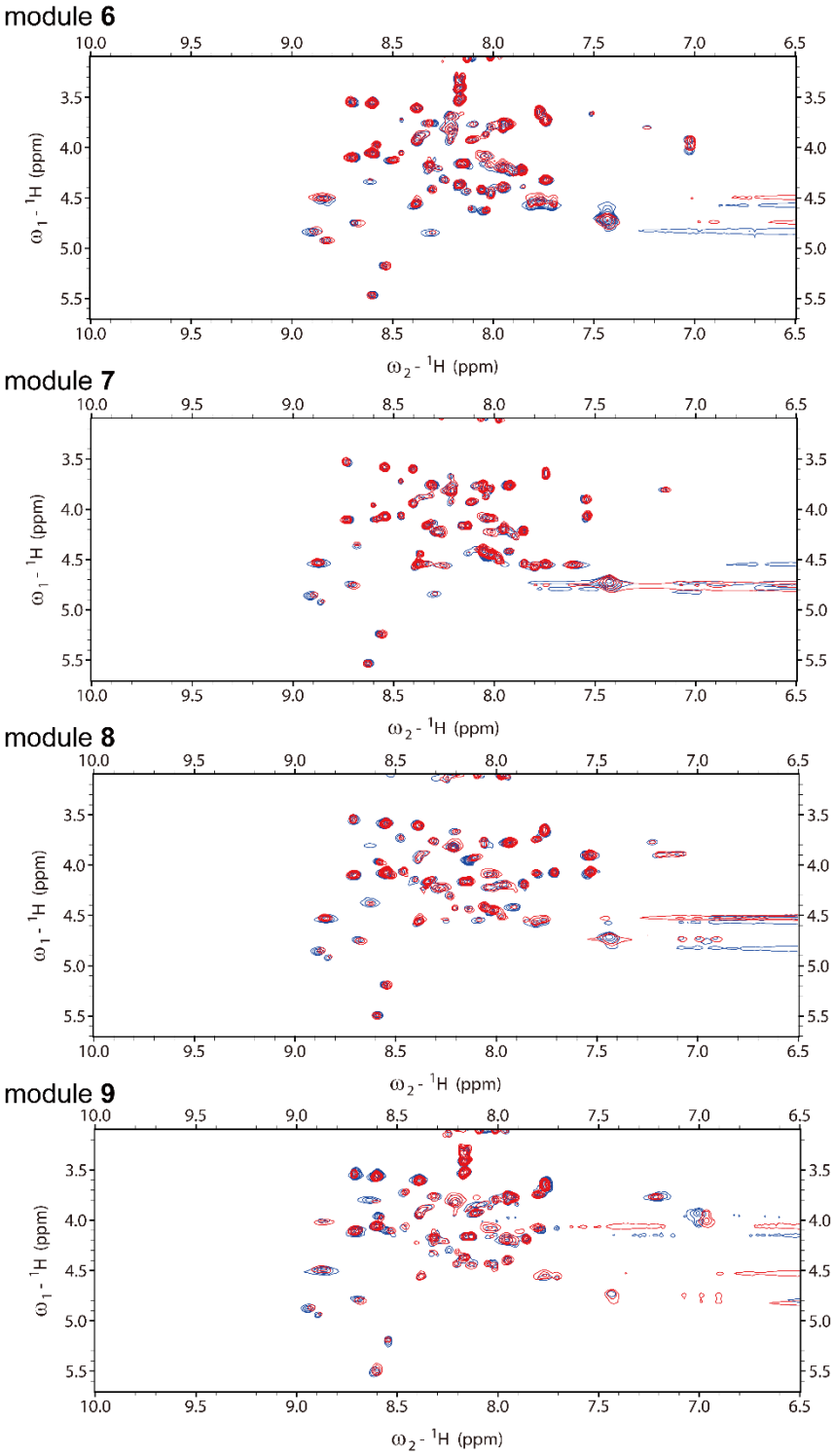


Table S3-1. ^{1}H-NMR chemical shifts of synthetic EGF11 modules **1**-**4**.

Residues		**1**	**2**	**3**	**4**	Residues		**1**	**2**	**3**	**4**
Asp412	HN	8.211	8.219	8.210	8.210	Gly419	HN	8.047	8.049	8.036	8.031
	Hα	4.532	4.536	4.532	4.532		Hα	3.606	3.622	3.582	3.594
	Hβ	2.308	2.346	2.243	2.252			3.817	3.808	3.825	3.825
		2.567	2.567	2.574	2.571	Ala420	HN	7.917	7.904	7.944	7.939
Val413	HN	7.888	7.913	7.833	7.839		Hα	4.149	4.148	4.140	4.137
	Hα	3.909	3.922	3.903	3.903		Hβ	1.130	1.123	1.124	1.127
	Hβ	1.812	1.830	1.784	1.784	Asn421	HN	8.400	8.381	8.457	8.454
	Hγ	0.704	0.717	0.694	0.703		Hα	4.532	4.532	4.526	4.520
		0.810	0.810	0.799	0.805		Hβ	2.746	2.739	2.749	2.746
Asp414	HN	8.708	8.696	8.723	8.715		γNH_2	6.895	6.868	6.868	6.853
	Hα	3.792	3.794	3.789	3.781			7.083	7.091	7.070	7.071
	Hβ	2.177	2.192	2.150	2.158	Pro422	Hα	4.296	4.315	4.273	4.285
		2.277	2.289	2.272	2.275		Hβ	1.870	1.897	1.875	1.888
Glu415	HN	9.554	9.489	9.920	9.907			2.273	2.298	2.267	2.289
	Hα	3.562	3.604	3.476	3.475		Hγ	2.001	2.041	1.989	2.029
	Hβ	1.244	1.330	1.130	1.146		Hδ	3.577	3.622	3.564	3.597
		1.475	1.523	1.384	1.397			4.124	4.143	4.144	4.148
	Hγ	2.013	2.067	1.981	1.993	Cys423	HN	7.957	7.959	7.985	7.983
		2.522	2.506	2.437	2.447		Hα	4.295	4.280	4.310	4.301
Cys416	HN	7.994	7.999	7.989	7.979		Hβ	2.213	2.218	2.214	2.208
	Hα	4.244	4.249	4.248	4.248			2.691	2.691	2.682	2.679
	Hβ	3.117	3.057	3.152	3.146	Glu424	HN	9.201	9.182	9.221	9.252
		-	3.119	-	-		Hα	5.017	5.022	5.008	5.011
Ser417	HN	7.367	7.396	7.353	7.359		Hβ	1.470	1.468	1.416	1.413
	Hα	4.536	4.523	4.567	4.558			1.593	1.596	1.586	1.594
	Hβ	3.672	3.697	3.649	3.652		Hγ	2.414	2.417	2.398	2.404
		3.726	-	3.737	3.731	His425	HN	7.844	7.829	7.855	7.836
Leu418	HN	7.321	7.356	7.289	7.295		Hα	4.022	4.064	4.012	4.053
	Hα	4.164	4.156	4.167	4.164		Hβ	3.030	3.031	3.027	3.024
	Hβ	1.466	1.459	1.454	1.451			3.420	3.415	3.442	3.433
	Hγ	1.517	1.521	1.527	1.527		Hδ2	7.293	7.297	7.285	7.288
	Hδ	0.671	0.682	0.659	0.659		Hε1	8.421	8.433	8.409	8.420
		0.719	0.726	0.703	0.709						

Residues		1	2	3	4	Residues		1	2	3	4
Ala426	HN	8.701	8.698	8.701	8.696	Leu433	HN	8.608	8.599	8.622	8.618
	Hα	3.683	3.697	3.673	3.678		Hα	4.065	4.059	4.064	4.055
	Hβ	1.331	1.335	1.334	1.334		Hβ	1.539	1.543	1.530	1.539
Gly427	HN	7.051	7.041	7.073	7.061		Hγ	1.473	1.477	1.486	1.463
	Hα	3.223	3.221	3.217	3.211		Hδ	0.759	0.761	0.753	0.753
		4.011	4.020	3.994	3.994			0.821	0.828	0.817	0.808
Lys428	HN	8.342	8.328	8.425	8.409	Gly434	HN	9.102	9.045	9.192	9.179
	Hα	3.912	3.918	3.901	3.901		Hα	3.715	3.724	3.690	3.684
	Hβ	1.590	1.587	1.562	1.556			4.124	4.117	4.196	4.193
	Hγ	2.140	2.170	2.074	2.100	Ser435	HN	6.873	6.941	6.734	6.748
	Hδ	1.397	1.406	1.381	1.396		Hα	4.434	4.443	5.111	5.108
	Hε	2.315	2.351	2.246	2.281		Hβ	3.416	3.432	3.699	3.699
	εNH_2	7.294	7.297	7.283	7.288			3.759	3.759	3.810	3.807
Cys429	HN	8.727	8.725	8.734	8.734	Phe436	HN	8.807	8.769	8.587	8.578
	Hα	4.995	4.991	4.976	4.976		Hα	5.137	5.163	5.105	5.108
	Hβ	2.866	2.863	2.886	2.880		Hβ	3.015	3.017	3.003	2.997
		2.972	2.973	2.977	2.974			3.387	3.344	3.450	3.445
Ile430	HN	9.545	9.522	9.613	9.594		Hδ	6.830	6.838	6.862	6.865
	Hα	4.222	4.232	4.205	4.205		Hε	6.961	6.965	6.935	6.935
	Hβ	1.765	1.759	1.784	1.784		Hζ	7.023	7.035	7.000	7.004
	Hγ1	0.762	0.770	0.747	0.7525	Glu437	HN	9.649	9.604	9.721	9.709
	Hγ2	1.101	1.101	1.150	1.147		Hα	4.576	4.580	4.520	4.529
		1.356	1.353	1.366	1.360		Hβ	1.827	1.847	1.784	1.808
	Hδ	0.678	0.678	0.665	0.665			1.933	1.949	1.916	1.922
Asn431	HN	9.007	8.970	9.041	9.038		Hγ	2.100	2.130	2.018	2.039
	Hα	4.805	4.805	4.824	4.810			2.177	2.209	2.106	2.135
	Hβ	2.567	2.585	2.541	2.547	Cys438	HN	8.739	8.763	8.720	8.720
		2.917	2.858	2.983	2.968		Hα	5.535	5.534	5.523	5.529
	γNH_2	7.686	7.615	7.754	7.746		Hβ	2.552	2.572	2.559	2.574
		7.897	7.841	7.955	7.951			2.756	2.739	2.752	2.740
Thr432	HN	7.919	7.944	7.934	7.934						
	Hα	4.558	4.541	4.570	4.564						
	Hβ	4.149	4.148	4.140	4.134						
	Hγ	0.922	0.916	0.910	0.902						

Residues		1	2	3	4	Residues		1	2	3	4
Gln439	HN	9.494	9.473	9.502	9.477	Thr445	HN	8.991	8.926	8.995	8.921
	Hα	4.576	4.585	4.570	4.576		Hα	4.700	4.885	4.704	4.871
	Hβ	1.943	1.953	1.945	1.945		Hβ	4.167	3.980	4.161	3.968
		2.042	2.042	2.044	2.039		Hγ	1.003	0.995	1.001	0.995
	Hγ	2.188	2.192	2.182	2.182	Gly446	HN	7.928	7.383	7.923	7.366
	δNH₂	6.730	6.730	6.721	6.718		Hα	3.905	3.870	3.895	3.857
		7.442	7.429	7.480	7.472	Pro447	Hα	4.141	4.180	4.115	4.137
Cys440	HN	8.872	8.871	8.866	8.870		Hβ	2.057	2.096	2.068	2.067
	Hα	4.630	4.647	4.637	4.671		Hγ	2.215	2.234	2.194	2.198
	Hβ	2.625	2.642	2.617	2.626			2.311	2.365	2.302	2.324
		3.168	3.119	3.167	3.108		Hδ	3.544	3.566	3.533	3.566
Leu441	HN	8.676	8.643	8.647	8.612			3.830	3.826	3.866	3.854
	Hα	4.393	4.399	4.395	4.389	Arg448	HN	8.216	8.242	8.246	8.270
	Hβ	1.652	1.636	1.653	1.638		Hα	4.321	4.333	4.310	4.313
	Hγ	1.743	1.741	1.738	1.738		Hβ	1.790	1.786	1.790	1.790
	Hδ	0.904	0.898	0.902	0.905		Hγ	1.608	1.605	1.609	1.606
		0.952	0.942	0.937	0.945		Hδ	2.884	2.885	2.883	2.883
Gln442	HN	8.262	8.270	8.267	8.270		εNH	7.485	7.487	7.488	7.488
	Hα	3.861	3.887	3.860	3.883	Cys449	HN	8.743	8.725	8.751	8.752
	Hβ	1.940	1.953	1.939	1.945		Hα	4.827	4.810	4.871	4.871
		2.013	2.006	2.009	2.006		Hβ	2.392	2.426	2.351	2.360
	Hγ	2.283	2.293	2.281	2.287			2.804	2.836	2.781	2.790
	δNH₂	6.770	6.767	6.762	6.768	Glu450	HN	10.68	10.64	10.68	10.62
		7.460	7.462	7.460	7.463		Hα	4.226	4.209	4.212	4.209
Gly443	HN	8.451	8.500	8.458	8.502		Hβ	1.273	1.283	1.343	1.350
	Hα	3.282	3.287	3.284	3.278			1.601	1.576	1.606	1.616
		3.847	3.839	3.845	3.833		Hγ	1.776	1.780	1.802	1.809
Tyr444	HN	7.711	7.734	7.711	7.727	Ile451	HN	8.756	8.634	8.764	8.641
	Hα	5.042	4.942	5.043	4.941		Hα	4.149	4.161	4.146	4.158
	Hβ	2.724	2.713	2.720	2.705		Hβ	1.867	1.600	1.875	1.586
		2.891	2.889	2.895	2.886		Hγ1	0.751	0.770	0.747	0.767
	Hδ	6.721	6.704	6.718	6.696		Hγ2	1.065	1.048	1.065	1.054
	Hε	6.227	6.222	6.225	6.216			1.386	1.437	1.384	1.439
							Hδ	0.678	0.686	0.662	0.682

Residues		1	2	3	4	Residues		1	2	3	4
GlcNAc	H1	-	4.474	-	4.480	Glc	H1	-	-	4.425	4.420
	H2	-	3.667	-	3.651		H2	-	-	2.846	2.840
	H3	-	3.470	-	3.450		H3	-	-	3.362	3.358
	H4	-	3.402	-	3.386		H4	-	-	3.454	3.446
	H5	-	3.372	-	3.357		H5	-	-	3.662	3.657
	H6	-	3.720	-	3.698		H6	-	-	3.761	3.758
		-	3.858	-	3.844			-	-	3.884	3.875
	NH	-	8.230	-	8.223						

Table S3-2. ^{13}C-NMR chemical shifts of synthetic EGF11 modules **1**-**4**.

Residues		**1**	**2**	**3**	**4**	Residues		**1**	**2**	**3**	**4**
Asp412	Cα	52.17	52.04	52.04	52.04	His425	Cα	51.95	52.13	51.99	52.04
	Cβ	41.22	40.53	41.23	41.14		Cβ	33.42	33.42	33.37	33.47
Val413	Cα	58.78	58.87	58.73	58.68		Cδ2	116.9	117.3	115.1	117.3
	Cβ	30.39	30.99	30.61	30.57		Cε1	131.4	131.9	132.3	131.5
	Cγ	18.15	18.45	18.12	18.26	Ala426	Cα	50.22	50.26	50.21	50.17
Asp414	Cα	54.55	54.47	54.66	54.57		Cβ	13.81	13.86	13.82	13.96
	Cβ	31.04	31.08	31.03	31.03	Gly427	Cα	42.74	43.15	42.78	42.87
Glu415	Cα	58.07	57.84	58.17	58.17	Lys428	Cα	53.03	52.88	53.11	53.11
	Cβ	26.43	26.59	27.29	27.06		Cβ	29.52	29.86	29.63	29.77
	Cγ	35.37	34.64	35.99	35.62		Cγ	31.04	31.36	32.90	32.25
Cys416	Cα	50.65	50.82	50.49	50.49		Cδ	26.81	26.78	25.65	25.23
	Cβ	33.42	33.51	33.33	33.33		Cε	33.75	33.61	32.90	32.39
Ser417	Cα	55.09	55.03	54.99	55.03	Cys429	Cα	54.98	54.94	55.13	55.08
	Cβ	61.49	61.21	61.58	61.58		Cβ	32.88	33.14	32.44	32.58
Leu418	Cα	52.71	52.97	52.69	52.69	Ile430	Cα	60.51	60.65	60.55	60.60
	Cβ	39.38	39.41	39.45	39.36		Cβ	37.54	37.82	37.21	37.40
	Cγ	24.11	24.16	23.73	23.88		Cγ1	14.35	14.24	14.28	14.00
	Cδ	20.31	20.60	20.23	20.23		Cγ2	24.11	24.34	24.34	24.34
Gly419	Cα	43.07	43.06	42.92	43.01		Cδ	9.045	10.21	9.232	10.21
Ala420	Cα	48.70	48.76	48.67	48.72	Asn431	Cα	51.84	51.85	51.85	51.90
	Cβ	16.63	16.67	16.76	16.62		Cβ	36.13	36.41	36.23	36.13
Asn421	Cα	48.16	48.30	48.25	48.20	Thr432	Cα	56.72	56.81	56.62	56.58
	Cβ	41.6	41.47	41.37	41.42		Cβ	68.64	68.23	68.37	68.27
Pro422	Cα	62.24	62.33	62.28	62.33		Cγ	23.13	23.22	23.13	23.22
	Cβ	29.20	29.58	29.44	29.54	Leu433	Cα	53.58	53.72	53.44	53.54
	Cγ	32.12	32.39	32.90	32.25		Cβ	37.11	37.26	37.63	37.49
	Cδ	47.29	47.27	47.27	47.22		Cγ	24.43	24.44	24.30	24.48
Cys423	Cα	50.65	50.73	50.45	50.49		Cδ	21.94	21.82	21.54	21.68
	Cβ	36.24	36.23	36.13	36.13	Gly434	Cα	43.07	43.06	43.20	43.24
Glu424	Cα	51.41	51.48	51.10	51.06	Ser435	Cα	53.36	53.44	53.11	53.16
	Cβ	26.60	26.96	27.52	27.24		Cβ	59.97	60.09	68.93	68.88
	Cγ	30.99	31.36	31.08	31.59						

Residues		1	2	3	4	Residues		1	2	3	4
Phe436	Cα	54.44	54.38	54.42	54.42	Pro447	Cα	55.31	55.03	55.45	55.27
	Cβ	37.22	37.54	36.13	37.07		Cβ	26.92	26.68	26.78	27.01
	Cδ	129.7	129.7	129.8	129.8		Cγ	34.07	33.75	33.89	33.79
	Cε	126.5	126.6	126.1	126.1		Cδ	47.62	48.11	47.97	47.97
	Cζ	127.9	127.9	127.9	127.8	Arg448	Cα	52.49	52.60	52.51	52.46
Glu437	Cα	51.52	51.57	51.62	51.57		Cβ	31.58	31.64	31.08	31.36
	Cβ	29.41	29.86	30.89	30.57		Cγ	26.38	26.50	26.40	26.45
	Cγ	32.12	31.27	32.72	32.20		Cδ	39.17	39.31	39.27	39.27
Cys438	Cα	50.76	50.82	50.78	50.78	Cys449	Cα	48.27	48.48	48.25	48.25
	Cβ	36.24	36.32	36.27	36.18		Cβ	36.35	36.04	36.41	36.23
Gln439	Cα	51.52	51.57	51.80	51.85	Glu450	Cα	57.26	57.47	57.14	57.14
	Cβ	25.84	26.03	25.93	26.03		Cβ	26.27	26.40	25.14	25.51
	Cγ	31.04	31.17	31.08	31.08		Cγ	31.15	31.64	31.45	31.92
Cys440	Cα	51.41	51.20	51.29	51.15	Ile451	Cα	57.48	57.75	57.42	57.56
	Cβ	35.37	35.76	35.29	35.71		Cβ	33.31	34.45	33.33	34.17
Leu441	Cα	52.06	52.04	51.99	51.90		Cγ1	14.35	14.24	14.28	14.10
	Cβ	40.36	40.44	40.34	40.30		Cγ2	24.32	24.44	24.34	24.34
	Cγ	25.30	25.37	25.33	25.23		Cδ	9.153	10.21	9.232	10.21
	Cδ	20.42	20.41	20.27	20.37	GlcNAc	C1	-	99.85	-	99.85
Gln442	Cα	54.98	54.94	55.03	54.94		C2	-	55.50	-	55.41
	Cβ	25.84	26.03	25.93	26.03		C3	-	73.28	-	73.28
	Cγ	31.04	31.08	31.08	31.13		C4	-	69.72	-	69.63
Gly443	Cα	41.66	41.56	41.47	41.65		C5	-	76.09	-	75.85
Tyr444	Cα	54.98	55.03	54.89	54.94		C6	-	60.74	-	60.65
	Cβ	37.32	38.00	37.58	38.05	Glc	C1	-	-	104.4	104.3
	Cδ	130.7	130.8	130.7	130.8		C2	-	-	72.86	72.90
	Cε	113.8	114.0	114.0	114.0		C3	-	-	75.80	75.90
Thr445	Cα	57.26	56.06	57.28	56.02		C4	-	-	69.30	69.35
	Cβ	68.64	74.68	68.79	74.68		C5	-	-	72.02	71.97
	Cγ	23.02	23.22	23.08	23.22		C6	-	-	60.04	59.99
Gly446	Cα	41.12	41.47	41.28	41.56						

Table S3-3. ^{1}H-NMR chemical shifts of synthetic EGF10 modules **5-9**.

Residues		**5**	**6**	**7**	**8**	**9**	Residues		**5**	**6**	**7**	**8**	**9**
Leu372	HN	8.140	8.143	8.132	8.135	8.135	Asn379	HN	7.807	7.806	7.849	7.856	7.854
	Hα	4.155	4.159	4.165	4.159	4.162		Hα	4.557	4.556	4.542	4.548	4.548
	Hβ	1.446	1.449	1.447	1.449	1.449		Hβ	2.729	2.735	2.733	2.756	2.759
	Hγ	1.510	1.514	1.523	1.519	1.519			2.843	2.849	2.867	2.873	2.870
	Hδ	0.773	0.774	0.789	0.780	0.780		γNH_2	6.717	6.720	6.701	6.714	6.717
Asn373	HN	8.380	8.379	8.378	8.389	8.388			7.406	7.408	7.401	7.388	7.385
	Hα	4.564	4.565	4.548	4.565	4.568	Pro380	Hα	4.435	4.447	4.442	4.491	4.493
	Hβ	2.624	2.621	2.633	2.627	2.633		Hβ	2.020	2.014	1.991	1.937	1.948
		2.732	2.738	2.750	2.747	2.750		Hγ	1.868	1.868	1.886	1.779	1.790
	γNH_2	6.836	6.853	6.849	6.853	6.818		Hδ	3.658	3.661	3.648	3.788	3.805
		7.468	7.478	7.473	7.474	7.482	Cys381	HN	8.125	8.127	8.060	8.093	8.093
Asp374	HN	8.057	8.059	8.055	8.058	8.050		Hα	4.602	4.606	4.577	4.533	4.536
	Hα	4.410	4.416	4.437	4.419	4.431		Hβ	2.621	2.615	2.610	2.648	2.639
	Hβ	2.486	2.498	2.487	2.499	2.499			3.109	3.112	3.092	3.101	3.101
		2.627	2.642	2.660	2.657	2.654	Asn382	HN	8.371	8.368	8.369	8.372	8.372
Ala375	HN	8.492	8.497	8.584	8.525	8.530		Hα	4.456	4.454	4.460	4.469	4.472
	Hα	4.120	4.124	4.101	4.095	4.106		Hβ	2.577	2.583	2.572	2.578	2.584
	Hβ	1.314	1.320	1.330	1.335	1.335			2.673	2.674	2.660	2.665	2.660
Cys376	HN	8.239	8.237	8.221	8.199	8.201		γNH_2	6.804	6.808	6.815	6.809	6.804
	Hα	4.316	4.322	4.232	4.189	4.195			7.488	7.492	7.485	7.488	7.489
	Hβ	2.884	2.887	2.955	2.888	2.888	Glu383	HN	8.577	8.579	8.601	8.581	8.581
		3.050	3.054	3.072	3.063	3.060		Hα	3.959	3.969	3.957	3.969	3.972
Ile377	HN	8.039	8.038	8.012	8.012	8.009		Hβ	1.881	1.887	1.876	1.879	1.881
	Hα	3.767	3.767	3.791	3.797	3.797		Hγ	2.173	2.191	2.186	2.218	2.209
	Hβ	1.772	1.769	1.787	1.758	1.770	Gly384	HN	8.700	8.708	8.734	8.711	8.715
	Hγ1	0.813	0.816	0.828	0.820	0.814		Hα	3.541	3.545	3.531	3.539	3.539
	Hγ2	1.123	1.120	1.140	1.115	1.127			4.082	4.097	4.103	4.092	4.103
		1.389	1.394	1.389	1.396	1.410	Ser385	HN	7.764	7.774	7.743	7.764	7.767
	Hδ	0.785	0.789	0.814	0.803	0.809		Hα	4.532	4.536	4.548	4.542	4.548
Ser378	HN	7.738	7.739	7.612	7.802	7.802		Hβ	3.653	3.641	3.642	3.648	3.642
	Hα	4.316	4.316	4.554	4.574	4.557							
	Hβ	3.717	3.723	3.680	3.744	3.747							
		-	-	4.215	4.077	4.074							

Residues		5	6	7	8	9	Residues		5	6	7	8	9
Asn386	HN	8.657	8.670	8.696	8.673	8.683	Asn393	HN	8.369	8.377	8.373	8.376	8.375
	Hα	4.749	4.749	4.752	4.749	4.761		Hα	4.547	4.547	4.548	4.539	4.545
	Hβ	2.542	2.545	2.537	2.548	2.543		Hβ	2.580	2.589	2.572	2.581	2.578
		2.650	2.653	2.657	2.654	2.651			2.670	2.680	2.662	2.665	2.665
	γNH$_2$	6.804	6.809	6.814	6.812	6.812		γNH$_2$	6.812	6.809	6.811	6.809	6.805
		7.455	7.464	7.451	7.457	7.464			7.455	7.462	7.450	7.469	7.463
Cys387	HN	8.540	8.548	8.572	8.550	8.554	Gly394	HN	8.372	8.379	8.405	8.385	8.391
	Hα	5.137	5.144	5.148	5.153	5.162		Hα	3.606	3.609	3.601	3.607	3.607
	Hβ	2.817	2.826	2.806	2.800	2.803			3.921	3.925	3.940	3.940	3.943
		2.945	2.943	2.946	2.955	2.946	Lys395	HN	7.851	7.855	7.853	7.859	7.857
Asp388	HN	8.869	8.880	8.895	8.913	8.915		Hα	4.225	4.229	4.217	4.188	4.188
	Hα	4.822	4.828	4.855	4.837	4.837		Hβ	1.644	1.651	1.651	1.651	1.657
	Hβ	2.428	2.440	2.443	2.443	2.440			1.709	1.709	1.721	1.712	1.704
		-	2.481	-	2.484	2.475		Hγ	1.197	1.203	1.192	1.204	1.204
Thr389	HN	8.605	8.610	8.682	8.601	8.611		Hδ	1.525	1.531	1.528	1.546	1.543
	Hα	4.310	4.316	4.346	4.349	4.352		Hε	2.866	2.862	2.868	2.871	2.865
	Hβ	3.822	3.831	3.806	3.820	3.814		εNH$_2$	6.687	6.692	6.676	6.679	6.679
	Hγ	0.927	0.926	0.938	0.938	0.926	Ala396	HN	8.295	8.300	8.364	8.207	8.207
Asn390	HN	8.112	8.126	8.122	8.132	8.131		Hα	4.410	4.416	4.440	4.555	4.558
	Hα	4.410	4.430	4.428	4.443	4.445		Hβ	1.159	1.163	1.151	1.160	1.157
	Hβ	2.542	2.560	2.554	2.563	2.563	Ile397	HN	7.947	7.952	7.954	7.954	7.954
	γNH$_2$	6.807	6.810	6.812	6.811	6.812		Hα	4.193	4.197	4.194	4.197	4.194
		7.452	7.462	7.450	7.457	7.458		Hβ	1.611	1.608	1.582	1.624	1.635
Pro391	Hα	4.299	4.305	4.308	4.308	4.308		Hγ1	0.746	0.751	0.746	0.741	0.758
	Hβ	2.199	2.196	2.193	2.182	2.185		Hγ2	0.939	0.938	0.930	0.931	0.937
	Hγ	1.895	1.892	1.892	1.881	1.893			1.371	1.371	1.377	1.366	1.378
	Hδ	3.779	3.773	3.803	3.776	3.777		Hδ	0.699	0.710	0.690	0.694	0.709
		3.967	3.969	3.966	3.962	3.960	Cys398	HN	8.573	8.575	8.611	8.582	8.578
Val392	HN	7.927	7.927	7.925	7.932	7.929		Hα	5.438	5.439	5.450	5.450	5.457
	Hα	3.767	3.770	3.767	3.776	3.773		Hβ	2.662	2.656	2.645	2.671	2.651
	Hβ	1.875	1.882	1.879	1.882	1.882			2.945	2.943	2.943	2.958	2.946
	Hγ	0.788	0.797	0.792	0.795	0.798							
		0.864	0.870	0.874	0.870	0.874							

Residues		5	6	7	8	9	Residues		5	6	7	8	9
Thr399	HN	8.849	8.851	8.877	8.866	8.864	Pro407	Hα	4.295	4.302	4.303	4.307	4.309
	Hα	4.503	4.495	4.530	4.519	4.513		Hβ	2.208	2.199	2.196	2.197	2.200
	Hβ	4.015	4.012	4.007	4.019	4.019		Hγ	1.915	1.932	1.909	1.916	1.934
	Hγ	1.094	1.093	1.094	1.086	1.103		Hδ	3.512	3.526	3.521	3.514	3.528
Cys400	HN	8.799	8.817	8.851	8.822	8.834			3.614	3.615	3.619	3.618	3.624
	Hα	4.901	4.910	4.916	4.904	4.904	Ala408	HN	8.274	8.314	8.283	8.286	8.313
	Hβ	2.799	2.820	2.806	2.803	2.820		Hα	4.220	4.223	4.226	4.226	4.212
		3.027	2.984	3.025	3.031	2.987		Hβ	1.276	1.291	1.280	1.283	1.295
Pro401	Hα	4.342	4.349	4.346	4.348	4.350	Cys409	HN	7.966	7.992	7.976	7.973	7.995
	Hβ	2.193	2.207	2.193	2.191	2.217		Hα	4.506	4.486	4.501	4.507	4.483
	Hγ	1.798	1.810	1.804	1.793	1.817		Hβ	2.893	2.884	2.890	2.896	2.888
	Hδ	3.364	3.340	3.378	3.372	3.347			3.103	3.098	3.104	3.104	3.098
		3.704	3.716	3.707	3.715	3.723	Ser410	HN	8.061	8.022	8.055	8.062	8.022
Ser402	HN	8.314	8.317	8.312	8.313	8.316		Hα	4.407	4.392	4.425	4.425	4.396
	Hα	4.129	4.112	4.127	4.139	4.118		Hβ	3.749	3.746	3.753	3.756	3.747
	Hβ	3.752	3.755	3.753	3.756	3.756			3.784	3.802	3.788	3.788	3.811
Gly403	HN	8.540	8.601	8.543	8.549	8.603	Gln411	HN	8.330	8.264	8.337	8.338	8.266
	Hα	3.579	3.559	3.583	3.586	3.563		Hα	4.155	4.167	4.159	4.162	4.171
		4.062	4.050	4.068	4.068	4.054		Hβ	1.840	1.732	1.844	1.847	1.738
Tyr404	HN	7.801	7.806	7.799	7.803	7.804			2.004	1.896	2.008	2.011	1.899
	Hα	4.570	4.565	4.577	4.574	4.571		Hγ	2.217	2.285	2.221	2.224	2.288
	Hβ	2.843	2.843	2.850	2.847	2.844		δNH_2	7.016	6.996	7.020	7.020	6.999
	Hδ	6.963	6.990	6.976	6.979	7,001			7.474	7.477	7.482	7.483	7.483
	Hε	6.682	6.698	6.690	6.693	6.718							
Thr405	HN	7.997	8.019	8.004	7.999	8.030							
	Hα	4.442	4.612	4.451	4.451	4.609							
	Hβ	4.082	4.167	4.086	4.086	4.162							
	Hγ	0.996	0.961	1.005	1.005	0.967							
Gly406	HN	7.488	6.954	7.501	7.508	6.963							
	Hα	3.892	3.928	3.902	3.899	3.934							
		4.053	3.986	4.063	4.065	4.001							

Residues		5	6	7	8	9	Residues		5	6	7	8	9
GlcNAc	H1	-	4.366	-	-	4.369	Xyl	H1	-	-	-	5.187	5.181
	H2	-	3.519	-	-	3.520		H2	-	-	-	3.428	3.429
	H3	-	3.397	-	-	3.403		H3	-	-	-	3.673	3.672
	H4	-	3.318	-	-	3.322		H4	-	-	-	3.685	3.691
	H5	-	3.271	-	-	3.275		H5	-	-	-	3.720	3.733
	H6	-	3.634	-	-	3.628	Xyl	H1	-	-	-	5.210	5.213
		-	3.732	-	-	3.747		H2	-	-	-	3.428	3.429
	NH	-	8.162	-	-	8.163		H3	-	-	-	3.552	3.553
Glc	H1	-	-	4.281	4.333	4.327		H4	-	-	-	3.486	3.492
	H2	-	-	3.074	3.217	3.214		H5	-	-	-	3.720	3.733
	H3	-	-	3.383	3.510	3.525							
	H4	-	-	3.226	3.337	3.331							
	H5	-	-	3.320	3.503	3.504							
	H6	-	-	3.579	3.599	3.593							
		-	-	3.783	3.788	3.787							

Table S3-4. ^{13}C-NMR chemical shifts of synthetic EGF10 modules **5**-**9**.

Residues		**5**	**6**	**7**	**8**	**9**	Residues		**5**	**6**	**7**	**8**	**9**
Leu372	Cα	52.69	52.69	52.79	52.79	52.74	Ser385	Cα	55.41	55.31	55.31	55.41	55.36
	Cβ	39.78	39.78	39.78	39.78	39.73		Cβ	61.96	62.14	62.14	62.14	62.19
	Cγ	24.25	24.34	24.34	24.44	24.48	Asn386	Cα	49.89	49.79	49.61	49.70	49.75
	Cδ	22.19	22.28	22.19	22.19	22.24		Cβ	38.00	38.10	38.00	38.00	38.05
Asn373	Cα	50.35	50.26	50.45	50.35	50.31	Cys387	Cα	53.63	53.63	53.72	54.00	53.86
	Cβ	36.32	36.23	36.32	36.23	36.37		Cβ	38.47	38.38	38.38	38.28	38.33
Asp374	Cα	51.48	51.48	51.57	51.66	51.62	Asp388	Cα	50.54	50.45	50.64	50.64	50.68
	Cβ	38.10	38.28	38.19	38.10	38.33		Cβ	41.19	41.37	41.75	41.47	41.79
Ala375	Cα	50.92	50.92	51.01	51.01	50.96	Thr389	Cα	55.59	55.31	55.59	55.69	55.83
	Cβ	16.67	16.58	16.58	16.58	16.72		Cβ	67.20	67.20	67.29	67.01	67.15
Cys376	Cα	53.16	53.26	52.97	52.88	52.93		Cγ	18.64	18.92	18.64	18.73	18.87
	Cβ	38.28	38.28	38.38	38.19	38.33	Asn390	Cα	50.45	50.45	50.73	50.64	50.68
Ile377	Cα	60.55	60.46	60.65	60.46	60.41		Cβ	36.32	36.32	36.41	36.32	36.37
	Cβ	35.20	35.29	35.20	35.20	35.15	Pro391	Cα	61.40	61.30	61.40	61.30	61.35
	Cγ1	14.80	14.80	14.80	14.89	14.85		Cβ	29.49	29.58	29.49	29.58	29.54
	Cγ2	25.56	25.56	25.65	25.65	25.70		Cγ	24.53	24.62	24.62	24.72	24.58
	Cδ	10.12	10.12	10.40	10.21	10.26		Cδ	48.20	48.11	48.20	48.20	48.25
Ser378	Cα	55.31	55.50	52.79	52.51	52.65	Val392	Cα	62.14	62.05	62.14	62.52	62.57
	Cβ	60.93	60.65	67.48	66.92	66.78		Cβ	21.63	22.28	21.54	21.63	21.77
Asn379	Cα	50.35	50.35	50.35	50.35	50.31		Cγ	18.26	18.07	18.17	18.45	18.31
	Cβ	36.13	36.13	36.23	36.32	36.27	Asn393	Cα	50.35	50.45	50.35	50.35	50.31
Pro380	Cα	61.02	61.02	61.11	60.93	60.97		Cβ	36.32	36.32	36.32	36.41	36.37
	Cβ	29.77	29.58	29.40	29.68	29.63	Gly394	Cα	43.24	43.52	43.34	43.15	43.38
	Cγ	24.53	24.62	24.62	24.72	24.86	Lys395	Cα	52.97	52.97	52.97	53.07	53.02
	Cδ	47.92	47.92	47.83	48.11	47.78		Cβ	30.71	30.80	30.71	30.80	30.75
Cys381	Cα	50.54	50.54	50.45	50.45	50.40		Cγ	22.66	22.85	22.75	22.94	22.89
	Cβ	37.63	37.44	37.63	38.19	38.14		Cδ	26.03	25.93	26.03	26.03	25.98
Asn382	Cα	50.45	50.45	50.92	51.57	51.62		Cε	39.22	39.22	39.22	39.31	39.17
	Cβ	36.13	36.23	36.23	36.41	36.27	Ala396	Cα	49.61	49.61	49.61	49.61	49.56
Glu383	Cα	55.22	55.41	55.41	55.41	55.45		Cβ	16.86	16.95	16.95	16.76	16.90
	Cβ	26.31	26.40	26.31	26.40	26.45							
	Cγ	32.67	32.76	32.86	32.86	33.09							
Gly384	Cα	42.40	42.31	42.31	42.31	42.26							

Residues		5	6	7	8	9	Residues		5	6	7	8	9
Ile397	Cα	57.75	57.65	57.65	57.84	57.70	Cys409	Cα	52.41	52.69	52.51	52.69	52.65
	Cβ	37.82	37.82	37.82	37.82	37.77		Cβ	38.28	38.28	38.19	38.19	38.24
	Cγ1	15.17	15.17	15.17	15.08	15.22	Ser410	Cα	55.59	55.69	55.59	55.59	55.55
	Cγ2	24.44	24.44	24.44	24.62	24.58		Cβ	61.21	61.30	61.02	60.93	61.26
	Cδ	10.78	10.78	10.87	10.78	10.92	Glu411	Cα	52.69	52.69	52.79	52.88	52.74
Cys398	Cα	49.79	49.79	49.98	50.17	50.03		Cβ	26.68	26.68	26.68	26.78	26.73
	Cβ	37.91	38.10	38.00	38.00	38.14		Cγ	31.37	31.17	31.36	31.36	31.11
Thr399	Cα	58.31	58.21	58.21	58.21	58.26	GlcNAc	C1	-	99.85	-	-	99.90
	Cβ	68.23	68.32	68.32	68.32	68.37		C2	-	55.50	-	-	55.55
	Cγ	19.10	19.01	19.10	19.01	19.06		C3	-	73.56	-	-	73.42
Cys400	Cα	50.92	51.01	50.92	50.92	51.06		C4	-	69.54	-	-	69.58
	Cβ	36.97	36.88	36.88	37.07	36.93		C5	-	75.71	-	-	75.76
Pro401	Cα	60.55	60.46	60.65	60.55	60.60		C6	-	60.55	-	-	60.69
	Cβ	29.49	29.49	29.49	29.58	29.63	Glc	C1	-	-	102.0	101.9	101.8
	Cγ	24.81	24.81	24.81	24.72	24.76		C2	-	-	73.18	71.78	71.83
	Cδ	47.92	48.11	48.02	48.11	48.16		C3	-	-	75.43	70.00	70.05
Ser402	Cα	56.81	57.00	56.81	56.90	57.04		C4	-	-	69.54	75.52	75.48
	Cβ	60.18	60.18	60.37	60.46	60.41		C5	-	-	75.90	82.07	82.21
Gly403	Cα	42.31	42.31	42.21	42.31	42.26		C6	-	-	60.83	60.55	60.51
Tyr404	Cα	55.13	55.03	55.22	55.03	55.08	Xyl	C1	-	-	-	99.10	99.06
	Cβ	36.51	36.60	36.51	36.60	36.55		C2	-	-	-	71.69	71.64
	Cδ	130.6	130.6	130.5	130.6	130.7		C3	-	-	-	79.17	79.31
	Cε	115.4	115.4	115.4	115.5	115.4		C4	-	-	-	69.82	69.86
Thr405	Cα	57.93	56.72	57.93	57.93	56.67		C5	-	-	-	61.30	61.35
	Cβ	67.48	75.06	67.38	67.48	75.10	Xyl	C1	-	-	-	98.92	99.06
	Cγ	18.73	15.36	18.54	18.64	15.41		C2	-	-	-	71.69	71.64
Gly406	Cα	42.31	42.21	42.40	42.31	42.45		C3	-	-	-	72.90	72.95
Pro407	Cα	61.40	61.30	61.30	61.30	61.35		C4	-	-	-	69.54	69.68
	Cβ	29.49	29.49	29.58	29.58	29.63		C5	-	-	-	61.30	61.35
	Cγ	24.53	24.53	24.62	24.62	24.58							
	Cδ	47.08	47.08	47.08	47.08	46.94							
Ala408	Cα	50.07	49.98	49.98	50.07	50.03							
	Cβ	16.39	16.39	16.39	16.48	16.53							

Chapter 4
Concluding Remarks

Glycosylation is one of the most remarkable posttranslational modifications because of their functional and structural diversity and close relationship with several diseases. Protein glycosylation, which is given in template-independent manner, can regulate several protein functions. Aberrant glycosylation as well as overexpression of some glycoproteins have been found in tumor cells. Therefore, glycoproteins as well as glycopeptides have been applied as cancer therapeutic and diagnostic targets. Particularly, glycopeptide carrying *O*-glycans have been developed as antigen determinant region, so called epitope.

To further understanding of glycosylation impacts on protein structure and function, and further application of glycopeptide to therapeutic and diagnostic targets, organic chemical approaches is necessary. In this thesis, unique *O*-glycans such as *O*-GlcNAcylation, *O*-fucosylation and *O*-glucosylation in human NOTCH1 receptor extracellular EGF-like domain are focused as target molecule to elucidate *O*-glycosylation impacts on protein structures as well as functions.

In chapter 2, chemical synthesis of variously *O*-glycosylated models for NOTCH1 EGF-like domains 10-12 uncovered functional roles of site-specific *O*-glycosylation of the individual domains, particularly in the modulation of unique "equilibria of folding states" under a general physiological reductant environment. We may speculate that the number and location of the correctly folded domains (canonical EGF folds) and misfolded domains will affect significantly the integrity of Notch ECD, which determines the ligand-binding modality of cell surface Notch receptors and subsequent stoichiometric signaling pathway. Although direct evidence on the presence of misfolded domains in the

intact NOTCH1 ECD has not been provided, our finding that domain-specific folding equilibrium in the Notch receptor ECD is tuned finely by site-specific *O*-glycosylation is highly suggestive.

In chapter 3, NMR studies of variously *O*-glycosylated models for NOTCH1 EGF-like domains 10~12 uncovered effects of site-specific *O*-glycosylation on the structural behavior of the individual domains, particularly, in the conformational stabilization by "sugar bridges". Comprehensive NMR studies of novel synthetic EGF modules **1**~**9** showed several interactions between sugar moiety and peptide chain contributed to conformational stabilization of overall EGF domains, although *O*-glycosylations have no significant influence on whole conformation of each EGF domain. Among them, we discovered the hydrophobic interaction between *O*-GlcNAc residue and C-terminus region of each EGF domain, indicating *O*-GlcNAcylation-mediated sugar bridge as well as Ca^{2+} coordination contributes to the regulation of flexibility and rigidity of the connecting regions of NOTCH1 extracellular domains. Our results are the first to characterize the structural significance of *O*-GlcNAcylation on EGF domain. As expected, the sugar bridge 378Ser-O-Glc-Xyl-Xyl---Lys395 affects the conformational stabilization of non-Ca^{2+}-binding domain 10, whereas no NOE correlation between the glucose attached to Ser435 and any amino acid residue is observed in Ca^{2+}-binding domain 11. The result suggests that the *O*-glucosylation at Ser435 within C3-X-N-T-X-G-S-F-X-C4 may have unidentified roles in Notch receptor function, while mutation of Ser435 to Ala affects NOTCH1 function.

These chemical approaches of *O*-glycosylated peptides will facilitate further investigations to reveal molecular mechanism of glycobiology and Etiology. Furthermore, these approaches will lead to understand the molecular mechanism of Notch signaling and find novel therapeutic approach to target the Notch signaling pathway of solid tumor.

Acknowledgement

I would like to express my sincere gratitude to Professor Shin-Ichiro Nishimura for his helpful suggestions, discussions, and encouragements. This work described herein would not have been successful without him.

I would like to express my special thanks to Professor Makoto Demura, Professor Tomoyasu Aizawa, Professor Kenji Monde and Professor Hiroshi Hinou in Hokkaido University for their critical suggestions, and tenderhearted encouragements.

I am deeply indebted to Assistant Professor Fayna Garcia-Martin and Dr. Shun Hayakawa for their helpful support, kind assistance and affectionate encouragements.

I want to express grateful thanks to my colleagues in Nishimura lab for daily supports, discussions, and encouragement.

Finally, I wish to express my sincere gratitude to my parents for their invaluable mental supports and affectionate encouragement.