Structural studies on cell surface receptors and lipopeptide antibiotics

On Notch1, epidermal growth factor receptor, and laspartomycin C analogues

Matthieu Raphy Zeronian

ISBN: 978-94-6423-433-6

Doctoral thesis
Structural studies on cell surface receptors and lipopeptide antibiotics
Matthieu Raphy Zeronian
Structural Biochemistry
Bijvoet Centre for Biomolecular Research, Faculty of Science
Utrecht University, the Netherlands

Cover illustration: Emma van Amerongen

Layout and print: ProefschriftMaken || www.proefschriftmaken.nl

Printed on FSC certified recycled paper

Structural studies on cell surface receptors and lipopeptide antibiotics

On Notch1, epidermal growth factor receptor, and laspartomycin C analogues

Onderzoek naar de structuur van celoppervlakreceptoren en lipopeptide-antibiotica

Over Notch1, epidermale groeifactor receptor, en laspartomycine C analogen
(met een samenvatting in het Nederlands)

Études structurales de récepteurs de surface cellulaire et d'antibiotiques lipopeptidiques : Notch1, récepteur du facteur de croissance épidermique, et analogues de laspartomycine C
(avec un résumé en français)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op donderdag 14 oktober 2021 des middags te 12.15 uur

door

Matthieu Raphy Zeronian
geboren op 25 september 1993
te Lyon, Frankrijk

Promotor:
Prof. dr. P. Gros
Copromotor:
Dr. B.J.C. Janssen

Table of contents

Chapter 1	General introduction	7
Chapter 2	Notch-Jagged signaling complex defined by an interaction mosaic	31
Chapter 3	Structural insights into the non-inhibitory mechanism of the anti-EGFR EgB4 nanobody	73
Chapter 4	Structure-activity studies and high-resolution crystal structures provide new mechanistic insights into C_{55}-P targeting lipopeptide antibiotics.	93
Chapter 5	General discussion	155
Summary		171
Samenvatting		175
Résumé		179
Acknowledgements		183
About the author		187

Chapter 1

General introduction

Matthieu R. Zeronian

Structural Biochemistry, Bijvoet Centre for Biomolecular Research, Utrecht University,
Padualaan 8, 3584 CH Utrecht, The Netherlands

At the cell surface, diverse molecules such as protein receptors and their associated ligands, but also specific lipids, are critical to the normal functioning and regulation of key cellular processes. This thesis describes the advances we made in understanding the mechanistic basis by which two distinct types of protein systems, cell surface receptors and lipopeptide calcium-dependent antibiotics, play a role in health and disease. We combine various structural biology approaches and use tools with therapeutic potential, such as nanobodies and antibiotics, to gain insights into these systems at the molecular and atomic levels, and to advance their efficacious utilization in a clinical context.

Cell surface receptors and signaling

Communication between cells is essential for various developmental and physiological processes, including cell growth, differentiation, migration and survival (1). To communicate, cells use highly specialized proteins termed cell surface receptors that are differentially expressed depending on the cell type. Cells can communicate via direct receptor contact, or more commonly by secretion of molecules, e.g. neurotransmitters or hormones, which bind to distant cell surface receptors. This leads to receptor activation, signal transduction, and induction of downstream receptor-dependent cellular responses. Different receptors recognize different ligands, which allows them to perform diverse tasks. Most cell surface receptors are composed of an extracellular region responsible for ligand recognition, a single transmembrane region, and an intracellular region that acts as an effector. Cell receptors can be divided into subgroups, such as the proteolytically activated receptors, enzyme-linked receptors, G-protein-coupled-receptors (GPCR), and ion channels. Receptor families have distinct signaling mechanisms, and generated signaling pathways can interact with each other. Here we will focus on two receptor families, namely the proteolytically activated Notch family and the enzyme-linked human epidermal growth factor receptor (HER) tyrosine kinase family, which were shown to influence each other during metazoan development (2).

Simple and complex: Notch signaling

More than a hundred years ago, genetic mutants that exhibited irregular notches of missing tissue at the tips of Drosophila wing blades were described and isolated (3). Later, the role of the aptly named Notch gene was further described in pioneer work from Donald Poulson, when the complete loss of Notch gene activity was found to be lethal at the embryonic stage due to neural hyperplasia (4). In the last decades, a wide consensus of scientific studies have convincingly demonstrated that Notch signaling acts as a central cell-cell communication system involved in a wide variety of processes in all metazoans, such as cell fate determination, stem cell maintenance, immune system regulation, and angiogenesis (5–8). Dysregulation of this system leads to a number of inherited and acquired diseases,

including congenital disorders and cancers (9–12). A striking feature of Notch signaling is the direct translocation of an active Notch fragment to the nucleus where it regulates transcription of target genes (Figure 1), differing from many other pathways that rely on signal amplification by secondary messengers, phosphorylation cascades or other signal-relaying mechanisms. Despite this deceptively simple framework, a remarkable complexity underlies Notch signaling, as it regulates an enormous number of cellular decisions during development (13) and in the adult (14).

The Notch signaling pathway is unusual in that most Notch ligands are transmembrane proteins instead of secreted molecules, therefore restricting signal to neighboring cells. Notch signals are transmitted using three main modes of action (15). Firstly, in lateral inhibition, a cell population sends an inhibitory signal to prevent other cells from adopting the same fate. This mechanism amplifies small differences in the levels of Notch signaling between neighboring populations of cells. Lateral inhibition can be limited in time, to prevent differentiation of a cell population and therefore maintain a pool of cell progenitors (16), or limited in space to control patterning through the differentiation of regularly spaced cells (17). Secondly, Notch also controls lineage fate of daughter cells by asymmetric inheritance of Notch regulators (e.g. Numb). For example, differential inheritance of Notch regulators determines whether stem cell progeny will adopt neural or glial fates (6). Thirdly, Notch signaling can control cell population boundaries by regulating the expression of ligands and Notch-associated enzymes (e.g. Fringe) in peripheral cells (18, 19).

In mammals, four Notch paralogs (Notch1-4) receive signals from their canonical ligands Jagged1, Jagged2, Delta-like (DLL) 1 and Delta-like4, in *trans* (from adjacent cells) or in *cis* (from the same cell) to activate or inhibit signaling, respectively. All Notch receptors and their canonical ligands are type I transmembrane proteins, i.e. they are composed of an extracellular N-terminal region, followed by a single transmembrane region and a cytosolic C-terminal segment. The highly modular extracellular segment of the Notch paralogs include variable numbers of epidermal growth factor (EGF) repeats (36 for Notch1-2, 34 for Notch3 and 29 for Notch4). EGF repeats are followed by the negative regulatory region (NRR), which contains three Lin12/Notch repeats (LNR) and a heterodimerization (HD) domain, and together prevent ligand-independent activation (20). On the cytosolic side, Notch receptors are composed of an RBP-Jkappa-associated module (RAM), followed by six ankyrin (ANK) repeats, two nuclear localization signals (NLS), a transactivation domain (TAD; for Notch1-2) and a PEST domain (rich in proline, glutamic acid, serine and threonine residues) which is targeted by ubiquitylation to regulate protein stability (21, 22). In contrast, the extracellular region of Notch ligands is characterized by a C2 domain, a Delta/Serrate/LAG-2 (DSL) domain, EGF repeats (16 repeats for Jagged1-2 ligands, 8 for Delta-like1 and 4 ligands), and a cysteine-rich domain (CRD; for Jagged1-2). Except for

Jagged2, the intracellular domains of Notch ligands contain a post-synaptic density protein ligand domain (PDZL).

In the endoplasmic reticulum (ER) Notch is extensively modified with oxygen-linked (O-linked) glycans, which are essential to Notch function (23–31). In the Golgi apparatus, the Notch NRR S1 site is cleaved by a furin-like convertase, which is required for signaling in mammals (Figure 1) (32, 33). The mature Notch receptor is then targeted to the cell surface as a heterodimer held together by non-covalent interactions in the NRR (34). The prevailing model for Notch activation states that upon ligand binding at Notch EGF8-12, ligand cell endocytosis generates a pulling force that triggers a conformational change in the Notch NRR (33, 35, 44, 36–43), leading to proteolytic cleavage by an ADAM-family metalloproteinase at the NRR S2 site (20, 45, 46), and subsequent cleavage by γ-secretase at the S3 site (Figure 1) (47–50). This triggers translocation of the free Notch intracellular domain (NICD) to the nucleus where it binds to the CBF1/Suppressor of Hairless/Lag-1 (CSL) complex, which acts as a repressor in the absence of Notch, and recruits Mastermind-like (MAML) proteins to activate downstream targets (Figure 1) (51–55). These include Hairy and Enhancer of Split (*HES*) and the related *HEY*/*HRT*/*HERP* genes, which all encode for transcriptional repressors (12, 15). Target genes are therefore repressed until additional NICD is produced. NICD activity is regulated by a rapid rate of protein turnover at the PEST degradation domain, which is targeted by ubiquitylation (22).

Notch signaling can be controlled at different stages of the signaling cascade, by regulating ligand-mediated protein cleavage, post-translational modifications, receptor and/or ligand clustering, or transcription factors activity and expression. Among these, O-glycosylation plays a critical role in Notch signaling regulation and specificity (26). Significant efforts have been made to try to understand the intricacies of this process. Three major types of O-glycosylation have been described in mammalian Notch: O-fucosylation by Protein O-Fucosyltransferase 1 (Pofut1), O-glucosylation by Protein O-Glucosyltransferase 1 (Poglut1) and O-GlcNAcylation by O-GlcNAc Transferase 1 (Eogt1). Notably, Notch was shown to contain more putative O-fucosylation and O-glucosylation sites than any other protein (56, 57). In mice and flies, the loss of Pofut1 (or its Drosophila homolog Ofut1) results in severe embryonic defects (58–60). Similar to that, knockout of Poglut1 in mice is lethal at the embryonic stage and displays defects in somitogenesis and cardiogenesis (57). A recent structural study has shown that O-fucose residues on Notch1 EGF8 and 12 directly interact with Jagged1, indicating that O-fucosylation regulates Notch signaling (37). O-fucose residues can be selectively extended by three Fringe enzymes (Manic, Lunatic and Radical Fringe), that add a GlcNAc residue, depending on their position in the Notch ectodomain (61). Lunatic Fringe mutant mice display severe somitogenesis defects and reduced viability at birth (62). Fringe elongation enhances Notch1 binding and activation by both DLL1 and Jagged1 when targeting O-fucose at EGF12, and by DLL1 when targeting

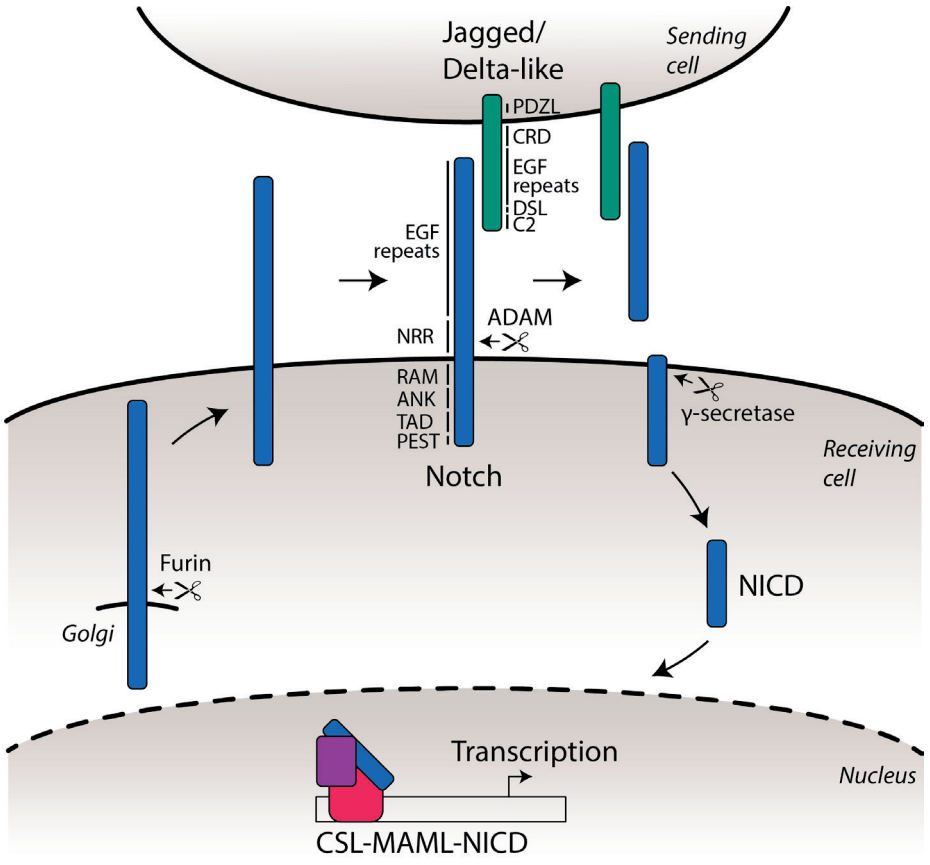


Figure 1. Schematic representation of Notch signaling activation. Following furin processing at the S1 site, the Notch heterodimer is targeted to the cell membrane where it interacts with one of its ligands Jagged or Delta-like in *trans* to activate signaling. Endocytosis from the sending cell generates a pulling force that exposes the S2 site to proteolytic cleavage by an ADAM-family metalloprotease, triggering cleavage at the S3 site by γ-secretase to release the NICD into the receiving cell. Subsequently, the NICD is translocated to the nucleus where it regulates the transcription of target genes by binding to the CSL repressor and recruiting MAML proteins. CRD is specific to Jagged, and PDZL is absent in Jagged2.

O-fucose at EGF8 (30, 37, 61), demonstrating that Fringe modifications mark specific Notch regions in order to fine-tune signaling.

In the last decades, impressive scientific progress has been achieved in understanding the intricacies of Notch signaling, and elucidating its implications in cell fate determination, embryogenesis, and adult tissue homeostasis. In 1991, the core Notch ligand recognition site EGF11-12 was identified (41), and recent crystallographic studies have revealed the atomic details of its interaction with the canonical ligands DLL and Jagged C2-EGF3

domains (37, 63). Besides the EGF11-12 and NRR activation sites, other regions in the Notch ectodomain, such as EGF6, EGF25-26 and EGF36, have also been proposed to play a role in signaling (24, 31, 61, 64–67). A low-resolution electron microscopy reconstruction indicates that the Notch1 ectodomain is dimeric, although the protein was purified unconventionally on affinity grids (68). Structural studies have indicated that flexibility is present to a certain extent in the Notch ectodomain (69, 70), and backfolding models have been suggested based on genetic and interaction studies (64–66). However, direct observation of Notch and ligand ectodomain flexibility is limited. Other than Notch engagement, the Jagged C2 domain has been shown to have additional functionalities such as membrane binding, which is required for optimal Notch activation (70, 71). In addition, in *Xenopus*, the CRD of Serrate-1 (a homolog of Jagged1) has been proposed to play a role in Notch activation in primary neurogenesis (72). Together, these studies indicate that regions asides from the core Notch and ligand activation sites may contribute to signaling and regulation. In this thesis, we studied how the Notch1 core activation sites and other functional regions collectively engage the canonical ligand Jagged1 in the Notch1-Jagged1 full extracellular complex.

EGFR: founding member of the HER tyrosine kinase family

The HER family

The human genome encodes 58 receptor tyrosine kinases (RTKs) divided into 20 subfamilies, the majority of which bind to growth factors and have the ability to auto-phosphorylate (73). Among the RTKs, the HER family (also called ErbB family) is one of the most studied categories of receptors due to its essential roles in key cellular processes including cellular growth, migration, differentiation, and oncogenesis (74–78). Members of the HER family include epidermal growth factor receptor (EGFR; also referred to as HER1 or ErbB1), HER2 (ErbB2), HER3 (ErbB3) and HER4 (ErbB4). In mice, knockouts of HER family members lead to embryonic lethality, with defects observed in brain, heart, bone, and various epithelia, such as skin, eyes and lung, illustrating the importance of these proteins in developmental processes (79, 80). By binding to specific ligands, the receptors initiate a signaling cascade to transmit information into the cell, which is critical to the development and homeostasis of metazoans (Figure 2) (73). Ligand binding is coupled to ectodomain dimerization, conformational rearrangement of the transmembrane region and asymmetric dimerization of the intracellular domains, one of which phosphorylates the other to initiate signaling (Figure 2) (73). The four members of the HER family can form heterodimers, and in particular HER2 and HER3 can exclusively signal through that mechanism as they do not form homodimers (73). Instead, they heterodimerize with one another, and with other HER family receptors to initiate signaling. Once phosphorylated, tyrosine residues stimulate several intracellular signaling pathways, including the mitogen-activated protein kinase (MAPK) and phosphatidylinositol 3-kinase (PI3K)/Akt pathways, both of which are implicated in a wide array of physiological and pathological processes,

and therefore constitute important therapeutic targets (Figure 2) (81). Except for HER2, the receptors are regulated by polypeptide extracellular ligands that all contain a conserved epidermal growth factor (EGF) domain. The 13 ligands that have been identified can be subdivided into three groups. Members of the first group only bind to EGFR and includes EGF, transforming growth factor a (TGF-α), amphiregulin and epigen. Ligands in the second group bind to both EGFR and HER4, and include betacellulin, HB-EGF and epiregulin. The third group includes neuregulins (NRG) 1-4, of which NRG1 and NRG2 bind to HER3 and HER4, while NRG3 and NRG4 bind only to HER4. HER family ligands are produced as membrane-bound precursors processed in a ligand-specific manner (82, 83). Although the role of EGF-like domains of HER ligands is sufficient to explain most of their biological effect, other regions within the full-length ligands probably also influence signaling via mechanisms that remain to be determined.

EGFR dysregulation and associated cancers

EGFR was the first family member shown to be overexpressed in cancers (84), and it is therefore an important therapeutic target (81, 85). EGFR knockout mice exhibit abnormalities in stem cell renewal, as well as in several organs, among which are brain, skin, lung, and the gastrointestinal tract (86, 87). Besides its role in development, EGFR also remains active in the mature nervous system (88). Mutations and dysregulations of EGFR are associated with growth and maintenance of various solid tumors, and specific genetic alterations lead to different types of tumors. Glioblastoma multiforme (GBM) is the most common type of malignant brain tumor in adults, and among the most lethal of all cancers, with current treatments resulting in a median survival of only 12-15 months (89). GBM cells were found to have amplified EGFR in 50% of the cases, and EGFR sequence alteration in 38% of the cases (90). Mutations of the extracellular domain that generate EGFR variants I, II and III (EGFRvI-III) are constitutively active, oncogenic, and frequently found in GBM (Figure 2) (91–95). Of all EGFR mutants, EGFRvIII is the most commonly observed in GBM, accounting for 60 to 70% of them (90). It is characterized by the deletion of amino acids 6-273 in the domains I and II of EGFR, addition of a glycine residue and of a free cysteine residue, together leading to increased homodimerization, impaired downregulation, and aberrant tyrosine kinase activity (90, 96, 97). While wild-type EGFR predominantly signals through the MAPK pathway, the EGFRvIII mutant preferentially actives the PI3K/Akt pathway (98). EGFRvII contains a deletion of amino acids 521-603, located in the cysteine-rich region of the EGFR extracellular domain, and accounts for 5% of EGFR mutations implicated in GBM (90). This mutation might confer a growth advantage to tumor cells (90). In addition, point mutations such as R108K, A289V/D/T and G598D, that keep EGFR in an active conformation, are found in 24% of GBM (94, 99).

In the intracellular segment of EGFR, activating mutations that occur in the membrane-proximal kinase domain promote development of non-small-cell lung cancers (NSCLC), in

particular adenocarcinoma (Figure 2) (75, 100–102). The L858R mutation, located in the activation loop, is the most commonly observed point mutation in the kinase domain, as it represents 45% of the mutations in that domain (81). This mutation destabilizes the domain auto-inhibitory conformation that is normally found in unliganded EGFR, and consequently stabilizes the active conformation to confer a 50-fold increase in kinase activity (103, 104). Various in-frame deletions in exon 19, or in-frame insertions in exon 20, both located in the kinase domain, are also frequently detected in NSCLC (94). Other alterations of the intracellular segment include mutations generating the EGFRvV and EGFRvIV mutants (Figure 2). EGFRvV is characterized by the truncation of most of the C-terminal tail, a region that mediates internalization and degradation, and represents 15% of EGFR mutants involved in GBM (90). An increased ligand-dependent kinase activity is associated with cells that present this mutation (90). EGFRvIV mutants are characterized by deletions in the exons 25-27, and although less frequent, they also have an oncogenic potential (105).

The race against EGFR-targeting drug resistance

To treat EGFR-associated cancers, monoclonal antibodies and tyrosine kinase inhibitors (TKIs) are the main molecules used. The most common EGFR alterations are the L858R single

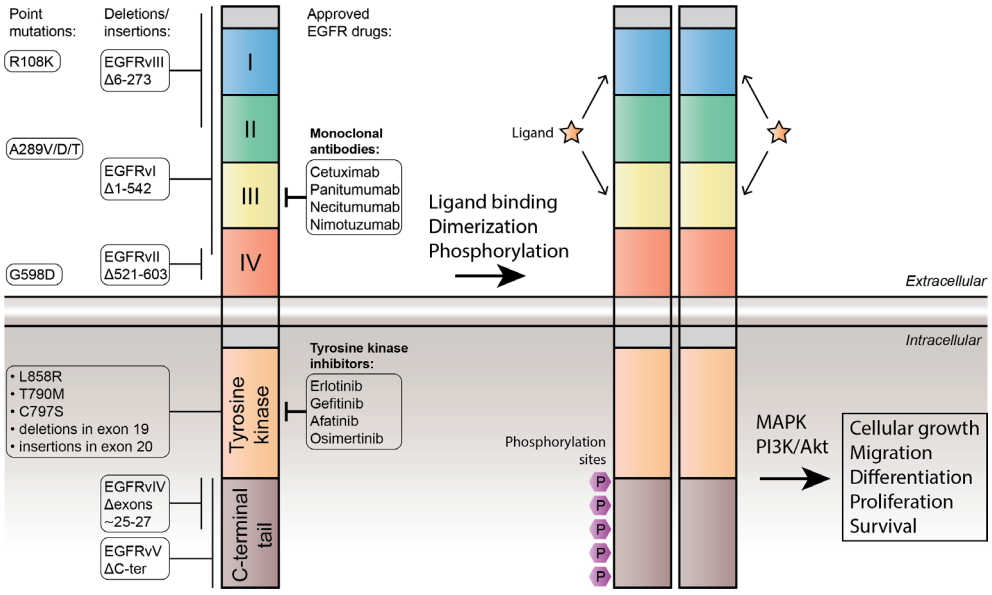


Figure 2. Mutations and available drugs targeting EGFR activation in cancer. All EGFR domains can be affected by genetic alterations such as point mutations, deletions and insertions. Approved treatments against EGFR-associated cancers include monoclonal antibodies, which bind to EGFR domain III, and tyrosine kinase inhibitors, which interact with the tyrosine kinase domain. Ligand binding to EGFR domains I and III is coupled to dimerization of the extracellular region, rearrangement of the transmembrane segment, and asymmetric dimerization of the intracellular region, one of which phosphorylates the other to initiate signaling.

point mutation and exon 19 in-frame deletions, occurring in the tyrosine kinase domain, and respectively accounting for 39% and 46% of all EGFR-activating mutations in lung cancer (106). TKIs are therefore therapeutic molecules of choice. Erlotinib and Gefitinib are approved TKIs used as first-line treatment of patients with locally advanced or metastatic NSCLC whose tumors harbor the L858R mutation or exon 19 deletions (Figure 2) (107). Erlotinib and Gefitinib bind to the ATP binding site in the EGFR kinase domain (108, 109), therefore blocking intracellular phosphorylation and interrupting downstream signaling pathways (107). This results in the inhibition of tumoral cell proliferation and cell death (110). The second-generation TKI Afatinib is an irreversible oral blocker that targets all members of the HER family (111). Like Erlotinib and Gefitinib, it is used in locally advanced and metastatic NSCLC. However, the response to these first- and second-generation TKI is drastically affected by the emergence of resistance to targeted therapy within a year of treatment (112). The T790M "gatekeeper" substitution is one of the most common acquired mutations, as it is observed in more than 50% of all cases (112). The term "gatekeeper" is used to describe the mutant EGFR methionine sidechain that sterically blocks binding of the first- and second-generation of TKI to EGFR. This mutation is located in the ATP binding site of EGFR and is also proposed to mediate TKI resistance by increasing the affinity for ATP (113). Osimertinib is a third-generation TKI, and is currently the gold standard for treatment of patients with NSCLC that acquire the T790M mutation (Figure 2). It has a high selectivity for the L858R and T790M mutant EGFR compared to the wild type (114–117). However, Osimertinib is also associated with the development of resistance after 6-17 months of treatment, for example by the acquisition of the C797S mutation (118), and therefore more research is needed to overcome these alterations. The fourth-generation of TKI is currently being developed, among which the candidate JBJ-04-125-02 was shown to overcome the triple mutant L858R/T790M/C797S *in vitro* and *in vivo* when used in combination with Osimertinib (119). Another candidate, CH7233163, was shown to overcome the triple mutant Del19/T790M/C797S (120).

To date, approved monoclonal antibodies include Cetuximab, Panitumumab, Necitumumab and Nimotuzumab (121). These antibodies target the ligand recognition site on EGFR extracellular domain III (Figure 2) (121–123). Cetuximab is a first-line treatment that competitively blocks ligand-mediated EGFR downstream signaling, and also binding of EGFR to other HER family members (121). It promotes EGFR internalization and degradation, causes cell cycle arrest, and inhibits the expression of pro-angiogenic factors (121). Cetuximab is used to treat head and neck squamous cell carcinoma (HNSCC) and metastatic colorectal cancer (121). Panitumumab, which binds to the same EGFR epitope as Cetuximab, may be effective in patients that acquire the S468R mutation after Cetuximab treatment (123). Panitumumab and Cetuximab inhibit EGFR signaling to similar levels, however Panitumumab is less effective than Cetuximab in mediating antitumor cell immune mechanisms, explaining differences in their clinical efficacy (124). Other approved

monoclonal antibodies include Necitumumab, a new first-line treatment for squamous NSCLC (125), and Nimotuzumab, used in some countries to treat HNSCC and advanced esophageal squamous cell carcinoma (121).

In this race against acquired tumor resistance, monoclonal antibodies are widely used but their large size (~150 kDa) leads to reduced tumor penetration and slow distribution (126–128). To overcome these limitations, the variable domains of heavy chain antibodies (VHH), also referred to as nanobodies in their isolated form, constitute an emerging tool in cancer diagnostics and therapy because of their small size (~15 kDa) and ability to bind to antigens with a high affinity (129–131). Nanobodies originate from Camelidae heavy-chain antibodies, which are composed of a homodimer of heavy chains while lacking light chains, and represent the smallest antigen-binding unit derived from natural sources (132). Although the use of nanobodies in research is a relatively recent occurrence, nanobody-based cancer treatments are currently under assessment in clinical trials (130). In 2019, for the first time a nanobody was approved for therapeutic uses by the United States Food and Drug Administration (FDA) and the European Medicines Agency (EMA), namely the 28-kDa bivalent nanobody Caplacizumab used to treat thrombotic thrombocytopenic purpura (133). EGFR-targeting nanobodies were developed for diagnostic and therapeutic applications, among which are the inhibitory nanobodies EgA1, 9G8 and 7D12 (134–138). Crystal structures of these nanobodies in complex with the inactive EGFR extracellular domain show that they prevent EGFR from adopting the extended conformation that is required for signaling (139). These nanobodies all bind to EGFR domain III, and while EgA1 and 9G8 bind to a cleft formed between domains II and III, 7D12 interacts with the ligand recognition site (139). In this thesis, we study the structure of the EgB4 nanobody, both alone and in complex with the active EGFR-EGF complex, and we describe the molecular mechanism of its non-inhibitory role.

Lipopeptide calcium-dependent antibiotics

The constant rise of antibiotic resistance is a worldwide threat that is considered one of the biggest global challenges by the World Health Organization (140). The identification and development of antibiotic molecules that operate using diverse and unexploited modes of action is key to addressing this growing problem (141). Due to the large costs and high risks associated with drug development, approval of new antibiotics for therapeutic use has not kept pace with the steep rise of antibiotic resistance, and in the last 40 years only two classes of antibiotics that are based on novel chemical scaffolds have obtained market approval (142). Among these, the cyclic lipopeptide Daptomycin is a calcium-dependent antibiotic (CDA) introduced in the clinic in 2003 and used for the management of multi-resistant bacteria, including methicillin-resistant *Staphylococcus aureus* (MRSA)

and vancomycin-resistant enterococci (VRE) (143). Over 40 CDAs have been described, with diverse mechanisms and varying antibacterial activities (144). The structure of most CDAs, including Laspartomycin C, Friulimicin B and Amphomycin, is characterized by a macrocycle constituted of 10 amino acids, and an exocyclic region, composed of at least one amino acid, N-terminally connected to a lipid (Figure 3). The macrocycle of CDAs includes conserved features such as D-amino acids located at specific positions, and an Asp-X-Asp-Gly motif involved in calcium binding that is required for antibacterial activity (145).

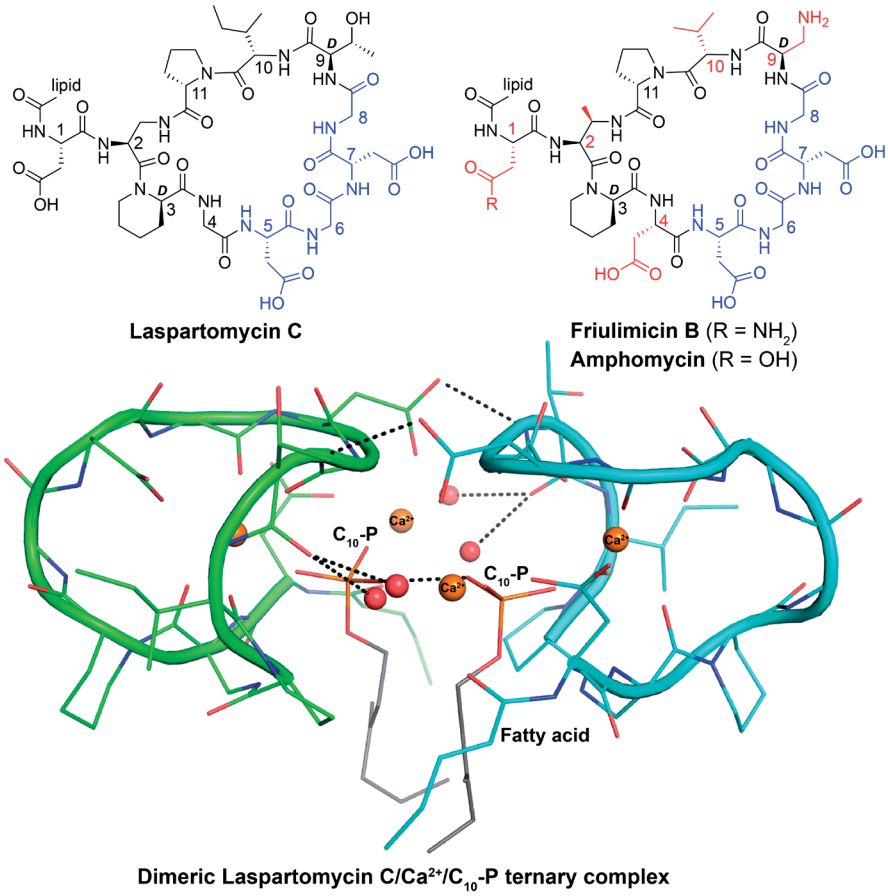


Figure 3. Structures of the CDAs Laspartomycin C, Friulimicin B and Amphomycin. (Top) The macrocycle of Laspartomycin C differs from that of Friulimicin B and Amphomycin at residues 1, 4, 9 and 10 (sidechains colored in red) while the Asp-X-Asp-Gly calcium-binding motif is conserved (blue). (Bottom) The Laspartomycin C/Ca^{2+}/C_{10}-P ternary complex is a dimer maintained by direct and indirect interactions (represented as dotted lines), in which the C_{10}-P ligands are sequestered from the solvent. Individual monomers are colored in green or in cyan. For clarity, calcium coordination is not shown.

Daptomycin has a bactericidal effect on Gram-positive germs by targeting the cell membrane, however its precise mode of action remains an ongoing debate (146–148). Possible mechanisms that could explain Daptomycin activity include inhibition of cell wall synthesis, membrane pore formation, and alteration of the membrane curvature leading to aberrant protein recruitment (147). In contrast, the mode of action of other CDAs such as Laspartomycin C, Friulimicin B and Amphomycin are better understood (149–152). These molecules share the feature of having their macrocycle closed by a lactam linkage and target the same molecule (undecaprenyl phosphate, or C_{55}-P) on the bacterial cell wall. C_{55}-P acts as a lipid carrier in cell wall biosynthesis, as reviewed in (153). Lipid I, a key intermediate of the Gram-positive bacteria cell wall biosynthesis, is formed by association of C_{55}-P with UDP-MurNAc-pentapeptide by the enzyme MraY specifically in the cytoplasmic leaflet of the bacterial membrane. Lipid I is subsequently converted to lipid II by addition of GlcNAc by the enzyme MurG. Lipid II is then translocated to the outer leaflet of the cytoplasmic membrane, where penicillin-binding proteins (PBP) incorporate the disaccharide-pentapeptide motif into the peptidoglycan layer. Finally, the pyrophosphorylated lipid carrier is dephosphorylated by UPP phosphatases to yield the initial C_{55}-P carrier. To start a new cycle, C_{55}-P must be flipped back to the inner side of the membrane, where it can be used again as a lipid carrier. Since these cyclic reactions represent the rate-limiting factor of the cell wall biosynthesis, they constitute important therapeutic targets (153). A wide range of antibiotics act to interfere with the lipid II cycle, by either inhibiting enzyme activity (e.g. PBP are blocked by beta-lactams), or sequestering intermediate carriers (e.g. C_{55}-P is bound by Laspartomycin C) (154).

In order to develop potent antibiotics that target the bacterial cell wall synthesis, studies are needed to provide information on the structure of the antibiotics and on the mechanisms by which they engage their bacterial target. Early structural insights for C_{55}-P-binding CDAs were provided by the structure of Tsushimycin, crystallized however in the absence of its bacterial target, that showed two calcium binding sites and a cavity potentially accommodating substrate binding in a Tsushimycin dimer (155). Recently, the structure of Laspartomycin C in complex with C_{10}-P (a soluble analogue of C_{55}-P) was solved by X-ray crystallography, for the first time providing structural information on a CDA bound to its biomolecular target (Figure 3) (151). The structure shows a saddle-shaped Laspartomycin C molecule bound to one C_{10}-P molecule and two calcium ions playing key roles in ligand engagement (151). The Laspartomycin C/Ca^{2+}/C_{10}-P ternary complex forms a symmetrical dimer stabilized by direct and indirect interactions between the two ternary subunits. As hypothesized from the structure of the unliganded Tsushimycin/Ca^{2+} complex (155), the C_{10}-P molecules insert into the cavity created by the dimeric arrangement of Laspartomycin C. A straightforward model can be derived from this structure, in which the Laspartomycin C fatty acid sidechains and the C_{10}-P isoprenyl tails are both oriented perpendicularly to a hydrophobic plane that is likely parallel to the bacterial membrane,

resulting in Laspartomycin being slightly submerged into the membrane. In this setting, the hydrophobic sidechains of D-Pip[3] and Pro[11], which belong to the lipopeptide macrocycle, could also contribute to interactions with the membrane. However, Laspartomycin C activity remains too low for clinical use, with a minimum inhibitory concentration (MIC) of 8 μg.mL^{-1} against MRSA, versus 0.5 μg.mL^{-1} for Daptomycin (151). Nonetheless, the Laspartomycin C structure provided valuable information for the design of CDA analogues with potentially enhanced activity. Notably, in this structure the macrocycle residues 4, 9 and 10 do not interact with the C_{55}-P head group or with the coordinating calcium ions. Structurally similar to Laspartomycin C, the CDAs Friulimicin B and Amphomycin also engage C_{55}-P. Subtle differences still distinguish them from Laspartomycin C, with changes in macrocycle residues 1, 2, 4, 9 and 10 (Figure 3), as well as in the fatty acid sidechain. This knowledge, coupled with the insights gained by the Laspartomycin C structure, prompted us to investigate the impact of introducing features from the friulimicin/amphomycin class into Laspartomycin C. To achieve this, we performed structure-activity studies and solved high-resolution crystal structures of the new lipopeptide analogues, which provides mechanistic insights into the mode of action of the C_{55}-P-targeting subfamily of CDAs.

Scope of the thesis

This thesis aims to investigate the molecular mechanisms by which two distinct protein systems, cell surface receptors and lipopeptide calcium-dependent antibiotics, control and regulate essential cellular processes. We use a combination of structural biology techniques and therapeutic tools to shed light into the biology of these systems, which opens new avenues in the design and development of future therapeutic molecules.

In **chapter 2**, we explore the molecular mechanisms of Notch1-Jagged1 activation. Using a combination of cross-linking mass spectrometry (XL-MS), and biophysical and structural techniques, we probe the molecular architecture of the Notch1-Jagged1 full extracellular complex. We identify five regions, two on Notch1 and three on Jagged1, that form an intra- and inter-molecular interaction network. We reveal that core Notch1 and Jagged1 activation sites are not distal, as previously thought, but engage directly to control Notch1 signaling. These data, coupled to small-angle X-ray scattering (SAXS) experiments showing Notch1 and Jagged1 ectodomain flexibility, support the formation of non-linear architectures. Collectively, this redefines the Notch1-Jagged1 activation mechanism and opens new avenues for therapeutic applications to treat Notch-associated diseases.

In **chapter 3**, we describe the non-inhibitory mechanism of the EGFR-targeting nanobody EgB4. We solve crystal structures of EgB4 alone, and in complex with the EGF-bound EGFR in the active conformation, revealing that EgB4 binds to a new epitope on EGFR domains I and

II. We compare our active EGFR structure with that of the inactive EGFR, and hypothesize that unlike inhibitory nanobodies, EgB4 can engage both the inactive and active EGFR. Together, this provides the molecular basis for the use of EgB4 as a biomarker to target EGFR-associated cancers, while not affecting EGFR function.

In **chapter 4**, we gain mechanistic insights into C_{55}-P-targeting lipopeptide antibiotics by solving high-resolution crystal structures of two CDA analogues in complex with C_{10}-P and performing structure-activity studies. Specifically, we evaluate the impact of introducing structural features from the friulimicin/amphomycin classes of CDAs into Laspartomycin C. We reveal that the two analogues form a higher-order arrangement, not observed for Laspartomycin C, that governs their interaction with the bacterial membrane and provides an explanation for their activity. In addition, we use live cell imaging to gain further insights into C_{55}-P-targeting lipopeptide antibiotics, and highlight a unique mode of action relative to the widely used Daptomycin.

Acknowledgements

Thanks to Alex Cloherty, Robert Englmeier, and Lucas Chataigner for critically proofreading this chapter. Their feedback was very useful and much appreciated.

References

1. C. H. Heldin, B. Lu, R. Evans, J. S. Gutkind, Signals and receptors. *Cold Spring Harb. Perspect. Biol.* **8** (2016).
2. D. B. Doroquez, I. Rebay, Signal integration during development: Mechanisms of EGFR and notch pathway function and cross-talk. *Crit. Rev. Biochem. Mol. Biol.* **41**, 339–385 (2006).
3. O. L. Mohr, Character Changes Caused by Mutation of an Entire Region of a Chromosome in Drosophila. *Genetics* **4**, 275–282 (1919).
4. D. F. Poulson, The effects of certain X-chromosome deficiencies on the embryonic development of Drosophila melanogaster. *J. Exp. Zool.* **83**, 271–325 (1940).
5. F. Radtke, H. R. MacDonald, F. Tacchini-Cottier, Regulation of innate and adaptive immunity by Notch. *Nat. Rev. Immunol.* **13**, 427–437 (2013).
6. A. Louvi, S. Artavanis-Tsakonas, Notch signalling in vertebrate neural development. *Nat. Rev. Neurosci.* **7**, 93–102 (2006).
7. J. L. Ables, J. J. Breunig, A. J. Eisch, P. Rakic, Not(ch) just development: Notch signalling in the adult brain. *Nat. Rev. Neurosci.* **12**, 269–83 (2011).
8. S. J. Bray, Notch signalling in context. *Nat. Rev. Mol. Cell Biol.* **9**, 722–735 (2016).
9. C. Siebel, U. Lendahl, Notch Signaling in Development, Tissue Homeostasis, and Disease. *Physiol. Rev.* **97**, 1235–1294 (2017).
10. A. P. Weng, *et al.*, Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science (80-.).* **306**, 269–271 (2004).
11. J. C. Aster, W. S. Pear, S. C. Blacklow, The Varied Roles of Notch in Cancer. *Annu. Rev. Pathol. Mech. Dis.* **12**, 245–275 (2017).
12. J. Mašek, E. R. Andersson, The developmental biology of genetic Notch disorders. *Development* **144**, 1743 LP – 1763 (2017).
13. S. Artavanis-Tsakonas, M. D. Rand, R. J. Lake, Notch signaling: Cell fate control and signal integration in development. *Science (80-.).* **284**, 770–776 (1999).
14. L. A. Milner, A. Bigas, Notch as a mediator of cell fate determination in hematopoiesis: Evidence and speculation. *Blood* **93**, 2431–2448 (1999).
15. S. Bray, Notch signalling: a simple pathway becomes complex. *Nat. Rev. Mol. Cell Biol.* **7**, 678–689 (2006).
16. A. Chitnis, D. Henrique, J. Lewis, D. Ish-Horowicz, C. Kintner, Primary neurogenesis in Xenopus embryos regulated by a homologue of the Drosophila neurogenic gene Delta. *Nature* **375**, 761–766 (1995).
17. P. Simpson, Lateral inhibition and the development of the sensory bristles of the adult peripheral nervous system of Drosophila. *Development* **109**, 509–519 (1990).
18. K. D. Irvine, Fringe, Notch, and making developmental boundaries. *Curr. Opin. Genet. Dev.* **9**, 434–441 (1999).
19. Y.-C. Cheng, *et al.*, Notch activation regulates the segregation and differentiation of rhombomere boundary cells in the zebrafish hindbrain. *Dev. Cell* **6**, 539–550 (2004).

20. W. R. Gordon, *et al.*, Structural basis for autoinhibition of Notch. *Nat. Struct. Mol. Biol.* **14**, 295–300 (2007).
21. S.-M. Kwon, C. Alev, S.-H. Lee, T. Asahara, "The Molecular Basis of Notch Signaling: A Brief Overview" in *Notch Signaling in Embryology and Cancer*, J. Reichrath, S. Reichrath, Eds. (Springer US, 2012), pp. 1–14.
22. E. R. Andersson, R. Sandberg, U. Lendahl, Notch signaling: Simplicity in design, versatility in function. *Development* **138**, 3593–3612 (2011).
23. H. Jafar-Nejad, J. Leonardi, R. Fernandez-Valdivia, Role of glycans and glycosyltransferases in the regulation of Notch signaling. *Glycobiology* **20**, 931–949 (2010).
24. R. Rampal, J. F. Arboleda-Velasquez, A. Nita-Lazar, K. S. Kosik, R. S. Haltiwanger, Highly conserved O-fucose sites have distinct effects on Notch1 function. *J. Biol. Chem.* **280**, 32133–32140 (2005).
25. R. Rampal, K. B. L. and R. S. Haltiwanger, Notch Signaling in Normal and Disease States: Possible Therapies Related to Glycosylation. *Curr. Mol. Med.* **7**, 427–445 (2007).
26. Y. Urata, H. Takeuchi, Effects of Notch glycosylation on health and diseases. *Dev. Growth Differ.* **62**, 35–48 (2020).
27. H. Takeuchi, R. S. Haltiwanger, Significance of glycosylation in Notch signaling. *Biochem. Biophys. Res. Commun.* **453**, 235–242 (2014).
28. P. Stanley, T. Okajima, Roles of glycosylation in notch signaling. *Curr. Top. Dev. Biol.* **92**, 131–164 (2010).
29. P. Stanley, Regulation of Notch signaling by glycosylation. *Curr. Opin. Struct. Biol.* **17**, 530–535 (2007).
30. P. Taylor, *et al.*, Fringe-mediated extension of O-linked fucose in the ligand-binding region of Notch1 increases binding to mammalian Notch ligands. *Proc. Natl. Acad. U.S.A.* **111**, 7290–7295 (2014).
31. B. C. Holdener, R. S. Haltiwanger, Protein O-fucosylation: structure and function. *Curr. Opin. Struct. Biol.* **56**, 78–86 (2019).
32. F. Logeat, *et al.*, The Notch1 receptor is cleaved constitutively by a furin-like convertase. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8108–8112 (1998).
33. J. T. Nichols, *et al.*, DSL ligand endocytosis physically dissociates Notch1 heterodimers before activating proteolysis can occur. *J. Cell Biol.* **176**, 445–458 (2007).
34. C. Sanchez-Irizarry, *et al.*, Notch Subunit Heterodimerization and Prevention of Ligand-Independent Proteolytic Activation Depend, Respectively, on a Novel Domain and the LNR Repeats. *Mol. Cell. Biol.* **24**, 9265–9273 (2004).
35. F. Chowdhury, *et al.*, Defining single molecular forces required for notch activation using nano yoyo. *Nano Lett.* **16**, 3892–3897 (2016).
36. K. N. Lovendahl, S. C. Blacklow, W. R. Gordon, "The Molecular Mechanism of Notch Activation" in *Molecular Mechanisms of Notch Signaling*, T. Borggrefe, B. D. Giaimo, Eds. (Springer International Publishing, 2018), pp. 47–58.
37. V. C. Luca, *et al.*, Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity. *Science* **355**, 1320–1324 (2017).

38. L. Meloty-Kapella, B. Shergill, J. Kuon, E. Botvinick, G. Weinmaster, Notch Ligand Endocytosis Generates Mechanical Pulling Force Dependent on Dynamin, Epsins, and Actin. *Dev. Cell* **22**, 1299–1312 (2012).
39. D. Seo, *et al.*, A mechanogenetic toolkit for interrogating cell signaling in space and time. *Cell* **165**, 1507–1518 (2016).
40. X. Wang, T. Ha, Defining single molecular forces required to activate integrin and Notch signaling. *Science (80-.).* **340**, 991–994 (2013).
41. I. Rebay, *et al.*, Specific EGF repeats of Notch mediate interactions with Delta and serrate: Implications for notch as a multifunctional receptor. *Cell* **67**, 687–699 (1991).
42. W. R. Gordon, *et al.*, Mechanical Allostery: Evidence for a Force Requirement in the Proteolytic Activation of Notch. *Dev. Cell* **33**, 729–736 (2015).
43. A. L. Parks, K. M. Klueg, J. R. Stout, M. A. T. Muskavitch, Ligand endocytosis drives receptor dissociation and activation in the Notch pathway. *Development* **127**, 1373–1385 (2000).
44. D. Henrique, F. Schweisguth, Mechanisms of Notch signaling: a simple logic deployed in time and space (2019) https:/doi.org/10.1242/dev.172148 (February 4, 2019).
45. C. Brou, *et al.*, A novel proteolytic cleavage involved in Notch signaling: The role of the disintegrin-metalloprotease TACE. *Mol. Cell* **5**, 207–216 (2000).
46. J. S. Mumm, *et al.*, A ligand-induced extracellular cleavage regulates γ-secretase-like proteolytic activation of Notch1. *Mol. Cell* **5**, 197–206 (2000).
47. R. Kopan, M. X. G. Ilagan, The Canonical Notch Signaling Pathway: Unfolding the Activation Mechanism. *Cell* **137**, 216–233 (2009).
48. G. Yang, *et al.*, Structural basis of Notch recognition by human γ-secretase. *Nature* **565**, 192–197 (2019).
49. B. De Strooper, *et al.*, A presenilin-1-dependent γ-secretase-like protease mediates release of notch intracellular domain. *Nature* **398**, 518–522 (1999).
50. M. Okochi, *et al.*, Presenilins mediate a dual intramembranous γ-secretase cleavage of Notch-1. *EMBO J.* **21**, 5408–5416 (2002).
51. K. Huenniger, *et al.*, Notch1 signaling is mediated by importins alpha 3, 4, and 7. *Cell. Mol. Life Sci.* **67**, 3187–3196 (2010).
52. S. J. Bray, M. Gomez-Lamarca, Notch after cleavage. *Curr. Opin. Cell Biol.* **51**, 103–109 (2018).
53. Y. Nam, A. P. Weng, J. C. Aster, S. C. Blacklow, Structural requirements for assembly of the CSL·intracellular Notch1·Mastermind-like 1 transcriptional activation complex. *J. Biol. Chem.* **278**, 21232–21239 (2003).
54. J. J. Wilson, R. A. Kovall, Crystal structure of the CSL-Notch-Mastermind Ternary complex bound to DNA. *Cell* **124**, 985–996 (2006).
55. R. A. Kovall, S. C. Blacklow, *Mechanistic insights into notch receptor signaling from structural and biochemical studies* (Elsevier Inc., 2010).
56. D. J. Moloney, *et al.*, Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. *J. Biol. Chem.* **275**, 9604–9611 (2000).

57. R. Fernandez-Valdivia, *et al.*, Regulation of mammalian Notch signaling and embryonic development by the protein O-glucosyltransferase Rumi. *Development* **138**, 1925–1934 (2011).
58. Y. Okamura, Y. Saga, Pofut1 is required for the proper localization of the Notch receptor during mouse development. *Mech. Dev.* **125**, 663–673 (2008).
59. S. Shi, P. Stanley, Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5234–5239 (2003).
60. T. Okajima, K. D. Irvine, Regulation of Notch signaling by O-linked fucose. *Cell* **111**, 893–904 (2002).
61. S. Kakuda, R. S. Haltiwanger, Deciphering the Fringe-Mediated Notch Code: Identification of Activating and Inhibiting Sites Allowing Discrimination between Ligands. *Dev. Cell* **40**, 193–201 (2017).
62. Y. A. Evrard, Y. Lun, A. Aulehla, L. Gan, R. L. Johnson, lunatic fringe is an essential mediator of somite segmentation and patterning. *Nature* **394**, 377–381 (1998).
63. V. C. Luca, *et al.*, Structural basis for Notch1 engagement of Delta-like 4. *Science (80-.).* **347**, 847–853 (2015).
64. Z. Pei, N. E. Baker, Competition between Delta and the Abruptex domain of Notch. *BMC Dev. Biol.* **8**, 4 (2008).
65. A. Sharma, A. Rangarajan, R. R. Dighe, Antibodies against the extracellular domain of human Notch1 receptor reveal the critical role of epidermal-growth-factor-like repeats 25-26 in ligand binding and receptor activation. *Biochem. J.* **449**, 519–30 (2013).
66. A. Xu, L. Lei, K. D. Irvine, Regions of Drosophila notch that contribute to ligand binding and the modulatory influence of Fringe. *J. Biol. Chem.* **280**, 30158–30165 (2005).
67. N. Lawrence, T. Klein, K. Brennan, A. Martinez Arias, Structural requirements for notch signalling with delta and serrate during the development and patterning of the wing disc of Drosophila. *Development* **127** (2000).
68. D. F. Kelly, *et al.*, Molecular Structure and Dimeric Organization of the Notch Extracellular Domain as Revealed by Electron Microscopy. *PLoS One* **5**, 1–6 (2010).
69. P. C. Weisshuhn, *et al.*, Non-Linear and Flexible Regions of the Human Notch1 Extracellular Domain Revealed by High-Resolution Structural Studies. *Structure* **24**, 555–566 (2016).
70. R. J. Suckling, *et al.*, Structural and functional dissection of the interplay between lipid and Notch binding by human Notch ligands. *EMBO J.* **36**, 2204–2215 (2017).
71. C. R. Chillakuri, *et al.*, Structural Analysis Uncovers Lipid-Binding Properties of Notch Ligands. *Cell Rep.* **5**, 861–867 (2013).
72. T. Kiyota, T. Kinoshita, Cysteine-rich region of X-Serrate-1 is required for activation of Notch signaling in Xenopus primary neurogenesis. *Int. J. Dev. Biol.* **46**, 1057–60 (2002).
73. M. A. Lemmon, J. Schlessinger, Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
74. T. Hunter, P. Blume-Jensen, Oncogenic kinase signalling. *Nature* **411**, 355 (2001).
75. C. L. Arteaga, J. A. Engelman, ERBB receptors: From oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* **25**, 282–303 (2014).

76. M. A. Lemmon, J. Schlessinger, K. M. Ferguson, The EGFR family: Not so prototypical receptor tyrosine kinases. *Cold Spring Harb. Perspect. Biol.* **6** (2014).
77. A. Citri, Y. Yarden, EGF-ERBB signalling: Towards the systems level. *Nat. Rev. Mol. Cell Biol.* **7**, 505–516 (2006).
78. Y. Yarden, G. Pines, The ERBB network: At last, cancer therapy meets systems biology. *Nat. Rev. Cancer* **12**, 553–563 (2012).
79. M. Sibilia, *et al.*, The epidermal growth factor receptor: From development to tumorigenesis. *Differentiation* **75**, 770–787 (2007).
80. A. W. Burgess, EGFR family: Structure physiology signalling and therapeutic target. *Growth Factors* **26**, 263–274 (2008).
81. P. Wee, Z. Wang, Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers (Basel).* **9**, 1–45 (2017).
82. R. C. Harris, E. Chung, R. J. Coffey, EGF receptor ligands. *Exp. Cell Res.* **284**, 2–13 (2003).
83. A. Buonanno, G. D. Fischbach, Neuregulin and ErbB receptor signaling pathways in the nervous system. *Curr. Opin. Neurobiol.* **11**, 287–296 (2001).
84. D. M. Thompson, G. N. Gill, The EGF receptor: structure, regulation and potential role in malignancy. *Cancer Surv.* **4**, 767–788 (1985).
85. P. Seshacharyulu, *et al.*, Targeting the EGFR signaling pathway in cancer therapy. *Expert Opin. Ther. Targets* **16**, 15–31 (2012).
86. I. D. Campbell, P. Bork, Epidermal growth factor-like modules. *Curr. Opin. Struct. Biol.* **3**, 385–392 (1993).
87. D. W. Threadgill, *et al.*, Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. *Science (80-.).* **269**, 230 LP – 234 (1995).
88. M. Yamada, T. Ikeuchi, H. Hatanaka, The neurotrophic action and signalling of epidermal growth factor. *Prog. Neurobiol.* **51**, 19–37 (1997).
89. T. F. Cloughesy, W. K. Cavenee, P. S. Mischel, Glioblastoma: From molecular pathology to targeted treatment. *Annu. Rev. Pathol. Mech. Dis.* **9**, 1–25 (2014).
90. L. Frederick, X. Y. Wang, G. Eley, C. D. James, Diversity and frequency of epidermal growth factor receptor mutations in human glioblastomas. *Cancer Res.* **60**, 1383–1387 (2000).
91. R. Nishikawa, *et al.*, Immunohistochemical analysis of the mutant epidermal growth factor, deltaEGFR, in glioblastoma. *Brain Tumor Pathol.* **21**, 53—56 (2004).
92. J. M. Francis, *et al.*, EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. *Cancer Discov.* **4**, 956–971 (2014).
93. P. A. Humphrey, *et al.*, Deletion-mutant epidermal growth factor receptor in human gliomas: Effect of type II mutation on receptor function. *Biochem. Biophys. Res. Commun.* **178**, 1413–1420 (1991).
94. F. B. Furnari, T. F. Cloughesy, W. K. Cavenee, P. S. Mischel, Heterogeneity of epidermal growth factor receptor signalling networks in glioblastoma. *Nat. Rev. Cancer* **15**, 302–310 (2015).

95. M. K. Nicholas, R. V. Lukas, N. F. Jafri, L. Faoro, R. Salgia, Epidermal growth factor receptor-mediated signal transduction in the development and therapy of gliomas. *Clin. Cancer Res.* **12**, 7261–7270 (2006).
96. S. I. Ymer, *et al.*, Glioma specific extracellular missense mutations in the first cysteine rich region of epidermal growth factor receptor (EGFR) initiate ligand independent activation. *Cancers (Basel).* **3**, 2032–2049 (2011).
97. R. Nishikawa, *et al.*, A mutant epidermal growth factor receptor common in human glioma confers enhanced tumorigenicity. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7727–7731 (1994).
98. H. K. Gan, A. N. Cvrljevic, T. G. Johns, The epidermal growth factor receptor variant III (EGFRvIII): Where wild things are altered. *FEBS J.* **280**, 5350–5370 (2013).
99. C. W. Brennan, *et al.*, The somatic genomic landscape of glioblastoma. *Cell* **155**, 462 (2013).
100. J. G. Paez, *et al.*, EGFR mutations in lung, cancer: Correlation with clinical response to gefitinib therapy. *Science (80-.).* **304**, 1497–1500 (2004).
101. W. Pao, *et al.*, EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13306–13311 (2004).
102. T. J. Lynch, *et al.*, Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004).
103. C. H. Yun, *et al.*, Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity. *Cancer Cell* **11**, 217–227 (2007).
104. X. Zhang, J. Gureasko, K. Shen, P. A. Cole, J. Kuriyan, An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell* **125**, 1137–1149 (2006).
105. G. Pines, P. H. Huang, Y. Zwang, F. M. White, Y. Yarden, EGFRvIV: A previously uncharacterized oncogenic mutant reveals a kinase autoinhibitory mechanism. *Oncogene* **29**, 5850–5860 (2010).
106. H. Shigematsu, *et al.*, Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J. Natl. Cancer Inst.* **97**, 339–346 (2005).
107. I. Solassol, F. Pinguet, X. Quantin, FDA- and EMA-approved tyrosine kinase inhibitors in advanced EGFR-mutated non-small cell lung cancer: Safety, tolerability, plasma concentration monitoring, and management. *Biomolecules* **9** (2019).
108. J. Stamos, M. X. Sliwkowski, C. Eigenbrot, Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.* **277**, 46265–46272 (2002).
109. S. Yoshikawa, *et al.*, Structural basis for the altered drug sensitivities of non-small cell lung cancer-associated mutants of human epidermal growth factor receptor. *Oncogene* **32**, 27–38 (2013).
110. J. Li, M. Zhao, P. He, M. Hidalgo, S. D. Baker, Differential metabolism of gefitinib and erlotinib by human cytochrome P450 enzymes. *Clin. Cancer Res.* **13**, 3731–3737 (2007).
111. G. M. Keating, Afatinib: A Review of Its Use in the Treatment of Advanced Non-Small Cell Lung Cancer. *Drugs* **74**, 207–221 (2014).

112. D. Westover, J. Zugazagoitia, B. C. Cho, C. M. Lovly, L. Paz-Ares, Mechanisms of acquired resistance to first-and second-generation EGFR tyrosine kinase inhibitors. *Ann. Oncol.* **29**, i10–i19 (2018).
113. C. H. Yun, *et al.*, The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2070–2075 (2008).
114. D. A. E. Cross, *et al.*, AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discov.* **4**, 1046–1061 (2014).
115. X. Zhai, R. A. Ward, P. Doig, A. Argyrou, Insight into the Therapeutic Selectivity of the Irreversible EGFR Tyrosine Kinase Inhibitor Osimertinib through Enzyme Kinetic Studies. *Biochemistry* **59**, 1428–1441 (2020).
116. Y. Yosaatmadja, *et al.*, Binding mode of the breakthrough inhibitor AZD9291 to epidermal growth factor receptor revealed. *J. Struct. Biol.* **192**, 539–544 (2015).
117. J. Niggenaber, *et al.*, Complex Crystal Structures of EGFR with Third-Generation Kinase Inhibitors and Simultaneously Bound Allosteric Ligands. *ACS Med. Chem. Lett.* **11**, 2484–2490 (2020).
118. S. Wang, Y. Song, F. Yan, D. Liu, Mechanisms of resistance to third-generation EGFR tyrosine kinase inhibitors. *Front. Med.* **10**, 383–388 (2016).
119. C. To, *et al.*, Single and dual targeting of mutant egfr with an allosteric inhibitor. *Cancer Discov.* **9**, 926–943 (2019).
120. K. Kashima, *et al.*, CH7233163 overcomes osimertinib-resistant EGFR-Del19/T790M/C797S Mutation. *Mol. Cancer Ther.* **19**, 2288–2297 (2020).
121. W. Q. Cai, *et al.*, The Latest Battles Between EGFR Monoclonal Antibodies and Resistant Tumor Cells. *Front. Oncol.* **10** (2020).
122. S. Li, *et al.*, Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* **7**, 301–311 (2005).
123. E. A. Sickmier, *et al.*, The panitumumab EGFR complex reveals a binding mechanism that overcomes cetuximab induced resistance. *PLoS One* **11**, 1–11 (2016).
124. S. Trivedi, *et al.*, Anti-EGFR Targeted Monoclonal Antibody Isotype Influences Antitumor Cellular Immunity in Head and Neck Cancer Patients. *Clin. Cancer Res.* **22**, 5229 LP – 5237 (2016).
125. A. Díaz-Serrano, A. Sánchez-Torre, L. Paz-Ares, Necitumumab for the treatment of advanced non-small-cell lung cancer. *Futur. Oncol.* **15**, 705–716 (2018).
126. S. I. Rudnick, G. P. Adams, Affinity and avidity in antibody-based tumor targeting. *Cancer Biother. Radiopharm.* **24**, 155–161 (2009).
127. C. G. Siontorou, Nanobodies as novel agents for disease diagnosis and therapy. *Int. J. Nanomedicine* **8**, 4215–4227 (2013).
128. J. H. E. Baker, *et al.*, Direct visualization of heterogeneous extravascular distribution of trastuzumab in human epidermal growth factor receptor type 2 overexpressing xenografts. *Clin. Cancer Res.* **14**, 2171–2179 (2008).
129. M. Kijanka, B. Dorresteijn, S. Oliveira, P. M. P. Van Bergen En Henegouwen, Nanobody-based cancer therapy of solid tumors. *Nanomedicine* **10**, 161–174 (2015).
130. E. Y. Yang, K. Shah, Nanobodies: Next Generation of Cancer Diagnostics and Therapeutics. *Front. Oncol.* **10** (2020).

131. E. De Genst, *et al.*, Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4586–4591 (2006).
132. C. Hamer-Casterman, Atarchouch, T, *et al.*, Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448 (1993).
133. M. Scully, *et al.*, Caplacizumab Treatment for Acquired Thrombotic Thrombocytopenic Purpura. *N. Engl. J. Med.* **380**, 335–346 (2019).
134. R. C. Roovers, *et al.*, Efficient inhibition of EGFR signalling and of tumour growth by antagonistic anti-EGFR Nanobodies. *Cancer Immunol. Immunother.* **56**, 303–317 (2007).
135. E. G. Hofman, *et al.*, EGF induces coalescence of different lipid rafts. *J. Cell Sci.* **121**, 2519–2528 (2008).
136. R. C. Roovers, *et al.*, A biparatopic anti-EGFR nanobody efficiently inhibits solid tumour growth. *Int. J. Cancer* **129**, 2013–2024 (2011).
137. S. Oliveira, *et al.*, Downregulation of EGFR by a novel multivalent nanobody-liposome platform. *J. Control. Release* **145**, 165–175 (2010).
138. R. Van Der Meel, *et al.*, Tumor-targeted Nanobullets: Anti-EGFR nanobody-liposomes loaded with anti-IGF-1R kinase inhibitor for cancer treatment. *J. Control. Release* **159**, 281–289 (2012).
139. K. R. Schmitz, A. Bagchi, R. C. Roovers, P. M. P. Van Bergen En Henegouwen, K. M. Ferguson, Structural evaluation of EGFR inhibition mechanisms for nanobodies/VHH domains. *Structure* **21**, 1214–1224 (2013).
140. P. Beyer, S. Paulin, Priority pathogens and the antibiotic pipeline: An update. *Bull. World Health Organ.* **98**, 151 (2020).
141. M. F. Chellat, L. Raguž, R. Riedl, Targeting Antibiotic Resistance. *Angew. Chemie–Int. Ed.* **55**, 6600–6626 (2016).
142. M. N. Gwynn, A. Portnoy, S. F. Rittenhouse, D. J. Payne, Challenges of antibacterial discovery revisited. *Ann. N. Y. Acad. Sci.* **1213**, 5–19 (2010).
143. M. Heidary, *et al.*, Daptomycin. *J. Antimicrob. Chemother.* **73**, 1–11 (2018).
144. T. M. Wood, N. I. Martin, The calcium-dependent lipopeptide antibiotics: Structure, mechanism, & medicinal chemistry. *Medchemcomm* **10**, 634–646 (2019).
145. M. Strieker, M. A. Marahiel, The structural diversity of acidic lipopeptide antibiotics. *ChemBioChem* **10**, 607–616 (2009).
146. P. T. Hart, *et al.*, A combined solid- and solution-phase approach provides convenient access to analogues of the calcium-dependent lipopeptide antibiotics. *Org. Biomol. Chem.* **12**, 913–918 (2014).
147. S. D. Taylor, M. Palmer, The action mechanism of daptomycin. *Bioorganic Med. Chem.* **24**, 6253–6268 (2016).
148. A. Müller, *et al.*, Daptomycin inhibits cell envelope synthesis by interfering with fluid membrane microdomains. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E7077–E7086 (2016).
149. T. Schneider, *et al.*, The lipopeptide antibiotic friulimicin B inhibits cell wall biosynthesis through complex formation with bactoprenol phosphate. *Antimicrob. Agents Chemother.* **53**, 1610–1618 (2009).

150. L. H. J. Kleijn, *et al.*, Total Synthesis of Laspartomycin C and Characterization of Its Antibacterial Mechanism of Action. *J. Med. Chem.* **59**, 3569–3574 (2016).
151. L. H. J. Kleijn, *et al.*, A High-Resolution Crystal Structure that Reveals Molecular Details of Target Recognition by the Calcium-Dependent Lipopeptide Antibiotic Laspartomycin C. *Angew. Chemie–Int. Ed.* **56**, 16546–16549 (2017).
152. A. Diehl, T. M. Wood, S. Gebhard, N. I. Martin, G. Fritz, The cell envelope stress response of bacillus subtilis towards laspartomycin C. *Antibiotics* **9**, 1–16 (2020).
153. G. Manat, *et al.*, Deciphering the metabolism of undecaprenyl-phosphate: The bacterial cell-wall unit carrier at the membrane frontier. *Microb. Drug Resist.* **20**, 199–214 (2014).
154. T. Schneider, H. G. Sahl, An oldie but a goodie–cell wall biosynthesis as antibiotic target pathway. *Int. J. Med. Microbiol.* **300**, 161–169 (2010).
155. G. Bunkóczi, L. Vértesy, G. M. Sheldrick, Structure of the lipopeptide antibiotic tsushimycin. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 1160–1164 (2005).

Chapter 2

Notch-Jagged signaling complex defined by an interaction mosaic

Matthieu R. Zeronian,[1,†] Oleg Klykov,[2,3,†,‡] Júlia Portell i de Montserrat,[1,§] Maria J. Konijnenberg,[1] Anamika Gaur,[1,‖] Richard A. Scheltema,[2,3,*] Bert J. C. Janssen[1,*]

[1]Structural Biochemistry, Bijvoet Centre for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3]Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

[†]These authors contributed equally to this work.

[‡]Present address: Department of Biochemistry and Molecular Biophysics, Columbia University, 650 West 168th Street, New York, NY 10032

[§]Present address: Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna BioCenter (VBC), Dr. Bohrgasse 3, 1030 Vienna, Austria

[‖]Present address: NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Department of Chemistry, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

*Corresponding author

Email: r.a.scheltema@uu.nl (R.A.S.), twitter@xlinkx_pd; b.j.c.janssen@uu.nl (B.J.C.J.)

This chapter has been published in *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2102502118 (2021).

Abstract

The Notch signaling system links cellular fate to that of its neighbors, driving proliferation, apoptosis, and cell differentiation in metazoans, whereas dysfunction leads to debilitating developmental disorders and cancers. Other than a five-by-five domain complex, it is unclear how the 40 extracellular domains of the Notch1 receptor collectively engage the 19 domains of its canonical ligand Jagged1 to activate Notch1 signaling. Here, using cross-linking mass spectrometry (XL-MS), biophysical and structural techniques on the full extracellular complex and targeted sites, we identify five distinct regions, two on Notch1 and three on Jagged1, that form an interaction network. The Notch1 membrane-proximal regulatory region individually binds to the established Notch1 epidermal growth factor (EGF) 8-13 and Jagged1 C2-EGF3 activation sites, as well as to two additional Jagged1 regions, EGF8-11 and cysteine-rich domain (CRD). XL-MS and quantitative interaction experiments show that the three Notch1 binding sites on Jagged1 also engage intramolecularly. These interactions, together with Notch1 and Jagged1 ectodomain dimensions and flexibility determined by small-angle X-ray scattering (SAXS), support the formation of non-linear architectures. Combined, the data suggest that critical Notch1 and Jagged1 regions are not distal, but engage directly to control Notch1 signaling, thereby redefining the Notch1-Jagged1 activation mechanism and indicating new routes for therapeutic applications.

Significance Statement

Communication between cells is essential for the development and homeostasis of tissues and prevents diseases, including cancers. The Notch and Jagged transmembrane proteins interact to regulate cell-cell communication in all multicellular animals. Defining their interactions is critical to understand Notch-associated disorders. While structural studies have focused on short regions of both proteins, it is unclear how their entire extracellular domains collectively engage to activate signaling. Here we identify several unreported interacting regions in the Notch1-Jagged1 full extracellular complex. We show that Notch1 and Jagged1 ectodomains are not fully extended and reveal that activation-determining regions, previously thought to be distal, engage directly to control signaling. This interaction network redefines our knowledge on Notch activation and provides new avenues for therapeutic advances.

Introduction

Notch signaling plays a central role in developmental processes by determining cell fate decisions in tissues during development. In adults, these signals both determine differentiation and maintenance of neuronal and hematopoietic stem cells as well as regulate the immune system (1–4). Dysregulation often leads to debilitating diseases in humans, including congenital disorders and cancers (5–8). The mammalian Notch1 receptor is the prototypical member of the Notch protein family, which consists of four paralogs (Notch1-4) that all receive signals from the associated ligands Jagged1, Jagged2, Delta-like1, and Delta-like4: in *trans* (from adjacent cells) to initiate signaling, or in *cis* (from the same cell) to inhibit signaling. The Notch1-Jagged1 receptor-ligand pair has been widely studied at functional, cellular, and molecular levels (4, 5). Both Notch1 and Jagged1 are type-I transmembrane proteins with large modular extracellular segments that determine interaction specificity and control the activation of signaling. Notch1 has an extracellular segment of 209 kDa composed of 36 EGF repeats followed by the negative regulatory region (NRR) at the membrane-proximal side, and differs from its paralogs in the number of EGF domains: from 36 for Notch2, 34 for Notch3 and 29 for Notch4. The Jagged1 ectodomain (139 kDa) is similar to that of Jagged2 and is composed of a C2 lipid-binding domain, a Delta/Serrate/Lag-2 (DSL) domain, 16 EGF repeats and a CRD at the membrane-proximal side.

The prevailing model for canonical Notch activation states that ligand binding at Notch1 EGF8-12 and an endocytosis-induced pulling force (9–16), generated by the signal-sending cell on the Notch-ligand complex (17, 18), triggers a conformational change and proteolytic processing in the Notch NRR located 24 EGF domains downstream of the ligand binding site (19–21). After Notch cleavage within the transmembrane domain (22, 23), the Notch intracellular domain translocates to the nucleus where it regulates transcription (24). At the N-terminal side of Jagged1, the C2-EGF3 region is important for Notch1 binding (11, 25–28). A recent structural study demonstrated that the Notch1 EGF8-12 region interacts in an antiparallel fashion through an extended interface with the Jagged1 C2-EGF3 region (11). Additional interactions add complexity to the mechanism of Notch activation and regulation. Notch-ligand, Notch-Notch and ligand-ligand interactions in *cis* can both inhibit (29–31) or activate (32) signaling. In addition to the canonical ligand binding site on EGF8-12 and the conformational change in the NRR, several other extracellular regions, such as EGF6, EGF25-26 and EGF36, seem to play a role in Notch function (33–39). Also, the Jagged1 extracellular segment harbors additional functionality other than the C2-EGF3 region interacting to Notch. It has been suggested that Jagged and Delta-like C2 domain binding to membranes has an important role in regulating ligand-dependent Notch signaling (26, 28). The CRD in *Xenopus* Serrate-1, a homolog of mammalian Jagged1, is required for Notch

activation in primary neurogenesis (40). These studies indicate that several sites in the Notch and Jagged extracellular segments may contribute to Notch signaling and regulation.

Structural studies have revealed details of key interaction sites (11, 41) and indicate that flexibility is present to a certain extent in the Notch and Jagged ectodomains (28, 42). A low-resolution negative stain electron microscopy reconstruction of the Notch1 ectodomain resolved distinct globular dimer states, although this protein was purified in an unconventional manner (43). Backfolded models for the Notch ectodomain have also been suggested based on genetic and interaction studies (33–35). Nonetheless, direct observations of ectodomain flexibility and non-extended architectures are limited. While Notch-Jagged interaction studies have focused predominantly on the well-established Notch1 EGF11-12–Jagged1 C2-EGF3 regions, other sites may play a direct role in this intermolecular interaction. Structural and biophysical studies on the full extracellular portions of Notch and Jagged have however been limited due to the size, flexibility and low expression levels of the proteins, hampering the identification of several interacting regions.

In this study, we combine cross-linking mass spectrometry, quantitative interaction assays and SAXS on purified Notch1 and Jagged1 full ectodomains, as well as shorter constructs, to probe the structure of the Notch1-Jagged1 complex and of the unliganded proteins (Fig. 1 *A-D*). This analysis reveals several, hitherto unreported, intra- and intermolecular interaction regions. We show that Jagged1 C2-EGF3, EGF8-11 and CRD can all interact with Notch1 EGF33-NRR and that the Notch1 NRR is sufficient for the interaction with Jagged1 C2-EGF3. In addition, the Notch1 EGF8-13 region directly interacts with Notch1 EGF33-NRR. XL-MS analysis suggested that four regions, C2-EGF1, EGF5-6, EGF9-12 and CRD, are in proximity within Jagged1, and we confirmed direct interactions for C2-EGF3 binding to EGF8-11 and to CRD. These data, together with SAXS analysis of the Notch1 and Jagged1 ectodomains, suggest that the proteins are not fully extended and indicate that regions in both proteins, *i.e.* Notch1 EGF8-13, Notch1 EGF33-NRR and Jagged1 C2-EGF3, previously shown to be important for Notch signaling, affect each other directly.

Results

XL-MS of the Notch1-Jagged1 complex reveals a mosaic of interaction sites

To determine which regions, beyond the canonical Notch1$^{EGF8\text{-}12}$-Jagged1$^{C2\text{-}EGF3}$ interaction site, are involved in receptor-ligand binding, we probed full ectodomains of Notch1 and Jagged1 (Notch1^{fe}-Jagged1^{fe}) with XL-MS (Figs. 1 *A-D*, 2 *A-B* and *SI Appendix*, Table S1 and Datasets S1,S2). Two variants of Jagged1 were used: a wild-type version (Jagged1^{fe,wt}), and one with five point mutations in the Jagged C2 region (Jagged1^{fe,HA}) that provide higher-

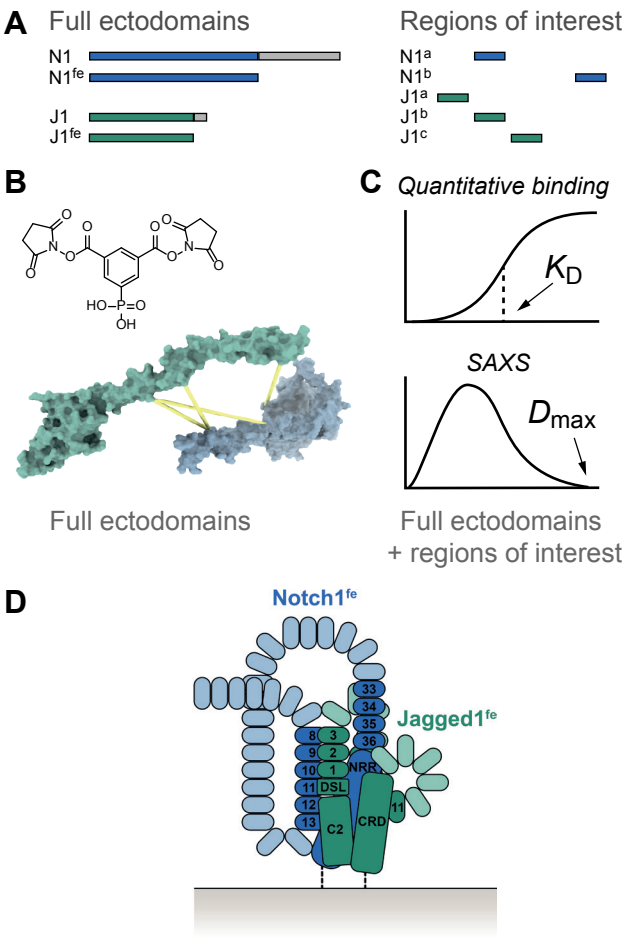


Fig. 1. XL-MS and biophysical studies reveal an interaction network in the Notch1-Jagged1 complex. (*A*) Notch1^{fe}, Jagged1^{fe} and targeted sites are expressed in HEK293 cells and purified by IMAC and SEC. (*B*) Identification of regions in proximity in the Notch1^{fe}-Jagged1^{fe} complex by XL-MS using PhoX (44). (*C*) The purified full ectodomain samples and shorter regions of interest are used in quantitative binding experiments to confirm direct interactions and SAXS studies. (*D*) The resulting data provides insights into the molecular architecture of the Notch1-Jagged1 complex, represented here as a schematic in a *cis* setting.

affinity binding to Notch1 EGF8-12 when incorporated in a Jagged1 C2-EGF3 construct (11). In surface plasmon resonance (SPR) experiments, where Notch1^{fe} is coupled at the C-terminus to the sensor surface to achieve a close-to-native topology (see Methods), Notch1^{fe}-Jagged1^{fe,HA} interact with a dissociation constant (K_D) of 1 μM and Jagged1^{fe,HA} interacts with similar affinity to the EGF8-13 portion of Notch1, while no interaction was measured between Jagged1^{fe,wt} and Notch1 EGF8-13 at 1 μM (Fig. 2 *C-D* and *SI Appendix*, Fig. S1 *A-B*).

Purified Notch1 and Jagged1 full ectodomain proteins were incubated at a 1 to 1 molar ratio to induce complex formation, *i.e.* Notch1^{fe}-Jagged1^{fe,wt} and Notch1^{fe}-Jagged1^{fe,HA}, and cross-linked with the lysine-targeting PhoX cross-linking reagent (44). In subsequent steps, the samples were subjected to deglycosylation, enriched for cross-linked peptides by immobilized metal affinity chromatography (IMAC) and finally analyzed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). From three independent replicates for each complex, we detected 166 unique distance restraints for Notch1^{fe}-Jagged1^{fe,wt} and 232 for Notch1^{fe}-Jagged1^{fe,HA}. As an additional step to reduce false positives and remove distance constraints arising from non-specific aggregation we solely retained restraints detected in at least two out of three replicates (45). This reduced the output to 113 and 164 restraints for Notch1^{fe}-Jagged1^{fe,wt} and Notch1^{fe}-Jagged1^{fe,HA} respectively (Fig. 2 *A and B*). For both complex samples, few intra-links were detected for Notch1^{fe} (9 for Notch1^{fe}-Jagged1^{fe,wt} and 12 for Notch1^{fe}-Jagged1^{fe,HA} and *SI Appendix*, Fig. S2 *A-D*). The number of intra-links for Jagged1^{fe} was however significantly larger and increased by 38% for the mutant (100 for Notch1^{fe}-Jagged1^{fe,wt} and 138 for Notch1^{fe}-Jagged1^{fe,HA}). A similar trend was visible in the number of intermolecular connections between Notch1 and Jagged1 where 3 inter-links were detected for Notch1^{fe}-Jagged1^{fe,wt} and 13 for Notch1^{fe}-Jagged1^{fe,HA}. This identification of intra- and inter-links suggests that the mutant protein, Jagged1^{fe,HA}, assisted by the stronger interaction between the two molecules, adopts a less flexible conformation compared to Jagged1^{fe,wt}, and provides more efficient complex formation that is beneficial for our approach (46).

The inter-links reveal that in the Notch1^{fe}-Jagged1^{fe} complex, three Jagged1 regions, C2-EGF1, EGF10 and CRD are in proximity to the Notch1 EGF29-NRR site with most inter-links arising from the Jagged C2-EGF1 region. The XL-MS experiments do not reveal any cross-links or mono-links between Notch1 EGF8-12 and Jagged1 C2-EGF3 (Fig. 2 *A-B* and *SI Appendix*, Fig. S2*A*), the well-established interaction site (11) for which we find a K_D of 0.3 μM by SPR, using the high-affinity variant of Jagged1 C2-EGF3 (Fig. 2*E* and *SI Appendix*, Fig. S1*C*). There are two possible explanations for the lack of links to Notch1 EGF8-12. **(I)** The two lysine residues in Notch1 EGF8-12, Lys395 and Lys428, are occluded in the Notch1^{fe}-Jagged1^{fe} complex or **(II)** the lysines are occluded from the cross-linking reaction by O-linked glycans such as O-fucose residues, which we show are present in our Notch1 sample (*SI Appendix*, Fig. S2*E* and Dataset S3). Shotgun mass spectrometric analysis of non-cross-linked Notch1^{fe} covers the segment containing the two lysine residues within the Notch1 EGF8-12 region, indicating that the relevant peptides can be identified (*SI Appendix*, Fig. S2*A*). A large part of the Notch1^{fe} EGF repeat region is decorated with O-linked glycosylation sites, with an average of 1.5 sites per EGF domain based on sequence prediction (47), and we cannot fully exclude the glycans prevent the cross-linking reaction. Notably, however, 25 cross-links are identified in the Notch1 EGF29-36 region, predicted to contain slightly less O-linked glycosylation sites, *i.e.* 1.1 sites per EGF domain (47). Combined, these observations suggest

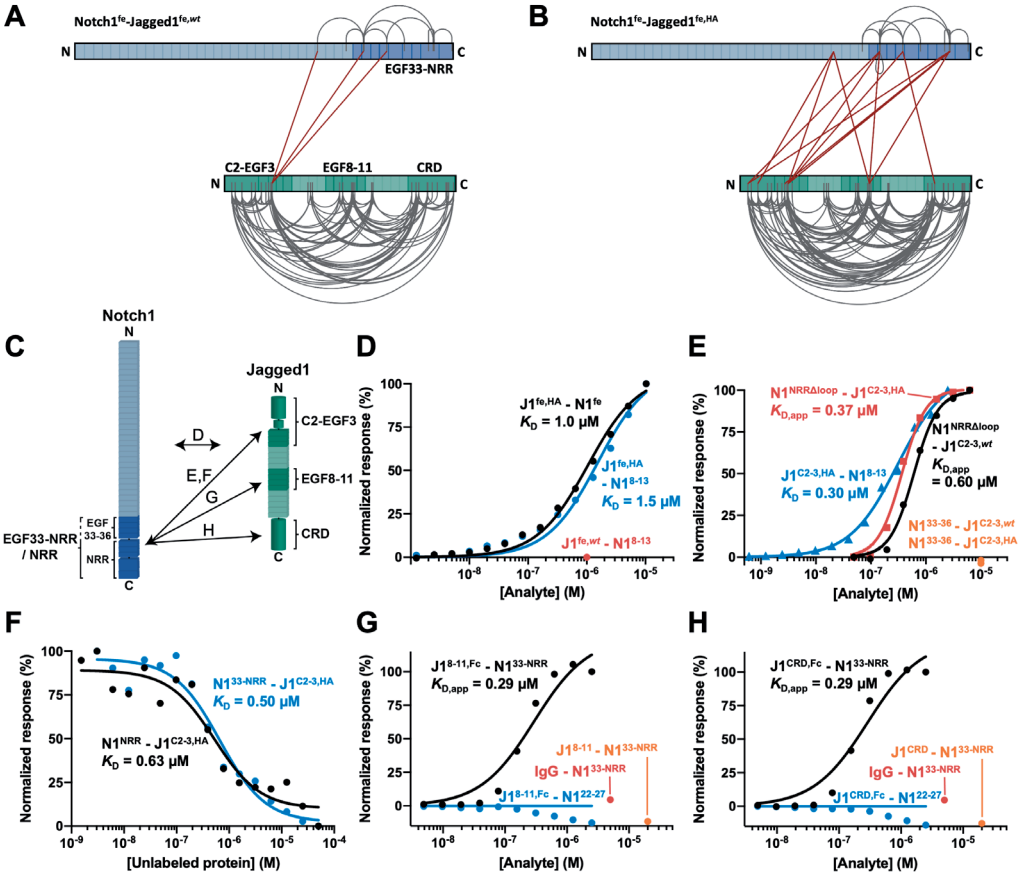


Fig. 2. The Notch1 C-terminal region interacts with Jagged1$^{C2\text{-}EGF3}$, Jagged1$^{EGF8\text{-}11}$ and Jagged1^{CRD} in the Notch1^{fe}-Jagged1^{fe} complex. (*A* and *B*) Overview of the detected distance constraints from the XL-MS experiments, for *wild-type* (*A*) and high-affinity (*B*) versions of Jagged1^{fe}. (*C*) Schematic representation of the interactions reported in panels (*D-H*), based on the XL-MS data and quantitative binding experiments. (*D*) SPR equilibrium binding plots of Jagged1^{fe,HA} to Notch1^{fe} (black) and to Notch1$^{EGF8\text{-}13}$ (blue). Jagged1^{fe,wt} does not interact with Notch1$^{EGF8\text{-}13}$ at 1 µM (red). (*E*) SPR equilibrium binding plots of Notch1$^{NRR\Delta loop}$ to *wild-type* (black) and to high-affinity (red) versions of Jagged1$^{C2\text{-}EGF3}$, and of Jagged1$^{C2\text{-}EGF3,HA}$ to Notch1$^{EGF8\text{-}13}$ (blue). Notch1$^{EGF33\text{-}36}$ does not bind to Jagged1$^{C2\text{-}EGF3}$ constructs (orange). A Hill coefficient of 2 is used to model the Notch1$^{NRR\Delta loop}$–Jagged1$^{C2\text{-}EGF3}$ interactions (see also Methods). (*F*) MST binding curve of Notch1^{NRR} (black) and Notch1$^{EGF33\text{-}NRR}$ (blue) to Jagged1$^{C2\text{-}EGF3,HA}$. (*G* and *H*) SPR equilibrium binding plots indicate interaction of dimerized Jagged1$^{EGF8\text{-}11,Fc}$ (*G*) and dimerized Jagged1^{CRD,Fc} (*H*) to Notch1$^{EGF33\text{-}NRR}$ (black) but not to Notch1$^{EGF22\text{-}27}$ that acts as negative control (blue). Non-dimerized versions do not interact at 20 µM (orange). The Fc domain does not interact with Notch1$^{EGF33\text{-}NRR}$ as shown by the IgG control at 5 µM (red).

that Notch1 EGF8-12 is hidden in the folded Notch1 full ectodomain. Although the XL-MS analysis has not revealed all the interacting regions on Notch1 in the Notch1^{fe}-Jagged1^{fe} complex, it does indicate that the Notch1 C-terminal region plays an important role in the interaction with Jagged1.

Notch1 NRR directly interacts with Jagged1 C2-EGF3

To further investigate interacting regions, we generated shorter Notch1 and Jagged1 constructs (*SI Appendix*, Fig. S3) and probed them by SPR and microscale thermophoresis (MST). The Notch1$^{EGF33\text{-}NRR}$ site interacts directly with Jagged1$^{C2\text{-}EGF3}$ in MST (Fig. 2*F* and *SI Appendix*, Fig. S1*G*) and in SPR (*SI Appendix*, Fig. S4 *A-C*), and this interaction is independent of the high-affinity mutations in the C2 domain of Jagged1 (Table 1 and *SI Appendix*, Fig. S4 *A-D*). Jagged1$^{C2\text{-}EGF3}$ is required and sufficient for the interaction with Notch$^{EGF33\text{-}NRR}$ (*SI Appendix*, Fig. S4 *A-F*). The Notch1$^{EGF33\text{-}NRR}$-Jagged1$^{C2\text{-}EGF3}$ binding site was further defined to Notch1^{NRR}, that binds with a K_D of 0.6 μM to Jagged1$^{C2\text{-}EGF3,HA}$, measured in solution by MST (Fig. 2*F* and *SI Appendix*, Fig. S1*F*), while Notch1$^{EGF33\text{-}36}$ by itself does not interact with either Jagged1$^{C2\text{-}EGF3,wt}$ or Jagged1$^{C2\text{-}EGF3,HA}$ (Fig 2*E*). In the NRR, a large unstructured loop (consisting of 38 residues) that contains the heterodimerization S1 cleavage site (21, 48) is not required for interaction (Fig. 2*E* and *SI Appendix*, Fig. S1*D*). In addition, the interaction is not affected by the high-affinity mutations in Jagged1$^{C2\text{-}EGF3}$, as the K_D values determined by SPR for Notch1$^{NRR\Delta loop}$ binding to Jagged1$^{C2\text{-}EGF3,wt}$ or to Jagged1$^{C2\text{-}EGF3,HA}$ are similar (Fig. 2*E* and *SI Appendix*, Fig. 1 *D-E*). Docking of the Notch1^{NRR}-Jagged1$^{C2\text{-}EGF3}$ complex, using the intermolecular cross-links as restraints, suggests that Jagged1 domains DSL-EGF1 engage Notch1 NRR (*SI Appendix*, Fig. S5) (11). However, given that the two cross-link sites on Notch1 NRR are both in the flexible heterodimerization loop, which we show is not involved in the interaction (Fig. 2*E*), there is ambiguity in the location of the Notch1^{NRR}-Jagged1$^{C2\text{-}EGF3}$ interaction site. Taken together, our interaction data on the smaller Notch1 and Jagged1 portions show that the Notch1 NRR is responsible for the interaction with the Jagged1 C2-EGF3 region.

Notch1 EGF33-NRR contains low affinity sites for Jagged1 EGF8-11 and Jagged1 CRD

The XL-MS data of Notch1^{fe}-Jagged1^{fe,HA} indicates that two additional regions in Jagged1, EGF10 and CRD, are in proximity to the Notch1 EGF33-NRR site (Fig. 2*B*). SPR binding experiments confirm the direct interactions to Notch1$^{EGF33\text{-}NRR}$, albeit with much lower affinity than the Jagged1 C2-EGF3 region, with no binding of Jagged1$^{EGF8\text{-}11}$ or Jagged1^{CRD} to Notch1$^{EGF33\text{-}NRR}$ observed at concentration of 20 μM (Fig. 2 *G* and *H*). To enhance a possible weak affinity, we employed a widely used strategy for cell and surface binding assays of artificially dimerizing proteins (49) that has previously been used to measure Notch interactions (25, 34). Fc-tagged versions of Jagged1$^{EGF8\text{-}11}$ and Jagged1^{CRD}, that are covalently

Table 1. Summary of measured affinities. All values are expressed in μM and derived from MST (*) or SPR experiments. n.d. = not determined.

	Analyte						
Ligand	**Notch1$^{NRR\Delta loop}$**	**Notch1^{NRR}**	**Notch1$^{EGF33\text{-}NRR}$**	**Jagged1^{fe,HA}**	**Jagged1$^{C2\text{-}EGF3,HA}$**	**Jagged1$^{EGF8\text{-}11,Fc}$**	**Jagged1^{CRD,Fc}**
Jagged1$^{C2\text{-}EGF3,wt}$	0.60 ± 0.03	n.d.	28 ± 2	n.d.	n.d.	0.34 ± 0.08	0.93 ± 0.13
Jagged1$^{C2\text{-}EGF3,HA}$	0.37 ± 0.02	0.63 ± 0.18*	0.50 ± 0.19* 15 ± 2	n.d.	n.d.	0.33 ± 0.07	0.57 ± 0.07
Jagged1$^{C2\text{-}EGF7,wt}$	n.d.	n.d.	19 ± 3	n.d.	n.d.	0.40 ± 0.11	0.89 ± 0.12
Jagged1$^{C2\text{-}EGF13,wt}$	n.d.	n.d.	8 ± 2	n.d.	n.d.	0.17 ± 0.03	0.26 ± 0.04
Jagged1^{fe,wt}	n.d.	n.d.	n.d.	n.d.	n.d.	0.22 ± 0.06	0.68 ± 0.10
Notch1^{fe}	n.d.	n.d.	n.d.	1.0 ± 0.1	n.d.	n.d.	n.d.
Notch1$^{EGF8\text{-}13}$	n.d.	n.d.	115 ± 8	1.5 ± 0.2	0.30 ± 0.03	n.d.	n.d.
Notch1$^{EGF33\text{-}NRR}$	n.d.	n.d.	n.d.	n.d.	n.d.	0.29 ± 0.09	0.29 ± 0.10

dimerized by the Fc tag, interact both with a $K_{D,app}$ of 0.29 μM to Notch1$^{EGF33\text{-}NRR}$ (Fig. 2 *G-H* and *SI Appendix*, Fig. S1 *H-I*).

Notch1^{fe} is flexible and has intramolecular interactions

SAXS analysis coupled to size-exclusion chromatography (SEC-SAXS) shows that monomeric Notch1^{fe} is a flexible molecule (Fig. 3 *A* and *B*), has a radius of gyration (R_g) of 105 ± 0.4 Å (Fig. 3*C*) and a maximum distance (D_{max}) of 380 Å (Fig. 3*D*). This suggests that Notch1^{fe} does not exist as an extended molecule, as it would have a D_{max} of 1,027 Å for a fully elongated Notch1^{fe} (see Methods), but instead has a non-linear architecture. Backfolded models were previously suggested based on genetic (35) and interaction data (33, 34), where the EGF domain connections were determined to confer flexibility to the Notch1 extracellular region (42). In addition, two parts in Notch1, EGF8-13 and EGF33-NRR, interact with a K_D of 115 μM (Fig. 3*E* and *SI Appendix*, Fig. S1*J*). While this is a relatively low affinity for an intermolecular interaction, *i.e.* as in a Notch1 dimer, it may be possible that these regions interact directly in an intramolecular fashion within the same Notch1 molecule. Overall, the non-linear architecture suggests that EGF domains may become buried in the fully folded molecule, providing further support to the data obtained by XL-MS.

Notch1 dimerizes through the NRR

Human Notch1^{fe} is a monomer at a concentration of 0.26 μM and has a molecular weight of 209 ± 2.4 kDa (Fig. 3*F*). This correlates well with the theoretical molecular weight of 200-220 kDa that is dependent on the glycosylation state (38, 50). Although mouse Notch1^{fe} has an additional cysteine at EGF25, it does not form a covalent homodimer (Fig. 3*G*). Interestingly, our XL-MS data showed that Notch1^{fe} can form dimers, which can be detected by XL-MS when the same residue in the protein sequence is linked by two different peptides induced by *e.g.* a missed cleavage. One self-link at lysine residue 1314 in EGF34 arises from an intermolecular Notch1-Notch1 interaction (Fig. 2*B*). In addition, the Notch1 NRR itself (Notch1^{NRR}) undergoes weak concentration-dependent dimerization during size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) analysis at concentrations ranging from 1.4 to 17 μM (Fig. 3*H*). Dimerization of the NRR has previously been reported for Notch3 and was predicted for the Notch1 NRR based on similarities in crystal packing comparing the NRR of Notch3 and Notch1 (51–53). The NRR-controlled dimerization of Notch3 may maintain the receptor in an autoinhibited state before ligand binding (53). We determined a crystal structure of the S1-cleaved mouse Notch1 NRR (*SI Appendix*, Fig. S6*A-C*; PDB: 7ABV) that shows the same dimerization interface as its human ortholog (51, 52). N-linked glycans, that do not seem to interfere with dimerization, are visible in the electron density at position N1489, as also reported previously (54), and additionally at position N1587 (*SI Appendix*, Fig. S6*A*). Taken together, the XL-MS analysis on Notch1^{fe} and dimerization of Notch1^{NRR} indicate that Notch1 can dimerize through the membrane proximal region.

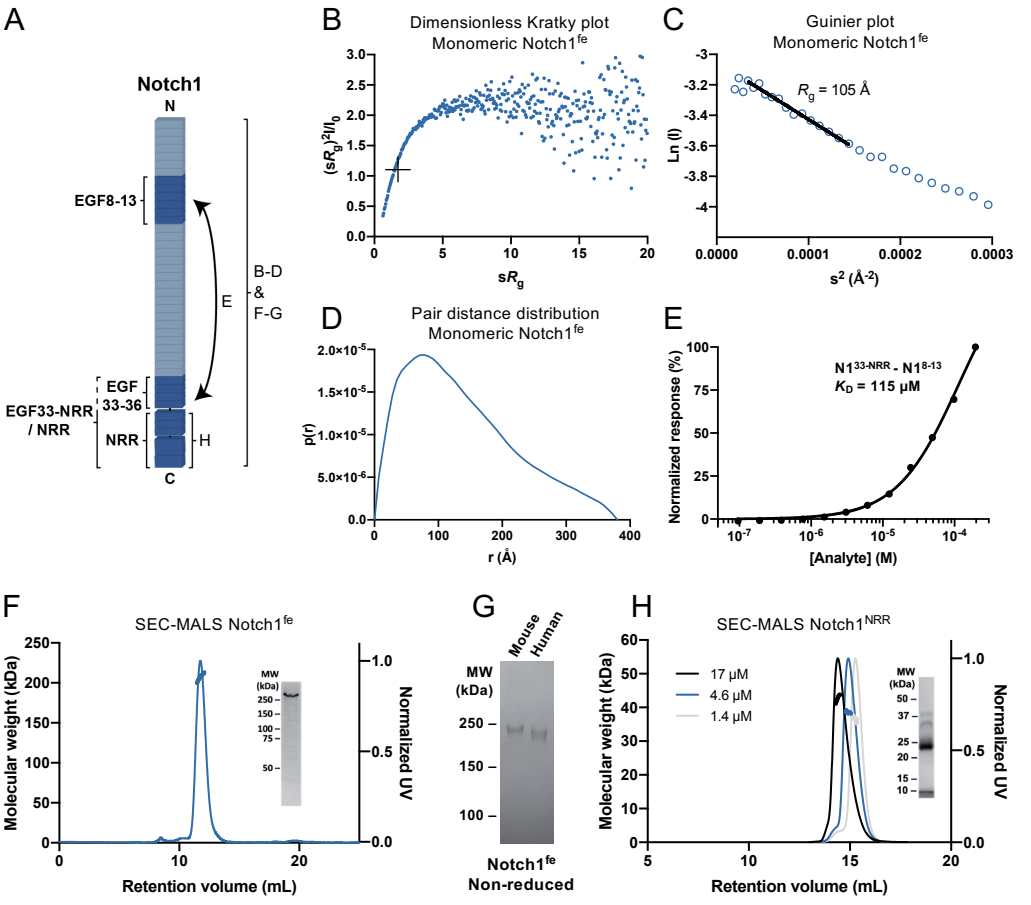
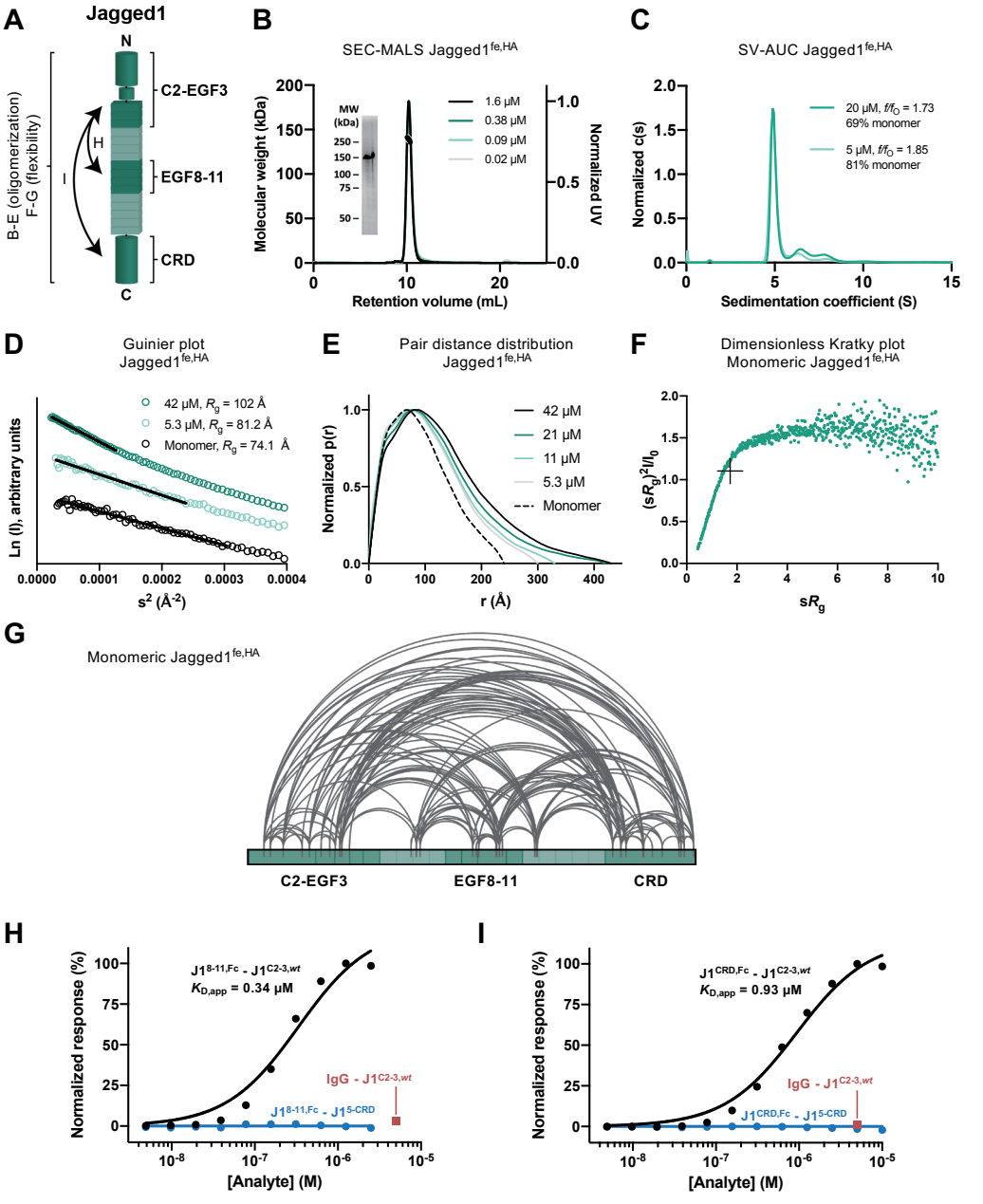


Fig. 3. Notch1^{fe} is flexible and the NRR dimerizes weakly. (*A*) Schematic representation of the interaction and biophysical experiments on regions reported in panels (*B-G*). (*B-D*) Structural analysis of monomeric Notch1^{fe} from SEC-SAXS, including Dimensionless Kratky plot with crosshairs indicating the peak position for a globular protein (*B*), Guinier plot with a black line indicating the fit used to derive the R_g (*C*) and pair distance distribution function (*D*). (*E*) SPR equilibrium binding plot of Notch1$^{EGF33\text{-}NRR}$ to Notch1$^{EGF8\text{-}13}$. (*F*) SEC-MALS analysis of Notch1^{fe} shows a monomeric and monodisperse sample (thick lines indicate the molecular weight, left axis). Inset: Coomassie-stained SDS-PAGE of purified Notch1^{fe} in reducing conditions. (*G*) Coomassie-stained SDS-PAGE of purified Notch1^{fe} in non-reducing conditions. (*H*) SEC-MALS analysis of Notch1^{NRR} at three concentrations determined at elution shows a monomer-dimer equilibrium (thick lines indicate the molecular weight, left axis). Inset: Coomassie-stained SDS-PAGE of purified Notch1^{NRR} in reducing conditions, note that Notch1^{NRR} is processed at the S1 cleavage site into two fragments of 8 kDa and 27 kDa.

Jagged1^{fe} has a non-linear architecture and oligomerizes

Jagged1^{fe,HA} has a weak propensity to dimerize. Up to a concentration of 1.6 µM, Jagged1^{fe,HA} is a monomer with a molecular weight of 137 ± 0.2 kDa (Fig. 4 *A* and *B*) that correlates well with the theoretical molecular weight of 120-140 kDa depending on the glycosylation state



(55, 56). At higher concentrations, Jagged1^{fe} forms oligomers (Fig. 4 *C-E*). In sedimentation velocity analytical ultracentrifugation (SV-AUC), at 5 µM 19% of Jagged1^{fe,HA} consists of oligomers, and this increases to 31% at 20 µM (Fig. 4*C*). Concentration-dependent dimerization is also supported by batch SAXS analysis. At 5 µM the R_g of Jagged1^{fe,HA} is 81.2 ± 0.8 Å and this increases to 102 ± 0.4 Å at 42 µM (Table 2 and Fig. 4*D*) indicating more Jagged1^{fe,HA} dimers or larger oligomeric species are present at higher concentration.

< Fig. 4. Jagged1^{fe,HA} is not fully extended, flexible and oligomerizes weakly. (*A*) Schematic representation of the interactions and biophysical experiments on regions reported in panels (*B-J*). (*B*) SEC-MALS analysis of Jagged1^{fe,HA} at four concentrations determined at elution shows overlapping monomeric and monodisperse peaks (thick lines indicate the molecular weight, left axis). Inset: Coomassie-stained SDS-PAGE of the purified sample in reducing conditions. (*C*) SV-AUC analysis shows that Jagged1^{fe,HA} oligomerizes in a concentration-dependent manner. (*D-F*) SAXS analysis of Jagged1^{fe,HA} in batch and from monomeric SEC-SAXS fractions including Guinier plot with black lines indicating the fits used to derive the R_g (*D*), pair distance distribution function (*E*) and dimensionless Kratky plot with crosshairs indicating the peak position for a globular protein (*F*). (*G*) Overview of the detected distance constraints from the XL-MS experiments for monomeric Jagged1^{fe,HA}. (*H* and *I*) SPR equilibrium binding plots indicate interaction of Jagged1$^{EGF8-11,Fc}$ (*H*) and Jagged1^{CRD,Fc} (*I*) to Jagged1$^{C2-EGF3}$ (black) but not to Jagged1$^{EGF5-CRD}$ that acts as negative control (blue). The Fc domain does not interact with Jagged1$^{C2-EGF3,wt}$ as shown by the IgG control at 5 μM (red).

Table 2. Structural parameters derived from SAXS experiments. SAXS batch data I_0 have been normalized by the sample concentration to allow for comparison between samples. Non-normalized I_0 values are available on SASBDB under the accession codes defined in "Data and materials availability". n/a = not applicable.

	Concentration (μM)	R_g (Å) Guinier	sR_g range used in Guinier for R_g	R_g (Å) P(r)	D_{max} (Å)	I_0 (cm^{-1})
Notch1^{fe} SEC-SAXS	n/a	105 ± 0.2	0.62–1.26	113	380	0.047 ± 5.7x10^{-4}
Jagged1^{fe,HA} SEC-SAXS	n/a	74.1 ± 0.6	0.44–1.29	74.3	240	0.07 ± 4.4x10^{-4}
Jagged1^{fe,HA} Batch	42	102 ± 0.4	0.49–1.15	110	430	0.26 ± 5.8×10^{-4}
	21	96.4 ± 0.7	0.49–1.08	103	430	0.23 ± 8.7×10^{-4}
	11	89.2 ± 1.0	0.49–1.10	90.2	330	0.19 ± 1.1×10^{-3}
	5.3	81.2 ± 0.8	0.45–1.25	85.3	300	0.16 ± 1.2×10^{-3}
Jagged1$^{EGF8-11}$ Batch	230	31.7 ± 0.1	0.62–1.12	32.7	120	0.044 ± 6.4×10^{-5}
	115	31.3 ± 0.1	0.69–1.22	32.7	115	0.045 ± 8.4×10^{-5}
	58	31.5 ± 0.1	0.56–1.26	32.8	115	0.046 ± 1.0×10^{-4}
	29	32.7 ± 0.4	0.64–1.16	32.5	110	0.047 ± 3.0×10^{-4}
Jagged1^{CRD} Batch	167	24.1 ± 0.0	0.40–1.09	24.3	90	0.036 ± 2.6×10^{-5}
	83	23.3 ± 0.0	0.18–1.16	23.3	82	0.036 ± 3.0×10^{-5}
	42	22.6 ± 0.1	0.21–1.29	22.8	78	0.036 ± 4.4×10^{-5}
	21	22.7 ± 0.1	0.21–1.30	22.8	75	0.035 ± 7.5×10^{-5}

We used SEC-SAXS to separate monomeric Jagged1^{fe,HA} from oligomeric species. The region at the right side of the Jagged1^{fe,HA} elution peak, *i.e.* at larger retention volume, was selected for further analysis as this region most likely represents a monomeric fraction. Jagged1^{fe,HA} has a R_g of 74.1 ± 0.6 Å (Fig. 4*D*) and a D_{max} of 240 Å (Fig. 4*E*). The normalized Kratky plot indicates that structural flexibility is present in the Jagged1 ectodomain (Fig. 4*F*). SAXS analysis of smaller Jagged1 portions, Jagged1$^{EGF8-11}$ and Jagged1^{CRD} (*SI Appendix*, Fig. S7 *A-H*), show both samples do not change their oligomeric state at different concentrations (Table 2 and *SI Appendix*, Fig. S7 *B* and *F*). While Jagged1$^{EGF8-11}$ is flexible (*SI Appendix*, Fig. S7*D*), Jagged1^{CRD} is compact and globular (*SI Appendix*, Fig. S7*H*). The measured D_{max} of 240 Å indicates monomeric Jagged1^{fe,HA} is not extended, as a fully elongated Jagged1 ectodomain would have a maximum dimension of 585 Å (see Methods). In agreement with the SAXS data, the XL-MS analysis suggest that the extracellular region of Jagged1 is not

fully extended (Fig. 2 *A* and *B*). The detected distance restraints arise from either intra- or intermolecular Jagged1^{fe} interactions, as Jagged1^{fe} may be dimerizing in this experiment. To isolate the intramolecular cross-links from the ambiguous intra- and intermolecular cross-links we repeated the cross-linking experiment with Jagged1^{fe,HA} and separated monomeric Jagged1^{fe,HA} from cross-linked Jagged1^{fe,HA} oligomers by SEC (size-exclusion chromatography; *SI Appendix*, Fig. S2*B*) and analyzed the cross-links of both fractions by MS. The data indicate that four regions of the Jagged1 extracellular segment (C2-EGF2, EGF5-6, EFG9-12 and CRD) are in proximity within the same Jagged1^{fe,HA} molecule, as most identified cross-links are present in the monomeric (as well as in the oligomeric) fraction (Fig. 4*G* and *SI Appendix*, Fig. S2*C*). Most of these intramolecular cross-links are also found in the Notch1^{fe}-Jagged1^{fe,wt} and Notch1^{fe}-Jagged1^{fe,HA} XL-MS datasets, indicating that these intramolecular cross-links are independent of Notch1^{fe} binding.

We used SPR to verify that the regions identified by XL-MS interact directly. Constructs consisting of the Jagged1 regions C2-EGF3, EGF8-11 and CRD reveal direct interactions between Jagged1$^{C2\text{-}EGF3}$ and Jagged1$^{EGF8\text{-}11}$, and between Jagged1$^{C2\text{-}EGF3}$ and Jagged1^{CRD}, supporting the XL-MS results. The interactions are weak as covalent dimerization by Fc-fusion was required to measure binding. Fc-Jagged1$^{EGF8\text{-}11}$ and Fc-Jagged1^{CRD} bound to Jagged1$^{C2\text{-}EGF3,wt}$ with a $K_{D,app}$ of 0.34 μM and 0.93 μM, respectively (Fig. 4*H-I* and *SI Appendix*, Fig. S1 *K-L*). The C2-EGF3 region is required and sufficient for these interactions. Both Fc-Jagged1$^{EGF8\text{-}11}$ and Fc-Jagged1^{CRD} do not interact with Jagged1$^{EGF5\text{-}CRD}$ that is lacking the C2-EGF3 region (Fig. 4*H-I*) and affinities are similar for larger constructs that include the C2-EFG3 region, i.e. Jagged1$^{C2\text{-}EGF7}$, Jagged1$^{C2\text{-}EGF13}$ and Jagged1^{fe} (Table 1). In addition, the Jagged1 high-affinity mutations (11) do not affect this interaction (Table 1). Taken together, the SPR and XL-MS data indicate that the EGF8-11 and CRD regions interact intramolecularly with the C2-EGF3 region within the Jagged1 molecule.

Discussion

Two regions in Notch, EGF11-12 and NRR, have been widely studied due to their critical role in Notch signaling (10, 15, 19, 20, 48, 57) and represent the minimal requirements for ligand-dependent Notch activation (21, 58). Transcellular ligand binding at the Notch1 EGF8-12 site, positioned far away from the NRR in the primary sequence, and subsequent Notch1-ligand endocytosis generate a mechanical pulling force (9–14, 16) that could be transmitted via EGF13-36 to the NRR where it triggers a conformational change to expose the S2 site to proteolytic cleavage (19–21). Ligand binding in *cis* can inhibit Notch activation (29–31), while it was recently shown that it could also stimulate Notch activation (32), although it is not clear if and how endocytosis plays a direct role in this setting. These studies raise the question of how the different regions within Notch1 and Jagged1 interact.

Here we show that the Jagged1 C2-EGF3 segment is in close proximity to the Notch1 NRR in the Notch1^{fe}-Jagged1^{fe} complex, that Notch1$^{EGF8\text{-}13}$ and Notch1^{NRR} can interact directly with the C2-EGF3 region in Jagged1, and that Notch1$^{EGF8\text{-}13}$ interacts with Notch1$^{EGF33\text{-}NRR}$ (Fig. 5*A*). We confirm that the Notch1 ectodomain has regions of flexibility (33, 34, 42), which suggests that the EGF8-13 and the EGF33-NRR segments in Notch1 can interact intramolecularly. In addition to the importance of the canonical ligand binding site, EGF8-12, and the proteolytic activation site, NRR, in Notch, other regions have previously been proposed to play a role in Notch function (33–38). Intramolecular interactions have been determined between Notch EGF8-12 and EGF22-27 (33), and were suggested to occur for Notch EGF8-12 and EGF25-26 by demonstrating that antibodies targeting EGF25-26 prevent Jagged1-mediated full length Notch activation (34). In a deletion study, Notch EGF25-36 was shown to play a role in the interaction with Serrate (35). Specific regions on Notch, namely EGF24-26 (36), O-linked fucosylation on EGF26 (37), and O-fucose extension with GlcNAc on EGF6 and EGF36 (38) play a role in Jagged/Serrate-mediated signaling. Some of these studies highlight the importance of the membrane-proximal region of the Notch ectodomain, e.g. EGF25-36 (35) and EGF36 (38). These sites are next to or include the region we identify in Notch1-Jagged1 and Notch1-Notch1 interactions by XL-MS and quantitative binding assays (Figs. 2*A*,*B*,*E*, 3*E* and *SI Appendix*, Fig. S4*B*). The interaction of Jagged1 C2-EGF3 with the membrane-proximal Notch1 NRR fits well with the previously shown lipid-binding role of the Jagged1 C2 domain and the requirement of C2-lipid binding for optimal Notch activation (26, 28). In addition, the interactions of Jagged1^{CRD} with Jagged1$^{C2\text{-}EGF3}$ and with Notch1$^{EGF33\text{-}NRR}$ (Figs. 2*B*,*H* and 4*I*) support the finding that the CRD is involved in signaling (40). Collectively, our work and that of others indicate that several sites in the Notch and Jagged extracellular segments contribute to Notch-Jagged interactions and signaling.

The various segments have different interaction strengths. The interaction of the Notch1 ectodomain and that of Jagged1 is weak but strengthened by a pulling force (11). The mutation of five residues in the Jagged1 C2 domain increases the affinity of the Jagged1 ectodomain for the Notch1 ectodomain to 1 μM (Fig. 2*D*), indicating that the Jagged1 C2 domain plays an important role in the interaction with Notch1. Surprisingly, the measured interaction between Notch1$^{NRR\Delta loop}$ and Jagged1$^{C2\text{-}EGF3}$ also has a K_D of about 1 μM and is not dependent on the high-affinity mutations (Fig. 2*E*). While this interaction may be influenced in the SPR experiment by an avidity effect, arising from dimerization of the NRR, the interaction measured between Notch1^{NRR} and Jagged1$^{C2\text{-}EGF3,HA}$ in solution using MST also shows a K_D of around 1 μM (Fig. 2*F*). The interaction of the larger Notch1$^{EGF33\text{-}NRR}$ with Jagged1$^{C2\text{-}EGF3}$ shows a similar affinity with a K_D of 0.5 μM measured by MST (Fig. 2*F*), whereas it is 30-fold weaker in the surface-based SPR method (Table 1 and *SI Appendix*, Fig. S4*B*), which indicates that the context of this interaction may be important. Taken together, these data show that the NRR in the Notch1 ectodomain is in direct contact to

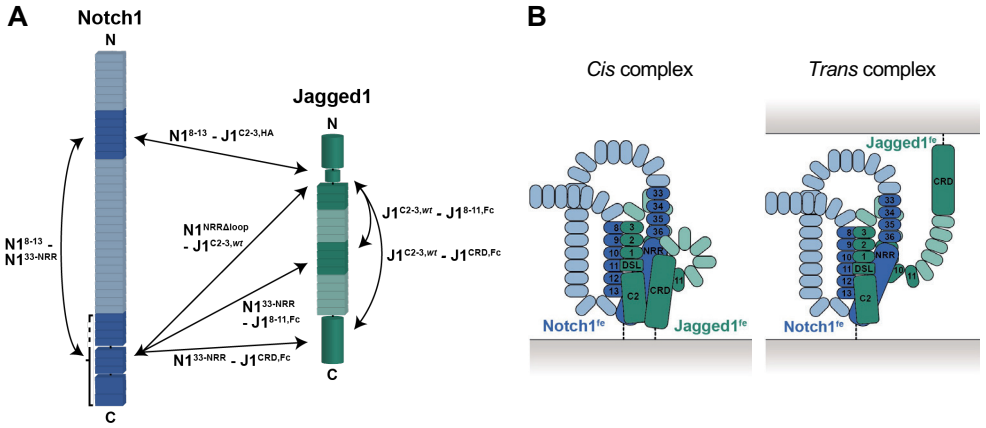


Fig. 5. Summary of the reported direct interactions and possible architectures of the complex. (*A*) Inter- and intra-molecular interactions based on the XL-MS and quantitative-interaction experiments are indicated by double arrows. (*B*) Schematic architectures of the Notch1-Jagged1 full ectodomain complex based on the interaction data shown in (*A*), represented in a *cis* or *trans* setting. Not all interactions might occur simultaneously, as reflected by the *trans* complex in which Jagged CRD is not contributing to interactions. The domains enabling backfolding have not been determined experimentally.

the Jagged1 C2-EGF3 region in the Notch1^{fe}-Jagged1^{fe} complex and suggest that ligand binding is directly coupled to Notch activation or regulation.

The setting at the cell surface or between two cells may dictate how Notch1 and Jagged1 interact. In our experiments we cannot discriminate between *cis* and *trans* interactions, and it may be possible we see both types of interactions simultaneously (Fig. 5*B*). For example, the interaction of the membrane proximal regions, *i.e.* Notch1 EGF33-NRR and Jagged1 CRD, seems more likely in a *cis* setting with both molecules expressed on the same cell. At the same time, the receptor and the ligand may undergo homomeric interactions on the cell surface which influences Notch signaling further (43, 53, 59–62). Besides the C2-EGF3 region, we have identified additional Jagged1 segments, namely EGF8-11 and CRD, that interact intermolecularly with Notch1 EGF33-NRR as well as intramolecularly with Jagged1 C2-EGF3 (Fig. 5*A*), and these regions could have a role in the clustering of Jagged1 and the Notch1-Jagged1 complex on, or between, cells. The interactions that we identify as intramolecular, i.e. Notch1 EGF8-13 with EGF33-NRR and Jagged1 C2-EGF3 with EFG8-11 and CRD may instead be used for intermolecular interactions when the proteins are expressed in a cell-surface setting. The role of the interactions in the function of Notch1 and Jagged1, whether they are intra- or intermolecular, occur in *cis* or in *trans*, and simultaneously or not, will need to be determined. In addition, it is currently not clear whether the Notch1 NRR–Jagged1 C2-EGF3 and Notch1 EGF8-13–NRR interactions are common features for the Notch and DSL family members. Interestingly, despite differences in domain composition, these three regions are present in all members, *i.e.* all Notch paralogs contain the EGF8-13 and

NRR segments and all DSL ligands have the C2-EGF3 region in common. Our data indicate that a mosaic of interaction sites is present, both on Notch1 and on Jagged1. Targeting these interactions may reveal their role in Notch signaling and could have potential for therapeutic applications to treat Notch-associated disorders.

Materials and Methods

Generation of constructs and mutagenesis

Notch1 and Jagged1 constructs were generated by polymerase chain reaction (PCR) using mouse Notch1 (Addgene 41728), human Notch1 (kind gift of Dr. Warren Pear, Univ. of Pennsylvania) and mouse Jagged1 (Image clone 6834418) as templates. All constructs are mouse version unless stated otherwise. Notch1^{fe} (residue numbers 19-1717) was subcloned in pUPE106.03 (U-Protein Express BV, cystatin secretion signal peptide, N-terminal His_6-tag), Notch1^{fe} (19-1728, human version), Notch1$^{EGF8-13}$ (294-526), Notch1$^{EGF22-27}$ (828-1058), Notch1$^{EGF33-36}$ (1267-1426), Notch1$^{EGF33-NRR}$ (1267-1717), Notch1^{NRR} (1446-1717) with and without its unstructured loop (1622-1659), Jagged1^{fe} (31-1067), Jagged1$^{C2-EGF3}$ (31-334), Jagged1$^{C2-EGF7}$ (31-485), Jagged1$^{C2-EGF13}$ (31-741), Jagged1$^{EGF5-13}$ (374-741), Jagged1$^{EGF5-CRD}$ (374-1067), Jagged1$^{EGF8-11}$ (487-665), Jagged1^{CRD} (857-1067) were subcloned in pUPE107.03 (U-Protein Express BV, cystatin secretion signal peptide, C-terminal His_6-tag). Jagged1 mutations (S32L, R68G, D72N, T87R, Q182R) based on Luca *et al*. (11) were introduced using Q5 Site-Directed Mutagenesis to generate Jagged1^{fe,HA} (31-1067) and Jagged1$^{C2-EGF3,HA}$ (31-334) constructs. In several figures, Notch1 and Jagged1 constructs are referred to as N1 and J1, respectively, and EGF repeats are referred to as their number, *i.e.* J1^{C2-3} for Jagged1$^{C2-EGF3}$.

Large-scale expression and purification

Constructs were transiently expressed in N-acetylglucoaminyltransferase I-deficient (GnTI-) Epstein-Barr virus nuclear antigen 1 (EBNA1)-expressing HEK293 cells growing in suspension (U-Protein Express BV), allowing for homogeneous N-glycosylations of the oligomannose type. With our open search approach (see Liquid Chromatography Mass Spectrometry and data analysis) we identified a core fucose modification (O-fucose) on four residues (T116, T194, T617, and T1362) (*SI Appendix*, Fig. S2*E* and Dataset S3). The medium was harvested six days after transfection, cells were spun down by 10 minutes of centrifugation at 1000x g, and cellular debris was spun down for 15 minutes at 4000x g. For human Notch1^{fe} used in the SEC-MALS experiment, the supernatant was concentrated fivefold and diafiltrated against 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 8.0, 500 mM NaCl and 2 mM $CaCl_2$ (IMAC A) using a Quixstand benchtop system (GE Healthcare) with a 10 kDa molecular weight cutoff (MWCO) membrane. Cellular debris were spun down for 10 min at 9500× g and the concentrate was filtered with a glass fiber prefilter (Minisart, Sartorius). Protein was purified by Nickel-nitrilotriacetic acid (Ni-NTA) affinity

chromatography and eluted with a mixture of 60% IMAC A and 40% of 25 mM HEPES pH 8.0, 500 mM NaCl, 500 mM imidazole, 2 mM $CaCl_2$ (IMAC B). For all other constructs and experiments, cells were spun down by 10 minutes of centrifugation at 1000x g, cellular debris was spun down for 15 minutes at 4000x g, and protein was directly purified by Ni Sepharose excel (GE Healthcare) affinity chromatography. Protein was eluted with a mixture of 60% of IMAC C (same as IMAC A, except pH 7.4) and 40% of IMAC D (same as IMAC B, except pH 7.4), or with 100% of IMAC D. SEC was performed on either a Superose6 10/300 increase (GE Healthcare) or a Superdex200 10/300 increase (GE Healthcare) equilibrated in SEC buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 2 mM $CaCl_2$). Protein purity was evaluated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie staining. Protein was concentrated and then stored at -80 °C.

Protein Cross-linking with PhoX

XL-MS was performed according to a previously optimized protocol (63). The optimal cross-linker concentration was established with SDS-PAGE. Cross-linking reactions were performed in triplicates with equimolar inputs of each protein for Notch1^{fe}-Jagged1^{fe,wt} and for Notch1^{fe}-Jagged1^{fe,HA}. 42 µL of protein solution, composed of the pre-incubated Notch1^{fe}-Jagged1^{fe,wt} or Notch1^{fe}-Jagged1^{fe,HA} complex at 5 µM in 20 mM HEPES pH 7.4, 150 mM NaCl and 2 mM $CaCl_2$, were mixed with 5 µL of the crosslinker solution composed of 10 mM PhoX in pure DMSO. Final concentrations of Ca^{2+} and PhoX during the XL-MS experiment were therefore 1.8 mM and 1.1 mM, respectively. The sample mixtures were filtered through MWCO 10 kDa filters (Vivaspin) into 10 mM Tris pH 7.5 in a 3:1 ratio (v:v) to a final volume of 25 µl. Prior to protein digestion, samples were deglycosylated overnight with Deglycosylation Mix II (NEBB), which predominantly targets N-linked glycans. After deglycosylation, urea was added to a final concentration of 8 M followed by addition of Tris(2-carboxyethyl)phosphine (TCEP) and chloroacetamide to a final concentration of 10 mM and 40 mM respectively. Samples were incubated at 37°C for 1 hour and then proteolytic digestion was performed with LysC (Wako) for 4 hours and trypsin (Promega) overnight. Resulting peptide mixtures were desalted with Oasis HLB plates (Waters), dried and stored at -80°C until further use.

Automated Fe(III)-IMAC-Based Enrichment

Cross-linked peptides were enriched with Fe(III)-NTA 5 µL in an automated fashion using the AssayMAP Bravo Platform (Agilent Technologies). Fe(III)-NTA cartridges were primed with 250 µL of 0.1% TFA in ACN and equilibrated with 250 µL of loading buffer (80% ACN/0.1% TFA). Samples were dissolved in 200 µL of loading buffer and loaded onto the cartridge. The columns were washed with 250 µL of loading buffer, and the cross-linked peptides were eluted with 25 µL of 10% ammonia directly into 25 µL of 10% formic acid. Samples were dried down and stored in 4 °C until subjected to LC-MS. For LC-MS analysis the samples were resuspended in 10% formic acid.

Liquid Chromatography Mass Spectrometry and data analysis

All mass spectrometry data was acquired using an UHPLC 1290 system (Agilent Technologies) coupled on-line to an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific). Peptides were trapped (Dr. Maisch Reprosil C_{18}, 3 µm, 2 cm × 100 µm) prior to separation on an analytical column (Agilent Poroshell EC-C_{18}, 2.7 µm, 50 cm × 75 µm). Trapping was performed by flushing in buffer A (0.1% v/v formic acid in water) for 10 min. Reversed phase separation was performed across a gradient of 10 % to 40 % buffer B (0.1% v/v formic acid in 80% v/v ACN) over 90 min at a flow-rate of approximately 300 nL/min. The instrument was operated in data-dependent MS^2 mode with MS^1 spectra recorded in the range 350-1400 Th and acquired in the Orbitrap at a resolution of 60,000 with an AGC of 4 x 10^5 and a maximum injection time of 50 ms. For MS^2, the cycle time was set to 3 s with charge state inclusion set to 3-8 for the enriched fraction and 2-8 for the flow-through. Dynamic exclusion was set to 12 s at 1.4 Th mass deviation. Stepped HCD was performed with the Ion Trap at NCE = 35 (+/- 10%) and acquired in the Orbitrap at a resolution of 30,000 with AGC set at 1 x 10^5 maximum injection time to 120 ms.

To quality control whether the O-linked glycans were successfully incorporated during protein expression, we performed an open search against the full sequence of Notch1. Identifications were filtered on whether they conform to the correct precursor mass offset for fucose, identify the peptide with high confidence, contain diagnostic ion(s) indicative for fucose, and match the expected sequence motif. This analysis does not exclude that other sites are modified by O-linked glycans as the data was not acquired in a mode geared towards glycan identification. The cross-linked peptides were analyzed with Thermo Proteome Discoverer (2.3.0.522) with incorporated XlinkX/PD nodes(63). The analysis was run with standard parameters in NonCleavable mode at 1 % False Discovery rate (FDR) at the level of the CSM and Cross-link tables against a manually created database with the target proteins and 200 random decoy entries. As fixed modification Carbamidomethyl (C) was set and as variable modification Oxidation (M), Acetyl (protein N-term), and Asn->Asp (N) (H_{-1} N_{-1} O). As cross-linking reagent PhoX (C_8 H_3 O_5 P) was set. Only cross-links detected in 2 out of 3 replicates were used for further analysis. The normal and mono-linked peptides were analyzed with MaxQuant (1.6.17.0)(64). The analysis was run with standard settings applied using the same database to search the spectra. As fixed modification Carbamidomethyl (C) was set and as variable modification Oxidation (M), Acetyl (protein N-term), PhoX Tris (K) (C_{12} H_{14} N O_8 P), PhoX H_2O (K) (C_8 H_5 O_6 P) and Asn->Asp (N) (H_{-1} N_{-1} O). Further downstream analysis and visual representation of the results was performed with the R scripting and statistical environment (65) using Circos (66) for data visualization.

Integrative modeling and docking of Notch1 NRR and Jagged1 C2-EGF3

To the crystal structure of Notch1 NRR described here (PDB: 7ABV), the missing flexible loop modelled with trRosetta (67) was added, *i.e.* residues 1622-1659. A structure of mouse

Jagged1 C2-EGF3 was generated by homology modelling in ITASSER (68) based on the rat high-affinity Jagged1 variant template (PDB: 5UK5) (11). Next, Notch1 NRR with the added loop and Jagged1 C2-EGF3 were docked together with three XL-MS based restraints from these regions and defined as 5-25 Å distance restraints in the HADDOCK2.4 webserver (69) (*SI Appendix*, Fig. S5). The loop was defined as fully flexible and the resulting outputs of the complex were examined in terms of scores with the emphasis on the biological relevance and restraints energy violations. UCSF ChimeraX (70) was used for visualization.

Surface plasmon resonance

SPR ligand constructs subcloned in-frame in pUPE107.62 (cystatin secretion signal peptide, C-terminal biotin acceptor peptide-tag followed by a C-terminal His_6-tag) were biotinylated in HEK293 cells by co-transfection with *E. coli* BirA biotin ligase with a sub-optimal secretion signal (in a pUPE5.02 vector), using a DNA ratio of 9:1 (sample:BirA, m/m). Additional sterile biotin (100 µL of 1 mg/mL HEPES-buffered biotin per 4 mL HEK293 culture) was supplemented to the medium. Protein was purified from the medium by Ni Sepharose excel (GE Healthcare) affinity chromatography. Purity was evaluated by SDS-PAGE and Coomassie staining. C-terminally biotinylated proteins were spotted on a P-STREP SensEye (Ssens) chip with a Continuous Flow Microspotter (CFM, Wasatch Microfluidics) using an 8x6 format. SEC buffer with 0.005% Tween-20 was used as a spotting buffer and the coupling was quenched using 1 mM biotin in SEC buffer. Proteins were therefore C-terminally coupled to the chip to ensure a native topology. Surface plasmon resonance experiments were performed on an MX96 SPRi instrument (IBIS Technologies). Analytes in SEC buffer were flowed over the sensor chip, and SEC buffer with 0.005% Tween-20 was used as a running buffer. Temperature was kept constant at 25 °C. The data was analyzed using SprintX (IBIS Technologies) and Prism (Graphpad) and modeled with a 1:1 Langmuir binding model to calculate the K_D and the maximum analyte binding (B_{max}). Since the NRR dimerizes, and bound with positive cooperativity to Jagged1$^{C2\text{-}EGF3}$ when it was used as an analyte, we fitted SPR equilibrium binding plots using a Hill equation with a Hill coefficient of 2. For the experiments in which full regeneration could not be achieved, the subsequent analyte injections were not zeroed in order to keep the B_{max} constant (*SI Appendix*, Fig. S1 *D,E,I,L*).

Microscale Thermophoresis

Jagged1$^{C2\text{-}EGF3,HA}$ in SEC buffer was labelled with NT-547 dye (NanoTemper Technologies) according to the manufacturer's instructions. Unlabelled Notch1$^{EGF33\text{-}NRR}$ and Notch1^{NRR} in SEC buffer were serially diluted from 50 µM to 3.0 nM (Notch1$^{EGF33\text{-}NRR}$) or 1.5 nM (Notch1^{NRR}) and incubated with 50 nM labelled Jagged1$^{C2\text{-}EGF3,HA}$ in the presence of 0.025% Tween-20 for 15 minutes at room temperature. Samples were transferred to Standard Treated Capillaries (NanoTemper Technologies) and run at 50 % excitation power on a Monolith NT.115 (NanoTemper Technologies) at a constant temperature of 25 °C. K_D was determined

according to the law of mass action using the program MO Affinity Analysis (NanoTemper Technologies) and results were plotted using Prism (Graphpad).

Small-angle X-ray scattering

Notch1^{fe} SEC-SAXS experiments were carried out at the European Synchrotron Radiation Facility (ESRF) beamline BM29. 500 µL of 8.1 µM human Notch1^{fe} were loaded on a Superose6 10/300 increase column (GE Healthcare) equilibrated in SEC buffer, via a high-performance liquid chromatography system (Shimadzu). A stable background signal was confirmed before measurement. Measurements were performed at room temperature at a flow rate of 0.5 mL/min. SAXS data was collected at a wavelength of 0.99 Å using a sample-to-detector (Pilatus 1M, Dectris) distance of 2.85 m. The scattering of pure water was used to calibrate the intensity to absolute units. 2000 frames of 2 s each were collected and data reduction was performed automatically using the EDNA pipeline (71). Frames with a stable R_g (± 10 %) and buffer frames were selected for further analysis using Chromixs (72). Data was analyzed in Primus (73) and Scatter (74), and results were plotted in Prism (Graphpad). The maximum dimension of 1027 Å for a theoretical elongated Notch1 ectodomain was calculated as follows: an average of 27 Å for the 36 EGF repeats (11) and 55 Å for the NRR (51).

Jagged1^{fe,HA} SEC-SAXS experiments were carried out at the Diamond Light Source (DLS) beamline B21 operating at an energy of 12.4 keV and using a sample-to-detector (Eigen 4M, Dectris) distance of 4.01 m. 45 µL of 42 µM Jagged1^{fe,HA} were loaded on a Superose6 3.2/300 increase (GE Healthcare) equilibrated in SEC buffer, via a HPLC system (Agilent). A stable background signal was confirmed before measurement. Measurements were performed at room temperature at a flow rate of 0.075 mL/min. The scattering of pure water was used to calibrate the intensity to absolute units. 620 frames of 3 s each were collected and data reduction was performed automatically using the DAWN pipeline (75). Frames with a stable R_g and buffer frames were selected for further analysis using Chromixs (72). Data was analyzed in Primus (73) and Scatter (74), and results were plotted in Prism (Graphpad).

Jagged1$^{EGF8\text{-}11}$, Jagged1^{CRD} and Jagged1^{fe} batch SAXS experiments were carried out the DLS beamline B21 operating at an energy of 12.4 keV and using a sample-to-detector (Eigen 4M, Dectris) distance of 4.01 m. The scattering of pure water was used to calibrate the intensity to absolute units. Data reduction was performed automatically using the DAWN pipeline (75). Frames were averaged after being manually inspected for radiation damage, the scattering of the SEC buffer was subtracted, and intensities were normalized by the concentration. Data was analyzed in Primus (73) and Scatter (74), and results were plotted in Prism (Graphpad). The maximum dimension of 585 Å for a theoretical elongated Jagged1 ectodomain was calculated as follows: 160 Å for the C2-EGF3 region as measured from its crystal structures (11, 26), an average of 27 Å for each of the remaining 13 EGF domains (11), and 75 Å as determined for the C-terminal CRD by SAXS (*SI Appendix*, Fig. S7*G*).

Multi-Angle Light Scattering

SEC-MALS was performed using a Superose6 10/300 increase (GE Healthcare) column for Notch1^{fe} (human version) or a Superdex 10/300 increase (GE Healthcare) column for Jagged1^{fe,HA} and Notch1^{NRR}, equilibrated in SEC buffer. For molecular weight characterization, light scattering was measured with a miniDAWN TREOS multi-angle light scattering detector (Wyatt Technology) connected to a RID-10A differential refractive index monitor (Shimadzu) for quantitation of the protein concentration. Chromatograms were collected, analyzed and processed on the ASTRA software suite (Wyatt Technology). A dn/dc of 0.1800 was calculated for Notch1^{fe} based on 6 N-glycosylation sites of the oligomannose type and 55 O-glycosylation sites (2 sugar moieties per site), 0.1814 for Jagged1^{fe,HA} based on 9 N-glycosylation sites and 16 O-glycosylation sites (4 O-glucosylation sites extended with 2 xylose moieties each, and 12 O-fucosylation sites), and 0.1828 for Notch1^{NRR} based on 2 N-glycosylation sites.

Crystallization and data collection

The Notch1 NRR was crystallized by sitting-drop vapour diffusion at 18 °C, by mixing 200 nL of protein solution containing a mixture of Notch1^{NRR} and Jagged1$^{C2\text{-}EGF3,HA}$ at 8.5 mg/mL in SEC buffer, and 100 nL of reservoir solution, composed of 2.0 M sodium chloride and 0.1 M sodium acetate pH 4.6. The protein solution was deglycosylated beforehand using EndoHf 1:100 (v/v) overnight at room temperature in SEC buffer. The crystal was harvested and flash-cooled in liquid nitrogen in the presence of reservoir solution supplemented with 25% glycerol. The dataset was collected at 100 K at the DLS beamline I03 (λ = 1.06998 Å).

Structure solution and refinement

The data was processed by the autoPROC pipeline (76) consisting of XDS (77), POINTLESS (78), AIMLESS (79), CCP4 (80) and STARANISO (81). The structure was solved by molecular replacement by searching for one copy of PDB ID 3ETO (51). After molecular replacement, the model was improved by manual model building in *Coot* (82) and refinement with REFMAC (83). Validation was performed using MolProbity (84).

Analytical ultracentrifugation

SV-AUC experiments were carried out in a Beckman Coulter Proteomelab XL-I analytical ultracentrifuge with An-60 Ti rotor (Beckman) at 40,000 revolutions per minute (r.p.m.). Jagged1^{fe,HA} at 5 µM and at 20 µM were measured in SEC buffer at 20 °C. Either 12 mm (5 µM sample) or 3 mm (20 µM sample) centerpieces with quartz windows were used. Absorbance was determined at 280 nm using SEC buffer as a reference. A total of 800 scans per cell were collected and analyzed in continuous c(s) mode in SEDFIT(85). Buffer density and viscosity were determined with SEDNTERP as 1.0061 g/mL and 0.010314 Pa·s, respectively.

Acknowledgments

We thank the staff of the DLS beamline B21, the staff of the ESRF beamline BM29, Dimphna H. Meijer and Mercedes Ramírez-Escudero for assistance with SAXS data collection, and the staff of the DLS beamline I03 for assistance with X-ray diffraction data collection. We thank Nadia Leloup for help with Notch1 SEC-MALS experiments and Dominique Thies-Weesie for help with SV-AUC experiments. This project has received funding from: the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme with grant agreement No. 677500 (to B.J.C.J.); the research programme TA with project number 741.018.201 (to R.A.S.), which is partly financed by the Dutch Research Council (NWO); and the European Union Horizon 2020 programme INFRAIA project Epic-XS project 823839 (to R.A.S.). This work benefited from access to the Amsterdam NKI, an Instruct-ERIC centre, with financial support provided by Instruct-ERIC (PID 10025). This work has been supported by iNEXT (PID 6764), funded by the Horizon 2020 programme of the European Union.

Data and materials availability

The mass spectrometry raw data, result/search files and the annotated spectra have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD023072. All SAXS data is made available at the Small Angle Scattering Biological Data Bank (SASBDB) with the accession codes SASDJG8 (Monomeric Notch1^{fe}), SASDJ38 (Monomeric Jagged1^{fe,HA}), SASDJ48 (5.3 μM Jagged1^{fe,HA}), SASDJ58 (11 μM Jagged1^{fe,HA}), SASDJ68 (21 μM Jagged1^{fe,HA}), SASDJ78 (42 μM Jagged1^{fe,HA}), SASDJ88 (29 μM Jagged1$^{EGF8-11}$), SASDJ98 (58 μM Jagged1$^{EGF8-11}$), SASDJA8 (115 μM Jagged1$^{EGF8-11}$), SASDJB8 (230 μM Jagged1$^{EGF8-11}$), SASDJC8 (21 μM Jagged1^{CRD}), SASDJD8 (42 μM Jagged1^{CRD}), SASDJE8 (83 μM Jagged1^{CRD}), SASDJF8 (167 μM Jagged1^{CRD}). Coordinates and structure factors for S1-cleaved mouse Notch1 NRR have been deposited to the Protein Data Bank (PDB) with accession code 7ABV.

Author Contributions

M.R.Z. and B.J.C.J. designed the experiments and interpreted all data; M.R.Z., J.P.M., M.J.K. and A.G. generated constructs and purified proteins; O.K. performed the MS experiments and data analysis; M.R.Z., J.P.M. and M.J.K performed the SPR experiments; M.R.Z. and M.J.K. performed the MST and crystallization experiments; M.R.Z. and J.P.M. performed the SAXS experiments; M.R.Z. performed the X-ray diffraction, SEC-MALS, and SV-AUC experiments; B.J.C.J. and R.A.S. supervised the project; M.R.Z. wrote the first draft of the manuscript (including figures), except for the XL-MS section; M.R.Z., O.K., R.A.S. and B.J.C.J. wrote the manuscript with input from all authors; B.J.C.J. conceived the project.

Competing Interest Statement

The authors declare no competing interests.

References

1. F. Radtke, H. R. MacDonald, F. Tacchini-Cottier, Regulation of innate and adaptive immunity by Notch. *Nat. Rev. Immunol.* **13**, 427–437 (2013).
2. A. Louvi, S. Artavanis-Tsakonas, Notch signalling in vertebrate neural development. *Nat. Rev. Neurosci.* **7**, 93–102 (2006).
3. J. L. Ables, J. J. Breunig, A. J. Eisch, P. Rakic, Not(ch) just development: Notch signalling in the adult brain. *Nat. Rev. Neurosci.* **12**, 269–83 (2011).
4. S. J. Bray, Notch signalling in context. *Nat. Rev. Mol. Cell Biol.* **9**, 722–735 (2016).
5. C. Siebel, U. Lendahl, Notch Signaling in Development, Tissue Homeostasis, and Disease. *Physiol. Rev.* **97**, 1235–1294 (2017).
6. A. P. Weng, *et al.*, Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science (80-.).* **306**, 269–271 (2004).
7. J. C. Aster, W. S. Pear, S. C. Blacklow, The Varied Roles of Notch in Cancer. *Annu. Rev. Pathol. Mech. Dis.* **12**, 245–275 (2017).
8. J. Mašek, E. R. Andersson, The developmental biology of genetic Notch disorders. *Development* **144**, 1743 LP – 1763 (2017).
9. F. Chowdhury, *et al.*, Defining single molecular forces required for notch activation using nano yoyo. *Nano Lett.* **16**, 3892–3897 (2016).
10. K. N. Lovendahl, S. C. Blacklow, W. R. Gordon, "The Molecular Mechanism of Notch Activation" in *Molecular Mechanisms of Notch Signaling*, T. Borggrefe, B. D. Giaimo, Eds. (Springer International Publishing, 2018), pp. 47–58.
11. V. C. Luca, *et al.*, Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity. *Science* **355**, 1320–1324 (2017).
12. L. Meloty-Kapella, B. Shergill, J. Kuon, E. Botvinick, G. Weinmaster, Notch Ligand Endocytosis Generates Mechanical Pulling Force Dependent on Dynamin, Epsins, and Actin. *Dev. Cell* **22**, 1299–1312 (2012).
13. D. Seo, *et al.*, A mechanogenetic toolkit for interrogating cell signaling in space and time. *Cell* **165**, 1507–1518 (2016).
14. X. Wang, T. Ha, Defining single molecular forces required to activate integrin and Notch signaling. *Science (80-.).* **340**, 991–994 (2013).
15. I. Rebay, *et al.*, Specific EGF repeats of Notch mediate interactions with Delta and serrate: Implications for notch as a multifunctional receptor. *Cell* **67**, 687–699 (1991).
16. W. R. Gordon, *et al.*, Mechanical Allostery: Evidence for a Force Requirement in the Proteolytic Activation of Notch. *Dev. Cell* **33**, 729–736 (2015).
17. J. T. Nichols, *et al.*, DSL ligand endocytosis physically dissociates Notch1 heterodimers before activating proteolysis can occur. *J. Cell Biol.* **176**, 445–458 (2007).
18. A. L. Parks, K. M. Klueg, J. R. Stout, M. A. T. Muskavitch, Ligand endocytosis drives receptor dissociation and activation in the Notch pathway. *Development* **127**, 1373–1385 (2000).

19. C. Brou, *et al.*, A novel proteolytic cleavage involved in Notch signaling: The role of the disintegrin-metalloprotease TACE. *Mol. Cell* **5**, 207–216 (2000).
20. J. S. Mumm, *et al.*, A ligand-induced extracellular cleavage regulates γ-secretase-like proteolytic activation of Notch1. *Mol. Cell* **5**, 197–206 (2000).
21. W. R. Gordon, *et al.*, Structural basis for autoinhibition of Notch. *Nat. Struct. Mol. Biol.* **14**, 295–300 (2007).
22. R. Kopan, M. X. G. Ilagan, The Canonical Notch Signaling Pathway: Unfolding the Activation Mechanism. *Cell* **137**, 216–233 (2009).
23. G. Yang, *et al.*, Structural basis of Notch recognition by human γ-secretase. *Nature* **565**, 192–197 (2019).
24. S. J. Bray, M. Gomez-Lamarca, Notch after cleavage. *Curr. Opin. Cell Biol.* **51**, 103–109 (2018).
25. K. Shimizu, *et al.*, Mouse Jagged1 physically interacts with Notch2 and other Notch receptors. Assessment by quantitative methods. *J. Biol. Chem.* **274**, 32961–32969 (1999).
26. C. R. Chillakuri, *et al.*, Structural Analysis Uncovers Lipid-Binding Properties of Notch Ligands. *Cell Rep.* **5**, 861–867 (2013).
27. J. Cordle, *et al.*, A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition. *Nat. Struct. Mol. Biol.* **15**, 849–857 (2008).
28. R. J. Suckling, *et al.*, Structural and functional dissection of the interplay between lipid and Notch binding by human Notch ligands. *EMBO J.* **36**, 2204–2215 (2017).
29. D. Del Álamo, H. Rouault, F. Schweisguth, Mechanism and significance of cis-inhibition in notch signalling. *Curr. Biol.* **21**, 40–47 (2011).
30. B. D'Souza, A. Miyamoto, G. Weinmaster, The many facets of Notch ligands. *Oncogene* **27**, 5148–5167 (2008).
31. D. Sprinzak, *et al.*, Cis-interactions between Notch and Delta generate mutually exclusive signalling states. *Nature* **465**, 86–90 (2010).
32. N. Nandagopal, L. A. Santat, M. B. Elowitz, Cis-activation in the Notch signaling pathway. *Elife* **8**, 1–34 (2019).
33. Z. Pei, N. E. Baker, Competition between Delta and the Abruptex domain of Notch. *BMC Dev. Biol.* **8**, 4 (2008).
34. A. Sharma, A. Rangarajan, R. R. Dighe, Antibodies against the extracellular domain of human Notch1 receptor reveal the critical role of epidermal-growth-factor-like repeats 25-26 in ligand binding and receptor activation. *Biochem. J.* **449**, 519–30 (2013).
35. A. Xu, L. Lei, K. D. Irvine, Regions of Drosophila notch that contribute to ligand binding and the modulatory influence of Fringe. *J. Biol. Chem.* **280**, 30158–30165 (2005).
36. N. Lawrence, T. Klein, K. Brennan, A. Martinez Arias, Structural requirements for notch signalling with delta and serrate during the development and patterning of the wing disc of Drosophila. *Development* **127** (2000).
37. R. Rampal, J. F. Arboleda-Velasquez, A. Nita-Lazar, K. S. Kosik, R. S. Haltiwanger, Highly conserved O-fucose sites have distinct effects on Notch1 function. *J. Biol. Chem.* **280**, 32133–32140 (2005).

38. S. Kakuda, R. S. Haltiwanger, Deciphering the Fringe-Mediated Notch Code: Identification of Activating and Inhibiting Sites Allowing Discrimination between Ligands. *Dev. Cell* **40**, 193–201 (2017).
39. B. C. Holdener, R. S. Haltiwanger, Protein O-fucosylation: structure and function. *Curr. Opin. Struct. Biol.* **56**, 78–86 (2019).
40. T. Kiyota, T. Kinoshita, Cysteine-rich region of X-Serrate-1 is required for activation of Notch signaling in Xenopus primary neurogenesis. *Int. J. Dev. Biol.* **46**, 1057–60 (2002).
41. V. C. Luca, *et al.*, Structural basis for Notch1 engagement of Delta-like 4. *Science (80-.).* **347**, 847–853 (2015).
42. P. C. Weisshuhn, *et al.*, Non-Linear and Flexible Regions of the Human Notch1 Extracellular Domain Revealed by High-Resolution Structural Studies. *Structure* **24**, 555–566 (2016).
43. D. F. Kelly, *et al.*, Molecular Structure and Dimeric Organization of the Notch Extracellular Domain as Revealed by Electron Microscopy. **5**, 1–6 (2010).
44. B. Steigenberger, R. J. Pieters, A. J. R. Heck, R. A. Scheltema, PhoX: An IMAC-Enrichable Cross-Linking Reagent. *ACS Cent. Sci.* **5**, 1514–1522 (2019).
45. O. Klykov, C. Van Der Zwaan, A. J. R. Heck, A. B. Meijer, R. A. Scheltema, Missing regions within the molecular architecture of human fibrin clots structurally resolved by XL-MS and integrative structural modeling. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 1976–1987 (2020).
46. J. Fürsch, K.-M. Kammer, S. G. Kreft, M. Beck, F. Stengel, Proteome-Wide Structural Probing of Low-Abundant Protein Interactions by Cross-Linking Mass Spectrometry. *Anal. Chem.* **92**, 4016–4022 (2020).
47. H. Takeuchi, R. S. Haltiwanger, Significance of glycosylation in Notch signaling. *Biochem. Biophys. Res. Commun.* **453**, 235–242 (2014).
48. F. Logeat, *et al.*, The Notch1 receptor is cleaved constitutively by a furin-like convertase. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8108–8112 (1998).
49. D. M. Czajkowsky, J. Hu, Z. Shao, R. J. Pleass, Fc-fusion proteins: New developments and future perspectives. *EMBO Mol. Med.* **4**, 1015–1028 (2012).
50. P. Taylor, *et al.*, Fringe-mediated extension of O-linked fucose in the ligand-binding region of Notch1 increases binding to mammalian Notch ligands. *Proc. Natl. Acad. U.S.A.* **111**, 7290–7295 (2014).
51. W. R. Gordon, *et al.*, Structure of the Notch1-negative regulatory region: implications for normal activation and pathogenic signaling in T-ALL. *Blood* **113**, 4381–90 (2009).
52. W. R. Gordon, *et al.*, Effects of S1 Cleavage on the Structure, Surface Export, and Signaling Activity of Human Notch1 and Notch2. *PLoS One* **4**, e6613 (2009).
53. X. Xu, *et al.*, Insights into Autoregulation of Notch3 from Structural and Functional Studies of Its Negative Regulatory Region. *Structure* **23**, 1227–1235 (2015).
54. Y. Wu, *et al.*, Therapeutic antibody targeting of individual Notch receptors. *Nature* **464**, 1052–1057 (2010).
55. A. Pandey, N. Niknejad, H. Jafar-Nejad, Multifaceted regulation of Notch signaling by glycosylation. *Glycobiology* (2020) https:/doi.org/10.1093/glycob/cwaa049.

56. S. M. Thakurdas, *et al.*, Jagged1 heterozygosity in mice results in a congenital cholangiopathy which is reversed by concomitant deletion of one copy of Poglut1 (Rumi). *Hepatology* **63**, 550–565 (2016).
57. P. A. Handford, B. Korona, R. Suckling, C. Redfield, S. M. Lea, "Structural Insights into Notch Receptor-Ligand Interactions" in *Molecular Mechanisms of Notch Signaling*, T. Borggrefe, B. D. Giaimo, Eds. (Springer International Publishing, 2018), pp. 33–46.
58. J. Cordle, *et al.*, Localization of the delta-like-1-binding site in human Notch-1 and its modulation by calcium affinity. *J. Biol. Chem.* **283**, 11785–11793 (2008).
59. M. Vooijs, E. H. Schroeter, Y. Pan, M. Blandford, R. Kopan, Ectodomain shedding and intramembrane cleavage of mammalian Notch proteins is not regulated through oligomerization. *J. Biol. Chem.* **279**, 50864–50873 (2004).
60. C. Hicks, *et al.*, A secreted Delta1-Fc fusion protein functions both as an activator and inhibitor of Notch1 signaling. *J. Neurosci. Res.* **68**, 655–667 (2002).
61. K. Shimizu, Integrity of intracellular domain of Notch ligand is indispensable for cleavage required for release of the Notch2 intracellular domain. *EMBO J.* **21**, 294–302 (2002).
62. K. Sakamoto, W. S. Chao, K. I. Katsube, A. Yamaguchi, Distinct roles of EGF repeats for the Notch signaling system. *Exp. Cell Res.* **302**, 281–291 (2005).
63. O. Klykov, *et al.*, Efficient and robust proteome-wide approaches for cross-linking mass spectrometry. *Nat. Protoc.*, 2964–2990 (2018).
64. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
65. R. Ihaka, R. Gentleman, R: A Language for Data Analysis and Graphics. *J. Comput. Graph. Stat.* **5**, 299–314 (1996).
66. M. Krzywinski, *et al.*, Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
67. J. Yang, *et al.*, Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1496–1503 (2020).
68. J. Yang, *et al.*, The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
69. G. C. P. van Zundert, *et al.*, The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428**, 720–725 (2016).
70. E. F. Pettersen, *et al.*, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* (2020) https:/doi.org/10.1002/pro.3943.
71. M.-F. Incardona, *et al.*, EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *J. Synchrotron Radiat.* **16**, 872–879 (2009).
72. A. Panjkovich, D. I. Svergun, CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944–1946 (2018).
73. P. V Konarev, V. V Volkov, A. V Sokolova, M. H. J. Koch, D. I. Svergun, PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).

74. S. Förster, L. Apostol, W. Bras, Scatter: software for the analysis of nano- and mesoscale small-angle scattering. *J. Appl. Crystallogr.* **43**, 639–646 (2010).
75. J. Filik, *et al.*, Processing two-dimensional X-ray diffraction and small-angle scattering data in DAWN 2. *J. Appl. Crystallogr.* **50**, 959–966 (2017).
76. C. Vonrhein, *et al.*, Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 293–302 (2011).
77. W. Kabsch, Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
78. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82 (2006).
79. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
80. M. D. Winn, *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).
81. I. J. Tickle, *et al.*, Staraniso. *Glob. Phasing Ltd., Cambridge, UK* (2018).
82. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
83. G. N. Murshudov, *et al.*, REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
84. V. B. Chen, *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).
85. P. Schuck, Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).

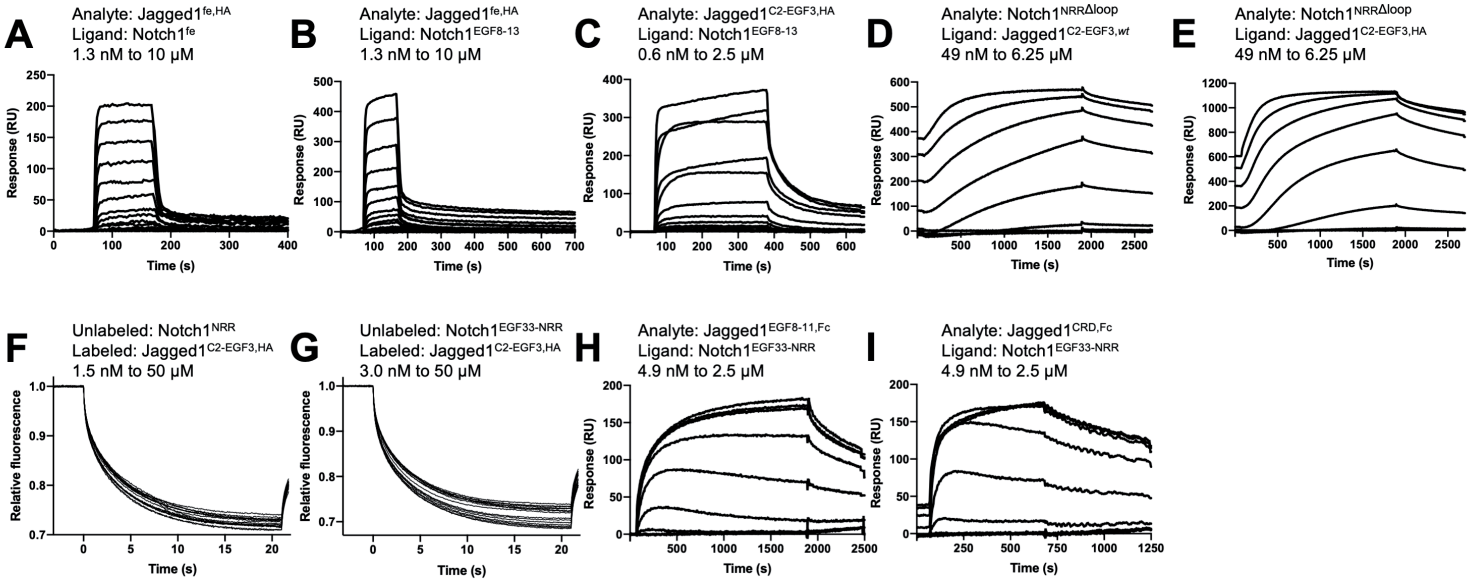
Supplementary Information for

Notch-Jagged signaling complex defined by an interaction mosaic

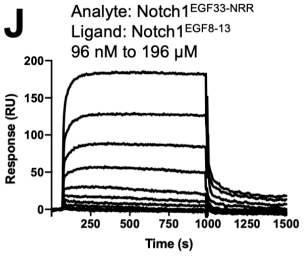
Matthieu R. Zeronian, Oleg Klykov, Júlia Portell i de Montserrat, Maria J. Konijnenberg, Anamika Gaur, Richard A. Scheltema, Bert J. C. Janssen

Email: r.a.scheltema@uu.nl (R.A.S.), twitter@xlinkx_pd; b.j.c.janssen@uu.nl (B.J.C.J.).

Sensorgrams and MST traces related to Fig. 2



Sensorgram related to Fig. 3



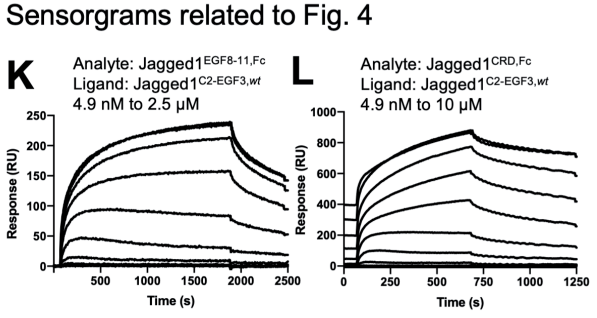
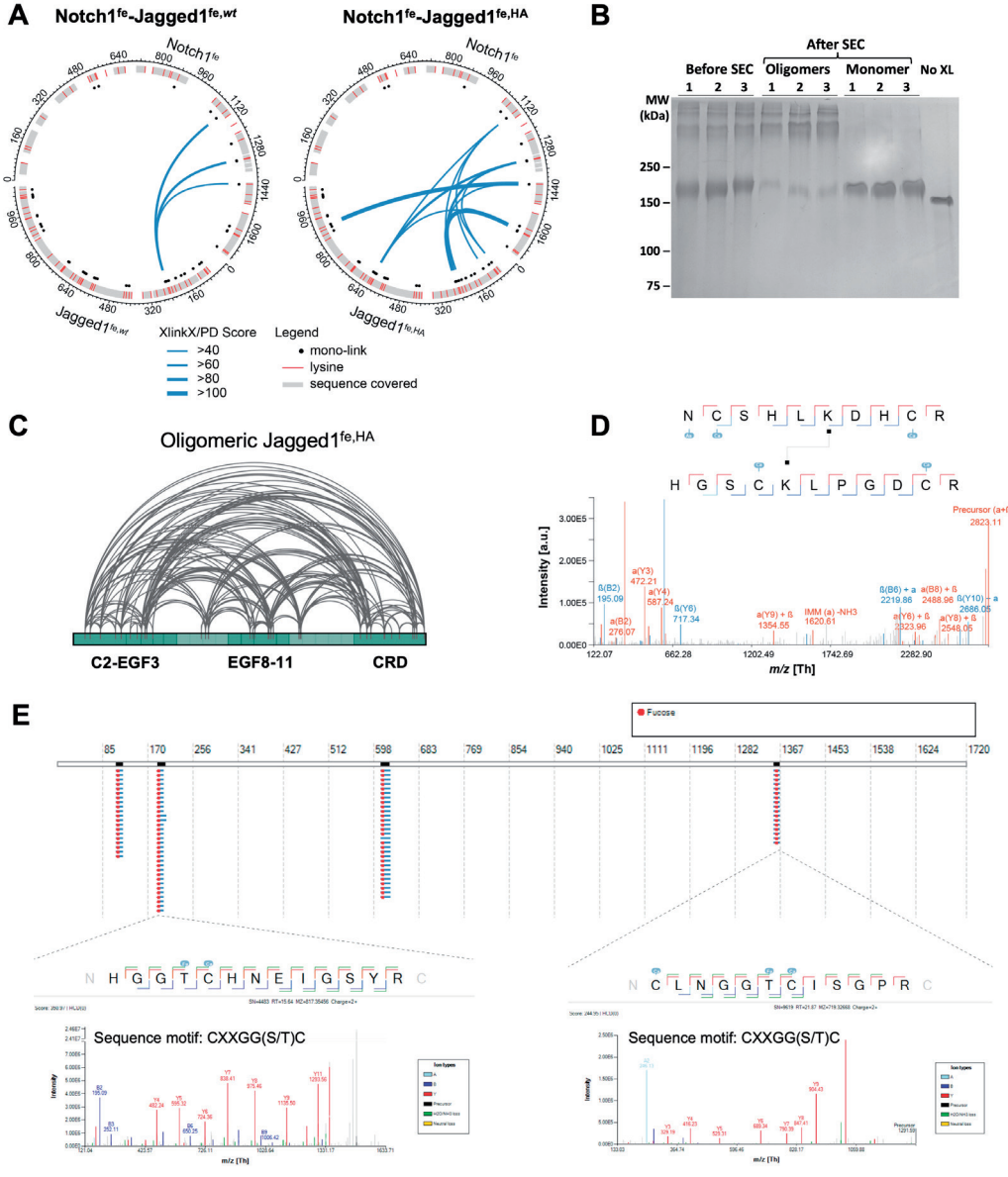


Fig. S1. Sensorgrams and MST traces. (*A-I*) Sensorgrams and MST traces related to Fig. 2, with sensorgrams of Jagged1^{fe,HA} binding to Notch1^{fe} (*A*) and to Notch1$^{EGF8-13}$ (*B*), Jagged1$^{C2-EGF3,HA}$ binding to Notch1$^{EGF8-13}$ (*C*), Notch1$^{NRR\Delta loop}$ binding to Jagged1$^{C2-EGF3,wt}$ (*D*) and to Jagged1$^{C2-EGF3,HA}$ (*E*), MST traces of Notch1^{NRR} binding to Jagged1$^{C2-EGF3,HA}$ (*F*) and Notch1$^{EGF33-NRR}$ binding to Jagged1$^{C2-EGF3,HA}$ (*G*), sensorgrams of Jagged1$^{EGF8-11,Fc}$ binding to Notch1$^{EGF33-NRR}$ (*H*) and Jagged1^{CRD,Fc} binding to Notch1$^{EGF33-NRR}$ (*I*). (*J*) Sensorgram related to Fig. 3, with Notch1$^{EGF33-NRR}$ binding to Notch1$^{EGF8-13}$. (*K* and *L*) Sensorgrams related to Fig. 4, with Jagged1$^{EGF8-11,Fc}$ binding to Jagged1$^{C2-EGF3,wt}$ (*K*) and Jagged1^{CRD,Fc} binding to Jagged1$^{C2-EGF3,wt}$ (*L*). The concentration range used in the experiment is indicated in all panels.


A
Notch1fe-Jagged1fe,wt
Notch1fe-Jagged1fe,HA
Notch1fe
Jagged1fe,wt
Jagged1fe,HA
XlinkX/PD Score
>40
>60
>80
>100
Legend
mono-link
lysine
sequence covered
B
After SEC
Before SEC
Oligomers
Monomer
No XL
1 2 3 1 2 3 1 2 3
MW (kDa)
250
150
100
75
C
Oligomeric Jagged1fe,HA
C2-EGF3
EGF8-11
CRD
D
N C S H L K D H C R
H G S C K L P G D C R
Precursor (a+ß) 2823.11
Intensity [a.u.]
3.00E5
2.00E5
1.00E5
0.00E0
122.07
662.28
1202.49
1742.69
2282.90
m/z [Th]
E
Fucose
85 170 256 341 427 512 598 683 769 854 940 1025 1111 1196 1282 1367 1453 1538 1624 1720
N H G G T C H N E I G S Y R C
N C L N G G T C I S G P R C
Sequence motif: CXXGG(S/T)C
Sequence motif: CXXGG(S/T)C

< Fig. S2. Additional information related to cross-linking mass-spectrometry experiments. (*A*) Circular plots indicating the inter-links, by XlinkX/Proteome Discoverer score, and mono-links identified in the cross-linking experiment of Notch1^{fe}-Jagged1^{fe,wt} (left) and Notch1^{fe}-Jagged1^{fe,HA} (right). The sequence covered in the peptide identification is indicated. (*B*) Coomassie-stained SDS-PAGE showing the cross-linked oligomeric and monomeric Jagged1 fractions purified by size exclusion chromatography in triplicate. The monomer fractions are well separated from the oligomeric fractions. (*C*) Overview of the detected distance constraints from the XL-MS experiments for oligomeric Jagged1^{fe,HA}. The detected distance constraints for monomeric Jagged1^{fe,HA} are shown in Fig. 4*G*. (*D*) Example mass spectrum of an identified cross-link. (*E*) Example of several identified core O-fucose residues on Notch1 shows that are our Notch1 protein is fucosylated. Positive identifications were made by an open peptide search and filtered based on the precursor mass difference, presence of glycan diagnostic ions, and conformance to the expected sequence motif. The blue stretches indicate where peptides were detected carrying a core fucose, with the red dot showing the precise position. The insets show representative spectra where the peptide was identified with high confidence, with a precursor mass difference indicative of a core fucose. A list of the identifications can be found in Dataset S3.

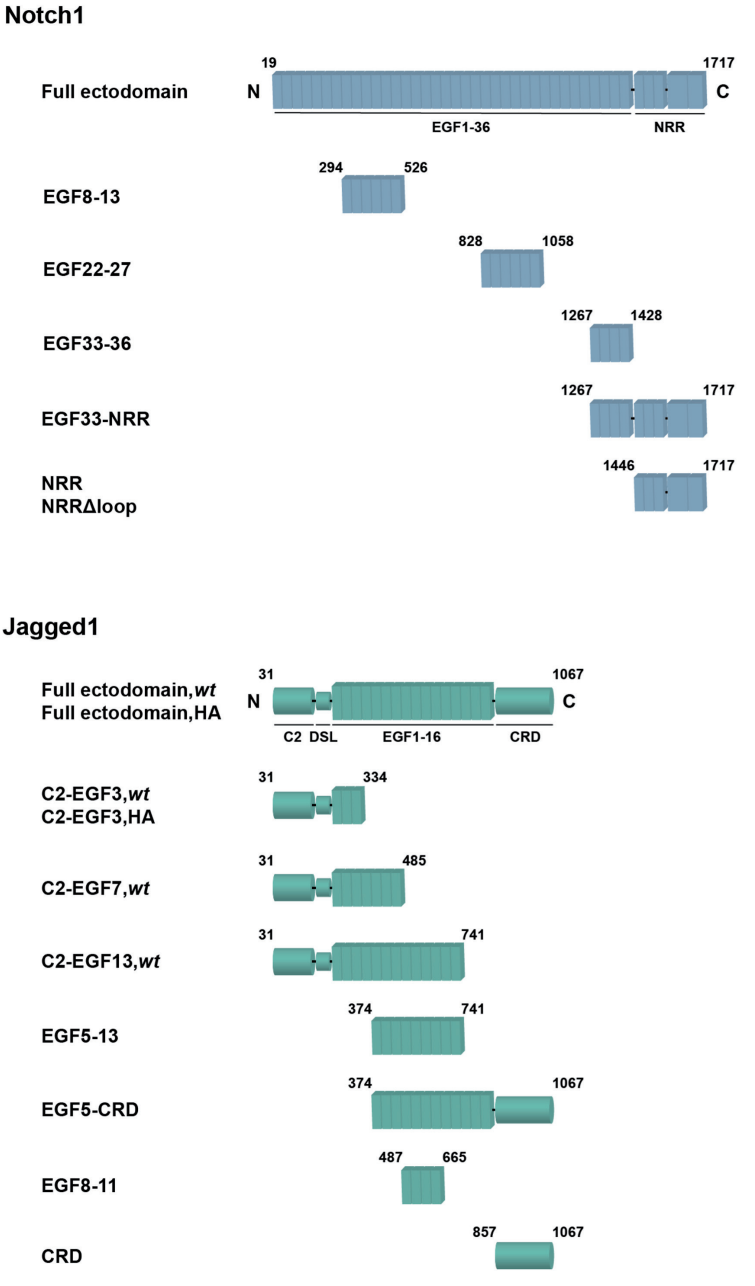


Fig. S3. Domain composition and main constructs generated.

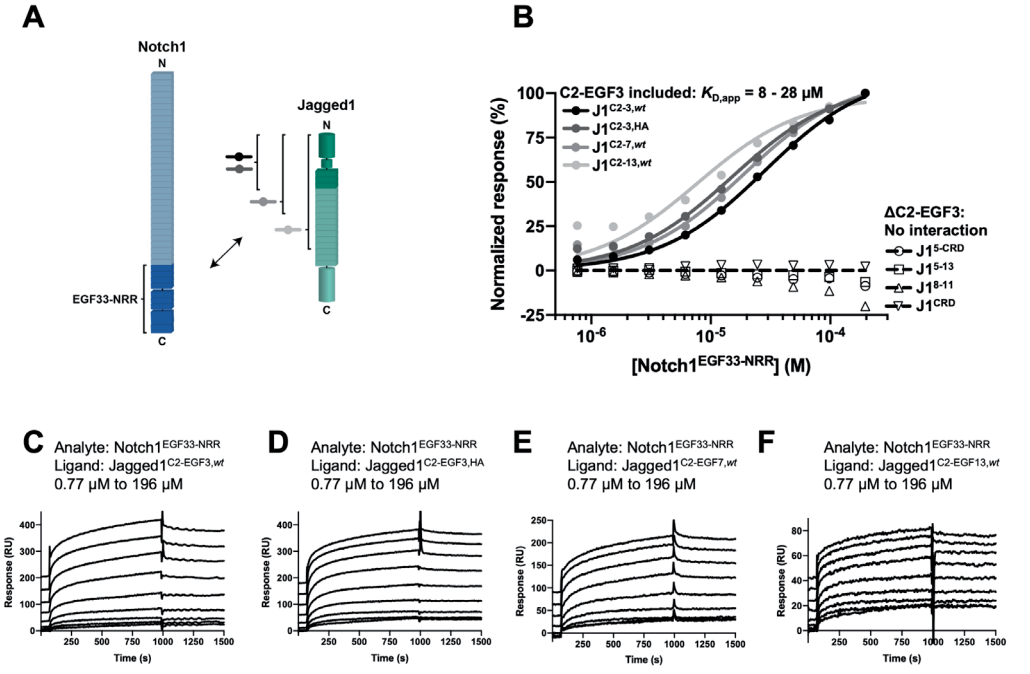


Fig. S4. The C2-EGF3 domain of Jagged1 is necessary and sufficient for Notch1 EGF33-NRR interaction. (*A*) Schematic representation of the interactions reported in panels (*B-F*). (*B*) SPR equilibrium binding plots of Notch1$^{EGF33\text{-}NRR}$ to Jagged1$^{C2\text{-}EGF3,wt}$ (black), Jagged1$^{C2\text{-}EGF3,HA}$ (dark grey), Jagged1$^{C2\text{-}EGF7,wt}$ (grey), Jagged1$^{C2\text{-}EGF13,wt}$ (light grey), Jagged1$^{EGF5\text{-}CRD}$ (open circle), Jagged1$^{EGF5\text{-}13}$ (open square), Jagged1$^{EGF8\text{-}11}$ (open triangle) and Jagged1^{CRD} (open inverted triangle). (*C-F*) Corresponding SPR sensorgrams, with Notch1$^{EGF33\text{-}NRR}$ binding to Jagged1$^{C2\text{-}EGF3,wt}$ (*C*), to Jagged1$^{C2\text{-}EGF3,HA}$ (*D*), to Jagged1$^{C2\text{-}EGF7,wt}$ (*E*), and to Jagged1$^{C2\text{-}EGF13,wt}$ (*F*).

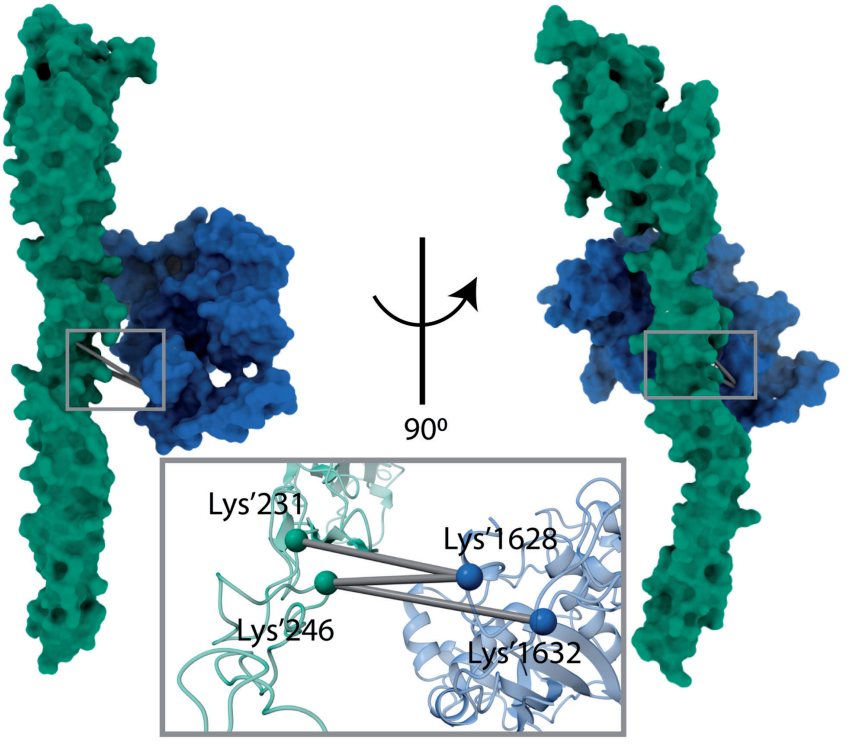
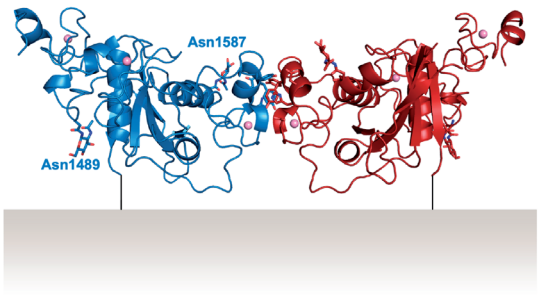
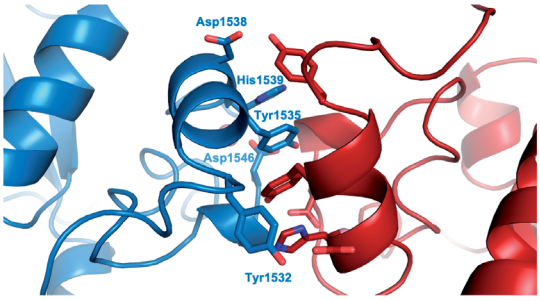


Fig. S5. Exploded model of the Notch1 NRR–Jagged1 C2-EGF3 complex. Docking of the Notch1 NRR–Jagged1 C2-EGF3 complex using the structure of Notch1 NRR described here (blue) and that of Jagged1 C2-EGF3 (green; PDB: 5UK5) and based on cross-links obtained by XL-MS. The two structures are slightly separated from each other to indicate the cross-links.

A



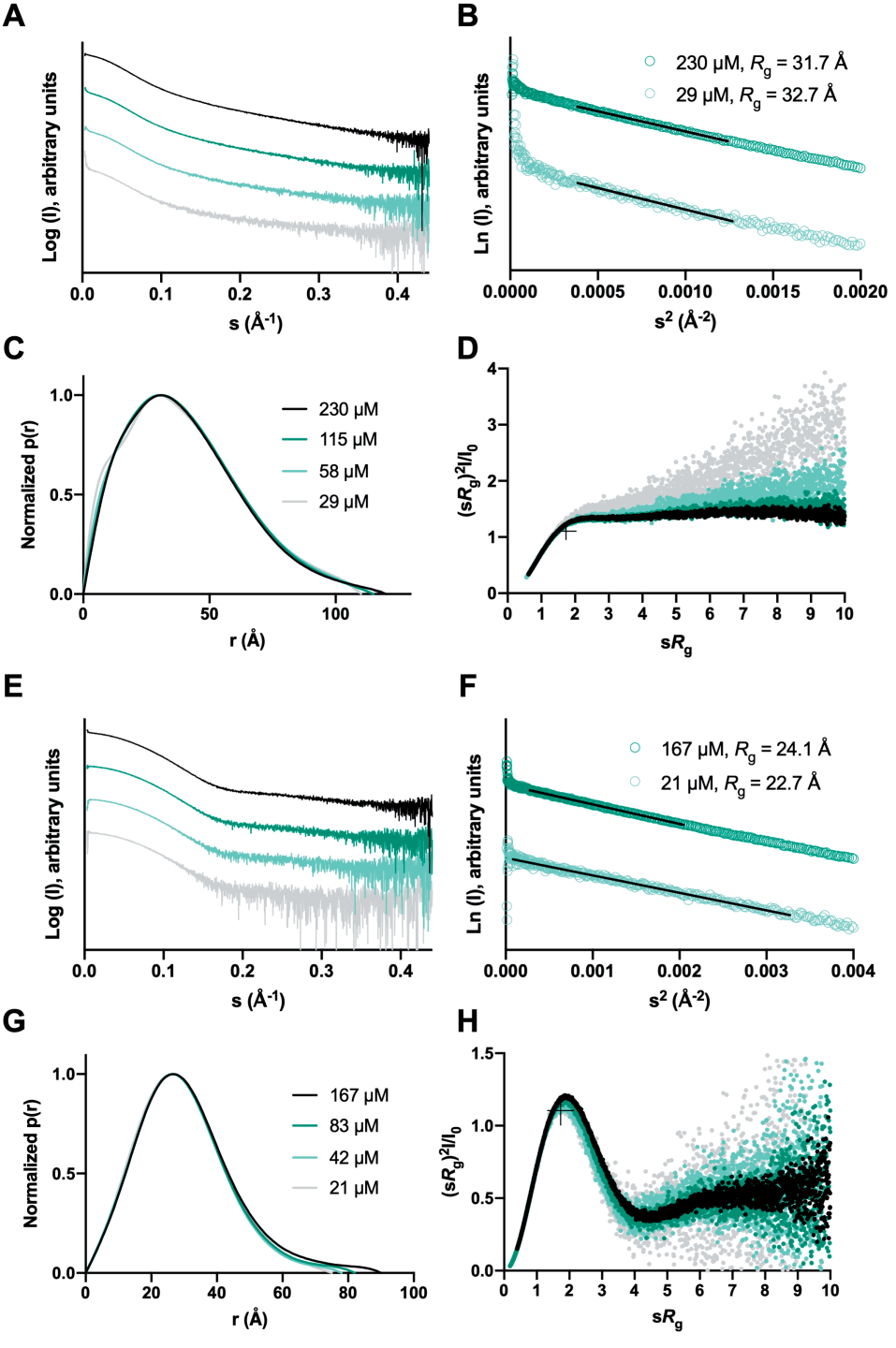
B



C

S1-cleaved mouse Notch1 NRR	
Data collection	
Space group	$P\,4_2\,2_1\,2$
Cell dimensions	
a, *b*, *c* (Å)	65.72, 65.72, 161.68
α, *β*, *γ* (°)	90, 90, 90
Resolution (Å)	60.89 - 2.06 (2.30 - 2.06)
No. observed reflections	195430 (8934)
No. unique reflections	15144 (757)
R_{merge}	0.105 (1.431)
Mean I/σI	14.6 (1.7)
$CC_{1/2}$	0.999 (0.770)
Completeness (spherical, %)	66.9 (12.6)
Completeness (ellipsoidal, %)	91.5 (64.6)
Ellipsoidal resolution limits (Å) [direction]	2.44 [a*] 2.44 [b*] 2.06 [c*]
Redundancy	12.9 (11.8)
Refinement	
Resolution (Å)	60.89 - 2.06
R_{work}/R_{free} (%)	19.53 / 23.39
Average *B*-factors (Å^2)	
Protein	57.7
Glycans/ions	94.5
Waters	46.6
R.M.S. deviations	
Bond lengths (Å)	0.0084
Bond angles (°)	1.52
Ramachandran (%)	
Favored	95.61
Allowed	4.39
Outliers	0
Molprobity score	2.06

Fig. S6. Structure of the S1-cleaved mouse Notch1 NRR. (*A*) Proposed orientation of the Notch1 NRR dimer with respect to the cell surface. (*B*) Key residues at the dimerization interface are indicated. (*C*) Data collection and refinement statistics. Highest resolution shell in parentheses.


A
Log (I), arbitrary units
s (Å-1)
B
Ln (I), arbitrary units
230 µM, Rg = 31.7 Å
29 µM, Rg = 32.7 Å
s2 (Å-2)
C
Normalized p(r)
230 µM
115 µM
58 µM
29 µM
r (Å)
D
(sRg)2I/I0
sRg
E
Log (I), arbitrary units
s (Å-1)
F
Ln (I), arbitrary units
167 µM, Rg = 24.1 Å
21 µM, Rg = 22.7 Å
s2 (Å-2)
G
Normalized p(r)
167 µM
83 µM
42 µM
21 µM
r (Å)
H
(sRg)2I/I0
sRg

< Fig. S7. Jagged1$^{EGF8-11}$ and Jagged1^{CRD} have distinct structural properties. (*A-D*) Structural analysis of Jagged1$^{EGF8-11}$ from batch SAXS, including Log (I) versus s plot (*A*), Guinier plot with black lines indicating the fits used to derive the R_g (*B*), pair distance distribution function (*C*) and dimensionless Kratky plot with crosshairs indicating the peak position for a globular protein (*D*). (*E-H*) Structural analysis of Jagged1^{CRD} from batch SAXS, including Log (I) versus s plot (*E*), Guinier plot with black lines indicating the fits used to derive the R_g (*F*), pair distance distribution function (*G*) and dimensionless Kratky plot with crosshairs indicating the peak position for a globular protein (*H*).

Table S1. Description of the files uploaded to the PRIDE repository.

Filename	Description
20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_MT1.raw 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_MT2.raw 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_MT3.raw	Triplicate measurements of the crosslinked and PhoX enriched fraction for the mouse Notch1 - mouse high-affinity mutant Jagged1 full ectodomain complex.
20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_MT_Asn-_AspSites.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_MT_evidence.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_MT_peptides.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_MT_PhoX H2OSites.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_MT_PhoX TrisSites.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_MT.pdf	MaxQuant output tables and annotated spectra.
20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_MT_Crosslinks.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_MT_CSMs.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_MT_CSMs.pdf	Proteome Discoverer XlinkX/PD output tables and annotated spectra.
20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_WT1.raw 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_WT2.raw 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_WT3.raw	Triplicate measurements of the crosslinked and PhoX enriched fraction for the mouse Notch1 - mouse wild-type Jagged1 full ectodomain complex.
20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_WT_Asn-_AspSites.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_WT_evidence.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_WT_peptides.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_WT_PhoX H2OSites.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_WT_PhoX TrisSites.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_WT.pdf	MaxQuant output tables and annotated spectra.
20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_WT_Crosslinks.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_WT_CSMs.txt 20190316_L1_Ag6_Klyko001_SA_NOTCH_PhoX_WT_CSMs.pdf	Proteome Discoverer XlinkX/PD output tables and annotated spectra.
20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_MT1.raw 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_MT2.raw 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_MT3.raw	Triplicate measurements of the flow-through (i.e. not cross-linked & mono-linked) fraction for the mouse Notch1 - mouse high-affinity mutant Jagged1 full ectodomain complex.
20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_MT_Asn-_AspSites.txt 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_MT_evidence.txt 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_MT_peptides.txt 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_MT_PhoX H2OSites.txt 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_MT_PhoX TrisSites.txt 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_MT.pdf	MaxQuant output tables and annotated spectra.
20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_WT1.raw 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_WT2.raw 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_WT3.raw	Triplicate measurements of the flow-through (i.e. not cross-linked & mono-linked) fraction for the mouse Notch1 - mouse wild-type Jagged1 full ectodomain complex.
20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_WT_Asn-_AspSites.txt 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_WT_evidence.txt 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_WT_peptides.txt 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_WT_PhoX H2OSites.txt 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_WT_PhoX TrisSites.txt 20190923_L1_Ag1_Klyko001_SA_JagNotch_PhoX_FT_WT.pdf	MaxQuant output tables and annotated spectra.

Dataset S1 (separate file). This annotated excel file contains the Crosslinks table of XlinkX/PD, broken up in inter- and intra-links. The original Crosslinks and CSM tables can be found in the PRIDE repository.

Dataset S2 (separate file). This annotated excel file contains the Site specific tables of MaxQuant for the PhoX:Tris and PhoX;H2O monolinks. The original tables plus the Evidence and Peptide tables can be found in the PRIDE repository.

Dataset S3 (separate file). Fucosylated peptides from Notch1, identified by open search. The position numbering is according to Uniprot.

Chapter 3

Structural insights into the non-inhibitory mechanism of the anti-EGFR EgB4 nanobody

Matthieu R. Zeronian,[1] Sofia Doulkeridou,[2] Paul M.P. van Bergen en Henegouwen[2] and Bert J.C. Janssen[1]*

[1]Structural Biochemistry, Bijvoet Center for Biomolecular Research, Department of Chemistry, Faculty of Science, Utrecht University, Utrecht, The Netherlands

[2]Department of Biology, Cell Biology, Neurobiology and Biophysics, Faculty of Science, Utrecht University, Utrecht, The Netherlands

*Corresponding author

Email: b.j.c.janssen@uu.nl (B.J.C.J.)

This chapter has been submitted for publication.

Abstract

The epidermal growth factor receptor (EGFR) is involved in various developmental processes, and alterations of its extracellular segment are associated with several types of cancers, in particular glioblastoma multiforme (GBM). The EGFR extracellular region is therefore a primary target for therapeutic agents, such as monoclonal antibodies and variable domains of heavy chain antibodies (VHH), also called nanobodies. Nanobodies have been previously shown to bind to EGFR, and to inhibit ligand-mediated EGFR activation. Here we present the X-ray crystal structures of the EgB4 nanobody, alone and bound to the full extracellular EGFR-EGF complex in its active conformation. We show that EgB4 binds to a new epitope located on EGFR domains I and II, and we describe the molecular mechanism by which EgB4 plays a non-inhibitory role in EGFR signaling. This work provides the structural basis for the application of EgB4 as a biomarker to locate EGFR-associated tumors, while not affecting EGFR activation.

Introduction

The human epidermal growth factor receptor (HER) tyrosine kinase family is essential to cellular growth, migration and differentiation, and is involved in a variety of cancers (1–5). Members of this family include EGFR, HER2, HER3 and HER4. Except for HER2, all members have been shown to bind to specific ligands (1), e.g. EGFR binds to epidermal growth factor (EGF) and transforming growth factor-α (TGF-α). EGFR was the first family member shown to be overexpressed in cancers (6) and it is therefore a primary target for anti-cancer therapies (7, 8). New tools may help to further characterize the role of EGFR in health and disease.

EGFR is a 170 kDa type I transmembrane receptor composed of an extracellular region characterized by four domains (I, II, III and IV), a transmembrane region, and an intracellular region composed of a kinase domain and a C-terminal tail. In the EGFR ectodomain, the leucin-rich domains I and III are related to one another and to similar domains in the insulin receptor, while domains II and IV are enriched in cysteine residues and share similarities with laminins and furin-like proteases (9). Ligand binding to the EGFR ectodomain is coupled to homodimerization (10), followed by a conformational rearrangement of the transmembrane region and asymmetric dimerization of the intracellular domains, one of which phosphorylates the other to initiate signaling (11–15). The EGFR ectodomain exists in a tethered, auto-inhibited conformation, in which the domain II dimerization arm interacts with domain IV (16). In the active, also called extended, conformation, domain II rotates 130° around domain III, therefore breaking the domain II – IV tether, creating a ligand binding pocket shared between domains I and III, and exposing the dimerization arm for intermolecular interaction (10, 17).

HER family members are expressed in all cell types and are critical to the embryogenesis of vertebrates (18). In EGFR null mice, lethality was shown to be due to abnormalities in several organs including brain, lung, skin and gastrointestinal tract, and the renewal of stem cells (19, 20). EGFR signaling remains also active in the mature central nervous system (21). Besides its critical role in development and homeostasis, EGFR is involved in the initiation and maintenance of several types of solid tumors. Notably, the epidermal growth factor variant III (EGFRvIII) is found in ~40 % of high-grade gliomas (22). The EGFRvIII ectodomain is characterized by the deletion of a stretch of 267 residues in domains I and II, addition of a glycine residue and a free cysteine residue, altogether leading to increased homodimerization, impaired downregulation, and aberrant tyrosine kinase activity (23, 24). EGFRvIII drives cancer proliferation through multiple mechanisms, although it preferentially activates the phosphatidylinositol 3-kinase (PI3K)/Akt signal transduction pathway, and it is involved in several types of cancers, including GBM, breast and lung cancer (25). The epidermal growth factor variant II (EGFRvII), characterized by the deletion of 83 residues in the membrane-proximal region of domain IV, is also oncogenic (26, 27). Other EGFR

alterations include mutations in the kinase domain that are involved in non-small-cell lung cancers (NSCLCs), especially adenocarcinoma (28–30).

To treat EGFR-associated cancers, monoclonal antibodies are increasingly used but their large size (~150 kDa) leads to reduced tumor penetration and slow distribution (31–33). Camelidae heavy-chain antibodies, discovered in 1993 (34), are composed of a homodimer of heavy chains that retains full binding capacity despite the lack of light chains. The VHH domain of heavy-chain antibodies, also referred to as nanobody in its isolated form, is the domain responsible for antigen binding, and constitutes the smallest (~15 kDa) antigen-binding unit derived from natural sources (34). Due to their small size and potential to bind to epitopes with a high affinity, nanobodies represent a valuable tool in cancer diagnostics and therapy (35, 36). Although the use of nanobodies in research is fairly recent, nanobody-based cancer therapies are currently assessed in clinical trials (36), and in 2019 a nanobody was approved for therapeutic use for the first time (37). Nanobodies that bind to EGFR with a nanomolar affinity were produced for diagnostic and therapeutic applications (38–42), and structures of three inhibitory nanobodies (EgA1, 9G8 and 7D12) were solved in complex with the EGFR ectodomain in its inactive conformation (43). All three nanobodies bind to domain III. The EgA1 and 9G8 nanobodies bind to a cleft formed between domains II and III, whereas the 7D12 interaction surface overlaps with the ligand binding site. These nanobodies prevent EGFR from adopting an extended conformation that is required for ligand-mediated receptor activation. The EgB4 nanobody was proposed to bind to EGFR domain I while not competing with EGF binding (39, 44), but no structural information is available on EgB4 or its interaction with EGFR. Here we report crystal structures of the EgB4 nanobody, alone and in complex with the full extracellular region of EGF-bound EGFR. The structures explain the non-inhibitory binding of EgB4 to EGFR, the specificity of EgB4 for EGFR domains I and II, and indicate that EgB4 can bind both EGF-bound and unliganded EGFR. This work provides the structural basis for the use of EgB4 as a biomarker to monitor EGFR expression in tissues and tumor imaging while not affecting EGFR function.

Results

Crystallization of the EgB4 nanobody

To investigate the structure of the EgB4 nanobody and its interaction with EGFR, we first determined a high-resolution structure of EgB4 (PDB: 7OM5). The EgB4 crystal diffracted to a maximum resolution of 1.48 Å (Table 1). The structure was solved by molecular replacement using the structure of the EgA1 nanobody (PDB: 4KRO) (43). Two EgB4 molecules are present in the asymmetric unit that align with a RMSD of 0.17 Å. Model building and refinement led to a final model with R_{work}/R_{free} of 0.177/0.205. The framework regions of EgB4, defined as the conserved segments of nanobodies, align with that of a typical VHH (45) with a

	EGFR-EgB4-EGF	EgB4
Data collection		
Space group	$P\,6_1\,2\,2$	$P\,1\,2_1\,1$
Cell dimensions		
a, *b*, *c* (Å)	307.61, 307.61, 135.14	38.54, 71.58, 53.20
α, *β*, *γ* (°)	90.0, 90.0, 120.0	90.0, 91.5, 90.0
Resolution (Å)	153.81 - 6.05 (7.15 - 6.05)	42.69 - 1.48 (1.50 - 1.48)
No. observed reflections	60948 (6230)	162212 (7549)
No. unique reflections	6321 (632)	47214 (2189)
R_{merge}	0.187 (1.995)	0.125 (1.681)
Mean I/σI	7.8 (1.6)	5.4 (1.1)
$CC_{1/2}$	0.996 (0.548)	0.990 (0.277)
Spherical completeness (%)	65.3 (17.0)	98.0 (93.0)
Ellipsoidal completeness (%)	91.8 (63.9)	n/a
Ellipsoidal resolution limits (Å) [direction]	7.25 [a*] 7.25 [b*] 6.02 [c*]	n/a
Redundancy	9.6 (9.9)	3.4 (3.4)
Refinement		
Resolution (Å)	153.81 - 6.05	42.69 - 1.48
R_{work}/R_{free} (%)	29.6 / 32.7	17.7 / 20.5
Average *B*-factors (Å^2)		
Protein	534	17.4
Glycans/ions/ligands	591	15.7
Water	n/a	26.5
R.M.S. deviations		
Bond lengths (Å)	0.0021	0.0122
Bond angles (°)	0.57	1.68
Ramachandran (%)		
Favored	94.0	97.2
Allowed	5.9	2.8
Outliers	0.1	0
Molprobity score	1.59	1.21

Table 1. Data collection and refinement statistics. Highest resolution shell in parentheses. n/a = not applicable.

RMSD of 0.51 Å, whereas a RMSD of 3.9 Å is measured when aligning the complementarity determining regions (CDR). Notably, the CDR3 of EgB4 is relatively short compared to that of other nanobodies (Figure 1) (43, 45).

EgB4 binds to domains I and II in the dimeric EGFR-EGF complex

To study the mechanism by which the EgB4 nanobody interacts with EGFR, we then determined the structure of the full ectodomain EGFR-EgB4-EGF ternary complex from a

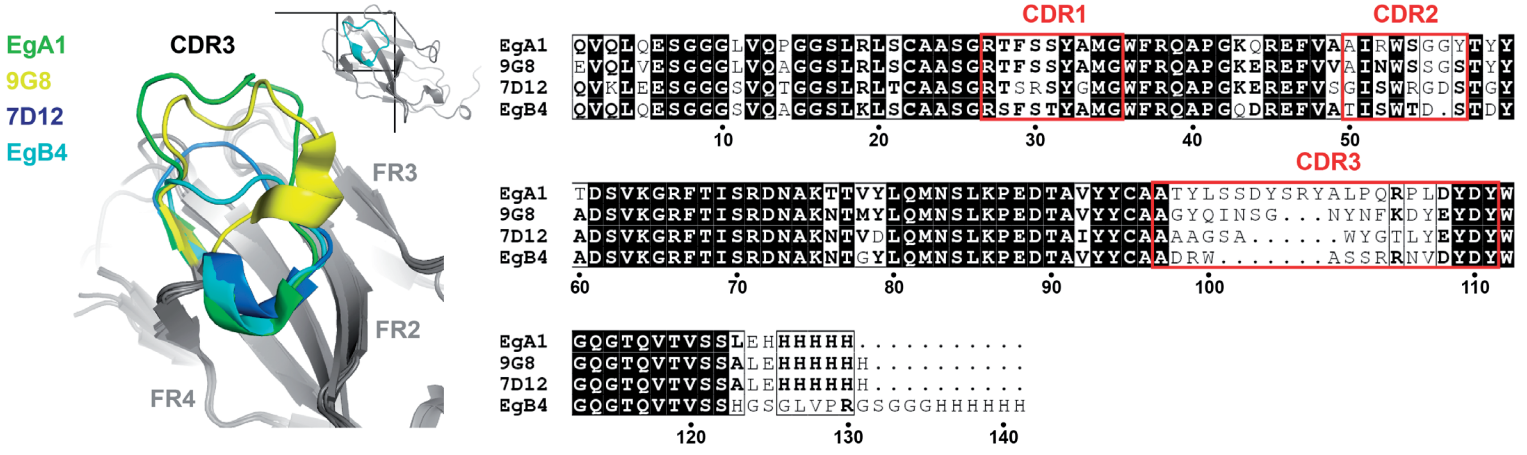


Figure 1. The CDR3 of EgB4 is substantially shorter than that of other anti-EGFR nanobodies. (Left) Structure alignment of four described anti-EGFR nanobodies showing the difference in CDR3 length. Inset shows the complete EgB4 nanobody. FR: framework region (grey). (Right) Corresponding sequence alignment with residues numbered according to the EgB4 sequence.

crystal that diffracted to 6.0 Å resolution (PDB: 7OM4; Table 1). The structure was solved by molecular replacement, using one monomer of the EGFR-EGF complex (PDB: 3NJP) (17) and one monomer of the EgB4 nanobody (described here) as search models. Model building and refinement of the complex led to a final model with R_{work}/R_{free} of 0.296/0.327. The structure shows a heart-shaped receptor-mediated dimer, on top of which EgB4 engages domains I and II, resulting in a physiological 2:2:2 complex (Figure 2). The structures of the EGFR-EGF part of the EgB4-bound complex and the previously determined EGFR-EGF complex (17) are very similar as they have a root mean square deviation (RMSD) of 0.98 Å, indicating that EgB4 binding does not induce conformational changes in the EGF-bound EGFR.

In the EGFR-EgB4-EGF ternary complex, EGFR domains I and III fold into right-handed β-helical barrels, similar to that of a previously solved structure of EGF-bound EGFR (17) with a RMSD of 0.71 Å and 0.82 Å for domains I and III, respectively. Domain II also has a similar structure, with a RMSD of 0.88 Å to previously reported domain II (17), with a protruding beta-hairpin that serves as a dimerization arm, engaging in a homo-interaction with domain II of the partnering receptor. While domains I through III form a compact C shape, domain IV extends from the base of domain III, pointing away from other domains, and curves back into the vertex of the heart, creating a secondary dimerization interface (Figure 2). Also domain IV is similar to that of the previously reported structure (17), it has however a slightly

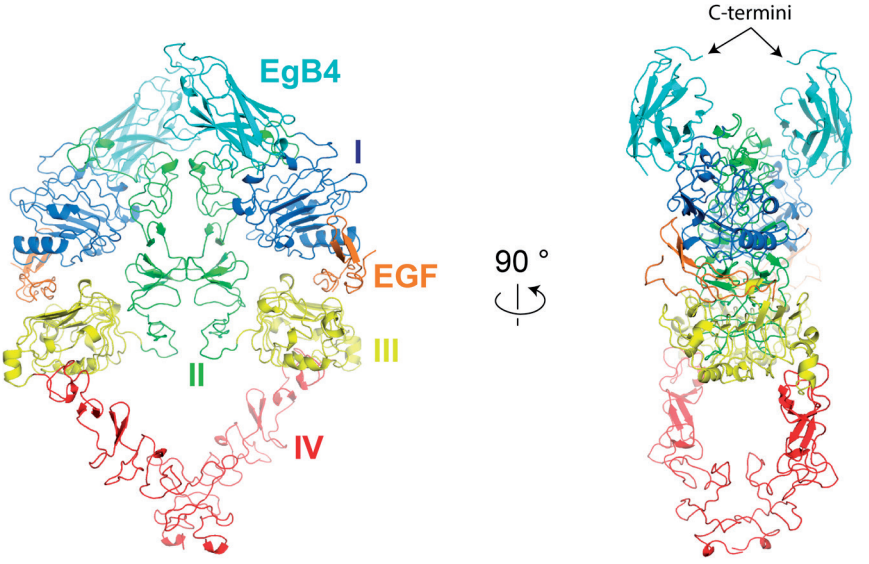


Figure 2. EgB4 nanobody binds to the active dimeric EGFR-EGF complex. The ternary complex, consisting of two EGFR, two EGF and two EgB4 molecules is shown in cartoon representation with domains colored differently. EgB4 binds mainly to EGFR domain I, with smaller contributions from EGFR domain II. EgB4 does not affect EGF binding nor does it change the structure of the EGFR-EGF complex.

larger domain-wise RMSD of 1.69 Å, most likely arising from some structural flexibility in this domain as reflected by the high B-factors in the refined structure.

EGF superimposes with that of the previously solved structure with a RMSD of 1.08 Å (17), and our structure shows EGF binding to EGFR domains I and III through three interfaces, as reported previously (10, 17, 46). The first interface is formed by a loop in the region of EGF residues 20-31 that engages EGFR domain I. EGFR domain III interacts with the region of EGF residues 6-19 and Arg41 in the second interface, and with the EGF C-terminal region in the third interface. Together, the data show that the EGFR-EGF complex is in a physiologically active conformation when bound to EgB4.

Although several ligands and nanobodies were shown to bind to EGFR, EgB4 interacts with a hitherto unreported EGFR epitope. As shown in Figure 3A, the CDR2 and CDR3 of EgB4 interact with the top of EGFR domain I, and CDR3 also interacts with residues at the domain I-II junction, together forming a buried surface area of 1403 Å^2 (Figure 3B). Specifically, a hydrophobic core is formed by the sidechains of Trp140 and Phe156 from the top of EGFR domain I, and by that of tryptophan residues in EgB4 CDR2 (Trp53) and CDR3

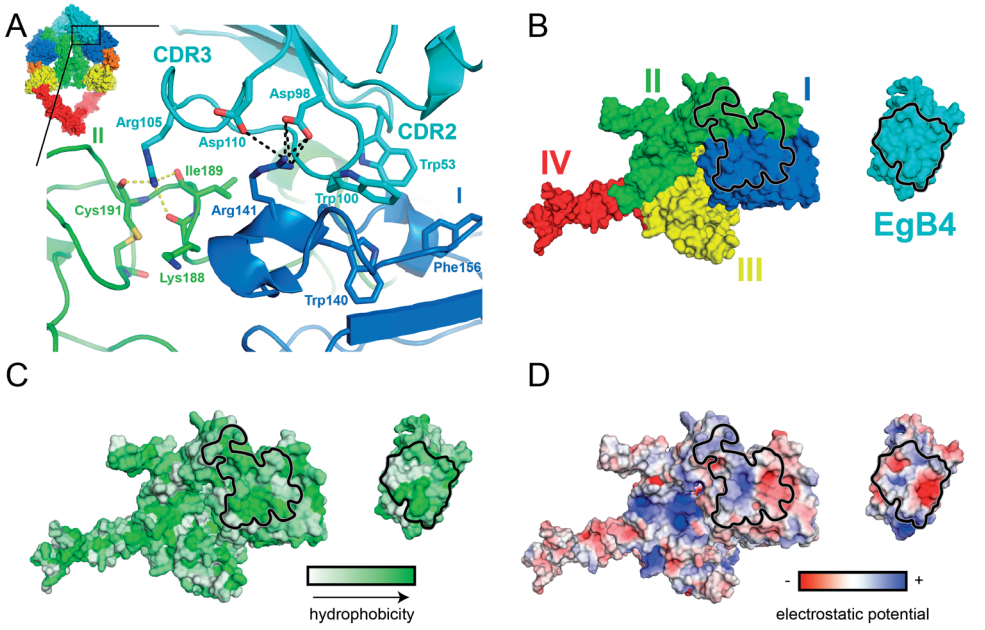


Figure 3. EgB4 nanobody interacts with EGFR domains I and II. A) The CDR3 of EgB4 engages EGFR domains I and II while CDR2 of EgB4 binds to EGFR domain I. Residues involved in the interaction are shown in stick representation. Electrostatic and hydrogen bonding interactions are indicated by black and yellow dotted lines, respectively. Inset shows the EGFR-EgB4-EGF complex in surface representation. B-D) Open book view of the EGFR-EgB4 complex with the interface delimited in black, colored by domains (B), hydrophobicity (C) and electrostatic potential (D).

(Trp100) (Figure 3A and 3C). Within the same region, the sidechain of EGFR Arg141 forms salt bridges with asparagine residues from EgB4 CDR3 (Asp98 and Asp110) (Figure 3A and 3D). The CDR3 Arg105 sidechain forms additional hydrogen bonding interactions with the backbone carbonyl groups of Lys188, Ile189 and Cys191 at the EGFR domain I-II junction, extending the EGFR-EgB4 interface towards EGFR domain II (Figure 3A). In the dimeric complex, although both EgB4 molecules bind at the top of EGFR domain I, they do not interact with each other (Figure 2). Collectively, the data show that EgB4 binds to EGFR domains I and II of the physiological dimeric EGFR-EGF complex, therefore targeting a new epitope that could be used for diagnostic and therapeutic applications.

Discussion

EGFR is a widely studied receptor involved in various cellular processes, such as cell differentiation and migration, and its overexpression in cancers makes it an important therapeutic target (1–5). EGFR-targeting drugs, including monoclonal antibodies (e.g. cetuximab), nanobodies (e.g. 9G8) or tyrosine kinase inhibitors (e.g. erlotinib) were designed to inhibit EGFR signaling by preventing conformational rearrangement of the receptor, competing with ligand binding, or blocking kinase activity (43, 47, 48). Structures of the inhibitory nanobodies 7D12, EgA1 and 9G8 in complex with EGFR show that they all bind to EGFR in its inactive conformation, blocking conformational rearrangement of the receptor and therefore preventing formation of the extended active conformation (43). All three nanobodies engage domain III, and while 7D12 interacts with the ligand binding region, EgA1 and 9G8 bind to a cleft created between domains II and III. Here we reveal the molecular details of EgB4 binding to EGFR by solving crystal structures of EgB4 alone, and in complex with EGF-bound EGFR, to provide structural information on the non-competing characteristics of EgB4. The data show that EgB4 binds to EGFR domains I and II through interactions with the variable regions CDR2 and CDR3. Most notably, a hydrophobic core constituted by tryptophan and phenylalanine residues at the top of EGFR domain I and tryptophan residues in CDR2 and CDR3, and electrostatic interactions between aspartic acid residues of EgB4 CDR3 and Arg141 on EGFR domain I, appear to be key to complex formation. The interaction is stabilized by additional hydrogen bonding interactions between the backbone carbonyl of Lys188, Ile189 and Cys191 on EGFR domain II and Arg105 in CDR3.

The residues involved in the interaction between the active EGFR and EgB4 superimpose with a RMSD of 0.96 Å with that of EGFR in its inactive conformation (49, 50), suggesting there is no EgB4 interface rearrangement within EGFR upon adoption of the active conformation and ligand binding (Figure 4A). Interactions that are formed between EgB4 and the active EGFR are therefore likely to be available in the inactive EGFR. As shown in

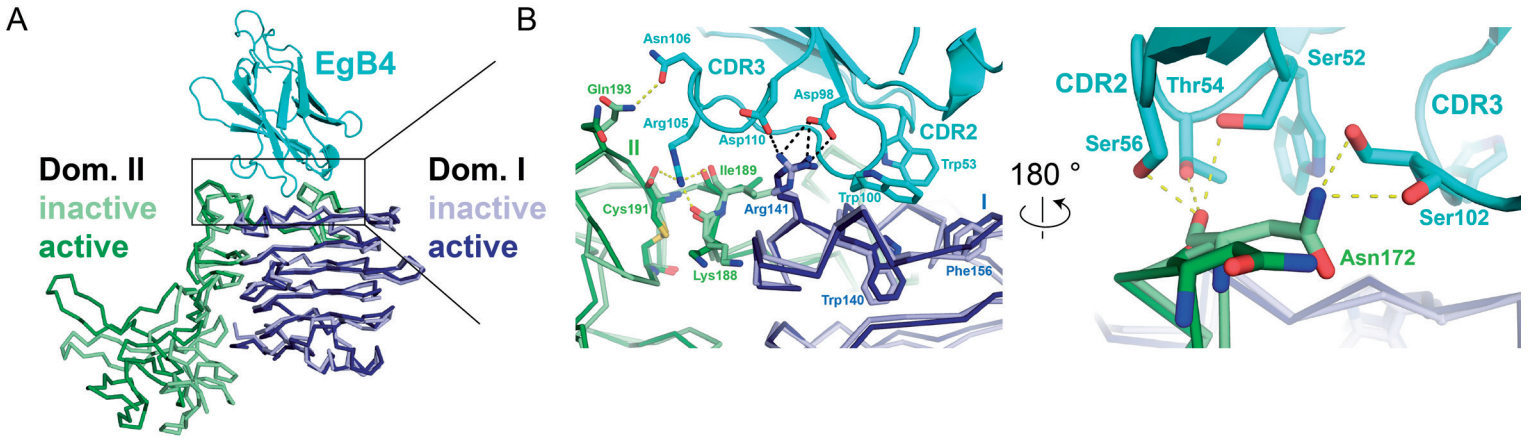


Figure 4. EgB4 nanobody can bind to both inactive and active EGFR. A) Structural alignment of the interface residues from EGFR in the inactive (light colors; PDB: 3qwq) (49) and active EgB4-bound conformation (dark colors) in ribbon representation. B) Inactive EGFR is also poised to interact with EgB4. Residues involved in the interaction are shown in stick representation. Electrostatic and hydrogen bonding interactions are indicated by black and yellow dotted lines, respectively. Potential steric clashes between EgB4 and residues Arg141 and Ile189 in the inactive EGFR are readily alleviated by selecting different rotamers.

Figure 4B, hydrophobic interactions, involving Trp140 and Phe156 from EGFR and Trp53 and Trp100 from EgB4, and salt bridges, involving Arg141 from EGFR and Asp98 and Asp110 from EgB4, can be maintained upon adoption of the EGFR inactive conformation. Hydrogen bonding interactions between Arg105 from EgB4 and Lys188, Ile189 and Cys191 from EGFR are also conserved with the inactive EGFR (Figure 4B). In the inactive EGFR-EgB4 complex, we observe possible additional hydrogen bonding interactions between the sidechains of Gln193 from EGFR and Asn106 from EgB4 CDR3, and between the Asn172 backbone carbonyl from EGFR and sidechains of Ser52, Thr54 and Ser56 from EgB4 CDR2 (Figure 4B). The sidechain of Asn172 is also involved in hydrogen bonding interactions with Ser102 from CDR3 (Figure 4B). It is possible that these additional interactions are also present in the active EGFR-EgB4 complex, but not observed in the structure due to the low resolution of the underlying data. While experimentally we only determined the structure of the active EGFR-EgB4-EGF complex, the data shows that the inactive EGFR conformation is also compatible with EgB4 engagement. It is therefore likely that EgB4 binds to the inactive EGFR and, unlike the 7D12, EgA1 and 9G8 nanobodies, allows the conformational change from the inactive to the active EGFR conformation (Figure 5).

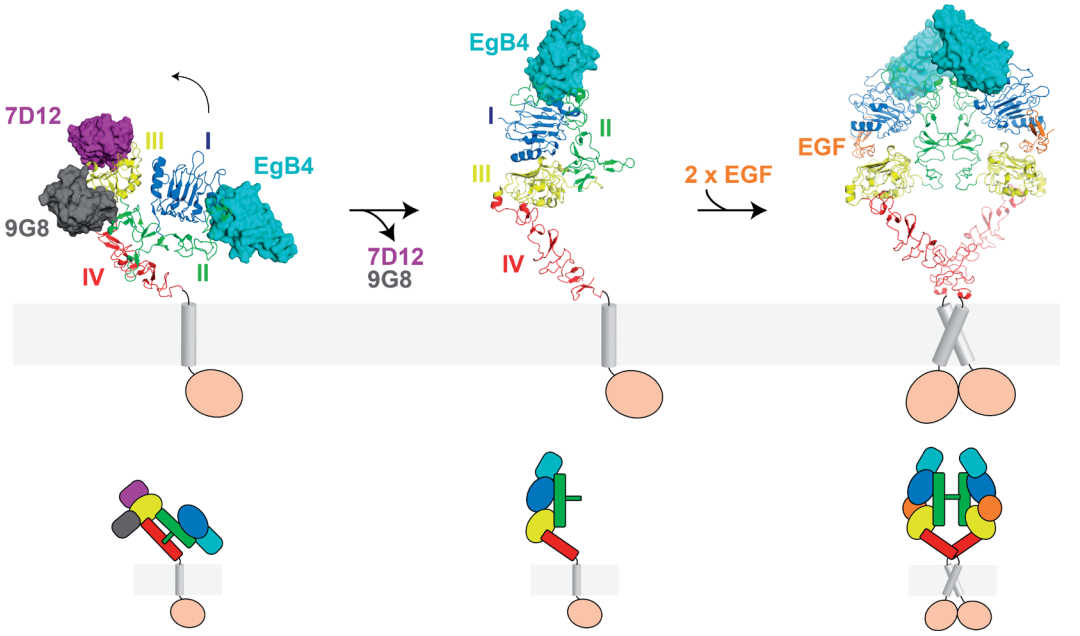


Figure 5. EgB4 nanobody does not affect EGFR signaling while 7D12 and 9G8/EgA1 nanobodies maintain EGFR in the inactive conformation. (Left) In the inactive conformation, four nanobodies can bind to EGFR; EgB4, 7D12, 9G8 and EgA1 (PDB: 3qwq, 4krm, 4krp) (43, 49). Here, EgB4 is modelled by superposition based on the EgB4-EGFR interface from our crystal structure. (Center) In the extended monomeric conformation, only EgB4 may be able to bind to the unliganded EGFR (model based on HER2; PDB: 1n8z) (50). (Right) Two EgB4 molecules can bind to the active dimeric EGFR. (Bottom) Corresponding schematic representations.

The potency of nanobodies can be extended by creating bivalent molecules, i.e. two nanobodies are fused by a flexible linker (38, 51). This can be used to create mono-specific nanobodies, i.e. fusion of two copies of the same nanobody, or biparatopic nanobodies, i.e. fusion of two different nanobodies targeting non-overlapping sites on the same target. As an example, multimerization of the 7D12 nanobody with other VHH domains has successfully led to the inhibition of tumor growth in vivo (40). The EgB4 binding site on EGFR domains I and II is located relatively far from the previously described 7D12, 9G8 and EgA1 binding sites (43), preventing the straightforward design of a dual-specific molecule that includes EgB4 in combination with one of these nanobodies. However, our crystal structure shows that in the active dimeric EGFR, the two EgB4 molecules are in proximity, with the C-termini facing each other on top of the complex, at a distance of 24.2 Å (Figure 2A). This provides the opportunity to design a bivalent mono-specific EgB4-EgB4 molecule in which the individual nanobodies are covalently linked by their C-termini (52–54). This molecule might have an increased affinity for the EGFR-EGF complex, and maintain it in a dimeric state, which could be used for diagnostic or therapeutic applications. Furthermore, the novel binding site of EgB4, located on EGFR domains I and II, could provide specificity on the type of EGFR variant that EgB4 can bind to. For example, EgB4 can probably bind to EGFRvII since that variant only lacks part of domain IV, but not to EGFRvIII that is truncated from most of its domains I and II. The use of EgB4 may help identify specific types of cancers that are characterized by the presence of EGFRvII rather than EGFRvIII, and therefore prove useful to target EGFR-associated cancers.

Methods

Expression and purification of EGFR, EgB4 and EGF

Codon-optimized DNA coding for human EGFR ectodomain (residues 1-621 of the mature protein) was purchased at GeneArt, subcloned in pUPE101.01 vector (U-Protein Express BV, C-terminal His6-tag) and transiently expressed in N-acetylglucoaminyltransferase I-deficient (GnTI-) Epstein-Barr virus nuclear antigen 1 (EBNA1)-expressing HEK293 cells growing in suspension (U-Protein Express BV). The medium was harvested six days after transfection and cells were spun down by 10 minutes of centrifugation at 1000x g. Protein was purified by Ni Sepharose excel (GE Healthcare) affinity chromatography, eluted with 500 mM imidazole in phosphate buffer saline (PBS) pH 7.4, and buffer-exchanged to PBS using the SnakeSkin™ Dialysis Tubing (Thermo Scientific). Size-exclusion chromatography (SEC) was performed on a Superdex200 10/300 increase column (GE Healthcare) equilibrated in SEC buffer (20 mM HEPES pH 7.4, 150 mM NaCl). Protein purity was evaluated by Coomassie-stained sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), concentrated, and stored at -80 °C.

Codon-optimized DNA coding for EgB4 was purchased at Integrated DNA Technologies (IDT BVA), cloned into a customized pHEN6 vector with pelB sequence for expression in the bacterial periplasm and thrombin cleavage site followed by a C-terminal His6-tag. Protein was expressed under IPTG induction in BL21-CodonPlus (DE3)-RIL *E. coli* bacteria cultured in Terrific Broth medium in a New Brunswick™ BioFlo®/CelliGen® 115 bioreactor (pH 7 ± 0.1 and dissolved oxygen 70%). The periplasm was extracted from the harvested bacteria via two rounds of freeze-thaw (-20°C) and was collected in PBS. The nanobody was purified from the isolated periplasm by Ni Sepharose™ High Performance chromatography, eluted in 500 mM imidazole in phosphate buffer saline (PBS) pH 7.4, and buffer-exchanged to PBS using a HiTrap™ Desalting column (GE Healthcare). The C-terminal His6-tag was removed by thrombin cleavage and SEC was performed on a Superdex75 10/300 increase column (GE Healthcare) equilibrated in SEC buffer. Protein purity was evaluated by Coomassie-stained SDS-PAGE, concentrated, and stored at -80 °C.

EGF was bought from Sino Biological Inc., reconstituted in Milli-Q® water, and purified by SEC on a Superdex75 10/300 increase column (GE Healthcare) equilibrated in SEC buffer. Protein purity was evaluated by Coomassie-stained SDS-PAGE, concentrated, and stored at -80 °C.

Crystallization and data collection

The EGFR-EgB4-EGF complex crystal grew by sitting-drop vapour diffusion at 20 °C, by mixing 150 nL of 10 mg/mL protein solution containing EGFR:EgB4:EGF in 1:1.1:1.1 molar ratio, respectively, with 150 nL of reservoir solution containing 0.1 M $LiSO_4$, 0.1 M glycine pH 10.5, 1.1 M sodium dihydrogen phosphate and 0.72 M dipotassium hydrogen phosphate. The crystal was harvested and flash-cooled in liquid nitrogen in presence of reservoir solution supplemented with 20 % glycerol. The dataset was collected at 100 K at the DLS beamline I24 ($\lambda = 0.9686$ Å).

The EgB4 crystal grew by sitting-drop vapour diffusion at 20 °C, by mixing 150 nL of protein solution at 22.4 mg/mL with 150 nL of reservoir solution containing 0.05 M zinc acetate and 20 % w/v PEG3350. The crystal was harvested and flash-cooled in liquid nitrogen in presence of reservoir solution supplemented with 25 % glycerol. The dataset was collected at 100 K at the DLS beamline I24 ($\lambda = 0.9688$ Å).

Structure solution and refinement

The EGFR-EgB4-EGF complex data was processed in the autoPROC pipeline (55), and additional anisotropic correction was done using the STARANISO server (56). The structure was solved by molecular replacement in PHASER (57), using one copy of the high-resolution EGFR-EGF complex (PDB: 3NJP) and one copy of the high-resolution EgB4 nanobody (described here). One copy of each molecule is present in the asymmetric unit. Refinement

was done in REFMAC and PHENIX using TLS groups, to model the B-factors (one group per EGFR domain, one group for EgB4 and one group for EGF), jelly-body and tight geometry restraints (58–60). Minimum manual rebuilding was done in COOT to correct Ramachandran outliers (61). MOLPROBITY (62) was used for validation. The final model has a R_{work}/R_{free} of 0.296/0.327 and was deposited to the Protein Data Bank under the accession code 7OM4.

The EgB4 data was processed in the XIA2 pipeline (63). The structure was solved by molecular replacement in PHASER (57), using one copy of the EgA1 nanobody as search model. Two copies of EgB4 are present in the asymmetric unit. Refinement was done in REFMAC (58), and MOLPROBITY (62) was used for validation. The final model has a R_{work}/R_{free} of 0.177/0.205 and was deposited to the Protein Data Bank under the accession code 7OM5. Sequence alignment was done in Clustal Omega (64) and represented with ESPript (65). Structure alignments were made in Pymol using the "align" command, and figures were made using Pymol (66).

Author contributions

S.D. and M.R.Z. designed the constructs, S.D. produced the proteins, M.R.Z. and S.D. purified the proteins, M.R.Z. did the X-ray diffraction experiments and data analysis, B.J.C.J. and P.M.P.v.B.H. supervised the project, M.R.Z. wrote the first draft of the manuscript (including figures), M.R.Z. and B.J.C.J. wrote the manuscript with input from all authors.

Competing Interest Statement

The authors declare no competing interests.

References

1. C. Arienti, S. Pignatta, A. Tesei, Epidermal Growth Factor Receptor Family and its Role in Gastric Cancer. *Front. Oncol.* **9**, 1–11 (2019).
2. K. Khan, *et al.*, Targeting EGFR pathway in metastatic colorectal cancer- tumour heterogeniety and convergent evolution. *Crit. Rev. Oncol. Hematol.* **143**, 153–163 (2019).
3. A. Citri, Y. Yarden, EGF-ERBB signalling: Towards the systems level. *Nat. Rev. Mol. Cell Biol.* **7**, 505–516 (2006).
4. C. L. Arteaga, J. A. Engelman, ERBB receptors: From oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* **25**, 282–303 (2014).
5. Y. Yarden, M. X. Sliwkowski, Untangling the ErbB network. *Nat. Rev. Mol. Cell Biol.* **2**, 127–137 (2001).
6. D. M. Thompson, G. N. Gill, The EGF receptor: structure, regulation and potential role in malignancy. *Cancer Surv.* **4**, 767–788 (1985).
7. P. Wee, Z. Wang, Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers (Basel).* **9**, 1–45 (2017).
8. P. Seshacharyulu, *et al.*, Targeting the EGFR signaling pathway in cancer therapy. *Expert Opin. Ther. Targets* **16**, 15–31 (2012).
9. C. W. Ward, P. A. Hoyne, R. H. Flegg, Insulin and epidermal growth factor receptors contain the cysteine repeat motif found in the tumor necrosis factor receptor. *Proteins Struct. Funct. Bioinforma.* **22**, 141–153 (1995).
10. H. Ogiso, *et al.*, Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110**, 775–787 (2002).
11. N. Jura, *et al.*, Mechanism for Activation of the EGF Receptor Catalytic Domain by the Juxtamembrane Segment. *Cell* **137**, 1293–1307 (2009).
12. A. Arkhipov, *et al.*, Architecture and membrane interactions of the EGF receptor. *Cell* **152**, 557–569 (2013).
13. N. F. Endres, *et al.*, Conformational coupling across the plasma membrane in activation of the EGF receptor. *Cell* **152**, 543–556 (2013).
14. M. A. Lemmon, J. Schlessinger, Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
15. X. Zhang, J. Gureasko, K. Shen, P. A. Cole, J. Kuriyan, An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell* **125**, 1137–1149 (2006).
16. K. M. Ferguson, *et al.*, EGF Activates Its Receptor by Removing Interactions that Autoinhibit Ectodomain Dimerization molecules. Crystallographic studies of several ligand-bound RTK extracellular fragments have allowed visual-ization of such dimers, including those from the Flt-1. *Mol. Cell* **11**, 507–517 (2003).
17. C. Lu, *et al.*, Structural Evidence for Loose Linkage between Ligand Binding and Kinase Activation in the Epidermal Growth Factor Receptor. *Mol. Cell. Biol.* **30**, 5432–5443 (2010).

18. C. M. Chia, R. M. L. Winston, A. H. Handyside, EGF, TGF-α and EGFR expression in human preimplantation embryos. *Development* **121**, 299–307 (1995).
19. I. D. Campbell, P. Bork, Epidermal growth factor-like modules. *Curr. Opin. Struct. Biol.* **3**, 385–392 (1993).
20. D. W. Threadgill, *et al.*, Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. *Science (80-.).* **269**, 230 LP – 234 (1995).
21. M. Yamada, T. Ikeuchi, H. Hatanaka, The neurotrophic action and signalling of epidermal growth factor. *Prog. Neurobiol.* **51**, 19–37 (1997).
22. N. Sugawa, A. J. Ekstrand, C. D. James, V. P. Collins, Identical splicing of aberrant epidermal growth factor receptor transcripts from amplified rearranged genes in human glioblastomas. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8602–8606 (1990).
23. S. I. Ymer, *et al.*, Glioma specific extracellular missense mutations in the first cysteine rich region of epidermal growth factor receptor (EGFR) initiate ligand independent activation. *Cancers (Basel).* **3**, 2032–2049 (2011).
24. R. Nishikawa, *et al.*, A mutant epidermal growth factor receptor common in human glioma confers enhanced tumorigenicity. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7727–7731 (1994).
25. H. K. Gan, A. N. Cvrljevic, T. G. Johns, The epidermal growth factor receptor variant III (EGFRvIII): Where wild things are altered. *FEBS J.* **280**, 5350–5370 (2013).
26. J. M. Francis, *et al.*, EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. *Cancer Discov.* **4**, 956–971 (2014).
27. E. R. Kastenhuber, *et al.*, Quantitative assessment of intragenic receptor tyrosine kinase deletions in primary glioblastomas: Their prevalence and molecular correlates. *Acta Neuropathol.* **127**, 747–759 (2014).
28. J. G. Paez, *et al.*, EGFR mutations in lung, cancer: Correlation with clinical response to gefitinib therapy. *Science (80-.).* **304**, 1497–1500 (2004).
29. W. Pao, *et al.*, EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13306–13311 (2004).
30. T. J. Lynch, *et al.*, Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004).
31. S. I. Rudnick, G. P. Adams, Affinity and avidity in antibody-based tumor targeting. *Cancer Biother. Radiopharm.* **24**, 155–161 (2009).
32. C. G. Siontorou, Nanobodies as novel agents for disease diagnosis and therapy. *Int. J. Nanomedicine* **8**, 4215–4227 (2013).
33. J. H. E. Baker, *et al.*, Direct visualization of heterogeneous extravascular distribution of trastuzumab in human epidermal growth factor receptor type 2 overexpressing xenografts. *Clin. Cancer Res.* **14**, 2171–2179 (2008).
34. C. Hamer-Casterman, Atarchouch, T, *et al.*, Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448 (1993).

35. M. Kijanka, B. Dorresteijn, S. Oliveira, P. M. P. Van Bergen En Henegouwen, Nanobody-based cancer therapy of solid tumors. *Nanomedicine* **10**, 161–174 (2015).
36. E. Y. Yang, K. Shah, Nanobodies: Next Generation of Cancer Diagnostics and Therapeutics. *Front. Oncol.* **10** (2020).
37. M. Scully, *et al.*, Caplacizumab Treatment for Acquired Thrombotic Thrombocytopenic Purpura. *N. Engl. J. Med.* **380**, 335–346 (2019).
38. R. C. Roovers, *et al.*, Efficient inhibition of EGFR signalling and of tumour growth by antagonistic anti-EGFR Nanobodies. *Cancer Immunol. Immunother.* **56**, 303–317 (2007).
39. E. G. Hofman, *et al.*, EGF induces coalescence of different lipid rafts. *J. Cell Sci.* **121**, 2519–2528 (2008).
40. R. C. Roovers, *et al.*, A biparatopic anti-EGFR nanobody efficiently inhibits solid tumour growth. *Int. J. Cancer* **129**, 2013–2024 (2011).
41. S. Oliveira, *et al.*, Downregulation of EGFR by a novel multivalent nanobody-liposome platform. *J. Control. Release* **145**, 165–175 (2010).
42. R. Van Der Meel, *et al.*, Tumor-targeted Nanobullets: Anti-EGFR nanobody-liposomes loaded with anti-IGF-1R kinase inhibitor for cancer treatment. *J. Control. Release* **159**, 281–289 (2012).
43. K. R. Schmitz, A. Bagchi, R. C. Roovers, P. M. P. Van Bergen En Henegouwen, K. M. Ferguson, Structural evaluation of EGFR inhibition mechanisms for nanobodies/VHH domains. *Structure* **21**, 1214–1224 (2013).
44. L. C. Zanetti-Domingues, *et al.*, The architecture of EGFR's basal complexes reveals autoinhibition mechanisms in dimers and oligomers. *Nat. Commun.* **9** (2018).
45. A. Desmyter, *et al.*, Crystal structure of a camel single- with lysozyme. **3** (1996).
46. T. P. J. Garrett, *et al.*, Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α. *Cell* **110**, 763–773 (2002).
47. S. Li, *et al.*, Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* **7**, 301–311 (2005).
48. J. Stamos, M. X. Sliwkowski, C. Eigenbrot, Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.* **277**, 46265–46272 (2002).
49. V. Ramamurthy, *et al.*, Structures of adnectin/protein complexes reveal an expanded binding footprint. *Structure* **20**, 259–269 (2012).
50. H. S. Cho, *et al.*, Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* **421**, 756–760 (2003).
51. K. E. Conrath, M. Lauwereys, L. Wyns, S. Muyldermans, Camel Single-domain Antibodies as Modular Building Units in Bispecific and Bivalent Antibody Constructs. *J. Biol. Chem.* **276**, 7346–7350 (2001).
52. M. D. Witte, *et al.*, Preparation of unnatural N-to-N and C-to-C protein fusions. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11993–11998 (2012).
53. M. Rashidian, *et al.*, Enzyme-Mediated Modification of Single-Domain Antibodies for Imaging Modalities with Different Characteristics. *Angew. Chemie–Int. Ed.* **55**, 528–533 (2016).

54. B. Zang, *et al.*, Freezing-assisted synthesis of covalent C-C linked bivalent and bispecific nanobodies. *Org. Biomol. Chem.* **17**, 257–263 (2019).
55. C. Vonrhein, *et al.*, Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 293–302 (2011).
56. I. J. Tickle, *et al.*, Staraniso. *Glob. Phasing Ltd., Cambridge, UK* (2018).
57. A. J. McCoy, *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
58. G. N. Murshudov, *et al.*, REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
59. D. Liebschner, *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in {\it Phenix}. *Acta Crystallogr. Sect. D* **75**, 861–877 (2019).
60. R. A. Nicholls, F. Long, G. N. Murshudov, Low-resolution refinement tools in REFMAC5. *Acta Crystallogr. Sect. D* **68**, 404–417 (2012).
61. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
62. V. B. Chen, *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).
63. G. Winter, XIA2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.–J APPL CRYST* **43**, 186–190 (2010).
64. F. Madeira, *et al.*, The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636—W641 (2019).
65. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320-4 (2014).
66. L. Schrödinger, "The PyMOL Molecular Graphics System, Version 2.3" (2015).

Chapter 4

Structure-activity studies and high-resolution crystal structures provide new mechanistic insights into C_{55}-P targeting lipopeptide antibiotics.

Thomas M. Wood,[1,2,†] Matthieu R. Zeronian,[3,†] Hanieh K. Abedian,[1] Aidan V. Johnson,[1] Nicolas M. Pearce,[3] Martin Lutz,[3] Johan Kemmink,[4] Tjalling Seirsma,[5] Leendert W. Hamoen,[5] Bert J. C. Janssen,[3,*] Nathaniel I. Martin[1,*]

[1]Biological Chemistry Group, Institute of Biology Leiden, Leiden University, Sylviusweg 72, 2333 BE Leiden, The Netherlands. *n.i.martin@biology.leidenuniv.nl

[2]Department of Chemical Biology & Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands.

[3]Structural Biochemistry, Bijvoet Centre for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. *b.j.c.janssen@uu.nl

[4]Faculty of Science and Engineering, University of Groningen, Groningen, The Netherlands

[5]Bacterial Cell Biology and Physiology Group, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

[†]These authors contributed equally to this work.

*Corresponding author.

Email: n.i.martin@biology.leidenuniv.nl (N.I.M), b.j.c.janssen@uu.nl (B.J.C.J.)

This chapter has been submitted for publication.

Abstract

The continued rise of antibiotic resistance is a global concern that threatens to undermine many aspects of modern medical practice. Key to addressing this threat is the discovery and development of new antibiotics that operate by unexploited modes of action. The so-called calcium-dependent lipopeptide antibiotics (CDAs) are an important emerging class of natural products that provides a source of new antibiotic agents rich in structural and mechanistic diversity. Notable in this regard is the subset of CDAs comprising the laspartomycins and amphomycins/friulimicins that specifically target the bacterial cell wall precursor undecaprenyl phosphate (C_{55}-P). In this study we describe the design and synthesis of new C_{55}-P-targeting CDAs with structural features drawn from both the laspartomycin and amphomycin/friulimicin classes. Antibacterial assessment of these lipopeptides reveals previously unknown and surprisingly subtle structural features that are required for potent activity. High-resolution crystal structures further indicate that the amphomycin/friulimicin-like lipopeptides adopt a unique crystal packing that governs their interaction with C_{55}-P and provides an explanation for their antibacterial effect. In addition, live-cell microscopy studies provide further insights into the biological activity of the C_{55}-P targeting CDAs highlighting their unique mechanism of action relative to the clinically used CDA daptomycin.

Introduction

The rapid emergence of multidrug resistant bacteria presents a growing threat to human health and is considered a top priority of the World Health Organization (1). The most effective way to address this threat is to identify antibiotics that operate by unique, unexploited mechanisms (2). While the so-called "golden age" of antibiotic discovery spanning the 1940s-1960s delivered a plethora of life-saving drugs, in the subsequent 50 years only two new antibiotic classes operating with truly novel modes of action have been introduced (3). Among these is the macrocyclic lipopeptide daptomycin, the preeminent calcium-dependent antibiotic (CDA), which entered the clinic as a first-in-class agent in 2004 (4, 5). Despite its clinical success, daptomycin's precise mechanism of action remains a topic of continued investigation (6–9). By comparison, the mode of action of other structurally similar CDAs like laspartomycin C, friulimicin B, and amphomycin (Fig. 1) are more fully understood (10–13). These CDAs specifically target the unique bacterial phospholipid undecaprenyl phosphate (C_{55}-P). In bacteria, C_{55}-P plays an essential role as a lipid carrier in cell wall biosynthesis (14). Specifically, on the inner surface of the bacterial membrane, the enzyme MraY couples C_{55}-P with UDP-MurNAc-pentapeptide to form lipid I. The membrane anchored lipid I is next converted to lipid II by action of MurG. Lipid II is then flipped to the periplasm where the disaccharide-pentapeptide motif is incorporated into the growing peptidoglycan layer and the phospholipid carrier is released, first as the pyrophosphate (C_{55}-PP) which is subsequently converted to C_{55}-P by action of the phosphatase UppP/BacA (15, 16). For another cycle to begin, C_{55}-P must first be flipped back to the cytoplasm where it can again be used as a membrane anchor for peptidoglycan synthesis. Compounds capable of binding to and sequestering C_{55}-P on the outer surface of the bacterial membrane therefore have the capacity to function as antibacterial agents. Notably, while C_{55}-P plays a central role in peptidoglycan synthesis, to date there are no clinically approved antibiotics that operate by directly binding C_{55}-P.

To date, more than forty structurally distinct CDAs have been reported with varying antibacterial activities and mechanisms of action (17). A number of structural features are common among the CDAs, including specifically positioned D-amino acids and the highly conserved Asp-X-Asp-Gly motif, essential for binding of calcium (Fig. 1) (18). Apart from the recently reported malacidins (19, 20), all CDAs contain 10 amino acids in their macrocycle. An interesting sub-class of CDAs are those wherein the peptide macrocycle is closed by a lactam linkage, a group comprised of laspartomycin C, friulimicin B, and amphomycin (Fig. 1).

Considering their structural similarities, it is perhaps not surprising that all three share the same target (C_{55}-P) as part of their antibacterial mechanisms. There are, however, a number of subtle structural features that distinguish the friulimicins/amphomycins from

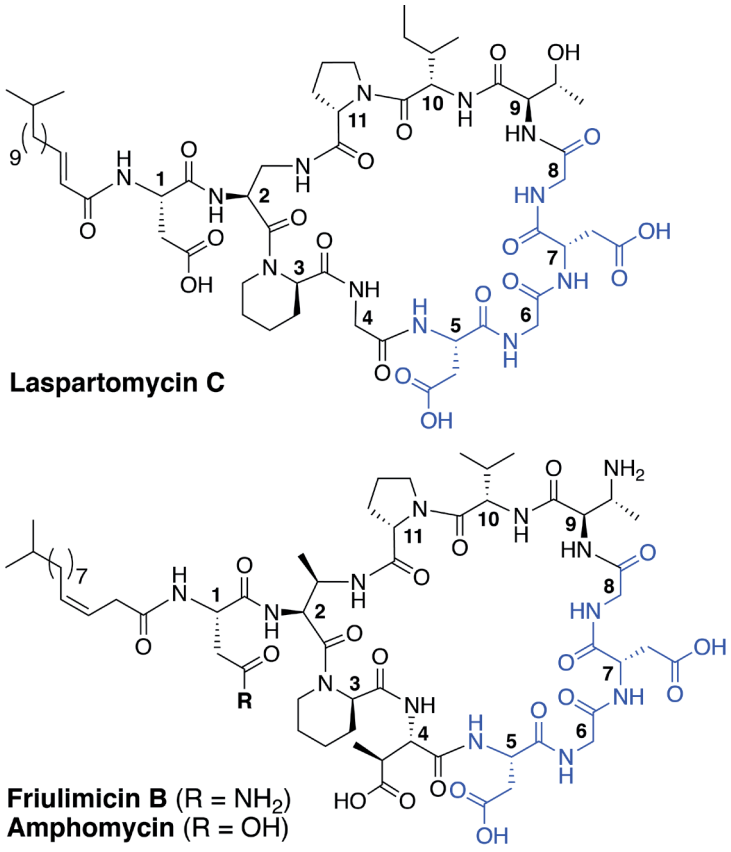


Figure 1. Structures of laspartomycin C, friulimicin B, and amphomycin. Highlighted in blue is the Asp-X-Asp-Gly calcium binding motif conserved throughout the CDAs. Laspartomycin C differs from friulimicin B and amphomycin at positions 2, 4, 9, and 10.

laspartomycin C. For example, the length and geometry of the N-terminal lipid in friulimicin B and amphomycin differs slightly from that found in laspartomycin C. In addition, while both laspartomycin C and amphomycin contain an Asp residue at position 1, in friulimicin B this is Asn. A more notable difference is observed within the peptide macrocycles of these CDAs. Laspartomycin C contains diamino-propionic acid (Dap), Gly, D-*allo*-Thr, and Ile at positions 2, 4, 9, and 10 respectively whereas in the friulimicin/amphomycin class the same position are filled by (*2S,3R*)-diamino-butyric acid (Dab), (*2S,3S*)-3-methyl-Asp, (*2R,3R*)-diamino-butyric acid (D-Dab), and Val respectively (Fig. 1).

Previous findings from our group revealed that for laspartomycin C the side chains of residues 4, 9, and 10 are not directly involved in coordinating the C_{55}-P phosphate head group or the bridging calcium ions (12). This knowledge, coupled with the structural differences between the laspartomycin and friulimicin/amphomycin class at these positions,

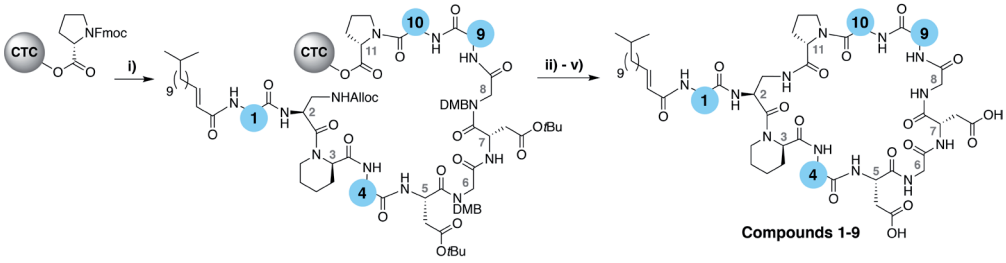
prompted us to investigate the impact of introducing structural features present in friulimicin/amphomycin into the laspartomycin C macrocycle. Specifically, the introduction of residues containing carboxylate and amino side chains at positions 4 and 9, respectively were first investigated providing analogues with rather diminished activity. Surprising, however, was the subsequent finding that in the same series, the additional substitution of Ile^{10} in laspartomycin C for Val (as in the friulimicin/amphomycin class) resulted in a significant enhancement of antibacterial activity. This subtle effect, wherein antibacterial activity is strongly dependent upon the absence of a single methyl group in the side chain of the residue at position 10, was subsequently investigated and explained by high-resolution X-ray crystal structures of the new lipopeptide variants in complexation with C_{10}-P and Ca^{2+}. Notably, these findings provide key new insights into the mechanism of action of the friulimicin/amphomycin class of CDAs and the subtle differences with that of the laspartomycin family. In addition, a series of live-cell imaging studies were also performed that shed new light on the effects that C_{55}-P targeting CDAs have on bacterial cell growth and division.

Results and Discussion

To evaluate the impact of introducing amino acids specific to the friulimicins/amphomycin class into laspartomycin C we applied a robust synthetic route wherein the linear peptide precursor, including the N-terminal lipid, was first assembled on solid support using the acid sensitive 2-chlorotrityl resin (Scheme 1) (11, 21, 22). Notably, Gly residues at positions 6 and 8 were incorporated using the corresponding Fmoc-DMB-Gly building block to avoid aspartamide formation. On resin removal of the Alloc group on the Dap^2 side chain was followed by mild acid cleavage to yield the protected peptide intermediate. Formation of the macrocycle was achieved by treatment with BOP/DIPEA under high-dilution conditions, after which global deprotection and RP-HPLC purification provided lipopeptides **1-9**. The first structural variation explored involved the swapping of the exocyclic Asp^1 found in laspartomycin C for Asn^1 as in friulimicin B. This analogue (**2**) showed no appreciable difference in minimum inhibitory concentration (MIC) when compared to laspartomycin C (Table 1). This is not surprising as this exocyclic amino acid is also the only difference between the amphomycin and friulimicin class of CDAs which are reported to have similar activities (11–13). We next focused our attention to the differing amino acids contained within the peptide macrocycles of the laspartomycin and friulimicins/amphomycin classes. To this end compounds **3**-**5** were prepared to assess the contribution of the acidic and basic residues unique to positions 4 and 9 in the amphomycin/friulimicin class. Interestingly, these new variants bearing either Asp^4 or D-Dap^9, or both substitutions, demonstrated severely reduced antibacterial activities relative to laspartomycin C and friulimicin B. Compound **6** was next synthesized to probe the role of Val^{10} present in the amphomycin/friulimicin class

compared to the slightly bulkier Ile reside found at the same position in laspartomycin C. Somewhat surprisingly, the subtle substitution of Val for Ile at position 10 led to a significant enhancement in the antibacterial activity of lipopeptide **6** relative to compound **5.** In the presence of 10 mM Ca^{2+}, **6** was found to exhibit an MIC of 1 µg/mL against MRSA, an activity on par/slightly better than that measured for both laspartomycin C and friulimicin B. Given the potent activity observed for **6**, analogue **7**, bearing Asn at position 1, was also prepared and found to also demonstrate a similarly enhanced antibacterial activity.

Our findings with lipopeptides **6** and **7** indicate that the potent antibacterial activity exhibited by these more friulimicin/amphomycin-like analogues is the product of a combined effect dependent on an acidic side chain at AA^4, a basic side chain at AA^9, and a slightly less bulky side chain in AA^{10}. This reasoning was further probed by the



Scheme 1: i) Fmoc SPPS; ii) $Pd[PPh_3]_4$, $PhSiH_3$, CH_2Cl_2, 1h; iii) HFIP, CH_2Cl_2, 1h; (v) BOP, DIPEA, CH_2Cl_2, 16h; v) TFA, TIS, H_2O, 1h.

Table 1. MIC[a] values for laspartomycin C, compounds **2-9**, and friulimicin B

Compound	AA^1	AA^4	AA^9	AA^{10}	[Ca2+] (mM)			
					0	**2.5**	**5**	**10**
1 (Laspartomycin C)	L-Asp	Gly	D-*allo*-Thr	L-Ile	>128	8	4	2
2	L-Asn	Gly	D-*allo*-Thr	L-Ile	>128	8	4	4
3	L-Asp	L-Asp	D-*allo*-Thr	L-Ile	>128	64	32	16
4	L-Asp	Gly	D-Dap	L-Ile	>128	16	8	4
5	L-Asp	L-Asp	D-Dap	L-Ile	>128	32	16	8
6	L-Asp	L-Asp	D-Dap	L-Val	>128	**4**	**2**	**1**
7	L-Asn	L-Asp	D-Dap	L-Val	>128	**4**	**4**	**2**
8	L-Asp	L-Asp	D-*allo*-Thr	L-Val	>128	8	4	4
9	L-Asp	Gly	D-Dap	L-Val	>128	32	16	8
Friulimicin B[b]	L-Asn	MeAsp	D-Dab	L-Val	>128	4	2	1-2

[a]Minimum inhibitory concentration reported in µg/mL against MRSA USA 300 at calcium concentration indicated.
[b]Natural product.
All compounds tested in triplicate

preparation of analogues **8** and **9** wherein the acidic and basic residues at positions 4 and 9 were independently exchanged for the uncharged Gly and D-*allo*-Thr residues found in laspartomycin C. The reduced activity measured for these compounds further confirms a role for both the acidic and basic residues at positions 4 and 9, in combination with the optimized sterics of Val^{10}, in achieving full antibacterial effect. Compounds **6** and **7** were further assessed against a range of Gram-positive pathogens including vancomycin-resistance and daptomycin-resistant isolates further demonstrating their potent antibacterial activities (supplemental tables S1-S3).

Mechanistic and crystallographic studies

The potent antibacterial activity observed for lipopeptides **6** and **7** led us to investigate the underlying mechanism responsible. To do so we first examined the ability of the compounds to interfere with bacterial cell wall synthesis. Specifically, an assay was used that detects accumulation of the cell wall precursor UDP-MurNAc pentapeptide in response to cell-wall targeting antibiotics. As the last soluble precursor in the lipid II cycle, UDP-MurNAc pentapeptide serves as a convenient diagnostic for compounds that disrupt cell wall synthesis. When *S. aureus* cells were treated with laspartomycin C and lipopeptides **6** and **7**, a clear accumulation of this precursor was observed (Supplementary Fig. S1). Interestingly, no such accumulation of the UDP-MurNAc pentapeptide species is observed for cells treated with daptomycin.

To gain molecular level insights into the differences in activity observed for lipopeptide **6** and **7** relative to analogue **5**, all three were taken forward for crystal structure determination. In doing so the lipopeptides were incubated with C_{10}-P, a more soluble analogue of C_{55}-P, in buffers containing Ca^{2+}. Under these conditions, compounds **5** and **7** gave crystals of sufficient quality for structural analysis, diffracting to a resolution of 1.04 Å and 1.14 Å, respectively. The structures of the complexes obtained for both **5** and **7** with C_{10}-P and Ca^{2+} share many similarities with the structure previously reported for the laspartomycin C complex (12). As illustrated in Figure 2A, the complex itself consists of one lipopeptide molecule, one geranyl phosphate ligand, and two calcium ions which play key roles both in establishing the conformation of the peptide as well as mediating binding of the phosphate head group. Notable interactions observed in the complex include hydrogen bonds formed between the C_{10} phosphate group and the backbone and side chain amides of Dap^2 as well as the backbone amide of Gly^8. Each calcium ion also provides an interaction with the phosphate moiety. Of the two calcium ions in the complex, one is more centrally coordinated via multiple interactions with the lipopeptide including four backbone carbonyls (Dap^2, Gly^6, Gly^8, Ile/Val^{10}) and one aspartic acid side chain (Asp^5). The peripheral Ca^{2+} is bound via interactions with the side chains of Asp/Asn^1 and Asp^7 and the N-terminal fatty acid carbonyl group along with one water molecule.

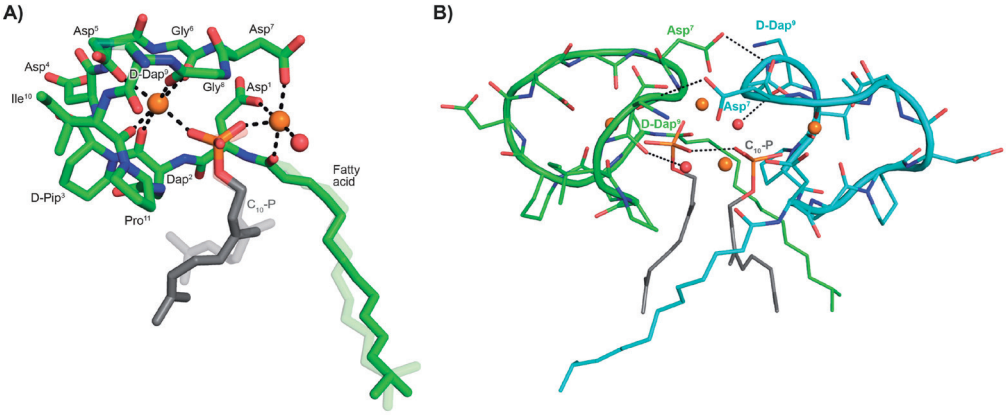


Figure 2. A) Structure of the ternary complex with lipopeptide **5** (green stick representation), two bound Ca^{2+} ions (orange spheres), a bound water molecule (red sphere), and the C_{10}-P ligand (lipid in grey). Major and minor conformations of the D-Dap^9 side chain, the C_{10}-P lipid and the lipopeptide fatty acid tail are shown in the structure (indicated with dark and light colouring respectively). **B)** Lipopeptide **5** adopts a saddle-shaped conformation when complexed with two Ca^{2+} ions and C_{10}-P that forms a dimer in the crystal. For clarity only major conformations are shown. Supplemental Figure S2 presents the same views for lipopeptide **7**.

Collectively, these interactions cause the lipopeptides to adopt a saddle-shaped fold wherein the cavity created envelops the C_{10}-P phosphate head group and the two calcium ions. As also observed for laspartomycin C, the complexes formed by both compounds **5** and **7** with C_{10}-P and Ca^{2+} organize as dimers stabilized by a number of intermolecular interactions. As shown in Figure 2B, dimerization is largely driven by hydrogen bonding interactions between the D-Dap^9 backbone amide of one lipopeptide molecule and the Asp^7 side chain carboxylate of the other. Additional indirect hydrogen bonding interactions are mediated by interaction of the D-Dap^9 backbone carbonyl and the water molecules coordinated by the peripheral Ca^{2+} of the other ternary unit. In this dimer complex, the two C_{10}-P phosphate head groups are fully coordinated and completely sequestered from the solvent. A comparison of the conformation of the peptide backbones and location of the C_{10}-P and Ca^{2+} in the dimers formed by **5** and **7** with that of laspartomycin C reveals a high degree of similarity (Supplemental Fig. S3). Notable, however, was the finding that the differing side chains at positions 4, 9, and 10 in compounds **5** and **7** induce and stabilize a unique, higher-ordered assembly not observed for laspartomycin C.

As noted above, the amphomycin/friulimicin class of lipopeptide antibiotics differs from the laspartomycin class at positions 4, 9, and 10. Compounds **5, 6,** and **7** were generated to specifically probe the roles played by the side chains of these different amino acids. The crystal structures obtained with **5** and **7** indeed reveal that the presence of Asp^4 and D-Dap^9 result in additional inter-dimer interactions not possible for laspartomycin C. Particularly striking was the finding that when coordinated with C_{10}-P and Ca^{2+}, lipopeptides **5** and **7**

both formed higher-ordered complexes that are not observed for laspartomycin under similar conditions (Fig. 3A). Specifically, interactions between Asp^4 and D-Dap^9 in **5** and **7** serve to stabilize this higher-ordered arrangement wherein the side chain carboxylate and backbone carbonyl of one Asp^4 residue in one dimer interacts with the amino side chain of a D-Dap^9 residue in an adjacent dimer (Fig. 3B). Furthermore, the same Asp^4 also interacts with the proximal calcium coordinated by the second dimer further stabilizing this arrangement. Interestingly, the dimer of dimers thus formed is precisely oriented so as to make the same interactions with the Asp^4 and D-Dap^9 side chains of a third dimer to generate a trimer of dimers. This repeating "trimer of dimers" motif is not observed in the crystal packing formed by laspartomycin C in complex with C_{10}-P and Ca^{2+} as it lacks the Asp^4 and D-Dap^9 required to do so. Also different from laspartomycin C is the finding that lipopeptides **5** and **7** form alternating layers in the crystal, with a peptide macrocycle layer inducing a strong packing in *cis* (within the same layer), sandwiched by a hydrophobic layer constituted of lipids (geranyl phosphate and the peptide N-terminal lipid) and by a hydrophilic layer composed of water molecules, both inducing a weak packing in *trans* (between adjacent layers) (Supplemental Fig. S4).

The higher-ordered trimer of dimers motif formed by both **5** and **7** in complex with C_{10}-P and Ca^{2+} also points to an explanation for the notable enhancement in the biological activity of lipopeptides **6** and **7** relative to **5**. As described above, the peptide macrocycles of **6** and **7** contain a Val residue at position 10 while in lipopeptide **5** the same position is filled by a slightly bulkier Ile residue. This subtle structural difference results in an 8-fold increase in the activity for **6** and **7** relative to **5**. Careful inspection of the trimer of dimers formed by both peptides **5** and **7** reveals a hydrophobic pocket at center of the trimer where the side chains of Val^{10}/Ile^{10} meet (Fig. 3C). This finding suggests that the Val^{10} side chain in compounds **6** and **7** (and as found naturally in the amphomycin/friulimicin class) allows for optimal packing of the trimer, enhancing the interaction with the C_{55}-P bacterial target, and drives the antibiotic activity observed. By comparison, the slightly bulkier Ile^{10} side chain in compound **5** may impinge upon the precise steric requirements of the hydrophobic pocket formed at the trimer interface and in doing so destabilize the interaction with C_{55}-P resulting in reduced antibacterial activity. Taken together, these findings provide new insight into the mechanism of action of amphomycin/friulimicin class of calcium dependent antibiotics and how they compare to the laspartomycin family.

Live cell imaging

To gain additional insights into the impact of these lipopeptide antibiotics on live bacteria, laspartomycin C, lipopeptide **6**, and daptomycin were evaluated in a set of comprehensive mode-of-action studies conducted using the model organism *Bacillus subtilis* and imaged using fluorescence light microscopy (8, 23). These studies reveal that laspartomycin C and **6** both interfere with bacterial membrane integrity by delocalizing key membrane proteins

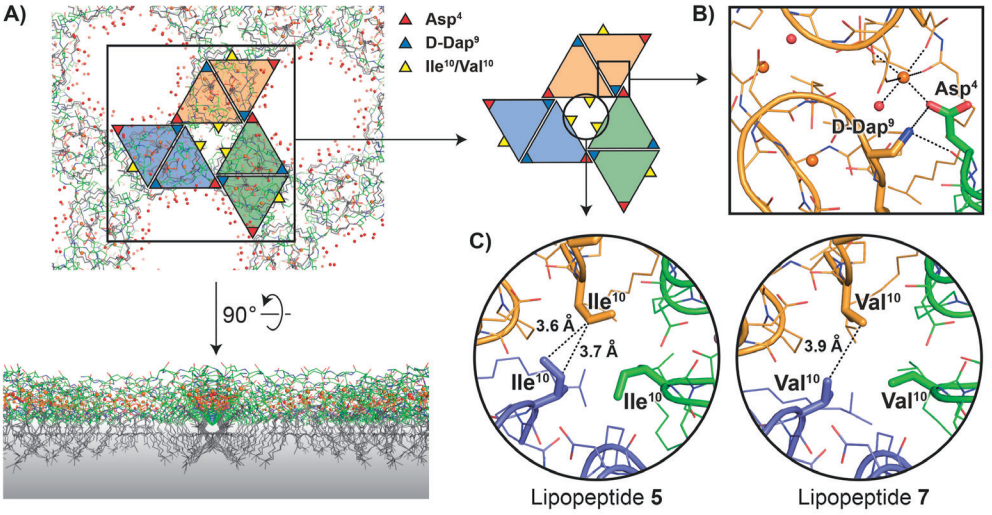


Figure 3. A) In the crystal packing lipopeptides **5** and **7** adopt higher-ordered structures not observed with laspartomycin C. In this arrangement the lipids of both the lipopeptides and C_{10}-P are oriented in the same direction while the peptide macrocycles interact to form a repeating trimer of dimers motif as indicated by the colored triangles. A proposed orientation of the multimeric assembly in the bacterial membrane (indicated with a grey gradient) is shown. **B)** Interactions between the D-Dap^9 and Asp^4 residues present in lipopeptides **5** and **7**, but absent in laspartomycin C, stabilize the trimer of dimers. **C)** The presence of a hydrophobic core formed at the center of the trimer of dimers motif suggests that the side chain of Val^{10} present in the biologically more active lipopeptide **7** more optimally suits the steric requirements of this motif vs. the slightly bulkier Ile^{10} in lipopeptide **5.** For clarity only major conformations are shown.

and/or interfering with lipid organization in a manner that is distinct from that observed for daptomycin. An extensive bacterial cytological profiling study previously showed that daptomycin causes the clustering of 'fluid lipids', i.e. lipids with short, branched or unsaturated fatty acyl chains, into large aggregates, lowering the membrane fluidity outside these aggregates (8). This has a severe effect on the binding of peripheral membrane proteins with essential functions, including the *N*-acetylglucosamine transferase MurG responsible for the last synthesis step of the peptidoglycan precursor lipid II. To further compare the effect of laspartomycin C with that of daptomycin, we performed bacterial cytological profiling using a broad set of *B. subtilis* reporter strains expressing GFP-tagged proteins involved in DNA replication (DNA polymerase subunit PolC), transcription (RNA polymerase subunit RpoC), translation (ribosome subunit RpsB), ATP synthesis (F1F0-ATPase subunit AtpA), cell division (FtsZ), cell wall synthesis coordination (MreB), cell division regulation (MinD) and peptidoglycan synthesis (MurG). The reporters MreB, MinD and MurG are all peripheral membrane proteins. Cells were incubated with the lipopeptide antibiotics at 2x MIC and observed by fluorescent light microscopy after 10 min and 30 min incubation. The reporter strains indicated that neither DNA, RNA and protein synthesis, nor cell division and the localization of F1F0-ATPase ATP were affected by laspartomycin

C or compound **6** (Fig. 4), findings that are in keeping with those previously observed for daptomycin (see Supplemental Fig. S6) (8). Likewise, the delocalization of MreB by laspartomycin C and compound **6** (see Supplemental Fig. S7) is similar to the effect seen with daptomycin (8). A notable difference, however, was the finding that the localization of MinD was unaffected by laspartomycin C and **6**, whereas this protein rapidly detaches from the membrane when treated with daptomycin (see Supplemental Fig. S8). Another clear difference is the delocalization of MurG, which detaches from the cell membrane in the presence of daptomycin, whereas laspartomycin C and compound **6** appear to dissolve the large MurG clusters so that the protein diffuses along the cell membrane (Fig. 4).

The differences observed for laspartomycin C and lipopeptide **6** vs daptomycin may be explained by the multifaceted mechanism of action attributed to daptomycin. Recent investigations have revealed that in the presence of phosphatidylglycerol, daptomycin can interact with C_{55}-P, C_{55}-PP, and the peptidoglycan precursor lipid II (9). As a result, the insertion of daptomycin in the cell membrane not only affects lipid II synthesis, but also causes a dramatic rearrangement of lipids in the cell membrane resulting in the detachment of peripheral membrane proteins, including MinD and MurG (8). Conversely, the binding of laspartomycin C to C_{55}-P is not facilitated by phosphatidylglycerol or any other phospholipids (11) and is therefore likely to more specifically interfere with lipid II synthesis and not with the distribution of other phospholipids in the membrane. However, the activity of proteins that rely on the C_{55}-P precursor, including MurG and MreB, will still be affected, explaining the dissolution of MurG clusters and the delocalization of MreB observed in our studies. This rationale is also in line with a recent report revealing MreB membrane association to be dependent on the presence of lipid-linked peptidoglycan precursors and that when such precursors are depleted, MreB filaments disassemble and peptidoglycan synthesis is disrupted (24).

Conclusion

In summary, a number of novel laspartomycin C variants were synthesized to probe the effects associated with structural differences specific to the friulimicin/amphomycin class of CDAs. The antibacterial activities measured and the high-resolution crystal structures obtained for these lipopeptide antibiotics reveal a previously unknown interplay between the side chains of residues at positions 4, 9, and 10 in the peptide macrocycle. Interestingly, the amino acid side chains present at these positions in the friulimicin/amphomycin class contribute to the formation of higher-order assemblies when in complex with Ca^{2+} and the bacterial target, an effect not seen for the other well-characterized C_{55}-P binding CDA laspartomycin C. In addition, live cell imaging studies reveal subtle differences in the activity of laspartomycin C and daptomycin. Compared to daptomycin, laspartomycin C

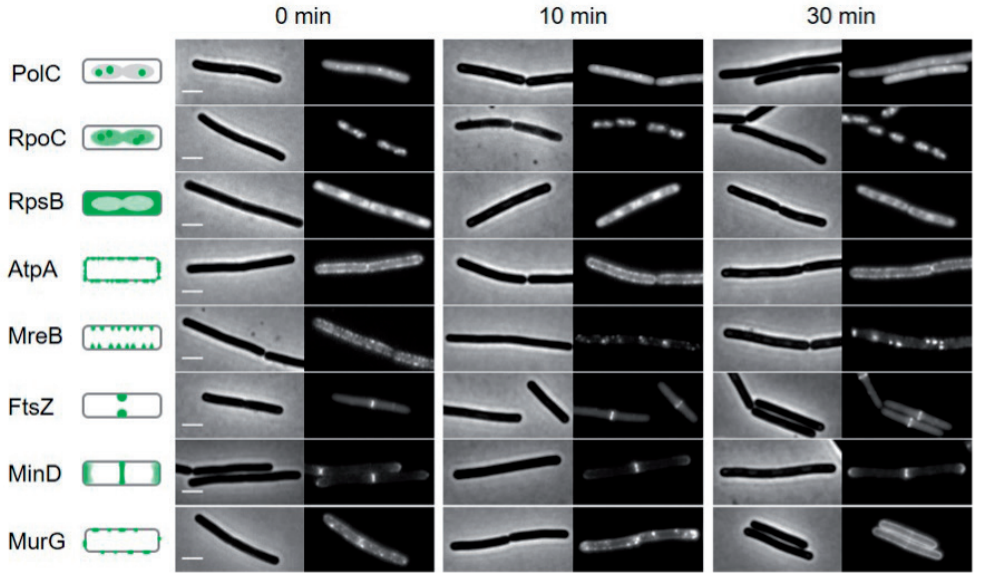


Figure 4. Bacterial cytological profiling analysis of Laspartomycin C. The GFP-tagged marker proteins represents the following cellular activities: DNA polymerization (PolC), RNA polymerization (RpoC), protein synthesis (RpsB), F0F1 ATPase (AtpA), lateral cell wall synthesis regulation (MreB), cell division (FtsZ), cell division regulation (MinD) and peptidoglycan precursor synthesis (MurG). Left panels schematically show the normal localization patterns of the different GFP fusions. Strains were grown in LB medium supplemented with 2 mM $CaCl_2$ at 30 °C. 2x MIC concentration was added (0 min) and samples for microscopy were taken after 10- and 30-min incubation, respectively. Scale bars indicate 2 µm. Lipopeptide **6** also showed a similar bacterial cytological profile (Supplemental Fig. S5).

and the other C_{55}-P targeting lipopeptides here studied appear to have a more narrowly defined range of cellular targets. Particularly notable is the ability of laspartomycin C to dissolve large clusters of MurG along the cell membrane, an effect not seen in daptomycin. Taken together, our results provide new insights into the mechanisms of action associated with the C_{55}-P-targeting subfamily of CDAs and expand our current understanding of this promising class of lipopeptide antibiotics.

Acknowledgements

We thank Paolo Innoccenti for performing HRMS analyses and Fons Lefeber for assistance in acquiring NMR spectra. Professor Tanja Schneider is kindly thanked for providing an authentic sample of friulimicin B. We thank the staff at beamlines I04-1 and I04 of the Diamond Light Source (Harwell, United Kingdom) for diffraction data collection. Financial support provided by the Netherlands Organization for Scientific Research (NWO PhD

grant to T.M.W.) and the European Research Council (ERC consolidator grant to N.I.M, grant agreement no. 725523, and ERC starting grant to B.J.C.J., grant agreement no. 677500).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

T.M.W. synthesized and tested the activity of all peptides, M.R.Z. performed the crystallization and X-ray diffraction experiments, M.R.Z., N.M.P and M.L. solved and refined the crystal structures, T.M.W. and J.K. performed the 2D NMR experiments, L.W.H. and T.S. performed the fluorescence microscopy experiments, N.I.M. and B.J.C.J. supervised the project. M.R.Z. wrote the first draft of the crystallography section (including figures). The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. †These authors contributed equally.

References

1. P. Beyer, S. Paulin, Priority pathogens and the antibiotic pipeline: An update. *Bull. World Health Organ.* **98**, 151 (2020).
2. M. F. Chellat, L. Raguž, R. Riedl, Targeting Antibiotic Resistance. *Angew. Chemie–Int. Ed.* **55**, 6600–6626 (2016).
3. M. N. Gwynn, A. Portnoy, S. F. Rittenhouse, D. J. Payne, Challenges of antibacterial discovery revisited. *Ann. N. Y. Acad. Sci.* **1213**, 5–19 (2010).
4. R. Sauermann, M. Rothenburger, W. Graninger, C. Joukhadar, Daptomycin: a review 4 years after first approval. *Pharmacology* **81**, 79–91 (2008).
5. R. M. Humphries, S. Pollett, G. Sakoulas, A current perspective on daptomycin for the clinical microbiologist. *Clin. Microbiol. Rev.* **26**, 759–780 (2013).
6. P. T. Hart, *et al.*, A combined solid- and solution-phase approach provides convenient access to analogues of the calcium-dependent lipopeptide antibiotics. *Org. Biomol. Chem.* **12**, 913–918 (2014).
7. S. D. Taylor, M. Palmer, The action mechanism of daptomycin. *Bioorganic Med. Chem.* **24**, 6253–6268 (2016).
8. A. Müller, *et al.*, Daptomycin inhibits cell envelope synthesis by interfering with fluid membrane microdomains. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E7077–E7086 (2016).
9. F. Grein, *et al.*, Ca2+-Daptomycin targets cell wall biosynthesis by forming a tripartite complex with undecaprenyl-coupled intermediates and membrane lipids. *Nat. Commun.* **11**, 1–11 (2020).
10. T. Schneider, *et al.*, The lipopeptide antibiotic friulimicin B inhibits cell wall biosynthesis through complex formation with bactoprenol phosphate. *Antimicrob. Agents Chemother.* **53**, 1610–1618 (2009).
11. L. H. J. Kleijn, *et al.*, Total Synthesis of Laspartomycin C and Characterization of Its Antibacterial Mechanism of Action. *J. Med. Chem.* **59**, 3569–3574 (2016).
12. L. H. J. Kleijn, *et al.*, A High-Resolution Crystal Structure that Reveals Molecular Details of Target Recognition by the Calcium-Dependent Lipopeptide Antibiotic Laspartomycin C. *Angew. Chemie–Int. Ed.* **56**, 16546–16549 (2017).
13. A. Diehl, T. M. Wood, S. Gebhard, N. I. Martin, G. Fritz, The cell envelope stress response of bacillus subtilis towards laspartomycin C. *Antibiotics* **9**, 1–16 (2020).
14. G. Manat, *et al.*, Deciphering the metabolism of undecaprenyl-phosphate: The bacterial cell-wall unit carrier at the membrane frontier. *Microb. Drug Resist.* **20**, 199–214 (2014).
15. M. El Ghachi, *et al.*, Crystal structure of undecaprenyl-pyrophosphate phosphatase and its role in peptidoglycan biosynthesis. *Nat. Commun.* **9**, 1078 (2018).
16. S. D. Workman, L. J. Worrall, N. C. J. Strynadka, Crystal structure of an intramembranal phosphatase central to bacterial cell-wall peptidoglycan biosynthesis and lipid recycling. *Nat. Commun.* **9**, 1159 (2018).
17. T. M. Wood, N. I. Martin, The calcium-dependent lipopeptide antibiotics: Structure, mechanism, & medicinal chemistry. *Medchemcomm* **10**, 634–646 (2019).

18. M. Strieker, M. A. Marahiel, The structural diversity of acidic lipopeptide antibiotics. *ChemBioChem* **10**, 607–616 (2009).
19. B. M. Hover, *et al.*, Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat. Microbiol.* **3**, 415–422 (2018).
20. Z. Sun, *et al.*, Total Synthesis of Malacidin A by β-Hydroxyaspartic Acid Ligation-Mediated Cyclization and Absolute Structure Establishment. *Angew. Chem. Int. Ed. Engl.* **59**, 19868–19872 (2020).
21. L. Corcilius, N. T. Elias, J. L. Ochoa, R. G. Linington, R. J. Payne, Total Synthesis of Glycinocins A-C. *J. Org. Chem.* **82**, 12778–12785 (2017).
22. T. M. Wood, K. Bertheussen, N. I. Martin, The contribution of achiral residues in the laspartomycin family of calcium-dependent lipopeptide antibiotics. *Org. Biomol. Chem.* **18**, 514–517 (2020).
23. M. Wenzel, *et al.*, A flat embedding method for transmission electron microscopy reveals an unknown mechanism of tetracycline. *Commun. Biol.* **4**, 306 (2021).
24. K. Schirner, *et al.*, Lipid-linked cell wall precursors regulate membrane association of bacterial actin MreB. *Nat. Chem. Biol.* **11**, 38–45 (2015).

Supplementary Information for

Mechanistic insights into the C55-P targeting lipopeptide antibiotics revealed by structure-activity studies and high-resolution crystal structures

Thomas M. Wood, Matthieu R. Zeronian, Hanieh K. Abedian, Aidan V. Johnson, Nicolas M. Pearce, Martin Lutz, Johan Kemmink, Tjalling Seirsma, Leendert W. Hamoen, Bert J. C. Janssen, Nathaniel I. Martin

Email: b.j.c.janssen@uu.nl (B.J.C.J.), n.i.martin@biology.leidenuniv.nl (N.I.M.)

SI Table of Contents

Reagents and General Methods 111

Instrumentation for Compound Characterization 111

Solid Phase Peptide Synthesis 112

Abbreviations: 113

Antibacterial Assays 114

UDP-MurNAc-pentapeptide Accumulation Assay 115

Characterization of Synthetic Peptides 118

Crystallization and data collection 145

Structure solution and refinement 145

Bacterial cytological profiling 148

SI References 153

Reagents and General Methods

All reagents employed were of American Chemical Society (ACS) grade or finer and were used without further purification unless otherwise stated. D-amino acids and 2-chlorotrityl resin was obtained from Iris Biotech GmbH, Egg PG and 0:6 PA was obtained from INstruchemie BV. C_{10}-P lithium salt was obtained from Sigma Aldrich and lyophilized from warm ^{t}BuOH:H_2O (1:1) to obtain a white powder with increased aqueous solubility.

Instrumentation for Compound Characterization

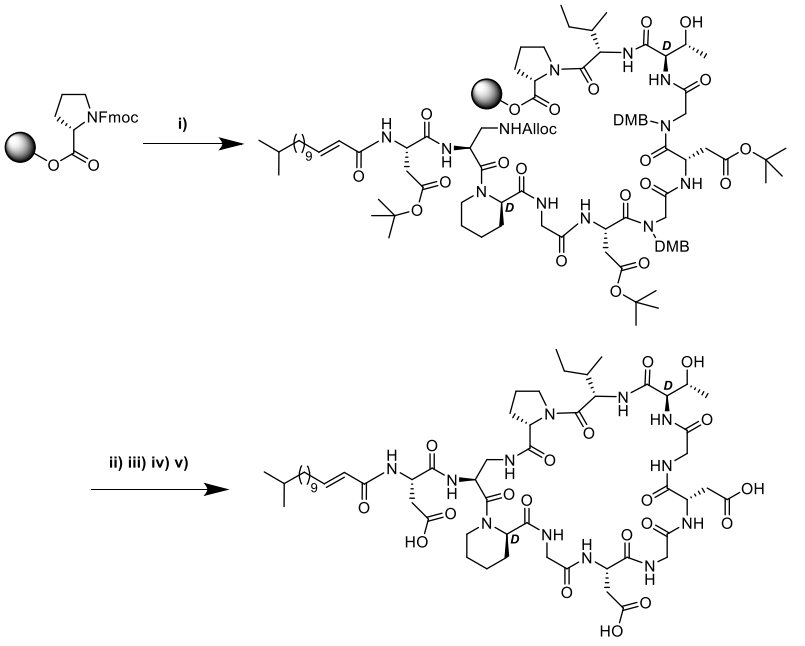
2D NMR experiments were performed on a 850 MHz instrument. HSQC, TOCSY and NOESY spectra were recorded for all peptides (5 mM in $DMSO_{d6}$) and the parent compound laspartomycin C matched pervious recorded spectra reported by our group.

HRMS analysis was performed on a Shimadzu Nexera X2 UHPLC system with a Waters Acquity HSS C18 column (2.1 × 100 mm, 1.8 µm) at 30 °C and equipped with a diode array detector. The following solvent system, at a flow rate of 0.5 mL/min, was used: solvent A, 0.1 % formic acid in water; solvent B, 0.1 % formic acid in acetonitrile. Gradient elution was as follows: 95:5 (A/B) for 1 min, 95:5 to 15:85 (A/B) over 6 min, 15:85 to 0:100 (A/B) over 1 min, 0:100 (A/B) for 3 min, then reversion back to 95:5 (A/B) for 3 min. This system was connected to a Shimadzu 9030 QTOF mass spectrometer (ESI ionisation) calibrated internally with Agilent's API-TOF reference mass solution kit (5.0 mM purine, 100.0 mM ammonium trifluoroacetate and 2.5 mM hexakis(1H,1H,3H-tetrafluoropropoxy)phosphazine) diluted to achieve a mass count of 10000.

Purity of the peptides was confirmed to be ≥ 95% by analytical RP-HPLC using a Shimadzu Prominence-i LC-2030 system with a Dr. Maisch ReproSil Gold 120 C18 column (4.6 × 250 mm, 5 µm) at 30 °C and equipped with a UV detector monitoring at 214 nm. The following solvent system, at a flow rate of 1 mL/min, was used: solvent A, 0.1 % TFA in water/acetonitrile, 95/5; solvent B, 0.1 % TFA in water/acetonitrile, 5/95. Gradient elution was as follows: 95:5 (A/B) for 2 min, 95:5 to 0:100 (A/B) over 55 min, 0:100 (A/B) for 2 min, then reversion back to 95:5 (A/B) over 1 min, 95:5 (A/B) for 2 min.

The compounds were purified via preparative HPLC using a BESTA-Technik system with a Dr. Maisch Reprosil Gold 120 C18 column (25 × 250 mm, 10 µm) and equipped with a ECOM Flash UV detector monitoring at 214 nm. The following solvent system, at a flow rate of 12 mL/min, was used: solvent A, 0.1 % TFA in water/acetonitrile 95/5; solvent B, 0.1 % TFA in water/acetonitrile 5/95. Gradient elution was as follows: 95:5 (A/B) for 2 min, 95:5 to 0:100 (A/B) over 55 min, 0:100 (A/B) for 2 min, then reversion back to 95:5 (A/B) over 1 min, 95:5 (A/B) for 2 min.

General Procedure for the Preparation of Laspartomycin C and Other Analogues



Scheme S1 (i) Fmoc SPPS; (ii) $Pd[(C_6H_5)_3P]_4$, $C_6H_5SiH_3$, CH_2Cl_2, 1 h; (iii) HFIP, CH_2Cl_2, 1 h; (iv) BOP, DIPEA, CH_2Cl_2, 16 h; (v) TFA, TIS, H_2O, 1 h

Solid Phase Peptide Synthesis

Chlorotrityl resin (5.0 g, 1.60 mmol/g) was loaded with Fmoc-Pro-OH. Resin loading was determined to be 0.41-0.62 $mmol.g^{-1}$. Linear peptide encompasing Pro11 to Asp1 were assembled manualy via standard Fmoc solid-phase peptide synthesis (SPPS) (resin bound AA:Fmoc-AA:BOP:DiPEA, 1:4:4:8 molar eq.) on a 0.1 mmol scale. DMF was used as solvent and Fmoc deprotections were carried out with piperidine:DMF (1:4 v:v). Amino acid side chains were protected as follows: ^{t}Bu for Asp, Alloc for DAP, and DMB for Gly6 and Gly8. D-*allo*-Thr was introduced without side chain protection. Following coupling and Fmoc deprotection of Asp1, N-terminal acylation was achieved by coupling (*E*)-13-methyltetradec-2-enoic acid using the same coupling conditions used for SPPS. The resin-bound, Alloc protected intermediate was next washed with CH_2Cl_2 and treated with $Pd(PPh_3)_4$ (30mg, 0.03 mmol) and $PhSiH_3$ (0.30 mL, 3.0 mmol) in CH_2Cl_2 (ca. 7 mL) under argon for 1 hour. The resin was subsequently washed with CH_2Cl_2 (5x10 mL), followed by a solution of diethyldithiocarbamic acid trihydrate sodium salt (5 mg mL-1 in DMF, 5x10 mL), and DMF (5x10 mL). The resin was treated with $(CF_3)_2CHOH$:CH_2Cl_2 (1:4, 10 mL) for 1 hour and rinsed with additional $(CF_3)_2CHOH$:CH_2Cl_2 and CH_2Cl_2. The combined washings were then

evaporated to yield the linear protected peptide with free C- and N-termini. The residue was dissolved in CH_2Cl_2 (150 mL) and treated with BOP (0.22 g, 0.5 mmol) and DiPEA (0.17 mL, 1.0 mmol) and the solution was stirred overnight after which TLC indicated complete cyclization. The reaction mixture was concentrated and directly treated with TFA:TIS:H_2O (95:2.5:2.5, 10 mL) for 90 minutes. The reaction mixture was added to MTBE:hexanes (1:1) and the resulting precipitate washed once more with MTBE:hexanes (1:1). The crude cyclic peptide was lyophilized from ^{t}BuOH:H_2O (1:1) and purified with reverse phase HPLC. Pure fractions were pooled and lyophilized to yield the desired cyclic lipopeptide products in >95% purity as white powders, typically in 10-45 mg quantities (4.2-30 % yield based on resin loading).

Abbreviations:

AA	Amino acid
Alloc	Allyloxycarbonyl
^{t}Bu	tert-butyl
^{t}BuOH	tert-butanol
BOP	(benzotriazole-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate
Dap	2,3-Diaminopropionic acid
DiPEA	*N,N*-diisopropylethylamine
DMB	2,4-dimethoxybenzyl
DMF	*N,N*-dimethylformamide
Fmoc	Fluorenylmethyloxycarbonyl
MTBE	Methyl tert-butyl ether
TFA	Trifluoroacetic acid
TIS	Triisopropylsilane

Antibacterial Assays

Minimum inhibitory concentrations (MICs) were determined by broth microdilution according to CLSI guidelines (1). Blood agar plates were inoculated with glycerol stocks of MRSA and *S. simulans* 22 followed by incubation for 16 hours at 37°C and 30°C respectively. Cation adjusted Mueller-Hinton broth (MHB) containing 10 mg L^{-1} Mg^{2+} was inoculated with individual colonies of MRSA and *S. simulans* and incubated for 16 hours at 220 RPM. The peptides were dissolved in MHB (10 mg L^{-1} Mg^{2+}) and serially diluted on

Table S1. MIC values (μg mL^{-1}) against MRSA and *S. simulans* at various Ca^{2+} concentrations.

	MRSA USA 300				*S. simulans 22*			
Compound	**0 mM**	**2.5 mM**	**5 mM**	**10 mM**	**0 mM**	**2.5 mM**	**5 mM**	**10 mM**
1 (LaspC)	>128	8	4	2	>128	8	4	2
2	>128	8	4	4	>128	16	8	8
3	>128	64	32	16	>128	>128	64	32
4	>128	16	8	4	>128	16	8	4
5	>128	32	16	8	>128	16	8	8
6	>128	4	2	1	>128	8	4	2
7	>128	8	4	2	>128	4	4	1
8	>128	8	4	4	>128	8	8	4
9	>128	32	16	8	>128	32	16	8
Friulimicin	>128	4	2	1-2	>128	2	1	1
Daptomycin	>128	1	0.5	0.25	>128	0.063	0.031	0.031

Table S2. MIC values (μg mL^{-1}) against VRSA and VISA at various Ca^{2+} concentrations.

	BR-VRSA				VISA LIM2			
Compound	**0 mM**	**2.5 mM**	**5 mM**	**10 mM**	**0 mM**	**2.5 mM**	**5 mM**	**10 mM**
Laspartomycin	>128	4	4	2	>128	4	4	2
6	>128	4	2	1	>128	4	2	1
7	>128	2	1	0.5	>128	4	2	2
Friulimicin	>128	4	2	1	>128	4	4	2
Daptomycin	>128	0.5	0.25	0.13	>128	0.25	0.13	0.13

Table S3. MIC values (μg mL^{-1}) against *E. faeceum* E7128 (daptomycin resistant) and VRE 155 at various Ca^{2+} concentrations.

	E7128				VRE 155			
Compound	**0 mM**	**2.5 mM**	**5 mM**	**10 mM**	**0 mM**	**2.5 mM**	**5 mM**	**10 mM**
Laspartomycin	>128	16	8	8	>128	4	4	2
6	>128	8	4	2	>128	2	1	0.5
7	>128	8	2	2	>128	1	1	0.5
Friulimicin	>128	8	4	2	>128	2	1	0.5
Daptomycin	>128	4	4	2	>128	0.25	0.25	0.13

polypropylene microtiter plates with a volume of 50 µL per well. Inoculated MHB ($2x10^5$ CFU.mL^{-1}) containing 10 mg L^{-1} Mg^{2+} and varying concentrations of Ca^{2+} was added to reach a total volume of 100 µL per well. The microtiter plates were sealed with an adhesive membrane and after 16 hours of incubation at 37°C or 30°C and 220 RPM the wells were visually inspected for bacterial growth. All reported MIC values result from three or more measurements. The following strains were obtained from BEI Resources, NIAID, NIH: *S. aureus* Strain 880 (BR-VRSA), NR-49120; *S. aureus* Strain LIM 2 (VISA), NR-45881.

UDP-MurNAc-pentapeptide Accumulation Assay

MRSA USA 300 was grown until OD_{600} = 0.5 in TSB supplemented with $CaCl_2$ (5.0 mM). Chloramphenicol (130 µg mL^{-1}) was added and after incubation for 15 minutes at 37°C, the culture was divided in 5 mL aliquots. Antibiotics were added at 10xMIC and one aliquot remained untreated. After 60 minutes, cells were separated from the medium and extracted with boiling d-H_2O (1 mL) for 15 minutes. The suspensions were spun down and the supernatant was lyophilized. The resulting material was analyzed by HPLC applying a gradient from 100% eluent A (50 mM $NaHCO_3$:5 mM Et_3N, pH = 8.3) to 75% eluent A over 15 minutes using a C18 column (eluent B: MeOH). Formation of UDP-MurNAc-pentapeptide was confirmed by comparison with authentic material by HPLC, and LC-MS analysis applying the same gradient with an adjusted eluent A (50 mM NH_4HCO_3:5 mM Et_3N, pH = 8.3).

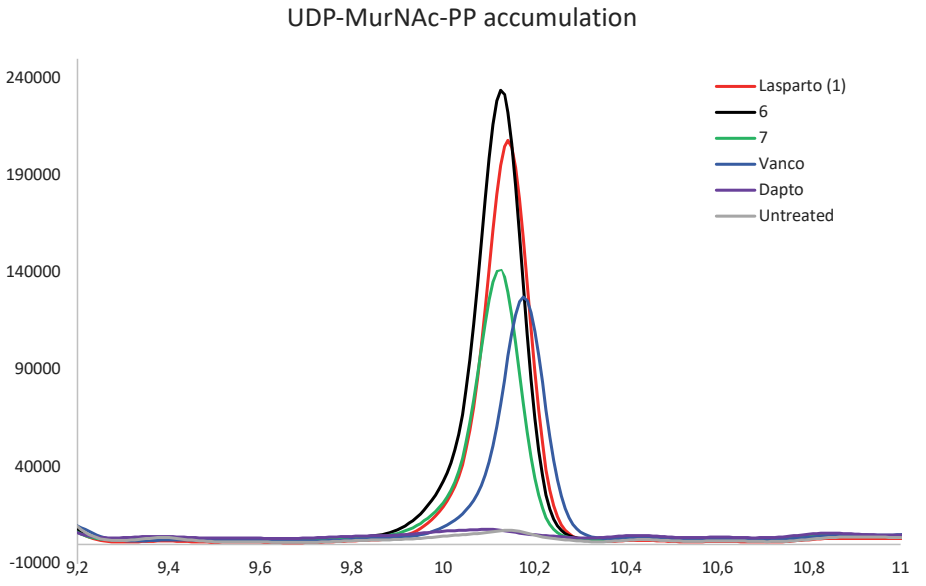
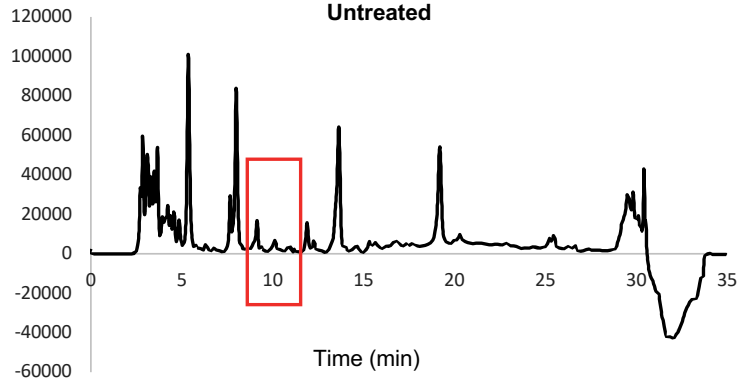
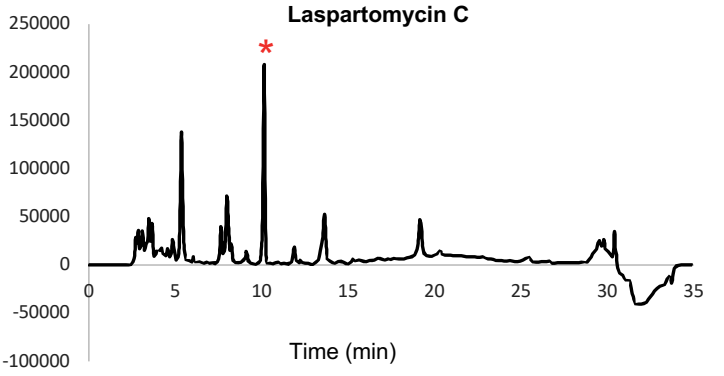
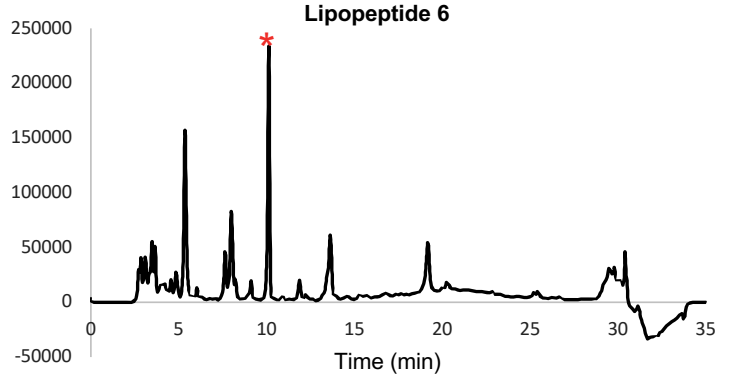


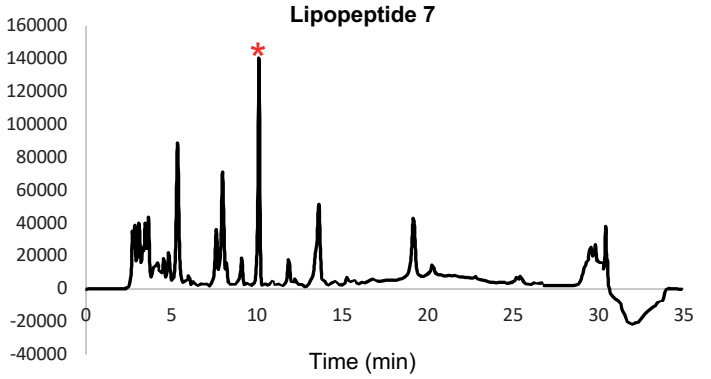
Figure S1. Analytical HPLC trace (zoom) for UDP-MurNAc-pentapeptide accumulation assay. Treatment of MRSA USA 300 with laspartomycin C (**1**), and lipopepeptides **6** and **7** results in accumulation of UDP-MurNAc-pentapeptide, an effect not observed with daptomycin. Vancomycin included as positive control.

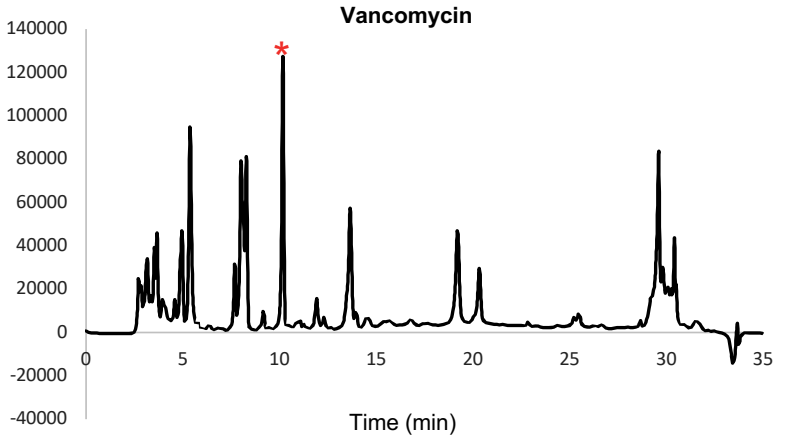
Full analytical HPLC traces for UDP-MurNAc-pentapeptide accumulation assays

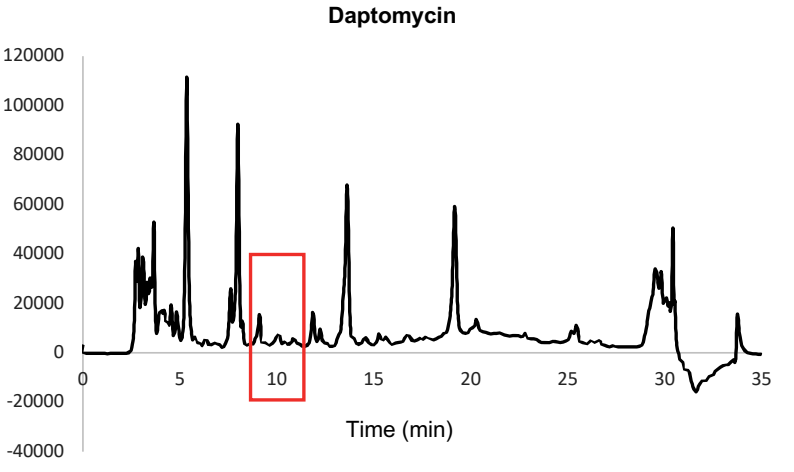






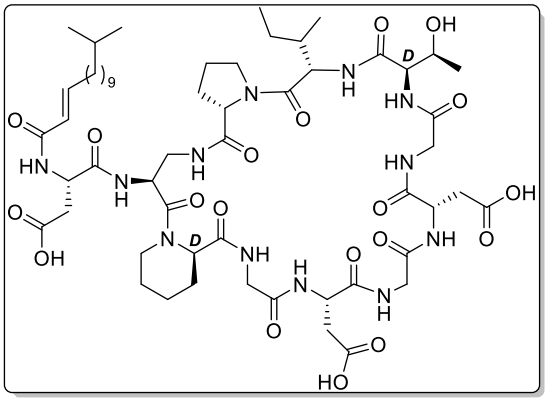

Lipopeptide 7
160000
140000
120000
100000
80000
60000
40000
20000
0
-20000
-40000
0
5
10
15
20
25
30
35
Time (min)


Vancomycin
140000
120000
100000
80000
60000
40000
20000
0
-20000
-40000
0
5
10
15
20
25
30
35
Time (min)


Daptomycin
120000
100000
80000
60000
40000
20000
0
-20000
-40000
0
5
10
15
20
25
30
35
Time (min)

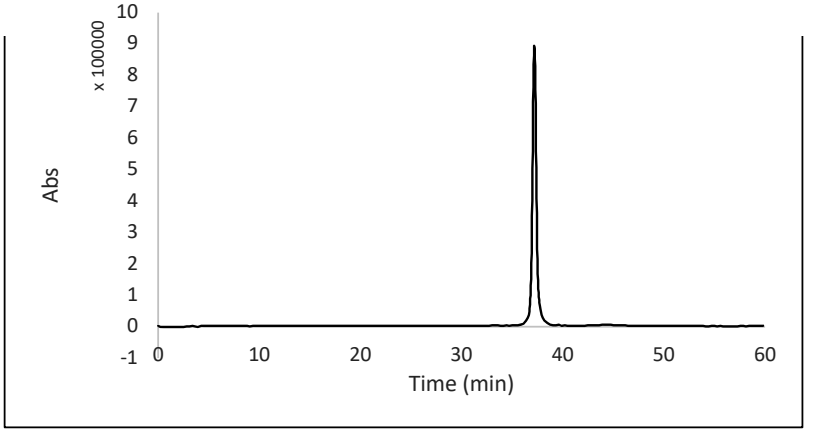
Characterization of Synthetic Peptides

Laspartomycin C (1)



Yield: 47.3 mg (18.7 umol, 15.3%)
HR-MS [M+H$^+$]: Calc.: 1247.6479, found: 1247.6522

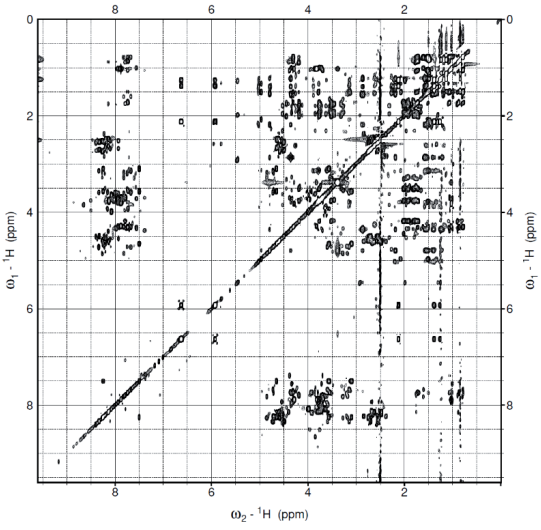
Analytical HPLC



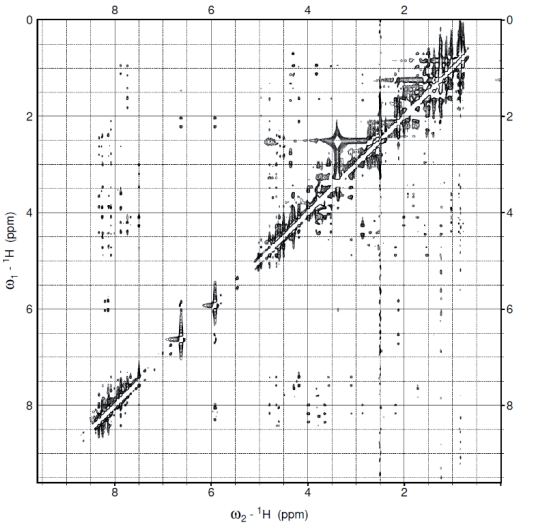
Laspartomycin C (1) NMR Chemical Shifts

Residue	NH	$H_\alpha(C_\alpha)$	Sidechain
Tail	-	5.92 (123.7)	$C_\beta H$ (6.63, 143.8), $C_\gamma H_2$ (2.12, 31.1), $C_\delta H_2$ (1.37, 27.5), $C_\varepsilon H_2$-$C_\iota H_2$ (1.23-1.27, 28.7), $C_\kappa H_2$ (1.23, 26.5), $C_\lambda H_2$ (1.12, 38.0), $C_\mu H$ (1.49, 27.1), $2C_\nu H_3$ (0.84, 22.4)
Asp-1	8.14	4.61 (48.9)	$C_\beta H_2$ (2.63/2.50, 35.9)
Dap-2	8.25	4.66 (48.5)	$C_\beta H_2$ (3.56/3.10, 39.5)
D-Pip-3	-	4.80 (55.9)	$C_\beta H_2$ (2.18/1.53, 26.4), $C_\gamma H_2$ (1.55/1.39, 20.1), $C_\delta H_2$ (1.51/1.22, 24.1), $C_\varepsilon H_2$ (4.35/2.86, 39.6)
Gly-4	8.08	4.00/3.65 (41.9)	-
Asp-5	8.25	4.61 (49.7)	$C_\beta H_2$ (2.74/2.52, 35.8)
Gly-6	8.13	3.76 (41.9)	-
Asp-7	8.33	4.50 (49.8)	$C_\beta H_2$ (2.70/2.54, 35.6)
Gly-8	7.87	3.80/3.67 (41.9)	-
D-*allo*-Thr-9	7.88	4.28 (58.1)	$C_\beta H$ (3.81, 66.6), $C_\gamma H_3$ (1.02, 19.3)
Ile-10	7.74	4.30 (54.0)	$C_\beta H$ (1.73, 35.8), $C_\gamma H_2$ (1.50/1.07, 24.0), $C_\gamma H_3$ (0.86, 14.5), $C_\delta H_2$ (0.78, 10.3)
Pro-11	-	4.18 (59.4)	$C_\beta H_2$ (2.00/1.74, 29.3), $C_\gamma H_2$ (1.92/1.80, 24.3), $C_\delta H_2$ (3.77/3.50, 46.9)

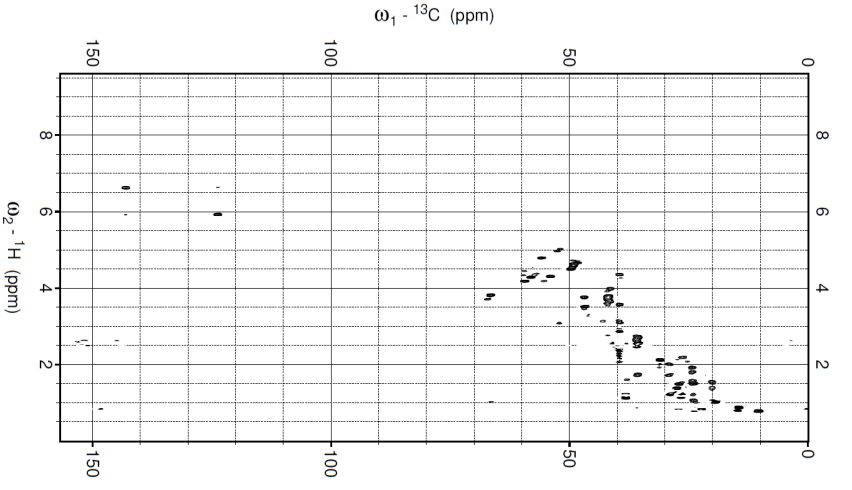
Laspartomycin C (1) 2D NMR Spectra



TOCSY

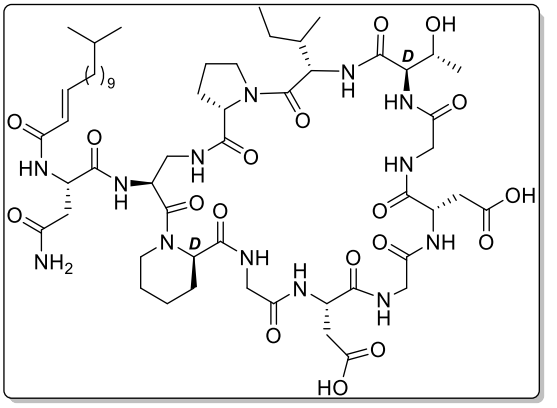


NOESY



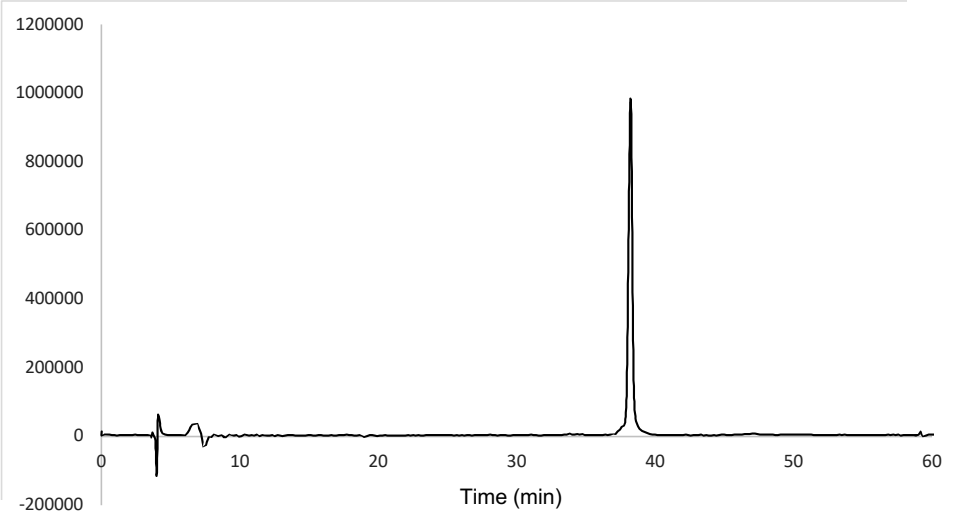
HSQC

Asn_1 containing lipopeptide (2)



Yield: 10.3 mg (8.3 umol, 3.3%)
HR-MS [M+H$^+$]: Calc.: 1246.6639, found: 1246.6605

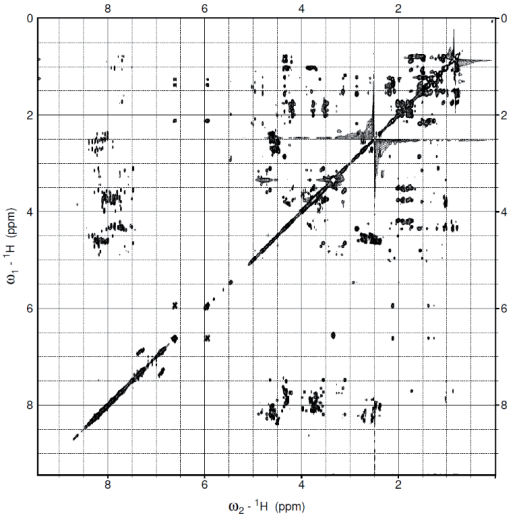
Analytical HPLC



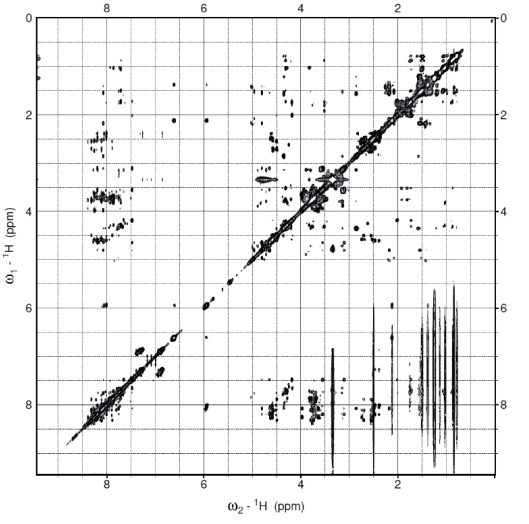
Asn_1 containing lipopeptide (2) NMR Chemical Shifts

Residue	NH	$H_\alpha(C_\alpha)$	Sidechain
Tail	-	5.95 (124.6)	$C_\beta H$ (6.63, 143.6), $C_\gamma H_2$ (2.12, 31.8), $C_\delta H_2$ (1.39, 28.3), $C_\varepsilon H_2$-$C_\iota H_2$ (1.25, 29.5), $C_\kappa H_2$ (1.24, 27.3), $C_\lambda H_2$ (1.14, 39.0), $C_\mu H$ (1.50, 28.0), $2C_\nu H_3$ (0.85, 22.9)
Asn-1	8.00	4.60 (50.1)	$C_\beta H_2$ (2.48/2.40, 37.7)
Dap-2	8.19	4.66 (48.9)	$C_\beta H_2$ (3.57/3.10, 40.3)
D-Pip-3	-	4.81 (56.6)	$C_\beta H_2$ (2.19/1.50, 27.3), $C_\gamma H_2$ (1.56/1.42, 20.8), $C_\delta H_2$ (1.57/1.23, 24.8), $C_\varepsilon H_2$ (4.37/2.87, 40.3)
Gly-4	8.07	3.98/3.66 (42.4)	-
Asp-5	8.26	4.58 (50.0)	$C_\beta H_2$ (2.75/2.53, 36.4)
Gly-6	8.11	3.77 (42.6)	-
Asp-7	8.32	4.50 (50.7)	$C_\beta H_2$ (2.70/2.55, 36.4)
Gly-8	7.89	3.79/3.70 (42.6)	-
D-*allo*-Thr-9	7.86	4.28 (58.9)	$C_\beta H$ (3.83, 67.3), $C_\gamma H_3$ (1.03, 20.5)
Ile-10	7.71	4.32 (54.8)	$C_\beta H$ (1.76, 36.5), $C_\gamma H_2$ (1.52/1.08, 24.8), $C_\gamma H_3$ (0.88, 15.2), $C_\delta H_2$ (0.79, 11.1)
Pro-11	-	4.20 (60.3)	$C_\beta H_2$ (2.02/1.74, 29.8), $C_\gamma H_2$ (1.93/1.83, 25.0), $C_\delta H_2$ (3.77/3.53, 47.7)

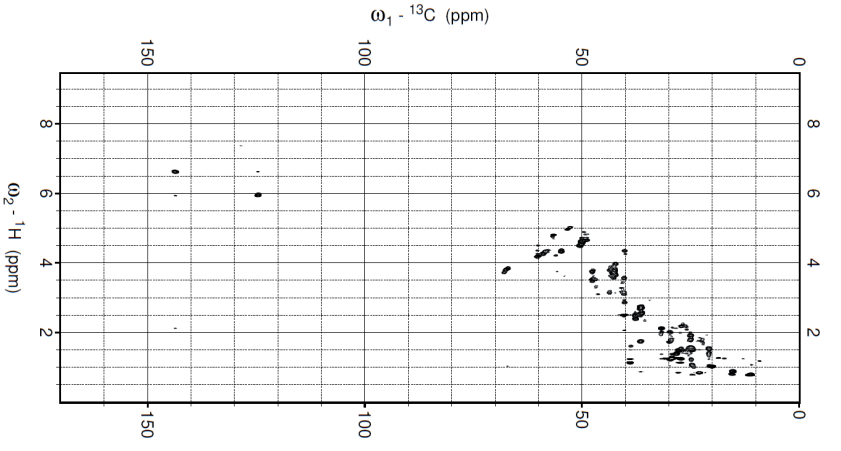
Asn_1 containing lipopeptide (2) 2D NMR Ppectra



TOCSY

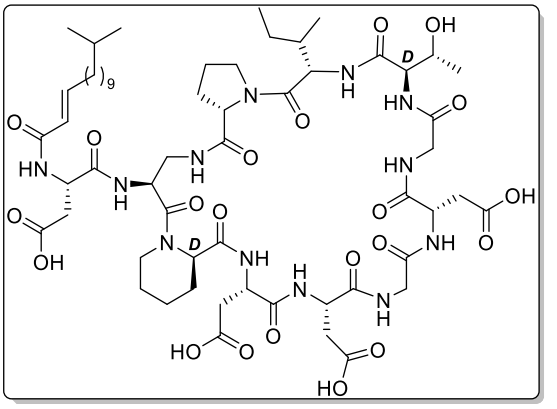


NOESY



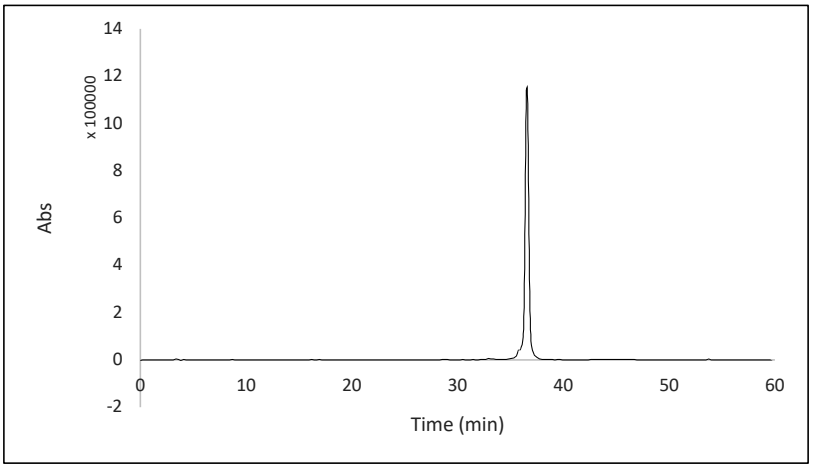
HSQC

Asp_4 containing lipopeptide (3)



Yield: 20.0 mg (15.3 umol, 6.1%)
HR-MS [M+H$^+$]: Calc.: 1305.6578, found: 1305.6583

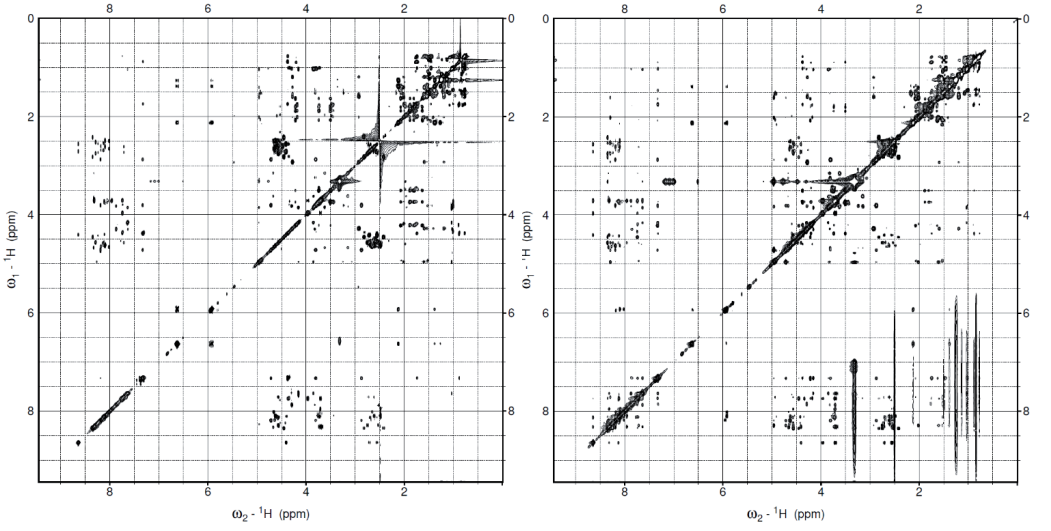
Analytical HPLC



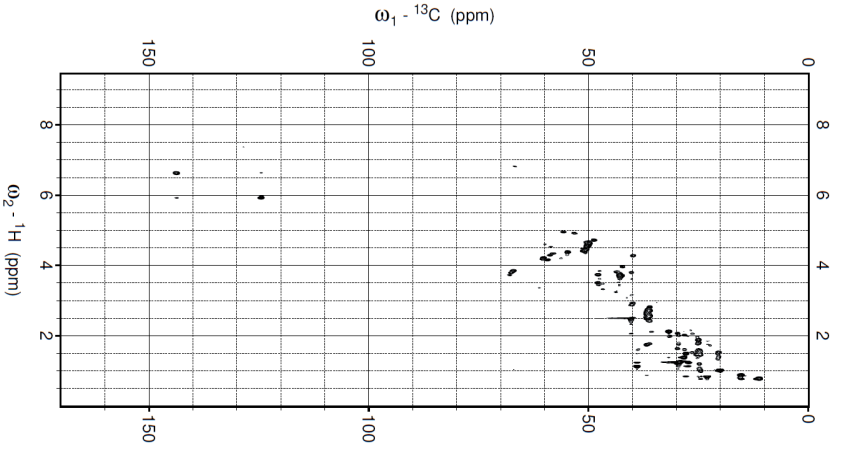
Asp_4 containing lipopeptide (3) NMR Chemical Shifts

Residue	NH	$H_\alpha(C_\alpha)$	Sidechain
Tail	-	5.92 (124.4)	$C_\beta H$ (6.63, 143.8), $C_\gamma H_2$ (2.12, 31.8), $C_\delta H_2$ (1.38, 28.3), $C_\varepsilon H_2$-$C_\iota H_2$ (1.25, 29.5), $C_\kappa H_2$ (1.24, 27.2), $C_\lambda H_2$ (1.13, 39.0), $C_\mu H$ (1.49, 27.9), $2C_\nu H_3$ (0.84, 23.0)
Asp-1	8.10	4.62 (49.9)	$C_\beta H_2$ (2.63/2.50, 36.7)
Dap-2	8.19	4.72 (48.7)	$C_\beta H_2$ (3.80/2.89, 40.3)
D-Pip-3	-	4.95 (55.7)	$C_\beta H_2$ (2.06/1.52, 26.4), $C_\gamma H_2$ (1.53/1.39, 20.4), $C_\delta H_2$ (1.59/1.19, 24.8), $C_\varepsilon H_2$ (4.28/2.93, 39.9)
Asp-4	8.64	4.41 (51.2)	$C_\beta H_2$ (2.71/2.56, 36.3)
Asp-5	8.29	4.56 (50.2)	$C_\beta H_2$ (2.75/2.52, 36.3)
Gly-6	7.96	3.72 (42.7)	-
Asp-7	8.35	4.47 (50.6)	$C_\beta H_2$ (2.81/2.47, 36.1)
Gly-8	7.74	3.96/3.71 (42.3)	-
D-*allo*-Thr-9	7.87	4.29 (58.7)	$C_\beta H$ (3.83, 67.3), $C_\gamma H_3$ (1.03, 20.0)
Ile-10	7.74	4.38 (54.7)	$C_\beta H$ (1.75, 36.8), $C_\gamma H_2$ (1.47/1.07, 24.6), $C_\gamma H_3$ (0.88, 15.3), $C_\delta H_2$ (0.78, 11.1)
Pro-11	-	4.16 (59.3)	$C_\beta H_2$ (2.00/1.75, 29.5), $C_\gamma H_2$ (1.88/1.80, 24.9), $C_\delta H_2$ (3.74/3.50, 47.9)

Asp_4 containing lipopeptide (3) 2D NMR Spectra

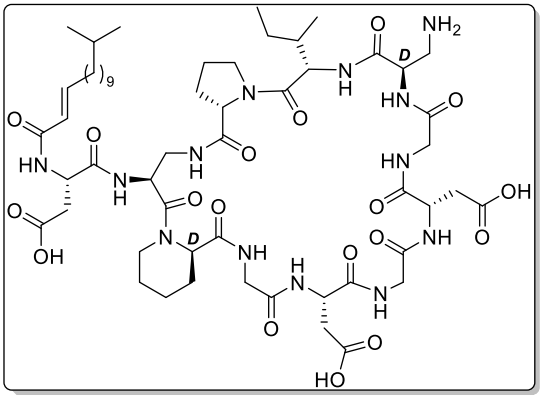


TOCSY **NOESY**



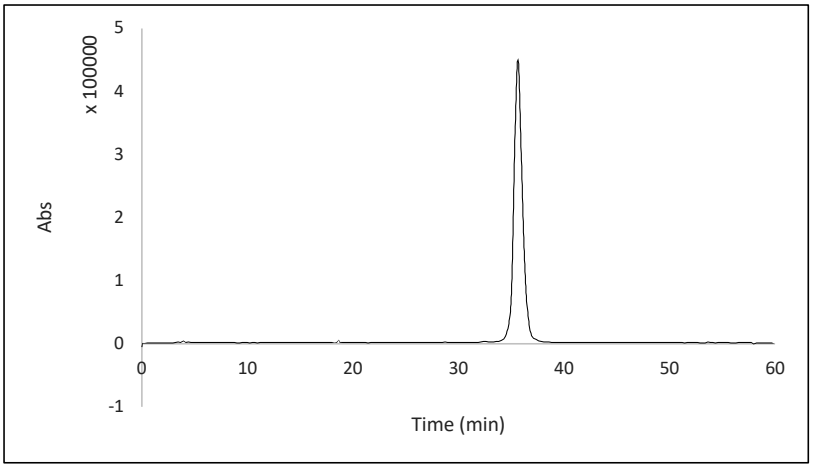
HSQC

d-Dap$_9$ containing lipopeptide (4)



Yield: 11.3 mg (9.1 umol, 3.6%)
HR-MS [M+H$^+$]: Calc.: 1232.6527, found: 1232.6531

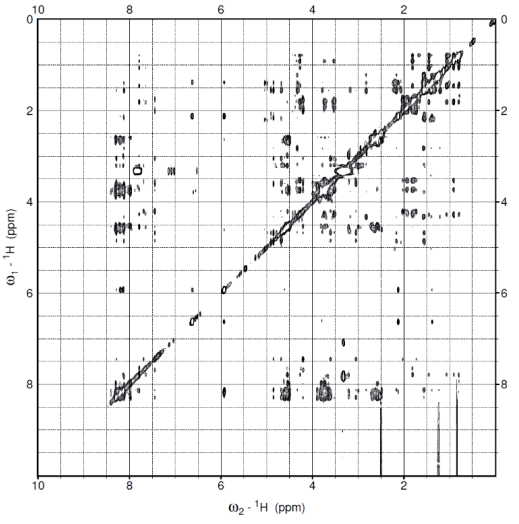
Analytical HPLC



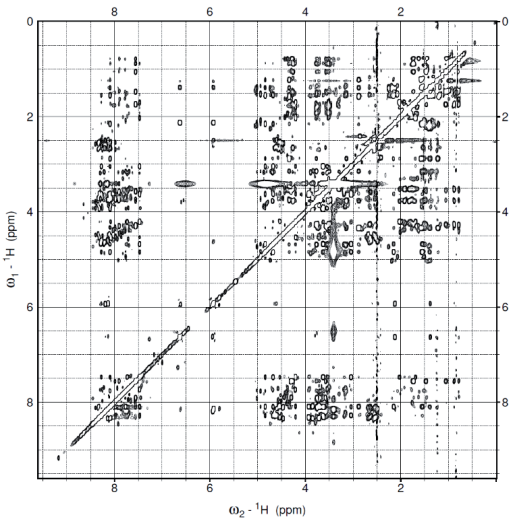
d-Dap$_9$ containing lipopeptide (4) NMR Chemical Shifts

Residue	NH	$H_\alpha(C_\alpha)$	Sidechain
Tail	-	5.93 (124.1)	$C_\beta H$ (6.63, 143.3), $C_\gamma H_2$ (2.12, 31.3), $C_\delta H_2$ (1.38, 27.8), $C_\varepsilon H_2$-$C_\iota H_2$ (1.25, 29.0), $C_\kappa H_2$ (1.23, 26.8), $C_\lambda H_2$ (1.13, 38.5), $C_\mu H$ (1.49, 27.4), $2C_\nu H_3$ (0.84, 22.5)
Asp-1	8.13	4.64 (49.4)	$C_\beta H_2$ (2.63/2.51, 36.1)
Dap-2	8.29	4.67 (48.5)	$C_\beta H_2$ (3.54/3.03, 39.8)
D-Pip-3	-	4.85 (55.9)	$C_\beta H_2$ (2.16/1.56, 28.5), $C_\gamma H_2$ (1.56/1.40, 20.2), $C_\delta H_2$ (1.57/1.21, 24.3), $C_\varepsilon H_2$ (4.34/2.83, 39.6)
Gly-4	8.21	3.80/3.63 (42.0)	-
Asp-5	8.21	4.60 (49.5)	$C_\beta H_2$ (2.73/2.57, 35.9)
Gly-6	8.14	3.78 (42.1)	-
Asp-7	8.30	4.50 (50.0)	$C_\beta H_2$ (2.71/2.48, 35.9)
Gly-8	7.97	3.73 (42.1)	-
D-Dap-9	7.43	4.67 (48.5)	$C_\beta H$ (3.60/3.05, 39.7)
Ile-10	7.45	4.26 (54.8)	$C_\beta H$ (1.81, 35.8), $C_\gamma H_2$ (1.46/1.06, 24.2), $C_\gamma H_3$ (0.91, 14.7), $C_\delta H_2$ (0.80, 10.6)
Pro-11	-	4.20 (59.7)	$C_\beta H_2$ (2.02/1.72, 29.4), $C_\gamma H_2$ (1.93/1.81, 24.5), $C_\delta H_2$ (3.74/3.53, 47.3)

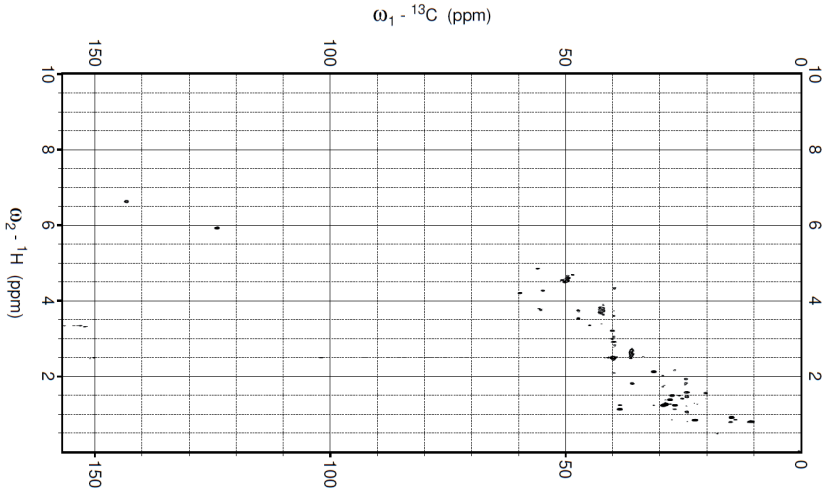
d-Dap$_9$ containing lipopeptide (4) 2D NMR Spectra



TOCSY

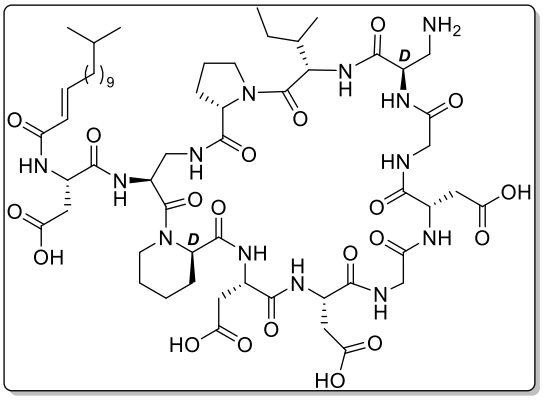


NOESY



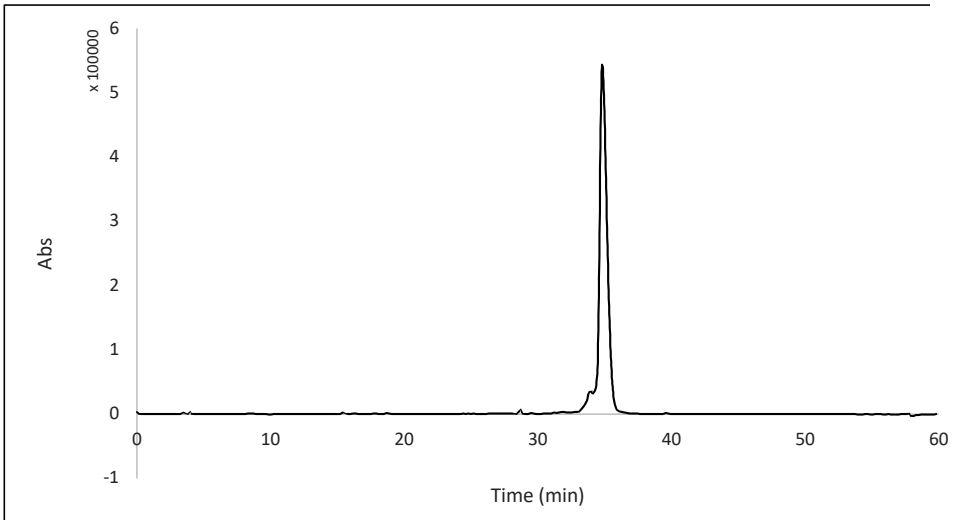
HSQC

Asp_4, D-Dap_9 containing lipopeptide (5)



Yield: 12.3 mg (9.5 umol, 3.8%)
HR-MS [M+H$^+$]: Calc.: 1290.6582, found: 1290.6603

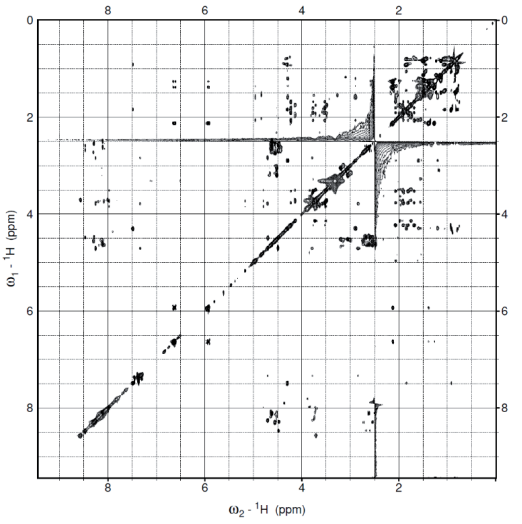
Analytical HPLC



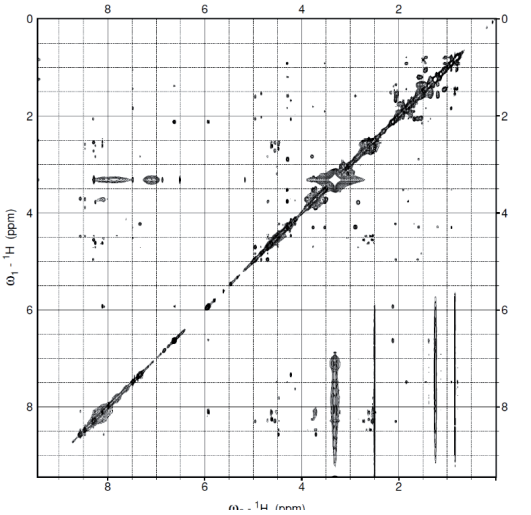
Asp_4, D-Dap_9 containing lipopeptide (5) NMR Chemical Shifts

Residue	NH	$H_\alpha(C_\alpha)$	Sidechain
Tail	-	5.94 (124.4)	$C_\beta H$ (6.63, 143.8), $C_\gamma H_2$ (2.12, 31.8), $C_\delta H_2$ (1.38, 28.2), $C_\varepsilon H_2$-$C_\iota H_2$ (1.25, 29.4), $C_\kappa H_2$ (1.24, 27.3), $C_\lambda H_2$ (1.13, 38.9), $C_\mu H$ (1.49, 27.8), $2C_\nu H_3$ (0.84, 23.0)
Asp-1	8.11	4.62 (50.1)	$C_\beta H_2$ (2.62/2.50, 36.6)
Dap-2	8.25	4.69 (48.9)	$C_\beta H_2$ (3.21/3.06, 40.4)
D-Pip-3	-	4.95 (55.8)	$C_\beta H_2$ (2.08/1.35, 28.2), $C_\gamma H_2$ (1.54/1.35, 20.4), $C_\delta H_2$ (1.58/1.20, 24.8), $C_\varepsilon H_2$ (4.29/2.94, 39.8)
Asp-4	8.47	4.48 (50.7)	$C_\beta H_2$ (2.68/2.60, 36.5)
Asp-5	8.28	4.49 (50.7)	$C_\beta H_2$ (2.53, 36.6)
Gly-6	8.09	3.71 (43.6)	-
Asp-7	8.30	4.56 (50.4)	$C_\beta H_2$ (2.72/2.55, 36.5)
Gly-8	8.20	3.74 (42.9)	-
D-Dap-9	7.34	4.74 (48.9)	$C_\beta H$ (3.19/3.05, 40.4)
Ile-10	7.49	4.30 (55.1)	$C_\beta H$ (1.83, 36.3), $C_\gamma H_2$ (1.46/1.05, 24.6), $C_\gamma H_3$ (0.92, 15.2), $C_\delta H_2$ (0.79, 11.0)
Pro-11	-	4.23 (60.3)	$C_\beta H_2$ (2.06/1.68, 29.8), $C_\gamma H_2$ (1.91/1.80, 25.0), $C_\delta H_2$ (3.76/3.52, 47.9)

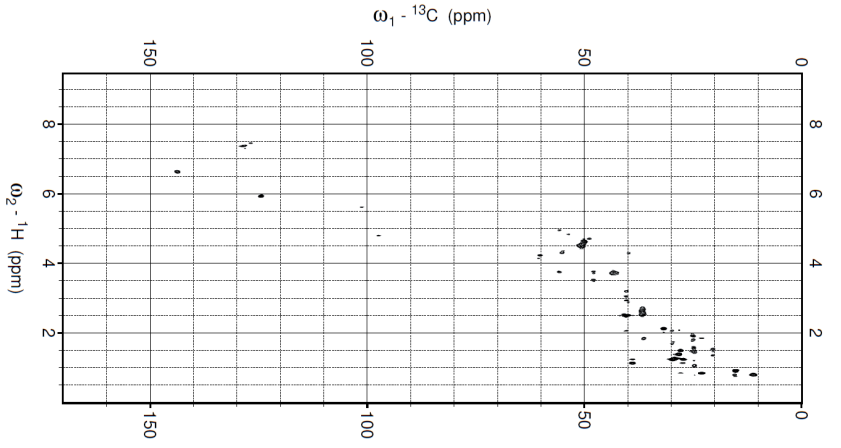
Asp_4, D-Dap_9 containing lipopeptide (5) 2D NMR Spectra



TOCSY

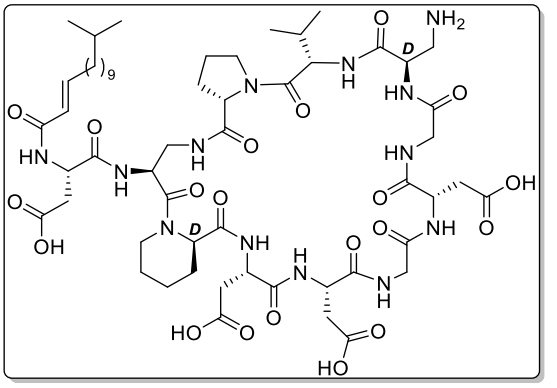


NOESY



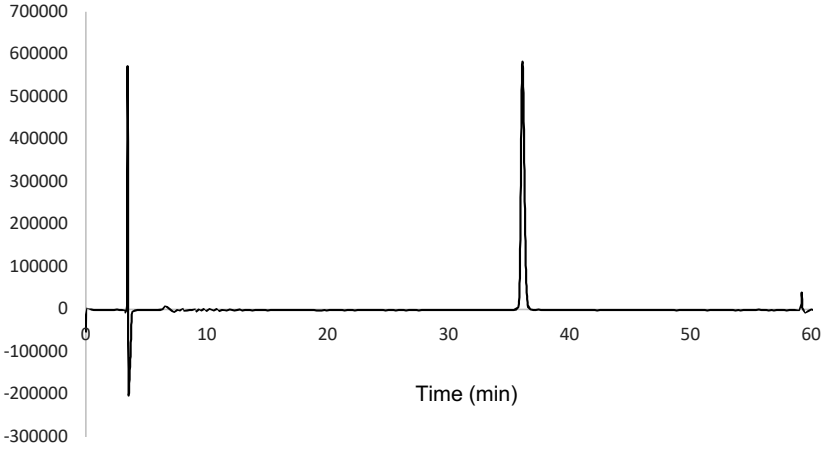
HSQC

Asp_4, d-Dap_9, Val_{10} containing lipopeptide (6)



Yield: 45 mg (32.9 umol, 32%)
HR-MS [M+H$^+$]: Calc.: 1276.6380, found: 1276.6395

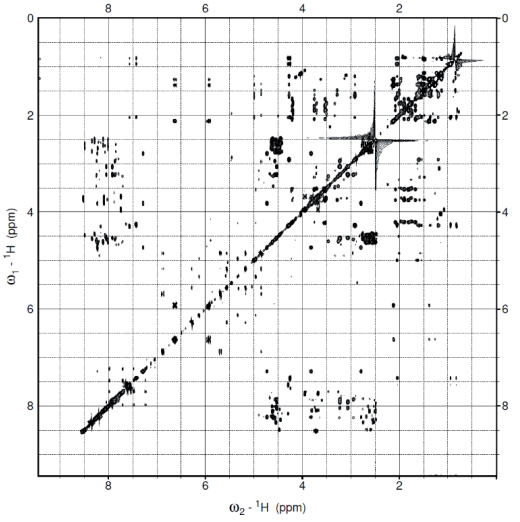
Analytical HPLC



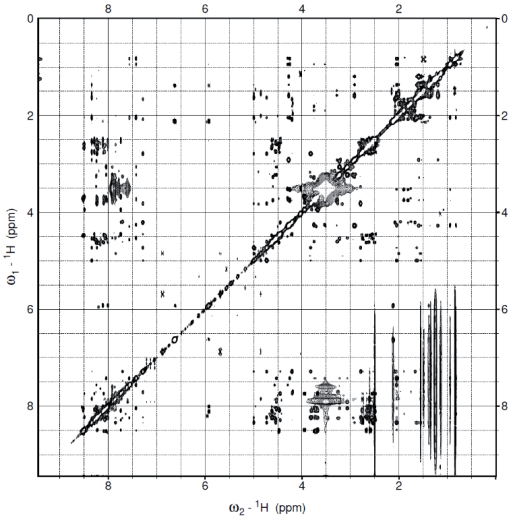
Asp_4, d-Dap_9, Val_{10} containing lipopeptide (6) NMR Chemical Shifts

Residue	NH	$H_\alpha(C_\alpha)$	Sidechain
Tail	-	5.94 (124.5)	$C_\beta H$ (6.63, 143.8), $C_\gamma H_2$ (2.12, 31.7), $C_\delta H_2$ (1.38, 28.3), $C_\varepsilon H_2$-$C_\iota H_2$ (1.25, 29.4), $C_\kappa H_2$ (1.24, 27.3), $C_\lambda H_2$ (1.13, 38.9), $C_\mu H$ (1.49, 27.8), $2C_\nu H_3$ (0.84, 23.0)
Asp-1	8.11	4.89 (49.8)	$C_\beta H_2$ (2.61/2.49, 36.5)
Dap-2	8.23	4.73 (48.8)	$C_\beta H_2$ (3.76/3.77, 40.3)
D-Pip-3	-	4.99 (55.7)	$C_\beta H_2$ (2.06/1.36, 28.4), $C_\gamma H_2$ (1.54/1.36, 20.3), $C_\delta H_2$ (1.60/1.20, 24.8), $C_\varepsilon H_2$ (4.28/2.92, 39.8)
Asp-4	8.34	4.55 (50.5)	$C_\beta H_2$ (2.76/2.54, 36.2)
Asp-5	8.22	4.61 (50.0)	$C_\beta H_2$ (2.68, 36.3)
Gly-6	8.09	3.76 (43.4)	-
Asp-7	8.29	4.47 (50.6)	$C_\beta H_2$ (2.51, 36.3)
Gly-8	7.74	3.94/3.67 (42.3)	-
D-Dap-9	7.28	4.72 (48.8)	$C_\beta H$ (3.81/2.80, 40.3)
Val-10	7.42	4.26 (56.6)	$C_\beta H$ (2.05, 30.4), $C\gamma 1H_3$ (0.94, 19.4), $C\gamma 2H_3$ (0.82, 18.9)
Pro-11	-	4.20 (60.3)	$C_\beta H_2$ (2.09/1.67, 29.8), $C_\gamma H_2$ (1.90/1.80, 24.9), $C_\delta H_2$ (3.75/3.53, 47.8)

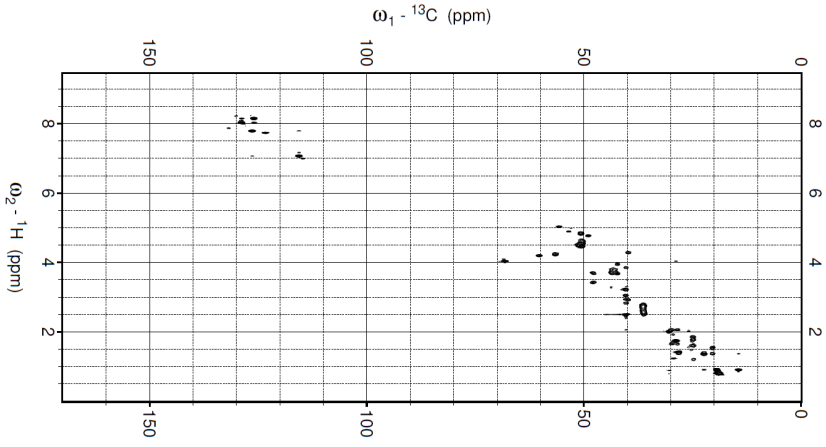
Asp_4, d-Dap_9, Val_{10} containing lipopeptide (6) 2D NMR Spectra



TOCSY



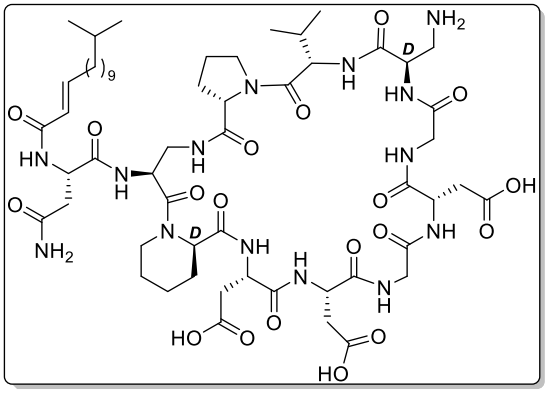
NOESY



HSQC

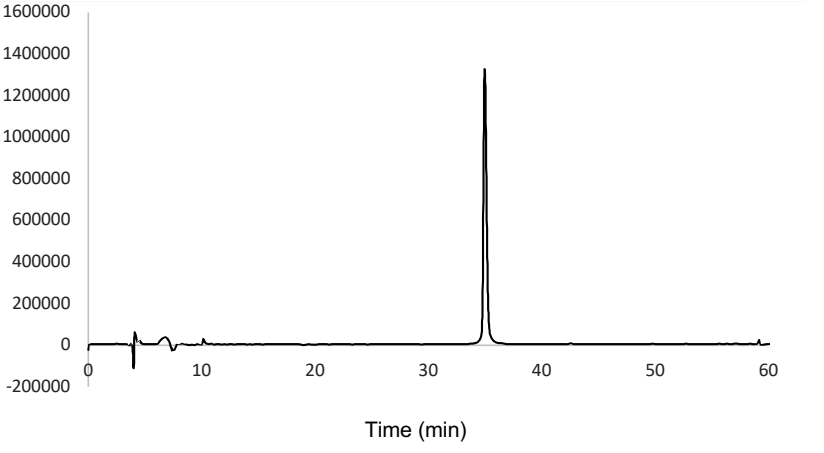
4

Asn_1, Asp_4, d-Dap_9, Val_{10} containing lipopeptide (7)



Yield: 12.5 mg (9.8 umol, 9.8%)
HR-MS [M+H$^+$]: Calc.: 1275.6585, found: 1275.6585

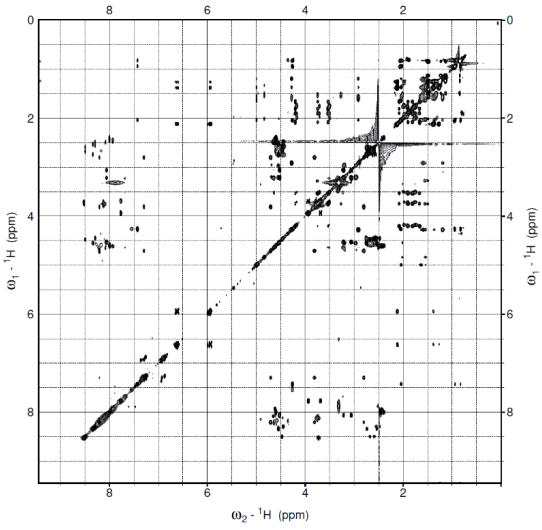
Analytical HPLC



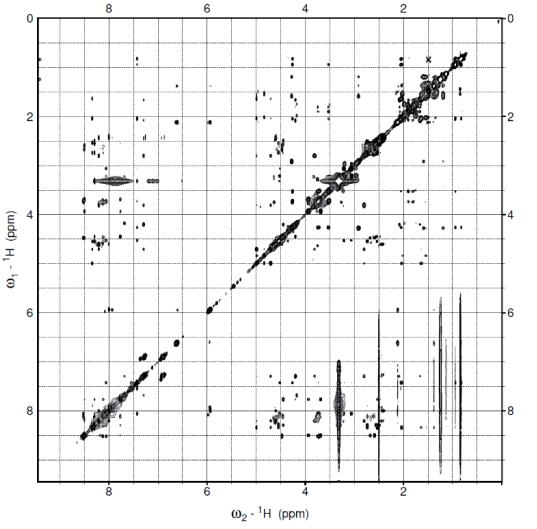
Asn_1, Asp_4, d-Dap_9, Val_{10} containing lipopeptide (7) NMR Chemical Shifts

Residue	NH	$H_\alpha(C_\alpha)$	Sidechain
Tail	-	5.93 (124.6)	$C_\beta H$ (6.62, 143.6), $C_\gamma H_2$ (2.12, 31.8), $C_\delta H_2$ (1.38, 28.4), $C_\varepsilon H_2$-$C_\iota H_2$ (1.25, 29.5), $C_\kappa H_2$ (1.24, 27.3), $C_\lambda H_2$ (1.13, 40.0), $C_\mu H$ (1.49, 27.9), $2C_\nu H_3$ (0.84, 23.0)
Asn-1	8.50	4.47 (50.9)	$C_\beta H_2$ (2.49/2.43, 37.7)
Dap-2	8.21	4.69 (48.7)	$C_\beta H_2$ (3.78/3.78, 40.3)
D-Pip-3	-	4.99 (55.7)	$C_\beta H_2$ (2.04/1.35, 28.3), $C_\gamma H_2$ (1.54/1.36, 20.4), $C_\delta H_2$ (1.58/1.20, 24.8), $C_\varepsilon H_2$ (4.28/2.91, 39.8)
Asp-4	8.34	4.55 (50.5)	$C_\beta H_2$ (2.68/2.60, 36.3)
Asp-5	8.00	4.60 (50.2)	$C_\beta H_2$ (2.68, 36.3)
Gly-6	8.12	3.73 (43.5)	-
Asp-7	8.29	4.45 (50.4)	$C_\beta H_2$ (2.51, 36.2)
Gly-8	7.77	3.93/3.69 (42.3)	-
D-Dap-9	7.29	4.71 (48.6)	$C_\beta H$ (3.81/2.81, 40.3)
Val-10	7.43	4.26 (56.7)	$C_\beta H$ (2.05, 30.4), $C\gamma 1H_3$ (0.94, 19.4), $C\gamma 2H_3$ (0.83, 19.1)
Pro-11	-	4.20 (60.4)	$C_\beta H_2$ (2.08/1.67, 29.8), $C_\gamma H_2$ (1.90/1.80, 25.0), $C_\delta H_2$ (3.74/3.53, 48.0)

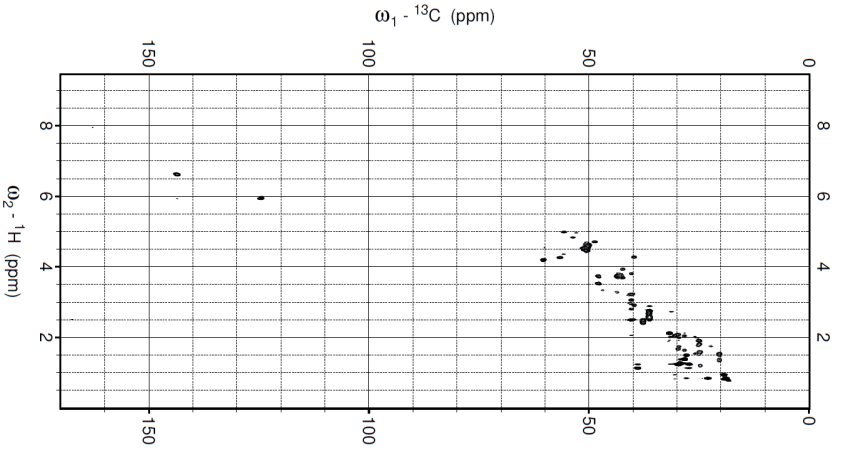
Asn_1, Asp_4, d-Dap_9, Val_{10} containing lipopeptide (7) 2D NMR Spectra



TOCSY

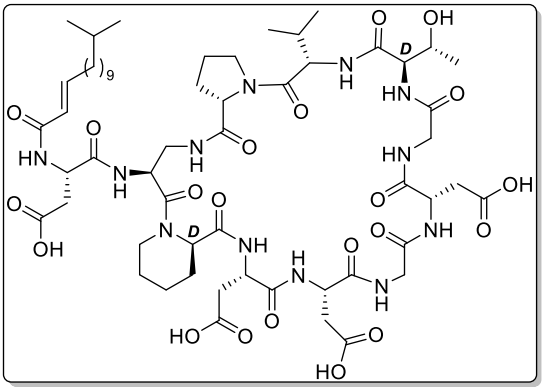


NOESY



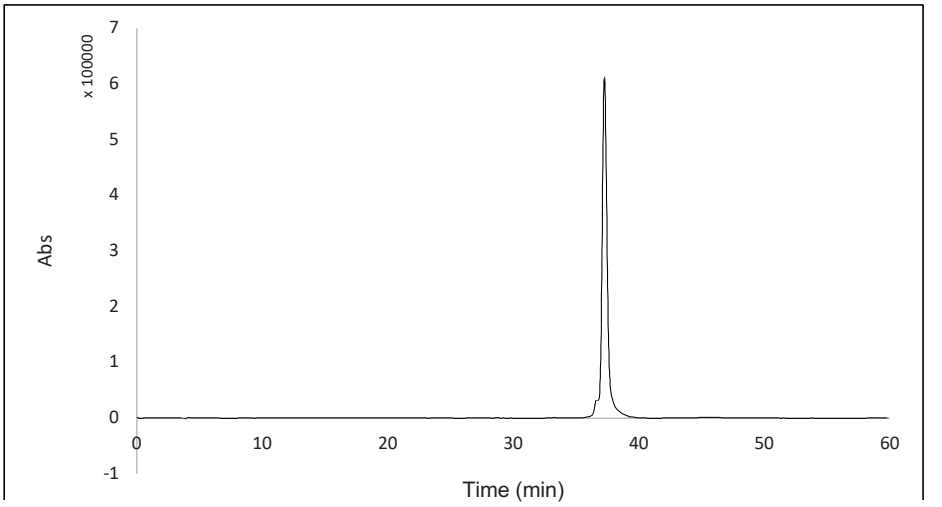
HSQC

Asp_4, Val_{10} containing lipopeptide (8)



Yield: 41 mg (32.9 umol, 32%)
HR-MS [M+H$^+$]: Calc.: 1291.6422, found: 1291.6483

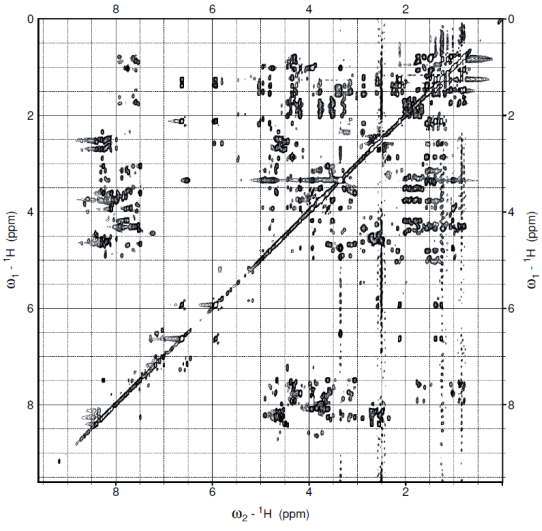
Analytical HPLC



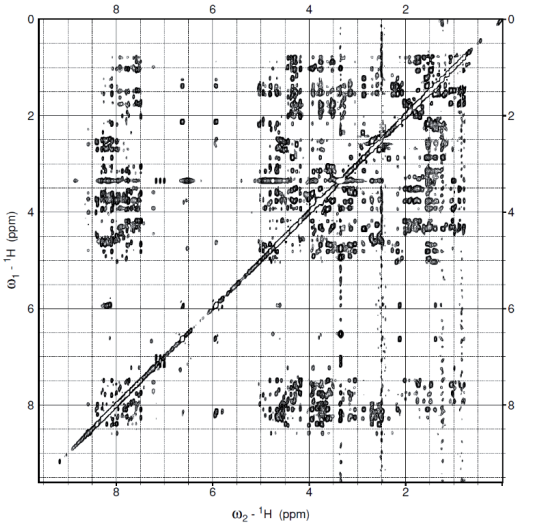
Asp_4, Val_{10} containing lipopeptide (8) NMR Chemical Shifts

Residue	NH	$H_\alpha(C_\alpha)$	Sidechain
Tail	-	5.94 (124.5)	$C_\beta H$ (6.63, 143.8), $C_\gamma H_2$ (2.12, 31.7), $C_\delta H_2$ (1.38, 28.3), $C_\varepsilon H_2$-$C_\iota H_2$ (1.25, 29.4), $C_\kappa H_2$ (1.24, 27.3), $C_\lambda H_2$ (1.13, 38.9), $C_\mu H$ (1.49, 27.8), $2C_\nu H_3$ (0.84, 23.0)
Asp-1	8.11	4.89 (49.8)	$C_\beta H_2$ (2.61/2.49, 36.5)
Dap-2	8.23	4.73 (48.8)	$C_\beta H_2$ (3.76/3.77, 40.3)
D-Pip-3	-	4.99 (55.7)	$C_\beta H_2$ (2.06/1.36, 28.4), $C_\gamma H_2$ (1.54/1.36, 20.3), $C_\delta H_2$ (1.60/1.20, 24.8), $C_\varepsilon H_2$ (4.28/2.92, 39.8)
Asp-4	8.34	4.55 (50.5)	$C_\beta H_2$ (2.76/2.54, 36.2)
Asp-5	8.22	4.61 (50.0)	$C_\beta H_2$ (2.68, 36.3)
Gly-6	8.09	3.76 (43.4)	-
Asp-7	8.29	4.47 (50.6)	$C_\beta H_2$ (2.51, 36.3)
Gly-8	7.74	3.94/3.67 (42.3)	-
D-*allo*-Thr-9	7.88	4.29 (58.7)	$C_\beta H$ (3.83, 67.2), $C_\gamma H_3$ (1.02, 20.1)
Val-10	7.43	4.26 (56.7)	$C_\beta H$ (2.03, 30.5), $C\gamma 1H_3$ (0.94, 19.4), $C\gamma 2H_3$ (0.82, 18.9)
Pro-11	-	4.20 (60.3)	$C_\beta H_2$ (2.09/1.67, 29.8), $C_\gamma H_2$ (1.90/1.80, 24.9), $C_\delta H_2$ (3.75/3.53, 47.8)

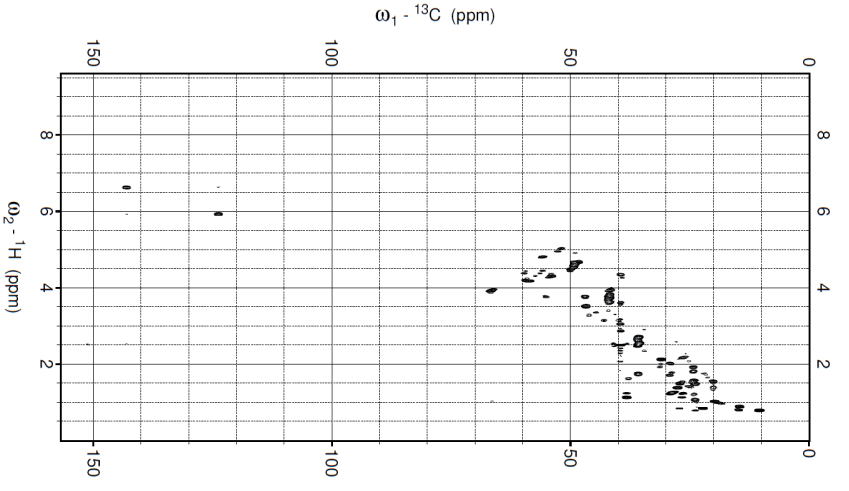
Asp_4, Val_{10} containing lipopeptide (8) 2D NMR Spectra



TOCSY



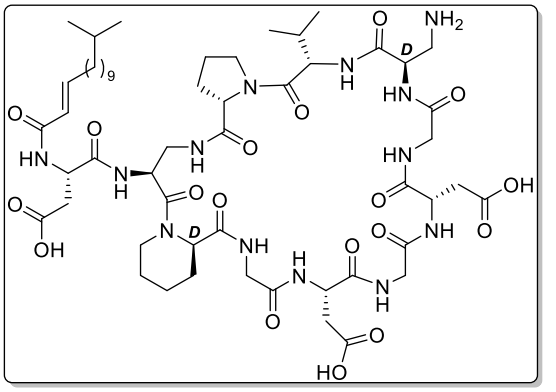
NOESY



HSQC

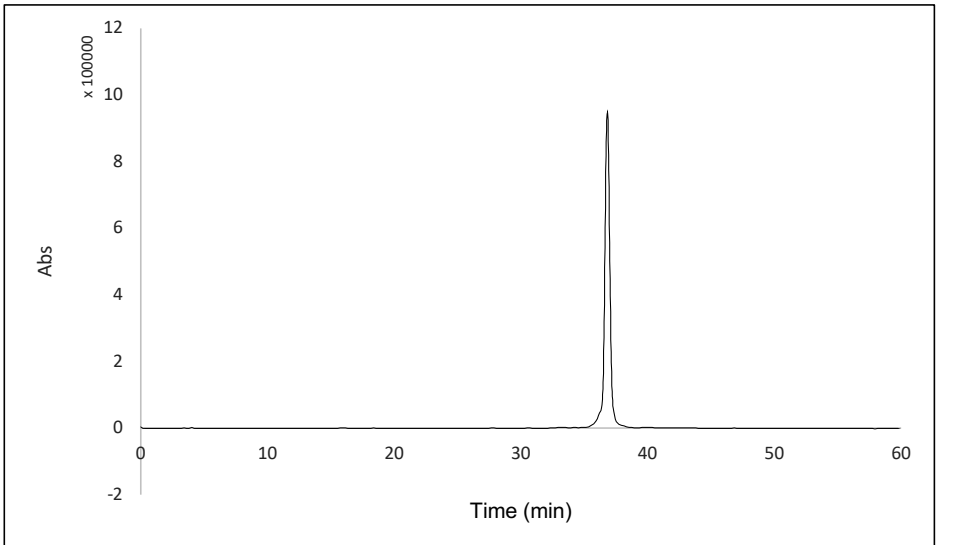
4

D-Dap_9, Val_{10} containing lipopeptide (9)



Yield: 38 mg (30.9 umol, 32%)
HR-MS [M+H$^+$]: Calc.: 1218.6370, found: 1218.6385

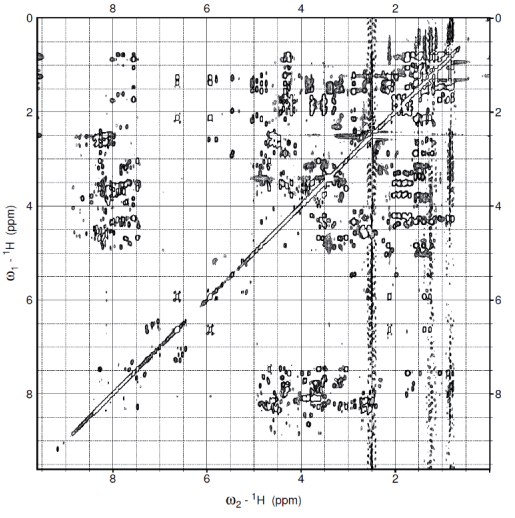
Analytical HPLC



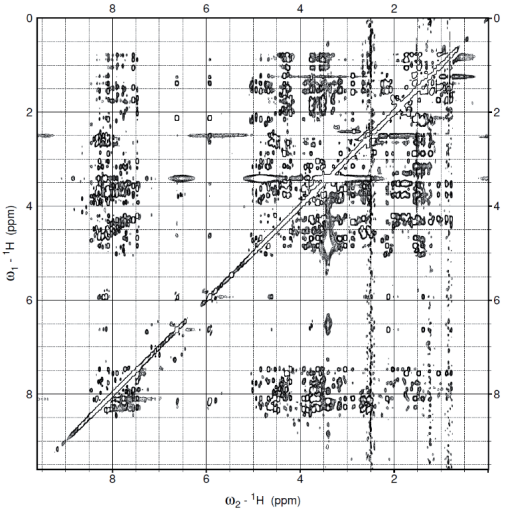
D-Dap_9, Val_{10} containing lipopeptide (9) NMR Chemical Shifts

Residue	NH	$H_\alpha(C_\alpha)$	Sidechain
Tail	-	5.94 (124.5)	$C_\beta H$ (6.63, 143.8), $C_\gamma H_2$ (2.12, 31.7), $C_\delta H_2$ (1.38, 28.3), $C_\varepsilon H_2$-$C_\iota H_2$ (1.25, 29.4), $C_\kappa H_2$ (1.24, 27.3), $C_\lambda H_2$ (1.13, 38.9), $C_\mu H$ (1.49, 27.8), $2C_\nu H_3$ (0.84, 23.0)
Asp-1	8.11	4.89 (49.8)	$C_\beta H_2$ (2.61/2.49, 36.5)
Dap-2	8.23	4.73 (48.8)	$C_\beta H_2$ (3.76/3.77, 40.3)
D-Pip-3	-	4.99 (55.7)	$C_\beta H_2$ (2.06/1.36, 28.4), $C_\gamma H_2$ (1.54/1.36, 20.3), $C_\delta H_2$ (1.60/1.20, 24.8), $C_\varepsilon H_2$ (4.28/2.92, 39.8)
Gly-4	8.20	3.81/3.63 (42.0)	-
Asp-5	8.22	4.61 (50.0)	$C_\beta H_2$ (2.68, 36.3)
Gly-6	8.09	3.76 (43.4)	-
Asp-7	8.29	4.47 (50.6)	$C_\beta H_2$ (2.51, 36.3)
Gly-8	7.74	3.94/3.67 (42.3)	-
D-Dap-9	7.26	4.74 (48.9)	$C_\beta H$ (3.83/2.82, 40.1)
Val-10	7.42	4.25 (56.4)	$C_\beta H$ (2.04, 30.4), $C\gamma 1H_3$ (0.95, 19.3), $C\gamma 2H_3$ (0.81, 19.0)
Pro-11	-	4.20 (60.3)	$C_\beta H_2$ (2.09/1.67, 29.8), $C_\gamma H_2$ (1.90/1.80, 24.9), $C_\delta H_2$ (3.75/3.53, 47.8)

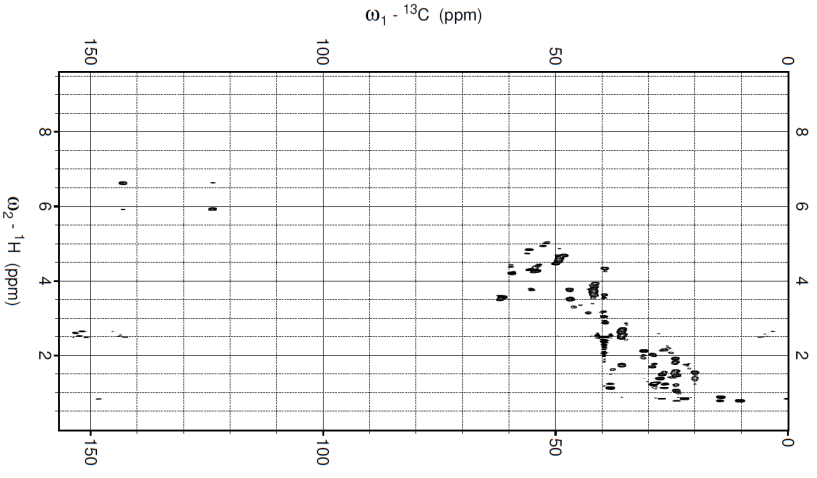
D-Dap$_9$, Val$_{10}$ containing lipopeptide (9) 2D NMR Spectra



TOCSY



NOESY



HSQC

Crystallization and data collection

Lipopeptide **5** or **7** was solubilized in 5 mM HEPES pH 7.5, 10 mM $CaCl_2$ and mixed 1 : 2 with C_{10}-P, to achieve a final concentration of 7.2 mM : 14.4 mM in presence of 10 % v/v PEG 200. Crystals were obtained by sitting drop vapour diffusion at 18 °C, by mixing 150 nL of the peptide solution with 150 nL of the reservoir solution, composed of 0.2 M sodium formate and 40 % v/v MPD for lipopeptide **5**, or 0.2 M cadmium chloride and 40 % v/v MPD for lipopeptide **7**, both supplemented by 10 % v/v PEG 200. Crystals were harvested without additional cryoprotectant and flash-cooled in liquid nitrogen. Datasets were collected at 100 K at the Diamond Light Source beamline I04-1 (lipopeptide **5**) or I04 (lipopeptide **7**).

Structure solution and refinement

The dataset of lipopeptide **5** was processed in the DIALS pipeline (2), whereas autoPROC (3) was used for lipopeptide **7**. The crystal of lipopeptide **7** was initially indexed in a hexagonal setting but based on the merging R-values the true symmetry appeared to be Primitive monoclinic with $\beta = 120°$. The reflection file was therefore re-indexed accordingly, and parameters for pseudo-merohedral twinning were included in the structure refinement. Additional anisotropic correction was done for the datasets of both analogues in STARANISO (3). Structures were solved by molecular replacement using PHASER (4), and one copy (lipopeptide **5**) or one dimer (lipopeptide **7**) of laspartomycin C in complex with geranyl phosphate (PDB: 5O0Z) (5) was used as a search model. Models were manually improved in *Coot* (6), refinement was performed using REFMAC (7) and Molprobity (8) was used for validation. Structures of lipopeptides **5** and **7** in complex with Ca^{2+} and C_{10}-P were deposited to the Protein Data Bank under the accession codes 7AG5 and 7ANY, respectively.

Table S2. Data collection and refinement statistics. Highest resolution shell in parentheses.

	Lipopeptide 5 (PDB: 7AG5)	**Lipopeptide 7 (PDB: 7ANY)**
Data collection		
Space group	*P* 6 2 2	*P* 2_1
Cell dimensions		
a, *b*, *c* (Å)	40.43, 40.43, 31.03	40.13, 68.32, 40.13
α, *β*, *γ* (°)	90, 90, 120	90, 120, 90
Resolution (Å)	35.01 - 1.03 (1.12 - 1.03)	34.76 - 1.14 (1.27 - 1.14)
No. observed reflections	74114 (4481)	119460 (3464)
No. unique reflections	6321 (421)	36022 (1799)
R_{merge}	0.185 (1.584)	0.087 (0.373)
Mean I/σI	8.0 (1.5)	6.2 (2.8)
$CC_{1/2}$	0.997 (0.726)	0.995 (0.852)
Completeness (spherical, %)	80.9 (26.4)	52.2 (9.1)
Completeness (ellipsoidal, %)	92.2 (53.1)	85.3 (31.6)
Ellipsoidal resolution limits (Å) [direction]	1.03 [a*] 1.03 [b*] 1.19 [c*]	1.14 [a*] 1.81 [b*] 1.20 [c*]
Redundancy	11.7 (10.5)	3.3 (1.9)
Refinement		
Resolution (Å)	35.01 - 1.04	34.76 - 1.14
R_{work}/R_{free} (%)	12.04 / 14.32	15.96 / 19.21
Average *B*-factors (Å^2)		
Protein	12.3	12.6
Ligands/ions	26.5	18.3
Waters	26.7	21.4
R.M.S. deviations		
Bond lengths (Å)	0.019	0.017
Bond angles (°)	1.72	2.19
No. atoms		
Protein	178	1080
Ligands/ions	34	237
Water	19	173

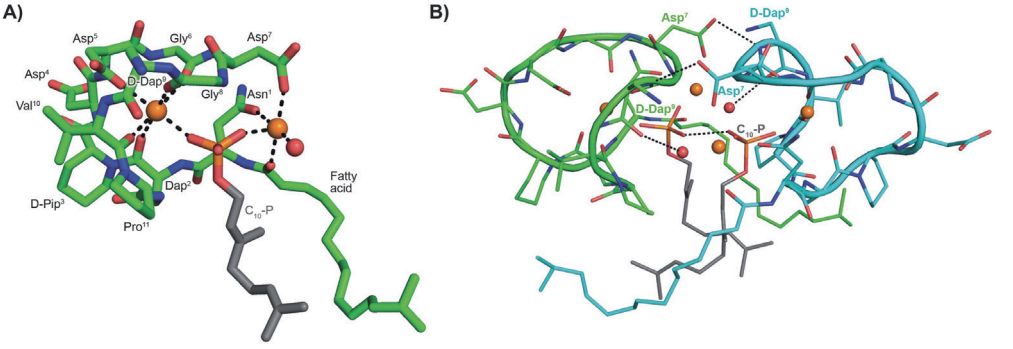
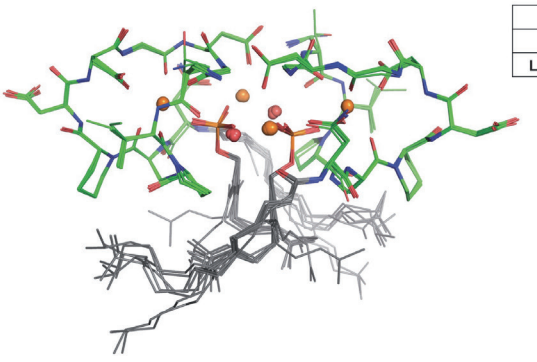


Figure S2. A) Structure of the ternary complex with lipopeptide **7** (green stick representation), two bound Ca^{2+} ions (orange spheres), a bound water molecule (red sphere), and the C_{10}-P ligand (lipid in grey). **B)** Lipopeptide **7** adopts a saddle-shaped conformation when complexed with two Ca^{2+} ions and C_{10}-P and forms a dimer in the crystal.



RMSD (Å)	Lipopeptide 5	Lipopeptide 7
Lipopeptide 7	0.100	X
Laspartomycin C	0.165	0.192

Figure S3: Lipopeptide 5 and Lipopeptide 7 dimers are similar to the Laspartomycin C dimer. Superposition of the dimer structures of lipopeptides **5** and **7** with that of laspartomycin C (PDB: 5O0Z) (5). Asymmetric units are composed of one monomer for lipopeptide **5** (the dimer shown for lipopeptide **5** generated by applying two-fold crystallographic symmetry), six dimers for lipopeptide **7** and one dimer for laspartomycin C. RMSD between dimers is indicated in the table. Ca^{2+} ions are represented by orange spheres and water molecules are represented by red spheres, C_{10}-P is also indicated.

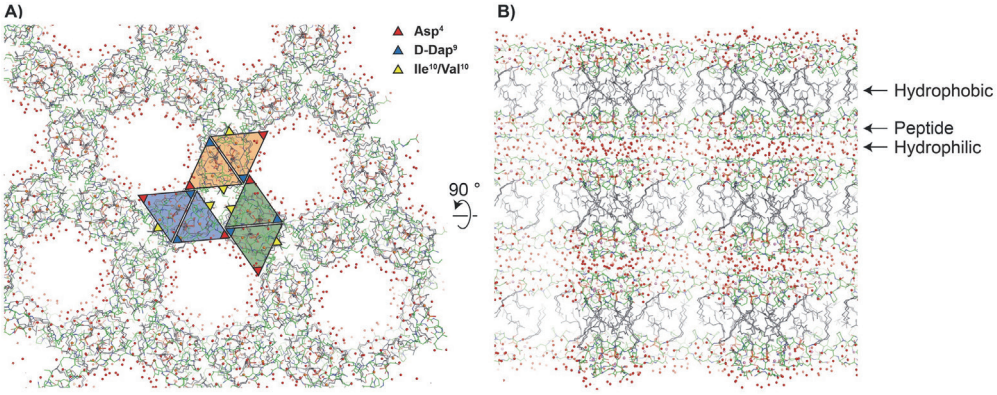


Figure S4. In the crystal state lipopeptide **5** forms a higher-ordered assembly when complexed with Ca^{2+} and C_{10}-P consisting of alternating hydrophobic (grey), peptidic (green), and hydrophilic (red) layers. A similar lattice is also observed for lipopeptide **7**. Notably, this higher ordered assembly is not seen for laspartomycin C.

Bacterial cytological profiling

B. subtilis reporter strains (Table S1) were aerobically grown at 30 °C in LB supplemented with 2mM $CaCl_2$ and antibiotic (5 µg/ml chloramphenicol or 100 µg/ml spectinomycin). Overnight cultures were diluted 100x without antibiotics and GFP-fusion protein expression induced with xylose (% in Table S3). At an OD_{600} of approximately 0.4 the cultures were diluted 10x in the same medium. At OD600 0.2-0.3 150 µl cells were incubated with 12.5 µg/ml laspartomycin C, 5 µg/ml lipopeptide **6,** or 2 µg/ml lipopeptide **6**. After 10 and 30 minutes 0.5 µl cells were immobilized on microscope slides covered with a 1% agarose film and imaged immediately.

Fluorescence microscopy was carried out using a Zeiss Axiovert 200M equipped with a Zeiss Neofluar 100x/1.30 Oil Ph3 objective, a Lambda S light source (Shutter Instruments), a Photometrics Coolnap HQ2 camera, and Metamorph 6 software (Molecular Devices). Images were analyzed using ImageJ (National Institutes of Health) v.1.52a.

Table S3: *B. subtilis* strains used in this study ref PMID: 27791134.

Strain	genotype	induction
1049	*amyE::spc Pxyl-rpsB-gfp*	1% xylose
1048	*cat rpoC-gfp Pxyl-rpoC*	1% xylose
YK405	*amyE::spc Pxyl-gfp-mreB*	0.3% xylose
4056	*amyE::spec Pxyl-gfp-pmut1-ftsZ*	0.1% xylose
TB35	*amyE::spc Pxyl-gfp-minD*	0.25% xylose
BS23	*atpA-gfp Pxyl-atpA cat*	0.1% xylose
TNVS91	*ΔamyE::specR-PxylR-PolC-4GS-msfGFP*	0.03% xylose
TNVS175	*amyE::spc-Pxyl-murG-msfgfp*	0.05% xylose

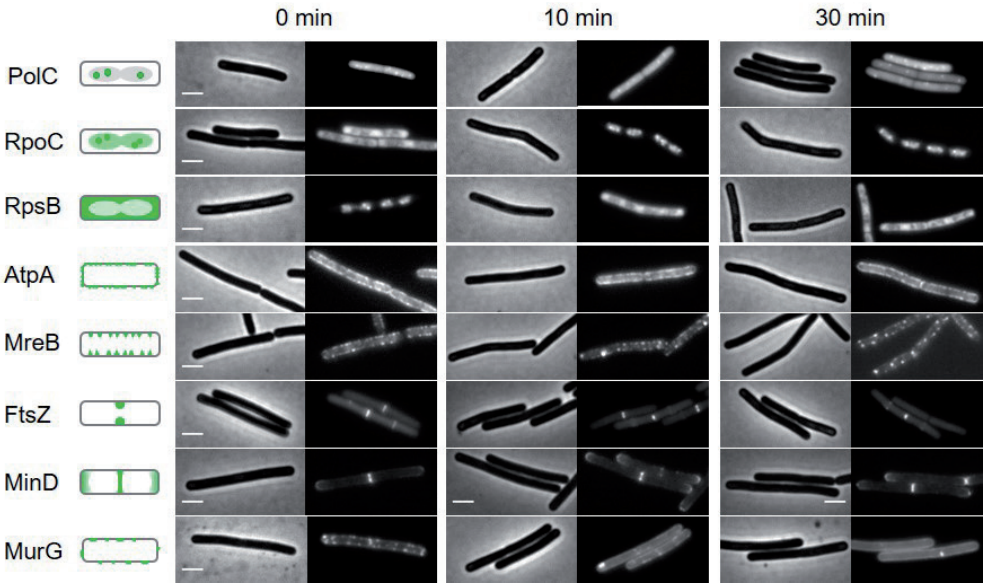


Figure S5. Bacterial cytological profiling analysis of lipopeptide **6**. The GFP-tagged marker proteins represents the following cellular activities: DNA polymerization (PolC), RNA polymerization (RpoC), protein synthesis (RpsB), F0F1 ATPase (AtpA), lateral cell wall synthesis regulation (MreB), cell division (FtsZ), cell division regulation (MinD) and peptidoglycan precursor synthesis (MurG). Left panels schematically show the normal localization patterns of the different GFP fusions. Strains were grown in LB medium supplemented with 2 mM $CaCl_2$ at 30 °C. 2x MIC concentration was added (0 min) and samples for microscopy were taken after 10- and 30-min incubation, respectively. Scale bars indicate 2 µm.

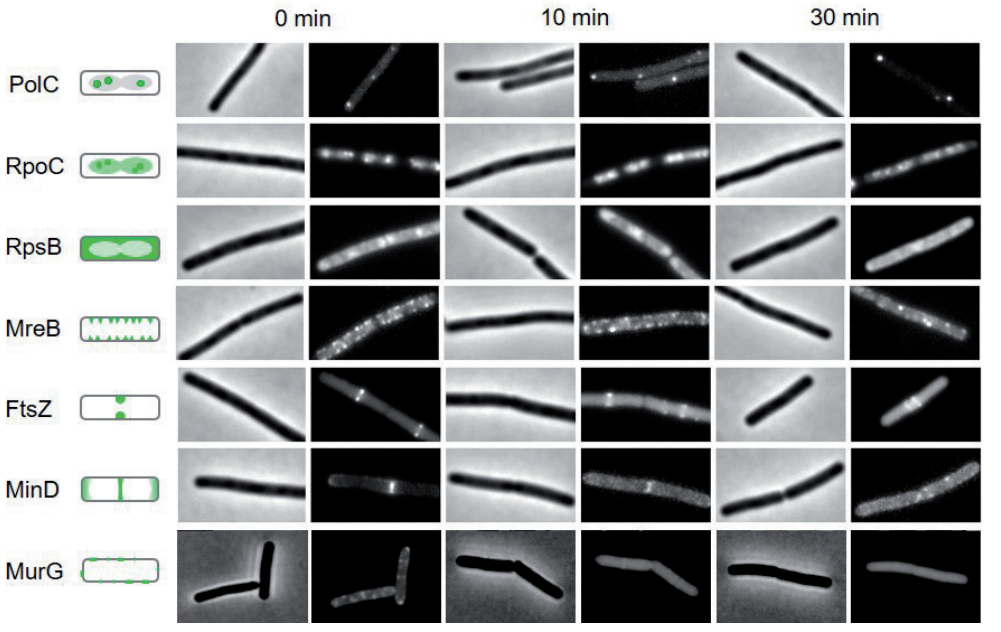


Figure S6. Daptomycin reference for the bacterial cytological profiling analyses. The figure was adapted from Müller et al. (9) The GFP-tagged marker proteins represents the following cellular activities: DNA polymerization (PolC), RNA polymerization (RpoC), protein synthesis (RpsB), lateral cell wall synthesis regulation (MreB), cell division (FtsZ), cell division regulation (MinD) and peptidoglycan precursor synthesis (MurG). Left panels schematically show the normal localization patterns of the different GFP fusions. Strains were grown in LB medium supplemented with 1.25 mM $CaCl_2$ and treated with 2 µg/mL daptomycin at 30 °C. Samples for microscopy were taken before (0 min) and after 10 and 30 min incubation.

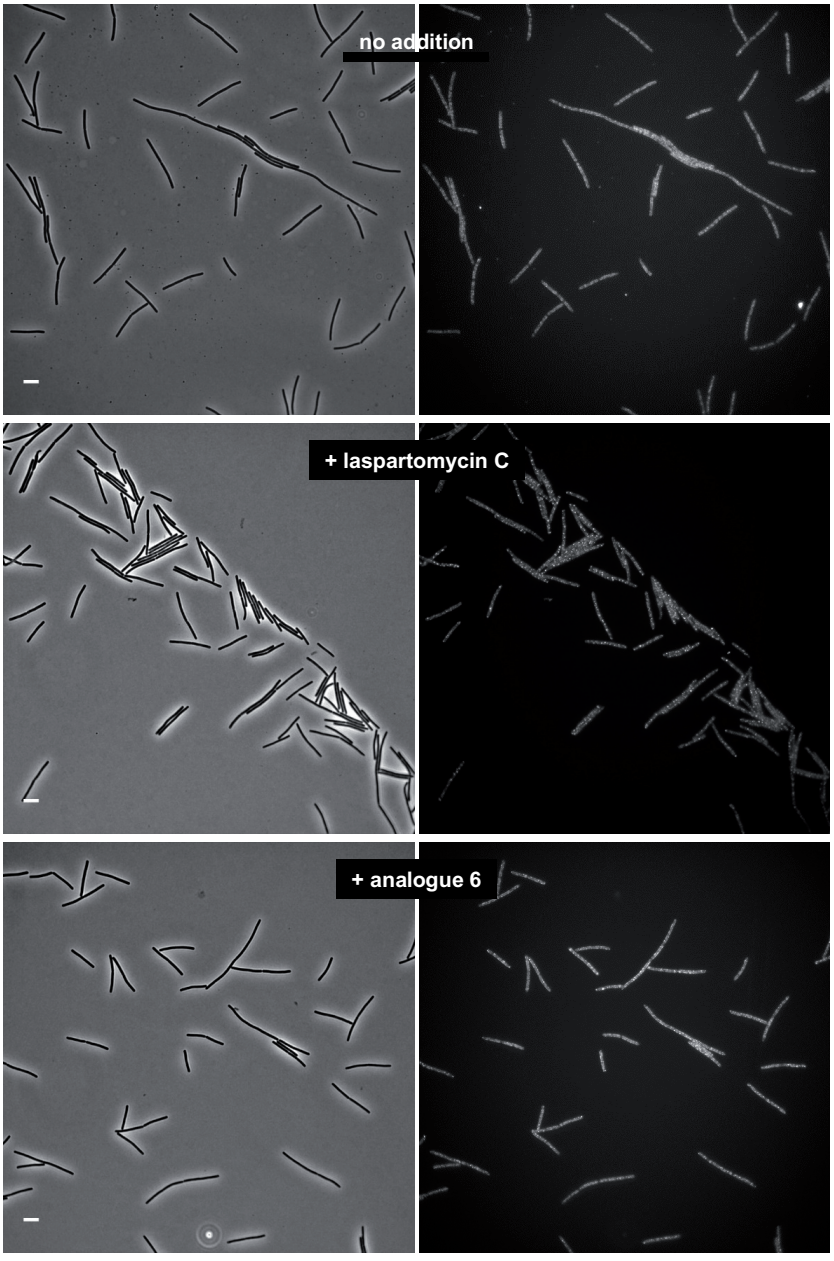


Figure S7. Large field phase contrast and fluorescent images showing the effect of 10 min incubation with 12.5 µg/ml laspartomycin C or with 5 µg/ml lipopeptide **6** on the localization of GFP-MreB. Strains were grown in medium supplemented with 2 mM $CaCl_2$ at 30 °C. Scale bars indicate 5 µm.

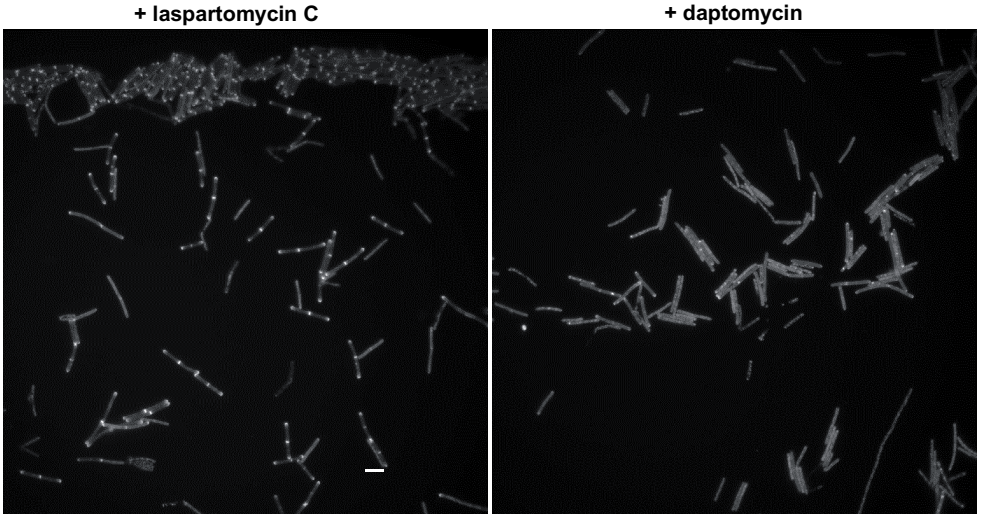


Figure S8. large field images showing the effect of laspartomycin C (12.5 µg/ml) or daptomycin (2 µg/ml) on the localization of GFP-MinD after 30 min incubation with the antibiotics. Localization of MinD is unaffected by laspartomycin C, whereas this protein detaches from the membrane when treated with daptomycin Strains were grown in medium supplemented with 2 mM $CaCl_2$ at 30 °C. Scale bars indicate 5 µm.

SI References

1. D. Abmm, D. Tamma, J. Kirn, S. K. Culle, performance standards for antimicrobial susceptibility testing. *CLSI Suppl. M100,* **20** (2020)
2. G. Winter, *et al.*, DIALS: implementation and evaluation of a new integration package. *Struct. Biol.* **74**, 85–97 (2018)
3. G. Vonrhein, *et al.*, Advances in automated data analysis and processing within autoPROC, combinded with improved characterisation, mitigation and visualisation of the anisotropy of diffraction limits using STARANISO. *Acta. Cryst.* **74**, 43537 (2018)
4. A. J. McCoy, *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007)
5. L. H. J. Kleijn, *et al.*, A high-resolution crystal strucutre that reveals molecular details of target recognition by the calcium-dependent lipopeptide antibiotic laspartomycin C. Angew. Chemie–Int. Ed. **56**(52), 16546–16549 (2017).
6. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010)
7. G. N. Murshudov, *et al.*, REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011)
8. V. B. Chen, *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010)
9. A. Müller, *et al.*, Daptomycin inhibits cell envelope synthesis by interfering with fluid membrane microdomains. *Proc. Natl. Acad. Sci. USA*. **113**(45), E7077-E7086 (2016)

Chapter 5

General discussion

Matthieu R. Zeronian

Structural Biochemistry, Bijvoet Centre for Biomolecular Research, Utrecht University,
Padualaan 8, 3584 CH Utrecht, The Netherlands

Due to their central roles in health and disease, cell surface receptors and calcium-dependent antibiotics represent top priorities for fundamental as well as applied research. Understanding the molecular mechanisms by which these molecules act on the human or bacterial cell membrane is key for ultimately developing therapeutic molecules against their associated diseases. In this thesis, we describe the progress we made towards understanding the role of the Notch1 and epidermal growth factor receptor (EGFR) cell surface proteins in development, homeostasis and disease, and we characterize new lipopeptide antibiotic analogues that constitute promising molecules in the fight against antibiotic resistance. In this chapter, I discuss the implications of our findings on better understanding the molecular mechanisms underlying these systems and their potential applications for therapeutics.

Structural insights into Notch activation

The Notch pathway is an evolutionary conserved signaling system that is present in all multicellular organisms. Normal Notch signaling plays a role in embryogenesis, as it controls cell differentiation, proliferation and determines cell fate decisions (1–3), however alterations of this system are associated with various congenital diseases and cancers affecting diverse organs such as the heart, lung, liver and skin (4–7). Despite the critical role of Notch signaling in development and homeostasis, the mechanism by which the Notch extracellular region engages its ligand Jagged to activate signaling remains unclear. An early study by the Artavanis-Tsakonas group identified epidermal growth factor (EGF)-like repeats 11-12 as the core Notch ligand recognition site (8), and later on the negative regulatory region (NRR) was shown to prevent ligand-independent Notch activation (9). These sites represent the minimal requirement for Notch activation and have been widely studied due to their key roles in Notch signaling (10–15). Transcellular ligand binding at Notch EGF8-12 and subsequent ligand cell endocytosis are proposed to generate a pulling force that mechanically triggers a conformational change in the NRR, located 24 domains away from EGF8-12 in the primary sequence, to expose the S2 site and further activate Notch signaling (10, 16–21). While ligand binding in *cis* is generally associated with signaling inhibition (22–24), it was recently proposed that ligand binding in *cis* can also activate Notch (25), although it is not clear if and how ligand cell endocytosis plays a role in this context. Together, these studies raise the questions of how the full extracellular region of Notch interacts with Jagged, and how this interaction impacts signaling.

Here, using a combination of cross-linking mass spectrometry (XL-MS), biophysical and structural techniques on the Notch1-Jagged1 full extracellular complex as well as defined sites, we identify several previously unreported regions that form an interaction network. We show that three regions in Jagged1, C2-EGF3, EGF8-11 and cysteine-rich domain (CRD),

engage intramolecularly which each other, and intermolecularly with Notch1 EGF33-NRR. We pinpoint the interaction of Jagged1 C2-EGF3 with Notch1 NRR, thus revealing that these two critical sites are not distal as previously thought, but engage directly to control signaling. Small-angle X-ray scattering (SAXS) experiments suggest that Notch1 ectodomain flexibility and dimensions are compatible with an intramolecular EGF8-13–NRR interaction, and with the formation of a non-linear Notch1 ectodomain architecture. Our work shows that both Notch1 EGF8-13 and NRR interact with Jagged1 C2-EGF3, raising the question of whether these regions form a ternary complex, and what influence such a complex would have on signaling.

Besides the core Notch EGF8-12 and NRR, other sites were suggested to play a role in Notch functioning, such as EGF6, EGF25-26 and EGF36 (26–31). In particular, the Notch EGF8-12 site was proposed to form intramolecular interactions with EGF22-27 and EGF25-26 (30, 31), and EGF25-36 was suggested to interact with Serrate (a homolog of Jagged) (28). Defined regions on Notch, such as EGF24-26 (29), O-linked fucosylation on EGF26 (27), and O-fucose elongation with GlcNAc on EGF6 and 36 (26), were previously shown to play a role in Jagged/Serrate-mediated Notch signaling. Some of these sites, including EGF36 (26) and EGF25-36 (28), are located in the Notch ectodomain C-terminal region. This concurs with our data that presents Notch1 EGF33-NRR as an interaction hub in the Notch1-Jagged1 complex, as it binds intermolecularly to Jagged1 C2-EGF3, EGF8-11 and CRD, and intramolecularly to Notch1 EGF8-13.

Specific regions in Jagged were also suggested to play a role in Notch function, such as the core C2-EGF3 binding site that activates Notch signaling (17, 32). In addition to binding to Notch, Jagged C2-EGF3 was proposed to interact with lipids, which would permit optimal Notch activation (33, 34). This correlates with our results that indicate binding between Jagged1 C2-EGF3 and the membrane-proximal Notch1 NRR, and opens the possibility of a ternary complex formed between these two sites and the cell membrane displaying Notch1. On the Jagged-presenting cell, the membrane-proximal CRD was suggested to play a role in ligand-mediated Notch activation (35), concurring with our finding that the CRD interacts intramolecularly with C2-EGF3, and intermolecularly with Notch1 EGF33-NRR.

Collectively, our work, together with that of others, indicates that aside from the core binding and activation sites, several additional regions in Notch and Jagged contribute to Notch function. We have, however, not been able to directly show the functional importance of the interactions we report, such as the Notch1 NRR–Jagged1 C2-EGF3 interaction, for Notch1 signaling. Lack of detailed information on the binding sites we identified prevented us from investigating the importance of these interactions in a cellular setting and to evaluate their impact on Notch signaling. More detailed structural studies, using the regions of interest we have identified, could potentially indicate sites or single residues that can be modified

to disrupt or strengthen interactions. In particular, the Notch1 NRR and Jagged1 C2-EGF3 sites are suited for such studies because we have shown that in Jagged1, the C2-EGF3 region is required and sufficient for Notch1 NRR engagement. Furthermore, based on a docking experiment we have suggested that the NRR and EGF8-13 sites on Notch1 bind to the same site on Jagged1 C2-EGF3. To prove this experimentally, it would be interesting to disrupt the already known Notch1 EGF8-13–Jagged1 C2-EGF3 interaction by point mutations, and to study whether these point mutations would also disturb the interaction of Jagged1 C2-EGF3 with Notch1 NRR. Once structure-guided mutations have been shown to modulate interactions in biophysical assays, such as surface plasmon resonance (SPR), these mutations can be tested in a functional cellular assay using full length proteins. This will show whether these interactions, such as Notch1 NRR–Jagged1 C2-EGF3, directly influence signaling, and if so, whether they activate or inhibit it.

In a cellular context, many parameters can influence how proteins bind to each other and initiate signaling. At the cell surface, Notch and ligand homomeric interactions were proposed to regulate signaling (36, 38–42). Furthermore, the Notch1-Jagged1 complex exhibits a catch-bond behavior (17), i.e. the complex bond lifetime increases with the tension force, which could explain how the low-affinity Notch-ligand interactions lead to significant Notch activation and cellular response. As Notch1 possesses an intrinsically low affinity for its ligands (36, 37), *in vitro* evolution studies were employed to produce point mutations on Jagged1 and generate a stable Notch1 EGF8-12–Jagged1 C2-EGF3 complex for structure determination (17). At the cellular level, ambiguity remains on whether this interaction occurs in *cis* (on the same cell) and/or in *trans* (on adjacent cells). Our experiments do not allow us to determine if the interactions we identify occur simultaneously or not, and whether they take place in *cis* and/or *trans*. It is possible that Jagged1 C2-EGF3 interacts in *trans* with the membrane-proximal Notch1 NRR, and can therefore engage with the membrane of the Notch1-presenting cell as previously suggested (33, 34). In contrast, the interaction between the two membrane-proximal sites, Jagged1 CRD and Notch1 EGF33-NRR, is more likely to occur in *cis,* as a direct Jagged1 CRD–Notch1 NRR *trans* interaction would possibly require very short intercellular distances.

In our experiments, we cannot conclude whether the interactions we report are intra- or intermolecular. To address this question, we separated the monomeric and oligomeric fractions of the cross-linked Jagged1 ectodomain by size-exclusion chromatography to discriminate between intra- and intermolecular cross-links. This shows that most Jagged1 cross-links identified in the oligomeric fraction are also present in the monomeric fraction, indicating that these represent intra-molecular interactions, i.e. within the same Jagged1 molecule. Furthermore, most of these cross-links are detected independently of the presence of Notch1 in the sample, suggesting that the Jagged1 intramolecular interactions take place prior to Notch1 engagement. In future research, cryo-electron tomography

could be employed on full-ectodomain or full-length complexes to identify *cis/trans* and homo-/heteromeric interactions in the Notch1-Jagged1 complex. For example, liposomes covalently decorated with Notch1 or Jagged1 ectodomains, using an alkyne-azide cycloaddition of azide-labeled proteins (43) into cyclooctyne-containing liposomes, could be used to study the Notch1-Jagged1 *trans*-cellular interaction, as well as the Notch1 and Jagged1 homomeric *cis* interactions.

This knowledge could be used in the development of molecules against pathologies involving Notch dysregulation by targeting the interactions we identify. For example, it is possible that the Notch1 NRR–Jagged1 C2-EGF3 interaction contributes to activate signaling by bringing the C2-EGF3 region of Jagged1 in proximity to the membrane of the Notch1-presenting cell, so that they can interact to optimally activate signaling (33, 34). Disrupting the Notch1 NRR–Jagged1 C2-EGF3 interaction, by using a therapeutic tool such as an anti-NRR antibody, would therefore block signaling, which would be beneficial in the case of a disease linked to Notch overactivation, such as T-cell acute lymphoblastic leukemia (T-ALL) (5). An antibody targeting the NRR was previously shown to stabilize the NRR closed conformation, therefore preventing signaling from an auto-activated mutant NRR (44). It would be interesting to test, using a biophysical technique such as SPR, whether this antibody could also block the Notch1 NRR–Jagged1 C2-EGF3 interaction, and to study the effect of this inhibition on signaling. Such experiments could also be performed on Notch-Jagged homologue complexes and validated in a cellular setting to ultimately understand the intricacies of Notch signaling and provide new therapeutic avenues for Notch-associated diseases.

Targeting cancer with anti-EGFR nanobodies

As a member of the human epidermal growth factor receptor (HER) tyrosine kinase family, the epidermal growth factor receptor (EGFR) is critical in cell proliferation, migration, and differentiation (45–47). EGFR activation is a relatively well understood mechanism in which ligand binding is coupled to the receptor ectodomain dimerization and asymmetric dimerization of the intracellular domains, one of which phosphorylates the other to initiate signaling (48). EGFR is the first HER family member shown to be overexpressed in cancer (49), and is therefore a key therapeutic target (50–52). New tools, such as nanobodies, may help to further characterize and treat EGFR-associated cancers. Here, we solve the structure of the EgB4 nanobody, both alone and in complex with the full ectodomain of EGF-bound EGFR. We reveal that EgB4 binds to a new epitope on EGFR domains I and II in the active dimeric EGFR-EGF complex, and we describe the non-inhibitory mechanism by which EgB4 interacts with EGFR.

EGFR-associated cancers are currently treated using monoclonal antibodies and tyrosine kinase inhibitors (TKIs). The most common EGFR oncogenic mutations are the L858R point mutation and exon 19 in-frame deletions, both frequently detected in lung cancer (53). These alterations affect the tyrosine kinase domain by destabilizing the auto-inhibited conformation and conferring an aberrant kinase activity (54, 55). TKIs are therefore molecules of choice in the treatment of such cancer (53). TKIs bind to the ATP recognition site in the EGFR kinase domain (56, 57), therefore blocking intracellular phosphorylation and interrupting downstream signaling pathways (58). By comparison, the EGFR variant III (EGFRvIII) is characterized by the deletion of amino acids 6-273 in the EGFR extracellular domain, and is the most common EGFR mutation observed in glioblastoma multiforme (GBM) (59). EGFRvIII leads to increased homodimerization, impaired downregulation and dysregulated tyrosine kinase activity (59–61). Monoclonal antibodies target the EGFR extracellular region and act by competitively binding to the ligand recognition site, therefore preventing the conformational rearrangement required to initiate downstream signaling (62–64). Among them, Cetuximab is administrated to treat metastatic colorectal cancer and head and neck squamous cell carcinoma (HNSCC), and constitutes the first anti-EGFR monoclonal antibody used in the clinic (64). However monoclonal antibodies fail to efficiently target oncogenic EGFR variants such as EGFRvIII (65).

Despite their wide clinical use, monoclonal antibodies are limited by their large size, leading to reduced tumor penetrator and slow distribution (66–68). To overcome such limitations, the variable domains of heavy chain antibodies (VHH), also called nanobodies in their isolated form, represent valuables tools due their small size and capacity to bind to antigens with a high affinity (69–71). Several nanobodies are currently being evaluated in clinical trials, and recently, a first nanobody was approved for clinical use, to treat thrombotic thrombocytopenic purpura (72).

Nanobodies binding to EGFR have previously been generated for diagnostic and therapeutic applications, including the nanobodies EgA1, 9G8 and 7D12 that inhibit ligand-dependent EGFR activation (73–77). While 7D12 competitively binds to the ligand recognition site located between EGFR domains I and III, EgA1 and 9G8 interact with EGFR domain III, sterically preventing the conformational rearrangement required for signaling (78). By comparison, our structure shows that the EgB4 nanobody binds to a new epitope located on EGFR domains I and II, through interactions with EgB4 complementarity determining regions (CDR) 2 and 3. In particular, a hydrophobic pocket is formed by tryptophan and phenylalanine residues from the top of EGFR domain I, and tryptophan residues in the CDR2 and CDR3 of EgB4. Furthermore, aspartic acid residues from the CDR3 of EgB4 are involved in electrostatic interactions with Arg141 from EGFR domain I. The EGFR-EgB4 interface is extended by hydrogen bonding interactions between Arg105 from EgB4 CDR3 and backbone carbonyl groups of Lys188, Ile189 and Cys191 at the EGFR domain I-II junction.

In addition to binding to the active EGFR, we hypothesize that EgB4 can also engage the inactive receptor. Indeed, a structural alignment of the interface residues from EGFR in the active and inactive conformations suggests that there is no interface rearrangement upon conformational change. The interactions we identify between EgB4 and the active EGFR are therefore likely to also occur with the inactive receptor. In the modeled inactive EGFR-EgB4 complex, we observe additional hydrogen bonding interactions involving Gln193 and Asn172, located on EGFR domain II, and residues from EgB4 CDR2 and CDR3. It is possible that these interactions are also present in the active EGFR-EgB4 complex, but not observed in the structure due to the low resolution of the underlying data.

The therapeutic potential of nanobodies can be improved by using protein engineering tools through three platforms (69). First, nanobodies can be fused to effector domains. For example, fusing a Fc domain to a nanobody can trigger antibody-dependent cell-mediated cytotoxicity (ADCC) of targeted cells for enhanced antitumor effect (79). Second, nanobodies can be conjugated to nanoparticles, such as liposomes, that encapsulate drugs, thereby directing drug delivery to specific target cells. Release of drugs from the internalized liposomes can then be achieved by intracellular degradation of the liposomes. As an example, nanoparticles encapsulating doxorubicin and coupled to the anti-EGFR EgA1 nanobody showed significantly improved antitumor effect and prolonged survival *in vivo* compared to the untargeted drug (80). Third, "naked" nanobodies can act as antagonists that interfere with receptor activation. For example, the 7D12 nanobody competitively inhibits EGFR activation (78). Nanobodies can also be linked to each other using flexible linkers to create multivalent molecules. This can be employed to create mono-specific nanobodies, e.g. two copies of the same nanobody fused to each other, or biparatopic nanobodies, e.g. two nanobodies targeting different epitopes on the same antigen. Illustrating the clinical potential of this strategy, the trivalent biparatopic anti-EGFR nanobody 7D12-9G8-Alb was shown to inhibit tumor growth *in vivo* (75).

In contrast to anti-EGFR antagonist molecules, the EgB4 nanobody does not inhibit EGFR activation. Indeed, EgB4 binds to the active EGFR-EGF complex as well as to the inactive receptor, while not sterically preventing EGFR from alternating between the active and inactive conformations. The EgB4 epitope on EGFR domains I and II is located relatively far from that of the other anti-EGFR nanobodies described (78), preventing the straightforward design of a multivalent molecule that includes EgB4 in combination with these other EGFR-targeting nanobodies. However, in the active dimeric EGFR-EgB4-EGF complex the two EgB4 molecules are in proximity, with their C-termini situated on top of the complex and separated by 24.2 Å. This provides the opportunity to create a bivalent mono-specific EgB4-EgB4 molecule, by linking the two EgB4 nanobodies through their C-termini (81–83). This molecule could exhibit an increased affinity for the EGFR-EGF complex, and maintain the complex in a dimeric state. In addition, the individual EgB4 nanobody could be used to

identify specific EGFR variants when used in combination with other anti-EGFR nanobodies that target a different epitope on EGFR. For example, EgB4 could bind to EGFRvII, as this variant is truncated from part of domain IV, while not interacting with EGFRvIII, which lacks most of the EGFR domains I and II. By comparison, a nanobody that binds to EGFR domain IV would be able to interact with EGFRvIII but not with EGFRvII, allowing us to distinguish between tumor cells which express either EGFRvIII or EGFRvII. Together, this shows that EgB4 could help identify tumors that are characterized by specific mutations, and might prove useful to target EGFR-associated cancers.

Functional comparison between Laspartomycin/Friulimicin and Daptomycin

The growing threat of multidrug resistant bacteria is a top priority of the World Health Organization (WHO) (84). An effective way to address this threat is to identify and develop antibiotics that operate using novel mechanisms of action as compared to long-used antibiotics (85). Notable in this regard are the calcium-dependent antibiotics (CDAs), which provide a remarkable source of molecules abundant in mechanistic diversity (86). Among these, the lipopeptide Daptomycin is widely used for the management of multidrug resistant bacteria since it entered the clinic in 2003 (87). Despite its clinical success, the mode of action of Daptomycin remains a topic of investigation (88–90). By comparison, the mechanism of other CDAs such as Laspartomycin C, Friulimicin B and Amphomycin are better explained (91–94). Recently, a crystal structure of Laspartomycin C in complex with C_{10}-P (a more soluble analogue of the C_{55}-P bacterial target) was solved by X-ray crystallography (93). This is the first structure reported for a CDA in complex with its biomolecular target. The structure shows a saddle-shaped Laspartomycin C molecule bound to one C_{10}-P molecule and to two calcium ions that mediate ligand engagement (93). This ternary complex symmetrically dimerizes through direct and indirect interactions, with the C_{10}-P molecules inserted into the dimer cavity and therefore sequestered from the solvent. From this, a simple model can be proposed in which the Laspartomycin C dimer fatty acid sidechain and the C_{10}-P isoprenyl tail point down towards the bacterial membrane and are oriented perpendicular to it, suggesting that Laspartomycin C is slightly embedded into the membrane. The Friulimicin/Amphomycin classes of CDAs share structural similarities with Laspartomycin C and also engage C_{55}-P. Subtle differences in the macrocycle between Laspartomycin C and Friulimicin/Amphomycin, coupled to the knowledge provided by the Laspartomycin C structure, prompted us to study the impact of introducing structural features from Friulimicin/Amphomycin into Laspartomycin C.

Here we present the design, synthesis and evaluation of such CDA analogues. Structure-activity studies combined with high-resolution crystal structures of two analogues in

complex with Ca^{2+} and C_{10}-P reveal an interplay between residues 4, 9 and 10 in the peptide macrocycle. These residues contribute to the formation of a higher-order arrangement in the crystal packing, which provides an explanation for the CDAs antibacterial effect. Specifically, interactions between Asp^4 from one dimer and D-Dap^9 from an adjacent dimer serve to form this higher-order arrangement. The same Asp^4 residue also interacts with a calcium ion coordinated by the second dimer, further stabilizing this arrangement. These two dimers form the same interactions with Asp^4 and D-Dap^9 from a third dimer, therefore creating a trimer of dimers not observed in the crystal packing of Laspartomycin C which was crystallized in similar conditions. This trimer of dimers is further stabilized by hydrophobic interactions between the sidechain of Ile^{10} or Val^{10} residues at the center of the trimer. Notably, the substitution of Ile by Val at position 10 of the peptide macrocycle results in an 8-fold increase in activity. Our structures show that the slightly less bulky Val^{10} sidechain might more optimally fulfill the steric requirements of the hydrophobic pocket, therefore providing a structural explanation for this difference in activity. In the crystal, this repeating "trimer of dimer" motif forms a two-dimensional layer not observed in Laspartomycin C. We observe alternating layers in the crystal, with one peptide macrocycle layer inducing a strong packing in *cis* (within the same layer), sandwiched by a hydrophobic layer composed of lipids (the lipopeptide fatty acid sidechain and C_{10}-P), and by a hydrophilic layer constituted of water molecules, both inducing a weak packing in *trans* (between adjacent layers). From this a straightforward model can be derived, in which a two-dimensional layer of lipopeptide is partially embedded into the bacterial membrane (represented by the hydrophobic layer in the crystal) where it sequesters C_{55}-P molecules to inhibit the bacterial cell wall biosynthesis.

To gain insights into the mode of action of Laspartomycin C and our lipopeptide analogue 6 with regards to Daptomycin, we performed a bacterial cytological profiling on the model organism *Bacillus subtilis* reporter strains expressing GFP-tagged proteins. These proteins are involved in various cellular processes such as DNA replication, transcription, translation, ATP synthesis, cell division, cell division regulation, cell wall synthesis coordination, and peptidoglycan synthesis. Proteins involved in cell division, and in the synthesis of DNA, RNA, protein and ATP, were not affected by Laspartomycin C, lipopeptide 6 or Daptomycin. Similarly, the MreB protein, involved in cell wall synthesis coordination, is delocalized by Laspartomycin C and lipopeptide 6, as previously reported with Daptomycin (90). By contrast, Daptomycin, but not Laspartomycin C or lipopeptide 6, delocalizes the MinD protein, which plays a role in cell division regulation, from the bacterial membrane. Further highlighting the differences in Laspartomycin/lipopeptide 6 versus Daptomycin mechanisms of action, the MurG protein, which is involved in peptidoglycan synthesis, is detached and delocalized from the membrane by Daptomycin, while Laspartomycin C and lipopeptide 6 appear to dissolve MurG clusters so that proteins diffuse along the membrane. This specificity of Laspartomycin C and lipopeptide 6 could be exploited in the design of

future antibiotics with modes of action differing from that of Daptomycin. In addition, a recent report suggested that in the presence of phosphatidylglycerol, Daptomycin can interact with C_{55}-P, C_{55}-PP and the peptidoglycan precursor lipid II (95), therefore causing a major displacement of lipids in the bacterial membrane. This correlates with the delocalization of the membrane proteins MurG and MinD by Daptomycin, as observed by live cell imaging (90). On the other hand, Laspartomycin C more specifically affects lipid II synthesis while probably not directly interfering with the cell membrane arrangement. Together, this shows that Laspartomycin C and lipopeptide 6 have a narrower range of cellular targets compared to Daptomycin. This is in line with our bacterial cytological profiling that indicates an alteration by Laspartomycin C of the MreB and MurG proteins, both involved in lipid II synthesis. Collectively, our data show that Laspartomycin and Friulimicin/Amphomycin constitute promising classes of CDAs that operate using a new mode of action, which may ultimately contribute to fight against multidrug resistant bacteria.

Acknowledgements

Thanks to Alex Cloherty and Robert Englmeier for critically proofreading this chapter. Their feedback was very useful and much appreciated.

References

1. A. Louvi, S. Artavanis-Tsakonas, Notch signalling in vertebrate neural development. *Nat. Rev. Neurosci.* **7**, 93–102 (2006).
2. S. J. Bray, Notch signalling in context. *Nat. Rev. Mol. Cell Biol.* **9**, 722–735 (2016).
3. E. C. Lai, Notch signaling: control of cell communication and cell fate. *Development* **131**, 965–73 (2004).
4. C. Siebel, U. Lendahl, Notch Signaling in Development, Tissue Homeostasis, and Disease. *Physiol. Rev.* **97**, 1235–1294 (2017).
5. A. P. Weng, *et al.*, Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science (80-.).* **306**, 269–271 (2004).
6. J. C. Aster, W. S. Pear, S. C. Blacklow, The Varied Roles of Notch in Cancer. *Annu. Rev. Pathol. Mech. Dis.* **12**, 245–275 (2017).
7. J. Mašek, E. R. Andersson, The developmental biology of genetic Notch disorders. *Development* **144**, 1743 LP – 1763 (2017).
8. I. Rebay, *et al.*, Specific EGF repeats of Notch mediate interactions with Delta and serrate: Implications for notch as a multifunctional receptor. *Cell* **67**, 687–699 (1991).
9. W. R. Gordon, *et al.*, Structural basis for autoinhibition of Notch. *Nat. Struct. Mol. Biol.* **14**, 295–300 (2007).
10. K. N. Lovendahl, S. C. Blacklow, W. R. Gordon, "The Molecular Mechanism of Notch Activation" in *Molecular Mechanisms of Notch Signaling*, T. Borggrefe, B. D. Giaimo, Eds. (Springer International Publishing, 2018), pp. 47–58.
11. P. A. Handford, B. Korona, R. Suckling, C. Redfield, S. M. Lea, "Structural Insights into Notch Receptor-Ligand Interactions" in *Molecular Mechanisms of Notch Signaling*, T. Borggrefe, B. D. Giaimo, Eds. (Springer International Publishing, 2018), pp. 33–46.
12. C. Brou, *et al.*, A novel proteolytic cleavage involved in Notch signaling: The role of the disintegrin-metalloprotease TACE. *Mol. Cell* **5**, 207–216 (2000).
13. J. S. Mumm, *et al.*, A ligand-induced extracellular cleavage regulates γ-secretase-like proteolytic activation of Notch1. *Mol. Cell* **5**, 197–206 (2000).
14. F. Logeat, *et al.*, The Notch1 receptor is cleaved constitutively by a furin-like convertase. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8108–8112 (1998).
15. J. Cordle, *et al.*, Localization of the delta-like-1-binding site in human Notch-1 and its modulation by calcium affinity. *J. Biol. Chem.* **283**, 11785–11793 (2008).
16. F. Chowdhury, *et al.*, Defining single molecular forces required for notch activation using nano yoyo. *Nano Lett.* **16**, 3892–3897 (2016).
17. V. C. Luca, *et al.*, Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity. *Science* **355**, 1320–1324 (2017).
18. L. Meloty-Kapella, B. Shergill, J. Kuon, E. Botvinick, G. Weinmaster, Notch Ligand Endocytosis Generates Mechanical Pulling Force Dependent on Dynamin, Epsins, and Actin. *Dev. Cell* **22**, 1299–1312 (2012).

19. D. Seo, *et al.*, A mechanogenetic toolkit for interrogating cell signaling in space and time. *Cell* **165**, 1507–1518 (2016).
20. X. Wang, T. Ha, Defining single molecular forces required to activate integrin and Notch signaling. *Science (80-.).* **340**, 991–994 (2013).
21. W. R. Gordon, *et al.*, Mechanical Allostery: Evidence for a Force Requirement in the Proteolytic Activation of Notch. *Dev. Cell* **33**, 729–736 (2015).
22. D. Del Álamo, H. Rouault, F. Schweisguth, Mechanism and significance of cis-inhibition in notch signalling. *Curr. Biol.* **21**, 40–47 (2011).
23. B. D'Souza, A. Miyamoto, G. Weinmaster, The many facets of Notch ligands. *Oncogene* **27**, 5148–5167 (2008).
24. D. Sprinzak, *et al.*, Cis-interactions between Notch and Delta generate mutually exclusive signalling states. *Nature* **465**, 86–90 (2010).
25. N. Nandagopal, L. A. Santat, M. B. Elowitz, Cis-activation in the Notch signaling pathway. *Elife* **8**, 1–34 (2019).
26. S. Kakuda, R. S. Haltiwanger, Deciphering the Fringe-Mediated Notch Code: Identification of Activating and Inhibiting Sites Allowing Discrimination between Ligands. *Dev. Cell* **40**, 193–201 (2017).
27. R. Rampal, J. F. Arboleda-Velasquez, A. Nita-Lazar, K. S. Kosik, R. S. Haltiwanger, Highly conserved O-fucose sites have distinct effects on Notch1 function. *J. Biol. Chem.* **280**, 32133–32140 (2005).
28. A. Xu, L. Lei, K. D. Irvine, Regions of Drosophila notch that contribute to ligand binding and the modulatory influence of Fringe. *J. Biol. Chem.* **280**, 30158–30165 (2005).
29. N. Lawrence, T. Klein, K. Brennan, A. Martinez Arias, Structural requirements for notch signalling with delta and serrate during the development and patterning of the wing disc of Drosophila. *Development* **127** (2000).
30. A. Sharma, A. Rangarajan, R. R. Dighe, Antibodies against the extracellular domain of human Notch1 receptor reveal the critical role of epidermal-growth-factor-like repeats 25-26 in ligand binding and receptor activation. *Biochem. J.* **449**, 519–30 (2013).
31. Z. Pei, N. E. Baker, Competition between Delta and the Abruptex domain of Notch. *BMC Dev. Biol.* **8**, 4 (2008).
32. J. Cordle, *et al.*, A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition. *Nat. Struct. Mol. Biol.* **15**, 849–857 (2008).
33. C. R. Chillakuri, *et al.*, Structural Analysis Uncovers Lipid-Binding Properties of Notch Ligands. *Cell Rep.* **5**, 861–867 (2013).
34. R. J. Suckling, *et al.*, Structural and functional dissection of the interplay between lipid and Notch binding by human Notch ligands. *EMBO J.* **36**, 2204–2215 (2017).
35. T. Kiyota, T. Kinoshita, Cysteine-rich region of X-Serrate-1 is required for activation of Notch signaling in Xenopus primary neurogenesis. *Int. J. Dev. Biol.* **46**, 1057–60 (2002).
36. C. Hicks, *et al.*, A secreted Delta1-Fc fusion protein functions both as an activator and inhibitor of Notch1 signaling. *J. Neurosci. Res.* **68**, 655–667 (2002).

37. J. T. Nichols, *et al.*, DSL ligand endocytosis physically dissociates Notch1 heterodimers before activating proteolysis can occur. *J. Cell Biol.* **176**, 445–458 (2007).
38. M. Vooijs, E. H. Schroeter, Y. Pan, M. Blandford, R. Kopan, Ectodomain shedding and intramembrane cleavage of mammalian Notch proteins is not regulated through oligomerization. *J. Biol. Chem.* **279**, 50864–50873 (2004).
39. D. F. Kelly, *et al.*, Molecular Structure and Dimeric Organization of the Notch Extracellular Domain as Revealed by Electron Microscopy. *PLoS One* **5**, 1–6 (2010).
40. X. Xu, *et al.*, Insights into Autoregulation of Notch3 from Structural and Functional Studies of Its Negative Regulatory Region. *Structure* **23**, 1227–1235 (2015).
41. K. Shimizu, Integrity of intracellular domain of Notch ligand is indispensable for cleavage required for release of the Notch2 intracellular domain. *EMBO J.* **21**, 294–302 (2002).
42. K. Sakamoto, W. S. Chao, K. I. Katsube, A. Yamaguchi, Distinct roles of EGF repeats for the Notch signaling system. *Exp. Cell Res.* **302**, 281–291 (2005).
43. J. Z. Yao, *et al.*, Fluorophore Targeting to Cellular Proteins via Enzyme-Mediated Azide Ligation and Strain-Promoted Cycloaddition. *J. Am. Chem. Soc.* **134**, 3720–3728 (2012).
44. Y. Wu, *et al.*, Therapeutic antibody targeting of individual Notch receptors. *Nature* **464**, 1052–1057 (2010).
45. C. L. Arteaga, J. A. Engelman, ERBB receptors: From oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* **25**, 282–303 (2014).
46. K. M. Ferguson, *et al.*, EGF Activates Its Receptor by Removing Interactions that Autoinhibit Ectodomain Dimerization molecules. Crystallographic studies of several ligand-bound RTK extracellular fragments have allowed visual-ization of such dimers, including those from the Flt-1. *Mol. Cell* **11**, 507–517 (2003).
47. Y. Yarden, M. X. Sliwkowski, Untangling the ErbB network. *Nat. Rev. Mol. Cell Biol.* **2**, 127–137 (2001).
48. M. A. Lemmon, J. Schlessinger, Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
49. D. M. Thompson, G. N. Gill, The EGF receptor: structure, regulation and potential role in malignancy. *Cancer Surv.* **4**, 767–788 (1985).
50. P. Wee, Z. Wang, Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers (Basel).* **9**, 1–45 (2017).
51. P. Seshacharyulu, *et al.*, Targeting the EGFR signaling pathway in cancer therapy. *Expert Opin. Ther. Targets* **16**, 15–31 (2012).
52. Y. Yarden, G. Pines, The ERBB network: At last, cancer therapy meets systems biology. *Nat. Rev. Cancer* **12**, 553–563 (2012).
53. H. Shigematsu, *et al.*, Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J. Natl. Cancer Inst.* **97**, 339–346 (2005).
54. C. H. Yun, *et al.*, Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity. *Cancer Cell* **11**, 217–227 (2007).

55. X. Zhang, J. Gureasko, K. Shen, P. A. Cole, J. Kuriyan, An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell* **125**, 1137–1149 (2006).
56. J. Stamos, M. X. Sliwkowski, C. Eigenbrot, Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.* **277**, 46265–46272 (2002).
57. S. Yoshikawa, *et al.*, Structural basis for the altered drug sensitivities of non-small cell lung cancer-associated mutants of human epidermal growth factor receptor. *Oncogene* **32**, 27–38 (2013).
58. I. Solassol, F. Pinguet, X. Quantin, FDA- and EMA-approved tyrosine kinase inhibitors in advanced EGFR-mutated non-small cell lung cancer: Safety, tolerability, plasma concentration monitoring, and management. *Biomolecules* **9** (2019).
59. L. Frederick, X. Y. Wang, G. Eley, C. D. James, Diversity and frequency of epidermal growth factor receptor mutations in human glioblastomas. *Cancer Res.* **60**, 1383–1387 (2000).
60. S. I. Ymer, *et al.*, Glioma specific extracellular missense mutations in the first cysteine rich region of epidermal growth factor receptor (EGFR) initiate ligand independent activation. *Cancers (Basel).* **3**, 2032–2049 (2011).
61. R. Nishikawa, *et al.*, A mutant epidermal growth factor receptor common in human glioma confers enhanced tumorigenicity. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7727–7731 (1994).
62. S. Li, *et al.*, Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* **7**, 301–311 (2005).
63. E. A. Sickmier, *et al.*, The panitumumab EGFR complex reveals a binding mechanism that overcomes cetuximab induced resistance. *PLoS One* **11**, 1–11 (2016).
64. W. Q. Cai, *et al.*, The Latest Battles Between EGFR Monoclonal Antibodies and Resistant Tumor Cells. *Front. Oncol.* **10** (2020).
65. Z. An, O. Aksoy, T. Zheng, Q. W. Fan, W. A. Weiss, Epidermal growth factor receptor and EGFRvIII in glioblastoma: Signaling pathways and targeted therapies. *Oncogene* **37**, 1561–1575 (2018).
66. S. I. Rudnick, G. P. Adams, Affinity and avidity in antibody-based tumor targeting. *Cancer Biother. Radiopharm.* **24**, 155–161 (2009).
67. C. G. Siontorou, Nanobodies as novel agents for disease diagnosis and therapy. *Int. J. Nanomedicine* **8**, 4215–4227 (2013).
68. J. H. E. Baker, *et al.*, Direct visualization of heterogeneous extravascular distribution of trastuzumab in human epidermal growth factor receptor type 2 overexpressing xenografts. *Clin. Cancer Res.* **14**, 2171–2179 (2008).
69. M. Kijanka, B. Dorresteijn, S. Oliveira, P. M. P. Van Bergen En Henegouwen, Nanobody-based cancer therapy of solid tumors. *Nanomedicine* **10**, 161–174 (2015).
70. E. Y. Yang, K. Shah, Nanobodies: Next Generation of Cancer Diagnostics and Therapeutics. *Front. Oncol.* **10** (2020).
71. E. De Genst, *et al.*, Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4586–4591 (2006).
72. M. Scully, *et al.*, Caplacizumab Treatment for Acquired Thrombotic Thrombocytopenic Purpura. *N. Engl. J. Med.* **380**, 335–346 (2019).

73. R. C. Roovers, *et al.*, Efficient inhibition of EGFR signalling and of tumour growth by antagonistic anti-EGFR Nanobodies. *Cancer Immunol. Immunother.* **56**, 303–317 (2007).
74. E. G. Hofman, *et al.*, EGF induces coalescence of different lipid rafts. *J. Cell Sci.* **121**, 2519–2528 (2008).
75. R. C. Roovers, *et al.*, A biparatopic anti-EGFR nanobody efficiently inhibits solid tumour growth. *Int. J. Cancer* **129**, 2013–2024 (2011).
76. S. Oliveira, *et al.*, Downregulation of EGFR by a novel multivalent nanobody-liposome platform. *J. Control. Release* **145**, 165–175 (2010).
77. R. Van Der Meel, *et al.*, Tumor-targeted Nanobullets: Anti-EGFR nanobody-liposomes loaded with anti-IGF-1R kinase inhibitor for cancer treatment. *J. Control. Release* **159**, 281–289 (2012).
78. K. R. Schmitz, A. Bagchi, R. C. Roovers, P. M. P. Van Bergen En Henegouwen, K. M. Ferguson, Structural evaluation of EGFR inhibition mechanisms for nanobodies/VHH domains. *Structure* **21**, 1214–1224 (2013).
79. P. Bannas, *et al.*, In vivo near-infrared fluorescence targeting of T cells: Comparison of nanobodies and conventional monoclonal antibodies. *Contrast Media Mol. Imaging* **9**, 135–142 (2014).
80. M. Talelli, *et al.*, Intrinsically active nanobody-modified polymeric micelles for tumor-targeted combination therapy. *Biomaterials* **34**, 1255–1260 (2013).
81. M. D. Witte, *et al.*, Preparation of unnatural N-to-N and C-to-C protein fusions. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11993–11998 (2012).
82. M. Rashidian, *et al.*, Enzyme-Mediated Modification of Single-Domain Antibodies for Imaging Modalities with Different Characteristics. *Angew. Chemie–Int. Ed.* **55**, 528–533 (2016).
83. B. Zang, *et al.*, Freezing-assisted synthesis of covalent C-C linked bivalent and bispecific nanobodies. *Org. Biomol. Chem.* **17**, 257–263 (2019).
84. P. Beyer, S. Paulin, Priority pathogens and the antibiotic pipeline: An update. *Bull. World Health Organ.* **98**, 151 (2020).
85. M. F. Chellat, L. Raguž, R. Riedl, Targeting Antibiotic Resistance. *Angew. Chemie–Int. Ed.* **55**, 6600–6626 (2016).
86. T. M. Wood, N. I. Martin, The calcium-dependent lipopeptide antibiotics: Structure, mechanism, & medicinal chemistry. *Medchemcomm* **10**, 634–646 (2019).
87. M. Heidary, *et al.*, Daptomycin. *J. Antimicrob. Chemother.* **73**, 1–11 (2018).
88. P. T. Hart, *et al.*, A combined solid- and solution-phase approach provides convenient access to analogues of the calcium-dependent lipopeptide antibiotics. *Org. Biomol. Chem.* **12**, 913–918 (2014).
89. S. D. Taylor, M. Palmer, The action mechanism of daptomycin. *Bioorganic Med. Chem.* **24**, 6253–6268 (2016).
90. A. Müller, *et al.*, Daptomycin inhibits cell envelope synthesis by interfering with fluid membrane microdomains. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E7077–E7086 (2016).
91. T. Schneider, *et al.*, The lipopeptide antibiotic friulimicin B inhibits cell wall biosynthesis through complex formation with bactoprenol phosphate. *Antimicrob. Agents Chemother.* **53**, 1610–1618 (2009).

92. L. H. J. Kleijn, *et al.*, Total Synthesis of Laspartomycin C and Characterization of Its Antibacterial Mechanism of Action. *J. Med. Chem.* **59**, 3569–3574 (2016).
93. L. H. J. Kleijn, *et al.*, A High-Resolution Crystal Structure that Reveals Molecular Details of Target Recognition by the Calcium-Dependent Lipopeptide Antibiotic Laspartomycin C. *Angew. Chemie–Int. Ed.* **56**, 16546–16549 (2017).
94. A. Diehl, T. M. Wood, S. Gebhard, N. I. Martin, G. Fritz, The cell envelope stress response of bacillus subtilis towards laspartomycin C. *Antibiotics* **9**, 1–16 (2020).
95. F. Grein, *et al.*, Ca2+-Daptomycin targets cell wall biosynthesis by forming a tripartite complex with undecaprenyl-coupled intermediates and membrane lipids. *Nat. Commun.* **11**, 1–11 (2020).

Summary

Proteins are universal molecular machines that are involved in the biological processes of all living organisms, from animals to bacteria. Proteins are constituted of building blocks, called amino acids, that are linearly attached to each other. Despite this apparent simplicity, proteins achieve an extraordinary diversity of functions such as cell differentiation and immune response. To do so, proteins adopt a unique three-dimensional shape that allows them to perform a specific task defining their function. A certain type of proteins, called cell surface receptors, play an important role by mediating communication to and from the cell. They sense the cell environment and transmit the received messages into the cell, but can also send messages to other cells. These messages can be exchanged between neighbor cells by direct protein contacts, or between distant cells by secretion of molecules that can travel from one cell to another. These processes are tightly regulated throughout the body, and if messages are incorrectly received or sent, diseases may develop. Proteins can also be used to treat pathologies arising from such dysfunction, or from infection by a pathogen. This is the case of the lipopeptide calcium-dependent antibiotics, which help fight bacterial infections. These compounds target specific molecules on the bacterial membrane and operate by destabilizing the membrane or by preventing the cell wall synthesis. In this thesis, we combine diverse structural biology approaches and use tools with clinical potential to gain insight into the structure and function of cell surface receptors and lipopeptide calcium-dependent antibiotics, to ultimately help develop therapeutic molecules to treat the associated diseases.

In chapter 1, I present the different cell surface receptors we focused on, namely Notch1 and epidermal growth factor receptor (EGFR), and I introduce the Laspartomycin and Friulimicin/Amphomycin classes of lipopeptide calcium-dependent antibiotics.

In chapter 2, we investigate the molecular mechanism of the Notch1 receptor activation by its canonical ligand Jagged1. The Notch signaling pathway is a central cell-cell communication system involved in various processes such as cell fate determination, stem cell maintenance, immune system regulation and angiogenesis. Dysregulation of Notch signaling leads to a variety of diseases, including congenital disorders and cancers. Structural investigations on the Notch1 extracellular region have been hampered due to difficulties in producing and isolating the recombinant protein in sufficient amount. The large Notch1 ectodomain is highly glycosylated and has a very modular architecture that

includes 36 epidermal growth factor (EGF)-like repeats containing 6 cysteine residues each. Here we recombinantly express and purify the Notch1 and Jagged1 ectodomains produced in mammalian cells. Using cross-linking mass spectrometry (XL-MS), and biophysical and structural techniques on the Notch1-Jagged1 full ectodomain complex as well as on targeted sites, we identify several unreported binding regions that form an interaction network in the Notch1-Jagged1 complex. Specifically, Notch1 EGF33-negative regulatory region (NRR) interacts intramolecularly with Notch1 EGF8-13, and intermolecularly with C2-EGF3, EGF8-11 and cysteine-rich domain (CRD) of Jagged1. XL-MS and quantitative binding experiments indicate that the three sites on Jagged1 also engage intramolecularly. These data, coupled to small-angle X-ray scattering (SAXS) experiments describing the dimensions and flexibility of Notch1 and Jagged1 ectodomains, support the formation of non-extended architectures. Together, this suggests that critical regions on Notch1 and Jagged1 are not distal as previously thought, but directly engage to control Notch1 activation, thereby redefining the Notch1-Jagged1 activation model and opening new avenues for therapeutic applications.

In chapter 3, we describe the non-inhibitory mechanism by which the EgB4 nanobody interacts with EGFR. The human epidermal growth factor receptor (HER) family, of which EGFR is the founding member, is critical in cell proliferation, migration, and differentiation. EGFR is overexpressed in various cancers including glioblastoma multiforme (GBM) and non-small-cell lung cancers (NSCLC) and is therefore an important therapeutic target. Anti-EGFR inhibitory nanobodies, which prevent EGFR activation, were developed for therapeutic purposes. A structural investigation showed that these nanobodies prevent EGFR activation either by competitively binding to the ligand recognition site on EGFR, or by sterically blocking the EGFR conformational change that is required for activation. Here we solve the X-ray crystal structure of the non-inhibitory EgB4 nanobody, both alone and in complex with the active EGF-bound EGFR. We show that EgB4 binds to a new epitope on EGFR domains I and II in the active dimeric EGF-bound EGFR. Structural alignment of the interface residues in the active and inactive EGFR indicates that no conformational rearrangement of the interface occurs when EGFR switches from one conformation to the other. This suggests that, unlike inhibitory nanobodies, EgB4 can engage both the active and inactive EGFR, while not preventing EGFR from alternating between these conformations. Together this shows that EgB4 could be used as a biomarker for tumor imaging while not affecting EGFR function, which might prove useful to develop medicines that treat EGFR-associated cancers.

In chapter 4, we present the design, synthesis, and structural and functional evaluations of lipopeptide calcium-dependent antibiotic (CDA) analogues. CDAs represent an emerging class of molecules used to treat infections by Gram-positive bacteria, as illustrated by the clinically used Daptomycin. While the mechanism by which Daptomycin targets

bacteria remains unclear, the mode of action of CDAs Laspartomycin C and Friulimicin/Amphomycin are better understood. Recently, the crystal structure of Laspartomycin C in complex with C_{10}-P (a more soluble analogue of the lipid carrier C_{55}-P) was solved, for the first time revealing the atomic details of a CDA bound to its bacterial target. Laspartomycin C sequesters C_{55}-P, which is involved in the synthesis of the peptidoglycan, therefore preventing the biosynthesis of the bacterial cell wall. The Friulimicin/Amphomycin classes of CDAs are structurally similar to Laspartomycin C and also target C_{55}-P, which prompted us to evaluate the impact of introducing features from Friulimicin/Amphomycin into Laspartomycin C. We produced and solved high-resolution structures of two CDA analogues in complex with C_{55}-P, revealing an interplay between residues 4, 9 and 10 in the peptide macrocycle. In the crystal, these residues, which differ from those of Laspartomycin C, control the formation of a higher-order assembly not seen in the Laspartomycin C crystal under similar conditions, thus providing an explanation for the antibacterial activity. In addition, live cell imaging provides insight into the C_{55}-P-targeting family of antibiotics, and highlights a unique mode of action of Laspartomycin and Friulimicin/Amphomycin relative to Daptomycin.

In chapter 5, I discuss the implications of our findings on the molecular mechanisms by which the cell surface receptors Notch1 and EGFR, as well as our lipopeptide CDA analogues, play a role in health and disease, and I elaborate on how this knowledge might ultimately help develop new therapeutics.

Together, these studies contribute to gain a better understanding of essential biological systems and may help to develop new molecules for therapeutic applications, therefore illustrating the importance of structural biology in fundamental as well as in applied research.

Samenvatting

Eiwitten zijn universele moleculaire machines die betrokken zijn bij de biologische processen van alle levende organismen, van dieren tot bacteriën. Eiwitten bestaan uit bouwblokken, aminozuren genaamd, die lineair met elkaar verbonden zijn. Ondanks deze ogenschijnlijke eenvoudigheid, bewerkstelligen eiwitten een buitengewone diversiteit van functies, zoals differentiatie van cellen en immuunrespons. Om dit te doen, nemen eiwitten unieke driedimensionale vormen aan die ze in staat stellen de specifieke taken uit te voeren die hun functie bepalen. Celoppervlaktereceptoren zijn belangrijke eiwitten die communicatie van en naar de cel bewerkstelligen. Ze voelen de omgeving van de cel aan en ontvangen signalen voor de cel, maar kunnen ook signalen naar andere cellen sturen. Deze berichten kunnen uitgewisseld worden tussen nabije cellen door direct contact tussen de receptoren of over langere afstanden via uitgescheiden moleculen die zich van de ene cel naar de andere kunnen verplaatsen. Deze processen zijn strak gereguleerd in het lichaam, en wanneer boodschappen incorrect worden verstuurd of ontvangen, kan dit aanleiding geven tot ziekten. Eiwitten kunnen ook gebruikt worden om aandoeningen ontstaan door zulk disfunctioneren of infectie door een pathogeen te behandelen. Dit is het geval bij calciumafhankelijke lipopeptide-antibiotica, die gebruikt worden voor het bestrijden van bacteriële infecties. Deze middelen binden specifieke moleculen op het bacteriële membraan en destabiliseren op die manier het membraan, of voorkomen zo synthese van de celwand. In deze thesis combineren we diverse structureel-biologische methodes en gebruiken we moleculaire gereedschappen met potentiële klinische toepassingen om inzicht te verwerven in de structuren en het functioneren van celoppervlaktereceptoren en calciumafhankelijke lipopeptide-antibiotica, om uiteindelijk bij te dragen aan de ontwikkeling van therapeutische moleculen om geassocieerde ziekten te kunnen behandelen.

In hoofdstuk 1 presenteer ik de verscheidene celoppervlaktereceptoren waarop we ons hebben gericht, namelijk Notch1 en epidermale groeifactor-receptor (EGFR) en introduceer ik de Laspartomycin en Friulimicin/Amphomycin klassen van calciumafhankelijke lipopeptide-antibiotica.

In hoofdstuk 2 onderzoeken we het moleculaire mechanisme van activatie van de receptor Notch1 door zijn canonieke ligand Jagged1. Signalering van Notch1 is een centraal communicatiesysteem tussen cellen dat betrokken is in verscheidene processen,

zoals celdifferentiatie, onderhoud van stamcellen, regulatie van het immuunsysteem en angiogenese. Disregulatie van signalering door Notch leidt tot verscheidene ziekten, waaronder aangeboren ziekten en kankers. Structureel-biologisch onderzoek op het extracellulaire deel van Notch wordt belemmerd door moeilijkheden in het verkrijgen van recombinant eiwit in voldoende hoeveelheden. Het grote ectodomein van Notch1 is zeer geglycosyleerd en heeft een erg modulaire structuur, bestaande uit 36 epidermale groeifactor (EGF)-achtige domeinen, die elk 6 cysteïneresiduen bevatten. Wij gebruiken recombinante expressie in humane cellen en zuiveren de ectodomeinen van Notch1 en Jagged1. Met behulp van *cross-linking* massaspectrometrie (XL-MS), en biofysische en structurele technieken op zowel het Notch1-Jagged1 complex, als ook gericht gekozen segmenten, identificeren we meerdere bindingsplaatsen die niet eerder gemeld waren. Samen vormen ze een interactienetwerk in het Notch1-Jagged1 complex. Specifiek gesproken interacteert het EGF33-negatief regulerende regio (NRR) van Notch1 intramoleculair met EGF8-13 van Notch1, en intermoleculair met C2-EGF3, EGF8-11 en het cysteïne-rijke domein (CRD) van Jagged1. XL-MS en kwantitatieve bindingsexperimenten geven aan dat deze drie delen van Jagged1 ook intramoleculair met elkaar interactie aangaan. Deze data, samengenomen met kleine hoek röntgenverstrooiing (SAXS) experimenten die de dimensies en flexibiliteit van Notch1 en Jagged1 beschrijven, ondersteunen de formatie van niet-uitgerekte structuren. Samengenomen suggereren deze data dat belangrijke delen van Notch1 en Jagged1 niet ver verwijderd van elkaar zijn, zoals eerder gedacht, maar elkaar direct binden om zo Notch1 activatie te contoleren. Hierdoor herdefiniëren we het model voor Notch1-Jagged1 activatie en openen we nieuwe wegen voor therapeutische toepassingen.

In hoofdstuk 3 beschrijven we het niet-remmende mechanisme waarbij het EgB4 nanolichaam interacteert met EGFR. De humane epidermale groeifactor receptor (HER) familie, waarvan EGFR de eerste is, is van kritiek belang in celproliferatie, -migratie en -differentiatie. EGFR is overgeëxpresseerd in verscheidene kankers, waaronder glioblastoma multiforme (GBM) en niet-kleincellige longkanker (NSCLC), en is daardoor een belangrijk therapeutisch doelwit. Anti-EGFR remmende nanolichamen, die EGFR-activatie verhinderen, waren al ontwikkeld voor therapeutische doeleinden. Een structureel onderzoek liet zien dat deze nanolichamen EGFR-activatie remmen door competitief te binden aan de ligandbindingsplaats op EGFR of door de conformatieverandering die nodig is voor EGFR-activatie sterisch te hinderen. Wij lossen de kristalstructuur van het niet-remmende EgB4 nanolichaam, op zichzelf en in complex met het actieve EGF-gebonden EGFR. Structuur uitlijning van residuen in het bindingsoppervlak in actief of inactief EGFR laat zien dat daar, bij EGFR activatie, geen conformatieveranderingen plaats vinden. Dit suggereert dat EgB4 kan binden aan zowel actief als inactief EGFR, zonder de overgang tussen deze twee vormen te voorkomen. Samengenomen laat dit zien dat EgB4 gebruikt kan worden om

EGFR te markeren, zonder de functie van EGFR te beïnvloeden. Dit kan nut hebben voor de ontwikkeling van medicijnen voor de behandeling van EGFR-gerelateerde kankers.

In hoofdstuk 4 presenteren we het ontwerp, synthese en structurele en functionele karakteriseringen van calciumafhankelijke lipopeptide-antibiotica (CDA) analogen. CDAs zijn een opkomende klasse moleculen die gebruikt worden voor de behandeling van infecties door Gram-positieve bacteriën, zoals geïllustreerd door het klinisch gebruikte Daptomycine. Hoewel het mechanisme waarmee Daptomycine bacteriën aanvalt onduidelijk blijft, zijn de werkingsmechanismen van CDAs Laspartomycine C en Friulimicine/Amphomycine beter begrepen. Recentelijk is de kristalstructuur van Laspartomycine C in complex met C_{10}-P (een beter oplosbare versie van lipidedrager C_{55}-P) opgelost. Dit liet voor het eerst de atomaire details zien van een CDA gebonden aan zijn bacteriële doelwit. Laspartomycine zondert C_{55}-P af, waardoor de rol van C_{55}-P in de synthese van peptidoglycanen wordt onderbroken, resulterend in preventie van synthese van de bacteriële celwand. De Friulimicine/Amphomycine klasses van CDAs zijn structureel vergelijkbaar met Laspartomycine C en richten zich ook op C_{55}-P. Dit gaf ons de aanleiding om te evalueren wat de impact is van het introduceren van kenmerken van Friulimicine/Amphomycine in Laspartomycine C. We helderde hoge-resolutie structuren op van twee CDA-analogen in complex met C_{55}-P. De structuren onthullen een wisselwerking tussen residuen 4, 9 en 10 in de macrocyclische peptide. In het kristal beheersen deze residuen, die anders zijn in Laspartomycine C, de formatie van een hogere-orde organisatie, welke niet gezien waren in het Laspartomine C kristal onder vergelijkbare omstandigheden. Dit verklaart de antibacteriële activiteit. Daarnaast gaf beeldopname van levende cellen een inzicht in de C_{55}-P-geassocieerde familie van antibiotica en werpt het een licht op het unieke mechanisme van Laspartomycine en Friulimicine/Amphomycine in vergelijking met Daptomycine.

In hoofdstuk 5, bediscussieer ik de implicaties van onze bevindingen over hoe de moleculaire mechanismen van de cel oppervlakte receptoren Notch1 en EGFR, en onze lipopeptide CDA analogen, een rol spelen in gezondheid en ziekte. Ik beschrijf kort hoe deze kennis uiteindelijk zou kunnen helpen om nieuwe therapeutische toepassingen te ontwikkelen.

Samengenomen dragen deze studies bij aan een beter begrip van essentiële biologische systemen en kunnen ze helpen bij de ontwikkeling van nieuwe moleculen voor therapeutische toepassingen. Dit werk illustreert het belang van structuurbiologie in zowel fundamenteel als toegepast onderzoek.

Résumé

Les protéines sont des machines moléculaires universelles impliquées dans les processus biologiques de tous les organismes vivants, des animaux aux bactéries. Les protéines sont constituées d'éléments de base nommés acides aminés, attachés les uns aux autres de manière linéaire. Malgré cette structure qui semble étonnamment simple, les protéines réalisent des tâches extrêmement variées, telle que la différenciation cellulaire et la réponse immunitaire. Pour remplir leur fonction, les protéines adoptent une structure tridimensionnelle unique qui leur permet de réaliser des tâches spécifiques, définissant leur rôle. D'importantes protéines permettent d'établir la communication entre les cellules : il s'agit des récepteurs de surface cellulaire. Ces récepteurs perçoivent l'environnement de la cellule et transmettent des messages vers, et depuis celle-ci. Ces messages peuvent être échangés entre cellules voisines par contacts directs entre protéines, ou entre des cellules distantes par le biais de molécules sécrétées qui peuvent voyager d'une cellule à une autre. Ces processus sont strictement régulés dans tout le corps, et des maladies peuvent se développer si des messages sont incorrectement reçus ou envoyés. Les protéines peuvent aussi servir à traiter des pathologies provenant de ce type de dysfonctionnement, ou à combattre des infections par un pathogène. C'est le cas des antibiotiques lipopeptidiques calcium-dépendants, qui permettent de lutter contre les infections bactériennes. Ces molécules ciblent la membrane bactérienne et agissent en la déstabilisant ou en empêchant la synthèse de la paroi bactérienne. Dans cette thèse, nous utilisons diverses approches de biologie structurale ainsi que des molécules à potentiel thérapeutique pour mieux connaître la structure et la fonction de récepteurs de surface cellulaire, et d'antibiotiques lipopeptidiques calcium-dépendants, afin de contribuer au développement de traitements thérapeutiques.

Dans le chapitre 1, je présente les différents récepteurs de surface cellulaire que nous avons étudié, à savoir Notch1 et le récepteur au facteur de croissance épidermique (EGFR, de l'anglais epidermal growth factor receptor), ainsi que les antibiotiques calcium-dépendants (CDAs, de l'anglais calcium-dependent antibiotics) lipopeptidiques appartenant aux classes Laspartomycine et Friulimicine/Amphomycine.

Dans le chapitre 2, nous étudions le mécanisme moléculaire d'activation du récepteur Notch1 par son ligand canonique Jagged1. La voie de signalisation Notch est un système de communication intercellulaire impliqué dans divers processus biologiques, comme

la détermination du destin cellulaire, le maintien des cellules souches, la régulation du système immunitaire et l'angiogenèse. Le dérèglement de la signalisation Notch donne lieu à une quantité de pathologies, telles que les maladies congénitales et les cancers. Les études structurales de la région extracellulaire de Notch1 sont compromises en raison de difficultés à produire et isoler cette protéine en quantité suffisante. L'ectodomaine de Notch1 est de grande taille, fortement glycosylé, et dispose d'une architecture très modulaire composée de 36 domaines analogues au facteur de croissance épidermique (EGF, de l'anglais epidermal growth factor) qui disposent de 6 résidus cystéine chacun. Dans ce travail, nous procédons à l'expression recombinante en cellules de mammifères et à la purification des ectodomaines de Notch1 et Jagged1. Nous utilisons le cross-linking couplé à la spectrométrie de masse (XL-MS, de l'anglais cross-linking mass spectrometry), ainsi que des techniques biophysiques et structurales, sur les ectodomaines entiers et sur des régions ciblées du complexe Notch1-Jagged1. Cela nous permet d'identifier plusieurs sites de liaison qui forment un réseau d'interactions dans le complexe Notch1-Jagged1. En particulier, le site « EGF33-région de régulation négative (NRR, de l'anglais negative regulatory region) » appartenant à Notch1 interagit de manière intramoléculaire avec le site EGF8-13, et de manière intermoléculaire avec les sites C2-EGF3, EGF8-11 et domaine riche en cystéines (CRD, de l'anglais cysteine-rich domaine) de Jagged1. Nos expériences de XL-MS et de dosage de liaison quantitatif indiquent que les trois sites sur Jagged1 interagissent également de façon intramoléculaire. Ces informations, couplées aux données de diffusion des rayons X aux petits angles (SAXS, de l'anglais small-angle X-ray scattering) qui décrivent les dimensions et la flexibilité des ectodomaines de Notch1 et Jagged1, indiquent que ces protéines adoptent une architecture partiellement repliée. D'après l'ensemble des données de l'étude, les régions critiques des ectodomaines de Notch1 et Jagged1 ne sont pas éloignées les unes des autres, contrairement à ce qui étaient pensé auparavant, mais en contact direct afin de contrôler l'activation de Notch1, redéfinissant ainsi le modèle d'activation Notch1-Jagged, et offrant de nouvelles possibilités d'applications thérapeutiques.

Dans le chapitre 3, nous décrivons le mécanisme non-inhibiteur par lequel le nanocorps EgB4 interagit avec EGFR. La famille de récepteurs du facteur de croissance épidermique humain (HER, de l'anglais human epidermal growth factor receptor), dont EGFR est le membre fondateur, est essentielle pour la prolifération, la migration et la différenciation cellulaires. EGFR est une cible thérapeutique prioritaire en raison de sa surexpression dans plusieurs types de cancers, dont le glioblastome multiforme et le cancer du poumon non à petites cellules. Des nanocorps anti-EGFR inhibiteurs, qui empêchent donc l'activation d'EGFR, ont été développés à des fins thérapeutiques. Une étude structurale a montré que ces nanocorps inhibent l'activation d'EGFR en se liant de façon compétitive au site de fixation du ligand EGF, ou en bloquant stériquement le changement conformationnel d'EGFR qui est nécessaire à son activation. Nous présentons les structures cristallographiques du

nanocorps non-inhibiteur EgB4, à la fois seul, et lié au complexe actif EGFR-EGF. Nous montrons qu'EgB4 se fixe à un nouvel épitope sur les domaines I et II d'EGFR lorsque celui-ci est actif, dimérique et lié à EGF. L'alignement structural des résidus d'interface de l'EGFR actif et inactif indique que l'interface ne subit pas de changement conformationnel lorsqu'EGFR alterne entre ces deux conformations. Cela suggère que, contrairement aux nanocorps inhibiteurs, EgB4 peut se lier à l'EGFR actif et inactif. De plus, EgB4 n'empêche pas EGFR d'alterner entre ces conformations. Ces données suggèrent qu'EgB4 pourrait être utilisé en tant que biomarqueur en imagerie des tumeurs sans altérer le fonctionnement d'EGFR, ce qui pourrait s'avérer utile pour contribuer au développement de molécules thérapeutiques contre les cancers associés à EGFR.

Dans le chapitre 4, nous présentons la conception, la synthèse, ainsi que l'évaluation structurale et fonctionnelle d'analogues de CDAs lipopeptidiques. Les CDAs constituent une classe émergente de molécules utilisées pour traiter les infections par les bactéries à Gram positif, comme illustré par la Daptomycine, qui est largement utilisée en milieu clinique. Alors que le mécanisme par lequel la Daptomycine cible les bactéries est peu clair, le mode d'action des CDAs Laspartomycine C et Friulimicine/Amphomycine est mieux compris. Récemment, la structure de Laspartomycin C en complexe avec C_{10}-P (un analogue soluble du transporteur lipidique C_{55}-P) a été résolue, ce qui a permis pour la première fois de révéler les détails atomiques d'un CDA lié à sa cible bactérienne. Laspartomycine C séquestre C_{55}-P, lui-même impliqué dans la synthèse du peptidoglycane, empêchant ainsi la construction de la paroi bactérienne. Les CDAs de classes Friulimicine/Amphomycine sont structuralement similaires à Laspartomycine C et ciblent également C_{55}-P, ce qui nous a incité à introduire des éléments provenant de Friulimilicine/Amphomycine dans Laspartomycine C. Nous avons produit deux analogues de CDAs et avons résolu leur structure cristallographique en complexe avec C_{55}-P, révélant ainsi la contribution des résidus 4, 9 et 10 du macrocycle peptidique. Dans le cristal, ces résidus, qui diffèrent de ceux présents dans Laspartomycine C, sont responsables de la formation d'un assemblage multimérique non-observé dans le cristal de Laspartomycine C, produit dans des conditions similaires, expliquant ainsi certaines différences d'activité antibactérienne. De plus, par imagerie des cellules vivantes nous apportons des informations complémentaires sur la famille d'antibiotiques ciblant C_{55}-P, mettant ainsi en lumière le mode d'action de Laspartomycine C et Friulimicine/Amphomycine, et ses différences avec Daptomycine.

Dans le chapitre 5, je discute des implications de nos recherches sur les mécanismes moléculaires utilisés par les récepteurs de surface cellulaire Notch1 et EGFR, ainsi que par nos lipopeptides analogues de CDAs. Nous évoquons le rôle de ces systèmes dans la lutte contre les maladies, et nous expliquons comment nos résultats pourront favoriser le développement de molécules thérapeutiques.

Ces études contribuent à approfondir la connaissance de systèmes biologiques essentiels et pourront aider au développement de nouvelles molécules au potentiel thérapeutique, illustrant ainsi l'importance de la biologie structurale dans la recherche fondamentale et appliquée.

Acknowledgements

Here we are, five years after I came to Utrecht for my PhD interview. At the time, I could not imagine how much I would learn during my PhD, both professionally and personally. Here I would like to thank the people that contributed to make this journey so special.

Bert, I am extremely grateful for the opportunity you gave me to do my PhD in your group. I would like to thank you for all the time and effort you have put to help me in my different projects. From the advice in the wet lab to manuscript writing, and including data analysis of all the techniques I had the chance to learn in your lab. Thank you for both the guidance and the freedom you gave me. Thank you for caring about my development as a researcher, but also as a person. I wish you many successes with the Structural Neurobiology group. As you always say: "Fingers crossed!". **Piet**, my promotor, thanks for the beautiful sculpture (now melted by the sun) you gave me as a Sinterklaas present in my first year. Hopefully, future sculptures will be safer in the David de Wied building.

I would like to thank the past and present members of the Structural Neurobiology group, which were great colleagues to work with. **Lucas**, mon poto, on aura fait tout notre doctorat ensemble, depuis l'entretien du premier jour, jusqu'aux derniers jours en tant que voisins. C'est fou qu'on ait pu franchir ce cap ensemble en pouvant tout se dire quotidiennement : nos angoisses, nos frustrations, mais aussi nos espoirs et nos réussites. Pouvoir tout partager avec toi en sachant que tu vas me comprendre ça n'a pas de prix, merci d'être là ! **Mercedes**, thank you for your kindness, your happiness, and for your help in the lab. I miss you! **Wouter**, thanks for always being helpful when I needed a colleague to rely on. Good luck with your projects, and I wish you all the happiness with your future wife! **Robin**, the young and promising. Good luck finishing your project! I would also like to warmly thank the three excellent master students I had the chance to supervise: **Anamika**, **Júlia**, **Maria**, I was lucky to work with such motivated and talented students, I wish you all three the best in life. **Nadia**, **Matti**, **Dimphna**, **Hedwich**, you guys set the bar high during your time in our group! Even though you finished your PhD's or postdoc soon after I started, it was fun working with you and learning from you. Special thanks to my collaborators: **Oleg** and **Richard** on the Notch project, **Tom**, **Laurens** and **Nathaniel** on the lipopeptides projects, and **Sofia** and **Paul** on the EGFR project; working together with all of you was one of the nicest part of my PhD. Good luck with your future projects!

I would also like to thank everyone from the Structural Biochemistry department, which played a big part in my life during my PhD time. **Robert (&Alex)**, a.k.a. the vegan couple, a.k.a. my walking buddies, you guys are the sweetest, most amazing friends. Looking forward to our future dinners at Vegitalian. **Juju**, **Gonza**, merci pour votre gentillesse sans limite et pour les soirées trop agréables qu'on a partagées, je dirais juste "Prouuughhhgh". Thanks to the Mexico team for the great time we had (although we did not make it to XXCAREETTT). **Karla**, or chlorophyll_isthenewblack, please don't bring all your plants to the David de Wied building, they will not fit. Besides that, thanks for treating us like kings and queens in Mexico, it was an amazing holiday thanks to you and your family. **Wout**, thanks for teaching me to say "ik wist niet dat je boos werd" and "Rakatakakataka AZ!" in my first year, and "kaduch" in my last year. Looking forward to hearing your next favorite word/sentence (in German). En parlant d'Allemagne, c'est au tour de **Dounia** (coucou Tobi): prépare toi, et surtout ta télé, car en 2022 c'est Rolland Garros + coupe du monde. Plus sérieusement, ton arrivée à Utrecht m'a fait tellement plaisir, reste encore un peu stp ! **Itziar**, "so much to learn from you", including that I should always carry anti histamines with me. Good luck with your house hunting! Hopefully you don't end up in Nunspeet. Talking about that, now it is **Ramon's** turn. I hope no one has called HR in your new group yet. In any case, we had a lot of fun at borrels, conferences, meat-eater watching sessions, and of course sometimes sharing a room ("please close the bathroom door"). **Jitse**, it was super nice being office mates in the last year (although I was often working from home due to the COVID-19 pandemic), looking forward to the future borrels! **Manu**, the true Akta expert, I wish you all the luck and success in your next career moves! Also, sorry that France beat Germany in the UEFA Euro 2020 group stage, but it was probably due to the French players from Bayern. **Viviana**, thanks for your kindness and for telling me about your post PhD experience! Butawhyy??? **Lena**, your never-ending enthusiasm is contagious, when are we going to eat sushis? **Max**, I always saw you walking in the lab (and everywhere in general) holding a styrofoam box, what's in there?? Good luck finishing up :). **Leanne**, the lab FIB-milling expert, I wish you the best for your PhD and all the happiness in den Bosch! **Rodrigo**, your calm and relaxed attitude is inspiring. I found out only recently that you speak German. What other skills do you hide?? **Louris**, thank you for always being available when I needed your help in the lab! **Camilla**, it was nice sharing an office with you, although that was not for long! Looking forward to going to future Le Guess Who festivals with you and Tim. **Xiaoguang**, thanks for your diverse advice on crosslinking and future career opportunities! **Nick**, **Martin**, **Toine**, thanks for helping me to solve and refine the antibiotic structures, and for the many borrel chats we had! **Joke**, **Arie**, **Mariska**, **Harma**, thanks for being the pillars of the lab! I don't know what we would do without you. **Noortje**, **Bianca**, **Lisa**, thank you for always being available to help with any kind of question I had. Besides that, I really enjoyed your company! **Eric**, **Loes**, it was fun teaching practicals and werkcollege with you, thanks for your help! **Tzviya**, thank you for systematically accepting to be the second reviewer of the master students I supervised, and for your suggestions regarding my project. **Stuart**,

thanks for kindly helping me a couple of times at the microscope. I really appreciated your company! **Friedrich**, thank you for the constructive suggestions during group meetings. **UPE members**, thanks for the amazing protein expression levels. Ravi, Gijs, Andreas, Miguel, Dan, Elisa, Jo, Nicoleta, Marten, Wenfei, and other colleagues, all the best for your projects!

I would like to thank other people that played a big role in my life these last few years. Je voudrais d'abord remercier la team Bouchon de liège, dont je tairai l'origine du nom, ainsi que la sous-team B&C. Alexis, Barthou, Boris, Clémence, Clotav, Dorian, Eliott, Franky, Gena (& Lotfi), Guillaume (a.k.a. Gilly), Jo, Laurent (BG), Nono, Thibaut, cœur sur vous. Je voudrais aussi remercier la team TS3. Jasmine, Clémence, Chaima, Thomas, Arthur (a.k.a. Chivers), Cyril, Arnaud, la classe crash test à tout jamais ! Thanks Jacopo, Julia, Davide, Camilo, Tania,… for all the fun we had so far in Utrecht!

I would like to warmly thank my Dutch family: Marjan, Toon, Simone, Thomas, Eline. I feel incredibly lucky to have such nice and caring in-laws, you made me feel at home right away. I am very grateful for all the nice weekends we had in Roosendaal and in the Dutch countryside, but also for our amazing USA trip. I look forward to many more nice times together!

Enfin, je voudrais remercier ma famille, et en particulier mes grands-parents et mes parents, pour tout ce qu'ils m'ont apporté jusqu'à présent. Maman, Papa, j'espère que votre tsakoug vous rend fiers !

I could not finish my thesis without acknowledging the most important person of all: my beautiful, smart, funny, caring, and amazing girlfriend. Daniëlle, thank you for your endless support, especially during the stressful times of my PhD. I look forward to the next steps in our life together!

About the author

Matthieu Raphy Zeronian was born on the 25th of September 1993 in Lyon, France. He obtained his scientific baccalaureate from the Edouard Herriot high school in Lyon, after which he joined the University Claude Bernard Lyon 1 to study Biochemistry in 2011. He received his bachelor's degree and enrolled a research Master in Structural and Functional Biochemistry in 2014. As part of this degree, he performed a six-month research project titled "Structural study of the BRICHOS-containing Bri2 protein" in the group of Prof. Stefan Knight at Uppsala University. In September 2016, he started his PhD in the Structural Biochemistry group at Utrecht University, under the supervision of Dr. Bert Janssen. The results of his research are presented in this thesis.