Investigation of the Structural Basis of Ligand Recognition of the C-Type Lectin Receptor Langerin

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (*Dr. rer. nat.*)

submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by

Jonas Hanske

from Frankfurt (Oder), Germany

May 2016

This dissertation was contrived between June 2012 and May 2016 under the supervision of Dr. Christoph Rademacher at the Department of Biomolecular Systems headed by Professor Dr. Peter H. Seeberger at the Max Planck Institute of Colloids and Interfaces in Potsdam-Golm, Germany.

1st Reviewer: Dr. Christoph Rademacher

2nd Reviewer: Professor Dr. Markus Wahl

Date of defense: 19.09.2016

Acknowledgements

Foremost, I would like to express my deepest gratitude to Dr. Christoph Rademacher for his mentorship and guidance, his trust and respect, his fairness, and his motivation and generous support. It was an honor for me to be the first PhD student under Dr. Rademacher's supervision and I am certain that he will be as great a mentor and teacher for many future generations of doctoral students as he was for me.

I would like to thank Professor Peter H. Seeberger for providing the institutional framework and the stimulating interdisciplinary scientific environment that enabled this dissertation. Moreover, I am very grateful to Professor Seeberger for for his valuable advice and insights about the academic world.

Furthermore, I would like to thank Professor Markus Wahl for reviewing this thesis. Professor Wahl invoked my interest for structural biology as early as in my second year of studies at Freie Universität Berlin. I am grateful for his lasting support over the entirety of my academic way.

Moreover, I thank Professor Angela Gronenborn and Professor Ekaterina Pletneva for their generous support and mentorship at different stages of my scientific career.

I would like to thank my fellow PhD students Jonas Aretz, Jessica Schulze, and Eike-Christian Wamhoff for the outstanding work atmosphere, their team spirit, and for their friendship. Furthermore, I thank them for reading large parts of this thesis very critical and providing invaluable suggestions to improve the quality of this work.

Also, I thank my students Andrina Braband, Jia Hui Li, Henrik Schmidt, Elena Shanina, and Daniel Tesolin. It was a great honor for me to teach and provide guidance and leadership for them. They are all outstandingly bright students and I am certain that they will have great impact on science in their futures. Their ideas and their hard work contributed greatly to this thesis.

Furthermore, I wish to thank the students who supported the daily work in our laboratory: Thomas Gruhl, Kim Silberreis, and Şafak Bayran. Their diligent work enabled smooth processes in the laboratory and reduced the experimental preparation and cleanup time significantly.

I wish to thank the administrative, IT, and technical staff at the MPI of Colloids and Interfaces, specifically Dorothee Böhme, René Genz, Eva Settels, Felix Hentschel, Uwe Vogel, and Anette Wahlbrink. I highly appreciate the work of Olaf Niemeyer who kept the

NMR machines running and who taught me a lot about the intricate hardware of Varian spectrometers.

Furthermore, I want to acknowledge the Funds of the Chemical Industry (FCI) and the DFG Collaborative Research Center SFB 765 for scholarships funding my PhD.

I am grateful to all my collaborators. Foremost, I thank Dr. Peter Schmieder, Dr. Martin Ballaschk, and Dr. Marcel Jurk for all their help with both the setup and the analysis of NMR experiments. I thank Monika Beerbaum for conducting the NMR assignment experiments and Nicole Holton and Jasmin Kurtzke for introduction and the generous measurement time at the ITC. I thank Professor James C. Paulson and Ryan McBride for conducting microarray screenings that contributed greatly to this work. I thank Dr. Daniel Varon Silva for providing mannoside sugars. I thank Dr. Yuriy A. Knirel for providing bacterial lipopolysaccharides and Dr. Wolfgang Rabsch for heat-killed bacteria. Furthermore, I want to thank Professor Markus Wahl, Professor Silke Leimkühler, and Professor Beate Koksch for granting me access to some of their instruments.

Moreover, I thank Professor Bettina Keller and Stevan Aleksic for the very productive collaboration on the Langerin allostery project. This thesis would have never reached the present quality without their theoretical work. Furthermore, I thank my colleague Felix Bröcker for two very successful collaboration projects. Also, I thank Professor Tom Großmann and Dr. Chang-Uk Lee for collaboration on a very interesting side project.

I am very grateful for the fruitful discussions about science and life with my colleagues Dr. Christopher Martin, Maurice Grube, Dana Michel, and all other members of the Department.

Finally, I would like to thank my family and my close friends for their outstanding support that was essential for the completion of this thesis. Their love lifted me up and carried me through the past years.

Publications

Parts of this work were or will be published and communicated.

Hanske, J.[+], Aleksic, S.[+], Ballaschk, M., Jurk, M., Shanina, E., Beerbaum, M., Schmieder, P., Keller, B., and Rademacher, C. (2016). Intradomain Allosteric Network Modulates Calcium Affinity of the C-Type Lectin Receptor Langerin. J Am Chem Soc *138 (37)*, 12176-12186.

Hanske, J., Aretz, J., Schulze, J., McBride, R, Schmidt, H., Knirel, Y., Rabsch, W., Paulson, J.C., and Rademacher, C. Murine and human Langerin do not share specificity for bacterial glycans. J Biol Chem, doi: 10.1074/jbc.M116.751750, *published online ahead of print*

Wamhoff, E.-C., Hanske, J., Schnirch, L., Aretz, J., Grube, M., Varón Silva, D., and Rademacher, C. (2016). ^{19}F NMR-Guided Design of Glycomimetic Langerin Ligands. ACS Chem Biol 11 (9), 2407-2413.

Aretz, J., Wamhoff, E.-C., Hanske, J., Heymann, D., and Rademacher, C. (2014). Computational and experimental prediction of human C-type lectin receptor druggability. Front Immunol, doi: 10.3389/fimmu.2014.00323

[+]These authors contributed equally to the work.

Summary

Glycan recognition by glycan-binding proteins is a fundamental process of life. Many members of the diverse family of the myeloid C-type lectin receptors serve as immune receptors recognizing complex glycan structures on pathogen surfaces and host cells and are involved in immune cell activation and regulation. Their restricted expression on defined subsets of antigen presenting cells as well as their exploitation for invasion by some adapted pathogens such as mycobacteria and HIV, and their function in immune cell activation render them attractive targets for novel therapies. A better insight into the underlying molecular processes that govern glycan binding and release as well as confer glycan specificity is a prerequisite to understand the onset of disease and subsequently develop new therapeutic approaches.

The endocytic C-type lectin receptor Langerin, which is expressed on Langerhans cells and specific dendritic cell subsets, was chosen as a model to gain insight into the molecular mechanism of calcium-dependent glycan binding and release in the early endosome that is required for recycling receptors. To this end, biomolecular NMR provided the tools to investigate the structure and dynamics changes of the protein upon binding and release of the essential co-factor calcium and carbohydrate ligands. Whereas the apo form exists in two distinct conformational states and experiences motions on timescales from pico- to nanosecond bond vector fluctuations to very slow cis/trans prolyl bond isomerization on the second timescale, the holo form shows highly reduced mobility on the slower timescales. Moreover, calcium binding is controlled by an allosteric mechanism that involves a large network of interacting residues. This allosteric network was probed by single-residue mutations thereby discovering the function of the network to down-regulate calcium affinity in a partially pH-dependent manner. These findings suggest that the allosteric network promotes fast ligand release in the early endosome. In contrast, binding of simple and complex carbohydrate ligand did not activate the allosteric mechanism suggesting that glycan recognition is controlled on a higher level of organization.

To assess whether the function of Langerin is conserved within the mammalian lineage, glycan binding propensities of human and murine Langerin were investigated side-by-side with respect to their capability to recognize simple carbohydrates and complex microbial polysaccharides. While recognition of simple mono- and disaccharides was almost identical in respect to affinity and binding mode, complex glycans were bound with both differential avidity and highly diverging specificities.

In conclusion, the findings presented within this dissertation demonstrate how intra-domain allostery and protein dynamics regulate ligand binding and release in an endocytic C-type lectin receptor in a co-factor dependent manner. Moreover, this work shows how closely related glycan binding proteins of the immune system readily adopt to the host ecological niche by evolving differential binding specificities for complex microbial glycan ligands.

Zusammenfassung

Erkennung von Glykanen durch ihre Rezeptoren ist ein fundamentaler Prozess des Lebens. Viele Mitglieder der vielfältigen Familie der myeloiden C-Typ-Lektinrezeptoren dienen als Immunrezeptoren, welche komplexe Glykanstrukturen auf Pathogenoberflächen und Wirtszellen erkennen, und damit häufig bei der Immunzellaktivierung und –regulierung beteiligt sind. Ihre eingeschränkte Expression auf definierten Untergruppen von Antigen-präsentierenden Zellen, die Nutzung dieser Rezeptoren zur Wirtszellinvasion durch angepasste Pathogen, wie Mykobakterien und HIV, und ihre Funktion in Immunzellaktivierung, ermöglichen vielfältige neue Therapieansätze. Für die Entwicklung neuartiger Ansätze ist es eine Voraussetzung, dass ein verbessertes Verständnis der grundlegenden molekularen Prozesse erlangt wird. Besonders in Hinblick auf die Glykanbindung und –freisetzung, sowie Glykanspezifität sind weitesgehend unverstanden.

Das endozytische C-Typ-Lectin Langerin, welches auf Langerhanszellen und anderen Untergruppen dendritischer Zellen exprimiert wird, wurde als Modell ausgewählt, um vertiefte Erkenntnis über den molekularen Mechanismus der Calcium-abhängigen Glykanbindung und –freisetzung im frühen Endosom zu gewinnen. Biomolekulare Kernspinresonanzspektroskopie wurde erfolgreich zur Untersuchung der Struktur- und Dynamikveränderungen des Proteins durch Bindung und Freisetzung von Calcium und Kohlenhydratliganden eingesetzt. Während die Apoform des Rezeptors in zwei unterschiedlichen Konformationen existiert und molekulare Bewegungen von Bindungsvektorfluktuation im Pico- bis Nanosekundenbereich bis hin zu sehr langsamer cis/trans-Isomerisierung einer zentralen Prolinpeptidbindung im Sekundenbereich erfährt, ist die Holoform in einer Konformation gefangen und zeigt reduzierte Mobilität auf der Micro- bis Millisekundenzeitskala. Zudem wird Calciumbindung durch einen allosterischen Mechanismus kontrolliert, welcher ein großes Netzwerk interagierender Reste umfasst. Einzelpunktmutationen haben gezeigt, dass das allosterische Netzwerk die Calciumaffinität abhängig vom pH-Wert der Umgebung reduziert. Die Ergebnisse suggerieren, dass das allosterische Netzwerk die schnelle Freisetzung von Liganden im frühen Endosom begünstigt. Dagegen hat die Bindung von einfachen und komplexen Kohlenhydratliganden das allosterische Netzwerk nicht aktiviert, was darauf hindeutet, dass Glykanerkennung auf einer höheren Organisationsebene kontrolliert wird.

Um zu ermitteln, ob die Funktion von Langerin in Säugetierspezies konserviert ist, wurden die Glykanbindungseigenschaften von humanen und murinen Langerin parallel untersucht. Dabei wurde insbesondere die Fähigkeit zur Erkennung von einfachen Kohlenhydraten und

komplexen bakteriellen Polysacchariden untersucht. Während die Erkennung von einfachen Strukturen in Bezug auf Affinität und Bindungsmodus fast identisch war, wurden komplexe Glykane mit unterschiedlicher Avidität und Spezifität erkannt.

Zusammenfassend zeigen die in dieser Dissertation präsentierten Erkenntnisse, wie Intradomänenallosterie und Proteindynamik Ligandenbindung und –freisetzung in einem endozytischen C-Typ-Lectin Kofaktor-abhängig steuern. Darüber hinaus zeigt diese Arbeit, wie im Immunsystem eng verwandte Glykan-bindende Proteine sich durch differenzierte Spezifität für mikrobielle Glykanliganden schnell an ihre ökologische Nische anpassen können.

Table of Contents

1 Introduction

All living cells are covered by a dense and complex matrix of glycans that is termed glycocalyx and plays a pivotal role in cellular recognition processes (Rambourg and Leblond, 1967; Varki, 2007). In the course of evolution, there have emerged many highly specialized glycan binding proteins (GBP) that decipher the information encoded in the complex oligosaccharide structures and confer a wide set of functions from intercellular adhesion over cellular signaling to host defense (Gagneux and Varki, 1999; Varki, 2006). Glycan structures vary highly among different lineages and multicellular host organisms have evolved a toolkit of innate defense mechanisms that protect from pathogen invasion by barcoding glycan structures to differentiate self from non-self (Bishop and Gagneux, 2007). However, many pathogens have adopted strategies to evade these mechanisms by hiding their surface glycans or even exploit these mechanisms by mimicking glycan structures of the host (Varki, 2006). This high evolutionary pressure imposed on both host and microbe has led to a rapid diversification of complex glycan structures which can be described as an evolutionary arms race between hosts and microbes (Bishop and Gagneux, 2007).

1.1 The myeloid C-type lectin receptors play a pivotal role pathogen recognition and shaping of the immune response

The innate immune system employs a large set of pattern recognition receptors (PPR) to detect pathogen-associated molecular patterns (PAMPs) and battle infectious disease at a very early stage (Akira et al., 2006). Many members of the heterogeneous protein family of the myeloid C-type lectin receptors (CTLRs) serve as glycan-binding PRRs and are involved in pathogen recognition and uptake as well as immune cell activation (Robinson et al., 2006; van Kooyk and Rabinovich, 2008; Osorio and Reis e Sousa, 2011). However, CTLRs also function in self-recognition and shape the immune response by dampening or altering myeloid cell activation. The majority of CTLRs are type II transmembrane receptors with a short regulatory intracellular domain, an extracellular stalk region that varies in length, and one or more carbohydrate recognition domains (CRD) that interact with glycans on both pathogens and self-structures. Many CTLRs oligomerize by coiled-coil interactions of the stalk region with the majority forming homotrimers (Weis and Drickamer, 1994; Head et al., 2003; Stambach and Taylor, 2003; Jégouzo et al., 2013) and few members forming homodimers (Brown et al., 2007), homotetramers (Feinberg et al., 2005), or even heterodimers (Furukawa et al., 2013). Moreover, some CTLRs organize in nanoclusters on the plasma membrane surface thereby regulating pathogen specificity (Manzo et al., 2012).

The expression of endocytic CTLRs is mainly restricted to antigen-presenting cells (APCs) such as dendritic cells (DCs) and macrophages where they serve as uptake receptors of pathogens (van Kooyk and Rabinovich, 2008). APCs bridge the gap between the innate and the adaptive immune responses by uptake of PAMP-carrying pathogens recognized by PRRs like CTLRs, subsequent antigen processing and antigen presentation on MHC class I and class II molecules by that eliciting a T-cell mediated adaptive immune response (Idoyaga et al., 2008; LeibundGut-Landmann et al., 2008; Sancho et al., 2009; Unger et al., 2012).

The regulatory properties of CTLRs are highly diverse with some members coupling to adaptors that carry immunoreceptor tyrosine-based activation motifs (ITAM) thereby activating the Syk kinase-dependent signaling pathways (Mocsai et al., 2010) or interacting directly with Syk via a hemITAM motif encoded in the sequence (Brown, 2006) resulting in immune cell activation. Other CTLRs carry a immunoreceptor tyrosin-based inhibitory motif (ITIM) that antagonizes Syk signaling by recruiting the phosphatases SHP-1, SHP-2, and SHIP resulting in a down-regulation of the immune response (Geijtenbeek and Gringhuis, 2009). Furthermore, coupling to Syk kinase pathway enables some CTLRs to transmit signals through the CARD9 caspase pathway leading to subsequent NF-κB activation independent of the SHP phosphatases (Drummond et al., 2011). In addition, some CTLRs also engage in non-tyrosine-based signaling thereby modulating pathways of other PRRs such as Toll-like receptor signaling (Geijtenbeek et al., 2009). Moreover, Raf1 signaling, which activates the transcription factor NF-κB independently from Syk kinase pathway, was demonstrated for the CTLR DC-SIGN (Gringhuis et al., 2007). Still, some CTLRs harbor no known signaling motif and their signaling capacities are still under debate (Osorio and Reis e Sousa, 2011). Notably, the ability to efficiently cross-present exogenous antigens in APCs depends on the presence of the signaling motifs (Fehres et al., 2014).

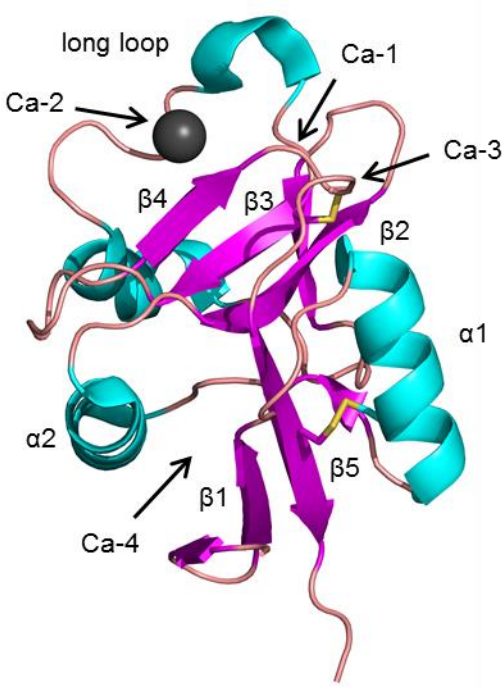


Figure 1-1: The CTDL fold is highly conserved. Cartoon representation of Langerin CRD exhibiting the canonic CTLD fold. The conserved long and short loop as well as the flanking α-helices and central antiparallel β-strands are labeled according to the nomenclature of Zelensky and Gready (2005). The conserved disulfide bridges are shown as sticks. Ca^{2+} is bound in the canonical site 2 and shown as black sphere. The other canonic Ca^{2+} binding sites are not present in Langerin but indicated by arrows (pdb entry: 3p5f, Feinberg et al., 2011).

1.2 The C-type lectin domain fold is an ancient ligand-binding feature

All CTLRs share a common structure motif - the C-type lectin domain (CTLD) fold (Drickamer, 1993a, 1999). The CTLD fold is an ancient conserved structural feature constituting a superfamily of extracellular metazoan proteins with diverse functions (Zelensky and Gready, 2005). All CTLDs consist of an overall looped structure with N- and C-terminal antiparallel β-sheets connected by two flanking α-helices and a three-stranded antiparallel β-sheet in the center of the domain (**Figure 1-1**) (Weis et al., 1991). Within the domain, there is a second loop that leaves and enters the core between the β2 and β3 strands. This long loop varies in length and can be structured by additional secondary structure elements (Zelensky and Gready, 2005). The fold is stabilized by two pairs of highly conserved cysteine residues forming disulfide bridges. Members of the protein family can accommodate up to four Ca^{2+} ions that are complexed by the carboxyl groups of glutamate or aspartate or the amide functionalities of asparagine and glutamine side chains of the protein. The upper loop region can bind up to three Ca^{2+} ions and also forms the Ca^{2+}-dependent carbohydrate binding site. The number of bound Ca^{2+} ions varies between none to four and differs even for closely related lectins (Zelensky and Gready, 2005) .

Although the name of the domain originates from the Ca^{2+}-dependent sugar-binding properties of the first identified members of the family (Drickamer, 1993b), the majority of its

over 1,000 known constituents neither bind Ca^{2+} nor carbohydrates (Weis et al., 1998; Drickamer, 1999). Hence, the CTLD fold is a ligand-binding motif with diverse functions recognizing proteins, lipids, inorganic molecules, and even ice (Geider et al., 1996; Ewart et al., 1998; Sano et al., 1998; Lorentsen et al., 2000; Natarajan et al., 2002). In addition to the metazoan superfamily of CTLD-containing proteins, CTLD-containing proteins are also found in bacteria (Stein et al., 1994), bacteriophages (McMahon et al., 2005) and viruses (Shaw et al., 2010). The CTLD fold is highly resistant to sequence variations (McMahon et al., 2005), a crucial feature of immune receptors that have to adapt rapidly to newly evolving pathogen antigens while still pertaining function.

1.3 A conserved proline serves as a regulatory switch in Ca^{2+} binding of CTLRs

Even though only a minority of CTLD-containing proteins appears to bind carbohydrates, Ca^{2+}-dependent glycan-binding is prevalent in myeloid CTLRs (Drickamer and Taylor, 2015). For the majority, a Ca^{2+} coordinated in the canonical site 2 is essential for carbohydrate binding with only few exceptions described for CTLRs that bind larger oligosaccharide structures (Bono et al., 2001; Brown and Gordon, 2001; Chabrol et al., 2012). The Ca^{2+} ligates a monosaccharide through two vicinal hydroxyl groups with additional interactions formed by amino acid residues of the protein (Weis et al., 1992). Two highly-conserved structure motifs entail ligand specificity: CTLRs containing an EPN motif in the binding site have preference for mannose-type sugars whereas a QPD motif is indicative of galactose-type sugar preference (Drickamer, 1992). In presence of Ca^{2+}, the prolyl bond conformation of the conserved central proline in both binding motifs is always cis. In absence of Ca^{2+}, however, the bond can flip back to the trans conformation (Ng and Weis, 1998; Ng et al., 1998; Poget et al., 2001; Nielbo et al., 2004; Onizuka et al., 2012) constituting a slow conformational switch with regulatory function (Ng and Weis, 1998). However, the release of Ca^{2+} from the low-affinity IgE receptor CD23 alters its conformation dramatically at distance positions in the CRD without cis-trans isomerization of the prolyl bond (Wurzburg et al., 2006). In most cases, the addition of Ca^{2+} has a stabilizing effect on the loop structure (Wurzburg et al., 2006). Furthermore, the binding of Ca^{2+} in the accessory sites 1 and 3 affects the geometry of the carbohydrate binding site entailing a regulatory function (Onizuka et al., 2012; Furukawa et al., 2013). In some CTLRs, basic amino acid side chains occupy the accessory Ca^{2+} sites (Thépaut et al., 2009; Nagae et al., 2013; Jegouzo et al., 2015). The number of Ca^{2+} binding sites influences the reaction order of the Ca^{2+} dependence of glycan binding (Drickamer and Taylor, 2015): CTLRs with accessory binding sites show a higher order Ca^{2+} dependence resulting in sharper transitions between active and inactive states. Many mammalian CTLRs

serve as endocytic receptors and may respond differentially to changes in Ca^{2+} concentration and pH along the endosomal pathway depending on the number of Ca^{2+} sites (Drickamer, 1993b; Ng and Weis, 1998). For instance, the three Ca^{2+}-binding sites in the carbohydrate recognition domain (CRD) of the human hepatic asialoglycoprotein receptor (ASGPR) show high allosteric cooperativity and a strong pH dependence resulting in a steep transition between active and inactive state (Blomhoff et al., 1982; Onizuka et al., 2012). Intriguingly, a strong pH dependence of binding has also been shown for other endocytic CTLRs like DC-SIGN, DC-SIGNR and Langerin that might be based on a conserved mechanism (Feinberg et al., 2001; Stambach and Taylor, 2003).

1.4 CTLR binding to carbohydrates is intrinsically weak and promiscous

The binding of monosaccharide ligands to mammalian CTLRs is intrinsically weak and promiscuous with dissociation constants in the millimolar range (Weis and Drickamer, 1996). Specificity and increasing affinity are conveyed on different levels of complexity: On the single domain level, for receptors that bind specifically to small rigid oligosaccharide ligands like sialyl-Lewis^x, affinity increase is majorly achieved due to a reduced entropy penalty as observed for the scavenger C-type lectin receptor (Coombs et al., 2005; Taylor and Drickamer, 2009). Both the ligand and the binding site are rigid and readily preformed (Zierke et al., 2013) thereby reducing the entropy cost for the replacement of the water molecules in the binding site. In addition, highly structured carbohydrates pay a reduced desolvation penalty because they resemble preformed water (Binder et al., 2012). Binding specificity and affinity towards more flexible oligosaccharide ligands like high-mannose structures are enhanced by additional contacts as observed in DC-SIGN (Guo et al., 2004), BDCA-2 (Jegouzo et al., 2015), DCIR-2 (Nagae et al., 2013), and Mincle (Furukawa et al., 2013). Interestingly, even for receptors that possess no additional contact sites, increased affinity for oligosaccharide ligands is observed (Feinberg et al., 2011; Probert et al., 2013). For galectins, another family of GBPs, it has been shown that entropy loss of the ligand is compensated by increased conformational entropy of the carbohydrate recognition domain (CRD) (Diehl et al., 2010), a mechanism that might also be at play in CTLRs.

1.5 Multivalent presentation of GBPs and glycan structures conveys affinity and specificity for glycan-protein interactions.

Multivalent display of both glycan and receptor enhances binding avidities by several orders of magnitude and has been proposed to be the underlying principle in glycan recognition to confer both sufficient affinity and fine tune specificity (Collins and Paulson, 2004). Indeed,

the concept of multivalency (Jencks, 1981; Kitov and Bundle, 2003), which describes the non-statistical addition of binding energies for systems with several physically connected interaction sites on both ligand and receptor, greatly aided the understanding of host-pathogen recognition, cell surface adhesion, and viral entry; processes that have underlying weak glycan-protein or even weaker glycan-glycan interactions (Kiessling and Lamanna, 2003). In the last two decades, the concept has been fruitfully applied in chemical biology for the development of molecular probes and even drug candidates (Mammen et al., 1998; Fasting et al., 2012; Cecioni et al., 2015). CTLRs enhance binding to ligands by multivalent display at two levels of organization:

On receptor level, most CTLRs form homo-oligomers thereby enhancing avidity towards multivalently presented ligands (Weis and Drickamer, 1996). The spacing of the binding sites, which is controlled by the geometry of CRD presentation and rigidity of the oligomers, plays an important role to convey specificity (Collins and Paulson, 2004). For example, ASGPR affinities for galactose-terminated glycans increase from high micromolar to nanomolar for mono-, di- and tri-antennary N-linked glycans because of the perfect geometry of CRD presentation (Lee et al., 1983; Bider et al., 1996). However, other CTRLs show a less dramatic increase in binding affinities upon multivalent display of ligands indicating a less optimal spacing of the CRDs (Stambach and Taylor, 2003). For uptake receptors, ligand affinity affects the fate of the cargo and its subsequent sorting in the early endosome and consequently entails a regulatory function (Grant and Donaldson, 2009).

On cell surface level, clustered presentation of both glycan structures and GBPs enhance binding avidities greatly (Collins and Paulson, 2004). By controlling receptor spacing on the surface ligand specificity can be fine-tuned. For example, the intrinsic nanoclustering of DC-SIGN on the cell surface conveys specificity for smaller viruses over larger bacteria or fungi (Manzo et al., 2012). Clustered presentation of monovalent selectins on microvilli is required for efficient leukocyte rolling (Schwarz and Alon, 2004).

Multivalent interactions are also employed by other mammalian GBPs: The secretion of multimeric Reg and galectin family proteins with antimicrobial activities forms a first-line defense against pathogens (Cash et al., 2006; Kohatsu et al., 2006). However, next to the afore described trans interactions of GBPs, multivalent glycan binding is also utilized by both membrane-anchored receptors and soluble GBPs for cis interaction: For example, galectins cross-link cell-surface glycoproteins creating functionally specialized lattice-type structures on the cell surface (Ahmad et al., 2004a; Ahmad et al., 2004b). Furthermore, the Siglec CD22, which regulates B-cell activity, has been shown to mask its glycan binding site for

trans ligands by interacting with cis ligands present on the cell membrane (Razi and Varki, 1998). However, preference for specific trans ligands results in re-distribution of the receptor to the interaction site and B-cell activation (Collins et al., 2004). Pathogens utilize multivalent display of GBPs for pathogenesis in form of toxins (Kitov et al., 2000) and surface receptors for bacterial adhesion and colonization (Karlsson, 1995) or viral entry (Smith and Helenius, 2004).

Taken together, GBPs fulfill diverse functions by means of multivalent glycan-protein interactions. These functions reach from cell-cell and extra-cellular matrix adhesion to host defense by pathogen recognition and uptake and by antimicrobial activity of secreted GBPs.

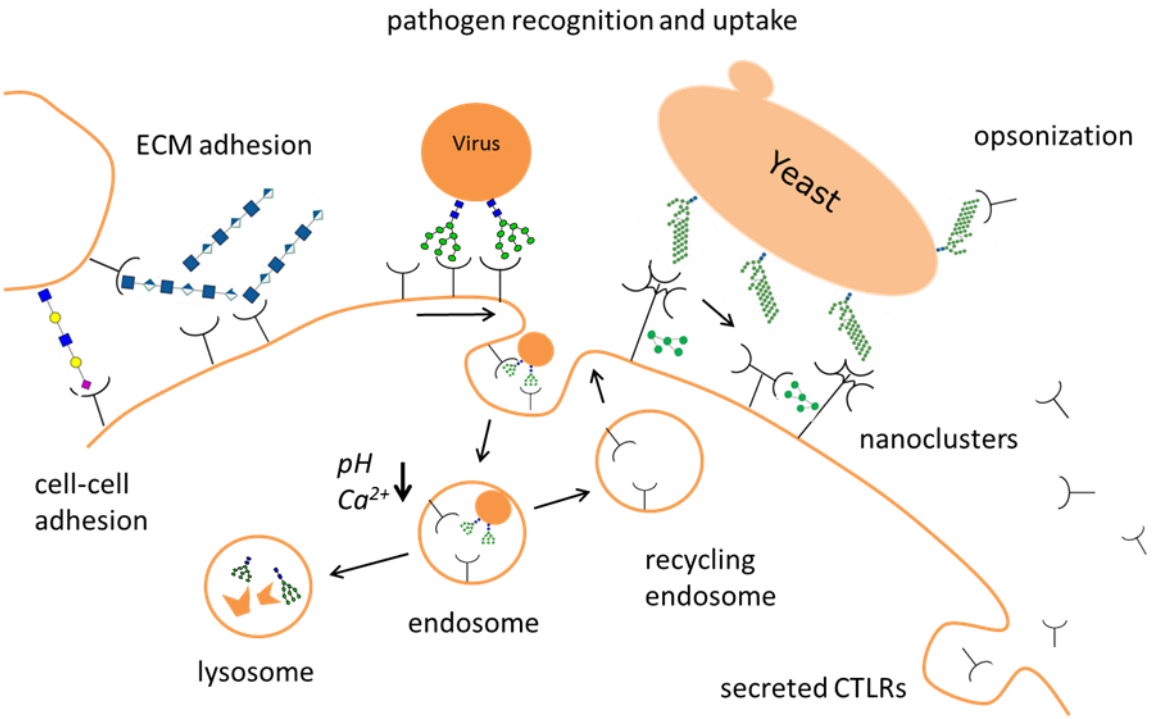


Figure 1-2: Some CTLRs facilitate cell-cell or extracellular matrix (ECM) adhesion by recognition of self-structures whereas others are involved in pathogen recognition and uptake. Some GBPs with antimicrobial activity are secreted and form a first-line-of-defense against pathogen invasion.

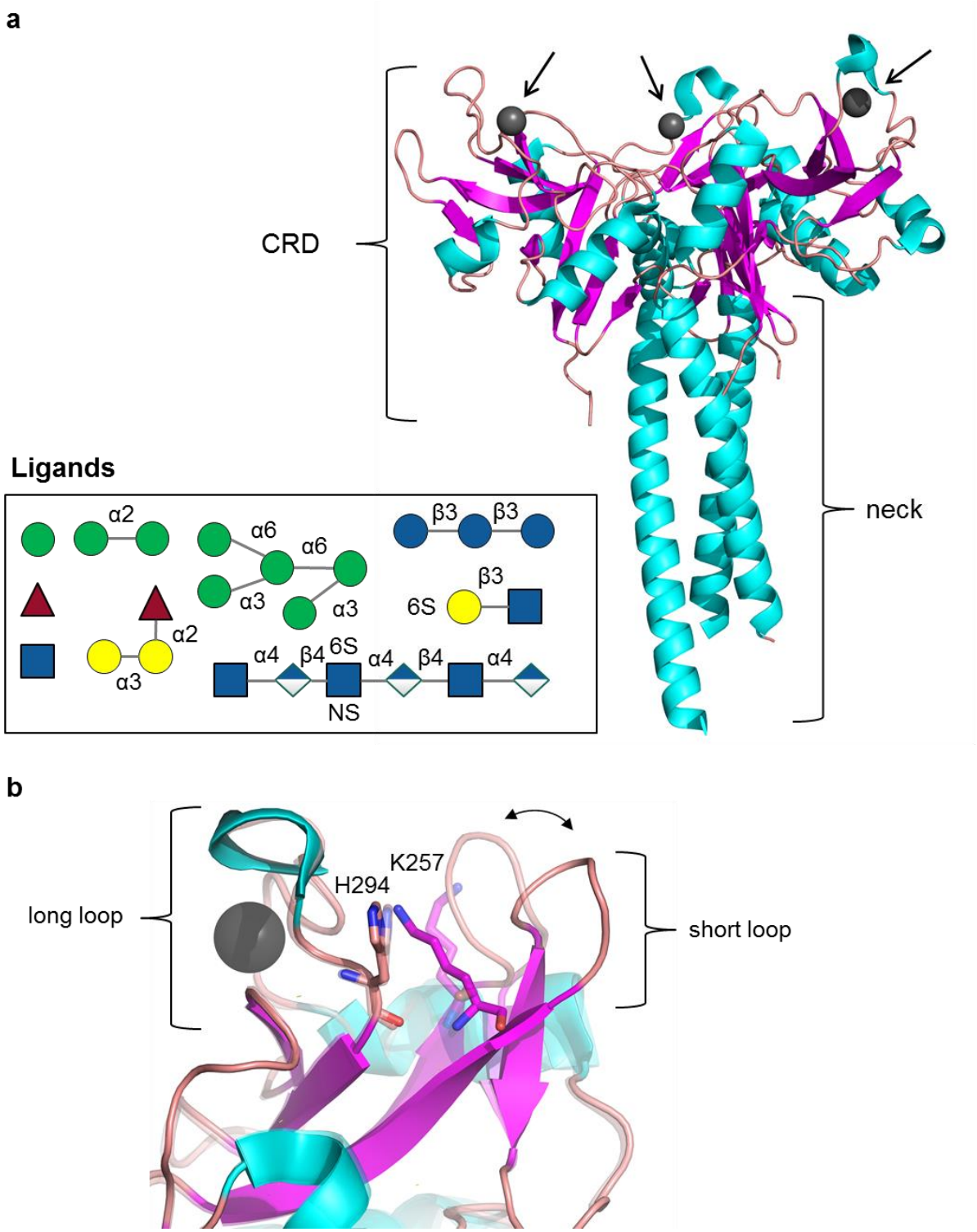


Figure 1-3: The myeloid CTLR Langerin is a trimeric receptor with a wide specificity for self and non-self glycan antigens. (**a**) Carton representation of trimer Langerin with truncated neck region (pdb entry: 3kqg, Feinberg et al., 2010). The CRD and part of the neck are labeled. Ca^{2+} is shown as black sphere and indicated by arrows additionally. The box shows known ligand structures according to the CFG code (s. Appendix). Next to mannose-containing glycan structures, β-glucans, 6-sulfated galactoses, heparin and fucose moieties are recognized. (**b**) Overlay of the crystal structure of the CRD (solid, pdb entry: 3p5f, Feinberg et al., 2011) and the trimer (transparent, pdb entry: 3kqg). The side chains of H294 and K257 occupy the canonic accessory Ca^{2+} binding sites 1 and 3. In the CRD structure both side chains form close contacts. In the trimeric structure (transparent), the short loop is flipped towards the long loop indicating the existence of alternate conformations.

1.6 The CTLR Langerin serves as uptake receptor in pathogen recognition

The trimeric CTLR Langerin (CD207) serves as a PRR and is mainly found on Langerhans cells (LC) in the epidermis and CD103+ dermal dendritic cells (DC) (Valladeau et al., 2000; Romani et al., 2010). It is involved in the detection of a wide set of pathogens including viruses such as HIV (de Witte et al., 2007) and measles (van der Vlist et al., 2011) as well as fungi (de Jong et al., 2010) and mycobacteria (Hunger et al., 2004), and recognizes antigens of endogenous origin like H and B blood-group antigens (Hsu et al., 2009) and heparin (Chabrol et al., 2012; Munoz-Garcia et al., 2015). Langerin has been shown to block HIV transmission and promote viral degradation (de Witte et al., 2007). Langerin is a recycling, endocytic receptor that accumulates in the recycling endosomes and the connected Birbeck granules, rod-shaped cellular compartments that are Langerin dependent (Valladeau et al., 2000; Mc Dermott et al., 2002; Stambach and Taylor, 2003). Langerin-mediated endocytosis has been shown to induce cross-presentation of exogenous ligands to both $CD8^+$ and $CD4^+$ T cells (Idoyaga et al., 2008). Interestingly, the ability to cross-present foreign antigens seems to depend on the nature of the ligand (van der Vlist et al., 2011). Notably, Langerin sorts its cargo into different endosomal compartments than the related CTLR Dectin-1 (Fehres et al., 2015). Although Langerin has not been connected to any signaling pathway so far, the intracellular domain harbors a proline rich motif that could serve as an SH3 domain adaptor. Deletion or mutations of the motif result in missorting of the receptor and Birbeck granule deficiency (Thépaut et al., 2009). Upon acidification, Langerin remains in trimeric form but releases its ligands likely in a Ca^{2+} dependent manner (Stambach and Taylor, 2003).

The 37.5 kD type II transmembrane receptor comprises a short intracellular domain, a transmembrane region, an α-helical neck domain (ND) and a CRD with the characteristic CTLD fold harboring a single Ca^{2+} ion (Thépaut et al., 2009; Feinberg et al., 2010). Several structures of the CRD interacting with different carbohydrate ligands have been solved by X-ray crystallography (Thépaut et al., 2009; Feinberg et al., 2011) and there is one structure of a truncated version of the trimeric extracellular domain (ECD) (Feinberg et al., 2010) (**Figure 1-3a**). Notably, no structure of the Ca^{2+}-free form is available. The trimeric structure shows a hydrophobic interaction surface of the three neighboring CRDs raising the question whether there is any crosstalk between the domains, which might result in allosteric cooperativity upon binding. Although allosteric cooperativity in lectin-carbohydrate interactions has been largely ruled out (Coombs et al., 2010), no detailed investigation on oligomeric CLTRs interacting with their primary ligand Ca^{2+} has been undertaken so far.

Langerin contains an EPN motif that entails mannose-type carbohydrate binding preference (Drickamer, 1992). Indeed, Langerin preferentially binds mannose and fucose but also shows affinity for N-acetyl glucosamine and glucose (Stambach and Taylor, 2003). Surprisingly, structural studies revealed that human Langerin can also accommodate 6-sulfated galactoses by forming additional contacts of K299 and K313 with the sulfate group thereby switching glycan orientation (Tateno et al., 2010; Feinberg et al., 2011). A common polymorphism (K313I) shifts glycan preference from terminal mannoses to terminal N-acetyl glucosamine structures (Ward et al., 2006; Feinberg et al., 2013). Langerin interacts with high mannose structures like Man9 from gp120 and yeast mannan (Stambach and Taylor, 2003). Its additional preference for β-glucans, typical fungal antigens, has led to its proposed role as major fungal recognition receptor next to Dectin-1 (de Jong et al., 2010). While the recognition of (1->6) branched β-glucans like laminarin can be explained by interactions with the 2- and 3-hydroxyl groups of the terminal glucose (Feinberg et al., 2011), the structural basis for the interaction with (1->3)β-linked linear β-glucans like curdlan remains elusive. Moreover, trimeric Langerin binds heparin in a Ca^{2+}-independent binding mode (Chabrol et al., 2012; Munoz-Garcia et al., 2015) thereby playing a potential role in self-recognition. As observed in other endocytic CTRLs, canonic carbohydrate binding in Langerin is pH dependent with ligand release occurring at endosomal pH (Stambach and Taylor, 2003; Ward et al., 2006). The ligand release has been attributed to a loss in Ca^{2+} affinity rather than to the carbohydrate itself (Ibid.). Naturally occurring polymorphisms have been shown to alter both protein stability and pH-dependence of binding while the mutants were still functional in uptake assays (Ward et al., 2006).

In Langerin, the long loop between the β2 and β3 strands is segmented by additional secondary structure elements. An additional β strand (β2', residues 264-266) forms an antiparallel β-sheet with the β2 strand (residues 255-257) thereby creating a loop with a sharp turn called short loop (residues 258-263). Moreover, two additional helices are observed in the long loop region: the α3 helix (residues 273-279) on the backside of the binding site and a 3:1 helix on top of the binding site (residues 289-292). Notably, several proline residues are present in the long loop region (residues 271, 283, 285, and 301) that may subdivide the region into distinct structural elements. Due to this segmented structure of the long loop region in Langerin, in this work, the term long loop will refer to the residues between the third proline residue and the β3 strand (residues 287-295).

Whereas several studies investigated the carbohydrate binding propensities of human Langerin (Stambach and Taylor, 2003; Takahara et al., 2004; Ward et al., 2006; Thépaut et

al., 2009; de Jong et al., 2010; Tateno et al., 2010; Feinberg et al., 2011; Holla and Skerra, 2011; Chabrol et al., 2012; Feinberg et al., 2013; Kim et al., 2014; Munoz-Garcia et al., 2015), there are no detailed reports on the effect of Ca^{2+} binding and its regulatory function to date. In the light of the previously described pH-dependence of glycan binding and differential cargo sorting in the early endosome, an in-depth analysis of the molecular mechanisms of Ca^{2+} binding may provide insight into the underlying principles of ligand release and sorting. Moreover, intra-domain allostery, which has been observed for CTLRs harboring multiple Ca^{2+} ions (vide supra), has not yet been investigated for a member of the family with a single Ca^{2+} binding site before.

1.7 Allostery – an undelaying organization principle conveying protein function

Allosteric regulation is a defining principle of life and has been named “the second secret of life” after the genetic code (Monod et al., 1963; Monod et al., 1965; Fenton, 2008). Although originally referring to enzymatic regulation mediated by small, non-substrate-like, molecules binding at distant site from the catalytic site (Ibid.), allosteric regulation is defined broader today as “the biological phenomenon where ligand binding or energetic perturbation at one molecular site results in structure or activity change at a second distinct site” (Hilser et al., 2012). In the last decade, the definition of allostery was expanded to changes in conformational dynamics that occur without major structural changes (Cooper and Dryden, 1984; Kern and Zuiderweg, 2003; Popovych et al., 2006; Boehr et al., 2009; Smock and Gierasch, 2009). Today, the ensemble view on allostery utilizing free energy landscapes of conformational equilibria in proteins has become the most fruitful approach to understand the underlying principles of allostery reconciling observations that challenged the previous purely structural models (Hilser et al., 2012; Motlagh et al., 2014).

Whereas X-ray crystallography has provided a vast number of three-dimensional structures of biomacromolecules and is unrivaled in atomic resolution, its abilities to observe dynamic processes are still limited although some studies demonstrated time-resolved observation of enzyme functions (Schotte et al., 2003; Keedy et al., 2015). In contrast, nuclear magnetic resonance spectroscopy (NMR) is to date the most powerful experimental techniques for the observation of dynamic processes at atomic resolution on all timescales (Torchia, 2011). To date, both NMR and X-ray crystallography are commonly viewed as complementary methods (Fenwick et al., 2014). Consequently, NMR has been employed widely to unravel dynamic allosteric processes (Nohaile et al., 1997; Wolf-Watz et al., 2004; Boehr et al., 2006; Kerns et al., 2015). Moreover, NMR provides the means to detect how information is propagated within the protein uncovering allosteric interaction networks (Akimoto et al., 2013; Boehr et

al., 2013; Cembran et al., 2014; Isogai et al., 2016). Lately, computational approaches have greatly contributed to unravel allosteric interaction pathways (Feher et al., 2014) either by employing topology analyses (Daily et al., 2008; Reynolds et al., 2011; van den Bedem et al., 2013) or by molecular dynamics simulation analyses (Sethi et al., 2009; VanWart et al., 2012). Recently, mutual information theory has been utilized to uncover correlated motions in proteins from MD simulations defining allosteric interaction networks (Keller et al., 2010; Bowman and Geissler, 2012).

1.8 An evolutionary perspective

Langerin is highly conserved in the mammalian lineage and has been described for 32 species where it evolved adaptively (Forni et al., 2014). Notably, P213 and A289 are positively selected sites. Whereas the former facilitates interactions at the trimerization interface, the latter is the only additional contact of high mannose sugars outside the binding site (vide supra). Moreover, haplotype analysis of the human Langerin gene showed a small group of individuals carrying two polymorphisms (N288D, K313I) that affect glycan specificity (vide supra). Surprisingly, this haplotype has been linked to an increased HIV resistance.

If subtle changes within one species already have pronounced impact on ligand binding propensities, the question arises whether function is conserved among species. Whereas most *in vitro* studies on Langerin were performed with the human homolog, *in vivo* studies in mice were based on the native murine homolog that shares 66% of the sequence (Valladeau et al., 2002). To date, there are only two cell-biological studies that report on ligand recognition of murine Langerin (Takahara et al., 2004; De Jesus et al., 2014). However, these studies entail no direct comparison with the human homolog. Moreover, the expression patterns of the two homologs are not identical: human Langerin is expressed on $CD1a^+$ APCs of the epithelium, epidermis and dermis whereas murine Langerin is additionally found on $CD8\square^+$ DCs that do not exist in human (Shortman and Liu, 2002; Valladeau et al., 2002; Chang and Kweon, 2010; Romani et al., 2010; De Jesus et al., 2014). So far, the two homologs have been treated functionally interchangeable in the literature (Drickamer and Taylor, 2015) although the differential expression in various tissues hints a diverging specificity and consequently an alternate function.

1.9 Enabling tools

In the last decade, glycan microarrays have enabled the screening of large sets of glycan structures to interrogate GBP specificities thereby paving the way for functional glycomics (Disney and Seeberger, 2004; Paulson et al., 2006; Rillahan and Paulson, 2011). Today,

glycan arrays comprising mammalian self-antigens and viral antigens (Blixt et al., 2004), heparin oligosaccharides (de Paz et al., 2006), and microbial glycan antigens (Stowell et al., 2014) are available to decipher protein-glycan interactions. The available arrays differ in origin of the glycans, synthetic versus purified from natural sources, the immobilization and linker chemistries, and the supporting surface (Rillahan and Paulson, 2011). However, the detection principle is common: a GBP is either directly conjugated with a fluorescent dye or labeled with a fluorescent antibody and incubated on the array slide. To enhance the intrinsically weak interactions, GBPs have been fused to Fc fragments for multivalent presentation (Paulson et al., 2006). Although glycan arrays are generally employed for qualitative screening, some groups quantified the strength of the interactions on the array (Liang et al., 2007).

As every screening platform, glycan arrays are prone to false-positive and false-negatives (Rillahan and Paulson, 2011). Therefore, hits have to be confirmed by orthogonal methods: Plate-based assays that either detect direct binding to the immobilized glycan directly or are performed in a competition format do not require advanced equipment and are widely utilized (Taylor and Drickamer, 2003). Furthermore, surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) are employed to verify binding and obtain both kinetics and thermodynamic data on the interactions (Karlsson and Fält, 1997; Freyer and Lewis, 2008). However, the latter method requires high amounts of glycan ligand (several 100 µg) that are often not available in case of complex glycans obtained from natural source thereby limiting its applicability. Moreover, both protein- and ligand-observed NMR spectroscopy provide powerful means to investigate the underlying molecular mechanism of glycan-protein interactions. Specifically, the intrinsically weak interactions of glycans with their receptors can be readily investigated by NMR methodologies (Meyer and Peters, 2003) (vide infra).

The interested reader is referred to (Christie et al., 2014) for a detailed review of biophysical methodologies to investigate glycan-lectin interactions.

1.10 Thesis objectives

The aim of this thesis was to gain a more profound understanding of the underlying molecular principles that govern glycan binding and release and convey glycan specificity in the endocytic C-type lectin receptor Langerin that serves as pattern recognition receptor at the interface of innate and adaptive immunity.

To this end, the development of a recombinant expression system to obtain sufficient amounts of isotope-labeled protein for biomolecular NMR studies was imperative. Moreover, the

backbone resonances of human Langerin had to be assigned to obtain site-specific information from NMR experiments. These two milestones for further studies are described in Chapter 4.1.

To understand the underlying molecular mechanisms regulating Ca^{2+} binding in the human Langerin carbohydrate recognition domain, I applied biomolecular NMR methodologies and isothermal titration calorimetry. Specifically, ^{1}H-^{15}N HSQC NMR titration experiments and ^{15}N relaxation NMR studies were chosen to interrogate changes in structure and dynamics induced by Ca^{2+} binding. Furthermore, site-directed mutagenesis was applied to aid the elucidation of allosteric interactions within the protein. To assess whether glycan binding is controlled by similar mechanisms as the interactions with Ca^{2+}, Langerin binding to simple oligosaccharide ligands was investigated by ^{1}H-^{15}N HSQC NMR titration experiments. The findings are presented in Chapter 4.2.

To gain insight into the evolutionary conservation of glycan specificity in two closely related homologs, the glycan binding propensities of the human and the murine Langerin were investigated side-by-side by glycan array analysis on an array comprising more than 300 pathogen-associated glycans, plate-based assays, ligand-based NMR, and direct binding to heat-killed bacteria. The findings are described in Chapter 4.3.

2 NMR Theory

Nuclear magnetic resonance spectroscopy serves as a powerful tool in structural biology providing both structural and dynamics information of biomacromolecules. First described by Bloch and Purcell (Bloch et al., 1946; Purcell et al., 1946), it exploits the intrinsic quantum physical property of the spin angular momentum of atomic nuclei, which is very sensitive to its chemical environment and its changes. In the context of biomolecules, the ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P isotopes possess the most favorable NMR properties with a nuclear spin angular momentum number *I* of ½. Although an inherently insensitive method, the implementation of pulsed FT-NMR (Ernst and Anderson, 1966), the concept of multidimensional NMR (Jeener, 1971), technological improvements in instrumentation, the design of powerful pulse sequences, and isotope labeling of biomacromolecules by means of biotechnology enabled NMR spectroscopy to become one of the two biophysical techniques that allow for both structure elucidation and observation of conformational dynamics with atomic resolution.

For a comprehensive introduction into NMR theory, the reader is referred to the textbook literature in the field (Cavanagh et al., 2007; Levitt, 2008; Keeler, 2010). In the following, a short introduction of the theoretical basis of NMR is given. Then, three important NMR phenomena are discussed that were exploited in this work: i) the chemical shift δ, ii) motional processes and time-scales, and iii) the nuclear spin relaxation. Finally, the NMR methodologies that were applied in this work will be reviewed in more detail.[1]

2.1 Basic quantum mechanics

2.1.1 The nuclear magnetic spin Hamiltonian

NMR theory follows the postulations of quantum physics and can be described in total by its equations. Therefore, the time-dependent Schrödinger equation can be formulated for the nuclear spin:

$$\frac{\partial}{\partial t}\left|\Psi(t)_{spin}\right\rangle \cong -\mathrm{i}\hat{H}_{spin}\left|\Psi(t)_{spin}\right\rangle \qquad (2\text{-}1)$$

With $\hat{H}_{spin}$ as the nuclear spin Hamiltonian which describes all direction changes in nuclear spin polarization under the assumption that both magnetic and electrical effects of the fast

[1] Most parts of this theory section are excerpted textbook knowledge and references for the relevant textbooks are given at the end of each paragraph. Nonetheless, seminal papers describing important phenomena or novel pulse sequences are referenced separately to give credit to the researchers.

moving electrons are canceled out. In a static magnetic field $\mathbf{B_0}$, the nuclear spin Hamiltonian $\hat{H}_{spin}$ can be described in terms of the angular momentum operator:

$$\hat{H}_{spin} = -\gamma \mathbf{B}_0 \hat{I}_z \tag{2-2}$$

with γ as the gyromagnetic ratio and $\hat{I}_z$ as the operator describing the z-component of the nuclear spin angular momentum **I** in a Cartesian coordinate system. The eigenfunctions of $\hat{I}_z$ depending on the spin quantum number *m* are:

$$\hat{I}_z \psi_m = m\hbar \psi_m \tag{2-3}$$

with the eigenvalue $m\hbar$ and $m = \pm\frac{1}{2}$ for spin-half nuclei.

The spin angular momentum is a vector property that also contains a x- and y- component besides the above mentioned z component. Whereas ψ_m is a eigenfunction of $\hat{I}_z$, it is not an eigenfunction of either $\hat{I}_x$ or $\hat{I}_y$. The spin angular momentum operators can be described for a spin-half nucleus in the matrix notation as follows:

$$\hat{I}_x = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \hat{I}_y = \frac{1}{2\mathrm{i}}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad \hat{I}_z = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2-3}$$

Inspecting the matrices in (2-8) shows that in $\hat{I}_x$ and $\hat{I}_y$, only off-diagonal elements are populated whereas in $\hat{I}_z$ only diagonal elements are populated (Levitt, 2008; Keeler, 2010).

2.1.2 Energy levels

The eigenfunction ψ_m of the spin angular momentum operator $\hat{I}_z$ are also eigenfunction of the nuclear spin Hamiltonian $\hat{H}_{spin}$:

$$\hat{H}_{spin}\psi_m = -\gamma \mathbf{B}_0 \left[\hat{I}_z \psi_m\right] = -m\hbar\gamma \mathbf{B}_0 \psi_m \tag{2-4}$$

with the eigenvalue:

$$E_m = -m\hbar\gamma \mathbf{B}_0 \tag{2-5}$$

with E_m as the energy level of the spin *m*.

For spin-half nuclei, according to (2-5) a single nuclear magnetic spin can occupy two energy levels that are usually termed α and β. Every spectroscopy mediates the transition of spins between energy states by exciting the ground state by means of electromagnetic irradiation with the wavelength that matches the energy difference of two different energy states. In

NMR spectroscopy, this is done by radio frequency (RF) pulses. However, unlike in other spectroscopies, the NMR spectrometer has to generate the energy difference of the eigenstates of the nuclear spin α and β first by applying the external magnetic field $\mathbf{B}_0$. According to (2-5), the energy difference of the states $\Delta E_{\alpha\to\beta}$ scales with the strength of $\mathbf{B}_0$:

$$\Delta E_{\alpha\to\beta} = E_\beta - E_\alpha = \frac{1}{2}\hbar\gamma\mathbf{B}_0 - (-\frac{1}{2}\hbar\gamma\mathbf{B}_0) = \hbar\gamma\mathbf{B}_0 \tag{2-6}$$

Thus, the transition of one spin from α to β can be facilitated by applying a radio frequency wave or pulse with the wavelength:

$$\omega_0 = -\gamma\mathbf{B}_0 \tag{2-7}$$

where ω_0 is termed the Larmor frequency in radians of the nuclear magnetic spin.

2.1.3 The density matrix

NMR spectroscopy always measures ensembles of nuclear spins. Thus, the system cannot be described by a single wavefunction. However, applying statistical quantum mechanics, the system can be divided in subensembles each described by a wavefunction Ψ and a probability density $P(\Psi)$ representing the contribution of the state to the ensemble. Thus, the expectation value $\langle\bar{A}\rangle$ for the ensemble can be described as follows:

$$\langle\bar{A}\rangle = \int P(\Psi)\langle\Psi|\hat{A}|\Psi\rangle d\tau \tag{2-8}$$

With $\langle\Psi|\hat{A}|\Psi\rangle$ as the eigenfunction of the operator $\hat{A}$ in the Dirac notation[2] assuming that the superposition state $\langle\Psi|\Psi\rangle$ is normalized i.e. equals one. Applying some basic quantum mechanics operations, equation (2-2) can be rewritten as:

$$\langle\bar{A}\rangle = \sum_{nm}\overline{c_n c_m^*}\langle m|\hat{A}|n\rangle \tag{2-9}$$

with $\overline{c_n c_m^*}$, the ensemble average of the coefficients, forming the density matrix, which is the matrix representation of the density operator ρ_{nm}. Whereas $\langle m|\hat{A}|n\rangle$ is constant, the density operator varies with the investigated system. The density operator can be expressed as a linear combination of Cartesian components of the spin angular momentum (2.8) and an additional operator $\hat{E}$:

[2] See standard textbooks for a detailed introduction e.g. Levitt, M.H. (2008). Spin dynamics: basics of nuclear magnetic resonance, 2nd edn (Chichester, U.K.; Hoboken, NJ: John Wiley & Sons).

$$\hat{\rho} = a_E \hat{E} + a_x \hat{I}_x + a_y \hat{I}_y + a_z \hat{I}_z \tag{2-10}$$

The above expression can also be described in the matrix forms of the operators:

$$\rho = a_E \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + a_x \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} + a_y \begin{pmatrix} 0 & -\frac{1}{2}\mathrm{i} \\ \frac{1}{2}\mathrm{i} & 0 \end{pmatrix} + a_z \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 2a_E + a_z & a_x - \mathrm{i}a_y \\ a_x + \mathrm{i}a_y & 2a_E - a_z \end{pmatrix} \tag{2-11}$$

Notably, the resulting matrix describes a mixed state with the diagonal elements describing the probabilities to be in one of the two basis states depending solely on $\hat{E}$ and $\hat{I}_z$, whereas the off-diagonal elements are a measure for the coherence between these two state depending merely on $\hat{I}_x$ and $\hat{I}_y$. The factors a_x, a_y, and a_z are time-dependent and proportional to the respective Cartesian components M_x, M_y, and M_z of the macroscopic bulk magnetization **M.**[3]

In equilibrium, all off-diagonal elements are zero and the equilibrium density operator is:

$$\hat{\rho}_{eq} = \hat{I}_z \tag{2-12}$$

In NMR experiments, coherence is generated by applying RF pulses orthogonal with respect to the $\mathbf{B}_0$ field.

Whereas the equilibrium density operator $\hat{\rho}_{eq}$ is time-independent, upon generation of coherence, the density operator evolves with time:

$$\hat{\rho}(t) = \exp(-\mathrm{i}\hat{H}t)\,\hat{\rho}(0)\exp(\mathrm{i}\hat{H}t) \tag{2-13}$$

with $\hat{\rho}(0)$ as the density operator at time 0, and $\hat{H}$ as the Hamiltonian that determines how the spins evolve in the time period $0 \rightarrow t$.

NMR experiments are a series of pulses and evolution times. Pulses generate coherence that evolves over time. In systems that contain different spins, coherence can be transferred between spins by means of pulses or coupling during evolution. The product operator formalism assumes that both pulses and delays, i.e. the different blocks of the pulse sequence, are commutative and thus allows for a straightforward analysis. However, this is only true for weakly coupled systems that do not experience significant relaxation over experimental time.[4]

[3] For a detailed introduction into the concept of the bulk magnetization, the reader is referred to the textbook literature.

[4] For a detailed treatise on the product operator, the reader is referred to the textbook literature.

2.2 Chemical shift, motions, and spin relaxation

2.2.1 The chemical shift

The chemical shift δ of a certain nucleus in a molecule describes the deviation from the precise Larmor frequency ω_L of a nucleus in a molecular environment (Knight, 1949). It arises from electronic motions caused by the external magnetic field that induce local magnetic fields. These local fields can diminish or augment the effect of the external field experienced by each nucleus in the molecule. This effect is called nuclear shielding and can be described as follows

$$\mathbf{B} = \mathbf{B}_0(1-\sigma) \tag{2-14}$$

where **B** is the net magnetic field experienced by each nucleus, $\mathbf{B}_0$ the bulk magnetic field applied to the sample, and σ the internal magnetic susceptibility of the electron distribution or shielding tensor. If the shielding tensor is anisotropic, the effect can also be described by the chemical shift anisotropy (CSA):

$$\sigma_{CSA} = \sigma_{11} - \frac{\sigma_{22}+\sigma_{33}}{2} \tag{2-15}$$

where σ_{11}, σ_{22}, σ_{33} are the principal components of the shielding tensor σ with σ_{11} as the principal value furthest from the isotropic value. As the molecule tumbles in solution, the local magnetic field fluctuates resulting in the CSA relaxation mechanism (*vide infra*). The chemical shift δ is directly proportional to the applied magnetic field. To compare shifts obtained at different fields and spectrometers, it is given relative to a reference resonance:

$$\delta = \frac{\omega - \omega_{\text{ref}}}{\omega_{\text{ref}}} \tag{2-16}$$

where ω is the Larmor frequency experienced by the nucleus and ω_{ref} the Larmor frequency of a reference compound. Due to its electronic origin, the chemical shift is very sensitive to the chemical environment of the nucleus and its changes. Therefore, it provides an excellent probe for both structural and dynamics analyses (Cavanagh et al., 2007).

2.2.2 Motional processes and chemical exchange

Motional processes are “visible” in NMR if they induce a change in the nuclear spin Hamiltonian. Thus, NMR is superior to many other techniques as it allows detection of molecular motion itself and does not derive the information solely from population changes. Molecular motions can occur on different timescales: from very fast molecular vibrations in the picosecond to nanosecond range, molecular rotation in the ns time scale up to chemical

exchange on the chemical shift timescale in the microsecond to second range. **Figure 2-1** provides an overview of the detectable timescales, the corresponding NMR observables, and some relevant processes occurring on these timescales. Rapid motions (fs to µs) cause motional averaged spin interactions and do not influence the chemical shift. However, slower molecular motions like chemical exchange that occurs on the micro- to millisecond timescale modulate the isotropic chemical shifts by altering their chemical environment.

For a two-state exchange system consisting of the states *A* and *B*:

$$A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} B \tag{2-17}$$

with the zero-order rate constants k_1 and k_{-1} describing the interconversion of the states. The exchange rate k_{ex} of the system is defined as the sum of the rate constants:

$$k_{ex} = k_1 + k_{-1} \tag{2-18}$$

The relation between the exchange rate k_{ex} and chemical shift difference of the two states $\Delta\omega = \omega_B - \omega_A$ [5] influence the motional averaging resulting in different lineshapes. Two limiting conditions can be described: i) slow exchange with $k_{ex} << \Delta\omega$, and ii) fast exchange with $k_{ex} >> \Delta\omega$. In the case of slow exchange, two distinct peaks can be observed with both peaks resonating at their corresponding Larmor frequencies. In this case, the equilibrium population fractions of the two states and thus the equilibrium constant $K = p_B p_A^{-1}$ can be readily quantified by comparing the peak integrals, with p_A and p_B as the population fraction of the two states and $1 = p_A + p_B$. If the states are exchanging fast, the peaks coalesce to one peak with an apparent chemical shift given by the mean of the two peaks weighted by the corresponding fractions of the two states:

$$\omega_{peak} = p_A \omega_A + p_B \omega_B \tag{2-19}$$

where ω_{peak} is the chemical shift of the coalesced peaks in the extreme narrowing limit and ω_A and ω_B the chemical shifts of the states without exchange. Thus, the equilibrium constant *K* can be determined:

$$\omega_{peak} = \frac{\omega_A + K\omega_B}{1 + K} \tag{2-20}$$

If the exchange rate is in the range of the chemical shift difference ($k_{ex} \approx \Delta\omega$), intermediate exchange is observed with broad lines. In the case of $k_{ex} < \Delta\omega/2$, slow intermediate exchange

[5] Note that the chemical shift is denoted here as ω indicating its absolute value in s^{-1}. The chemical shift time-scale is dependent on the $\mathbf{B}_0$ field.

is observed with exchange broadening with increasing k_{ex}. In the case of $k_{ex} > \Delta\omega/2$, fast intermediate exchange is observed with exchange narrowing occurring with increasing k_{ex}. In practice, it is often difficult to determine the exact exchange regime especially to distinguish between fast intermediate and fast exchange because the chemical shift differences are not directly accessible as in the case of slow exchange. Lineshape analysis might provide useful insight since fast intermediate exchanging peaks are broadened by the exchange contribution (Levitt, 2008; Keeler, 2010).

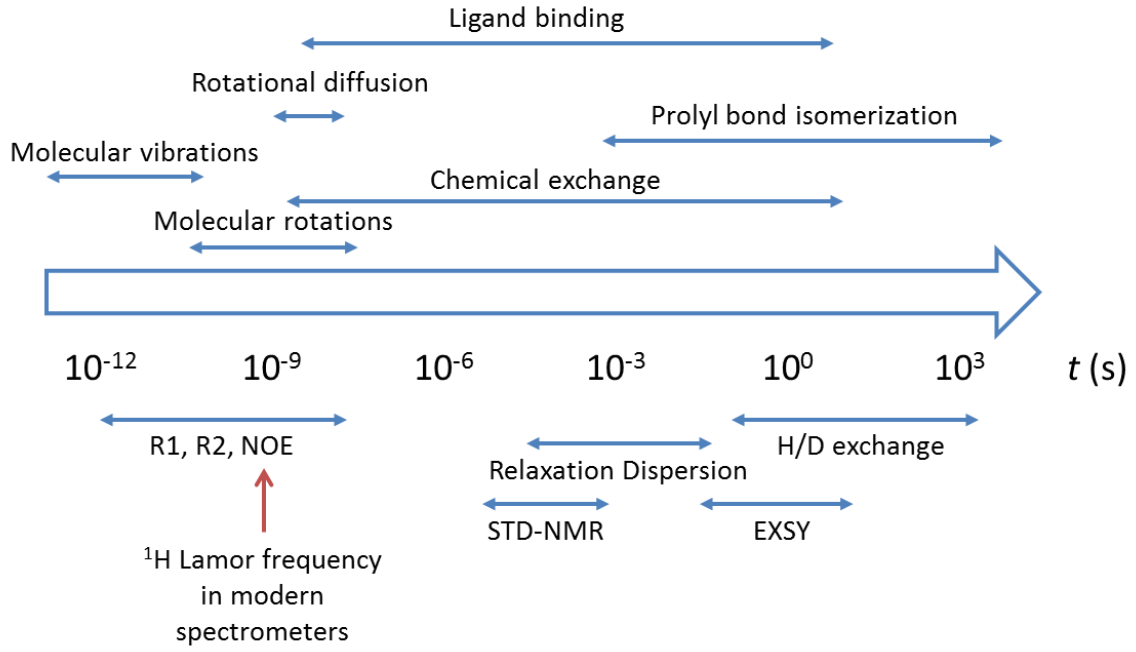


Figure 2-1: Biomacromolecules undergo motions on several timescales. An overview of the timescales of motions experienced by biomacromolecules, the related biophysical processes (above the arrow) and the NMR observables reporting on motions in the respective time regime (below the arrow).

2.2.3 Nuclear spin relaxation

In NMR, RF pulses applied to the sample disturb the equilibrium and create coherence. Nuclear spin relaxation describes the return of the system to thermal equilibrium. In the context of NMR, this entails that all coherences are absent and the populations of the energy levels are in Boltzmann equilibrium. Two types of relaxation processes control the return to equilibrium: i) longitudinal or R_1 relaxation describes the return of the populations to the Boltzmann equilibrium whereas ii) transverse or R_2 relaxation describes the decay of single quantum coherences. Using of density operator formalism, R_1 relaxation only affects the diagonal elements whereas R_2 affects the off-diagonal elements. An additional type of relaxation is the cross-relaxation via other spins which gives rise to the NOE. The study of relaxation processes in molecules gives insight into many motional processes. Furthermore,

relaxation processes are sensitive to non-secular spin interactions[6] which cannot be observed in ordinary NMR spectra. The origin of relaxation of spin ½ nuclei are local fluctuating magnetic fields caused by random thermal motion of the molecules. These fields either originate from the interactions of dipolar fields of two molecules (dipole-dipole interaction), the chemical shift anisotropy (CSA), or spin rotation. The afore-mentioned relaxation mechanisms are listed in order of their importance. CSA only becomes an important contributor at very high fields (>600 MHz). Other very strong relaxation mechanisms are quadrupole relaxation or paramagnetic relaxation enhancement (PRE) induced by unpaired electrons.

The random thermal motion can be described and quantified by the autocorrelation function *G*:

$$G(t,\tau) = \left\langle \mathbf{B}_{\text{loc}}(t)\mathbf{B}_{loc}(t+\tau) \right\rangle \neq 0 \qquad (2\text{-}21)$$

where $\mathbf{B}_{\text{loc}}$ is the local magnetic field, t the time and τ the time constant that describes how fast the autocorrelation functions decays. Fast fluctuating fields have short decays whereas slowly fluctuating fields imply long decays of the autocorrelation function. At τ = 0 the autocorrelation function is:

$$G(0) = \left\langle \mathbf{B}_{\text{loc}}{}^{2}(t) \right\rangle \qquad (2\text{-}22)$$

Assuming that the autocorrelation is independent of the time point t (stationary assumption), one can describe the autocorrelation function as a simple decay function only depending on τ:

$$G(\tau) = \left\langle \mathbf{B}_{\text{loc}}{}^{2} \right\rangle e^{-|\tau|/\tau_c} \qquad (2\text{-}23)$$

with τ_c as the correlation time. The Fourier transform of the autocorrelation function yields the spectral density function:

$$J(\omega_0) = \left\langle \mathbf{B}_{\text{loc}}{}^{2} \right\rangle \frac{2\tau_{\text{c}}}{1+\omega_0^2\tau_{\text{c}}^2} \qquad (2\text{-}24)$$

where ω_0 is the Larmor frequency of the nucleus.

Following classical relaxation theory, the relaxation rate constants R_1, R_2 and the cross-correlation rate σ_{NOE} of a HX two-spin system like ^{15}N and ^{1}H of the amide group of the

[6] Non-secular spin interactions are the relaxation components of the internal spin Hamiltonian that are lost during the secular approximation and motional averaging. For a detailed treatise of the internal spin Hamiltonian see Levitt, M.H. (2008). Spin dynamics: basics of nuclear magnetic resonance, 2nd edn (Chichester, U.K.; Hoboken, NJ: John Wiley & Sons).

protein backbone can be described in terms of their respective spectral density functions as follows (Abragam, 1961):

$$R_1 = d\left[J(\omega_H - \omega_x) + 3J(\omega_X) + 6J(\omega_H + \omega_X)\right] + cJ(\omega_X) \quad (2\text{-}25)$$

$$R_2 = \frac{d}{2}\left[4J(0) + J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H) + 6J(\omega_H + \omega_X)\right] + \frac{c}{6}\left[3J(\omega_X) + 4J(0)\right] + R_{ex} \quad (2\text{-}26)$$

$$\sigma_{NOE} = d\left[6J(\omega_H + \omega_X) - J(\omega_H - \omega_X)\right] \quad (2\text{-}27)$$

with $d = \frac{1}{4}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_H^2 \gamma_X^2 \hbar^2}{\left\langle r_{HX}^6 \right\rangle}$ and $c = \frac{(\omega_X \sigma_{CSA})^2}{3}$.

In the equations (2-23-25), X is the hetero nucleus, $J(\omega_i)$ is the spectral density function at the corresponding Larmor frequency, R_{ex} is the chemical exchange contribution, γ_i is the gyromagnetic ratio, μ_0 is the magnetic susceptibility, B_0 is the external magnetic field, $\hbar$ is the Planck constant and r_{HX} is the average H-X bond length and σ_{CSA} the CSA. For ^{15}N backbone amides, experimental data showed that the CSA is about -170 ppm (Tjandra et al., 1996). More recent studies, however, revealed a variability of the CSA depending on the amino acid type and the protein structure (Fushman et al., 1998) as well as on the static magnetic field (Millet et al., 2000). Nonetheless, a uniform value for the CSA is often applied because direct assessment on a per residue resolution require extensive measurements.

The steady-state NOE relates to the cross-relaxation rate σ_{NOE} as follows:

$$NOE = 1 + \frac{\gamma_H}{\gamma_X}\frac{\sigma_{NOE}}{R_1} \quad (2\text{-}28)$$

The equations (2-25-27) above suffice to describe the experimentally obtained relaxation data and their relations, but do not explain the microscopic origin of relaxation that can only be understood by the use of spin internal Hamiltonians. Again, the reader is referred to standard textbooks (Cavanagh et al., 2007; Levitt, 2008).

A closer inspection of eq (2-25-27) shows that the relaxation rate constants only depend on the spectral density functions, which occur on the ps-ns timescale and are insensitive for slower motions. Only the transverse relaxation rate R_2 contains a contribution from slower µs-ms motions. Thus, exchange rates can be extracted from observed transverse relaxation rate constants $R_{2,eff}$ if the intrinsic relaxation rate constant without exchange $R_2^0$is known:

$$R_{2,eff} = R_2^0 + R_{ex} \quad (2\text{-}29)$$

For slowly exchanging systems as described above, R_{ex} is equal to k_1 or k_{-1} depending on the observed state. If the equilibrium distribution is known, the exchange rate can thus be readily determined:

$$R_{ex,A} = k_1 = (1 - p_A)k_{ex} \tag{2-30}$$

2.2.3.1 Model-free analysis of relaxation data

The spectral density function $J(\omega_i)$ for a protein in solution consists of two components: i) the overall motion of the biomacromolecule as a whole and ii) the internal motion of the N-H bond vector. Assuming decoupled correlation functions for overall Brownian diffusion G_O and internal motions G_I (i.e. the motions occur independently):

$$G(\tau) = G_O(\tau)G_I(\tau) \tag{2-31}$$

Lipari and Szabo introduced a formalism that allows for dissecting the internal motions and the overall tumbling of the biomacromolecule without any physical model assumption for the nature of the N-H bond (Lipari and Szabo, 1982a, b). Thus, they called their approach model-free analysis. According to their formalism, the spectral density function $J(\omega_i)$ can be described as:

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_m}{1+(\tau_m\omega)^2} + \frac{(1-S^2)\tau}{1+(\tau\omega)^2} \right) \tag{2-32}$$

with S^2 as the generalized order parameter, τ_m the overall correlation time of the biomacromolecule and $\tau^{-1} = \tau_m^{-1} + \tau_e^{-1}$, with τ_e as the effective correlation time of each N-H bond vector. The generalized order parameter S^2 can be interpreted as a measure for the amplitude of internal motions and τ_e as the timescale of these motions (Yang and Kay, 1996).

Furthermore, the model assumes isotropic tumbling for the biomacromolecule. It was later extended to account to the possible existence of two motional regimes acting on the N-H bind vector (Clore et al., 1990b):

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_m}{1+(\tau_m\omega)^2} + \frac{(1-S_f^2)\tau_f}{1+(\tau_f\omega)^2} + \frac{(S_f^2-S^2)\tau_s}{1+(\tau_s\omega)^2} \right) \tag{2-33}$$

with τ_f and τ_s as the fast and slow correlation times and S_f^2 and S_s^2 as the corresponding generalized order parameters and $S^2 = S_s^2 S_f^2$.

2.2.4 ^3J couplings

Besides the above discussed field-dependent CSA, electrons also mediate scalar bilinear spin-spin couplings that affect the nuclear spin Hamiltonian:

$$\hat{H}_{spin} = 2\pi \mathbf{J}_{jk} \hat{I}_j \hat{I}_k \tag{2-34}$$

with $\mathbf{J}_{jk}$ as the scalar coupling tensor of the two spins *j* and *k*. In isotropic liquids, the tensor is averaged out simplifying to a J_{jk}, the scalar coupling constant. The *J* coupling is independent of the $\mathbf{B}_0$ field strength and has a sign depending on the gyromagnetic ratios of the involved nuclei. Scalar couplings can act over several chemical bonds within one molecule and their magnitude is very sensitive to the bond geometries. In biomolecular NMR, the three-bond ^{3}J coupling of the amide proton $^1H^N$ with the C_α proton $^1H^\alpha$ can serve as probe for the H^N-N-C_α-H^α torsional angle θ according to the a semi-empirical Karplus relation (Karplus, 1959):

$$^3J(H^N H^\alpha) = \left[6.4\cos^2\left(\phi - \frac{\pi}{3}\right) - 1.4\cos\left(\phi - \frac{\pi}{3}\right) + 1.9 \right] \text{Hz} \tag{2-35}$$

The relation is simulated in Figure 2-2. Using Ramachandran plot analysis, the secondary structure of a protein can be predicted based on the distribution of the torsion angles. $^3J(H^N H^\alpha)$ coupling constants can be obtained from the HNHA experiment.

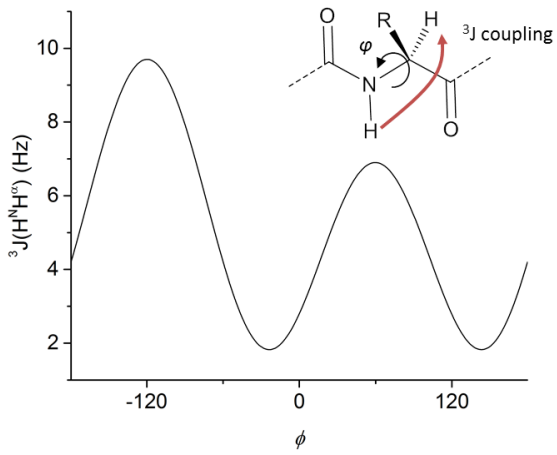


Figure 2-2: ^3J scalar coupling constants of the amide and C_α proton are sensitive to the torsion angle of the N-C_α bond following an semi-empirical Karplus relation that is simulated above.

2.3 Experimental approaches

Every NMR experiment consists of an excitation pulse on a nucleus of interest that generates magnetic coherence, which can be manipulated and transferred by a sequence of pulses and delays, and finally recorded in form of the time-dependent free induction decay (FID). Consecutive Fourier transformation yields a spectrum of frequency-resolved signals. Incrementing delays before transfers allows for more-dimensional spectroscopy.

2.3.1 The HSQC experiment

Throughout this work a basic two-dimensional NMR experiment was used: the ^{1}H-^{15}N hetero single quantum coherence transfer (HSQC) experiment correlates the chemical shifts of ^{1}H and ^{15}N nuclei directly connected by a chemical bond and is widely used in structural investigations of proteins. The basic pulse sequence (Figure 2-3) consists of a excitation of the spin *I*, a consecutive INEPT transfer of coherence to the hetero spin *S*[7], in incremented evolution time t_1 where the magnetization evolves on *S* while it is refocused on *I*, a second INEPT coherence transfer back to spin *I* and recording of the FID on the proton channel while decoupling the hetero channel during t_2.

For protein backbone analyses, uniform ^{15}N-labeled protein is required. Even without assignment of the backbone resonances, the spectrum readily provides information on the protein folding just by spread of the resonances over the chemical shift scale. Well-spread resonances with chemical shifts distributed between 90 ppm and 135 ppm in the ^{15}N- and 6 ppm and 11 ppm in the ^{1}H-spectral dimension, respectively, are the fingerprints of a folded protein whereas in unstructured proteins the peaks cluster around 8 ppm in $\delta(^{1}\mathrm{H})$. Due to the strong dependence of R_2 relaxation on the overall correlation time τ_m, only isotropic proteins below 25 kD MW provide satisfying spectra. However, the size limit has been pushed to larger proteins and protein complexes >100 kD MW by the introduction of transverse relaxation optimized NMR spectroscopy (TROSY) and deuteration of proteins exploiting favorable relaxation properties at high field (Pervushin et al., 1997).

[7] The INEPT transfer describes the coherence transfer using anti-phase states. For a detailed treatise the reader is referred to standard textbook literature.

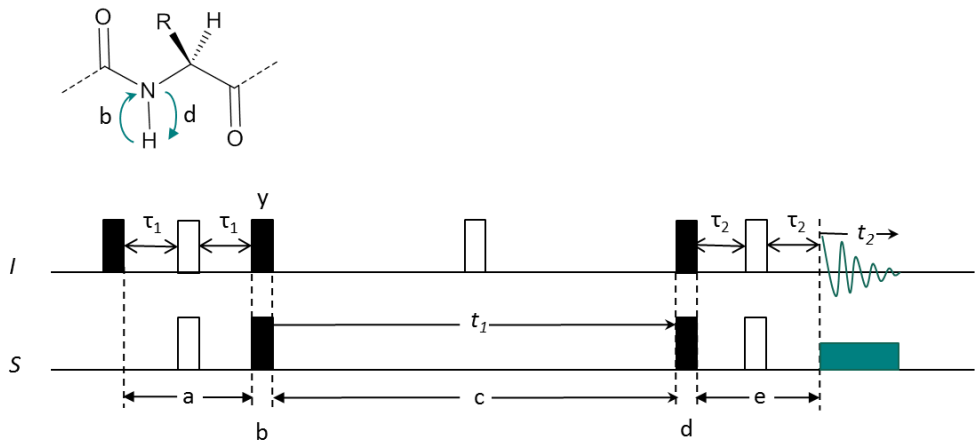


Figure 2-3: Schematic representation of a standard HSQC pulse sequence. A 90° pulse on the proton nuclei (*I*) generates coherence which is transferred via an INEPT transfer (b) while the chemical shift is refocused during (a). During the incremented mixing period t_1 (c) the chemical shift on the hetero nucleus (*S*) evolves while it is refocused in the proton dimension. A second INEPT transfer (d) transfers the coherence back to the protons. After refocusing (e) it is recorded during t_2 while the hetero channel is decoupled. The coherence transfers occurring during steps (b) and (d) are marked on the structure above the sequence. Empty rectangular boxes indicate 180° (π) pulses and filled boxes 90° (π/2) pulses (Keeler, 2010).

2.3.1.1 Chemical shift perturbations as measure for ligand binding and conformational change

As discussed above, line shape and peak width readily provide information about the protein flexibility. Broad lines with lower intensities indicate exchange processes on the chemical shift timescale. Moreover, chemical shift perturbations of the protein resonances induced by ligand binding provide means to quantify the affinity of the interactions. Assuming a simple one-to-one binding of a ligand *L* to a protein receptor *P*, the equilibrium with the resulting receptor-ligand complex *PL* can be described as follows:

$$P + L \rightleftharpoons PL$$

All resonances of residues that experience a change in chemical environment upon ligand binding will be perturbed. These changes may result from either direct interactions of the residues with the ligand or conformational rearrangement of the binding pocket (e.g. induced fit) or in remote areas of the protein (conformational allostery). Moreover, changes in internal dynamics can also result in chemical shift changes due to shifting population ensembles (Cooper and Dryden, 1984; Kern and Zuiderweg, 2003). Depending on the exchange regime on the NMR timescale, the resonances can either move along a vector as the equilibrium populations of unbound and bound protein (fast and fast-intermediate exchange regime) change, or one peak diminishes while another peak appears (slow and slow intermediate exchange regime). Due to the non-trivial nature of the origin of chemical shift changes observed for protein backbone resonances, there is no rigorous method to quantify chemical shift changes (Williamson, 2013). In two-dimensional ^{1}H-^{15}N HSQC NMR, the change in

chemical shift is often quantified by calculating the averaged Euclidean distance using a scaled ^{15}N chemical shift:

$$\Delta\delta = \sqrt{\frac{1}{2}\left[\Delta\delta_{\mathrm{H}}^2 + \left(\alpha\Delta\delta_{\mathrm{N}}\right)^2\right]} \tag{2-36}$$

with $\Delta\delta_i$ as the difference in chemical shift (in ppm) and α an empirical weighing factor. The optimal value for α is disputed and several approaches can be found in the literature (Williamson, 2013). Throughout this work, a weighing factor of 0.2 for glycine and 0.14 for all other amino acid backbone resonances is used. In the case of low-affinity binders showing exchange in the fast limit, it is often difficult to reach saturation and the chemical shift of the bound fraction remains unknown. Therefore, the K_d cannot be determined as done in eq. (2-17). However, eq. (2-18) can be transformed to include the change in chemical shift to determine the dissociation constant K_d for the receptor ligand complex:

$$\Delta\delta_{obs} = \Delta\delta_{max}\frac{\left([P]_t + [L]_t + K_d\right) - \sqrt{\left([P]_t + [L]_t + K_d\right)^2 - 4[P]_t[L]_t}}{2[P]_t} \tag{2-37}$$

with $\Delta\delta_{obs}$ as the observed chemical shift difference, $\Delta\delta_{max}$ the maximal chemical shift difference, and $[P]_t$ and $[L]_t$ as the total protein and ligand concentrations. $\Delta\delta_{max}$ and K_d can be fitted as free parameters when $\Delta\delta_{obs}$ is measured at different ligand concentrations. For slow exchanging populations, the peak of the unbound population diminishes while a new peak for the bound population arises and the K_d is readily accessible by comparing the peak integrals (*vide supra*). However, the afore-mentioned is only true for the limiting cases of very fast and very slow exchange, respectively. Care must be taken to discern intermediate fast and intermediate slow exchange from the limiting conditions which can usually be undertaken by line shape analysis (*vide supra*).

As in all experimental fields, the differentiation between signal and experimental noise is always a tradeoff between sensitivity and specificity. Therefore, a cut-off value has to be estimated that allows for discrimination of meaningful data from background noise. In NMR practice, the definition of a cut-off value for the CSP remains challenging since the experimental noise in the chemical shift is difficult to assess. Setting a low cut-off will result in high sensitivity while increasing the rate of false positive signals whereas setting a high value will result in high specificity while potentially losing relevant signals due to the reduced sensitivity. In context of the CSP, this problem is especially challenging because there is no linear relation between the magnitude and the significance of the CSP. In many cases, arbitrary definitions of cut-off values based on the magnitude of shift observed in the

spectrum are made. However, there are also methods that try to determine the cut-off value based on the experimental data. Schumann et al. presented a method that uses the standard deviation of all CSP from zero (σ_0) in one recorded spectrum assuming a normal distribution of the noise (Schumann et al., 2007):

$$\sigma_0 = \sqrt{\frac{1}{N-1}\sum_{i=1}^{n} CSP_i^2} \qquad (2\text{-}38)$$

In order to reduce the bias introduced by strong signals, all CSP with values above $3\sigma_0$ are iteratively removed from the calculation. The final σ_0 is used as cut-off. Alternatively, undisturbed regions or high mobility regions with fast exchanging peaks like the resonances of the sequence termini or tags should only report on the changes introduced by temperature variations or buffer conditions and can therefore serve for noise estimation and definition of the cut-off value. To this end, the mean of the CSP in addition of two standard deviations can be defined as a cut-off. The latter method is less prone to artifacts introduced by many shifting residues with similar amplitude as can be observed in highly allosteric proteins which would result in an artificially high σ_0. However, if several experiments with different preparation batches at different spectrometers are recorded, the definition of the CSP of standard deviation of the mean of all chemical shifts can serve as cut-off value. This should be the most reliable method because it is based on independent replica datasets and not biased by strong CSP within the dataset of one titration as the Schumann approach. Moreover, it is independent from the assumption that the termini behave independently from the remainder of the protein.

2.3.1.2 Alternative methods to obtain structural information from chemical shift changes

It has been noted that spatial information of chemical shift changes can be interpreted to derive conclusion on ensemble changes or structural rearrangements upon binding of allosteric inhibitors or mutations of certain residues (Nohaile et al., 1997; Boehr et al., 2013; Chi et al., 2015). The Melacini group proposed the Chemical Shift Projection Analysis (CHESPA) approach to quantify similarity of changes and the extent of the changes induced by mutation or addition of inhibitors to a system (Selvaratnam et al., 2012). The authors analyzed whether a perturbation induced by means of mutations or allosteric inhibitors would cause chemical shift changes related to the original shift changes observed upon activation or binding of the natural ligand. Hence, two vectors originating from the inactive or apo state of the WT are defined: i) a reference vector **A** that connects inactive and active or apo and holo forms, respectively, and ii) a vector **B** that connects the resonances of the perturbed state to

the inactive or apo state of the WT. The cosine of the angle θ between these vectors as well as the relative shift projected on the reference vector are interpreted as measure of the conformational similarity and the extent of the induced changes (**Figure 2-4**) thereby allowing to discern local changes introduced by mutation or binding of the allosteric effector from global allosteric effects. The cosine of θ is derived from the scalar product of the two vectors:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{A}||\mathbf{B}|} \tag{2-39}$$

If the absolute value of the cosine approaches unity, the chemical shift changes induced by the induced perturbation are located on the same vector defined by the reference chemical shifts of ligand binding. The sign of this cosine can be interpreted in the sense whether the induced chemical change shifts the new conformation towards a more holo- or active-state-like conformation, thus having an activating effect, or if it shifts it further away towards a more deactivated state. The effect can be scaled by the projection *X* of vector **B** on **A**:

$$X = \frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{A}|^2} \tag{2-40}$$

If *X* approaches unity the conformational or dynamics changes induced by the mutation or the allosteric regulator are the same as those induced by binding.

Other approaches used principal component analyses to detect correlated chemical shift changes and map allosteric interaction networks (Akimoto et al., 2013; Boulton et al., 2014; Cembran et al., 2014).

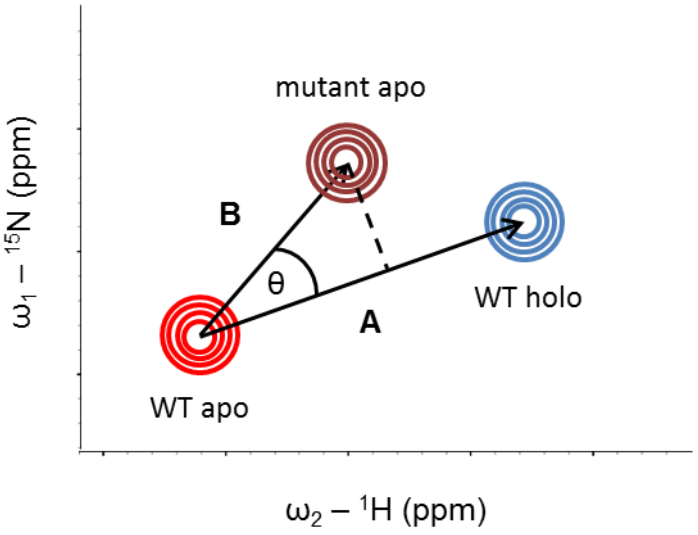


Figure 2-4: Principle of CHESPA analysis. The chemical shift vector **A** describes the change in chemical shift induced by ligand binding in the WT. The angle and the projection of a second vector **B** between the apo state of the WT and the apo state of a mutant allows for conclusions on the similarity of the induced changes by ligand

binding and mutation. If |cos(θ)|≈1 it can be assumed that the changes induced by mutation shift the conformational ensemble populations between the bound and unbound state with the projection of **B** on **A** as the measure of similarity. Figure adopted from (Selvaratnam et al., 2012).

2.3.2 Protein backbone resonance assignment

To obtain structural information from NMR experiments, the backbone resonances of the protein of interest have to be assigned. Today, a set of complementary three-dimensional experiments is available that allows for a sequential assignment of the backbone resonances: HNCA, HNCO, HN(CA)CO, CBCANNH and CBCA(CO)NNH. They require uniformly ^{15}N and ^{13}C isotope labeled protein. Moreover, they are inherently insensitive compared to the HSQC experiment, thus requiring long measurement times.

All these experiments generate coherence on protons (either $^{1}H^{N}_{i}$ or aliphatic $^{1}H^{\square}_{i}$ and the $^{1}H^{\square}_{i}$), transfer the coherence using INEPT transfers, followed by an evolution of the first hetero atom chemical shift, while decoupling the other channels, during t_1. An additional INEPT step transfers the coherence to the second hetero nucleus where the chemical shift evolves for t_2 while again decoupling the other channels. One or two INEPT steps transfer the coherence to the $^{1}H^{N}_{i}$ where it is detected during t_3 while decoupling the hetero channels. The coherence transfer direction is controlled by the INEPT delays □ corresponding to the inverse of the ^{1}J scalar coupling. By incrementing the evolution times t_1 and t_2 chemical shift information of the hetero nuclei chemical shifts is included in the proton FID allowing for the construction of a three-dimensional spectrum.

The HNCA and HNCO experiments (Kay et al., 1990) are out-and-back experiments that correlate the chemical shifts within the spin system *i* of the backbone atoms $^{1}H^{N}_{i}$, $^{15}N_{i}$, $^{13}C^{\square}_{i}$, $^{13}C^{\square}_{i-1}$, and $^{13}CO_{i-1}$. The HNCA transfers coherence from $^{1}H^{N}_{i}$ to the connected backbone $^{15}N_{i}$ and after chemical shift evolution during t_1 to both of the connected α-carbons $^{13}C^{\square}_{i}$ or $^{13}C^{\square}_{i-1}$ where the chemical shifts evolve during t_2. Coherence is transferred back via two INEPT steps and detected on $^{1}H^{N}_{i}$ during t_3. The HNCO transfers coherence only to the sequentially previous carbonyl $^{13}CO_{i-1}$. The HNCA can be modified as a constant time experiment (CT-HNCA) introducing a constant-time period where the ^{15}N chemical shift and ^{15}N-^{13}C scalar couplings evolve simultaneously, in contrast to the standard HNCA, allowing for sharper peaks in the ^{15}N dimension by reducing relaxation losses and enhancing sensitivity (Farmer II et al., 1992).

The CBCA(CO)NNH (Grzesiek and Bax, 1992a) and CBCANNH (Grzesiek and Bax, 1992b) experiments excite the $^{1}H^{\beta}_{i}$ and the $^{1}H^{\alpha}_{i}$ protons and transfer coherence via the $^{13}C^{\beta}_{i}$ and $^{13}C^{\alpha}_{i}$, where the carbon chemical shift evolves during t_1, and then either via the carbonyl to the

sequentially following backbone amide $^{15}N_{i+1}$ or to the directly connected $^{15}N_i$, where the nitrogen chemical shift evolves during t_2. Finally, coherence is transferred to the amide protons $^{1}H^{N}_{i+1}$ or $^{1}H^{N}_{i}$ where it is then acquired during t_3. The two experiments can also be carried out as out-and-back experiments with excitation and detection of the $^{1}H^{N}$ nucleus with the advantage of higher sensitivity due to shorter transverse magnetization times on the ^{13}C nuclei (Wittekind and Mueller, 1993). In addition, a CC(CO)NNH can be recorded to obtain $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C^{\gamma}$ and side chain resonances of linear or branched amino acids (Grzesiek et al., 1993). Moreover, the experiment also yields $^{13}C_{\delta}$ resonances of closed-ring proline residues. Coherence is generated on the side chain $^{1}H^{\alpha}_{i-1}$, $^{1}H^{\beta}_{i-1}$ and $^{1}H^{\gamma}_{i-1}$, transferred to the connected ^{13}C nuclei where it evolves and is transferred by isotropic ^{13}C mixing between the carbon nuclei. Then, coherence is transferred via the carbonyl to the $^{15}N_i$ and the attached $^{1}H^{N}_{i}$ with evolution times on both nuclei resulting in a three-dimensional spectrum.

All the experiments described above, can be run as BEST-type experiments allowing for shorter recycle delays by applying selective shaped pulses on the band regions of interest ($^{1}H^{N}$ region for HNCA or HNCO experiments or the $^{1}H^{\alpha}$ and the $^{1}H^{\beta}$ region for CBCA(CO)NNH and CBCANNH experiments, respectively) and then allowing faster R_1 relaxation via dipolar interactions with the unperturbed aliphatic protons of the protein (Schanda et al., 2006; Lescop et al., 2007).

Once the data has been collected, the experiments provide complementary information on the $^{13}C^{\beta}_{i}$ and $^{13}C^{\alpha}_{i}$ chemical shifts connected with the correlated ($^{1}H^{N}_{i}$; $^{15}N_i$ and ($^{1}H^{N}_{i+1}$; $^{15}N_{i+1}$) chemical shifts of the root ^{1}H-^{15}N HSQC NMR spectrum allowing for a sequential walk through the backbone's spin systems i.e. the amino acids. The higher sensitivity of the HNCA and improved ^{15}N-chemical shift resolution of the CT-HNCA help to resolve overlapping peaks and the HNCO provides the $^{13}CO_i$ chemical shifts. Together, the hetero nuclei resonances assigned to one spin system *i* can be used to predict the amino acid type based on empirical chemical shifts taken from all deposited assignments in the BRMB which enables the assignment of connected spin systems to the amino acid primary sequence.

The ^{1}H,^{15}N-NOESY-HSQC experiment (Marion et al., 1989a; Marion et al., 1989b; Zuiderweg and Fesik, 1989) can guide the assignment of poorly resolved regions in the triple resonance datasets or verify the assignment by identifying cross-peaks of sequential amide protons or those spatially close in the tertiary structure. This is especially helpful if a crystal structure of the protein is available. The experiment setup is the straightforward combination of a homonuclear ^{1}H,^{1}H-NOESY during t_1 while decoupling the hetero channel followed by ^{1}H,^{15}N-HSQC pulse sequence evolving the ^{15}N chemical shift in t_2 and detection in the proton

channel in t_3 while decoupling the hetero channel. Today, the software packages SPARKY (Goddard and Kneller) and especially CCPN Analysis (Vranken et al., 2005) provide the means to visualize and connect the spectra allowing for a straightforward assignment. CCPN Analysis also provides useful empirical data from the BMRB to predict single amino acid types and secondary structure elements from chemical shifts.

2.3.3 NMR Relaxation experiments

Proteins undergo motions on a variety of timescales (Karplus and Mccammon, 1986). Protein backbone ^{15}N-relaxation rate constants provide a wealth of insight into internal dynamics on the ps-ns timescale. Today, the standard experiment to measure the relaxation rate constants R_1 and R_2 in the protein backbone are HSQC-type experiments involving a relaxation period T before t_1 (Farrow et al., 1994). To measure R_1 relaxation, the ^{1}H channel is decoupled while the ^{15}N spins are allowed to relax unperturbed during T. To observe solely the effect of R_2 relaxation, a CPMG pulse train (Carr and Purcell, 1954; Meiboom and Gill, 1958) is applied on the ^{15}N channel during T, refocusing the chemical shift while allowing transverse relaxation to occur. Incrementing T gives rise to a series of spectra with varying peak intensities. These can be fitted to a mono-exponential decay function thus obtaining the relaxation rate constants:

$$I(\mathrm{T}) = I_0 e^{-R_x \mathrm{T}} \tag{2-41}$$

where I(T) is the peak intensity depending on the relaxation period T, I_0 the peak intensity at T=0 and R_x the corresponding rate constant.

The hetero steady-state NOE is measured by recording on spectrum with saturation of the ^{1}H magnetization and a reference spectrum without saturation. The NOE is than given as:

$$NOE = \frac{I_{sat}}{I_{ref}} \tag{2-42}$$

with I_{sat} and I_{ref} as the peak intensities in the saturated and reference spectrum, respectively.

2.3.3.1 Estimation of the overall correlation time from R_1 and R_2 relaxation rate constants

For small proteins, the overall correlation time of a protein τ_c can be estimated empirically from the ratio of the transverse and longitudinal relaxation rate constants R_2 and R_1 (Kay et al., 1989). However, local internal motions highly affect the transverse relaxation rate constant either decreasing it if the motions are fast (pico- to nanosecond timescale) or increasing it if chemical exchange on the chemical shift timescale occurs. To account for

these motions, a robust estimate of the global correlation time is the average R_2/R_1 ratio determined for each residue trimmed by the standard deviation (Clore et al., 1990a):

$$\tau_c = \left\langle \frac{R_2}{R_1} \right\rangle - \sigma_{\langle R_2/R_1 \rangle} \tag{2-43}$$

2.3.3.2 Reduced spectral density mapping and consistency testing

To compare relaxation data collected at different fields or different machines, special care must be taken for temperature control and buffer conditions. It is good practice to calibrate the temperature using a 1% methanol sample in D_2O and measure the highly temperature sensitive difference in chemical shift between the methyl peak of the methanol and the HDO peak. Moreover, the quality of a data set consisting of relaxation rate constants and the NOE collected at different fields can be tested for consistency. One way is to test R_2 consistency by back-calculating the reduced spectral density function at zero field $J(0)$ for each residue (Farrow et al., 1995):

$$J(0) = \frac{-1.5}{3d + c}\left(\frac{R_1}{2} - R_2 + 0.6\left[(NOE - 1) R_1 \frac{\gamma_N}{\gamma_H} \right] \right) \tag{2-44}$$

with the dipolar constant and the CSA component defined above. If the data obtained at two fields are consistent, they correlate linearly. The function is field independent if µs-ms motions are absent. However, if slow motions on the µs-ms timescale are present and data is collected at two fields, the exchange rate R_{ex} can be readily quantified using the relation:

$$J(0)_{eff} = J(0) + \lambda R_{ex} \tag{2-45}$$

where $J(0)_{eff}$ is the calculated spectral density at the respective field, $J(0)$ the spectral density without chemical exchange and a scaling factor $\lambda = (3/2)[1/(3d + c)]$.

Two other functions, F_{R2} and F_η, have been proposed probing R_2 consistency which follow the same principle as the afore mentioned (Fushman et al., 1998).

2.3.3.3 Relaxation dispersion

Another way to interrogate slow motional processes on a micro- to millisecond timescale, CPMG relaxation dispersion measurements have become an invaluable tool (Loria et al., 1999; Mulder et al., 2001; Tollinger et al., 2001). Although model-free analysis of the relaxation rate constants can provide insight on slower motional processes, the approach is highly susceptible to anisotropic molecular diffusion and fast internal motions (see above). However, the CPMG relaxation dispersion experiment readily detects and quantifies

conformational exchange independent from these phenomena. The chemical exchange is measured as a function of the effective CPMG field strength that is varied in a series of constant time (CT) CPMG experiments. For slow exchanging processes, the observed transverse relaxation rate $R_{2,\mathrm{eff}}$ is the sum of the intrinsic transverse relaxation rate R_2^0, which is only attributed to dipolar coupling and CSA, and the exchange rate R_{ex} that varies with the CPMG field:

$$R_{2,eff} = R_2^0 + R_{ex}(\upsilon_{CPMG}) \qquad (2\text{-}46)$$

By either solving the Bloch-McConnell equation (Mcconnell, 1958) numerically or using its analytical solutions for different limiting cases (Luz and Meiboom, 1963; Carver and Richards, 1972), the relaxation dispersion data potentially provide the exchange rate k_{ex}, the fractional populations of the interchanging states, and thereby the K_d and free energy difference ΔG, the chemical shift difference of the states, and the respective R_2^0 values.

2.3.3.4 Model-free analysis implementation

Variations of R_2 and NOE along the sequence are already good indicators for local changes in internal dynamics. However, to obtain a more pronounced understanding, model-free analysis should be undertaken. The dissection of internal and overall motion remains far from trivial and several complex models were developed to account for the optimization of the model-free parameters and to estimate the diffusion tensor governing global correlation time τ_m (Mandel et al., 1995; Fushman et al., 1997; Orekhov et al., 1999; Korzhnev et al., 2001; Zhuravleva et al., 2004). To find the universal solution for such a problem, several models have to be optimized and then a model selection must be undertaken. Classic approaches follow the diffusion seeded model-free paradigm optimizing first the diffusion tensor before finding solutions for the model-free parameters. Herein, I follow the analysis protocol developed by d'Auvergne and Gooley (d'Auvergne and Gooley, 2008) that first minimizes nine model-free models to obtain the model-free parameters while generating local τ_m values and then performs model selection. Only afterwards, the diffusion tensor is optimized for a sphere, two spheroids and an ellipsoid model. Finally, model selection is obtained to find the universal solution. For a more detailed description of the approach, see below. The reader is also referred to the extensive literature on the d'Auvergne protocol (d'Auvergne and Gooley, 2003, 2006, 2007, 2008).

2.3.4 Ligand-observed NMR methods

In the case of large proteins or protein complexes interacting with small molecules, ligand-observed NMR is a fast alternative providing both affinity data and structural information of the binding site. Saturation transfer difference (STD) NMR (Mayer and Meyer, 1999) is a cross relaxation-based experiment that allows for the identification of small ligands of high MW proteins. After selective saturation of the protein resonances with a series of selective pulses, magnetization spreads within the protein via spin diffusion i.e. internal cross relaxation σ_{intra} and is then transferred to bound ligands by an intermolecular cross-relaxation σ_{inter}. Thereby, the ligand is progressively saturated. After the saturation period, a ^{1}H 90° pulse generates ^{1}H coherence on all non-saturated spins that is then detected. If the off-rate of the ligand is fast and therefore a substantial fraction of the ligand population visits the protein during the time of saturation, the difference spectra of off- and on-resonance saturation experiments will show increased peak intensities for ligand protons that are spatially close to the protein in the binding pocket. The STD effect can be quantified to obtain ligand epitope information. To account for R_1 occurring during the saturation transfer period, a series of STD NMR spectra with different saturation transfer times is recorded. Plotting the STD intensities against the saturation time will result in a STD built-up curve described by:

$$I_{STD} = I_{STD,\max}(1 - e^{-k_{sat}\tau_{sat}}) \qquad (2\text{-}47)$$

where I_{STD} is the STD intensity, $I_{STD,max}$ the maximal STD intensity, k_{sat} the saturation transfer rate, and τ_{sat} the saturation transfer time. The initial slope of the curve is given by the product of $I_{STD,max}$ and k_{sat} and corresponds to the STD intensity in absence of R_1 relaxation (Mayer and James, 2004). Especially in the context of screening small molecule fragments for drug development, STD NMR has become an invaluable tool (Meyer and Peters, 2003).

3 Materials and Methods

3.1 Materials

All standard chemicals and buffers used within these work were purchased from Sigma Aldrich (St. Louis, MO, USA) or Carl Roth (Karlsruhe, Germany) if not indicated otherwise. For all aqueous solutions ultrapure water with 18.1 MΩ resistance prepared with an Integra UV Plus (Siemens, Munich, Germany) was used. For DNA plasmid purification the GeneJet Plasmid Miniprep kit (ThermoFisher Scientific, Waltham, MA, USA) was used. All standard PCR reactions were performed using a Phusion High Fidelity DNA polymerase (ThermoFisher Scientific) according to the manufacturer's instructions. All DNA oligonucleotide primers were synthesized by Invitrogen (ThermoFisher Scientific). Sequence identities were verified by Sanger sequencing performed by either GATC Biotech (Konstanz, Germany) or SeqLab Sequence Laboratories (Göttingen, Germany). All data analysis, plotting and curve fitting was performed with OriginPro 2015 (OriginLab, Northampton, MA) if not indicated otherwise.

3.2 Protein design, expression, and purification

After initial tries to reproduce Langerin CRD expression according to the literature awarded only poor yields (Chatwell et al., 2008), full length human Langerin ECD (residues 67-328) (Stambach and Taylor, 2003) was codon optimized for expression in *E. coli* and synthesized by GenScript (Piscataway, NJ, USA) (**s. Appendix**). The protein sequence was C-terminally fused with a TEV cleavage site and a Strep-tag II for purification (Schlapschy et al., 2006) and provided in a pUC57 vector. Insoluble expression, refolding from inclusion bodies and purification of the full length Langerin ECD construct performed according to a published protocol (Stambach and Taylor, 2003) from the pUC57 vector resulted only in poor yields. Therefore, full length Langerin ECD, a truncated version of Langerin ECD with improved solubility (residues 148-328) (Feinberg et al., 2010), and Langerin CRD (residues 193-328) were cloned into a pET30a expression vector (Merck Millipore, Darmstadt, Germany) resulting in the vectors pET30a:*flhLangerin ECD:ST2*, pET30a:*trhLangerin ECD:ST2*, and pET30a:*hLangerin CRD:ST2*. The PCR primers are listed in **Table 3-1**. Identity was confirmed by sequencing using the T7 promotor as starting point. However, only poor yields of less than 2 mg L^{-1} culture were achieved following the literature. Therefore, a different refolding method was established by conducting a commercially available screen (PROTEOSTAT©, Enzo Life Science, Farmingdale, NY, USA) to optimize refolding conditions for rapid dilution following the manufacturer's protocol.

After several optimization steps, protocols for both truncated Langerin ECD and Langerin CRD with protein yields of up to 15 mg L^{-1} culture medium were developed. Langerin ECD was expressed insolubly in *E. coli* BL21* (ThermoFisher Scienific) in LB medium at 37 °C. Protein production was induced by adding 0.5-1 mM IPTG at $OD_{600\ nm}$ of 0.5-0.8 and cells were harvested by centrifugation at 4,000*g* for 30 min. Pellets were lysed by incubation with 1 mg mL^{-1} lysozyme (Sigma Aldrich) and 100 µg mL^{-1} DNase I (Applichem, Darmstadt, Germany) in a detergent containing lysis buffer (50 mM Tris, 150 mM NaCl, 10 mM $MgCl_2$, 0.1% Triton-X, pH 8) for at least 3 h at RT. Inclusion bodies were washed once with lysis buffer and three time with 20 mM Tris, pH 8 with centrifugation steps of 10 min at 10,000*g*. IB pellets were solubilized in 6 M guanidinium hydrochloride in 100 mM Tris, pH 8, and 1 mM DTT at 37 °C for at least 2 h by adding 40 mL L^{-1} culture. Langerin CRD was produced as described above. To achieve uniform ^{15}N and ^{13}C isotope labeling, cells were grown in M9 minimal medium (**Table 3-2**) supplemented with [^{15}N, 99%] NH_4Cl for uniform ^{15}N- or with both [^{15}N, 99%] NH_4Cl and [$^{13}C_6$, 99%] D-Glucose (both SigmaAldrich). For solubilization, 0.5 mM TCEP were used instead of DTT.

Solubilized IBs were centrifuged at 15,000*g* for 90 min to remove insoluble cell debris. Langerin ECD was refolded by adding it drop wise into 0.4 M L-arginine in 50 mM Tris, 20 mM NaCl,, 0.8 mM KCl, pH 7.5 in a ratio of 1:10 while rapidly stirring. Reduced and oxidized glutathione was used in concentrations of 1 mM and 0.2 mM, respectively, to allow for cysteine redox shuffling. Yields were highly improved when performing all steps starting from refolding at 4 °C. After refolding, refolded Langerin ECD was dialyzed against the mannan chromatography buffer (50 mM Tris, 150 mM NaCl, 20 mM $CaCl_2$, pH 7.5). Precipitates were removed by centrifugation at 15,000*g*, for 90 min. Protein purification was achieved by affinity chromatography using mannan coupled sepharose beads (Sigma Aldrich) as resin. After application, the resin was washed with at least 5 column volumes chromatography buffer and protein was subsequent eluted with 50 mM Tris, 150 mM NaCl, 5 mM EDTA (pH 7.5). Elution fractions were tested for protein by Bradford (Applichem) and protein-containing fractions were pooled. Pooled protein solution was concentrated to about 2 mL using centrifugal filters (10,000 MWCO, Corning, Corning, NY, USA). Buffer was exchanged against TBS or HBS with centrifugal desalting columns (ThermoFisher Scientific).

Langerin CRD was refolded as described for Langerin ECD using 0.8 M L-arginine as the only parameter that differed, dialyzed against 50 mM Tris, 150 mM NaCl, 1 mM EDTA, pH 8, centrifuged, and purified using a StrepTactin column (IBA, Göttingen, Germany). Loaded resin was washed with at least 5 CV chromatography buffer and protein was eluted

with 2.5 mM d-desthiobiotin in chromatography buffer. EDTA was later excluded from the elution buffer because it interfered in subsequent experiments. Protein-containing fractions were pooled and dialyzed against 25 mM MES, 40 mM NaCl, pH 6 to remove soluble protein aggregates by precipitation. Insoluble aggregates were removed by centrifugation at 15,000*g* for 90 min. The protein solution was concentrated to target concentration by centrifugal filters (10,000 MWCO, Corning). Final protein concentrations were determined by UV absorption at 280 nm using the calculated extinction coefficients of 56170 M^{-1} cm^{-1} for both Langerin constructs (Gasteiger et al., 2005) and the corresponding molecular weights of 18.0 kD for Langerin CRD and 22.9 kD for truncated Langerin ECD. Proteins were stable in concentrations of about 200°µM at 4 °C for about 14-21 days. For long-term storage, 100 µM protein solutions were shock-frozen in 50 µL aliquots liquid N_2 and kept at -80 °C. However, thawed protein showed reduced activity and increased tendency to aggregate. Therefore, all experiments were carried out with a freshly prepared protein batch. Murine truncated Langerin ECD was generously provided by Jonas Aretz.

All Langerin CRD mutants were generated by introducing single-codon mutations in the pET30a:*Langerin CRD:ST2* vector using the QuikChange Lightning kit (Agilent, Santa Clara, CA, USA) according to the manufacturer's protocol. The primers are listed in **Table 3-1**. Mutant protein constructs were expressed and purified as described for the WT above. Purity was assessed by SDS-PAGE analysis and proper folding by ^{1}H-^{15}N HSQC NMR.

Table 3-1: Primers for cloning of Langerin constructs into pET30a expression vector

Construct	**Primer**
hLangerin CRD	forward: 5' GGTGGTCATATGGCCCAGGTGGTTAGCCAAGGCTGGAAATAC 3'
	reverse: 5' ACCACCAAGCTTTTATTTTTCAAACTGCGGATG 3'
hLangerin ECD	forward: 5' GGTGGTCATATGGCCTCGACGCTGAATGCCCAGATTCCGG 3'
(truncated)	reverse: 5' ACCACCAAGCTTTTATTTTTCAAACTGCGGATG 3'
hLangerin ECD	forward: 5' GGTACCATATGGCACGCTTCATGGGC 3'
(full length)	reverse: 5' ACCACCAAGCTTTTATTTTTCAAACTGCGGATG 3'

Table 3-2: Composition of M9 minimal medium used for expression of isotope labeled protein. Asterix indicates components that are used with 99.9% ^{15}N or ^{13}C isotope labeling, respectively.

Component	**Concentration**
Na_2HPO_4	33.7 mM
KH_2PO_4	22.0 mM
NaCl	8.55 mM
NH_4Cl*	9.35 mM
Glucose*	2.2 mM
$MgSO_4$	1 mM
$CaCl_2$	0.3 mM
$FeCl_3$	31 µM
$ZnCl_2$	6.20 µM
$CuCl_2$	760 nM
$CoCl_2$	420 nM
H_3BO_3	1.62 µM
$MnCl_2$	81 nM
Biotin	1 µg L^{-1}
Thiamin	1 µg L^{-1}
Riboflavin	0.5 µg L^{-1}
Niacinamide	0.5 µg L^{-1}
Pyridoxine monohydrate	0.5 µg L^{-1}

Table 3-3: List of primers utilized for single-residue mutants of human Langerin CRD.

Mutant	Primer
N205A	Forward: 5'-atcagactgaagtagtagaaggcacccttgaagtatttccagcc-3'
	Reverse: 5'-ggctggaaatacttcaagggtgccttctactacttcagtctgat-3'
E220A	Forward: 5'-ctcacgcaaaactgtgcggcggagtaccagg
	Reverse: 5'- cctggtactccgccgcacagttttgcgtgag-3'
L230A	Forward: 5'-gtaacgctggtcgcatgggagttacgactcacgca-3'
	Reverse: 5'- tgcgtgagtcgtaactcccatgcgaccagcgttac-3'
Q239A	Forward: 5'- ggttttatacagaaattctgcttcagattcgctcgtaacgctggt-3'
	Reverse: 5'- accagcgttacgagcgaatctgaagcagaatttctgtataaaacc-3'
K257A	Forward: 5'-cctccatgcccgctgccgtcagaccgatcc-3'
	Reverse: 5'-ggatcggtctgacggcagcgggcatggagg-3'
G259I	Forward: 5'- ccaatcaccctccatgatcgctttcgtcagaccg-3'
	Reverse: 5'- cggtctgacgaaagcgatcatggagggtgattgg-3'
D263A	Forward: 5'-caacccatgaccaagcaccctccatgccc-3'
	Reverse: 5'- gggcatggagggtgcttggtcatgggttg-3'
S265A	Forward: 5'- catggagggtgattgggcatgggttgatgacac-3'
	Reverse: 5'- gtgtcatcaacccatgcccaatcaccctccatg-3'
P286A	Forward: 5'- cggcattgttcgcttcgcccggaatcca-3'
	Reverse: 5'- tggattccgggcgaagcgaacaatgccg-3'
H294A	Forward: 5'-gatgttgccacaggcttcattgttaccggcattgttcgg-3'
	Reverse: 5'-ccgaacaatgccggtaacaatgaagcctgtggcaacatc-3'
N297A	Forward: 5'- caggctcggtgctttgatggcgccacagtgttcattgtta-3'
	Reverse: 5'- taacaatgaacactgtggcgccatcaaagcaccgagcctg-3'
K299A	Forward: 5'- ctgcaggctcggtgctgcgatgttgccacagtgt-3'
	Reverse: 5'- acactgtggcaacatcgcagcaccgagcctgcag-3'
K320A	Forward: 5'- gaacgtacgggcgtgcacaaatgaacagaaacgttttgtcgca-3'
	Reverse: 5'- tgcgacaaaacgtttctgttcatttgtgcacgcccgtacgttc-3'

3.3 Quality control of protein constructs

Expression constructs were characterized by SDS-PAGE, MALDI-TOF MS, CD spectroscopy. In addition, Langerin CRD trimerization was tested by chemical crosslinking and activity was tested by a plate-based binding assay (see below). Langerin CRD quality was accessed by ^{15}N-HSQC NMR (see below). SDS-PAGE was conducted using a Tris-Tricine buffer system according to (Schagger, 2006). Protein loading was adjusted to about 1 µg and 1 mm 12% PA-gels were run at 160 V for 45 min after a 10 min stacking phase at 50 V. Staining was performed with Coomassie Brilliant Blue R250 (Carl Roth) (0.1% in 40% methanol, 10% acetic acid). Chemical crosslinking of Langerin ECD was performed by addition of 0.1 to 5 mM of the amine reactive BS3 (Sigma Aldrich) (1 mg L^{-1}, freshly dissolved in DMSO) to 100 µL of protein solution (5 mg mL^{-1}) in HBS and subsequent incubation for 1a JASCO J-810 CD spectrometer with scan region from 240 to 190 nm, 0.5

nm spacing and averaged over 3 scans before buffer baseline subtraction. MALDI-TOF MS was performed by Ms. Eva Settels on a Bruker Autoflex Speed using a DHAP matrix according to standard protocols.

Single-residue mutants of Langerin were analyzed by SDS-PAGE after purification and subsequent ^{1}H-^{15}N HSQC NMR.

3.4 Fluorescence protein labeling

Both the murine and the human Langerin ECD constructs were fluorescently labeled with fluorescein isothiocyanate (ThermoFisher Scientific) by adding slowly 100 µL of the dye (1 mg mL^{-1} in DMSO) to 2 mL of protein (2 mg mL^{-1}) in HEPES-buffered saline (HBS), pH 7.2 containing 20 mM D-mannose and 5 mM $CaCl_2$ and stirred at room temperature for 90 min in the dark. Reaction was quenched by adding 50 mM ethanolamine (1 M, pH 8.5). The superfluous dye was removed by buffer exchange using a Zeba spin column. The labeled proteins were purified over a mannan affinity column and active protein was pooled and shock frozen in liquid nitrogen. Labeling efficiencies were assessed by absorption spectroscopy and amounted to 0.9 and 0.6 for the human and the murine Langerin construct, respectively.

3.5 Polysaccharides

Mannan, laminarin, and curdlan were obtained from Sigma-Aldrich. Bacterial glycans were provided generously by Yuriy Knirel (N.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Moscow, Russia). The dry powder glycans were dissolved in D_2O to a final concentration of 1 mg mL^{-1}.

3.6 Plate-based assays

NUNC maxisorb© 96-well plates (ThermoFisher Scientific) were coated overnight with 1 µg mL^{-1} to 100 µg mL^{-1} of the corresponding polysaccharide in 50 mM carbonate buffer, pH 9.6, and blocked with 2% BSA in TBS-T for 1 h. Wells were washed three times with TBS-T + 5 mM $CaCl_2$, incubated with protein in the same buffer for at least 90 min at RT, washed again, and incubated with StrepTactin-HRP conjugate (IBA, Göttingen, Germany) in 2% BSA in TBS-T + 5 mM $CaCl_2$ for 1 h. Plates were developed by adding TMB solution (tebu-bio, Versailles, France). The reaction was quenched by addition of 0.18 M H_2SO_4. Absorbance was measured at 450 nm on a SpectraMax M5 plate reader (Molecular Devices, Sunnyvale, CA). Protein concentration was kept constant at 50 nM if not titrated for EC_{50}

determination. For inhibition assays, inhibitor was added prior to incubation on the plate. Data was fitted to a dose-response model according to the equation:

$$A_{450nm} = \frac{A_{450nm,\,max}}{1 + \left(\frac{IC_{50}}{[I]} \right)^{p}} \qquad (3\text{-}01)$$

with A_{450nm} as the absorption at 450 nm corresponding to the readout of the assay, [I] as the inhibitor concentration and p the Hill coefficient. The same equation was utilized to obtain the effective concentration of binding by substitution of the inhibitor concentration with the protein concentration.

3.7 Glycan array screening

Fluorescently-labeled murine and human Langerin ECD constructs were screened against the previously reported microbial glycan array at the Scripps Research Institute (La Jolla, CA, USA) (Stowell et al., 2014). The measurements were performed by Ryan McBride according to standard routines described previously (Ibid.).

3.8 Bacterial binding assay

Direct binding of fluorescently-labeled murine and human Langerin ECD constructs to heat-killed E. coli O106 (generously provided by Dr. Wolfgang Rabsch, Robert Koch Institute, Wernigerode Branch, National Reference Centre for Salmonellae and other Bacterial Enteric Pathogens, Wernigerode, Germany) was observed in a flow-cytometry assay using an Attune NxT flow cytometer (ThermoFisher). Bacterial suspensions (≈$6x10^8$ cells mL^{-1}) were incubated with the fluorescently-labeled protein at concentration ranging from 18.75 nM to 600 nM in HBSS containing 2 mM $CaCl_2$ for 30 min at RT. The suspension was washed three times. The measurements were conducted at a flow rate of 100 µL min^{-1} and a laser powers of 420 (FSC), 400 (SSC), and 260 (BL1-A). Gating was performed based on the forward and side scatter of the bacteria. For the inhibition assay, the competitors EDTA (20 mM), laminarin, curdlan, and mannan (all 0.05 mg mL^{-1}) were incubated together with the protein and the bacteria.

3.9 ITC measurements

All isothermal titration calorimetry experiments were performed using a MicroCal iTC200 (Malvern Instruments, Malvern, UK). The titrant, either $CaCl_2$ (5 mM) or mannose (50 mM), was added in defined steps of 1-2.5 µL to the protein solution (150 µM to 250 µM) at 298 K while stirring at 750 rpm. The differential heat of each injection was measured and plotted

against the molar ratio. The data was fitted to a one-set of sites binding model assuming a Hill coefficient of 1 (Freyer and Lewis, 2008). Due to the low c-values of the measurements ($c < 5$), the enthalpy could not be determined reliably (Pierce et al., 1999).

3.10 NMR measurements

All NMR assignment and relaxation measurements were performed on a 600 MHz Bruker Avance III spectrometer (Bruker, Billerica, MA, USA) equipped with a triple resonance cryogenic probe. Relaxation data was additionally collected on a 750 MHz Bruker Avance II spectrometer equipped with a triple resonance cryogenic probe. ^{15}N-HSQC measurements were either performed on the Bruker 600 MHz (see above) or an Agilent 600 MHz vnmrs machine equipped with a room-temperature double resonance onenmr probe.

For Langerin CRD backbone resonance assignment, a series of standard triple resonance experiments was collected (**Table 3-4**)using 350 µM (U:^{15}N;^{13}C) protein in 25 mM MES, 40 mM NaCl, 5 mM $CaCl_2$, pH 6 containing 100 µM DSS, 0.05% NaN_3, and 10% D_2O.

^{1}H-^{15}N HSQC NMR titrations experiments were performed with 120-250 µM of (U:^{15}N) Langerin CRD in 25 mM MES, 40 mM NaCl, pH 6 containing 100 µM DSS, and 10% D_2O. For titration of Ca^{2+}, $CaCl_2$ was added stepwise to the sample (600 µL initial volume) from a buffer-matched stock (0.1 M or 1 M depending on target concentration) until 10 mM final concentration as endpoint was reached. For WT and mutant Langerin CRD at pH 6, a five-step protocol of 0, 100, 500, 1500, and 10000 µM $CaCl_2$ was established as standard. Depending on protein concentration and implemented machine, 128 increments and 4-16 transients were recorded at 298 K. For measurements at pH 7, the buffer system was changed to 25 mM HEPES, 150 mM NaCl, pH 7. Ca^{2+} titrations were performed accordingly. Carbohydrate titrations were performed analogously to Ca^{2+} titration with 10 mM $CaCl_2$ in the buffer. Carbohydrates were either dissolved directly in the sample or diluted from high-molarity stocks into the sample depending on the respective solubility. Ligand concentrations varied due to solubility restrictions. Mannose and glucose were added up to 60 mM and 62.5 mM, respectively. Final concentration of laminaritriose was 21.6 mM, of laminarihexaose 5 mM, and of mannopentaose 12 mM. Mannan was added to 100 µg mL^{-1} final concentration. Measurements were performed as described for Ca^{2+} titrations.

^{15}N-relaxation experiments were performed with (U:^{15}N)-labeled apo Langerin CRD (200 µM) in 25 mM MES, 40 mM NaCl, pH 6 containing 100 µM DSS and 10% D_2O for apo Langerin CRD. Measurement of the holo form was performed on the same sample after addition of 10 mM $CaCl_2$. The temperature was calibrated by measuring a 1% CD_3OH in D_2O

prior to the relaxation measurements. The chemical shift difference between methanol and HDO signal is highly temperature sensitive and can be used to calibrate the sample's absolute temperature according to following empirical equation (Findeisen et al., 2007):

$$T = -23.832\Delta\delta^2 - 29.46\Delta\delta + 403 \quad (3\text{-}1)$$

The measurements were conducted on both field strengths at 299.2 K.

All STD NMR experiments were performed on the Agilent 600 MHz spectrometer. Samples were prepared with 50 µM protein concentration, 1 mM α-methyl (1-2α) mannobioside in 25 mM d_{11}-Tris (Eurisotop, Saint-Aubin Cedex, France), 150 mM NaCl, pH 7.0 in D_2O. Spectra were recorded with an STD pulse sequence containing a WATERGATE solvent suppression. Selective gauss-shaped saturation pulses with an -0.5 ppm offset for on-resonance and -80 ppm for off-resonance spectra were applied with saturation transfer times of 0.5, 1, 2, and 4 s with a total of 1024 scans (2048 at 0.5 s saturation transfer time). Difference spectra were processed and analyzed with MestreNova (MestreLab, Santiago de Compostela, Spain) and the epitope was mapped as described above.

Table 3-4: List of ^{1}H-detected NMR experiments, pulse sequences, and parameters for backbone resonance assignment of Langerin CRD. Number of transients (nt), and complex points in each channel are given corresponding to ^{1}H (TD_3), ^{13}C (TD_2), and ^{15}N (TD_1), respectively.

Sample	Experiment	Pulse sequence	*nt*	TD_3	TD_2	TD_1
Langerin CRD WT (U:^{15}N;^{13}C, 350 µM)	HNCA	MFhnca_best	16	1024	96	128
	CT-HNCA	MFhnca_best	16	1024	128	270
	HNCO	MFhnco_best	4	1024	116	128
	HN(CA)CO	MFhncaco_best	24	1024	96	128
	HNCACB	MFhncacb_best	32	1024	116	110
	HN(CO)CACB	MFhncocacb_best	32	1024	116	110
	CC(CO)NH	MFccconnhwg	8	1024	128	128
	NOESY-HSQC	MFnoehsqcwtgf3	8	1024	128	128
	HNHA	MFhnhawg	8	1024	128	128
	NCO	MFnco	128	-	1024	128

Table 3-5: List of ^{1}H,^{15}N-HSQC NMR experiments, pulse sequences and parameters. Number of transients (nt), and complex points in each channel are given corresponding to ^{1}H (TD_2) and ^{15}N (TD_1), respectively.

Sample	Experiment	Pulse sequence	*nt*	TD_2	TD_1
Langerin CRD WT and mutants (U:^{15}N; 100-250 µM)	^{1}H,^{15}N-HSQC	HSQC (WET)$^{+}$	8-12	1404	96-128
		gNfhsqc (WATERGATE)$^{+}$	4-8	1404	128
		MFhsqcwtgf3*	4	1024	128

$^{+}$Performed on Agilent 600 MHz spectrometer with RT probe

*Performed on Bruker 600 MHz spectrometer with cryogenic probe.

Table 3-6: List of ^{15}N relaxation NMR experiments, pulse sequences, and parameters. Number of transients (nt), and complex points in each channel are given corresponding to ^{1}H (TD_2), ^{15}N (TD_1), and number of planes respectively.

Sample	Experiment	Pulse sequence	*nt*	TD_2	TD_1	n_{planes}
Langerin CRD WT (U:^{15}N; 300 µM)	T1	mb15nt1wtg_3d_11	24	1024	160	10
	T2	mb15nt2wtg_3d_6	56	1024	160	10
	hetero NOE	pF15nnoewtg_3d_6	96	1024	160	2

3.11 NMR data processing

All assignment spectra were processed in Bruker TopSpin 3.2 and assignment was performed in CCPN Analysis (Vranken et al., 2005) according to the above described strategies.

NMR titration data was processed with NMRpipe (Delaglio et al., 1995) applying Lorentz-to-Gauss transformation in f2 with 20 Hz line-broadening factor and sine bell functions in f1 dimension. 4x zero-filling and polynomial baseline corrections were applied in both dimensions. The spectra were visualized, referenced and analyzed in CCPN. All spectra were referenced using the internal spectrometer reference and assignments were transferred from the reference spectrum (root spectrum from the resonance backbone assignment) to the nearest neighbor. In case of ambiguities caused by strongly overlapping or disappearing peaks, the resonance assignments were not transferred. For titration experiments, the assignment transfer was facilitated by following shifting peaks in respect to the reference spectrum. Peak lists for each mutant and titration point were exported for CSP calculation according to eq. (2-33). Dissociation constants were obtained by global fit of peaks that appeared to be near fast exchange limit using equation (2-34).

CHESPA analysis was performed as described in the literature (Selvaratnam et al., 2012) defining the vector between apo and holo WT as reference and calculating the vector product and projection of the vector between apo mutant and apo WT. The same cut-off as defined for the CSP measurements was applied after conversion to vector length (0.025 ppm).

The relaxation data was processed in NMRPipe (Delaglio et al., 1995) and peaks were picked and assignment transferred in Sparky (Goddard and Kneller). To obtain R_1 and R_2 relaxation rate constants, peak heights were fitted against the relaxation period using a single exponential decay model (eq. 2-41) with a Newton minimization algorithm implemented in the relax analysis package (d'Auvergne and Gooley, 2008). Errors were estimated by using a Monte Carlo simulation with 500 iterations that tests back-calculated data from the model against the experimental values. The hNOE was determined by the difference in peak intensities of the saturated and reference spectra. The error was estimated based on the experimental noise given by the RMSD of spectral regions without signals.

The relaxation data obtained from both fields was tested for consistency by calculating the spectral densities at zero field $J(0)$ (see above). The values obtained from both datasets were plotted against each other and fitted to a linear equation with zero offset to check for linear correlation. Outliers and overlapping peaks were excluded from further analysis. Distributions

of the ratios of the spectral densities obtained on both fields were analyzed with the inbuilt statistics function of OriginPro 2015 assuming normal distribution of the data.

All figures displaying two-dimensional spectra were prepared and visualized with CCPN or SPARKY. All one-dimensional spectra were processed, analyzed and prepared for figure depiction in MestreNova.

3.12 Hydrodynamic Calculations

The rotational diffusion tensor for holo Langerin CRD (pdb entry: 3p5f) was calculated using HYDRONMR 7 (Garcia de la Torre et al., 2000) according to the authors' protocol.

3.13 Structure representations

All structure representations within this work were generated with the program PyMol 1.7.2.1 (Schrodinger, 2015). CSP and conservation scores were color-coded on the structure by substituting the B-factors in the pdb file.

3.14 Analysis of evolutionary conservation

All mammalian Langerin sequences deposited in UniProtKB were chosen for multiple sequence alignment using CLUSTAL O 1.2.1 (Sievers et al., 2011) and subsequent scoring of residue conservation (Valdar, 2002).

4 Results and Discussion

4.1 Protein production and backbone resonance assignment of human Langerin

4.1.1 Protein expression, purification and quality control

To obtain insight into the molecular mechanisms of ligand recognition in human Langerin by means of biophysical methodologies, the protein needs to be accessible in high purity and high amount. Moreover, isotope enriched protein is required for NMR studies. Recombinant bacterial expression systems provide both high yields as well as straight-forward and inexpensive means for isotope labeling. However, bacteria lack an elaborated secretory and processing system that ensures correct formation of disulfide bridges by means of chaperones and redox-shuffling systems (Madigan et al., 2003). As typical for all CTLD-containing proteins, two highly conserved disulfide bonds ensure the stability of the fold (Stambach and Taylor, 2003; Zelensky and Gready, 2005). Expression of Langerin CRD into the bacterial periplasm, where an oxidative environment favors disulfide formation, was described before but was not reproducible (Chatwell et al., 2008). An alternative approach utilized insoluble expression into inclusion bodies that required refolding of the protein into functional form (Stambach and Taylor, 2003; Thépaut et al., 2009). However, poor yields were obtained using the published protocols. Therefore, a new refolding and purification strategy based on a refolding screen was contrived: Solubilized inclusion bodies are refolded by rapid dilution into L-arginine-containing buffer with a glutathione redox-shuffling system and subsequently purified via affinity chromatography. The CRD suffered from severe aggregation which could be overcome by adopting a low-pH precipitation of soluble aggregates that has been described for the purification of the related CTLR DC-SIGN (Pederson et al., 2014). This newly established methods awarded both recombinant Langerin monomeric CRD and trimeric truncated ECD (**Figure 4-1a**) in high yields of 10 to 20 mg L^{-1} of expression medium. The two constructs allow for discerning interactions that are inherent to the single domain or require the trimeric form of the receptor.

To assess whether the protein constructs had the correct identity and met quality parameters necessary for protein NMR, a series of quality controls were conducted. The protein constructs had the expected molecular weights of 18 kD and 23 kD, respectively, as was confirmed by SDS-PAGE (**Figure 4-1b**). Moreover, one-step purification was sufficient because no major protein contaminants were observed. Chemical crosslinking of Langerin ECD resulted in monomer, dimer, and trimer species at low concentrations of the cross linker. At cross linker concentrations above 1 mM, only the trimer species and some higher

molecular weight species were observed (**Figure 4-1b**). However, the band corresponding to the trimeric protein was dominant indicating that the higher molecular weight species might be attributable to unspecific interactions of several protein trimers. Overall, the data confirmed that recombinant truncated Langerin ECD formed trimers in solution.

The secondary structure was confirmed by CD spectroscopy (**Figure 4-1c**). The two constructs were folded with well-defined α-helical and β-strand elements. As expected, Langerin ECD displayed a larger α-helical content due to the α-helical neck region. Isotope-labeling efficiency was confirmed by MALDI-TOF mass spectrometry. In the region of interest, one single major peak was found corresponding to the single-charged protein. ^{13}C-isotope-labeling efficiency was 98% as calculated by mass difference (**Figure 4-1d**). In the ^{15}N-istope labeled sample, the measured mass deviated by 112 D from the predicted value of the uniform ^{15}N-labeled full-length primary structure of the construct (18,186 D). Since the second residue in the nascent chain is an alanine, the bacterial processing machinery likely cleaved the initial N-formyl methionine post-translationally, which would account for a mass difference of 177 D (Hirel et al., 1989). The resulting mass difference of 65 D either accounts for non-uniform ^{15}N-isotope labeling or non-uniform cleavage of the formyl methionine. The truncated human Langerin ECD showed one major mass peak corresponding to 22.77 kD (**Figure 4-1e**). The observed mass difference of 169 D to the calculated mass of the full-length sequence (22,936 D) indicated an almost complete cleavage of the N-formyl methionine.

Protein aggregation was inspected by dynamic light scattering. Samples passed quality control if they displayed only one monodisperse peak below 10 nm apparent radius of gyration (**Figure 4-1f**). Protein activity of trimeric Langerin ECD was ensured by the purification method; inactive protein would not bind to the mannan column and is thus washed away before elution (Taylor and Drickamer, 2003). However, the low affinity of the CRD to the mannan resin rendered activity assessment more challenging. Based on the structure-function paradigm correct folding of Langerin CRD should entail its functionality. ^{1}H-^{15}N HSQC NMR spectra were used as fingerprint of folding (**Figure 4-1g**). A well spread out spectrum indicates both differential hydrogen bond patterns (^{1}H dimension) and backbone torsion angles (^{15}N dimension) and thus a defined tertiary structure (Cavanagh et al., 2007). In addition, activity was confirmed at a later point by binding studies with Ca^{2+} and mannose (vide infra). Taken together, functional forms of both recombinant Langerin CRD and ECD were readily obtained in high yields and high purity thus providing the necessary material to elucidate the molecular basis of ligand recognition.

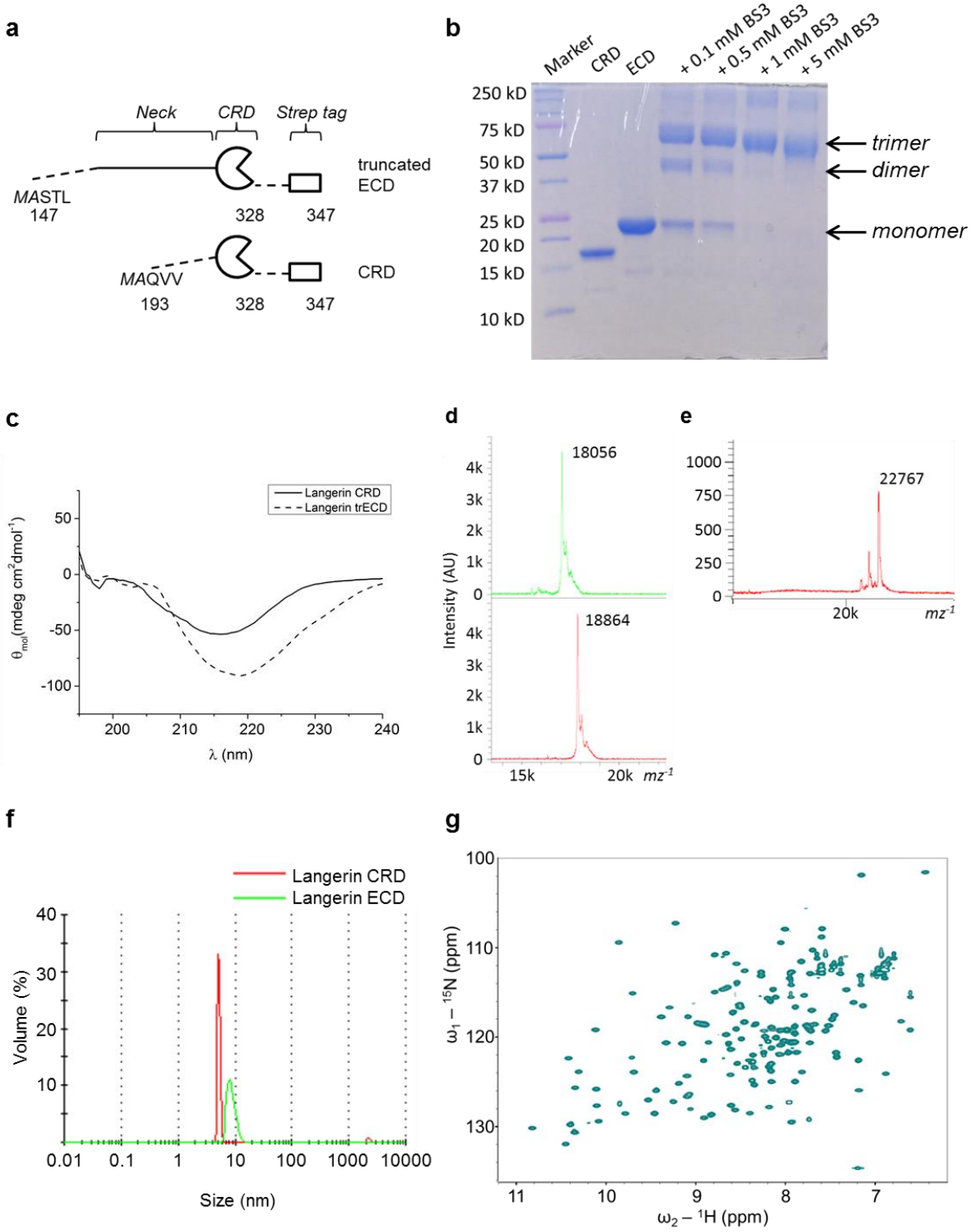


Figure 4-1: Quality control of recombinant Langerin proteins. (**a)** Recombinant expression construct of Langerin truncated ECD and CRD used in this study. Both constructs were expressed with a C-terminal Strep tag. Sequence numbers following the full-length protein numbering scheme are given. (**b**) SDS-PAGE of denatured CDR and ECD as well as cross-linked ECD. (**c**) Overlay of CD spectra of CRD and ECD. (**d**) MALDI-TOF mass spectrogram of (U:15N) and (U:^{15}N;^{13}C) isotope labeled human Langerin CRD (top and bottom respectively). (**e**) MALDI-TOF mass spectrogram of human Langerin truncated ECD. (**f**) Dynamic light scattering size distribution based on the estimated hydrodynamic radius of Langerin CRD and ECD. (**g**) ^{1}H-^{15}N HSQC NMR spectrum of (U:^{15}N;^{13}C) isotope labeled Langerin CRD showing a well-spread resonances over the entire spectrum indicating proper folding of the domain.

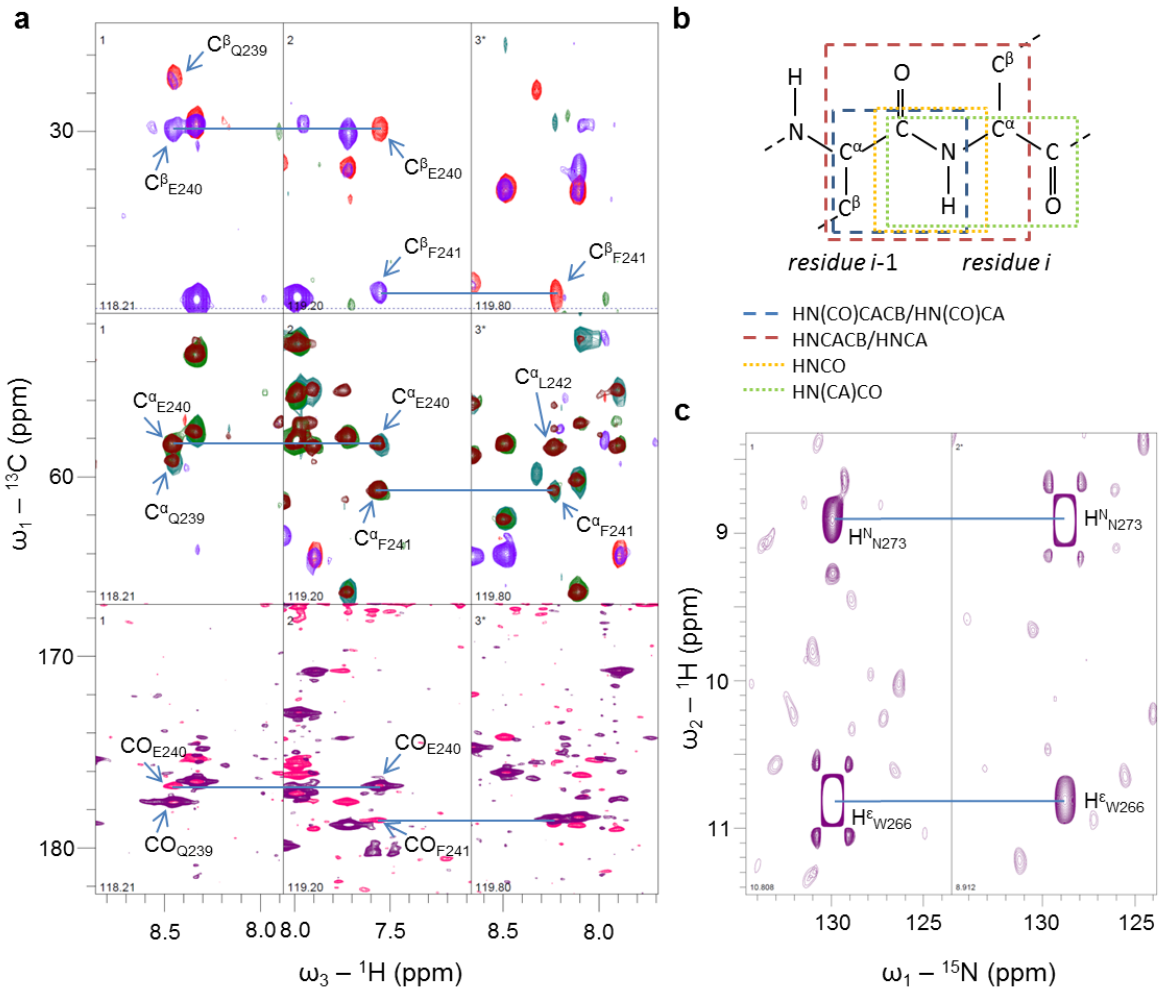


Figure 4-2: Backbone resonance assignment strategy for Langerin CRD. (**a**) Sequential walk through the protein backbone via the ^{13}C resonances either connected with the N and H^N resonances of the same or the previous residue. Strips in the ω_1-ω_3 plane are shown. The level of the planes corresponds to the ^{15}N chemical shifts (ω_2) of the corresponding backbone resonances. The example shows the connection of residues 240-242. (**b**) Graphical representation of the complementary 3D experiments utilized for backbone assignment. Whereas HNCACB, HNCA, and HN(CA)CO report both on ^{13}C resonances of the same (*i*) and previous residue (*i*-1), HN(CO)CACB, HN(CO)CA, and HNCO have selective transfers over the carbonyl group thus only reporting on $^{13}C_{i-1}$ resonances allowing for a sequential walk-through in *i*-1 direction. (**c**) A ^{1}H-^{1}H NOESY-^{15}N HSQC spectrum allows for the assignment of spatially close protons (r<5 Å). The example shows the interactions of backbone H^N_{N273} with the proximal side chain H^{ε}_{W266}. The assignment was guided by the X-ray crystal structure (pdb entry: 3p5f, Feinberg et al., 2011).

4.1.2 Backbone assignment of human Langerin CRD

To obtain structural information from NMR spectra, the observed resonances were assigned to single residues of the protein using a set of complementary three-dimensional NMR experiments that allowed for a sequential walk through the protein backbone (vide supra**, Figure 4-2a-b**)[8]. In addition, analysis of NOESY-HSQC data aided to resolve ambiguous assignments due to peak overlap and verified assignments based on the Langerin CRD X-ray crystal structure (Feinberg et al., 2011) (**Figure 4-2c**). In total, 91% of the backbone NH resonances were assigned covering the majority of the domain (**Figure 4-3**, **Table A-1, BMRB entry: 26791**). Additionally, analysis of the CC(CO)NH experiment awarded assignments of several side chain ^{13}C resonances including all ^{13}C resonances of eight out of nine proline residues. Furthermore, by combing CC(CO)NH and NHCACB data, some glutamine and asparagine side chain amide resonances were also assigned. Moreover, a NOESY cross peak of H^{N}_{N273} to a resonance in the indol H^{ε}-N^{ε} region of the ^{1}H-^{15}N HSQC allowed for the assignment of the W266 side chain H^{ε}-N^{ε} resonances. According to the crystal structure (pdb entry: 3p5f, Feinberg et al., 2011), W266 is the only tryptophan residue with a N^{ε} within 5 Å of the backbone N of N273 (**Figure 4-2c)**. Finally, analysis of K257A and N297A mutants (vide infra) aided to assign these two residues.

Unassigned regions were located in the center of the long loop (residues 286-288, 293-295) and the β4 strand (residues 307-311) that are both in proximity of the Ca^{2+} cage. Assignment attempts failed presumably due to unfavorable dynamics in this region as observed for other CTLRs (Probert et al, 2014, Pederson et al. 2014). The signal intensities of the assignment experiments decreased in the flanking residues and the sequential walk could not be continued. Only residues 289-292, which form a transient 3/10-helix in the crystal structure, could be assigned. An alternative approach to assign the backbone amides of the β4 strand using NOE cross peaks from spatially close β3 strand residues was unsuccessful due to missing peaks in the ^{1}H-^{1}H NOESY-^{15}N HSQC NMR spectrum.

In recent years, ^{13}C-detected NMR has provided invaluable insight into intrinsically disordered proteins (IDP) exploiting the greater chemical shift distributions of the heteronuclei and different relaxation properties (Felli and Pierattelli, 2012). Therefore, a set of ^{13}CO-detected experiments was recorded to investigate whether the unassigned regions could be found in these spectra. However, the same effect was observed as in the three-dimensional

[8]Data was collected and spectra were processed by Monika Beerbaum, Leibniz Institute for Molecular Pharmacology FMP, Berlin. Data analysis was conducted by me with fruitful discussions with Dr. Marcel Jurk, Max Planck Institute of Molecular Genetics, Berlin.

$^{1}H^{N}$-detected assignment spectra. In the CON spectrum (**Figure 4-3b**), which is equivalent to the ^{13}C-^{15}N projection of the HNCO, signals in the regions of interest were abrogated as well. However, the spectra allowed the assignment of all but one proline amide N resonances. Unfortunately, P286 that constitutes the center part of the EPN motif in Langerin was not among them.

Notably, assignment of the C-terminal Strep-tag revealed the existence of two distinct conformational states (**Table A-2**). One population exhibited strong signal intensities indicating free rotation in solution whereas the other population showed weak signals in the ^{1}H-^{15}N HSQC root spectrum. The latter indicates re-binding of the tag to other parts of the domain. This could have influenced the overall molecular tumbling and thereby protein relaxation. Unfortunately, the re-bound species was only discovered after the majority of the experiments comprising this work were already carried out.

Taken together, atoms in 94% of all residues and 62% of all ^{13}C and ^{15}N resonances of the Ca^{2+}-bound CRD were assigned providing the foundation for site-specific analysis of Langerin-ligand interactions by biomolecular NMR (**Table 4-1**). The loss of signals in the long loop region indicates dynamics in the micro- to millisecond timescale in the region.

Table 4-1: Assignment summary for hLangerin CRD for the residue 193 to 328 (excluding Strep tag II and TEV cleavage site).

Category	Available	Assigned	% Assigned
Backbone residues	136	128	91
Backbone N+H	263	238	90
CO	136	122	90
C_{α}	136	127	93
C_{β}	126	116	92
H^{α}	97	146	66
H^{N}	116	127	91

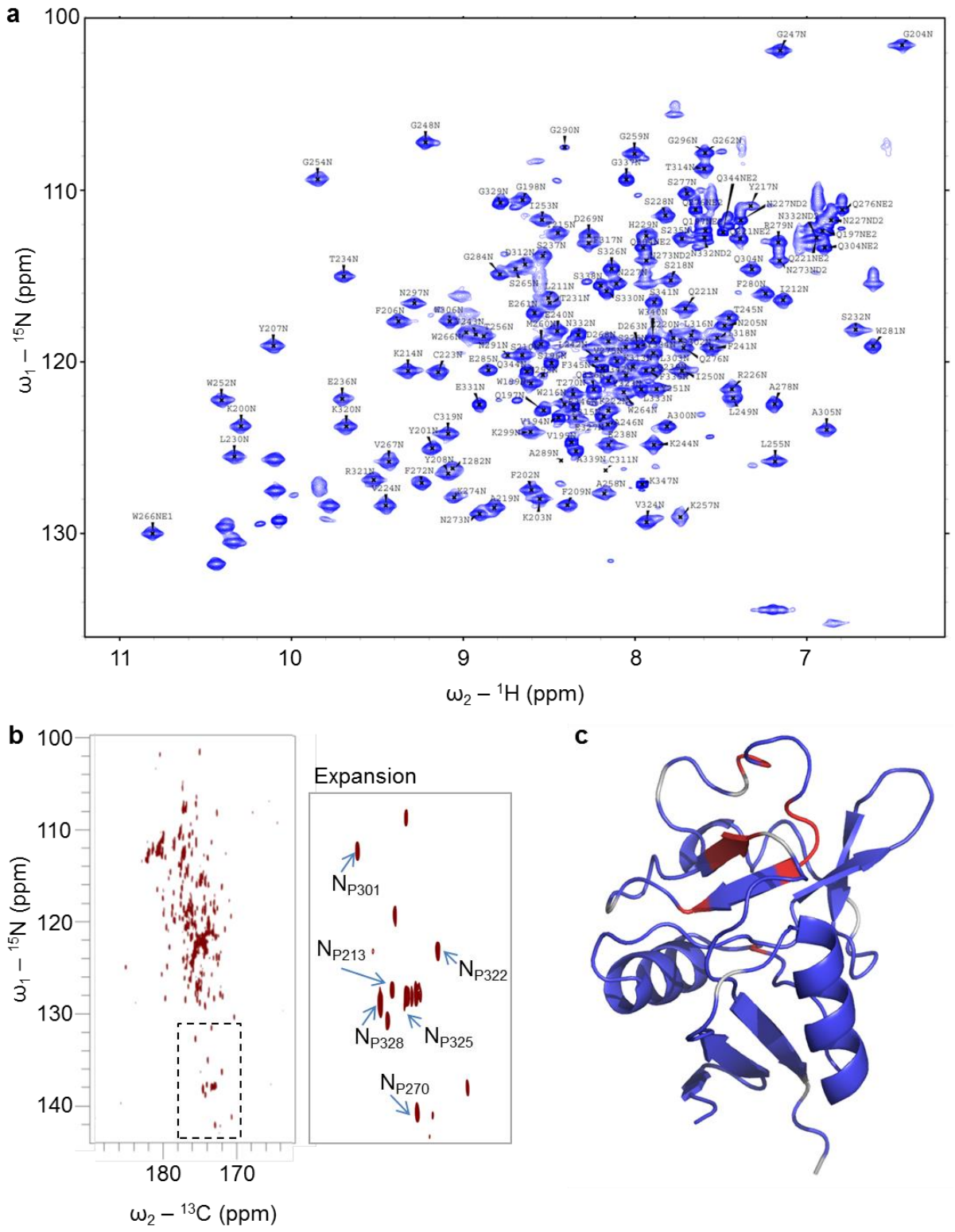


Figure 4-3: Backbone resonance assignment of Langerin CRD. (**a**) 92% of all backbone amide resonances were assigned using standard triple resonance spectra, NOESY spectra, and mutant analysis. In addition, some side chain amides and W266 indol NH resonances were assigned. Backbone resonances are marked with the corresponding residue and N, side chain resonances with the corresponding nitrogen NE1 for tryptophane indole, ND2, or NE2 for asparagine and glutamine side chain amides, respectively. All assignment spectra were recorded at pH 6, 298 K, and in presence of 5 mM $CaCl_2$. (**b**) CON spectrum with an expansion of the prolyl backbone nitrogen region with six assigned resonances. (**c**) Structure representation of Langerin CRD (pbd entry: 3p5f, Feinberg et al., 2011) color coded according to assignment status (blue – assigned, red – unassigned, light grey – prolines).

4.2 Langerin Ca^{2+} interactions

Langerin requires the binding of one Ca^{2+} ion as cofactor for canonic carbohydrate binding. To understand more about the structure and dynamics of Langerin-Ca^{2+} interactions at physiological relevant pH, I conducted plate-based assays, isothermal titration calorimetry and biomolecular NMR measurements.

4.2.1 Langerin binding is pH dependent

To initiate the study, pH dependence of carbohydrate binding to human Langerin was reproduced in a plate-based assay[9]. Langerin ECD affinity to coated mannan decreased steeply with decreasing pH. The data was fitted to a dose-response model with the turning point at pH 5.4±0.1 (pH_{50}) and a Hill coefficient of 1.7±0.3. Below pH 5, binding was completely abolished (**Figure 4-4a**). These findings agreed with previously published work by Stambach and Taylor (2003). To dissect whether this loss of affinity was indeed solely attributed to Ca^{2+} affinity, I set out to evaluate binding at both pH 7 and pH 6 which reflect the conditions encountered on the cell surface and in the early endosomal compartment were ligand release occurs (Ibid.).

4.2.2 Langerin pH-dependent glycan binding can be delineated to a decrease of Ca^{2+} affinity

First, affinities of both Ca^{2+} and mannose with Langerin CRD and ECD were assessed by isothermal titration calorimetry (**Figure 4-4b-c**). At pH 7, both recombinant proteins showed comparable affinities for Ca^{2+} of 105±15 µM and 130±50 µM for CRD and ECD, respectively. The same was true for mannose with affinities of 6.2±0.5 mM and 5.6±1.5 mM for CRD and ECD, respectively. The findings for mannose were in good agreement with previously published data (Wamhoff et al., 2016; Feinberg et al., 2013). Interestingly, Ca^{2+} affinities, which had not been reported for Langerin before, agreed well with data published for ASGPR (Onizuka et al., 2012) indicating a conserved mechanism among related CTLRs. Unfortunately, the thermodynamic properties of change in free enthalpy and entropy during binding could not reliably be extracted from the data due to the unfavorable c-value of the measurements (Turnbull and Daranas, 2003).[10] For the same reason, the number of binding sites for the interactions had to be set to one and could not be determined independently. However, all reliable published X-ray crystal structures only show one coordinated Ca^{2+} ion

[9] Measurements were partially performed and analyzed by Henrik Schmidt under my supervision.

[10] The c-vvalue is the ratio of receptor concentration and K_d. Only at $c > 5$, binding enthalpy can be obtained with confidence because otherwise, the lower asymptote of the calometrigram cannot be directly obtained from the experimental data. For low binding affinities, receptor concentration has to be increased. However, high risk of aggregation at protein concentrations above 400 µM imposed restrictions on increasing the C value.

at saturating concentrations of the ion (Thépaut et al., 2009; Feinberg et al., 2010) therefore justifying the assumption of only one binding site. The good agreement in K_d for both CRD and ECD indicated that there is no major inter-domain allostery at play in both in mannose and Ca^{2+} binding at pH 7. This is in accordance with previously published work on DC-SIGN and DC-SIGNR (Coombs et al., 2010). To quantify the pH dependence of binding, ITC measurements were performed at pH 6. Due to protein precipitation, measurement quality was very low and only for Langerin CRD interacting with Ca^{2+} binding affinities could be obtained awarding a K_d of 780±150 µM corresponding to an about 6-fold decrease of affinity (**Figure 4-4d**).

To assess the effect of pH on binding of mannose, I performed ^{1}H-^{15}N HSQC NMR titration experiments at pH 6 at saturating Ca^{2+} concentrations of 10 mM. The resulting K_d of 11.4±1.4 mM is about twofold increased compared to the K_d for mannose at pH 7[11]. At pH 6, however, a concentration of 10 mM $CaCl_2$ would only result in about 92% receptor occupancy based on the observed K_d. Increasing the $CaCl_2$ concentration further was unfavorable for protein stability likely due to destabilizing effects of high ionic strength. However, comparing the magnitude of change in Ca^{2+} and carbohydrate affinity, 5 kJ mol^{-1} and 1.5 kJ mol^{-1} for Ca^{2+} and mannose binding, respectively, the observed pH dependence in ligand release is majorly attributable to a loss of Ca^{2+} affinity rather than to the carbohydrate itself which is in line with previously published work (Stambach and Taylor, 2003).

[11] The data is presentedbelow in context of carbohydrate recognition of Langerin CRD.

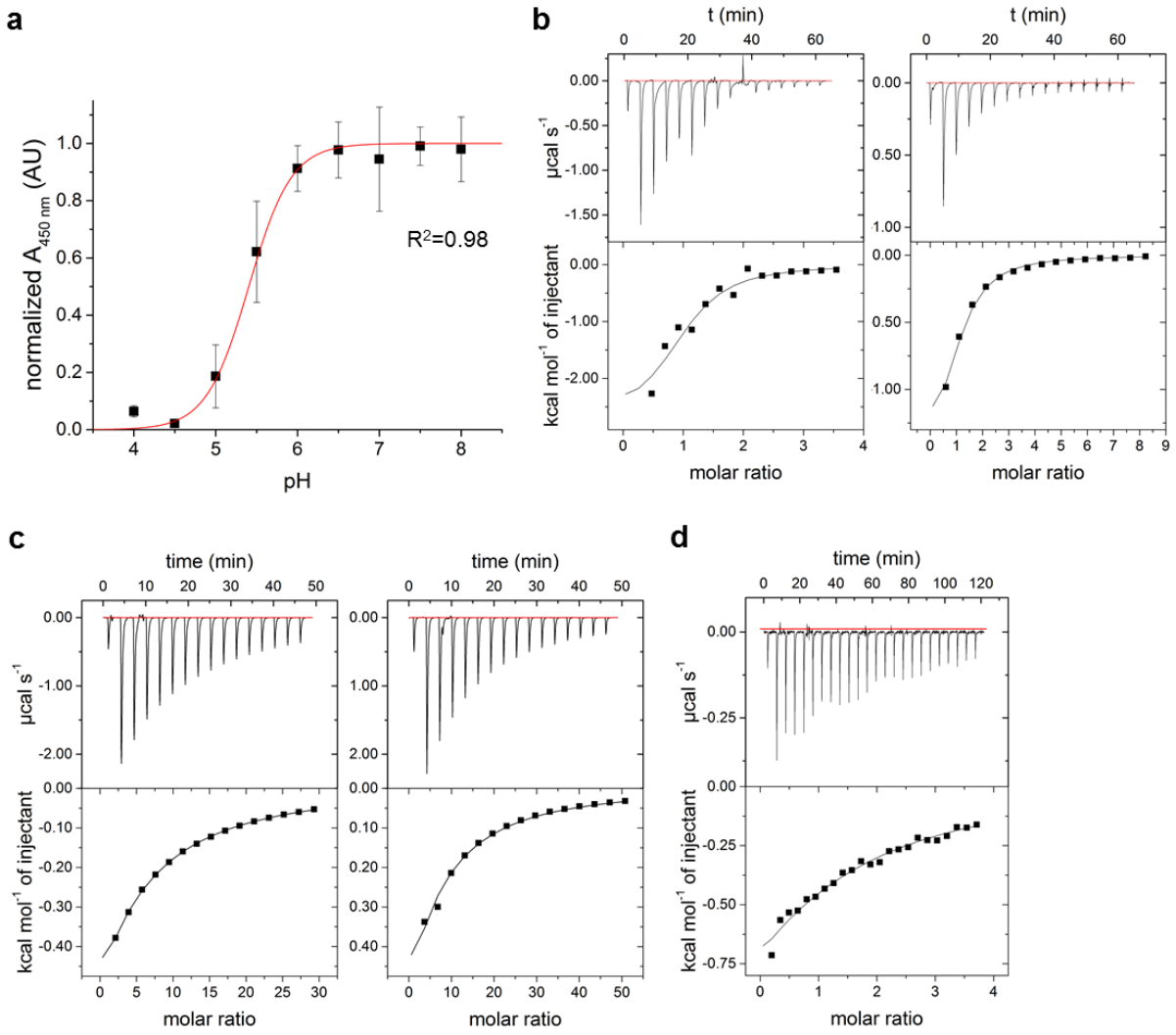


Figure 4-4: Langerin binds Ca^{2+} in a pH-dependent fashion. (**a**) Binding of Langerin ECD to mannan-coated plates decreases rapidly upon acidification. Fitting the data to a dose-response model resulted in a pH_{50} of 5.4±0.1 and a Hill coefficient of 1.7±0.3. (**b**) ITC thermograms and reaction heats of Ca^{2+} interacting with Langerin CRD and ECD at pH 7. Both proteins share the same affinity for Ca^{2+} with a K_d of of 105±15 µM and 130±50 µM for CRD and ECD, respectively. (**c**) ITC thermogram and reaction heats of mannose interacting with Langerin CRD and ECD at pH 7. Both proteins share the same affinity for Ca^{2+} with a K_d of of 6.2±0.5 mM and 5.6±1.5 mM for CRD and ECD, respectively. (**d**) ITC thermogram and reaction heats for Ca^{2+} binding to Langerin CRD at pH 6 with a K_d of 780±150 µM. All measurements were recorded at 298 K.

4.2.3 Apo Langerin undergoes prolyl cis/trans isomerization

Having observed pronounced differences in Ca^{2+} and sugar affinity at extracellular and endosomal pH, I set out to unravel the underlying mechanisms by means of biomolecular NMR. Since I was specifically interested in the mechanism of pH-dependent ligand release, I started the investigation with ^{1}H-^{15}N HSQC NMR measurements of the apo form of human Langerin CRD.[12] The backbone resonance assignment was performed with the holo form at pH 6 (vide supra) and assignment was transferred to the apo spectra by following the peak changes during Ca^{2+} titration (Williamson, 2013)[13]. Accordingly, 84% of the amide backbone resonances in the apo form were assigned. Surprisingly, additional peaks in the ^{1}H-^{15}N HSQC spectrum of the apo form of the CRD pointed towards a second conformation of the protein in slow exchange. Published work on other CTLRs that acknowledged cis/trans isomerization of the highly conserved proline residue of the binding motif in the apo state (Ng and Weis, 1998; Ng et al., 1998; Poget et al., 2001; Nielbo et al., 2004; Onizuka et al., 2012) hinted for the existence of a slowly interchanging trans state in apo Langerin as well. Unfortunately, chemical shift data of the ^{13}C side-chain resonances of the suspected isomerizing proline P286 was not available, which would report on the conformation of the prolyl bond (Schubert et al., 2002). Therefore, a P286A mutant was generated that ought to force the long loop into a conformation resembling the P286 prolyl trans state. Indeed, the mutant was folded and the mutation caused several resonances to shift in the proximity of the newly arisen peaks in the WT apo form (**Figure 4-5a-b**). Consequently, it can be assumed these peaks belong to the trans state of the protein and undergo slow exchange with the cis state. Hence, 23 resonances were assigned to the trans state. These residues were located in the upper region of the fold comprising both the long and short loop, the α3 helix, and the β-strands β2, β2', β3 and β4 (**Figure 4-5c-d**). Notably, the chemical shift differences between the trans and cis states of residues in the β2/β2'-sheet region and the enclosed short loop were greater than those observed for residues proximal to P286 indicating that the short loop regions is allosterically affected by events occurring in the binding site.

[12] 1H-15N HSQC-TROSY NMR experiments did not afford interpretable spectra of the trimeric Langerin ECD construct likely due to anisotropic tumbling of the 66 kD receptor (Stambach and Taylor, 2003).

[13] According to the conventional practice for naming of cofactor-binding proteins, apo and holo refer to the Ca^{2+}-free and Ca^{2+}-bound forms of the protein (Berg, J.M., Tymoczko, J.L., and Stryer, L. (2007). Biochemistry, 6th edn (New York: W.H. Freeman). To describe carbohydrate binding to the holo form, the terms carbohydrate-bound or carbohydrate-free will be used.

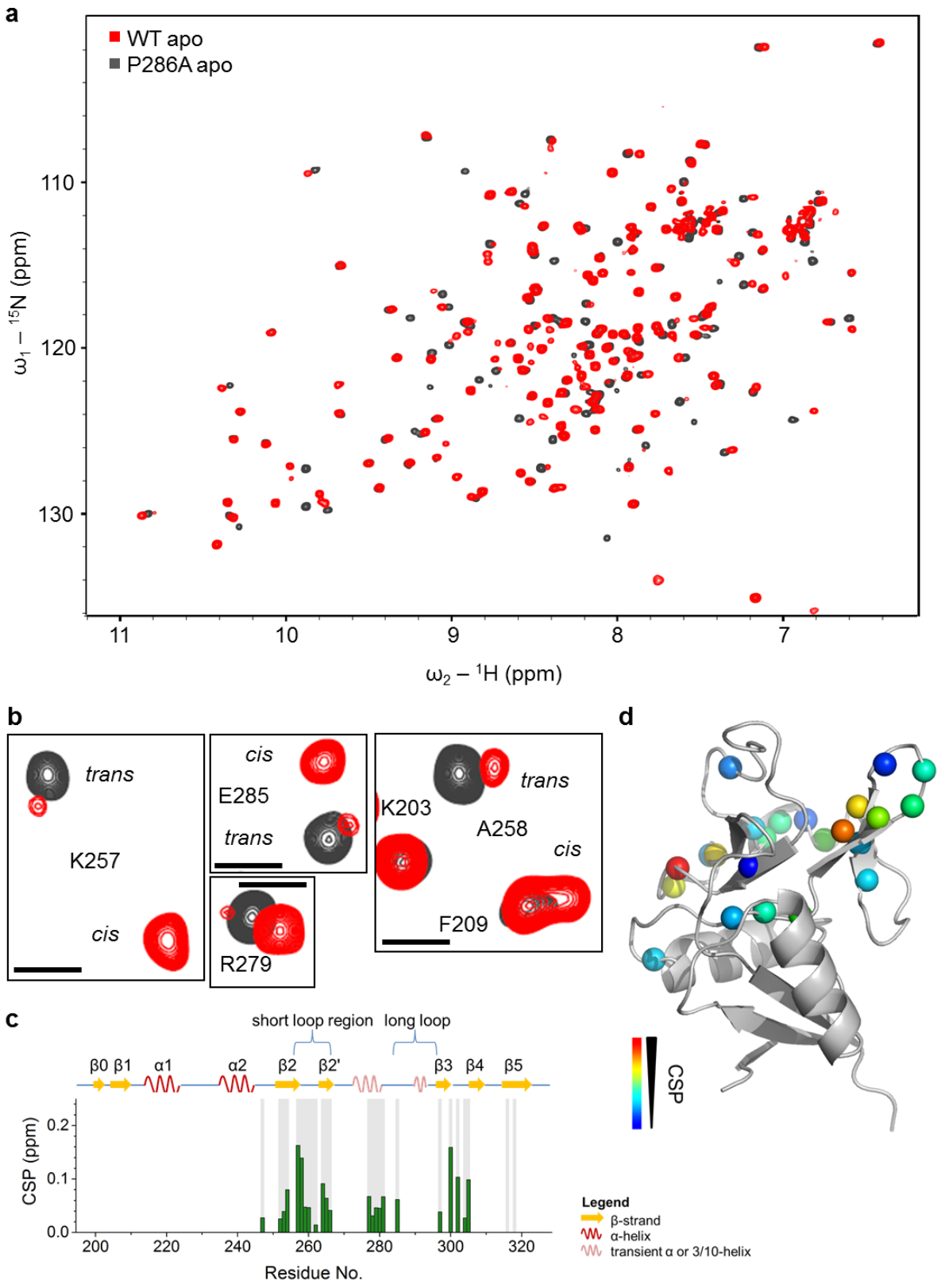


Figure 4-5: Apo Langerin CRD experiences cis/trans isomerization of P286 prolyl bond. (**a**) Overlay of ^{1}H-^{15}N HSQC NMR spectra of Langerin CRD WT apo and P286A mutant at pH 6. (**b**) Expansions of the backbone resonances of representative residues undergoing cis/trans isomerization. In total, 23 residues were identified. Black bar represents 0.1 ppm and 0.75 ppm on the ^{1}H and ^{15}N chemical shift scale, respectively. (**c**) Chemical shift differences between trans and cis states at pH 6. Grey bars indicate all residues identified to undergo cis/trans isomerization. In some cases, chemical shift differences could not be extracted due to peak overlap. (**d**) Chemical shift differences between a cis and trans apo states plotted on Langerin structure (pdb entry: 3p5f, Feinberg et al., 2011). All spectra were recorded at pH 6, 298 K.

The chemical shift differences of the ^{15}N resonances provide information on the timescale of the inter-conversion of the cis and trans conformations because they are more sensitive to backbone rearrangements than the ^{1}H chemical shift (vide supra). The ^{15}N chemical shift differences ranged from 1 to 80 Hz. For slow exchanging systems, the exchange rate is expected to be at least one order of magnitude slower than the chemical shift difference (vide supra). Accordingly, the exchange rate is slower than 0.1 s^{-1} that is in the range expected for not catalyzed prolyl cis/trans isomerization (Andreotti, 2003) and has been described for MBP, a related CTLR before (Ng and Weis, 1998).

Surprisingly, the ratio of the peak intensities of the cis and trans states revealed that about 75±10% of the apo form remained in the cis conformation (**Table B-1**), which is in contrast to other CTLRs were the dominant species in the apo form occupies the trans state (Ng and Weis, 1998; Ng et al., 1998; Poget et al., 2001; Nielbo et al., 2004; Onizuka et al., 2012). Upon addition of Ca^{2+}, only the resonances in the cis conformation experienced chemical shift perturbations while the peaks of the trans population decreased in size. This is indicative for conformational selection of the cis state for Ca^{2+} binding. The peak intensities of the trans apo state can serve as a direct reporter for the fraction of unbound protein. Fitting of the peak intensities of trans A258 against the ligand concentration assuming a one-site binding model awarded a K_d of 810±110 µM, which is identical with the value observed in ITC (**Figure B-3**). However, the poor signal-to-noise ratio and difficult error assessment of most trans state peaks did not allow for a general application within this method.

In conclusion, the apo form of Langerin undergoes cis/trans isomerization but is predominantly in the cis prolyl conformational. This cis/trans isomerization affects next to the long loop also the short loop substantiating an allosteric coupling of the two loops.

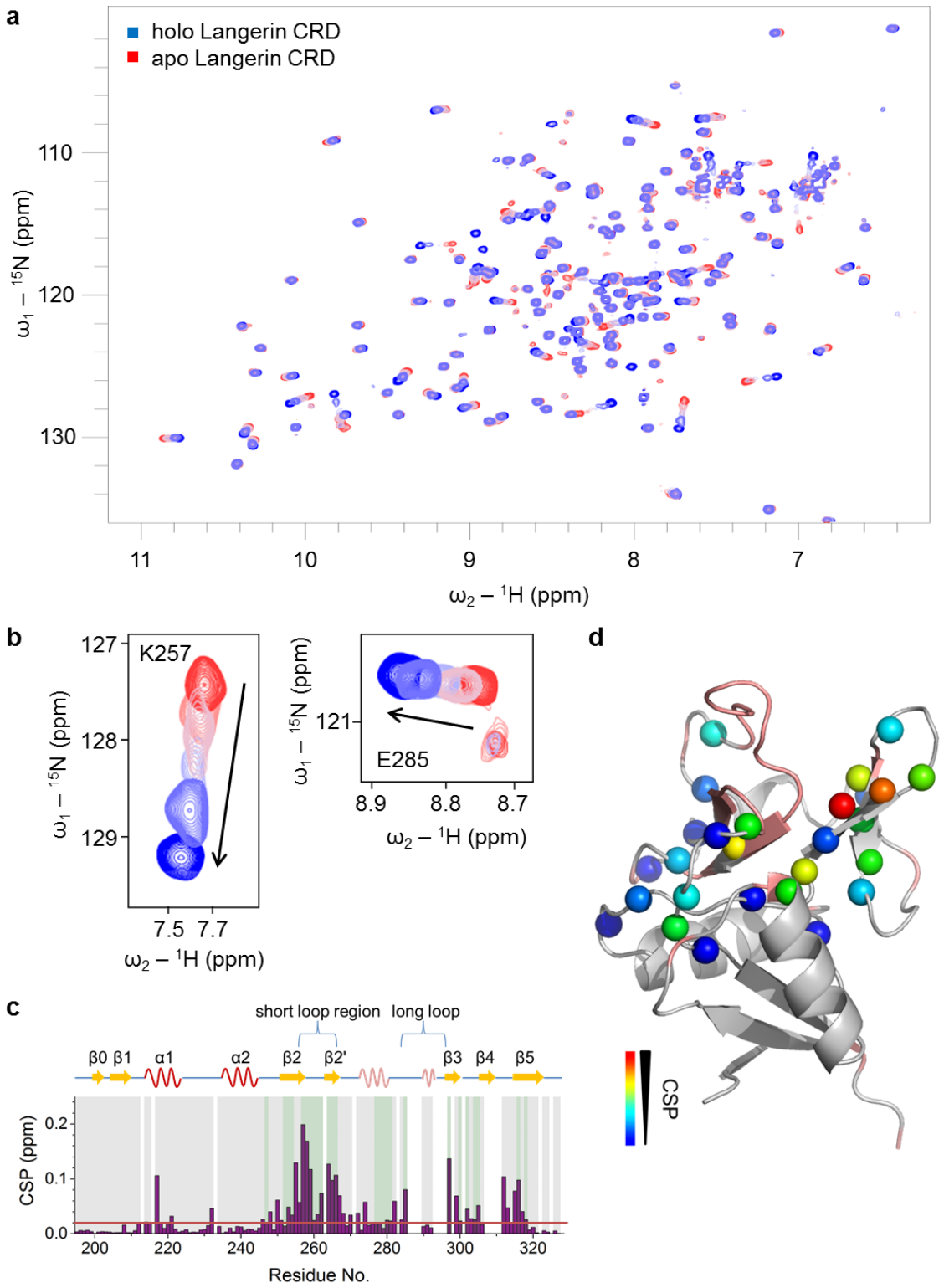


Figure 4-6: Ca^{2+} binding results affects several residues causing them to undergo chemical shift changes in intermediate exchange. (**a**) Overlay of ^{1}H-^{15}N HSQC NMR spectra of Langerin CRD recorded during a Ca^{2+} titration series at pH 6, 298 K. The measurement points are labeled on a scale from red to blue with increasing Ca^{2+} concentrations. (**b**) Expansions show the backbone resonances of K257 and E285 – representative residues of the short and long loop region, respectively. Both peaks experience severe line broadening during Ca^{2+} titration indicative for intermediate exchange processes. (**c**) Chemical shift perturbations caused by addition of Ca^{2+} at pH 6. Applying a cutoff of 0.02 ppm (red line) defines 44 residues that undergo change upon addition of Ca^{2+}. Olive bars show residues undergoing cis/trans isomerization in the apo form; grey bars indicate assignment in both holo and apo form. (**d**) Ca^{2+} interaction network plotted on Langerin structure (pdb entry: 3p5f, Feinberg et al., 2011). Unassigned residues are colored salmon.

4.2.4 Ca^{2+} binding activates an allosteric network of residues

Surprisingly, the addition of Ca^{2+} to Langerin CRD resulted in pronounced chemical shift perturbations of the cis conformation of the apo form in the ^{1}H-^{15}N HSQC NMR spectra at pH 6 (**Figure 4-6a**). To discern signal from noise, different methods were tested to define the CSP cut-off value. First, the standard deviation within one dataset was determined by consecutively eliminating all residues with CSP>$3\sigma_0$ using eq. (2-35) (Schumann et al., 2007). This resulted in a cut-off value of 0.035 ppm for Langerin CRD WT Ca^{2+} titration. Using the variation of the tag, a cut-off value of 0.02 ppm corresponding to the average and two standard deviations was defined. Based on three independent measurements with different protein batches measured on three spectrometers with three different HSQC pulse sequences awarded a standard deviation of 0.02 ppm for the average over all residues. Since the latter approach was least biased, the value was chosen as cut-off for all consequent CSP analyses.

Using this cut-off value, 44 residues were identified to be affected by Ca^{2+} addition. (**Figure 4-6c**). Notably, 43 of these residues were outside a 5 Å radius of the binding site which is unexpected for a rigid protein with a single binding site and indicative for an underlying allosteric mechanism (Zuiderweg, 2002). Hence, the perturbed residues constitute an allosteric network responsive to Ca^{2+} binding (Boehr et al., 2013). Plotting the perturbed residues on the protein structure revealed that predominantly the upper part of the fold was affected with the most prominent effects in the β2 and β2' strands and the enclosed small loop (**Figure 4-6d**). These findings parallel the already observed communication of the binding loop with the adjacent short loop region observed for the isomerization of the apo form.

Following the chemical shift changes upon the addition of Ca^{2+} awarded a dissociation constant of 625±35 µM (**Figure B-1**). However, many peaks showed severe line broadening during Ca^{2+} titration indicating intermediate exchange on the NMR time scale (**Figure 4-6b**, **Figure B-2**). This is indicative for allosteric rearrangements occurring on the chemical shift timescale (Isogai et al., 2016) but could also originate from slow kinetics of binding that occur independently from allosteric rearrangements in the protein (Feeney et al., 1979; Williamson, 2013). In the intermediate fast exchange regime, the observed chemical shift change does not necessarily reflect on the weighted population average as in the very fast exchange limiting condition (Feeney et al., 1979). Hence, the values obtained by fitting the data to equation (2-37) tend to underestimate the K_d to an extent depending on the exchange rate in relation to the affinity. Moreover, slow off-rates induce a distortion of the binding isotherm similar to cooperative binding (Williamson, 2013). However, the data fitted well to a one-site binding isotherm (**Figure B-1**) indicating that the off-rate is sufficiently fast (>1,000 s^{-1} for a K_d of

between 10^{-4} and 10^{-3} M, Ibid.). Comparison with dissociation constant obtained from ITC showed no systematic underestimation of the K_d from the NMR data but indicated that the divergence is rather attributable to the different measurement principles of the two methods. In fact, variations in the magnitude of the K_d value are commonly observed when different biophysical techniques were applied (Fielding, 2007).

Taken together, the data evidenced that although cis apo state is the active conformation selected for Ca^{2+} binding, it undergoes additional changes resulting in exchange broadening upon the addition of Ca^{2+}. Hence, these allosteric conformational rearrangements occur on the chemical shift timescale (micro- to millisecond) and are faster than the cis/trans isomerization of the loop which occurs on a s timescale.

4.2.5 The short loop adopts different conformations in the holo form depending on pH

To assess whether the different Ca^{2+} affinities observed at pH 6 and pH 7 are reflected in changes of the underlying allosteric recognition mechanism, ^{1}H-^{15}N HSQC NMR spectra of Langerin CRD were recorded at pH 7 (**Figure 4-7a**). The resonance assignment was transferred from the holo form at pH 6 to the holo form at pH 7 by assigning the closest neighbor (Williamson, 2013). The apo form was assigned by following the chemical shift changes in course of a titration experiment. In case of ambiguities, resonances were not assigned. Thus, 84% and 79% of the resonances of the apo and holo form were assigned, respectively. The differences in chemical shift of each form in respect to change in pH revealed that the apo form is mainly unaffected by change in pH whereas the holo form experienced pronounced chemical shift changes with the most prominent effects in the small loop region (**Figure 4-7b**). The chemical shift differences of the holo forms at pH 6 and pH 7 were mainly localized in the β2-strand and the short loop (**Figure 4-7d**). Plotting the chemical shift perturbations induced by Ca^{2+} binding at pH 7 awarded a similar pattern to that observed at pH 6 although the magnitude of the changes was higher (**Figure 4-7c)**. Using the same cut-off value as before, an allosteric network of 48 residues was identified. Comparison with the network defined at pH 6 showed 95% identity (41 residues). However, the increased size of the network and the magnitude of the chemical shift perturbations indicated that the allosteric mechanism at pH 7 is more prominent than at pH 6. These findings might relate to the observed decrease in Ca^{2+} affinity upon reduction of the pH. Accordingly, ^{1}H-^{15}N HSQC NMR titration awarded a dissociation constant of 160±25 µM for Ca^{2+} binding, which is in agreement with the ITC data (Supplementary figure).

The chemical shift differences between the apo trans and apo cis conformation were unaffected by pH change. Moreover, the cis/trans ratio also remained constant thereby excluding the possibility that a higher preformation of the cis state at pH 7 is the reason for the increased Ca^{2+} affinity.

The increased magnitude of the chemical shift perturbation could be explained by the fact, that the receptor is not saturated with Ca^{2+} at pH 6 and thus the equilibrium populations not in the final bound state. Indeed, with the measured dissociation constant of about 800 µM and a protein concentration of 200 µM the receptor is only 92% saturated at 10 mM $CaCl_2$.

Surprisingly, perturbations in that region were also observed in response to change in pH in the holo forms. In the light of the different affinities, the observed shifts induced by pH. If that was true, the observed changes in the holo form at pH 7 ought to occur along the vector defined by the holo and apo form at pH 6 (Selvaratnam et al., 2012). However, chemical shift projection analysis (CHESPA) of the holo form at pH 7 referenced to the chemical shifts observed for Ca^{2+} binding at pH 6 showed deviation in both the angle defined by the two vectors as well as non-uniform variations of the fractional length of the vector projections X (**Figure 4-4e**). The majority of the residues experiencing a chemical shift perturbation over 0.02 ppm show both the same vector orientation and lengths indicating that there was no major shift in the binding equilibrium towards the unbound state. Notably, several residues of the β2-strand and the sequential short loop deviate both in vector lengths and angle showing different effects: Whereas some residues experienced a decrease in the vector lengths, in neighboring residues the vector lengths increased. The same was observed for the angle distribution. Hence, variation of the pH results in a highly heterogeneous response in the β2-strand and the short loop indicating differential conformations or dynamics for each pH state.

Summarizing these findings, the decrease in Ca^{2+} affinity upon acidification is accompanied by conformational changes in the short loop regions of the holo form. These changes cannot be delineated to mere shifts between bound and free conformation, but originate from distinct conformations of the holo form at both pH values. Hence, there exists likely a pH-responsive allosteric mechanism involving the short loop region.

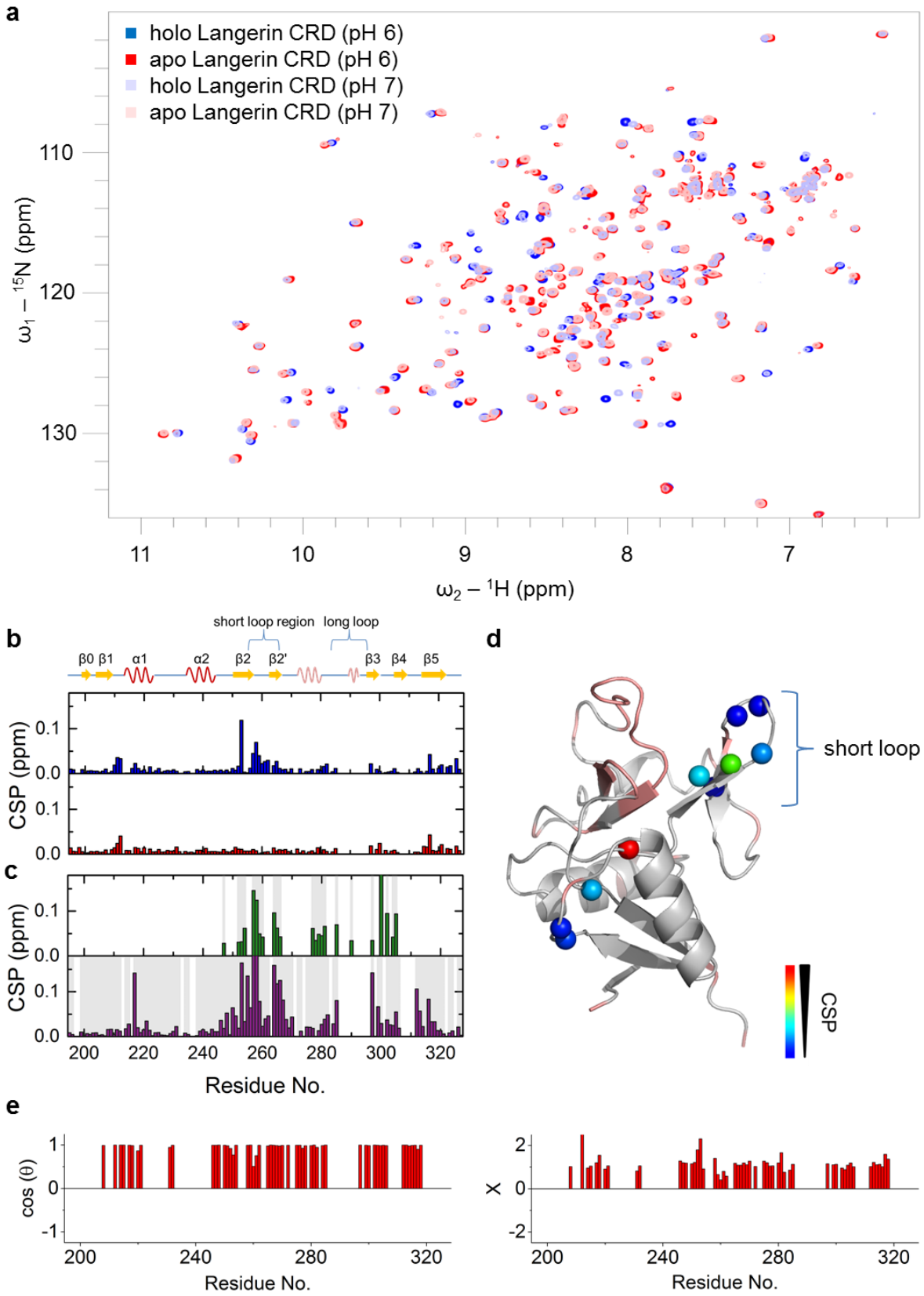


Figure 4-7: The short loop in holo Langerin CRD alters conformations in a pH-dependent manner. (**a**) Overlay of ^{1}H-^{15}N HSQC NMR spectra of apo and holo Langerin CRD at pH 6 and pH 7 at 298 K. Holo forms were generated by addition of 10 mM $CaCl_2$. (**b**) Chemical shift perturbations for apo (lower panel) and holo Langerin CRD (upper panel) upon change of pH. While the apo form remained largely unaffected, in the holo form pronounced CSP were observed in the β2 region and the following small loop. (**c**) Chemical shift differences induced by the addition of Ca^{2+} (upper panel) and observed for the trans and cis conformations of the

apo form (lower panel). Grey bars indicate assigned resonances in the respective two forms used for CSP calculations. (**d**) Chemical shift perturbations above 0.02 ppm are heat-map labeled on the protein structure (pdb entry: 3p5f, Feinberg et al., 2011). Unassigned regions are colored salmon. The short loop experiencing most-pronounced changes is labeled. (**e**) CHESPA analysis of pH-dependent chemical shift changes in Langerin CRD holo form reveal a non-linear relationship. The vector defined by holo and apo Langerin CRD at pH 6 was defined as reference and cos(θ) and relative length of the projection X of the vector defined by the holo form at pH 7 and the apo form at pH 6 calculated. Only residues with a vector length above 0.02 ppm are shown. Although the majority of changes occur along the same axis (cos(θ)≈1), this is not true for some residues in the β2 strand and the sequential short loop. Of note, in the same region the vector projections differ highly in relative length.

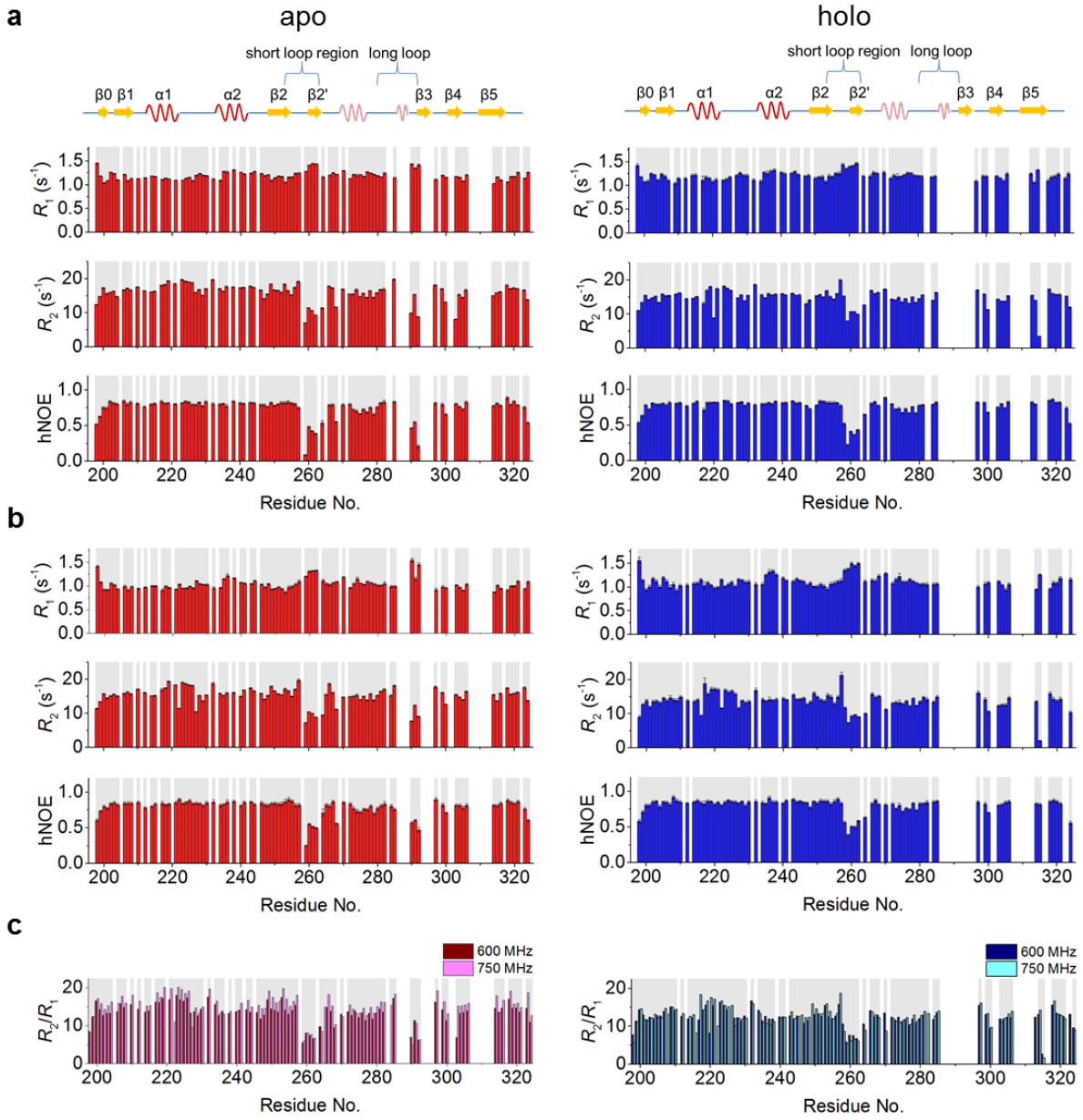


Figure 4-8: Langerin CRD is overall rigid with increased dynamics in the loop regions. (**a**) ^{15}N backbone relaxation rate constants R_1 and R_2 and hetero NOE of apo (left panel) and holo (right panel) Langerin CRD obtained at 600 MHz. The protein exhibits rather uniform relaxation rate constants with major exceptions in the small loop and long loop region. (**b**) Relaxation data collected at 750 MHz. (**c**) R_2/R_1 ratio per residue to estimate local correlation times τ_m at 600 MHz and 750 MHz field strengths apo and holo (left and right panel, respectively). The data confirmed fast internal motions on the pico- to nanosecond timescale in the short loop. In the apo form, an increased ratio at 750 MHz is observed for the α1-helix and β2-strand indicating motions on the micro- to millisecond timescale, grey bars indicate resolved residues used for analysis. All spectra were recorded at 299.2 K, pH 6. Holo forms were generated by addition of 10 mM $CaCl_2$.

4.2.6 Langerin CRD exhibits dynamics on several timescales

The chemical shift and the line widths of a resonance are sensitive to changes in conformational dynamics if they occur on the chemical shift timescale. To assess whether altered conformational dynamics account for the observed chemical shift perturbations, a full set of NMR ^{15}N backbone relaxation data was collected at 600 MHz and 750 MHz field strength for both holo and apo Langerin CRD (**Figure 4-8a-b**).[14]

The two protein forms exhibited uniform relaxation rate constants in most parts of the domain. However, a local reduction of the transverse relaxation rate constant R_2 and the hetero NOE was observed in the short loop region in both holo and apo forms. The hetero NOE is very sensitive to changes in local correlation times and a reduction is therefore indicative for an increased mobility on the pico- to nanosecond timescale (Lipari and Szabo, 1982c). The magnitude of the rate constants in this region was insensitive to Ca^{2+} addition. In apo Langerin CRD, a decrease in R_2 and hetero NOE was additionally observed for residues of the long loop that constitute the 3/10 helix in the holo form. Notably, the resonances of these residues are severely broadened in the holo form prohibiting the extraction of relaxation data. This change in signal intensities upon Ca^{2+} addition indicated a change on the timescale of motions of the long loop between the holo and apo state with the apo state undergoing faster motions. Motions on the ns timescale account for loop flexibility (Henzler-Wildman and Kern, 2007). In fact, the B-factors observed in the crystal structures of holo Langerin CRD agree with the observed increased fast dynamics of both long and short loops and indicate that they are preserved in the holo form (**Figure B-4**). Moreover, the described alternative conformation of the short loop (vide supra) indicates that the loop samples several energetically closely related conformations.

Table 4-2: Average ^{15}N relaxation rate constants obtained from NMR measurements at 600 MHz and 750 MHz field strength. Average values with standard deviations are given. The rotational correlation time was estimated from R_2 and R_1 according to eq. (2-43).

	apo		**holo**	
	600 Mhz	**750 MHz**	**600 Mhz**	**750 MHz**
R_1 **(s^{-1})**	1.2±0.1	1.0±0.1	1.2±0.1	1.1±0.1
R_2 **(s^{-1})**	16.0±2.6	15.2±2.5	14.8±2.3	13.7±2.5
hetero NOE	0.75±0.14	0.79±0.11	0.74±0.2	0.81±0.1
τ_c **(10^{-9} s)**	10.8	11.6	10.0	9.6

[14]Measurements, data processing, and data analysis were conducted with the help of Dr. Martin Ballaschk, Leibniz Institute of Molecular Pharmacology, Berlin.

To estimate the local correlation times, the ratios of the transverse and longitudinal relaxation rate constants, R_2 and R_1, were determined (**Figure 4-8c**). The data showed decreased τ_c values for the short loop in both forms and the residues 290-292 of the long loop in the apo form confirming the increased flexibility on the on the pico- to nanosecond timescale in these regions (Kay et al., 1989). Furthermore, increased local τ_c values were observed for the α1-helix and β2-strand in the apo form due to an local increase in R_2. This was even more pronounced at 750 MHz. This field-dependent increase in R_2 is indicative of motions on the micro- to millisecond timescale in these regions (Kay et al., 1989). The trimmed average ratio of R_2 and R_1 awarded global correlation times of about 11 ns for the apo form at both fields. A value of similar magnitude was described for DC-SIGNR, a closely-related CTRL (Probert et al., 2013). In presence of Ca^{2+}, the local τ_c values decreased in the entire domain, which is due to an overall reduction in R_2 observed at both field strengths (**Table 4-2**). This resulted in a reduction of about 1 ns of the overall correlation times at both fields. However, the restricted increase in local τ_c values in the α1-helix was preserved whereas for the β2-strand the picture becomes less clear than in the apo form. The data indicate that upon the loss of Ca^{2+} the overall structure of the CRD becomes more flexible with motions occurring on the micro- to millisecond timescale. In fact, this has been observed by NMR for other CTLRs before (Nielbo et al., 2004; Onizuka et al., 2012; Probert et al., 2014) supporting the notion that Ca^{2+} binding has an overall stabilizing effect on the fold (Ng et al., 1998).

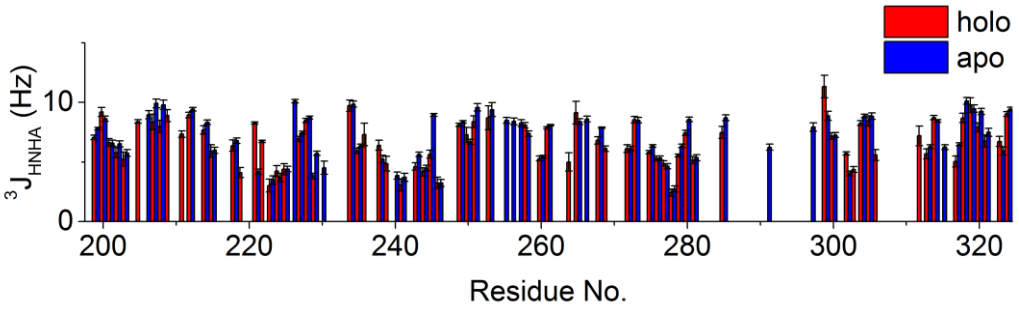


Figure 4-9: Comparison $^3J_{HNHA}$ coupling constants for both apo (red) and holo (blue) Langerin CRD confirms that the overall structure observed in the holo X-ray structure is preserved in the apo form of the protein. $^3J_{HNHA}$ coupling constants were obtained from peak intensity ratios from HNHA experiments by fitting to eq. (2-35). All spectra were recorded at 299.2 K, pH 6. Holo forms were generated by addition of 10 mM $CaCl_2$.

4.2.7 $^3J_{HNHA}$ couplings of holo and apo form indicate over similar structure

The $^3J_{HNHα}$ coupling constants, which report on torsion angle distributions, suggest that the overall structure of the cis holo and cis apo forms is almost identical (**Figure 4-9**). However, the dataset contains several gaps due to poor signal resolution and specifically in the long loop region data for a large strip of residues is missing. Moreover, the only substantial difference noted was for residue Q221 in the α1-helix: an about twofold difference in the $^3J_{HNHα}$ coupling

constants implied a change of the φ torsion angle of about 30° (-90° in holo versus -60° in apo) according to the Karplus relation (vide supra). However, according to the crystal structure of the holo form (pdb entry: 3p5f), the φ angle is -60° as expected for a residue constituting an α-helix. Therefore, the deviation of the coupling constants is likely an error of the measurement and it can be assumed that the holo and apo states share the same overall conformation.

4.2.8 Slow motions act on several timescales in apo Langerin

If the apo and holo forms of Langerin share the same overall structure, the observed differences in R_2 only account for exchange contributions caused by motions on the micro- to millisecond timescale (vide supra). Hypothesizing that the increased transverse relaxation rate constants in the apo forms solely originate from the cis/trans isomerization of P286 prolyl bond (vide supra), the exchange rate can be estimated according to eq (2-30). Based on the averaged differences in R_2 between the holo and apo forms (1.2 s^{-1} and 1.5 s^{-1} at 600 MHz and 750 MHz field, respectively) together with the known equilibrium populations of both the two states (vide supra), the exchange rate was determined to be about 5 s^{-1}. This rate is about the same magnitude as the observed chemical shift differences for resonances in the cis and trans conformation (vide supra). However, for residues in slow exchange the exchange rate is at least 10-times slower than the chemical shift differences (vide supra). Moreover, the literature reports consensually on slower rates, which are less than 0.1 s^{-1} for not catalyzed prolyl cis/trans isomerization as has been observed in the CTLR MBP (Stein, 1993; Ng and Weis, 1998). Consequently, the observed increase in the transverse relaxation rate in the apo form originates from additional faster motions on the micro- to millisecond timescale.

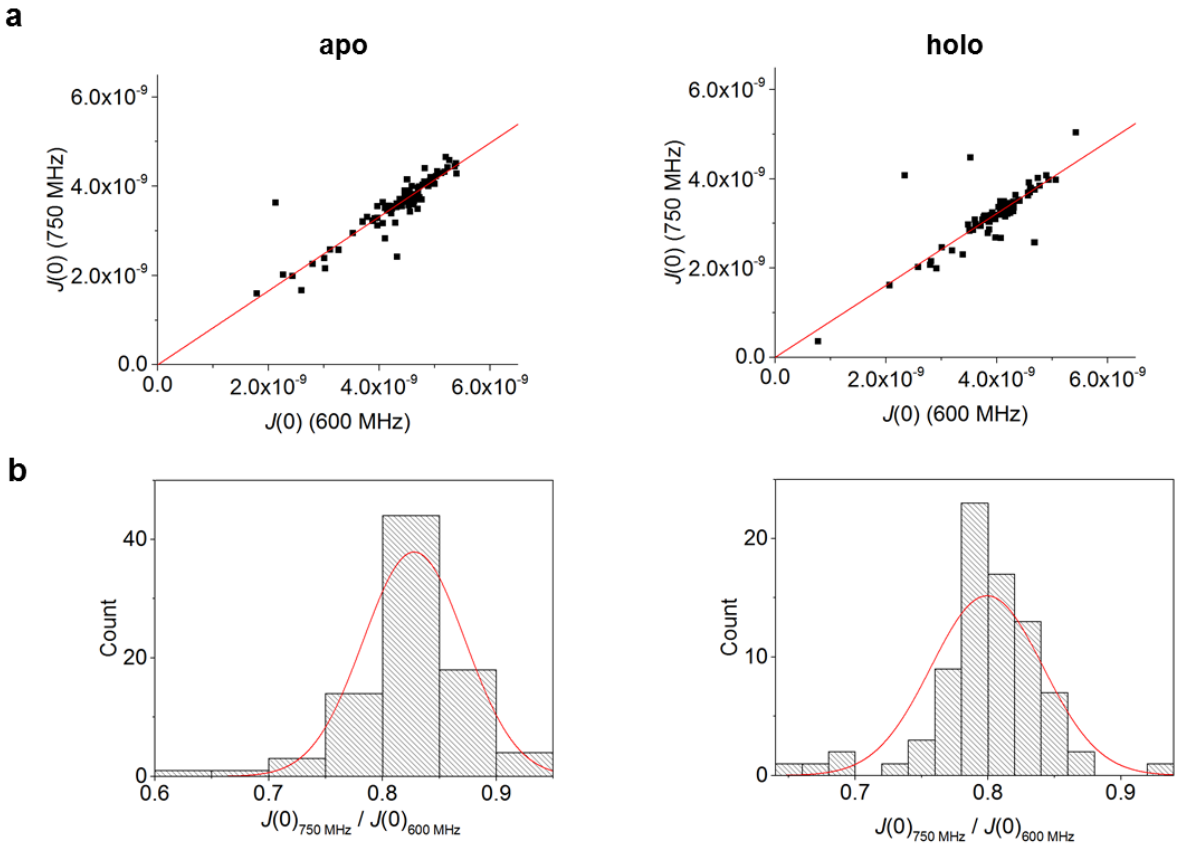


Figure 4-10: A systematic shift of the reduced spectral densities observed for data collected at two fields of both apo and holo Langerin indicates a global exchange process on the millisecond time scale. (**a**) Correlation plot of the reduced spectral densities at zero field $J(0)$ per residue from the datasets collected at 600 MHz and 750 MHz field strength. The left panel shows the apo and the right panel the holo form. The datasets for both forms correlate well linearly. However, the slope is about 0.83 and 0.80 for apo and holo form, respectively. (**b**) Distributions of the ratios of the reduced spectral densities at zero field $J(0)$. Outlier identified from (**a**) were excluded. The distributions show a maximum of 0.83±0.05 and 0.80±0.05 for the apo and holo form, respectively.

4.2.9 Relaxation data displays field dependent shift indicating slow global motions in apo and holo form

To evaluate the quality of the relaxation data, consistency of the collected datasets at different fields was tested by calculation of the reduced spectral densities at zero field $J(0)$. The two datasets were highly correlated with few outliers resulting in Pearson correlation coefficients of 0.96 ($p<0.05$) and 0.95 ($p<0.05$) for apo and holo forms respectively obtained from a pairwise correlation of the datasets collected at 600 MHz and 750 MHz. However, plotting the correlations per residue and fitting the data to a linear model revealed a slope of 0.83 and 0.81 for apo and holo forms respectively indicating that the correlation was skewed (**Figure 4-10**). The same was observed when analyzing the distribution of the ratios of $J(0)$ with the mean of the distribution at 0.83±0.05 and 0.80±0.05 for apo and holo form, respectively. Thus, there was a systematic field-dependent shift in the relaxation data. Temperature variations were excluded by careful calibration before the measurements and are hence

unlikely to account for the shift. Moreover, RF inhomogeneity or changes in viscosity would result in a less homogenous shift of the entire data set. The relaxation rate constants are in principal field independent assuming a homogeneous $\mathbf{B}_0$ field (Abragam, 1961). However, a global exchange in a protein that caused a field dependence of the spectral density functions had been described before in the literature (Peng and Wagner, 1995; Vis et al., 1998). Based on the reduced spectral densities at zero field $J(0)$, this exchange contribution was estimated as 5 s^{-1} for both apo and holo form on the chemical shift timescale of the 600 MHz field (Error! Reference source not found.). The physical origin of this exchange process, however, r emains elusive.

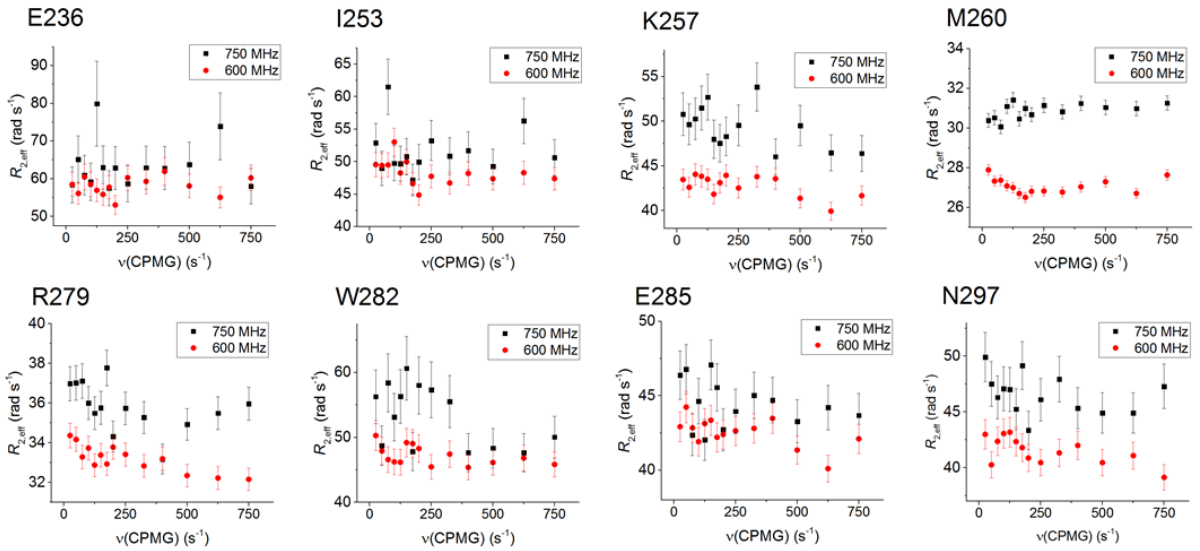


Figure 4-11: Relaxation dispersion measurements hint for slow motions in apo Langerin CRD. Shown are effective transverse relaxation rate constants $R_{2,eff}$ depending on the CPMG frequencies ν_{CPMG} of selected residues at both 750 MHz and 600 MHz field strength. In residues of the short (K257, M260) and long loop (R297, W282, E285, N297) show a clear filed dependence of $R_{2,eff}$. Moreover, several residues exhibit an frequency modulated amplitude of $R_{2,eff}$ indicating exchange processes on the micro- to millisecond timescale.

4.2.10 Relaxation dispersion measurements reveal several slowly interchanging residues

To further dissect the local slow motions on the micro- to millisecond timescale in the apo form of Langerin CRD, relaxation dispersion measurements were conducted at both 600 MHz and 750 MHz field strength [15]. Some residues exhibited a CPMG frequency-dependent variation in the effective transverse relaxation rate indicating chemical exchange on the timescale of the applied CPMG frequencies (750 Hz to 50 Hz) (**Figure B-5**). Moreover, the abrupt changes in $R_{2,eff}$ observed for some residues hint for exchange processes (Vis et al.,

[15] Measurements were performed by Dr. Peter Schmieder, Leibniz Institute of Molecular Pharmacology, Berlin. Spectra processing and data analysis were conducted by me.

1998). A closer inspection of some of the residues revealed that residues of the long and short loop experienced not only a high $\mathbf{B}_0$ field dependency in their effective transverse relaxation rate constants (**Figure 4-11**) but also exhibited a CPMG-field dependent amplitude of the $R_{2,\mathrm{eff}}$ (Millet et al., 2000). This was not observed for residues in the lower part of the domain. The data suggest that upon loss of the Ca^{2+}, the entire upper domain becomes more flexible on a timescale faster than the cis/trans isomerization. Unfortunately, the large error on the data collected at 750 MHz prohibited subsequent data fitting to extract the exchange rate and other parameters.

4.2.11 The model-free approach is not applicable to the dataset

To gain more insight on the nature of the observed fast internal motions of the short and long loop, model-free analysis of the relaxation data was considered to obtain the order parameter S^2 that can be interpreted as a measure for internal entropy (Lipari and Szabo, 1982a; Yang and Kay, 1996). However, the observed slow motions in the apo form due to cis/trans isomerization and the field-dependent exchange contribution of unknown origin described above were already suggested that the model-free assumption of independent motions of the NH-bond vector and the overall domain might not be valid.

Nevertheless, rotational diffusion, which is essential for model-free analysis of the data, was assessed by hydrodynamic calculations on the holo Langerin CRD structure (pdb entry: 3p5f, Feinberg et al. 2011). The analysis yielded an anisotropic rotational diffusion tensor with an effective rotational correlation time of 7.9 ns (**Table B-2**), which was lower than the values estimated from the R_2/R_1 ratios (vide supra). The divergence likely derives from the fact that the crystal structure used for the hydrodynamic calculations contains no C-terminal tag. Therefore, it is questionable whether the calculated diffusion model is suitable to describe the construct used for the relaxation measurements. Besides, anisotropic diffusion represents an additional difficulty for the model-free approach since it increases the parameter space of the model calculations.

Regardless of the above-mentioned concerns, model-free analysis was performed and the algorithm selected an anisotropic rotational diffusion model and optimized the model-free parameters together with the tensor parameters until convergence. However, closer inspection of the shape of the diffusion tensors for both holo and apo Langerin revealed a highly stretched tensor with an axis that did not reflect the protein symmetry (). Moreover, the model-free parameters S^2 and τ_e highly diverged from the initial estimates of the program that assumed independent correlation times τ_m for each residue. These observations hinted for

over-fitting of the data into an artificial minimum of the model space. Likewise, providing the algorithm with the diffusion tensor obtained from the hydrodynamic calculations as a starting point for optimization resulted in the same distorted diffusion model and increase in parameter values after only few rounds of iteration. Hence, the model-free approach was not applicable to the obtained dataset.

Taken together, relaxation measurements revealed motions on different timescales from fast bond vector fluctuation on the pico- to nanosecond in the small and long loop, over micro- to millisecond timescale motions affecting the entire domain as observed by an overall $\mathbf{B}_0$ field-dependency of both holo and apo form. However, the apo form shows low frequency motions absent in the holo form occurring on a faster timescale than the additional very slow cis/trans isomerization. Ligand-induced changes in conformational dynamics are indicative for an underlying allosteric mechanism (Boehr et al., 2009).

4.2.12 Langerin CRD exhibits pronounced intra-domain allostery

The binding of Ca^{2+} induced major chemical shift changes in residues remote from the binding site indicating the existence of an allosteric network (**Figure 4-6**). If such a network exists, it can be probed by introducing mutations at different sites that result in alike chemical shift perturbations as caused by ligand binding (Boehr et al., 2013). To this end, twelve single-residue mutants were designed probing different regions of the fold[16]. The choice of the mutation sites was guided by both the observation that the small loop was highly responsive to pH and Ca^{2+} binding and the desire to interrogate all parts of the domain. Thus, about half of the sites were located in the small loop or the adjacent beta strands. The remainder was distributed over the entire fold. To keep the core structure of the fold intact, I deterred from mutating bulky hydrophobic residues as well as the highly conserved cysteine residues (Ward et al., 2006).

4.2.13 Mutations in Langerin CRD affect distant residues within domain

To assess whether the mutants were folded and still bound Ca^{2+}, ^{1}H-^{15}N HSQC NMR spectra in the absence and presence of Ca^{2+} were recorded. The ten mutants that were soluble after refolding and purification (**Figure 4-12a**, **Figure B-7**) exhibited a ^{1}H-^{15}N HSQC NMR spectrum that indicated a similar fold to the WT (**Figure 4-12b**). Moreover, addition of Ca^{2+} resulted in chemical shift changes as observed in the WT indicating that Ca^{2+} binding was preserved. However, closer inspection of the peak distribution showed individual chemical

[16] Mutation sites were contrived by me. The generation of the mutants, protein expression and protein purification were performed by Elena Shanina under my supervision.

shifts for many residues in the mutants (**Figure 4-12c**). Especially residues in the β-strands and loops of the upper region of the fold experienced pronounced changes in both apo and holo forms. Although some residues experienced chemical shift changes along certain vectors whereas for other no clear pattern could be discerned (**Figure 4-12c**). Such shifts would indicate changes in the ensembles equilibrium populations that are already preformed in the WT (Boehr et al., 2009). Depending on the position of the mutant peaks relative to the WT bound and unbound species, the mutations could then be categorized as activating (holo-like) or deactivating (apo-like) (Nohaile et al., 1997; Selvaratnam et al., 2012). Notably, the two mutations, which awarded no functional protein (N205A and K320A), altered residues at the stem of the domain where the terminal β-strands interact indicating the importance of conserved interactions at this site for correct folding.

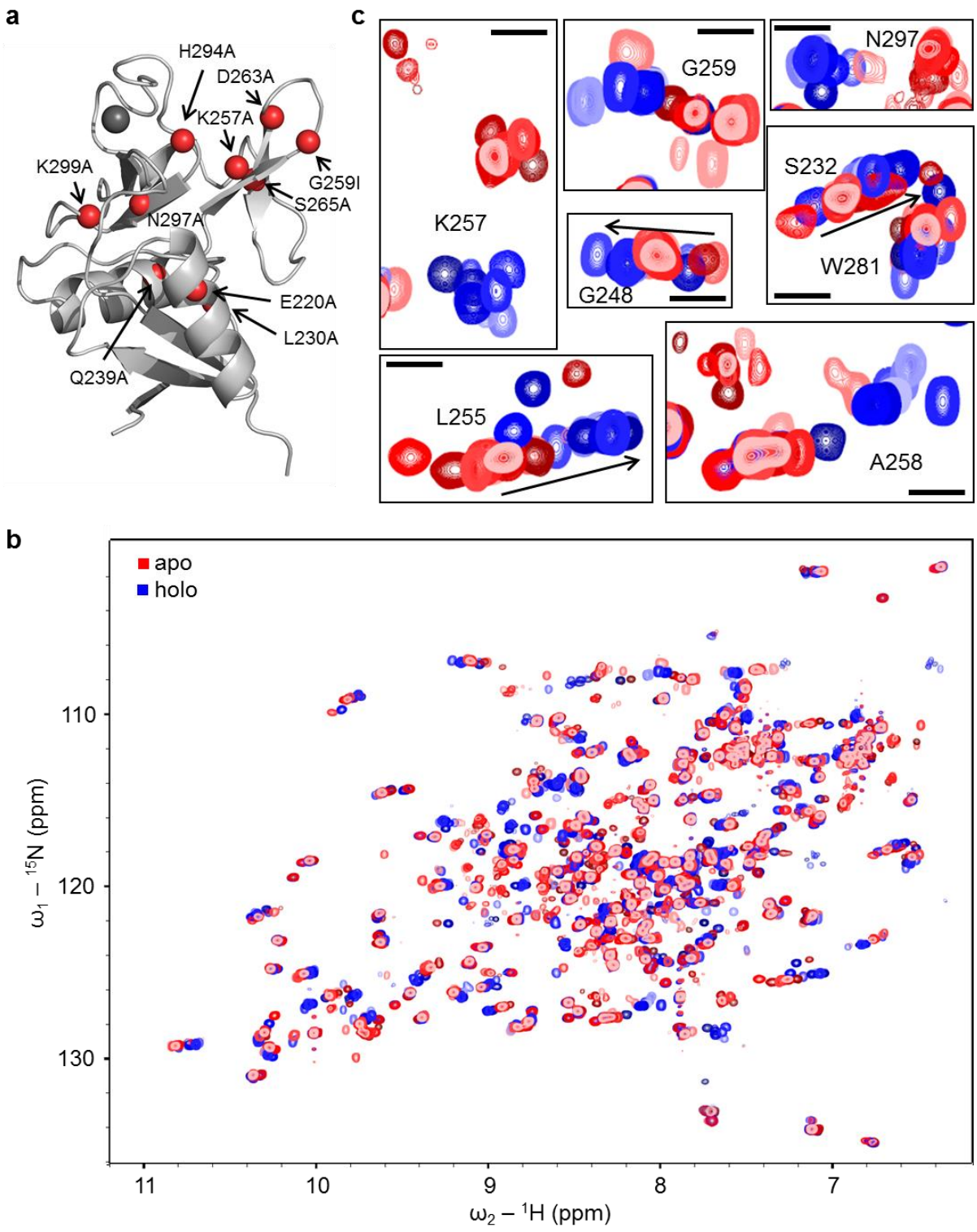


Figure 4-12: Probing Langerin CRD allosteric network by single point mutations. (**a**) Sites of the single-residue mutations that were folded correctly shown on cartoon representation of Langerin CRD structure (pdb entry: 3p5f, Feinberg et al., 2011). Backbone nitrogen atoms of mutated residues are shown as red spheres, Ca^{2+} is shown as black sphere. (**b**) Overlay of ^{1}H-^{15}N HSQC NMR spectra of the holo and apo forms (blue and red, respectively) of Langerin CRD WT and ten single residue mutants. All mutants were folded and bound Ca^{2+}. All spectra were recorded at pH 6, 298 K. Holo forms were generated by addition of 10 mM $CaCl_2$. (**c**) Closer inspection of resonances of residues in the short and long loop revealed pronounced chemical shift changes in both apo and holo form caused by mutation. Some of residues shift along the same vector upon mutation (indicated as black arrows) indicating shifting equilibrium populations.

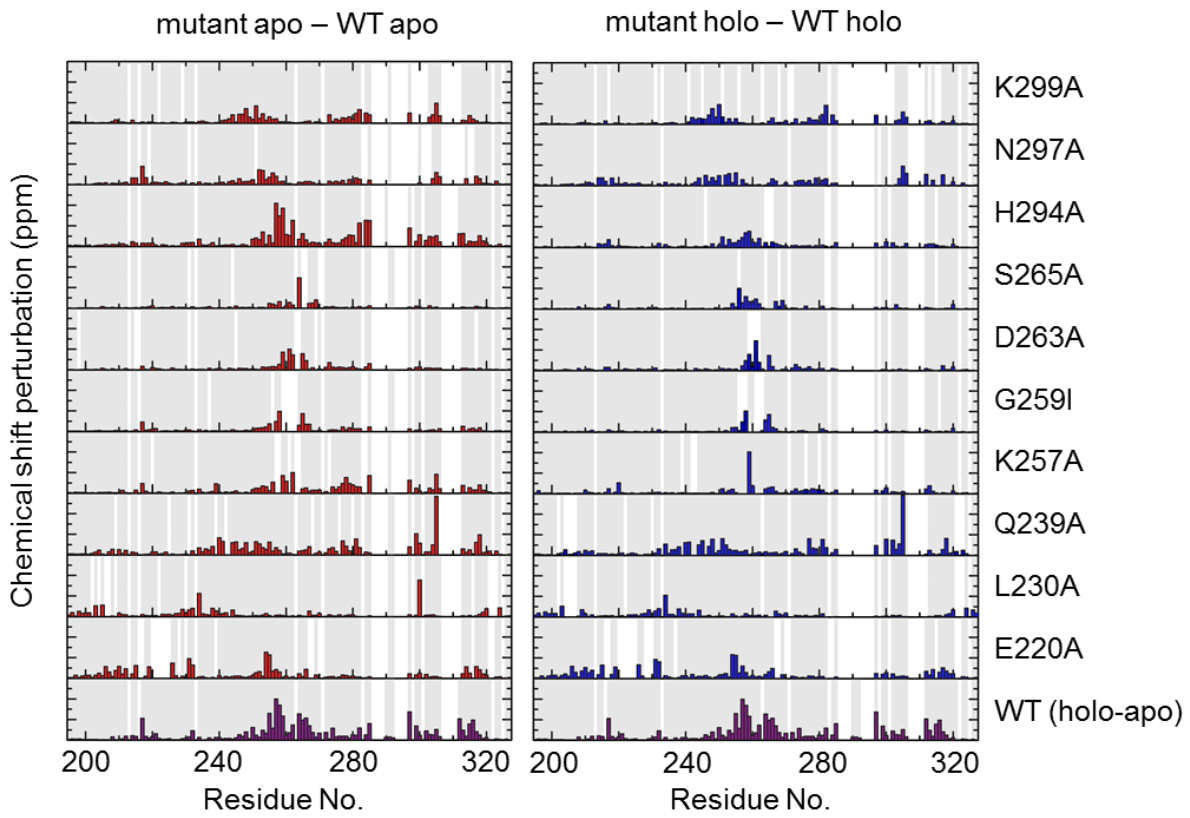


Figure 4-13: Single-residue mutations in Langerin CRD introduce pronounced chemical shift changes remote from mutation site. CSP of mutant apo and holo forms (left and right panel respectively) in respect to the respective WT form show a distinct pattern of induced chemical shift changes. Bottom row shows the WT binding to Ca^{2+} to compare the similarities of the induced CSP patterns. CSP scale ranges from 0 to 0.3 ppm in every graph. Grey bars indicate assigned residues in the tow respective spectra used for CSP determination. All data was collected at pH 6, 298 K.

4.2.14 Mutation of H294 and K257 differentially affect the holo and apo forms

To quantify the observed changes induced by the mutations and to assess whether they resemble the pattern observed for Ca^{2+} binding, the CPS of the mutant apo and holo forms with respect to the WT were plotted per residue (**Figure 4-13**). In the apo form, all mutations induced chemical shift perturbations in residues distributed over the entire sequence. However, differences could be observed both in the number of the affected residues and in the magnitude of the induced changes. Whereas mutations located in the center of the fold induced more widely spread changes, mutation of the small loop region resulted in rather local effects. Interestingly, K257A mutation also affected residues in the long loop region suggesting a communication between the two loops. Notably, H294A exhibited the most pronounced changes in the apo form. Although the affected residues were distributed over the entire domain, most prominent changes occurred in the small loop region and the long loop indicating a prominent role of H294 side chain in facilitating the communication between the two regions. Analysis of the size of both the apo and holo network confirmed that the residues of the small loop except K257A only induce minor changes. Performing the same analysis as

described above with the mutant holo forms in respect to the holo WT yielded the same picture for most mutants indicating that the network persisted independent of Ca^{2+} binding. Notably, H294A induced less variation from the WT in the holo form compared to the pronounced effect in the apo form. The same seems to be true for K257A suggesting that the two residues play a differential role in the apo and holo form of the WT, respectively. For the other mutants, Ca^{2+} binding had no sizeable effect on network size and identity.

Using a cutoff value of 0.02 ppm and excluding all residues within 5 Å of the mutation site provided numbers for differently sized networks of affected residues for each mutant in apo and holo form (**Table 4-3**). Moreover, identity with the WT Ca^{2+} interaction network varied and was overall low with the exception of H294A and G259I. Interestingly, identity increased in H294A holo form probably resulting from the lower number of affected residues, which were all located in the small loop region. Comparison of the network sizes in apo and holo showed for three mutants (H294A, K257A, G259I)[17] an about two-fold reduction of the number of affected residues in the holo form compared to the apo form. These findings imply that the mutations shift the conformational ensemble into a more holo-like state pointing towards a functional role of these residues in the Ca^{2+}-induced allosteric network.

Table 4-3: Allosteric networks defined for each mutant based on shifting residues upon Ca^{2+} binding (Ca^{2+} network) or shifted residues compared to the WT in apo and holo forms (apo and holo networks), respectively. All peak experiencing CSP greater than 0.02 ppm were counted. Overlap with the original network defined by Ca^{2+} binding to the WT is given in percent. The ratio of the apo and holo networks of each mutant is given to assess the sensitivity towards Ca^{2+} binding. All data was recorded at pH 6, 298 K.

Mutant	Ca^{2+} network	Overlap	apo network	Overlap	holo network	Overlap	Ratio apo /holo
E220A	36	89%	28	57%	31	58%	0.90
L230A	40	85%	13	15%	10	10%	1.30
Q239A	39	85%	33	67%	28	54%	1.18
K257A*	45	67%	23	65%	13	62%	1.77
G259I*	42	74%	10	80%	5	100%	2.00
D263A^{+}	41	73%	4	50%	4	75%	1.00
S265A*	38	74%	4	50%	8	75%	0.50
H294A^{+}	49	71%	39	74%	15	87%	2.60
N297A*	28	71%	18	61%	23	52%	0.78
K299A*	36	78%	22	50%	17	47%	1.29

*Residues are part of the Ca^{2+} interaction network of the WT.

$^{+}$Resonances of these residues are not assigned in the WT spectrum.

[17] The small number of affected residues by S265A mutation did not allow for any conclusions based on the observed increase in network size upon Ca^{2+} addition.

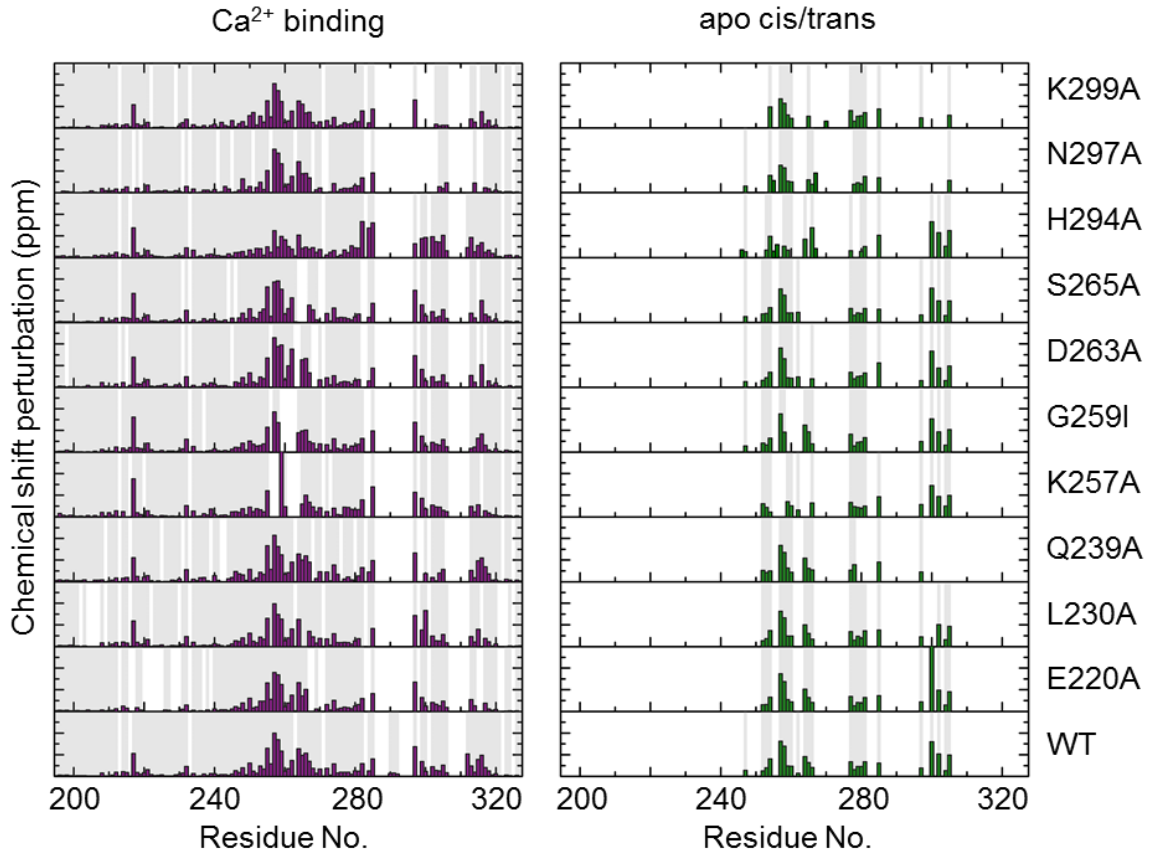


Figure 4-14: Single-residue mutations in Langerin CRD pertain Ca^{2+} interaction network and exhibit similar cis/trans pattern in apo. CSP of mutant peaks shifted upon addition Ca^{2+} and chemical shift differences between trans and cis state resonances in the apo form (left and right panel respectively) show a high conservation of both properties compared to the WT (bottom row). Only H294A exhibits noticeable differences in both properties. CSP scale ranges from 0 to 0.3 ppm in every graph. Grey bars indicate assigned residues in the tow respective spectra used for CSP determination. All data was collected at pH 6, 298 K.

4.2.15 The interaction network activated by Ca^{2+} binding is robust against mutations

To assess the robustness of the Ca^{2+} interaction network, chemical shift perturbations induced upon Ca^{2+} binding were plotted for each mutant (**Figure 4-14**) and the interaction networks were quantified using the same cutoff value as applied for the WT (**Table 4-3**). The analysis revealed a high conservation of the Ca^{2+} response networks both in size and identity. Mutations in the small and long loop region, however, shifted the network towards the region of the α3 helix that experienced greater chemical shift perturbations in these mutants. Moreover, the residues prior to the binding site (280-285) experienced strong chemical shift changes in H294A whereas the short loop exhibited smaller perturbations indicating a partial decoupling of the two loops. The same is observed in K257A although to lesser extent. These findings indicate a redundancy of allosteric interactions in the upper fold that can be perturbed by single mutations but not completely abolished (Boehr et al., 2013).

Based on the WT and mutant CSP observed upon Ca^{2+} binding, I could define a core network of 24 residues that were outside 5 Å of the respective mutation site and experienced chemical shift perturbation over 0.02 ppm both in the WT and in over 80% of the mutants (**Figure 4-15**). This core network comprises residues in the entire upper fold with highest conservations observed for residues in the proximity of the binding site and the small loop region evidencing a robust communication of the two regions.

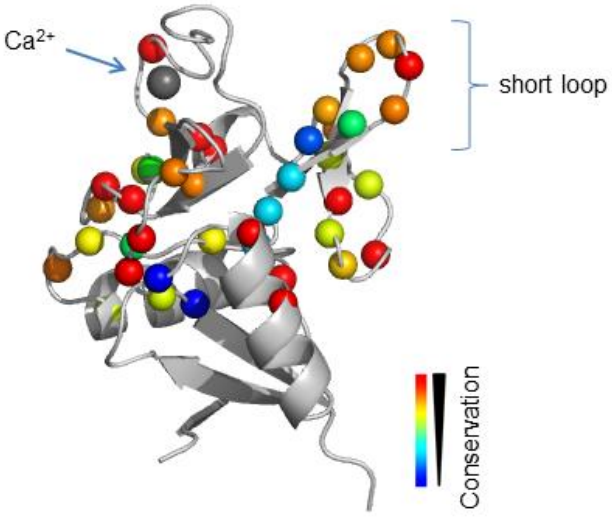


Figure 4-15: Langerin CRD harbors a robust allosteric network responsive to Ca^{2+} binding. Cartoon representation of Langerin CRD (pdb entry: 3p5f, Feinberg et al., 2011) with members of the Ca^{2+} interaction network shown as spheres. The color code represents the resistance against mutation with the highest conserved members of the network in red. The bound Ca^{2+} ion is shown as grey sphere. Residues in the vicinity of the binding site, in the short loop and the α2 helix are highly conserved members of the interaction network.

4.2.16 Mutation of H294 and K257 enhance Ca^{2+} binding

In the apo forms, quantitation of the peak intensities of trans and cis resonances, respectively, resulted in P286 prolyl cis/trans ratios within the error of the WT ratio of 75±10% (**Table B-3**). Moreover, the chemical shift differences between the two states were highly conserved indicating that mutations affect both states alike. Thus, the observed conformational changes in the mutants did not act on cis/trans isomerization of most mutants. Again, differences were observed for H294A that showed decreased chemical shift differences for the apo cis and trans resonance in the small loop region indicating that the states are more alike in the mutant than in the WT.

To assess whether the introduced mutations also alter the protein function, the dissociation constants of Ca^{2+} binding were determined for seven mutants by ^{1}H-^{15}N HSQC NMR titration experiments at pH 6 (**Table 4-4**). Intriguingly, only three mutants (H294A, K257A, K299A) induced changes in affinity outside the error margin of the WT. For these mutants, an increase

in affinity was observed with H294A exhibiting the strongest effect with an about 8-fold increase in affinity compared to the WT. Taken together, these findings suggest, that a robust allosteric network down-regulates Ca^{2+} binding in Langerin at endosomal pH.

Table 4-4: Binding affinities of Langerin CRD to Ca^{2+} at pH 6 and pH 7 obtained by ITC and NMR measurements. Error was estimated by goodness of fit if not indicated otherwise.

	***Kd* at pH 6 (µM)**		***Kd* at pH 7 (µM)**	
Mutant	**ITC**	**NMR**	**ITC**	**NMR**
WT	780 ± 160	620 ± 50*	105 ± 15*	160 ± 30
Q239A	-	640 ± 40	-	-
K257A	-	265 ± 30	-	-
D263A	-	620 ± 30	-	-
S265A	-	570 ± 30	-	-
H294A	125 ± 5	110 ± 10	35 ± 15	-
N297A	-	770 ± 30	-	-
K299A	-	210 ± 20	-	-

*Error given as standard deviation of three independent measurements.

4.2.17 H294 is a key player in the allosteric network and partially conveys the pH response in Ca^{2+} binding

Having observed that Ca^{2+} binding evoked allosteric changes in an extended network of residues and that altering the network by means of mutation partially affected the protein function, I became curious whether the pH sensitivity could be ascribed to a single residue within the network. In the case of the CTLR ASGPR, pH sensitivity is conveyed by a single histidine side chain in the vicinity of the Ca^{2+} binding site (Wragg and Drickamer, 1999). Langerin CRD only contains two histidine residues of which one (H229) is located in the lower part of the fold and not part of the Ca^{2+} interaction network. Based on the prominent role of H294 in the allosteric network and the increased Ca^{2+} affinity in the H294A mutant at low pH, I hypothesized that the residue could serve as the central pH switch governing Ca^{2+} binding.

In NMR titration experiments, the H294A mutant displayed increased peak intensities and additional peaks at both pH 6 and pH 7 compared to the WT. These peaks were found in the spectral region with a high incidence of asparagine and acidic residues and most likely belonged to the unassigned residues in the long loop (**Figure 4-16a**). This data suggests that removal of H294 side chain altered the dynamics of the loop region thereby shifting the exchange regime on the NMR timescale. Moreover, whereas at pH 6 both apo and holo forms of H294A mutant differed in chemical shift from the WT, at pH 7 the holo forms of WT and

H294A almost completely overlapped (**Figure 4-16b**). Notably, only I253 showed a chemical shift perturbation above the threshold of 0.02 ppm. At the same time, comparison of the holo forms of H294A at pH 6 and pH 7 showed that the magnitude of the chemical shift changes is reduced compared to the effect observed for the pH-response in the WT holo forms (vide supra). Consequently, the holo forms of H294A mutant at pH 6 and pH 7 are energetically more similar than the respective WT forms. These findings further evidence that the pH responsiveness of the allosteric network can be partially delineated to the side chain of H294.

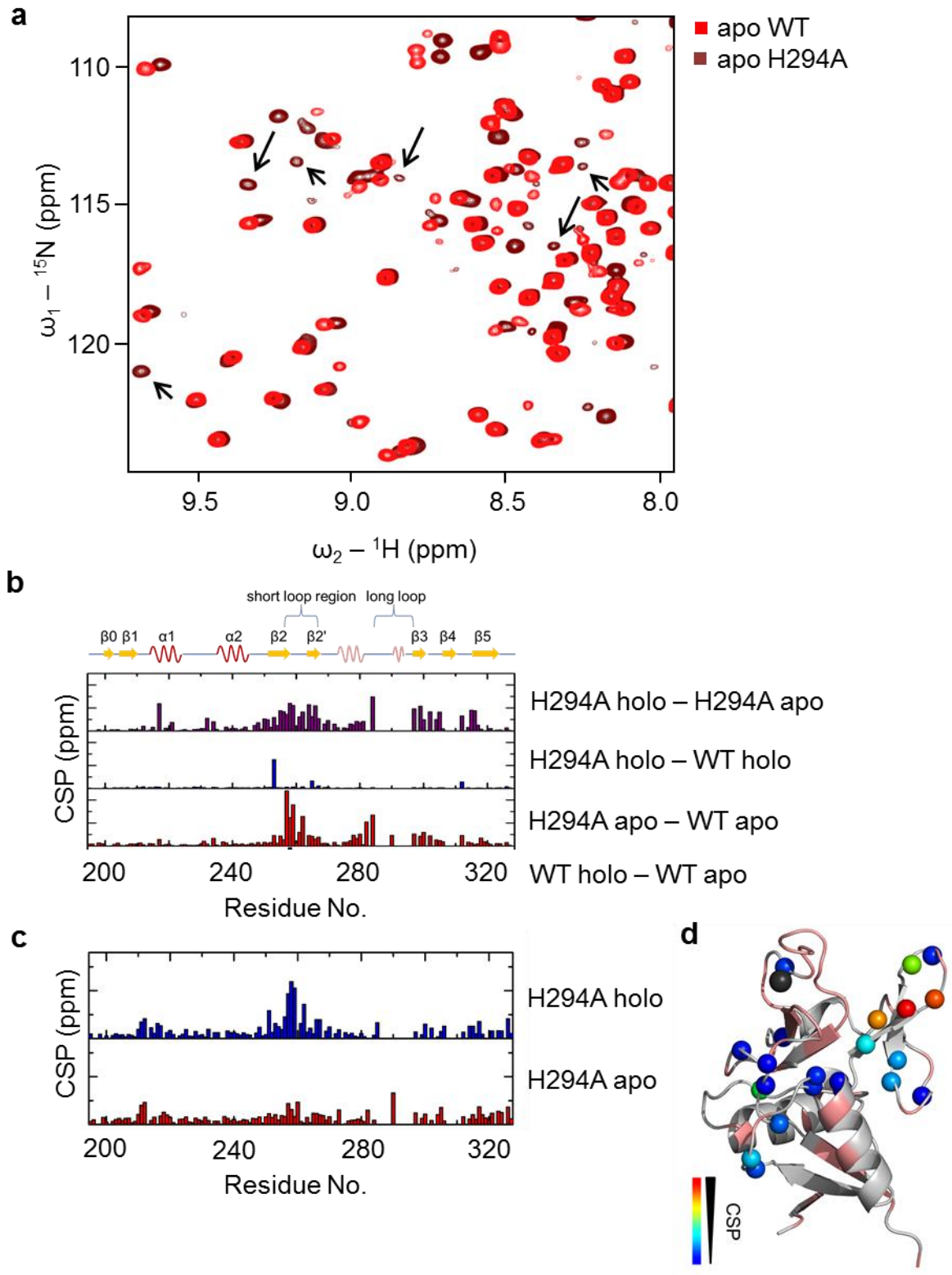


Figure 4-16: H294A mutant of Langerin CRD alters dynamics of apo form and decouples the long and short loop. (**a**) Overlay of ^{1}H-^{15}N HSQC NMR spectra of the apo forms of Langerin CRD WT and H294A at pH 6, 298 K. In the spectrum of H294A several additional peaks appear that could belong to the unassigned loop residues. (**b**) Chemical shift perturbations of WT and H294A at pH 7. While mutant and WT apo forms still differ, the holo forms are almost identical. (**c**) Chemical shift differences observed among the apo and holo forms of H294A at pH 6 and pH 7, respectively. The holo form showed a strong pH-dependent shift in the short loop region. Mjor ticks on the CSP axis represent 0.1 ppm. (**d**) Residues experiencing chemical shift perturbations upon change in pH in the holo form of H294A mutant are located majorly in the β2-strand. Cartoon representation of Langerin CRD (pdb entry: 3p5f, Feinberg et al., 2011) with heat map color-coded spheres of backbone nitrogens of residues according to the magnitude of the CSP >0.02 ppm.

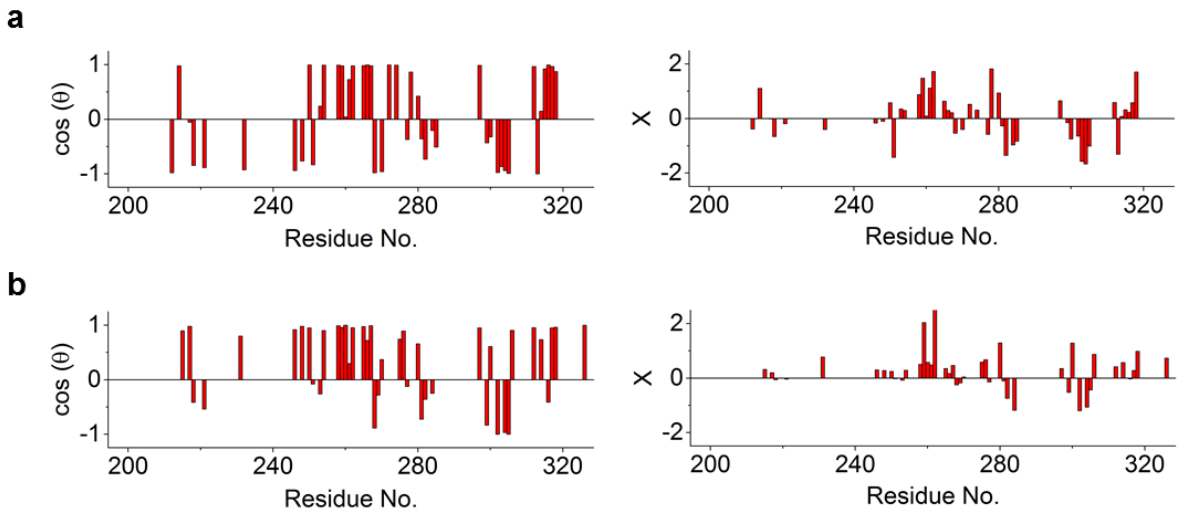


Figure 4-17: The pH dependent chemical shift changes in the holo form of H294A are more pronounced than in the WT. (**a**) CHESPA analysis of H294A apo at pH 6. Left panel shows cos(θ) of the mutant vectors. Most induced changes align very well with the reference vector indicated by |cos(θ)|≈1. Right panel shows the projection of the vector lengths (**b**) CHESPA analysis of H294A apo form at pH 7.

To gain more insight into the structural origin of the pronounced chemical shift perturbations observed in the H294A apo mutant, a chemical shift projection analysis was conducted (Selvaratnam et al., 2012). The CHESPA analysis of the apo mutant form at pH 6 revealed a set of activating shifts along the WT apo-holo axis for the short loop region, whereas the residues flanking the long loop and the Ca^{2+} binding site shifted in the opposite direction indicating an overall decoupling of the two loops (**Figure 4-16c**). CHESPA analysis of the apo CSP data collected at pH 7 conveyed a similar picture as at pH 6 for the short and long loop region (**Figure 4-16d**). Moreover, the number of activating shifts in other regions of the protein was higher at pH 7 than at pH 6 further evidencing that there is a pH response encoded in H294 that affects remote regions of the domain. Unfortunately, CHESPA analysis failed to provide a clear pattern for the other mutants (**Figure B-9**).

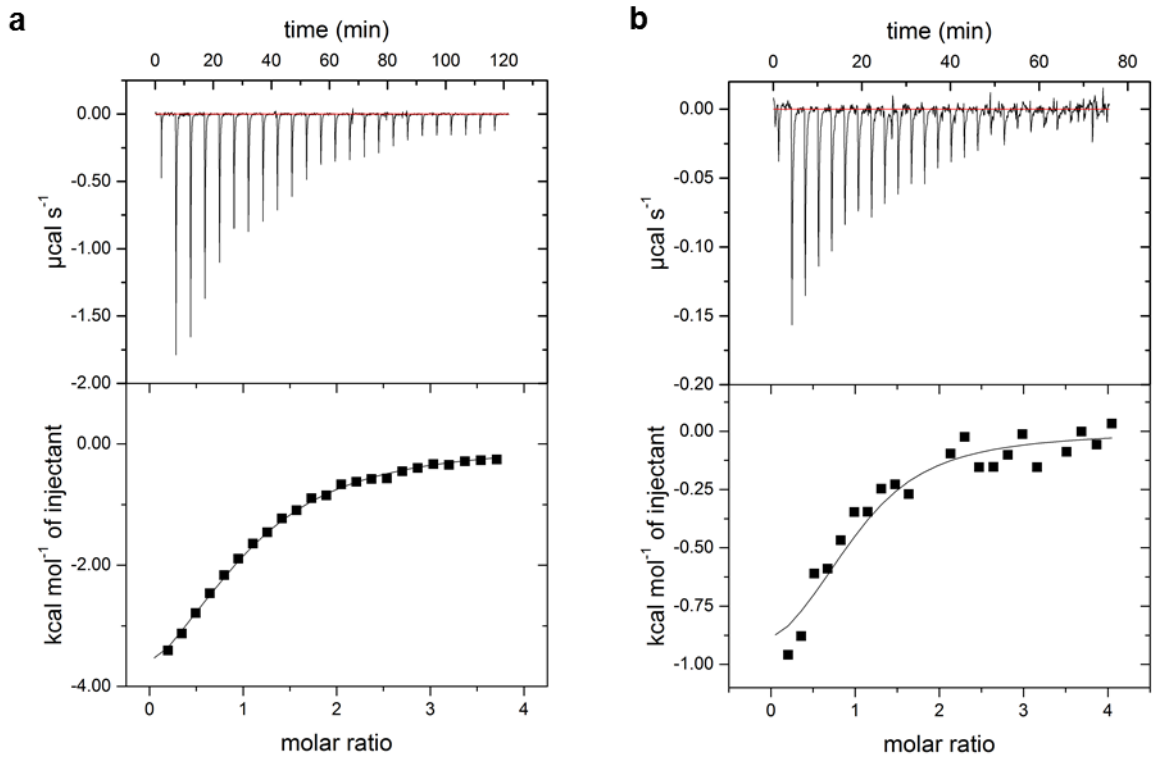


Figure 4-18: H294A binds Ca^{2+} with pH-dependent affinity. Isothermal titration calorimetry of Langerin CRD H294A mutant awarded a K_d of 126±5 µM at pH 6 (**a**) and 36±15 µM at pH 7 (**b**) using a one-set-of-sites binding model with a 1:1 stoichiometry. Measurements were performed at 298 K titrating $CaCl_2$ into the protein solution.

If protonation of H294 was solely responsible for the pH-dependent decrease of Ca^{2+} affinity observed in the WT, the Ca^{2+} affinity of the mutant should be unaffected by pH. To test this hypothesis, I determined the K_d of the H294A mutant by ITC at both pH 6 (126±5 µM) and pH 7 (36±15 µM) (**Figure 4-18**). The K_d at pH 6 agreed well with the value obtained from NMR titrations. Unfortunately, several attempts to obtain a K_d at pH 7 from ^{1}H-^{15}H HSQC NMR failed. Hence, Ca^{2+} affinity was still altered by changes in pH in the mutant ruling out the hypothesis that H294 side chain functions as the sole pH modulator in Langerin CRD. However, comparing the magnitudes of the pH-dependent changes reveals that the WT experiences a higher pH-dependent shift in affinity than the H294A mutant implying that the pH effect is reduced in the mutant.

Taken together, H294 plays an important role in coupling the dynamics of the long and short loop region. Moreover, it partially entails the pH responsiveness of Ca^{2+} binding though it serves not as the sole pH switch as observed in ASGPR (vide supra).

Table 4-5: Energy landscape model describing the pH dependence of Ca^{2+} binding in Langerin CRD. The model consists of three effects whose free energy contributions (ΔΔG) are defined by the WT and H294A mutant Ca^{2+} affinities at pH 6 and pH 7. The presence or the absence of the effects are indicated by 1 and 0, respectively. Experimental free binding energies were determined from ITC and NMR titrations based on eq. (3-02). The predicted values were calculated from the binding free energies of the WT at pH 6. Prediction of the free energies that were used to construct the model is redundant and therefore fields are colored grey.

Switch	ΔΔG (kJ mol^{-1})	WT pH 7	WT pH 6	H294A pH 7	H294A pH 6	K257A pH 7	K257A pH 6	K299A pH 6	Q239A pH 6
Loop-loop interactions	**2.7**	1	1	0	0	0	0	1	1
H294-mediated pH response	**1.9**	0	1	0	0	0	1	1	1
Charge close to Ca^{2+} site	**3.1**	0	1	0	1	0	1	0	1
ΔG_{exp} (kJ mol^{-1})		**-22.9**	**-17.8**	**-25.5**	**-22.4**	**N.D.**	**-20.6**	**-21.1**	**-18.3**
ΔG_{pred} (kJ mol^{-1})						***-25.5***	***-20.5***	***-21.0***	***-17.8***

4.2.18 A thermodynamic model explains changes in Ca^{2+} affinity

To dissect the contributions of the single components to the pH-dependent regulation of Ca^{2+} binding by the allosteric network, I constructed a free energy landscape model (**Table 4-5**), comprising three switches:

i) a pH-independent coupling of the long loop and short loop facilitated by physical contacts of the K257 and H294 side chains,

ii) a pH-dependent effect attributable to H294 H-bond and/or charge, and

iii) a pH-dependent charge introduced in the proximity of the Ca^{2+} cage.

Based on the ITC affinity data for the WT and the H294A mutant at pH 6 and pH 7, free energy changes (ΔΔG) were assigned to all three effects. To test our model, the binding free energies ΔG of three mutants (K257A, K299A and Q239A) were predicted and the outcome compared to the experimental data. Removing K257 side chain by mutation should abolish the physical interaction with H294 side chain but leave the other two switches untouched. By mutating K299, we removed a positive charge proximal to the Ca^{2+} cage. Therefore, at pH 6 the pH-dependent charge gain by protonation should be balanced by the mutation. The Q239A mutant served as control since as it is not part of the core network and far away from the binding site. Thus, it should have no major effect on Ca^{2+} affinity. Within an error of 0.5 kJ mol^{-1}, the model predicted the correct free energies for all three mutants. The model error is in the same range as the differences observed between affinity data determined by ITC

and NMR, respectively. The Ca^{2+} affinities of K257A and possibly also K299A and Q239A at pH 7 are useful controls to falsify the model. However, preliminary experimental attempts failed so far due to protein aggregation at pH 7. Nonetheless, the model satisfies the experimental data of three mutants and thus can be accepted until falsified by additional experimental insight.

The additional protonation site responsible for the major pH effect could be the Ca^{2+} binding cage itself: The presence of several negatively charged side chains in close proximity might induce an upward shift in pK_a of the carboxylic acid groups facilitating protonation of these sites (**Figure 4-19**). Indeed, pK_a shifts of up to 6.8 for glutamate side chains induced by neighboring negatively charged side chains have been described in the literature before (Davoodi et al., 1995). Although there are several computational approaches to derive pK_a values from the protein structure, the accuracy is still low and results vary highly depending on the employed algorithm (Alongi and Shields, 2010). For Langerin, four different algorithms applying both empirical and simulation approaches awarded pK_a values from 1.1 to 6.3 for the H294 side chain in the holo form (pdb entry: 3p5f, Feinberg et al., 2011) (**Table B-4**). Therefore, NMR pK_a titrations (Tollinger et al., 2002) or alternative experimental data are a necessary prerequisite to further support this hypothesis.

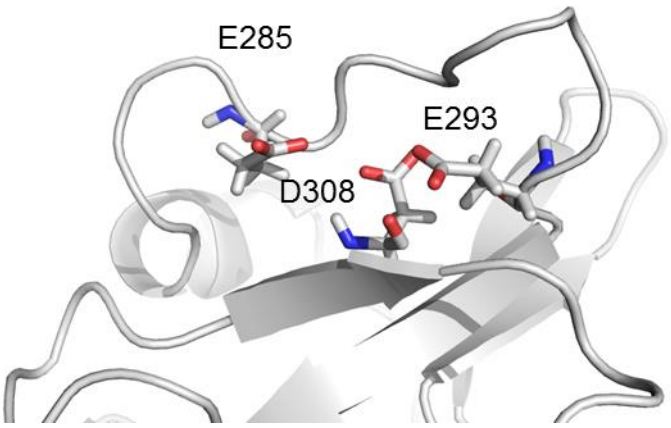


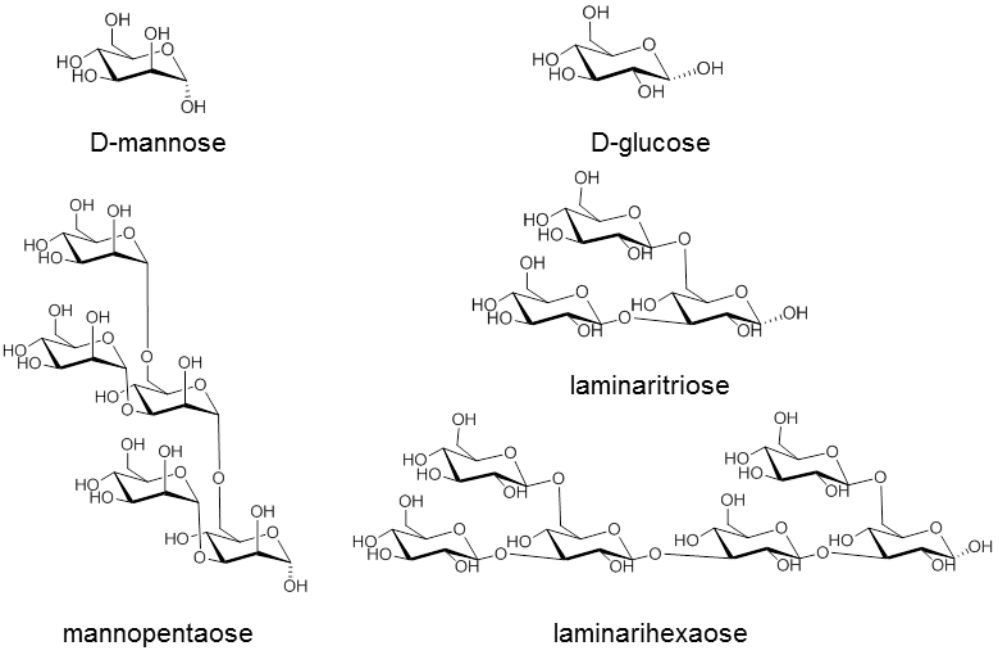
Figure 4-19: Acidic side chains are localized in close proximity in apo Langerin likely causing an increase in pK_a. The side chains of E285, E293, and D308 point towards the same direction in the apo form. Cartoon representation of Langerin with the protonable residues shown as sticks. Structure is a snapshot of a 2 µs MD simulation of apo Langerin performed by Stevan Aleksic, Freie Universität Berlin.

4.2.19 Carbohydrate binding to human Langerin does not activate the allosteric network

Previous literature reports on Langerin interacting with complex oligosaccharides observed by X-ray crystallography concluded that the interactions of the CRD with glycans are limited to the Ca^{2+} coordination of the terminal carbohydrate moiety and some hydrogen bonds of the this moiety with protein residues in the vicinity of the binding site (Feinberg et al., 2011). However, having observed allosteric mechanisms underlying the binding of Ca^{2+} which were invisible to X-ray crystallography, I became curious whether this would hold true for carbohydrate recognition as well. To assess whether carbohydrate binding would induce similar allosteric changes as observed for Ca^{2+} binding, the interactions of a series of mannose- and glucose-type carbohydrates with Langerin CRD were investigated by biomolecular NMR. The investigated ligands (

Scheme 1) were chosen based on previously published X-ray structures and commercial availability. In addition, mannan was used to assess whether polysaccharide binding would be observable by protein NMR methodologies.

Scheme 1: Structures of the carbohydrate ligands that were employed in this study. Reducing end hydroxyl groups are given in their preferred conformation in solution, respectively.



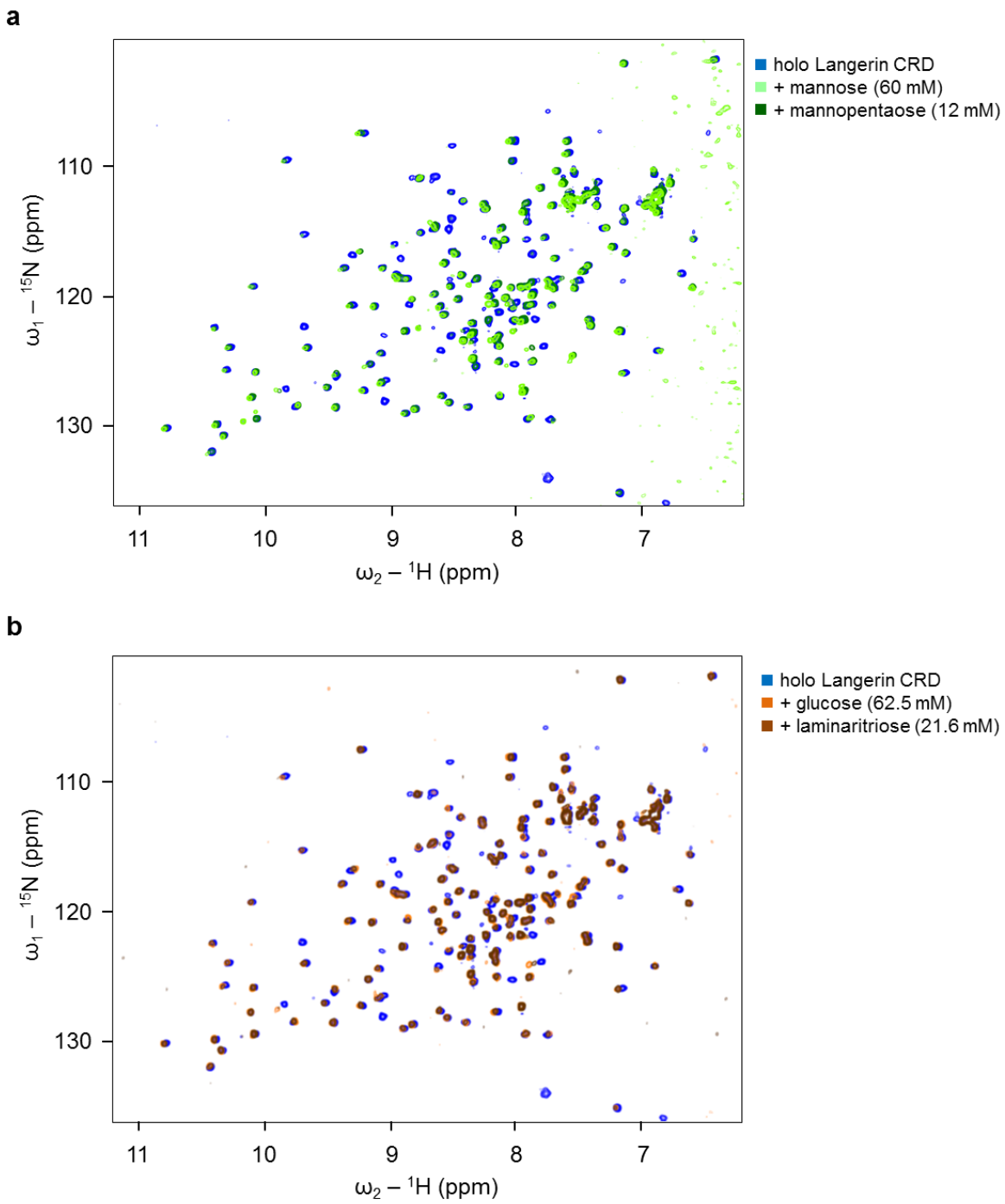


Figure 4-20: Mannose and glucose-based carbohydrate ligands induce similar chemical shift changes. Overlay of ^{1}H-^{15}N-HSQC NMR spectra of holo Langerin CRD interacting (**a**) with mannose-based carbohydrate ligands and (**b**) glucose based carbohydrate ligands. Ligand concentrations are according to the legend. Spectra were recorded in presence of 10 mM $CaCl_2$ at pH 6, 298 K.

The ^{1}H-^{15}N HSQC NMR spectra were recorded at pH 6 to resemble the situation of ligand release in the early endosome (**Figure 4-20**). The overlay of the spectra revealed that only few peaks were shifted upon addition of the carbohydrate ligand. No major differences were observed for the mannose and glucose type ligands. Mannan, however, induced changes in almost all residues likely due to overall changes in the proteins chemical environment (**Figure B-11**). In contrast, laminarihexaose showed no chemical shift perturbations up to its solubility limit (5 mM). Therefore, both glycans were excluded from further analysis.

To assess which parts of the domain were affected by glycan binding, the CSP were plotted per residue. The overall changes were low in magnitude (<0.1 ppm) and confined to the binding site area in the long loop, and the small loop between the β3 and β4 strands with K299 and A300 as the most prominently affected residue. (**Figure 4-21a**). The same affected residues were observed for all four glycans. However, glucose and laminaritriose induced less pronounced shifts likely due to decreased affinity of the ligands. Moreover, the unspecific background during titration was higher than observed for Ca^{2+} binding (vide supra) likely due to changes in viscosity caused by high sugar concentrations. Plotting the CSP on the structure of Langerin CRD confirmed the findings from the crystal structures (**Figure 4-21b**): The residues in the in the vicinity of K299 experienced chemical shift perturbations which can be attributed to the fact that K299 side chain is in close proximity to the bound carbohydrate (Feinberg et al., 2011). Moreover, E285 undergoes some changes likely due to electronic effects transmitted via the Ca^{2+} ion in the binding site. Therefore, growing complexity or switching the class of the ligand showed no additional binding sites and revealed no extended allosteric changes in the domain. Unfortunately, the protein contact site of the penultimate carbohydrate moiety of (1→2)α branched high mannose structures, A289 (Ibid.), was not visible in the NMR due to the unfavorable exchange regime.

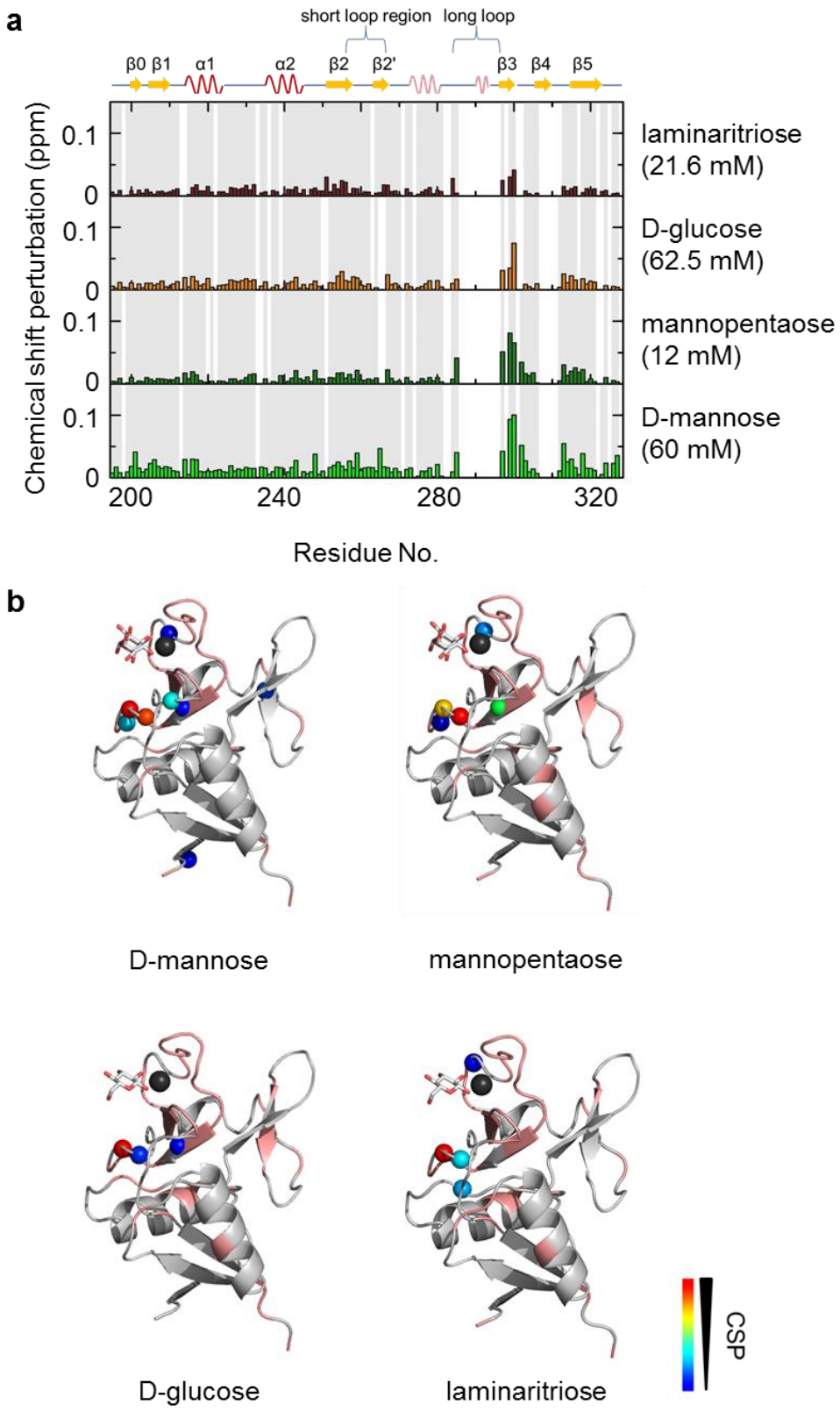


Figure 4-21: Langerin CRD interactions with complex glycan ligands does not reveal a second binding site or induce any allosteric effects. (**a**) Chemical shift perturbations measured for Langerin CRD interacting with the respective carbohydrate. Grey bars indicate assigned residues. (**b**) Langerin CRD interacting with glycan ligands. Residues experiencing CSP are labeled on Langerin structure (pdb entries: 3p5f (mannose-based ligands) and 3p5h (glucose-based ligands), Feinberg et al., 2011). Calcium is labeled as dark grey sphere and the ligand from the crystal structure is shown in sticks for orientation. Color code follows is a heat map with a lower cutoff value of 0.4-0.6 of the maximum CSP value.

Table 4-6: Dissociation constants of carbohydrate ligands interacting with Langerin CRD obtained by ^{1}H-^{15}N HSQC NMR titration experiments. Spectra were recorded at pH 6, 298 K.

Inhibitor	K_d (mM)
D-mannose	11.4±1.4
Mannopentaose	6.8±1.0
D-glucose	21.4±4.2
Laminaritriose	12.1±3.0

4.2.20 Increased size and complexity of the carbohydrate ligands does not result in an increased affinity

To assess whether increasing size and complexity of the carbohydrate ligand would result in increased affinity as has been observed in other CTRLs before (Probert et al., 2013; Talaga et al., 2014), the dissociation constants were determined by ^{1}H-^{15}N HSQC NMR titration experiments (**Figure B-12**). The resulting dissociation constants were in the lower mM range (**Table 4-3**). However, a linear increase of the titration curve indicated a non-specific interaction with the protein at higher ligand concentrations. Therefore, the observed dissociation might be underestimated due to the fitting procedure that aims to accommodate the linear part of the curve. Indeed, the dissociation constants obtained for mannose and glucose were about threefold higher than previously published K_i values from a plate-based competition assay (Stambach and Taylor, 2003). Moreover, the affinity for mannose was about twofold decreased compared to affinities obtained from NMR assays at pH 7 for both Langerin CRD and ECD applied in our lab (Wamhoff et al., 2016) indicating the possibility of a pH dependence of binding to the carbohydrate itself (vide supra).

Taken together, the interactions with both mannose- and glucose-based ligands were confined to the proximity of the binding site and did not activate the allosteric network. Moreover, both carbohydrate types interacted with the same residues indicating a similar orientation in the binding site.

4.2.21 Conclusion

Human Langerin is a trimeric recycling endocytic receptor mainly found on Langerhans cells in the skin where it serves as pattern recognition receptor in innate host-defense. Langerin binds its cargo in a Ca^{2+} dependent manner and releases it in the early endosome where a rapid decrease of both the pH and the Ca^{2+} concentration occurs. Unlike the closely related CTLR DC-SIGN, Langerin remains in the multimeric form upon acidification, therefore excluding a potential loss of multivalency as the main release mechanism (Stambach and Taylor, 2003; Tabarani et al., 2009). Moreover, here I show that binding to monovalent ligands is not under inter-domain allosteric regulation, which is also in contrast to DC-SIGN (Coombs et al., 2010). Whereas carbohydrate affinity is only little affected by changes in pH, binding to the essential cofactor Ca^{2+} is strongly pH dependent (**Figure 4-4**). Hence, Ca^{2+} affinity and pH were singled out as the two main contributors of ligand release guiding the subsequent investigations of the interactions of the carbohydrate recognition domain with Ca^{2+} by biomolecular NMR spectroscopy.

The present work reveals that Ca^{2+} binding activates a sizeable intra-domain allosteric network that downregulates Ca^{2+} affinity in a pH responsive manner. This network comprises the majority of residues in the upper region of the fold, specifically the long and short loop as well as the adjacent β sheets 2, 2', 3 and 4 and the connecting loops between them (**Figure 4-6**). The network was successfully probed by single point mutants that induced changes in chemical environment of distant residues partially resembling the Ca^{2+} binding thereby providing further evidence for the existence of such an allosteric network (Boehr et al., 2013). Moreover, the network is robust since long-range communication upon Ca^{2+} binding was not abrogated in the mutants (**Figure 4-15**). Surprisingly, mutation of the two side chains that physically connect the short and long loop by H-bonds and van-der-Waals interactions, K257 and H294, resulted in a pronounced increase of Ca^{2+} affinity (**Table 4-4**) suggesting that the coupling of the two loops might by an essential feature of the allosteric mechanism controlling the loop dynamics.

Intriguingly, the canonic Ca^{2+} binding sites 1 and 3, which serve as allosteric regulation sites in other CTLRs (Drickamer, 1993b; Ng and Weis, 1998), are occupied by K257 and a histidine side chain H294, respectively (**Figure B-10**). Notably, the two residues interact closely in most crystal structures mediating a contact between the short and the long loop. In the trimeric crystal structure, this short loop is flipped towards the long loop indicating the existence of alternate conformations (**Figure 1-3b**). The dynamics of such alternate conformations could elicit regulatory function on ligand binding (Kern and Zuiderweg, 2003).

These observations are further substantiated by the herein observed strong effects of mutating these residues on both the allosteric network (**Figure 4-13**, **Table 4-3**) and the dynamics of the long loop (**Figure 4-16**).

To further support this notion, Ca^{2+} binding influences the dynamics of large parts of the domain which is an additional indicator for allostery (**Figure 4-8**). Whereas the apo form undergoes very slow cis/trans isomerization of the P286 prolyl bond and additional motions on the micro- to millisecond timescale in different regions of the domain, the holo form is locked in the cis conformation and also shows reduced mobility on the micro- to millisecond timescale evidenced by a change in the overall correlation times (**Table 4-2**). Notably, both the long and short loop show increased mobility on the pico- to nanosecond timescale in both the holo and the apo form. These findings suggest that receptor dynamics are closely associated with ligand binding and release. The receptor is rather rigid under extracellular conditions and its mobility is restricted to the coupled movement of the long and short loop, which likely resides on the nanosecond timescale as inferred from NMR measurements. Upon endocytosis and hence Ca^{2+} release, additional micro- to millisecond dynamics appear, likely promoting ligand release or preventing rebinding (**Table 4-2**). Similar observations have been reported for other CTLRs (Poget et al., 2001; Probert et al., 2014).

In parallel to this work, molecular dynamics simulations were conducted by Stevan Aleksic. By applying mutual information theory to interrogate correlated motions, which are indicative for allosteric interactions, a vast network of interacting amino acid residues was identified (Keller et al., 2010). Moreover, MD analysis identified the physical nature of the interactions among the members of the network and revealed that the observed chemical shift changes originate from subtle backbone rearrangements communicated mainly via interactions of the side chains. Furthermore, the simulations showed a correlated motion of the long and short loop on a 10 ns timescale which was not altered by addition of Ca^{2+} indicating a coupling of the loops. However, mutation of H294 resulted in a loss of this correlated movement reflecting the missing link between the two loops after removal of H294 side chain. Thus, the MD data supports the findings presented herein and hence provide further evidence for the existence of an allosteric regulatory mechanism.

To further dissect the pH dependent mechanism for Ca^{2+} release, H294 was identified as a potential regulator due to its role as central player of the network and the observations reported for ASGPR, a hepatic CTLR, where the protonation state of H202 was previously reported to play such role (Feinberg et al., 2000). To further delineate the pH-dependent and -independent contributions to Ca^{2+} affinity, a thermodynamic model was constructed based on

the affinities of the WT and mutant data at both pH 6 and 7 (**Table 4-5**). Surprisingly, H294 only partially contributes to the pH dependence of the Ca^{2+} affinity as could be seen from the persisting pH dependence of the H294A mutant. Other potential pH sensors are the members of the Ca^{2+} cage E285, E293, D308 for which remarkable high pK_a values were calculated (**Figure 4-19**, **Table B-4**). In particular, the former two residues show similarity to an aspartyl dyad with elevated pK_a values similar to residues found in BACE-1 and HIV-1 protease (Yamazaki et al., 1994; Kim et al., 2015).

In addition, the findings presented herein allows for dismissing several potential functions of the allosteric network in Langerin. First, there were no substantial differences in the cis/trans prolyl bond isomerization state of P286 in any of the prepared mutants at pH 6 (**Table B-3**), indicating that the cis/trans isomerization equilibrium is not under control of the allosteric network. Furthermore, the spatial restriction of the network to the upper part of the fold suggests that the network is not involved in transmitting information from the binding site through the protein to the intracellular potential signaling domain (Drickamer and Taylor, 2015). Finally, as several carbohydrate binding proteins have been proposed to compensate loss of conformational entropy upon ligand binding by increasing conformational entropy at remote sites of the protein, the herein presented model does not allow for any conclusive remarks at this state (MacRaild et al., 2007; Diehl et al., 2010; Nesmelova et al., 2010). However, additional data might substantiate such hypothesis but is beyond the scope of this work.

Allosteric networks have been observed in various classes of proteins (Henzler-Wildman and Kern, 2007). Although allostery has been appreciated as a fundamental principle underlying many dynamic processes essential for protein function (Kern and Zuiderweg, 2003; Swain and Gierasch, 2006; Smock and Gierasch, 2009; Motlagh et al., 2014), no detailed studies on any member of the CTLR family have been reported to date. The few available NMR studies on CTLRs draw a blurry picture: Whereas in ASGPR, which binds two Ca^{2+} ions with high cooperativity at physiological conditions in a pH-dependent manner, chemical shift perturbations were highly confined to the binding site regions (Onizuka et al., 2012), for other CTLRs like DC-SIGNR, a myeloid endocytic CTLR that binds three Ca^{2+} ions in a pH-dependent fashion, changes in both chemical shift and line widths have been reported although not discussed in context of intra-domain allostery (Probert et al., 2014). The same is true for tetranectin that undergoes structure and dynamics changes upon Ca^{2+} binding (Nielbo et al., 2004). Intriguingly, the identified allosteric side in Langerin is located where allosteric regulation in other CTLRs is conferred by accessory Ca^{2+} binding, which is likely ancestrally

encoded into the fold (Zelensky and Gready, 2005). Current investigations in our laboratory show that in DC-SIGN also a series of residues experience chemical shift changes upon addition of Ca^{2+} suggesting that the allosteric mechanism is also present in at least some CTLRs with more than one Ca^{2+} binding site. These findings show that an ancient fold can adopt different strategies to fulfill similar functions.

To assess whether the allosteric network is also involved in carbohydrate recognition, the interactions of Langerin with mannose and glucose-based carbohydrate ligands resembling deletion sequences of known Langerin ligands man9 from gp120 and laminarin, representative β-glucan found in brown algae, were investigated by protein-observed NMR. Both mannose and glucose-based ligands interact with the same residues of the protein only varying in magnitude of the induced chemical shift changes that can likely be attributed to the different degree of saturation (Williamson, 2013), a notion supported by the affinity data (**Table 4-6**). It has been suggested that in DC-SIGN signaling is altered based on the interactions with the non-reducing-end carbohydrate moiety (Gringhuis et al., 2009). The findings of this work, however, imply that in Langerin there is unlikely any structural regulation of possible different signaling outcomes of mannose- and glucose-based carbohydrates induced by different monosaccharide units (**Figure 4-20**). Moreover, only the terminal monosaccharide units at the non-reducing end appeared to interact with the protein in the resolved regions while the remainder pointed away. This is supported by the rather small affinity increase observed for the complex oligosaccharides in respect to the monosaccharides which is likely due to re-binding of equivalent sites and consequently an entropic contribution not originating from addition contacts. Notably, the same was observed for the related CTLR DC-SIGNR that also interacts with high-mannose structures from gp120 (Probert et al., 2013). Based on these findings, selectivity is likely controlled by a kinetics- or affinity-based binding mechanism that regulates the tightness of the multivalent interactions of several CRDs with the multivalent ligand. Hence, specificity for high-mannose and high-glucose structures is generated on a higher level of complexity either by the geometry of the domain presentation on the trimer or spatial-temporal organization of the receptor on the cell surface as observed for other CTLRs (Cambi and Figdor, 2005).

Intriguingly, Langerin interactions with the self-antigen heparin involve a second binding site independent from the primary Ca^{2+} site (Chabrol et al., 2012; Munoz-Garcia et al., 2015). The absence of this secondary site for the fungal and viral antigens investigated in this work might entail a regulatory function for the receptor to discern self from non-self. However, cell biological studies are required to draw this conclusion.

Additional experiments could further substantiate the changes in dynamics induced by Ca^{2+} binding. First, relaxation data for the WT collected at pH 7 would reveal whether the timescales of the observed motions are altered by pH in a fashion that promotes Ca^{2+} release. As a change in dynamics upon removal of H294 side chain was already indicated by changes in lineshape in the ^{1}H-^{15}N HSQC NMR spectra, relaxation measurements of H294A mutant would provide an additional wealth of insight into the nature of these changes. Finally, a K257A/H294A double mutant would allow for further interrogating the coupled loop motions. As mentioned above, ^{1}H-^{15}N HSQC-TROSY NMR measurements of the trimeric ECD awarded no interpretable spectra thereby ruling further NMR studies out. Alternatively, electron paramagnetic resonance (EPR) and time-resolved fluorescence spectroscopy provide powerful means to interrogate inter-domain cooperativity in multivalent binding. However, initial trials to install a paramagnetic ion in the Ca^{2+} binding site, which would provide a powerful probe for both afore mentioned techniques, failed. On cellular level, our laboratory is currently developing cell-based assays to investigate the spatial-temporal organization of Langerin in uptake-competent cells. However, the assay was not available in the timeframe of this work.

4.3 Differential recognition of complex glycans by the human and the murine Langerin

Human and murine Langerin have been object of several studies aiming to elucidate the structure and function of CTLRs. Although similar function has been assumed for the two homologs and their interchangeable treatment in the literature (Drickamer and Taylor, 2015), so far there are no functional studies comparing ligand recognition of the two receptors. To close this gap, I aimed to elucidate the glycan specificity of human and murine Langerin by means of biophysical ligand-interaction studies and glycan microarrays.

4.3.1 Murine and human Langerin show similar affinity for simple mannose ligands, but differ in recognition of a fungal mannan polysaccharide.

To compare the functional homology of murine and human Langerin, the recombinant trimeric extracellular domain (ECD) of both proteins were employed to investigate the affinity towards simple mannose-type structures, a commonly reported ligand for Langerin (Stambach and Taylor, 2003). Isothermal titration calorimetry (ITC) measurements revealed a dissociation constant of 7.0 ± 0.2 mM for the murine homolog, which is identical with the observed affinity of 5.6 ± 1.7 mM for the human Langerin (**Figure 4-22A**). To further substantiate the similarity of the mannose recognition, binding epitope information of the simple disaccharide ligand α-methyl-(α-1,2)-mannobioside **1** was obtained by STD NMR for both homologs (**Figure 4-22B**, **Figure B-13**, **Figure B-14**). Notably, the binding epitopes for both were identical: The non-reducing end mannose receives most saturation with the highest effects on protons H2', H3', and H4'. These data are very well in agreement with the canonic binding mode of simple glycans to EPN motif carrying CTLRs (Drickamer, 1992) and the crystal structure of human Langerin binding to mannobiose (Feinberg et al., 2011). Moreover, the H5 received a strong STD effect when interacting with either of the two homologs, which is in agreement with the crystal structure of the human CRD that revealed a close packing of C5 to A289 if the terminal mannose is coordinated to the Ca^{2+} (Feinberg et al., 2011). Although the two homologs recognize the same epitope, the STD built-up curves of the murine Langerin had a higher magnitude of the magnetization transfer indicating slower off-rates than in the human Langerin.

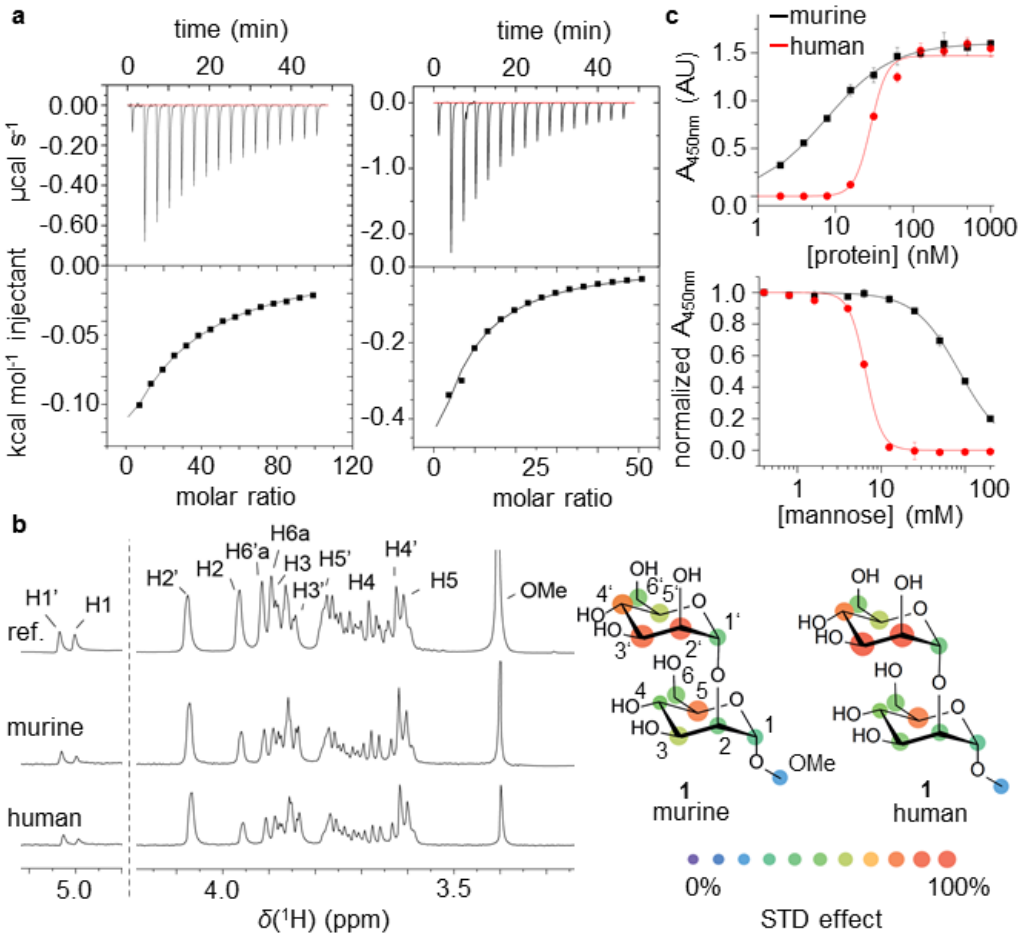


Figure 4-22: Murine and human Langerin recognize simple ligands with similar affinity and binding epitopes but highly differ in recognition of complex carbohydrates. (**a**) Isothermal titration calorimetry of murine Langerin ECD awarded a K_d of 7.0±0.2 mM fitting the data to a one-set-of-binding-sites model. (**b**) STD NMR spectra of man2 in presence of murine or human Langerin. Resonances are assigned in the reference spectrum. STD spectra shown were recorded with a saturation transfer delay of 4 s. Epitopes of **1** were mapped for both proteins based on STD built-up curves (**Figure B-14**). Both proteins recognize an identical epitope. (**c**) Plate-based assays revealed differential behavior of both trimers towards mannan-coated surface. EC_{50} values for plate binding were 7.4±0.2 nM and 28.5±1.4 nM with Hill slopes of 1.0±0.1 and 4.1±0.3 for murine and human Langerin, respectively (upper panel). IC_{50} values for mannose competition were 85.2±1.3 mM and 6.5±0.1 mM with Hill slopes of 1.6±0.1 and 4.5±0.5 for murine and human Langerin, respectively (lower panel). Data are represented as mean ± STDEV.

To assess whether the observed similarities for interactions of small mannose-type ligands are conserved for larger polysaccharide structures, the analysis was expanded to mannan, a fungal branched (1→2)α;(1→6)α D-mannose polysaccharide, by comparison of the binding of the homologs in a plate-binding assay (**Figure 4-22C**). In the present assay, the ECD was titrated to the immobilized mannan resulting in an effective concentration of binding (EC_{50}) of 7.4±0.2 nM and 28.5±1.4 nM for the murine and human Langerin, respectively. Besides this avidity difference, both proteins show distinct cooperative binding behavior. Whereas the murine Langerin association to the plate fits to a dose-response model with a Hill coefficient of 1.0±0.1, suggesting no cooperativity, the human Langerin experiences a higher degree of positive cooperativity with a Hill coefficient of 4.1±0.3. To exclude effects resulting from

varying lectin activity, the assay was converted to a competition experiment: D-Mannose was titrated into a solution of Langerin, keeping the protein at a fixed concentration, yielding inhibitory concentrations (IC_{50}) of 85.2±1.3 mM and 6.5±0.1 mM with Hill coefficients of 1.6±0.1 and 4.5±0.5 for murine and human Langerin, respectively. Consequently, since both proteins share the same affinity to mannose these data strongly suggest that murine Langerin has a higher avidity to the mannose polysaccharide than its human homolog resulting from differential multivalent interactions.

Overall, the data suggest that murine Langerin has a higher avidity for the mannan polysaccharide compared to the murine homolog, while both share the same affinity for monovalent mannose and the same epitope for a dimannoside. However, the murine Langerin ECD caused overall stronger STD effects indicating decreased off-rates of the ligand which might explain the observed avidity increase.

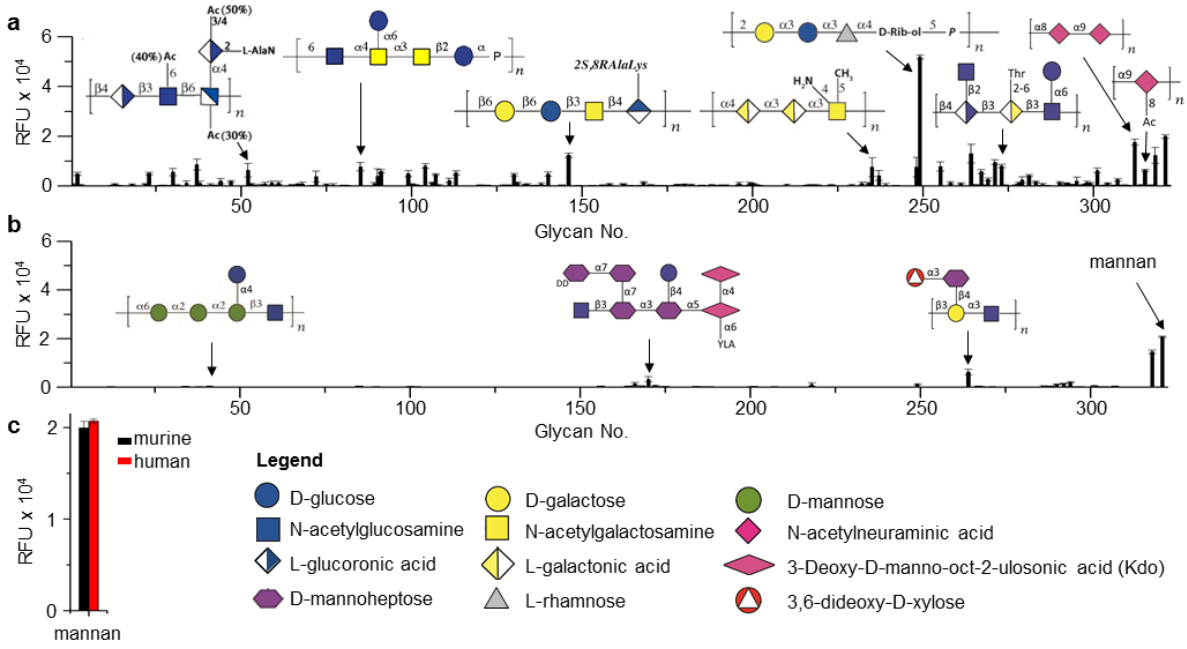


Figure 4-23: Murine and human Langerin recognize defined subsets of bacterial polysaccharides on printed glycan microarrays. (**a**) Murine Langerin binding to microbial glycan array. To highlight the diversity of the hits, representative structures are shown. (**b**) Human Langerin binding to the same array. (**c**) Comparison of relative fluorescence units (RFU) of murine and human Langerin binding to yeast mannan. Complete microarray data are found in supplementary table S3. Data are represented as mean ± S.E.M.

4.3.2 Human and murine Langerin differentially recognize bacterial glycans

To investigate whether the differences observed in mannan binding translate to other complex polysaccharides, murine and human Langerin were tested against a microbial glycan microarray comprising more than 300 bacterial glycans (**Figure 4-23**, **Table A-3**) (Stowell et al., 2014)[18]. Under saturating amounts of lectin, both homologs exhibited very similar fluorescence intensities when binding to mannan, which was used as positive control (**Figure 4-23c**). Remarkably, the recognition of bacterial glycans highly differed between the two homologs (**Figure 4-23a-b**). Whereas murine Langerin bound 56 of the 313 immobilized glycan structures using a cut-off value of 1,000 RFU, human Langerin only bound ten glycan structures above the cut-off value of which six were also recognized by the murine homolog. Murine Langerin recognized a diverse set of species of the bacterial genii presented on the array. Notably, murine Langerin showed the highest fluorescence intensities binding to *S. pneumoniae* serotype 6B with a signal almost threefold above the fluorescence intensities observed for mannan. The only original hits of human Langerin were two *Y. pestis* antigens, *S. flexneri* 6a, and *E. coli* 112ab. Therefore, the full analysis of all microbial glycans suggests a highly diverging scope of epitopes being recognized by these two homologous lectins, except for the fungal polysaccharide mannan . Notably, some previously reported bacterial glycan structures binding to human Langerin were not identified by this analysis: *E. coli* 0106, *E. coli* 180/C3, and *S. boydii* B10 (Feinberg et al., 2011).

[18] The measurements were performed by Ryan McBride under supervision of Prof. James Paulson at The Scripps Research Institute, La Jolla, CA, USA. The FITC-labeled protein constructs were prepared by me. The data was analyzed jointly by Ryan McBride, my supervisor Dr. Christoph Rademacher, and me.

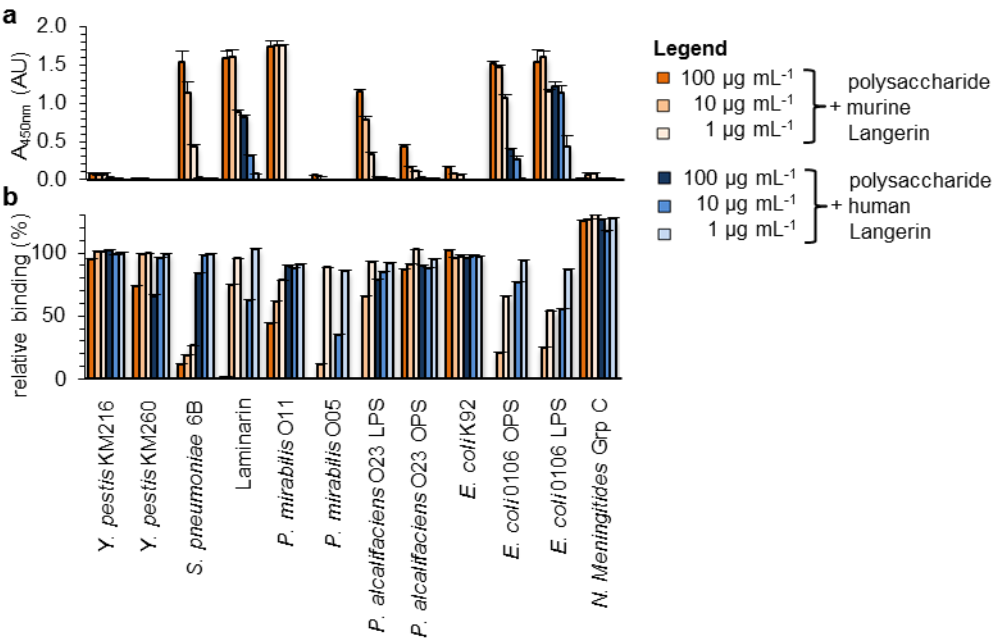


Figure 4-24: Murine and human Langerin recognize bacterial glycans differentially in plate-based assays. (**a**) Direct binding of murine (orange) and human (blue) Langerin to immobilized polysaccharides on microtiter plates (upper panel). (**b**) Competition assay with polysaccharides titrated to compete with binding of both proteins to mannan-coated microtiter plate (lower panel). Glycan concentrations with the corresponding protein in the assay are given in the legend. Data are represented as mean ± S.D.M.

To validate the identified target structures from the array, nine epidemiologically relevant and structurally diverse hits were chosen to investigate the differences of the interactions in more detail (). The purified lipopolysaccharide (LPS) and some O polysaccharide (OPS) structures[19] were tested in a plate-based, direct binding assay (**Figure 4-24a**)[20]. Murine Langerin exhibited affinity towards seven of the bacterial polysaccharides and against laminarin, a yeast (1→3)β glucan polymer that served as positive control as it binds to human Langerin (de Jong et al., 2010). The signal intensities highly varied from saturation to merely above background. However, in all but one case a concentration-dependent response was observed. Only *P. mirabilis* O11 awarded a saturated signal at all tested protein concentrations. In contrast, human Langerin only showed high signal intensities for *Y. pestis* KM260, and reduced signals for laminarin and *S. pneumoniae* 6B.

Because immobilization efficiencies of the polysaccharides on the surfaces of the microtiter plates are poorly accessible and may vary highly between different structures, the assay was reversed to a competition format. In this format, binding of the protein to the immobilized

[19] Natural microbial polysaccharides were generously provided by Dr. Yuriy Knirel, Zelinsky Institute of Organic Chemistry, Moscow, Russia.
[20] Measurements were performed by Henrik Schmidt under my supervision.

mannan was competed with the polysaccharides in solution in different concentrations (**Figure 4-24b**), thereby allowing for a better quantitative assessment of the interactions. Hence, binding to murine Langerin was confirmed for six of the structures and laminarin as observed by a dose-dependent inhibition of binding to the mannan plate. Although weak signals for *P. alcalifaciens* O23 OPS were detected in the direct binding assay, no competition was observed consequently excluding this structure as false positive. On the other hand, *P. mirabilis* O05 showed a dose-dependent competition indicating that in the direct binding assay the immobilization of the polysaccharide failed. Human Langerin exhibited inhibition of binding for *Y. pestis* KM260, *P. mirabilis* O05 and *E. coli* O106 LPS and OPS structures as well as to a positive control laminarin. These findings suggest that the competition assay possesses a higher sensitivity compared to the direct binding format that failed to identify *E. coli* O106 as ligand. Notably, this antigen has been reported previously as a ligand of human Langerin (Feinberg et al., 2011). The polysialic acids of *E. coli* K92 and *N. meningitides* failed to produce signals both in the direct binding assay and the competition assay and were therefore excluded. Sialic acids are often identified as false positives in glycan microarray screenings due to the tendency of the moiety to hydrolyze. Finally, direct binding to *P. mirabilis* O05 could probably not be detected due to insufficient immobilization of the glycan on the microtiter plate. Overall, the plate-based assays verified the glycan array results for seven glycans and identified four false-positive hits.

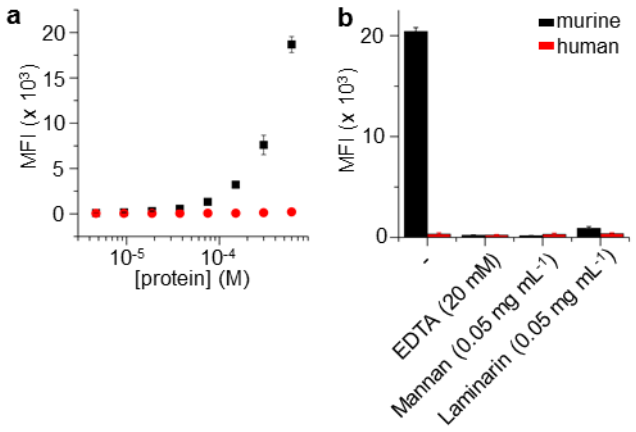


Figure 4-25: Human and murine Langerin bind E.coli 0106 OPS differentially. (**a**) Binding of FITC-labeled murine and human Langerin to *E. coli* 0106 investigated by flow cytometry (n=3). Murine Langerin exhibits stronger avidity for the bacteria than human Langerin. Raw data is shown in **Figure B-16**. (**b**) Competition with known inhibitors confirmed binding specificity to bacteria.

4.3.3 Human and murine Langerin bind to E.coli O106 with differential avidity

To assess whether the results from the microarray and plate-based assays could be translated to a higher level of complexity, direct binding of fluorescently labeled murine and human Langerin to heat-inactivated *E. coli* O106 was investigated[21]. Both proteins bound to the bacteria as observed by flow cytometry (**Figure 4-25**,). However, murine Langerin exhibited a stronger response indicating a higher avidity of binding in accordance with the plate data for both *E. coli* O106 antigens. Moreover, competition with the Ca^{2+} chelator EDTA and known glycan ligands of Langerin abolished binding indicating specific interactions of the protein constructs with the bacterial polysaccharides. Notably, murine Langerin binding was not inhibited by the β-glucan curdlan, a known ligand of human Langerin, thereby identifying another difference in binding specificity (**Figure 4-25c**). In sum, elevated levels of murine Langerin binding to bacteria compared to the human homolog provide additional line of evidence that inter-species difference is profound and can be translated to direct microbial recognition.

4.3.4 Langerin amino acid sequence is highly conserved among mammalian species with the exception of the extended binding site

To assess whether sequence variations among the mammalian Langerin homologs might entail. To this end, a multiple sequence alignment on 28 mammalian Langerin amino acid sequences found in the UniprotKB was perfomed (**Figure B-17**) and the residue conservation score calculated (Valdar, 2002) (**Figure 4-26a**). The residues of the secondary structure elements, with exception of the β4 strand, are almost identical in all structures. Surprisingly, the only major deviations in the sequence were found adjacent to the binding site, which is highly conserved, in the region comprising the interaction site for the sulfate group (K299 and K313) and stretching from there towards the center of the trimer. Notably, the highly variable residues are part of a surface facing the shallow cavity in the center of the trimer where a larger polysaccharide could be accommodated. Moreover, A289 which was implicated with additional interactions with high-mannose structures (Feinberg et al., 2011) is not conserved in the other mammalian Langerin homologs possibly altering specificity for this glycan class. Taken together, the evolutionary conservation analysis revealed that the region adjacent to the binding site and facing towards the trimerization cavity is highly variable likely influencing the receptor's specificity for complex glycans (**Figure 4-26b**).

[21] Experiments were designed, performed, and analyzed by Jessica Schulze and Henrik Schmidt. Bacteria were provided by Dr. Wolfgang Rabsch, Robert Koch Institute, Wernigerode.

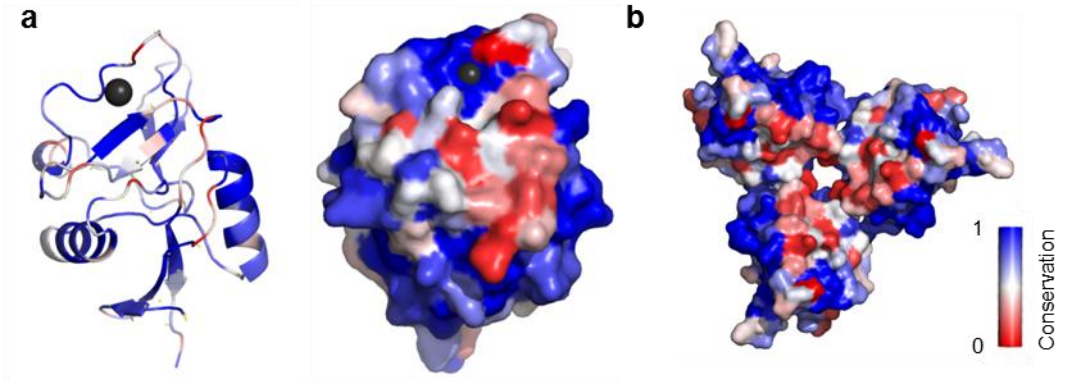


Figure 4-26: Langerin sequence is highly conserved among mammalian species with exception of a probable secondary interaction interface. (**a**) Residue conservation scores in Langerin derived from 28 mammalian homologs mapped as heat map on human Langerin structure shown as cartoon (left) and surface view (right) The surface view shows a highly variable region in the vicinity of the Ca^{2+} site forming a potential secondary interaction site for complex glycans. (**b**) Surface representation of trimeric human Langerin color-coded as in (A). The highly variable region faces towards the shallow groove in center of the trimer potentially forming a secondary interaction site for complex glycans.

4.3.5 Conclusion

Human and murine Langerin were compared side-by-side regarding their carbohydrate ligand recognition. While monovalent, simple mannose ligands were recognized almost identical, significant difference between the two homologous receptors were observed for the recognition of complex polysaccharides (**Figure 4-22**). Both binding avidities and specificities differ highly between the two homologs suggesting that the interaction with large polysaccharides is not determined by the presence of simple molecular patterns per se, but by their arrangement in the polysaccharide. Hence, the molecular details of this differential behavior cannot be understood by analysis of the monovalent interaction and must be embedded in the overall architecture of the ECD consequently questioning the practice to predict interactions with relevant polysaccharide ligands from binding data for simple glycans. Moreover, the differences in cooperativity of binding are even more difficult to address since they are most likely an effect of inter-domain cooperativity within the trimer or might result from receptor clustering. The ITC results show that in the case of small ligands in solution, receptor clustering can be excluded since this could easily be detected by dramatic alterations in the binding curve shape. Inter-domain cooperativity could be investigated by incorporation of spin labels or fluorescent probes. However, this would be far beyond of the scope of this work.

This differential behavior became dramatically apparent when the two proteins were tested on a microbial glycan array (**Figure 4-23**). While the murine homolog binds a large set of structures with strong signal intensities, the human homolog only recognizes a small set of structurally related glycans. In fact, subtle changes in the binding site may result in the

creation or blocking of additional weak interaction site. This effect has been described for other CTLRs before (Guo et al., 2004; Pederson et al., 2014) and might serve as explanation for diverging specificity here. Especially in the case of weak multivalent interactions, subtle changes in subsites can radically skew binding equilibria. Only recently, the crystal structure of murine Langerin has been solved in our laboratory and the changes that were observed in the vicinity of the binding site could be involved in the altered specificity. Specifically, the side chains of K299 and K313, which have been shown to play a role in the binding of 6-sulfated galactoses (Tateno et al., 2010; Feinberg et al., 2011), are replaced by arginine and asparagine residues in the murine homolog. An evolutionary analysis revealed that these residues are highly variable among all mammalian Langerin homologs and form part of an extended stretch of residues facing the shallow cavity of the trimer (**Figure 4-26**). Notably, this region has also been suggested before as an extended binding site for heparin in human Langerin (Munoz-Garcia et al., 2015). Overall, these data provide insight into how changes in secondary binding sites as well as in the inter-domain communication combine into significantly altered polysaccharide specificities that cannot be inferred from the analysis of simple glycans.

In the case of Langerin, a recent account identified both positive and balancing selection on inter- and intra-species level (Forni et al., 2014). The need of fast adaptation of receptor specificity might be explained by the high evolutionary pressure on PRRs to coevolve with the microbial threat thereby adapting to the hosts ecological niche. Along these lines, it was pointed out that CTLRs involved in pathogen recognition exhibit higher interspecies diversity in binding specificity than the CTLRs specialized to recognize self-antigens (Drickamer and Taylor, 2015). Another indicator for such an adaptive evolutionary behavior of Langerin differing in mice and men is the divergence in the expression pattern between the two mammals. Human Langerin^{+} APCs are majorly encountered in the skin and mucosa, where the evolutionary pressure might have driven specificity towards opportunistic fungi, mycobacteria like *M. leprea* and viruses invading via the mucosa similar to HIV. Conversely, murine Langerin is also found in APC subsets of the blood, lung, and gut where it encounters a very different set of pathogens such as *S. pneumoniae* in the lung and *Providencia* and *Proteus* genii in the gut. The low evolutionary conservation of surface exposed amino acids is in line with these findings and highlights the interspecies differences.

Although their basic biophysical properties are well conserved, the herein presented findings on the diverging specificity together with the previously known differential expression pattern suggest that human and murine Langerin also differ on a functional level. Especially in the

context of biomedical research, where mouse models serve for proof-of-principle of therapeutic approaches, already minor differences might impede translation to the human system (Mestas and Hughes, 2004). Further functional assays using native dendritic cells and the elucidation of the signaling pathway are needed to shed more light on the biology of this receptor. Targeted delivery approaches addressing the receptor might prove as useful tools to elucidate the biological function in both systems.

5 Final Conclusion and Perspectives

Glycan recognition by glycan-binding proteins is a fundamental process of life. In innate immunity, the recognition of glycan structures on pathogen surfaces by pattern recognition receptors can induce pathogen uptake and result in immune cell activation. The CTLR Langerin serves as an uptake receptor for viral, bacterial, and fungal pathogens, which are subsequently processed for antigen presentation to T cells.

The aim of this dissertation was to ascertain the underlying structural mechanism of ligand binding and release as well as ligand specificity of the CTLR Langerin using biomolecular NMR methodologies and isothermal titration calorimetry as well as glycan array screenings, plate-based assays, and direct binding assays to heat-killed bacteria. Hence, increasing levels of complexity from monovalent to multivalent interactions on both receptor and ligand site were interrogated providing novel insight on the molecular foundations of these interactions.

A detailed interrogation of the interaction of Langerin with Ca^{2+}, its essential co-factor of binding, revealed an intra-domain allosteric mechanism that acts to down-regulate Ca^{2+} affinity in response to pH likely promoting ligand release in the early endosome. The underlying robust allosteric network possibly fulfills the same function as the allosteric Ca^{2+} sites in other CTLRs. In addition, the apo form of human Langerin exists in two distinct conformations and experiences dynamics ranging from very slow cis/trans prolyl bond isomerization on the second timescale over slow motions on the micro- to millisecond timescale to fast motions on pico- to nanosecond timescale (**Figure 5-1**). In the holo form, only the fast motions are preserved that likely reflect the correlated movement of two loops. The coupling of the two loops is an essential contributor to the allosteric control mechanism. To date, no intra-domain allosteric ligand binding mechanism has been described to this extent for a member of the myeloid C-type lectin receptor family and the findings presented herein might guide future drug-design campaigns for this important receptor family (Aretz et al., 2014).

Intriguingly, the interactions of mannose- and glucose-based oligosaccharide were not governed by intra-domain allostery. Mono- and oligosaccharide ligands induced changes in the same, limited number of protein residues in the proximity of the binding site thereby confirming previously reported crystal structure data (Feinberg et al., 2011).

Finally, the assessment of the glycan specificity of the murine and human homologs of Langerin showed that while simple glycans are bound with the same affinity and in the same binding mode, complex glycans are recognized with both differential avidity and specificity.

This discrepancy is attributable to the interplay of inter-domain cooperativity and subtle changes in the amino acid composition of the binding site, both resulting from a high evolutionary pressure imposed by pathogenic threat. These differences alongside with the altered expression pattern of the receptor on different immune cell subsets shed more light on the evolution of innate immunity receptors. These findings have important implications for the use of murine and human infection models to study human disease.

To conclude, Ca^{2+} binding induces intra-domain allosteric structural and dynamics changes that retain regulatory functions whereas specificity for complex carbohydrates is generated on a higher level of complexity by inter-domain cooperativity and multivalent receptor presentation. These findings suggest that there exists a robust allosteric release mechanism utilizing the expelling of Ca^{2+} from the binding site to facilitate glycan release. The subsequent slow structural rearrangements of the apo form reduce re-binding of the ligand. Glycan specificity, however, is subject to subtle changes in the amino acid composition on the receptor surface thereby allowing fast evolutionary adaptation to a rapidly evolving pathogen threat. Hence, the example of Langerin elegantly demonstrates how the immune system utilizes an ancient fold for a conserved function as uptake receptor while adapting quickly to the pathogen threat faced by the host in his respective ecological niche.

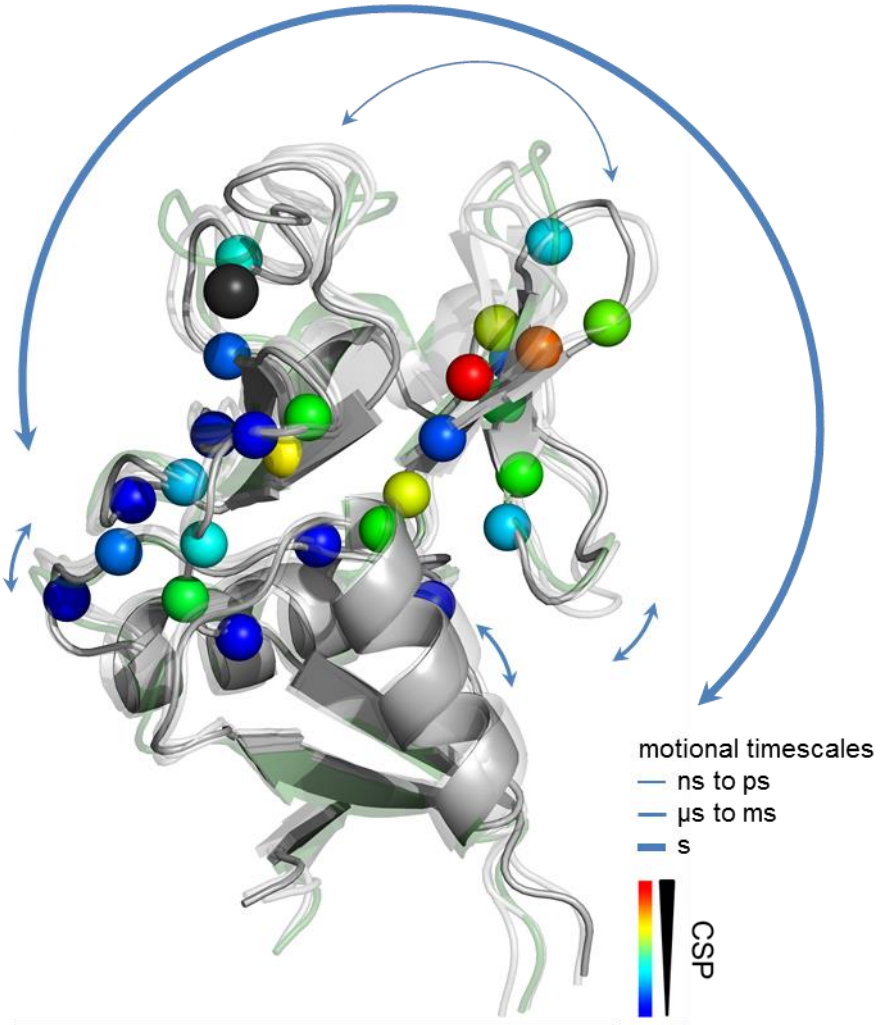


Figure 5-1: Human apo Langerin is highly dynamic on all timescales and undergoes intra-domain allosteric changes upon addition of Ca^{2+}. The apo form of Langerin CRD experiences motions on all timescales: Fast pico- to nanosecond motions affect the short and long loop region. Slower micro- to millisecond motions additionally affect the α-helices. Very slow motions on the s timescale induced by the cis/trans isomerization affect the entire upper loop. The arrows indicated the affected regions with the line width as an indicator of the motional regime. An intra-domain allosteric network of connected residues conveys information from the binding site in the long loop to the short loop. Binding of Ca^{2+} causes allosteric changes in an extended network of residues majorly affecting the short loop region. Cartoon representation of the cis holo form (pdb entry 3p5f, Feinberg et al., 2011) In solid grey, three snapshots from an MD simulation of the cis apo form are shown in transparent grey. One MD snapshot of the trans apo form is shown in transparent green. Color-coded spheres are backbone nitrogens of residues that experience significant chemical shifts perturbations upon addition of Ca^{2+} (heat map for CSP>0.02 ppm). Ca^{2+} is shown as black sphere.

6 References

Abragam, A. (1961). The principles of nuclear magnetism (Oxford: Clarendon Press).

Ahmad, N., Gabius, H.J., Andre, S., Kaltner, H., Sabesan, S., Roy, R., Liu, B.C., Macaluso, F., and Brewer, C.F. (2004a). Galectin-3 precipitates as a pentamer with synthetic multivalent carbohydrates and forms heterogeneous cross-linked complexes. J Biol Chem *279*, 10841-10847.

Ahmad, N., Gabius, H.J., Sabesan, S., Oscarson, S., and Brewer, C.F. (2004b). Thermodynamic binding studies of bivalent oligosaccharides to galectin-1, galectin-3, and the carbohydrate recognition domain of galectin-3. Glycobiology *14*, 817-825.

Akimoto, M., Selvaratnam, R., McNicholl, E.T., Verma, G., Taylor, S.S., and Melacini, G. (2013). Signaling through dynamic linkers as revealed by PKA. Proc Natl Acad Sci USA *110*, 14231-14236.

Akira, S., Uematsu, S., and Takeuchi, O. (2006). Pathogen recognition and innate immunity. Cell *124*, 783-801.

Alongi, K.S., and Shields, G.C. (2010). Chapter 8 - Theoretical Calculations of Acid Dissociation Constants: A Review Article. In Annual Reports in Computational Chemistry, A.W. Ralph, ed. (Elsevier), pp. 113-138.

Andreotti, A.H. (2003). Native state proline isomerization: An intrinsic molecular switch. Biochemistry *42*, 9515-9524.

Aretz, J., Wamhoff, E.-C., Hanske, J., Heymann, D., and Rademacher, C. (2014). Computational and experimental prediction of human C-type lectin receptor druggability. Front Immunol *5*, 323.

Berg, J.M., Tymoczko, J.L., and Stryer, L. (2007). Biochemistry, 6th edn (New York: W.H. Freeman).

Bider, M.D., Wahlberg, J.M., Kammerer, R.A., and Spiess, M. (1996). The oligomerization domain of the asialoglycoprotein receptor preferentially forms 2:2 heterotetramers in vitro. J Biol Chem *271*, 31996-32001.

Binder, F.P.C., Lemme, K., Preston, R.C., and Ernst, B. (2012). Sialyl Lewisx: A "Pre-Organized Water Oligomer"? Angew Chem Int Edit *51*, 7327-7331.

Bishop, J.R., and Gagneux, P. (2007). Evolution of carbohydrate antigens - microbial forces shaping host glycomes? Glycobiology *17*, 23r-34r.

Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M.E., Alvarez, R., Bryan, M.C., Fazio, F., Calarese, D., Stevens, J., *et al.* (2004). Printed covalent glycan array for ligand profiling of

diverse glycan binding proteins. Proceedings of the National Academy of Sciences of the United States of America *101*, 17033-17038.

Bloch, F., Hansen, W.W., and Packard, M. (1946). Nuclear Induction. Phys Rev *69*, 127-127.

Blomhoff, R., Tolleshaug, H., and Berg, T. (1982). Binding of calcium ions to the isolated asialo-glycoprotein receptor. Implications for receptor function in suspended hepatocytes. J Biol Chem *257*, 7456-7459.

Boehr, D.D., McElheny, D., Dyson, H.J., and Wright, P.E. (2006). The dynamic energy landscape of dihydrofolate reductase catalysis. Science *313*, 1638-1642.

Boehr, D.D., Nussinov, R., and Wright, P.E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol *5*, 789-796.

Boehr, D.D., Schnell, J.R., McElheny, D., Bae, S.H., Duggan, B.M., Benkovic, S.J., Dyson, H.J., and Wright, P.E. (2013). A Distal Mutation Perturbs Dynamic Amino Acid Networks in Dihydrofolate Reductase. Biochemistry *52*, 4605-4619.

Bono, P., Rubin, K., Higgins, J.M.G., and Hynes, R.O. (2001). Layilin, a novel integral membrane protein, is a hyaluronan receptor. Mol Biol Cell *12*, 891-900.

Boulton, S., Akimoto, M., Selvaratnam, R., Bashiri, A., and Melacini, G. (2014). A Tool Set to Map Allosteric Networks through the NMR Chemical Shift Covariance Analysis. Sci Rep-Uk *4*.

Bowman, G.R., and Geissler, P.L. (2012). Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. Proceedings of the National Academy of Sciences of the United States of America *109*, 11681-11686.

Brown, G.D. (2006). Dectin-1: a signalling non-TLR pattern-recognition receptor. Nature Reviews in Immunology *6*, 33-43.

Brown, G.D., and Gordon, S. (2001). Immune recognition - A new receptor for beta-glucans. Nature *413*, 36-37.

Brown, J., O'Callaghan, C.A., Marshall, A.S.J., Gilbert, R.J.C., Siebold, C., Gordon, S., Brown, G.D., and Jones, E.Y. (2007). Structure of the fungal β-glucan-binding immune receptor dectin-1: Implications for function. Protein Sci *16*, 1042-1052.

Cambi, A., and Figdor, C.G. (2005). Levels of complexity in pathogen recognition by C-type lectins. Curr Opin Immunol *17*, 345-351.

Carr, H.Y., and Purcell, E.M. (1954). Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. Phys Rev *94*, 630-638.

Carver, J.P., and Richards, R.E. (1972). General 2-Site Solution for Chemical Exchange Produced Dependence of T2 Upon Carr-Purcell Pulse Separation. J Magn Reson *6*, 89.

Cash, H.L., Whitham, C.V., Behrendt, C.L., and Hooper, L.V. (2006). Symbiotic bacteria direct expression of an intestinal bactericidal lectin. Science *313*, 1126-1130.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G., Rance, M., and Skelton, N.J. (2007). Protein NMR Spectroscopy: Principles and Practice, 2nd edn (Burlington, San Diego, London: Elsevier Academic Press).

Cecioni, S., Imberty, A., and Vidal, S. (2015). Glycomimetics versus multivalent glycoconjugates for the design of high affinity lectin ligands. Chem Rev *115*, 525-561.

Cembran, A., Kim, J., Gao, J., and Veglia, G. (2014). NMR mapping of protein conformational landscapes using coordinated behavior of chemical shifts upon ligand binding. Phys Chem Chem Phys *16*, 6508-6518.

Chabrol, E., Nurisso, A., Daina, A., Vassal-Stermann, E., Thepaut, M., Girard, E., Vivès, R.R., and Fieschi, F. (2012). Glycosaminoglycans Are Interactants of Langerin: Comparison with gp120 Highlights an Unexpected Calcium-Independent Binding Mode. PLoS ONE *7*, e50722.

Chang, S.Y., and Kweon, M.N. (2010). Langerin-expressing dendritic cells in gut-associated lymphoid tissues. Immunol Rev *234*, 233-246.

Chatwell, L., Holla, A., Kaufer, B.B., and Skerra, A. (2008). The carbohydrate recognition domain of Langerin reveals high structural similarity with the one of DC-SIGN but an additional, calcium-independent sugar-binding site. Mol Immunol *45*, 1981-1994.

Chi, C.N., Vogeli, B., Bibow, S., Strotz, D., Orts, J., Guntert, P., and Riek, R. (2015). A Structural Ensemble for the Enzyme Cyclophilin Reveals an Orchestrated Mode of Action at Atomic Resolution. Angew Chem Int Edit *54*, 11657-11661.

Christie, M.P., Toth, I., and Simerska, P. (2014). Biophysical characterization of lectin–glycan interactions for therapeutics, vaccines and targeted drug delivery. Future Medicinal Chemistry *6*, 2113-2129.

Clore, G.M., Driscoll, P.C., Wingfield, P.T., and Gronenborn, A.M. (1990a). Analysis of the Backbone Dynamics of Interleukin-1-Beta Using 2-Dimensional Inverse Detected Heteronuclear N-15-H-1 Nmr-Spectroscopy. Biochemistry *29*, 7387-7401.

Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C., and Gronenborn, A.M. (1990b). Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. J Am Chem Soc *112*, 4989–4991.

Collins, B.E., Blixt, O., DeSieno, A.R., Bovin, N., Marth, J.D., and Paulson, J.C. (2004). Masking of CD22 by cis ligands does not prevent redistribution of CD22 to sites of cell

contact. Proceedings of the National Academy of Sciences of the United States of America *101*, 6104-6109.

Collins, B.E., and Paulson, J.C. (2004). Cell surface biology mediated by low affinity multivalent protein-glycan interactions. Curr Opin Chem Biol *8*, 617-625.

Coombs, P.J., Graham, S.A., Drickamer, K., and Taylor, M.E. (2005). Selective binding of the scavenger receptor C-type lectin to Lewis(x) trisaccharide and related glycan ligands. J Biol Chem *280*, 22993-22999.

Coombs, P.J., Harrison, R., Pemberton, S., Quintero-Martinez, A., Parry, S., Haslam, S.M., Dell, A., Taylor, M.E., and Drickamer, K. (2010). Identification of novel contributions to high-affinity glycoprotein-receptor interactions using engineered ligands. J Mol Biol *396*, 685-696.

Cooper, A., and Dryden, D.T.F. (1984). Allostery without conformational change - A plausible model. Eur Biophys J *11*, 103-109.

d'Auvergne, E.J., and Gooley, P.R. (2003). The use of model selection in the model-free analysis of protein dynamics. J Biomol NMR *25*, 25-39.

d'Auvergne, E.J., and Gooley, P.R. (2006). Model-free model elimination: A new step in the model-free dynamic analysis of NMR relaxation data. J Biomol NMR *35*, 117-135.

d'Auvergne, E.J., and Gooley, P.R. (2007). Set theory formulation of the model-free problem and the diffusion seeded model-free paradigm. Molecular Biosystems *3*, 483-494.

d'Auvergne, E.J., and Gooley, P.R. (2008). Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. J Biomol NMR *40*, 107-119.

d'Auvergne, E.J., and Gooley, P.R. (2008). Optimisation of NMR dynamic models. J Biomol NMR *40*, 107–133.

Daily, M.D., Upadhyaya, T.J., and Gray, J.J. (2008). Contact rearrangements form coupled networks from local motions in allosteric proteins. Proteins-Structure Function and Bioinformatics *71*, 455-466.

Davoodi, J., Wakarchuk, W.W., Campbell, R.L., Carey, P.R., and Surewicz, W.K. (1995). Abnormally High Pk(a) of an Active-Site Glutamic-Acid Residue in Bacillus-Circulans Xylanase - the Role of Electrostatic Interactions. Eur J Biochem *232*, 839-843.

De Jesus, M., Ostroff, G.R., Levitz, S.M., Bartling, T.R., and Mantis, N.J. (2014). A population of Langerin-positive dendritic cells in murine Peyer's patches involved in sampling beta-glucan microparticles. PLoS One *9*, e91002.

de Jong, M.A.W.P., Vriend, L.E.M., Theelen, B., Taylor, M.E., Fluitsma, D., Boekhout, T., and Geijtenbeek, T.B.H. (2010). C-type lectin Langerin is a β-glucan receptor on human Langerhans cells that recognizes opportunistic and pathogenic fungi. Mol Immunol *47*, 1216-1225.

de Paz, J.L., Noti, C., and Seeberger, P.H. (2006). Microarrays of synthetic heparin oligosaccharides. J Am Chem Soc *128*, 2766-2767.

de Witte, L., Nabatov, A., Pion, M., Fluitsma, D., de Jong, M.A.W.P., de Gruijl, T., Piguet, V., van Kooyk, Y., and Geijtenbeek, T.B.H. (2007). Langerin is a natural barrier to HIV-1 transmission by Langerhans cells. Nat Med *13*, 367-371.

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277-293.

Diehl, C., Engstrom, O., Delaine, T., Hakansson, M., Genheden, S., Modig, K., Leffler, H., Ryde, U., Nilsson, U.J., and Akke, M. (2010). Protein Flexibility and Conformational Entropy in Ligand Design Targeting the Carbohydrate Recognition Domain of Galectin-3. J Am Chem Soc *132*, 14577-14589.

Disney, M.D., and Seeberger, P.H. (2004). The use of carbohydrate microarrays to study carbohydrate-cell interactions and to detect pathogens. Chem Biol *11*, 1701-1707.

Drickamer, K. (1992). Engineering galactose-binding activity into a C-type mannose-binding protein. Nature *360*, 183-186.

Drickamer, K. (1993a). Evolution of Ca2+-Dependent Animal Lectins. Prog Nucleic Acid Re *45*, 207-232.

Drickamer, K. (1993b). Recognition of Complex Carbohydrates by Ca2+-Dependent Animal Lectins. Biochem Soc Trans *21*, 456-459.

Drickamer, K. (1999). C-type lectin-like domains. Curr Opin Struct Biol *9*, 585-590.

Drickamer, K., and Taylor, M.E. (2015). Recent insights into structures and functions of C-type lectins in the immune system. Curr Opin Struct Biol *34*, 26-34.

Drummond, R.A., Saijo, S., Iwakura, Y., and Brown, G.D. (2011). The role of Syk/CARD9 coupled C-type lectins in antifungal immunity. Eur J Immunol *41*, 276-281.

Ernst, R.R., and Anderson, W.A. (1966). Application of Fourier Transform Spectroscopy to Magnetic Resonance. Rev Sci Instrum *37*, 93.

Ewart, K.V., Li, Z.J., Yang, D.S.C., Fletcher, G.L., and Hew, C.L. (1998). The ice-binding site of Atlantic herring antifreeze protein corresponds to the carbohydrate-binding site of C-type lectins. Biochemistry *37*, 4080-4085.

Farmer II, B.T., Venters, R.A., Spicer, L.D., Wittekind, M.G., and Müller, L. (1992). A refocused and optimized HNCA: Increased sensitivity and resolution in large macromolecules. J Biomol NMR *2*, 195-202.

Farrow, N., Zhang, O., Szabo, A., Torchia, D., and Kay, L. (1995). Spectral density function mapping using 15N relaxation data exclusively. J Biomol NMR *6*, 153-162.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Formankay, J.D., and Kay, L.E. (1994). Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology-2 Domain Studied by N-15 Nmr Relaxation. Biochemistry *33*, 5984-6003.

Fasting, C., Schalley, C.A., Weber, M., Seitz, O., Hecht, S., Koksch, B., Dernedde, J., Graf, C., Knapp, E.-W., and Haag, R. (2012). Multivalency as a Chemical Organization and Action Principle. Angewandte Chemie International Edition *51*, 10472-10498.

Feeney, J., Batchelor, J.G., Albrand, J.P., and Roberts, G.C.K. (1979). The effects of intermediate exchange processes on the estimation of equilibrium constants by NMR. J Magn Reson *33*, 519-529.

Feher, V.A., Durrant, J.D., Van Wart, A.T., and Amaro, R.E. (2014). Computational approaches to mapping allosteric pathways. Curr Opin Struct Biol *25*, 98-103.

Fehres, C.M., Duinkerken, S., Bruijns, S.C., Kalay, H., van Vliet, S.J., Ambrosini, M., de Gruijl, T.D., Unger, W.W., Garcia-Vallejo, J.J., and van Kooyk, Y. (2015). Langerin-mediated internalization of a modified peptide routes antigens to early endosomes and enhances cross-presentation by human Langerhans cells. Cell Mol Immunol.

Fehres, C.M., Unger, W.W., Garcia-Vallejo, J.J., and van Kooyk, Y. (2014). Understanding the biology of antigen cross-presentation for the design of vaccines against cancer. Front Immunol *5*, 149.

Feinberg, H., Guo, Y., Mitchell, D.A., Drickamer, K., and Weis, W.I. (2005). Extended Neck Regions Stabilize Tetramers of the Receptors DC-SIGN and DC-SIGNR. J Biol Chem *280*, 1327-1335.

Feinberg, H., Mitchell, D.A., Drickamer, K., and Weis, W.I. (2001). Structural Basis for Selective Recognition of Oligosaccharides by DC-SIGN and DC-SIGNR. Science *294*, 2163-2166.

Feinberg, H., Powlesland, A.S., Taylor, M.E., and Weis, W.I. (2010). Trimeric Structure of Langerin. J Biol Chem *285*, 13285-13293.

Feinberg, H., Rowntree, T.J.W., Tan, S.L.W., Drickamer, K., Weis, W.I., and Taylor, M.E. (2013). Common polymorphisms in human langerin change specificity for glycan ligands. J Biol Chem *288*, 36762-36771.

Feinberg, H., Taylor, M.E., Razi, N., McBride, R., Knirel, Y.A., Graham, S.A., Drickamer, K., and Weis, W.I. (2011). Structural Basis for Langerin Recognition of Diverse Pathogen and Mammalian Glycans through a Single Binding Site. J Mol Biol *405*, 1027-1039.

Feinberg, H., Torgersen, D., Drickamer, K., and Weis, W.I. (2000). Mechanism of pH-dependent N-acetylgalactosamine binding by a functional mimic of the hepatocyte asialoglycoprotein receptor. J Biol Chem *275*, 35176-35184.

Felli, I.C., and Pierattelli, R. (2012). Recent progress in NMR spectroscopy: toward the study of intrinsically disordered proteins of increasing size and complexity. IUBMB Life *64*, 473-481.

Fenton, A.W. (2008). Allostery: an illustrated definition for the 'second secret of life'. Trends Biochem Sci *33*, 420-425.

Fenwick, R.B., van den Bedem, H., Fraser, J.S., and Wright, P.E. (2014). Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. Proceedings of the National Academy of Sciences of the United States of America *111*, E445-E454.

Fielding, L. (2007). NMR methods for the determination of protein–ligand dissociation constants. Prog Nucl Mag Res Sp *51*, 219-242.

Findeisen, M., Brand, T., and Berger, S. (2007). A H-1-NMR thermometer suitable for cryoprobes. Magn Reson Chem *45*, 175-178.

Forni, D., Cagliani, R., Tresoldi, C., Pozzoli, U., De Gioia, L., Filippi, G., Riva, S., Menozzi, G., Colleoni, M., Biasin, M., *et al.* (2014). An evolutionary analysis of antigen processing and presentation across different timescales reveals pervasive selection. PLoS Genet *10*, e1004189.

Freyer, M.W., and Lewis, E.A. (2008). Isothermal titration calorimetry: Experimental design, data analysis, and probing Macromolecule/Ligand binding and kinetic interactions. Method Cell Biol *84*, 79-113.

Furukawa, A., Kamishikiryo, J., Mori, D., Toyonaga, K., Okabe, Y., Toji, A., Kanda, R., Miyake, Y., Ose, T., Yamasaki, S., *et al.* (2013). Structural analysis for glycolipid recognition by the C-type lectins Mincle and MCL. Proceedings of the National Academy of Sciences of the United States of America *110*, 17438-17443.

Fushman, D., Cahill, S., and Cowburn, D. (1997). The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: Analysis of N-15 relaxation with monomer/dimer equilibration. J Mol Biol *266*, 173-194.

Fushman, D., Tjandra, N., and Cowburn, D. (1998). Direct measurement of N-15 chemical shift anisotropy in solution. J Am Chem Soc *120*, 10947-10952.

Gagneux, P., and Varki, A. (1999). Evolutionary considerations in relating oligosaccharide diversity to biological function. Glycobiology *9*, 747-755.

Garcia de la Torre, J., Huertas, M.L., and Carrasco, B. (2000). HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. Journal of Magnetic Resonance (1969) *147*, 138-146.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, J.M. Walker, ed. (Humana Press), pp. 571-607.

Geider, S., Baronnet, A., Cerini, C., Nitsche, S., Astier, J.P., Michel, R., and Verdier, J.M. (1996). Pancreatic lithostathine as a calcite habit modifier. J Biol Chem *271*, 26302-26306.

Geijtenbeek, T.B., den Dunnen, J., and Gringhuis, S.I. (2009). Pathogen recognition by DC-SIGN shapes adaptive immunity. Future microbiology *4*, 879-890.

Geijtenbeek, T.B., and Gringhuis, S.I. (2009). Signalling through C-type lectin receptors: shaping immune responses. Nature Reviews in Immunology *9*, 465-479.

Goddard, T.D., and Kneller, D.G. SPARKY 3 (University of California, San Francisco).

Grant, B.D., and Donaldson, J.G. (2009). Pathways and mechanisms of endocytic recycling. Nat Rev Mol Cell Bio *10*, 597-608.

Gringhuis, S.I., den Dunnen, J., Litjens, M., van der Vlist, M., and Geijtenbeek, T.B.H. (2009). Carbohydrate-specific signaling through the DC-SIGN signalosome tailors immunity to Mycobacterium tuberculosis, HIV-1 and Helicobacter pylori. Nat Immunol *10*, 1081-1088.

Gringhuis, S.I., den Dunnen, J., Litjens, M., van Het Hof, B., van Kooyk, Y., and Geijtenbeek, T.B. (2007). C-type lectin DC-SIGN modulates Toll-like receptor signaling via Raf-1 kinase-dependent acetylation of transcription factor NF-kappaB. Immunity *26*, 605-616.

Grzesiek, S., Anglister, J., and Bax, A. (1993). Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in 13C/15N-Enriched Proteins by Isotropic Mixing of 13C Magnetization. Journal of Magnetic Resonance, Series B *101*, 114-119.

Grzesiek, S., and Bax, A. (1992a). An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins. J Magn Reson *99*, 201-207.

Grzesiek, S., and Bax, A. (1992b). Improved 3d Triple-Resonance Nmr Techniques Applied to a 31-Kda Protein. J Magn Reson *96*, 432-440.

Guo, Y., Feinberg, H., Conroy, E., Mitchell, D.A., Alvarez, R., Blixt, O., Taylor, M.E., Weis, W.I., and Drickamer, K. (2004). Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. Nat Struct Mol Biol *11*, 591-598.

Head, J.F., Mealy, T.R., McCormack, F.X., and Seaton, B.A. (2003). Crystal Structure of Trimeric Carbohydrate Recognition and Neck Domains of Surfactant Protein A. J Biol Chem *278*, 43254-43260.

Henzler-Wildman, K., and Kern, D. (2007). Dynamic personalities of proteins. Nature *450*, 964-972.

Hilser, V.J., Wrabl, J.O., and Motlagh, H.N. (2012). Structural and Energetic Basis of Allostery. In Annual Review of Biophysics, Vol 41, D.C. Rees, ed., pp. 585-609.

Hirel, P.H., Schmitter, J.M., Dessen, P., Fayat, G., and Blanquet, S. (1989). Extent of N-Terminal Methionine Excision from Escherichia-Coli Proteins Is Governed by the Side-Chain Length of the Penultimate Amino-Acid. Proceedings of the National Academy of Sciences of the United States of America *86*, 8247-8251.

Holla, A., and Skerra, A. (2011). Comparative analysis reveals selective recognition of glycans by the dendritic cell receptors DC-SIGN and Langerin. Protein Engineering, Design & Selection *24*, 659-669.

Hsu, T.-L., Cheng, S.-C., Yang, W.-B., Chin, S.-W., Chen, B.-H., Huang, M.-T., Hsieh, S.-L., and Wong, C.-H. (2009). Profiling Carbohydrate-Receptor Interaction with Recombinant Innate Immunity Receptor-Fc Fusion Proteins. J Biol Chem *284*, 34479-34489.

Hunger, R.E., Sieling, P.A., Ochoa, M.T., Sugaya, M., Burdick, A.E., Rea, T.H., Brennan, P.J., Belisle, J.T., Blauvelt, A., Porcelli, S.A., *et al.* (2004). Langerhans cells utilize CD1a and langerin to efficiently present nonpeptide antigens to T cells. J Clin Invest *113*, 701-708.

Idoyaga, J., Cheong, C., Suda, K., Suda, N., Kim, J.Y., Lee, H., Park, C.G., and Steinman, R.M. (2008). Cutting Edge: Langerin/CD207 Receptor on Dendritic Cells Mediates Efficient Antigen Presentation on MHC I and II Products In Vivo. J Immunol *180*, 3647-3650.

Isogai, S., Deupi, X., Opitz, C., Heydenreich, F.M., Tsai, C.J., Brueckner, F., Schertler, G.F., Veprintsev, D.B., and Grzesiek, S. (2016). Backbone NMR reveals allosteric signal transduction networks in the beta1-adrenergic receptor. Nature *530*, 237-241.

Jeener, J. (1971). In Ampère Summer School (Basko Polje, Yugoslavia).

Jegouzo, S.A.F., Feinberg, H., Dungarwalla, T., Drickamer, K., Weis, W.I., and Taylor, M.E. (2015). A Novel Mechanism for Binding of Galactose-terminated Glycans by the C-type

Carbohydrate Recognition Domain in Blood Dendritic Cell Antigen 2. J Biol Chem *290*, 16759-16771.

Jégouzo, S.A.F., Quintero-Martínez, A., Ouyang, X., dos Santos, Á., Taylor, M.E., and Drickamer, K. (2013). Organization of the extracellular portion of the macrophage galactose receptor: A trimeric cluster of simple binding sites for N-acetylgalactosamine. Glycobiology *23*, 853-864.

Jencks, W.P. (1981). On the attribution and additivity of binding energies. Proceedings of the National Academy of Sciences *78*, 4046-4050.

Karlsson, K.A. (1995). Microbial Recognition of Target-Cell Glycoconjugates. Curr Opin Struct Biol *5*, 622-635.

Karlsson, R., and Fält, A. (1997). Experimental design for kinetic analysis of protein-protein interactions with surface plasmon resonance biosensors. J Immunol Methods *200*, 121-133.

Karplus, M. (1959). Contact Electron-Spin Coupling of Nuclear Magnetic Moments. J Chem Phys *30*, 11-15.

Karplus, M., and Mccammon, J.A. (1986). The Dynamics of Proteins. Sci Am *254*, 42-&.

Kay, L.E., Ikura, M., Tschudin, R., and Bax, A. (1990). 3-Dimensional Triple-Resonance Nmr-Spectroscopy of Isotopically Enriched Proteins. J Magn Reson *89*, 496-514.

Kay, L.E., Torchia, D.A., and Bax, A. (1989). Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry *28*, 8972–8979.

Keedy, D.A., Kenner, L.R., Warkentin, M., Woldeyes, R.A., Hopkins, J.B., Thompson, M.C., Brewster, A.S., Van Benschoten, A.H., Baxter, E.L., Uervirojnangkoorn, M., *et al.* (2015). Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. eLife *4*.

Keeler, J. (2010). Understanding NMR spectroscopy, 2nd edn (Chichester, U.K.: John Wiley and Sons).

Keller, B., Gattin, Z., and van Gunsteren, W.F. (2010). What stabilizes the 3(14)-helix in beta3-peptides? A conformational analysis using molecular simulation. Proteins *78*, 1677-1690.

Kern, D., and Zuiderweg, E.R.P. (2003). The role of dynamics in allosteric regulation. Curr Opin Struct Biol *13*, 748-757.

Kerns, S.J., Agafonov, R.V., Cho, Y.J., Pontiggia, F., Otten, R., Pachov, D.V., Kutter, S., Phung, L.A., Murphy, P.N., Thai, V., *et al.* (2015). The energy landscape of adenylate kinase during catalysis. Nat Struct Mol Biol *22*, 124-131.

Kiessling, L.L., and Lamanna, A.C. (2003). Multivalency in Biological Systems. In Chemical Probes in Biology: Science at the Interface of Chemistry, Biology and Medicine, M.P. Schneider, ed. (Dordrecht: Springer Netherlands), pp. 345-357.

Kim, H.J., Brennan, P.J., Heaslip, D., Udey, M.C., Modlin, R.L., and Belisle, J.T. (2014). Carbohydrate Dependent Binding of Langerin to SodC, a Cell Wall Glycoprotein of Mycobacterium leprae. J Bacteriol *197*, 615-325.

Kim, M.O., Blachly, P.G., and McCammon, J.A. (2015). Conformational Dynamics and Binding Free Energies of Inhibitors of BACE-1: From the Perspective of Protonation Equilibria. PLoS Comp Biol *11*, e1004341.

Kitov, P.I., and Bundle, D.R. (2003). On the Nature of the Multivalency Effect: A Thermodynamic Model. J Am Chem Soc *125*, 16271-16284.

Kitov, P.I., Sadowska, J.M., Mulvey, G., Armstrong, G.D., Ling, H., Pannu, N.S., Read, R.J., and Bundle, D.R. (2000). Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands. Nature *403*, 669-672.

Knight, W.D. (1949). Nuclear Magnetic Resonance Shift in Metals. Phys Rev *76*, 1259-1260.

Kohatsu, L., Hsu, D.K., Jegalian, A.G., Liu, F.T., and Baum, L.G. (2006). Galectin-3 induces death of Candida species expressing specific beta-1,2-linked Mannans. J Immunol *177*, 4718-4726.

Korzhnev, D.M., Billeter, M., Arseniev, A.S., and Orekhov, V.Y. (2001). NMR studies of Brownian tumbling and internal motions in proteins. Prog Nucl Mag Res Sp *38*, 197-266.

Lee, Y.C., Townsend, R.R., Hardy, M.R., Lonngren, J., Arnarp, J., Haraldsson, M., and Lonn, H. (1983). Binding of Synthetic Oligosaccharides to the Hepatic Gal Galnac Lectin - Dependence on Fine-Structural Features. J Biol Chem *258*, 199-202.

LeibundGut-Landmann, S., Osorio, F., Brown, G.D., and Sousa, C.R.E. (2008). Stimulation of dendritic cells via the dectin-1/Syk pathway allows priming of cytotoxic T-cell responses. Blood *112*, 4971-4980.

Lescop, E., Schanda, P., and Brutscher, B. (2007). A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. J Magn Reson *187*, 163-169.

Levitt, M.H. (2008). Spin dynamics: basics of nuclear magnetic resonance, 2nd edn (Chichester, U.K.; Hoboken, NJ: John Wiley & Sons).

Liang, P.H., Wang, S.K., and Wong, C.H. (2007). Quantitative analysis of carbohydrate-protein interactions using glycan microarrays: Determination of surface and solution dissociation constants. J Am Chem Soc *129*, 11177-11184.

Lipari, G., and Szabo, A. (1982a). Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. J Am Chem Soc *104*, 4546-4559.

Lipari, G., and Szabo, A. (1982b). Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. J Am Chem Soc *104*, 4559-4570.

Lipari, G., and Szabo, A. (1982c). Protein dynamics and NMR relaxation: comparison of simulations with experiment. Nature *300*, 197-198.

Lorentsen, R.H., Graversen, J.H., Caterer, N.R., Thogersen, H.C., and Etzerodt, M. (2000). The heparin-binding site in tetranectin is located in the N-terminal region and binding does not involve the carbohydrate recognition domain. Biochem J *347*, 83-87.

Loria, J.P., Rance, M., and Palmer, A.G. (1999). A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. J Am Chem Soc *121*, 2331-2332.

Luz, Z., and Meiboom, S. (1963). Nuclear Magnetic Resonance Study of Protolysis of Trimethylammonium Ion in Aqueous Solution - Order of Reaction with Respect to Solvent. J Chem Phys *39*, 366.

MacRaild, C.A., Daranas, A.H., Bronowska, A., and Homans, S.W. (2007). Global changes in local protein dynamics reduce the entropic cost of carbohydrate binding in the arabinose-binding protein. J Mol Biol *368*, 822-832.

Madigan, M.T., Martinko, J.M., and Parker, J. (2003). Brock biology of microorganisms, 10th edn (Upper Saddle River, NJ: Prentice Hall/Pearson Education).

Mammen, M., Choi, S.K., and Whitesides, G.M. (1998). Polyvalent interactions in biological systems: Implications for design and use of multivalent ligands and inhibitors. Angew Chem Int Edit *37*, 2755-2794.

Mandel, A.M., Akke, M., and Palmer, A.G. (1995). Backbone Dynamics of Escherichia-Coli Ribonuclease Hi - Correlations with Structure and Function in an Active Enzyme. J Mol Biol *246*, 144-163.

Manzo, C., Torreno-Pina, J.A., Joosten, B., Reinieren-Beeren, I., Gualda, E.J., Loza-Alvarez, P., Figdor, C.G., Garcia-Parajo, M.F., and Cambi, A. (2012). The neck region of the C-type lectin DC-SIGN regulates its surface spatiotemporal organization and virus-binding capacity on antigen presenting cells. J Biol Chem *287*, 38946-38955.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M., and Clore, G.M. (1989a). Overcoming the Overlap Problem in the Assignment of H-1-Nmr Spectra of

Larger Proteins by Use of 3-Dimensional Heteronuclear H-1-N-15 Hartmann-Hahn Multiple Quantum Coherence and Nuclear Overhauser Multiple Quantum Coherence Spectroscopy - Application to Interleukin-1-Beta. Biochemistry *28*, 6150-6156.

Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A., and Bax, A. (1989b). 3-Dimensional Heteronuclear Nmr of N-15-Labeled Proteins. J Am Chem Soc *111*, 1515-1517.

Mayer, M., and James, T.L. (2004). NMR-Based Characterization of Phenothiazines as a RNA Binding Scaffold†. J Am Chem Soc *126*, 4453-4460.

Mayer, M., and Meyer, B. (1999). Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. Angewandte Chemie International Edition *38*, 1784-1788.

Mc Dermott, R., Ziylan, U., Spehner, D., Bausinger, H., Lipsker, D., Mommaas, M., Cazenave, J.P., Raposo, G., Goud, B., de la Salle, H., *et al.* (2002). Birbeck granules are subdomains of endosomal recycling compartment in human epidermal Langerhans cells, which form where langerin accumulates. Mol Biol Cell *13*, 317-335.

Mcconnell, H.M. (1958). Reaction Rates by Nuclear Magnetic Resonance. J Chem Phys *28*, 430-431.

McMahon, S.A., Miller, J.L., Lawton, J.A., Kerkow, D.E., Hodes, A., Marti-Renom, M.A., Doulatov, S., Narayanan, E., Sali, A., Miller, J.F., *et al.* (2005). The C-type lectin fold as an evolutionary solution for massive sequence variation. Nat Struct Mol Biol *12*, 886-892.

Meiboom, S., and Gill, D. (1958). Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. Rev Sci Instrum *29*, 688-691.

Mestas, J., and Hughes, C.C.W. (2004). Of Mice and Not Men: Differences between Mouse and Human Immunology. The Journal of Immunology *172*, 2731-2738.

Meyer, B., and Peters, T. (2003). NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. Angewandte Chemie International Edition *42*, 864-890.

Millet, O., Loria, J.P., Kroenke, C.D., Pons, M., and Palmer, A.G. (2000). The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale. J Am Chem Soc *122*, 2867-2877.

Mocsai, A., Ruland, J., and Tybulewicz, V.L.J. (2010). The SYK tyrosine kinase: a crucial player in diverse biological functions. Nature Reviews Immunology *10*, 387-402.

Monod, J., Changeux, J.P., and Jacob, F. (1963). Allosteric Proteins and Cellular Control Systems. J Mol Biol *6*, 306.

Monod, J., Wyman, J., and Changeux, J.P. (1965). On Nature of Allosteric Transitions - a Plausible Model. J Mol Biol *12*, 88.

Motlagh, H.N., Wrabl, J.O., Li, J., and Hilser, V.J. (2014). The ensemble nature of allostery. Nature *508*, 331-339.

Mulder, F.A.A., Skrynnikov, N.R., Hon, B., Dahlquist, F.W., and Kay, L.E. (2001). Measurement of slow (mu s-ms) time scale dynamics in protein side chains by N-15 relaxation dispersion NMR spectroscopy: Application to Asn and Gln residues in a cavity mutant of T4 lysozyme. J Am Chem Soc *123*, 967-975.

Munoz-Garcia, J.C., Chabrol, E., Vives, R.R., Thomas, A., de Paz, J.L., Rojo, J., Imberty, A., Fieschi, F., Nieto, P.M., and Angulo, J. (2015). Langerin-heparin interaction: two binding sites for small and large ligands as revealed by a combination of NMR spectroscopy and cross-linking mapping experiments. J Am Chem Soc *137*, 4100-4110.

Nagae, M., Yamanaka, K., Hanashima, S., Ikeda, A., Morita-Matsumoto, K., Satoh, T., Matsumoto, N., Yamamoto, K., and Yamaguchi, Y. (2013). Recognition of Bisecting N-Acetylglucosamine STRUCTURAL BASIS FOR ASYMMETRIC INTERACTION WITH THE MOUSE LECTIN DENDRITIC CELL INHIBITORY RECEPTOR 2. J Biol Chem *288*, 33598-33610.

Natarajan, K., Dimasi, N., Wang, J., Mariuzza, R.A., and Margulies, D.H. (2002). Structure and function of natural killer cell receptors: Multiple molecular solutions to self, nonself discrimination. Annu Rev Immunol *20*, 853-885.

Nesmelova, I.V., Ermakova, E., Daragan, V.A., Pang, M., Menendez, M., Lagartera, L., Solis, D., Baum, L.G., and Mayo, K.H. (2010). Lactose binding to galectin-1 modulates structural dynamics, increases conformational entropy, and occurs with apparent negative cooperativity. J Mol Biol *397*, 1209-1230.

Ng, K.K., and Weis, W.I. (1998). Coupling of prolyl peptide bond isomerization and Ca2+ binding in a C-type mannose-binding protein. Biochemistry *37*, 17977-17989.

Ng, K.K.S., Park-Snyder, S., and Weis, W.I. (1998). Ca2+-Dependent Structural Changes in C-type Mannose-Binding Proteins†,‡. Biochemistry *37*, 17965-17976.

Nielbo, S., Thomsen, J.K., Graversen, J.H., Jensen, P.H., Etzerodt, M., Poulsen, F.M., and Thogersen, H.C. (2004). Structure of the plasminogen kringle 4 binding calcium-free form of the C-type lectin-like domain of tetranectin. Biochemistry *43*, 8636-8643.

Nohaile, M., Kern, D., Wemmer, D., Stedman, K., and Kustu, S. (1997). Structural and functional analyses of activating amino acid substitutions in the receiver domain of NtrC: Evidence for an activating surface. J Mol Biol *273*, 299-316.

Onizuka, T., Shimizu, H., Moriwaki, Y., Nakano, T., Kanai, S., Shimada, I., and Takahashi, H. (2012). NMR study of ligand release from asialoglycoprotein receptor under solution conditions in early endosomes. FEBS J *279*, 2645-2656.

Orekhov, V.Y., Korzhnev, D.M., Diercks, T., Kessler, H., and Arseniev, A.S. (1999). H-1-N-15 NMR dynamic study of an isolated alpha-helical peptide (1-36)- bacteriorhodopsin reveals the equilibrium helix-coil transitions. J Biomol NMR *14*, 345-356.

Osorio, F., and Reis e Sousa, C. (2011). Myeloid C-type lectin receptors in pathogen recognition and host defense. Immunity *34*, 651-664.

Paulson, J.C., Blixt, O., and Collins, B.E. (2006). Sweet spots in functional glycomics. Nat Chem Biol *2*, 238-248.

Pederson, K., Mitchell, D.A., and Prestegard, J.H. (2014). Structural Characterization of the DC-SIGN-Lewis Complex. Biochemistry *53*, 5700-5709.

Peng, J.W., and Wagner, G. (1995). Frequency Spectrum of NH Bonds in Eglin c from Spectral Density Mapping at Multiple Fields. Biochemistry *34*, 16733-16752.

Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997). Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proceedings of the National Academy of Sciences of the United States of America *94*, 12366-12371.

Pierce, M.M., Raman, C.S., and Nall, B.T. (1999). Isothermal titration calorimetry of protein-protein interactions. Methods *19*, 213-221.

Poget, S.F., Freund, S.M., Howard, M.J., and Bycroft, M. (2001). The ligand-binding loops in the tunicate C-type lectin TC14 are rigid. Biochemistry *40*, 10966-10972.

Popovych, N., Sun, S., Ebright, R.H., and Kalodimos, C.G. (2006). Dynamically driven protein allostery. Nat Struct Mol Biol *13*, 831-838.

Probert, F., Mitchell, D.A., and Dixon, A.M. (2014). NMR evidence for oligosaccharide release from the DC-SIGNR (CLEC4M) carbohydrate recognition domain at low pH. FEBS J *281*, 3739-3750.

Probert, F., Whittaker, S.B.-M., Crispin, M., Mitchell, D.A., and Dixon, A.M. (2013). Solution NMR analyses of the C-type carbohydrate recognition domain of DC-SIGNR reveal different binding modes for HIV-derived oligosaccharides and smaller glycan fragments. J Biol Chem *288*, 22745-22757.

Purcell, E.M., Torrey, H.C., and Pound, R.V. (1946). Resonance Absorption by Nuclear Magnetic Moments in a Solid. Phys Rev *69*, 37-38.

Rambourg, A., and Leblond, C.P. (1967). Electron Microscope Observations on Carbohydrate-Rich Cell Coat Present at Surface of Cells in Rat. J Cell Biol *32*, 27.

Razi, N., and Varki, A. (1998). Masking and unmasking of the sialic acid-binding lectin activity of CD22 (Siglec-2) on B lymphocytes. Proceedings of the National Academy of Sciences of the United States of America *95*, 7469-7474.

Reynolds, K.A., McLaughlin, R.N., and Ranganathan, R. (2011). Hot Spots for Allosteric Regulation on Protein Surfaces. Cell *147*, 1564-1575.

Rillahan, C.D., and Paulson, J.C. (2011). Glycan Microarrays for Decoding the Glycome. Annu Rev Biochem *80*, 797-823.

Robinson, M.J., Sancho, D., Slack, E.C., LeibundGut-Landmann, S., and Reis e Sousa, C. (2006). Myeloid C-type lectins in innate immunity. Nat Immunol *7*, 1258-1265.

Romani, N., Clausen, B.E., and Stoitzner, P. (2010). Langerhans cells and more: langerin-expressing dendritic cell subsets in the skin. Immunol Rev *234*, 120-141.

Sancho, D., Joffre, O.P., Keller, A.M., Rogers, N.C., Martinez, D., Hernanz-Falcon, P., Rosewell, I., and Reis e Sousa, C. (2009). Identification of a dendritic cell receptor that couples sensing of necrosis to immunity. Nature *458*, 899-903.

Sano, H., Kuroki, Y., Honma, T., Ogasawara, Y., Sohma, H., Voelker, D.R., and Akino, T. (1998). Analysis of chimeric proteins identifies the regions in the carbohydrate recognition domains of rat lung collectins that are essential for interactions with phospholipids, glycolipids, and alveolar type II cells. J Biol Chem *273*, 4783-4789.

Schagger, H. (2006). Tricine-SDS-PAGE. Nature Protocols *1*, 16-22.

Schanda, P., Van Melckebeke, H., and Brutscher, B. (2006). Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes. J Am Chem Soc *128*, 9042-9043.

Schlapschy, M., Grimm, S., and Skerra, A. (2006). A system for concomitant overexpression of four periplasmic folding catalysts to improve secretory protein production in Escherichia coli. Protein Eng Des Sel *19*, 385-390.

Schotte, F., Lim, M.H., Jackson, T.A., Smirnov, A.V., Soman, J., Olson, J.S., Phillips, G.N., Wulff, M., and Anfinrud, P.A. (2003). Watching a protein as it functions with 150-ps time-resolved X-ray crystallography. Science *300*, 1944-1947.

Schrodinger, L. (2015). The PyMOL Molecular Graphics System, Version 1.8.

Schubert, M., Labudde, D., Oschkinat, H., and Schmieder, P. (2002). A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. J Biomol NMR *24*, 149-154.

Schumann, F.H., Riepl, H., Maurer, T., Gronwald, W., Neidig, K.P., and Kalbitzer, H.R. (2007). Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. J Biomol NMR *39*, 275-289.

Schwarz, U.S., and Alon, R. (2004). L-selectin-mediated leukocyte tethering in shear flow is controlled by multiple contacts and cytoskeletal anchorage facilitating fast rebinding events. Proceedings of the National Academy of Sciences of the United States of America *101*, 6940-6945.

Selvaratnam, R., VanSchouwen, B., Fogolari, F., Mazhab-Jafari, M.T., Das, R., and Melacini, G. (2012). The Projection Analysis of NMR Chemical Shifts Reveals Extended EPAC Autoinhibition Determinants. Biophys J *102*, 630-639.

Sethi, A., Eargle, J., Black, A.A., and Luthey-Schulten, Z. (2009). Dynamical networks in tRNA: protein complexes. Proceedings of the National Academy of Sciences of the United States of America *106*, 6620-6625.

Shaw, P.L., Kirschner, A.N., Jardetzky, T.S., and Longnecker, R. (2010). Characteristics of Epstein-Barr virus envelope protein gp42. Virus Genes *40*, 307-319.

Shortman, K., and Liu, Y.J. (2002). Mouse and human dendritic cell subtypes. Nature Reviews in Immunology *2*, 151-161.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., *et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol *7*, 539.

Smith, A.E., and Helenius, A. (2004). How viruses enter animal cells. Science *304*, 237-242.

Smock, R.G., and Gierasch, L.M. (2009). Sending Signals Dynamically. Science *324*, 198-203.

Stambach, N.S., and Taylor, M.E. (2003). Characterization of carbohydrate recognition by langerin, a C-type lectin of Langerhans cells. Glycobiology *13*, 401-410.

Stein, P.E., Boodhoo, A., Armstrong, G.D., Cockle, S.A., Klein, M.H., and Read, R.J. (1994). The Crystal-Structure of Pertussis Toxin. Structure *2*, 45-57.

Stein, R.L. (1993). Mechanism of Enzymatic and Nonenzymatic Prolyl cis-trans Isomerization. Advances in Protein Chemistry *44*, 1-24.

Stowell, S.R., Arthur, C.M., McBride, R., Berger, O., Razi, N., Heimburg-Molinaro, J., Rodrigues, L.C., Gourdine, J.P., Noll, A.J., von Gunten, S., *et al.* (2014). Microbial glycan microarrays define key features of host-microbial interactions. Nat Chem Biol *10*, 470-476.

Swain, J.F., and Gierasch, L.M. (2006). The changing landscape of protein allostery. Curr Opin Struct Biol *16*, 102-108.

Tabarani, G., Thepaut, M., Stroebel, D., Ebel, C., Vives, C., Vachette, P., Durand, D., and Fieschi, F. (2009). DC-SIGN neck domain is a pH-sensor controlling oligomerization: SAXS and hydrodynamic studies of extracellular domain. J Biol Chem *284*, 21229-21240.

Takahara, K., Yashima, Y., Omatsu, Y., Yoshida, H., Kimura, Y., Kang, Y.S., Steinman, R.M., Park, C.G., and Inaba, K. (2004). Functional comparison of the mouse DC-SIGN, SIGNR1, SIGNR3 and Langerin, C-type lectins. Int Immunol *16*, 819-829.

Talaga, M.L., Fan, N., Fueri, A.L., Brown, R.K., Chabre, Y.M., Bandyopadhyay, P., Roy, R., and Dam, T.K. (2014). Significant Other Half of a Glycoconjugate: Contributions of Scaffolds to Lectin-Glycoconjugate Interactions. Biochemistry *53*, 4445-4454.

Tateno, H., Ohnishi, K., Yabe, R., Hayatsu, N., Sato, T., Takeya, M., Narimatsu, H., and Hirabayashi, J. (2010). Dual Specificity of Langerin to Sulfated and Mannosylated Glycans via a Single C-type Carbohydrate Recognition Domain. J Biol Chem *285*, 6390-6400.

Taylor, M.E., and Drickamer, K. (2003). Structure–Function Analysis of C-Type Animal Lectins. Methods Enzymol *363*, 3-15.

Taylor, M.E., and Drickamer, K. (2009). Structural insights into what glycan arrays tell us about how glycan-binding proteins interact with their ligands. Glycobiology *19*, 1155-1162.

Thépaut, M., Valladeau, J., Nurisso, A., Kahn, R., Arnou, B., Vivès, C., Saeland, S., Ebel, C., Monnier, C., Dezutter-Dambuyant, C., *et al.* (2009). Structural Studies of Langerin and Birbeck Granule: A Macromolecular Organization Model. Biochemistry *48*, 2684-2698.

Tjandra, N., Szabo, A., and Bax, A. (1996). Protein Backbone Dynamics and 15N Chemical Shift Anisotropy from Quantitative Measurement of Relaxation Interference Effects. J Am Chem Soc *118*, 6986-6991.

Tollinger, M., Forman-Kay, J.D., and Kay, L.E. (2002). Measurement of Side-Chain Carboxyl pKa Values of Glutamate and Aspartate Residues in an Unfolded Protein by Multinuclear NMR Spectroscopy. J Am Chem Soc *124*, 5714-5717.

Tollinger, M., Skrynnikov, N.R., Mulder, F.A.A., Forman-Kay, J.D., and Kay, L.E. (2001). Slow dynamics in folded and unfolded states of an SH3 domain. J Am Chem Soc *123*, 11341-11352.

Torchia, D.A. (2011). Dynamics of biomolecules from picoseconds to seconds at atomic resolution. J Magn Reson *212*, 1-10.

Turnbull, W.B., and Daranas, A.H. (2003). On the value of c: Can low affinity systems be studied by isothermal titration calorimetry? J Am Chem Soc *125*, 14859-14866.

Unger, W.W.J., van Beelen, A.J., Bruijns, S.C., Joshi, M., Fehres, C.M., van Bloois, L., Verstege, M.I., Ambrosini, M., Kalay, H., Nazmi, K., *et al.* (2012). Glycan-modified

liposomes boost CD4+ and CD8+ T-cell responses by targeting DC-SIGN on dendritic cells. J Controlled Release *160*, 88-95.

Valdar, W.S. (2002). Scoring residue conservation. Proteins *48*, 227-241.

Valladeau, J., Clair-Moninot, V., Dezutter-Dambuyant, C., Pin, J.-J., Kissenpfennig, A., Mattéi, M.-G., Ait-Yahia, S., Bates, E.E.M., Malissen, B., Koch, F., *et al.* (2002). Identification of Mouse Langerin/CD207 in Langerhans Cells and Some Dendritic Cells of Lymphoid Tissues. J Immunol *168*, 782-792.

Valladeau, J., Ravel, O., Dezutter-Dambuyant, C., Moore, K., Kleijmeer, M., Liu, Y., Duvert-Frances, V., Vincent, C., Schmitt, D., Davoust, J., *et al.* (2000). Langerin, a Novel C-Type Lectin Specific to Langerhans Cells, Is an Endocytic Receptor that Induces the Formation of Birbeck Granules. Immunity *12*, 71-81.

van den Bedem, H., Bhabha, G., Yang, K., Wright, P.E., and Fraser, J.S. (2013). Automated identification of functional dynamic contact networks from X-ray crystallography. Nat Methods *10*, 896-U110.

van der Vlist, M., de Witte, L., de Vries, R.D., Litjens, M., de Jong, M.A.W.P., Fluitsma, D., de Swart, R.L., and Geijtenbeek, T.B.H. (2011). Human Langerhans cells capture measles virus through Langerin and present viral antigens to CD4+ T cells but are incapable of cross-presentation. Eur J Immunol *41*, 2619-2631.

van Kooyk, Y., and Rabinovich, G.A. (2008). Protein-glycan interactions in the control of innate and adaptive immune responses. Nat Immunol *9*, 593-601.

VanWart, A.T., Eargle, J., Luthey-Schulten, Z., and Amaro, R.E. (2012). Exploring Residue Component Contributions to Dynamical Network Models of Allostery. J Chem Theory Comput *8*, 2949-2961.

Varki, A. (2006). Nothing in Glycobiology Makes Sense, except in the Light of Evolution. Cell *126*, 841-845.

Varki, A. (2007). Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature *446*, 1023-1029.

Vis, H., Vorgias, C.E., Wilson, K.S., Kaptein, R., and Boelens, R. (1998). Mobility of NH bonds in DNA-binding protein HU of Bacillus stearothermophilus from reduced spectral density mapping analysis at multiple NMR fields. J Biomol NMR *11*, 265-277.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687-696.

Ward, E.M., Stambach, N.S., Drickamer, K., and Taylor, M.E. (2006). Polymorphisms in Human Langerin Affect Stability and Sugar Binding Activity. J Biol Chem *281*, 15450-15456.

Weis, W.I., and Drickamer, K. (1994). Trimeric structure of a C-type mannose-binding protein. Structure *2*, 1227-1240.

Weis, W.I., and Drickamer, K. (1996). Structural Basis of Lectin-Carbohydrate Recognition. Annu Rev Biochem *65*, 441-473.

Weis, W.I., Drickamer, K., and Hendrickson, W.A. (1992). Structure of a C-type mannose-binding protein complexed with an oligosaccharide. Nature *360*, 127-134.

Weis, W.I., Kahn, R., Fourme, R., Drickamer, K., and Hendrickson, W.A. (1991). Structure of the calcium dependent lectin domain from a rat mannose-bindnig protein determined by MAD phasing. Science *254*, 1608-1615.

Weis, W.I., Taylor, M.E., and Drickamer, K. (1998). The C-type lectin superfamily in the immune system. Immunol Rev *163*, 19-34.

Williamson, M.P. (2013). Using chemical shift perturbation to characterise ligand binding. Prog Nucl Mag Res Sp *73*, 1-16.

Wittekind, M., and Mueller, L. (1993). Hncacb, a High-Sensitivity 3d Nmr Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins. J Magn Reson Ser B *101*, 201-205.

Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E.Z., and Kern, D. (2004). Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nat Struct Mol Biol *11*, 945-949.

Wragg, S., and Drickamer, K. (1999). Identification of amino acid residues that determine pH dependence of ligand binding to the asialoglycoprotein receptor during endocytosis. J Biol Chem *274*, 35400-35406.

Wurzburg, B.A., Tarchevskaya, S.S., and Jardetzky, T.S. (2006). Structural changes in the lectin domain of CD23, the low-affinity IgE receptor, upon calcium binding. Structure *14*, 1049-1058.

Yamazaki, T., Nicholson, L.K., Wingfield, P., Stahl, S.J., Kaufman, J.D., Eyermann, C.J., Hodge, C.N., Lam, P.Y.S., and Torchia, D.A. (1994). NMR and X-ray Evidence That the HIV Protease Catalytic Aspartyl Groups Are Protonated in the Complex Formed by the Protease and a Non-Peptide Cyclic Urea-Based Inhibitor. J Am Chem Soc *116*, 10791-10792.

Yang, D.W., and Kay, L.E. (1996). Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: Application to protein folding. J Mol Biol *263*, 369-382.

Zelensky, A.N., and Gready, J.E. (2005). The C-type lectin-like domain superfamily. FEBS J *272*, 6179-6217.

Zhuravleva, A.V., Korzhnev, D.M., Kupce, E., Arseniev, A.S., Billeter, M., and Orekhov, V.Y. (2004). Gated electron transfers and electron pathways in azurin: A NMR dynamic study at multiple fields and temperatures. J Mol Biol *342*, 1599-1611.

Zierke, M., Smiesko, M., Rabbani, S., Aeschbacher, T., Cutting, B., Allain, F.H., Schubert, M., and Ernst, B. (2013). Stabilization of branched oligosaccharides: Lewis(x) benefits from a nonconventional C-H...O hydrogen bond. J Am Chem Soc *135*, 13464-13472.

Zuiderweg, E.R.P. (2002). Mapping protein-protein interactions in solution by NMR Spectroscopy. Biochemistry *41*, 1-7.

Zuiderweg, E.R.P., and Fesik, S.W. (1989). Heteronuclear 3-Dimensional Nmr-Spectroscopy of the Inflammatory Protein C5a. Biochemistry *28*, 2387-2391.

Appendix A Tables

A.1 Table of Abbreviations

Abbreviation	Full name
ASGPR	Asialoglycoprotein receptor
BDCA-2	Blood dendritic cell antigen 2
BEST	Band-selective excitation short-transient
BMRB	Biological Magnetic Resonance Databank
BS3	Bis(sulfosuccinimidyl)suberate
CD	Circular dichroism (biophysics) / cluster of differentiation (immunology)
CHESPA	Chemical shift projection analysis
CPMG	Carr-Purcell-Meiboom-Gill
CRD	Carbohydrate recognition domain
CSA	Chemical shift anisotropy
CTLD	C-type lectin domain
CTLR	C-type lectin receptor
CV	Column volume
DC	Dendritic cells
DCIR	Dendritic cell immunoreceptor
DC-SIGN(R)	Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (related)
DHAP	2,5-Dihydroxyacetophenone
DTT	Dithiothreitol
EC_{50}	Half maximal effective concentration
ECD	Extracellular domain
EDTA	Ethylenediaminetetraacetic acid
FT	Fourier transform
GalNAc	N-acetyl galactosamin
GBP	Glycan-binding protein
Glc	Glucose
GlcA	Glucoronic acid

GlcNAc	N-acetyl glucosamine
HBS	HEPES-buffered saline
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human immunodeficiency virus
hLang	Human Langerin
HRP	Horse-radish peroxidase
HSQC	Hetero single quantum coherence transfer
IB	Inclusion bodies
IC_{50}	Half maximal inhibitory concentration
IDP	Intrinsically disordered proteins
INEPT	Insensitive nuclei enhanced by polarization transfer
ITAM	Immunoreceptor tyrosine-based activation motifs
ITC	Isothermal titration calorimetry
ITIM	Immunoreceptor tyrosin-based inhibitory motif
Kdo	3-Deoxy-D-manno-oct-2-ulosonic acid
Ko	D-manno-oct-2-ulosonic acid
LC	Langerhans cells
LPS	Lipopolysaccharide
MALDI-TOF MS	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometer
Man	Mannose
MES	2-(N-morpholino)ethanesulfonic acid
mLang	Murine Langerin
MW	Molecular weight
ND	Neck domain
Neu5Ac	N-Acetylneuraminic acid
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NMR	Nuclear magnetic resonance spectroscopy
NOE	Nulcear Overhauser enhancement
NOESY	Nuclear Overhauser enhancement spectroscopy
OPS	O-specific polysaccharide
PAMP	Pathogen-associated molecular pattern
pH_{50}	Half maximal effective pH
PRR	Pattern recognition receptor

RF	Radio frequency
RT	Room temperature
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SH3	SRC homology 3 domain
SPR	Surface Plasmon Resonance
STD NMR	Saturation transfer difference NMR
Syk	Spleen tyrosine kinase
TBS	Tris-buffered saline
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
Tris	Tris(hydroxymethyl)aminomethane
TROSY	Transverse relaxation optimized spectroscopy
WATERGATE	Water suppression by gradient tailored excitation

A.2 Table of Constants

Avogadros's number	N_A	6.022×10^{23}
Boltzmann constant	k_B	1.381×10^{-23} J K^{-23}
Gas constant	R	8.3145 J mol^{-1} K^{-1}
Gyromagnetic ratio of proton	γ_H	2.6752×10^{8} T^{-1} s^{-1}
Gyromagnetic ratio of nitrogen	γ_N	-2.7116×10^{7} T^{-1} s^{-1}
Permeability of free space	μ_0	$4\pi \times 10^{-7}$ T^2 m^3 J^{-1}
Planck's constant	h	6.626×10^{-34} J s
	$\hbar = h/2\pi$	1.055×10^{-34} J s

A.3 Data Tables

Table A-1: Resonance assignment of holo Langerin CRD with chemical shifts of all assigned peaks of the major conformation (in ppm).

No.	Type	H^N	N	N^{δ}	N^{ε}	H^{α}	H^{δ}	H^{ε}	C	C^{α}	C^{β}	C^{γ}	C^{δ}	C^{ε}
193	Q	-	-	-	-	-	-	-	175.7	55.8	29.8	-	-	-
194	V	8.45	123.3	-	-	4.17	-	-	176.1	62.5	32.9	-	-	-
195	V	8.37	124.6	-	-	4.22	-	-	176.1	62.2	33	-	-	-
196	S	8.49	120.1	-	-	4.48	-	-	174.2	58.3	64	-	-	-
197	Q	8.55	122.9	-	112.4	4.51	-	7.59	176.3	56.4	29.7	33.8	180.5	-
								6.90						
198	G	8.68	110.6	-	-	4.05	-	-	174.1	45.4	-	-	-	-
						4.31								
199	W	8.6	121.2	-	-	5.26	-	-	176.7	57.4	30.1	-	-	-
200	K	10.3	123.8	-	-	5.09	-	-	175.9	55.4	36	-	-	-
201	Y	9.2	125.1	-	-	5.07	-	-	175.2	57.8	40.5	-	-	-
202	F	8.61	127.5	-	-	3.91	-	-	173.8	58.8	40.4	-	-	-
203	K	8.56	128	-	-	3.32	-	-	175	57	31.8	24.7	29.2	42
204	G	6.45	101.5	-	-	3.71	-	-	173	45.7	-	-	-	-
205	N	7.49	117.9	-	-	5.24	-	-	171	52.3	44.1	-	-	-
206	F	9.4	117.7	-	-	5.79	-	-	175.7	56.7	42.9	-	-	-
207	Y	10.11	119	-	-	6.02	-	-	175.1	57	43.4	-	-	-
208	Y	9.09	126.5	-	-	4.44	-	-	173.1	56	39.5	-	-	-
209	F	8.4	128.3	-	-	4.33	-	-	174.7	56.1	38	-	-	-
210	S	8.67	119.6	-	-	-	-	-	172	58.9	64.1	-	-	-
211	L	8.5	116.2	-	-	4.52	-	-	177.4	54.5	43.3	-	-	-
212	I	7.14	116.4	-	-	4.79	-	-	174.1	56.5	40.8	-	-	-
213	P	-	137.7	-	-	-	-	-	176.5	61.5	32.3	26.6	51.2	-
214	K	9.33	120.5	-	-	4.9	-	-	177.4	55.6	41.4	-	-	-
215	T	8.45	112.5	-	-	5.53	-	-	174.4	61.9	71	-	-	-
216	W	8.4	122.5	-	-	-	-	-	177.3	63.4	30.7	-	-	-
217	Y	7.34	110.9	-	-	-	-	-	179.3	60.7	38.7	-	-	-
218	S	7.8	115.2	-	-	4.27	-	-	177	61.2	63.2	-	-	-
219	A	8.82	128.5	-	-	2.85	-	-	177.2	55	19.2	-	-	-
220	E	7.77	118.6	-	-	-	-	-	178.6	58.3	26.6	-	-	-
221	Q	7.71	116.9	-	112.8	3.77	-	7.39	179.1	59.1	28.1	34.1	180.1	-
								6.99						
222	F	8.16	122.9	-	-	4.67	-	-	179.6	61.5	37.9	-	-	-
223	C	9.15	120.6	-	-	-	-	-	179.7	56.4	33.5	-	-	-
224	V	9.45	128.4	-	-	3.95	-	-	179.9	67.1	31.8	-	-	-
225	S	8.06	119.1	-	-	4.38	-	-	174.3	61.3	63	-	-	-
226	R	7.44	121.6	-	-	-	-	-	174.9	54.3	29	-	-	-
227	N	8.11	115.4	111.7	-	4.46	7.39	-	173.4	54.2	36.8	178.5	-	-
							6.86							
228	S	7.83	111.4	-	-	4.95	-	-	171.5	57.3	67.4	-	-	-
229	H	7.94	112.6	-	-	5.16	-	-	174.9	55.2	35.7	-	-	-

No.	Type	H^N	N	N^δ	N^ε	H^α	H^δ	H^ε	C	C^α	C^β	C^γ	C^δ	C^ε
230	L	10.34	125.5	-	-	-	-	-	180.8	56.3	43.4	-	-	-
231	T	8.5	116.6	-	-	-	-	-	173.5	64.3	69.9	-	-	-
232	S	6.72	118.1	-	-	-	-	-	-	56.3	64.5	-	-	-
233	V	-	-	-	-	-	-	-	177	62.8	33.8	20.7	-	-
234	T	9.7	115	-	-	4.73	-	-	174.4	61.6	70.9	-	-	-
235	S	7.73	112.8	-	-	4.91	-	-	174	57.6	66.2	-	-	-
236	E	9.71	122.1	-	-	4.3	-	-	177.9	60.3	29.4	-	-	-
237	S	8.54	113.8	-	-	-	-	-	176.8	61.6	62.7	-	-	-
238	E	8.16	124.8	-	-	4.12	-	-	176.6	60	30.6	-	-	-
239	Q	7.89	120.4	-	-	4.35	-	-	177.6	59.2	27.2	-	-	-
240	E	8.46	118.2	-	-	-	-	-	176.8	58.3	29.9	-	-	-
241	F	7.56	119.2	-	-	3.71	-	-	178.5	60.7	38.5	-	-	-
242	L	8.23	119.8	-	-	3.57	-	-	178.7	58.5	41.9	-	-	-
243	Y	8.93	118.4	-	-	4.44	-	-	179.5	60.5	35.8	-	-	-
244	K	7.9	124.8	-	-	3.74	-	-	180	59.6	31.7	-	-	-
245	T	7.46	117.4	-	-	3.35	-	-	175.5	65.7	67.3	-	-	-
246	A	8.17	123.6	-	-	3.94	-	-	180.4	54.5	19.3	-	-	-
247	G	7.16	101.8	-	-	3.94	-	-	175.5	47.7	-	-	-	-
248	G	9.22	107.2	-	-	4.29	-	-	174.6	45	-	-	-	-
						3.52								
249	L	7.43	122.1	-	-	4.51	-	-	175.6	54.1	42.3	-	-	-
250	I	7.73	120.5	-	-	4.56	-	-	176	61.2	38.1	-	-	-
251	Y	7.88	121.5	-	-	4.94	-	-	177.1	55.2	43.1	-	-	-
252	W	10.41	122.2	-	-	-	-	-	176.5	58.8	32.8	-	-	-
253	I	8.55	111.7	-	-	5.12	-	-	176.3	60.1	40.8	-	-	-
254	G	9.85	109.3	-	-	-	-	-	173.4	49.6	-	-	-	-
255	L	7.18	125.8	-	-	-	-	-	175.4	54	46	-	-	-
256	T	8.89	118.5	-	-	5.12	-	-	171.4	59.3	70	-	-	-
257	K	7.75	129	-	-	-	-	-	175.5	55.1	33.3	-	-	-
258	A	8.18	127.6	-	-	4.2	-	-	177.3	52	19.4	-	-	-
259	G	8.01	107.9	-	-	3.81	-	-	174.3	44.8	-	-	-	-
						4.01								
260	M	8.55	118.9	-	-	4.17	-	-	177.6	57.3	32.5	-	-	-
261	E	8.59	117.1	-	-	4.31	-	-	176.9	56.2	29.6	36.5	-	-
262	G	-	-	-	-	-	-	-	172.9	-	-	-	-	-
262	G	7.59	107.8	-	-	3.81	-	-	-	45.4	-	-	-	-
263	D	7.98	119.1	-	-	-	-	-	176.2	53.1	42.8	-	-	-
264	W	8.07	121.7	-	-	4.14	-	-	174.5	58.2	29.5	-	-	-
265	S	8.7	114.6	-	-	4.49	-	-	171.3	57.1	66.7	-	-	-
266	W	8.98	118.3	-	129.9	5.91	-	10.81	180.2	54.3	31.7	-	-	-
267	V	9.44	125.8	-	-	-	-	-	174.6	64.1	32.6	-	-	-
268	D	8.16	118.8	-	-	3.73	-	-	176.1	52.7	40.2	-	-	-
269	D	8.27	112.7	-	-	4.51	-	-	176.2	56.6	39.2	-	-	-
270	T	8.36	121.9	-	-	-	-	-	172.9	63.3	69.4	-	-	-

No.	Type	H^N	N	N^δ	N^ε	H^α	H^δ	H^ε	C	C^α	C^β	C^γ	C^δ	C^ε
271	P	-	142.1	-	-	-	-	-	176.6	63.9	31.9	28.3	51.3	-
272	F	9.24	127	-	-	4.83	-	-	175	58.9	40.7	-	-	-
273	N	8.91	128.8	114	-	4.91	7.16 7.94	-	175.6	51.8	38.8	177.5	-	-
274	K	9.05	127.9	-	-	-	-	-	178.4	60.2	33.1	-	-	-
275	V	8.11	120	-	-	3.85	-	-	178.8	66	32	-	-	-
276	Q	7.72	119.2	-	111.1	3.91	-	6.80 7.65	177.6	57.9	30.1	34.5	180	-
277	S	7.7	110.2	-	-	3.89	-	-	174.5	60.7	63.7	-	-	-
278	A	7.19	122.4	-	-	4.4	-	-	179.7	55.3	18.6	-	-	-
279	R	7.17	113	-	-	4.23	-	-	175.4	56.5	30	26.4	43.5	-
280	F	7.24	116	-	-	5.17	-	-	176	53.9	36.4	-	-	-
281	W	6.61	119	-	-	4.65	-	-	177.1	57.3	31.1	-	-	-
282	I	9.06	126.2	-	-	-	-	-	-	61.5	40.1	-	-	-
283	P	-	-	-	-	-	-	-	177.6	64.8	31.3	28.1	51.1	-
284	G	8.79	114.9	-	-	4.44 3.59	-	-	173.9	45	-	-	-	-
285	E	8.86	120.4	-	-	4.74	-	-	-	52.7	29.4	-	-	-
288	N	-	-	-	-	-	-	-	-	58.3	-	-	-	-
289	A	8.43	125.7	-	-	-	-	-	178.2	52.8	19.2	-	-	-
290	G	8.41	107.5	-	-	3.99	-	-	176.8	45.7	-	-	-	-
291	N	8.75	119.6	-	-	4.61	-	-	175.2	53.2	39	-	-	-
292	N	8.54	120.8	-	-	-	-	-	-	53	38.9	-	-	-
296	G	7.6	107.8	-	-	-	-	-	171.5	44.5	-	-	-	-
297	N	9.29	116.6	-	-	-	-	-	-	50.9	-	-	-	-
298	I	-	-	-	-	-	-	-	173.7	61.9	40.2	16.7	14.9	-
299	K	8.61	124	-	-	4.57	-	-	175.8	56.5	35.8	-	-	-
300	A	7.81	123.7	-	-	4.87	-	-	175.6	49.8	20.4	-	-	-
301	P	-	132.8	-	-	-	-	-	174.3	61.2	27.7	-	49.6	-
302	S	7.73	118.8	-	-	3.57	-	-	170.7	55.5	64.1	-	-	-
303	L	7.9	119.5	-	-	4.56	-	-	179	58.4	42.4	27.2	25.3	-
304	Q	7.32	114.6	-	113.2	3.92	-	6.90 7.95	173.1	55.2	27.5	34.5	175	-
305	A	6.89	123.9	-	-	4.73	-	-	180	53.6	24.3	-	-	-
306	W	9.08	117.6	-	-	5.92	-	-	176.3	55.4	31.5	-	-	-
310	P	-	-	-	-	-	-	-	176.3	63.6	32.1	27.4	50.9	-
311	C	8.17	126.3	-	-	-	-	-	173.8	57.5	46	-	-	-
312	D	8.65	114.3	-	-	4.81	-	-	176.7	53.9	40.1	-	-	-
313	K	8.02	120.3	-	-	4.36	-	-	175.4	-	32.6	-	-	-
314	T	7.6	108.7	-	-	4.55	-	-	174	58.4	69.6	22	-	-
315	F	8.36	123.3	-	-	5.08	-	-	176.5	55.7	43.5	-	-	-
316	L	7.68	118.4	-	-	-	-	-	173.9	56.3	41.3	-	-	-
317	F	8.27	113	-	-	5.87	-	-	172.6	56.5	43.4	-	-	-
318	I	7.53	118.6	-	-	5.65	-	-	177.3	59.3	41.6	17.5	15.2	-

No.	Type	H^N	N	N^δ	N^ε	H^α	H^δ	H^ε	C	C^α	C^β	C^γ	C^δ	C^ε
319	C	9.09	124.2	-	-	5.72	-	-	174.4	52	40.5	-	-	-
320	K	9.68	123.7	-	-	5.63	-	-	175.5	55.5	36.5	-	-	-
321	R	9.52	126.8	-	-	5.17	-	-	172	53.2	33.3	-	-	-
322	P	-	136.4	-	-	-	-	-	177	63.1	32.5	27.6	51.2	-
323	Y	8.04	120.7	-	-	4.71	-	-	174.3	57.8	39	-	-	-
324	V	7.92	129.3	-	-	4.23	-	-	173.4	58.9	33.5	-	-	-
325	P	-	137.9	-	-	-	-	-	176.7	63.1	32	27.2	51	-
326	S	8.14	114.6	-	-	4.38	-	-	174.3	58.5	63.8	-	-	-
327	E	8.18	123.2	-	-	-	-	-	174.6	54.4	29.8	-	-	-
328	P	-	138.2	-	-	-	-	-	177.8	64	31.9	27.6	50.9	-
329	G	8.79	110.7	-	-	4.17 3.97	-	-	175	45.5	-	-	-	-
330	S	8.17	115.9	-	-	4.48	-	-	175.1	59	64	-	-	-
331	E	8.91	122.5	-	-	-	-	-	176.5	57.6	29.7	-	-	-
332	N	8.34	118.5	112.8	-	-	7.58 6.93	-	175.4	53.6	38.8	177	-	-
333	L	7.97	121.6	-	-	-	-	-	177.2	55.7	42.2	-	-	-
334	Y	7.98	119.1	-	-	-	-	-	175.7	58.2	38.7	-	-	-
335	F	7.93	120.6	-	-	-	-	-	175.6	57.8	39.5	-	-	-
336	Q	8.24	121.6	-	-	-	-	-	176.3	56.2	29.3	-	-	-
337	G	8.05	109.4	-	-	4	-	-	174.3	45.5	-	-	-	-
338	S	8.2	115.6	-	-	-	-	-	174.7	58.5	64	-	-	-
339	A	8.35	125.2	-	-	-	-	-	177.5	53	19	-	-	-
340	W	7.9	118.7	-	-	-	-	-	176.2	57.2	29.7	-	-	-
341	S	7.89	116.6	-	-	-	-	-	173.5	58.2	64	-	-	-
342	H	8.16	121.1	-	-	-	-	-	172.8	54	29.5	-	-	-
343	P	-	137.9	-	-	-	-	-	176.8	63.4	32.1	27.5	50.9	-
344	Q	8.64	120.7	-	112.4	-	-	7.49	175.6	56.1	29.5	33.8	-	-
345	F	8.21	120.3	-	-	4.75	-	-	175.5	57.4	39.7	-	-	-
346	E	8.36	122.6	-	-	-	-	-	175.2	56.6	30.7	-	-	-
347	K	7.96	127.1	-	-	-	-	-	171.8	57.8	33.8	-	-	-

Table A-2: Alternate resonance assignment for residues of the TEV site and the Strep tag region.

No.	Type	H^N	N	N^δ	N^ε	H^α	H^δ	H^ε	C	C^α	C^β	C^γ	C^δ	C^ε
329	G								174.32					
330	S	8.28	115.36						174.77	58.52	63.91			
331	E	8.79	122.21						176.19	57.12	29.93			
332	N	8.38	118.84								38.87			
333	L	8.04	122.43						175.92	55.23			26.83	25.04

Tables

No.	Type	H^N	N	N^δ	N^ε	H^α	H^δ	H^ε	C	C^α	C^β	C^γ	C^δ	C^ε
334	Y	7.57	124.38						180.45	58.96	39.66			
335	F													
336	Q													
337	G	8.45	109.47						173.86	45.26				
338	S	8.28	116.06						173.44	58.3	64.25			
339	A	8.14	131.61						176.18	54.05	20.22			
340	W	8.09	124.88						176.19	57.23	29.65			
341	S	7.97	117.59						173.44	58.13				
342	H													
343	P													
344	Q													
345	F	8.27	120.78						175.4	57.62	39.86			
346	E	8.25	122.39						176.01	55.99	29.63			
347	K													

Table A-3: Microbial glycan array data of fluorescently labeled recombinant murine and human Langerin ECD.

Chart#	Murine (RFU)	[SEM] (RFU)	Human (RFU)	[SEM] (RFU)	Species / Origin	Type of structure /Cat.No.	BPS#
313	306	34	38	23	1-2 Mannan Acetobacter methanolieus MB135		322
307	-50	5	-4	12	Davanat		316
305	259	315	-8	10	Escherichia coli K13	CPS	314
18	859	94	23	8	Escherichia coli K235 LPS	L2143	22
304	75	156	-21	13	Escherichia coli K5	CPS	313
303	17615	1165	17	11	Escherichia coli K92	CPS	312
32	8634	2173	73	7	Escherichia coli O106		37
298	2504	550	420	46	Escherichia coli O107	PS	307
222	15	11	-4	5	Escherichia coli O111:B4 LPS- solution at 1 mg/mL	L5293-2ML (LPS) (Sigma)	231
210	317	238	1204	707	Escherichia coli O112ab	OPS	218
211	145	307	41	24	Escherichia coli O118	OPS	219
212	2	18	16	10	Escherichia coli O125	OPS	220
225	1005	373	-14	17	Escherichia coli O127:B8 LPS- solution at 1 mg/mL	L5668-2ML (LPS) (Sigma)	234
19	5039	541	6	5	Escherichia coli O128-B12 LPS	L2755	23
33	1056	921	183	21	Escherichia coli O130		38
297	320	59	-21	28	Escherichia coli O145	LPSOH	306
34	201	13	38	6	Escherichia coli O148		39
35	-537	323	20	9	Escherichia coli O150		40
213	-68	68	12	6	Escherichia coli O151	OPS	221
214	10	6	46	10	Escherichia coli O168	OPS	222
36	522	395	496	27	Escherichia coli O180		41
223	1097	483	14	24	Escherichia coli O26:B6 LPS- solution at 1 mg/mL	L5543-2ML (LPS) (Sigma)	232
30	52	20	20	5	Escherichia coli O29		35
31	-3731	575	59	17	Escherichia coli O40		36
205	161	31	163	5	Escherichia coli O49	OPS	213
206	10	4	-4	11	Escherichia coli O52	OPS	214
224	-364	221	-2	8	Escherichia coli O55:B5 LPS- solution at 1 mg/mL	L5418-2ML (LPS) (Sigma)	233
207	-131	54	25	7	Escherichia coli O58	OPS	215
208	181	149	38	16	Escherichia coli O61	LPSOH	216
294	282	86	209	103	Escherichia coli O71	PS	303
209	500	82	72	2	Escherichia coli O73	OPS	217
295	1041	221	265	30	Escherichia coli O85	PS	304
310	40	34	-33	4	Escherichia coli O86		319

296	10	46	13	21	Escherichia coli O99	PS	305
185	-220	100	-2	2	Francisella novicida OPS	OPS	193
186	16	2	-9	2	Francisella tularensis OPS	OPS	194
311	-122	27	7	20	Galactomannan DAVANT (160102) Pro-Pharmacenti		320
187	19	35	9	7	Klebsiella O1 OPS	OPS	195
194	191	27	82	16	Klebsiella O12 OPS	OPS	202
188	1001	520	67	9	Klebsiella O2a OPS	OPS	196
189	20	109	62	3	Klebsiella O2ac OPS	OPS	197
190	10	5	2	3	Klebsiella O3 OPS	OPS	198
191	995	576	31	4	Klebsiella O4 OPS	OPS	199
192	958	198	206	35	Klebsiella O5 OPS	OPS	200
193	420	160	248	119	Klebsiella O8 OPS	OPS	201
308	424	70	-12	17	Laminarin		317
306	6280	416	-6	4	Neisseria meningitidis Group C	CPS	315
39	2004	1036	30	4	Proteus mirabilis O10 (HJ4320)		44
264	8065	663	15	9	Proteus mirabilis O11 (9B-m)	PS	273
92	4964	1268	13	18	Proteus mirabilis O11 (PrK 24/57)	OPS	99
93	724	255	313	56	Proteus mirabilis O13 (PrK 26/57)	OPS	100
94	319	237	45	6	Proteus mirabilis O14a,14b (PrK 29/57)	OPS	101
95	216	47	208	43	Proteus mirabilis O16 (4652)	OPS	102
96	63	13	49	9	Proteus mirabilis O17 (PrK 32/57)	OPS	103
266	90	23	10	4	Proteus mirabilis O18 (PrK 34/57)	LPSOH	275
267	-229	162	152	17	Proteus mirabilis O20 (PrK 38/57)	LPSOH	276
174	499	245	-21	9	Proteus mirabilis O23a, 23b, 23c (CCUG 10701)	OPS	182
97	8033	947	10	28	Proteus mirabilis O23a,b,d (PrK 42/57)	OPS	104
98	1280	227	167	9	Proteus mirabilis O26 (PrK 49/57)	OPS	106
99	4658	333	126	14	Proteus mirabilis O27 (PrK 50/57)	OPS	107
100	249	305	36	11	Proteus mirabilis O28 (PrK 51/57)	OPS	108
101	-60	84	6	4	Proteus mirabilis O29a (PrK 52/57)	OPS	109
40	-181	55	25	2	Proteus mirabilis O29a, 29b (2002)		45
269	1167	330	22	5	Proteus mirabilis O33 (D52)	PS	278
37	304	93	42	11	Proteus mirabilis O3a, 3c (G1)		42
261	55	5	-29	5	Proteus mirabilis O3ab (S1959)	PS	270
102	43	29	51	8	Proteus mirabilis O40 (10703)	OPS	110
103	2259	866	-7	9	Proteus mirabilis O41 (PrK 67/57)	OPS	111
270	2463	1200	52	9	Proteus mirabilis O43 (PrK 69/57)	PS	279
272	4209	348	61	12	Proteus mirabilis O49 (PrK 75/57)	PS	281
262	9551	971	200	14	Proteus mirabilis O5 (PrK 12/57)	PS	271
41	-47	80	12	12	Proteus mirabilis O50 (TG332)		46

104	-61	74	57	4	Proteus mirabilis O51 (19011)*	LPS	112
42	1831	427	17	15	Proteus mirabilis O54a, 54b (10704)		47
273	-2	21	24	7	Proteus mirabilis O54ab (OE)	PS	282
43	-51	88	34	38	Proteus mirabilis O57 (TG319)		48
91	63	133	56	9	Proteus mirabilis O6 (PrK 14/57)	OPS	98
105	5323	821	70	18	Proteus mirabilis O74 (10705, OF)	OPS	113
106	-145	56	-28	30	Proteus mirabilis O75 (10702, OC)	OPS	114
107	-68	10	28	24	Proteus mirabilis O77 (3 B-m)	OPS	115
38	-252	80	-1	6	Proteus mirabilis O8 (TG326)		43
263	107	16	29	11	Proteus mirabilis O9 (PrK 18/57)	PS	272
121	153	108	-9	13	Proteus myxofaciens O60	OPS	129
122	4570	506	37	23	Proteus O56 (genomospecies 4)	OPS	130
265	220	315	36	23	Proteus penneri O17 (16)	PS	274
108	-39	51	46	13	Proteus penneri O31a (26)	OPS	116
268	203	263	56	8	Proteus penneri O31ab (28)	PS	277
109	-232	150	23	10	Proteus penneri O52 (15)	OPS	117
110	10	39	58	19	Proteus penneri O58 (12)	OPS	118
111	-195	30	-12	9	Proteus penneri O59 (9)	OPS	119
112	324	195	21	8	Proteus penneri O61 (21)	OPS	120
113	77	75	-15	17	Proteus penneri O62 (41)	OPS	121
114	166	205	33	9	Proteus penneri O63 (22)	OPS	122
45	64	30	-6	10	Proteus penneri O64a, 64b, 64d (39)		50
115	-179	76	-11	7	Proteus penneri O64a,b,c (27)	OPS	123
116	-125	71	-21	4	Proteus penneri O65 (34)	OPS	124
46	35	7	14	9	Proteus penneri O66 (2)		51
117	-8	20	43	19	Proteus penneri O67 (8)	OPS	125
118	-125	40	-12	27	Proteus penneri O68 (63)	OPS	126
47	6316	2783	30	13	Proteus penneri O69 (25)		52
119	244	210	-6	12	Proteus penneri O70 (60)	OPS	127
48	767	283	148	12	Proteus penneri O71 (42)		53
49	255	228	12	8	Proteus penneri O72a, 72b (4)		54
120	294	180	50	19	Proteus penneri O73a,b (103)	OPS	128
274	1606	349	20	11	Proteus penneri O73ac (75)	PS	283
44	-2	13	12	8	Proteus penneri O8 (106)		49
72	73	18	-3	7	Proteus vulgaris O1 (18984)*	LPS	78
74	207	136	27	8	Proteus vulgaris O12 (PrK 25/57)	OPS	80
75	100	126	21	9	Proteus vulgaris O13 (8344)	OPS	81
76	-231	89	59	29	Proteus vulgaris O15 (PrK 30/57)	OPS	82
77	52	56	52	20	Proteus vulgaris O17 (PrK 33/57)	OPS	83
78	-253	199	30	11	Proteus vulgaris O19a (PrK 37/57)	OPS	84

260	2700	575	36	6	Proteus vulgaris O2 (OX2)	PS	269
79	7721	1653	451	47	Proteus vulgaris O21 (PrK 39/57)*	LPS	85
80	-41	103	30	6	Proteus vulgaris O22 (PrK 40/57)	OPS	86
175	-40	55	-5	5	Proteus vulgaris O24 (PrK 47/57)	LPSOH	183
81	-110	69	25	4	Proteus vulgaris O25 (PrK 48/57)	OPS	88
82	769	676	14	5	Proteus vulgaris O34 (4669)*	LPS	89
83	3906	2987	242	17	Proteus vulgaris O37a,b (PrK 63/57)	OPS	90
84	5865	719	21	8	Proteus vulgaris O37a,c (PrK 72/57)	OPS	91
73	-113	66	38	6	Proteus vulgaris O4 (PrK 9/57)	OPS	79
85	-129	113	-6	9	Proteus vulgaris O44 (PrK 67/57)	OPS	92
86	-23	96	-1	5	Proteus vulgaris O45 (4680)	OPS	93
271	-200	84	49	14	Proteus vulgaris O47 (PrK 73/57)	Not stated	280
87	-76	96	39	7	Proteus vulgaris O53 (TG 276-10)	OPS	94
88	-64	45	1	5	Proteus vulgaris O54a,54c (TG 103)	OPS	95
89	-71	79	11	18	Proteus vulgaris O55 (TG 155)	OPS	96
90	13	18	29	5	Proteus vulgaris O65 (TG 251)	OPS	97
275	130	78	17	5	Proteus vulgaris O76 (HSC438)	PS	284
136	191	248	65	16	Providencia alcalifaciens O19	LPS/NaOH	144
135	87	12	9	3	Providencia alcalifaciens O19	LPS	143
134	15	20	-57	12	Providencia alcalifaciens O19	OPS	142
137	-20	9	30	7	Providencia alcalifaciens O21	OPS	145
138	12353	876	-20	20	Providencia alcalifaciens O23	OPS	146
139	7	13	11	11	Providencia alcalifaciens O27	OPS	147
140	-7	16	26	11	Providencia alcalifaciens O29	OPS	148
141	225	60	62	1	Providencia alcalifaciens O30	OPS	149
142	18	13	-10	32	Providencia alcalifaciens O32	OPS	150
143	55	7	35	11	Providencia alcalifaciens O36*	LPS-NH4OH	151
144	33	32	11	32	Providencia alcalifaciens O39	OPS	152
132	4802	894	-4	22	Providencia alcalifaciens O5	OPS	140
133	-57	34	5	29	Providencia alcalifaciens O6*	LPS	141
145	163	38	-14	21	Providencia rustigianii O14	OPS	153
146	-74	43	-33	19	Providencia rustigianii O16	OPS	154
147	18	21	13	9	Providencia rustigianii O34	OPS	155
124	93	47	5	12	Providencia stuartii O18	OPS	132
125	377	375	50	15	Providencia stuartii O20*	LPS	133
123	1329	324	-34	25	Providencia stuartii O4	OPS	131
126	594	309	-3	11	Providencia stuartii O43	OPS	134
127	28	25	17	16	Providencia stuartii O44	OPS	135
128	246	22	-3	11	Providencia stuartii O47	OPS	136
129	221	100	4	17	Providencia stuartii O47, Core 9	OPS	137

1	496	426	15	6	Providencia stuartii O49	PO49 Core-linked	1
130	810	437	25	24	Providencia stuartii O49, Core 1	OPS	138
2	4887	666	56	5	Providencia stuartii O52	PO52 Core-linked	2
131	-2	20	0	8	Providencia stuartii O57	OPS	139
4	3	10	-8	1	Pseudomonas aerguinosa O1 (Fisher immunotype 4)	PO1 Core-linked	4
13	20	14	6	5	Pseudomonas aerguinosa O10	L8643	15
65	185	44	-2	6	Pseudomonas aerguinosa O10 10a,10b	IATS 10, OPS	71
66	3836	2153	2	22	Pseudomonas aerguinosa O10 10a,10c	IATS 19, OPS	72
67	5	143	15	13	Pseudomonas aerguinosa O11 11a,11b	IATS 11, OPS	73
68	59	11	-28	13	Pseudomonas aerguinosa O12 12	IATS 12, OPS Habs 12	74
6	65	39	2	4	Pseudomonas aerguinosa O13 (Sandvik serotype II)	PO13 Core-linked	6
69	558	302	107	12	Pseudomonas aerguinosa O13 13a,13c	IATS 14, OPS	75
70	136	19	-21	8	Pseudomonas aerguinosa O14 14	IATS 17,OPS Meitert X	76
71	425	552	16	8	Pseudomonas aerguinosa O15 15	IATS 15, OPS	77
50	-6	27	8	7	Pseudomonas aerguinosa O2 (2a),2d,2f	IATS 10 , OPS	55
5	-65	27	8	5	Pseudomonas aerguinosa O2 (Fisher immunotype 3)	PO2 Core-linked	5
51	329	379	10	1	Pseudomonas aerguinosa O2 2a,2b	IATS 16 OPS	56
52	856	692	22	8	Pseudomonas aerguinosa O2 2a,2b,2e	IATS NO, OPS	57
53	-41	27	14	7	Pseudomonas aerguinosa O2 2a,2d	IATS 5 OPS	58
54	1243	749	7	5	Pseudomonas aerguinosa O2 Immuno 7	IATS 18, OPS	59
55	-18	10	8	6	Pseudomonas aerguinosa O3 3a,3b	IATS NO, OPS	60
56	1224	646	16	5	Pseudomonas aerguinosa O3 3a,3b,3c	IATS 3, OPS	61
57	-45	22	25	8	Pseudomonas aerguinosa O3 3a,3d	IATS NO, OPS	62
3	344	247	5	2	Pseudomonas aerguinosa O4 (Habs serotype 4)	PO4 Core-linked	3
58	-78	272	22	5	Pseudomonas aerguinosa O4 4a,4c	IATS NO, OPS	63
59	-4	10	16	6	Pseudomonas aerguinosa O6 6a	IATS 6, OPS	64
60	517	105	74	12	Pseudomonas aerguinosa O6 6a,6c	IATS NO, OPS	65
61	166	160	15	13	Pseudomonas aerguinosa O6 Immuno 1	IATS NO, OPS	66
8	120	66	11	5	Pseudomonas aerguinosa O6a (Habs serotype6, fraction IIa)	PO6a Core-linked-O-unit	8
9	0	50	-3	2	Pseudomonas aerguinosa O6a (Habs serotype6, fraction IIb)	PO6a unsubstituted core	9
62	824	140	16	9	Pseudomonas aerguinosa O7 7a,7b,7c	IATS 7,LPS	67
63	867	801	24	10	Pseudomonas aerguinosa O7 7a,7b,7d	IATS 8,LPS	68
64	34	63	21	4	Pseudomonas aerguinosa O7 7a,7d	IATS NO, LPS	69
7	44	44	24	3	Pseudomonas aerguinosa O9 (9a, 9b, 9d)	PO9 Core-linked	7
20	24	17	21	3	Salmonella enterica abortus equi LPS	L5886	25
299	227	128	101	8	Salmonella enterica O17	PS	308

300	463	53	127	5	Salmonella enterica O28	PS	309
301	169	42	41	31	Salmonella enterica O47	PS	310
302	44	30	28	14	Salmonella enterica O55	PS	311
15	-44	37	-6	6	Salmonella enteritidis dodeca saccharide	1262	17
22	308	86	53	31	Salmonella enteritidis LPS	L2012	27
14	11	3	-3	4	Salmonella typhimurium dodeca saccharide	4809	16
16	551	266	114	36	Salmonella typhimurium LPS	L2262	18
10	168	66	214	23	Salmonella typhimurium SL 11881 (Re mut)	LPS-L9516	12
12	243	37	30	3	Salmonella typhimurium SL 684 (Rc mut)	LPS-L5891	14
11	677	237	31	7	Salmonella typhimurium TV 119 (Ra mut)	LPS-L6016	13
21	150	120	23	20	Salmonella typhosa LPS	L2387	26
17	66	30	20	14	Serratia marcescens LPS	L6136	20
25	5598	1474	29	13	Shigella bodyii type10		30
23	25	8	138	50	Shigella bodyii type2		28
24	181	48	36	9	Shigella bodyii type4		29
195	78	32	48	4	Shigella boydii type 1	LPSOH	203
199	-49	67	142	88	Shigella boydii type 11	OPS	207
200	105	15	66	12	Shigella boydii type 12	OPS	208
292	6238	981	532	164	Shigella boydii type 13	LPSOH	301
293	129	11	83	12	Shigella boydii type 14	LPSOH	302
201	86	68	13	6	Shigella boydii type 15	OPS	209
202	431	244	75	33	Shigella boydii type 16	OPS	210
203	14	17	-8	5	Shigella boydii type 17	OPS	211
204	47	16	97	46	Shigella boydii type 18	OPS	212
196	42	21	19	7	Shigella boydii type 3	OPS	204
197	26	3	14	2	Shigella boydii type 5	OPS	205
289	1455	485	30	9	Shigella boydii type 6	PS	298
290	1035	387	140	16	Shigella boydii type 7	PS	299
291	34	21	93	10	Shigella boydii type 8	PS	300
198	155	42	8	3	Shigella boydii type 9	OPS	206
28	-25	29	-2	3	Shigella dysennteriae type 11		33
29	1163	975	274	60	Shigella dysennteriae type 13		34
26	328	103	28	5	Shigella dysennteriae type 3		31
27	31	10	4	4	Shigella dysennteriae type 8 (batch 12)		32
288	491	254	232	80	Shigella dysenteriae type 1	PS	297
215	437	184	34	3	Shigella dysenteriae type 2	LPSOH	223
216	14	7	18	14	Shigella dysenteriae type 4	OPS	224
217	21	4	56	8	Shigella dysenteriae type 5	OPS	225
218	-81	108	39	12	Shigella dysenteriae type 6 SR-strain	SR-strain	226
219	-3	9	19	6	Shigella dysenteriae type 7	OPS	227

220	-1088	878	-8	11	Shigella dysenteriae type 8 (Russian)	OPS	228
221	323	312	14	7	Shigella dysenteriae type 9	OPS	229
276	580	534	37	5	Shigella flexneri type 1a	PS	285
277	205	25	602	51	Shigella flexneri type 1b	PS	286
278	-70	46	-29	5	Shigella flexneri type 2a	PS	287
279	132	8	452	67	Shigella flexneri type 2b	PS	288
280	79	32	200	13	Shigella flexneri type 3a	PS	289
281	1420	320	1190	74	Shigella flexneri type 3b	PS	290
282	357	112	149	42	Shigella flexneri type 4a	PS	291
283	1144	141	1493	114	Shigella flexneri type 4b	PS	292
284	519	215	195	8	Shigella flexneri type 5b	PS	293
286	2005	1599	147	66	Shigella flexneri type 6	PS	295
285	682	74	2029	303	Shigella flexneri type 6a	PS	294
287	131	49	197	39	Shigella flexneri type X	PS	296
226	7605	3760	29	7	Streptococcus pneumoniae type 1 (Danish type 1)	161-X // Capsular PS	235
233	97	171	-11	12	Streptococcus pneumoniae type 12 (Danish type 12F)	193-X// Capsular PS	242
234	-14	4	43	8	Streptococcus pneumoniae type 14 (Danish type 14)	197-X// Capsular PS	243
235	-69	58	-8	10	Streptococcus pneumoniae type 17 (Danish type 17F)	201-X// Capsular PS	244
236	3	21	-8	3	Streptococcus pneumoniae type 19 (Danish type 19F)	205-X// Capsular PS	245
227	-188	117	9	17	Streptococcus pneumoniae type 2 (Danish type 2)	165-X// Capsular PS	236
237	-47	94	6	4	Streptococcus pneumoniae type 20 (Danish type 20)	209-X// Capsular PS	246
238	-44	50	15	5	Streptococcus pneumoniae type 22 (Danish type 22F)	213-X// Capsular PS	247
239	7610	3930	12	5	Streptococcus pneumoniae type 23 (Danish type 23F)	217-X// Capsular PS	248
240	51802	794	1123	333	Streptococcus pneumoniae type 26 (Danish type 6B)	225-X// Capsular PS	249
228	4197	1970	1	6	Streptococcus pneumoniae type 3 (Danish type 3)	169-X// Capsular PS	237
241	-36	5	27	7	Streptococcus pneumoniae type 34 (Danish type 10A)	229-X// Capsular PS	250
229	-44	127	-3	16	Streptococcus pneumoniae type 4 (Danish type 4)	173-X// Capsular PS	238
242	-119	175	4	5	Streptococcus pneumoniae type 43 (Danish type 11A)	233-X// Capsular PS	251
230	490	588	-25	7	Streptococcus pneumoniae type 5 (Danish type 5)	177-X// Capsular	239

						PS	
243	-28	39	19	10	Streptococcus pneumoniae type 51 (Danish type 7F)	237-X// Capsular PS	252
244	-397	202	1	13	Streptococcus pneumoniae type 54 (Danish type 15B)	241-X// Capsular PS	253
245	-27	6	18	19	Streptococcus pneumoniae type 56 (Danish type 18C)	245-X// Capsular PS	254
246	7824	1874	-1	7	Streptococcus pneumoniae type 57 (Danish type 19A)	249-X// Capsular PS	255
247	-37	32	-10	4	Streptococcus pneumoniae type 68 (Danish type 9V)	253-X// Capsular PS	256
248	-32	32	86	13	Streptococcus pneumoniae type 70 (Danish type 33F)	257-X// Capsular PS	257
231	-97	76	4	15	Streptococcus pneumoniae type 8 (Danish type 8)	185-X// Capsular PS	240
232	20	8	14	9	Streptococcus pneumoniae type 9 (Danish type 9N)	189-X// Capsular PS	241
309	12264	3236	14535	771	Yeast Mannan		318
312	19942	737	20693	265	Yeast Mannan Sigma M-3640		321
153	98	32	-8	4	Yersinia pestis 1146-25	Core oligo saccharide	161
177	217	102	223	32	Yersinia pestis 260(11)-37C-186	LPS	185
178	76	41	239	7	Yersinia pestis 260(11)-37C-187	LPS	186
179	54	52	64	16	Yersinia pestis 260(11)-37C-416	LPS	187
180	466	158	159	51	Yersinia pestis 260(11)-37C-417	LPS	188
183	239	246	5	24	Yersinia pestis I-2377-25C	OS	191
184	312	167	145	9	Yersinia pestis I-2377-37C	LPS	192
249	1337	452	38	19	Yersinia pestis KM218-6C	OS	258
176	43	99	398	21	Yersinia pestis KM260(11)-6C	LPS	184
252	982	308	153	28	Yersinia pestis KM260(11)-wabC/waaL	OS	261
251	-93	90	15	12	Yersinia pestis KM260(11)-wabD/waaL	OS	260
250	-68	58	7	10	Yersinia pestis KM260(11)-yjhW-6C	OS	259
181	-18	3	31	4	Yersinia pestis P-1680-25C	OS	189
182	103	203	89	38	Yersinia pestis P-1680-37C	LPS	190
156	152	54	421	75	Yersinia pestis, 0KM218-37	LPS	164
152	-54	46	155	24	Yersinia pestis, 1146-25	LPS	160
154	870	110	128	4	Yersinia pestis, 1146-37	LPS	162
155	142	78	45	20	Yersinia pestis, 1146-37	Core oligo saccharide	163
172	702	122	15	25	Yersinia pestis, 11M-25	LPS	180
173	-48	88	24	7	Yersinia pestis, 11M-37	LPS	181
171	8	3	26	8	Yersinia pestis, KIMD1-25	Core oligo	179

						saccharide		
170	616	122	-17	8	Yersinia pestis, KIMD1-37	Core saccharide	oligo	178
159	83	146	7	7	Yersinia pestis, KM218-25	Core saccharide	oligo	167
158	16	53	1361	598	Yersinia pestis, KM218-25	LPS		166
157	15	15	18	2	Yersinia pestis, KM218-37	Core saccharide	oligo	165
162	475	178	3275	1102	Yersinia pestis, KM260(11)-Δ0186	LPS		170
163	35	16	-9	9	Yersinia pestis, KM260(11)-Δ0186	Core saccharide	oligo	171
149	7	22	24	14	Yersinia pestis, KM260(11)-Δ0187	Core saccharide	oligo	157
148	378	72	452	43	Yersinia pestis, KM260(11)-Δ0187	LPS		156
161	109	40	26	23	Yersinia pestis, KM260(11)-ΔpmrF	Core saccharide	oligo	169
160	1460	308	575	82	Yersinia pestis, KM260(11)-ΔpmrF	LPS		168
150	755	300	91	28	Yersinia pestis, KM260(11)-Δrfe	LPS		158
151	-7	18	-62	11	Yersinia pestis, KM260(11)-Δrfe	Core saccharide	oligo	159
166	186	85	174	59	Yersinia pestis, KM260(11)-ΔwaaL	LPS		174
165	-32	21	-5	9	Yersinia pestis, KM260(11)-ΔwaaQ	Core saccharide	oligo	173
164	164	146	577	200	Yersinia pestis, KM260(11)-ΔwaaQ	LPS		172
167	141	44	390	34	Yersinia pestis, KM260(11)-25	LPS		175
168	105	13	33	22	Yersinia pestis, KM260(11)-25	Core saccharide	oligo	176
169	108	109	-13	6	Yersinia pestis, KM260(11)-37	Core saccharide	oligo	177
254	112	53	1	5	Yersinia pseudotuberculosis 85pCad-20C	OS		263
253	-71	70	-10	13	Yersinia pseudotuberculosis 85pCad-37C	OS		262
255	13011	3710	6265	1156	Yersinia pseudotuberculosis O:2a	PS		264
256	456	176	134	8	Yersinia pseudotuberculosis O:2a-dhmA	PS		265
257	421	116	2	6	Yersinia pseudotuberculosis O:2c	PS		266
258	5731	784	380	62	Yersinia pseudotuberculosis O:3	PS		267
259	456	30	225	19	Yersinia pseudotuberculosis O:4b	PS		268

A.4 Table of Figures

Appendix B Supplementary Data

B.1 Codon-optimized Langerin gene

ATGGCACGCTTCATGGGCACGATTAGCGATGTTAAAACCAATGTTCAACTGCTGAAAGGCCGTGTGGACAATA
TC
TCGACCCTGGATAGTGAAATTAAGAAAAACAGCGATGGCATGGAAGCGGCCGGTGTCCAGATCCAAATGGTG
AAT
GAATCACTGGGTTATGTCCGTTCGCAGTTTCTGAAACTGAAAACCAGTGTTGAAAAAGCAAACGCTCAGATTC
AA
ATCCTGACCCGCTCATGGGAAGAAGTGTCGACGCTGAATGCCCAGATTCCGGAACTGAAAAGCGATCTGGAAA
AA
GCGTCTGCCCTGAACACGAAAATCCGTGCACTGCAGGGCAGCCTGGAAAACATGTCTAAACTGCTGAAACGCC
AA
AATGACATTCTGCAGGTGGTTAGCCAAGGCTGGAAATACTTCAAGGGTAACTTCTACTACTTCAGTCTGATCCC
G
AAAACCTGGTACTCCGCCGAACAGTTTTGCGTGAGTCGTAACTCCCATCTGACCAGCGTTACGAGCGAATCTGA
A
CAAGAATTTCTGTATAAAACCGCTGGCGGTCTGATTTACTGGATCGGTCTGACGAAAGCGGGCATGGAGGGTG
AT
TGGTCATGGGTTGATGACACCCCGTTTAATAAAGTCCAGTCGGCGCGTTTCTGGATTCCGGGCGAACCGAACA
AT
GCCGGTAACAATGAACACTGTGGCAACATCAAAGCACCGAGCCTGCAGGCATGGAATGATGCTCCGTGCGACA
AA
ACGTTTCTGTTCATTTGTAAACGCCCGTACGTTCCGTCTGAACCGGGATCCGAAAATCTGTACTTCCAAGGTTC
G
GCGTGGTCCCATCCGCAGTTTGAAAAATAA

B.2 Estimation of dissociation constant from shifting peaks in ^{1}H-^{15}N HSQC NMR titrations

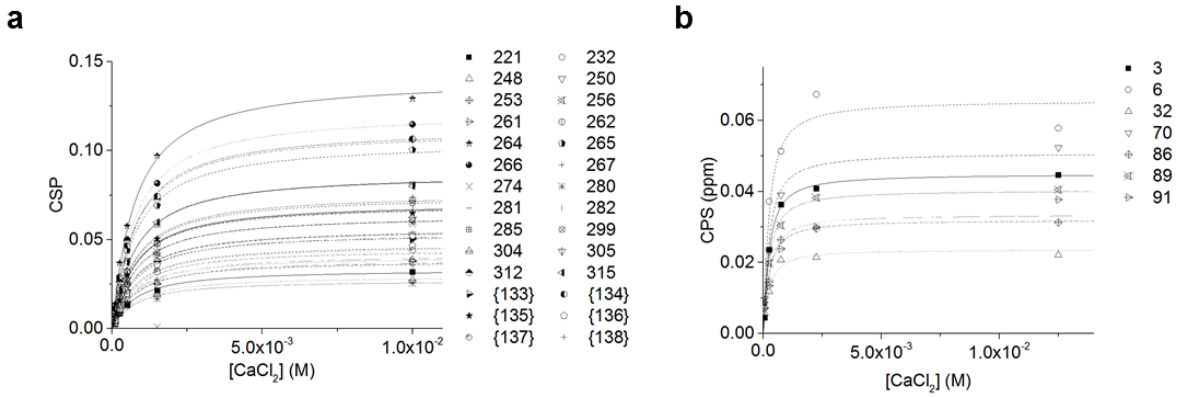


Figure B-1: ^{1}H-^{15}N HSQC NMR titrations of Ca^{2+} interacting with Langerin CRD at pH 6 (**a**) and pH 7 (**b**). Data was fitted to eq. (2-37) to obtain K_d values. Values are given **Table 4-4**.

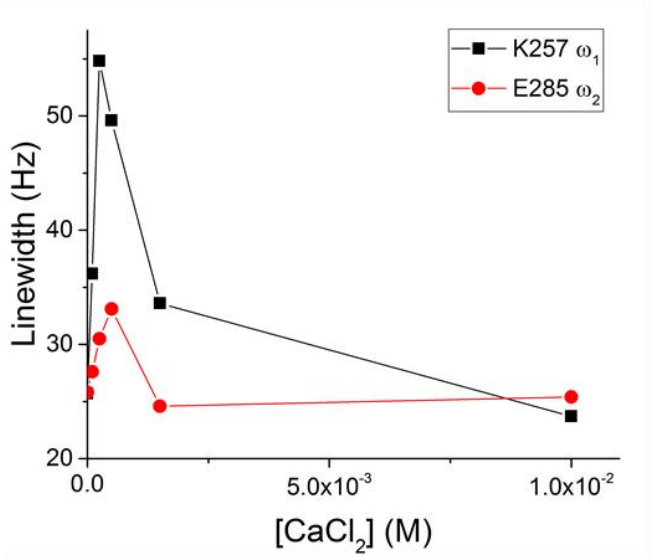
B.3 Slow exchanging populations in cis and trans conformation in apo Langerin

Table B-1: Residue in Langerin CRD apo form that undergo cis trans isomerization as indicated by a second peak. Fraction population in cis conformation was calculated from the peak integrals of cis and apo population. STDEV was determined from three independent measurements.

Residue	Fraction in cis
G247	76%
W252	71%
I253	75%
G254	72%
K257	70%
A258	63%
G259	71%
M260	78%
G262	N.A.
W264	69%
S265	68%
W266	68%
S277	79%
A278	76%
R279	80%
F280	77%
W281	82%
E285	69%
N297	63%
A300	75%
S302	79%
Q304	75%
A305	78%
Average	**73%**
STDEV*	**10%**

B.4 Ca^{2+} binding induces intermediate exchange broadening of several shifting peaks

a



b

	K257		E285	
$[CaCl_2]$ (M)	Line width ω_2 (Hz)	Line width ω_1 (Hz)	Line width ω_2 (Hz)	Line width ω_1 (Hz)
0.0E-00	24.9	25.7	25.8	21.0
1.0E-04	21.3	36.2	27.6	21.6
2.5E-04	34.4	54.8	30.5	21.6
5.0E-04	22.4	49.6	33.1	20.5
1.5E-03	23.0	33.6	24.6	20.2
1.0E-02	29.1	23.7	25.4	20.9

Figure B-2: Binding of Ca^{2+} results in intermediate exchange broadening in several shifting peaks. (**a**) Linewidths of K257 ^{15}N peak dimension and E285 ^{1}H peak dimension during Ca^{2+} titration. Both peaks have a maximum between 250 µM and 500 µM ligand concentration roughly corresponding to 1/3 K_d. (**b**) The table summarizes the linewidths of both investigated peaks at different ligand concentrations.

B.5 Estimation of the dissociation constant from decrease in apo trans population

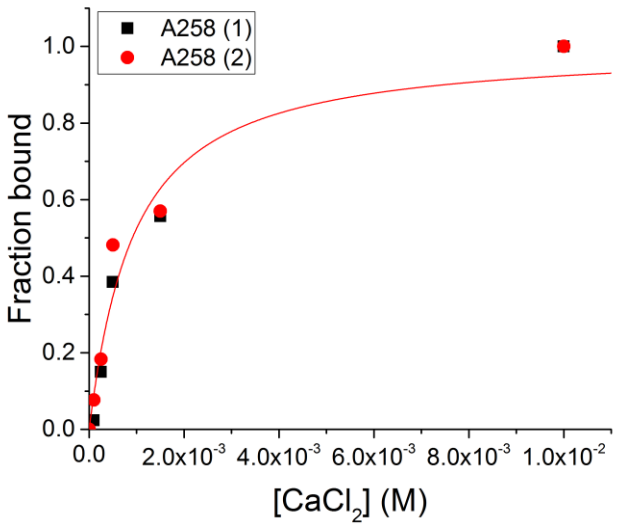


Figure B-3: Binding of Ca^{2+} results in decrease of peak intensities of trans population which can be used to obtain the K_d. Plot of peak intensities of A258 trans peak from two independent measurements at pH 6. The fraction bound was calculated by subtraction of the normalized peak intensities (representing free protein) from 1. The data was fitted globally to a modified version of eq. (2-37) and awarded a K_d of 810±110 µM.

B.6 B-factors observed in crystal structure support the observed increase in flexibility in the short and long loops

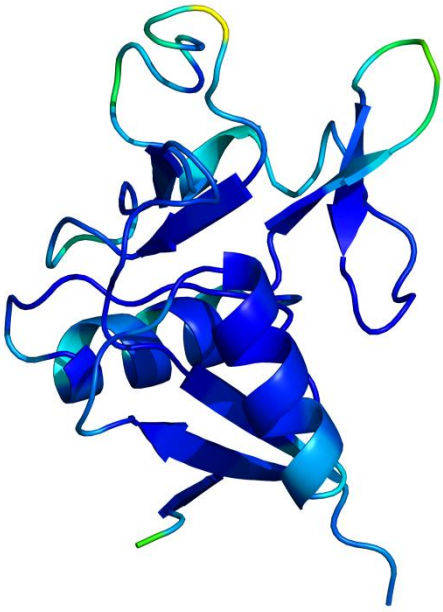


Figure B-4: B-factors heat-map color-coded on the structure of holo Langerin CRD (pdb entry: 3p5f) indicate increased flexibility in the long and short loop region.

B.7 Relaxation dispersion measurements of apo and holo Langerin at pH 6

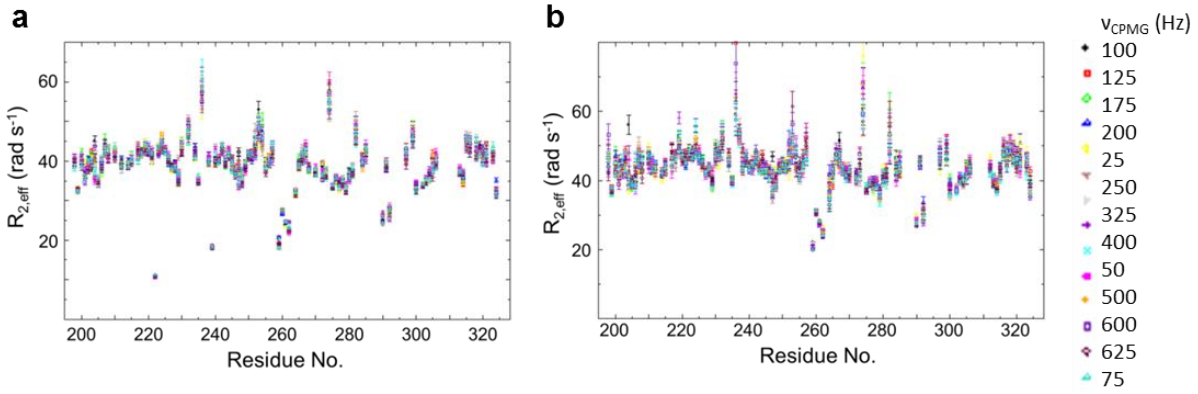


Figure B-5: Relaxation dispersion measurements reveal local field-dependent variations in R2 indicative for exchange processes on the micro- to millisecond timescale.

B.8 Model-Free Approach

Table B-2: Anisotropic rotational diffusion tensor determined for holo Langerin CRD using HYDRONMR 7

Diffusion Parameter	Value
τ_m (s)	7.9E-09
D_z (s^{-1})	2.381E+07
D_y (s^{-1})	1.927E+07
D_x (s^{-1})	1.998E+07

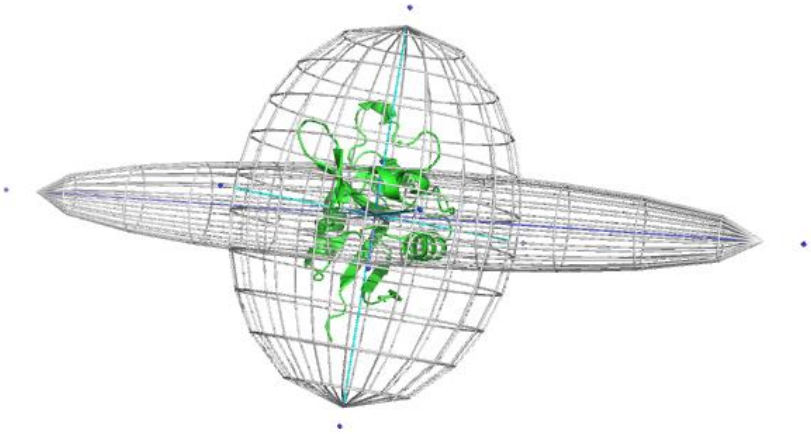


Figure B-6: Anisotropic rotational diffusion tensor for holo Langerin CRD (pdb entry: 3p5f) estimated by HYDRONMR 7 (round shaped) and after six iterations using the relax algorithm (long shape).

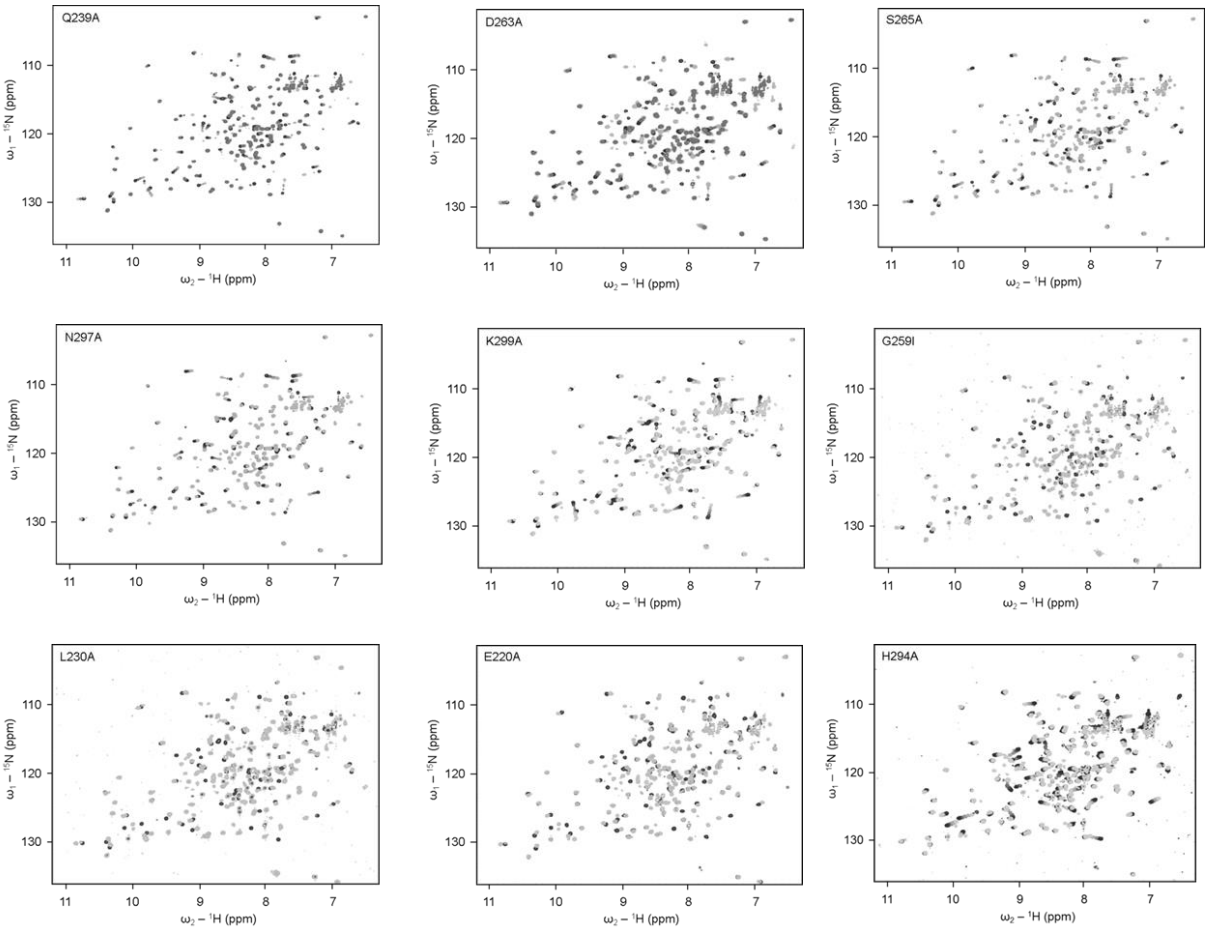


Figure B-7: ^{1}H-^{15}N HSQC spectra of Langerin CRD mutants as indicated in each spectra overlay. Overlay shows apo (light grey) and holo (dark grey) form, respectively. In case of titrations, titration spectra are shown in intermediate greyscale. All proteins were folded and bound Ca2+. Spectra were recorded at pH 6 and at 298 K.

B.9 Prolyl cis/trans Ratio of Langerin CRD Mutants

Table B-3: P286 prolyl cis/trans ratio remains unaffected by single-residue mutations at pH 6

Langerin CRD construct	Fraction in cis	STDEV
WT*	73%	10%
E220A	78%	4%
L230A	76%	3%
Q239A	65%	9%
K257A	75%	10%
G259I	77%	5%
D263A	78%	5%
S265A	77%	5%
H294A	80%	4%
N297A	83%	5%
K299A	67%	5%

*cis/trans ratio was determined as average from three independent measurements. Error was propagated from the standard deviations of each measurement.

B.10 Binding Isotherms of Langerin CRD Mutants Binding to Ca^{2+}

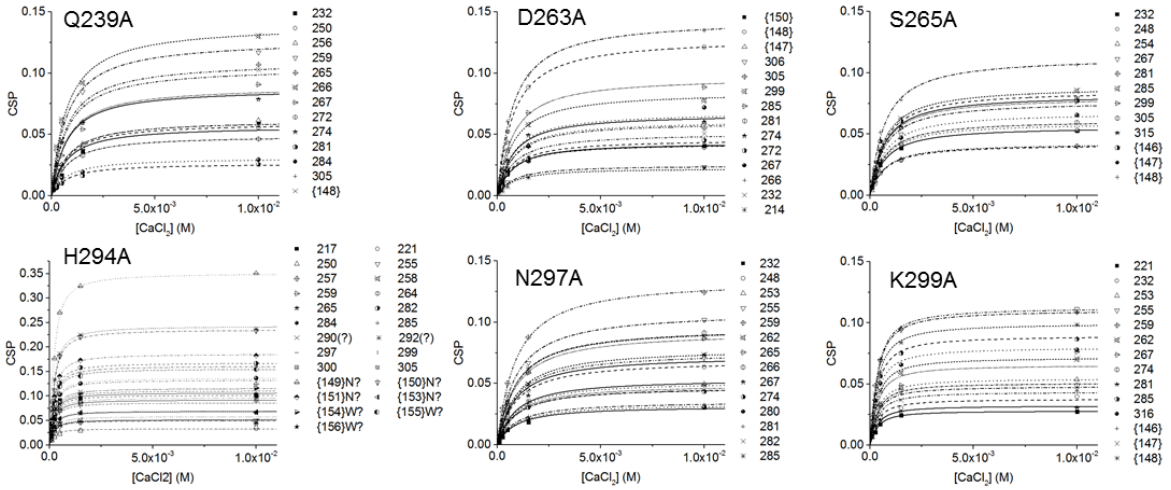


Figure B-8: ^{1}H-^{15}N HSQC NMR titrations of Langerin CRD mutants at pH 6. Data was fitted to eq. (2-37) to obtain K_d values. Values are listed in **Table 4-4**.

B.11 CHESPA Analysis of Langerin CRD Mutants at pH 6

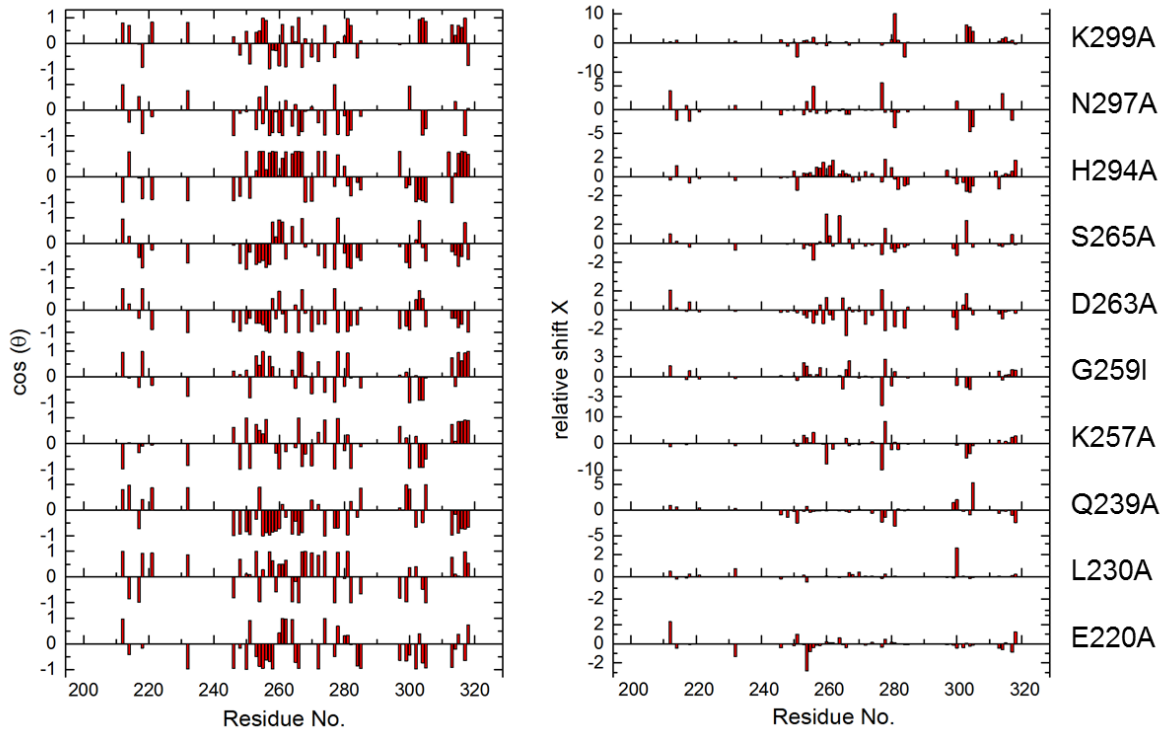


Figure B-9: CHESPA analysis of ten Langerin CRD mutants derived from ^{1}H-^{15}N HSQC NMR measurements at pH 6. With the exception of H294A, no clear pattern of activating and deactivating shifts was discerned.

B.12 pK_A Calculations

Table B-4: pKa calculations for H294 side chain of holo Langerin CRD (pdb entry: 3p5f, Feinberg et al., 2011) awarded highly diverging values.

Method	Principle	pKa	Citation
MOE (PROPKA)	Empirical	**6.3**	(Li et al., 2005)
Rosetta CL	Includes neighboring side chains, electrostatic calculations	**4.1**	(Kilambi and Gray, 2012)
H++ (v3.2)	AMBER force field to optimize protons, electrostatic calculations	**1.1**	(Anandakrishnan et al., 2012)
PROPKA3	Empirical including side chain effects	**5.85**	(Olsson, 2011)

Table B-5: pKa calculations for H294, E285 and E293 side chains of apo and holo Langerin CRD performed with different snapshots from states observed in the MD simulations by Stevan Aleksic using two different computational methods.

Structure	Method	pKa (H294)	pK_a (E285)	pK_a (E293)
apo open	PROPKA3[a]	5.87	6.35	5.29
	Rosetta CL[b]	3.6	6.3	5.9
apo intermediate	PROPKA3	5.8	5.42	6.00
	Rosetta CL	4.8	5.9	6.2
apo closed	PROPKA3	6.56	5.47	5.17
	Rosetta CL	6.5	6.5	4.7
holo open	PROPKA3	5.99	5.08	4.74
	Rosetta CL	-	-	-
holo intermediate	PROPKA3	5.71	5.46	4.35
	Rosetta CL	5.7	3.1	2.4
holo closed	PROPKA3	6.05	5.00*	5.59*
	Rosetta CL	6.3	2.8	1.9

[a] (Olsson, 2011)

[b] (Kilambi and Gray, 2012)

B.13 H294 and K257 occupy the canonic accessory Ca^{2+} sites.

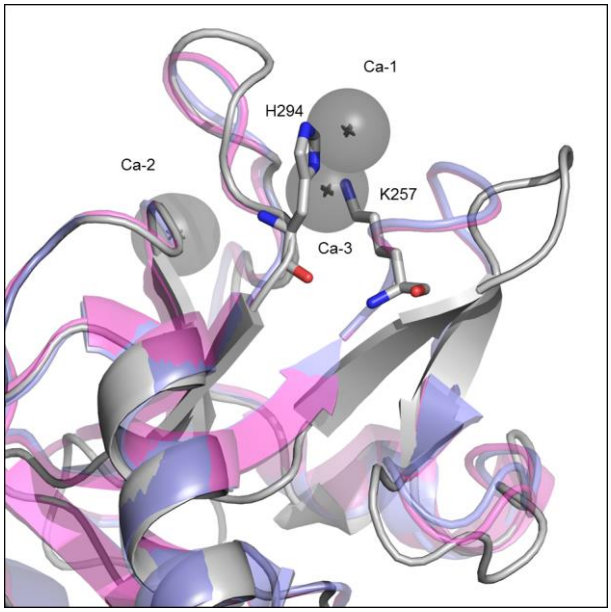


Figure B-10: Cartoon representation of Langerin CRD (light grey) with residues H294 and K257 as sticks. Transparent cartoon representations of aligned CRD structures of DC-SIGN (slate, pdb entry: 1SL4 (Guo et al., 2004)) and DC-SIGNR (light magenta, pdb entry: 1K9J (Feinberg et al., 2001)). The conserved Ca-1 and Ca-3 positions in DC-SIGN and DC-SIGNR are occupied by H294 and K257 side chains in Langerin facilitating the interactions of the small and long loop.

B.14 Carbohydrate Binding to Langerin CRD

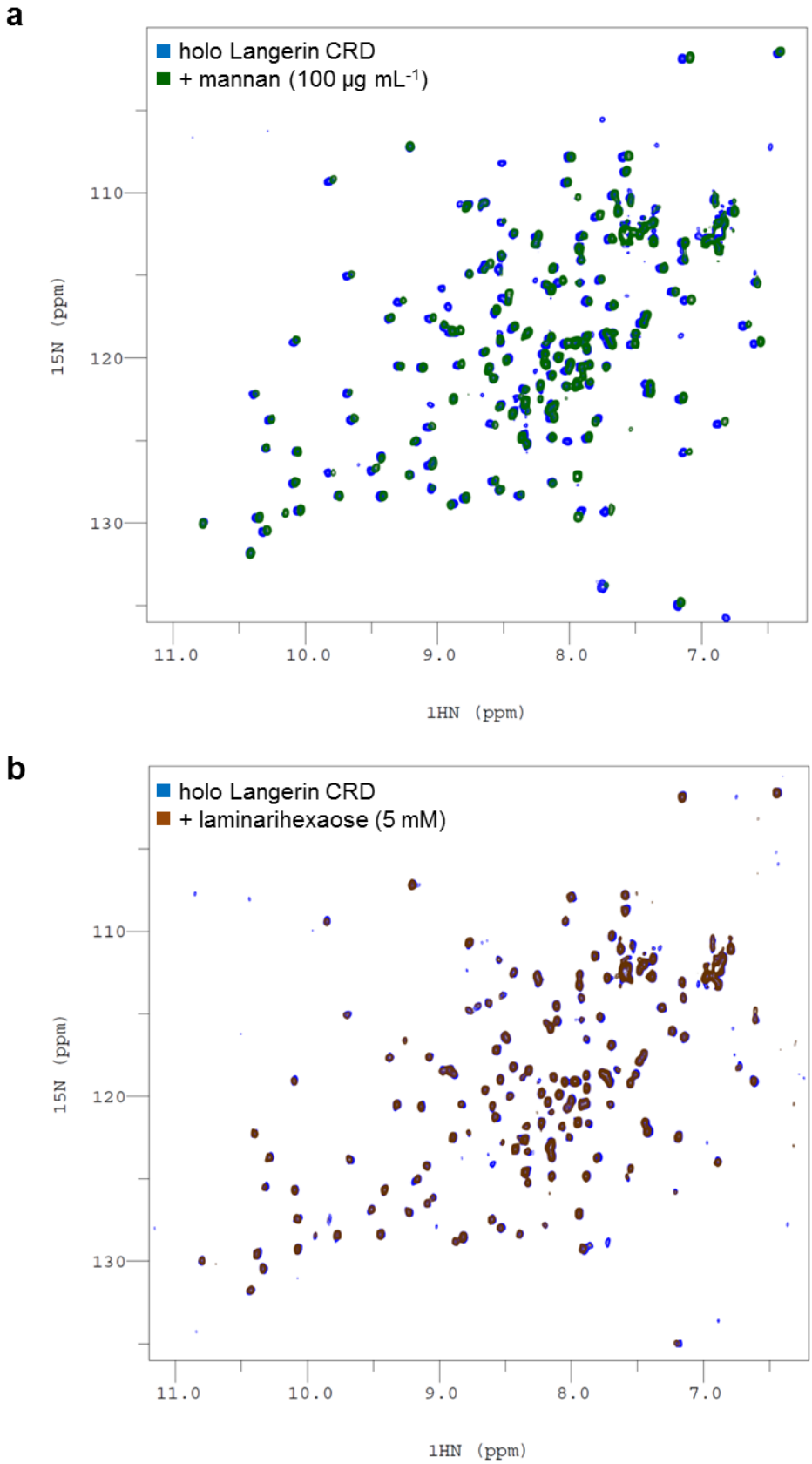


Figure B-11: Langerin interacting with complex carbohydrates. (**a**) Binding of mannan (100 µg mL^{-1}) induces minor shifts over the entire spectrum of Langerin CRD (200 µM). (**b**) Addition of 5 mM laminarihexaose showed no chemical shift perturbation in Langerin CRD spectrum. Spectra were recorded at 200 µM protein concentration, 10 mM $CaCl_2$, pH 6, 298 K.

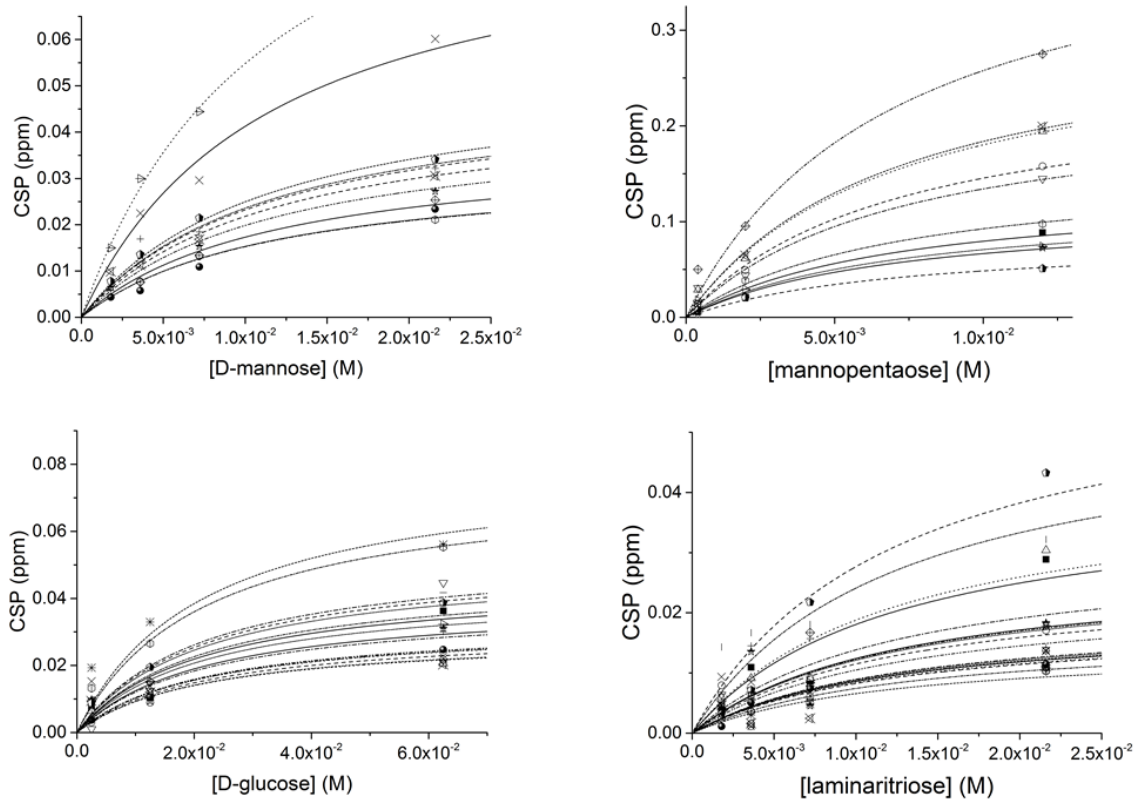


Figure B-12: ^{1}H-^{15}N HSQC NMR titrations of carbohydrate ligands interacting with Langerin CRD. Data was fitted to eq. (2-37) to obtain K_d values. Titration curves did not reach saturation. Spectra were recorded at 200 µM protein concentration stepwise adding the carbohydrate in presence of 10 mM $CaCl_2$, pH 6, 298 K.

B.15 STD NMR Measurements of mannobiose 1

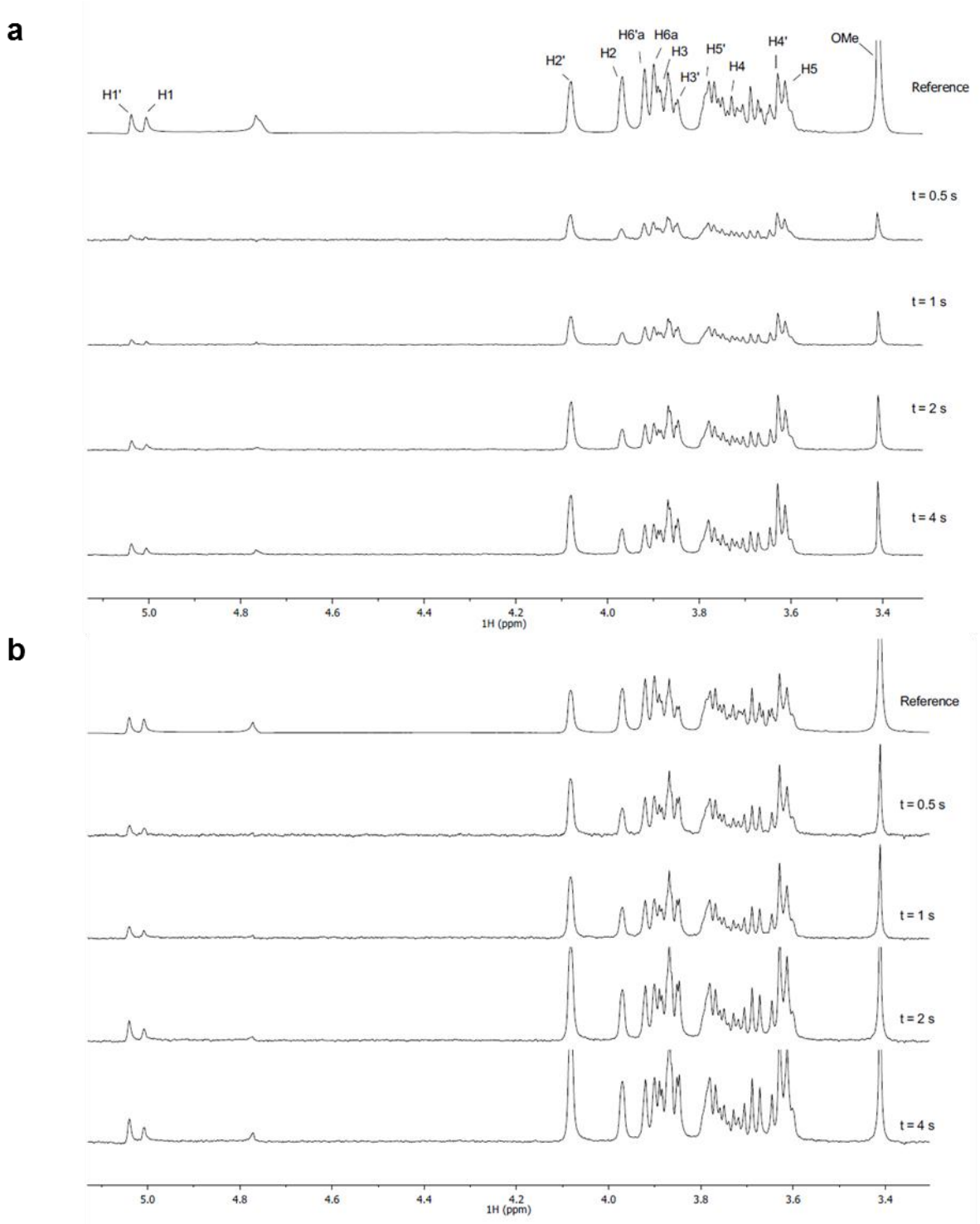


Figure B-13: ^{1}H STD NMR spectra of man2 (1 mM) interacting with (**a**) human and (**b**) murine Langerin ECD (each 50 µM). Saturation transfer times are given next to the spectrum.

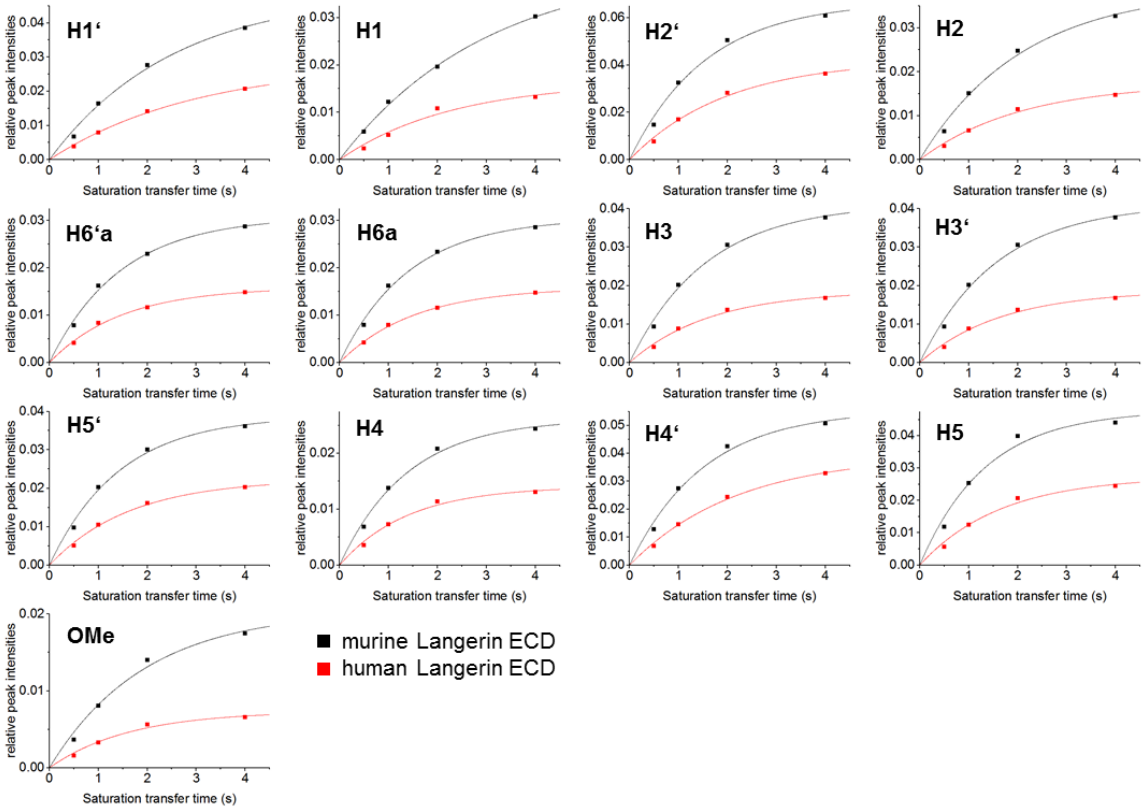


Figure B-14: STD built-up curves of man2 non-exchanging protons interacting with murine and human Langerin. Data was fitted to eq. (2-46) to obtain saturation transfer rates.

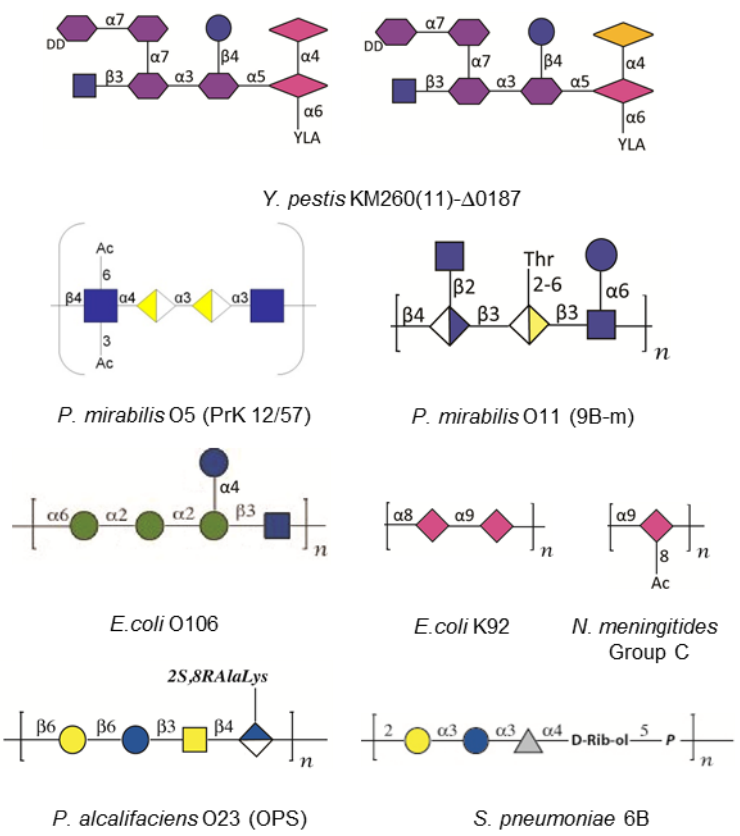


Figure B-15: Structures of selected bacterial polysaccharide hits for validation of glycan array results.

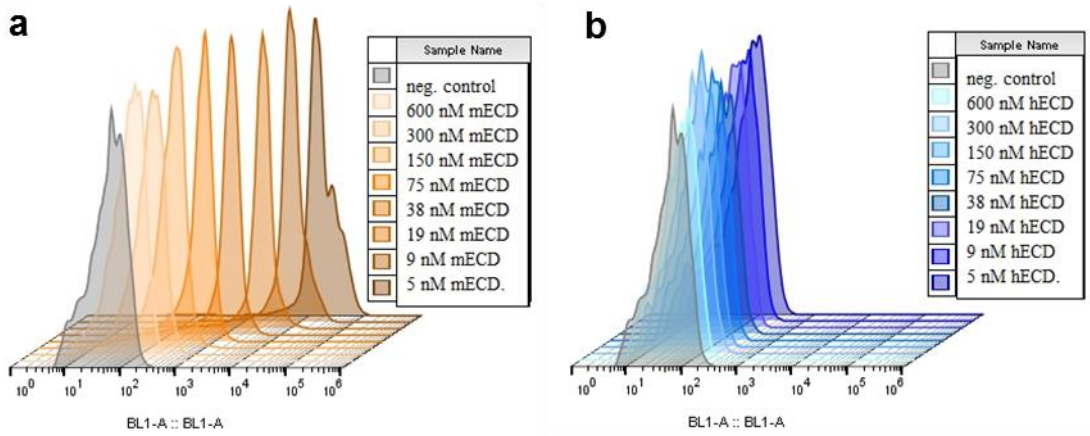


Figure B-16: Flow cytometry assay showing the interactions of murine (a) and human (b) Langerin with heat-killed E. coli O106 bacteria. Experiment was designed, performed, and analyzed by Jessica Schulze.

B.16 Multiple Sequence Alignment 28 Mammalian Langerin Homologs

```
SP|Q9UJ71|CLC4K_HUMAN          ------MTVEKEAPDAHFTVDKQNISLWPRE-PPPKSGPSLVPGKTPTVRAALICLTLVL 53
SP|Q8VBX4|CLC4K_MOUSE          --MPEAEMKE-EAPEAHFTVDKQNISLWPREP-PPKQDLSPVLRKPLCICVAFTCLALVL 56
TR|D3ZBX0|D3ZBX0_RAT           --MPEVEMKETEVPDAHFTVDKQNISLWPREP-PPKQDLTPVLRKPHCICAAFICLALVL 57
TR|B3FVQ1|B3FVQ1_PIG           -----MKAVESEVHDAHFTVDKQNISLWPREP-PPKTGPYLVLGRCLTVRAAVVLLTLVL 54
TR|E1BB51|E1BB51_BOVIN         -----MKTADSEVLDAHFTVDKQNVSLWPPDP-PPKTCPSQVLGRLLTVRAVVVFLMLVT 54
TR|F7EFC9|F7EFC9_CALJA         ------MTAEREVPDAHFTVDKQNISLWPRAEPPPKSGSSLVPGKTSTVCAALICLTLVL 54
TR|H0WH23|H0WH23_OTOGA         ---LRMKGTEGEIPDAHFTVDKQNISLWPQAEPPPKSGLCVVLRKAPAIGAVLVCLTLVL 57
TR|G3WUZ1|G3WUZ1_SARHA         VAQARMKENVMDAQVAHFTVDNQDISLWSRDP-PPKKESPIDPRTLPSVYVFIILLALGL 59
TR|H0V494|H0V494_CAVPO         -----MKPEEKEAPDAHFFVDKQNIALWPRDP-PSKQDPPLGPRKSFTIRVVLICLSLIM 54
TR|I3N0C4|I3N0C4_ICTTR         -----MKAAEMEVPDAHFTVDKQNISLWPREP-PPKQDPSLVLRKPLILRVALICLSLVL 54
TR|F6TZE3|F6TZE3_HORSE         -------AAESEVPDAHFTVDKQNISLWPRAPPPPKMGPSLVLRKLLTVRAAVIFLLLVL 53
TR|F6U7K2|F6U7K2_MONDO         VAEARMKENEVDVQVAHFTVDNQDISLWPRDS-PPKKESSISLRTLTSVHVVIILMALVL 59
TR|H2R8S3|H2R8S3_PANTR         ------MTVEREAPDAHFTVDKQNISLWPRE-PPPKSGPSLVPGKTPTVRAALICLTLVL 53
TR|G1MBS3|G1MBS3_AILME         -----MKAAESEVPDVHFTVDKQNVSLWPREP-PPKTGPCLVLRKPLTVQAAVIILTLGL 54
TR|M3WIV5|M3WIV5_FELCA         -----MKAAESEVSDVHFSVDKQNISLWPREP-PPKTGPSLVLRKSLTVRAAVIILTLGL 54
TR|G1RGW2|G1RGW2_NOMLE         ------MTVEREAPDAHFTVDKQNISLWPRE-PPPKSGPSLLPGKTPTVRAALICLTLVL 53
TR|M3Y7D9|M3Y7D9_MUSPF         -----MKAAESEVPDAHFTVDKQNISLWPRDP-PPKTDPCLVLRKPLMVRAAVITLILGF 54
TR|G1PJQ2|G1PJQ2_MYOLU         --------------------------LLPTEP-PPKMGSSPVLRKPCTVRAAFVILTLVL 33
TR|W5PNL3|W5PNL3_SHEEP         -----MKAADSEVPDAHFRVDKQNVSLWPPDP-PPKTCPSQVLGRLLTVHAVVVFLTLVT 54
TR|G3QPX8|G3QPX8_GORGO         ------MTVEREAPDAHFTVDKQNISLWPRE-PPPKSGPSLVPGKTPTVRAALICLTLVL 53
TR|F7HNH2|F7HNH2_MACMU         ------MTAEREVPDAHFTVDKQNISLWPRATPPPKSGPSLFPGKTPTVRAALICLTLVL 54
TR|G3TJF3|G3TJF3_LOXAF         -----RMTAGEELPCAHFEVDKQNISLWPREP-PSKMDRPLALRKPSTLCAVLTFLTLVL 54
TR|A0A0P6J2W2|A0A0P6J2W2_HETGA -----MKAAEKEAHDAHFTVDKQNIALWPREP-PPKQALSLHPRKPLTIRVVLICLSLVV 54
TR|G1TDE9|G1TDE9_RABIT         -----MQAADREALDAHFTVDKQNISLWPREP-SPKPAPSPLLREPVSVRAALICLSLVL 54
TR|A0A0D9RRP0|A0A0D9RRP0_CHLSB ------MTAEREVPDAHFTVDKQNISLWPRAEPPPKSGPSLFSGKTPTVRAALICLTLVL 54
TR|F7A5P2|F7A5P2_MACMU         ------MTAEREVPDAHFTVDKQNISLWPRE-PPPKSGPSLFPGKTPTVRAALICLTLVL 53
TR|F6Y539|F6Y539_CANLF         -----MKAAESEGPDAHFAVDKQSISLRPREP-PPDTGPCPVPRKPLVVQAAAVVLLLGL 54
TR|A0A096N0R1|A0A096N0R1_PAPAN ------------------------------------------------------------

SP|Q9UJ71|CLC4K_HUMAN          VASVLLQAVLYPRFMGTISD--VKTNVQLLKGRVDNIST-LDSEIKKNSDGMEAAGVQIQ 110
SP|Q8VBX4|CLC4K_MOUSE          VTSIVLQAVFYPRLMGKILD--VKSDAQMLKGRVDNIST-LGSDLKTERGRVDDAEVQMQ 113
TR|D3ZBX0|D3ZBX0_RAT           VTSIVLQAVFYPRLMGKILD--VKSDAQMLRGRVDNIST-LGSDLKRERGRVDDAEVQMR 114
TR|B3FVQ1|B3FVQ1_PIG           VASILLQAILYPWFMGTMSD--VKTSAQLLQGRVDNIST-LSSEIKRNSGGVEAAGVQVR 111
TR|E1BB51|E1BB51_BOVIN         ITSVLLQAILYPWFMGTISD--VKTNAQLLKGRVDNISS-LSSEIKRNRGGLVAVGIQVR 111
TR|F7EFC9|F7EFC9_CALJA         VASVLLQAVLYPRFMGTISD--VKTNVQLLKGRVDNIST-LDSEIKKNSDGMEVAGTQIQ 111
TR|H0WH23|H0WH23_OTOGA         VASLLLQGILYPWFMGTISD--VKTNAQLLKGRVDNISA-LGSEIKKNSGSVEAAGLQFH 114
TR|G3WUZ1|G3WUZ1_SARHA         VASLLATFILYPQLLGRIDD--VKTSVHMLKGHAENASVFLSSEIKTLRNKLKDSRTQIQ 117
TR|H0V494|H0V494_CAVPO         VTSIVLQAILYPWFMGTISD--VKANAQWLKGRVDNISI-LGSEMKKDSGGMKAVSVQIE 111
TR|I3N0C4|I3N0C4_ICTTR         VASVVLQAILYPWFMGTISD--IKTDAQLLKGRVNNVSA-LSSKVEKNSRDVEAAGVQIR 111
TR|F6TZE3|F6TZE3_HORSE         VASVTLQAILYSWFMGTISD--VKSNAQLLKGRVDNIST-LSSEIKRNRGGVAATGVQVQ 110
TR|F6U7K2|F6U7K2_MONDO         AAFLLAMAILYPQLLGKIDD--VKTSVHMLKGRAENVSVFLRSEIRKLRNKLEDSHTELQ 117
TR|H2R8S3|H2R8S3_PANTR         VASVLLQAVLYPRFMGTISD--VKTNVQLLKGRVDNIST-LDSEIKKNSDGMEAAGVQIQ 110
TR|G1MBS3|G1MBS3_AILME         TASVLLQAIFYPWFMRTISD--VKTNAQLLKGHVDNIST-LGSEVKRNRGGVEATGIQVR 111
TR|M3WIV5|M3WIV5_FELCA         AASILLQAVLYPWFMRTISN--VKNNAQLLKGRVDNISA-LGSEIKRNRGGVEATGIQVQ 111
TR|G1RGW2|G1RGW2_NOMLE         VASILLQAVLYPWFMGTISD--VKTNVQLLKGRVDNIST-LDSEIKKNSDGMEAAGVQIQ 110
TR|M3Y7D9|M3Y7D9_MUSPF         VASILLQATLYPWFMRTISD--VKTNAQLLKGSVDNIST-LGSEIKRNRGSVEATSLQVR 111
TR|G1PJQ2|G1PJQ2_MYOLU         LASVVLQAILYPWFMGTLSD--VKTNAQLLNGRVDNIST-LSSEVKRNRGGVKAAGIQVQ 90
TR|W5PNL3|W5PNL3_SHEEP         VTSVLLQAILYPWFMGTISD--VKTNAQLLKGRVDNISS-LSSEVKRNRGALVAVGFQVR 111
TR|G3QPX8|G3QPX8_GORGO         VASVLLQAILYPRFMGTISD--VKTNVQLLKGRVDNIST-LDSEIKKNSDGMEAAGVQIQ 110
TR|F7HNH2|F7HNH2_MACMU         VASILLQAVLYPRFMGNISD--VKTNVQLLKGRVDNIST-LDSEIKKNSDGMEAAGVQIQ 111
TR|G3TJF3|G3TJF3_LOXAF         VTSILLQSILYSWLLGTISD--VRINAQLLKGRVDNIST-EGSEIKRNQGGLKTASVQIQ 111
TR|A0A0P6J2W2|A0A0P6J2W2_HETGA VASIVLQAILYPWFMGTISD--VNTNAQMLKSRVDNMST-LGSEIEKNSGSVKVAGVQIQ 111
TR|G1TDE9|G1TDE9_RABIT         AASVVLQAVLYPWFMGRISD--VKTDAQLLKDRVGNISS-LGSEVRKNGVGVQAAGAQIH 111
TR|A0A0D9RRP0|A0A0D9RRP0_CHLSB VASVLLQAVLYPRFMGSISD--VKTNVQLLKGRVDNIST-LDSEIKKNSDGMEAAGVQIQ 111
TR|F7A5P2|F7A5P2_MACMU         VASILLQAVLYPRFMGNISD--VKTNVQLLKGRVDNIST-LDSEIKKNSDGMEAAGVQIQ 110
TR|F6Y539|F6Y539_CANLF         SASVLLQAALCESLGFTVPAGAFSTASQLPQ-HPDPLLP----AVQV----SPTSGFSVG 105
TR|A0A096N0R1|A0A096N0R1_PAPAN --------------MGNISD--VKTNVQLLKGRVDNIST-LDSEIKKNSNGMEAAGVQIQ 43
                                             :    .    : .            :.          ..

SP|Q9UJ71|CLC4K_HUMAN          MVNESLGYVRSQFLKLKTSVEKANAQIQILTRSWEEVSTLNA--QIPELKSDLEKASALN 168
SP|Q8VBX4|CLC4K_MOUSE          IVNTTLKRVRSQILSLETSMKIANDQLQILTMSWGEVDSLSA--KIPELKRDLDKASALN 171
TR|D3ZBX0|D3ZBX0_RAT           IVNTSLGRVRSKILSLEASMKIVSNQLQVLTMNWGEVDNLNA--KIPELQKDLDKASALN 172
TR|B3FVQ1|B3FVQ1_PIG           MVNASLDHVRSQIQRLEISVKEADARLQKLTSSWEEVDKLNA--QIPELKRDLDKASALN 169
TR|E1BB51|E1BB51_BOVIN         MVNASLDRISSQIRRLETGLKEASAQLQVLTSSWKAVDELNA--QIPELKQDLDKASALN 169
TR|F7EFC9|F7EFC9_CALJA         LVNVSLGYVRSQFLKLKTSVEKANAQIQILTRSWEDVGTLNT--QIPELKRDLEKASALN 169
TR|H0WH23|H0WH23_OTOGA         LVNASLSHVRSQILKLQTSVEKANAQIQTLTRNWEEVDTLNA--QIPELKRDLDKASALN 172
TR|G3WUZ1|G3WUZ1_SARHA         MLNFSMQYSKVHMQMLNAMLQENSNHLQELMSNWEELEHLNA--QIPMLKQDMKKTGELN 175
TR|H0V494|H0V494_CAVPO         TVNASLDSVHSQILILKTNVEKASAQIDNLRKKWEEADHLNT--QIPELKRDLDKASALN 169
TR|I3N0C4|I3N0C4_ICTTR         MVDASLDSVNAQILTLTASIEKATAEIQSLTRNWEEVGHLNA--QIPELKRDLDKASALN 169
TR|F6TZE3|F6TZE3_HORSE         MVNASLDRVRSQMRRLETDVKEANARIQTLTKSWEEVDNLNA--QIPELKRDLDKASTLN 168
TR|F6U7K2|F6U7K2_MONDO         MLNGTLRVSNAQIQLLNSKLKNANNQLQTLMRGWQELDDLNS--QIPMLKQDLAKTGELK 175
TR|H2R8S3|H2R8S3_PANTR         MVNVSLGYVRSQFLKLKTSVEKANAQIQILTRSWEEVSTLNA--QIPELKSDLEKASALN 168
TR|G1MBS3|G1MBS3_AILME         MVNASLDHVRSQIGKLEAGVKKADAQIQILTRRWEEVIDLNA--QIPALKRDLDKASSLN 169
TR|M3WIV5|M3WIV5_FELCA         KVNASLDHVRSQIWKLETGVKEANAQIQMLTRSWEAVNDLNA--QIPELKRNLDKASSLN 169
TR|G1RGW2|G1RGW2_NOMLE         MVNVSLGYVLSQFLELKTSVEKANAQIQILTRSWEEVSTLNA--QIPELKSDLEKASALN 168
TR|M3Y7D9|M3Y7D9_MUSPF         MVNASLAHVHSQIGKLEAGVRDAAAQIQMLTRRWEEVNDLNA--QIPALKRDLGKASSLN 169
TR|G1PJQ2|G1PJQ2_MYOLU         MVNASLDHVHSQIRRLETGMKAVNAQIQMLTRSWEEVYSLNA--QLPELKRDLDKASALN 148
TR|W5PNL3|W5PNL3_SHEEP         MMNASLGRVSSQIRKLETGLKEASAQLHVLTSSWEAVDELNA--QIPGLKQDLDKASALN 169
TR|G3QPX8|G3QPX8_GORGO         MVNVSLGYVRSQFLKLKTSVEKANAQIQILTRSWEEVSTLNA--QIPELKSDLEKASALN 168
TR|F7HNH2|F7HNH2_MACMU         TVNVSLDYVRSQFLKLKTSVEKANAQIQILTRSWEEVGTLNA--QIPELKSDLEKASALN 169
```

```
TR|G3TJF3|G3TJF3_LOXAF          RVNVSLNHVHSQLLTLETGLEKANAQIQMLTKSWEEVDALNA--QIPELKRDLDKASVLN 169
TR|A0A0P6J2W2|A0A0P6J2W2_HETGA MVNASLDNVHSQVLILKTSVEKANAQIQSLRKTWEEVDHLNA--KIPELKRDLDKASALN 169
TR|G1TDE9|G1TDE9_RABIT          TVNASLDRVYSQILTLKTSVERASAQIQILTESWEEVHSLNA--RIPELKSDLDKASTLH 169
TR|A0A0D9RRP0|A0A0D9RRP0_CHLSB TVNVSLGYVRSQFLKLKTSVEKANAQIQILTRSWEEVGTLNA--QIPELKSDLEKASALN 169
TR|F7A5P2|F7A5P2_MACMU          TVNVSLDYVRSQFLKLKTSVEKANAQIQILTRSWEEVGTLNA--QIPELKSDLEKASALN 168
TR|F6Y539|F6Y539_CANLF          L---------------MSPLSRCSHRVPCLQKQREGTSVLSLQGFGMYLVSDLEPLSRVS 150
TR|A0A096N0R1|A0A096N0R1_PAPAN MVNVSLGYVRSQFLKLKTSVEKANAQIQILTRSWEEVGTLNA--QIPELKSDLEKASALN 101
                                          :     .:  *         *.       *  ::   . :

SP|Q9UJ71|CLC4K_HUMAN           TKIRALQGSLENMSKLLKRQNDILQVVSQGWKYFKGNFYYFSLIPKTWYSAEQFCVSRNS 228
SP|Q8VBX4|CLC4K_MOUSE           TKVQGLQNSLENVNKLLKQQSDILEMVARGWKYFSGNFYYFSRTPKTWYSAEQFCISRKA 231
TR|D3ZBX0|D3ZBX0_RAT            TKVQGLQNSLENINKLLKEQSDILEMMSRGWKYFMGNFYYFSRTPKTWYSAEQYCISRKA 232
TR|B3FVQ1|B3FVQ1_PIG            AKVWDLQSGLGNISKMLKQQNDILQMMSQGWKYFRGNVYYFSVMKKTWYSAQQFCVSRES 229
TR|E1BB51|E1BB51_BOVIN          AKVRELQSGLESISKLLQQQKDILQVVSQGWKYFGGHFYYFSKISKTWYSAQQICISRDS 229
TR|F7EFC9|F7EFC9_CALJA          TKIRELQSSLENMSKLLKRQNDILQVVSQGWNFFRGNFYYFSRIPKTWYSAQQFCISRNS 229
TR|H0WH23|H0WH23_OTOGA          AKIRNLQSSLENTSKLLKQQSDILQMVSRGWRYFNGNFYYFSPVLKTWYSAEQFCVSKNS 232
TR|G3WUZ1|G3WUZ1_SARHA          AKIEQLKKDLQNLGASLSQQRYILEMASQNWKFFEGNFYYFSTMKKSWYSAEQFCTTQDS 235
TR|H0V494|H0V494_CAVPO          GMVRTLQSRLDNINKSLRQQRDILQMVSQGWEFFKDNFYYFSRISKTWYSAEQFCVSRDS 229
TR|I3N0C4|I3N0C4_ICTTR          AKVRGLQSSLETISKSLKQQNDILQMVSQDWKYFKGNFYYFSHTPKTWYSAQQFCVSRNS 229
TR|F6TZE3|F6TZE3_HORSE          AKVRALQSSLENVSKLLKQQNDILQVVSQGWKYFKGNFYYFSRVPKTWYSAQQFCISRNS 228
TR|F6U7K2|F6U7K2_MONDO          TKVEQLRKDLQTLGTSLSRQRYILEMASQDWKYFQGNFYYFSTKRKSWYSAEQFCTTRDS 235
TR|H2R8S3|H2R8S3_PANTR          TKIRALQGSLENMSKLLKRQNDILQVVSQGWKYFKGNFYYFSLISKTWYSAEQFCVSRNS 228
TR|G1MBS3|G1MBS3_AILME          AKVRGLQSNLENISKLFKRQNDILQMVSQGWKYFKENFYYFSQIPKTWYSAQQFCMSRDS 229
TR|M3WIV5|M3WIV5_FELCA          AKVQGLQTSLENISKVLKQQNDILQMVSQGWKYFRGNFYYFSQVPKTWYSAQQFCVSRDS 229
TR|G1RGW2|G1RGW2_NOMLE          TKIRALQGSLENMSKLLKRQNDILQVVSQGWKYFKGNFYYFSLIPKTWYSAEQFCVSRNS 228
TR|M3Y7D9|M3Y7D9_MUSPF          AKIQGLQSHLEDISKLLRRQNDILQMVSQGWKYFKENFYYFSQVPKTWYSAQQFCMSRDS 229
TR|G1PJQ2|G1PJQ2_MYOLU          VKVRGLQSSLENISKLLKQQNDTLQMVSQGWKYFKENFYYFSHVQKTWYSAQQFCVSRNS 208
TR|W5PNL3|W5PNL3_SHEEP          AKVRELQSGLESISKLLQQQNDILQVVSQGWKYFRGHFYYFSQISKTWYSAQQMCISRDS 229
TR|G3QPX8|G3QPX8_GORGO          TKIRALQGSLENMSKLLKRQNDILQVVSQGWKYFKGNFYYFSLIPKTWYSAEQFCVSRNS 228
TR|F7HNH2|F7HNH2_MACMU          TKIRALQGSLDNMSKLLKRQNDILQVVSQGWKYFKGNFYYFSLITKTWYSAQQFCVSRNS 229
TR|G3TJF3|G3TJF3_LOXAF          AKVRGLQSSLENMGKSLRQQNDILQMVSQGWKYFKGNFYYFSRVQKTWYSAQQFCKSRNS 229
TR|A0A0P6J2W2|A0A0P6J2W2_HETGA AMVRALQSRLDSINKLLRRQSDILQMVSQGWKFFRENFYYFSHISKTWYSAEQFCMSKDS 229
TR|G1TDE9|G1TDE9_RABIT          AKVRGLQSSLENISEVLRQQNDILQMVYQGWKYFKGNFYYFSRIAKTWYSAQQFCMSRDS 229
TR|A0A0D9RRP0|A0A0D9RRP0_CHLSB TKIRALQGSLDNMSKLLKRQNDILQVVSQGWKYFKGNFYYFSLVTKTWYSAQQFCVSRNS 229
TR|F7A5P2|F7A5P2_MACMU          TKIRALQGSLDNMSKLLKRQNDILQVVSQGWKYFKGNFYYFSLITKTWYSAQQFCVSRNS 228
TR|F6Y539|F6Y539_CANLF          ------SWVVVGLSYLAPKPDDILQMVSQGWKYFRGNFYYFSQVSKTWYSAQQFCMSRDS 204
TR|A0A096N0R1|A0A096N0R1_PAPAN TKIRALQGSLDNMSKLLKRQNDILQAVSQGWKYFKGNFYYFSLVTKTWYSAQQFCVSRNS 161
                                         :        .    *:   : *.:*  ..****   *:****:* * ::.:

SP|Q9UJ71|CLC4K_HUMAN           HLTSVTSESEQEFLYKTAGGLIYWIGLTKAGMEGDWSWVDDTPFNKV--QSVRFWIPGEP 286
SP|Q8VBX4|CLC4K_MOUSE           HLTSVSSESEQKFLYKAADGIPHWIGLTKAGSEGDWYWVDQTSFNKE--QSRRFWIPGEP 289
TR|D3ZBX0|D3ZBX0_RAT            HLTSVSSESEHEFLYKVADGIPHWIGLTKAGSEGDWYWVDQTSFNKE--QSRRFWIPGEP 290
TR|B3FVQ1|B3FVQ1_PIG            HLTSVTSESEQEFLYKMTDGLIYWIGLTKAGSEWDWHWVDDTPFDKL--QSARFWIPGEP 287
TR|E1BB51|E1BB51_BOVIN          HLTSVTSEREQEFLYRTAGGLPYWIGLTKAGSEGDWHWVDGTPYNKV--QSEKFWIPGEP 287
TR|F7EFC9|F7EFC9_CALJA          HLTSVTSESEQEFLYKTAGGRIYWIGLTKAGMEGEWSWVDDTPFNKV--QSERFWIPGEP 287
TR|H0WH23|H0WH23_OTOGA          HLTSVTSDSEQAFLSKSAAEVIRWIGLTKAGTEGGWHWIDDTPFNKA--QTERYWIPGEP 290
TR|G3WUZ1|G3WUZ1_SARHA          HLTSVTSTKEQEFLYKVANGVPYWIGLTKIGSSGTWHWADGTIFIER--ENARFWIKGEP 293
TR|H0V494|H0V494_CAVPO          HLTSVTSQNEQEFLYKTAGGLPHWIGLTKAGSEGDWYWVDDTPFNKD--SVSRFWLPGEP 287
TR|I3N0C4|I3N0C4_ICTTR          HLTSVTSKSEQEFLYKTASDFPFWIGLTKTGSEGNWYWMDNTPFNDPSLRFGRFWIPGEP 289
TR|F6TZE3|F6TZE3_HORSE          HLTSVTSESEQEFLYKAAGGLLHWIGLTKAGSEGDWHWVDDTPFDKV--QSARFWISGEP 286
TR|F6U7K2|F6U7K2_MONDO          HLTSVTSAKEQEFLYKTAAGISYWIGLTKTGSSGTWHWIDGTLYIER--ENARFWAKGEP 293
TR|H2R8S3|H2R8S3_PANTR          HLTSVTSESEQEFLYKTAGGLIYWIGLTKAGMEGDWSWVDDTPFNKV--QSARFWIPGEP 286
TR|G1MBS3|G1MBS3_AILME          HLTSVASENEQEFLYKMAGGLFYWIGLTKAGSEGDWYWVDNTPFDKD--LSARFWIPGEP 287
TR|M3WIV5|M3WIV5_FELCA          HLTSVASESEQEFLYKMAGGLFYWIGLTKAGSEGDWYWVDDTPFNKV--QSARFWIPGEP 287
TR|G1RGW2|G1RGW2_NOMLE          HLTSVTSESEQEFLYKTAGGFIYWIGLTKAGMEGDWSWVDDTPFNKV--QSARFWIPGEP 286
TR|M3Y7D9|M3Y7D9_MUSPF          QLTSVASESEQEFLYKMAGGLFYWIGLTKAGTEGDWYWVDNTPFDKV--QSARFWIPGEP 287
TR|G1PJQ2|G1PJQ2_MYOLU          QLTSVTSDSEQEFLYKTAGGLFYWIGLTKAGTDGDWYWVDETPFNKV--QSDRFWIPGEP 266
TR|W5PNL3|W5PNL3_SHEEP          HLTSVTSEREQEFLYRTAGGLPYWIGLTKAGSEGDWHWVDGTPFNKV--QSEKFWIPGEP 287
TR|G3QPX8|G3QPX8_GORGO          HLTSVTSESEQEFLYKTAGGLIYWIGLTKAGMEGDWSWVDDTPFNKV--QSARFWIPGEP 286
TR|F7HNH2|F7HNH2_MACMU          HLTSVTSESEQEFLYKTAGGLTYWIGLTKAGTEGDWFWVDDTPFDKV--QSAKFWIPGEP 287
TR|G3TJF3|G3TJF3_LOXAF          QLTSVTSESEQEFLYKTAGGISYWIGLTKAGSEGDWSWVDDTPFNKV--QSAKFWIPGEP 287
TR|A0A0P6J2W2|A0A0P6J2W2_HETGA HLTSVTSRDEQEFLYKGAGGLPNWIGLTKAGSEGKWYWVDDTPFNKE--SV-RFWLPGEP 286
TR|G1TDE9|G1TDE9_RABIT          HLTSVTSESEQEFLYKTAGGLVYWIGLSKAGSEGHWYWADGTSFNEA--QSVRFWIPGEP 287
TR|A0A0D9RRP0|A0A0D9RRP0_CHLSB HLTSVTSESEQEFLYKTAGGLTYWIGLTKAGTEGDWFWVDDTPFNKV--QSAKFWIPGEP 287
TR|F7A5P2|F7A5P2_MACMU          HLTSVTSESEQEFLYKTAGGLTYWIGLTKAGTEGDWFWVDDTPFDKV--QSAKFWIPGEP 286
TR|F6Y539|F6Y539_CANLF          QLTSVTSESEQEFLYRTAGGLSYWIGLTKAGSEGDWSWVDDTPFDKVQ-SAHRFWIPGEP 263
TR|A0A096N0R1|A0A096N0R1_PAPAN HLTSVTSESEQEFLYKTAGGLTYWIGLTKAGTEGDWFWVDDTPFNKV--QSAKFWIPGEP 219
                                :****:*  *: ** : :     ****:* * .  * * * * : .      ::*  ***
```

```
SP|Q9UJ71|CLC4K_HUMAN          NNAGNNEHCGNIKAPSLQAWNDAPCDKTFLFICKRPYVPSEP- 328
SP|Q8VBX4|CLC4K_MOUSE          NNAGNNEHCANIRVSALKCWNDGPCDNTFLFICKRPYVQTTE- 331
TR|D3ZBX0|D3ZBX0_RAT           NNVRNNEHCANIRVSALKCWNDSPCDNVYSFICKMPYIRMIT- 332
TR|B3FVQ1|B3FVQ1_PIG           NNSENNEHCANIKRSSLRSWNDAPCDIELLFICKRPYVPSEP- 329
TR|E1BB51|E1BB51_BOVIN         NNVGNNEHCVTLKTSLLRSWNDASCDNTFLFICKRSYKPSEP- 329
TR|F7EFC9|F7EFC9_CALJA         NNYGNNEHCANLKASSLQSWNDAPCDQTFLFICKRPYIPSEP- 329
TR|H0WH23|H0WH23_OTOGA         NNSGFNEHCVSIRVLSLQSWNDSPCDIKYSFICKRPYIPSEP- 332
TR|G3WUZ1|G3WUZ1_SARHA         NDSGQNEHCVTLEQTSLMSWNDVTCDRGLQFICKKGPKSPVID 336
TR|H0V494|H0V494_CAVPO         NNVGNSEHCANIKVSSLRSWNDDSCDIKLFFICKRPYIQSEP- 329
TR|I3N0C4|I3N0C4_ICTTR         NNAGNNEHCADMRFPLLMSWNDASCDKTLPFVCKQPYVPSEP- 331
TR|F6TZE3|F6TZE3_HORSE         NNYGSNEHCANIKLFSLQSWNDASCDITLLFICKRPYTPSEA- 328
TR|F6U7K2|F6U7K2_MONDO         NNGYGNEHCVNLEKSSLMSWNDVNCDGELQFICKKNPKPLEMD 336
TR|H2R8S3|H2R8S3_PANTR         NNAGNNEHCGNIKAPSLQAWNDAPCDITFLFICKRPYVPSEP- 328
TR|G1MBS3|G1MBS3_AILME         NNFGSNEHCANIKASSLQSWNDASCDNKLFFICKRPYIPSEP- 329
TR|M3WIV5|M3WIV5_FELCA         NNFGNNEHCANIKMSSLQSWNDASCDNKLLFICKRPYIPSEP- 329
TR|G1RGW2|G1RGW2_NOMLE         NNAGNNEHCGNIKALSLQAWNDAPCDKTFLFICKRPYVPSEP- 328
TR|M3Y7D9|M3Y7D9_MUSPF         NNYGNNEHCANIKMSSLQSWNDASCDNKLLFICKRPYITSEP- 329
TR|G1PJQ2|G1PJQ2_MYOLU         NNLGNNEHCANLKMSSLQSWNDAPCDHPFLFICKRRYIPSEP- 308
TR|W5PNL3|W5PNL3_SHEEP         NNYGNNEHCVNLKTSSLRSWNDASCDNTFPFICKRSYKPSEP- 329
TR|G3QPX8|G3QPX8_GORGO         NNAGNNEHCGNIKAPSLQAWNDAPCDKTFLFICKRPYVPSEP- 328
TR|F7HNH2|F7HNH2_MACMU         NNAGNNEHCGNIRVSSLQAWNDAQCDKTFLFICKRPYIPSEP- 329
TR|G3TJF3|G3TJF3_LOXAF         NNSGYSEHCAHIRMASLQSWNDASCDNTLPFICKQLYIPSEP- 329
TR|A0A0P6J2W2|A0A0P6J2W2_HETGA NNVGNSEHCASIKVSSLQSWNDDSCDNKLLFICKQPYIQSEP- 328
TR|G1TDE9|G1TDE9_RABIT         NNMGNGEHCVSIKVSLLQSWNDASCDDKLLFVCKRPYTAAGP- 329
TR|A0A0D9RRP0|A0A0D9RRP0_CHLSB NNAGNNEHCGNIRASSLQAWNDAQCDKTFLFICKRPYVPSEP- 329
TR|F7A5P2|F7A5P2_MACMU         NNAGNNEHCGNIRVSSLQAWNDAQCDKTFLFICKRPYIPSEP- 328
TR|F6Y539|F6Y539_CANLF         N----NEHCADIKVSSLQSWNDVSCDSTLFFICKRPYSPSEP- 301
TR|A0A096N0R1|A0A096N0R1_PAPAN NNVGNSEHCGNIRASSLQAWNDAQCDTTFLFICKRPYVPSEP- 261
                               *     ***  :.   * .***  **    *:**
```

Figure B-17: Multiple sequence alignment of the primary structures of 28 mammalian Langerin sequences (associated in uniprotKB with CLC4K). Alignment was performed using CLUSTAL O(1.2.1) (Sievers et al., 2011).

Appendix C Supplementary References

Anandakrishnan, R., Aguilar, B., and Onufriev, A.V. (2012). H++3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res *40*, W537-W541.

Feinberg, H., Mitchell, D.A., Drickamer, K., and Weis, W.I. (2001). Structural Basis for Selective Recognition of Oligosaccharides by DC-SIGN and DC-SIGNR. Science *294*, 2163-2166.

Guo, Y., Feinberg, H., Conroy, E., Mitchell, D.A., Alvarez, R., Blixt, O., Taylor, M.E., Weis, W.I., and Drickamer, K. (2004). Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. Nat Struct Mol Biol *11*, 591-598.

Kilambi, Krishna P., and Gray, Jeffrey J. (2012). Rapid Calculation of Protein pK(a) Values Using Rosetta. Biophys J *103*, 587-595.

Li, H., Robertson, A.D., and Jensen, J.H. (2005). Very fast empirical prediction and rationalization of protein pK(a) values. Proteins-Structure Function and Bioinformatics *61*, 704-721.

Olsson, M.H.M. (2011). Protein electrostatics and pK(a) blind predictions; contribution from empirical predictions of internal ionizable residues. Proteins *79*, 3333-3345.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., *et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol *7*, 539.