Dimerization of the fungal defense lectin CCL2 is essential for its toxicity against nematodes

Author(s):
Bleuler-Martinez, Silvia; Stutz, Katrin; Sieber, Ramon; Collot, Mayeul; Mallet, Jean-Maurice; Hengartner, Michael; Schubert, Mario; Varrot, Annabelle; Künzler, Markus

Publication date:
2017-05

Permanent link:
https://doi.org/10.3929/ethz-b-000123693

Originally published in:
Glycobiology 27(5), https://doi.org/10.1093/glycob/cww113

Funding acknowledgement:
130671 - Molekulare Charakterisierung der protein-vermittelten Abwehr von Pilzen gegen Frassfeinde und Parasiten und deren Anwendung zur Bekämpfung von tierparasitischen Nematoden (SNF)

OXFORD UNIVERSITY PRESS | Glycobiology

Dimerization of the fungal defense lectin CCL2 is essential for its toxicity against nematodes

Journal:	*Glycobiology*
Manuscript ID	GLYCO-2016-00100.R1
Manuscript Type:	Regular Manuscripts
Date Submitted by the Author:	n/a
Complete List of Authors:	Bleuler-Martinez, Silvia; ETH Zürich, Institute of Microbiology Stutz, Katrin; University of Zürich, Institute of Molecular Life Sciences Sieber, Ramon; ETH Zürich, Institute of Microbiology Collot, Mayeul; ENS, UMR CNRS 7203, Chimie Mallet, Jean-Maurice; ENS, UMR CNRS 7203, Chimie Hengartner, Michael; University of Zürich, Institute of Molecular Life Sciences Schubert, Mario; University of Salzburg, Molecular Biology Varrot, Annabelle; CERMAV-CNRS, Grenoble, Chimie Künzler, Markus; ETH Zürich, Institute of Microbiology
Key Words:	toxin, carbohydrate recognition, fucose, N-glycan core, glycan binding

Dimerization of the fungal defense lectin CCL2 is essential for its toxicity against nematodes

Silvia Bleuler-Martinez[2], Katrin Stutz[3], Ramon Sieber[2], Mayeul Collot[4,8], Jean-Maurice Mallet[4], Michael Hengartner[3], Mario Schubert[1,5,6], Annabelle Varrot[1,7] and Markus Künzler[1,2]

[2] Institute of Microbiology, ETH Zürich, Vladimir-Prelog-Weg 4, 8093 Zürich, Switzerland

[3] Institute of Molecular Life Sciences, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[4] Laboratoire des Biomolécules, UPMC Université Paris 06, Ecole Normale Supérieure, Paris, France

[5] Institute of Molecular Biology and Biophysics, ETH Zürich, Schafmattstr. 20, 8093 Zürich, Switzerland

[6] Department of Molecular Biology, University of Salzburg, Billrothstrasse 11, 5020 Salzburg, Austria

[7] CERMAV, UPR5301, CNRS and Université Grenoble Alpes, 38041 Grenoble, France

[8] *Current address:* Laboratoire de Biophotonique et Pharmacologie, Université de Strasbourg, 74 route du Rhin, CS 60024, 67401 Illkirch Cédex, France

[1] To whom correspondence should be addressed:
Tel: +41 44 6324925; e-mail: mkuenzle@ethz.ch (MK); Tel: +33 4 76037634; e-mail: varrot@cermav.cnrs.fr (AV); Tel: +43 662 80447243; e-mail: mario.schubert@sbg.ac.at (MS)

Running head: *Coprinopsis cinerea* fruiting body lectin CCL2 forms dimers

Keywords: Toxin / dimer / glycan binding / carbohydrate recognition / fucose / N-glycan core / nematodes / insects

Supplementary data:

Table S1

Figures S1-S8

Abstract

Lectins are used as defense effector proteins against predators, parasites and pathogens by animal, plant and fungal innate defense systems. These proteins bind to specific glycoepitopes on the cell surfaces and thereby interfere with the proper cellular functions of the various antagonists. The exact cellular toxicity mechanism is in many cases unclear. Lectin CCL2 of the mushroom *Coprinopsis cinerea* was previously shown to be toxic for *Caenorhabditis elegans* and *Drosophila melanogaster*. This toxicity is dependent on a single, high-affinity binding site for the trisaccharide GlcNAc(Fucα1,3)β1,4GlcNAc, which is a hallmark of nematode and insect N-glycan cores. The carbohydrate-binding site is located at an unusual position on the protein surface when compared to other β-trefoil lectins. Here, we show that CCL2 forms a compact dimer in solution and in crystals. Substitution of two amino acid residues at the dimer interface, R18A and F133A, interfered with dimerization of CCL2 and reduced toxicity but left carbohydrate-binding unaffected. These results, together with the positioning of the two carbohydrate-binding sites on the surface of the protein dimer, suggest that crosslinking of N-glycoproteins on the surface of intestinal cells of invertebrates is a crucial step in the mechanism of CCL2-mediated toxicity. Comparisons of the number and positioning of carbohydrate-binding sites among different dimerizing fungal β-trefoil lectins revealed a considerable variability in the carbohydrate-binding patterns of these proteins which are likely to correlate with their respective functions.

Introduction

A common feature of animal, plant and fungal innate defense systems is the use of lectins as effector proteins (Arthur, C.M., Cummings, R.D., et al. 2014, Gallo, R.L. and Hooper, L.V. 2012, Kunzler, M. 2015, Lannoo, N. and Van Damme, E.J. 2014). These proteins act by binding to specific glycoepitopes on the surface of the cells and tissues of the antagonist and thereby fulfill a dual defense function: (1) they distinguish between self and non-self based on the organism-specific glycomes and (2) they affect the viability of the antagonist either directly or in conjunction with additional effector proteins or protein domains. Examples of lectin-mediated toxicity in innate defense are bactericidal galectins recognizing blood group antigens in the intestinal mucosa of mammals (Mukherjee, S., Zheng, H., et al. 2014, Stowell, S.R., Arthur, C.M., et al. 2010), insecticidal vacuolar lectins produced and stored in specific plant tissues recognizing terminal mannose residues on the brush border membrane of midgut epithelial cells of insects (Caccia, S., Van Damme, E.J., et al. 2012) and nematicidal and insecticidal cytoplasmic lectins abundant in the fruiting bodies and sclerotia of dikaryotic fungi (Sabotic, J., Ohm, R.A., et al. 2015). The cellular toxicity mechanism of these lectins is often unclear since most of these proteins lack additional domains or interacting proteins carrying established toxicity functions e.g. proteolysis or pore-formation. It appears that, in these cases, mere binding to glycoconjugates on cell surfaces leads to malfunctioning and eventually death of cells.

A key requirement for the toxicity of lectins appears to be the capacity to crosslink glycoconjugates (glycoproteins, glycolipids, polysaccharides) on the cell surface due to multivalency of the lectins i.e. the presence of multiple binding sites for the same or different glycoepitopes on the lectin surface (Boscher, C., Dennis, J.W., et al. 2011, Brewer, C.F., Miceli, M.C., et al. 2002, Kunzler, M. 2015). Experimental evidence for this mechanism stems from studies with mammalian galectins, which were shown to crosslink specific

glycoconjugates on the surface of mammalian cells (Mockl, L., Horst, A.K., et al. 2015, Pace, K.E., Lee, C., et al. 1999, Sacchettini, J.C., Baum, L.G., et al. 2001). The resulting lattices between lectins and glycoconjugates on the cell surface were shown to interfere with cell function. Two mechanisms for the coupling of lattice formation and cell function are currently discussed. (1) Lectin-mediated multimerization of glycosylated cell surface receptors activates signaling pathways leading to apoptosis (Hamshou, M., Van Damme, E.J., et al. 2012, Pace, K.E., Hahn, H.P., et al. 2000, Stillman, B.N., Hsu, D.K., et al. 2006); (2) Lectin-mediated clustering of glycoconjugates on cell surfaces interferes with the spatiotemporal dynamics, i.e. the residence time and localization of the glycoconjugates on the cell surface (Cha, S.K., Ortega, B., et al. 2008, Lakshminarayan, R., Wunder, C., et al. 2014, Partridge, E.A., Le Roy, C., et al. 2004, Torreno-Pina, J.A., Castro, B.M., et al. 2014). Indirect evidence for the requirement of lattice formation for lectin-mediated toxicity comes from a report where a fungal β-trefoil lectin, CNL from the mushroom *Clitocybe nebularis*, lost its toxicity due to a loss in multivalency but not carbohydrate-binding (Pohleven, J., Renko, M., et al. 2012).

An exception to these proposed mechanisms seemed to be the recently identified β-trefoil lectin CCL2 that is highly expressed in *Coprinopsis cinerea* fruiting bodies (Plaza, D.F., Lin, C.W., et al. 2014). CCL2 has been implicated in the defense of this mushroom against invertebrate predators based on its toxicity towards nematodes and insects (Heim, C., Hertzberg, H., et al. 2015, Schubert, M., Bleuler-Martinez, S., et al. 2012). Toxicity of CCL2 towards the model nematode *Caenorhabditis elegans* is dependent on the specific recognition of the α1,3fucosylated cores of N-glycans on the nematode intestinal epithelium but does not involve endocytosis of the lectin (Stutz, K., Kaech, A., et al. 2015). The previously determined three-dimensional NMR structure of monomeric CCL2 explained how the specificity and high affinity of this protein for this unusual glycan target is achieved

(Schubert, M., Bleuler-Martinez, S., et al. 2012) but gave, as a monovalent monomer, no hints at the toxicity mechanism. Here, further characterization with notably protein X-ray crystallography and NMR revealed that CCL2 forms in fact a compact dimer and, thus, is bivalent for its ligand. Dimerization occurs in a way that the two carbohydrate binding sites are located on one face of the dimer. Mutations in the dimerization interface did not affect binding of CCL2 to the glycan target in solution but interfered with the dimeric structure and led to a reduction of toxicity. These results suggest that the cross-linking of N-glycoproteins on the surface of intestinal epithelial cells is an essential part of the toxicity mechanism of this fungal β-trefoil lectin towards invertebrates.

Results

Mutation of a conserved residue of CCL2 (R18A) significantly reduces nematotoxicity

An alignment of CCL2 with its paralog CCL1 from *C. cinerea* and homologs encoded in other fungal genomes available via the Mycocosm of the US Department of Energy Joint Genome Institute (DOE-JGI) (Fig. 1) revealed a number of highly conserved amino acid residues. Some of these residues, e.g. Y57, W78, L87, Y92 and W94, were previously shown to be involved in binding of the trisaccharide GlcNAc(Fucα1,3)β1,4GlcNAc by the single, high-affinity binding site of CCL2 (Schubert, M., Bleuler-Martinez, S., et al. 2012). The conservation of other conserved residues, e.g. R18, K26 and R136, could not be explained by a role in ligand-binding, however. We speculated that these positively charged residues may be involved in binding of an additional ligand, e.g. phospholipids, and that this interaction would contribute to the nematotoxicity of the protein. In order to test this hypothesis, we generated individual alanine mutants of these residues and assessed the toxicity of *Escherichia coli* cells expressing untagged versions of these protein variants compared to the analogous wildtype protein towards *Caenorhabditis elegans* N2. As a control, we used a

protein variant with a dysfunctional carbohydrate binding site, CCL2(Y92A). The CCL2(R18A) protein variant showed thereby a significant drop in nematotoxicity while it was expressed in *E. coli* in soluble form and at similar levels as the wildtype and the CCL2(Y92A) protein variant (Fig. 2).

R18A mutation does not affect carbohydrate binding

To elucidate the molecular basis behind the effect of the R18A mutation on nematotoxicity, we first verified that R18 is not involved in carbohydrate-binding. We performed an NMR titration experiment with recombinant, His-tagged protein in which we monitored binding of the previously used, synthetic ligand GlcNAc(Fucα1,3)β1,4GlcNAcβ1O$(CH_2)_5$COOH (Schubert, M., Bleuler-Martinez, S., et al. 2012) on the protein side at residue-specific precision using 2D NMR spectroscopy in combination with ^{15}N labeled protein (Fig. 3). In the presented 2D ^{1}H,^{15}N-HSQC spectrum the NH group of each amino acid is represented by a cross-peak whose spectral position reflects the local chemical environment of that NH group. In the case of CCL2 WT and CCL2(R18A), ligand binding leads to the appearance of new cross-peaks and the disappearance of some initial signals, as ligand concentration is increased. The new cross-peaks originate from amino acids in tight contact to the ligand whereas the initial signals from the same amino acids reporting the absence of ligand disappear. The simultaneous observation of the free and bound form (Fig. 3 on the right) indicates a slow exchange kinetics between the free and bound form of the protein in regard to the NMR time scale (slow k_{ON} and slow k_{OFF}). This is in contrast to fast exchanging interactions that are typical for weak protein-carbohydrate interactions, in which one average signal position is observed for each amino acid with contributions of the free and the bound form leading to a movement of signal positions with increasing ligand concentration. In the case of the CCL2(Y92A) variant with the dysfunctional carbohydrate binding site, neither the appearance

of new signals nor a movement of signals occurred, confirming that the oligosaccharide does not bind. The fact that almost identical chemical shift changes and slow exchange kinetics were observed for the R18A variant and the wild type protein shows that the influence of the ligand on the protein resonances at the carbohydrate-protein interface is comparable. Taken together, these results confirmed that the ligand binding in CCL2(R18A) is intact and has structural features that are comparable to the wild type protein.

X-ray crystallography reveals the formation of CCL2 dimers

Based on previously published NMR and size exclusion chromatography (SEC) data, it was concluded that CCL2 was monomeric (Schubert, M., Bleuler-Martinez, S., et al. 2012). In contrast, six other single domain β-trefoil fungal lectins have been structurally characterized so far and been shown to form dimers such as *Clitocybe nebularis* lectin CNL (Pohleven, J., Renko, M., et al. 2012), *Boletus edulis* beta-trefoil lectin BEL (Bovi, M., Cenci, L., et al. 2013), *Sclerotinia sclerotiorum* agglutinin SSA (Sulzenbacher, G., Roig-Zamboni, V., et al. 2010) and *Rhizoctonia solani* agglutinin RSA (Skamnaki, V.T., Peumans, W.J., et al. 2013). The MPL lectin from *Macrolepiota procera* was crystallized only under its monomeric form but was shown to form dimers in solution (Zurga, S., Pohleven, J., et al. 2014). Based on these structures and above results regarding the CCL2(R18A) mutant, we reexamined the possibility of CCL2 oligomerization by determining the crystal structure of the recombinant, His-tagged protein.

The crystallization trials with the Midas screen, which is a screen based on alternative polymeric precipitants (Grimm, C., Chari, A., et al. 2010), resulted in many hits. The CCL2 apoprotein produced clusters of thin needles with a diffraction limited to about 2.2 Å. In the presence of the ligand sLe^X tetrasaccharide (Neu5Acα2,3Galβ1,4(Fuc1,3)GlcNAc) (Schubert, M., Bleuler-Martinez, S., et al. 2012), several crystal forms were obtained but the best

diffraction was observed for crystals grown from solutions containing glycerol ethoxylate. All crystals were quite susceptible to radiation damage.

Two and four protein chains were found by molecular replacement in the asymmetric unit of the apoprotein and sLe^X co-crystals, respectively. The structure of the apoprotein was determined and refined to 2.25 Å resolution to values of R = 17.2 % and R_{free} = 21.0 %. The structure comprised residues 8 to 142 in both protein chains and has excellent geometry with no outliers in the Ramachandran plot (Table 1). The sLe^X complex structure was determined and refined to 1.95 Å resolution to values of R = 17.5 % and R_{free} = 20.4 %. The protein structure, comprising residues 7–142 in molecule A, 8-142 in molecule B, 10-141 in molecule C and 12-141 in molecule D, present excellent geometry (Table 1 and validation report available on wwPDB). Protein chains B, C and D presented much higher temperature factors than protein chain A probably as a result of thermal motion as shown from the anisotropic displacement observed with the TLS refinement (Winn, M.D., Isupov, M.N., et al. 2001) (Table 1). The electron density was poor for the side chains of some surface loops so some atoms had to be omitted. As a result, those protein chains present most of the real-space R-value Z-score (RSRZ) outliers and very few water molecules were found associated to them contrary to protein chain A for which the thermal motion is more limited (Table 1 and validation report available on wwPDB). In both structures, the N-terminal His-tag and the first amino acids could not be modelled due to the absence of clear electron density and probably as the result of disorder. In the NMR structures, those residues adopted numerous conformations. Protein chain A is the best defined and small differences to the other protein chains in the asymmetric unit are only observed at the level of surface loops which leads to a rmsd between 0.17 to 0.21 Å. Thus, in the following, we will refer to protein chain A unless otherwise stated.

CCL2 apoprotein adopts the same β-trefoil fold in the crystal than the one observed by NMR spectroscopy in solution (Fig. 4A). It consists of three β-β-β-β repeats with a pseudo C3 symmetry (Schubert, M., Bleuler-Martinez, S., et al. 2012). When compared to the NMR structures, differences are mainly observed at the level of flexible regions such as the surface loops, in particular loops β1-β2, β6-β7, β8-β9 and β11-β12 as defined in (Schubert, M., Bleuler-Martinez, S., et al. 2012) (Fig. S1). Most importantly, the crystal structures of both the apo- and the liganded protein clearly indicate a dimeric arrangement of the protein (Fig. 4A).

Analysis of the quaternary crystal structure with PISA (Krissinel, E. and Henrick, K. 2007) confirmed the presence of a dimer interface with a total buried surface of 1344, 1356 and 1371 $Å^2$ between protein chains A and B in the apoprotein and sLe^X complex and protein chains C and D in the sLe^X complex, respectively. This represents around 12 % of the solvent accessible surface area 11774 $Å^2$ (AB apo), 11631 $Å^2$ (AB sLe^X complex) and 11121 $Å^2$ (CD sLe^X complex).

Surface loops, in particular loops β1−β2, β10−β11, β11−β12, are involved in the formation of the dimer interface. The interface is mainly stabilized by hydrogens bonds formed by the L20 main chain in one protomer and the R18 main and side chain in the other protomer of the dimer (Fig. 4B). The side chain of R18 interacts also with the main chain oxygen of N118 and A119 of the other chain. The NH_3 group of K26 is bonded to the main chain oxygen of G24 of the other protomer of the dimer. There are also two Van der Waals stacking interactions implicating the aromatic ring of F133 and the side chain of R18 of one protomer with the side chains of V19 and P22 of the other protomer respectively. Thus, PISA detailed analysis reveals two main residues involved in dimerization: R18 and F133.

NMR relaxation and deuterium exchange confirm dimerization of CCL2

Two different NMR experiments confirm the X-ray crystallography data and suggest that CCL2 forms also dimers in solution. On the one hand, we measured ^{15}N longitudinal and transverse relaxation times using standard experiments (Farrow, N.A., Muhandiram, R., et al. 1994) and used the ratio between T_1 and T_2 to estimate the rotational correlation time τ_c according to Fushman at el. (Fushman, D., Weisemann, R., et al. 1994) (Supplementary Fig. S2). Using a double logarithmic plot of τ_c versus the molecular weight (Serdyuk, I.N., Zaccai, N.R., et al. 2007), the obtained τ_c value of 16.4 ns points to a molecular weight in the range of 30 to 40 kDa (Fig. 5A), which agrees with a dimer of CCL2.

Dimerization of CCL2 is further supported by a deuterium exchange experiment. Dissolving a lyophilized CCL2 sample in 100% D_2O resulted in an NMR spectrum in which only ^{1}H-^{15}N correlations of amides show up that were protected from $^{1}H/^{2}H$ exchange typically due the involvement into a hydrogen bond (Fig. 5B). Although arginine side chains normally exchange very rapidly, one arginine side chain remained protected for > 50 min. That signal originates from R18 Hε of which a protection cannot be explained by the determined monomeric NMR structure. However, in the crystal structure R18 is located in the center of the dimer interface and Nε forms a hydrogen bond to the backbone carbonyl of L20 of the partner molecule (Fig. 4B). The exchange protected side chain NεHε of R18 indicates thus the presence of an intermolecular hydrogen bond at the dimer interface and a very tight dimer interaction because any dissociation event would lead to immediate exchange. A fast equilibrium between dimer and monomer can be excluded because no significant dissociation of the dimer occurs within 50 min. In conclusion, the exchange protected intermolecular hydrogen bond at the dimer interface NMR undoubtedly confirms the presence of a very tight dimer in solution.

CCL2 dimerizes in solution in a R18- and F133-dependent fashion

Based on the above results, we produced untagged and His-tagged versions of CCL2(F133A) single and CCL2(R18A_F133A) double protein variants for further analysis. To assess the oligomerization state of CCL2 wildtype protein and R18A, F133A and R18A_F133A variants in solution, several experimental approaches were used. On the one hand, we analyzed the proteins using a previously used, partially denaturing gel electrophoresis procedure (Schutz, M., Weiss, E.M., et al. 2010, Singh, S., Pal, K., et al. 2015). For this procedure, identical samples of purified His-tagged proteins were mixed with Laemmli sample buffer. One of these samples was heated to 95°C for 5 min whereas the other one was left at room temperature. Both samples were analyzed on a regular denaturing SDS-PAGE gel and stained with Coomassie brilliant blue. In case of the CCL2 wildtype protein, the gel showed a clear protein band at the expected size of the monomeric protein in the heated sample (Fig. 6A). In the non-heated sample, however, most of the CCL2 protein ran at an apparent MW of 50 kDa indicating oligomerization interpreted as dimerization. In contrast, the non-heated sample of the R18A, F133A and R18A_F133A variants did not show any formation of dimers suggesting that both the single and double substitutions of these residues sufficiently weaken dimerization under these conditions.

As a second approach to assess the oligmeric state of the WT and protein variants, we used Diffusion Ordered nuclear magnetic resonance SpectroscopY (DOSY) (Johnson, C.S. 1999) to measure the translational diffusion coefficients (D) of CCL2 variants at 298K (Fig. S3) and correlated the obtained values with the molecular weight using protein standards ranging from 6.5 to 66 kDa according to Groves et al. (Groves, P., Palczewska, M., et al. 2004). The measured logD value of CCL2 wild type falls in the expected range for a dimer (Fig. 6B). The R18A and F133A variants yielded diffusion constants not significantly different from wild type, indicating that those point mutants diffused still as dimers. In contrast, the

CCL2(R18A_F133A) double variant displayed a significantly smaller logD value indicative of a monomer.

As a third approach to demonstrate the dimerization of CCL2 in solution, size exclusion chromatography (SEC) using a ENrich SEC 70 column (Bio-Rad Ltd) was performed in a buffer composed of 50 mM KH_2PO_4/K_2HPO_4 pH 5.7 and 150 mM NaCl (Fig. 6C). According to the calibration curve (Fig. S4A), the calculated molecular weights were 30.8 kDa for CCL2 WT, 29.8 kDa for CCL2(R18A), 31.5 kDa for CCL2(F133A) and 17.1 kDa for CCL2(R18A_F133A). The expected molecular weight being of 33.1 kDa for a dimer and 16.6 kDa for a monomer, the elution volumes are in accordance with CCL2 WT, F133A and R18A forming dimers and CCL2 (R18A_F133A) forming monomers. Similar results were obtained when the buffer was changed to MES pH 6.0 or Hepes 7.5 (data not shown). For comparison, the same samples were subjected to SEC on a Superdex 75 column (GE Healthcare) in KH_2PO_4/K_2HPO_4 pH 5.7 and 150 mM NaCl (Fig. S4B). According to the calibration curve (Fig. S4C), the calculated molecular weight was 17.8 kDa for CCL2 WT, 16.2 kDa for CCL2(R18A), 18.3 kDa for CCL2(F133A) and 11.7 kDa for CCL2(R18A_F133A). On this resin, the elution volumes of the WT and the single variant proteins are rather in accordance with a monomeric protein as described previously (Schubert, M., Bleuler-Martinez, S., et al. 2012). The difference between the two resins is likely due to a weak interaction between the lectin CCL2 and the carbohydrate-based Superdex resin which leads to a delayed elution of the proteins and gives rise to erroneous conclusions on their molecular weight and hence on their oligomeric state. The ENrich resin is not carbohydrate-based and, hence, no interactions with the CCL2 protein appear to occur. Despite this difference between the resins, both resins reveal a clear difference in the elution profile of the WT and single variants compared to the double variant (Fig. 6C and Fig. S4B). Finally, size measurements by dynamic light scattering (DLS) were performed before and after SEC. SEC fractions were analyzed directly if the

concentration allowed it (>0.2 mg ml^{-1}) or after their concentration. The hydrodynamic radius (R_h) measured for the various samples of CCL2 WT, CCL2(F133A) and CCL2(R18A) was between 2.81 and 3.08 nm which is in agreement with the radius of 3.45 nm measured for the elongated CCL2 dimer from the X-ray structure (Fig. 6D). These results were also obtained for the CCL2 WT and single variant proteins eluted from the Superdex column (data not shown), supporting above hypothesis about the abnormal mobility of the CCL2 protein on this resin. The R_h measured for CCL2 (R18A_F133A) at a concentration below 2 mg ml^{-1} was between 2.03 and 2.12 nm which is in agreement with the monomer radius of 2.05 measured from the X-ray structure. At concentrations higher than 2 mg ml^{-1}, however, the R_h was in the range of the dimer and the solution was polydisperse indicating a mix of monomers and dimers.

To further analyze the monomer-dimer equilibrium we used concentration dependent NMR measurements of the monomeric CCL2(R18A_F133A) double variant (Fig. S5). At a protein concentration of 50 μM narrow line widths are observed, typical for a monomer and at 450 μM concentration some signals move, broaden or disappear. These effected signals originate from the dimer interface indicating a fast to intermediate exchange equilibrium between monomers and dimers.

Taken together, these results demonstrate that CCL2 forms dimers in solution and that this dimerization is weakened in the R18A and F133A single variants and practically abolished in the CCL2(R18A_F133A) double variant.

Interference with CCL2 dimerization goes along with reduction of nematotoxicity

Based on above results, we concluded that the reduced toxicity of the CCL2(R18A) variant was due to a reduced stability of the protein dimer rather than reduced binding to a second ligand.

To confirm this hypothesis, we analyzed the nematotoxicity of the dimerization-deficient CCL2(F133A) single and CCL2(R18A_F133A) double variants. In agreement with our conclusion, the nematotoxicity of the F133A single variant was also reduced in comparison to the WT protein, albeit not as much as the R18A variant, and the double variant protein was further reduced in comparison to the R18A single variant since a 50%/50% mixture of 'empty' vector-containing and CCL2(R18A_F133A)-expressing bacteria showed hardly any toxicity any more (Fig. 2). In order to exclude any effects by general folding defects, we confirmed that purified, His-tagged versions of the R18A and R18A_F133A variant proteins folded properly according to ^{15}N-HSQC finger print spectra (Fig. S6 and S7) and bound to their carbohydrate target comparably to WT as monitored by NMR titration experiments (Fig. 3). The concentration dependence of the NMR spectra of CCL2(R18A_F133A) described above, indicates a monomer at concentrations of 50 μM or lower. In summary, it appears that the degree of dimerization of CCL2 correlates with its nematotoxicity.

The carbohydrate-binding sites of CCL2 are oriented towards one face of the dimer

For the simultaneous binding of multiple binding sites of the lectin to multiple ligands displayed on a cell surface, e.g. in a nematode intestinal epithelium, the location of the binding sites on the lectin surface may be crucial. For this reason, we studied the characteristics and the location of the two carbohydrate binding sites on the CCL2 dimer and compared their location with other structurally characterized fungal β-trefoil lectins.

Inspection of the electron density map after rebuilding with Arp-Warp (Langer, G., Cohen, S.X., et al. 2008) revealed without ambiguity the presence of the Gal and GlcNAc moieties of sLe^X bound to the molecule A of CCL2. The sialic acid moiety appeared in the next refinement step and showed some disorder especially for the glycerol substituent for whose only clear density was observed at 0.7 σ and was therefore modelled with a reduced

occupancy (Fig. 7).

The sLe^X is found in the binding site previously described for GlcNAcβ1,4(Fucα1,3)GlcNAcβ1O$(CH_2)_5$COOH (Schubert, M., Bleuler-Martinez, S., et al. 2012) and all interactions are summarized in Table 2, Fig. 7 and Fig. S8. Upon binding of sLe^X to CCL2, only minor differences are observed with the apoprotein structure especially at the level of the binding site. There is a small rigid body movement of the β9-β10 loop which moves away by 0.8 Å in order to avoid close contact and form correct H-bonding between G108 and the fucose O4 hydroxyl. They are also slight changes at the level of loop β6- β7 that lead to optimal Van der Waals interaction between W78 side chain and the sialic acid moiety. The Fuc, Gal and GlcNAc moieties present orientations equivalent to the one observed in GlcNAcβ1,4(Fucα1,3)GlcNAcβ1O(CH2)5COOH (Fig. S8). The rotamer of the hydroxymethyl of the GlcNAc is however different in both complexes. There are 8 direct hydrogen bonds and 10 water-mediated hydrogen bonds. The Van der Waals interactions involve Y57, W8, Y92, and W94 like in the case of GlcNAcβ1,4(Fucα1,3)GlcNAcβ1O(CH2)5COOH. The binding site accommodates the galactose moiety very well and in the absence of the acetamido group, the side chain of Y92 reorients a little since there is no more stacking interaction with it. Comparison of the two complex structures shows clearly that there are some rearrangements of the loop β6-β7 as a result of movements of G79 and G81 (Fig. S8). The orientation observed in the fucosylated chitobiose complex cannot permit the accommodation of the Sialyl moiety and would result in steric conflict. The new orientation allows the side chain of W78 to optimize stacking with the sialyl moiety.

The sLe^X is bound in a similar conformation observed in many other lectin complexes with this oligosaccharide such as the staphylococcal superantigen-like protein 11 (Chung, M.C., Wines, B.D., et al. 2007). The torsion angles for the different glycosidic linkages all fall in

low energy regions, in particular the glycosidic linkage torsional angles Φ (O6-C2-O-C'x) and Ψ (C2-O-C'x-C'x+1) for the αNeu5Ac(2-3)Gal linkage with values of 48.03º and 100º, respectively (Table S1) (Imberty, A. and Perez, S. 2000). The Le^x trisaccharide is present in a closed conformation, in which the galactose moiety is stacked onto the fucose moiety as described earlier (Zierke, M., Smiesko, M., et al. 2013). This rigid low energy conformation is the one found bound to most lectins apart of two special cases: the lectins RSL from *Ralstonia solanacearum* and BambL from *Burkholderia ambifaria* (Topin, J., Lelimousin, M., et al. 2016).

The residues involved in the glycan-binding interface residues of CCL2 are spatially well separated from those implicated in the dimerization. By using the analogy between the β-trefoil structure with a tree (Renko, M., Sabotic, J., et al. 2010), the ligand binding occurs mainly in the upper stem, the lower and upper crown at the interface between subdomains β and γ and therefore on the side of CCL2 (Fig. 8A). This is very different when compared to the other dimerizing fungal β-trefoil lectins like CNL fom *C. nebularis* (Pohleven, J., Renko, M., et al. 2012), BEL from *B. edulis* (Bovi, M., Cenci, L., et al. 2013), SSA from *S. sclerotiorum* (Sulzenbacher, G., Roig-Zamboni, V., et al. 2010) and RSA from *R. solani* (Skamnaki, V.T., Peumans, W.J., et al. 2013) where binding occurs at the top, in the upper crown of one or several subdomains according to the lectin (Fig. 8A).

Dimeric β-trefoil lectins from fungi differ considerably in their three-dimensional arrangements

Since a considerable number of dimerizing fungal β-trefoil lectins have been structurally characterized, we compared the three-dimensional arrangements of the carbohydrate-binding sites and the dimerization interfaces in these lectins with the one of CCL2. This comparison revealed striking differences and a large variety among those lectins (Fig. 8). They all present

different dimerization interfaces, which often involve some surface loops. Those are dissimilar in length and orientation from one lectin to the other whilst the beta-strands are structurally well conserved. In CCL2, the β1−β2 loop for example plays a critical role in the dimer formation. It includes an insertion of four to five residues compared to the above lectins, which allows this loop to have the proper orientation and to bring some necessary residues for CCL2 dimerization like L20, P22 and G24. The stem and the lower crown of subdomains α and γ are associated with CCL2 dimer formation whereas other subdomains are used by the other β-trefoil dimers. Thus, the CCL2 dimer is not only unique in the localization of its dimer interface but also in the orientation of the two molecules: one molecule is in a side-by-side contact with another up-side down oriented one (Fig. 8). The pseudo-symmetry axes of both molecules are perpendicular to the axis connecting both molecules and are thus not tilted. However, both monomers are twisted by a torsion angle of 49°. The two glycan-binding sites are located at the extremities of the dimer in a front view (Fig. 8B, left panel), but are facing the same side in a side view (Fig. 8B, right panel), which will enable binding to two ligands exposed on a surface. The orientations of the molecules reflected by tilt and torsion angles vary extremely. Like SSA, CCL2 presents only one binding surface whilst the others present two but the distances between the binding sites are distinct with 55 Å in CCL2 compared to 38 Å for SSA (Fig. 8B).

Discussion

In retrospect, we are surprised that we obtained such high quality NMR data with a 30 kDa dimer complex (Schubert, M., Bleuler-Martinez, S., et al. 2012). Normally deuteration and TROSY techniques are required to study proteins of a size > 25 kDa. The compactness of the dimer and the elevated temperature of 37ºC are likely the reasons for the fairly sharp NMR signals. The previously obtained, misleading size exclusion results seem to be due to very

weak interactions of the lectin with the polysaccharide column material resulting in slightly longer elution time and thus smaller apparent size. Although we did not notice the presence of a dimer in the previous NMR structure, the main conclusions of that study, namely the β-trefoil fold, the unusual binding location and the structural details of the carbohydrate recognition interface responsible for specificity and affinity are not questioned with the present paper. However, the results of this study clearly demonstrate that the formation of dimers is essential for the function of the lectin as a toxin.

Previous hypotheses with regard to the toxicity mechanism of CCL2 towards invertebrates were based on the assumption of a monomeric state of CCL2 and the absence of additional carbohydrate binding sites on the protomer. Here, we show that mutations in the dimerization interface, revealed by X-ray crystallography and confirmed by NMR, interfere with dimer formation as demonstrated by size exclusion chromatography combined with dynamic light scattering, partially denaturing gel electrophoresis and NMR analysis. Although these mutations do not affect the binding of the protein to the carbohydrate ligand, they clearly reduce the nematotoxicity of the protein. These results suggest that the bivalency of CCL2 is a requirement for its nematotoxicity and that the toxicity mechanism involves crosslinking of N-glycoproteins on the surface of intestinal epithelial cells of the nematode. The orientation of the carbohydrate binding sites pointing towards the same face of the protein dimer may thereby facilitate the simultaneous binding to two N-glycoproteins that are fixed to the same cell surface. As mentioned in the introduction, the crosslinking of N-glycoproteins of the plasma membrane or the glycocalyx may either lead to lectin-induced activation of intracellular signaling pathways or to defects in plasma membrane or glycocalyx function possibly by changed vertical (in/out) and lateral (rafts) trafficking of the target glycoproteins. According to a recent study, this mechanism does not appear to involve surface binding but not internalization of CCL2 (Stutz, K., Kaech, A., et al. 2015).

The major characteristics i.e. the structure (β-trefoil, homodimer), the orientation and valency (one binding site per monomer, both binding sites located on one face of the dimer) and the toxicity mechanism (dependency on carbohydrate-binding, binding to microvilli but no apparent internalization) of CCL2 are very similar to another fungal lectin SSA from *S. sclerotiorum* which shows strong toxicity towards insects (Hamshou, M., Smagghe, G., et al. 2010, Sulzenbacher, G., Roig-Zamboni, V., et al. 2010). The only difference is in the architecture of the dimer interface (different geometry and residues involved), the location of the carbohydrate binding site (canonical site in case of SSA) and the carbohydrate specificity since SSA is specific for β-1,3 or β-1,4-linked terminal Gal or GalNAc residues. In contrast, the β-trefoil lectin BEL from *Boletus edulis* with a different dimer geometry and six carbohydrate binding sites was shown to be internalized into mammalian epithelial Caco-2 cells (Bovi, M., Cenci, L., et al. 2013) although it has the same carbohydrate-specificity as SSA. Thus, it is possible that toxic β-trefoil lectins with a low number of carbohydrate binding sites may exert toxicity by mere crosslinking of glycoconjugates at the cell surface whereas the toxicity mechanism of lectins with a higher valency may involve internalization. On the other hand, the observed differences in toxicity mechanisms might also be due to differences in the target glycoconjugate between the different β-trefoil lectins. Nematotoxicity of CCL2 was previously shown to be dependent on the binding of N-glycoproteins on the surface of intestinal epithelial cells of *C. elegans* (Schubert, M., Bleuler-Martinez, S., et al. 2012, Stutz, K., Kaech, A., et al. 2015). In contrast, the chimeric β-trefoil lectin MOA from *Marasmius oreades* was previously shown to exert nematotoxicity by binding to glycosphingolipids of these cells (Wohlschlager, T., Butschi, A., et al. 2011). Accordingly, bacterial toxins and entire bacteria were recently shown to trigger their own endocytosis by binding and clustering of glycolipids (lipid zipper mechanism) (Aigal, S., Claudinon, J., et al. 2015, Eierhoff, T., Bastian, B., et al. 2014). Similarly, binding of galectin-3 to glycoproteins

was recently shown to trigger endocytosis of these proteins by additional interaction of galectin-3 with glycolipids in a mechanism independent of clathrin (CLIC mechanism) (Lakshminarayan, R., Wunder, C., et al. 2014). Based on the differences in the number and the arrangement of the carbohydrate-binding sites between the characterized dimerizing β-trefoil lectins from fungi, we hypothesize that these lectins act by simultaneous binding of multiple glycans and forcing these glycans into a spatial arrangement that is determined by discrete distances of the carbohydrate binding sites on the lectin surface. If these glycans are part of membrane-bound glycoconjugates, this lectin-enforced spatial arrangement may have a profound effect on the function of the respective biological membrane and the underlying cell. The effect likely depends on the nature of the glycoconjugate (glycoproteins, glycolipids) and the specific spatial arrangement of the glycoconjugates enforced by the lectin.

In summary, after CNL (Pohleven, J., Obermajer, N., et al. 2009, Pohleven, J., Renko, M., et al. 2012), CCL2 is only the second example of a lectin for which multivalency was shown to be essential for biological function (toxicity). The significance of this finding might go beyond basic research as CCL2 was recently shown to bind specifically to cancerous epithelia from pancreatic tumors which are characterized by the display of sialyl-Lewis^X (Singh, S., Pal, K., et al. 2015). CCL2 has not been tested for toxicity towards these cells yet but an application of this lectin in cancer diagnostics or therapy may be feasible.

Materials and methods

Cultivation conditions

E. coli was cultivated on LB or NGM medium as described (Sambrook, J. and Russell, D.W. 2001, Stiernagle, T. 2006). Cultivation of *C. elegans* was performed according to Stiernagle, 2006 (Stiernagle, T. 2006).

Site-directed mutagenesis

Mutant versions of the CCL2-expressing plasmids pET24-CCL2 and pET24-His8-CCL2 were generated using the QuikChange™ site-directed mutagenesis kit according to the manufacturer's recommendations (Stratagene).

Protein expression and purification

Untagged and His8-tagged versions of CCL2 were expressed in the cytoplasm of *E. coli* BL21(DE3) and His8-tagged versions were purified by metal-affinity chromatography as described previously (Schubert, M., Bleuler-Martinez, S., et al. 2012).

Crystallization and structures determination

Crystals of CCL2 were obtained by the hanging drop vapor diffusion method using 2 µl of drops containing a 50:50 (v/v) mix of protein and reservoir solution at 20 °C. For the complex with sLe^X, the protein at 5 mg/ml in 10 mM Hepes/NaOH buffer pH 7.5 and 150 mM NaCl was incubated with sLe^X 1.2 mM during 1h at room temperature prior to co-crystallization. Crystals were obtained in two days from solution 2-28 from the Midas screen (Molecular Dimensions Ltd): 30% glycerol ethoxylate and 100mM Tris/HCl pH 8.5 (Grimm, C., Chari, A., et al. 2010). Crystals of the apoprotein were obtained from a protein solution at 5.1 mg/ml in 100 mM sodium citrate pH 5.8 150 mM NaCl. Clusters of long and thin needles

were obtained in 24 hours from solution 1-14 of the Midas screen: 15% pentaerythritol propoxylate 5/4 PO/OH, 0.2 M sodium thiocyanate and 0.1 M Hepes pH 7.5 (Grimm, C., Chari, A., et al. 2010). A single crystal or broken needle was directly mounted in a cryoloop and flash-freezed in liquid nitrogen. Diffraction data for the apoprotein and sLe^X complex were collected at 100 K at the European Synchrotron Radiation Facility (Grenoble, France) at beamlines BM30A-FIP and ID23-1 using a ADSC Q315r CCD and a Pilatus detector respectively. The data were processed using XDS (Kabsch, W. 2010). All further computing was performed using the CCP4 suite (Winn, M.D., Ballard, C.C., et al. 2011). Data quality statistics are summarized in Table 1. Both structures were solved by molecular replacement using PHASER (McCoy, A.J., Grosse-Kunstleve, R.W., et al. 2007). For the sLe^X complex, the 2LIE coordinates of model 1 comprising residues 20-153 were used as search model to find four molecules in the asymmetric unit (Schubert, M., Bleuler-Martinez, S., et al. 2012). The dimer coordinates of the sLe^X complex were then used as search model to solve the apoprotein structure. Five percent of the observations were set aside for cross-validation analysis, and hydrogen atoms were added in their riding positions and used for geometry and structure-factor calculations. After initial rebuilding using ARP/WARP (Langer, G., Cohen, S.X., et al. 2008), the structures were refined with TLS and restrained maximum likelihood refinement using REFMAC 5.8 (Murshudov, G.N., Skubak, P., et al. 2011) iterated with manual rebuilding in Coot (Emsley, P., Lohkamp, B., et al. 2010). Incorporation of the ligand was performed after inspection of the 2mFo-DFc weighted maps. Water molecules, introduced automatically using Coot, were inspected manually. The stereochemical quality of the models was assessed with the program Molprobity (Chen, V.B., Arendall, W.B., 3rd, et al. 2010), and coordinates were deposited in the Protein Data Bank under codes 4USO and 4USP for the sLe^X complex and apo CCL2 structure respectively.

SDS-PAGE-based oligomerization assay

The oligomerization assay employing partially denatured protein samples for SDS-PAGE was performed as described previously (Singh, S., Pal, K., et al. 2015). In short, protein samples were mixed with 2x SDS-PAGE (Laemmli) loading dye and left at RT or heated to 95°C for 5 min, run on regular SDS-PAGE and stained with Coomassie Brilliant blue R-250.

Size exclusion chromatography and dynamic light scattering

Size exclusion chromatography (SEC) was performed on an ENrich SEC 70 10 x 300 column (Bio-Rad) at a flow rate of 1 ml min^{-1} in a buffer composed of 50 mM sodium KH_2PO_4/K_2HPO_4 pH 5.7 and 150 mM NaCl. 100 µl of protein solutions at concentrations of 6 mg ml^{-1} for R18A_F133A, 4.2 mg ml^{-1} for F133A, 9 mg ml^{-1} for R18A and 3 mg ml^{-1} for WT were injected and the eluate was monitored at 280 nm. Another size exclusion chromatography was performed on an Superdex 75 10 x 300 column (GE Healthcare) at a flow rate of 0.5 ml min^{-1} in the same buffer. 100 µl of protein solutions at concentrations of 5 mg ml^{-1} for R18A_F133A, 4 mg ml^{-1} for F133A, 4 mg ml^{-1} for R18A and 3 mg ml^{-1} for WT were injected and the eluate was monitored at 280 nm. A calibration curve was done on each column with the following standards: Bovine serum albumine 66.5 kDa, ovalbumin 43 kDa and RNAse A 13.7 kDa. The CCL2 elution peaks were concentrated on a Vivaspin 500 (5 kDa cut off, Sartorius) to a volume around 90µl and size measurement were done using dynamic light scattering (DLS) on a zetasizer Nano ZS (Malvern instrument Ltd) using small size quartz cuvettes. Proteins were also subjected to DLS analysis prior to SEC.

Carbohydrates

Chemical synthesis of GlcNAcβ1,4(Fucα1,3)GlcNAcβ1O$(CH_2)_5$COOH was described previously (Schubert, M., Bleuler-Martinez, S., et al. 2012). sLe^X was purchased from Elicityl

(Crolles, France).

NMR Spectroscopy

All NMR spectra were recorded on Bruker Avance HD spectrometer operating at 500, 600, 700, or 900 MHz equipped with triple-resonance cryogenetic or room temperature probes at 313 K in a buffer of 50 mM KH_2PO_4/K_2HPO_4 pH 5.8 and 100 mM NaCl. Standard ^{15}N-HSQC spectra of WT and mutant proteins were recorded at protein concentrations between 240 and 485 μM using 1024×256 points. The detection of exchange protected amide protons was achieved with a ^{15}N-HSQC recorded on a lyophilized CCL2 sample after dissolving it in D_2O at a concentration of 280 μM (buffer conditions identical to samples in H_2O/D_2O). 4 number of scans and 1024×128 points were recorded. Binding to α1,3-fucosylated chitobiose was tested by titrating concentrated chitobiose solution (28 mM in identical buffer) into protein solutions of WT, R18A and R18A_F133A at concentrations of 400 μM, 240 μM and 330 μM, respectively.2D DOSY experiments were measured with aqueous samples of WT, R18A, F133A and R18A/F133A protein variants (unlabeled) at 300 μM concentration. The standard Bruker stebpgp1s19 pulse sequence, with stimulated echo, bipolar gradient pulses, one spoil gradient and 3-9-19 water suppression, was used with a linear gradient (53.5 G cm^{-1}) stepped between 2 and 95%. Protein standards were measured at concentrations ranging from 200 μM to 400 μM. They included aprotinin, lysozyme, carbonic anhydrase, ovalbumin and bovine serum albumin albumin. Typical parameters were: td = 32, a gradient duration of δ=1.0-1.4 ms and an echo delay Δ between 200 ms and 400 ms. T_1 and T_2 relaxation times were measured with ^{13}C/^{15}N-labeld CCL2 WT at a concentration of 1mM using the Bruker pulse sequences hsqct1etf3gpsi.2 and hsqct2etf3gpsi, respectively. Seven T_1 relaxation delays

ranging from 1 ms to 2 sec and six T_2 relaxation delays ranging from 4 to 130 ms were applied using ns=32 and td=200.

Spectra were processed using Topspin 3.2 (Bruker Biospin) and analyzed by Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco, USA). ^{1}H chemical shifts are referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS). ^{15}N chemical shifts are indirectly referenced using a scaling factor of 0.101329118 (Markley, J.L., Bax, A., et al. 1998).

Nematotoxicity assays

Toxicity assays using CCL2-expressing *E. coli* BL21(DE3) and *C. elegans* N2 were performed as described previously (Kunzler, M., Bleuler-Martinez, S., et al. 2010). In brief, ~30 L1 larvae were added to 100 µl of bacterial suspension (final A_{600} = 2) in sterile PBS in 96-well plates. In order to detect subtle differences in nematotoxicity between the CCL2 variants, CCL2-expressing *E. coli* were mixed at different percentages with vector control-containing *E. coli* as described previously (Yan, S., Bleuler-Martinez, S., et al. 2012). The percentage of individuals reaching L4 stage was quantified after 48 h of incubation at 20 °C. Each assay was performed in quintuplicate. Welch's-t test was performed to validate statistical significance of the differences observed.

Comparing dimer structures of β-trefoil lectins

The arrangements of the dimers were analyzed using Molmol (Koradi, R., Billeter, M., et al. 1996). Molecules were oriented till a symmetry axis was visible, followed by x-, y- and z-rotations till the pseudo-symmetry axis of an individual β-trefoil molecule was visible. The derived angles have an estimated error of 1°.

Supplementary data

Supplementary data for this article are available online http://glycob.oxfordjournals.org/.

Funding

This work was supported by funds of the Swiss National Science Foundation (Grant. No. 31003A_130671 to M. K.), University of Zürich, ETH Zürich, University of Salzburg and CNRS.

Acknowledgements

We thank Markus Aebi for helpful discussions and critical reading of the manuscript. We are grateful to the European Synchrotron Radiation Facility (Grenoble, France) for access and technical support to beamlines BM30A-FIP and ID23-1, and to ETH Zürich for access to the biomolecular NMR facility. We thank the US Department of Energy Joint Genome Institute (DOE JGI) and the individual project leaders from academia for sharing unpublished fungal genome and transcriptome data. We are grateful to the laboratories of Johann Brandstetter and Gerhard Obermeyer (Univ. Salzburg) for kindly providing protein standards for the DOSY calibration curve, and Jesús-Jiménez Barbero (BioGune Bilbao) and Dolores Díaz (Univ. Cologne) for helpful discussions concerning DOSY.

Conflict of interest statement

None declared

Abbreviations

2D-[^{1}H-^{15}N]-HSQC, two-dimensional [^{1}H-^{15}N] Heteronuclear Single Quantum Coherence;

sLe^{X}, Sialyl-Lewis^{X}

References

Aigal S, Claudinon J, Romer W. 2015. Plasma membrane reorganization: A glycolipid gateway for microbes. *Biochim et Biophys Acta*, 1853:858-871.

Arthur CM, Cummings RD, Stowell SR. 2014. Using glycan microarrays to understand immunity. *Curr Opin Chem Biol*, 18:55-61.

Boscher C, Dennis JW, Nabi IR. 2011. Glycosylation, galectins and cellular signaling. *Curr Opin Cell Biol*, 23:383-392.

Bovi M, Cenci L, Perduca M, Capaldi S, Carrizo ME, Civiero L, Chiarelli LR, Galliano M, Monaco HL. 2013. BEL beta-trefoil: A novel lectin with antineoplastic properties in king bolete (*Boletus edulis*) mushrooms. *Glycobiology*, 23:578-592.

Brewer CF, Miceli MC, Baum LG. 2002. Clusters, bundles, arrays and lattices: novel mechanisms for lectin-saccharide-mediated cellular interactions. *Curr Opin Struct Biol*, 12:616-623.

Caccia S, Van Damme EJ, De Vos WH, Smagghe G. 2012. Mechanism of entomotoxicity of the plant lectin from Hippeastrum hybrid (Amaryllis) in *Spodoptera littoralis* larvae. *J Insect Physiol*, 58:1177-1183.

Cha SK, Ortega B, Kurosu H, Rosenblatt KP, Kuro OM, Huang CL. 2008. Removal of sialic acid involving Klotho causes cell-surface retention of TRPV5 channel via binding to galectin-1. *Proc Natl Acad Sci U S A*, 105:9805-9810.

Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D*, 66:12-21.

Chung MC, Wines BD, Baker H, Langley RJ, Baker EN, Fraser JD. 2007. The crystal structure of staphylococcal superantigen-like protein 11 in complex with sialyl Lewis X reveals the mechanism for cell binding and immune inhibition. *Mol Microbiol*, 66:1342-1355.

Dayie KT, Wagner G, Lefevre JF. 1996. Theory and practice of nuclear spin relaxation in proteins. *Annu Rev Phys Chem*, 47:243-282.

Eierhoff T, Bastian B, Thuenauer R, Madl J, Audfray A, Aigal S, Juillot S, Rydell GE, Muller S, de Bentzmann S*, et al.* 2014. A lipid zipper triggers bacterial invasion. *Proc Natl Acad Sci U S A*, 111:12895-12900.

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, 66:486-501.

Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, 33:5984-6003.

Fushman D, Weisemann R, Thuring H, Ruterjans H. 1994. Backbone dynamics of ribonuclease T1 and its complex with 2'GMP studied by two-dimensional heteronuclear NMR spectroscopy. *J Biomol NMR*, 4:61-78.

Gallo RL, Hooper LV. 2012. Epithelial antimicrobial defence of the skin and intestine. *Nat Rev Immunol*, 12:503-516.

Grimm C, Chari A, Reuter K, Fischer U. 2010. A crystallization screen based on alternative polymeric precipitants. *Acta Crystallogr D Biol Crystallogr*, 66:685-697.

Groves P, Palczewska M, Molero MD, Batta G, Canada FJ, J J-B. 2004. Protein molecular weight standards can compensate systematic errors in diffusion-ordered spectroscopy. *Anal Biochem*, 331:395-397.

Hamshou M, Smagghe G, Shahidi-Noghabi S, De Geyter E, Lannoo N, Van Damme EJ. 2010. Insecticidal properties of *Sclerotinia sclerotiorum* agglutinin and its interaction with insect tissues and cells. *Insect Biochem Mol Biol*, 40:883-890.

Hamshou M, Van Damme EJ, Vandenborre G, Ghesquiere B, Trooskens G, Gevaert K, Smagghe G. 2012. GalNAc/Gal-binding *Rhizoctonia solani* agglutinin has antiproliferative activity in *Drosophila melanogaster* S2 cells via MAPK and JAK/STAT signaling. *PLoS ONE*, 7:e33680.
Heim C, Hertzberg H, Butschi A, Bleuler-Martinez S, Aebi M, Deplazes P, Kunzler M, Stefanic S. 2015. Inhibition of *Haemonchus contortus* larval development by fungal lectins. *Parasit Vectors*, 8:425.
Imberty A, Perez S. 2000. Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations. *Chem Rev*, 100:4567-4588.
Johnson CS. 1999. Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Prog Nucl Mag Res Sp*, 34:203-256.
Kabsch W. 2010. Xds. *Acta Crystallogr D Biol Crystallogr*, 66:125-132.
Koradi R, Billeter M, Wuthrich K. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Gr*, 14:51-55, 29-32.
Krissinel E, Henrick K. 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, 372:774-797.
Kunzler M. 2015. Hitting the Sweet Spot: Glycans as Targets of Fungal Defense Effector Proteins. *Molecules*, 20:8144-8167.
Kunzler M, Bleuler-Martinez S, Butschi A, Garbani M, Luthy P, Hengartner MO, Aebi M. 2010. Biotoxicity assays for fruiting body lectins and other cytoplasmic proteins. *Methods Enzymol*, 480:141-150.
Lakshminarayan R, Wunder C, Becken U, Howes MT, Benzing C, Arumugam S, Sales S, Ariotti N, Chambon V, Lamaze C*, et al.* 2014. Galectin-3 drives glycosphingolipid-dependent biogenesis of clathrin-independent carriers. *Nat Cell Biol*, 16:595-606.
Langer G, Cohen SX, Lamzin VS, Perrakis A. 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc*, 3:1171-1179.
Lannoo N, Van Damme EJ. 2014. Lectin domains at the frontiers of plant defense. *Front Plant Sci*, 5:397.
Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wuthrich K. 1998. Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on the Standardization of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy. *J Biomol NMR*, 12:1-23.
McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *JApp Crystallogr*, 40:658-674.
Mockl L, Horst AK, Kolbe K, Lindhorst TK, Brauchle C. 2015. Microdomain Formation Controls Spatiotemporal Dynamics of Cell-Surface Glycoproteins. *Chembiochem : a Eur JChemical Biol*, 16:2023-2028.
Mukherjee S, Zheng H, Derebe MG, Callenberg KM, Partch CL, Rollins D, Propheter DC, Rizo J, Grabe M, Jiang QX*, et al.* 2014. Antibacterial membrane attack by a pore-forming intestinal C-type lectin. *Nature*, 505:103-107.
Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr*, 67:355-367.
Nicholas KB, Nicholas Jr. HB, Deerfield II DW. 1997. GeneDoc - Analysis and Visualization of Genetic Variation. *embnet.news*, 4:1-4.
Pace KE, Hahn HP, Pang M, Nguyen JT, Baum LG. 2000. CD7 delivers a pro-apoptotic signal during galectin-1-induced T cell death. *J Immunol*, 165:2331-2334.

Pace KE, Lee C, Stewart PL, Baum LG. 1999. Restricted receptor segregation into membrane microdomains occurs on human T cells during apoptosis induced by galectin-1. *J Immunol*, 163:3801-3811.
Partridge EA, Le Roy C, Di Guglielmo GM, Pawling J, Cheung P, Granovsky M, Nabi IR, Wrana JL, Dennis JW. 2004. Regulation of cytokine receptors by Golgi N-glycan processing and endocytosis. *Science*, 306:120-124.
Plaza DF, Lin CW, van der Velden NS, Aebi M, Kunzler M. 2014. Comparative transcriptomics of the model mushroom *Coprinopsis cinerea* reveals tissue-specific armories and a conserved circuitry for sexual development. *BMC Genomics*, 15:492.
Pohleven J, Obermajer N, Sabotic J, Anzlovar S, Sepcic K, Kos J, Kralj B, Strukelj B, Brzin J. 2009. Purification, characterization and cloning of a ricin B-like lectin from mushroom *Clitocybe nebularis* with antiproliferative activity against human leukemic T cells. *Biochim Biophys Acta*, 1790:173-181.
Pohleven J, Renko M, Magister S, Smith DF, Kunzler M, Strukelj B, Turk D, Kos J, Sabotic J. 2012. Bivalent carbohydrate binding is required for biological activity of *Clitocybe nebularis* lectin (CNL), the N,N'-diacetyllactosediamine (GalNAcbeta1-4GlcNAc, LacdiNAc)-specific lectin from basidiomycete *C. nebularis*. *J Biol Chem*, 287:10602-10612.
Renko M, Sabotic J, Mihelic M, Brzin J, Kos J, Turk D. 2010. Versatile Loops in Mycocypins Inhibit Three Protease Families. *J Biol Chem*, 285:308-316.
Sabotic J, Ohm RA, Kunzler M. 2015. Entomotoxic and nematotoxic lectins and protease inhibitors from fungal fruiting bodies. *App Microbiol Biotechnol*, 100:91-111.
Sacchettini JC, Baum LG, Brewer CF. 2001. Multivalent protein-carbohydrate interactions. A new paradigm for supermolecular assembly and signal transduction. *Biochemistry*, 40:3009-3015.
Sambrook J, Russell DW. 2001. *Molecular Cloning: A Laboratory Manual*. 3 ed. Cold Spring Harbor Laboratory Press:New York.
Schubert M, Bleuler-Martinez S, Butschi A, Walti MA, Egloff P, Stutz K, Yan S, Collot M, Mallet JM, Wilson IB*, et al.* 2012. Correction: Plasticity of the beta-Trefoil Protein Fold in the Recognition and Control of Invertebrate Predators and Parasites by a Fungal Defence System. *PLoS Pathog*, 8:e1002706.
Schutz M, Weiss EM, Schindler M, Hallstrom T, Zipfel PF, Linke D, Autenrieth IB. 2010. Trimer stability of YadA is critical for virulence of Yersinia enterocolitica. *Infect Immun*, 78:2677-2690.
Serdyuk IN, Zaccai NR, Zaccai J. 2007. Methods in Molecular Biophysics: Structure, Dynamics, Function. *Methods in Molecular Biophysics: Structure, Dynamics, Function*:1-1120.
Singh S, Pal K, Yadav J, Tang H, Partyka K, Kletter D, Hsueh P, Ensink E, Kc B, Hostetter G*, et al.* 2015. Upregulation of Glycans Containing 3' Fucose in a Subset of Pancreatic Cancers Uncovered Using Fusion-Tagged Lectins. *J Proteome Res*, 14:2594-2605.
Skamnaki VT, Peumans WJ, Kantsadi AL, Cubeta MA, Plas K, Pakala S, Zographos SE, Smagghe G, Nierman WC, Van Damme EJ*, et al.* 2013. Structural analysis of the *Rhizoctonia solani* agglutinin reveals a domain-swapping dimeric assembly. *FEBS J*, 280:1750-1763.
Stiernagle T. 2006. Maintenance of C. elegans. *WormBook*:1-11.
Stillman BN, Hsu DK, Pang M, Brewer CF, Johnson P, Liu FT, Baum LG. 2006. Galectin-3 and galectin-1 bind distinct cell surface glycoprotein receptors to induce T cell death. *J Immunol*, 176:778-789.
Stowell SR, Arthur CM, Dias-Baruffi M, Rodrigues LC, Gourdine JP, Heimburg-Molinaro J, Ju T, Molinaro RJ, Rivera-Marrero C, Xia B*, et al.* 2010. Innate immune lectins kill bacteria expressing blood group antigen. *Nat Med*, 16:295-301.

Stutz K, Kaech A, Aebi M, Kunzler M, Hengartner MO. 2015. Disruption of the *C. elegans* Intestinal Brush Border by the Fungal Lectin CCL2 Phenocopies Dietary Lectin Toxicity in Mammals. *PLoS One*, 10:e0129381.

Sulzenbacher G, Roig-Zamboni V, Peumans WJ, Rouge P, Van Damme EJ, Bourne Y. 2010. Crystal structure of the GalNAc/Gal-specific agglutinin from the phytopathogenic ascomycete *Sclerotinia sclerotiorum* reveals novel adaptation of a beta-trefoil domain. *J Mol Biol*, 400:715-723.

Topin J, Lelimousin M, Arnaud J, Audfray A, Perez S, Varrot A, Imberty A. 2016. The Hidden Conformation of Lewis x, a Human Histo-Blood Group Antigen, Is a Determinant for Recognition by Pathogen Lectins. *ACS Chem Biol*, 11:2011-2020.

Torreno-Pina JA, Castro BM, Manzo C, Buschow SI, Cambi A, Garcia-Parajo MF. 2014. Enhanced receptor-clathrin interactions induced by N-glycan-mediated membrane micropatterning. *Proc Natl Acad Sci U S A*, 111:11037-11042.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A*, et al.* 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*, 67:235-242.

Winn MD, Isupov MN, Murshudov GN. 2001. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr*, 57:122-133.

Wohlschlager T, Butschi A, Zurfluh K, Vonesch SC, Auf dem Keller U, Gehrig P, Bleuler-Martinez S, Hengartner MO, Aebi M, Kunzler M. 2011. Nematotoxicity of *Marasmius Oreades* Agglutinin (Moa) Depends on Glycolipid-Binding and Cysteine Protease Activity. *J Biol Chem*, 286:30337-30343.

Yan S, Bleuler-Martinez S, Plaza Gutierrez DF, Kuenzler M, Aebi M, Joachim A, Razzazi-Fazeli E, Jantsch V, Geyer R, Wilson IB*, et al.* 2012. Galactosylated fucose epitopes in nematodes: increased expression in a *Caenorhabditis* mutant associated with altered lectin sensitivity and occurrence in parasitic species. *J Biol Chem*, 287:28276-28290.

Zierke M, Smiesko M, Rabbani S, Aeschbacher T, Cutting B, Allain FH, Schubert M, 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.

Zurga S, Pohleven J, Renko M, Bleuler-Martinez S, Sosnowski P, Turk D, Kunzler M, Kos J, Sabotic J. 2014. A novel beta-trefoil lectin from the parasol mushroom (*Macrolepiota procera*) is nematotoxic. *FEBS J*, 281:3489-3506.

Figure Legends

Figure 1. ***Multiple sequence alignment of CCL2 homologs encoded in different fungal genomes.*** Sequences were retrieved from the DOE JGI Mycocosm via a BLASTP search against the filtered model proteins of all available fungal genomes and carry the following abbreviations and JGI gene model numbers: CCL2_A: CCL2 of *C. cinerea* strain AmutBmut; CCL2_O: CCL2 of *C. cinerea* strain Okayama7; CCL1_A: CCL1 of *C. cinerea* strain AmutBmut; CCL1_O: CCL1 of *C. cinerea* strain Okayama7; CV_L1: *Crepidotus variabilis* lectin 1 (Crevar1_905694); CV_L2: *Crepidotus variabilis* lectin 2 (Crevar1_109630); PP_L1: *Postia placenta* lectin 1 (Pospl1_130016); PP_L2: *Postia placenta* lectin 2 (Pospl1_121916); SL_L1: *Serpula lacrymans* lectin 1 (SerlaS7_144703); CP_L1: *Coniophora puteana* lectin 1 (Conpu1_119225); PO_L1: *Pleurotus ostreatus* lectin 1 (PleosPC9_89828); PO_L2: *Pleurotus ostreatus* lectin 2 (PleosPC15_1043947); PO_L3: *Pleurotus ostreatus* lectin 3 (PleosPC9_64199); PO_L4: *Pleurotus ostreatus* lectin 4 (PleosPC15_1065820); GC_L1: *Gymnopilus chrysopellus* lectin 1 (Gymch1_1574453); CU_L1: *Cerrena unicolor* lectin 1 (Cerun2_380560); AP_L1: *Agrocybe pediades* lectin 1 (Agrped1_766896); AP_L2: *Agrocybe pediades* lectin 2 (Agrped1_766895); RV_L1: *Rhizopogon vinicolor* lectin 1 (Rhivi1_768162); RV_L2: *Rhizopogon vinicolor* lectin 2 (Rhivi1_700206); ArO_L1: *Armillaria ostoyae* lectin 1 (Armost1_894609); ArO_L2: *Armillaria ostoyae* lectin 2 (Armost1_934509); ArO_L3: *Armillaria ostoyae* lectin 3 (Armost1_1033526); SS_L1: *Sphaerobolus stellatus* lectin 1 (Sphst1_256542); SH_L1: *Suillus hirtellus* lectin 1 (Suihi1_915777); HR_L1: *Hydnum rufescens* lectin 1 (Hydru2_1299653); DS_L1: *Dichomitus squalens* lectin 1 (Dicsq1_137073); AO_L1: *Arthrobotrys oligospora* lectin 1 (Artol1_3664); LB_L1: *Laccaria bicolor* lectin 1 (Lbic_330799); LB_L2: *Laccaria bicolor* lectin 2 (Lbic_327918). Light grey, dark grey and black shading indicate 60, 80 and 95%

sequence identity. On top the secondary structure of CCL2 according to NMR (Schubert, M., Bleuler-Martinez, S., et al. 2012) and X-ray crystallography (this work) is indicated. Residues involved in carbohydrate binding are marked with green triangle, residues involved in the dimerization with yellow filled circles. The figure was generated using GeneDoc (Nicholas, K.B., Nicholas Jr., H.B., et al. 1997).

Figure 2. ***Amino acid substitutions in the N- and C-terminal loops of the β-trefoil fold abolish the nematotoxicity of the fungal lectin CCL2.*** (A) Soluble expression levels of CCL2 wildtype, R18A single, F133A single and R18A_F133A double variant proteins in *E. coli*. Soluble protein extracts of *E. coli* BL21(DE3) transformants expressing wildtype CCL2, CCL2(Y92A), CCL2(R18A), CCL2(F133A) and CCL2(R18A_F133A), respectively, were separated by SDS-PAGE and stained with Coomassie brilliant blue (R-250). Transformants containing pET24 without insert (Vector Control) and expressing CCL2(Y92A), previously shown to be defective in carbohydrate-binding, served as controls. Sizes of marker proteins are indicated. (B) Development of *Caenorhabditis elegans* L1 larvae fed with *E. coli* BL21(DE3) transformants expressing wildtype CCL2, CCL2(R18A), CCL2(F133A) and CCL2(R18A_F133A). CCL2-expressing transformants were mixed in different ratios with bacteria containing pET24 vector. The percentage of CCL2-expressing bacteria in these mixtures is indicated on the X-axis. Pure cultures of pET24-containing transformants (Vector Control) and transformants expressing CCL2(Y92A) served as controls. Error bars indicate the standard deviation from the average of three biological replicates.

Figure 3. ***Amino acid substitutions in the N- and C-terminal loops of CCL2 do not affect carbohydrate-binding.*** In the presented contour plots of ^{1}H-^{15}N NMR correlations every

amino acid residue contributes one signal of its backbone amide (except proline), any change in the environment, e.g. ligand binding, typically leads to changes of the ^{1}H and ^{15}N resonance frequencies resulting in a change of position in this 2D plot. Binding of CCL2 wildtype, R18A single, R18A_F133A double variant and Y92A proteins to GlcNAcβ1,4(Fucα1,3)GlcNAcβ1O$(CH_2)_5$COOH was studied using NMR, as shown in panels A, B, C and D, respectively. ^{15}N-^{1}H HSQC spectra of ^{15}N-labeled proteins in the absence (blue) and presence of GlcNAcβ1,4(Fucα1,3)GlcNAcβ1O$(CH_2)_5$COOH (red) are shown on the left. Most prominent chemical shift changes are labeled at the peak positions of the WT. The behavior of the signal of T111 as one representative, is shown on the right in the course of a titration. The simultaneous presence of the free and bound state of T111 indicates slow exchange kinetics compared to the NMR time scale. The slow k_{OFF}, which is only seen for tight binders, is unchanged in the R18A and the R18A_F133A double variant, indicating an intact binding site. No chemical shift changes were observed for Y92A, indicating that the protein is not able to bind the ligand anymore.

Figure 4. ***CCL2 forms dimers in crystals.*** (A) Representation of the crystallographic CCL2 dimer complexed with sLe^X (4USO). The molecules are colored by chain and sLe^X is depicted as balls and sticks. (B) Wall-eyed stereo representation of the CCL2 dimeric interface as revealed by X-ray crystallography. Residues R18 and F133 are part of the CCL2 dimerization interface. Carbon atoms and labels are colored according to their chain (A in forest and B in marine). Hydrogen bonds are represented as dotted lines.

Figure 5. ***CCL2 forms dimers in solution.*** (A) NMR relaxation data support a protein dimer in solution. Shown is a correlation of the rotational correlation time τ_c and the molecular

weight of 34 proteins. Values of τ_c scaled to 310 K were taken from Dayie et al. (Dayie, K.T., Wagner, G., et al. 1996), molecular weights were calculated with the protein sequence. A linear regression resulted in τ_c = 0.887*MW–0.182 (R=0.89). The rotational correlation time of CCL2 estimated from the average T_1/T_2 ratio (Fig. S2) and the corresponding approximate molecular weight are shown in red. (B) The side chain of R18 is protected from proton-deuterium exchange. Left panel: ^{15}N-^{1}H HSQC spectrum of WT CCL2 in a solution containing 95% H_2O recorded at 900 MHz. The Nε/Hε signals of the arginines are aliased and appear with a different sign (maroon). Right panel: Comparable ^{15}N-^{1}H HSQC spectrum of wildtype CCL2 after 50 min in 100% D_2O. Only amides that do not exchange their proton to deuterium within that time show up – they are protected from exchange typically due to their involvement in a hydrogen bond. The Hε of R18 is protected and labeled with R18sc. Folded signals a colored in magenta.

Figure 6. ***Amino acid substitutions in the CCL2 dimer interface impair dimerization.*** (A) Affinity-purified wildtype CCL2, CCL2(R18A), CCL2(F133A) and CCL2(R18A_F133A) proteins were dissolved in 2x Laemmli sample buffer and run out on a denaturing SDS-PAGE without (RT) and with preheating (95°C) of the samples. Sizes of marker proteins (M) are indicated in kDa. (B) Diffusion Ordered nuclear magnetic resonance SpectroscopY (DOSY) measured at 298 K. Shown is a correlation between experimental translational diffusion coefficients D of CCL2 variants and the molecular weight (MW). A linear fit of five protein standards is indicated. Solutions of 0.31 mM aprotinin (6.5 kDa), 0.3 mM lysozyme (14.3 kDa), 0.3 mM carbonic anhydrase (29 kDa), 0.23 mM ovalbumin (44 kDa) and 0.34 mM bovine serum albumin (BSA, 68 kDa) were used as standards. The measured logD of the CCL2 variants are indicated in red together with their apparent logMW. The molecular

weights of monomer and dimer of CCL2 are indicated as dotted lines. (C) Size exclusion chromatograms of wildtype CCL2, CCL2(R18A), CCL2(F133A) and CCL2(R18A_F133A) proteins on an ENrich SEC 70 10 x 300 mm column equilibrated with 20 mM KH_2PO_4/K_2HPO_4 pH 5.7 and 150 mM NaCl. Elution volumes and calculated MW based on a calibration curve with standard proteins (see Fig. S2A) is indicated. (D) Size distribution determined by dynamic light scattering. Hydrodynamic radius (Rh) measured for wildtype CCL2, CCL2(R18A), CCL2(F133A) and CCL2(R18A_F133A) proteins in 50 mM KH_2PO_4/K_2HPO_4 pH 5.7 and 150 mM NaCl. Measurements at several concentrations (0.5, 2.1 and 12 mg ml^{-1}) are shown for CCL2(R18A_F133A). Each measurement was done at least three times on a Zetasizer nano ZS using a small size quartz cuvette but only one is represented here for clarity.

Figure 7. ***Coordination of sLeX by CCL2.*** Representation of the maximum-likelihood weighed 2mFo-DFc electron density around the sLe^X complexed to CCL2 contoured at 0.9 Å (0.205 eA^3). Hydrogen bonds are represented as dotted lines and waters as magenta spheres.

Figure 8. ***Orientation of the binding sites in the dimer and three-dimensional arrangements of single-domain β-trefoil lectin dimers of fungi.*** (A) Locations of the dimerization interface and the carbohydrate-binding sites in CCL2 as compared to other β-trefoil fold lectins from fungi. CLN, *Clitocybe nebularis* lectin; BEL, *Boletus edulis* β-trefoil lectin; SSA, *Sclerotinia sclerotiorum* agglutinin; RSA, *Rhizoctonia solani* agglutinin. See text for references. As an insert the monomers with the pseudo C_3 symmetry are shown schematically indicating which subdomains (α, β, γ) are mainly involved at the dimerization interface. This scheme is very simplified, any tilt and twist between the two molecules is ignored. The bottom of the β-

trefoil fold (root in tree model) is indicated in grey, ligands in green and the dimer interface in yellow. (B) Schematic representation with the lectin name and the corresponding PDB accession code in brackets. Tilt and torsion angles are derived from visual inspection and x- , y- and z-rotations in MOLMOL (Koradi, R., Billeter, M., et al. 1996). Distances between different ligand binding sites were measured between the O1 atoms of the monosaccharide that is in closest contact with the protein.

Table 1. Data Collection and Refinement Statistics for CCL2 sLe^X complex and apoprotein structures.

	sLe^X complex				**Apo**	
Data						
Beamline (wavelength, Å)	ID23-1/0.9732				BM30A/0.9205	
Spacegroup	$P6_122$				$P4_12_12$	
Unit cell dimensions, a, b,c Å	121.7, 121.7 144.9				59.92, 59.92, 202.9	
Resolution (outer shell), Å	29.23-1.95 (2.00-1.95)				44.85-2.25 (2.33-2.25)	
Measured/ Unique reflections	292542/46087				102330/18296	
Average multiplicity	6.3 (6.6)				5.6 (5.7)	
R_{merge}	0.047 (0.424)				0.093 (0.543)	
R_{pim}	0.029 (0.271)				0.061 (0.359)	
Completeness (%)	99.2 (100)				99.5 (99.8)	
Mean I / σI	18 (4.4)				15.7 (3.7)	
CC1/2	0.999 (0.892)				0.998 (0.900)	
Wilson B (truncate)	34.1				29.5	
Refinement						
R_{cryst}/ R_{free}	17.5/ 20.4				17.2/ 21.00	
nb reflections/free reflections	43754/2327				17392/892	
R_{msd} bonds, Å	0.015				0.015	
R_{msd} angles,°	1.61				1.69	
Rmsd chiral, $Å^3$	0.094				0.096	
Atoms (chain)	A	B	C	D	A	B
Protein	1052	1022	962	967	1033	1033
Bfac $Å^2$	27.4	49.2	58.9	66.6	26.1	34.4
Water molecules	152	39	28	10	123	72
Bfac, $Å^2$	39.8	46.1	45.5	52.5	34.8	36.0
Ligand	56					
Bfac, $Å^2$	43.1					
Ramachandran (Molprobity) Allowed	100 %				100 %	
Favoured	97 %				96.6 %	
Outliers :	0 %				0 %	
PDBcode	4USO				4USP	

Table 2. Contacts between CCL2 and sLe^X

Sugar atom	Protein atom	Distances (Å)
Fucose site		
O2	Wat2149- GalO6	3.11-2.59
	Wat2149- GlcNAcO3	3.11-2.87
O3	Gly108 N	3.05
	Wat2117 - Asp107 OD2	2.60 - 2.74
O4	Gly108 N	2.76
	Val93 O	2.60
O5	Val93 N	3.08
Van der Waals	Trp94	
GlcNAc site		
O1	Wat2099-Asn91 ND2	2.62-3.02
	Wat2099-Asn90 OD1	2.62-2.84
N2	Asn91 O	2.67
O3	Wat2149-FucO2	2.67-3.11
Van der Waals	Tyr57, Tyr92	
Gal site		
O2	Wat2091-Trp78 O	2.72-2.87
	Wat2091- Trp78 N	2.72-3.12
O4	Wat2148-Wat2120-Thr111 OG1	3.16-3.07-2.89
O6	Wat2148-Wat2117-Asp107 OD2	3.16-3.14-2.74
Van der Waals	Trp78	
Sia site		
O4	Trp78 O	3.18
	Gly79 O	3.20
	Ala80 N	3.12
Van der Waals	Trp78	

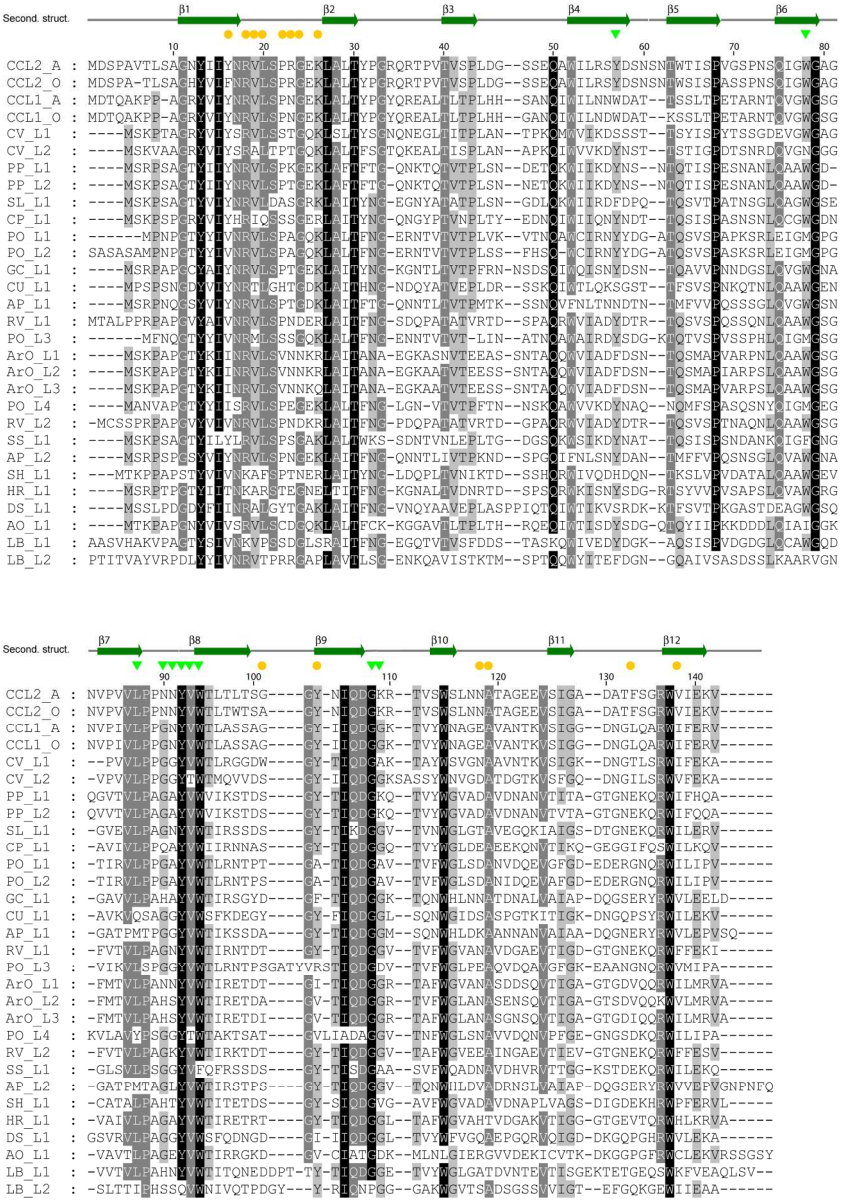

Figure 1. Multiple sequence alignment of CCL2 homologs encoded in different fungal genomes.

154x222mm (300 x 300 DPI)

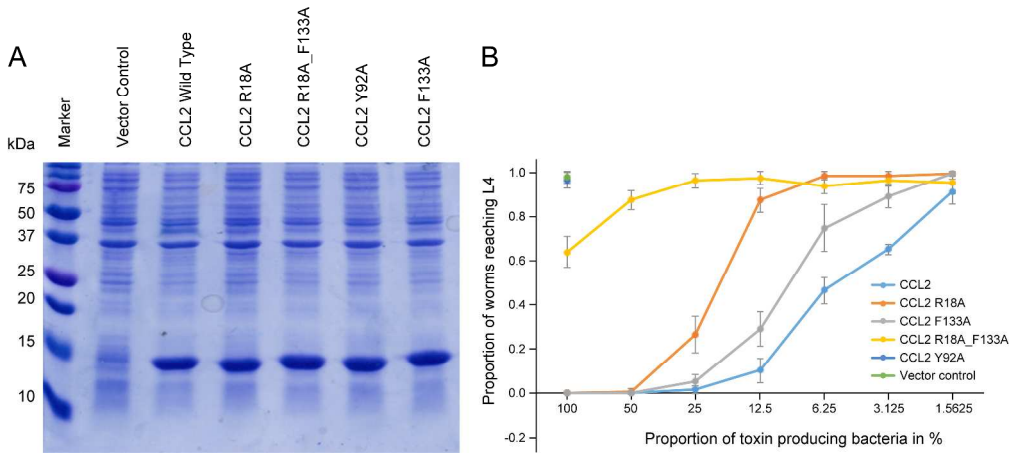

Figure 2. Amino acid substitutions in the N- and C-terminal loops of the β-trefoil fold abolish the nematotoxicity of the fungal lectin CCL2.

401x178mm (300 x 300 DPI)

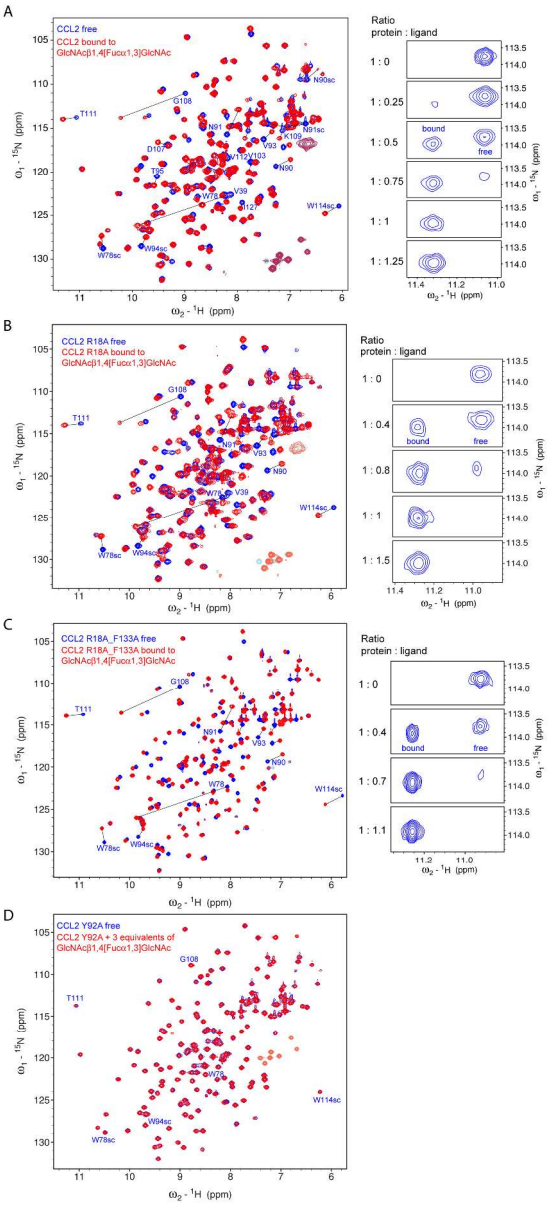

Figure 3. Amino acid substitutions in the N- and C-terminal loops of CCL2 do not affect carbohydrate-binding.

168x370mm (300 x 300 DPI)

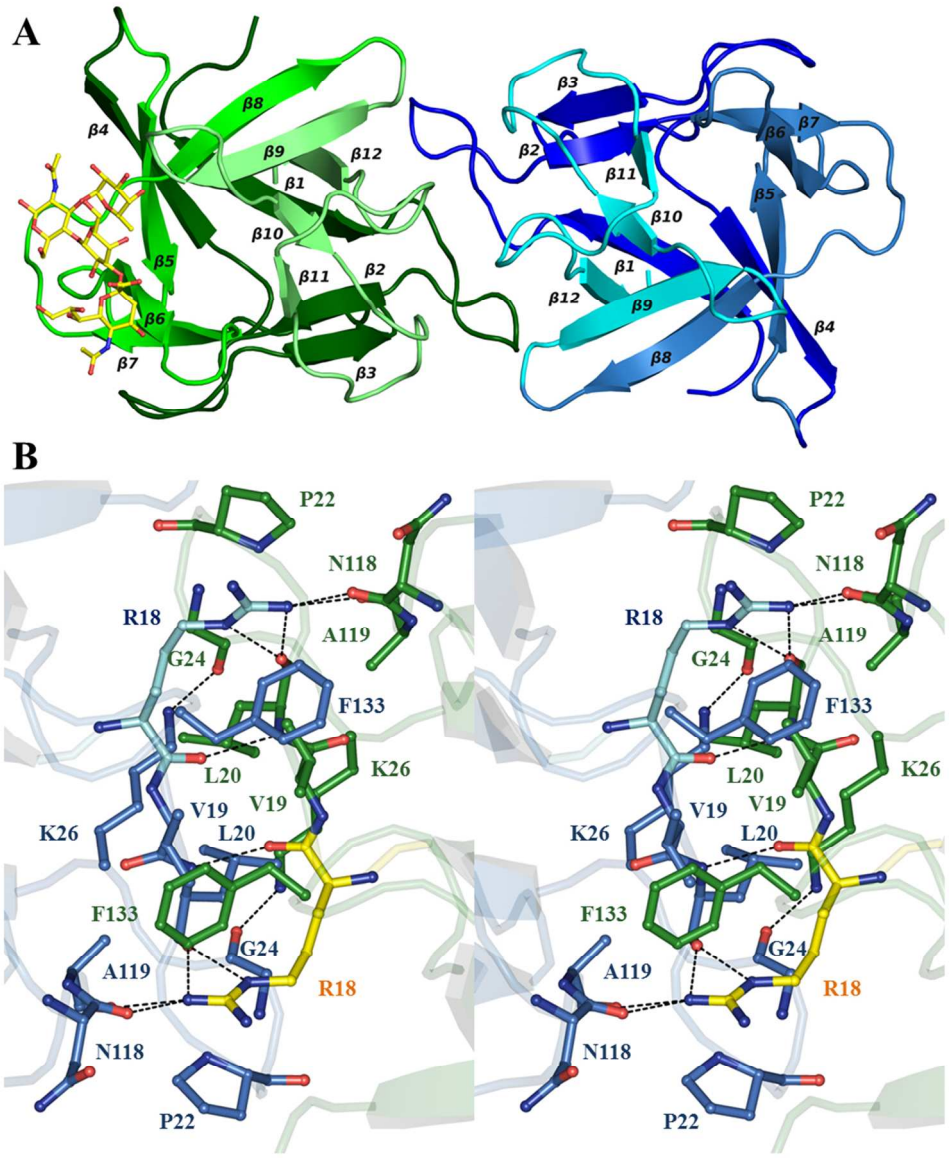

Figure 4. CCL2 forms dimers in crystals.

269x339mm (96 x 96 DPI)

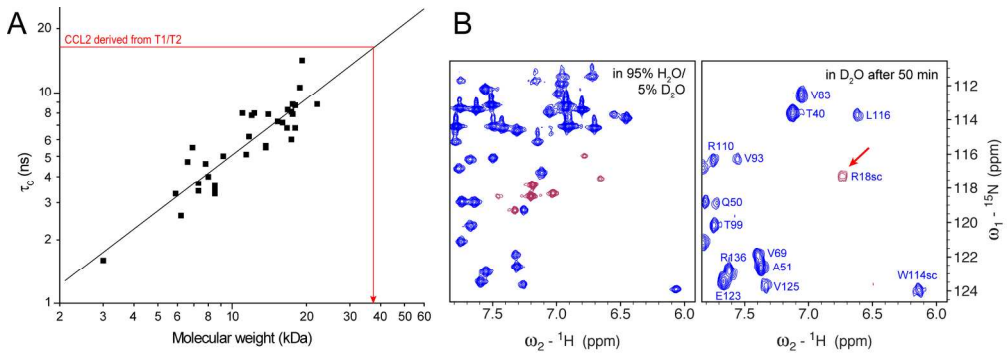

Figure 5. CCL2 forms dimers in solution.

188x65mm (300 x 300 DPI)

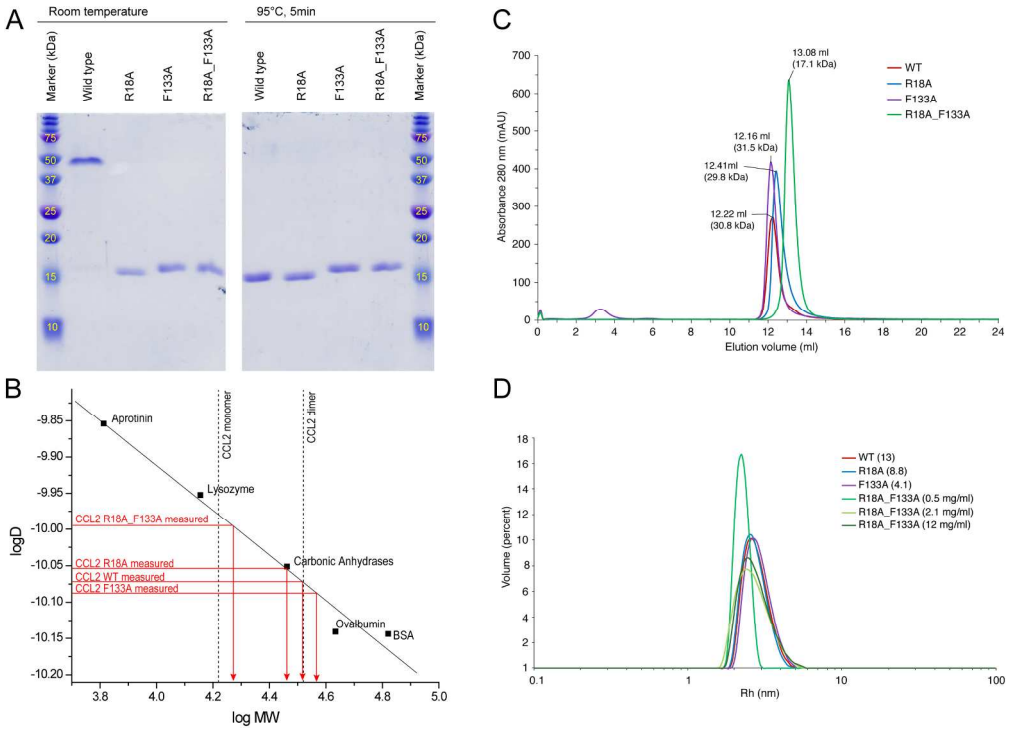

Figure 6. Amino acid substitutions in the CCL2 dimer interface impair dimerization.

207x151mm (300 x 300 DPI)

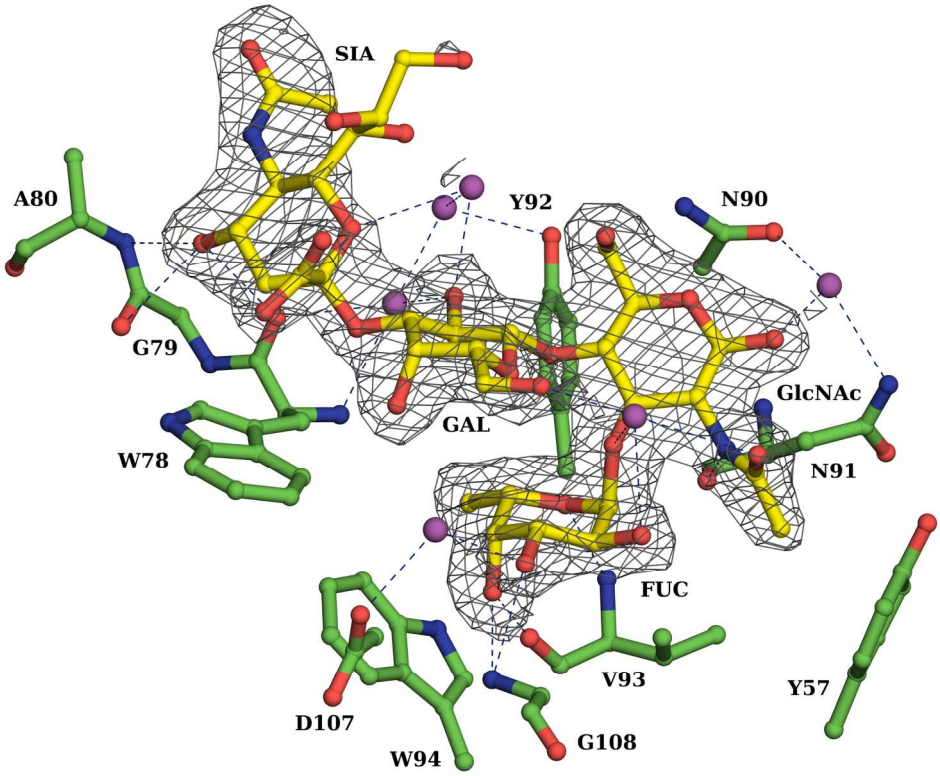

Figure 7. Coordination of sLeX by CCL2.

186x152mm (300 x 300 DPI)

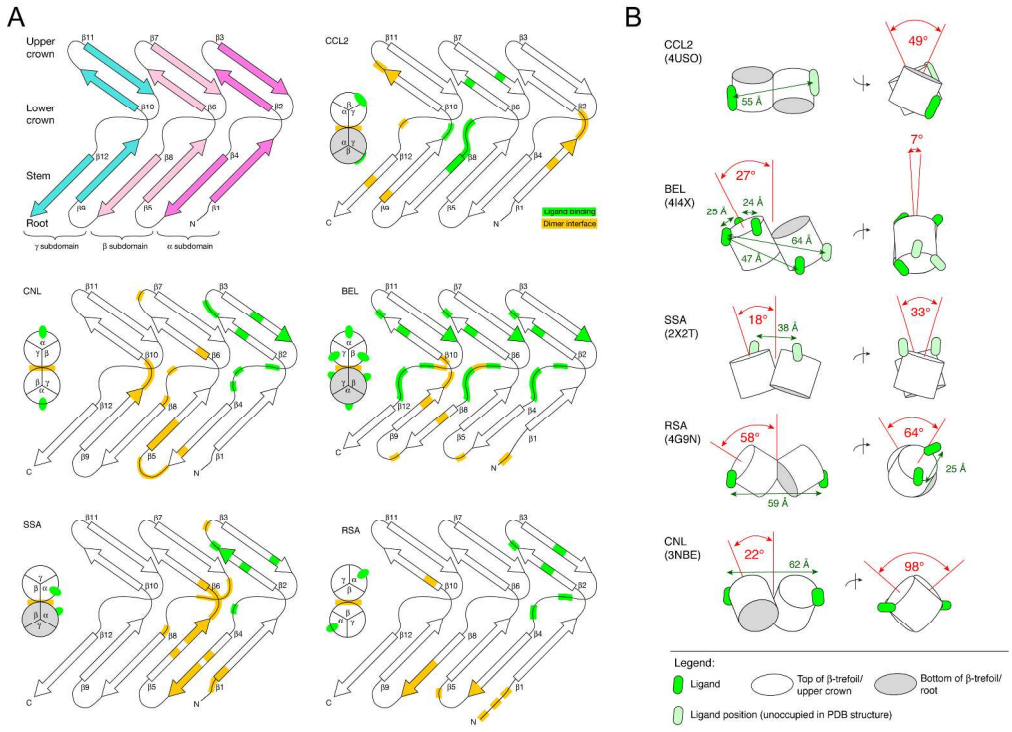

Figure 8. Orientation of the binding sites in the dimer and three-dimensional arrangements of single-domain β-trefoil lectin dimers of fungi.

273x197mm (300 x 300 DPI)

Supplementary online material

Dimerization of the fungal defense lectin CCL2 is essential for its toxicity against invertebrates

Silvia Bleuler-Martinez, Katrin Stutz, Ramon Sieber, Mayeul Collot, Jean-Maurice Mallet, Michael Hengartner, Mario Schubert, Annabelle Varrot and Markus Künzler

Supplementary Tables

Supplementary Figures

Table S1. Torsion angles observed in sLe^x bound to CCL2. Phi (Φ) = O5-C1-O'x-C'x and Psi (Ψ) = C1-O'x-C'x-C'(x+1) for a (1-> x) linkage and Φ = O6-C2-O'x-C'x and Ψ = C2-O'x-C'x-C'(x+1) for a (2-> x) linkage

Sugar	Φ	Ψ
Fuc	-66.54	143.48
Gal	-73.83	246.53
Sia	48.03	100

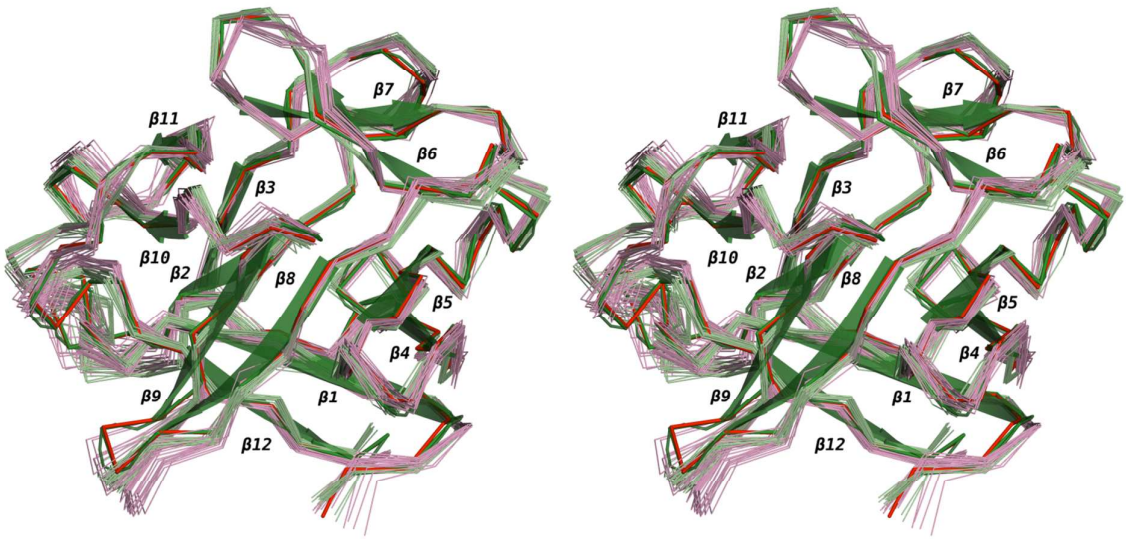

Figure S1. ***Superposition of CCL2 overall fold as determined by X-ray and NMR.*** Ribbon representation of the X-ray structures of CCL2 apoprotein (4USP, green), CCL2 complexed with sLe^X (4USO, red) and of the NMR ensemble for CCL2 apoprotein (2LIE, limegreen), and for CCL2 complexed with GlcNAcβ1,4(Fucα1,3)GlcNAcβ1O$(CH_2)_5$COOH (2LIQ, pink). The cartoon representation has been added to visualize the secondary structure elements for 4USP coordinates.

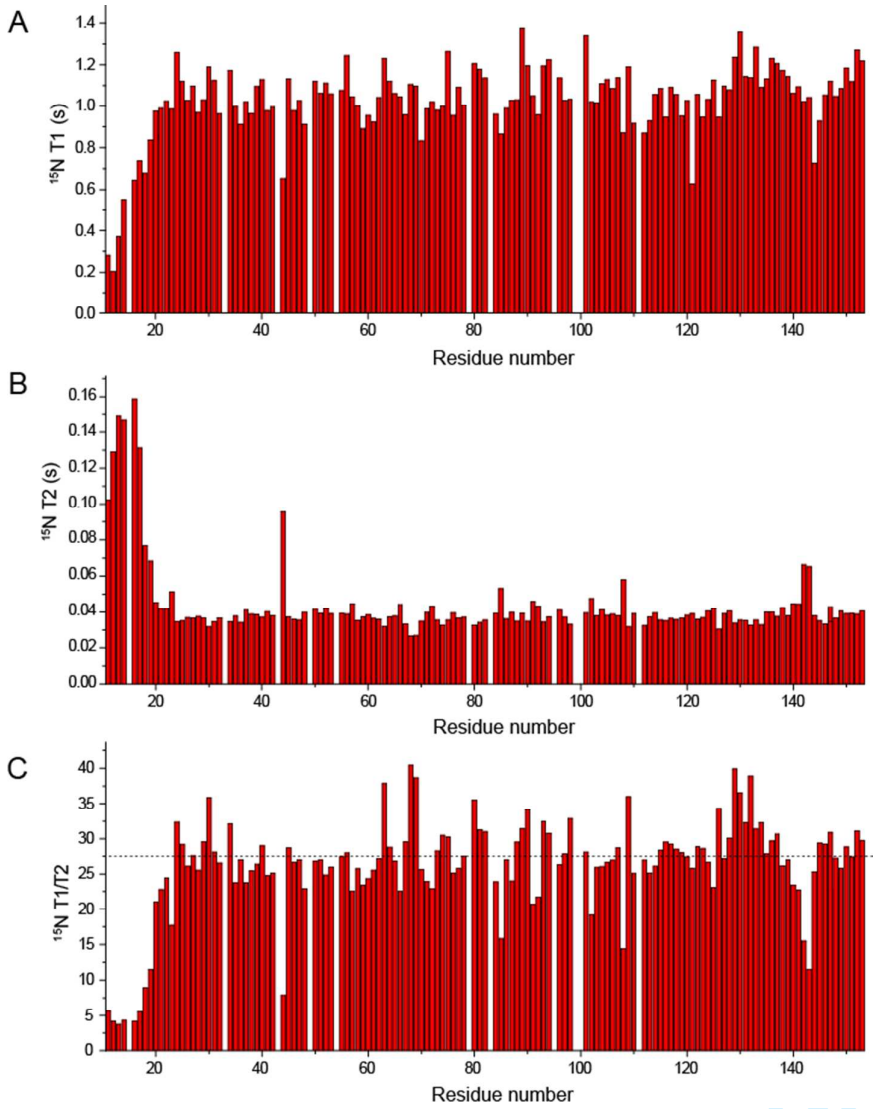

Figure S2. ***Backbone ^{15}N relaxation times of CCL2 (^{13}C/^{15}N-labeled, 1 mM) measured at a 600 MHz spectrometer at 310 K.*** (A) ^{15}N longitudinal relaxation times (T_1) plotted against residue number. (B) ^{15}N transverse relaxation times (T_2) plotted against residue number. (C) T_1 divided by T_2 with the average value of 27.4 for the structured region of residue 20-153 indicated as a dotted line.

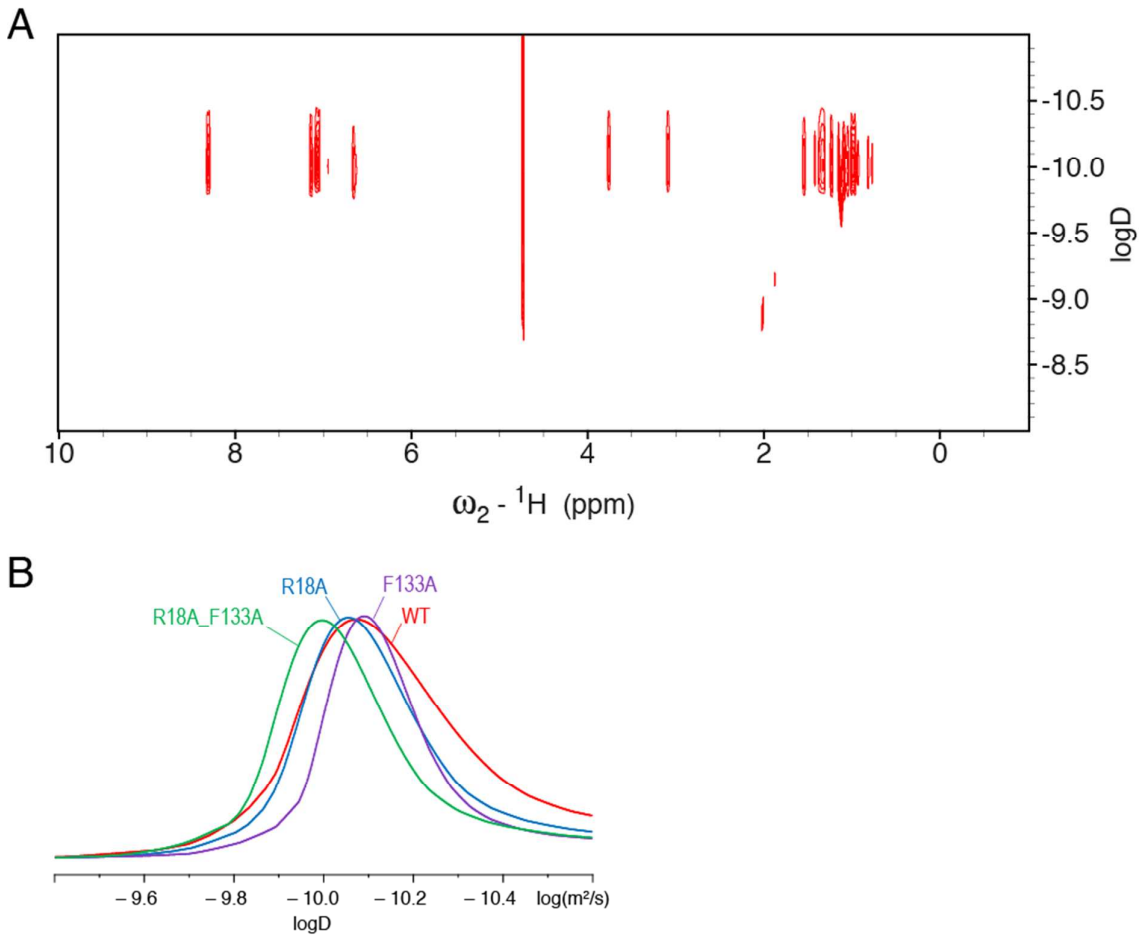

Figure S3. ***Diffusion-ordered SpectroscopY (DOSY).*** (A) 2D DOSY spectrum of CCL2 (R18A_F133A) in aqueous buffer at a concentration of 0.3 mM recorded at 298 K and 600 MHz using 128 scan, 32 increments, a gradient duration of δ=1.2ms and an echo delay of Δ=400ms. The standard Bruker stebpgp1s19 pulse sequence, with stimulated echo, bipolar gradient pulses, one spoil gradient and 3-9-19 water suppression, was run with a linear gradient (53.5 G cm^{-1}) stepped between 2 and 95%. (B) Projections of logD of CCL2 WT, CCL2(R18A), CCL2(F133A) and CCL2(R18A_F133A) recorded with comparable parameters and concentrations. The region of methyl groups (1.04–0.5 ppm) was used for the projection.

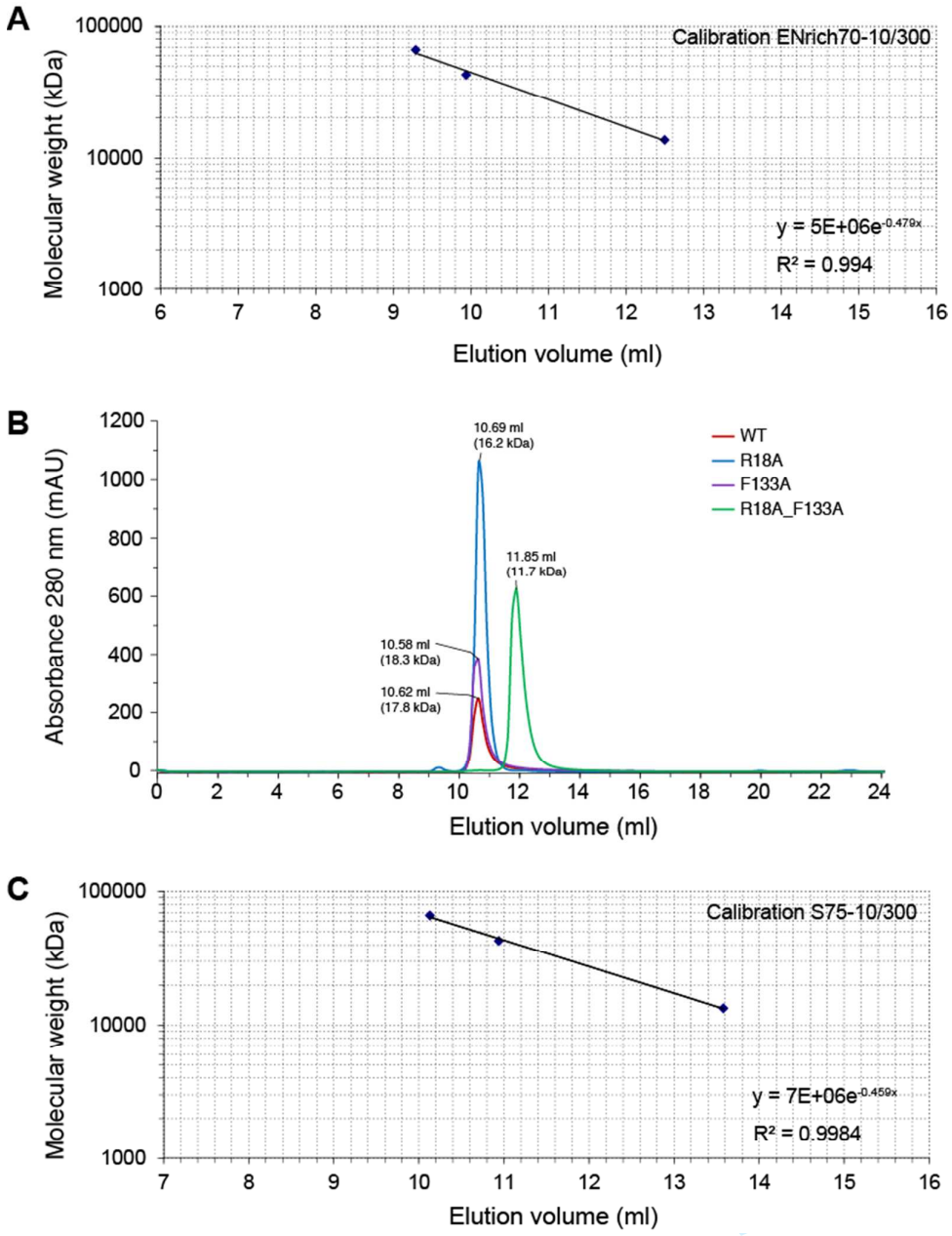

Figure S4. ***Size exclusion chromatography of CCL2 variants.*** (A) Calibration curve of the ENrich SEC 70 10 x 300 mm column was done with the following standards: bovine serum albumin 66.5 kDa (9.29 ml), ovalbumin 43 kDa (9.94 ml) and RNase A 13.7 kDa (12.5 ml). (B) SEC Chromatograms for CCL2 wild-type, CCL2(R18A), CCL2(F133A) and CCL2(R18A_F133A) proteins on a Superdex 75 10 x 300 mm column (C) Calibration curve of the Superdex 75 10 x 300 column with the same standards as above. The elution volumes were 10.13 ml (BSA), 10.94 (Ova) and 13.58 ml (RNase A). All experiments were done using 50 mM KH_2PO_4/K_2HPO_4 pH 5.7 and 150 mM NaCl as buffer.

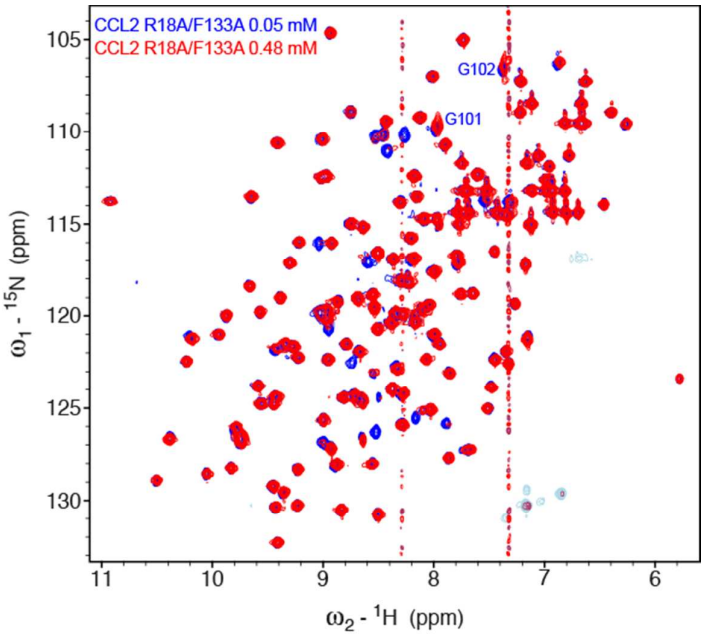

Figure S5. ***The R18A/F133A double mutant shows concentration dependent chemical shift deviations.*** The ^{15}N-^{1}H HSQC spectrum at a concentration of 50 μM (blue) shows the expected number of sharp signals whereas an identical spectrum recorded at 480 μM (red) indicates line broadening (signals of G101 and G102 at the dimer interface) and the disappearance of some signals from the dimer interface. The latter signals (blue contours without red contours in front) showed large chemical shift changes in the R18A F133A double mutant compared to WT indicating their involvement in the dimer interface.

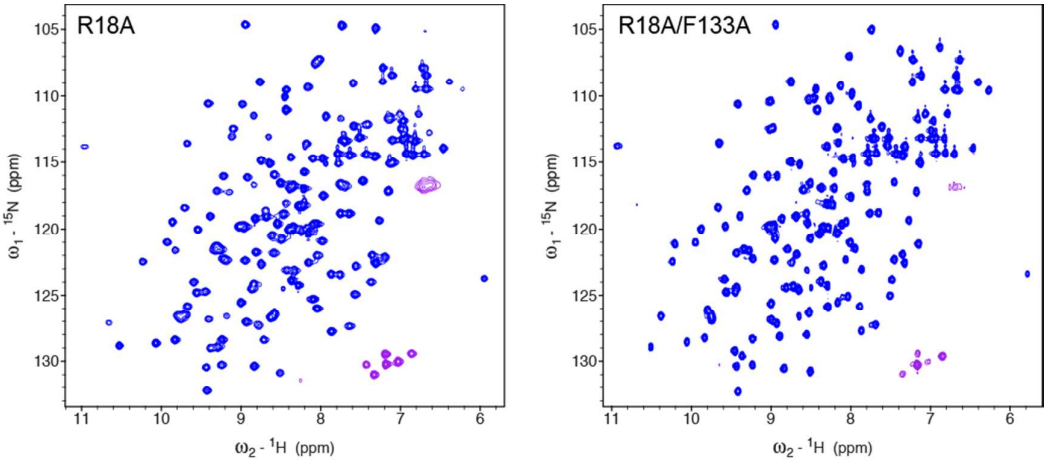

Figure S6. ***All point mutant proteins are folded.*** ^{15}N-^{1}H HSQC spectra of ^{15}N-labeled proteins show a nicely dispersed signal distribution which is typical for a β-sheet protein and comparable to WT.

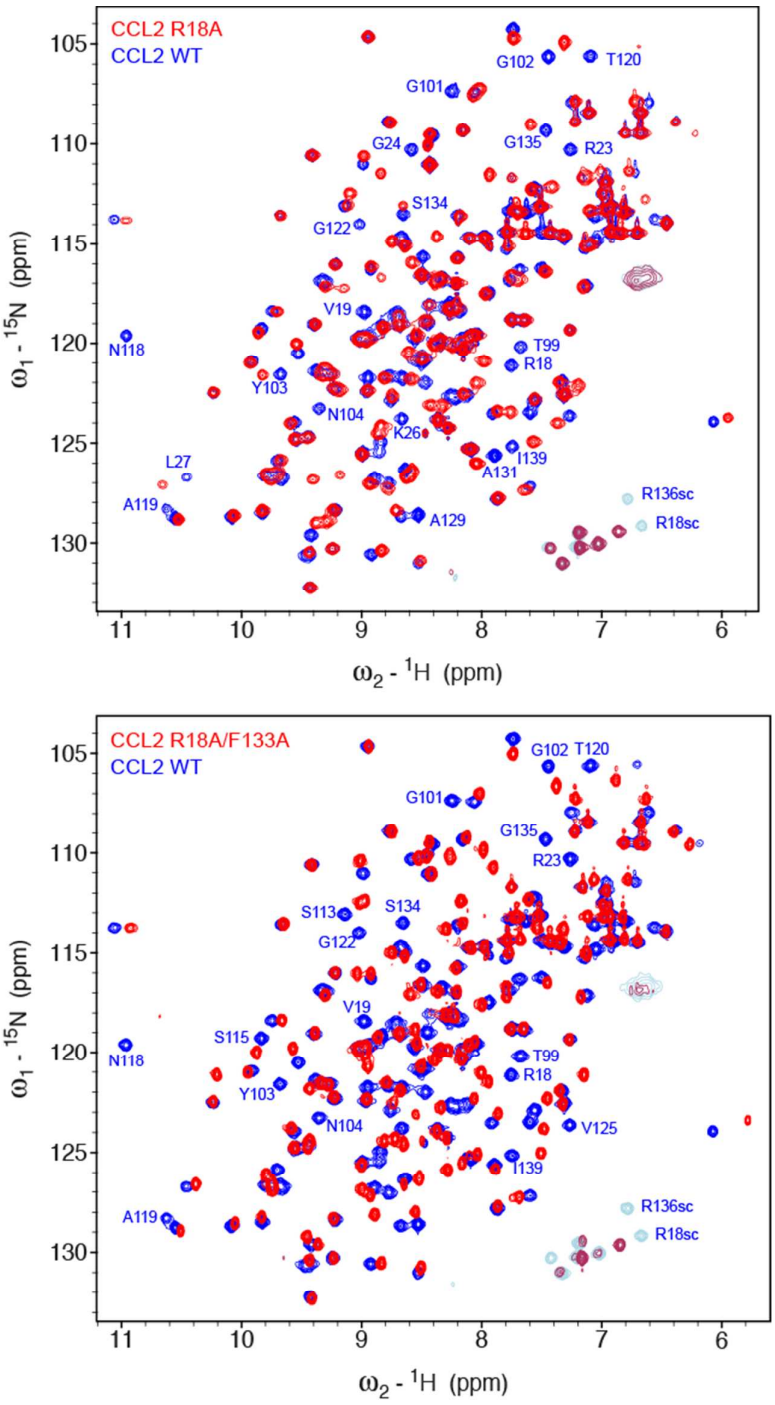

Figure S7. ***Protein regions affected by point mutations.*** ^{15}N-^{1}H HSQC spectra of ^{15}N-labeled mutant proteins (red) in comparison to WT (blue). Most prominent residues influenced by the point mutations are labeled at the peak positions of the WT. Signals from side chains are indicated by 'sc'.

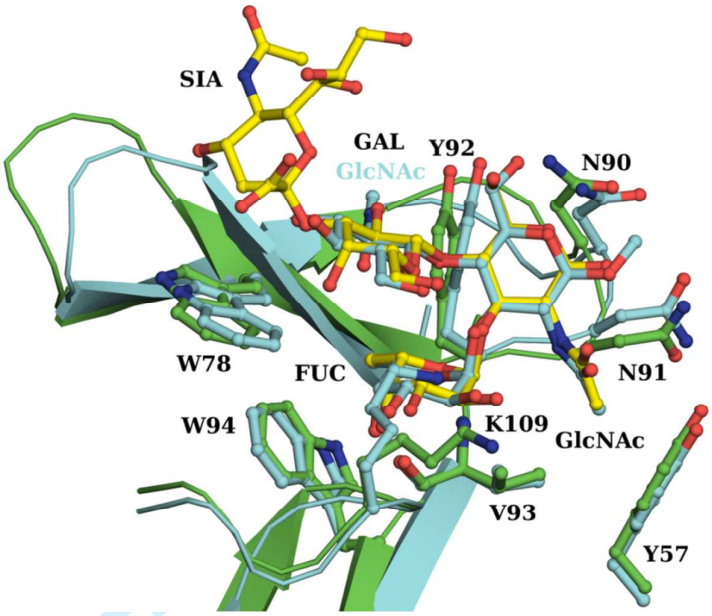

Figure S8. ***Superposition of the carbohydrate-binding site of CCL2 for sLe^x and GlcNAcβ1,4(Fucα1,3)GlcNAcβ1O(CH$_2$)$_5$COOH.*** The sLe^X complex (4USO) is represented in green/yellow and the GlcNAcβ1,4(Fucα1,3)GlcNAcβ1O(CH$_2$)$_5$COOH complex (2LIQ) in cyan.