



Computational and experimental prediction of human C-type lectin receptor druggability

Jonas Aretz^{1,2}, Eike-Christian Wamhoff^{1,2}, Jonas Hanske^{1,2}, Dario Heymann¹ and Christoph Rademacher^{1,2*}

¹ Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Potsdam, Germany

² Department of Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Berlin, Germany

Edited by:

Elizabeth Yuriev, Monash University, Australia

Reviewed by:

Anthony G. Coyne, University of Cambridge, UK

Stephen James Headey, Monash University, Australia

*Correspondence:

Christoph Rademacher, Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, Potsdam 14424, Germany
e-mail: christoph.rademacher@mpikg.mpg.de

Mammalian C-type lectin receptors (CTLRs) are involved in many aspects of immune cell regulation such as pathogen recognition, clearance of apoptotic bodies, and lymphocyte homing. Despite a great interest in modulating CTLR recognition of carbohydrates, the number of specific molecular probes is limited. To this end, we predicted the druggability of a panel of 22 CTLRs using DoGSiteScorer. The computed druggability scores of most structures were low, characterizing this family as either challenging or even undruggable. To further explore these findings, we employed a fluorine-based nuclear magnetic resonance screening of fragment mixtures against DC-SIGN, a receptor of pharmacological interest. To our surprise, we found many fragment hits associated with the carbohydrate recognition site (hit rate = 13.5%). A surface plasmon resonance-based follow-up assay confirmed 18 of these fragments (47%) and equilibrium dissociation constants were determined. Encouraged by these findings we expanded our experimental druggability prediction to Langerin and MCL and found medium to high hit rates as well, being 15.7 and 10.0%, respectively. Our results highlight limitations of current *in silico* approaches to druggability assessment, in particular, with regard to carbohydrate-binding proteins. In sum, our data indicate that small molecule ligands for a larger panel of CTLRs can be developed.

Keywords: C-type lectin receptors, druggability, inhibitor, DC-SIGN, langerin, MCL, fragment screening, NMR screening

INTRODUCTION

Glycans are present in a large diversity on cell surfaces and are essential in many aspects of life such as embryonic development, cell–cell communication, and regulation of the immune system (1). In particular, our understanding of the role of glycans in immunobiology has grown significantly during the last decades. Three major families of secreted or membrane-bound lectins recognize carbohydrates. Complementary to other receptors of the innate and adaptive immune system, Galectins, Siglecs, and C-type lectins shape the response to incoming signals (2, 3). Among many other processes, they are involved in pathogen recognition and killing, antigen processing, and tumor progression (2, 4, 5).

Mammalian C-type lectin receptors (CTLRs) represent a large family of lectins, which is subdivided into 17 groups based on their phylogenetic relationships and domain structure (6). CTLRs are present in a variety of tissues and the glycan specificity of receptors present on cells of the innate immune system has been studied extensively. For example, they function as homing receptors on leukocytes as well as pattern recognition receptors (2, 3, 7). A particularly well-studied pattern recognition receptor is the dendritic cell-specific intercellular adhesion molecules-3-grabbing non-integrin (DC-SIGN) (8, 9). This CTLR is expressed on dendritic cells and macrophages and is involved in the recognition of a large array of pathogens such as *Mycobacterium tuberculosis*, *Leishmania*, HCV, Ebola, and HIV (3, 10–15). It was demonstrated that DC-SIGN promotes HIV *trans*-infection of T cells and has since

then drawn attention as a therapeutic target in anti-viral therapy (10, 16, 17).

Aside from interference with pathogen recognition, leukocyte homing has been a target for small molecule inhibition of CTLR function. To this end, Selectins, a group of three CTLRs, have been in the focus as anti-inflammatory drug targets since the mid-90s (18). Only recently, the glycomimetic GMI-1070 has entered clinical trials for the treatment of sickle cell anemia (19). Likewise, agonistic CTLR ligands hold promise to serve as adjuvants for immune stimulation (20). However, despite increasing interest in CTLRs as pharmacological targets, there is only a limited set of small molecule agonists or antagonists available (17). Partially, this can be attributed to the limited success of previous attempts to find lead structures from classical drug discovery campaigns.

All CTLRs share a C-type lectin domain (CTLD) that has a conserved fold with a characteristic double-loop stabilized by two disulfide bridges (7, 21). This domain is often referred to as carbohydrate recognition domain (CRD) for those CTLRs involved in glycan binding. Additional domains are frequently present and in particular, heptad-repeats and collagen-like neck domains promote oligomerization, resulting in high avidity glycan binding. In transmembrane CTLRs, CRD, and neck domain are referred to as extracellular domain (ECD). Canonical carbohydrate recognition is mediated by a calcium ion and although there are four Ca²⁺ binding sites, only the second site (Ca²⁺-2) is described to be involved in coordinating glycans (21). While Ca²⁺-4 has

not been associated with carbohydrate binding, positive cooperative effects are observed between the other sites (22, 23). Not all potential Ca^{2+} sites are occupied in every CTLD, which reflects the fine-tuned physiological role of this interaction. For endocytic CTLRs the pH sensitivity of the heptad-repeat neck formation and Ca^{2+} coordination as well as active Ca^{2+} export from the endosome are major contributors to endosomal ligand release (23, 24). Some CTLRs bind carbohydrates in a Ca^{2+} -independent, non-canonical binding site with Dectin-1 being the prime example (25). All CRDs share a carbohydrate recognition site that is largely flat and hydrophilic. This is a consequence of glycans being highly hydrophilic themselves (17, 26). Hence, binders are also often hydrophilic and do not suffice the requirements for orally available drugs (27).

Whether a protein is a suitable candidate for drug development is of major concern during the drug discovery process. Considering the expenses involved in the development of a pharmacologically active small molecule, target selection has to be done carefully (28). The modulation of a suitable drug target with a rule of five compliant molecules should result in a therapeutic effect (29). The term druggability, however, refers to the ability of a protein to bind a drug-like ligand with high affinity and specificity (29–31). Furthermore, this interaction has to result in a modulation of the protein function. Importantly, a high druggability does not infer the protein being a good drug target. The latter definition includes a therapeutic effect induced by small molecule binding (32, 33). Methods to assess the druggability of a target protein have become good predictors prior to starting a drug discovery campaign, as low scores are indicators for a high failure rate during later stages of the project (30, 33).

The availability of structural information enables computational assessment of druggability. Limited resources are required and many computational tools have been developed to deduce druggability scores from crystallographic information (34, 35). In a two-step process, pockets on the protein surface are first identified and then scored (28, 32, 34, 36). Large sets of proteins can be analyzed and predictions have been found to correlate well with experimental data (31, 34, 37, 38). To the best of our knowledge, there are only two studies on the computational druggability assessments of glycan-binding proteins, both reporting low scores (39, 40).

Experimental assessment of target druggability can be pursued even in the absence of structural information. For this, screening of drug-like molecules in a high-throughput screening format can be performed. Previous reports on micromolar inhibitors of DC-SIGN resulting from a screening campaign highlight the success of this approach (41, 42). Alternatively, a diverse library of fragments of drug-like molecules is screened against the target. The molecular weight of these fragments ranges between 150 and 300 Da. Estimates propose that 1000 fragments can cover a similar chemical space as 10 trillion drug-sized molecules would (33). This in turn allows applying smaller libraries to test the druggability of a candidate protein (31, 33). The low complexity of fragments increases their likelihood of binding a receptor and consequently hit rates of 5–15% are regularly observed for druggable targets (31, 37, 43).

Small molecule fragments have low affinities with dissociation constants in the upper micro- to lower millimolar range. Hence, sensitive biophysical techniques are necessary to monitor this interaction and nuclear magnetic resonance (NMR) spectroscopy has established itself as one of the major techniques used for fragment screening (31, 33, 37–39, 44, 45). In particular, hit rates from NMR-based screenings have proven to be reliable measures of druggability (31, 37, 44). In ligand-observed NMR, mixtures of fragments are screened against a target and changes in NMR observables such as chemical shift, line width, and signal intensity upon binding allow hit identification. Notably, deconvolution of the fragment mixtures is not necessary. The use of fluorine atoms in drug-like fragments has proven to be instrumental (38, 46). As fluorine is rare in biological samples, ^{19}F NMR spectra of fragment cocktails are not perturbed by background resonances. Moreover, the fluorine spin is highly susceptible to changes in its chemical environment and allows sensitive identification of hits.

To predict the druggability of human CTLRs, we compiled a set of 22 crystal structures and analyzed it by applying computational methods. We then chose DC-SIGN and conducted experimental fragment screening to compare these findings. Low druggability scores derived *in silico* did not match the moderate to high fragment hit rates during experimental evaluation. Hence, we expanded our screening by two additional CTLRs, namely Langerin and MCL and discovered similarly high experimental druggability estimates. Taken together, our results highlight the limitations of *in silico* druggability prediction for CTLRs while our fragment screening present promising grounds for inhibitor design against this family.

MATERIALS AND METHODS

STRUCTURE-BASED MULTIPLE SEQUENCE ALIGNMENT AND CONSENSUS STRUCTURE

The scope of structural data on human CTLRs was assessed using the protein family (Pfam) database (accession code: PF00059) (47). Natural killer (NK) cell lectin-like receptors were treated as a closely related, yet physiologically distinct subfamily according to the Pfam annotation and were not included in the analysis. Furthermore, CTLRs crystallized as a domain swap dimer, namely blood dendritic cell antigen 2 (BDCA-2) and mannose receptor (MR), were omitted (48, 49). Murine Dectin-1 was included in the selection as it has an unusual Ca^{2+} -independent carbohydrate-binding mode and no structural information of the human ortholog is available (25). All structures considered for analysis are listed (Table 1). If available, a structure in complex with a carbohydrate ligand was selected. Prior to the calculations, all structures were trimmed down to the respective CRD domain as inferred from the Pfam domain definition. A structure-based multiple sequence alignment was performed in molecular operating environment (MOE) (50). Pairwise root mean square deviation (RMSD) values were determined for all pairs of C_α atoms unless a gap was found in one of the compared sequences. Next, a phylogenetic analysis based on the pairwise sequence similarities was conducted in R (51, 52). Hierarchical clustering was performed based on the Manhattan metric and via the complete linkage criterion. To complement the phylogenetic analysis, MOE was used

to predict a consensus structure of all CRDs. During model construction, up to 20 gaps and RMSD values of C_{α} up to 10 Å were allowed for a single position in the multiple sequence alignment.

BINDING SITE PREDICTION AND *IN SILICO* DRUGGABILITY ASSESSMENT

Initially, CTLR structures were superposed in MOE. For superposition and the subsequent druggability assessment, physiologically relevant oligomerization states were assumed (Table 1). The EGF domains of Selectin structures were removed. The resulting files served as input data for binding site prediction with DoGSite (72). The predicted binding sites were mapped on the structure and classified into four categories following the reported nomenclature of secondary structure elements and Ca^{2+} binding sites (21): (i) Ca^{2+} -2-binding sites, (ii) Ca^{2+} -associated binding sites in long loop, (iii) Ca^{2+} -independent carbohydrate-binding sites, and (iv) other binding sites. A binding site was assigned to category (i) if the Ca^{2+} -2 ion was part of the predicted binding site. For category (ii), the criteria were less restrictive and all binding sites with residues within a 6 Å radius of either Ca^{2+} -1, 2, or 3 were included (Figure S1 in Supplementary Material). Binding sites in category (iii) are located in close proximity to the experimentally determined Ca^{2+} -independent carbohydrate-binding site. The druggability of all binding sites was scored with DoGSiteScorer (73). Finally, category (i), (ii), or (iii) binding sites that displayed the highest score for a receptor were selected and this selection served to determine a mean druggability score for the analyzed CTLRs.

CLONING

Codon optimized genes for DC-SIGN and human Langerin for expression in *E. coli* were purchased from Life Technologies (Carlsbad, CA, USA) and GenScript (Piscataway, NJ, USA), respectively. The DC-SIGN gene included a C-terminal TEV (tobacco etch virus) cleavage site and a Strep-tag II for affinity purification. The ECD and CRD, ranging from amino acids 62 to 404 and 250 to 404 (Figure S3 in Supplementary Material), respectively, were cloned into a pUC19 vector using primers including a T7 promoter and ribosomal binding site (RBS) upstream of the gene (Table 2). Human Langerin truncated ECD, ranging from amino acids 148 to 328, was cloned with a C-terminal TEV cleavage site and a Strep-tag II into a pET32a expression vector (EMD Millipore, Billerica, MA, USA). The MCL gene was obtained from the DNASU Plasmid Repository (HsCD00507041, Arizona State University, Phoenix, AZ, USA) and the ECD was cloned into a pUC19 vector already carrying a Strep-tag II, a T7 promoter and an RBS. For MCL ECD, amino acids ranging from 61 to 215 were used (65).

PROTEIN EXPRESSION AND PURIFICATION

All growth media or chemicals used for protein expression and purification were purchased from Carl Roth (Karlsruhe, Germany) if not stated otherwise. Proteins were expressed insoluble in *E. coli* BL21(DE3) (New England Biolabs, Ipswich, MA, USA) or KRX (Promega, Fitchburg, WI, USA). Precultures were grown in 50 mL Luria-Bertani (LB) medium supplemented with 100 mg L⁻¹ carbenicillin for DC-SIGN and MCL expression or 35 mg L⁻¹ kanamycin for Langerin expression at 37°C in 250 mL baffled shaking flasks at 220 rpm shaking frequency. The precultures of DC-SIGN and MCL were centrifuged (3,000 × g, 10 min,

Table 1 | List of analyzed CTLR structures.

Receptor	PDB code	Domain	Oligo-merization	Reference
ASGPR	1DV8	CRD	Monomer	Meier et al. (53)
CD23a	2H2T	CRD	Monomer	Wurzberg et al. (54)
Clec1b	2C6U	CRD	Monomer	Watson et al. (55)
Clec5a	2YHF	CRD	Monomer	Watson et al. (56)
Clec9a	3VPP	CRD	Monomer	Zhang et al. (57)
DC-SIGN	2XR6	CRD	Monomer	Unpublished
DC-SIGNR	1K9J	CRD	Monomer	Feinberg et al. (58)
mDectin-1	2CL8	CRD	Monomer	Brown et al. (25)
E-Selectin	1ESL	CRD-EGF	Monomer	Graves et al. (59)
EMBP	1H8U	CRD	Monomer	Swaminathan et al. (60)
Langerin	3P5F	CRD	Monomer	Feinberg et al. (61)
	3KQG	ECD	Trimer	Feinberg et al. (62)
L-Selectin	3CFW	CRD-EGF	Monomer	Unpublished
LOX-1	1YPO	ECD	Dimer	Park et al. (63)
MBP-C	1HUP	ECD	Trimer	Sheriff et al. (64)
MCL	3WHD	ECD	Monomer	Furukawa et al. (65)
Mincle	3WH3	ECD	Monomer	Furukawa et al. (65)
P-Selectin	1G1S	CRD-EGF	Monomer	Somers et al. (66)
Reg3a	1UV0	CRD	Monomer	Abergel et al. (67)
Reg1a	1QDD	CRD	Monomer	Gerbaud et al. (68)
SCARA4	2OX8	CRD	Monomer	Feinberg et al. (48)
SP-D	3IKN	ECD	Trimer	Shrive et al. (69)
Tetranectin	1HTN	CRD	Monomer	Nielsen et al. (70)
	1TN3	ECD	Trimer	Kastrup et al. (71)

Table 2 | Primer sequences used for cloning.

Primer	Sequence
DC-SIGN ECD F	GCCGCCTCTAGAGAGTAATACGACTCACTATAGGGACT AGAGAAAGAGGAGAAAAGTAGATGGC- CAAAGTCCGAGCAGCATT
DC-SIGN CRD F	GCCGCCTCTAGAGGAGTAATACGACTCACTATAGGGAC TAGAGAAAGAGGAGAAAAGTAGATGGCT- GAACGTCTGTGCATCCGTG
Langerin F	GGTGGTCATATGGCC- TCGACGCTGAATGCCAGATTCCGG
Langerin R	ACCACC-AAGCTTTTATTTTCAAAGCTGCGGATG
MCL F	GGCGGCGCCGGC-CATGCAAAGCTCAAATGCAT
MCL R	GCCGCCCTGCAG-GTTCAATGTTGTTCCAGGTA

4°C), the supernatant was discarded, and the sediment was resuspended in 500 mL LB medium supplemented with 50 mg L⁻¹ carbenicillin. The cells were afterwards cultivated at 37°C in 2.5 L baffled shaking flasks at 220 rpm shaking frequency. Protein expression was induced with 1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) at OD₆₀₀ = 0.4–0.6 for additional 4 h at 37°C. The preculture of Langerin trECD was diluted directly to OD₆₀₀ = 0.1 into 500 mL of LB medium supplemented with 35 mg

L⁻¹ kanamycin, 0.01% D-glucose, and 0.05% L-rhamnose for autoinduction of expression. Bacteria were harvested (4,000 × g, 20 min, 4°C), frozen, and resuspended in lysis buffer (50 mM Tris–HCl, pH 7.5, 10 mM magnesium chloride, 0.1% Triton X-100, 4 mg lysozyme [Sigma-Aldrich, St. Louis, MO, USA] and 500 U DNaseI (Applichem, Darmstadt, Germany) per gram of wet biomass] and incubated on ice for 4 h. Inclusion bodies were harvested by centrifugation (10,000 × g, 10 min, 4°C) and washed thrice with 20 mL washing buffer (50 mM Tris–HCl, pH 8.0, 4 M urea, 500 mM sodium chloride, 1 mM EDTA) to remove soluble proteins.

For DC-SIGN ECD and Langerin ECD purification, the washed inclusion bodies were resuspended and denatured in 40 mL denaturation buffer (6 M guanidine hydrochloride, 100 mM Tris–HCl, pH 8.0, 1 mM DTT) and incubated at 30°C for 1 h or at 4°C over night, following a centrifugation (42,000 × g, 1 h, 4°C). The denatured inclusion bodies were slowly diluted threefold with cold binding buffer (TBS, pH 7.8 with 25 mM calcium chloride), supplemented with 1 mM reduced glutathione (GSH, Applichem) and 0.1 mM oxidized glutathione (GSSG, Applichem), and afterwards dialyzed twice against 2 L of this buffer for 24 h at 4°C. After another 2 L dialysis against binding buffer, proteins were purified according to previously published protocols using a mannan agarose affinity chromatography (Sigma-Aldrich) (74).

The washed inclusion bodies of DC-SIGN CRD were resuspended and denatured in 10 mL denaturation buffer and incubated at 30°C for 1 h or at 4°C over night, following a centrifugation (42,000 × g, 1 h, 4°C). The solubilized inclusion bodies in the supernatant were refolded by rapid dilution into 50 mL of cold refolding buffer (100 mM Tris–HCl, pH 8.0, 1 M L-arginine, 150 mM sodium chloride, 120 mM sucrose) while stirring at 4°C. After 2 days, protein solution was dialyzed against 2 L of cold buffer W (100 mM Tris–HCl, pH 8.0, 150 mM sodium chloride, 1 mM EDTA) and aggregated protein was removed by centrifugation (42,000 × g, 1.5 h, 4°C). The protein was purified using a Streptactin affinity chromatography (IBA, Goettingen, Germany) according to the manufacturer's instructions.

MCL refolding and purification was performed according to Furukawa and coworkers introducing minor changes in the protocol. Briefly, purification was performed via Streptactin affinity chromatography after dialysis against 2 L of buffer W.

FRAGMENT LIBRARY

Fragments were selected from a pool of commercially available compounds from different manufacturers (Sigma-Aldrich, St. Louis, MO, USA; KeyOrganics, Camelford, UK; ACB Blocks, Toronto, ON, Canada; Santa Cruz Biotechnology, Santa Cruz, CA, USA; Vistas-MLab, Moscow, Russia; LifeChemicals, Kyiv, Ukraine; Alfa Aesar, Ward Hill, MA, USA; TCI, Tokyo, Japan; Apollo Scientific, Stockport, UK) using chemoinformatic tools as implemented in MOE and KNIME (75). Only compounds with <23 non-hydrogen atoms and at least one ring were PAINS-filtered and consecutively included in the diversity selection (76). Fragment selection was based on normalized moments of inertia for shape diversity, Tanimoto coefficient (<0.8) using MACCS fingerprint for chemical diversity and scaffold diversity was ensured following definitions given by Murcko and coworkers (77, 78).

Maximum pairwise similarities were calculated in MOE using three-point pharmacophore-based fingerprints (GpiDAPH3) as descriptors and Tanimoto coefficient as similarity metric. The same descriptor was used to assess the chemical complexity of the fragments (31).

Fragments were dissolved in d₆-DMSO (Euriso-Top, Saint-Aubin, France) to 100 mM stock solutions under a nitrogen atmosphere in Matrix plates (Thermo Scientific, Waltham, MA, USA) followed by shaking at room temperature for 18 h at 140 rpm. Fragments were stored at –20°C. Next, each fragment was dissolved under nitrogen atmosphere at 1 mM in 500 μL 10 mM deuterated phosphate buffer, pH 7.0, containing 50 μM d₄-TSP [(3-(trimethylsilyl)-2,2',3,3'-tetra deuterio propionic acid, Sigma-Aldrich), 50 μM TFA (trifluoroacetic acid, Sigma-Aldrich), and 0.01% sodium azide (Carl Roth)]. A ¹⁹F and ¹H NMR spectrum of each fragment was recorded for quality control. All NMR studies were measured at 298 K in Norell SP5000-7 5 mm tubes (Norell, Landisville, NJ, USA) on a Varian PremiumCOMPACT 600 MHz spectrometer equipped with an oneNMR probe (Agilent, Santa Clara, CA, USA) with TSP and TFA as internal references. All spectra were analyzed in MestReNova 9.0.0 (Mestrelab Research, Santiago de Compostela, Spain) for identity and for solubility in D₂O of at least 200 μM. Substances, that did not fulfill these quality criteria (17%), were removed from the library. Chemical shifts were used to design 8 screening mixtures consisting of 36 compounds each. A genetic algorithm was used to solve the optimization problem of mixture prediction (unpublished data). Prior to screening, all mixtures were analyzed in ¹⁹F NMR spectra after 18–24 h incubation at room temperature to ensure stability of the mixtures. Compounds experiencing precipitation or changes in chemical shift were removed from the following experiments. The quality control left 281 compounds (83%) to be prepared in mixtures of 100 μM compound each, 100 μM TFA, 150 mM sodium chloride in 20 mM Tris–HCl, pH 7.8, in 20% D₂O (Euriso-Top) that were stored at –20°C as aliquots until used.

NMR SCREENING

All protein samples were prepared at 20 μM of final concentration in 20 mM Tris–HCl, pH 7.8, with 150 mM sodium chloride and 1 mM EDTA and mixed 1:1 with the screening mixture aliquots resulting in a final protein and compound concentration of 10 and 50 μM, respectively, in 500 μL final volume. Fluorine spectra were recorded with a spectral width of 140 ppm and a transmitter offset at –120 ppm, acquiring 128 scans, with an acquisition time of 0.8 and 2 s relaxation time. T₂-filtered spectra were recorded using a CPMG pulse sequence with a 180° pulse repetition rate of 50 Hz and duration of 1.0 s using same acquisition and relaxation times (79, 80). Two CPMG spectra were recorded per mixture to cover the full spectral width. A spectrum ranging from –50 to –100 ppm and from –100 to –150 ppm was recorded with 96 and 256 scans, respectively. Screening was performed first in the presence and absence of protein including 0.5 mM EDTA. Next, calcium chloride was added to a final concentration of 10 mM and measurements were repeated. All spectra were analyzed for changes in peak intensity and chemical shift. As an additional quality control, frequent hitters identified during unrelated screening campaigns were removed.

SPR FOLLOW-UP SCREENING

All surface plasmon resonance (SPR) measurements were performed on a Biacore® T100 (GE Healthcare, Chalfont St. Giles, UK) with a flow-rate of $10 \mu\text{L}\cdot\text{min}^{-1}$ using HBS-P buffer [10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.6, 150 mM sodium chloride, 0.05% Tween-20] at 298 K. DC-SIGN ECD was immobilized on a CM7 Series S sensor chip in a density of 3317 RU using 0.2 M EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, Sigma-Aldrich) and 0.05 M NHS (*N*-hydroxysuccinimide, Merck, Hohenbrunn, Germany) as coupling reagents. The activated surface was saturated with 1 M ethanolamine (Sigma-Aldrich), pH 8.5, after immobilization. The reference flow cell was treated in the same manner without immobilizing protein. Prior to measurements, the solubility of each compound in SPR buffer was determined by recording absorption spectra at different concentrations between 400 and 800 nm in clear 96-well plates (Nalge Nunc International, Penfield, NY, USA) in a SpectraMax M5 plate reader (Molecular Devices, Sunnyvale, CA, USA). During SPR measurements, fragments were injected for 30 s following a dissociation time of 120 s at $10 \mu\text{L}\cdot\text{min}^{-1}$ flow-rate omitting regeneration as fast off-rates were observed for all ligands. To estimate the apparent affinity of a compound, at least three dilutions between 0.1 and 1 mM depending on the solubility were run in triplicates, blanking the data against a corresponding DMSO control. A positive control was included during screening to ensure stability of the sensorgrams. A 1:1 binding model was applied for data fitting:

$$\text{RU} = \text{RU}_{\text{max}} \frac{[L]}{K_{\text{D, app}} + [L]} \quad (1)$$

with the fragment concentration $[L]$, the measured relative response units RU, the apparent dissociation constant $K_{\text{D, app}}$, and the maximal relative response units RU_{max} using Origin8.6Gpro (OriginLab, Northampton, MA, USA). The maximal relative response units were estimated using:

$$\text{RU}_{\text{max}} = A \cdot \text{RU}_{\text{immobilized}} \frac{\text{MW}_{\text{compound}}}{\text{MW}_{\text{protein}}} \quad (2)$$

with the immobilization level of protein $\text{RU}_{\text{immobilized}}$, the molecular weight of the compound and protein $\text{MW}_{\text{compound}}$ and $\text{MW}_{\text{protein}}$, respectively, and the remaining activity of the protein on the chip A. The latter was determined to be 0.6 using 4 as positive control (Figure S4 in Supplementary Material). The apparent affinity constant for each compound was determined under two conditions, either in the presence of 0.5 mM EDTA or 2 mM calcium chloride included in the running and sample buffer. Ligand efficiencies (LE) were calculated applying

$$\text{LE} = -\frac{RT \ln(K_{\text{D, app}})}{\text{HA}} \quad (3)$$

using the apparent dissociation constant $K_{\text{D, app}}$, the temperature T, the gas constant R, and the number of non-hydrogen atoms HA (81).

RESULTS

STRUCTURE-BASED SEQUENCE ALIGNMENT IDENTIFIES CANONICAL CARBOHYDRATE-BINDING SITES

A comparative framework between the CTLRs served as the starting point of our druggability prediction. To this end, a structure-based sequence alignment was performed for 22 CRDs (Figure S3 in Supplementary Material). With an average of 41%, the global sequence similarity within the set of receptors is low. It spans a range from 26 to 86% (Figure 1A). A phylogenetic analysis based on this alignment yields a dendrogram that resembles the canonical classifications of CTLRs, in particular with respect to the correct assignment of members of the groups II, III, IV, V, and VII (1). Collectin-12 deviates from this classification, as it is part of the group II cluster. Moreover, Tetranectin and eosinophil major basic protein (EMBP) are the only representatives of group IX and XII used in this study. Both display elevated distances to other branches. EMBP and Tetranectin as well as Clec9a, Lox-1, Clec1b, and Reg1a have been reported to interact with non-carbohydrate ligands and all of these CTLRs were assigned to cluster B. Strikingly, CRDs known to recognize carbohydrates via the Ca^{2+} -2-binding site are exclusively present in cluster A (Figure 1A).

CONSENSUS STRUCTURE PREDICTION REVEALS ELEVATED STRUCTURAL VARIABILITY IN THE LONG LOOP

Contrasting the low global sequence similarity, the overall structure of the CTLD is highly conserved. RMSD values of C_{α} atoms obtained from the structure-based multiple sequence alignment are uniformly low and do not exceed 3.2 Å (Figure 1A). To visualize the conservation of the domain architecture, we calculated a consensus structure (Figure 1B). While the core of the CTLD displays only minor deviations, a higher level of structural variability characterizes the two loop regions. The long loop is of particular interest as it harbors the Ca^{2+} -1, -2, and -3 sites and thus plays a fundamental role in Ca^{2+} -dependent carbohydrate recognition (21).

COMPUTATIONAL ANALYSIS PREDICTS LOW DRUGGABILITY FOR THE MAJORITY OF CTLRS

The initial identification of binding sites with DoGSite yielded between three and nine sites for CRDs and 9–19 for ECDs. Next, DoGSiteScorer was applied to calculate druggability scores. In the scoring scheme of this program, scores over 0.5 are indicative of a druggable binding site (73). At least one site that meets this criterion is found for the majority of the analyzed CTLRs (Figure 2A). However, targeting these sites with drug-like molecules will not necessarily exert an effect on the physiological function of the respective CTLR.

We propose that binding sites in proximity to Ca^{2+} ions located in the long loop region are relevant to carbohydrate recognition. Therefore, we assumed that small molecule-binding to these sites potentially modulates CTLR function. To this end, binding sites were assigned to four categories: (i) Ca^{2+} -2-dependent, (ii) Ca^{2+} -associated binding sites, (iii) Ca^{2+} -independent carbohydrate-binding sites, and (iv) other binding sites (Figure 2A). Ca^{2+} -associated binding sites (i, ii) were identified by DoGSite for all CTLRs coordinating a Ca^{2+} -2 ion except for Mincle and the Langerin trimer. Experimentally determined Ca^{2+} -independent

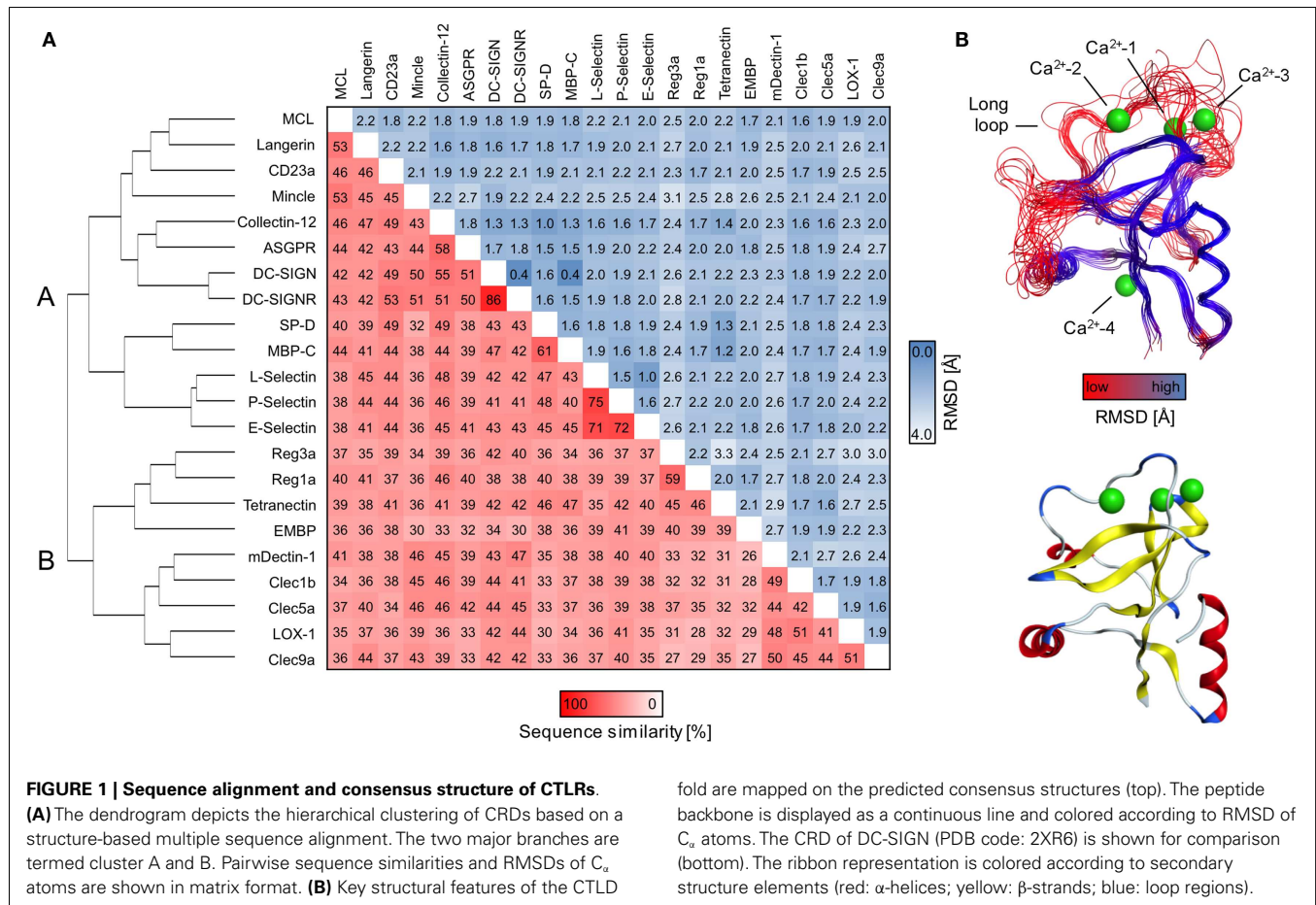


FIGURE 1 | Sequence alignment and consensus structure of CTLRs.

(A) The dendrogram depicts the hierarchical clustering of CRDs based on a structure-based multiple sequence alignment. The two major branches are termed cluster A and B. Pairwise sequence similarities and RMSDs of C_α atoms are shown in matrix format. (B) Key structural features of the CTLD

fold are mapped on the predicted consensus structures (top). The peptide backbone is displayed as a continuous line and colored according to RMSD of C_α atoms. The CRD of DC-SIGN (PDB code: 2XR6) is shown for comparison (bottom). The ribbon representation is colored according to secondary structure elements (red: α-helices; yellow: β-strands; blue: loop regions).

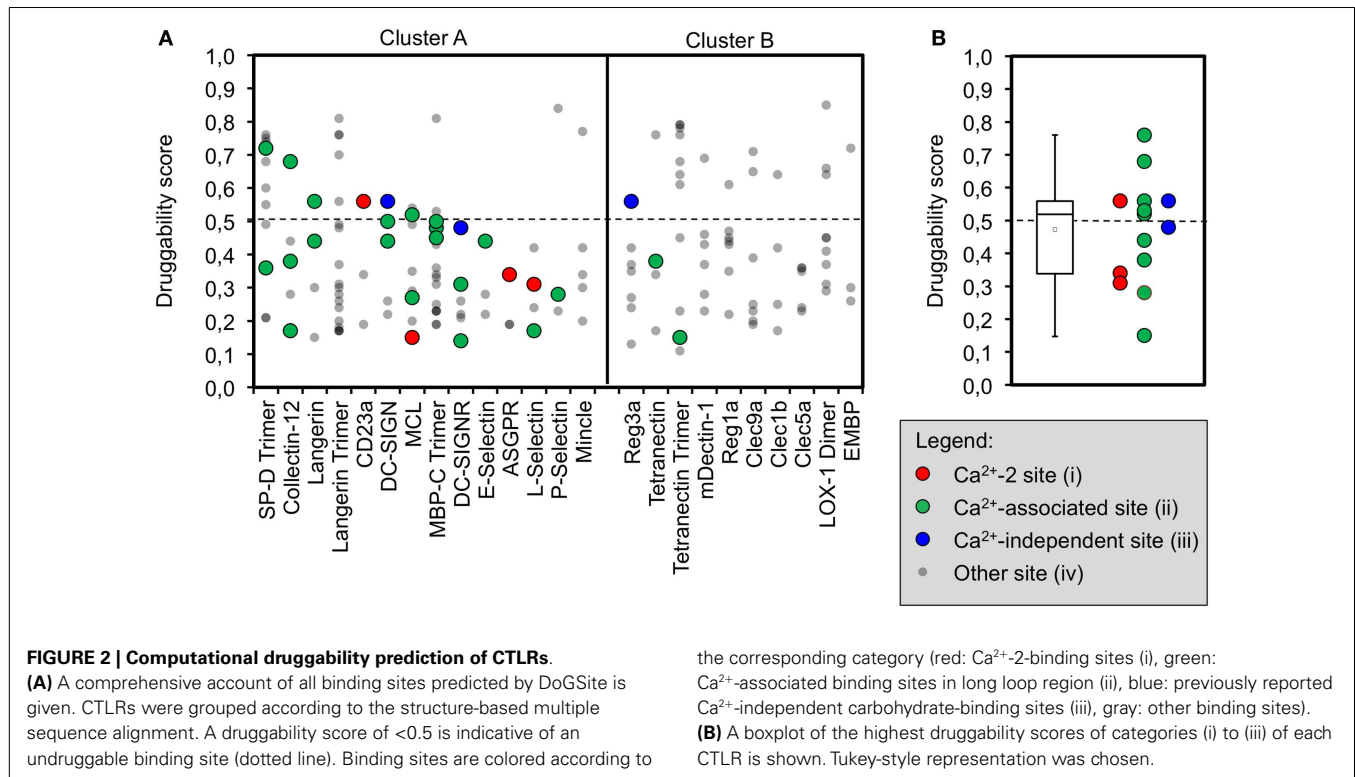
carbohydrate-binding sites (iii) were identified for DC-SIGN, DC-SIGNR, and Reg3a (58, 82). The existence of a single druggable site is sufficient to render a target druggable. Accordingly, for each CTLR, sites assigned to categories (i) and (ii) displaying the highest score were selected for statistical analysis and a mean druggability score of 0.47 was calculated (Figure 2B). This classifies CTLRs as “difficult” or even “undruggable” targets (73). Notably, individual receptors such as SP-D and Collectin-12 possess favorable pockets in the long loop region. Other targets such as E-Selectin display druggability values well below the mean.

FRAGMENT SCREENING REVEALS HIGH HIT RATES AGAINST DC-SIGN, LANGERIN, AND MCL

The existence of pockets on the surface of a receptor that are suitable to accommodate drug-like ligands can be assessed experimentally using fragment screening. The resulting hit rate serves as a predictor for druggability. Therefore, we composed a chemical library of fragments to be used in a homogeneous, label-free NMR-based screening assay. All fragments carry a fluorine atom, which allows for ¹⁹F NMR spectroscopy-based assessment of fragment binding. After quality control, 281 fragments were available for screening in 8 mixtures of maximum 36 fragments. The fragment library displays high shape and chemical diversity (Figures 3A,B).

DC-SIGN CRD and ECD were screened against the fragment library using ¹⁹F and T₂-filtered ¹⁹F NMR spectra. Fragment binding to DC-SIGN was observed monitoring changes in chemical shift, line broadening, and T₂ relaxation. Moreover, three spectra were recorded per fragment mixture. First, a spectrum was recorded in the absence of protein to exclude false positives such as Ca²⁺ chelators. The second spectrum was acquired in the presence of 10 μM protein to monitor fragment binding. Finally, Ca²⁺ was added in excess to the protein-fragment mixture, hypothesizing that metal binding to DC-SIGN modulates interaction of those fragments that are good candidates for inhibition of carbohydrate recognition (*vide supra*). Hits for DC-SIGN CRD and ECD were combined and frequent hitters were removed. Consequently, we identified 38 hits (13.5%) from mixtures binding to DC-SIGN in a Ca²⁺-dependent manner (Figure 3C). Out of these hits, 16 were found in both screenings and 21 hits were identified only during the CRD screening. Only one fragment was found while the ECD was used for screening.

To further validate these hits, SPR spectroscopy was employed as an orthogonal biophysical assay. This method not only detects binding of small molecules to macromolecules, but also allows for the determination of equilibrium dissociation constants. DC-SIGN ECD was immobilized on the chip surface and two experimental setups were utilized to differentiate Ca²⁺-mediated fragment binding from Ca²⁺-fragment competition. Fragments



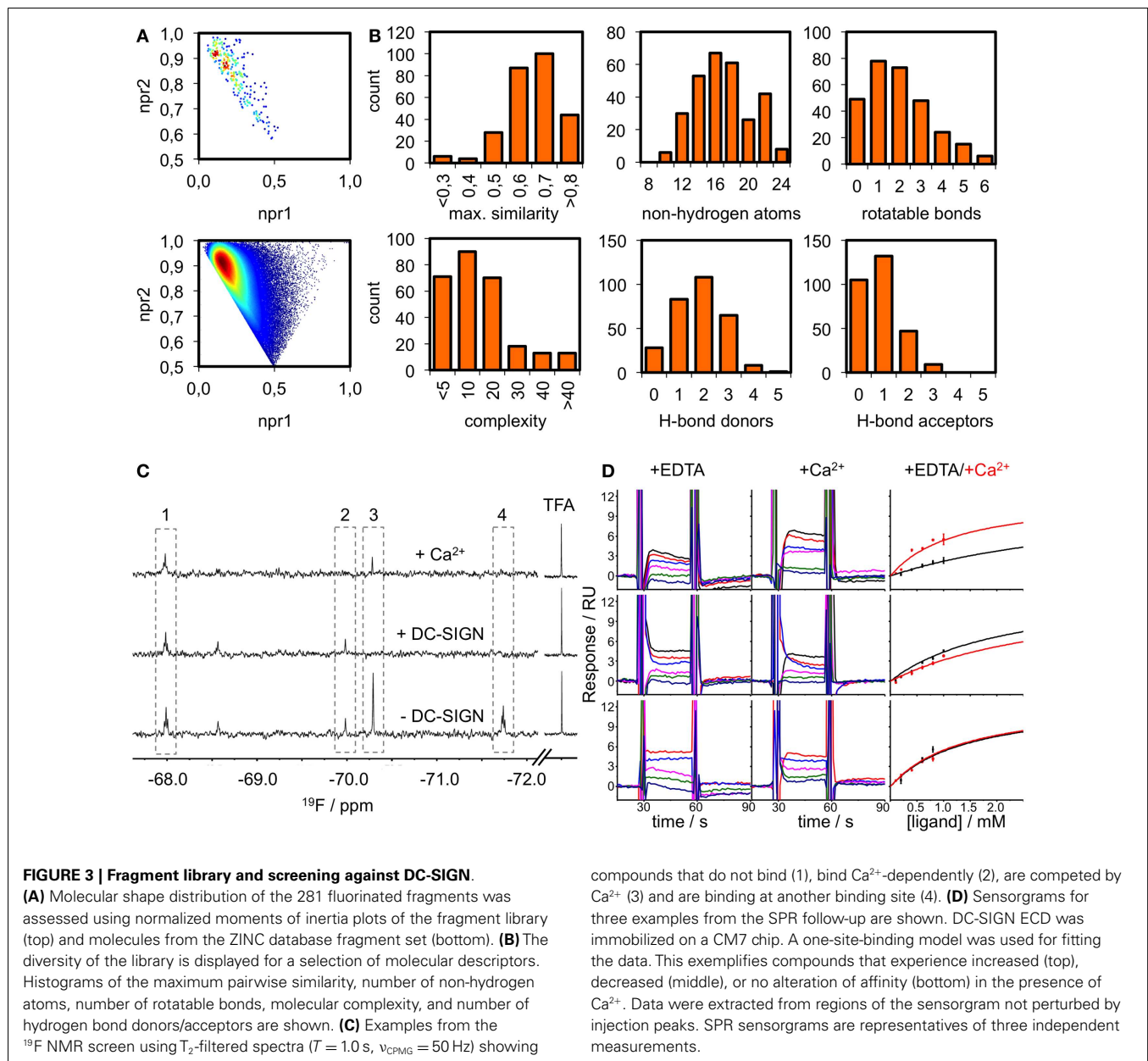
were injected either in the presence of 0.5 mM EDTA or 2 mM calcium chloride (**Figure 3D**), confirming a 1:1 binding model for 18 fragments (47%). Five fragments (13%) bound with a higher stoichiometry, 3 experienced no change in response in presence or absence of Ca^{2+} (8%), and 12 fragments (32%) did not give rise to detectable signals. The highest affinities measured were in the upper micromolar to lower millimolar range ($0.6 \text{ mM} < K_{D,app} > 1.3 \text{ mM}$). Of the 18 fragments confirmed by SPR, 9 showed increase in affinity upon Ca^{2+} addition and 9 displayed competitive behavior. Moreover, fragments similar to substructures of an already published submicromolar DC-SIGN inhibitor were identified (41, 42) (**Figure 4**). While fragments 1 and 2 bound competitive with the polysaccharide mannan in a ^{19}F NMR competition assay, fragment 3 showed no such behavior upon addition of the natural carbohydrate ligand of DC-SIGN (data not shown).

In light of our computational analysis, we were surprised to find such a high fragment hit rate for DC-SIGN, and decided to expand our ^{19}F NMR-based druggability prediction against the ECDs of two further CTLRs. We decided to screen our fragment library against Langerin being sufficiently distant to DC-SIGN in our structural sequence alignment (**Figure 1A**). To compare these findings to a CTLR more closely related to Langerin, we also included MCL in our analysis. Both proteins were expressed in *E. coli* and screened following the same protocol as for DC-SIGN. Again, Ca^{2+} was utilized as a competitor (Figures S5A,B in Supplementary Material) and several hits associated with Ca^{2+} binding were identified (**Table 3**). The pairwise overlap between the three CTLRs was low and none of the fragment hits bound to all CTLRs (Figure S5C in Supplementary Material).

DISCUSSION

In this report, we assessed the potential of human CTLRs to be targeted with drug-like molecules. Therefore, we explored the ability of a set of CTLRs to accommodate inhibitors to modulate the receptor–carbohydrate interaction. This druggability prediction is an important part of the decision on whether a drug discovery campaign should be pursued (28–30). Despite a large body of recent research highlighting the importance of CTLRs in immune cell regulation, pathogen uptake, and as targets for adjuvants, only a few drug-like molecules have been developed for the CTLR family (2, 16, 17, 20). Herein, we aimed to rationalize why these receptors are considered challenging targets.

To start our investigations, CTLR druggability was predicted by computational methods. No data focusing on CTLRs are available and more general reports on glycan-binding proteins presented low druggability scores (39, 40). Unfortunately, the exact structures were either not disclosed or highly redundant and no CTLR was explicitly included. We assembled a set of 21 human CTLRs, and the murine Dectin-1. The latter was included as a reference as it is a well-studied CTLR and harbors a potential non-canonical, calcium-independent carbohydrate recognition site. The druggability prediction was performed using DoGSiteScorer, recently released software to predict the druggability of protein targets based on structural and physicochemical properties (73). Here, potential pockets on the protein surface were identified first, and then scored according to their physicochemical properties. Major determinants of druggability are depth, volume, and amino acid composition of the pocket (28, 32, 34, 36, 73). Generally, highly hydrophilic binding sites are considered undruggable (36).



Between three and nine binding sites were identified for CRDs, which is in accordance with values reported for other protein families (32). For Langerin, MBP-C, and Tetranectin, data on the homo-trimeric form were available. Here, the algorithm identified more potential sites, which is not surprising due to the larger surface area and symmetry of the assemblies (Figure 2A). Yet, targeting this initial set of binding sites does not necessarily interfere with carbohydrate recognition. Therefore, we categorized pockets according to their potential to modulate glycan binding. We argue that a druggable pocket located in close proximity of the long loop renders it a potential binding site for an inhibitor. The loop exhibits considerable movement in the absence of calcium as observed for other CRDs (65, 67, 83, 84) and adjacent sites have been proposed to communicate with the primary carbohydrate

recognition site (22, 23). Four categories of sites were defined out of which only two, namely categories (i) and (ii), are either directly or indirectly associated with calcium ion binding.

The success-rate of detecting the canonical Ca^{2+} -2 site (i) was low. Only 4 out of the 14 structures known to harbor such a site were identified (Figure 2A). This low number reflects a limitation of the employed pocket prediction, potentially due to shallow architecture of the Ca^{2+} -2 sites. The low druggability score of the successfully identified Ca^{2+} -2 sites corroborates this finding. Overall, these findings suggest that identification of carbohydrate recognition sites with computational algorithms such as DoGSite is challenging (*vide infra*).

Moreover, we analyzed a larger panel of sites associated with either the Ca^{2+} -1, -2, or -3 site, summarized in category (ii)

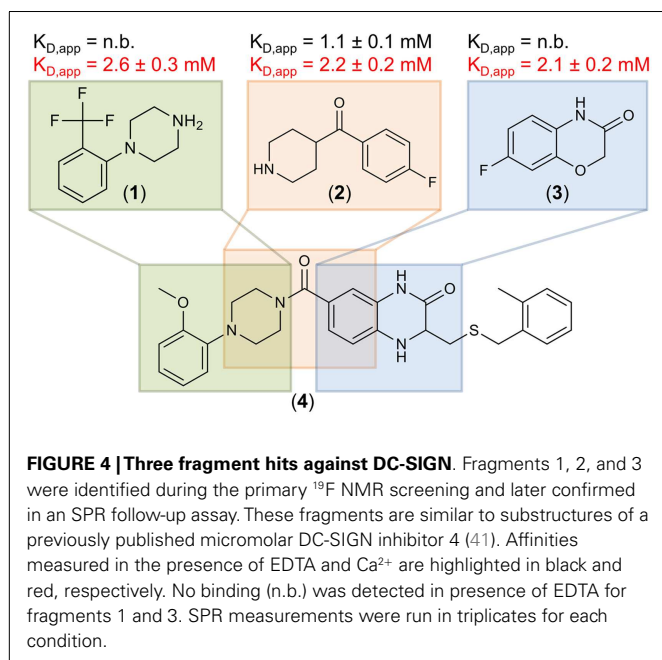


Table 3 | Hit rates for three C-type lectin receptors from ^{19}F NMR screening against a library of 281 compounds and hits confirmed by SPR for DC-SIGN.

C-type lectin receptor	Ca^{2+} -associated hits	Validated hits by SPR
DC-SIGN	38 (13.5%)	18 (6.4%)
Langerin	44 (15.7%)	
MCL	28 (10.0%)	

(Figure 2A). The criteria of this category were less stringent and based on an extended definition of sites potentially interfering with carbohydrate binding. Again, druggable sites were sparse. Collectin-12 and SP-D, both members of the Collectin group (CTLR group III), represent notable exceptions. Furthermore, our data on Langerin, for which monomer and trimer were analyzed side by side, highlight that subtle changes in the long loop region upon oligomerization abrogate the recognition of these sites by DoGSite (62).

Low scores for category (ii) sites are also found for members of cluster B of the sequence alignment. This cluster exclusively contains CTLRs not known to bind carbohydrates with their Ca^{2+} -2 site (Figure 2A). The Ca^{2+} -independent carbohydrate-binding sites of category (iii) found for Reg3a (group VII) is located in other regions of the CRD fold and has druggability scores of 0.56, predicting this CTLR to be challenging (82). Overall, only a few members of the CTLR family were predicted to be druggable (Figure 2B), which is in line with previous reports on glycan-binding proteins (39, 40).

To substantiate the computational studies, a ^{19}F NMR-based fragment screening against one of the analyzed CTLRs was conducted. We chose DC-SIGN because as a viral uptake receptor it is of pharmacological interest and has been targeted in a high-throughput screening (41). While the successful HTS was

already an indicator of DC-SIGN being amenable to fragment binding, the low druggability assessment by our computational analysis predicted a low hit rate of fragments interfering with any of the three DC-SIGN calcium sites. To our surprise, a high hit rate of 13.5% of the fragments from our library bound to DC-SIGN in Ca^{2+} -associated sites during the NMR screening. The follow-up screening via SPR validated 18 (47%) of these fragments, a value not unusual for these two assay systems (85). Hits that were not validated by the SPR screening were either superstoichiometric binders (13%), not competitive with Ca^{2+} (8%), or had affinities below the detection limit of the SPR assay. The latter can be attributed to the high sensitivity of ^{19}F NMR as a primary screen (38). Together, NMR and SPR result in a hit rate of 6.4%, which is in the expected range for fragment-based screenings and does not suggest a low likelihood to bind drug-like molecules (31, 37, 43, 86).

We performed the primary NMR screen against the CRD and the tetrameric ECD of DC-SIGN. Notably, only one fragment was uniquely identified during the screening of the ECD compared to 21 in the CRD screening. Conversely, many fragments binding to the ECD were later discovered to be false positives, such as frequent hitters from unrelated screening campaigns against non-CTLR targets. Hence, we conclude that screening for inhibitors has a lower false positive rate in absence of the neck region of DC-SIGN.

Another indicator for the validity of our screen to discover fragments inhibiting carbohydrate binding to DC-SIGN was the identification of the three fragments 1, 2, and 3. These hits are similar to substructures of the previously reported micromolar DC-SIGN inhibitor 4 (Figure 4) (41). In this respect, four has been shown to compete with carbohydrate binding and antagonized the DC-SIGN-mediated cell adhesion and particle uptake (41, 42). Direct competition between four and the three fragments was hampered by direct interaction of the fragments with four in absence of DC-SIGN (data not shown). Thus, mannan was employed to compete with fragments 1–3 and resulted in reproducible competition with fragments 1 and 2 (data not shown). Although, fragment 3 did not experience competition with the natural ligand, it can be speculated that it is associated with the binding site, as recognition was detected in SPR only in presence of Ca^{2+} (Figure 4). Moreover, other fragment hits showed even higher LE ranging from 0.30 to 0.37, which is a good starting point for further fragment evolution. A subsequent expansion of our ^{19}F NMR-based screening to Langerin and MCL, also revealed similarly high hit rates (Table 3). Following up on these initial hits is subject of current research in the laboratory.

These encouraging experimental results are in contrast to our computational predictions. We attribute this conflict to the limitations of the DoGSiteScorer algorithm, which on the one hand is not parameterized for carbohydrate or metal binding sites (72) and on the other does not account for protein flexibility. Currently, there is no single software for druggability prediction available that is able to overcome these limitations.

Throughout the experimental evaluation, we employed competition with calcium ions as an indicator for the inhibition of carbohydrate recognition. We assumed the existence of allosteric

sites originating from the flexibility of the long loop and cooperativity between the adjacent sites as previously described for other CTLRs (22, 23, 65, 67, 83, 84). In this context, it should be noted that accounting for conformational dynamics is recognized as a particular challenge for the development of improved algorithms (34).

To summarize, we report high *in silico* druggability scores for group III and V CTLRs as well as high experimental hit rates from fragment screenings against group II CTLRs. These data stand alongside with a successful drug design campaign that has already been launched against group IV CTLRs (19). Hence, we conclude that our data, while highlighting the limitations of current computational methods, support the assessment of CTLRs as suitable targets for drug-like molecules.

ACKNOWLEDGMENTS

The authors thank the Max Planck Society, the German Research Foundation (DFG, Emmy Noether program RA1944/2-1), and the Chemical Industry Fund for financial support. Dr. Andrea Volkamer is gratefully acknowledged for support using DoGSiteScorer and Olaf Niemeyer for NMR technical support. Jonas Hanske is supported by a fellowship from the Chemical Industry Fund and Eike Christian Wamhoff acknowledges funding from the Collaborative Research Centre 765. We thank Prof. Dr. Peter H. Seeberger for support and helpful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/Journal/10.3389/fimmu.2014.00323/abstract>

REFERENCES

- Varki A. *Essentials of Glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (2009).
- van Kooyk Y, Rabinovich GA. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat Immunol* (2008) **9**:593–601. doi:10.1038/ni.f.203
- Geijtenbeek TB, Gringhuis SI. Signalling through C-type lectin receptors: shaping immune responses. *Nat Rev Immunol* (2009) **9**:465–79. doi:10.1038/nri2569
- Fuster MM, Esko JD. The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat Rev Cancer* (2005) **5**:526–42. doi:10.1038/nrc1649
- Stowell SR, Arthur CM, Dias-Baruffi M, Rodrigues LC, Gourdine JP, Heimburg-Molinario J, et al. Innate immune lectins kill bacteria expressing blood group antigen. *Nat Med* (2010) **16**:295–301. doi:10.1038/nm.2103
- Cummings RD, McEver RP. C-type lectins. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (2009). p. 439–58.
- Weis WI, Taylor ME, Drickamer K. The C-type lectin superfamily in the immune system. *Immunol Rev* (1998) **163**:19–34. doi:10.1111/j.1600-065X.1998.tb01185.x
- Curtis BM, Scharnowske S, Watson AJ. Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. *Proc Natl Acad Sci U S A* (1992) **89**:8356–60. doi:10.1073/pnas.89.17.8356
- Geijtenbeek TB, Torensma R, Van Vliet SJ, Van Duijnhoven GC, Adema GJ, Van Kooyk Y, et al. Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell* (2000) **100**:575–85. doi:10.1016/S0092-8674(00)80693-5
- Geijtenbeek TB, Kwon DS, Torensma R, Van Vliet SJ, Van Duijnhoven GC, Middel J, et al. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* (2000) **100**:587–97. doi:10.1016/S0092-8674(00)80694-7
- Appelmelk BJ, Van Die I, Van Vliet SJ, Vandenbroucke-Grauls CM, Geijtenbeek TB, Van Kooyk Y. Cutting edge: carbohydrate profiling identifies new pathogens that interact with dendritic cell-specific ICAM-3-grabbing nonintegrin on dendritic cells. *J Immunol* (2003) **170**:1635–9. doi:10.4049/jimmunol.170.4.1635
- Lozach PY, Lortat-Jacob H, de Lacroix de Lavalette A, Staropoli I, Foug S, Amara A, et al. DC-SIGN and L-SIGN are high affinity binding receptors for hepatitis C virus glycoprotein E2. *J Biol Chem* (2003) **278**:20358–66. doi:10.1074/jbc.M301284200
- Simmons G, Reeves JD, Grogan CC, Vandenberghe LH, Baribaud F, Whitbeck JC, et al. DC-SIGN and DC-SIGNR bind ebola glycoproteins and enhance infection of macrophages and endothelial cells. *Virology* (2003) **305**:115–23. doi:10.1006/viro.2002.1730
- Tailleux L, Schwartz O, Herrmann JL, Pivert E, Jackson M, Amara A, et al. DC-SIGN is the major *Mycobacterium tuberculosis* receptor on human dendritic cells. *J Exp Med* (2003) **197**:121–7. doi:10.1084/jem.20021468
- Sancho D, Reis e Sousa C. Signaling by myeloid C-type lectin receptors in immunity and homeostasis. *Annu Rev Immunol* (2012) **30**:491–529. doi:10.1146/annurev-immunol-031210-101352
- van Kooyk Y, Geijtenbeek TB. DC-SIGN: escape mechanism for pathogens. *Nat Rev Immunol* (2003) **3**:697–709. doi:10.1038/nri1182
- Ernst B, Magnani JL. From carbohydrate leads to glycomimetic drugs. *Nat Rev Drug Discov* (2009) **8**:661–77. doi:10.1038/nrd2852
- Kansas GS. Selectins and their ligands: current concepts and controversies. *Blood* (1996) **88**:3259–87.
- Chang J, Patton JT, Sarkar A, Ernst B, Magnani JL, Frenette PS. GMI-1070, a novel pan-selectin antagonist, reverses acute vascular occlusions in sickle cell mice. *Blood* (2010) **116**:1779–86. doi:10.1182/blood-2009-12-260513
- Lang R, Schoenen H, Desel C. Targeting Syk-Card9-activating C-type lectin receptors by vaccine adjuvants: findings, implications and open questions. *Immunobiology* (2011) **216**:1184–91. doi:10.1016/j.imbio.2011.06.005
- Zelensky AN, Gready JE. The C-type lectin-like domain superfamily. *FEBS J* (2005) **272**:6179–217. doi:10.1111/j.1742-4658.2005.05031.x
- Blomhoff R, Tolleshaug H, Berg T. Binding of calcium ions to the isolated asialoglycoprotein receptor. Implications for receptor function in suspended hepatocytes. *J Biol Chem* (1982) **257**:7456–9.
- Onizuka T, Shimizu H, Moriwaki Y, Nakano T, Kanai S, Shimada I, et al. NMR study of ligand release from asialoglycoprotein receptor under solution conditions in early endosomes. *FEBS J* (2012) **279**:2645–56. doi:10.1111/j.1742-4658.2012.08643.x
- Tabarani G, Thepaut M, Stroebel D, Ebel C, Vives C, Vachette P, et al. DC-SIGN neck domain is a pH-sensor controlling oligomerization: SAXS and hydrodynamic studies of extracellular domain. *J Biol Chem* (2009) **284**:21229–40. doi:10.1074/jbc.M109.021204
- Brown J, O'Callaghan CA, Marshall AS, Gilbert RJ, Siebold C, Gordon S, et al. Structure of the fungal beta-glucan-binding immune receptor dectin-1: implications for function. *Protein Sci* (2007) **16**:1042–52. doi:10.1110/ps.072791207
- Weis WI, Drickamer K. Structural basis of lectin-carbohydrate recognition. *Annu Rev Biochem* (1996) **65**:441–73. doi:10.1146/annurev.bi.65.070196.002301
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* (2001) **46**:3–26. doi:10.1016/S0169-409X(00)00129-0
- Egner U, Hillig RC. A structural biology view of target drugability. *Expert Opin Drug Discov* (2008) **3**:391–401. doi:10.1517/17460441.3.4.391
- Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov* (2002) **1**:727–30. doi:10.1038/nrd892
- Cheng AC, Coleman RG, Smyth KT, Cao Q, Soular P, Caffrey DR, et al. Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol* (2007) **25**:71–5. doi:10.1038/nbt1273
- Chen IJ, Hubbard RE. Lessons for fragment library design: analysis of output from multiple screening campaigns. *J Comput Aided Mol Des* (2009) **23**:603–20. doi:10.1007/s10822-009-9280-5
- Sheridan RP, Maiorov VN, Holloway MK, Cornell WD, Gao YD. Drug-like density: a method of quantifying the “bindability” of a protein target based on a very large set of pockets and drug-like ligands from the Protein Data Bank. *J Chem Inf Model* (2010) **50**:2029–40. doi:10.1021/ci100312t

33. Edfeldt FN, Folmer RH, Breeze AL. Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discov Today* (2011) **16**:284–7. doi:10.1016/j.drudis.2011.02.002
34. Henrich S, Salo-Ahen OM, Huang B, Rippmann FF, Cruciani G, Wade RC. Computational approaches to identifying and characterizing protein binding sites for ligand design. *J Mol Recognit* (2010) **23**:209–19. doi:10.1002/jmr.984
35. Perot S, Sperandio O, Miteva MA, Camproux AC, Villoutreix BO. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov Today* (2010) **15**:656–67. doi:10.1016/j.drudis.2010.05.015
36. Halgren TA. Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* (2009) **49**:377–89. doi:10.1021/ci800324m
37. Hajduk PJ, Huth JR, Fesik SW. Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem* (2005) **48**:2518–25. doi:10.1021/jm049131r
38. Jordan JB, Poppe L, Xia X, Cheng AC, Sun Y, Michelsen K, et al. Fragment based drug discovery: practical implementation based on 19F NMR spectroscopy. *J Med Chem* (2012) **55**:678–87. doi:10.1021/jm201441k
39. Hajduk PJ, Huth JR, Tse C. Predicting protein druggability. *Drug Discov Today* (2005) **10**:1675–82. doi:10.1016/s1359-6446(05)03624-x
40. Schmidtke P, Barril X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J Med Chem* (2010) **53**:5858–67. doi:10.1021/jm100574m
41. Borrok MJ, Kiessling LL. Non-carbohydrate inhibitors of the lectin DC-SIGN. *J Am Chem Soc* (2007) **129**:12780–5. doi:10.1021/ja072944v
42. Mangold SL, Prost LR, Kiessling LL. Quinoxalinone inhibitors of the lectin DC-SIGN. *Chem Sci* (2012) **3**:772–7. doi:10.1039/c2sc00767c
43. Hann MM, Leach AR, Harper G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci* (2001) **41**:856–64. doi:10.1021/ci000403i
44. Barelier S, Pons J, Gehring K, Lancelin JM, Krimm I. Ligand specificity in fragment-based drug design. *J Med Chem* (2010) **53**:5256–66. doi:10.1021/jm100496j
45. Davis BJ, Erlanson DA. Learning from our mistakes: the ‘unknown knowns’ in fragment screening. *Bioorg Med Chem Lett* (2013) **23**:2844–52. doi:10.1016/j.bmcl.2013.03.028
46. Dalvit C, Fagerness PE, Hadden DT, Sarver RW, Stockman BJ. Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. *J Am Chem Soc* (2003) **125**:7696–703. doi:10.1021/ja034646d
47. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. *Nucleic Acids Res* (2012) **40**:D290–301. doi:10.1093/nar/gkr1065
48. Feinberg H, Park-Snyder S, Kolatkar AR, Heise CT, Taylor ME, Weis WI. Structure of a C-type carbohydrate recognition domain from the macrophage mannose receptor. *J Biol Chem* (2000) **275**:21539–48. doi:10.1074/jbc.M002366200
49. Nagae M, Ikeda A, Kitago Y, Matsumoto N, Yamamoto K, Yamaguchi Y. Crystal structures of carbohydrate recognition domain of blood dendritic cell antigen-2 (BDCA2) reveal a common domain-swapped dimer. *Proteins* (2014) **82**:1512–8. doi:10.1002/prot.24504
50. Chemical Computing Group I. *Molecular Operating Environment (MOE)*. Montreal, QC: Chemical Computing Group Inc. (2013).
51. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* (2004) **20**:289–90. doi:10.1093/bioinformatics/btg412
52. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (2014).
53. Meier M, Bider MD, Malashkevich VN, Spiess M, Burkhard P. Crystal structure of the carbohydrate recognition domain of the H1 subunit of the asialoglycoprotein receptor. *J Mol Biol* (2000) **300**:857–65. doi:10.1006/jmbi.2000.3853
54. Wurzburg BA, Tarchevskaya SS, Jardetzky TS. Structural changes in the lectin domain of CD23, the low-affinity IgE receptor, upon calcium binding. *Structure* (2006) **14**:1049–58. doi:10.1016/j.str.2006.03.017
55. Watson AA, Brown J, Harlos K, Eble JA, Walter TS, O’Zallaghan CA. The crystal structure and mutational binding analysis of the extracellular domain of the platelet-activating receptor CLEC-2. *J Biol Chem* (2007) **282**:3165–72. doi:10.1074/jbc.M610383200
56. Watson AA, Lebedev AA, Hall BA, Fenton-May AE, Vagin AA, Dejnirattisai W, et al. Structural flexibility of the macrophage dengue virus receptor CLEC5A: implications for ligand binding and signaling. *J Biol Chem* (2011) **286**:24208–18. doi:10.1074/jbc.M111.226142
57. Zhang JG, Czabotar PE, Policheni AN, Caminschi I, Wan SS, Kitsoulis S, et al. The dendritic cell receptor Clec9A binds damaged cells via exposed actin filaments. *Immunity* (2012) **36**:646–57. doi:10.1016/j.immuni.2012.03.009
58. Feinberg H, Mitchell DA, Drickamer K, Weis WI. Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science* (2001) **294**:2163–6. doi:10.1126/science.1066371
59. Graves BJ, Crowther RL, Chandran C, Rumberger JM, Li S, Huang KS, et al. Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains. *Nature* (1994) **367**:532–8. doi:10.1038/367532a0
60. Swaminathan GJ, Weaver AJ, Loegering DA, Checkel JL, Leonidas DD, Gleich GJ, et al. Crystal structure of the eosinophil major basic protein at 1.8 Å. An atypical lectin with a paradigm shift in specificity. *J Biol Chem* (2001) **276**:26197–203. doi:10.1074/jbc.M100848200
61. Feinberg H, Taylor ME, Razi N, McBride R, Knirel YA, Graham SA, et al. Structural basis for langerin recognition of diverse pathogen and mammalian glycans through a single binding site. *J Mol Biol* (2011) **405**:1027–39. doi:10.1016/j.jmb.2010.11.039
62. Feinberg H, Powlesland AS, Taylor ME, Weis WI. Trimeric structure of langerin. *J Biol Chem* (2010) **285**:13285–93. doi:10.1074/jbc.M109.086058
63. Park H, Adsit FG, Boyington JC. The 1.4 angstrom crystal structure of the human oxidized low density lipoprotein receptor lox-1. *J Biol Chem* (2005) **280**:13593–9. doi:10.1074/jbc.M500768200
64. Sheriff S, Chang CY, Ezekowitz RA. Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil. *Nat Struct Biol* (1994) **1**:789–94. doi:10.1038/nsb1194-789
65. Furukawa A, Kamishikiryo J, Mori D, Toyonaga K, Okabe Y, Toji A, et al. Structural analysis for glycolipid recognition by the C-type lectins Mincle and MCL. *Proc Natl Acad Sci U S A* (2013) **110**:17438–43. doi:10.1073/pnas.1312649110
66. Somers WS, Tang J, Shaw GD, Camphausen RT. Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1. *Cell* (2000) **103**:467–79. doi:10.1016/S0092-8674(00)00138-0
67. Abergel C, Chenivresse S, Stinnakre MG, Guasco S, Brechot C, Claverie JM, et al. Crystallization and preliminary crystallographic study of HIP/PAP, a human C-lectin overexpressed in primary liver cancers. *Acta Crystallogr D Biol Crystallogr* (1999) **55**:1487–9. doi:10.1107/S0907444999007969
68. Gerbaud V, Pignol D, Loret E, Bertrand JA, Berland Y, Fontecilla-Camps JC, et al. Mechanism of calcite crystal growth inhibition by the N-terminal undecapeptide of lithostathine. *J Biol Chem* (2000) **275**:1057–64. doi:10.1074/jbc.275.2.1057
69. Shrive AK, Martin C, Burns I, Paterson JM, Martin JD, Townsend JB, et al. Structural characterisation of ligand-binding determinants in human lung surfactant protein D: influence of Asp325. *J Mol Biol* (2009) **394**:776–88. doi:10.1016/j.jmb.2009.09.057
70. Nielsen BB, Kastrup JS, Rasmussen H, Holtet TL, Graversen JH, Etzerodt M, et al. Crystal structure of tetranectin, a trimeric plasminogen-binding protein with an alpha-helical coiled coil. *FEBS Lett* (1997) **412**:388–96. doi:10.1016/S0014-5793(97)00664-9
71. Kastrup JS, Nielsen BB, Rasmussen H, Holtet TL, Graversen JH, Etzerodt M, et al. Structure of the C-type lectin carbohydrate recognition domain of human tetranectin. *Acta Crystallogr D Biol Crystallogr* (1998) **54**:757–66. doi:10.1107/S0907444997016806
72. Volkamer A, Griewel A, Grombacher T, Rarey M. Analyzing the topology of active sites: on the prediction of pockets and subpockets. *J Chem Inf Model* (2010) **50**:2041–52. doi:10.1021/ci100241y
73. Volkamer A, Kuhn D, Grombacher T, Rippmann F, Rarey M. Combining global and local measures for structure-based druggability predictions. *J Chem Inf Model* (2012) **52**:360–72. doi:10.1021/ci200454v
74. Stambach NS, Taylor ME. Characterization of carbohydrate recognition by langerin, a C-type lectin of Langerhans cells. *Glycobiology* (2003) **13**:401–10. doi:10.1093/glycob/cwg045
75. Berthold MR, Cebron N, Dill F, Gabriel TR, Kotter T, Meinel T, et al. *KNIME: The Konstanz Information Miner*. Berlin: Springer (2008).
76. Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* (2010) **53**:2719–40. doi:10.1021/jm901137j

77. Bemis GW, Murcko MA. The properties of known drugs. 1. Molecular frameworks. *J Med Chem* (1996) **39**:2887–93. doi:10.1021/jm9602928
78. Sauer WH, Schwarz MK. Molecular shape diversity of combinatorial libraries: a prerequisite for broad bioactivity. *J Chem Inf Comput Sci* (2003) **43**:987–1003. doi:10.1021/ci025599w
79. Carr HY, Purcell EM. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys Rev* (1954) **94**:630–8. doi:10.1103/Physrev.94.630
80. Meiboom S, Gill D. Modified spin-echo method for measuring nuclear relaxation times. *Rev Sci Instrum* (1958) **29**:688–91. doi:10.1063/1.1716296
81. Hopkins AL, Groom CR, Alex A. Ligand efficiency: a useful metric for lead selection. *Drug Discov Today* (2004) **9**:430–1. doi:10.1016/S1359-6446(04)03069-7
82. Lehotzky RE, Partch CL, Mukherjee S, Cash HL, Goldman WE, Gardner KH, et al. Molecular basis for peptidoglycan recognition by a bactericidal lectin. *Proc Natl Acad Sci U S A* (2010) **107**:7722–7. doi:10.1073/pnas.0909449107
83. Ng KK, Park-Snyder S, Weis WI. Ca²⁺-dependent structural changes in C-type mannose-binding proteins. *Biochemistry* (1998) **37**:17965–76. doi:10.1021/bi981972a
84. Nielbo S, Thomsen JK, Graversen JH, Jensen PH, Etzerodt M, Poulsen FM, et al. Structure of the plasminogen kringle 4 binding calcium-free form of the C-type lectin-like domain of tetranectin. *Biochemistry* (2004) **43**:8636–43. doi:10.1021/bi049570s
85. Wielens J, Headey SJ, Rhodes DI, Mulder RJ, Dolezal O, Deadman JJ, et al. Parallel screening of low molecular weight fragment libraries: do differences in methodology affect hit identification? *J Biomol Screen* (2013) **18**:147–59. doi:10.1177/1087057112465979
86. Mashalidis EH, Sledz P, Lang S, Abell C. A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat Protoc* (2013) **8**:2309–24. doi:10.1038/nprot.2013.130

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 May 2014; accepted: 26 June 2014; published online: 10 July 2014.

Citation: 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. doi: 10.3389/fimmu.2014.00323

This article was submitted to *Immunotherapies and Vaccines*, a section of the journal *Frontiers in Immunology*.

Copyright © 2014 Aretz, Wamhoff, Hanske, Heymann and Rademacher. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.