Trimeric structure of the mouse Kupffer cell C-type lectin receptor Clec4f

Zhenlin Ouyang, Jan Felix, Jinhong Zhou, Yingmei Pei, Bohan Ma, Peter Hwang, M Joanne Lemieux, Irina Gutsche, Fang Zheng, Yurong Wen

FEBS Letters

FEBS PRESS science publishing by scientists

Trimeric structure of the mouse Kupffer cell C-type lectin receptor Clec4f

Zhenlin Ouyang[1,2], Jan Felix[3], Jinhong Zhou[2], Yingmei Pei[1], Bohan Ma[1], Peter M. Hwang[4], M. Joanne Lemieux[4,5], Irina Gutsche[3], Fang Zheng[2] and Yurong Wen[1,2,4,5]

1 The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, China
2 Department of Biochemistry and Molecular Biology, The Key Laboratory of Environment and Genes Related to Disease of Ministry of Education, Health Science Center, Xi'an Jiaotong University, China
3 Institut de Biologie Structurale, Univ. Grenoble Alpes, CEA, CNRS, IBS, Grenoble, France
4 Department of Biochemistry, Faculty of Medicine & Dentistry, Edmonton, Canada
5 Membrane Protein Disease Research Group, University of Alberta, Edmonton, Canada

Correspondence
Y. Wen, The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China
Tel: +86 13572905082
E-mail: Yurong.Wen@xjtu.edu.cn
and
F. Zheng, Department of Biochemistry and Molecular Biology, The Key Laboratory of Environment and Genes Related to Disease of Ministry of Education, Health Science Center, Xi'an Jiaotong University, Xi'an 710061, China
Tel: +86 15002900174
E-mail: Fang.Zheng@xjtu.edu.cn

(Received 8 April 2019, revised 13 July 2019, accepted 30 July 2019, available online 9 August 2019)

doi:10.1002/1873-3468.13565

Edited by Michael Bubb

The C-type lectin receptor Clec4f has been identified as a specific surface marker for Kupffer cells, although its ortholog is absent in humans and its biological function remains elusive. Here, we report the crystal structure of a truncated mouse trimeric Clec4f. The orientation between the carbohydrate-recognition domain of Clec4f and its neck region differs from other C-type lectins, resulting in an observed distance of 45 Å between the glycan-binding sites within the Clec4f trimer. Interestingly, the trimeric coiled-coil interface within its heptad neck region contains multiple polyglutamine interactions instead of the predominantly hydrophobic leucine zipper found in other C-type lectin receptors. The Clec4f trimeric structure displays unique features regarding its assembly and ligand recognition, shedding light on the evolution and diversity of the C-type lectin family.

Keywords: Clec4f; crystallography; C-type lectin; Kupffer cell receptor; trimer

C-type lectins play important roles in a diverse range of physiologic functions because of their ability to recognize a wide range of ligands. Particularly, C-type lectins contribute to both innate and adaptive antimicrobial immune responses by recognizing surface polysaccharides of specific pathogens [1]. The C-type lectin domain family 4 member f (Clec4f) was first characterized as the Kupffer cell receptor (KCR) because of its exclusive expression on Kupffer cells, and it was originally identified as a hepatic fucose binding protein [2,3]. An investigation of rat Clec4f glycan binding preferences demonstrated high affinities for galactose- and N-acetylgalactosamine-terminated glycans [4] as well as desialylated and complex N-linked glycans [5]. Similarly, a systematic glycan array screen for murine Clec4f showed strong binding to oligosaccharides containing either GalNAc or galactose with exposed 3- and 4-OH groups [6]. Furthermore, Clec4f was indicated to be involved in

Abbreviations
Clec4f, C-type lectin domain family 4 member f; CRD, carbohydrate-recognition domain; KCR, Kupffer cell receptor.

alpha galactosylceramide presentation and *Listeria monocytogenes* infection in mouse liver [6]. The exact biological functional of Clec4f in the Kupffer cell is still elusive, with most Clec4f investigations performed in mouse, whereas the Clec4f ortholog is absent in humans because of aberrant splicing as reported by Taylor *et al*. [7].

Kupffer cells represent the largest population of macrophages (mϕ) in mammals and the predominant macrophage population in the liver under homeostatic conditions. Recent studies have shown that Kupffer cells must self-maintain via *in situ* proliferation [8]. Recent studies have shown that Clec4f is Kupffer cell-specific, since it is exclusively expressed by Kupffer cells and is absent in infiltrating MoMϕs [8,9]. A subsequent study has shown that exposure to lipopolysaccharides (LPS) can activate a Clec4f-positive Kupffer cell subset, but not other activated macrophages or monocytes in the liver [10]. Although the exact function of Clec4f in Kupffer cell is not yet clearly delineated, Clec4f is widely used as a highly specific Kupffer cell marker to study the roles of distinct mϕ populations present in the liver. Loss of Clec4f-positive Kupffer cells was characterized in Crimean-Congo hemorrhagic fever virus (CCHFV) in mice, where it was shown to be associated with severe hepatic damage [11]. Li and colleagues have shown that Clec4f is necessary for the clearance of desialylated platelets due to recognition of exposed galactose epitopes [12].

Clec4f is classified as a type II C-type lectin family member, with a carbohydrate-recognition domain (CRD) that recognizes glycans in a Ca^{2+}-dependent (C-type) manner. Besides the CRD, Clec4f contains an N-terminal cytoplasmic signaling domain, a transmembrane hydrophobic helix, and a heptad neck region, which stabilizes trimer formation. Here we report the structure of a truncated trimeric mouse Clec4f containing the CRD and part of the heptad neck domain. The Clec4f trimeric structure reveals two conserved calcium-binding sites. The distance between the glycan-binding site and the orientation between the CRD and neck region differs from that of Langerin and other C-type lectins, which may contribute to the unique recognition pattern of Clec4f. The Clec4f structure presented here enables further investigation of its biological function and helps to shed light on the evolution and diversity of type II C-type lectin families.

Materials and methods

Protein expression, refolding, and purification

The mouse Clec4f protein sequence was aligned with other type II C-type lectin families in which the trimeric neck region was observed to stabilize the trimeric state. Multiple different length mouse Clec4f constructs were tested with expression, refolding, and crystallization, and the construct spanning residue 389 to the C terminus (residue 581) yielded a good diffraction pattern. A DNA fragment corresponding to this region in mouse Clec4f was inserted into the pBAD28a plasmid. The resulting pBAD-mClec4f$_{389–581}$ plasmid was transformed into the *Escherichia coli* BL21(DE3) star cell line. Ensuing expression of recombinant truncated mClec4f resulted in insoluble inclusion bodies. One liter of bacterial culture was harvested by centrifugation and lysed by sonication in a buffer containing 50 mM Tris pH 8.0, 200 mM NaCl, 2 mM EDTA, 0.5% Triton X-100, 0.1 mM PMSF, and 1 mM DTT. Inclusion bodies were collected and washed several times by centrifugation in a buffer containing 50 mM Tris pH 8.0, 200 mM NaCl, 2 mM DTT. The inclusion bodies were further dissolved in 25 mM Tris pH 7.0, 6 M guanidine HCl, and 1 mM DTT. The solubilized supernatant was diluted dropwise into 1 L 25 mM Tris pH 8.0, 150 mM NaCl, 0.8 M arginine, 2.5 mM reduced glutathione, 0.5 mM oxidized glutathione, and 25 mM $CaCl_2$ at 4 °C. After 12 h incubation, the mClec4f protein was concentrated to 2 mL and subsequently loaded onto a Superdex 75 16/60 (GE Healthcare) size-exclusion chromatography column pre-equilibrated with 25 mM Tris pH 8.0, 150 mM NaCl, 5% glycerol, and 25 mM $CaCl_2$. The fractions containing mClec4f were confirmed by SDS/PAGE and concentrated for further use.

Crystallization and X-ray diffraction data collection

Size exclusion chromatography-purified mClec4f fractions were concentrated to 4–5 mg·mL^{-1} and used for crystallization trials. Initial crystallization screens were set up using 384-well plates via sitting drop vapor diffusion against commercially available screen kits at 20 °C. Crystals were grown in 0.1 M potassium thiocyanate, 30% polyethylene glycol monomethyl ether 2000. The mClec4f crystals were cryo-protected with reservoir solution supplemented with 30% glycerol and subsequently cryo-cooled in liquid nitrogen. X-ray diffraction data were collected at Beamline BL19U1 at Shanghai Synchrotron Radiation Facility (SSRF) with a rotation of 360°, with 0.5° per image. The X-ray diffraction data were processed and scaled using XDS [13], data collection details and statistics are summarized in Table 1.

Structure determination and refinement

The dataset strongly indicated the crystal belonged to the trigonal spacegroup P321, with one monomeric mClec4f molecule in the asymmetric unit and 49% solvent content. Molecular replacement was performed with Phaser using truncated human Langerin C-type carbohydrate-recognition domain (PDB: 3KQG) as a model [14,15], and resulted in a single solution with a TFZ score of 31 and R_{free} of 0.44

Table 1. X-ray data collection and refinement statistics

Crystal	Clec4f
Data collection	
Spacegroup	P1
a, b, c (Å)	75.15, 75.19, 61.2
α, β, γ (°)	90.01, 90.04, 120.06
Resolution (Å)	32.6–2.8 (2.9–2.8)
R_{merge}	0.073 (1.4)
R_{meas}	0.088 (1.8)
Multiplicity	3.4 (3.5)
CC(1/2)	0.998 (0.61)
CC*	0.999 (0.80)
I/σ(I)	7.93 (1.0)
Completeness (%)	93.82 (88.6)
Wilson B-factor (Å^2)	97.63
Refinement	
Total Reflections	94367 (9975)
Unique Reflections	27730 (2577)
R_{work}/R_{free}	0.2502/0.2895
Number of atoms	
Protein	7370
Ligands	12
Average B-factor (Å^2)	118.8
Protein ADP (Å^2)	118.8
Ligands (Å^2)	111.4
Ramachandran plot	
Favored/Allowed (%)	92.3/7.6
Root-Mean-Square-Deviation	
Bond lengths (Å)	0.005
Bond Angle (°)	0.75
PDB code	6JJJ

Statistics for the highest resolution shell are shown in parentheses.

after an initial round of refinement. Continued refinement of the Clec4f model after tracing the missing residues revealed an extended N-terminal α helix involved in mClec4f trimer formation. However, further refinement of the model in spacegroup P321 resulted in stalling of R_{free} around 0.39, likely due to a deviation from perfect 3-fold symmetry in the Clec4f trimer. To overcome this, the dataset was rescaled in P1 and the ensuing refinement eventually reduced the R_{free} to 0.34, while still giving a relative complete dataset with overall completeness of 93.82%. The final dataset was processed to 2.8 Å with a = 75.15 Å, b = 75.19 Å, c = 61.2 Å and α = 90.01, β = 90.04, γ = 120.06. Model building and refinement were performed in Coot and Phenix [16,17]. Individual B-factor, NCS torsion-angle and TLS parameters were used in refinement strategies. The final model demonstrated an R_{work}/R_{free} value of 0.245/0.2895 in spacegroup P1. Six Clec4f monomers are present in the crystal asymmetric unit, of which three form a trimer, while the remaining three form trimers with symmetry mates. The final mouse Clec4f model includes residues from 393 to 543, two calcium ions and one water molecule per monomer.

Microscale thermophoresis assay

Glycan binding of reconstituted Clec4f trimer was assessed with saturation binding experiment using the Microscale Thermophoresis Assay as described before [18,19]. The Microscale Thermophoresis Assay was carried on a Bio-Rad CFX384™ RTPCR system, and the protein Thermal Shift™ dye (Thermo Fisher Scientific, Shanghai, China) was used to detect the unfolding transition during melting curves. In detail, 10 μM of trimeric mouse Clec4f recombinant protein was incubated for 30 min in varying concentrations of GalNAc ranging from 0 to 700 μM. The melting curve was recorded by a ramp rate of 0.2 degree·min^{-1}. Data were generated with the imbedded program and further analyzed using the GRAPHPAD PRISM 8 software (La Jolla, CA, USA).

Dynamic light scattering

Dynamic Light Scattering (DLS) measurements were performed with a Malvern Zetasizer Nano ZS (Malvern, Herrenberg, Germany) equipped with a 633-nm He-Ne laser and operating at an angle of 173°. Each protein sample was prepared in Tris-NaCl buffer (25 mM Tris, 150 mM NaCl, 25 mM CaCl2, pH 8.0) at a concentration of 20 μM mClec4f trimers, and filtered through a pre-rinsed 0.2-μm filter followed by equilibration to 25°C before making a minimum of three measurements per sample. The apparent molecular weight was calculated using the manufacturer-provided software from Malvern.

Results and Discussion

Evolution and sequence analysis of mouse Clec4f

Of the 16 members in the type II C-type lectin family classified by the genomic animal lectin resource (http://www.imperial.ac.uk/research/animallectins/), Clec4f has a close evolutionary relationship with Langerin (CD207), as indicated by phylogeny analysis (Fig. 1A) [20]. Despite their high degree of relatedness, the protein length of Clec4f is 117 residues longer than Langerin (Fig. 1B). Clec4f was recently suggested to be the result of gene duplication of Langerin, which is expressed on Langerhans cells and binds to pathogens like human immunodeficiency virus (HIV) or opportunistic fungi such as *Candida* [21,22]. Although Clec4f and Langerin are expressed on distinct cell surfaces and represent different cell surface markers with potentially diverse biological functions, both mouse Clec4f and Langerin consist of an N-terminal cytoplasmic signaling domain, a hydrophobic single transmembrane helix, a heptad repeat neck domain, and a CRD (Figs 1B, 2A). A sequence alignment between mouse Clec4f and Langerin

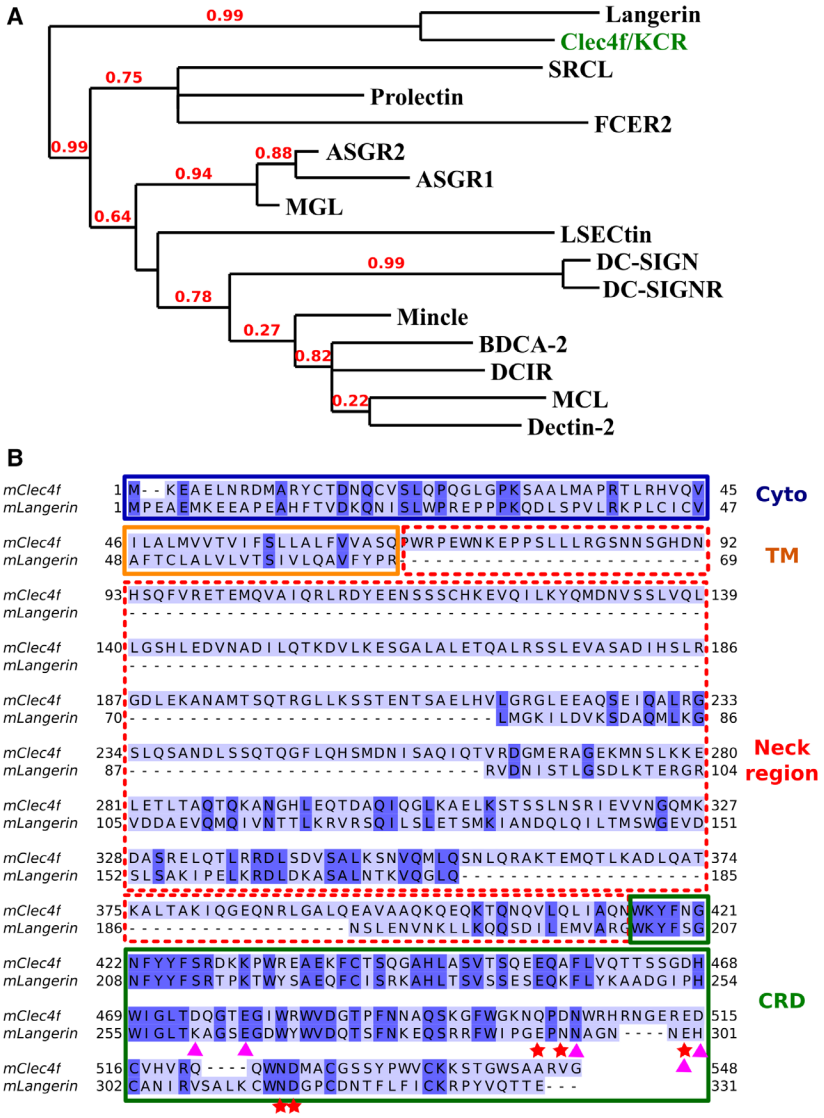

Fig. 1. Evolution and sequence analysis of mouse Clec4f. (A) Evolutionary relationship of mouse Clec4f within the Type II C-type lectin receptor family using the human source sequences from the Genomic Animal Lectin Resource (http://www.imperial.ac.uk/research/animallectins/) and generated using the Phylogeny.fr program based on maximum likelihood tree construction. Clec4f shows a close evolutionary relatedness to Langerin. Branch support values are colored red, and represent the evolutionary similarity between branches. (B) Sequence alignment between mouse Clec4f and Langerin. The carbohydrate-recognition domains (CRD), Neck region, transmembrane helix and cytoplasmic signaling domain are indicated, respectively. Although the CRD of Clec4f and Langerin are 54% identical, Clec4f contains a much longer neck region as compared to Langerin. The red stars and pink triangles represent residues involved in the first and second calcium binding, respectively.

demonstrated a sequence identity of 54% in the CRD (Fig. 1B) and a sequence identity of 25% in the cytoplasmic signaling domain. Both proteins lack the signature immunoreceptor tyrosine-based inhibition (ITIM) motifs observed in other members of the family (Fig. 1B). Particularly noteworthy, the heptad repeat neck domain of Clec4f is over 100 residues longer than in Langerin (Fig. 1B).

Crystal structure of trimeric mouse Clec4f

We designed, cloned, purified, and refolded recombinant truncated Clec4f spanning residues 389–581, including the whole CRD and part of the heptad repeat neck region (Fig. 2A). Dynamic light scattering (DLS) analysis indicates that recombinant truncated Clec4f exists as a stable trimeric form in solution, with an apparent diameter of 6.0 ± 0.2 nm and an apparent molecular weight of around 56 ± 2 kDa (theoretical monomer molecular weight: 18.3 kDa) (Fig. 2B). Characterization of glycan binding to recombinant Clec4f using Microscale Thermophoresis Assay revealed that supplementation with GalNAc molecules increased the melting temperature of Clec4f in a concentration-dependent manner (Fig. 2C). Plotting this effect as a function of GalNAc concentration indicated a Kd of 89 μM (Fig. 2C). Thus, the Clec4f trimer binds GalNAc with an affinity in the range of 10^{-5} M, similar to other lectins [23].

To gain insight into the assembly of trimeric mouse Clec4f, recombinant truncated $Clec4f_{389–581}$ was crystallized and the structure was solved to 2.8 Å in space

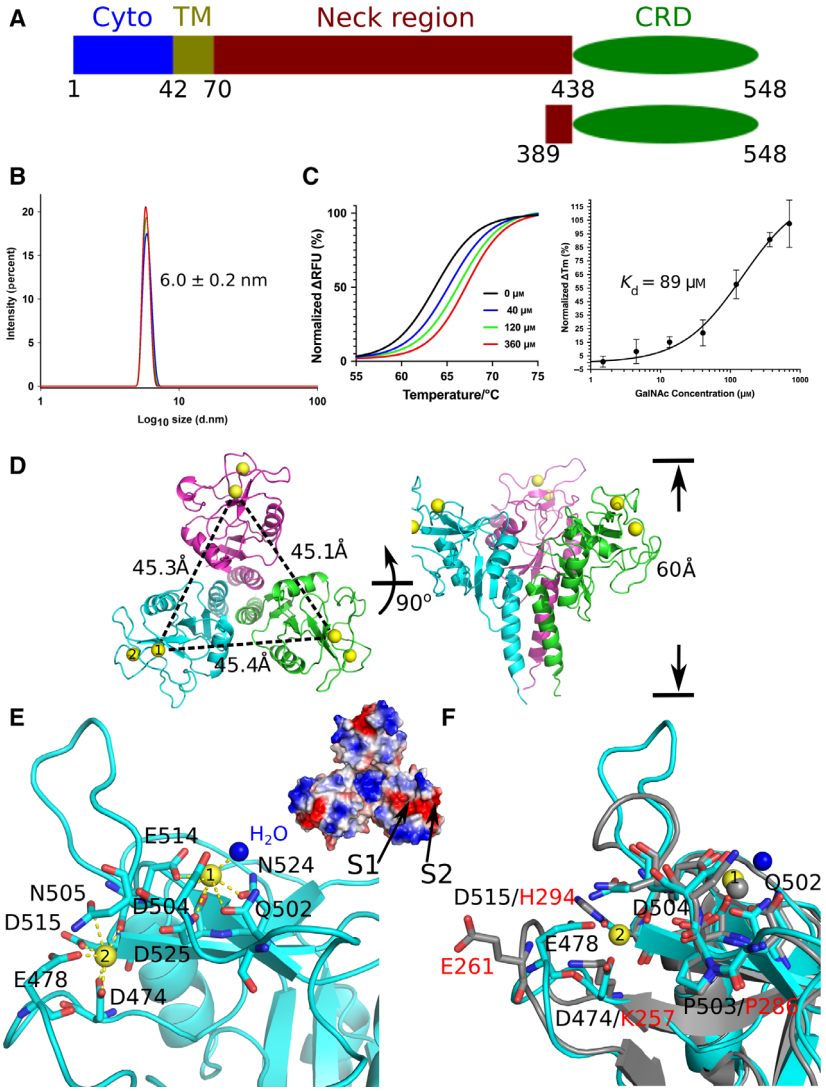

Fig. 2. Overall crystal structure of mouse Clec4f trimer. (A) Schematic representation of the domains present in full-length and truncated mouse Clec4f. (B) Dynamic light scattering characterization of truncated mouse Clec4f shows an assembly with a diameter of 6 nm. (C) Microscale Thermophoresis Assay of mouse Clec4f supplemented with varying concentrations of GalNAc. The left panel is a plot of normalized ΔRFU versus temperature, indicating that binding of GalNAc increases the melting temperature of Clec4f. The right panel is a plot of ΔTm versus GalNAc concentration, demonstrating a Kd of 89 μM. (D) Overall crystal structure of the trimeric mouse Clec4f in top and side view. Calcium ions are shown as yellow spheres, while the distance between two calcium ions in CRDs from neighboring monomers in the trimer is indicated using a dotted line. The primary and secondary binding Ca^{2+} are labeled with 1 and 2, respectively. The S1 and S2 indicated the charged pockets of the calcium binding. (E) The residues involved in calcium ion binding are shown as sticks, while the calcium ion and water molecule are represented as yellow and blue spheres, respectively. Top right: Charge smoothed electrostatic surface view of trimeric mouse Clec4f; the arrow indicates the positively charged calcium-binding pocket potentially involved in glycan recognition. (F) Alignment of the Ca^{2+} ion-binding motif between mouse Clec4f (cyan) and human Langerin (PDB: 3KQG) (gray). Labels of the residues of Clec4f and Langerin are shown in black and red, respectively. Residues D515 and D474 in the secondary calcium-binding site of Clec4f are replaced by H294 and K257 in Langerin, which strongly disfavor binding of calcium site, and cause residue E261 to be flipped outside the CRD. Residues P503 in Clec4f and P286 in Langerin are present in the cis-proline conformation at the center of the QPD and EPN motifs involved in calcium binding.

group P1 (Table 1, Fig. 2D). In the Clec4f structure, two Ca^{2+} ions are present in calcium-binding motifs that are relatively conserved within the C-type lectin family. This clearly differs from Langerin, which has only a single calcium-binding site (Fig. 2D,E) [24] that forms part of the primary glycan-binding site. The

corresponding Ca^{2+}-binding site in Clec4f recruits the sidechains of residues Q502, D504, E514, N524, D525, and one water molecule, which is presumably displaced upon glycan binding (Fig. 2E). This primary glycan-Ca^{2+} binding site is largely conserved between Clec4f and Langerin, except for the ^{502}QPD^{504} sequence in Clec4f that replaces the ^{285}EPN^{287} sequence found in Langerin (Fig. 2F). This difference has been reported to be crucial for different glycan-binding specificities in C-type lectins, with the QPD sequence favoring binding to galactose and GalNAc and the EPN sequence favoring mannose, glucose, and GlcNAc [25–28]. Interestingly, the central cis-proline residue in this motif (Q502-**P503**-D504 in mouse Clec4f and E285-**P286**-N287 in Langerin) plays an important structural role, and mutation of this proline to serine in the human clec4f gene is proposed to result in a loss of function, combined with an additional mutation that prevents the last exon of the CRD from being properly spliced [7].

The secondary Ca^{2+} ion-binding site of Clec4f involves the side chains of E478, D474, N505, D515, and the main chain carbonyl of E514. The sequence alignment shown in Fig. 1 demonstrates that in the second calcium-binding motif, D474 and D515 in mouse Clec4f are replaced by K257 and H294 in Langerin, respectively (Figs 2F and 1B), which strongly disfavors the coordination of a calcium ion. Langerin is thus similar to mouse DCIR2 and BDCA-2, which lack a secondary calcium ion-binding site, because residues typically coordinating the Ca^{2+} ion are replaced with basic amino acid side chains [29,30]. In Clec4f, the Ca^{2+} ion in the secondary binding site is unlikely to be directly involved in glycan binding, being completely coordinated by the protein. However, it does modify the structure of the region, which may contribute to the midpoint shift of the transition between the sugar-binding and inactive conformations [28].

Interaction interface forming trimeric Clec4f

The main interface responsible for forming Clec4f trimers is the coiled-coil structure found in the neck region. Interestingly, the coiled-coil structure of Clec4f is quite distinct from Langerin. Instead of forming a canonical hydrophobic leucine zipper as found in Langerin, MBP, and many other C-type lectins, the Clec4f trimer coiled-coil has a leucine/isoleucine core at the first heptad preceding the CRD, but is followed by a polyglutamine stack interface (Fig. 3A). Glutamine residues at positions A and D of the second heptad stack together through a network of hydrogen bonds and salt bridges (Fig. 3A,B). Similar polyglutamine coiled-coil structures are postulated in (Q/N)-rich prions and polyQ proteins involved in neurodegenerative disorders, where they are thought to contribute to the formation of pathogenic stable aggregates [31,32]. The glutamine residues present in the polyglutamine coiled-coil stack in mouse Clec4f are highly conserved between other species, including rat, pig and chimpanzee (Fig. 3C). Notably, in the region of the trimeric coiled-coil interface, the frequency of glutamine residues is as high as 40% (Fig. 3C). Besides the glutamine residues in the core of the coiled coil, there are many conserved glutamine and other polar residues lining the outer surface of the neck region as well.

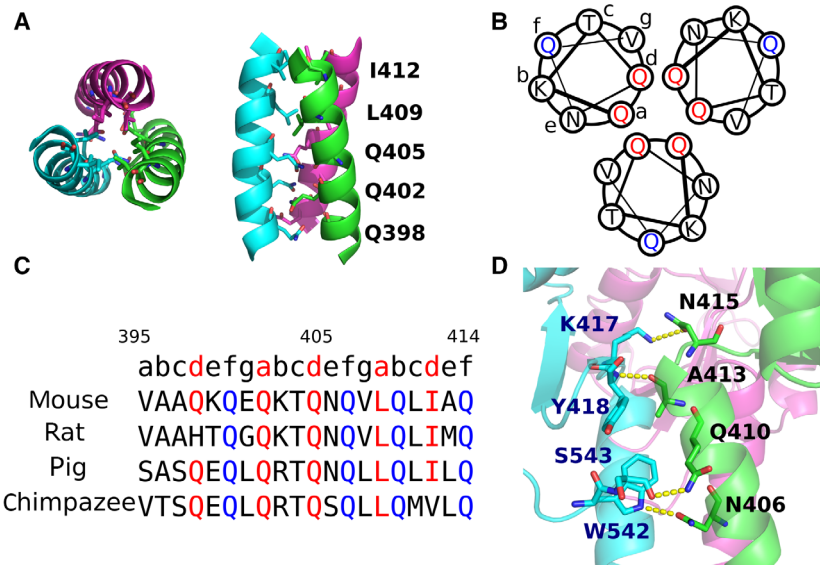

Fig. 3. Trimeric interface of mouse Clec4f. (A) Interaction interface of the coiled-coil neck region of mouse Clec4f in top and side view. (B) A helical wheel representation of the secondary heptad repeat reveals a glutamine stack interface within the coiled-coil structure. (C) Sequence alignment between the coiled-coil of Clec4f in the neck region of mouse, rat, pig, and chimpanzee demonstrates a strong conservation of the glutamine residues involved in the trimeric interface. (D) View of other interaction interfaces within the C terminus of Clec4f involved in the stability and orientation of Clec4f trimers.

Some of these form intermolecular interactions with the CRD of the adjacent monomer, priming and fixing the orientations between the CRDs and the central neck region. The coiled coil of the neck region is connected to the first β-strand of the CRD by a very short linker comprised of N415-W416. Inter-monomer hydrogen bonds are formed between side chains of K417 and N415 and between main chain atoms of Y418 and A413 (Fig. 3D). The C-terminal tail from the CRD also finds its way down to interact with the neck region, forming multiple inter-monomer hydrogen bonds, including S543-Q410 and W542-N406 interactions (Fig. 3D). The orientation between the CRD and neck region may be fixed, which would in turn fix the distances between glycan-binding sites within the trimer and potentially define Clec4f specificity for certain classes of glycosylated glycoproteins (based on the spatial relationships between individual sugars within each oligosaccharide) and recognition of particular pathogens. On the other hand, if the CRD is able to move relative to the neck region, this would broaden the range of oligosaccharides that Clec4f is able to recognize. Compared to the hydrophobic coiled-coil structure and more abundant set of hydrogen bonds and hydrophobic interactions between the CRD and neck region in the structure of trimeric Langerin [15], the interface forming the mouse Clec4f trimer displays fewer and more polar interactions within the neck region and between the neck region and CRD, possibly resulting in more flexibility between individual domains.

Comparison of Clec4f with other C-type lectins

The mouse Clec4f trimer structure we present here is the first oligomeric form of a galactose binding lectin, as far as we know. By comparison of the Clec4f CRD with the galactose binding lectin CRD monomeric structure of mouse SRCL [33], it can be seen that Clec4f contains a highly conserved galactose binding

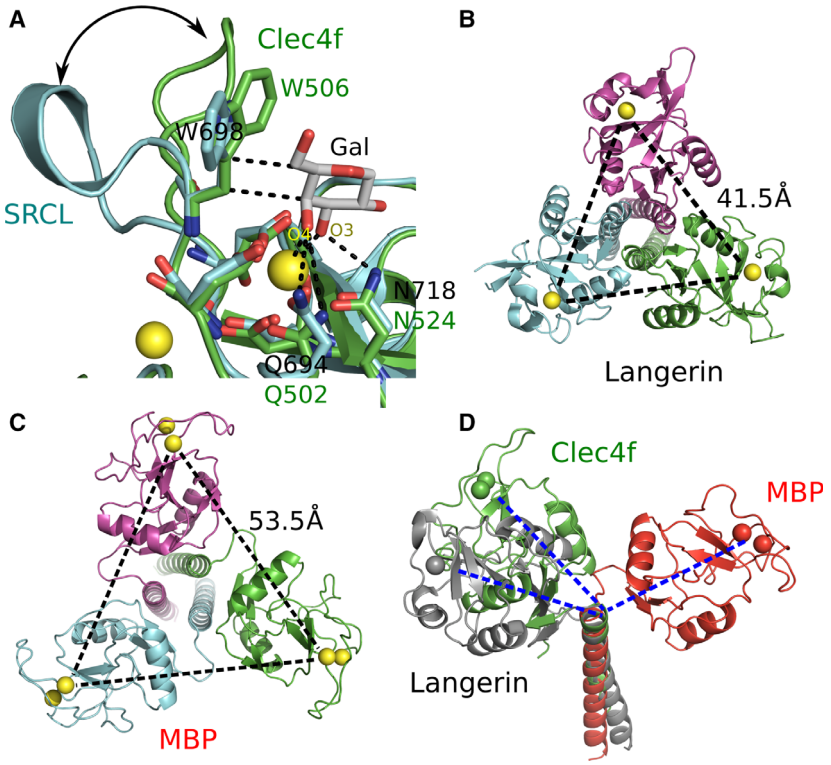

Fig. 4. Comparison of Clec4f trimer assembly to Langerin and other C-type lectin. (A) The alignment of Clec4f (green) with the galactose binding lectin mouse SRCL (PDB: 2OX9) (cyan) in complex with galactose. The residues involved in the calcium and galactose binding has been shown as stick and colored with O red, N blue, Ca2 + yellow. The dash line indicated the interaction between the galactose and the residues. The W506 in Clec4f and W698 in SRCL stack against the hydrophobic face of galactose. B and C represent the trimeric assembly of human Langerin (PDB: 3KQG) and rat MBP protein (PDB: 1BCH). Distances between the bound calcium ions within the trimer are indicated using dotted lines. (D) Structural alignment of the Clec4f, Langerin (PDB: 3KQG) and MBP coiled-coil neck region (PDB: 1BCH) demonstrates the distinct orientations between the CRD and the neck region. MBP has a much longer loop region between the coiled-coil neck and the first β strand resulting in a dramatically different orientation when compared to Clec4f and Langerin.

pocket (Fig. 4A). The residues involved in both calcium and galactose binding are highly conserved, with Q502 and N524 in Clec4f correlating with Q694 and N718 in SRCL, which were shown to interact with the O4 and O3 atoms in galactose, respectively (Fig. 4A). The aromatic Clec4f W506 residue also matches in position with W698 in SRCL, which stacks against the hydrophobic face of galactose and is conserved in human and mouse Gal/GalNAc binding C-type lectins (Fig. 4A). Interestingly, the loop in which this tryptophan residue is located demonstrates a distinct conformational flip in Clec4f when compared to SRCL (Fig. 4A), the significance of which is unknown at this time.

The CRDs in the Clec4f trimer structure deviate slightly from a perfect 3-fold symmetry, with distances between the primary glycan-binding Ca^{2+} being 45.1 Å, 45.3 Å, and 45.4 Å, respectively. The distance between the calcium-coordinating carbohydrate-binding sites is unique to each lectin receptor, thought to recognize pathogens selectively through multivalent binding to repetitive glycan patterns on specific pathogenic cell surfaces [34,35]. The distance between the carbohydrate-binding sites in mouse Clec4f is distinct from that of human Langerin (41.5 Å, Fig. 4B) and rat mannose binding protein (MBP) (53.5 Å, Fig. 4C). This distance can also be different in orthologs from different species. For instance, the primary Ca^{2+} binding distance in MBP is 53.5 Å in rat, while it has a shorter distance of 46.3 Å in human MBP [15,34,36]. Furthermore, a structural alignment of the coiled-coil neck region in Clec4f with the well-characterized Langerin and MBP reveals a rotation of the CRD (Fig. 4D). This orientation may also play a role in classification of the C-type lectin family, since MBP demonstrates a totally different orientation resulting from a longer loop between the neck region and the first β strand, while mouse Clec4f and Langerin have more similar orientations between the CRD and neck region (Fig. 4D).

Interestingly, although Clec4f is most related to Langerin, it has rapidly evolved features that make its functional properties quite different from Langerin and in fact more similar to other C-type lectins. First of all, its central neck region is more than 100 amino acids longer, likely extending its reach further away from the Kupffer cell surface or allowing it to penetrate deeper into the outer surface layer of target cells. The central coil region is more hydrophilic, particularly within the coiled coil core. The functional significance of this is unknown, although it may allow for more flexibility through transient dissociation of the coiled coil. The interactions between the CRD and central neck region are also different between Clec4f and Langerin, changing the distance and geometry between the glycan-binding sites of each monomer. Finally, the glycan-binding sites are very different. As previously pointed out, Clec4f has a classic QPD sequence that binds calcium and galactose, whereas Langerin has the classic EPN motif with different specificity. The surface area immediately surrounding the glycan-binding site is also very different in Clec4f. In Clec4f, there are a number of aromatic and positively charged residues in the immediate vicinity of the glycan-binding site: K500, W506, H508, H518, and R520. None of these residues are present in Langerin. Thus, the glycan-binding site in Clec4f harbors a positively charged patch with some hydrophobic character, suggesting that it may have an affinity for anionic glycans or glycolipids.

In conclusion, contrasting with a sequence comparison suggesting a high level of similarity between Clec4f and Langerin, our structural study reveals that they are quite different in terms of trimerization interface, distance between glycan-binding sites, and binding surface character. Taken together, our structural study of Clec4f sheds light on the understanding of evolution and diversity of type II C-type lectin receptors and will pave the way for further functional investigation of Clec4f.

Accession number

The structure factors and coordinates of trimeric mouse Clec4f have been deposited in the Protein Data Bank under the accession code 6JJJ.

Acknowledgements

We thank the staff from the BL18U1/BL19U1 beamline for technical support during data collection at the National Center for Protein Sciences Shanghai (NCPSS) at Shanghai Synchrotron Radiation Facility. This work was supported by the National Natural Science Foundation of China (NO. 31870132, NO. 81741088 and NO. 31500051 to YW).

Author contributions

YW and FZ designed and conducted the project. ZO, JZ, and YP carried out the protein overexpression, purification, and refolding experiments under the supervision of YW. YW and ZO collected the X-ray diffraction data. YW and JF solved the structure. ZO, BM, PH performed all other experiments. YW wrote the paper with contribution from all the authors.

References

1 Brown GD, Willment JA and Whitehead L (2018) C-type lectins in immunity and homeostasis. *Nat Rev Immunol* **18**, 374–389.

2 Haltiwanger RS, Lehrman MA, Eckhardt AE and Hill RL (1986) The distribution and localization of the fucose-binding lectin in rat tissues and the identification of a high affinity form of the mannose/N-acetylglucosamine-binding lectin in rat liver. *J Biol Chem* **261**, 7433–7439.

3 Lehrman MA, Pizzov SV, Imber MJ and Hill RL (1986) The binding of fucose-containing glycoproteins by hepatic lectins. *J Biol Chem* **261**, 7412–7418.

4 Coombs PJ, Taylor ME and Drickamer K (2006) Two categories of mammalian galactose-binding receptors distinguished by glycan array profiling. *Glycobiology* **16**, 1c–7c.

5 Fadden AJ, Holt OJ and Drickamer K (2003) Molecular characterization of the rat Kupffer cell glycoprotein receptor. *Glycobiology* **13**, 529–537.

6 Yang CY, Chen JB, Tsai TF, Tsai YC, Tsai CY, Liang PH, Hsu TL, Wu CY, Netea MG, Wong CH *et al.* (2013) CLEC4F is an inducible C-type lectin in F4/80-positive cells and is involved in alpha-galactosylceramide presentation in liver. *PLoS ONE* **8**, 1–14.

7 Taylor ME, Snelling T, Smith DF & Drickamer K (2018) Absence of a human ortholog of rodent Kupffer cell galactose-binding receptor encoded by the CLEC4f gene. *Glycobiology* **29**, 1–14.

8 Scott CL, Zheng F, De Baetselier P, Martens L, De Prijck S, Lippens S, Saeys Y, Abels C, Schoonooghe S, Raes G *et al.* (2016) Bone marrow-derived monocytes give rise to self-renewing and fully differentiated Kupffer cells. *Nat Commun* **7**, 1–10.

9 Zheng F, Sparkes A, De Baetselier P, Schoonooghe S, Stijlemans B, Muyldermans S, Flamand V, Van Ginderachter JA, Devoogdt N, Raes G *et al.* (2017) Molecular imaging with Kupffer cell-targeting nanobodies for diagnosis and prognosis in mouse models of liver pathogenesis. *Mol Imaging Biol* **19**, 49–58.

10 Tuin SJL van der, Li Z, Berbee immy FP, Verkouter I, Ringnalda LE, Neele AE, Klinken JB van, Rensen SS, Fu J, de Winther MPJ *et al.* (2018) Lipopolysaccharide lowers cholesteryl ester transfer protein by activating F4/80 + Clec4f + Vsig4 + Ly6C_ Kupffer Cell Subsets. *J Am Hear Assoc* **7**, e008105.

11 Lindquist ME, Zeng X, Altamura LA, Daye SP, Delp KL, Blancett C, Coffin KM, Koehler JW, Coyne S, Shoemaker CJ *et al.* (2018) Exploring Crimean-Congo hemorrhagic fever virus-induced hepatic injury using antibody-mediated type I interferon blockade in mice. *J Virol* **92**, JVI.01083-18.

12 Li Y, Fu J, Ling Y, Yago T, McDaniel JM, Song J, Bai X, Kondo Y, Qin Y, Hoover C *et al.* (2017) Sialylation on O-glycans protects platelets from clearance by liver Kupffer cells. *Proc Natl Acad Sci USA* **114**, 8360–8365.

13 Kabsch W (2010) Xds. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 125–132.

14 Mccoy AJ, Grosse-kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.

15 Feinberg H, Powlesland AS, Taylor ME and Weis WI (2010) Trimeric structure of Langerin. *J Biol Chem* **285**, 13285–13293.

16 Emsley P and Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* **60**, 2126–2132.

17 Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al.* (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 213–221.

18 Moonens K, Bouckaert J, Coddens A, Tran T, Panjikar S, De Kerpel M, Cox E, Remaut H and De Greve H (2012) Structural insight in histo-blood group binding by the F18 fimbrial adhesin FedF. *Mol Microbiol* **86**, 82–95.

19 Wienken CJ, Baaske P, Rothbauer U, Braun D and Duhr S (2010) Protein-binding assays in biological liquids using microscale thermophoresis. *Nat Commun* **1**, 1–7.

20 Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M *et al.* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**, 465–469.

21 De Witte L, Nabatov A, Pion M, Fluitsma D, De Jong MAWP, De Gruijl T, Piguet V, Van Kooyk Y and Geijtenbeek TBH (2007) Langerin is a natural barrier to HIV-1 transmission by Langerhans cells. *Nat Med* **13**, 367–371.

22 de Jong MAWP, Vriend LEM, Theelen B, Taylor ME, Fluitsma D, Boekhout T and Geijtenbeek TBH (2010) C-type lectin Langerin is a β-glucan receptor on human Langerhans cells that recognizes opportunistic and pathogenic fungi. *Mol Immunol* **47**, 1216–1225.

23 Goldstein IJ & Hayes CE (1978) The lectins: Carbohydrate-binding proteins of plants and animals. In Advances in Carbohydrate Chemistry and Biochemistry (Tipson RS & Horton D, eds), pp. 127–340. Academic Press, Cambridge, MA.

24 Feinberg H, Taylor ME, Razi N, McBride R, Knirel YA, Graham SA, Drickamer K and Weis WI (2011) Structural basis for langerin recognition of diverse

pathogen and mammalian glycans through a single binding site. *J Mol Biol* **405**, 1027–1039.
25 Drickamer K (1992) Engineering galactose-binding activity into a C-type mannose-binding protein. *Nature* **360**, 183–186.
26 Iobst ST and Drickamer K (1994) Binding of Sugar Ligands to Ca2+-dependent Animal Lectins. *Biochemistry* **269**, 15505–15511.
27 Kolatkar AR and Weis WI (1996) Structural basis of galactose recognition by C-type animal lectins. *Biochemistry* **271**, 6679–6685.
28 Drickamer K and Taylor ME (2015) Recent insights into structures and functions of C-type lectins in the immune system. *Curr Opin Struct Biol* **34**, 26–34.
29 Nagae M, Yamanaka K, Hanashima S, Ikeda A, Morita-Matsumoto K, Satoh T, Matsumoto N, Yamamoto K and Yamaguchi Y (2013) Recognition of bisecting N -acetylglucosamine structural basis for asymmetric interaction with the mouse lectin. *J Biol Chem* **288**, 33598–33610.
30 Jégouzo SAF, Feinberg H, Dungarwalla T, Drickamer K, Weis WI and Taylor ME (2015) A novel mechanism for binding of galactose-terminated glycans by the C-type carbohydrate recognition domain in blood dendritic cell antigen 2. *J Biol Chem* **290**, 16759–16771.
31 Grigoryan G and Keating AE (2008) Structural specificity in coiled-coil interactions. *Curr Opin Struct Biol* **18**, 477–483.
32 Fiumara F, Fioriti L, Kandel ER and Hendrickson WA (2010) Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* **143**, 1121–1135.
33 Feinberg H, Taylor ME and Weis WI (2007) Scavenger receptor C-type lectin binds to the leukocyte cell surface glycan Lewisx by a novel mechanism. *J Biol Chem* **282**, 17250–17258.
34 Weis WI and Drickamer K (1994) Trimeric structure of a C-type mannose-binding protein. *Structure* **2**, 1227–1240.
35 Ng KKS, Kolatkar AR, Park-Snyder S, Feinberg H, Clark DA, Drickamer K and Weis WI (2002) Orientation of bound ligands in mannose-binding proteins. Implications for multivalent ligand recognition. *J Biol Chem* **277**, 16088–16095.
36 Sheriff S, Chang CY and Ezekowitz RAB (1994) Human mannose binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil. *Nat Struct Biol* **1**, 789–794.