The structure of SALSA/DMBT1 SRCR domains reveal the conserved ligand-binding mechanism of the ancient SRCR-fold

Martin P. Reichhardt[1,*], Vuokko Loimaranta[2], Susan M. Lea[1,3] and Steven Johnson[1,*]

[1]Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, UK

[2]Institute of Dentistry, University of Turku, Turku, Finland

[3]Central Oxford Structural Molecular Imaging Centre, University of Oxford, Oxford, OX1 3RE, UK

[*]Correspondence: steven.johnson@path.ox.ac.uk or martin.reichhardt@path.ox.ac.uk

Abstract

The scavenger receptor cysteine-rich (SRCR) family of proteins comprise more than 20 membrane-associated and secreted molecules. Characterised by the presence of one or more copies of the ~110 amino acid SRCR domain, this class of proteins have widespread functions as anti-microbial molecules, scavenger- and signalling-receptors. Despite the high level of structural conservation of SRCR domains, no molecular basis for ligand interaction has been described. The SRCR protein SALSA, also known as dmbt1/gp340, is a key player in mucosal immunology. Based on detailed structures of the SALSA SRCR domains 1 and 8, we here reveal a novel universal ligand binding mechanism for SALSA ligands. The binding interface incorporates a dual cation binding site, which is highly conserved across the SRCR super family. Along with the well-described cation dependency on most SRCR domain-ligand interactions, our data suggest that the binding mechanism described for the SALSA SRCR domains is applicable to all SRCR domains. We thus propose to have identified in SALSA a conserved functional mechanism for ligand recognition by the SRCR class of proteins.

Keywords

SRCR / SALSA / dmbt1 / gp340 / mucosal immunology

Introduction

The Salivary Scavenger and Agglutinin (SALSA), also known as gp340, 'deleted in malignant brain tumors 1' (DMBT1) and salivary agglutinin (SAG), is a multifunctional molecule found in high abundance on human mucosal surfaces [1-4]. SALSA has widespread functions in innate immunity, inflammation, epithelial homeostasis and tumour suppression [5-7]. SALSA binds and agglutinates a broad spectrum of pathogens including, but not limited to, human immunodeficiency virus type 1, *Helicobacter pylori, Salmonella enterica* serovar *Typhimurium,* and many types of streptococci [8-11]. In addition to its microbial scavenging function, SALSA has been suggested to interact with a wide array of endogenous immune defence molecules. These include secretory IgA, surfactant proteins A (SP-A) and D (SP-D), lactoferrin, mucin-5B, and components of the complement system [1, 2, 12-18]. SALSA thus engages innate immune defence molecules and has been suggested to cooperatively mediate microbial clearance and maintenance of the integrity of the mucosal barrier.

The 300-400 kDa SALSA glycoprotein is encoded by the *DMBT1* gene. The canonical form of the gene encodes 13 highly conserved scavenger receptor cysteine-rich (SRCR) domains, followed by two C1r/C1s, urchin embryonic growth factor and bone morphogenetic protein-1 (CUB) domains that surround a 14th SRCR domain, and finally a zona pellucida (ZP) domain at the C-terminus [19, 20]. The first 13 SRCRs are 109 amino acid (aa) domains found as "pearls on a string" separated by SRCR interspersed domains (SIDs) (Figure 1a) [1, 21]. The SIDs are 20-23 aa long stretches of predicted disorder containing a number of glycosylation sites which have been proposed to force them into an extended conformation of roughly 7 nm [7]. In addition to this main form, alternative splicing and copy number variation mechanisms lead to expression of variants of SALSA containing variable numbers of SRCRs in the N-terminal region.

The SRCR protein super family include a range of secreted and membrane-associated molecules, all containing one or more SRCR domains. For a number of these molecules the SRCR domains have been directly implicated in ligand-binding. These include CD6 signalling via CD166, CD163 mediated clearance of the Haemoglobin-Haptoglobin complex, Mac-2 binding protein's (M2bp) interaction with matrix components, and the binding of microbial ligands by the scavenger receptors SR-A1, SPα and MARCO [22-27]. While the multiple SALSA SRCR domains likewise have been implicated in ligand-binding, the molecular basis for its diverse interactions remain unknown.

To understand the multiple ligand binding properties of the SALSA molecule, we undertook an X-ray crystallographic study to provide detailed information of the SALSA interaction surfaces. We here provide the atomic resolution structures of SALSA SRCR domains 1 and 8. We identify cation-binding sites and demonstrate their importance for ligand binding. By comparing our data to previously published structures of SRCR domains, we propose a generalised binding mechanism for this ancient, evolutionarily conserved, fold.

Methods

Expression of recombinant proteins

Insect cell expression. Codon-optimized DNA (GeneArt, ThermoFisher) was cloned into a modified pExpreS2-2 vector (ExpreS2ion Biotechnologies, Denmark) with a C-terminal His-6 tag. The purified plasmid was transformed into S2 cells grown in EX-CELL 420 (Sigma) with 25 µL ExpreS2 Insect-TR 5X (ExpreS2ion Biotechnologies). Selection for stable cell lines (4 mg/mL geneticin (ThermoFisher)) and expansion were carried out according to the manufacturer's instructions. **E. coli expression.** DNA strings (GeneArt, ThermoFisher) were cloned into pETM-14 and transformed into M15pRep cells. Protein expression was carried out in LB media (with 30 µg/mL kanamycin). Cells were induced with 1 mM IPTG. The cultures were centrifuged (3,220 x g, 15 min) and the cell pellets resuspended and lysed in PBS containing 1 mg/mL DNase and 1 mg/mL lysozyme.

Protein purification

SRCR-domains: Insect culture supernatant was collected by centrifugation (1000 x g at 30 min), filtered and loaded onto a Ni^{2+}-chromatography column (1 ml, Roche), washed in 20 CV buffer (50 mM Tris, pH: 9.0, 200 mM NaCl). Bound protein was eluted with 250 mM imidazole. Following this size-exclusion chromatography (SEC) was carried out on an S75 16/60 HR column(GE) equilibrated in 10 mM Tris, pH: 7.5, 200 mM NaCl. **Spy-2:** Lysed cell pellets were homogenized and centrifuged at 20,000 x g for 30 min. The filtered supernatant was loaded onto a Ni^{2+}-chromatography column (5 ml, Qiagen) and washed in 20 CV buffer (50 mM Tris, pH: 8.5, 200 mM NaCl, 20 mM imidazole). Bound protein was eluted with 250 mM imidazole, concentrated and subjected to SEC (S75 16/60 HR column, GE).

Crystallization, X-ray data collection, and structure determination

Purified SRCR1 and SRCR8 were concentrated to 20 mg/ml. SRCR1 was mixed with an equal volume of mother liquor containing 0.2 M $MgCl_2$ hexahydrate, 10 % (w/v) PEG8000, 0.1 M Tris, pH: 7.0, and crystallised in 400 nl drops by the vapor diffusion method at 21 °C. SRCR8 was mixed with an equal volume of mother liquor containing 0.1 M $LiSO_4$, 20 % (w/v) PEG6000, 0.01 M HEPES, pH: 6.5, and crystallised in 800 nl drops. For SRCR8 + cations crystals were grown in 0.2 M $MgCl_2$ hexahydrate, 20 % (v/v) isopropanol, 0.1 M HEPES, pH: 7.5, and crystallised in 400 nl drops. The crystallization buffer was supplemented with 10 mM Mg^{2+} and 10 mM Ca^{2+} 24 hours prior to freezing. All crystals were cryoprotected in mother liquor supplemented with 30% glycerol and flash frozen in liquid N_2. Data were collected at a temperature of 80 K on beamlines I03 and I04 at the Diamond Light Source (Harwell, UK), as specified in Table 1. The Structure of SRCR8 was solved by molecular replacement using MolRep within CCP4 [28] with the structure of CD6 SRCR domain 3 (PDB ID 5a2e [27]). The structure of SRCR1 and SRCR8 soaked in cations were solved by molecular replacement using the structure of SRCR8. Refinement and rebuilding was carried out in Phenix and Coot [29, 30]. The structures are characterized by the statistics shown in Table 1 with no Ramachandran outliers. Protein structure figures were prepared using Pymol Version 2.0 (Schrödinger, LLC).

Hydroxyapatite binding assay

150 µl Hydroxyapatite nanoparticle suspension (Sigma) was washed into buffer (10 mM HEPES, pH: 7.5, 150 mM NaCl, 1 mM Ca^{2+}). Beads were incubated in 80 ul SRCR8, SRCR8 D34A or SRCR8 D35A (all at 0.5 mg/ml) with shaking for 1 hour at RT. Beads were spun and washed 6x in 1 ml buffer. Bound protein was eluted in 100 µl 0.5 M EDTA, and visualized by SDS-PAGE (4-20 %, Biorad) and Coomassie staining (Instant Blue, Expedeon, UK).

Heparin binding assay

SRCR8, SRCR8 D34A or SRCR8 D35A in 10 mM HEPES, pH: 7.5, 10 mM NaCl, 1 mM Ca^{2+} were loaded onto a HiTrap Heparin HP column (1 ml, GE). Bound protein was then eluted with 10 mM HEPES, pH: 7.5, 10 mM NaCl, 20 mM EDTA.

Spy-2 binding assay

On a Maxisorp plate (Nunc, Denmark) 100 µl purified Spy-2 was coated O/N at 4 ^{o}C in a concentration ranging from 0.032 – 3.2 µM in coating buffer (100 mM $NaHCO_3$-buffer, pH: 9.5). The plate was blocked in 1 % gelatine in PBS, and SRCR8, SRCR8 D34A, and SRCR8 D35A were added (all at 7.1 µM in 10 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM Ca^{2+}, 0.05 % Tween20). Bound protein was detected with monoclonal anti-SALSA antibody diluted 1:10,000 (1G4, Novus Biologicals, UK) and HRP-conjugated rabbit anti-mouse antibody 1:10,000 (Promega, W4028). The plate was developed with 2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (Sigma), and analysed by spectrophotometry at 405 nm. To test calcium-specific dependency of the interaction the WT-assay above was repeated in a buffer containing: 10 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM Mg^{2+}, 1 mM EGTA, 0.05 % Tween20.

Data availability

Structure factors and coordinates from this publication have been deposited to the PDB database https://www.wwpdb.org, and assigned the identifiers: SRCR1 pdbid: 6sa4, SRCR8 pdbid: 6sa5, SRCR8 with calcium pdbid: 6san.

Results

The scavenger receptor SALSA has a very wide range of described ligands, including microbial-, host innate immune- and extracellular matrix (ECM)-molecules. To understand the very broad ligand-binding abilities of the SALSA molecule, we applied a crystallographic approach to determine the structure of the ligand binding SRCR domains of SALSA. SRCR domains 1 and 8 (SRCR1 and SRCR8) were expressed in *Drosophila melanogaster* Schneider S2 cells with a C-terminal His-tag. The domains were purified by Ni-chelate and size-exclusion chromatography, crystallised, and the structures were solved by molecular replacement. This yielded the structures of SRCR1 and SRCR8 at 1.77 Å and 1.29 Å, respectively (Fig 1). For crystallographic details, see Table 1.

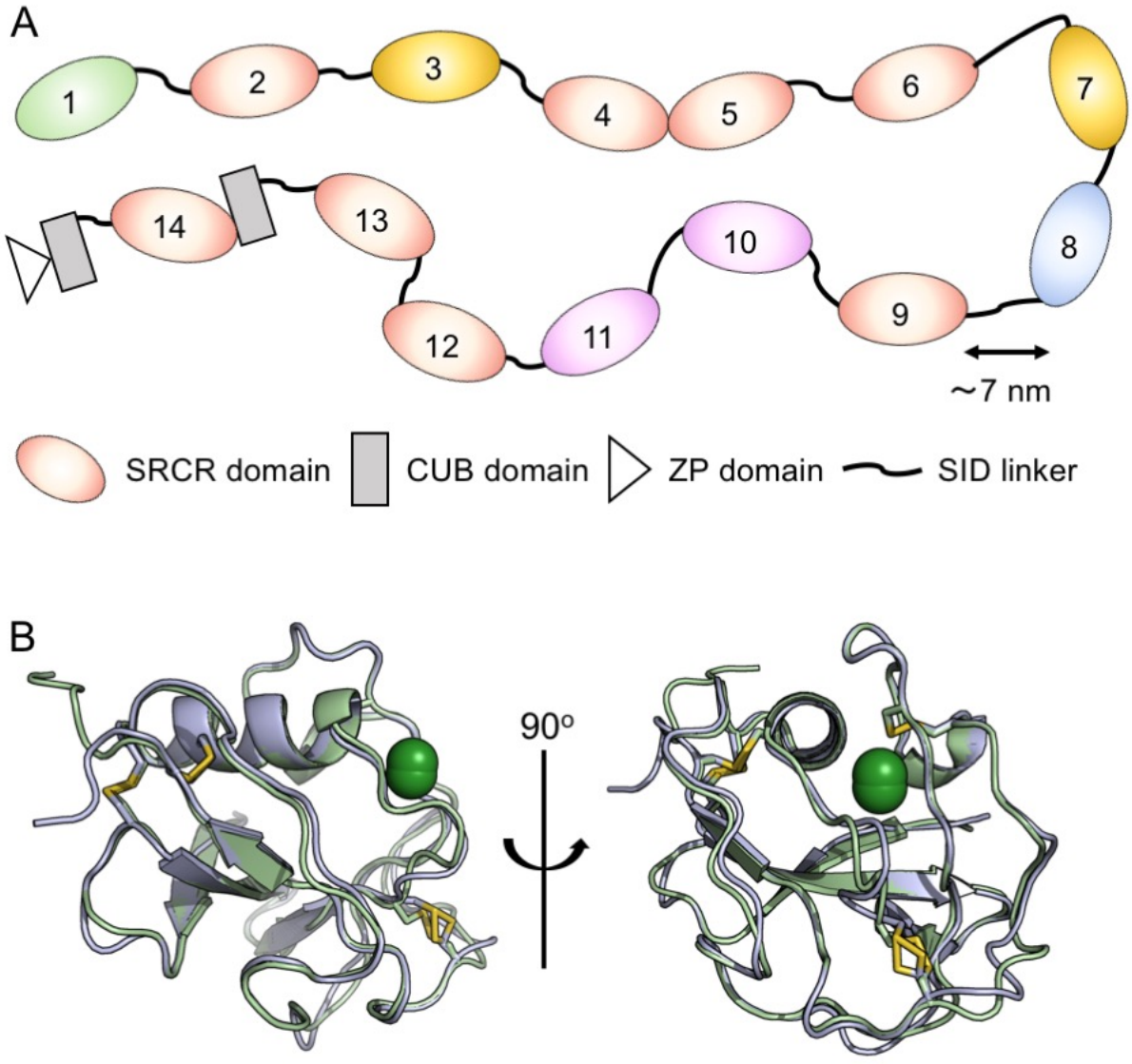

Figure 1: Crystal structure of SALSA domains SRCR1 and SRCR8. (A) Schematic representation of the domain organization of full-length SALSA. SRCR1 and SRCR8 are highlighted in green and blue, respectively. All SRCR domains share > 88 % sequence identity. 100 % identity is shared by SRCR3 and 7 (yellow) and SRCR10 and 11(purple). **(B)** Front and side views of an overlay of SRCR1 (green) and SRCR8 (blue), showing 4 conserved disulphide bridges (yellow). Both SRCR1 and SRCR8 were found to coordinate a metal, modelled as Mg^{2+} (dark green). The limited structural variation observed between SRCR1 and SRCR8 (92 % sequence identity) imply that these are appropriate representations of all SALSA SRCR domains 1-13.

Table 1 Data collection and refinement statistics (molecular replacement)

	SRCR1 (pdbid 6sa4)	SRCR8 (pdbid 6sa5)	SRCR8MgCa (pdbid: 6san)
Data collection			
Space group	P 21 21 21	P 21 21 21	P 1 21 1
Cell dimensions			
a, *b*, *c* (Å)	36.77, 45.19, 69.37	32.82, 40.82, 62.99	27.24, 46.64, 93.63
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 97.37, 90.00
Resolution (Å)	28.52-1.77 (1.80-1.77)	40.82-1.29 (1.31-1.29)	46.64-1.36 (1.39-1.36)
R_{merge}	0.17 (1.36)	0.117 (1.12)	0.074 (0.801)
$I / \sigma I$	8.1 (1.1)	8.6 (0.9)	14.9 (2.2)
Completeness (%)	99.8 (99.3)	100 (99.6)	98.3 (96.9)
Redundancy	6.3 (6.6)	11.4 (8.0)	6.6 (6.4)
Refinement			
Resolution (Å)	28.52-1.77 (1.95-1.77)	34.27-1.29 (1.35-1.29	30.95-1.36 (1.39-1.36)
No. reflections	11730	21958	49168
R_{work} / R_{free}	0.186/0.229 (0.266/0.348)	0.155/0.188 (0.326-0.279)	0.196/0.245 (0.326-0.265)
No. atoms			
Protein	824	829	1675
Ligand/ion	26	18	32
Water	109	124	271
B-factors			
Protein	22.36	16.80	18.04
Ligand/ion	49.33	54.03	30.77
Water	31.15	33.66	36.69
R.m.s. deviations			
Bond lengths (Å)	0.006	0.009	0.005
Bond angles (°)	0.809	1.026	0.78

*Number of xtals was one for each structure. *Values in parentheses are for highest-resolution shell. Data from SRCR1 and SRCR8 crystals were collected on Diamond beamline I04, while data for SRCR8MgCa were collected on beamline I03.

The SALSA SRCR domains reveal a classic globular SRCR-fold, with 4 conserved disulphide bridges, as described for the SRCR type B domains. The fold contains one alpha-helix and one additional single helical turn. The N- and C-termini come together in a four-stranded beta-sheet. SRCR1-13 are highly conserved, with 88-100 % identity. Variation is only observed in 9 out of the 109 amino acid residues, all of these observed in peripheral loops, without apparent structural significance. Combined, the data from SRCR1 and SRCR8 are thus valid representations of all SALSA SRCR domains. Both SRCR1 and SRCR8 are stabilized by a metal-ion buried in the globular fold. The placement suggests the ion is bound during the folding of the domain, and is modelled as Mg^{2+}, which is present in the original expression medium and in the crystallization conditions of SRCR1.

So far, all described ligand-binding interactions of SALSA have been shown to be Ca^{2+}-dependent. We therefore proceeded to address the cation binding potential of the SRCR domains in further detail. Crystals were grown of purified SRCR8, and Ca^{2+} and Mg^{2+} were soaked in before freezing. This yielded the crystal structure of SRCR8 with the original Mg^{2+} ion, site 1, and two additional cations bound, sites 2 and 3, SRCR8MgCa (Fig 2). All three sites are class three cation binding sites, with the coordination obtained from residues in distant parts of the sequence [31].

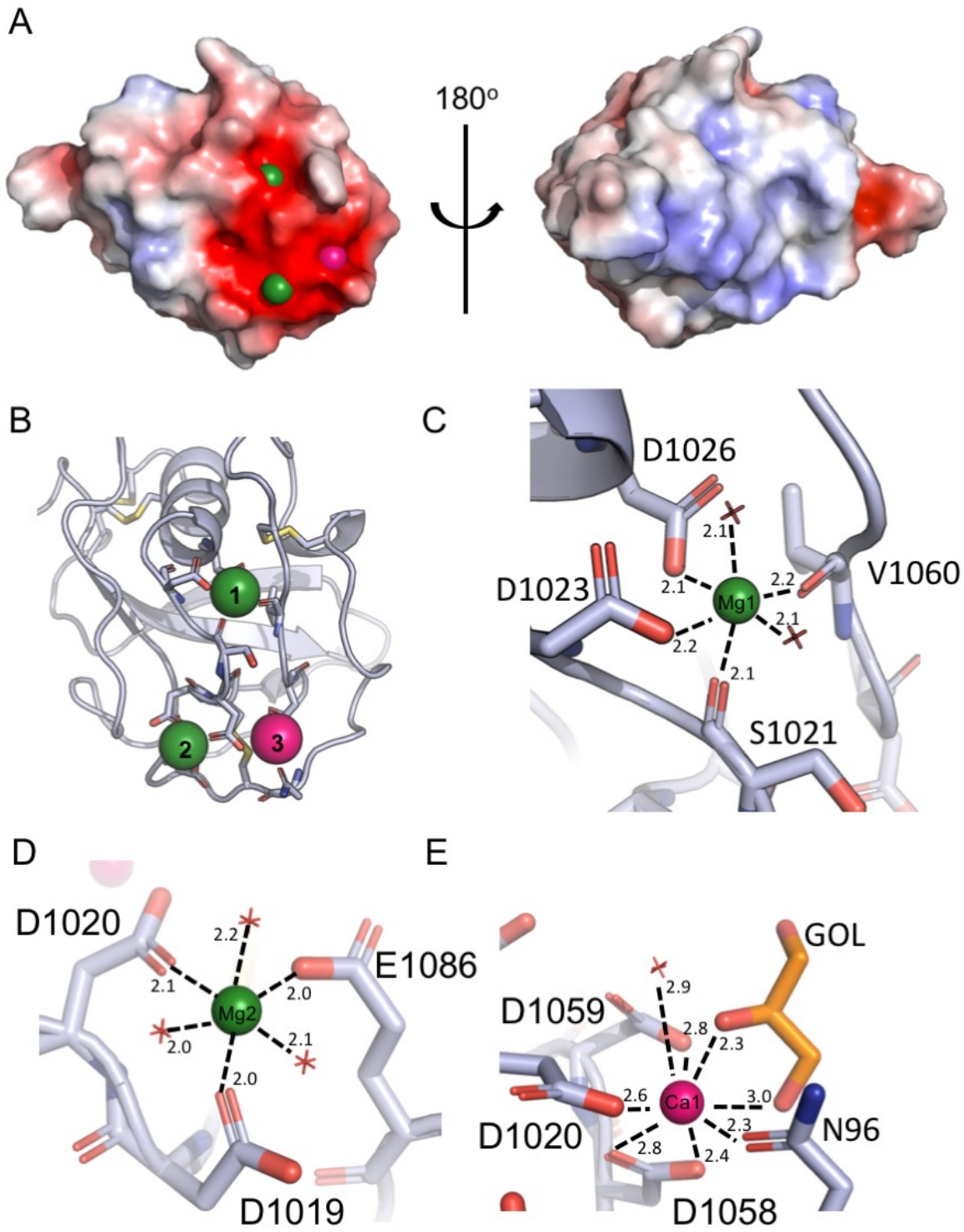

Figure 2: Crystal structure of SRCR8MgCa bound indicates mechanism of ligand-binding. (**A)** Surface charge distribution of SRCR8MgCa (calculated without the presence of cations) shows a positive cluster on one side with a strong negative cluster on the other. The negative cluster expands across approximately 300 Å^2 and mediates the binding of three cations. **(B)** Representation of the residues coordinating the three cations. The upper Mg^{2+}, site 1, sits somewhat buried in the structure, and may be essential for structural stability. The lower cations at sites 2 and 3 are more exposed. **(C)** Detailed view of the coordination of the upper Mg^{2+}, site 1. The coordination number of six is achieved by two waters, two backbone carbonyls and two sidechain carboxylates. **(D)** Detailed view of the coordination of the cation at site 2, modelled as Mg^{2+} based on bond length and coordination number. Here the coordination number of six is achieved by three waters and three sidechain carboxylates. **(e)** Detailed view of the coordination of the cation at site 3, modelled as Ca^{2+}, based on bond length and the higher coordination number. The coordination number of eight is achieved by one water, four sidechain carboxylates, one sidechain amide and a glycerol.

A surface charge analysis of the SRCR domain showed a strong electronegative surface generated by a number of loop extensions. Soaking of the SRCR8-crystals in cations revealed that this surface binds three cations, a Mg^{2+} at the base of the alpha helix (also observed in the initial SRCR1 and SRCR8 structures), and two in close proximity to each other, tying the superficial loop at aa1078-1085 to the rest of the domain. The Mg^{2+} at site 1 is coordinated by the backbone carbonyl groups of S1021 and V1060, as well as the side chains of D1023 and D1026, and 2 waters, and is buried in the domain fold. The Mg^{2+} at site 2 is coordinated by D1019, D1020 and E1086 plus 3 waters, while the Ca^{2+} at site 3 is coordinated by the sidechains of D1020, D1058, D1059 and N1081 with additional contributions from a water and a glycerol molecule from the cryoprotection. In contrast to the Mg^{2+} at site 1, these cations are exposed on the surface of the domain. Assignment of the ions at the paired site was carried out based on analysis of the bond lengths and coordination number of the two sites. Mg^{2+}- bonds are typically shorter than those of Ca^{2+}, and Ca^{2+} is often observed to have a higher coordination number [32]. However, it is worth noting that either site could accommodate either cation depending on local concentration.

A key feature of the cation binding sites 2 and 3 is that the protein only contributes a fraction of the coordination sphere, with the remainder contributed by waters or hydroxyl containing small molecules from the crystallisation solution. According to the literature, the majority of described SALSA ligands are negatively charged. Thus, the surface-exposed cations likely provide a mechanism for ligand binding for the SALSA SRCR domains, whereby the anions of the ligand substitute for the waters or the glycerol at site 3 observed in our structure. To test this hypothesis, site directed mutagenesis was employed targeting the key residues coordinating site 2 and 3. Included in the further analysis were single mutations D1019A and D1020A. While mutation of D1019 is expected to only disrupt binding of cations at site 2, mutation of the shared D1020, will likely affect binding of both cations.

As SALSA recognizes a very broad range of biological ligands, we set out to test the effect of SRCR domain point mutations on interactions with a wide array of biological ligands. These included binding to: (1) hydroxyapatite, a phosphate rich mineral essential for the binding of SALSA to the teeth-surface, where it mediates anti-microbial effects [33]; (2) heparin, a sulphated glycosaminoglycan as a mimic for the ECM/cell surface, for which binding of SALSA has been described to affect cellular differentiation as well as microbial colonisation [34]; (3) Group A Streptococcus surface protein, Spy0843, a Leucine-rich repeat (Lrr) protein demonstrated to bind to SALSA [35] (Fig 3).

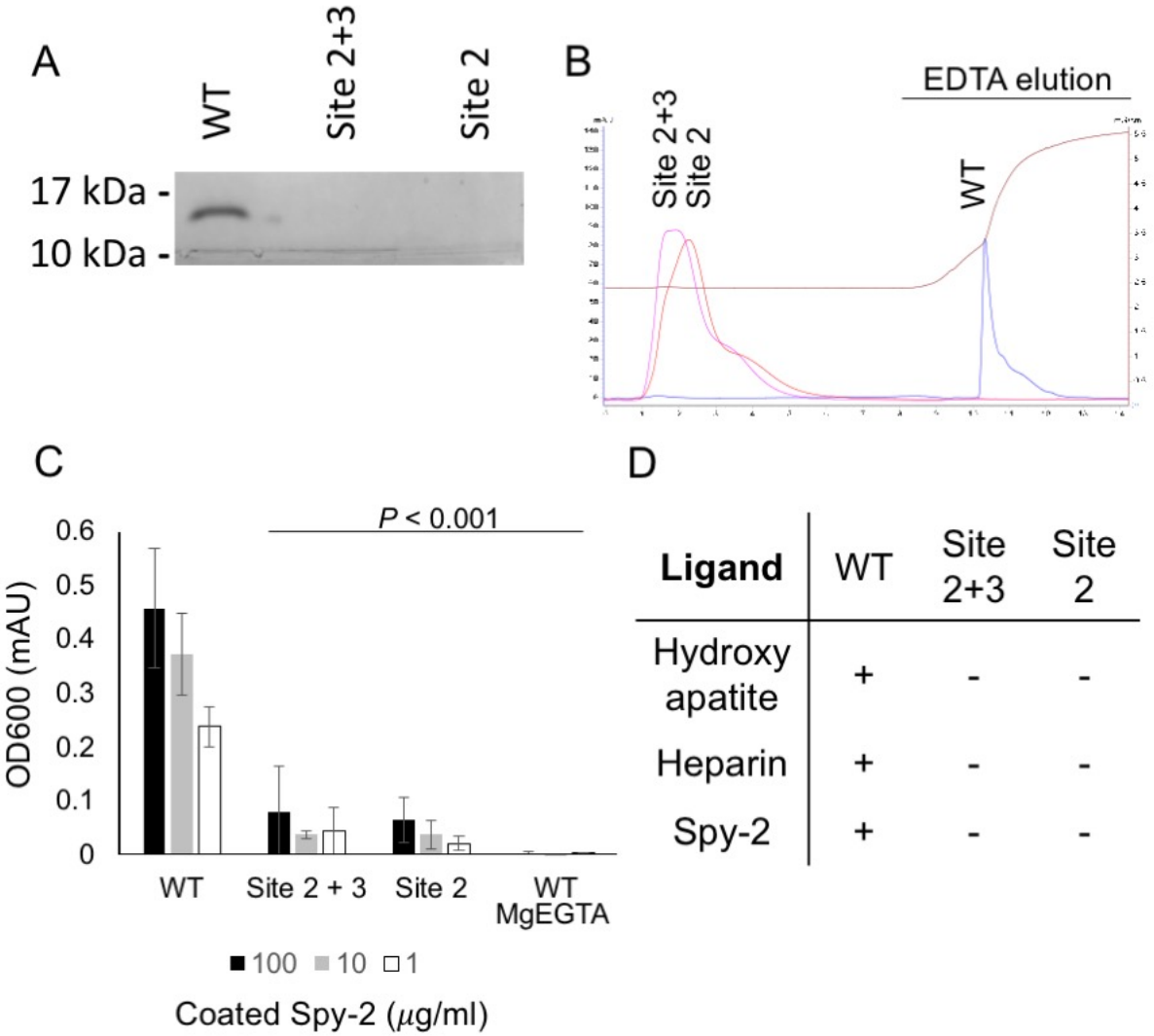

Figure 3: Mutating the cation-binding residues of SRCR domains abolish function. Through multiple ligand binding assays, we demonstrated the functional importance of calcium binding by the SRCR domains. Mutations affecting site 2 (D1019A) and mutations affecting sites 2 and 3 (D1020A) both abolish function. **(A)** WT and mutant forms of SRCR8 were incubated with hydroxyapatite beads in a Ca^{2+} containing buffer. After extensive washing, bound protein was eluted with EDTA. Eluted fractions were run on a 4 - 20 % SDS-PAGE gel and visualized by Coomassie staining. Only WT SRCR8 bound hydroxyapatite. **(B)** WT and mutant forms of SRCR8 were flown over a Heparin (HiTrap HP, 1 ml) column in a Ca^{2+}-containing buffer. Protein bound to the column was eluted with 0.5 M EDTA. Only WT SRCR8 bound the Heparin-column. Traces: SRCR8 (blue), D1020A (pink), D1019A (red), conductivity (brown). **(C)** In an ELISA-based setup, a concentration range of the Spy-2 domain of Spy0843 was coated (0.032 – 3.2 μM). WT and mutant SRCR8 domains were added in molar excess (7.1 μM), and binding was detected with a monoclonal anti-SALSA antibody. Binding was only observed for WT SRCR8. **(D)** Overview of ligand binding studies, + denotes binding, - denotes no binding.

The different binding assays provide an understanding of a generalized binding mechanism of the SALSA SRCR domains. While the WT SRCR domain bound to all three ligands, both of the cation binding site mutations, D1019A and D1020A, abolished binding. This is consistent with the bound cations acting as a bridge for ligand interaction, and thus provides a mechanistic explanation for the binding properties of the SALSA SRCR domains. In the literature, SALSA ligand binding has been described as specifically calcium-dependent. To verify this, we conducted binding assays in a MgEGTA containing buffer (Figure 3C). The exchange of magnesium for calcium abolished ligand interactions. This confirms the calcium-specific mediation of binding. As mutation of site 2 alone (modelled as Mg^{2+} in our

structure) abolish binding, our data suggest that Ca^{2+} will occupy both site 2 and site 3 under physiological conditions.

All known members of the SRCR-super family share a very high degree of identity, both at sequence and structural level. An FFAS search [36] of the SRCR8 sequence showed highest similarities to CD163 SRCR5 (score: -65.4, 46 % identity), Mac-2-binding protein (M2bp) (score: -64.4, 54 % identity), neutrotrypsin (score: -61.5, 50 % identity), MARCO (score: -59.4, 50 % identity), CD5 SRCR1 (score: -48.8, 26 % identity), and CD6 SRCR2 (score: -45.6, 60 % identity). Using the Dali server [37], searches for the cation-binding SRCR8MgCa structure identified the two top hits as M2bp (pdbid: 1by2) and CD6 SRCR3 (pdbid: 5a2e). These were identified with respective Z-scores of 21.4 (r.m.s.d. of 1.1 Å with 106 out of 112 residues aligned) and 20.4 (r.m.s.d. of 1.5 Å with 109 out of 109 residues aligned). Despite the classical division of SRCR-super family proteins into group A and B, based on the conserved 3 versus 4 cysteine bridges, the SRCR fold is very highly conserved, and the SALSA SRCR domain structures correlate closely to both group A and group B SRCR-super family domains (Fig 4A).

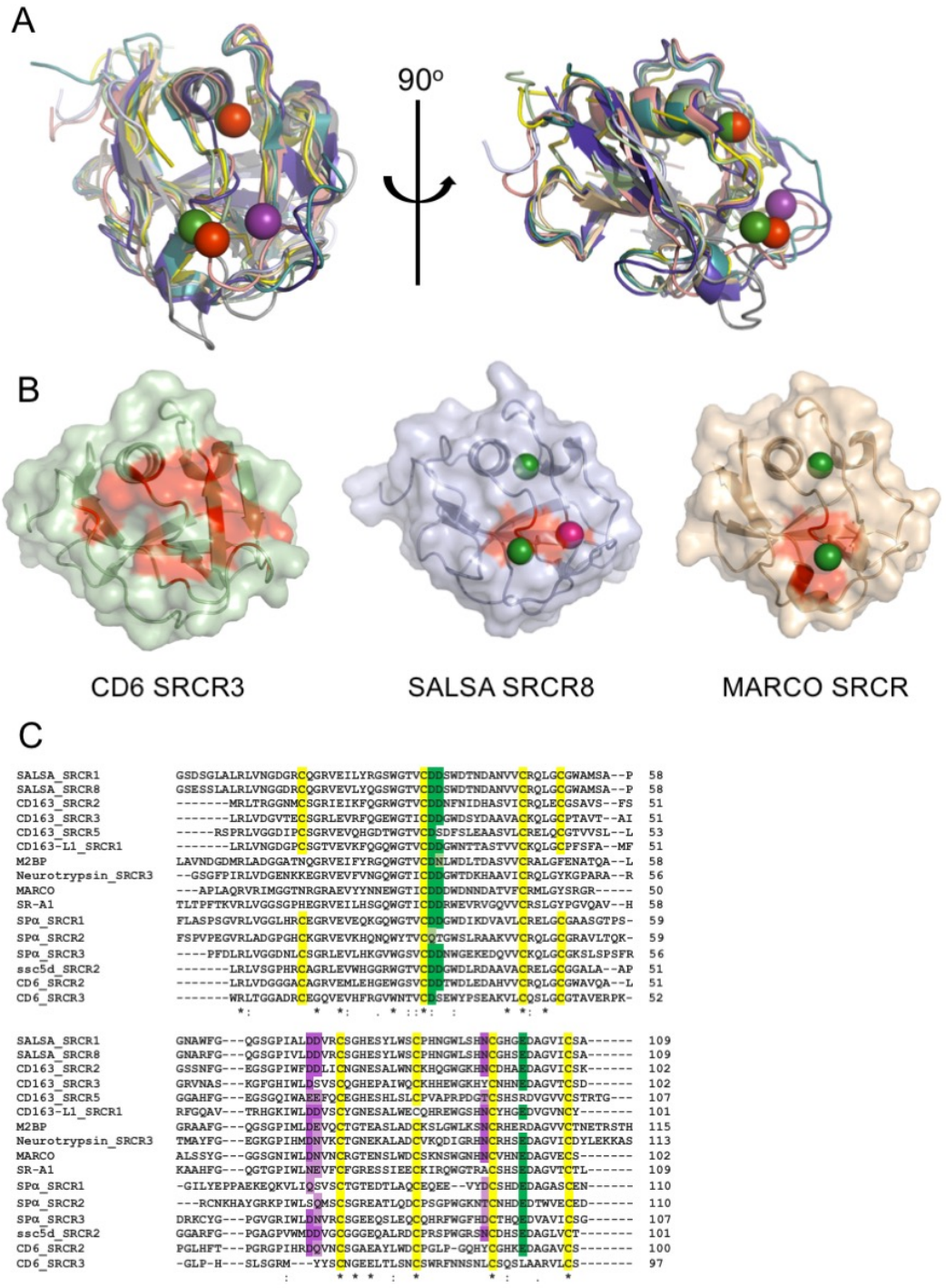

Figure 4: Conserved ligand binding motif across SRCR domains. SRCR-domains from multiple proteins engage in cation-dependent ligand interactions. **(A)** Structural overlay of domains from seven SRCR-super family proteins, all with ligand binding mediated through the SRCR domain. This reveals a highly conserved fold across both type A and type B SRCR domains. SRCR1 (pale green), SRCR8MgCa (light blue), MARCO (pdbid: 2oy3, sand), CD163 (pdbid: 5jfb, purple), CD5 (pdbid: 2OTT, grey), CD6 (pdbid: 5a2e, pink), M2bp (pdbid: 1by2, yellow), and murine neurotrypsin (pdbid: 6h8m, teal) [38]. SALSA Magnesium: dark green, MARCO Magnesium: red, Calcium: dark purple. **(B)** Surface representation of CD6 SRCR3, MARCO and SALSA SRCR8 in same orientation. Point mutations with a verified impact on ligand-binding are highlighted in red, indicating a conserved surface involved in ligand binding. **(C)** Clustal Omega (EMBL-EBI) sequence alignment of SRCR domains from ten SRCR-super family proteins. Conservation of the cation binding sites are displayed in green (site 2) and purple (site 3). Dark colouring indicates 100 % identity with the SALSA sites, and lighter colouring indicates conservation of residues commonly implicated in calcium-binding (D, E, Q or N). Cysteines are highlighted in yellow, and overall sequence identity is denoted by: ***** (100 %), **:** (strongly similar chemical properties) and **.** (weakly similar chemical properties).

For members of the SRCR-super family where the SRCR domain directly partake in ligand binding, both microbial and endogenous protein ligands have been described. For MARCO, crystallographic structures identified a cation-binding site exactly corresponding to site 2 in the SALSA SRCR8 domain [26] and point mutations of this site abolished function. Common to the MARCO and SALSA SRCR domains is the cluster of negatively charged residues coordinating the functionally important cations. A similar cluster is also observed in the SRCR3 domain of CD6, and have been show by mutagenesis to be directly involved in binding to the human surface receptor CD166. Indeed, a point mutation of D291A (corresponding to D1019 of SALSA-SRCR8) reduced the ligand-binding potential of CD6 to less than 10 % [27]. Furthermore, mutational studies of SRCR domains 2 and 3 from CD163 proved an involvement of this specific site in binding of the haemoglobin-haptoglobin complex [39]. All structural evidence from mutational studies of SRCR domains thus indicate a conserved surface mediating ligand binding (Fig 4B).

Interestingly, various levels of calcium-dependency on ligand interactions have been described for all SRCR domains directly involved with binding. SR-A1, Spα, MARCO, CD5 and CD6 all rely on calcium for interactions with microbial ligands [24-26, 40-42]. Furthermore, the binding of CD163 to the haemoglobin-haptoglobin complex is calcium-dependent, while CD6 also recognize endogenous surface structures (other than CD166) in a calcium-dependent manner [43]. This suggests that the binding mechanism identified for SALSA is a general conserved feature of all SRCR domains. Indeed, sequence alignment of SRCR domains from ten different SRCR-super family proteins, all with SRCR domains directly involved with ligand binding, reveal a very high level of conservation of the two cation binding sites identified in the SALSA domains (Fig 4C). A search with the SRCR8 model on the Consurf server [44], show that D1019, D1020, D1058 and E1086 all score 7 (out of 9, highly conserved), while N1081 and D1059 score 6 and 4, respectively. Whenever sequence identity is not conserved, substitutions are observed with other residues overrepresented in calcium-binding sites (D, E, Q and N) [31, 32]. The cation binding sites identified in the SALSA SRCR-domains, thus appear to be a highly conserved feature of the general SRCR-fold. Structural inspection of the overlay of SALSA SRCR8 with the corresponding region in the other known SRCR-domain structures, clearly show the potential for cation binding at these sites. We thus propose that the cation-binding sites identified here are an essential feature of the ancient SRCR-fold, and is a conserved mechanism responsible for mediating ligand-binding in the class of SRCR-super family proteins.

Discussion

While SALSA has previously been described to interact with a wide range of biological ligands, little has been known of the binding mechanisms. Furthermore, it has not been known if SALSA interacts with the various ligands in a similar way, or if distinct binding sites are utilized. Here we demonstrate that mutations of a dual cation-binding site interrupts interactions with a representative selection of very different types of ligands. Specific disruption of site 2 was sufficient to abolish ligand binding. We modelled the cation at site 2 in our crystal as Mg^{2+}, based on an analysis of bond length, coordination number and behaviour of crystallographic refinements with different cations modelled. However, experimental data demonstrated that binding to the ligands tested was only dependent on the presence of Ca^{2+}, and not Mg^{2+}. This is in line with previous descriptions of most SALSA-ligand interactions [5, 6, 45]. The generation of the SRCR8-crystal form with multiple cations present, was achieved by soaking the crystals in 10 mM Mg^{2+} and 10 mM Ca^{2+}. In the extracellular compartment, however, the molar concentration of Ca^{2+} is higher than Mg^{2+}, and it is therefore likely that both sites 2 and 3 will be occupied by Ca^{2+} in a physiological setting. The identification of this dual Ca^{2+}-binding site thus provides an explanation for the Ca^{2+}-dependency of all SALSA-ligand interactions described in the literature, suggesting this mechanism of binding is applicable to all SALSA SRCR ligands. Multiple studies have proposed a role for the motif GRVEVxxxxxW in ligand binding [46-48]. The crystal data show that this peptide sequence is buried in the SRCR fold, and we found no validation for a role in ligand binding although mutations within this sequence are likely to perturb the overall fold. This motif thus does not appear to have any physiological relevance as defining a ligand-binding site.

The conserved usage of a single ligand-binding area for multiple interactions, suggests that each SALSA SRCR domain engages in one ligand interaction. A common feature of the ligands described here, as well as a number of other ligands such as DNA and LPS etc., is the presence of repetitive negatively charged motifs [34].We analysed ligand binding by individual SRCR domains in surface-plasmon resonance and isothermal calorimetry assays, but interactions were observed to be very low affinity, making reliable measurements unfeasible. This is not surprising for a molecule such as SALSA, where the molecular make-up with the full extension of 13 repeated units, interspersed by predicted non-structured flexible SIDs, provide a molecule that can generate high avidity interactions with repetitive ligands, despite having only low affinity interactions for an individual domain. Furthermore, it has been suggested that SALSA in body secretions may oligomerize into larger complexes [5, 49-51], probably via the C-terminal CUB and ZP domains. The repetitive nature, and possible oligomerization, allow SALSA to not only engage with a repetitive ligand on one surface (e.g. LPS or Spy0843 on microbes), but also engage in multiple ligand interactions simultaneously. This would be relevant for its interactions with other endogenous defence molecules, such as IgA, surfactant proteins and complement components, where a cooperative effect on microbial clearance has been demonstrated [12, 16, 17, 52]. In addition, this model of multiple ligand-binding would be relevant for microbes described to utilize SALSA for colonization of the teeth or the host epithelium [10, 53, 54] (Fig 5).

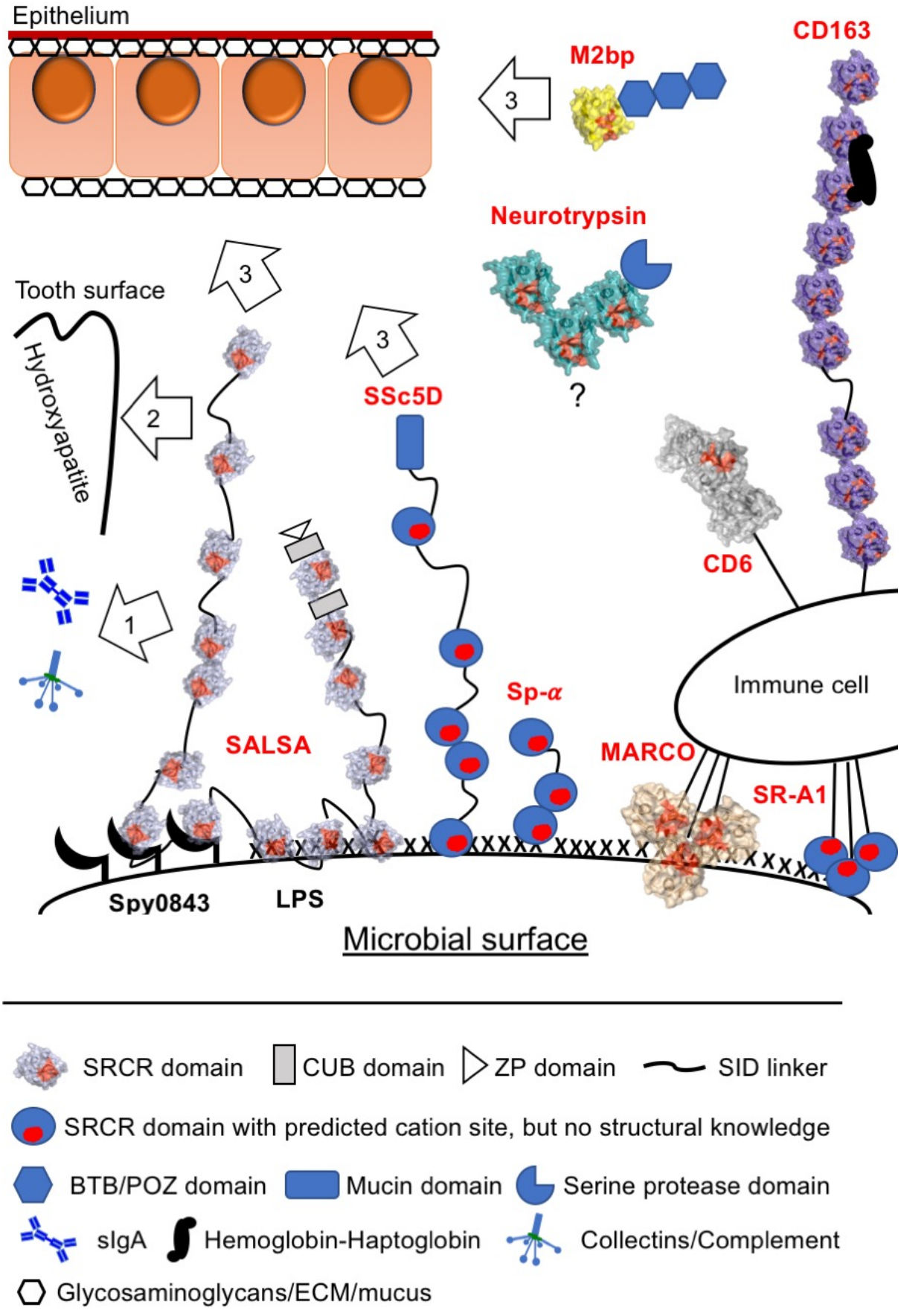

Figure 5: The SALSA SRCR cation binding motif reveal a conserved mechanism for broad-spectrum ligand interactions of SRCR-super family molecules. Based on mutational studies and structural information across SRCR proteins, we propose a generalised mechanism of ligand interaction mediated by the cation-binding surface motif of the evolutionarily ancient SRCR fold. **(Left side)** SALSA has been described to bind a broad range of ligands, incorporating into a complex network of binding-partners on the body surfaces as well as the colonizing microbiota. The multiple SRCR domains of full-length SALSA bind repetitive targets (both protein- and carbohydrate-structures) on the surface of microbes. The secreted fluid-phase molecule may thus lead to microbial agglutination and clearance. However, the repetitive form of binding-sites will allow for simultaneous binding to endogenous targets as well. This being e.g. 1) Binding of IgA, collectins and complement components, to induce a cooperative anti-microbial effect. 2) Binding of hydroxyapatite on the tooth surface. 3) ECM proteins and glycosaminoglycans, as well as mucus components of the epithelial surface (such as Heparin, galectin 3 and mucins). **(Right side)** The cation-binding motif described in SALSA is conserved in most other SRCR proteins. For CD6, CD163 and MARCO mutational studies support a crucial role for this area in ligand interactions. CD163 binds the haemoglobin-haptoglobin complex as well as microbial surfaces. CD6 binds endogenous ligands, but also engage in microbial binding. MARCO forms multimers, and binds microbial surface structures. Other SRCR proteins with similar functions and conserved cation-sites include SR-A1, Sp-alpha, SSc5D and M2bp. The functional role of the Neurotrypsin SRCR domains is not known. The remarkable repetitive formation of multiple SRCR domains in many SRCR super family proteins, with several domains containing a binding-site with a broad specificity, would supposedly allow for interactions with multiple ligands simultaneously. The SRCR-fold thus appear to be an important functional component of scavenging molecules engaging in complex network of interactions. The multiple SRCR domains shown for SALSA, CD163 and Neurotrypsin are represented as copies of protein-specific SRCR domains with known structure. Conserved cation-coordinating residues are highlighted in red. SRCR8MgCa (light blue), MARCO (pdbid: 2oy3, sand), CD163 (pdbid: 5jfb, purple), CD6 (pdbid: 5a2e, grey), M2bp (pdbid: 1by2, yellow), and murine neurotrypsin (pdbid: 6h8m, teal).

SALSA belongs to the SRCR-super family, a family of proteins characterised by the presence of one or more copies of the ancient and evolutionarily highly conserved SRCR-fold [55]. Though a couple of SRCR-domains, such as the ones found in complement Factor I and Hepsin, have not been described to bind ligands directly, most others have [56, 57]. SRCR-super family members, such as SALSA, SR-A1, Spα, SSc5D, MARCO, CD6 and CD163 have broad scavenger-receptor functions, recognizing a broad range of microbial surface structures and mediate clearance [24-26, 40]. While this potentially is relevant for all SRCR-super family proteins, some members of the family have distinct protein ligands, such as CD6, CD163, and M2bp [22, 23, 27, 58]. With the exception of the CD6-CD166 interactions, most described SRCR-ligand interactions are calcium-dependent, irrespective of the ligand [24-26, 40-43]. A cation-binding site is conserved across SRCR domains, and multiple studies support a role for this site in ligand binding. Even the specialised CD6-CD166 interaction utilise the same surface for binding, despite 'having lost' the calcium dependency [27].

Our studies have thus identified a dual cation binding site as essential for SALSA-ligand interactions. Analysis of SRCR-folds from various ligand-binding domains reveal a very high level of conservation of the residues at this dual site. The conservation of this site, along with the well-described cation dependency on most SRCR-ligand interactions suggest that the binding mechanism described for the SALSA SRCR domains is applicable to all SRCR domains. We thus propose to have identified in SALSA a conserved functional mechanism for the SRCR class of proteins. This notion is further supported by the specific lack of conservation of these residues observed in the SRCR domains of complement Factor I and Hepsin, where no ligand binding has been shown. The SRCR domains in these two molecules may thus represent an evolutionary diversion from the common broad ligand binding potential of the SRCR fold. The novel understanding of the SRCR domain generated here, will allow for an interesting future targeting of other SRCR-super family proteins, with the potential of modifying function.

Acknowledgements

We acknowledge Diamond Light Source and the staff of beamlines I03 and I04 for access under proposal MX18069. The Central Oxford Structural Microscopy and Imaging Centre is supported by the Wellcome Trust (201536). M.P.R was financially supported by grants from the Wihuri foundation and the Finnish Cultural foundation. Staff and experimental costs in S.M.L. lab were supported by a Wellcome Investigator Award (100298) and an MRC programme grant (M011984).

Author contributions

M.P.R.: Designed and performed experiments. Protein purification, characterisation of protein complexes, X-ray crystallography, structure determination and analysis. Wrote paper with S.J. and S.M.L.

S.J.: Designed, supervised and performed experiments. Characterisation of protein complexes. Wrote manuscript with M.P.R. and S.M.L.

V.L.: Performed experiments. Strain and plasmid construction, protein purification.

S.M.L.: Designed and supervised experiments. Wrote paper with M.P.R. and S.J.

References

1. Holmskov U, Lawson P, Teisner B, Tornoe I, Willis AC, Morgan C, Koch C, Reid KB (1997) Isolation and characterization of a new member of the scavenger receptor superfamily, glycoprotein-340 (gp-340), as a lung surfactant protein-D binding molecule. J Biol Chem **272:** 13743-13749
2. Thornton DJ, Davies JR, Kirkham S, Gautrey A, Khan N, Richardson PS, Sheehan JK (2001) Identification of a nonmucin glycoprotein (gp-340) from a purified respiratory mucin preparation: evidence for an association involving the MUC5B mucin. Glycobiology **11:** 969-977
3. Schulz BL, Oxley D, Packer NH, Karlsson NG (2002) Identification of two highly sialylated human tear-fluid DMBT1 isoforms: the major high-molecular-mass glycoproteins in human tears. Biochem J **366:** 511-520
4. Reichhardt MP, Jarva H, de Been M, Rodriguez JM, Jimenez Quintana E, Loimaranta V, Meindert de Vos W, Meri S (2014) The Salivary Scavenger and Agglutinin in Early Life: Diverse Roles in Amniotic Fluid and in the Infant Intestine. J Immunol
5. Madsen J, Mollenhauer J, Holmskov U (2010) Review: Gp-340/DMBT1 in mucosal innate immunity. Innate Immun **16:** 160-167
6. Reichhardt MP, Meri S (2016) SALSA: A Regulator of the Early Steps of Complement Activation on Mucosal Surfaces. Front Immunol **7:** 85
7. Reichhardt MP, Holmskov U, Meri S (2017) SALSA-A dance on a slippery floor with changing partners. Mol Immunol **89:** 100-110
8. Prakobphol A, Xu F, Hoang VM, Larsson T, Bergstrom J, Johansson I, Frangsmyr L, Holmskov U, Leffler H, Nilsson C et al. (2000) Salivary agglutinin, which binds Streptococcus mutans and Helicobacter pylori, is the lung scavenger receptor cysteine-rich protein gp-340. J Biol Chem **275:** 39860-39866
9. Hartshorn KL, White MR, Mogues T, Ligtenberg T, Crouch E, Holmskov U (2003) Lung and salivary scavenger receptor glycoprotein-340 contribute to the host defense against influenza A viruses. Am J Physiol Lung Cell Mol Physiol **285:** L1066-76
10. Loimaranta V, Jakubovics NS, Hytonen J, Finne J, Jenkinson HF, Stromberg N (2005) Fluid- or surface-phase human salivary scavenger protein gp340 exposes different bacterial recognition properties. Infect Immun **73:** 2245-2252
11. Rosenstiel P, Sina C, End C, Renner M, Lyer S, Till A, Hellmig S, Nikolaus S, Folsch UR, Helmke B et al. (2007) Regulation of DMBT1 via NOD2 and TLR4 in intestinal epithelial cells modulates bacterial recognition and invasion. J Immunol **178:** 8203-8211
12. Rundegren J, Arnold RR (1987) Differentiation and interaction of secretory immunoglobulin A and a calcium-dependent parotid agglutinin for several bacterial strains. Infect Immun **55:** 288-292
13. Boackle RJ, Connor MH, Vesely J (1993) High molecular weight non-immunoglobulin salivary agglutinins (NIA) bind C1Q globular heads and have the potential to activate the first complement component. Mol Immunol **30:** 309-319
14. Tino MJ, Wright JR (1999) Glycoprotein-340 binds surfactant protein-A (SP-A) and stimulates alveolar macrophage migration in an SP-A-independent manner. Am J Respir Cell Mol Biol **20:** 759-768
15. Oho T, Bikker FJ, Nieuw Amerongen AV, Groenink J (2004) A peptide domain of bovine milk lactoferrin inhibits the interaction between streptococcal surface protein antigen and a salivary agglutinin peptide domain. Infect Immun **72:** 6181-6184
16. Leito JT, Ligtenberg AJ, van Houdt M, van den Berg TK, Wouters D (2011) The bacteria binding glycoprotein salivary agglutinin (SAG/gp340) activates complement via the lectin pathway. Mol Immunol **49:** 185-190

17. Reichhardt MP, Loimaranta V, Thiel S, Finne J, Meri S, Jarva H (2012) The salivary scavenger and agglutinin binds MBL and regulates the lectin pathway of complement in solution and on surfaces. Front Immunol **3:** 205
18. Madsen J, Sorensen GL, Nielsen O, Tornoe I, Thim L, Fenger C, Mollenhauer J, Holmskov U (2013) A variant form of the human deleted in malignant brain tumor 1 (DMBT1) gene shows increased expression in inflammatory bowel diseases and interacts with dimeric trefoil factor 3 (TFF3). PLoS One **8:** e64441
19. Holmskov U, Mollenhauer J, Madsen J, Vitved L, Gronlund J, Tornoe I, Kliem A, Reid KB, Poustka A, Skjodt K (1999) Cloning of gp-340, a putative opsonin receptor for lung surfactant protein D. Proc Natl Acad Sci U S A **96:** 10794-10799
20. Mollenhauer J, Holmskov U, Wiemann S, Krebs I, Herbertz S, Madsen J, Kioschis P, Coy JF, Poustka A (1999) The genomic structure of the DMBT1 gene: evidence for a region with susceptibility to genomic instability. Oncogene **18:** 6233-6240
21. Mollenhauer J, Wiemann S, Scheurlen W, Korn B, Hayashi Y, Wilgenbus KK, von Deimling A, Poustka A (1997) DMBT1, a new member of the SRCR superfamily, on chromosome 10q25.3-26.1 is deleted in malignant brain tumours. Nat Genet **17:** 32-39
22. Inohara H, Akahani S, Koths K, Raz A (1996) Interactions between galectin-3 and Mac-2-binding protein mediate cell-cell adhesion. Cancer Res **56:** 4530-4534
23. Kristiansen M, Graversen JH, Jacobsen C, Sonne O, Hoffman HJ, Law SK, Moestrup SK (2001) Identification of the haemoglobin scavenger receptor. Nature **409:** 198-201
24. Santiago-Garcia J, Kodama T, Pitas RE (2003) The class A scavenger receptor binds to proteoglycans and mediates adhesion of macrophages to the extracellular matrix. J Biol Chem **278:** 6942-6946
25. Sarrias MR, Rosello S, Sanchez-Barbero F, Sierra JM, Vila J, Yelamos J, Vives J, Casals C, Lozano F (2005) A role for human Sp alpha as a pattern recognition receptor. J Biol Chem **280:** 35391-35398
26. Ojala JR, Pikkarainen T, Tuuttila A, Sandalova T, Tryggvason K (2007) Crystal structure of the cysteine-rich domain of scavenger receptor MARCO reveals the presence of a basic and an acidic cluster that both contribute to ligand recognition. J Biol Chem **282:** 16654-16666
27. Chappell PE, Garner LI, Yan J, Metcalfe C, Hatherley D, Johnson S, Robinson CV, Lea SM, Brown MH (2015) Structures of CD6 and Its Ligand CD166 Give Insight into Their Interaction. Structure **23:** 1426-1436
28. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A et al. (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr **67:** 235-242
29. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr **66:** 486-501
30. Adams PD, Baker D, Brunger AT, Das R, DiMaio F, Read RJ, Richardson DC, Richardson JS, Terwilliger TC (2013) Advances, interactions, and future developments in the CNS, Phenix, and Rosetta structural biology software systems. Annu Rev Biophys **42:** 265-287
31. Pidcock E, Moore GR (2001) Structural characteristics of protein binding sites for calcium and lanthanide ions. J Biol Inorg Chem **6:** 479-489
32. Dokmanic I, Sikic M, Tomic S (2008) Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination. Acta Crystallogr D Biol Crystallogr **64:** 257-263
33. Haukioja A, Loimaranta V, Tenovuo J (2008) Probiotic bacteria affect the composition of salivary pellicle and streptococcal adhesion in vitro. Oral Microbiol Immunol **23:** 336-343

34. End C, Bikker F, Renner M, Bergmann G, Lyer S, Blaich S, Hudler M, Helmke B, Gassler N, Autschbach F et al. (2009) DMBT1 functions as pattern-recognition molecule for poly-sulfated and poly-phosphorylated ligands. Eur J Immunol **39:** 833-842
35. Loimaranta V, Hytonen J, Pulliainen AT, Sharma A, Tenovuo J, Stromberg N, Finne J (2009) Leucine-rich repeats of bacterial surface proteins serve as common pattern recognition motifs of human scavenger receptor gp340. J Biol Chem **284:** 18614-18623
36. Zhou YF, Eng ET, Zhu J, Lu C, Walz T, Springer TA (2012) Sequence and structure relationships within von Willebrand factor. Blood **120:** 449-458
37. Holm L, Laakso LM (2016) Dali server update. Nucleic Acids Res **44:** W351-5
38. Canciani A, Catucci G, Forneris F (2019) Structural characterization of the third scavenger receptor cysteine-rich domain of murine neurotrypsin. Protein Sci **28:** 746-755
39. Nielsen MJ, Andersen CB, Moestrup SK (2013) CD163 binding to haptoglobin-hemoglobin complexes involves a dual-point electrostatic receptor-ligand pairing. J Biol Chem **288:** 18834-18841
40. Sarrias MR, Farnos M, Mota R, Sanchez-Barbero F, Ibanez A, Gimferrer I, Vera J, Fenutria R, Casals C, Yelamos J et al. (2007) CD6 binds to pathogen-associated molecular patterns and protects from LPS-induced septic shock. Proc Natl Acad Sci U S A **104:** 11724-11729
41. Vera J, Fenutria R, Canadas O, Figueras M, Mota R, Sarrias MR, Williams DL, Casals C, Yelamos J, Lozano F (2009) The CD5 ectodomain interacts with conserved fungal cell wall components and protects from zymosan-induced septic shock-like syndrome. Proc Natl Acad Sci U S A **106:** 1506-1511
42. Nonaka M, Ma BY, Imaeda H, Kawabe K, Kawasaki N, Hodohara K, Kawasaki N, Andoh A, Fujiyama Y, Kawasaki T (2011) Dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin (DC-SIGN) recognizes a novel ligand, Mac-2-binding protein, characteristically expressed on human colorectal carcinomas. J Biol Chem **286:** 22403-22413
43. Patel DD, Wee SF, Whichard LP, Bowen MA, Pesando JM, Aruffo A, Haynes BF (1995) Identification and characterization of a 100-kD ligand for CD6 on human thymic epithelial cells. J Exp Med **181:** 1563-1568
44. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res **44:** W344-50
45. Ligtenberg AJ, Karlsson NG, Veerman EC (2010) Deleted in malignant brain tumors-1 protein (DMBT1): a pattern recognition receptor with multiple binding sites. Int J Mol Sci **11:** 5212-5233
46. Bikker FJ, Ligtenberg AJ, Nazmi K, Veerman EC, van't Hof W, Bolscher JG, Poustka A, Nieuw Amerongen AV, Mollenhauer J (2002) Identification of the bacteria-binding peptide domain on salivary agglutinin (gp-340/DMBT1), a member of the scavenger receptor cysteine-rich superfamily. J Biol Chem **277:** 32109-32115
47. Bikker FJ, Ligtenberg AJ, End C, Renner M, Blaich S, Lyer S, Wittig R, van't Hof W, Veerman EC, Nazmi K et al. (2004) Bacteria binding by DMBT1/SAG/gp-340 is confined to the VEVLXXXXW motif in its scavenger receptor cysteine-rich domains. J Biol Chem **279:** 47699-47703
48. Leito JT, Ligtenberg AJ, Nazmi K, de Blieck-Hogervorst JM, Veerman EC, Nieuw Amerongen AV (2008) A common binding motif for various bacteria of the bacteria-binding peptide SRCRP2 of DMBT1/gp-340/salivary agglutinin. Biol Chem **389:** 1193-1200
49. Young A, Rykke M, Smistad G, Rolla G (1997) On the role of human salivary micelle-like globules in bacterial agglutination. Eur J Oral Sci **105:** 485-494

50. Oho T, Yu H, Yamashita Y, Koga T (1998) Binding of salivary glycoprotein-secretory immunoglobulin A complex to the surface protein antigen of Streptococcus mutans. Infect Immun **66:** 115-121
51. Crouch EC (2000) Surfactant protein-D and pulmonary host defense. Respir Res **1:** 93-108
52. White MR, Crouch E, van Eijk M, Hartshorn M, Pemberton L, Tornoe I, Holmskov U, Hartshorn KL (2005) Cooperative anti-influenza activities of respiratory innate immune proteins and neuraminidase inhibitor. Am J Physiol Lung Cell Mol Physiol **288:** L831-40
53. Jonasson A, Eriksson C, Jenkinson HF, Kallestal C, Johansson I, Stromberg N (2007) Innate immunity glycoprotein gp-340 variants may modulate human susceptibility to dental caries. BMC Infect Dis **7:** 57
54. Stoddard E, Cannon G, Ni H, Kariko K, Capodici J, Malamud D, Weissman D (2007) gp340 expressed on human genital epithelia binds HIV-1 envelope protein and facilitates viral transmission. J Immunol **179:** 3126-3132
55. Sarrias MR, Gronlund J, Padilla O, Madsen J, Holmskov U, Lozano F (2004) The Scavenger Receptor Cysteine-Rich (SRCR) domain: an ancient and highly conserved protein module of the innate immune system. Crit Rev Immunol **24:** 1-37
56. Roversi P, Johnson S, Caesar JJ, McLean F, Leath KJ, Tsiftsoglou SA, Morgan BP, Harris CL, Sim RB, Lea SM (2011) Structural basis for complement factor I control and its disease-associated sequence polymorphisms. Proc Natl Acad Sci U S A **108:** 12839-12844
57. Koschubs T, Dengl S, Durr H, Kaluza K, Georges G, Hartl C, Jennewein S, Lanzendorfer M, Auer J, Stern A et al. (2012) Allosteric antibody inhibition of human hepsin protease. Biochem J **442:** 483-494
58. Burgueno-Bucio E, Mier-Aguilar CA, Soldevila G (2019) The multiple faces of CD5. J Leukoc Biol **105:** 891-904