THE PROTEIN SOCIETY

Structural characterization of the third scavenger receptor cysteine-rich domain of murine neurotrypsin

Anselmo Canciani,[1] Gianluca Catucci,[2] and Federico Forneris [1]*

[1]The Armenise-Harvard Laboratory of Structural Biology, Department of Biology and Biotechnology, University of Pavia, Via Ferrata 9/A, 27100 Pavia, Italy

[2]Department of Life Sciences and Systems Biology, University of Torino, 10123 Turin, Italy

Received 3 October 2018; Accepted 11 February 2019
DOI: 10.1002/pro.3587
Published online 12 February 2019 proteinscience.org

Abstract: **Neurotrypsin (NT) is a multi-domain serine protease of the nervous system with only one known substrate: the large proteoglycan Agrin. NT has seen to be involved in the maintenance/turnover of neuromuscular junctions and in processes of synaptic plasticity in the central nervous system. Roles which have been tied to its enzymatic activity, localized in the C-terminal serine-protease (SP) domain. However the purpose of NT's remaining 3–4 scavenger receptor cysteine-rich (SRCR) domains is still unclear. We have determined the crystal structure of the third SRCR domain of murine NT (mmNT-SRCR3), immediately preceding the SP domain and performed a comparative structural analysis using homologous SRCR structures. Our data and the elevated degree of structural conservation with homologous domains highlight possible functional roles for NT SRCRs. Computational and experimental analyses suggest the identification of a putative binding region for Ca^{2+} ions, known to regulate NT enzymatic activity. Furthermore, sequence and structure comparisons allow to single out regions of interest that, in future studies, might be implicated in Agrin recognition/binding or in interactions with as of yet undiscovered NT partners.**

Keywords: scavenger receptor cysteine-rich domain; SRCR; protease; neurotrypsin; neuromuscular junctions

Introduction

Neurotrypsin (NT), also known as PRSS12 and motopsin, is a multi-domain extracellular serine protease of the nervous system first described in the late 1990s.[1,2] Spanning 761 amino acids, the murine variant encompasses an N-terminal proline-rich segment, a kringle domain (Kr), three scavenger receptor cysteine rich (SRCR) domains, and a C-terminal serine-protease (SP) domain.[1,2] In comparison with the human homolog (875 a.a.), sharing an overall sequence identity of 87%, the murine ortholog is shorter in virtue of an additional N-terminal SRCR domain[3] [Figs. 1(A) and Supporting Information Fig. S1].

Produced predominantly by neurons of the central and peripheral nervous systems, NT is stocked in synaptic vesicles and released in an activity-dependent manner.[4,5] Upon secretion, NT cleaves the large proteoglycan Agrin, its only known substrate, generating two C-terminal fragments. This event is likely responsible for the roles played by NT in neuromuscular junction maintenance/turnover and in neuronal plasticity.[4–7] Such is the importance of this cleavage event that de-regulation and/or inactivation result, respectively, in Sarcopoenia (muscle wasting disease of the elderly)[7] and non-syndromic mental retardation.[4]

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Fondazione Cariplo 2014-0881; Grant sponsor: Giovanni Armenise-Harvard Foundation Career Development Award 2013; Grant sponsor: Ministero dell'Istruzione, dell'Università e della Ricerca Programma Dipartimenti di Eccellenza 2018-2022 (Dipartimento di Biologia e Biotecnologie, Università di Pavia) and Programma Giovani Ricercatori "Rita Levi-Montalcini.

*Correspondence to: Federico Forneris, The Armenise-Harvard Laboratory of Structural Biology, Department of Biology and Biotechnology, University of Pavia, Via Ferrata 9/A, 27100 Pavia, Italy. E-mail: federico.forneris@unipv.it

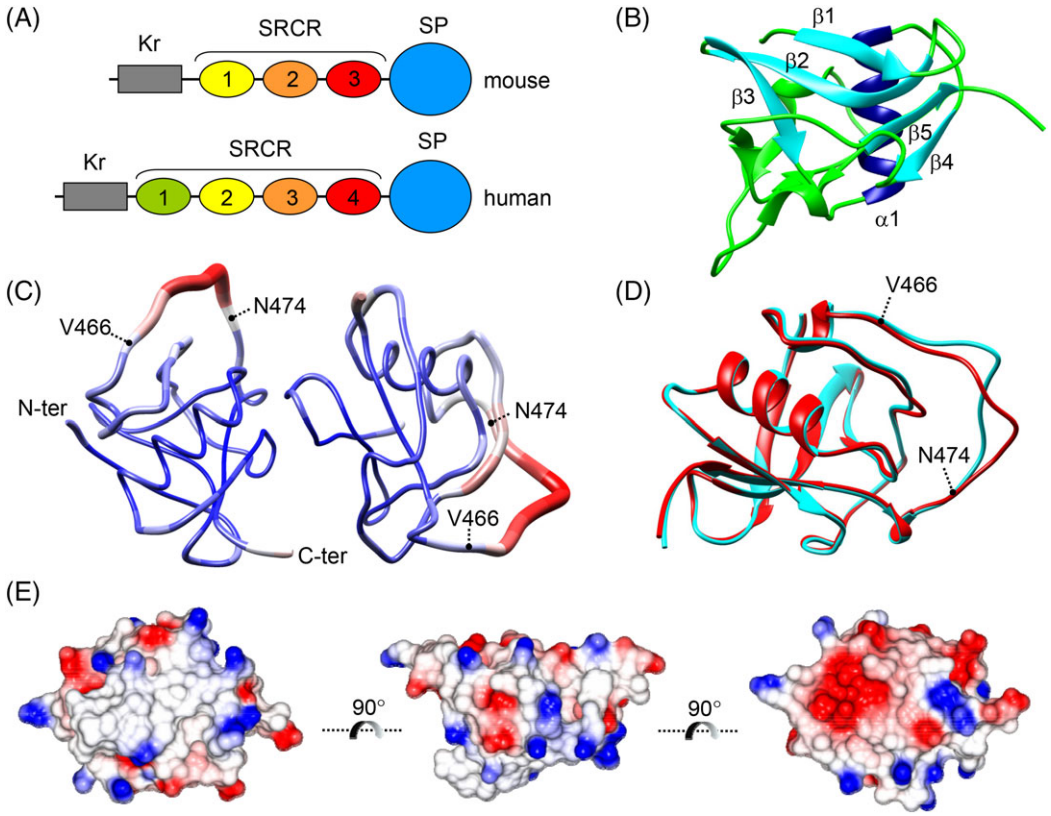

Figure 1. Three-dimensional structure of the third SRCR domain of murine NT. (A) Domain organization of mouse and human NT, highlighting the localization of mmNT-SRCR3 within the multidomain enzyme architecture. Coloring of the domains highlights sequence conservation between homologous enzymes, as described in Supporting Information Fig. S1. (B) Cartoon representation of mmNT-SRCR3 displaying its main secondary structure elements, α-helix (α1) in blue and β-strands (β1–β5) in cyan. (C) Cartoon "putty" representation of the two mmNT-SRCR3 monomers found in the crystallographic asymmetric unit, colored blue-red by B-factors (34.2–162.5 Å^2). (D) Superposition of the two mmNT-SRCR3 molecules found in the asymmetric unit. Chain A is shown in red, Chain B is shown in cyan. (E) Surface representation with coloring by electrostatic potential (−2.0 − +2.0 V), negative charges in red, positive charges in blue and neutral areas in white. Monomer is presented in orientations highlighting main features: uncharged bottom, mixed charges on sides and negatively charged patch on top.

While the significance of the catalytic domain has been touched upon, the role and relevance of the accessory domains remain unclear. Nonetheless, it would be reasonable to attribute them with a role in protein–protein interaction and/or substrate recognition. Indeed, literature reports two potential NT interactors, in addition to the substrate Agrin, identified through yeast two-hybrid screening experiments: the product of seizure-related gene 6 (Sez-6)[8] and the Integral Membrane Protein 2A (Itm2a).[8] These have been seen to influence, respectively, neuronal development and function,[9,10] and the differentiation of different tissue/cell types,[11–13] among which skeletal muscle.[14] NT's interaction with Sez-6 has been tied to one of the accessory domains: the Kr domain.[8] Conversely, the domain(s) responsible for interacting with Itm2a has yet to be identified.

The scavenger receptor cysteine-rich (SRCR) domain is a small (90–110 residues) distinguishing element of a broad superfamily of proteins whose members span different functional roles.[15–17] It displays a highly conserved structure with five anti-parallel beta strands cradling an alpha helix and a distinct disulfide bridge architecture.[18,20–23] A feature which has been used to identify two types of SRCR domains, A and B, bearing, respectively, 6 and 8 cysteines. While the number of cysteines differs, the relative pairing is surprisingly consistent and tends to form as follows: C1–C4, C2–C7, C3–C8, and C5–C6. The first pair distinguishes Types A and B as it is present exclusively in the latter.[16] Members of the SRCR superfamily can share little functional similarity playing roles in immune response,[18] cell differentiation,[24] apoptosis,[25] and tumor suppression alike. While it has been difficult to assign a univocal function to SRCR domains, a commonality can be drawn across several SRCR superfamily members, broadly attributing the SRCR domain with a role in protein–protein/protein–substrate interaction.[26] It is, therefore, likely that NT SRCR domains may mediate interactions with the surrounding environment. They could be responsible for the reported glycosaminoglycan and heparin binding capabilities,[27] contribute to its specificity, mediate its localized lingering at the synapse[19] and/or be involved in binding to,

as of yet, unidentified partners. Of NT's accessory domains, only the structure of the Kr domain has been successfully determined and assigned a possible role.[8,20] The lack of structural information pertaining to the remaining domains limits our functional understanding of NT. As such new structural data on the SP or SRCR domains might provide key insights on the mechanisms underlying NT's biological role.

Here we report the recombinant production, purification, crystallization, and crystal structure determination of the third SRCR domain of murine NT (mmNT-SRCR3), and analysis of its molecular architecture based on comparisons with homologous SRCR domains. These data constitute a solid additional step toward the understanding of NT's molecular interactions and biological functions.

Results

Purification of mmNT-SRCR3

Owing to the high cysteine content of extracellular SRCR domains, we produced the mmNT-SRCR3 in soluble form using a *Escherichia coli* strain particularly adapted to facilitate disulfide bond formation. The purification of mmNT-SRCR3 was performed using a two-step immobilized Ni-affinity chromatography (Ni-IMAC) followed by size-exclusion chromatography (SEC) (Fig. S2). The purified material was used for crystallization experiments and protein characterization studies in solution.

Crystal structure of mmNT-SRCR3

A single mmNT-SRCR3 crystal allowed complete X-ray data collection and structure determination [Fig. 1(B)]; data processing and structure refinement statistics are reported in Table I. This process highlighted the presence of two mmNT-SRCR3 monomers in the asymmetric unit assembling into what appeared to be a crystallographic dimer [Fig. 1(C)]. Each monomer presents a very compact fold with a central alpha helix (α1) nested in five beta strands (β1–β5) forming a twisted anti-parallel cradle [Fig. 1(B)]. The B-factor distribution across the structure highlights the high stability of the central core and progressive flexibility of the peripheral loops interconnecting the secondary structure elements [Fig. 1(C)]. Of particular interest is the Val466–Asn474 segment, whose flexibility is such that the electron density map of that region is very poor despite the 1.7-Å resolution (Fig. S3).

Each mmNT-SRCR3 model spans 107 residues and lacks the final C-terminal "KKASS" sequence, owing to the high mobility of this five-residue tail. Superposition of the two monomers found in the asymmetric unit and analysis of the root mean square deviation (r.m.s.d) shows that the only visible significant difference is located in the aforementioned Val466–Asn474 loop, further supporting its flexible nature [Fig. 1(D)]. Calculations of surface electrostatic potentials using CCP4mg[21] displayed a disc-like shape with mixed charge distribution along its sides and two opposite faces, one almost fully hydrophobic and the other displaying a patch of negative charges [Fig. 1(E)].

Table I. *X-Ray Data Collection and Refinement Statistics*

Data Collection	
Diffraction source	ESRF ID30A-3
Wavelength (Å)	0.983
Space group	P $2_12_12_1$
Cell parameters a, b, c (Å)	a = 57.8 Å, b = 62.9 Å, c = 84.5 Å, $\alpha = \beta = \gamma = 90°$
Resolution range (Å)[a,b]	42.54–1.70 (1.73–1.70)
Unique reflections	34576 (1785)
Completeness (%)	100.0 (100.0)
Multiplicity	6.9 (6.9)
Mean ($I/\sigma[I]$)	15.5 (0.5)
CC(1/2)	0.999 (0.434)
R_{sym}[c]	0.049 (3.37)
Refinement	
R_{work}/R_{free}	0.19 / 0.21
No. of non-H atoms	1747
Protein	1662
Ligands	0
Waters	85
r.m.s. deviations	
Bonds (Å)	0.016
Angles (°)	1.699
Average B factors (Å[b])	
Protein	65.3
Water	54.0
Ramachandran favored (%)	97.6
Ramachandran allowed (%)	2.4

[a]Values for the outer shell are given in parentheses.
[b]Resolution limits were determined by applying a cut-off based on the mean intensity correlation coefficient of half-datasets (CC1/2) approximately of 0.5[49].
[c]$R_{sym} = \Sigma \mid I - \langle I \rangle \mid / \Sigma I$, where I is the observed intensity for a reflection and $\langle I \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections.

SEC-SAXS analysis

In order to assess the oligomeric state of the purified protein sample in solution, and thus ascertain the nature of the mmNT-SRCR3 dimer observed in the crystal packing [Fig. 1(C)], we performed a SEC-small angle X-ray scattering (SEC-SAXS) experiment. The initial chromatogram displayed a well-defined peak with good scattering intensities and a very homogeneous mass distribution [Fig. 2(A)]. mmNT-SRCR3 presents as a monodisperse species with an MW of 12 kDa calculated from averaged peak intensities (Table II). This is in line with the 12.5 kDa calculated for a monomer of this construct and incompatible with higher order oligomeric species. *CRYSOL*[22] was used to compare the averaged scattering curve of the

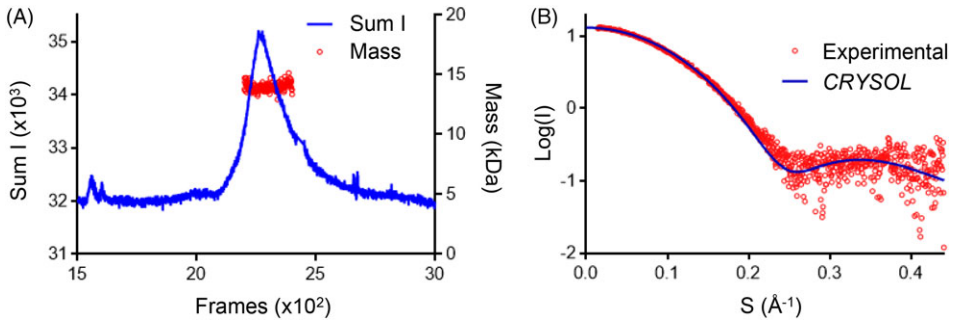

Figure 2. mmNT-SRCR3 SEC-SAXS analysis. (A) SEC-SAXS chromatogram showing mmNT-SRCR3 scattering intensities and profile of mass distribution across the peak. (B) Averaged experimental peak scattering curve, after buffer subtraction, plotted, and superimposed to a theoretical, *CRYSOL*-derived, scattering curve for an mmNT-SRCR3 monomer (χ^2 = 1.4).

SEC-SAXS peak to a theoretically calculated scattering curve for our crystallographic structure. The single mmNT-SRCR3 monomer displayed a very good fit (χ^2 = 1.4) with the in-solution data [Fig. 2(B)], ruling the dimer as induced by the crystallization process.

Comparison with other SRCR domains

To better understand the functional role of NT's SRCR domains, we used the *DALI* server[23] to compare this structure with other entries in the Protein Data Bank (PDB). This search returned several high scoring matches which were manually filtered to remove redundancies resulting in a list of only eight hits (Table III). While limited in sequence identity, structural superpositions with mmNT-SRCR3 [Fig. 3(A)] evidenced a high degree of structural conservation, strongest within the central core and with greater variability in the peripheral loops. Comparison with an expanded pool of homologs using *ConSurf*[24] highlighted a similar trend for sequence conservation. When mapped to the structure of mmNT-SRCR3, it was possible to see how the areas of greater sequence identity, for the most part, corresponded to highly conserved secondary structure elements forming the domain core [Fig. 3(B)]. Conversely, the more flexible external loops displayed a significantly greater compositional variability. Of particular note is the β4 strand that, while being one of five core β strands displays little to no amino acid conservation [Fig. 3(C)].

Table II. *Summary of SEC-SAXS Data Collection and Analysis*

Data collection	
Light source and beamline	ESRF BM29
Beam energy (keV)	12.5
Sample-detector distance (m)	2.867
Exposure time (s)	1
Sample cell thickness (mm)	1
Temperature (°)	20
Final q range (nm^{-1})	0.01–4
Data analysis	
Points used for Guinier analysis	27–173
Guinier qRg limits	0.24
Guinier Rg (nm)	1.54 ± 0.03
$I(0)$ (nm^{-1})	13.48 ± 0.02
D_{max} (nm)	5.8
MW estimation (V_c based) (kDa)	12

DALI matches with >30% identity evidenced several perfectly conserved residues which were mirrored in ConSurf alignments. Among these is a six-cysteine network (residues 411, 424, 455, 465, 475, and 485) responsible for the archetypical SRCR domain disulfide bridges [Fig. 3(B)]. This canonical pairing is expressed relative to type B SRCR domains, and while Type A domains lack the C1–C4 pair, the relative numbering of the other pairs remains conventionally and structurally unaltered. Therefore, in mmNT-SRCR3, we found that Cys411–Cys475 corresponds to the C2–C7 pair, while Cys424–Cys485 and Cys455–Cys465 correspond to the C3–C8 and C5–C6 pairs, respectively. Residues Gly391, Gly397, Glu400, Ala420, Val422, Leu427, Gly457, Glu459, and Val483 are also fully conserved [Fig. 3(C)]. Of these, Ala420, Val422, and Leu427 can be mapped to the central helix, while Gly397, Glu400, and Val483 can be found on strands β2 and β5, respectively [Fig. 3(B)]. Conversely, the remaining Gly391, Gly457, and Glu459 locate on the more flexible, and generally less conserved, external loops [Fig. 3(B)].

Comparison of mmNT-SRCR3 with other mmNT SRCR domains evidenced a certain extent of conservation (44–58% sequence identity) comparable with that of the highest *DALI* matches [Supporting Information Fig. S1(B)]. Extending this analysis to include human NT evidenced how its SRCR Domains 2–4 correspond, respectively, to mmNT SRCR Domains 1–3, while the first human SRCR domain is not conserved across the two homologs [Supporting Information Fig. S1(B)].

Evidence for Ca^{2+} binding

Given the significant contributions of Ca^{2+} to NT activity,[5] and the reported Ca^{2+}-based modulation of SRCR-mediated interactions,[18,25] we investigated whether mmNT-SRCR3 might contribute in similar fashion. Therefore, an isothermal titration calorimetry

Table III. *List of Homologous SRCR Structures Identified by DALI*

Protein	PDB	*Z*-score	r.m.s.d.	% id.
Lysyl oxidase Homolog 2	5ZE3	18.2	2.1	49
CD6	5A2E	18.0	1.5	49
Mac-2 binding protein (M2BP)	1BY2	18.0	1.8	50
Scavenger receptor cysteine-rich Type 1 protein m	5JFB	16.6	1.8	38
Macrophage receptor (MARCO)	2OYA	13.7	1.7	52
Human complement Factor I	2XRC	11.2	2.2	31
T-cell surface glycoprotein CD5	2JA4	7.5	2.4	27
Serine protease Hepsin	3T2N	7.2	3.0	16

(ITC) experiment was conducted in presence of $CaCl_2$. This evidenced a Ca^{2+} binding to mmNT-SRCR3 [Fig. 4(A)], with K_d = 10.1 mM, ΔH = −7.7 kcal/mol, ΔS = −16.8 cal/mol/deg.

A comparison of mmNT-SRCR3 with the structure of the monomeric MARCO SRCR domain (PDB: 2OY3)[18] highlighted a cluster of highly conserved residues co-ordinating a bivalent metal-ion in MARCO [Fig. 4(B)]. These correspond to amino acids Asp412–Asp413, His473, Asn474, and Glu479 that significantly contribute mmNT-SRCR3's negatively charged patch. Interestingly, the same residues were seen to

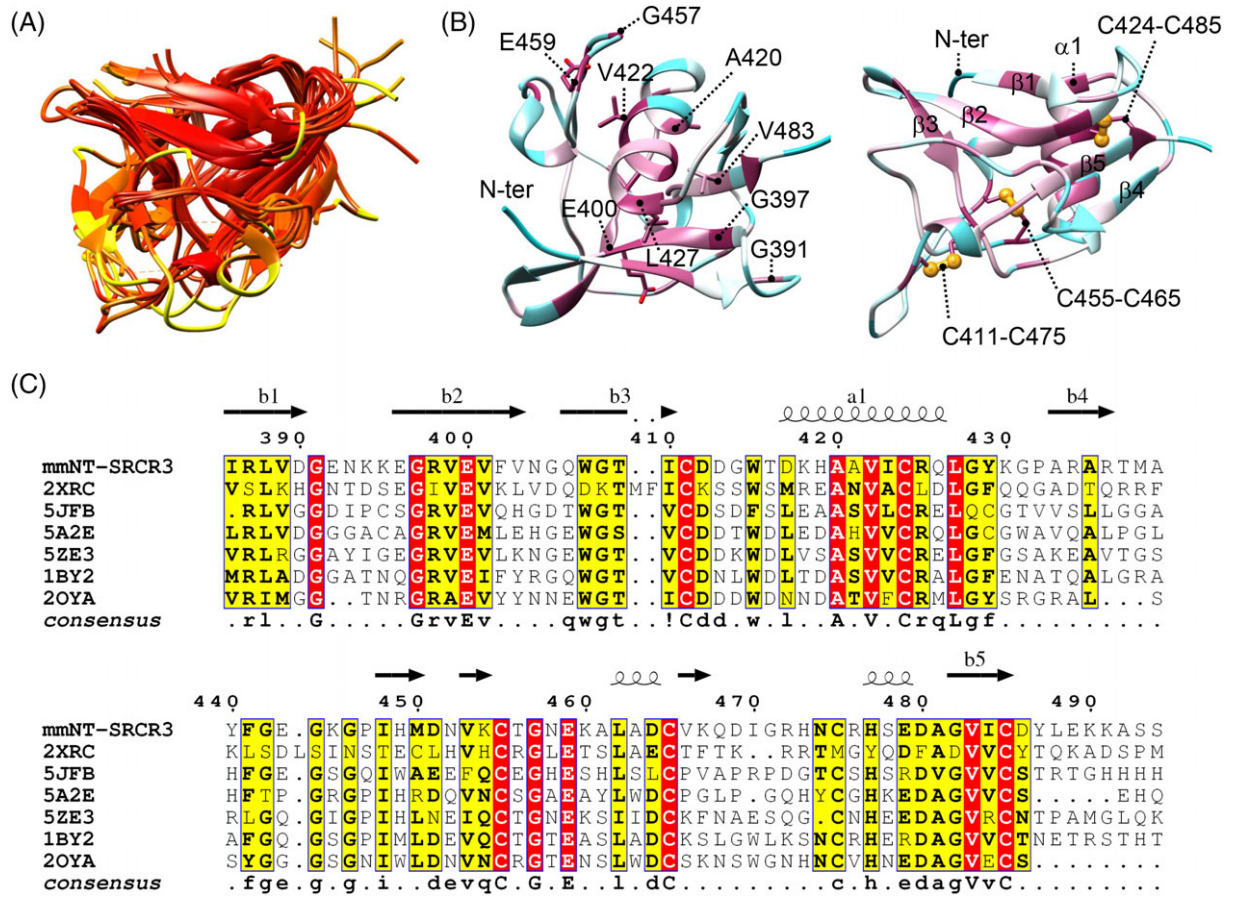

Figure 3. Structural comparison of mmNT-SRCR3 with other SRCR domains. (A) Superposition with non-redundant *DALI* matches (Table III) colored red-yellow by r.m.s.d. (0.39–6.32 Å). (B) Cartoon representation of mmNT-SRCR3 in two orientations, colored cyan-brown (0–100%) by residue conservation after *ConSurf* analysis. The left panel maps highly conserved residues from SRCR structure sequence alignments, while the right panel maps the principal secondary structure elements and the archetypical SRCR disulfide bridge organization. The C1–C4 pair is absent in SRCR type a domains, residues Cys411–Cys475 correspond to the C2–C7 pair, Cys424–Cys485 to the C3–C8 pair and Cys455–Cys465 to the C5–C6 pair. (C) Sequence alignment of SRCR structures identified by *DALI* analysis of mmNT-SRCR3. Identical residues boxed in red, residues with 70% conservation boxed in yellow. mmNT-SRCR3 secondary structure elements displayed above alignment and consensus sequence with 70% conservation under alignment. Image created using *ESPRIPT3.*[50]

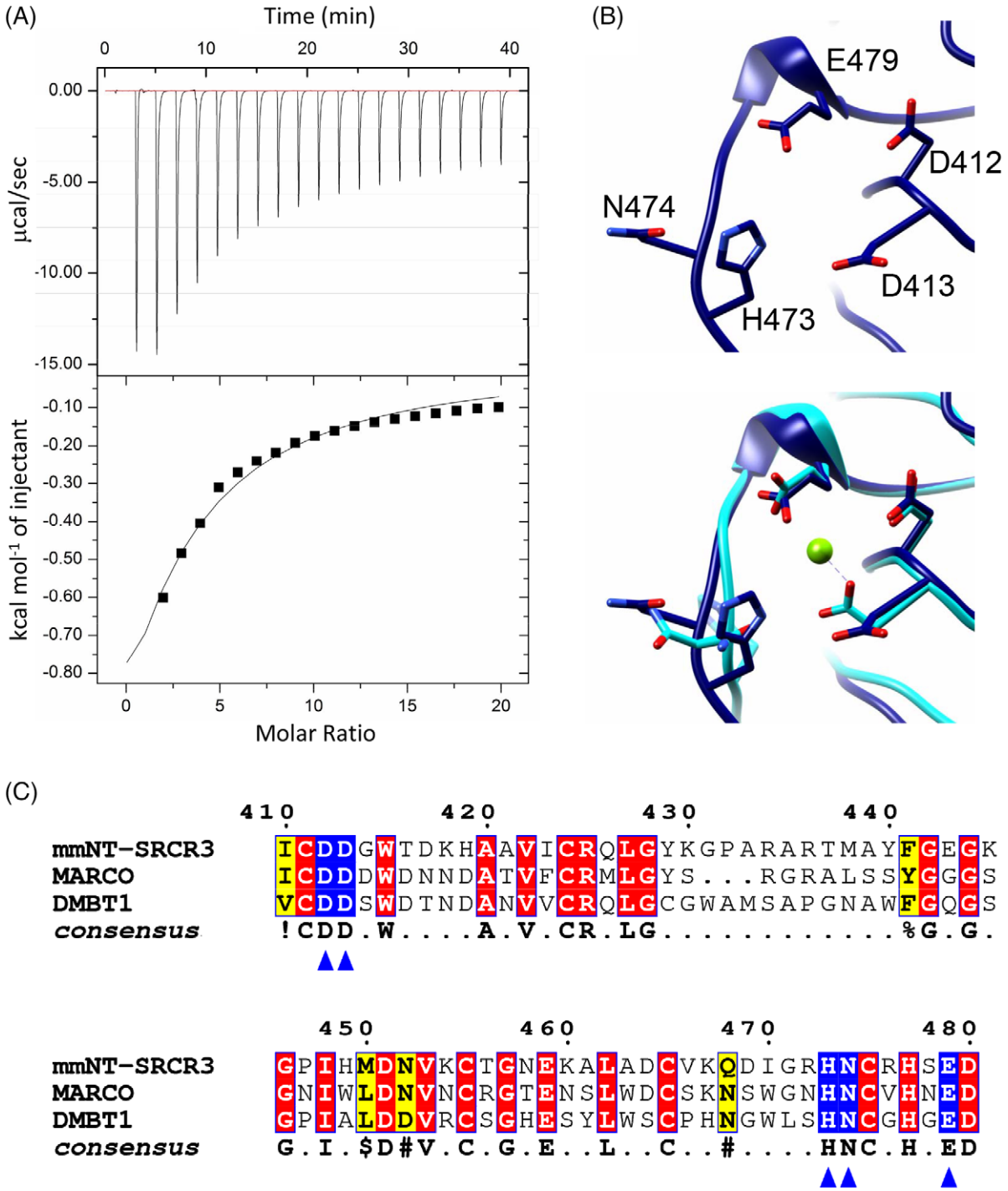

Figure 4. Evidence for mmNT-SRCR3 Ca^{2+} binding. (A) Isothermal calorimetric titration of $CaCl_2$ (120 m*M*) with purified mmNT-SRCR3. Heat variation generated by each injection of titrant at each time interval (top panel) and the integration of each peak and the amount of heat produced (lower panel). (B) Structural inference of mmNT-SRCR3 binding of metal ions. Top panel, residues contributing to the formation of the negatively charged patch on mmNT-SRCR3 likely bind metal ions. Bottom panel, superposition of mmNT-SRCR3 (blue) with monomeric MARCO (cyan, PDB: 2OY3) co-ordinating Mg^{2+} (green sphere) via those same amino acids. Side chains of residues Asp412, Asp413, and Glu479 display almost identical conformations, while Asn474 and His473 show higher flexibility. (C) Sequence alignment of mmNT-SRCR3 against MARCO and DMBT1 SRCR domains known to bind Ca^{2+}. Blue arrows and boxes indicate residues evidenced in (B) which are perfectly conserved across all three proteins. Perfectly conserved residues boxed in red, residues with 70% conservation are boxed in yellow. Evaluation of sequence conservation is shown under alignment; symbols indicate:! = Ile or Val, $ = Leu or Met, % = Phe or Tyr, # = Asn, Asp, Glu, or Gln. Residue numbering refers to mmNT-SRCR3.

be perfectly conserved in the DMBT1 SRCR-1 domain, also known to display Ca^{2+} binding[26] [Fig. 4(C)].

Discussion

The scavenger receptor cysteine-rich domain groups numerous proteins with different functions in a large superfamily.[15] However, the understanding of the specific function of SRCR domains is, to date, limited to a general consensus of protein–protein/protein–ligand interaction. The solution of the structure of the third SRCR domain of murine Neurotrypsin (mmNT-SRCR3) and its analysis highlighted several interesting features that might provide a functional insight both in regards to NT and SRCR domains in general. Most striking of all was the degree of similarity with the available homologous structures despite a generally poor sequence conservation (Table III).

At the core of mmNT-SRCR3 is an extremely stable secondary structure element organization which seems to be perfectly conserved across available structures. Greater variability was observed for the more flexible external loops, most notably the Val466–Asn474 loop. This region not only displayed the highest mobility across homologs but also within the mmNT-SRCR3 monomers observed in the crystallographic dimer. Mapping of sequence conservation to the structure of mmNT-SRCR3 showed that the least conserved residues were located on this and other peripheral flexible loops, while the core secondary structure elements bore the majority of highly conserved amino acids. Of particular note are, a six-cysteine network responsible for the typical disulfide bond architecture of SRCR domains, and several perfectly conserved residues (Gly397, Glu400, Ala420, Val422, Leu427, and Val483) mapped to the highly stable core. These, owing to their localization and conservation, are likely to provide essential contributions to the fold of SRCR domains. Conversely, three additional highly conserved residues (Gly391, Gly457, and Glu459), mapped to the more flexible loop regions, and the β4 secondary structure element, which displays structural but not compositional conservation, might contribute to SRCR domain function.

Among the more conserved regions two stretches, spanning Gly397–Val401 and Trp407–Asp412, respectively, should be mentioned owing to their documented evolutionary conservation.[27] The Gly397–Val401 stretch likely contributes functionally, as similar consensus sequences were seen to mediate protein–target interaction in other SRCR-SF members including MARCO,[28] DMBT1,[35] CD163,[29] as well as structurally, given its mapping to the highly conserved β1 secondary structure element. Conversely, the Trp407–Asp412 stretch has yet to be assigned with a functional role, but it is thought that its conservation is structurally related.

Owing to structural conservation, it is possible to suppose a protein-target role for mmNT-SRCR3. Additional observations drawn from our structure and those of other SRCR-SF members might hint at more specific functions in relation to NT. Notably, analysis of surface charge distribution of mmNT-SRCR3 evidenced a negatively charged patch corresponding to a cluster of residues (Asp412, Asp413, His473, Asn474, and Glu479), conserved in the metal-ion binding SRCR domain of MARCO.[18] In NT, these residues might be involved in binding of Ca^{2+}, known to be important for NT enzymatic activity.[30] A plausible hypothesis, given that ITC experiments showed Ca^{2+} binding for mmNT-SRCR3 [Fig. 4(A)] and that other SRCR-SF members (MARCO, DMBT1, and CD163) also display Ca^{2+}-dependent modulation mapped to their SRCR domains.[18,29,31] Further analogies can be speculated regarding the nature of possible binding partners. Several SRCR-SF members are known to bind to bacteria,[26,31] lipopolysaccarides (LPS) directly[28,32] or glycoproteins[33] via their SRCR domains. Similarly, NT SRCR domains might mediate interactions with the heavily glycosylated substrate Agrin.[19]

Finally, cross-comparison of human and murine NT evidenced how the first SRCR domain in the human NT sequence represents the principal source of divergence between the two homologs. Surprisingly, also the residues putatively coordinating Ca^{2+} in mmNT-SRCR3 are well conserved across all NT SRCR domains, except for human NT-SRCR1 (Fig. S4). An observation which hints at a possible lack of Ca^{2+} binding for that SRCR domain and that, in conjunction with its absence in murine NT, opens to speculation regarding the function and relevance of the human NT SRCR1.

In conclusion, it seems plausible that the broad protein-target interaction attributed to SRCR domains might be mediated by the highly conserved nature of their core structure. While the more structurally and compositionally variable regions could be responsible for more nuanced and specific functional aspects such as ligand recognition. Finally, in regard to NT, our characterization of mmNT-SRCR3 allowed us to identify a Ca^{2+} binding site and infer, by comparison, a possible role for this domain in Agrin recognition.

Materials and Methods

Molecular cloning, recombinant expression, and purification

The cDNA encoding for the third SRCR domain of murine Neurotrypsin (UniProt id O08762, residues 383–494) was obtained from Source Bioscience (I.M.A.G.E. clone ID 3665834), amplified with a Phusion DNA polymerase (Thermo Fisher Scientific) using oligos AAAGATCTGGTTTTCCCATCAGACTAGTGG ATG (forward) and AAGCGGCCGCACTTGATGCTTT

CTTCTCTAAATAG (reverse), and inserted into the pCIOX recombinant expression vector (Addgene), yielding the final construct bearing an N-terminal 8-His-SUMO tag. This plasmid was transformed into chemically competent *E. coli* SHuffle T7 cells (New England Biolabs), which were used for protein production.

A small culture of transformed cells was grown over night shaking at 30 °C in LB + kanamycin (50 μg/mL) and used to inoculate, in a 1:100 ratio, a larger volume (1–6 L) of autoinducing ZYP-5052 media[34] for large-scale production. Inoculated media was incubated shaking at 30 °C for 5 h, after which the temperature was lowered to 20 °C to induce recombinant protein expression, and further incubated for 20 h.

Cells were harvested by centrifugation (5000*g*, 15 min, 4 °C) with a swinging bucket centrifuge (Beckman Coulter). The supernatant was discarded and the cells were resuspended in buffer A (25 m*M* 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES)/NaOH, 0.5 M NaCl, pH 8) in a 1:5 (*w*/*v*) wet cell pellet-to-buffer ratio. This suspension was placed on ice for 30 minutes and then sonicated (80% amplitude, 8 cycles with 40 s on/20 s off pulses) to lyse the cells. Cell debris was removed by centrifugation (10,000 *g*, 50 min, 4 °C) and the supernatant was filtered with a 0.45 μm syringe filter (Sartorius). The clarified lysate was loaded on a 5 mL HisTrap (GE Healthcare) column, pre-conditioned with buffer A, at a rate of 1 mL/min. Unbound material was washed from the column with buffer A and non-specifically bound contaminants were removed with buffer A supplemented with 25 m*M* imidazole. mmNT-SRCR3 was eluted by further increasing the imidazole concentration to 250 m*M*. The eluted fractions were pooled, supplemented with SUMO protease (1.2 mg/mL stock, 1:300 *v*/*v*) and dialyzed overnight at 4 °C against buffer A. Following dialysis the sample was subject to centrifugation (5000*g*, 4 °C, 15 min) and the supernatant was loaded onto a 1 mL HisTrap column (GE Healthcare). Successful removal of the His-SUMO tag was evaluated through sample recovery in the flow through fractions, which were further purified by gel filtration on a Superdex 75 10/300 GL column (GE Healthcare) equilibrated with 25 m*M* HEPES/NaOH, 100 m*M* NaCl, pH 8. At each step of protein purification, samples were collected and analyzed using SDS-PAGE. The final yield was ≈2 mg of pure protein per gram of bacterial cells.

Crystallization, X-ray data collection, and processing

For crystallization screening, the protein was concentrated to 20 mg/mL with a 5 kDa MWCO Vivaspin12 concentrator (Sartorius). Crystallization was performed using the sitting-drop vapor diffusion method at both 4 and 20 °C. Drops were set up in a volumetric 1:1 protein-to-precipitant solution ratio using an Oryx8 crystallization robot (Douglas Instruments) with MRC 96-well PS plates (SwissSci). Initial screening performed with commercial screens yielded thin microcrystals, not suitable for diffraction testing, in numerous conditions. Optimization of initial crystallization hits was performed by hand using the sitting drop method. The best crystals were obtained at 4 °C, over the span of a month, by mixing mmNT-SRCR3 concentrated at 19 mg/mL with a solution composed of 16% PEG3350, 0.1 M Tri-sodium citrate, pH 7.5. Prior to flash-cooling in liquid nitrogen, crystals were harvested with nylon cryoloops (Hampton Research) and briefly soaked into a cryo-protectant solution (19% PEG3350, 0.1 M Tri-sodium citrate, 20% Glycerol). X-ray diffraction data were collected at 100 K at the ID30A-3 beamline of the ESRF synchrotron. Data were indexed and integrated using *XDS*[35] and scaled using *Aimless*.[36] Data collection statistics are summarized in Table I.

Structure determination and refinement

The structure of mmNT-SRCR3 was solved at 1.7 Å by molecular replacement using the structure of the Mac-2 binding protein (M2BP) SRCR domain (pdb: 1BY2)[37] as search model. This model (50% sequence identity to mmNT-SRCR3) was selected based on conservation of sequence identity as evaluated using *NCBI BLAST*.[38] Prior to molecular replacement, the sequences of mmNT-SRCR3 and M2BP SRCR were aligned using *EBI MUSCLE*[39] and non-conserved residues were adjusted using *CHAINSAW*.[40] The resulting model was used in molecular replacement with *PHASER*.[41] Two copies of the search model were found constituting the asymmetric unit, with a V_m of 3.11 Å^3 and 60% solvent content. The structure was refined with successive steps of manual building in *COOT*[42] and automated refinement with *REFMAC5*.[43] Model validation was performed with *MolProbity*.[44] Refinement statistics for the final model are reported in Table I as deposited to the PDB under accession code 6H8M. Electrostatic surface calculations and representations were generated with *CCP4mg*.[21] Other structural images were generated with *UCSF Chimera*.[45]

SEC-SAXS analysis

Size exclusion chromatography coupled to small-angle X-ray scattering (SEC-SAXS) analysis was performed at the BM29 beamline of the ESRF synchrotron in Grenoble (France) using a protocol adapted from.[46] The protein was concentrated to 15 mg/mL and a 15 μL sample was run on a Superdex 75 Increase 3.2/300 column (GE Healthcare) equilibrated in 25 m*M* HEPES/NaOH, 0.1 M NaCl, pH 8 and mounted on a Nexera High Pressure Liquid/Chromatography (HPLC; Shimadzu) system. SAXS data were collected from the sample capillary mounted on line with the HPLC system, using a Pilatus 1 M detector (Dectris) positioned at distance of 2.87 m allowing a global q range of 0.03–4.5 nm with

12.5 keV energy. Analysis of scattering intensities was performed using *CHROMIXS*[47] and the *ATSAS* suite.[48] Comparison of in-solution scattering to crystallographic data was carried out with CRYSOL.[22] Details of SEC-SAXS data collection and analysis are summarized in Table II as deposited in SASBDB under accession code SASDES5.

ITC

Titrations were carried out at 25 °C in 100 m*M* NaCl, 25 m*M* HEPES/NaOH, pH 8 using the high feedback mode of a MicroCal ITC200 instrument (Malvern Instruments, Worcestershire, UK). The concentration of mmNT-SRCR3 in the cell was 1.2 m*M*, whereas the $CaCl_2$ solution in the syringe was at 120 m*M*. The first injection was kept to the minimum volume of 0.1 μL to allow complete equilibration of the cell. The following 19 titrations of $CaCl_2$ were performed using a 2 μL injection volume with a 120 s. Time interval while maintaining stirring at 750 rpm. The heat of dilution of 120 m*M* $CaCl_2$ was determined by performing a second titration in which the cell was filled only with buffer while all other experimental parameters remained unaltered. The net contribution of binding of $CaCl_2$ to mmNT-SRCR3 was calculated by subtracting the heat of dilution of $CaCl_2$. Data were analyzed using the "*One Set of Sites*" curve fitting model (MicroCal ITC200 Origin).

Acknowledgments

We thank the European Synchrotron Radiation Facility (ESRF) for the provision of synchrotron radiation facilities and beamline scientists of the ESRF and the European Molecular Biology Laboratory for assistance. This work was supported by a Cariplo Foundation Grant (id. 2014-0881), the Giovanni Armenise-Harvard Career Development Award (2013), the "Programma Rita Levi-Montalcini" from the Italian Ministry of University and Research (MIUR), and by the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Program (2018–2022) – Department of Biology and Biotechnology "L. Spallanzani", University of Pavia.

Conflict of interest statement

The authors declare no competing interests.

Author Contributions

F.F. conceived and supervised the project. A.C. cloned, expressed, purified, crystallized, and solved the structure of mmNT-SRCR3. A.C. and F.F. analyzed crystallographic and SEC-SAXS data. G.C. performed ITC measurements and associated data processing. A.C. prepared the figures. A.C. and F.F. wrote the paper. All authors read and approved the final manuscript.

References

1. Gschwend TP, Krueger SR, Kozlov SV, Wolfer DP, Sonderegger P (1997) Neurotrypsin, a novel multidomain serine protease expressed in the nervous system. Mol Cell Neurosci 9:207–219.
2. Yamamura Y, Yamashiro K, Tsuruoka N, Nakazato H, Tsujimura A, Yamaguchi N (1997) Molecular cloning of a novel brain-specific serine protease with a kringle-like structure and three scavenger receptor cysteine-rich motifs. Biochem Biophys Res Commun 239:386–392.
3. Proba K, Gschwend TP, Sonderegger P (1998) Cloning and sequencing of the cDNA encoding human neurotrypsin. Biochim Biophys Acta 1396:143–147.
4. Molinari F, Rio M, Meskenaite V, Encha-Razavi F, Auge J, Bacq D, Briault S, Vekemans M, Munnich A, Attie-Bitach T, Sonderegger P, Colleaux L (2002) Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation. Science 298:1779–1781.
5. Reif R, Sales S, Hettwer S, Dreier B, Gisler C, Wolfel J, Luscher D, Zurlinden A, Stephan A, Ahmed S, Baici A, Ledermann B, Kunz B, Sonderegger P (2007) Specific cleavage of agrin by neurotrypsin, a synaptic protease linked to mental retardation. FASEB J 21:3468–3478.
6. Stephan A, Mateos JM, Kozlov SV, Cinelli P, Kistler AD, Hettwer S, Rulicke T, Streit P, Kunz B, Sonderegger P (2008) Neurotrypsin cleaves agrin locally at the synapse. FASEB J 22:1861–1873.
7. Butikofer L, Zurlinden A, Bolliger MF, Kunz B, Sonderegger P (2011) Destabilization of the neuromuscular junction by proteolytic cleavage of agrin results in precocious sarcopenia. FASEB J 25:4378–4393.
8. Mitsui S, Hidaka C, Furihata M, Osako Y, Yuri K (2013) A mental retardation gene, motopsin/prss12, modulates cell morphology by interaction with seizure-related gene 6. Biochem Biophys Res Commun 436:638–644.
9. Shimizu-Nishikawa K, Kajiwara K, Sugaya E (1995) Cloning and characterization of seizure-related gene, SEZ-6. Biochem Biophys Res Commun 216:382–389.
10. Gunnersen JM, Kim MH, Fuller SJ, De Silva M, Britto JM, Hammond VE, Davies PJ, Petrou S, Faber ES, Sah P, Tan SS (2007) Sez-6 proteins affect dendritic arborization patterns and excitability of cortical pyramidal neurons. Neuron 56:621–639.
11. Boeuf S, Borger M, Hennig T, Winter A, Kasten P, Richter W (2009) Enhanced ITM2A expression inhibits chondrogenic differentiation of mesenchymal stem cells. Differentiation 78:108–115.
12. Lagha M, Mayeuf-Louchart A, Chang T, Montarras D, Rocancourt D, Zalc A, Kormish J, Zaret KS, Buckingham ME, Relaix F (2013) Itm2a is a Pax3 target gene, expressed at sites of skeletal muscle formation in vivo. PLoS One 8:e63143.
13. Tai TS, Pai SY, Ho IC (2014) Itm2a, a target gene of GATA-3, plays a minimal role in regulating the development and function of T cells. PLoS One 9:e96535.
14. Van den Plas D, Merregaert J (2004) Constitutive overexpression of the integral membrane protein Itm2A enhances myogenic differentiation of C2C12 cells. Cell Biol Int 28:199–207.
15. Freeman M, Ashkenas J, Rees DJ, Kingsley DM, Copeland NG, Jenkins NA, Krieger M (1990) An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II murine macrophage scavenger receptors. Proc Natl Acad Sci USA 87:8810–8814.
16. Resnick D, Chatterton JE, Schwartz K, Slayter H, Krieger M (1996) Structures of class A macrophage scavenger receptors. Electron microscopic study of flexible, multidomain, fibrous proteins and determination of the disulfide

bond pattern of the scavenger receptor cysteine-rich domain. J Biol Chem 271:26924–26930.
17. Martinez VG, Moestrup SK, Holmskov U, Mollenhauer J, Lozano F (2011) The conserved scavenger receptor cysteine-rich superfamily in therapy and diagnosis. Pharmacol Rev 63:967–1000.
18. 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.
19. Gisler C, Luscher D, Schatzle P, Durr S, Baici A, Galliciotti G, Reif R, Bolliger MF, Kunz B, Sonderegger P (2013) Zymogen activation of neurotrypsin and neurotrypsin-dependent agrin cleavage on the cell surface are enhanced by glycosaminoglycans. Biochem J 453:83–100.
20. Ozhogina OA, Grishaev A, Bominaar EL, Patthy L, Trexler M, Llinas M (2008) NMR solution structure of the neurotrypsin Kringle domain. Biochemistry 47: 12290–12298.
21. McNicholas S, Potterton E, Wilson KS, Noble ME (2011) Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr D67:386–394.
22. Svergun D, Barberato C, Koch MHJ (1995) CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Cryst 28:768–773.
23. Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38:W545–W549.
24. 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–W350.
25. Purushotham S, Deivanayagam C (2014) The calcium-induced conformation and glycosylation of scavenger-rich cysteine repeat (SRCR) domains of glycoprotein 340 influence the high affinity interaction with antigen I/II homologs. J Biol Chem 289:21877–21887.
26. Bikker FJ, Ligtenberg AJ, End C, Renner M, Blaich S, Lyer S, Wittig R, van't Hof W, Veerman EC, Nazmi K, de Blieck-Hogervorst JM, Kioschis P, Nieuw Amerongen AV, Poustka A, Mollenhauer J (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.
27. Yap NV, Whelan FJ, Bowdish DM, Golding GB (2015) The evolution of the scavenger receptor cysteine-rich domain of the class a scavenger receptors. Front Immunol 6:342.
28. Brannstrom A, Sankala M, Tryggvason K, Pikkarainen T (2002) Arginine residues in domain V have a central role for bacteria-binding activity of macrophage scavenger receptor MARCO. Biochem Biophys Res Commun 290: 1462–1469.
29. Madsen M, Moller HJ, Nielsen MJ, Jacobsen C, Graversen JH, van den Berg T, Moestrup SK (2004) Molecular characterization of the haptoglobin. Hemoglobin receptor CD163. Ligand binding properties of the scavenger receptor cysteine-rich domain region. J Biol Chem 279:51561–51567.
30. Reif R, Sales S, Dreier B, Luscher D, Wolfel J, Gisler C, Baici A, Kunz B, Sonderegger P (2008) Purification and enzymological characterization of murine neurotrypsin. Protein Expr Purif 61:13–21.
31. 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.
32. Sankala M, Brannstrom A, Schulthess T, Bergmann U, Morgunova E, Engel J, Tryggvason K, Pikkarainen T (2002) Characterization of recombinant soluble macrophage scavenger receptor MARCO. J Biol Chem 277:33378–33385.
33. Bowen MA, Bajorath J, Siadak AW, Modrell B, Malacko AR, Marquardt H, Nadler SG, Aruffo A (1996) The amino-terminal immunoglobulin-like domain of activated leukocyte cell adhesion molecule binds specifically to the membrane-proximal scavenger receptor cysteine-rich domain of CD6 with a 1:1 stoichiometry. J Biol Chem 271:17390–17396.
34. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41: 207–234.
35. Kabsch W (2010) Xds. Acta Crystallogr D66:125–132.
36. Evans PR (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr D67:282–292.
37. Hohenester E, Sasaki T, Timpl R (1999) Crystal structure of a scavenger receptor cysteine-rich domain sheds light on an ancient superfamily. Nat Struct Biol 6:228–232.
38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
39. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.
40. Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. J Appl Cryst 41:641–643.
41. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Cryst 40:658–674.
42. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D66: 486–501.
43. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D67:355–367.
44. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D66:12–21.
45. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.
46. Brennich ME, Round AR, Hutin S (2017) Online size-exclusion and ion-exchange chromatography on a SAXS beamline. J Vis Exp 2017:54861.
47. Panjkovich A, Svergun DI (2017) CHROMIXS: automatic and interactive analysis of chromatography-coupled small angle X-ray scattering data. Bioinformatics 34: 1944–1946.
48. Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT, Kikhney AG, Hajizadeh NR, Franklin JM, Jeffries CM, Svergun DI (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J Appl Cryst 50:1212–1225.
49. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. Science 336:1030–1033.
50. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42:W320–W324.