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

(Article begins on next page)

20 April 2024

Structural characterization of the

third scavenger receptor cysteine-rich domain

of murine Neurotrypsin

Anselmo Canciani[1], Gianluca Catucci[2], 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).

http://fornerislab.unipv.it; [2]Department of Life Sciences and Systems Biology,

University of Torino, 10123 Turin, Italy

* Correspondence to: Federico Forneris, Tel. +39 0382 985228, Fax +39 0382 528496, e-mail: federico.forneris@unipv.it

Running title: *Crystal structure of Neurotrypsin SRCR3*

Total Number of Manuscript Pages: 26

Total Number of Supplementary Material Pages: 5

Total Number of Tables: 3

Total Number of Figures: 4

Description of Supplementary Materials:

Supplementary materials consist of a single PDF file including 4 supplementary figures. All supplementary materials are cited within the manuscript.

25 **Abstract**

26 Neurotrypsin (NT) is a multi-domain serine-protease of the nervous system with only one

27 known substrate: the large proteoglycan Agrin. NT has seen to be involved in the

28 maintenance/turnover of neuromuscular junctions and in processes of synaptic plasticity in

29 the central nervous system. Roles which have been tied to its enzymatic activity, localized in

30 the C-terminal serine-protease (SP) domain. However the purpose of NT's remaining 3-4

31 scavenger receptor cysteine-rich (SRCR) domains is still unclear. We have determined the

32 crystal structure of the third SRCR domain of murine NT (mmNT-SRCR3), immediately

33 preceding the SP domain and performed a comparative structural analysis using homologous

34 SRCR structures. Our data and the elevated degree of structural conservation with

35 homologous domains highlight possible functional roles for NT SRCRs. Computational and

36 experimental analyses suggest the identification of a putative binding region for Ca^{2+} ions,

37 known to regulate NT enzymatic activity. Furthermore, sequence and structure comparisons

38 allow to single out regions of interest that, in future studies, might be implicated in Agrin

39 recognition/binding or in interactions with as of yet undiscovered NT partners.

40

44

45

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 90's (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 domain (SP) (1; 2). In comparison to 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) (Figure 1 A, Supplementary Fig. 1).

Produced predominantly by neurons of the central and peripheral nervous systems, NT is stocked in synaptic vesicles and released in an activity-dependant 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).

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

70   been tied to one of the accessory domains: the Kr domain (8). Conversely, the domain(s)

71   responsible for interacting with Itm2a has yet to be identified.

72   The scavenger receptor cysteine-rich (SRCR) domain is a small (90-110 residues)

73   distinguishing element of a broad superfamily of proteins whose members span different

74   functional roles (15-17). It displays a highly conserved structure with five anti-parallel beta

75   strands cradling an alpha helix and a distinct disulfide bridge architecture (18; 20-23). A

76   feature which has been used to identify two types of SRCR domains, A and B, bearing

77   respectively 6 and 8 cysteines. While the number of cysteines differs, the relative pairing is

78   surprisingly consistent and tends to form as follows: C1-C4, C2-C7, C3-C8 and C5-C6. The

79   first pair distinguishes types A and B as it is present exclusively in the latter (16). Members

80   of the SRCR superfamily can share little functional similarity playing roles in immune

81   response (18), cell differentiation (24), apoptosis (25) and tumor suppression alike. While it

82   has been difficult to assign a univocal function to SRCR domains, a commonality can be

83   drawn across several SRCR superfamily members, broadly attributing the SRCR domain with

84   a role in protein-protein/protein-substrate interaction (26). It is therefore likely NT SRCR

85   domains may mediate interactions with the surrounding environment. They could be

86   responsible for the reported glycosaminoglycan and heparin binding capabilities (27),

87   contribute to its specificity, mediate its localized lingering at the synapse (19), contribute to

88   its specificity, mediate its localized lingering at the synapse (20) and/or be involved in

89   binding to, as of yet, unidentified partners. Of NT's accessory domains only the structure of

90   the Kr domain has been successfully determined and assigned a possible role (21; 8). The

91   lack of structural information pertaining to the remaining domains limits our functional

92   understanding of NT. As such new structural data on the SP or SRCR domains might provide

93   key insights on the mechanisms underlying NT's biological role.

94   Here we report the recombinant production, purification, crystallization and crystal structure

95   determination of the third SRCR domain of murine NT (mmNT-SRCR3), and analysis of its

96   molecular architecture based on comparisons with homologous SRCR domains. Our

97   structural data constitute a solid additional step towards the understanding of NT's molecular

98   interactions and biological functions.

99

100   **Results**

101   *Purification of mmNT-SRCR3*

102   Owing to the high cysteine content of extracellular SRCR domains, we produced the mmNT-

103   SRCR3 in soluble form using a *E. coli* strain particularly adapted to facilitate disulfide bond

104   formation. Purification of mmNT-SRCR3 was performed using a two-step immobilized Ni-

105   affinity chromatography (Ni-IMAC) followed by size-exclusion chromatography (SEC)

106   (Supplementary Fig. 2). The purified material was used for crystallization experiments and

107   protein characterization studies in solution.

108

109   *Crystal structure of mmNT-SRCR3*

110   A single mmNT-SRCR3 crystal allowed complete X-ray data collection and structure

111   determination (Figure 1 B); data processing and structure refinement statistics are reported in

112   table 1. This process highlighted the presence of two mmNT-SRCR3 monomers in the

113   asymmetric unit assembling into what appeared to be a crystallographic dimer (Figure 1 C).

114   Each monomer presents a very compact fold with a central alpha helix (α1) nested in five

115   beta strands (β1-β5) forming a twisted anti-parallel cradle (Figure 1 B). The B-factor

116   distribution across the structure highlights the high stability of the central core and

117   progressive flexibility of the peripheral loops interconnecting the secondary structure

118   elements (Figure 1 C). Of particular interest is the Val466–Asn474 segment, whose flexibility

119 is such that the electron density map of that region is very poor despite the 1.7-Å resolution

120 (Supplementary Fig. 3).

121 Each mmNT-SRCR3 model spans 107 residues and lacks the final C-terminal "KKASS"

122 sequence, owing to the high mobility of this five-residue tail. Superposition of the two

123 monomers found in the asymmetric unit and analysis of the root mean square deviation

124 (r.m.s.d) shows that the only visible significant difference is located in the aforementioned

125 Val466–Asn474 loop, further supporting its flexible nature (Figure 1 D). Calculations of

126 surface electrostatic potentials using CCP4mg (22) displayed a disc-like shape with mixed

127 charge distribution along its sides and two opposite faces, one almost fully hydrophobic and

128 the other displaying a patch of negative charges (Figure 1 E).

129

130 *SEC-SAXS analysis*

131 In order to assess the oligomeric state of the purified protein sample in solution, and thus

132 ascertain the nature of the mmNT-SRCR3 dimer observed in the crystal packing (Figure 1 C),

133 we performed a SEC-small angle x-ray scattering (SEC-SAXS) experiment. The initial

134 chromatogram displayed a well-defined peak with good scattering intensities and a very

135 homogeneous mass distribution (Figure 2 A). mmNT-SRCR3 presents as a monodisperse

136 species with an MW of 12 kDa calculated from averaged peak intensities (Table 2). This is in

137 line with the 12.5 kDa calculated for a monomer of this construct and incompatible with

138 higher order oligomeric species. *CRYSOL* (23) was used to compare the averaged scattering

139 curve of the SEC-SAXS peak to a theoretically calculated scattering curve for our

140 crystallographic structure. The single mmNT-SRCR3 monomer displayed a very good fit (χ^2

141 = 1.4) with the in-solution data (Figure 2 B), ruling the dimer as induced by the

142 crystallization process.

143

Comparison with other SRCR domains

To better understand the functional role of NT's SRCR domains, we used the *DALI* server (24) 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 8 hits (Table 3). While limited in sequence identity, structural superpositions with mmNT-SRCR3 (Figure 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* (25) 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 (Figure 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 (Figure 3 C).

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 (Figure 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 (Figure 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 (Figure 3 B).

169   Conversely, the remaining Gly391, Gly457 and Glu459 locate on the more flexible, and

170   generally less conserved, external loops (Figure 3 B).

171   Comparison of mmNT-SRCR3 with other mmNT SRCR domains evidenced a certain extent

172   of conservation (44-58% sequence identity) comparable to that of the highest *DALI* matches

173   (Supplementary Fig. 1 B). Extending this analysis to include human NT evidenced how its

174   SRCR domains 2-4 correspond, respectively, to mmNT SRCR domains 1-3, while the first

175   human SRCR domain is not conserved across the two homologs (Supplementary Fig. 1B).

176

177   *Evidence for Ca^{2+} binding*

178   Given the significant contributions of Ca^{2+} to NT activity (5), and the reported Ca^{2+}-based

179   modulation of SRCR-mediated interactions (18; 26), we investigated whether mmNT-SRCR3

180   might contribute in similar fashion. Therefore, an Isothermal Titration Calorimetry (ITC)

181   experiment was conducted in presence of $CaCl_2$. This evidenced a Ca^{2+} binding to mmNT-

182   SRCR3 (Figure 4 A), with K_d = 10.1 mM, ΔH = -7.7 Kcal/mol, ΔS = -16.8 cal/mol/deg.

183   A comparison of mmNT-SRCR3 with the structure of the monomeric MARCO SRCR

184   domain (18) highlighted a cluster of highly conserved residues coordinating a bivalent metal-

185   ion in MARCO (Figure 4 B). These correspond to amino acids Asp412-Asp413, His473,

186   Asn474 and Glu479, that significantly contribute mmNT-SRCR3's negatively charged patch.

187   Interestingly the same residues were seen to be perfectly conserved in the DMBT1 SRCR-1

188   domain, known as well to display Ca^{2+} binding (27) (Figure 4 C).

189

190   **Discussion**

191   The scavenger receptor cysteine-rich domain groups numerous proteins with different

192   functions in a large superfamily (15). However, the understanding of the specific function of

193   SRCR domains is, to date, limited to a general consensus of protein-protein/protein-ligand

194   interaction. The solution of the structure of the third SRCR domain of murine Neurotrypsin

195   (mmNT-SRCR3) and its analysis highlighted several interesting features that might provide a

196   functional insight both in regards to NT and SRCR domains in general. Most striking of all

197   was the degree of similarity with the available homologous structures despite a generally

198   poor sequence conservation (Table 3).

199   At the core of mmNT-SRCR3 is an extremely stable secondary structure element

200   organization which seems to be perfectly conserved across available structures. Such

201   variability could be traced to the more flexible external loops, most notably the Val466-

202   Asn474 loop. This region not only displayed the highest mobility across homologs but also

203   within the mmNT-SRCR3 monomers observed in the crystallographic dimer. Mapping of

204   sequence conservation to the structure of mmNT-SRCR3 showed that the least conserved

205   residues were located on this and other peripheral flexible loops, while the core secondary

206   structure elements bore the majority of highly conserved amino acids. Of particular note are:

207   a six-cysteine network responsible for the typical disulfide bond architecture of SRCR

208   domains, and several perfectly conserved residues (Gly397, Glu400, Ala420, Val422,

209   Leu427, and Val483) mapped to the highly stable core. These, owing to their localization and

210   conservation, are likely to provide essential contributions to the fold of SRCR domains.

211   Conversely, three additional highly conserved residues (Gly391, Gly457 and Glu459),

212   mapped to the more flexible loop regions, and the β4 secondary structure element, which

213   displays structural but not compositional conservation, might contribute to SRCR domain

214   function.

215   Among the more conserved regions two stretches, spanning Gly397–Val401 and Trp407–

216   Asp412 respectively, should be mentioned owing to their documented evolutionary

217   conservation (28). The first likely contributes functionally, as similar consensus sequences

218   were seen to mediate protein-target interaction in other SRCR-SF members including

219 MARCO (29), DMBT1 (35), CD163 (30), as well as structurally, given its mapping to the

220 highly conserved β1 secondary structure element. Conversely, the second has yet to be

221 assigned with a functional role but it is thought that the conservation of this stretch is

222 structurally related.

223 Owing to structural conservation it is possible to suppose a protein-target role for mmNT-

224 SRCR3. Additional observations drawn from that structure and other SRCR-SF members

225 might hint at its more specific function in relation to NT. Notably, analysis of surface charge

226 distribution of mmNT-SRCR3 evidenced a negatively charged patch corresponding to a

227 cluster of residues (Asp412, Asp413, His473, Asn474 and Glu479), conserved in the metal-

228 ion binding MARCO SRCR domain (18), that might be involved in Ca^{2+} ion binding known

229 to be important for NT enzymatic activity (31). A plausible hypothesis, given that ITC

230 experiments showed Ca^{2+} binding for mmNT-SRCR3 and that other SRCR-SF members

231 (MARCO, DMBT1 and CD163) also display a Ca^{2+} dependant modulation mapped to their

232 SRCR domains (32; 30; 18). Further analogies can be speculated regarding the nature of

233 possible binding partners. As several SRCR-SF members are known to bind to bacteria (32;

234 27), lipopolysaccarides (LPS) directly (29; 33) or glycoproteins (34) via their SRCR domains.

235 Thus it is possible that NT SRCR domains might mediate similar interactions between NT

236 and its heavily glycosylated substrate Agrin (19).

237 Finally, cross-comparison of human and murine NT evidenced how the first SRCR domain in

238 the human NT sequence represents the principal source of divergence between the two

239 homologs. Surprisingly, also the residues putatively coordinating Ca^{2+} in mmNT-SRCR3 are

240 well conserved across all NT SRCR domains, except for human NT-SRCR1 (Supplementary

241 Fig. 4). An observation which hints at a possible lack of Ca^{2+} binding for that SRCR domain

242 and that, in conjunction with its absence in murine NT, opens to speculation regarding the

243 function and relevance of the human NT SRCR1.

244   In conclusion, it seems plausible that the broad protein-target interaction attributed to SRCR

245   domains might be mediated by the highly conserved nature of their core structure. While the

246   more structurally and compositionally variable regions could be responsible for more

247   nuanced and specific functional aspects such as ligand recognition. Finally in regards to NT,

248   our characterization of mmNT-SRCR3 allowed us to identify a Ca^{2+} binding site and infer, by

249   comparison, a possible role for this domain in Agrin recognition.

250

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 AA<u>AGATCT</u>GGTTTTCCCATCAGACTAGTGGATG (forward) and AA<u>GCGGCCGC</u>ACTTGATGCTTTCTTCTCTAAATAG (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 *Escherichia 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 (35) 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 x 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 mM 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 x *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

276 column with buffer A and non-specifically bound contaminants were removed with buffer A

277 supplemented with 25 mM imidazole. mmNT-SRCR3 was eluted by further increasing the

278 imidazole concentration to 250 mM. The eluted fractions were pooled, supplemented with

279 SUMO protease (1.2 mg/mL stock, 1:300 *v/v*) and dialyzed overnight at 4 °C against buffer

280 A. Following dialysis the sample was subject to centrifugation (5000 x g, 4 °C, 15 min) and

281 the supernatant was loaded onto a 1 mL HisTrap column (GE Healthcare). Successful

282 removal of the His-SUMO tag was evaluated through sample recovery in the flow through

283 fractions, which were further purified by gel filtration on a Superdex 75 10/300 GL column

284 (GE Healthcare) equilibrated with 25 mM HEPES/NaOH, 100 mM NaCl, pH 8. At each step

285 of protein purification, samples were collected and analysed using SDS-PAGE. The final

286 yield was ≈ 2 mg of pure protein per gram of bacterial cells.

287

288 *Crystallization, X-ray Data collection and processing*

289 For crystallization screening the protein was concentrated to 20 mg/mL with a 5 kDa MWCO

290 Vivaspin12 concentrator (Sartorius). Crystallization was performed using the sitting-drop

291 vapor diffusion method at both 4 and 20 °C. Drops were set up in a volumetric 1:1 protein-to-

292 precipitant solution ratio using an Oryx8 crystallization robot (Douglas Instruments) with

293 MRC 96-well PS plates (SwissSci). Initial screening performed with commercial screens

294 yielded thin microcrystals, not suitable for diffraction testing, in numerous conditions.

295 Optimization of initial crystallisation hits was performed by hand using the sitting drop

296 method. The best crystals were obtained at 4 °C, over the span of a month, by mixing

297 mmNT-SRCR3 concentrated at 19 mg/mL with a solution composed of 16% PEG3350, 0.1

298 M Tri-Sodium citrate, pH 7.5. Prior to flash-cooling in liquid nitrogen, crystals were

299 harvested with nylon cryoloops (Hampton Research) and briefly soaked into a cryo-

300 protectant solution (19% PEG3350, 0.1 M Tri-Sodium citrate, 20% Glycerol). X-ray

301 diffraction data were collected at 100 K at the ID30A-3 beamline of the ESRF synchrotron.

302 Data were indexed and integrated using *XDS* (36) and scaled using *Aimless* (37). Data

303 collection statistics are summarized in Table 1.

304

305 *Structure determination and refinement*

306 The structure of mmNT-SRCR3 was solved at 1.7 Å by molecular replacement using the

307 structure of the Mac-2 binding protein (M2BP) SRCR domain (pdb: 1BY2) (38) as search

308 model. This model (50% sequence identity to mmNT-SRCR3) was selected based on

309 conservation of sequence identity as evaluated using *NCBI BLAST* (39). Prior to molecular

310 replacement, the sequences of mmNT-SRCR3 and M2BP SRCR were aligned using *EBI*

311 *MUSCLE* (40) and non-conserved residues were adjusted using *CHAINSAW* (41). The

312 resulting model was used in molecular replacement with *PHASER* (42). Two copies of the

313 search model were found constituting the asymmetric unit, with a V_m of 3.11 Å^3 and 60%

314 solvent content. The structure was refined with successive steps of manual building in *COOT*

315 (43) and automated refinement with *REFMAC5* (44). Model validation was performed with

316 *MolProbity* (45). Refinement statistics for the final model are reported in Table 1 as

317 deposited to the PDB under accession code 6H8M. Electrostatic surface calculations and

318 representations were generated with *CCP4mg* (22). Other structural images were generated

319 with *UCSF Chimera* (46) .

320

321 *SEC-SAXS analysis*

322 Size exclusion chromatography coupled to small-angle X-ray scattering (SEC-SAXS)

323 analysis was performed at the BM29 beamline of the ESRF synchrotron in Grenoble (France)

324 using a protocol adapted from (47). The protein was concentrated to 15 mg/mL and a 15 µL

325 sample was run on a Superdex 75 Increase 3.2/300 column (GE Healthcare) equilibrated in

326 25 mM HEPES/NaOH, 0.1 M NaCl, pH 8 and mounted on a Nexera High Pressure

327 Liquid/Chromatography (HPLC; Shimadzu) system. SAXS data were collected from the

328 sample capillary mounted on line with the HPLC system, using a Pilatus 1 M detector

329 (Dectris) positioned at distance of 2.87 m allowing a global q range of 0.03–4.5 nm with 12.5

330 keV energy. Analysis of scattering intensities was performed using *CHROMIXS* (48) and the

331 *ATSAS* suite (49). Comparison of in-solution scattering to crystallographic data was carried

332 out with CRYSOL (23). Details of SEC-SAXS data collection and analysis are summarized

333 in Table 2 as deposited in SASBDB under accession code SASDES5.

334

335 *Isothermal Titration Calorimetry*

336 Titrations were carried out at 25°C in 100 mM NaCl 25 mM HEPES/NaOH pH8 using the

337 high feedback mode of a MicroCal ITC200 instrument (Malvern Instruments, Worcestershire,

338 UK). The concentration of mmNT-SRCR3 in the cell was 1.2 mM, whereas the $CaCl_2$

339 solution in the syringe was at 120 mM. The first injection was kept to the minimum volume

340 of 0.1 uL to allow complete equilibration of the cell. The following 19 titrations of $CaCl_2$

341 were performed using a 2 uL injection volume with a 120 sec. time interval while

342 maintaining stirring at 750 rpm. The heat of dilution of 120 mM $CaCl_2$ was determined by

343 performing a second titration in which the cell was filled only with buffer while all other

344 experimental parameters remained unaltered. The net contribution of binding of $CaCl_2$ to

345 mmNT-SRCR3 was calculated by subtracting the heat of dilution of $CaCl_2$. Data were

346 analyzed using the "One Set of Sites" curve fitting model (MicroCal ITC200 Origin).

347

348 **Supplementary Material**

349 Supplementary figures and tables are available and linked through the text.

350

Acknowledgements

Conflict of interest statement

The authors declare no competing financial interests.

Author Contributions

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

370 **Table 1. X-ray data collection and refinement statistics**

Data Collection	
Diffraction source	ESRF ID30A-3
Wavelength (Å)	0.983
Space group	P $2_1 2_1 2_1$
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)[1,2]	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} [3]	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^2)	
Protein	65.3
Water	54.0
Ramachandran favoured (%)	97.6
Ramachandran allowed (%)	2.4

371

372 [1] Values for the outer shell are given in parentheses.

373 [2] Resolution limits were determined by applying a cut-off based on the mean intensity correlation coefficient of

374 half-datasets (CC1/2) approximately of 0.5 (50).

375 [3] $R_{sym} = \Sigma \mid I - <I> \mid / \Sigma I$, where I is the observed intensity for a reflection and $<I>$ is the average intensity

376 obtained from multiple observations of symmetry-related reflections.

377

378   **Table 2. 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
Dmax (nm)	5.8
MW estimation (V_c based) (kDa)	12

379

380

381 **Table 3** 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

382

Figure Legends:

384 *Figure 1* - Three-dimensional structure of the third SRCR domain of murine NT. (A) domain

385 organization of mouse and human NT, highlighting the localization of mmNT-SRCR3 within

386 the multidomain enzyme architecture. Colouring of the domains highlights sequence

387 conservation between homologous enzymes, as described in Supplementary Fig. 1. (B)

388 Cartoon representation of mmNT-SRCR3 displaying its main secondary structure elements,

389 α-helix (α1) in blue and β-strands (β1 – β5) in cyan. (C) Cartoon "putty" representation of

390 the two mmNT-SRCR3 monomers found in the crystallographic asymmetric unit, coloured

391 blue-red by B-factors (34.2 – 162.5 $Å^2$). (D) Superposition of the two mmNT-SRCR3

392 molecules found in the asymmetric unit. chain A is shown in red, chain B is shown in cyan.

393 (E) Surface representation with colouring by electrostatic potential (-2.0 – +2.0 V), negative

394 charges in red, positive charges in blue and neutral areas in white. Monomer is presented

395 orientations highlighting main features: uncharged bottom, mixed charges on sides and

396 negatively charged patch on top.

397

398 *Figure 2* – mmNT-SRCR3 SEC-SAXS analysis. (A) SEC-SAXS chromatogram showing

399 mmNT-SRCR3 scattering intensities and profile of mass distribution across the peak. (B)

400 Averaged experimental peak scattering curve, after buffer subtraction, plotted superimposed

401 to a theoretical, *CRYSOL*-derived, scattering curve for an mmNT-SRCR3 monomer (χ^2 =

402 1.4).

403

404 *Figure 3* - Structural comparison of mmNT-SRCR3 with other SRCR domains. (A)

405 Superposition with non-redundant *DALI* matches (Table 3) coloured red-yellow by r.m.s.d.

406 (0.39 – 6.32 Å). (B) Cartoon representation of mmNT-SRCR3 in two orientations, coloured

407 cyan-brown (0 – 100%) by residue conservation after *ConSurf* analysis. The left panel maps

408   highly conserved residues from SRCR structure sequence alignments, while the right panel

409   maps the principal secondary structure elements and the archetypical SRCR disulfide bridge

410   organization. The C1-C4 pair is absent in SRCR type A domains, residues Cys411-Cys475

411   correspond to the C2-C7 pair, Cys424-Cys485 to the C3-C8 pair and Cys455-Cys465 to the

412   C5-C6 pair. (C) Sequence alignment of SRCR structures identified by *DALI* analysis of

413   mmNT-SRCR3. Identical residues boxed in red, residues with 70% conservation boxed in

414   yellow. mmNT-SRCR3 secondary structure elements displayed above alignment and

415   consensus sequence with 70% conservation under alignment. Image created using *ESPRIPT3*

416   (51).

417

418   *Figure 4* – Evidence for mmNT-SRCR3 Ca^{2+} binding. (A) Isothermal calorimetric titration of

419     $CaCl_2$ (120 mM) with purified mmNT-SRCR3. Heat variation generated by each injection of

420   titrant at each time interval (top panel) and the integration of each peak and the amount of

421   heat produced (lower panel). (B) Structural inference of mmNT-SRCR3 binding of metal

422   ions. Top panel, residues contributing to the formation of the negatively charged patch on

423   mmNT-SRCR3 likely bind metal ions. Bottom panel, superopsition of mmNT-SRCR3 (blue)

424   with monomeric MARCO (cyan, PDB: 2OY3) coordinating Mg^{2+} (green sphere) via those

425   same amino acids. Side chains of residues Asp412, Asp413 and Glu479 display almost

426   identical conformations, while Asn474 and His473 show higher flexibility. (C) Sequence

427   alignment of mmNT-SRCR3 against MARCO and DMBT1 SRCR domains known to bind

428   Ca^{2+}. Blue arrows and boxes indicate residues evidenced in (B) which are perfectly conserved

429   across all three proteins. Perfectly conserved residues boxed in red, residues with 70%

430   conservation are boxed in yellow. Evaluation of sequence conservation is shown under

431   alignment; symbols indicate: ! = Ile or Val, $ = Leu or Met, % = Phe or Tyr, # = Asn, Asp,

432   Glu or Gln. Residue numbering refers to mmNT-SRCR3.

433

21

434   **References**

435   1. Gschwend TP, Krueger SR, Kozlov SV, Wolfer DP, Sonderegger P (1997) Neurotrypsin, a

436   novel multidomain serine protease expressed in the nervous system. Mol Cell Neurosci

437   9:207-219. PMID: 9245503

438   2. Yamamura Y, Yamashiro K, Tsuruoka N, Nakazato H, Tsujimura A, Yamaguchi N (1997)

439   Molecular cloning of a novel brain-specific serine protease with a kringle-like structure and

440   three scavenger receptor cysteine-rich motifs. Biochem Biophys Res Commun 239:386-392.

441   PMID: 9344839

442   3. Proba K, Gschwend TP, Sonderegger P (1998) Cloning and sequencing of the cDNA

443   encoding human neurotrypsin. Biochim Biophys Acta 1396:143-147. PMID: 9540828

444   4. Molinari F, Rio M, Meskenaite V, Encha-Razavi F, Auge J, Bacq D, Briault S, Vekemans

445   M, Munnich A, Attie-Bitach T, Sonderegger P, Colleaux L (2002) Truncating neurotrypsin

446   mutation in autosomal recessive nonsyndromic mental retardation. Science 298:1779-1781.

447   PMID: 12459588

448   5. Reif R, Sales S, Hettwer S, Dreier B, Gisler C, Wolfel J, Luscher D, Zurlinden A, Stephan

449   A, Ahmed S, Baici A, Ledermann B, Kunz B, Sonderegger P (2007) Specific cleavage of

450   agrin by neurotrypsin, a synaptic protease linked to mental retardation. FASEB J 21:3468-

451   3478. PMID: 17586728

452   6. Stephan A, Mateos JM, Kozlov SV, Cinelli P, Kistler AD, Hettwer S, Rulicke T, Streit P,

453   Kunz B, Sonderegger P (2008) Neurotrypsin cleaves agrin locally at the synapse. FASEB J

454   22:1861-1873. PMID: 18230682

455   7. Butikofer L, Zurlinden A, Bolliger MF, Kunz B, Sonderegger P (2011) Destabilization of

456   the neuromuscular junction by proteolytic cleavage of agrin results in precocious sarcopenia.

457   FASEB J 25:4378-4393. PMID: 21885656

458   8. Mitsui S, Hidaka C, Furihata M, Osako Y, Yuri K (2013) A mental retardation gene,

459   motopsin/prss12, modulates cell morphology by interaction with seizure-related gene 6.

460   Biochem Biophys Res Commun 436:638-644. PMID: 23770366

461   9. Shimizu-Nishikawa K, Kajiwara K, Sugaya E (1995) Cloning and characterization of

462   seizure-related gene, SEZ-6. Biochem Biophys Res Commun 216:382-389. PMID: 7488116

463   10. Gunnersen JM, Kim MH, Fuller SJ, De Silva M, Britto JM, Hammond VE, Davies PJ,

464   Petrou S, Faber ES, Sah P, Tan SS (2007) Sez-6 proteins affect dendritic arborization patterns

465   and excitability of cortical pyramidal neurons. Neuron 56:621-639. PMID: 18031681

466   11. Boeuf S, Borger M, Hennig T, Winter A, Kasten P, Richter W (2009) Enhanced ITM2A

467   expression inhibits chondrogenic differentiation of mesenchymal stem cells. Differentiation

468   78:108-115. PMID: 19541402

469   12. Lagha M, Mayeuf-Louchart A, Chang T, Montarras D, Rocancourt D, Zalc A, Kormish J,

470   Zaret KS, Buckingham ME, Relaix F (2013) Itm2a is a Pax3 target gene, expressed at sites of

471   skeletal muscle formation in vivo. PLoS One 8:e63143. PMID: 23650549

472   13. Tai TS, Pai SY, Ho IC (2014) Itm2a, a target gene of GATA-3, plays a minimal role in

473   regulating the development and function of T cells. PLoS One 9:e96535. PMID: 24831988

474   14. Van den Plas D, Merregaert J (2004) Constitutive overexpression of the integral

475   membrane protein Itm2A enhances myogenic differentiation of C2C12 cells. Cell Biol Int

476   28:199-207. PMID: 14984746

477   15. Freeman M, Ashkenas J, Rees DJ, Kingsley DM, Copeland NG, Jenkins NA, Krieger M

478   (1990) An ancient, highly conserved family of cysteine-rich protein domains revealed by

479   cloning type I and type II murine macrophage scavenger receptors. Proc Natl Acad Sci U S A

480   87:8810-8814. PMID: 1978939

481   16. Resnick D, Chatterton JE, Schwartz K, Slayter H, Krieger M (1996) Structures of class A

482   macrophage scavenger receptors. Electron microscopic study of flexible, multidomain,

483   fibrous proteins and determination of the disulfide bond pattern of the scavenger receptor

484   cysteine-rich domain. J Biol Chem 271:26924-26930. PMID: 8900177

485   17. Martinez VG, Moestrup SK, Holmskov U, Mollenhauer J, Lozano F (2011) The

486   conserved scavenger receptor cysteine-rich superfamily in therapy and diagnosis. Pharmacol

487   Rev 63:967-1000. PMID: 21880988

488   18. Ojala JR, Pikkarainen T, Tuuttila A, Sandalova T, Tryggvason K (2007) Crystal structure

489   of the cysteine-rich domain of scavenger receptor MARCO reveals the presence of a basic

490   and an acidic cluster that both contribute to ligand recognition. J Biol Chem 282:16654-

491   16666. PMID: 17405873

492   19. Gisler C, Luscher D, Schatzle P, Durr S, Baici A, Galliciotti G, Reif R, Bolliger MF,

493   Kunz B, Sonderegger P (2013) Zymogen activation of neurotrypsin and neurotrypsin-

494   dependent agrin cleavage on the cell surface are enhanced by glycosaminoglycans. Biochem

495   J 453:83-100. PMID: 23560819

496   20. Frischknecht R, Fejtova A, Viesti M, Stephan A, Sonderegger P (2008) Activity-induced

497   synaptic capture and exocytosis of the neuronal serine protease neurotrypsin. J Neurosci

498   28:1568-1579. PMID: 18272678

499   21. Ozhogina OA, Grishaev A, Bominaar EL, Patthy L, Trexler M, Llinas M (2008) NMR

500   solution structure of the neurotrypsin Kringle domain. Biochemistry 47:12290-12298. PMID:

501   18956887

502   22. McNicholas S, Potterton E, Wilson KS, Noble ME (2011) Presenting your structures: the

503   CCP4mg molecular-graphics software. Acta Crystallogr D Biol Crystallogr 67:386-394.

504   PMID: 21460457

505   23. Svergun D, Barberato C, Koch MHJ (1995) CRYSOL - a Program to Evaluate X-ray

506   Solution Scattering of Biological Macromolecules from Atomic Coordinates. Journal of

507   Applied Crystallography 28:768-773.

508   24. Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. Nucleic Acids

509   Res 38:W545-549. PMID: 20457744

510   25. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N (2016)

511   ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation

512   in macromolecules. Nucleic Acids Res 44:W344-350. PMID: 27166375

513   26. Purushotham S, Deivanayagam C (2014) The calcium-induced conformation and

514   glycosylation of scavenger-rich cysteine repeat (SRCR) domains of glycoprotein 340

515   influence the high affinity interaction with antigen I/II homologs. J Biol Chem 289:21877-

516   21887. PMID: 24923446

517   27. Bikker FJ, Ligtenberg AJ, End C, Renner M, Blaich S, Lyer S, Wittig R, van't Hof W,

518   Veerman EC, Nazmi K, de Blieck-Hogervorst JM, Kioschis P, Nieuw Amerongen AV,

519   Poustka A, Mollenhauer J (2004) Bacteria binding by DMBT1/SAG/gp-340 is confined to

520   the VEVLXXXXW motif in its scavenger receptor cysteine-rich domains. J Biol Chem

521   279:47699-47703. PMID: 15355985

522   28. Yap NV, Whelan FJ, Bowdish DM, Golding GB (2015) The Evolution of the Scavenger

523   Receptor Cysteine-Rich Domain of the Class A Scavenger Receptors. Front Immunol 6:342.

524   PMID: 26217337

525   29. Brannstrom A, Sankala M, Tryggvason K, Pikkarainen T (2002) Arginine residues in

526   domain V have a central role for bacteria-binding activity of macrophage scavenger receptor

527   MARCO. Biochem Biophys Res Commun 290:1462-1469. PMID: 11820786

528   30. Madsen M, Moller HJ, Nielsen MJ, Jacobsen C, Graversen JH, van den Berg T, Moestrup

529   SK (2004) Molecular characterization of the haptoglobin.hemoglobin receptor CD163.

530   Ligand binding properties of the scavenger receptor cysteine-rich domain region. J Biol

531   Chem 279:51561-51567. PMID: 15448162

532   31. Reif R, Sales S, Dreier B, Luscher D, Wolfel J, Gisler C, Baici A, Kunz B, Sonderegger P

533   (2008) Purification and enzymological characterization of murine neurotrypsin. Protein Expr

534   Purif 61:13-21. PMID: 18577456

535   32. Bikker FJ, Ligtenberg AJ, Nazmi K, Veerman EC, van't Hof W, Bolscher JG, Poustka A,

536   Nieuw Amerongen AV, Mollenhauer J (2002) Identification of the bacteria-binding peptide

537   domain on salivary agglutinin (gp-340/DMBT1), a member of the scavenger receptor

538   cysteine-rich superfamily. J Biol Chem 277:32109-32115. PMID: 12050164

539   33. Sankala M, Brannstrom A, Schulthess T, Bergmann U, Morgunova E, Engel J,

540   Tryggvason K, Pikkarainen T (2002) Characterization of recombinant soluble macrophage

541   scavenger receptor MARCO. J Biol Chem 277:33378-33385. PMID: 12097327

542   34. Bowen MA, Bajorath J, Siadak AW, Modrell B, Malacko AR, Marquardt H, Nadler SG,

543   Aruffo A (1996) The amino-terminal immunoglobulin-like domain of activated leukocyte cell

544   adhesion molecule binds specifically to the membrane-proximal scavenger receptor cysteine-

545   rich domain of CD6 with a 1:1 stoichiometry. J Biol Chem 271:17390-17396. PMID:

546   8663238

547   35. Studier FW (2005) Protein production by auto-induction in high density shaking cultures.

548   Protein Expr Purif 41:207-234. PMID: 15915565

549   36. Kabsch W (2010) Xds. Acta Crystallogr D Biol Crystallogr 66:125-132. PMID:

550   20124692

551   37. Evans PR (2011) An introduction to data reduction: space-group determination, scaling

552   and intensity statistics. Acta Crystallogr D Biol Crystallogr 67:282-292. PMID: 21460446

553   38. Hohenester E, Sasaki T, Timpl R (1999) Crystal structure of a scavenger receptor

554   cysteine-rich domain sheds light on an ancient superfamily. Nat Struct Biol 6:228-232.

555   PMID: 10074941

556   39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment

557   search tool. J Mol Biol 215:403-410. PMID: 2231712

558   40. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high

559   throughput. Nucleic Acids Res 32:1792-1797. PMID: 15034147

560   41. Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in

561   molecular replacement. Journal of Applied Crystallography 41:641-643.

562   42. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007)

563   Phaser crystallographic software. J Appl Crystallogr 40:658-674. PMID: 19461840

564   43. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot.

565   Acta Crystallographica Section D 66:486-501.

566   44. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn

567   MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal

568   structures. Acta Crystallogr D Biol Crystallogr 67:355-367. PMID: 21460454

569   45. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray

570   LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for

571   macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66:12-21. PMID:

572   20057044

573   46. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE

574   (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J

575   Comput Chem 25:1605-1612. PMID: 15264254

576   47. Brennich ME, Round AR, Hutin S (2017) Online Size-exclusion and Ion-exchange

577   Chromatography on a SAXS Beamline. J Vis Exp. PMID: 28117806

578   48. Panjkovich A, Svergun DI (2017) CHROMIXS: automatic and interactive analysis of

579   chromatography-coupled small angle X-ray scattering data. Bioinformatics. PMID: 29300836

580   49. Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT,

581   Kikhney AG, Hajizadeh NR, Franklin JM, Jeffries CM, Svergun DI (2017) ATSAS 2.8: a

582   comprehensive data analysis suite for small-angle scattering from macromolecular solutions.

583   J Appl Crystallogr 50:1212-1225. PMID: 28808438

584   50. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality.

585   Science 336:1030-1033. PMID: 22628654

586   51. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new

587   ENDscript server. Nucleic Acids Res 42:W320-324. PMID: 24753421