

The three-dimensional structure of the first domain of neural cell adhesion molecule

The structure of the N-terminal Ig-superfamily (IgSF) domain of NCAM has a high similarity with the intermediate set of Ig-like domains: the first domain of the vascular cell adhesion molecule (VCAM-1), the M¹ domain in titin, and telokin.

The three-dimensional structure of the first domain of neural cell adhesion molecule

The structure of the N-terminal Ig-superfamily (IgSF) domain of NCAM has a high similarity with the intermediate set of Ig-like domains: the first domain of the vascular cell adhesion molecule (VCAM-1), the M5 domain in titin, and telokin.

Sir—The neural cell adhesion molecule (NCAM) is a cell surface glycoprotein^{1–3}, which plays a key role in development of the nervous system and in learning^{4,5}. NCAM contains five immunoglobulin (Ig) and two fibronectin type III (FIII) homology domains. It is known to mediate cell-cell interactions by homophilic and heterophilic binding mechanisms^{6–10}. NCAM transduces extracellular signals probably by interaction with the fibroblast growth factor receptor, leading to tyrosine phosphorylation and an increase in intracellular calcium concentration¹¹. It has been shown that the first domain of NCAM promotes attachment of neurons, stimulates migration of neural cells and affects intracellular calcium concentration¹². No binding interactions of the first Ig-like domain have been reported. However, surface plasmon resonance studies show that the first Ig-like NCAM domain binds in a homophilic interaction to the second Ig-like NCAM domain (V.V.K. *et al.* unpublished results).

Domain 1 structure

The three-dimensional structure of NCAM domain 1 (99 residues) was determined by ¹H, ¹⁵N and ¹³C NMR spectroscopy (Table 1). The structure has the classical Ig-SF fold with two β -sheets forming a β -sandwich (Fig. 1a,b). It consists of nine β -strands sequentially labelled A,A',B,C,C',D,E,F,G, that are arranged with strands A,B,D,E in one sheet and strands A',C,C',E,G in the other sheet. The strands are antiparallel, except the A' strand and C-terminal part of the G strand, which run in a parallel fashion. A cysteine bridge between Cys 24 in strand B and Cys 79 in strand F connects the two sheets. A type II

β -turn from Ser 15 to Glu 18 connects the A' strand and the B strand. A strong hydrogen bond between HN of Ala 74 to CO of Asn 70 between strands E and F indicates that this is one turn of an α -helix. The carboxylate of Asp 73 in the α -helix forms a salt bridge with the amino group of the otherwise buried Lys 20. Another salt bridge connects Lys 33 in the B–C loop to Asp 60 in the D–E loop. The A and A' strands are connected by a *cis*-proline, Pro 9, which makes it possible for the A strand to make an antiparallel β -sheet with the B strand and then bend to make a parallel β -sheet with the C-terminal G strand. The hydrophobic core consists, in particular, of the side chains of Ile 7, Lys 20, Phe 22, Trp 37, Leu 64 and Tyr 77.

Related structures

In a search for the best structural alignment with other members of known IgSF structures, the N-terminal domain of VCAM-1¹³, the M5 domain of titin¹⁴, and telokin¹⁵ clearly superimposed with very small r.m.s. deviation of the C α atoms (Fig. 1c). These three domains are all members of the recently identified I-set of the IgSF¹⁶. The structure of NCAM domain 1 is therefore a new member of the I-set.

The pairwise sequence homology between these four IgSF I-set domains is below 30%. Conserved in all four molecules are only five residues corresponding in NCAM domain 1 to Gly 17 in the tight turn between strands A' and B, Cys 24 and Gly 28 in the B strand, Trp 37 in the C strand and Tyr 77 in the F strand. In spite of the low sequence homology between the four members of the I-set a pairwise structure comparison¹⁷ shows that approxi-

mately 80 % of the C α atoms occupy equivalent positions in the aligned structures. For 58 C α atoms, mainly found in the A',B,C,D,E and F strands, positions in all four structures were equivalent and formed a core that is common for the I-set fold (Fig. 2). The five residue turn of α -helix (70–74 in NCAM domain 1) is present in all four structures between strands E and F. Two of the I-set core residues are located around the α -helix; these are Ala 69 and Asp 73 in NCAM domain 1. In NCAM domain 1 the α -helix is a regular N-cap box motif¹⁸. The A and C' strands vary more in spatial positions and in the titin M5 domain the C' strand could not be detected as a regular β -strand. Consequently the A and C' strands were not included in the I-set core.

Although the four structures are closely similar, there are distinct differences that divide the four molecules into two groups. The two cell adhesion proteins in the I-set, NCAM domain 1 and VCAM-1 domain 1, differ from the two muscle proteins in at least three places. The canonical cysteine bridge of the IgSF is missing in titin and telokin. In titin the cysteine corresponding to Cys 79 in NCAM domain 1 is absent, and in telokin the cysteine bridge is not formed, although the two cysteine residues are both present¹⁵. The *cis*-prolines, Pro 9 and Pro 7 in NCAM domain 1 and VCAM-1 domain 1 respectively, make a sharp bend between strand A in the first β -sheet and strand A' in the other β -sheet. This *cis*-proline is absent in the two muscle proteins titin and telokin, where a smooth connection is made between the two respective β -sheets. However, in the B–C loop, titin and telokin have a *cis*-proline

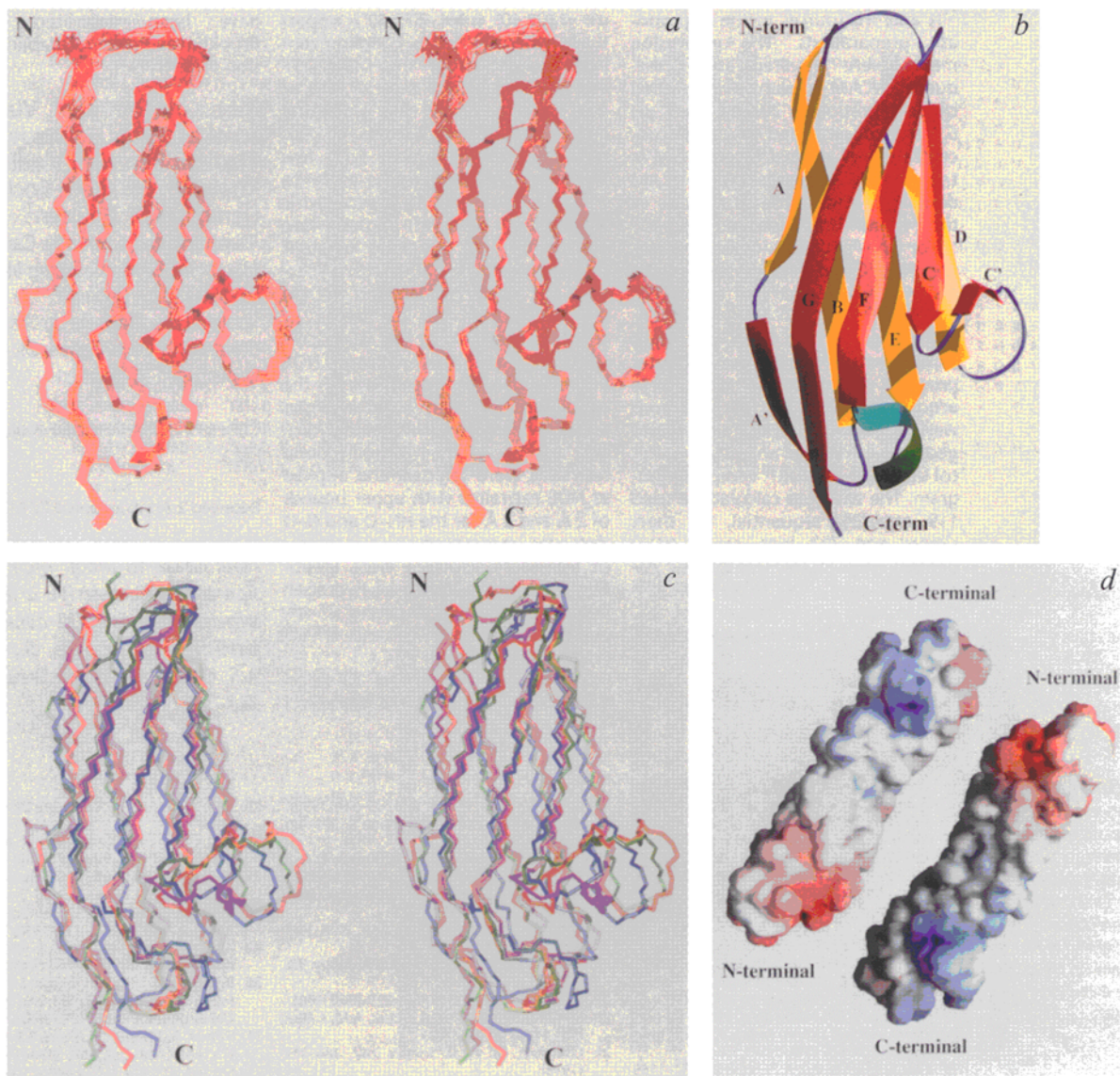


Fig. 1 Three-dimensional structure of NCAM domain 1. *a*, Superposition of C α , N and C from 20 structures representing NCAM domain 1. *b*, View of the backbone N, C α and C atoms and β -strands of NCAM domain 1 residues 1–99 facing the A'CC'FG sheet and C'-D loop. The ABDE (yellow) and A'CC'FG (red) β -sheets consist of residues 3–8, 19–28, 52–57, 61–67 and 11–15, 34–39, 43–45, 75–83, 87–98, respectively. The α -helix (green) consists of residue 70–74. (*b*) and (*c*) were produced using the computer program MOLMOL. *c*, Alignment of the four Ig I-set structures. View of the N, C α and C atoms of VCAM-1 domain 1 residues 1–90 (purple), telokin residues 40–134 (green) and the M5 domain of titin residues 1–91 (blue) facing the A'CC'FG sheet and C'-D loop, superimposed on NCAM domain 1 residues 1–99 (red) by a rigid body transformation. Pairs of C α atoms with a distance after the transformation less than 3 Å were used in the target function of the transformation. The Protein Brookhaven Databank entries of VCAM-1, telokin and the M5 domain of titin are 1vca, 1tlk, and 1tnm, respectively. *d*, Model of the antiparallel homophilic binding of the two N-terminal Ig domains of NCAM. Surface potential maps were generated with GRASP³³

at equivalent positions, Pro 27 and Pro 69, respectively, that is followed by an intervening residue and then a *trans*-proline. This *cis*-proline-X-*trans*-proline motif makes a sharp turn possible in this loop.

A new sequence profile based only on the four known I-set structures was subsequently constructed according to previously described principles^{19,20} and used

in a pattern search of the SWISS-PROT protein sequence data bank for potential members of the I-set. The results of this pattern search are in excellent agreement with the study by Harpaz and Chothia¹⁶, who based their sequence profile only on the core of one I-set domain, telokin, and selected V domains. In the new selection a majority of the matching IgSF

domains belongs to cell adhesion molecules of the nervous system (Fig. 2) and a number of these—for example NCAM, neural cell adhesion molecule L1, axonin-1 and contactin—have been shown to be involved in axonal growth and fasciculation. Apparently the new I-set sequence profile identifies preferentially IgSF I-set domains of neural origin.

Apart from the large number of IgSF domains of neural origin, a few domains of muscular and other origins were found by the profile. So far no common functional properties have been assigned to the I-set domains and it remains to be seen whether the common features of the members of the IgSF I-set comprise a structural platform for a common function, or whether they underlie a diversity of molecular functions. The α -helix between strands E and F is a conserved motif of the V-set and I-set domains, and are present

in some of the C2-set structures. In E-cadherin, a distant member of the IgSF, a similar turn of α -helix is present between strands E and F²¹. Most of the domains have the proline-X-proline motif in the B-C loop and are therefore expected to be similar to telokin and titin in this region. The C'-D loop, the D strand and the A strand are the regions where major differences in the fold are expected.

Of the four I-set Ig-domain structures known so far only VCAM-1 domain 1 has a potential binding site—the integrin binding consensus sequence (Q38IDSPL) located in the region between the C and D strands¹³. For the M5 domain of titin it has been proposed that the binding site for myomesin and M-protein also are in this region¹⁴. In a search for potential homophilic binding sites in NCAM a model was made of NCAM domain 2 with the computer program Modeller²² using the I-set as a template (NCAM domain 2 contains most of the I-set sequence profile characteristics shown in Fig. 2). The two N-terminal domains of NCAM were then assembled using VCAM-1 domain 1 and 2 as a template. The surface charge map of the model suggests an antiparallel homophilic binding between the acidic cluster of residues at the N-terminal of domain 1: Asp 29, 32, 34 at loop BC, Asp 58, 59 at loop DE, Asp 84, Glu 85 at loop FG, and a basic cluster of residues of domain 2: Lys 135, 137 at the C strand, Arg 139 at the C' strand, Arg 171 at the F strand, Lys 185 at the G strand (Fig. 1*d*). This is in agreement with the finding that NCAM domain 1 and 2 bind to each other in plasmon surface resonance analysis experiments (V.V.K. *et al.*, unpublished data).

The first structure determination of a neural IgSF domain shows that NCAM domain 1 is related to a number of neural IgSF domains all predicted to be members of the I-set. A precise description at the molecular level of cell adhesion in neuron-neuron and neuron-glia interactions is a prerequisite to an understanding of neural cell differentiation and the formation of complex neural networks.

Table 1 Structural statistics of the 20 structures of NCAM domain 1

Distance restraints (All) ¹	1445
intraresidue	119
sequential ($ i-j =1$)	353
medium range ($1< i-j \leq 5$)	161
long range ($ i-j >5$)	722
hydrogen bonds	90
Dihedral angle restraints (All)	91
ϕ	53
χ^1	38
Deviations from experimental derived restraints	
Distance restraints (Å)	
0.1–0.2	24.45
0.2–0.3	3.85
0.3–0.4	0.2
R.m.s. deviation	0.027 (± 0.002)
Hydrogen bonds >0.4	0
Dihedral angle restraints	
>5°	0
R.m.s. deviation	0.49 (± 0.108)
Deviations from idealized geometry	
Impropers (°)	
>5	0.35
R.m.s. deviation	0.59 (± 0.045)
Bonds (Å)	
>0.05	0
R.m.s. deviation	0.0075 (± 0.0002)
Angles (°)	
>5	10.9
R.m.s. deviation	0.82 (± 0.016)
Energies (kcal mol ⁻¹) ²	
NOE	51.35 (± 8.40)
Dihedral angle restraint	1.54 (± 0.63)
Bond	84.79 (± 4.15)
Angle	276.68 (± 10.78)
Improper	41.00 (± 6.34)
Repel	0.30 (± 0.22)
van der Waals	-460.98 (± 4.26)
Hydrogen bond restraint	94.85 (± 4.19)
Hydrogen bond	-39.62 (± 2.95)
R.m.s. deviations of atomic positions (Å) ³	
Backbone (C α , N, C)	0.33 (± 0.20)
All heavy atoms	0.59 (± 0.38)

The number of experimentally derived restraints used in the structure calculations and average values per structure of the deviations between the 20 accepted structures and the experimental restraints or idealized geometry are shown.

¹Number of non-redundant distance restraints.

²The energies were calculated using final force constants of: $k_{\text{NOE}} = 50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, $k_{\text{cdih}} = 200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$, $k_{\text{bond}} = 1000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, $k_{\text{angle}} = 500 \text{ kcal mol}^{-1} \text{ rad}^{-2}$, $k_{\text{prp}} = 500 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, $k_{\text{repel}} = 4.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ with the van der Waals hard sphere radius set to 0.825 of the values calculated from the X-PLOR parallhdg.pro parameter file. The van der Waals energy and hydrogen bond energies were calculated using the X-PLOR switched Lennard-Jones van der Waals energy function and hydrogen bond function using the parmllh3x.pro parameter file, but they were not included in the target functions used in the simulated annealing procedure.

³R.m.s. deviations from an average structure of the 20 structures of NCAM domain 1.

^{15}N and ^{13}C sources in the appropriate preparations. The expression media were desalted and subsequently NCAM domain 1 was purified by gel filtration in 20 mM NaCl, pH 6.0 and concentrated to a final concentration of approximately 2 mM in the sample of unlabelled NCAM domain 1, and 1 mM in samples of the ^{15}N and $^{15}\text{N}/^{13}\text{C}$ labelled protein.

For the structure determination, the ^1H , ^{13}C and ^{15}N resonances were assigned from spectra of DQF-COSY, TOCSY, ^{15}N TOCSY-HMQC, HNCA, HN(CO)CA and HCCH-TOCSY NMR experiments^{23–28} using the computer program Pronto²⁹. All data were acquired at 298 K. 100 structures were generated with a distance geometry/simulated annealing protocol using the X-PLOR³⁰ computer program. The structure calculations used 119 intra, 353 sequential, 161 short range, and 722 long range NOEs derived from 80 ms and 200 ms 2D-NOESY³¹ and ^{15}N NOESY-HMQC³² spectra with upper bounds of 2.7, 3.3, 5.0 and 7.0 Å increased by 0.5 Å if the NOE restraint included a methyl group. 53 ϕ dihedral angle restraints were applied with bounds

of $-120\pm 40^\circ$ and $-57\pm 40^\circ$, respectively, when the $^3J_{\text{HNH}\alpha}$ coupling constant derived from the DQF-COSY and NOESY spectra²⁹ were >7.5 Hz or <5.5 Hz, respectively. Nine ϕ dihedral angle restraints with bounds $-90\pm 90^\circ$ were applied in residues with few NMR restraints showing no evidence of positive ϕ angles. 38 stereospecific assigned χ^1 dihedral angles were derived by $^3J_{\text{NH}\alpha}$ coupling constant measurements in 2D-NOESY spectrum of the ^{15}N labelled NCAM domain 1, estimates of the $^3J_{\text{H}\alpha\text{H}\beta}$ coupling constants in the DQF-COSY spectrum, and the relevant NOE intensities. After inspection of the calculated hydrogen bond energies and/or evidence of slow amide hydrogen exchange, 45 hydrogen bond restraints were selected and applied as NOE restraints with upper bounds of 2 Å and 3 Å for the HN–O and N–O distances, respectively. The elements of secondary structure were identified from the evidence of hydrogen bonds and the typical strong $\text{H}^\alpha\text{--H}^\alpha$, and the weaker $\text{H}^{\text{N}}\text{--H}^\alpha$ and $\text{H}^{\text{N}}\text{--HN}$ cross-strand NOEs.

The coordinates of the 20 structures representing NCAM domain 1

have been deposited in the Brookhaven Protein Databank (accession code 1NCM).

Niels Kirk Thomsen¹, Vladislav Soroka², Peter Holme Jensen¹, Vladimir Berezin², Vladislav V. Kiselyov², Elisabeth Bock² and Flemming M. Poulsen¹

¹Department of Chemistry, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby, Denmark

²Protein Laboratory, Panum Institute, University of Copenhagen, DK-2200 Copenhagen N, Denmark

Correspondence should be addressed to F.M.P.

Received 2 April; accepted 22 May 1996.

Note added in proof: After submission of this manuscript, a new I-set structure, the titin I27 domain, was published by Improte, S., Politou, A.S. & Pastore, A. in *Structure* 4, 323–337 (1996).

Acknowledgements

This research project was supported by the Danish Biotechnological Research and Development Programme. We thank Pia Skovgaard for skilled technical assistance.

- Barthels, D. *et al.* *EMBO J.* **6**, 907–914 (1987).
- Cunningham, B.A. *et al.* *Science* **236**, 799–806 (1987).
- Harrelson A.L. & Goodman C.S. *Science* **242**, 700–708 (1988).
- Cremer, H. *et al.* *Nature* **367**, 455–459 (1994).
- Rønn, L.C.B., Bock, E., Linnemann, D. & Jahnsen, H. *Brain Res.* **677**, 145–151 (1995).
- Rao, Y., We, X.-F., Yip, P., Gariepy, J. & Siu, C.-H. *J. Biol. Chem.* **268**, 20630–20638 (1993).
- Grumet, M., Flaccus, A. & Margolis, R.U. *J. Cell. Biol.* **120**, 815–824 (1993).
- Nybroe, O., Moran, N. & Bock, E. *Neurochem.* **52**, 1947–1949 (1989).
- Probstmeier, R., Kühn, K. & Schachner, M. *J. Neurochem.* **53**, 1794–1801 (1989).
- Cole, G.J. & Akeson, R. *Neuron* **2**, 1157–1165 (1989).
- Williams, E.J., Furness, J., Walsh, F.S. & Doherty, P. *Neuron* **13**, 583–594 (1994).
- Frei, T., von Bohlen und Halbach, F., Wille W. & Schachner M. *J. Cell Biol.* **118**, 177–194 (1992).
- Jones, E. Y. *et al.* *Nature* **373** 539–544 (1995).
- Pfuhl, M. & Pastore A. *Structure* **3**, 391–401 (1995).
- Holden, H.M., Ito, M., Hartshorne, D.J. & Rayment, I. *J. Mol. Biol.* **227**, 840–851 (1992).
- Harpaz, Y. & Chothia, C. *J. Mol. Biol.* **238**, 528–539 (1994).
- Lesk, A.M. & Chothia, C. *J. Mol. Biol.* **160**, 325–342 (1982).
- Haper, E.T. & Rose, G.D. *Biochemistry* **32**, 7605–7610 (1993).
- Taylor, R.W. *J. Mol. Biol.* **188**, 233–258 (1986).
- Bashford, D., Chothia, C. & Lesk, A.M. *J. Mol. Biol.* **196**, 199–216 (1987).
- Overduin, M. *et al.* *Science* **267**, 386–389 (1995).
- Šali, A. & Blundell, T.L. *J. Mol. Biol.* **234**, 779–815 (1993).
- Rance *et al.* *Biochem. Biophys. Res. Commun.* **2**, 479–485 (1983).
- Bax, A. & Davis, D.G. *J. Magn. Reson.* **65**, 355–360 (1985).
- Marion, D. *et al.* *Biochemistry* **28**, 6150–6156 (1989).
- Kay, L.E., Ikura, M., Tschudin, R. & Bax, A. *J. Magn. Reson.* **89**, 496–514 (1990).
- Bax, A. & Ikura, M. *J. Biomolec. NMR* **1**, 99–104 (1991).
- Bax, A., Clore, M. & Gronenborn, A.M. *J. Magn. Reson.* **88**, 425–431 (1990).
- Kjær, M., Andersen, K.V. & Poulsen, F.M. *Meth. Enzym.* **239**, 288–307 (1994).
- Brünger A.T. X-PLOR Version 3.1. A system for X-ray Crystallography and NMR. Yale University, New Haven CT. (1992).
- Kumar, A., Wagner, G., Ernst, R.R. & Wütrich, K. *J. Am. Chem. Soc.* **103**, 3654–3658 (1981).
- Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. & Bax, A. *J. Am. Chem. Soc.* **111**, 1515–1517 (1989).
- Nicholls, A. GRASP: Graphical representation and analysis of surface properties (Columbia University, NY, 1993).