J Biomol NMR (2007) 39:331–336
DOI 10.1007/s10858-007-9201-7

NMR STRUCTURE NOTE

The high resolution NMR structure of the third SH3 domain of CD2AP

Jose L. Ortega Roldan · M. Luisa Romero Romero · Ari Ora · Eiso AB · Obdulio Lopez Mayorga · Ana I. Azuaga · Nico A. J. van Nuland

Received: 14 August 2007 / Accepted: 14 September 2007 / Published online: 9 October 2007

Abstract CD2 associated protein (CD2AP) is an adaptor protein that plays an important role in cell to cell union needed for the kidney function. CD2AP interacts, as an adaptor protein, with different natural targets, such as CD2, nefrin, c-Cbl and podocin. These proteins are believed to interact to one of the three SH3 domains that are positioned in the N-terminal region of CD2AP. To understand the network of interactions between the natural targets and the three SH3 domains (SH3-A, B and C), we have started to determine the structures of the individual SH3 domains. Here we present the high-resolution structure of the SH3-C domain derived from NMR data. Full backbone and side-chain assignments were obtained from triple-resonance spectra. The structure was determined from distance restraints derived from high-resolution 600 and 800 MHz NOESY spectra, together with *phi* and *psi* torsion angle restraints based on the analysis of ^{1}HN, ^{15}N, ^{1}Hα, ^{13}Cα, ^{13}CO and ^{13}Cβ chemical shifts. Structures were calculated using CYANA and refined in water using RECOORD. The three-dimensional structure of CD2AP SH3-C contains all the features that are typically found in other SH3 domains, including the general binding site for the recognition of polyproline sequences.

Keywords Adaptor protein · CD2AP · NMR · Protein structure · SH3 domain

Abbreviations

SH3	Src-homology domain 3
SH3-C	The third SH3 domain of CD2AP
CD2AP	CD2 associated protein
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser enhancement spectroscopy
TOCSY	Total correlation spectroscopy
RMSD	Root mean square deviation

Biological context

The CD2 associated protein (CD2AP) is a cytosolic protein expressed in podocytes, the cells of kidney glomerulus, which form the barrier between blood and urine and therefore plays a crucial role in renal ultra-filtration (Dustin et al. 1998). CD2AP acts as an adaptor protein, which facilitates interaction of its targets and is involved in cytosolic transport along microtubules. Among others, CD2, nephrine, podocine and c-Cbl have so far been identified as natural targets. Its possible important role in renal functionality is underlined by its involvement in several renal diseases like the hereditary nephritic syndrome (Wolf and Stahl 2003).

The domain structure of CD2 consists of three N-terminal SH3 (Src-Homology) domains followed by a poly-proline region, and a globular domain linked to a C-terminal coiled-coil region. Intermolecular interactions

J. L. Ortega Roldan · M. L. Romero Romero · O. Lopez Mayorga · A. I. Azuaga · N. A. J. van Nuland (✉)
Departamento de Química Física e Instituto de Biotecnología, Facultad de Ciencias, Universidad de Granada, Campus Fuentenueva s/n, Granada 18071, Spain
e-mail: najvan@ugr.es

A. Ora
Department of Biosciences/Institute of Biotechnology, University of Helsinki, Biocenter 3, Viikinkaari 1, Helsinki 00014, Finland

E. AB
Gorlaeus Laboratory, Leiden University, Einsteinweg 55, Leiden 2333 CC, The Netherlands

known so far are mostly mediated via the three N-terminal SH3 domains, named A, B and C, and also via the poly-proline linker region, which is possibly also involved in intramolecular interactions.

SH3 domains are a very good example of the modular structure of nature: they are found in various proteins where they exert their typical function: molecular recognition and subsequent binding (Mayer 2001). Although they form a highly conserved family of domains, their amino acid composition varies at a few key sites, allowing for a wide range of molecular targets. To understand the network of interactions between the natural targets of CD2AP and the three SH3 domains (SH3-A, B and C), we have started to determine the structures of the individual SH3 domains. The structures of the first and second SH3 domains of the CD2AP human analogue, CMS, have recently been solved by X-ray diffraction (Moncalián et al. 2006) and NMR (Yao et al. 2007), respectively. Here we present the high resolution structure of the third SH3 domain of CD2AP.

Methods and results

Protein expression and purification

The plasmid pETM-11 (Protein Expression and Purification Core Facility, EMBL Heidelberg, Germany) containing the SH3-C gene covalently linked to a N-terminal 6xHis tag and a TEV protease cleavage site was expressed in *Escherichia coli* Rosetta(DE3) strain (Novagen). Cells were grown on LB medium four times the volume of the M9 minimal medium at 37°C till an OD_{600} of 0.7. After two washing steps cells were resuspended for 1 h in M9 medium at 20°C. Protein expression was induced with 1 mM IPTG at 20°C during 15–20 h. Cells were resuspended after centrifugation at 4000 RPM for 20 min in 50 mM phosphate, 0.3 M NaCl, pH 8.0 buffer (Column Buffer: CB) and broken in a French press. After centrifugation at 30,000 RPM during 30 min, the supernatant was loaded on 5 ml of Ni-sepharose resin (Pharmacia) equilibrated with CB. The protein was eluted with a gradient of CB from 0 to 500 mM imidazole. Fractions containing SH3-C were dialyzed against TEV buffer (50 mM Tris–HCl pH 8.0, 0.5 mM EDTA) to eliminate imidazole. The His-tag was cleaved by incubation at room temperature overnight with TEV protease in presence of 1 mM ditiothreitol. The cleaved SH3-C was recovered by a final chromatography step on a Superdex G75 gel-filtration column. ^{15}N- and ^{13}C/^{15}N-labelled proteins were purified from cells grown in M9 minimal medium containing either $^{15}NH_4Cl$ (Cortecnet) or $^{15}NH_4Cl$ and $(^{13}C_6)_D$-glucose (Spectra Stable Isotopes and Cortecnet) as the sole sources of nitrogen and carbon, respectively. Sample concentration was determined spectrophotometrically using an extinction coefficient of 13,980 M^{-1} cm^{-1}, determined using the Gill and von Hippel method (Gill and von Hippel 1989).

NMR Spectroscopy

^{15}N-labelled and ^{15}N/^{13}C-labelled samples of CD2AP SH3-C were prepared for NMR experiments at 1 mM in 93% H_2O/7% D_2O, 50 mM NaPi, 100 mM NaCl, 1 mM DTT at pH 2.0. All NMR spectra were recorded at 25°C on a Varian NMR Direct-Drive Systems 600 MHz spectrometer (^{1}H frequency of 600.25 MHz) equipped with a triple-resonance PFG-XYZ probe. Two-dimensional nuclear Overhauser enhancement spectroscopy (2D NOESY, 125 ms mixing time) spectra were acquired on an unlabelled sample in D_2O. 2D heteronuclear single quantum correlation (^{15}N-HSQC), ^{15}N-TOCSY-HSQC (mixing time of 80 ms) and ^{15}N-NOESY-HSQC (mixing time of 125 ms) were acquired on a ^{15}N-labelled sample both at 600 MHz as well as on a Varian Inova 800 MHz spectrometer equipped with a triple-resonance PFG-Z cold probe. Triple resonance spectra CBCA(CO)NH, HNCA CB, HNCO, HBHA(CO)NH, aromatic (HB)CB(CGCD) HD/(HB)CB(CGCDCE)HE and HCCH-TOCSY were recorded at 600 MHz on a ^{13}C/^{15}N-labelled sample.

All NMR data were processed using NMRPipe (Delaglio et al. 1995) and analyzed by NMRView (Johnson and Blevins 1994).

Assignment of the SH3-C domain of CD2AP

A modified version of the SmartNotebook 3.2 tool integrated in NMRView was used to semi-automatically assign the protein backbone. The SmartNotebook module allows visual inspection of the backbone sequential connectivities and provides tools to assign segments of residues to the primary sequence based on characteristic carbon chemical shifts. Peaklists containing $^{13}C\alpha$ and $^{13}C\beta$ frequencies taken from the HNCACB and CBCA(CO)NH at the ^{1}H,^{15}N frequency of every peak in the HSQC spectrum were used as input for SmartNotebook, which then creates automatically connections between the ^{1}H,^{15}N pairs. ^{13}CO frequencies were taken from the HNCO at the assigned ^{1}H,^{15}N frequencies. $^1H\alpha$ and $^1H\beta$ were assigned using the HBHA(CO)NH. The obtained $^1H\alpha$-$^{13}C\alpha$ and $^1H\beta$-$^{13}C\beta$ frequencies were then used for the assignment of the ^{1}H and ^{13}C frequencies of aliphatic side chains from the HCCH-TOCSY spectrum. ^{1}H resonances were checked by the combined use of the ^{15}N-edited 3D-TOCSY and 3D-NOESY spectra. Aromatic $^1H\delta$ and $^1H\varepsilon$ frequencies

were obtained from the combined use of the (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE spectra, where the aromatic ^{1}H frequencies are correlated with the assigned $^{13}C\beta$ frequencies of the aromatic residue. Missing aromatic ^{1}H frequencies (e.g. those of Trp side chains) were assigned from a ^{1}H,^{1}H 2D-NOESY in D_2O.

Figure 1 shows an assigned ^{1}N-^{15}N HSQC spectrum of SH3-C that resulted from the assignment procedure ascribed above. The ^{1}H, ^{15}N and ^{13}C assigned resonances have been deposited in the BioMagResBank database with accession code 15407.

The structure of the SH3-C domain of CD2AP

NOE cross peaks were obtained by manual peak picking in each strip of the 800 MHz 3D ^{15}N-NOESY-HSQC spectra taken at the assigned ^{1}H,^{15}N frequencies and by automatic peak picking of the 600 MHz 2D NOESY in D_2O followed by removal of diagonal peaks and peaks arising from artifacts (e.g. residual water). NOEs were assigned using

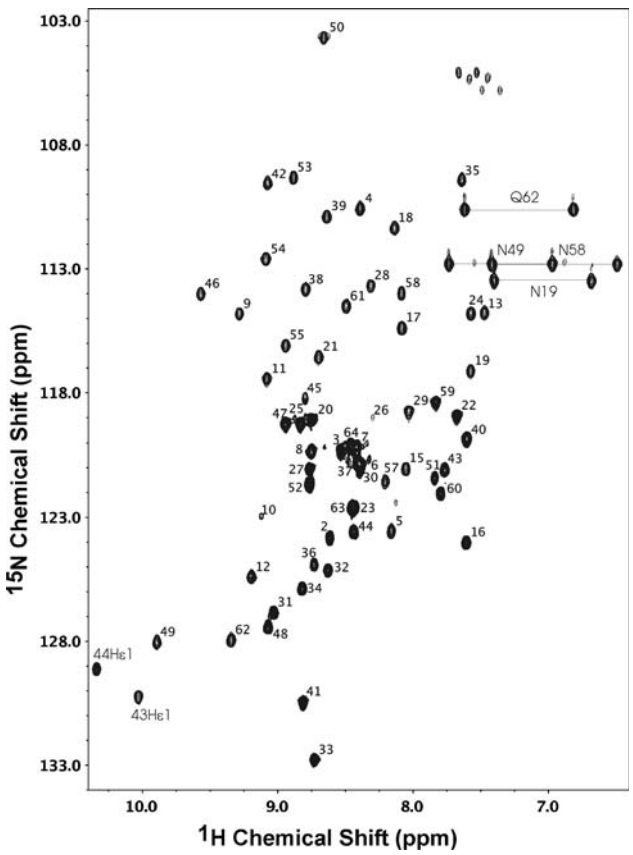

Fig. 1 *Assigned 600 MHz ^{1}H-^{15}N HSQC of CD2AP SH3-C at pH 2.0, 25°C.* All backbone ^{1}H-^{15}N pairs are indicated by their corresponding number in the amino acid sequence. The Hε1 protons of Trp43 and Trp44 as well as the side-chains of the single glutamine and three asparagines are indicated

the automated NOE assignment procedure of CYANA version 2.1 (Güntert et al. 1997; Hermann et al. 2002).

Phi and *psi* torsion angle restraints were included based on analysis of ^{1}HN, ^{15}N, $^{1}H\alpha$, $^{13}C\alpha$, ^{13}CO and $^{13}C\beta$ chemical shifts using the program PREDITOR (Berjanskii et al. 2006). Only *phi* and *psi* angle restraints were used with a confidence factor higher than 0.5. The standard protocol was used with seven cycles of combined automated NOE assignment and structure calculation of 100 conformers in each cycle. This protocol was repeated ten times using different seeding numbers. Restraints that were unambiguously assigned in at least 5 out of 10 CYANA runs were used for a final structure calculation run using CNS (Brunger et al. 1998) in explicit solvent using the RECOORD protocol (Nederveen et al. 2005). The twenty lowest-energy structures were used for the final analysis. Input data and structure calculation statistics are summarized in Table 1. PROCHECK-NMR (Laskowski et al. 1996) and WHATIF (Vriend 1990) were used to analyze the quality of the structures. MOLMOL (Koradi et al. 1996) and PYMOL (http://www.pymol.org/) were used for visualization. The structural coordinates and experimentally derived restraints have been deposited in the Protein Data Bank with accession number 2JTE.

Figure 2a shows a stereo representation of the superposition of the ensemble of 20 lowest-energy water-refined structures derived from the automatic NOE assignment protocol in CYANA and the subsequent refinement using RECOORD. The ensemble fulfils the experimental data very well (Table 1). Moreover, the ensemble of structures shows excellent Ramachandran statistics reflecting the high quality of the 3D structures. The success of the approach can be attributed to the use of high-quality 800 MHz 3D NOESY and 600 MHz 2D NOESY spectra and the almost complete set of assignments of almost all proton frequencies (>95%) used for the automatic NOE assignment stage and the addition of torsion angle restraints derived form the combination of selected ^{1}H, ^{15}N and ^{13}C chemical shifts. The structural disorder observed for the N-terminus is consistent with the higher flexibility in this region of the protein as indicated by the smaller heteronuclear NOE values (Fig. 3a).

Backbone dynamics from ^{15}N relaxation data

^{15}N NMR relaxation data are widely used to probe both molecular rotational diffusion and local backbone dynamics of a protein. We have measured the relaxation parameters ^{15}N R_1, R_2 and ^{1}H-^{15}N steady-state NOE at 600.25 MHz and 25°C of CD2AP SH3-C. Relaxation values were obtained from series of 2D experiments with coherence selection achieved by pulse field gradients at

Table 1 Structural statistics for the 20 conformers of CD2AP SH3-C[a]

Experimental restraints[b]	
Distance restraints	
Intraresidue (i-j = 0)	241
Sequential (|i-j| = 1)	297
Medium range (2≤|i-j|≤4)	109
Long range (|i-j|≥5)	394
Dihedral restraints	
Torsion angle (phi/psi)	2 × 24
Total number of restraints	1089
Restraints statistics	
NOE violations > 0.5 Å	0
Dihedral violations > 5°	0
CNS energies (Kcal/mol)	
E_{total}	–2567.7 ± 33.2
E_{vdw}	–344.0 ± 10.6
E_{elec}	–2646.4 ± 35.3
RMSD from average for residues 8–62 (Å)[c]	
Backbone N, CA, C′	0.37
Heavy atoms	0.91
Ramachandran plot[d]	
Most favored regions (%)	89.1 ± 2.0
Additional allowed regions (%)	10.5 ± 2.0
Generously allowed regions (%)	0.2 ± 0.6
Disallowed regions (%)	0.2 ± 0.6

[a] The statistics is obtained from an ensemble of 20 lowest-energy water-refined structures for SH3-C

[b] Restraint statistics reported for unique, unambiguous assigned NOEs

[c] Coordinate precision is given as the pair-wise Cartesian coordinate Root Mean Square Deviations from the average structure over the ensemble

[d] Values obtained from the PROCHECK-NMR analysis over all residues including the first seven highly flexible residues

600.25 MHz using the experiments described previously (Farrow et al. 1994) on ^{15}N-labelled SH3-C. The ^{1}H-^{15}N heteronuclear NOEs were determined from the ratio of peak intensities (*I*on/*I*off) with and without the saturation of the amide protons for 3 s. ^{15}N R_1 and ^{15}N R_2 relaxation rates were measured from spectra with different relaxation delays: 10, 100, 200, 300, 400, 500, 700, 900, 1200 and 1500 ms for R_1 and 10, 30, 50, 70, 90, 110, 150, 190 and 250 ms for R_2. Relaxation parameters and their corresponding errors were extracted with the program NMRView.

High NOE values (about 0.8) are observed for most of the SH3 domain chain, except for the 7 N-terminal residues and the C-terminal residue 64 (Fig. 3a). Analysis of the relaxation data of SH3-C using TENSOR2 (Dosset et al. 2000) indicate average ^{15}N R_2/R_1 ratios (Fig. 3b) for 32 residues in the most ordered regions corresponding to an apparent rotational correlation time τ_m of 3.79 ns within the expected range for a monomeric 7 kDa protein. We also used TENSOR2 to perform a model-free analysis of the local internal mobility affecting the backbone amides in the presence of an isotropic tensor. All measured relaxation data can be adequately described by a single parameter, the order parameter S^2, for most residues (Fig. 3c). Besides the seven N-terminal residues, the local internal mobility of residues 24 and 60 can only be described using an additional term for fast internal motions (τ_e) or for exchange contributions (R_{ex}), respectively.

Discussion and conclusions

The 3D structure of CD2AP SH3-C shows the five β-strands and the 3^{10} helix that are common in the family of SH3 domains (Fig. 2b). It also contains all the features that are typically found in other SH3 domains, including the general binding site for the recognition of polyproline

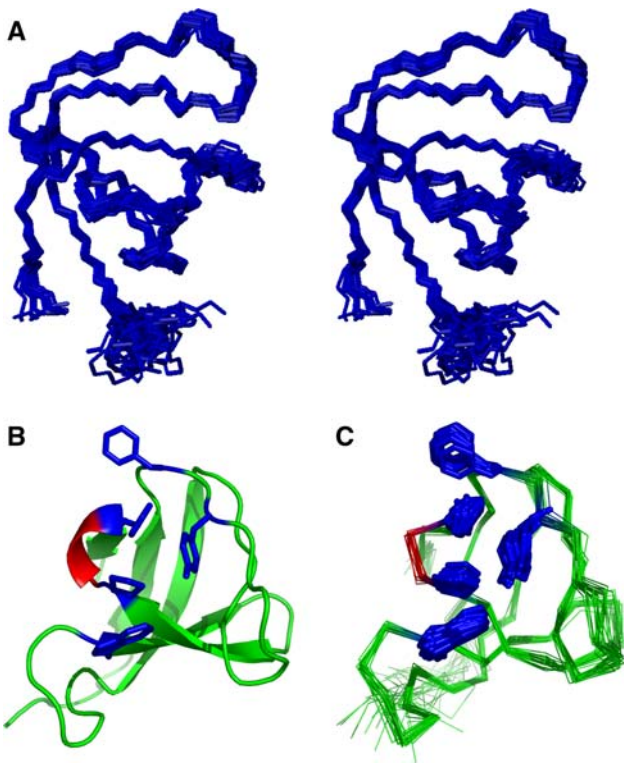

Fig. 2 *The high resolution NMR structure of CD2AP SH3-C.* (**a**). Stereo picture of the ensemble of 20 lowest energy structures superimposed on the backbone atoms of residues 8–62. (**b**). Cartoon representation of the lowest energy structure. The five β-strands and loops and turns are colored green. The single 3^{10}-helix is colored red, the residues that are involved in polyproline binding are presented in blue sticks. (**c**). Ribbon diagram of the ensemble of 20 structures. Color codes as in **b**

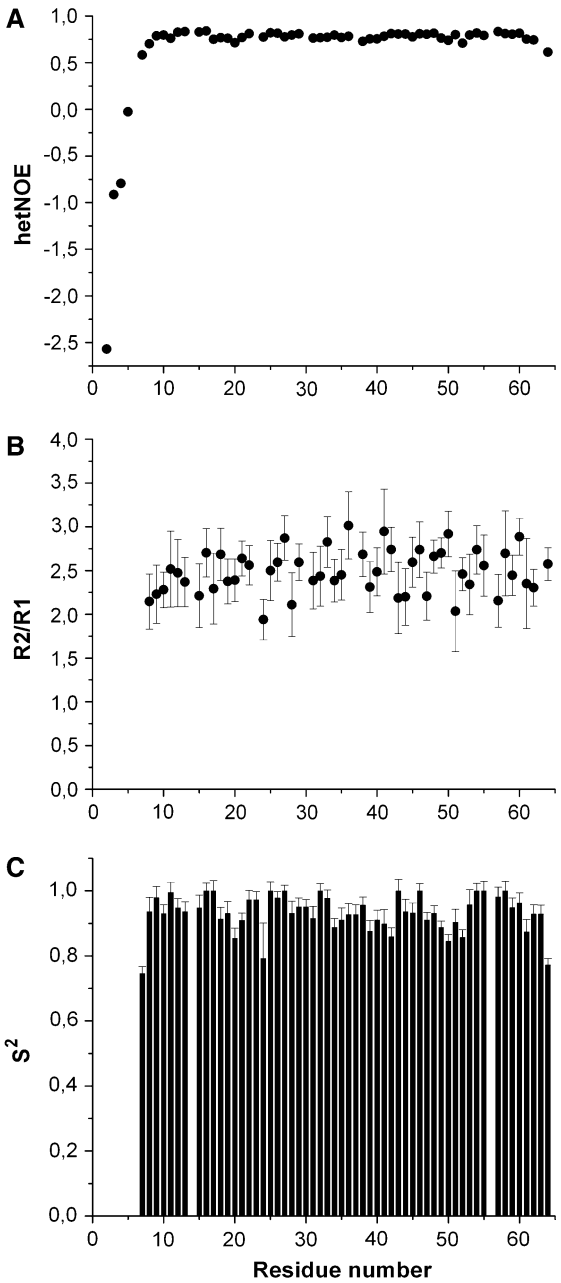

Fig. 3 *Backbone dynamics of CD2AP SH3-C measured at 600.25 MHz and 25°C.* (**a**). ^{1}H-^{15}N steady-state NOEs (hetNOE). (**b**). ^{15}N R_2 over R_1 ratios (R2/R1); analysis of relaxation data using TENSOR2 show average R2/R1 ratios from residues in secondary structure elements that correspond to an apparent rotational correlation time τ_m of 3.79 ns. (**c**). Order parameters (S^2) obtained from the TENSOR2 analysis using the model-free approach to describe the rotational diffusion of an isotropic tumbling SH3-C molecule with a τ_m of 3.79 ns

sequences that consists of a hydrophobic surface with three shallow pockets defined by the side chains of preserved aromatic residues that are indicated in Fig. 2C.

The high resolution structure of the third SH3 domain of CD2AP presented here completes the structure determination of the three SH3 domains that are present in CD2AP and in the human analogue CMS. As these three SH3 domains are believed to interact with different natural targets, it is interesting to compare our structure with the first and second SH3 domains of CMS. CD2AP SH3-C shares 41% sequence identity with both CMS SH3-A and SH3-B. A homology search through the sequence databank also shows a 41% sequence identity with the first SH3 domain of CIN85 (Jozic et al. 2005), a protein related to the human CMS. The CD2AP SH3-C structure described in our work is similar to the X-ray structure of the first SH3 domain of CIN85 (PDB entry 2BZ8), with a RMSD of 0.91 Å for all backbone atoms excluding the flexible seven N-terminal and two C-terminal residues and the residues in the n-Src loop, which is different in length for the two proteins, but deviates more from the NMR structure of the CMS SH3-B domain (PDB entry 2FEI, RMSD of 1.16 Å). CD2AP SH3-C, however, is closely similar to the X-ray structure of the CMS SH3-A domain (PDB entry 2J6F) with a RMSD of 0.78 Å. Moreover, the aromatic residues that are involved in recognition of polyproline sequences (shown in Fig. 2b and c) are very similar positioned in the two structures. As these two closely similar SH3 domains are part of the same adaptor protein, their binding specificity must come from small differences at specific sites in the two domains, enabling to recognize and bind to a specific partner. Future studies directed to resolve the structures of the complexes between these SH3 domains and their natural targets will hopefully lead to a detailed understanding on how these similar domains recognize different targets providing insight on how signaling is regulated by CD2AP. This may ultimately lead towards the design of specific drugs that might be used in treatments of renal diseases.

Acknowledgements This research is funded by grant BIO2005-04650 from the Spanish Ministry of Education and Science (MEC), the European Science Foundation EUROCORES Programme and contract ERAS-CT-2003-980409 from the European Union. We thank Adrien Favier for the use of SmartNotebook, Salvador Casares for useful discussions and Pedro L. Mateo for his continuous support. J.L.O.R. and M.L.R.R. are supported by a FPU and FPI pre-doctoral grant from the MEC, respectively. A.I.A. is a recipient of a Return research contract from the Junta de Andalucia and N.A.J.v.N. is a recipient of a Ramón y Cajal research contract from the MEC. The 600 MHz spectra were recorded in the Centre for Scientific Instrumentation (CIC) of the University of Granada. The 800 MHz spectra were recorded at the RALF Large Scale Facility in Grenoble, which is funded by the 'Access to Research Infrastructures' program of the European Union.

References

Berjanskii MV, Neal S, Wishart DS (2006) PREDITOR: a web server for predicting protein torsion angle restraints. Nucleic Acids Res 34:W63–W69

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J, Kuszewski J, Nilges M, Pannu NS, Read

RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D54:905–921

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) Nmrpipe - a multidimensional spectral processing system based on unix pipes. J Biomol NMR 6:277–293

Dosset P, Hus JC, Blackledge M, Marion D (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. J Biomol NMR 16:23–28

Dustin ML, Olszowy MW, Holdorf AD, Li J, Bromley S, Desai N, Widder P, Rosenberger F, van der Merwe PA, Allen PM, Shaw AS (1998) A novel adaptor protein orchestrates receptor patterning and cytoskeletal polarity in T-cell contacts. Cell 94:667–677

Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology-2 domain studied by N-15 NMR relaxation. Biochemistry 33:5984–6003

Gill SC, Von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182:319–326

Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298

Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209–227

Johnson BA, Blevins RA (1994) NMR view - a computer-program for the visualization and analysis of NMR data. J Biomol NMR 4:603–614

Jozic D, Cárdenes N, Deribe YL, Moncalián G, Hoeller D, Groemping Y, Dikic I, Rittinger K, Bravo J (2005) Cbl promotes clustering of endocytic adaptor proteins. Nat Struct Mol Biol 12:972–979

Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:51–55, 29–32

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486

Mayer BJ (2001) SH3 domains: complexity in moderation. J Cell Sci 114:1253–1263

Moncalián G, Cárdenes N, Deribe YL, Spínola-Amilibia M, Dikic I, Bravo J (2006) Atypical polyproline recognition by the CMS N-terminal Src homology 3 domain. J Biol Chem 281:38845–38853

Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CA, Nabuurs SB, Güntert P, Livny M, Markley JL, Nilges M, Ulrich EL, Kaptein R, Bonvin AM (2005). Proteins 59:662–672

Vriend G (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graph 8:52–56

Wolf G, Stahl RAK (2003) CD2-associated protein and glomerular disease. Lancet 362:1746–1748

Yao B, Zhang J, Dai H, Sun J, Jiao Y, Tang Y, Wu J, Shi Y (2007) Solution structure of the second SH3 domain of human CMS and a newly identified binding site at the C-terminus of c-Cbl. Biochim Biophys Acta 1774:35–43